Applied Statistics and Probability for Engineers

Third Edition

Douglas C. Montgomery
Arizona State University

George C. Runger
Arizona State University



John Wiley & Sons, Inc.

ACQUISITIONS EDITOR *Wayne Anderson*
ASSISTANT EDITOR *Jenny Welter*
MARKETING MANAGER *Katherine Hepburn*
SENIOR PRODUCTION EDITOR *Norine M. Pigliucci*
DESIGN DIRECTOR *Maddy Lesure*
ILLUSTRATION EDITOR *Gene Aiello*
PRODUCTION MANAGEMENT SERVICES *TechBooks*

This book was set in Times Roman by TechBooks and printed and bound by Donnelley/Willard. The cover was printed by Phoenix Color Corp.

This book is printed on acid-free paper.

Copyright 2003 © John Wiley & Sons, Inc. All rights reserved.

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 Sections 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, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM. To order books please call 1(800)-225-5945.

Library of Congress Cataloging-in-Publication Data

Montgomery, Douglas C.
Applied statistics and probability for engineers / Douglas C. Montgomery, George C. Runger.—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-20454-4 (acid-free paper)
1. Statistics. 2. Probabilities. I. Runger, George C. II. Title.

QA276.12.M645 2002
519.5—dc21

2002016765

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To:

Meredith, Neil, Colin, and Cheryl

Rebecca, Elisa, George, and Taylor

Preface

This is an introductory textbook for a first course in applied statistics and probability for undergraduate students in engineering and the physical or chemical sciences. These individuals play a significant role in designing and developing new products and manufacturing systems and processes, and they also improve existing systems. Statistical methods are an important tool in these activities because they provide the engineer with both descriptive and analytical methods for dealing with the variability in observed data. Although many of the methods we present are fundamental to statistical analysis in other disciplines, such as business and management, the life sciences, and the social sciences, we have elected to focus on an engineering-oriented audience. We believe that this approach will best serve students in engineering and the chemical/physical sciences and will allow them to concentrate on the many applications of statistics in these disciplines. We have worked hard to ensure that our examples and exercises are engineering- and science-based, and in almost all cases we have used examples of real data—either taken from a published source or based on our consulting experiences.

We believe that engineers in all disciplines should take at least one course in statistics. Unfortunately, because of other requirements, most engineers will only take one statistics course. This book can be used for a single course, although we have provided enough material for two courses in the hope that more students will see the important applications of statistics in their everyday work and elect a second course. We believe that this book will also serve as a useful reference.

ORGANIZATION OF THE BOOK

We have retained the relatively modest mathematical level of the first two editions. We have found that engineering students who have completed one or two semesters of calculus should have no difficulty reading almost all of the text. It is our intent to give the reader an understanding of the methodology and how to apply it, not the mathematical theory. We have made many enhancements in this edition, including reorganizing and rewriting major portions of the book.

Perhaps the most common criticism of engineering statistics texts is that they are too long. Both instructors and students complain that it is impossible to cover all of the topics in the book in one or even two terms. For authors, this is a serious issue because there is great variety in both the content and level of these courses, and the decisions about what material to delete without limiting the value of the text are not easy. After struggling with these issues, we decided to divide the text into two components; a set of core topics, many of which are most

likely to be covered in an engineering statistics course, and a set of supplementary topics, or topics that will be useful for some but not all courses. The core topics are in the printed book, and the complete text (both core and supplementary topics) is available on the CD that is included with the printed book. Decisions about topics to include in print and which to include only on the CD were made based on the results of a recent survey of instructors.

The *Interactive e-Text* consists of the complete text and a wealth of additional material and features. The text and links on the CD are navigated using Adobe Acrobat™. The links within the *Interactive e-Text* include the following: (1) from the Table of Contents to the selected *eText* sections, (2) from the Index to the selected topic within the *e-Text*, (3) from reference to a figure, table, or equation in one section to the actual figure, table, or equation in another section (all figures can be enlarged and printed), (4) from end-of-chapter Important Terms and Concepts to their definitions within the chapter, (5) from in-text **boldfaced terms** to their corresponding Glossary definitions and explanations, (6) from in-text references to the corresponding Appendix tables and charts, (7) from boxed-number end-of-chapter exercises (essentially most odd-numbered exercises) to their answers, (8) from some answers to the complete problem solution, and (9) from the opening splash screen to the textbook Web site.

Chapter 1 is an introduction to the field of statistics and how engineers use statistical methodology as part of the engineering problem-solving process. This chapter also introduces the reader to some engineering applications of statistics, including building empirical models, designing engineering experiments, and monitoring manufacturing processes. These topics are discussed in more depth in subsequent chapters.

Chapters 2, 3, 4, and 5 cover the basic concepts of probability, discrete and continuous random variables, probability distributions, expected values, joint probability distributions, and independence. We have given a reasonably complete treatment of these topics but have avoided many of the mathematical or more theoretical details.

Chapter 6 begins the treatment of statistical methods with random sampling; data summary and description techniques, including stem-and-leaf plots, histograms, box plots, and probability plotting; and several types of time series plots. Chapter 7 discusses point estimation of parameters. This chapter also introduces some of the important properties of estimators, the method of maximum likelihood, the method of moments, sampling distributions, and the central limit theorem.

Chapter 8 discusses interval estimation for a single sample. Topics included are confidence intervals for means, variances or standard deviations, and proportions and prediction and tolerance intervals. Chapter 9 discusses hypothesis tests for a single sample. Chapter 10 presents tests and confidence intervals for two samples. This material has been extensively rewritten and reorganized. There is detailed information and examples of methods for determining appropriate sample sizes. We want the student to become familiar with how these techniques are used to solve real-world engineering problems and to get some understanding of the concepts behind them. We give a logical, heuristic development of the procedures, rather than a formal mathematical one.

Chapters 11 and 12 present simple and multiple linear regression. We use matrix algebra throughout the multiple regression material (Chapter 12) because it is the only easy way to understand the concepts presented. Scalar arithmetic presentations of multiple regression are awkward at best, and we have found that undergraduate engineers are exposed to enough matrix algebra to understand the presentation of this material.

Chapters 13 and 14 deal with single- and multifactor experiments, respectively. The notions of randomization, blocking, factorial designs, interactions, graphical data analysis, and fractional factorials are emphasized. Chapter 15 gives a brief introduction to the methods and applications of nonparametric statistics, and Chapter 16 introduces statistical quality control, emphasizing the control chart and the fundamentals of statistical process control.

Each chapter has an extensive collection of exercises, including end-of-section exercises that emphasize the material in that section, supplemental exercises at the end of the chapter that cover the scope of chapter topics, and mind-expanding exercises that often require the student to extend the text material somewhat or to apply it in a novel situation. As noted above, answers are provided to most odd-numbered exercises and the *e-Text* contains complete solutions to selected exercises.

USING THE BOOK

This is a very flexible textbook because instructors' ideas about what should be in a first course on statistics for engineers vary widely, as do the abilities of different groups of students. Therefore, we hesitate to give too much advice but will explain how we use the book.

We believe that a first course in statistics for engineers should be primarily an applied statistics course, not a probability course. In our one-semester course we cover all of Chapter 1 (in one or two lectures); overview the material on probability, putting most of the emphasis on the normal distribution (six to eight lectures); discuss most of Chapters 6 though 10 on confidence intervals and tests (twelve to fourteen lectures); introduce regression models in Chapter 11 (four lectures); give an introduction to the design of experiments from Chapters 13 and 14 (six lectures); and present the basic concepts of statistical process control, including the Shewhart control chart from Chapter 16 (four lectures). This leaves about three to four periods for exams and review. Let us emphasize that the purpose of this course is to introduce engineers to how statistics can be used to solve real-world engineering problems, not to weed out the less mathematically gifted students. This course is not the "baby math-stat" course that is all too often given to engineers.

If a second semester is available, it is possible to cover the entire book, including much of the *e-Text* material, if appropriate for the audience. It would also be possible to assign and work many of the homework problems in class to reinforce the understanding of the concepts. Obviously, multiple regression and more design of experiments would be major topics in a second course.

USING THE COMPUTER

In practice, engineers use computers to apply statistical methods to solve problems. Therefore, we strongly recommend that the computer be integrated into the class. Throughout the book we have presented output from Minitab as typical examples of what can be done with modern statistical software. In teaching, we have used other software packages, including Statgraphics, JMP, and Statisticia. We did not clutter up the book with examples from many different packages because how the instructor integrates the software into the class is ultimately more important than which package is used. All text data is available in electronic form on the *e-Text* CD. In some chapters, there are problems that we feel should be worked using computer software. We have marked these problems with a special icon in the margin.



In our own classrooms, we use the computer in almost every lecture and demonstrate how the technique is implemented in software as soon as it is discussed in the lecture. Student versions of many statistical software packages are available at low cost, and students can either purchase their own copy or use the products available on the PC local area networks. We have found that this greatly improves the pace of the course and student understanding of the material.

USING THE WEB

Additional resources for students and instructors can be found at www.wiley.com/college/montgomery/.

ACKNOWLEDGMENTS

We would like to express our grateful appreciation to the many organizations and individuals who have contributed to this book. Many instructors who used the first two editions provided excellent suggestions that we have tried to incorporate in this revision. We also thank Professors Manuel D. Rossetti (University of Arkansas), Bruce Schmeiser (Purdue University), Michael G. Akritas (Penn State University), and Arunkumar Pennathur (University of Texas at El Paso) for their insightful reviews of the manuscript of the third edition. We are also indebted to Dr. Smiley Cheng for permission to adapt many of the statistical tables from his excellent book (with Dr. James Fu), *Statistical Tables for Classroom and Exam Room.* John Wiley and Sons, Prentice Hall, the Institute of Mathematical Statistics, and the editors of Biometrics allowed us to use copyrighted material, for which we are grateful. Thanks are also due to Dr. Lora Zimmer, Dr. Connie Borror, and Dr. Alejandro Heredia-Langner for their outstanding work on the solutions to exercises.

Douglas C. Montgomery

George C. Runger

Contents

CHAPTER 1 The Role of Statistics in Engineering 1

1-1 The Engineering Method and Statistical Thinking 2
1-2 Collecting Engineering Data 5
1-2.1 Basic Principles 5
1-2.2 Retrospective Study 5
1-2.3 Observational Study 6
1-2.4 Designed Experiments 6
1-2.5 A Factorial Experiment for the Connector Pull-Off Force Problem (CD Only) 8
1-2.6 Observing Processes Over Time 8
1-3 Mechanistic and Empirical Models 11
1-4 Probability and Probability Models 14

CHAPTER 2 Probability 16

2-1 Sample Spaces and Events 17
2-1.1 Random Experiments 17
2-1.2 Sample Spaces 18
2-1.3 Events 22
2-1.4 Counting Techniques (CD Only) 25
2-2 Interpretations of Probability 27
2-2.1 Introduction 27
2-2.2 Axioms of Probability 30
2-3 Addition Rules 33
2-4 Conditional Probability 37
2-5 Multiplication and Total Probability Rules 42
2-5.1 Multiplication Rule 42
2-5.2 Total Probability Rule 43
2-6 Independence 46
2-7 Bayes' Theorem 51
2-8 Random Variables 53

CHAPTER 3 Discrete Random Variables and Probability Distributions 59

3-1 Discrete Random Variables 60
3-2 Probability Distributions and Probability Mass Functions 61
3-3 Cumulative Distribution Functions 63
3-4 Mean and Variance of a Discrete Random Variable 66
3-5 Discrete Uniform Distribution 70
3-6 Binomial Distribution 72
3-7 Geometric and Negative Binomial Distributions 78
3-7.1 Geometric Distribution 78
3-7.2 Negative Binomial Distribution 80
3-8 Hypergeometric Distribution 84
3-9 Poisson Distribution 89

CHAPTER 4 Continuous Random Variables and Probability Distributions 97

4-1 Continuous Random Variables 98
4-2 Probability Distributions and Probability Density Functions 98
4-3 Cumulative Distribution Functions 102
4-4 Mean and Variance of a Continuous Random Variable 105
4-5 Continuous Uniform Distribution 107
4-6 Normal Distribution 109
4-7 Normal Approximation to the Binomial and Poisson Distributions 118
4-8 Continuity Corrections to Improve the Approximation (CD Only) 122
4-9 Exponential Distribution 122
4-10 Erlang and Gamma Distribution 128
4-10.1 Erlang Distribution 128
4-10.2 Gamma Distribution 130
4-11 Weibull Distribution 133
4-12 Lognormal Distribution 135

CHAPTER 5 Joint Probability Distributions 141

5-1 Two Discrete Random Variables 142
5-1.1 Joint Probability Distributions 142
5-1.2 Marginal Probability Distributions 144
5-1.3 Conditional Probability Distributions 146
5-1.4 Independence 148
5-2 Multiple Discrete Random Variables 151
5-2.1 Joint Probability Distributions 151
5-2.2 Multinomial Probability Distribution 154
5-3 Two Continuous Random Variables 157
5-3.1 Joint Probability Distributions 157
5-3.2 Marginal Probability Distributions 159
5-3.3 Conditional Probability Distributions 162
5-3.4 Independence 164
5-4 Multiple Continuous Random Variables 167
5-5 Covariance and Correlation 171
5-6 Bivariate Normal Distribution 177
5-7 Linear Combinations of Random Variables 180
5-8 Functions of Random Variables (CD Only) 185
5-9 Moment Generating Functions (CD Only) 185
5-10 Chebyshev's Inequality (CD Only) 185

CHAPTER 6 Random Sampling and Data Description 189

6-1 Data Summary and Display 190
6-2 Random Sampling 195
6-3 Stem-and-Leaf Diagrams 197
6-4 Frequency Distributions and Histograms 203
6-5 Box Plots 207
6-6 Time Sequence Plots 209
6-7 Probability Plots 212
6-8 More About Probability Plotting (CD Only) 216

CHAPTER 7 Point Estimation of Parameters 220

7-1 Introduction 221
7-2 General Concepts of Point Estimation 222
7-2.1 Unbiased Estimators 222
7-2.2 Proof that S is a Biased Estimator of $\square$ (CD Only) 224
7-2.3 Variance of a Point Estimator 224
7-2.4 Standard Error: Reporting a Point Estimator 225
7-2.5 Bootstrap Estimate of the Standard Error (CD Only) 226
7-2.6 Mean Square Error of an Estimator 226
7-3 Methods of Point Estimation 229
7-3.1 Method of Moments 229
7-3.2 Method of Maximum Likelihood 230
7-3.3 Bayesian Estimation of Parameters (CD Only) 237
7-4 Sampling Distributions 238
7-5 Sampling Distribution of Means 239

CHAPTER 8 Statistical Intervals for a Single Sample 247

8-1 Introduction 248
8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known 249
8-2.1 Development of the Confidence Interval and Its Basic Properties 249
8-2.2 Choice of Sample Size 252
8-2.3 One-sided Confidence Bounds 253
8-2.4 General method to Derive a Confidence Interval 253
8-2.5 A Large-Sample Confidence Interval for $\square$ 254
8-2.6 Bootstrap Confidence Intervals (CD Only) 256
8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown 257
8-3.1 The t Distribution 258
8-3.2 Development of the t Distribution (CD Only) 259
8-3.3 The t Confidence Interval on $\square$ 259

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution 261
8-5 A Large-Sample Confidence Interval for a Population Proportion 265
8-6 A Prediction Interval for a Future Observation 268
8-7 Tolerance Intervals for a Normal Distribution 270

CHAPTER 9 Tests of Hypotheses for a Single Sample 277

9-1 Hypothesis Testing 278
9-1.1 Statistical Hypotheses 278
9-1.2 Tests of Statistical Hypotheses 280
9-1.3 One-Sided and Two-Sided Hypotheses 286
9-1.4 General Procedure for Hypothesis Testing 287
9-2 Tests on the Mean of a Normal Distribution, Variance Known 289
9-2.1 Hypothesis Tests on the Mean 289
9-2.2 *P*-Values in Hypothesis Tests 292
9-2.3 Connection Between Hypothesis Tests and Confidence Intervals 293
9-2.4 Type II Error and Choice of Sample Size 293
9-2.5 Large Sample Test 297
9-2.6 Some Practical Comments on Hypothesis Tests 298
9-3 Tests on the Mean of a Normal Distribution, Variance Unknown 300
9-3.1 Hypothesis Tests on the Mean 300
9-3.2 *P*-Value for a *t*-Test 303
9-3.3 Choice of Sample Size 304
9-3.4 Likelihood Ratio Approach to Development of Test Procedures (CD Only) 305
9-4 Tests on the Variance and Standard Deviation of a Normal Distribution 307
9-4.1 The Hypothesis Testing Procedures 307
9-4.2 □-Error and Choice of Sample Size 309
9-5 Tests on a Population Proportion 310
9-5.1 Large-Sample Tests on a Proportion 310
9-5.2 Small-Sample Tests on a Proportion (CD Only) 312
9-5.3 Type II Error and Choice of Sample Size 312
9-6 Summary of Inference Procedures for a Single Sample 315
9-7 Testing for Goodness of Fit 315
9-8 Contingency Table Tests 320

CHAPTER 10 Statistical Inference for Two Samples 327

10-1 Introduction 328
10-2 Inference For a Difference in Means of Two Normal Distributions, Variances Known 328
10-2.1 Hypothesis Tests for a Difference in Means, Variances Known 329
10-2.2 Choice of Sample Size 331
10-2.3 Identifying Cause and Effect 333
10-2.4 Confidence Interval on a Difference in Means, Variances Known 334
10-3 Inference For a Difference in Means of Two Normal Distributions, Variances Unknown 337
10-3.1 Hypothesis Tests for a Difference in Means, Variances Unknown 337
10-3.2 More About the Equal Variance Assumption (CD Only) 344
10-3.3 Choice of Sample Size 344
10-3.4 Confidence Interval on a Difference in Means, Variances Unknown 345
10-4 Paired *t*-Test 349
10-5 Inference on the Variances of Two Normal Distributions 355
10-5.1 The *F* Distribution 355
10-5.2 Development of the *F* Distribution (CD Only) 357
10-5.3 Hypothesis Tests on the Ratio of Two Variances 357
10-5.4 □-Error and Choice of Sample Size 359
10-5.5 Confidence Interval on the Ratio of Two Variances 359

10-6 Inference on Two Population Proportions 361
10-6.1 Large-Sample Test for $H_0: p_1 = p_2$ 361
10-6.2 Small Sample Test for $H_0: p_1 = p_2$ (CD Only) 364
10-6.3 -Error and Choice of Sample Size 364
10-6.4 Confidence Interval for $P_1 - P_2$ 365
10-7 Summary Table for Inference Procedures for Two Samples 367

CHAPTER 11 Simple Linear Regression and Correlation 372

11-1 Empirical Models 373
11-2 Simple Linear Regression 375
11-3 Properties of the Least Squares Estimators 383
11-4 Some Comments on Uses of Regression (CD Only) 384
11-5 Hypothesis Tests in Simple Linear Regression 384
11-5.1 Use of *t*-Tests 384
11-5.2 Analysis of Variance Approach to Test Significance of Regression 387
11-6 Confidence Intervals 389
11-6.1 Confidence Intervals on the Slope and Intercept 389
11-6.2 Confidence Interval on the Mean Response 390
11-7 Prediction of New Observations 392
11-8 Adequacy of the Regression Model 395
11-8.1 Residual Analysis 395
11-8.2 Coefficient of Determination (R^2) 397
11-8.3 Lack-of-Fit Test (CD Only) 398
11-9 Transformations to a Straight Line 400
11-10 More About Transformations (CD Only) 400
11-11 Correlation 400

CHAPTER 12 Multiple Linear Regression 410

12-1 Multiple Linear Regression Model 411
12-1.1 Introduction 411
12-1.2 Least Squares Estimation of the Parameters 414
12-1.3 Matrix Approach to Multiple Linear Regression 417
12-1.4 Properties of the Least Squares Estimators 421
12-2 Hypothesis Tests in Multiple Linear Regression 428
12-2.1 Test for Significance of Regression 428
12-2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients 432
12-2.3 More About the Extra Sum of Squares Method (CD Only) 435
12-3 Confidence Intervals in Multiple Linear Regression 437
12-3.1 Confidence Intervals on Individual Regression Coefficients 437
12-3.2 Confidence Interval on the Mean Response 438
12-4 Prediction of New Observations 439
12-5 Model Adequacy Checking 441
12-5.1 Residual Analysis 441
12-5.2 Influential Observations 444
12-6 Aspects of Multiple Regression Modeling 447
12-6.1 Polynomial Regression Models 447
12-6.2 Categorical Regressors and Indicator Variables 450
12-6.3 Selection of Variables and Model Building 452
12-6.4 Multicollinearity 460
12-6.5 Ridge Regression (CD Only) 461
12-6.6 Nonlinear Regression Models (CD Only) 461

CHAPTER 13 Design and Analysis of Single-Factor Experiments: The Analysis of Variance 468

13-1 Designing Engineering Experiments 469
13-2 The Completely Randomized Single-Factor Experiment 470
13-2.1 An Example 470
13-2.2 The Analysis of Variance 472
13-2.3 Multiple Comparisons Following the ANOVA 479

13-2.4 More About Multiple Comparisons (CD Only) 481
13-2.5 Residual Analysis and Model Checking 481
13-2.6 Determining Sample Size 482
13-2.7 Technical Details about the Analysis of Variance (CD Only) 485
13-3 The Random Effects Model 487
13-3.1 Fixed Versus Random Factors 487
13-3.2 ANOVA and Variance Components 487
13-3.3 Determining Sample Size in the Random Model (CD Only) 490
13-4 Randomized Complete Block Design 491
13-4.1 Design and Statistical Analysis 491
13-4.2 Multiple Comparisons 497
13-4.3 Residual Analysis and Model Checking 498
13-4.4 Randomized Complete Block Design with Random Factors (CD Only) 498

CHAPTER 14 Design of Experiments with Several Factors 505

14-1 Introduction 506
14-2 Some Applications of Designed Experiments (CD Only) 506
14-3 Factorial Experiments 506
14-4 Two-Factor Factorial Experiments 510
14-4.1 Statistical Analysis of the Fixed-Effects Model 511
14-4.2 Model Adequacy Checking 517
14-4.3 One Observation Per Cell 517
14-4.4 Factorial Experiments with Random Factors: Overview 518
14-5 General Factorial Experiments 520
14-6 Factorial Experiments with Random Factors (CD Only) 523
14-7 2^k Factorial Designs 523
14-7.1 2^2 Design 524
14-7.2 2^k Design for $k \square 3$ Factors 529
14-7.3 Single Replicate of the 2^k Design 537
14-7.4 Addition of Center Points to a 2^k Design (CD Only) 541
14-8 Blocking and Confounding in the 2^k Design 543
14-9 Fractional Replication of the 2^k Design 549
14-9.1 One Half Fraction of the 2^k Design 549
14-9.2 Smaller Fractions: The $2^{k \square p}$ Fractional Factorial 555
14-10 Response Surface Methods and Designs (CD Only) 564

CHAPTER 15 Nonparametric Statistics 571

15-1 Introduction 572
15-2 Sign Test 572
15-2.1 Description of the Test 572
15-2.2 Sign Test for Paired Samples 576
15-2.3 Type II Error for the Sign Test 578
15-2.4 Comparison to the *t*-Test 579
15-3 Wilcoxon Signed-Rank Test 581
15-3.1 Description of the Test 581
15-3.2 Large-Sample Approximation 583
15-3.3 Paired Observations 583
15-3.4 Comparison to the *t*-Test 584
15-4 Wilcoxon Rank-Sum Test 585
15-4.1 Description of the Test 585
15-4.2 Large-Sample Approximation 587
15-4.3 Comparison to the *t*-Test 588
15-5 Nonparametric Methods in the Analysis of Variance 589
15-5.1 Kruskal-Wallis Test 589
15-5.2 Rank Transformation 591

CHAPTER 16 Statistical Quality Control 595

16-1 Quality Improvement and Statistics 596
16-2 Statistical Quality Control 597
16-3 Statistical Process Control 597
16-4 Introduction to Control Charts 598
16-4.1 Basic Principles 598
16-4.2 Design of a Control Chart 602
16-4.3 Rational Subgroups 603
16-4.4 Analysis of Patterns on Control Charts 604
16-5 $\bar{X}$ and *R* or *S* Control Chart 607
16-6 Control Charts for Individual Measurements 615

16-7 Process Capability 619
16-8 Attribute Control Charts 625
16-8.1 *P* Chart (Control Chart for Proportion) 625
16-8.2 *U* Chart (Control Chart for Defects per Unit) 627
16-9 Control Chart Performance 630
16-10 Cumulative Sum Control Chart 632
16-11 Other SPC Problem-Solving Tools 639
16-12 Implementing SPC 641

APPENDICES 649

APPENDIX A: Statistical Tables and Charts 651

Table I Summary of Common Probability Distributions 652
Table II Cumulative Standard Normal Distribution 653
Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution 655
Table IV Percentage Points $t_{\alpha,\nu}$ of the *t*-distribution 656
Table V Percentage Points f_{α,ν_1,ν_2} of the *F*-distribution 657
Chart VI Operating Characteristic Curves 662
Table VII Critical Values for the Sign Test 671
Table VIII Critical Values for the Wilcoxon Signed-Rank Test 671
Table IX Critical Values for the Wilcoxon Rank-Sum Test 672
Table X Factors for Constructing Variables Control Charts 673
Table XI Factors for Tolerance Intervals 674

APPENDIX B: Bibliography 677

APPENDIX C: Answers to Selected Exercises 679

GLOSSARY 689

INDEX 703

PROBLEM SOLUTIONS

1 The Role of Statistics in Engineering

CHAPTER OUTLINE

1-1 THE ENGINEERING METHOD AND STATISTICAL THINKING

1-2 COLLECTING ENGINEERING DATA

- 1-2.1 Basic Principles
- 1-2.2 Retrospective Study
- 1-2.3 Observational Study
- 1-2.4 Designed Experiments
- 1-2.5 A Factorial Experiment for the Pull-off Force Problem (CD Only)
- 1-2.6 Observing Processes Over Time

1-3 MECHANISTIC AND EMPIRICAL MODELS

1-4 PROBABILITY AND PROBABILITY MODELS

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Identify the role that statistics can play in the engineering problem-solving process
2. Discuss how variability affects the data collected and used for making engineering decisions
3. Explain the difference between enumerative and analytical studies
4. Discuss the different methods that engineers use to collect data
5. Identify the advantages that designed experiments have in comparison to other methods of collecting engineering data
6. Explain the differences between mechanistic models and empirical models
7. Discuss how probability and probability models are used in engineering and science

CD MATERIAL

8. Explain the factorial experimental design.
9. Explain how factors can Interact.

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also

available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

1-1 THE ENGINEERING METHOD AND STATISTICAL THINKING

An engineer is someone who solves problems of interest to society by the efficient application of scientific principles. Engineers accomplish this by either refining an existing product or process or by designing a new product or process that meets customers' needs. The **engineering,** or **scientific, method** is the approach to formulating and solving these problems. The steps in the engineering method are as follows:

1. Develop a clear and concise description of the problem.
2. Identify, at least tentatively, the important factors that affect this problem or that may play a role in its solution.
3. Propose a model for the problem, using scientific or engineering knowledge of the phenomenon being studied. State any limitations or assumptions of the model.
4. Conduct appropriate experiments and collect data to test or validate the tentative model or conclusions made in steps 2 and 3.
5. Refine the model on the basis of the observed data.
6. Manipulate the model to assist in developing a solution to the problem.
7. Conduct an appropriate experiment to confirm that the proposed solution to the problem is both effective and efficient.
8. Draw conclusions or make recommendations based on the problem solution.

The steps in the engineering method are shown in Fig. 1-1. Notice that the engineering method features a strong interplay between the problem, the factors that may influence its solution, a model of the phenomenon, and experimentation to verify the adequacy of the model and the proposed solution to the problem. Steps 2–4 in Fig. 1-1 are enclosed in a box, indicating that several cycles or iterations of these steps may be required to obtain the final solution. Consequently, engineers must know how to efficiently plan experiments, collect data, analyze and interpret the data, and understand how the observed data are related to the model they have proposed for the problem under study.

The field of **statistics** deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes. Because many aspects of engineering practice involve working with data, obviously some knowledge of statistics is important to any engineer. Specifically, statistical techniques can be a powerful aid in designing new products and systems, improving existing designs, and designing, developing, and improving production processes.

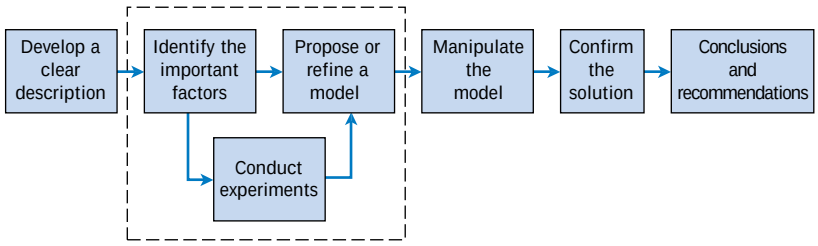


Figure 1-1 The engineering method.

Statistical methods are used to help us describe and understand **variability.** By variability, we mean that successive observations of a system or phenomenon do not produce exactly the same result. We all encounter variability in our everyday lives, and **statistical thinking** can give us a useful way to incorporate this variability into our decision-making processes. For example, consider the gasoline mileage performance of your car. Do you always get exactly the same mileage performance on every tank of fuel? Of course not—in fact, sometimes the mileage performance varies considerably. This observed variability in gasoline mileage depends on many factors, such as the type of driving that has occurred most recently (city versus highway), the changes in condition of the vehicle over time (which could include factors such as tire inflation, engine compression, or valve wear), the brand and/or octane number of the gasoline used, or possibly even the weather conditions that have been recently experienced. These factors represent potential **sources of variability** in the system. Statistics gives us a framework for describing this variability and for learning about which potential sources of variability are the most important or which have the greatest impact on the gasoline mileage performance.

We also encounter variability in dealing with engineering problems. For example, suppose that an engineer is designing a nylon connector to be used in an automotive engine application. The engineer is considering establishing the design specification on wall thickness at 3□32 inch but is somewhat uncertain about the effect of this decision on the connector pull-off force. If the pull-off force is too low, the connector may fail when it is installed in an engine. Eight prototype units are produced and their pull-off forces measured, resulting in the following data (in pounds): 12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1. As we anticipated, not all of the prototypes have the same pull-off force. We say that there is variability in the pull-off force measurements. Because the pull-off force measurements exhibit variability, we consider the pull-off force to be a **random variable.** A convenient way to think of a random variable, say X, that represents a measurement, is by using the model

$$X = \square + \square \tag{1-1}$$

where □ is a constant and □ is a random disturbance. The constant remains the same with every measurement, but small changes in the environment, test equipment, differences in the individual parts themselves, and so forth change the value of □. If there were no disturbances, □ would always equal zero and X would always be equal to the constant □. However, this never happens in the real world, so the actual measurements X exhibit variability. We often need to describe, quantify and ultimately reduce variability.

Figure 1-2 presents a **dot diagram** of these data. The dot diagram is a very useful plot for displaying a small body of data—say, up to about 20 observations. This plot allows us to see easily two features of the data; the **location,** or the middle, and the **scatter** or **variability.** When the number of observations is small, it is usually difficult to identify any specific patterns in the variability, although the dot diagram is a convenient way to see any unusual data features.

The need for statistical thinking arises often in the solution of engineering problems. Consider the engineer designing the connector. From testing the prototypes, he knows that the average pull-off force is 13.0 pounds. However, he thinks that this may be too low for the



Figure 1-2 Dot diagram of the pull-off force data when wall thickness is 3/32 inch.

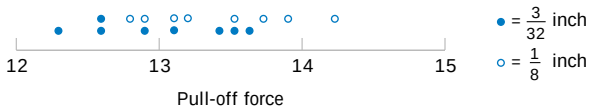


Figure 1-3 Dot diagram of pull-off force for two wall thicknesses.

intended application, so he decides to consider an alternative design with a greater wall thickness, 1□8 inch. Eight prototypes of this design are built, and the observed pull-off force measurements are 12.9, 13.7, 12.8, 13.9, 14.2, 13.2, 13.5, and 13.1. The average is 13.4. Results for both samples are plotted as dot diagrams in Fig. 1-3, page 3. This display gives the impression that increasing the wall thickness has led to an increase in pull-off force. However, there are some obvious questions to ask. For instance, how do we know that another sample of prototypes will not give different results? Is a sample of eight prototypes adequate to give reliable results? If we use the test results obtained so far to conclude that increasing the wall thickness increases the strength, what risks are associated with this decision? For example, is it possible that the apparent increase in pull-off force observed in the thicker prototypes is only due to the inherent variability in the system and that increasing the thickness of the part (and its cost) really has no effect on the pull-off force?

Often, physical laws (such as Ohm's law and the ideal gas law) are applied to help design products and processes. We are familiar with this reasoning from general laws to specific cases. But it is also important to reason from a specific set of measurements to more general cases to answer the previous questions. This reasoning is from a sample (such as the eight connectors) to a population (such as the connectors that will be sold to customers). The reasoning is referred to as **statistical inference.** See Fig. 1-4. Historically, measurements were obtained from a sample of people and generalized to a population, and the terminology has remained. Clearly, reasoning based on measurements from some objects to measurements on all objects can result in errors (called sampling errors). However, if the sample is selected properly, these risks can be quantified and an appropriate sample size can be determined.

In some cases, the sample is actually selected from a well-defined population. The sample is a subset of the population. For example, in a study of resistivity a sample of three wafers might be selected from a production lot of wafers in semiconductor manufacturing. Based on the resistivity data collected on the three wafers in the sample, we want to draw a conclusion about the resistivity of all of the wafers in the lot.

In other cases, the population is conceptual (such as with the connectors), but it might be thought of as future replicates of the objects in the sample. In this situation, the eight prototype connectors must be representative, in some sense, of the ones that will be manufactured in the future. Clearly, this analysis requires some notion of **stability** as an additional assumption. For example, it might be assumed that the sources of variability in the manufacture of the prototypes (such as temperature, pressure, and curing time) are the same as those for the connectors that will be manufactured in the future and ultimately sold to customers.

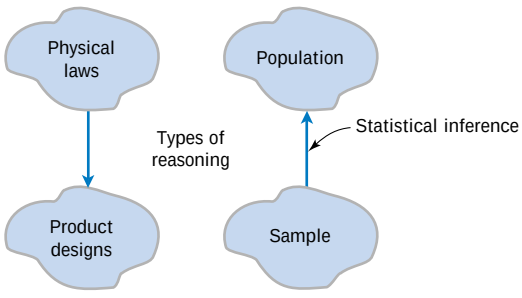


Figure 1-4 Statistical inference is one type of reasoning.

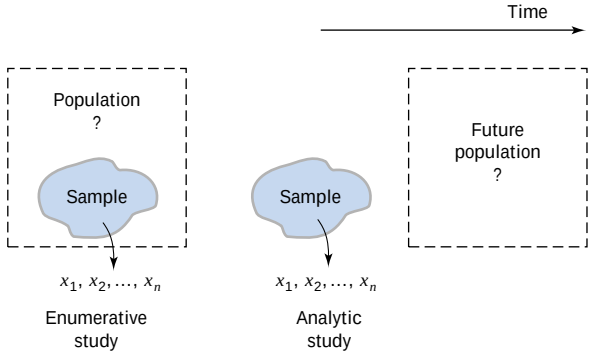


Figure 1-5 Enumerative versus analytic study.

The wafers-from-lots example is called an **enumerative** study. A sample is used to make an inference to the population from which the sample is selected. The connector example is called an **analytic** study. A sample is used to make an inference to a conceptual (future) population. The statistical analyses are usually the same in both cases, but an analytic study clearly requires an assumption of stability. See Fig. 1-5, on page 4.

1-2 COLLECTING ENGINEERING DATA

1-2.1 Basic Principles

In the previous section, we illustrated some simple methods for summarizing data. In the engineering environment, the data is almost always a sample that has been selected from some population. Three basic methods of collecting data are

- A **retrospective** study using historical data
- An **observational** study
- A **designed experiment**

An effective data collection procedure can greatly simplify the analysis and lead to improved understanding of the population or process that is being studied. We now consider some examples of these data collection methods.

1-2.2 Retrospective Study

Montgomery, Peck, and Vining (2001) describe an acetone-butyl alcohol distillation column for which concentration of acetone in the distillate or output product stream is an important variable. Factors that may affect the distillate are the reboil temperature, the condensate temperature, and the reflux rate. Production personnel obtain and archive the following records:

- The concentration of acetone in an hourly test sample of output product
- The reboil temperature log, which is a plot of the reboil temperature over time
- The condenser temperature controller log
- The nominal reflux rate each hour

The reflux rate should be held constant for this process. Consequently, production personnel change this very infrequently.

A retrospective study would use either all or a sample of the historical process data archived over some period of time. The study objective might be to discover the relationships among the two temperatures and the reflux rate on the acetone concentration in the output product stream. However, this type of study presents some problems:

1. We may not be able to see the relationship between the reflux rate and acetone concentration, because the reflux rate didn't change much over the historical period.
2. The archived data on the two temperatures (which are recorded almost continuously) do not correspond perfectly to the acetone concentration measurements (which are made hourly). It may not be obvious how to construct an approximate correspondence.

3. Production maintains the two temperatures as closely as possible to desired targets or set points. Because the temperatures change so little, it may be difficult to assess their real impact on acetone concentration.
4. Within the narrow ranges that they do vary, the condensate temperature tends to increase with the reboil temperature. Consequently, the effects of these two process variables on acetone concentration may be difficult to separate.

As you can see, a retrospective study may involve a lot of **data,** but that data may contain relatively little useful **information** about the problem. Furthermore, some of the relevant data may be missing, there may be transcription or recording errors resulting in **outliers** (or unusual values), or data on other important factors may not have been collected and archived. In the distillation column, for example, the specific concentrations of butyl alcohol and acetone in the input feed stream are a very important factor, but they are not archived because the concentrations are too hard to obtain on a routine basis. As a result of these types of issues, statistical analysis of historical data sometimes identify interesting phenomena, but solid and reliable explanations of these phenomena are often difficult to obtain.

1-2.3 Observational Study

In an observational study, the engineer observes the process or population, disturbing it as little as possible, and records the quantities of interest. Because these studies are usually conducted for a relatively short time period, sometimes variables that are not routinely measured can be included. In the distillation column, the engineer would design a form to record the two temperatures and the reflux rate when acetone concentration measurements are made. It may even be possible to measure the input feed stream concentrations so that the impact of this factor could be studied. Generally, an observational study tends to solve problems 1 and 2 above and goes a long way toward obtaining accurate and reliable data. However, observational studies may not help resolve problems 3 and 4.

1-2.4 Designed Experiments

In a designed experiment the engineer makes *deliberate* or *purposeful changes* in the controllable variables of the system or process, observes the resulting system output data, and then makes an inference or decision about which variables are responsible for the observed changes in output performance. The nylon connector example in Section 1-1 illustrates a designed experiment; that is, a deliberate change was made in the wall thickness of the connector with the objective of discovering whether or not a greater pull-off force could be obtained. Designed experiments play a very important role in engineering design and development and in the improvement of manufacturing processes. Generally, when products and processes are designed and developed with designed experiments, they enjoy better performance, higher reliability, and lower overall costs. Designed experiments also play a crucial role in reducing the lead time for engineering design and development activities.

For example, consider the problem involving the choice of wall thickness for the nylon connector. This is a simple illustration of a designed experiment. The engineer chose two wall thicknesses for the connector and performed a series of tests to obtain pull-off force measurements at each wall thickness. In this simple **comparative experiment,** the

engineer is interested in determining if there is any difference between the 3□32- and 1□8-inch designs. An approach that could be used in analyzing the data from this experiment is to compare the mean pull-off force for the 3□32-inch design to the mean pull-off force for the 1□8-inch design using statistical **hypothesis testing,** which is discussed in detail in Chapters 9 and 10. Generally, a **hypothesis** is a statement about some aspect of the system in which we are interested. For example, the engineer might want to know if the mean pull-off force of a 3□32-inch design exceeds the typical maximum load expected to be encountered in this application, say 12.75 pounds. Thus, we would be interested in testing the hypothesis that the mean strength exceeds 12.75 pounds. This is called a **single-sample hypothesis testing problem.** It is also an example of an **analytic study.** Chapter 9 presents techniques for this type of problem. Alternatively, the engineer might be interested in testing the hypothesis that increasing the wall thickness from 3□32- to 1□8-inch results in an increase in mean pull-off force. Clearly, this is an **analytic study;** it is also an example of **a two-sample hypothesis testing problem.** Two-sample hypothesis testing problems are discussed in Chapter 10.

Designed experiments are a very powerful approach to studying complex systems, such as the distillation column. This process has three factors, the two temperatures and the reflux rate, and we want to investigate the effect of these three factors on output acetone concentration. A good experimental design for this problem must ensure that we can separate the effects of all three factors on the acetone concentration. The specified values of the three factors used in the experiment are called **factor levels.** Typically, we use a small number of levels for each factor, such as two or three. For the distillation column problem, suppose we use a "high," and "low," level (denoted +1 and □1, respectively) for each of the factors. We thus would use two levels for each of the three factors. A very reasonable experiment design strategy uses every possible combination of the factor levels to form a basic experiment with eight different settings for the process. This type of experiment is called a **factorial experiment.** Table 1-1 presents this experimental design.

Figure 1-6, on page 8, illustrates that this design forms a cube in terms of these high and low levels. With each setting of the process conditions, we allow the column to reach equilibrium, take a sample of the product stream, and determine the acetone concentration. We then can draw specific inferences about the effect of these factors. Such an approach allows us to proactively study a population or process. Designed experiments play a very important role in engineering and science. Chapters 13 and 14 discuss many of the important principles and techniques of experimental design.

Table 1-1 The Designed Experiment (Factorial Design) for the Distillation Column

Reboil Temp.	Condensate Temp.	Reflux Rate
□1	□1	□1
□1	□1	□1
□1	□1	□1
□1	□1	□1
□1	□1	□1
□1	□1	□1
□1	□1	□1
□1	□1	□1

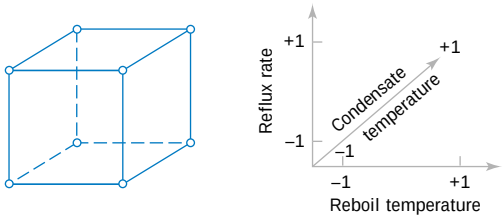


Figure 1-6 The factorial design for the distillation column.

1-2.5 A Factorial Experiment for the Connector Pull-off Force Problem (CD Only)

1-2.6 Observing Processes Over Time

Often data are collected over time. In this case, it is usually very helpful to plot the data versus time in a **time series plot.** Phenomena that might affect the system or process often become more visible in a time-oriented plot and the concept of stability can be better judged.

Figure 1-7 is a dot diagram of acetone concentration readings taken hourly from the distillation column described in Section 1-2.2. The large variation displayed on the dot diagram indicates a lot of variability in the concentration, but the chart does not help explain the reason for the variation. The time series plot is shown in Figure 1-8, on page 9. A shift in the process mean level is visible in the plot and an estimate of the time of the shift can be obtained.

W. Edwards **Deming,** a very influential industrial statistician, stressed that it is important to understand the nature of variability in processes and systems over time. He conducted an experiment in which he attempted to drop marbles as close as possible to a target on a table. He used a funnel mounted on a ring stand and the marbles were dropped into the funnel. See Fig. 1-9. The funnel was aligned as closely as possible with the center of the target. He then used two different strategies to operate the process. (1) He never moved the funnel. He just dropped one marble after another and recorded the distance from the target. (2) He dropped the first marble and recorded its location relative to the target. He then moved the funnel an equal and opposite distance in an attempt to compensate for the error. He continued to make this type of adjustment after each marble was dropped.

After both strategies were completed, he noticed that the variability of the distance from the target for strategy 2 was approximately 2 times larger than for strategy 1. The adjustments to the funnel increased the deviations from the target. The explanation is that the error (the deviation of the marble's position from the target) for one marble provides no information about the error that will occur for the next marble. Consequently, adjustments to the funnel do not decrease future errors. Instead, they tend to move the funnel farther from the target.

This interesting experiment points out that adjustments to a process based on random disturbances can actually *increase* the variation of the process. This is referred to as **overcontrol**

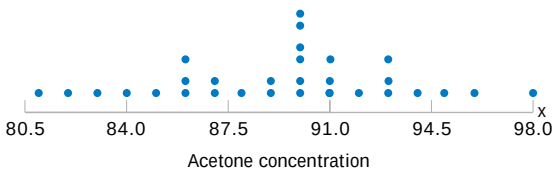


Figure 1-7 The dot diagram illustrates variation but does not identify the problem.

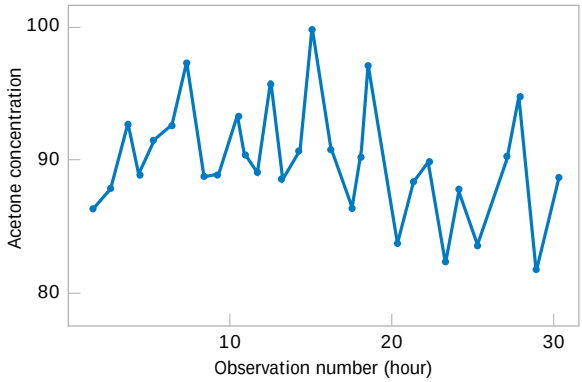


Figure 1-8 A time series plot of concentration provides more information than the dot diagram.

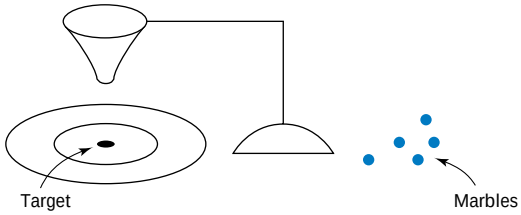


Figure 1-9 Deming's funnel experiment.

or **tampering.** Adjustments should be applied only to compensate for a nonrandom shift in the process—then they can help. A computer simulation can be used to demonstrate the lessons of the funnel experiment. Figure 1-10 displays a time plot of 100 measurements (denoted as y) from a process in which only random disturbances are present. The target value for the process is 10 units. The figure displays the data with and without adjustments that are applied to the process mean in an attempt to produce data closer to target. Each adjustment is equal and opposite to the deviation of the previous measurement from target. For example, when the measurement is 11 (one unit above target), the mean is reduced by one unit before the next measurement is generated. The overcontrol has increased the deviations from the target.

Figure 1-11 displays the data without adjustment from Fig. 1-10, except that the measurements after observation number 50 are increased by two units to simulate the effect of a shift in the mean of the process. When there is a true shift in the mean of a process, an adjustment can be useful. Figure 1-11 also displays the data obtained when one adjustment (a decrease of

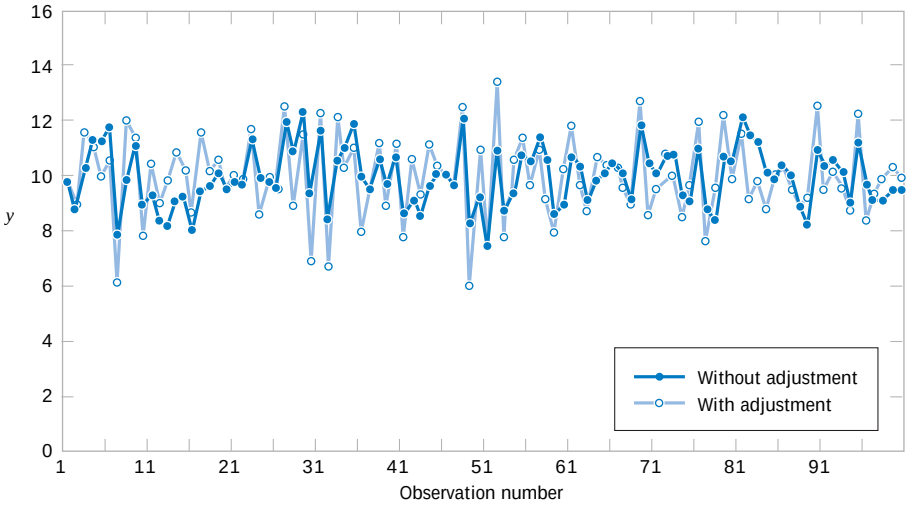


Figure 1-10 Adjustments applied to random disturbances overcontrol the process and increase the deviations from the target.

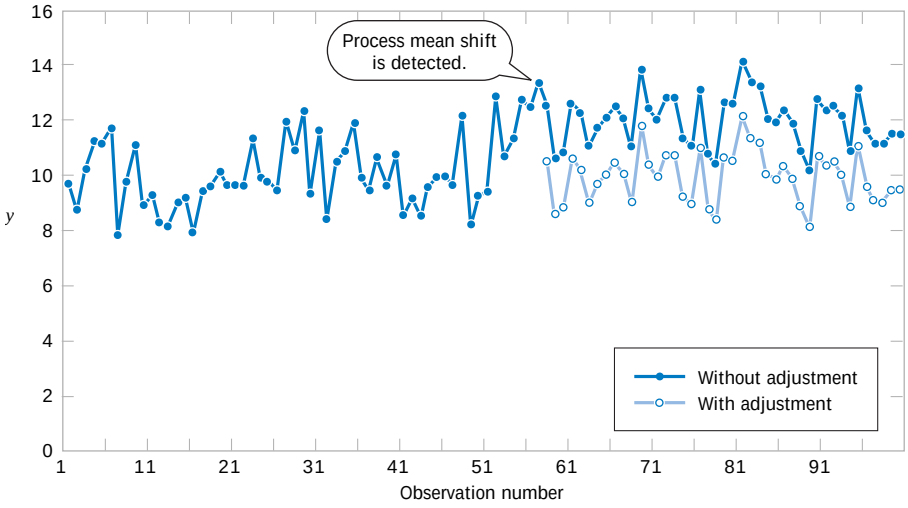


Figure 1-11 Process mean shift is detected at observation number 57, and one adjustment (a decrease of two units) reduces the deviations from target.

two units) is applied to the mean after the shift is detected (at observation number 57). Note that this adjustment decreases the deviations from target.

The question of when to apply adjustments (and by what amounts) begins with an understanding of the types of variation that affect a process. A **control chart** is an invaluable way to examine the variability in time-oriented data. Figure 1-12 presents a control chart for the concentration data from Fig. 1-8. The **center line** on the control chart is just the average of the concentration measurements for the first 20 samples ($\bar{x}$ □ 91.5 g□l) when the process is stable. The **upper control limit** and the **lower control limit** are a pair of statistically derived limits that reflect the inherent or natural variability in the process. These limits are located three standard deviations of the concentration values above and below the center line. If the process is operating as it should, without any external sources of variability present in the system, the concentration measurements should fluctuate randomly around the center line, and almost all of them should fall between the control limits.

In the control chart of Fig. 1-12, the visual frame of reference provided by the center line and the control limits indicates that some upset or disturbance has affected the process around sample 20 because all of the following observations are below the center line and two of them

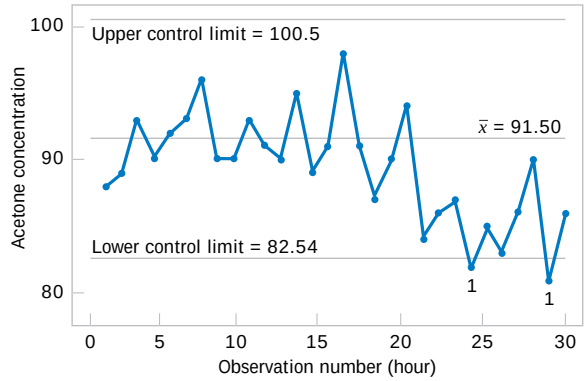


Figure 1-12 A control chart for the chemical process concentration data.

actually fall below the lower control limit. This is a very strong signal that corrective action is required in this process. If we can find and eliminate the underlying cause of this upset, we can improve process performance considerably.

Control charts are a very important application of statistics for monitoring, controlling, and improving a process. The branch of statistics that makes use of control charts is called **statistical process control,** or **SPC.** We will discuss SPC and control charts in Chapter 16.

1-3 MECHANISTIC AND EMPIRICAL MODELS

Models play an important role in the analysis of nearly all engineering problems. Much of the formal education of engineers involves learning about the models relevant to specific fields and the techniques for applying these models in problem formulation and solution. As a simple example, suppose we are measuring the flow of current in a thin copper wire. Our model for this phenomenon might be Ohm's law:

$$\text{Current} = \text{voltage}/\text{resistance}$$

or

$$I = E/R \tag{1-2}$$

We call this type of model a **mechanistic model** because it is built from our underlying knowledge of the basic physical mechanism that relates these variables. However, if we performed this measurement process more than once, perhaps at different times, or even on different days, the observed current could differ slightly because of small changes or variations in factors that are not completely controlled, such as changes in ambient temperature, fluctuations in performance of the gauge, small impurities present at different locations in the wire, and drifts in the voltage source. Consequently, a more realistic model of the observed current might be

$$I = E/R + \varepsilon \tag{1-3}$$

where ε is a term added to the model to account for the fact that the observed values of current flow do not perfectly conform to the mechanistic model. We can think of ε as a term that includes the effects of all of the unmodeled sources of variability that affect this system.

Sometimes engineers work with problems for which there is no simple or well-understood mechanistic model that explains the phenomenon. For instance, suppose we are interested in the number average molecular weight (M_n) of a polymer. Now we know that M_n is related to the viscosity of the material (V), and it also depends on the amount of catalyst (C) and the temperature (T) in the polymerization reactor when the material is manufactured. The relationship between M_n and these variables is

$$M_n = f(V, C, T) \tag{1-4}$$

say, where the *form* of the function f is unknown. Perhaps a working model could be developed from a first-order Taylor series expansion, which would produce a model of the form

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T \tag{1-5}$$

where the □'s are unknown parameters. Now just as in Ohm's law, this model will not exactly describe the phenomenon, so we should account for the other sources of variability that may affect the molecular weight by adding another term to the model; therefore

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T + \varepsilon \tag{1-6}$$

is the model that we will use to relate molecular weight to the other three variables. This type of model is called an **empirical model;** that is, it uses our engineering and scientific knowledge of the phenomenon, but it is not directly developed from our theoretical or first-principles understanding of the underlying mechanism.

To illustrate these ideas with a specific example, consider the data in Table 1-2. This table contains data on three variables that were collected in an observational study in a semiconductor manufacturing plant. In this plant, the finished semiconductor is wire bonded to a frame. The variables reported are pull strength (a measure of the amount of force required to break the bond), the wire length, and the height of the die. We would like to find a model relating pull strength to wire length and die height. Unfortunately, there is no physical mechanism that we can easily apply here, so it doesn't seem likely that a mechanistic modeling approach will be successful.

Table 1-2 Wire Bond Pull Strength Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

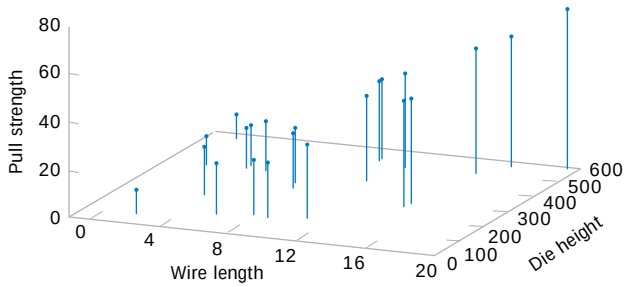


Figure 1-13 Three-dimensional plot of the wire and pull strength data.

Figure 1-13 presents a three-dimensional plot of all 25 observations on pull strength, wire length, and die height. From examination of this plot, we see that pull strength increases as both wire length and die height increase. Furthermore, it seems reasonable to think that a model such as

$$\text{Pull strength} = \beta_0 + \beta_1(\text{wire length}) + \beta_2(\text{die height}) + \varepsilon$$

would be appropriate as an empirical model for this relationship. In general, this type of empirical model is called a **regression model.** In Chapters 11 and 12 we show how to build these models and test their adequacy as approximating functions. We will use a method for estimating the parameters in regression models, called the method of **least squares,** that traces its origins to work by Karl Gauss. Essentially, this method chooses the parameters in the empirical model (the β's) to minimize the sum of the squared distances between each data point and the plane represented by the model equation. Applying this technique to the data in Table 1-2 results in

$$\widehat{\text{Pull strength}} = 2.26 + 2.74(\text{wire length}) + 0.0125(\text{die height}) \tag{1-7}$$

where the "hat," or circumflex, over pull strength indicates that this is an estimated or predicted quantity.

Figure 1-14 is a plot of the predicted values of pull strength versus wire length and die height obtained from Equation 1-7. Notice that the predicted values lie on a plane above the wire length–die height space. From the plot of the data in Fig. 1-13, this model does not appear unreasonable. The empirical model in Equation 1-7 could be used to predict values of pull strength for various combinations of wire length and die height that are of interest. Essentially, the empirical model could be used by an engineer in exactly the same way that a mechanistic model can be used.

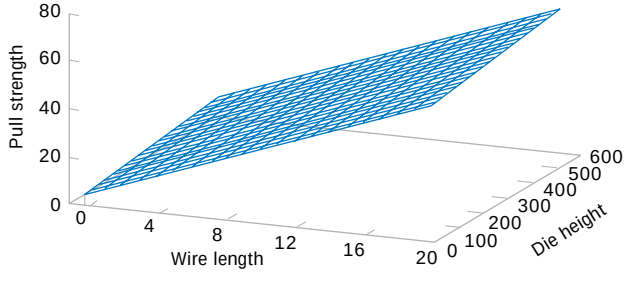


Figure 1-14 Plot of predicted values of pull strength from the empirical model.

1-4 PROBABILITY AND PROBABILITY MODELS

In Section 1-1, it was mentioned that decisions often need to be based on measurements from only a subset of objects selected in a sample. This process of reasoning from a sample of objects to conclusions for a population of objects was referred to as statistical inference. A sample of three wafers selected from a larger production lot of wafers in semiconductor manufacturing was an example mentioned. To make good decisions, an analysis of how well a sample represents a population is clearly necessary. If the lot contains defective wafers, how well will the sample detect this? How can we quantify the criterion to "detect well"? Basically, how can we quantify the risks of decisions based on samples? Furthermore, how should samples be selected to provide good decisions—ones with acceptable risks? **Probability** models help quantify the risks involved in statistical inference, that is, the risks involved in decisions made every day.

More details are useful to describe the role of probability models. Suppose a production lot contains 25 wafers. If all the wafers are defective or all are good, clearly any sample will generate all defective or all good wafers, respectively. However, suppose only one wafer in the lot is defective. Then a sample might or might not detect (include) the wafer. A probability model, along with a method to select the sample, can be used to quantify the risks that the defective wafer is or is not detected. Based on this analysis, the size of the sample might be increased (or decreased). The risk here can be interpreted as follows. Suppose a series of lots, each with exactly one defective wafer, are sampled. The details of the method used to select the sample are postponed until randomness is discussed in the next chapter. Nevertheless, assume that the same size sample (such as three wafers) is selected in the same manner from each lot. The proportion of the lots in which the defective wafer is included in the sample or, more specifically, the limit of this proportion as the number of lots in the series tends to infinity, is interpreted as the probability that the defective wafer is detected.

A probability model is used to calculate this proportion under reasonable assumptions for the manner in which the sample is selected. This is fortunate because we do not want to attempt to sample from an infinite series of lots. Problems of this type are worked in Chapters 2 and 3. More importantly, this probability provides valuable, quantitative information regarding any decision about lot quality based on the sample.

Recall from Section 1-1 that a population might be conceptual, as in an analytic study that applies statistical inference to future production based on the data from current production. When populations are extended in this manner, the role of statistical inference and the associated probability models becomes even more important.

In the previous example, each wafer in the sample was only classified as defective or not. Instead, a continuous measurement might be obtained from each wafer. In Section 1-2.6, concentration measurements were taken at periodic intervals from a production process. Figure 1-7 shows that variability is present in the measurements, and there might be concern that the process has moved from the target setting for concentration. Similar to the defective wafer, one might want to quantify our ability to detect a process change based on the sample data. Control limits were mentioned in Section 1-2.6 as decision rules for whether or not to adjust a process. The probability that a particular process change is detected can be calculated with a probability model for concentration measurements. Models for continous measurements are developed based on plausible assumptions for the data and a result known as the central limit theorem, and the associated normal distribution is a particularly valuable probability model for statistical inference. Of course, a check of assumptions is important. These types of probability models are discussed in Chapter 4. The objective is still to quantify the risks inherent in the inference made from the sample data.

Throughout Chapters 6 through 15, decisions are based statistical inference from sample data. Continuous probability models, specifically the normal distribution, are used extensively to quantify the risks in these decisions and to evaluate ways to collect the data and how large a sample should be selected.

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Analytic study
Designed experiment
Empirical model
Engineering method
Enumerative study
Mechanistic model
Observational study
Overcontrol
Population
Probability model
Problem-solving method
Retrospective study
Sample
Statistical inference
Statistical Process Control
Statistical thinking
Tampering
Variability

CD MATERIAL

Factorial Experiment
Fractional factorial experiment
Interaction

1-2.5 A Factorial Experiment for the Connector Pull-off Force Problem (CD only)

Much of what we know in the engineering and physical-chemical sciences is developed through testing or experimentation. Often engineers work in problem areas in which no scientific or engineering theory is directly or completely applicable, so experimentation and observation of the resulting data constitute the only way that the problem can be solved. Even when there is a good underlying scientific theory that we may rely on to explain the phenomena of interest, it is almost always necessary to conduct tests or experiments to confirm that the theory is indeed operative in the situation or environment in which it is being applied. We have observed that statistical thinking and statistical methods play an important role in planning, conducting, and analyzing the data from engineering experiments.

To further illustrate the factorial design concept introduced in Section 1-2.4, suppose that in the connector wall thickness example, there are two additional factors of interest, time and temperature. The cure times of interest are 1 and 24 hours and the temperature levels are 70°F and 100°F. Now since all three factors have two levels, a factorial experiment would consist of the eight test combinations shown at the corners of the cube in Fig. S1-1. Two trials, or **replicates,** would be performed at each corner, resulting in a 16-run factorial experiment. The observed values of pull-off force are shown in parentheses at the cube corners in Fig. S1-1. Notice that this experiment uses eight 3□32-inch prototypes and eight 1□8-inch prototypes, the same number used in the simple comparative study in Section 1-1, but we are now investigating *three* factors. Generally, factorial experiments are the most efficient way to study the joint effects of several factors.

Some very interesting tentative conclusions can be drawn from this experiment. First, compare the average pull-off force of the eight 3□32-inch prototypes with the average pull-off force of the eight 1□8-inch prototypes (these are the averages of the eight runs on the left face and right face of the cube in Fig. S1-1, respectively), or 14.1 □ 13.45 □ 0.65. Thus, increasing the wall thickness from 3□32 to 1□8-inch increases the average pull-off force by 0.65 pounds. Next, to measure the effect of increasing the cure time, compare the average of the eight runs in the back face of the cube (where time □ 24 hours) with the average of the eight runs in the front face (where time □ 1 hour), or 14.275 □ 13.275 □ 1. The effect of increasing the cure time from 1 to 24 hours is to increase the average pull-off force by 1 pound; that is, cure time apparently has an effect that is larger than the effect of increasing the wall

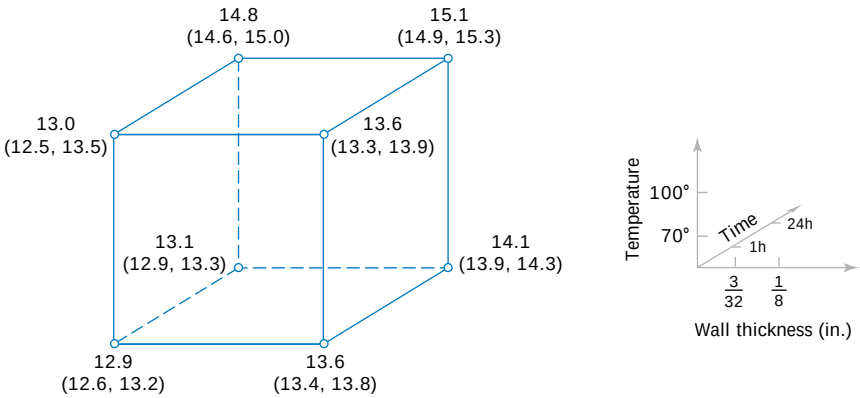


Figure S1-1 The factorial experiment for the connector wall thickness problem.

Time	Temp.	Avg. Force
1 h	70°F	13.25
1 h	100°F	13.30
24 h	70°F	13.60
24 h	100°F	14.95

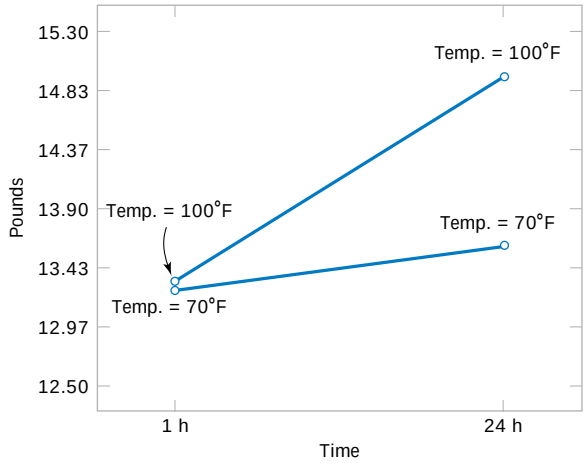


Figure S1-2 The two-factor interaction between cure time and cure temperature.

thickness. The cure temperature effect can be evaluated by comparing the average of the eight runs in the top of the cube (where temperature □ 100°F) with the average of the eight runs in the bottom (where temperature □ 70°F), or 14.125 □ 13.425 □ 0.7. Thus, the effect of increasing the cure temperature is to increase the average pull-off force by 0.7 pounds. Thus, if the engineer's objective is to design a connector with high pull-off force, there are apparently several alternatives, such as increasing the wall thickness and using the "standard" curing conditions of 1 hour and 70°F or using the original 3□32-inch wall thickness but specifying a longer cure time and higher cure temperature.

There is an interesting relationship between cure time and cure temperature that can be seen by examination of the graph in Fig. S1-2. This graph was constructed by calculating the average pull-off force at the four different combinations of time and temperature, plotting these averages versus time and then connecting the points representing the two temperature levels with straight lines. The slope of each of these straight lines represents the effect of cure time on pull-off force. Notice that the slopes of these two lines do not appear to be the same, indicating that the cure time effect is *different* at the two values of cure temperature. This is an example of an **interaction** between two factors. The interpretation of this interaction is very straightforward; if the standard cure time (1 hour) is used, cure temperature has little effect, but if the longer cure time (24 hours) is used, increasing the cure temperature has a large effect on average pull-off force. Interactions occur often in physical and chemical systems, and factorial experiments are the only way to investigate their effects. In fact, if interactions are present and the factorial experimental strategy is not used, incorrect or misleading results may be obtained.

We can easily extend the factorial strategy to more factors. Suppose that the engineer wants to consider a fourth factor, type of adhesive. There are two types: the standard adhesive and a new competitor. Figure S1-3 illustrates how all four factors, wall thickness, cure time, cure temperature, and type of adhesive, could be investigated in a factorial design. Since all four factors are still at two levels, the experimental design can still be represented geometrically as a cube (actually, it's a *hypercube*). Notice that as in any factorial design, all possible combinations of the four factors are tested. The experiment requires 16 trials.

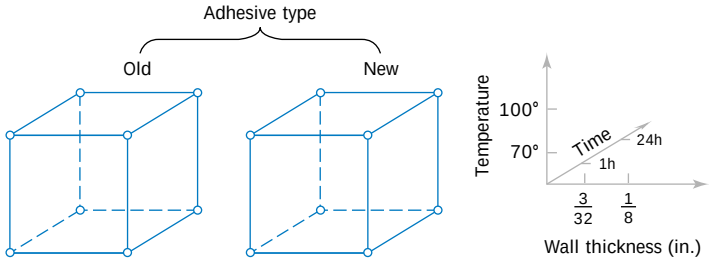


Figure S1-3 A four-factorial experiment for the connector wall thickness problem.

Generally, if there are k factors and they each have two levels, a factorial experimental design will require 2^k runs. For example, with $k = 4$, the 2^4 design in Fig. S1-3 requires 16 tests. Clearly, as the number of factors increases, the number of trials required in a factorial experiment increases rapidly; for instance, eight factors each at two levels would require 256 trials. This quickly becomes unfeasible from the viewpoint of time and other resources. Fortunately, when there are four to five or more factors, it is usually unnecessary to test all possible combinations of factor levels. A **fractional factorial experiment** is a variation of the basic factorial arrangement in which only a subset of the factor combinations are actually tested. Figure S1-4 shows a fractional factorial experimental design for the four-factor version of the connector experiment. The circled test combinations in this figure are the only test combinations that need to be run. This experimental design requires only 8 runs instead of the original 16; consequently it would be called a **one-half fraction.** This is an excellent experimental design in which to study all four factors. It will provide good information about the individual effects of the four factors and some information about how these factors interact.

Factorial and fractional factorial experiments are used extensively by engineers and scientists in industrial research and development, where new technology, products, and processes are designed and developed and where existing products and processes are improved. Since so much engineering work involves testing and experimentation, it is essential that all engineers understand the basic principles of planning efficient and effective experiments. We discuss these principles in Chapter 13. Chapter 14 concentrates on the factorial and fractional factorials that we have introduced here.

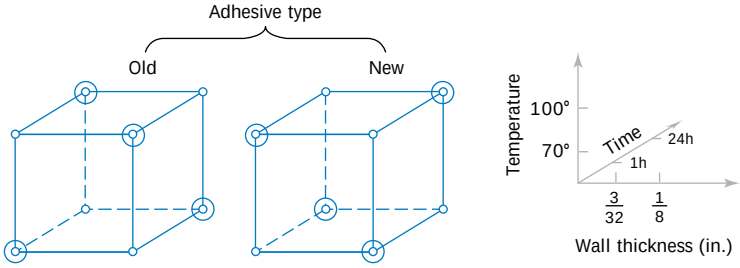


Figure S1-4 A fractional factorial experiment for the connector wall thickness problem.

2 Probability

CHAPTER OUTLINE

2-1 SAMPLE SPACES AND EVENTS
- 2-1.1 Random Experiments
- 2-1.2 Sample Spaces
- 2-1.3 Events
- 2-1.4 Counting Techniques (CD Only)

2-2 INTERPRETATIONS OF PROBABILITY
- 2-2.1 Introduction
- 2-2.2 Axioms of Probability

2-3 ADDITION RULES

2-4 CONDITIONAL PROBABILITY

2-5 MULTIPLICATION AND TOTAL PROBABILITY RULES
- 2-5.1 Multiplication Rule
- 2-5.2 Total Probability Rule

2-6 INDEPENDENCE

2-7 BAYES' THEOREM

2-8 RANDOM VARIABLES

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Understand and describe sample spaces and events for random experiments with graphs, tables, lists, or tree diagrams
2. Interpret probabilities and use probabilities of outcomes to calculate probabilities of events in discrete sample spaces
3. Calculate the probabilities of joint events such as unions and intersections from the probabilities of individual events
4. Interpret and calculate conditional probabilities of events
5. Determine the independence of events and use independence to calculate probabilities
6. Use Bayes' theorem to calculate conditional probabilities
7. Understand random variables

CD MATERIAL

8. Use permutation and combinations to count the number of outcomes in both an event and the sample space.

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

2-1 SAMPLE SPACES AND EVENTS

2-1.1 Random Experiments

If we measure the current in a thin copper wire, we are conducting an experiment. However, in day-to-day repetitions of the measurement the results can differ slightly because of small variations in variables that are not controlled in our experiment, including changes in ambient temperatures, slight variations in gauge and small impurities in the chemical composition of the wire if different locations are selected, and current source drifts. Consequently, this experiment (as well as many we conduct) is said to have a **random** component. In some cases, the random variations, are small enough, relative to our experimental goals, that they can be ignored. However, no matter how carefully our experiment is designed and conducted, the variation is almost always present, and its magnitude can be large enough that the important conclusions from our experiment are not obvious. In these cases, the methods presented in this book for modeling and analyzing experimental results are quite valuable.

Our goal is to understand, quantify, and model the type of variations that we often encounter. When we incorporate the variation into our thinking and analyses, we can make informed judgments from our results that are not invalidated by the variation.

Models and analyses that include variation are not different from models used in other areas of engineering and science. Figure 2-1 displays the important components. A mathematical model (or abstraction) of the physical system is developed. It need not be a perfect abstraction. For example, Newton's laws are not perfect descriptions of our physical universe. Still, they are useful models that can be studied and analyzed to approximately quantify the performance of a wide range of engineered products. Given a mathematical abstraction that is validated with measurements from our system, we can use the model to understand, describe, and quantify important aspects of the physical system and predict the response of the system to inputs.

Throughout this text, we discuss models that allow for variations in the outputs of a system, even though the variables that we control are not purposely changed during our study. Figure 2-2 graphically displays a model that incorporates uncontrollable inputs (noise) that combine with the controllable inputs to produce the output of our system. Because of the

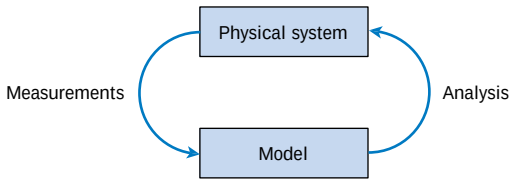


Figure 2-1 Continuous iteration between model and physical system.

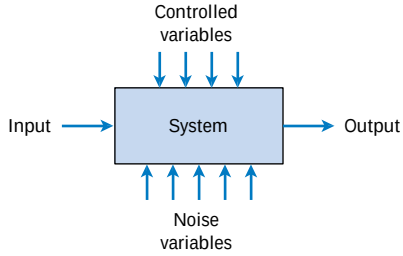


Figure 2-2 Noise variables affect the transformation of inputs to outputs.

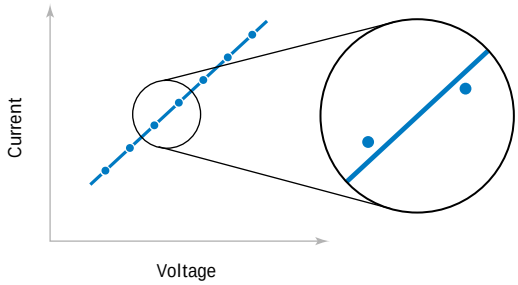


Figure 2-3 A closer examination of the system identifies deviations from the model.

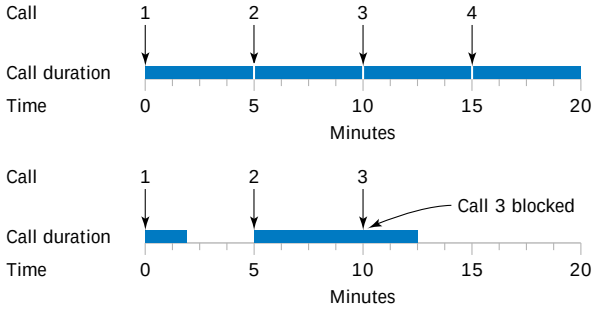


Figure 2-4 Variation causes disruptions in the system.

uncontrollable inputs, the same settings for the controllable inputs do not result in identical outputs every time the system is measured.

Definition

An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a **random experiment.**

For the example of measuring current in a copper wire, our model for the system might simply be Ohm's law. Because of uncontrollable inputs, variations in measurements of current are expected. Ohm's law might be a suitable approximation. However, if the variations are large relative to the intended use of the device under study, we might need to extend our model to include the variation. See Fig. 2-3.

As another example, in the design of a communication system, such as a computer or voice communication network, the information capacity available to service individuals using the network is an important design consideration. For voice communication, sufficient external lines need to be purchased from the phone company to meet the requirements of a business. Assuming each line can carry only a single conversation, how many lines should be purchased? If too few lines are purchased, calls can be delayed or lost. The purchase of too many lines increases costs. Increasingly, design and product development is required to meet customer requirements *at a competitive cost.*

In the design of the voice communication system, a model is needed for the number of calls and the duration of calls. Even knowing that on average, calls occur every five minutes and that they last five minutes is not sufficient. If calls arrived precisely at five-minute intervals and lasted for precisely five minutes, one phone line would be sufficient. However, the slightest variation in call number or duration would result in some calls being blocked by others. See Fig. 2-4. A system designed without considering variation will be woefully inadequate for practical use. Our model for the number and duration of calls needs to include variation as an integral component. An analysis of models including variation is important for the design of the phone system.

2-1.2 Sample Spaces

To model and analyze a random experiment, we must understand the set of possible **outcomes** from the experiment. In this introduction to probability, we make use of the basic

concepts of sets and operations on sets. It is assumed that the reader is familiar with these topics.

Definition

The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. The sample space is denoted as S.

A sample space is often defined based on the objectives of the analysis.

EXAMPLE 2-1

Consider an experiment in which you select a molded plastic part, such as a connector, and measure its thickness. The possible values for thickness depend on the resolution of the measuring instrument, and they also depend on upper and lower bounds for thickness. However, it might be convenient to define the sample space as simply the positive real line

$$S = R^1 = \{x \mid x > 0\}$$

because a negative value for thickness cannot occur.

If it is known that all connectors will be between 10 and 11 millimeters thick, the sample space could be

$$S = \{x \mid 10 < x < 11\}$$

If the objective of the analysis is to consider only whether a particular part is low, medium, or high for thickness, the sample space might be taken to be the set of three outcomes:

$$S = \{low, medium, high\}$$

If the objective of the analysis is to consider only whether or not a particular part conforms to the manufacturing specifications, the sample space might be simplified to the set of two outcomes

$$S = \{yes, no\}$$

that indicate whether or not the part conforms.

It is useful to distinguish between two types of sample spaces.

Definition

A sample space is **discrete** if it consists of a finite or countable infinite set of outcomes. A sample space is **continuous** if it contains an interval (either finite or infinite) of real numbers.

In Example 2-1, the choice $S = R^1$ is an example of a continuous sample space, whereas $S = \{yes, no\}$ is a discrete sample space. As mentioned, the best choice of a sample space

depends on the objectives of the study. As specific questions occur later in the book, appropriate sample spaces are discussed.

EXAMPLE 2-2

If two connectors are selected and measured, the extension of the positive real line R is to take the sample space to be the positive quadrant of the plane:

$$S = R^2 = R \times R$$

If the objective of the analysis is to consider only whether or not the parts conform to the manufacturing specifications, either part may or may not conform. We abbreviate *yes* and *no* as *y* and *n*. If the ordered pair *yn* indicates that the first connector conforms and the second does not, the sample space can be represented by the four outcomes:

$$S = \{yy, yn, ny, nn\}$$

If we are only interested in the number of conforming parts in the sample, we might summarize the sample space as

$$S = \{0, 1, 2\}$$

As another example, consider an experiment in which the thickness is measured until a connector fails to meet the specifications. The sample space can be represented as

$$S = \{n, yn, yyn, yyyn, yyyyn, \text{and so forth}\}$$

In random experiments in which items are selected from a batch, we will indicate whether or not a selected item is replaced before the next one is selected. For example, if the batch consists of three items $\{a, b, c\}$ and our experiment is to select two items **without replacement,** the sample space can be represented as

$$S_{\text{without}} = \{ab, ac, ba, bc, ca, cb\}$$

This description of the sample space maintains the order of the items selected so that the outcome *ab* and *ba* are separate elements in the sample space. A sample space with less detail only describes the two items selected $\{\{a, b\}, \{a, c\}, \{b, c\}\}$. This sample space is the possible subsets of two items. Sometimes the ordered outcomes are needed, but in other cases the simpler, unordered sample space is sufficient.

If items are replaced before the next one is selected, the sampling is referred to as **with replacement.** Then the possible ordered outcomes are

$$S_{\text{with}} = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$$

The unordered description of the sample space is $\{\{a, a\}, \{a, b\}, \{a, c\}, \{b, b\}, \{b, c\}, \{c, c\}\}$. Sampling without replacement is more common for industrial applications.

Sometimes it is not necessary to specify the exact item selected, but only a property of the item. For example, suppose that there are 5 defective parts and 95 good parts in a batch. To study the quality of the batch, two are selected without replacement. Let *g* denote a good part and *d* denote a defective part. It might be sufficient to describe the sample space (ordered) in terms of quality of each part selected as

$$S = \{gg, gd, dg, dd\}$$

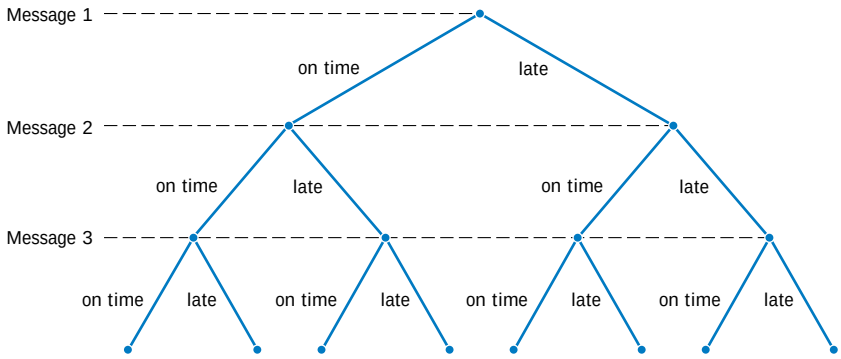


Figure 2-5 Tree diagram for three messages.

One must be cautious with this description of the sample space because there are many more pairs of items in which both are good than pairs in which both are defective. These differences must be accounted for when probabilities are computed later in this chapter. Still, this summary of the sample space will be convenient when conditional probabilities are used later in this chapter. Also, if there were only one defective part in the batch, there would be fewer possible outcomes

$$S = \{gg, gd, dg\}$$

because dd would be impossible. For sampling questions, sometimes the most important part of the solution is an appropriate description of the sample space.

Sample spaces can also be described graphically with **tree diagrams.** When a sample space can be constructed in several steps or stages, we can represent each of the n_1 ways of completing the first step as a branch of a tree. Each of the ways of completing the second step can be represented as n_2 branches starting from the ends of the original branches, and so forth.

EXAMPLE 2-3

Each message in a digital communication system is classified as to whether it is received within the time specified by the system design. If three messages are classified, use a tree diagram to represent the sample space of possible outcomes.

Each message can either be received on time or late. The possible results for three messages can be displayed by eight branches in the tree diagram shown in Fig. 2-5.

EXAMPLE 2-4

An automobile manufacturer provides vehicles equipped with selected options. Each vehicle is ordered

- With or without an automatic transmission
- With or without air-conditioning
- With one of three choices of a stereo system
- With one of four exterior colors

If the sample space consists of the set of all possible vehicle types, what is the number of outcomes in the sample space? The sample space contains 48 outcomes. The tree diagram for the different types of vehicles is displayed in Fig. 2-6.

EXAMPLE 2-5

Consider an extension of the automobile manufacturer illustration in the previous example in which another vehicle option is the interior color. There are four choices of interior color: red, black, blue, or brown. However,

- With a red exterior, only a black or red interior can be chosen.
- With a white exterior, any interior color can be chosen.

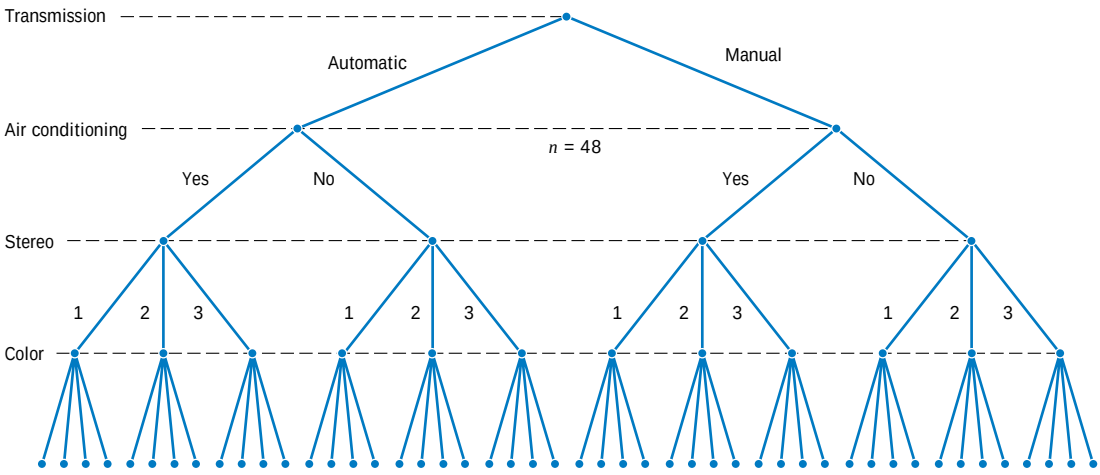


Figure 2-6 Tree diagram for different types of vehicles.

With a blue exterior, only a black, red, or blue interior can be chosen.

With a brown exterior, only a brown interior can be chosen.

In Fig. 2-6, there are 12 vehicle types with each exterior color, but the number of interior color choices depends on the exterior color. As shown in Fig. 2-7, the tree diagram can be extended to show that there are 120 different vehicle types in the sample space.

2-1.3 Events

Often we are interested in a collection of related outcomes from a random experiment.

Definition

An **event** is a subset of the sample space of a random experiment.

We can also be interested in describing new events from combinations of existing events. Because events are subsets, we can use basic set operations such as unions, intersections, and

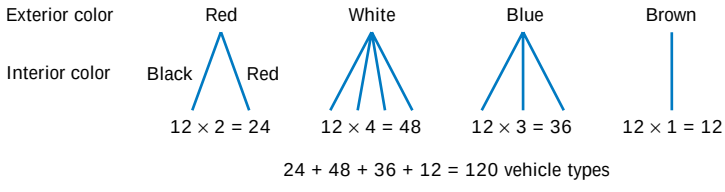


Figure 2-7 Tree diagram for different types of vehicles with interior colors.

complements to form other events of interest. Some of the basic set operations are summarized below in terms of events:

- The **union** of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$.
- The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.
- The **complement** of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the component of the event E as E'.

EXAMPLE 2-6

Consider the sample space $S = \{yy, yn, ny, nn\}$ in Example 2-2. Suppose that the set of all outcomes for which at least one part conforms is denoted as E_1. Then,

$$E_1 = \{yy, yn, ny\}$$

The event in which both parts do not conform, denoted as E_2, contains only the single outcome, $E_2 = \{nn\}$. Other examples of events are $E_3 = \emptyset$, the null set, and $E_4 = S$, the sample space. If $E_5 = \{yn, ny, nn\}$,

$$E_1 \cup E_5 = S \qquad E_1 \cap E_5 = \{yn, ny\} \qquad E_1' = \{nn\}$$

EXAMPLE 2-7

Measurements of the time needed to complete a chemical reaction might be modeled with the sample space $S = R^+$, the set of positive real numbers. Let

$$E_1 = \{x \mid 1 < x < 10\} \quad \text{and} \quad E_2 = \{x \mid 3 < x < 11\}$$

Then,

$$E_1 \cup E_2 = \{x \mid 1 < x < 11\} \quad \text{and} \quad E_1 \cap E_2 = \{x \mid 3 < x < 10\}$$

Also,

$$E_1' = \{x \mid x > 10\} \quad \text{and} \quad E_1' \cap E_2 = \{x \mid 10 < x < 11\}$$

EXAMPLE 2-8

Samples of polycarbonate plastic are analyzed for scratch and shock resistance. The results from 50 samples are summarized as follows:

		shock resistance	
		high	low
scratch resistance	high	40	4
	low	1	5

Let A denote the event that a sample has high shock resistance, and let B denote the event that a sample has high scratch resistance. Determine the number of samples in $A \cap B$, A', and $A \cup B$.

The event $A \cap B$ consists of the 40 samples for which scratch and shock resistances are high. The event A' consists of the 9 samples in which the shock resistance is low. The event $A \cup B$ consists of the 45 samples in which the shock resistance, scratch resistance, or both are high.

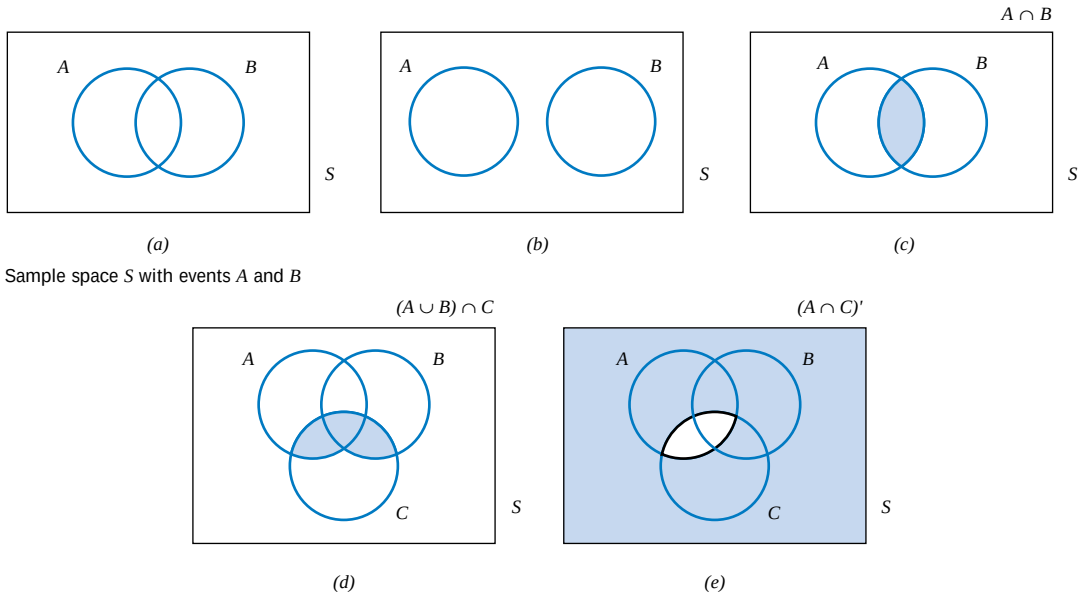


Figure 2-8 Venn diagrams.

Diagrams are often used to portray relationships between sets, and these diagrams are also used to describe relationships between events. We can use **Venn diagrams** to represent a sample space and events in a sample space. For example, in Fig. 2-8(a) the sample space of the random experiment is represented as the points in the rectangle S. The events A and B are the subsets of points in the indicated regions. Figure 2-8(b) illustrates two events with no common outcomes; Figs. 2-8(c) to 2-8(e) illustrate additional joint events.

Two events with no outcomes in common have an important relationship.

Definition

Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \emptyset$$

are said to be **mutually exclusive.**

The two events in Fig. 2-8(b) are mutually exclusive, whereas the two events in Fig. 2-8(a) are not.

Additional results involving events are summarized below. The definition of the complement of an event implies that

$$(E')' = E$$

The distributive law for set operations implies that

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C), \quad \text{and} \quad (A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

DeMorgan's laws imply that

$$(A \cup B)' = A' \cap B' \quad \text{and} \quad (A \cap B)' = A' \cup B'$$

Also, remember that

$$A \cap B = B \cap A \quad \text{and} \quad A \cup B = B \cup A$$

2-1.4 Counting Techniques (CD Only)

As sample spaces become larger, complete enumeration is difficult. Instead, counts of the number outcomes in the sample space and in various events are often used to analyze the random experiment. These methods are referred to as **counting techniques** and described on the CD.

EXERCISES FOR SECTION 2-1

Provide a reasonable description of the sample space for each of the random experiments in Exercises 2-1 to 2-18. There can be more than one acceptable interpretation of each experiment. Describe any assumptions you make.

2-1. Each of three machined parts is classified as either above or below the target specification for the part.

2-2. Each of four transmitted bits is classified as either in error or not in error.

2-3. In the final inspection of electronic power supplies, three types of nonconformities might occur: functional, minor, or cosmetic. Power supplies that are defective are further classified as to type of nonconformity.

2-4. In the manufacturing of digital recording tape, electronic testing is used to record the number of bits in error in a 350-foot reel.

2-5. In the manufacturing of digital recording tape, each of 24 tracks is classified as containing or not containing one or more bits in error.

2-6. An ammeter that displays three digits is used to measure current in milliamperes.

2-7. A scale that displays two decimal places is used to measure material feeds in a chemical plant in tons.

2-8. The following two questions appear on an employee survey questionnaire. Each answer is chosen from the five-point scale 1 (never), 2, 3, 4, 5 (always).

Is the corporation willing to listen to and fairly evaluate new ideas?

How often are my coworkers important in my overall job performance?

2-9. The concentration of ozone to the nearest part per billion.

2-10. The time until a tranaction service is requested of a computer to the nearest millisecond.

2-11. The pH reading of a water sample to the nearest tenth of a unit.

2-12. The voids in a ferrite slab are classified as small, medium, or large. The number of voids in each category is measured by an optical inspection of a sample.

2-13. The time of a chemical reaction is recorded to the nearest millisecond.

2-14. An order for an automobile can specify either an automatic or a standard transmission, either with or without air-conditioning, and any one of the four colors red, blue, black or white. Describe the set of possible orders for this experiment.

2-15. A sampled injection-molded part could have been produced in either one of two presses and in any one of the eight cavities in each press.

2-16. An order for a computer system can specify memory of 4, 8, or 12 gigabytes, and disk storage of 200, 300, or 400 gigabytes. Describe the set of possible orders.

2-17. Calls are repeatedly placed to a busy phone line until a connect is achieved.

2-18. In a magnetic storage device, three attempts are made to read data before an error recovery procedure that repositions the magnetic head is used. The error recovery procedure attempts three repositionings before an "abort" message is sent to the operator. Let

s denote the success of a read operation

f denote the failure of a read operation

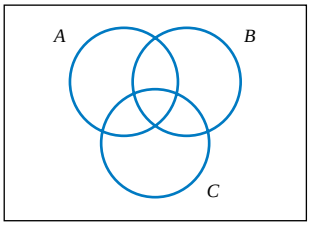
F denote the failure of an error recovery procedure

S denote the success of an error recovery procedure

A denote an abort message sent to the operator.

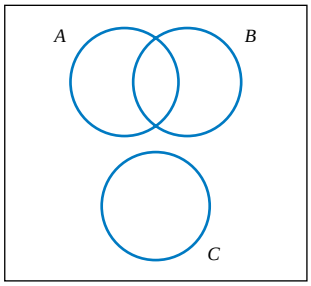
Describe the sample space of this experiment with a tree diagram.

2-19. Three events are shown on the Venn diagram in the following figure:



Reproduce the figure and shade the region that corresponds to each of the following events.
(a) A' (b) $A \cup B$
(c) $(A \cup B)' \cap C$ (d) $(B \cap C)'$
(e) $(A \cup B)' \cap C$

2-20. Three events are shown on the Venn diagram in the following figure:



Reproduce the figure and shade the region that corresponds to each of the following events.
(a) A' (b) $(A \cup B)' \cap (A \cup B)'$
(c) $(A \cup B)' \cap C$ (d) $(B \cap C)'$
(e) $(A \cup B)' \cap C$

2-21. A digital scale is used that provides weights to the nearest gram.
(a) What is the sample space for this experiment?
Let A denote the event that a weight exceeds 11 grams, let B denote the event that a weight is less than or equal to 15 grams, and let C denote the event that a weight is greater than or equal to 8 grams and less than 12 grams.
Describe the following events.
(b) $A \cap B$ (c) $A \cup B$
(d) A' (e) $A \cap B \cap C$
(f) $(A \cap C)'$ (g) $A \cup B \cup C$
(h) $B' \cup C$ (i) $A \cap (B \cup C)$

2-22. In an injection-molding operation, the length and width, denoted as X and Y, respectively, of each molded part are evaluated. Let

A denote the event of $48 \le X \le 52$ centimeters

B denote the event of $9 \le Y \le 11$ centimeters

C denote the event that a critical length meets customer requirements.

Construct a Venn diagram that includes these events. Shade the areas that represent the following:
(a) A (b) $A \cup B$
(c) $A' \cap B$ (d) $A \cap B$
(e) If these events were mutually exclusive, how successful would this production operation be? Would the process produce parts with $X \le 50$ centimeters and $Y \le 10$ centimeters?

2-23. Four bits are transmitted over a digital communications channel. Each bit is either distorted or received without distortion. Let A_i denote the event that the ith bit is distorted, $i = 1, \ldots, 4$.
(a) Describe the sample space for this experiment.
(b) Are the A_i's mutually exclusive?
Describe the outcomes in each of the following events:
(c) A_1 (d) A_1'
(e) $A_1 \cup A_2 \cup A_3 \cup A_4$ (f) $(A_1 \cup A_2) \cap (A_3 \cup A_4)$

2-24. A sample of three calculators is selected from a manufacturing line, and each calculator is classified as either defective or acceptable. Let A, B, and C denote the events that the first, second, and third calculators respectively, are defective.
(a) Describe the sample space for this experiment with a tree diagram.
Use the tree diagram to describe each of the following events:
(b) A (c) B
(d) $A \cup B$ (e) $B \cap C$

2-25. A wireless garage door opener has a code determined by the up or down setting of 12 switches. How many outcomes are in the sample space of possible codes?

2-26. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized below:

		shock resistance	
		high	low
scratch	high	70	9
resistance	low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch

resistance. Determine the number of disks in $A \cup B$, A', and $A \cap B$.

2-27. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		edge finish	
		excellent	good
surface finish	excellent	80	2
	good	10	8

(a) Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent edge finish. Determine the number of samples in $A' \cup B$, B', and $A \cap B$.

(b) Assume that each of two samples is to be classified on the basis of surface finish, either excellent or good, edge finish, either excellent or good. Use a tree diagram to represent the possible outcomes of this experiment.

2-28. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		conforms	
		yes	no
supplier	1	22	8
	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. Determine the number of samples in $A' \cup B$, B', and $A \cap B$.

2-29. The rise time of a reactor is measured in minutes (and fractions of minutes). Let the sample space be positive, real numbers. Define the events A and B as follows:

$$A = \{x \mid x > 72.5\}$$

and

$$B = \{x \mid x < 52.5\}$$

Describe each of the following events:

(a) A' (b) B'

(c) $A \cup B$ (d) $A \cap B$

2-30. A sample of two items is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

(a) The batch contains the items $\{a, b, c, d\}$.

(b) The batch contains the items $\{a, b, c, d, e, f, g\}$.

(c) The batch contains 4 defective items and 20 good items.

(d) The batch contains 1 defective item and 20 good items.

2-31. A sample of two printed circuit boards is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

(a) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 2 boards with major defects.

(b) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 1 board with major defects.

2-32. Counts of the Web pages provided by each of two computer servers in a selected hour of the day are recorded. Let A denote the event that at least 10 pages are provided by server 1 and let B denote the event that at least 20 pages are provided by server 2.

(a) Describe the sample space for the numbers of pages for two servers graphically.

Show each of the following events on the sample space graph:

(b) A (c) B

(d) $A \cup B$ (e) $A \cap B$

2-33. The rise time of a reactor is measured in minutes (and fractions of minutes). Let the sample space for the rise time of each batch be positive, real numbers. Consider the rise times of *two* batches. Let A denote the event that the rise time of batch 1 is less than 72.5 minutes, and let B denote the event that the rise time of batch 2 is greater than 52.5 minutes.

Describe the sample space for the rise time of two batches graphically and show each of the following events on a two-dimensional plot:

(a) A (b) B'

(c) $A \cup B$ (d) $A \cap B$

2-2 INTERPRETATIONS OF PROBABILITY

2-2.1 Introduction

In this chapter, we introduce probability for **discrete sample spaces**—those with only a finite (or countably infinite) set of outcomes. The restriction to these sample spaces enables us to simplify the concepts and the presentation without excessive mathematics.

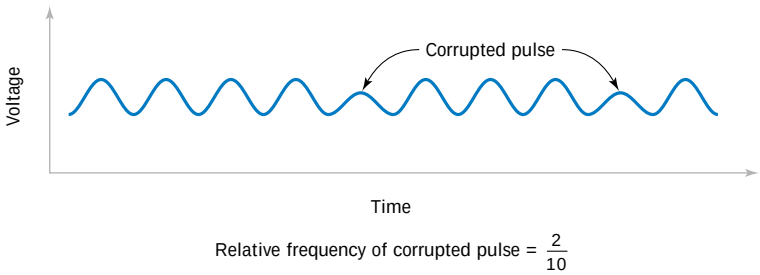


Figure 2-9 Relative frequency of corrupted pulses sent over a communication channel.

Probability is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur. "The chance of rain today is 30%" is a statement that quantifies our feeling about the possibility of rain. The likelihood of an outcome is quantified by assigning a number from the interval [0, 1] to the outcome (or a percentage from 0 to 100%). Higher numbers indicate that the outcome is more likely than lower numbers. A 0 indicates an outcome will not occur. A probability of 1 indicates an outcome will occur with certainty.

The probability of an outcome can be interpreted as our subjective probability, or **degree of belief,** that the outcome will occur. Different individuals will no doubt assign different probabilities to the same outcomes. Another interpretation of probability is based on the conceptual model of repeated replications of the random experiment. The probability of an outcome is interpreted as the limiting value of the proportion of times the outcome occurs in n repetitions of the random experiment as n increases beyond all bounds. For example, if we assign probability 0.2 to the outcome that there is a corrupted pulse in a digital signal, we might interpret this assignment as implying that, if we analyze many pulses, approximately 20% of them will be corrupted. This example provides a **relative frequency** interpretation of probability. The proportion, or relative frequency, of replications of the experiment that result in the outcome is 0.2. Probabilities are chosen so that the sum of the probabilities of all outcomes in an experiment add up to 1. This convention facilitates the relative frequency interpretation of probability. Figure 2-9 illustrates the concept of relative frequency.

Probabilities for a random experiment are often assigned on the basis of a reasonable model of the system under study. One approach is to base probability assignments on the simple concept of equally likely outcomes.

For example, suppose that we will select one laser diode **randomly** from a batch of 100. The sample space is the set of 100 diodes. *Randomly* implies that it is reasonable to assume that each diode in the batch has an equal chance of being selected. Because the sum of the probabilities must equal 1, the probability model for this experiment assigns probability of 0.01 to each of the 100 outcomes. We can interpret the probability by imagining many replications of the experiment. Each time we start with all 100 diodes and select one at random. The probability 0.01 assigned to a particular diode represents the proportion of replicates in which a particular diode is selected.

When the model of **equally likely outcomes** is assumed, the probabilities are chosen to be equal.

> Whenever a sample space consists of N possible outcomes that are equally likely, the probability of each outcome is $1/N$.

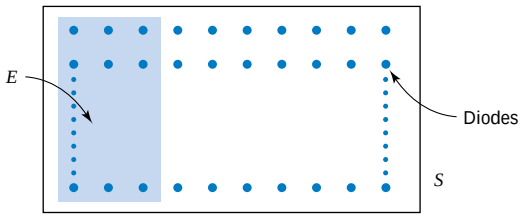


Figure 2-10
Probability of the event E is the sum of the probabilities of the outcomes in E.

It is frequently necessary to assign probabilities to events that are composed of several outcomes from the sample space. This is straightforward for a discrete sample space.

EXAMPLE 2-9

Assume that 30% of the laser diodes in a batch of 100 meet the minimum power requirements of a specific customer. If a laser diode is selected randomly, that is, each laser diode is equally likely to be selected, our intuitive feeling is that the probability of meeting the customer's requirements is 0.30.

Let E denote the subset of 30 diodes that meet the customer's requirements. Because E contains 30 outcomes and each outcome has probability 0.01, we conclude that the probability of E is 0.3. The conclusion matches our intuition. Figure 2-10 illustrates this example.

For a discrete sample space, the probability of an event can be defined by the reasoning used in the example above.

Definition

For a discrete sample space, the *probability of an event E*, denoted as $P(E)$, equals the sum of the probabilities of the outcomes in E.

EXAMPLE 2-10

A random experiment can result in one of the outcomes $\{a, b, c, d\}$ with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let A denote the event $\{a, b\}$, B the event $\{b, c, d\}$, and C the event $\{d\}$.Then,

$$P(A) = 0.1 + 0.3 = 0.4$$

$$P(B) = 0.3 + 0.5 + 0.1 = 0.9$$

$$P(C) = 0.1$$

Also, $P(A') = 0.6$, $P(B') = 0.1$, and $P(C') = 0.9$. Furthermore, because $A \cup B = \{b\}$, $P(A \cup B) = 0.3$. Because $A \cap B = \{a, b, c, d\}$, $P(A \cap B) = 0.1 + 0.3 + 0.5 + 0.1 = 1$. Because $A \cup C$ is the null set, $P(A \cup C) = 0$.

EXAMPLE 2-11

A visual inspection of a location on wafers from a semiconductor manufacturing process resulted in the following table:

Number of Contamination Particles	Proportion of Wafers
0	0.40
1	0.20
2	0.15
3	0.10
4	0.05
5 or more	0.10

If one wafer is selected randomly from this process and the location is inspected, what is the probability that it contains no particles? If information were available for each wafer, we could define the sample space as the set of all wafers inspected and proceed as in the example with diodes. However, this level of detail is not needed in this case. We can consider the sample space to consist of the six categories that summarize the number of contamination particles on a wafer. Then, the event that there is no particle in the inspected location on the wafer, denoted as E, can be considered to be comprised of the single outcome, namely, $E = \{0\}$. Therefore,

$$P(E) = 0.4$$

What is the probability that a wafer contains three or more particles in the inspected location? Let E denote the event that a wafer contains three or more particles in the inspected location. Then, E consists of the three outcomes {3, 4, 5 or more}. Therefore,

$$P(E) = 0.10 + 0.05 + 0.10 = 0.25$$

EXAMPLE 2-12

Suppose that a batch contains six parts with part numbers $\{a, b, c, d, e, f\}$. Suppose that two parts are selected without replacement. Let E denote the event that the part number of the first part selected is a. Then E can be written as $E = \{ab, ac, ad, ae, af\}$. The sample space can be enumerated. It has 30 outcomes. If each outcome is equally likely, $P(E) = 5/30 = 1/6$.

Also, if E_2 denotes the event that the second part selected is a, $E_2 = \{ba, ca, da, ea, fa\}$ and with equally likely outcomes, $P(E_2) = 5/30 = 1/6$.

2-2.2 Axioms of Probability

Now that the probability of an event has been defined, we can collect the assumptions that we have made concerning probabilities into a set of **axioms** that the probabilities in any random experiment must satisfy. The axioms ensure that the probabilities assigned in an experiment can be interpreted as relative frequencies and that the assignments are consistent with our intuitive understanding of relationships between relative frequencies. For example, if event A is contained in event B, we should have $P(A) \le P(B)$. The **axioms do not determine probabilities;** the probabilities are assigned based on our knowledge of the system under study. However, the axioms enable us to easily calculate the probabilities of some events from knowledge of the probabilities of other events.

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

(2) $0 \le P(E) \le 1$

(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

The property that $0 \le P(E) \le 1$ is equivalent to the requirement that a relative frequency must be between 0 and 1. The property that $P(S) = 1$ is a consequence of the fact that an outcome from the sample space occurs on every trial of an experiment. Consequently, the relative frequency of S is 1. Property 3 implies that if the events E_1 and E_2 have no outcomes in common, the relative frequency of outcomes in $E_1 \cup E_2$ is the sum of the relative frequencies of the outcomes in E_1 and E_2.

These axioms imply the following results. The derivations are left as exercises at the end of this section. Now,

$$P(\emptyset) = 0$$

and for any event E,

$$P(E') = 1 - P(E)$$

For example, if the probability of the event E is 0.4, our interpretation of relative frequency implies that the probability of E' is 0.6. Furthermore, if the event E_1 is contained in the event E_2,

$$P(E_1) \le P(E_2)$$

EXERCISES FOR SECTION 2-2

2-34. Each of the possible five outcomes of a random experiment is equally likely. The sample space is $\{a, b, c, d, e\}$. Let A denote the event $\{a, b\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) $P(A)$ (b) $P(B)$

(c) $P(A')$ (d) $P(A \cup B)$

(e) $P(A \cap B)$

2-35. The sample space of a random experiment is $\{a, b, c, d, e\}$ with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event $\{a, b, c\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) $P(A)$ (b) $P(B)$

(c) $P(A')$ (d) $P(A \cup B)$

(e) $P(A \cap B)$

2-36. A part selected for testing is equally likely to have been produced on any one of six cutting tools.

(a) What is the sample space?

(b) What is the probability that the part is from tool 1?

(c) What is the probability that the part is from tool 3 or tool 5?

(d) What is the probability that the part is not from tool 4?

2-37. An injection-molded part is equally likely to be obtained from any one of the eight cavities on a mold.

(a) What is the sample space?

(b) What is the probability a part is from cavity 1 or 2?

(c) What is the probability that a part is neither from cavity 3 nor 4?

2-38. A sample space contains 20 equally likely outcomes. If the probability of event A is 0.3, how many outcomes are in event A?

2-39. Orders for a computer are summarized by the optional features that are requested as follows:

	proportion of orders
no optional features	0.3
one optional feature	0.5
more than one optional feature	0.2

(a) What is the probability that an order requests at least one optional feature?
(b) What is the probability that an order does not request more than one optional feature?

2-40. If the last digit of a weight measurement is equally likely to be any of the digits 0 through 9,
(a) What is the probability that the last digit is 0?
(b) What is the probability that the last digit is greater than or equal to 5?

2-41. A sample preparation for a chemical measurement is completed correctly by 25% of the lab technicians, completed with a minor error by 70%, and completed with a major error by 5%.
(a) If a technician is selected randomly to complete the preparation, what is the probability it is completed without error?
(b) What is the probability that it is completed with either a minor or a major error?

2-42. A credit card contains 16 digits between 0 and 9. However, only 100 million numbers are valid. If a number is entered randomly, what is the probability that it is a valid number?

2-43. Suppose your vehicle is licensed in a state that issues license plates that consist of three digits (between 0 and 9) followed by three letters (between A and Z). If a license number is selected randomly, what is the probability that yours is the one selected?

2-44. A message can follow different paths through servers on a network. The senders message can go to one of five servers for the first step, each of them can send to five servers at the second step, each of which can send to four servers at the third step, and then the message goes to the recipients server.
(a) How many paths are possible?
(b) If all paths are equally likely, what is the probability that a message passes through the first of four servers at the third step?

2-45. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		shock resistance	
		high	low
scratch resistance	high	70	9
	low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. If a disk is selected at random, determine the following probabilities:
(a) $P(A)$ (b) $P(B)$
(c) $P(A')$ (d) $P(A \cup B)$
(e) $P(A \cap B)$ (f) $P(A' \cap B)$

2-46. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		edge finish	
		excellent	good
surface finish	excellent	80	2
	good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent length. If a part is selected at random, determine the following probabilities:
(a) $P(A)$ (b) $P(B)$
(c) $P(A')$ (d) $P(A \cup B)$
(e) $P(A \cap B)$ (f) $P(A' \cap B)$

2-47. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		conforms	
		yes	no
supplier	1	22	8
	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. If a sample is selected at random, determine the following probabilities:
(a) $P(A)$ (b) $P(B)$
(c) $P(A')$ (d) $P(A \cup B)$
(e) $P(A \cap B)$ (f) $P(A' \cap B)$

2-48. Use the axioms of probability to show the following:
(a) For any event E, $P(E') \leq 1$ or $P(E) \geq 0$.
(b) $P(\emptyset) = 0$
(c) If A is contained in B, then $P(A) \leq P(B)$

2-3 ADDITION RULES

Joint events are generated by applying basic set operations to individual events. Unions of events, such as $A \cup B$; intersections of events, such as $A \cap B$; and complements of events, such as A', are commonly of interest. The probability of a joint event can often be determined from the probabilities of the individual events that comprise it. Basic set operations are also sometimes helpful in determining the probability of a joint event. In this section the focus is on unions of events.

EXAMPLE 2-13

Table 2-1 lists the history of 940 wafers in a semiconductor manufacturing process. Suppose one wafer is selected at random. Let H denote the event that the wafer contains high levels of contamination. Then, $P(H) = 358/940$.

Let C denote the event that the wafer is in the center of a sputtering tool. Then, $P(C) = 626/940$. Also, $P(H \cap C)$ is the probability that the wafer is from the center of the sputtering tool and contains high levels of contamination. Therefore,

$$P(H \cap C) = 112/940$$

The event $H \cup C$ is the event that a wafer is from the center of the sputtering tool or contains high levels of contamination (or both). From the table, $P(H \cup C) = 872/940$. An alternative calculation of $P(H \cup C)$ can be obtained as follows. The 112 wafers that comprise the event $H \cap C$ are included once in the calculation of $P(H)$ and again in the calculation of $P(C)$. Therefore, $P(H \cup C)$ can be found to be

$$P(H \cup C) = P(H) + P(C) - P(H \cap C)$$
$$= 358/940 + 626/940 - 112/940 = 872/940$$

The preceding example illustrates that the probability of A or B is interpreted as $P(A \cup B)$ and that the following general **addition rule** applies.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (2\text{-}1)$$

EXAMPLE 2-14

The wafers such as those described in Example 2-13 were further classified as either in the "center" or at the "edge" of the sputtering tool that was used in manufacturing, and by the degree of contamination. Table 2-2 shows the proportion of wafers in each category. What is

Table 2-1 Wafers in Semiconductor Manufacturing Classified by Contamination and Location

	Location in Sputtering Tool		
Contamination	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	

Table 2-2 Wafers Classified by Contamination and Location

Number of Contamination Particles	Center	Edge	Totals
0	0.30	0.10	0.40
1	0.15	0.05	0.20
2	0.10	0.05	0.15
3	0.06	0.04	0.10
4	0.04	0.01	0.05
5 or more	0.07	0.03	0.10
Totals	0.72	0.28	1.00

the probability that a wafer was either at the edge or that it contains four or more particles? Let E_1 denote the event that a wafer contains four or more particles, and let E_2 denote the event that a wafer is at the edge.

The requested probability is $P(E_1 \cup E_2)$. Now, $P(E_1) = 0.15$ and $P(E_2) = 0.28$. Also, from the table, $P(E_1 \cap E_2) = 0.04$. Therefore, using Equation 2-1, we find that

$$P(E_1 \cup E_2) = 0.15 + 0.28 - 0.04 = 0.39$$

What is the probability that a wafer contains less than two particles or that it is both at the edge and contains more than four particles? Let E_1 denote the event that a wafer contains less than two particles, and let E_2 denote the event that a wafer is both from the edge and contains more than four particles. The requested probability is $P(E_1 \cup E_2)$. Now, $P(E_1) = 0.60$ and $P(E_2) = 0.03$. Also, E_1 and E_2 are mutually exclusive. Consequently, there are no wafers in the intersection and $P(E_1 \cap E_2) = 0$. Therefore,

$$P(E_1 \cup E_2) = 0.60 + 0.03 = 0.63$$

Recall that two events A and B are said to be mutually exclusive if $A \cap B = \emptyset$. Then, $P(A \cap B) = 0$, and the general result for the probability of $A \cup B$ simplifies to the third axiom of probability.

If A and B are mutually exclusive events,

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

Three or More Events

More complicated probabilities, such as $P(A \cup B \cup C)$, can be determined by repeated use of Equation 2-1 and by using some basic set operations. For example,

$$P(A \cup B \cup C) = P[(A \cup B) \cup C] = P(A \cup B) + P(C) - P[(A \cup B) \cap C]$$

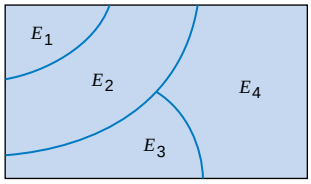


Figure 2-11 Venn diagram of four mutually exclusive events.

Upon expanding $P(A \cup B)$ by Equation 2-1 and using the distributed rule for set operations to simplify $P[(A \cup B) \cap C]$, we obtain

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

We have developed a formula for the probability of the union of three events. Formulas can be developed for the probability of the union of any number of events, although the formulas become very complex. As a summary, for the case of three events

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) + P(A \cap B) + P(A \cap C) \\ &\quad + P(B \cap C) + P(A \cap B \cap C) \end{aligned} \tag{2-3}$$

Results for three or more events simplify considerably if the events are mutually exclusive. In general, a collection of events, $E_1, E_2, \ldots, E_k$, is said to be mutually exclusive if there is no overlap among any of them.

The Venn diagram for several mutually exclusive events is shown in Fig. 2-11. By generalizing the reasoning for the union of two events, the following result can be obtained:

A collection of events, $E_1, E_2, \ldots, E_k$, is said to be **mutually exclusive** if for all pairs,

$$E_i \cap E_j = \emptyset$$

For a collection of mutually exclusive events,

$$P(E_1 \cup E_2 \cup \cdots \cup E_k) = P(E_1) + P(E_2) + \cdots + P(E_k) \tag{2-4}$$

EXAMPLE 2-15 A simple example of mutually exclusive events will be used quite frequently. Let X denote the pH of a sample. Consider the event that X is greater than 6.5 but less than or equal to 7.8. This

probability is the sum of any collection of mutually exclusive events with union equal to the same range for X. One example is

$$P(6.5 \le X \le 7.8) = P(6.5 \le X \le 7.0) + P(7.0 \le X \le 7.5) + P(7.5 \le X \le 7.8)$$

Another example is

$$P(6.5 \le X \le 7.8) = P(6.5 \le X \le 6.6) + P(6.6 \le X \le 7.1) + P(7.1 \le X \le 7.4) + P(7.4 \le X \le 7.8)$$

The best choice depends on the particular probabilities available.

EXERCISES FOR SECTION 2-3

2-49. If $P(A) = 0.3$, $P(B) = 0.2$, and $P(A \cup B) = 0.1$, determine the following probabilities:

(a) $P(A')$ (b) $P(A \cap B)$
(c) $P(A' \cup B)$ (d) $P(A \cup B')$
(e) $P[(A \cap B)']$ (f) $P(A' \cap B)$

2-50. If A, B, and C are mutually exclusive events with $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.4$, determine the following probabilities:

(a) $P(A \cap B \cap C)$ (b) $P(A \cup B \cup C)$
(c) $P(A \cup B)$ (d) $P[(A \cap B) \cup C]$
(e) $P(A' \cup B' \cup C')$

2-51. If A, B, and C are mutually exclusive events, is it possible for $P(A) = 0.3$, $P(B) = 0.4$, and $P(C) = 0.5$? Why or why not?

2-52. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		shock resistance	
		high	low
scratch resistance	high	70	9
	low	16	5

(a) If a disk is selected at random, what is the probability that its scratch resistance is high and its shock resistance is high?
(b) If a disk is selected at random, what is the probability that its scratch resistance is high or its shock resistance is high?
(c) Consider the event that a disk has high scratch resistance and the event that a disk has high shock resistance. Are these two events mutually exclusive?

2-53. The analysis of shafts for a compressor is summarized by conformance to specifications.

		roundness conforms	
		yes	no
surface finish conforms	yes	345	5
	no	12	8

(a) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements?
(b) What is the probability that the selected shaft conforms to surface finish requirements or to roundness requirements?
(c) What is the probability that the selected shaft either conforms to surface finish requirements or does not conform to roundness requirements?
(d) What is the probability that the selected shaft conforms to both surface finish and roundness requirements?

2-54. Cooking oil is produced in two main varieties: mono- and polyunsaturated. Two common sources of cooking oil are corn and canola. The following table shows the number of bottles of these oils at a supermarket:

		type of oil	
		canola	corn
type of unsaturation	mono	7	13
	poly	93	77

(a) If a bottle of oil is selected at random, what is the probability that it belongs to the polyunsaturated category?
(b) What is the probability that the chosen bottle is monounsaturated canola oil?

2-55. A manufacturer of front lights for automobiles tests lamps under a high humidity, high temperature environment using intensity and useful life as the responses of interest. The following table shows the performance of 130 lamps:

		useful life	
		satisfactory	unsatisfactory
intensity	satisfactory	117	3
	unsatisfactory	8	2

(a) Find the probability that a randomly selected lamp will yield unsatisfactory results under any criteria.
(b) The customers for these lamps demand 95% satisfactory results. Can the lamp manufacturer meet this demand?

2-56. The shafts in Exercise 2-53 are further classified in terms of the machine tool that was used for manufacturing the shaft.

Tool 1

		roundness conforms	
		yes	no
surface finish	yes	200	1
conforms	no	4	2

Tool 2

		roundness conforms	
		yes	no
surface finish	yes	145	4
conforms	no	8	6

(a) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or to roundness requirements or is from Tool 1?

(b) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or does not conform to roundness requirements or is from Tool 2?

(c) If a shaft is selected at random, what is the probability that the shaft conforms to both surface finish and roundness requirements or the shaft is from Tool 2?

(d) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or the shaft is from Tool 2?

2-4 CONDITIONAL PROBABILITY

A digital communication channel has an error rate of one bit per every thousand transmitted. Errors are rare, but when they occur, they tend to occur in bursts that affect many consecutive bits. If a single bit is transmitted, we might model the probability of an error as 1□1000. However, if the previous bit was in error, because of the bursts, we might believe that the probability that the next bit is in error is greater than 1□1000.

In a thin film manufacturing process, the proportion of parts that are not acceptable is 2%. However, the process is sensitive to contamination problems that can increase the rate of parts that are not acceptable. If we knew that during a particular shift there were problems with the filters used to control contamination, we would assess the probability of a part being unacceptable as higher than 2%.

In a manufacturing process, 10% of the parts contain visible surface flaws and 25% of the parts with surface flaws are (functionally) defective parts. However, only 5% of parts without surface flaws are defective parts. The probability of a defective part depends on our knowledge of the presence or absence of a surface flaw.

These examples illustrate that probabilities need to be reevaluated as additional information becomes available. The notation and details are further illustrated for this example.

Let D denote the event that a part is defective and let F denote the event that a part has a surface flaw. Then, we denote the probability of D given, or assuming, that a part has a surface flaw as $P(D|F)$. This notation is read as the **conditional probability** of D given F, and it is interpreted as the probability that a part is defective, given that the part has a surface flaw. Because 25% of the parts with surface flaws are defective, our conclusion can be stated as $P(D|F) = 0.25$. Furthermore, because F' denotes the event that a part does not have a surface flaw and because 5% of the parts without surface flaws are defective, we have that $P(D|F') = 0.05$. These results are shown graphically in Fig. 2-12.

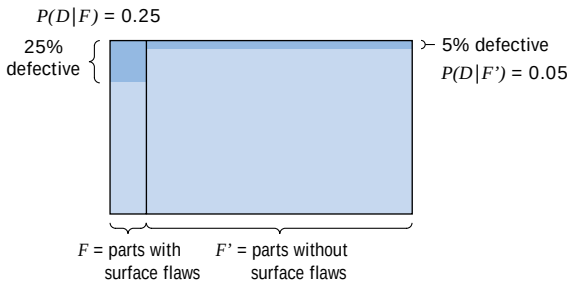


Figure 2-12 Conditional probabilities for parts with surface flaws.

Table 2-3 Parts Classified

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	10	18	38
	No	30	342	362
	Total	40	360	400

EXAMPLE 2-16

Table 2-3 provides an example of 400 parts classified by surface flaws and as (functionally) defective. For this table the conditional probabilities match those discussed previously in this section. For example, of the parts with surface flaws (40 parts) the number defective is 10. Therefore,

$$P(D \mid F) = 10/40 = 0.25$$

and of the parts without surface flaws (360 parts) the number defective is 18. Therefore,

$$P(D \mid F') = 18/360 = 0.05$$

In Example 2-16 conditional probabilities were calculated directly. These probabilities can also be determined from the formal definition of conditional probability.

Definition

The **conditional probability** of an event B given an event A, denoted as $P(B \mid A)$, is

$$P(B \mid A) = P(A \cap B)/P(A) \tag{2-5}$$

for $P(A) > 0$.

This definition can be understood in a special case in which all outcomes of a random experiment are equally likely. If there are n total outcomes,

$$P(A) = (\text{number of outcomes in } A)/n$$

Also,

$$P(A \cap B) = (\text{number of outcomes in } A \cap B)/n$$

Consequently,

$$P(A \cap B)/P(A) = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } A}$$

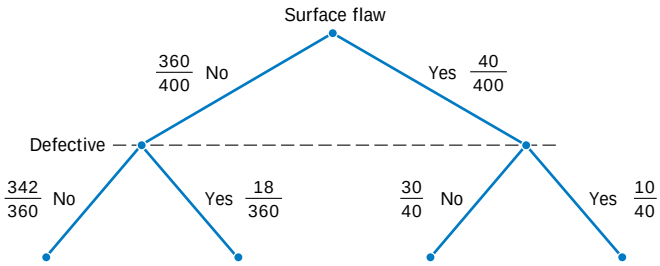


Figure 2-13 Tree diagram for parts classified

Therefore, $P(B \mid A)$ can be interpreted as the relative frequency of event B among the trials that produce an outcome in event A.

EXAMPLE 2-17

Again consider the 400 parts in Table 2-3. From this table

$$P(D \mid F) = P(D \cap F)/P(F) = \frac{10}{400} \Big/ \frac{40}{400} = \frac{10}{40}$$

Note that in this example all four of the following probabilities are different:

$$P(F) = 40/400 \qquad P(F \mid D) = 10/28$$
$$P(D) = 28/400 \qquad P(D \mid F) = 10/40$$

Here, $P(D)$ and $P(D \mid F)$ are probabilities of the same event, but they are computed under two different states of knowledge. Similarly, $P(F)$ and $P(F \mid D)$ are computed under two different states of knowledge.

The tree diagram in Fig. 2-13 can also be used to display conditional probabilities. The first branch is on surface flaw. Of the 40 parts with surface flaws, 10 are functionally defective and 30 are not. Therefore,

$$P(D \mid F) = 10/40 \quad \text{and} \quad P(D' \mid F) = 30/40$$

Of the 360 parts without surface flaws, 18 are functionally defective and 342 are not. Therefore,

$$P(D \mid F') = 342/360 \quad \text{and} \quad P(D' \mid F') = 18/360$$

Random Samples from a Batch

Recall that to select one item randomly from a batch implies that each item is equally likely. If more than one item is selected, *randomly* implies that each element of the sample space is equally likely. For example, when sample spaces were presented earlier in this chapter, sampling with and without replacement were defined and illustrated for the simple case of a batch with three items $\{a, b, c\}$. If two items are selected randomly from this batch without replacement, each of the six outcomes in the ordered sample space

$$S_{\text{without}} = \{ab, ac, ba, bc, ca, cb\}$$

has probability 1/6. If the unordered sample space is used, each of the three outcomes in $\{\{a, b\}, \{a, c\}, \{b, c\}\}$ has probability 1/3.

What is the conditional probability that *b* is selected second given that *a* is selected first? Because this question considers the results of each pick, the ordered sample space is used. The definition of conditional probability is applied as follows. Let E_1 denote the event that the first item selected is *a* and let E_2 denote the event that the second item selected is *b*. Then,

$$E_1 = \{ab, ac\} \quad \text{and} \quad E_2 = \{ab, cb\} \quad \text{and} \quad E_1 \cap E_2 = \{ab\}$$

and from the definition of conditional probability

$$P(E_2 \mid E_1) = P(E_1 \cap E_2)/P(E_1) = \frac{1/6}{1/3} = 1/2$$

When the sample space is larger, an alternative calculation is usually more convenient. For example, suppose that a batch contains 10 parts from tool 1 and 40 parts from tool 2. If two parts are selected randomly, without replacement, what is the conditional probability that a part from tool 2 is selected second given that a part from tool 1 is selected first? There are 50 possible parts to select first and 49 to select second. Therefore, the (ordered) sample space has $50 \times 49 = 2450$ outcomes. Let E_1 denote the event that the first part is from tool 1 and E_2 denote the event that the second part is from tool 2. As above, a count of the number of outcomes in E_1 and the intersection is needed.

Although the answer can be determined from this start, this type of question can be answered more easily with the following result.

> To select randomly implies that at each step of the sample, the items that remain in the batch are equally likely to be selected.

If a part from tool 1 were selected with the first pick, 49 items would remain, 9 from tool 1 and 40 from tool 2, and they would be equally likely to be picked. Therefore, the probability that a part from tool 2 would be selected with the second pick given this first pick is

$$P(E_2 \mid E_1) = 40/49.$$

In this manner, other probabilities can also be simplified. For example, let the event *E* consist of the outcomes with the first selected part from tool 1 and the second part from tool 2. To determine the probability of *E*, consider each step. The probability that a part from tool 1 is selected with the first pick is $P(E_1) = 10/50$. The conditional probability that a part from tool 2 is selected with the second pick, given that a part from tool 1 is selected first is $P(E_2 \mid E_1) = 40/49$. Therefore,

$$P(E) = P(E_2 \mid E_1)P(E_1) = \frac{40}{49} \cdot \frac{10}{50} = 0.163$$

Sometimes a partition of the question into successive picks is an easier method to solve the problem.

EXAMPLE 2-18

A day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Two parts are selected randomly without replacement from the batch. What is the probability that the second part is defective given that the first part is defective?

Let A denote the event that the first part selected is defective, and let B denote the event that the second part selected is defective. The probability needed can be expressed as $P(B|A)$. If the first part is defective, prior to selecting the second part, the batch contains 849 parts, of which 49 are defective, therefore

$$P(B|A) = 49/849$$

EXAMPLE 2-19

Continuing the previous example, if three parts are selected at random, what is the probability that the first two are defective and the third is not defective? This event can be described in shorthand notation as simply $P(ddn)$. We have

$$P(ddn) = \frac{50}{850} \cdot \frac{49}{849} \cdot \frac{800}{848} = 0.0032$$

The third term is obtained as follows. After the first two parts are selected, there are 848 remaining. Of the remaining parts, 800 are not defective. In this example, it is easy to obtain the solution with a conditional probability for each selection.

EXERCISES FOR SECTION 2-4

2-57. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		shock resistance	
		high	low
scratch resistance	high	70	9
	low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$
(c) $P(A|B)$ (d) $P(B|A)$

2-58. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		length	
		excellent	good
surface finish	excellent	80	2
	good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent length. Determine:

(a) $P(A)$ (b) $P(B)$
(c) $P(A|B)$ (d) $P(B|A)$
(e) If the selected part has excellent surface finish, what is the probability that the length is excellent?
(f) If the selected part has good length, what is the probability that the surface finish is excellent?

2-59. The analysis of shafts for a compressor is summarized by conformance to specifications:

		roundness conforms	
		yes	no
surface finish conforms	yes	345	5
	no	12	8

(a) If we know that a shaft conforms to roundness requirements, what is the probability that it conforms to surface finish requirements?
(b) If we know that a shaft does not conform to roundness requirements, what is the probability that it conforms to surface finish requirements?

2-60. The following table summarizes the analysis of samples of galvanized steel for coating weight and surface roughness:

		coating weight	
		high	low
surface roughness	high	12	16
	low	88	34

(a) If the coating weight of a sample is high, what is the probability that the surface roughness is high?
(b) If the surface roughness of a sample is high, what is the probability that the coating weight is high?
(c) If the surface roughness of a sample is low, what is the probability that the coating weight is low?

2-61. Consider the data on wafer contamination and location in the sputtering tool shown in Table 2-2. Assume that one wafer is selected at random from this set. Let A denote the event that a wafer contains four or more particles, and let B denote the event that a wafer is from the center of the sputtering tool. Determine:
(a) $P(A)$ (b) $P(A|B)$
(c) $P(B)$ (d) $P(B|A)$
(e) $P(A \cup B)$ (f) $P(A \cap B)$

2-62. A lot of 100 semiconductor chips contains 20 that are defective. Two are selected randomly, without replacement, from the lot.
(a) What is the probability that the first one selected is defective?
(b) What is the probability that the second one selected is defective given that the first one was defective?
(c) What is the probability that both are defective?
(d) How does the answer to part (b) change if chips selected were replaced prior to the next selection?

2-63. A lot contains 15 castings from a local supplier and 25 castings from a supplier in the next state. Two castings are selected randomly, without replacement, from the lot of 40. Let A be the event that the first casting selected is from the local supplier, and let B denote the event that the second casting is selected from the local supplier. Determine:
(a) $P(A)$ (b) $P(B|A)$
(c) $P(A \cup B)$ (d) $P(A \cap B)$

2-64. Continuation of Exercise 2-63. Suppose three castings are selected at random, without replacement, from the lot of 40. In addition to the definitions of events A and B, let C denote the event that the third casting selected is from the local supplier. Determine:
(a) $P(A \cup B \cup C)$
(b) $P(A \cup B \cup C')$

2-65. A batch of 500 containers for frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement from the batch.
(a) What is the probability that the second one selected is defective given that the first one was defective?
(b) What is the probability that both are defective?
(c) What is the probability that both are acceptable?

2-66. Continuation of Exercise 2-65. Three containers are selected, at random, without replacement, from the batch.
(a) What is the probability that the third one selected is defective given that the first and second one selected were defective?
(b) What is the probability that the third one selected is defective given that the first one selected was defective and the second one selected was okay?
(c) What is the probability that all three are defective?

2-67. A maintenance firm has gathered the following information regarding the failure mechanisms for air conditioning systems:

		evidence of gas leaks	
		yes	no
evidence of electrical failure	yes	55	17
	no	32	3

The units without evidence of gas leaks or electrical failure showed other types of failure. If this is a representative sample of AC failure, find the probability
(a) That failure involves a gas leak
(b) That there is evidence of electrical failure given that there was a gas leak
(c) That there is evidence of a gas leak given that there is evidence of electrical failure

2-68. If $P(A|B) = 1$, must $A = B$? Draw a Venn diagram to explain your answer.

2-69. Suppose A and B are mutually exclusive events. Construct a Venn diagram that contains the three events A, B, and C such that $P(A|C) = 1$ and $P(B|C) = 0$?

2-5 MULTIPLICATION AND TOTAL PROBABILITY RULES

2-5.1 Multiplication Rule

The definition of conditional probability in Equation 2-5 can be rewritten to provide a general expression for the probability of the intersection of two events. This formula is referred to as a **multiplication rule** for probabilities.

Multiplication Rule

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B) \quad (2\text{-}6)$$

The last expression in Equation 2-6 is obtained by interchanging A and B.

EXAMPLE 2-20

The probability that an automobile battery subject to high engine compartment temperature suffers low charging current is 0.7. The probability that a battery is subject to high engine compartment temperature is 0.05.

Let C denote the event that a battery suffers low charging current, and let T denote the event that a battery is subject to high engine compartment temperature. The probability that a battery is subject to low charging current and high engine compartment temperature is

$$P(C \cap T) = P(C|T)P(T) = 0.7 \times 0.05 = 0.035$$

2-5.2 Total Probability Rule

The multiplication rule is useful for determining the probability of an event that depends on other events. For example, suppose that in semiconductor manufacturing the probability is 0.10 that a chip that is subjected to high levels of contamination during manufacturing causes a product failure. The probability is 0.005 that a chip that is not subjected to high contamination levels during manufacturing causes a product failure. In a particular production run, 20% of the chips are subject to high levels of contamination. What is the probability that a product using one of these chips fails?

Clearly, the requested probability depends on whether or not the chip was exposed to high levels of contamination. We can solve this problem by the following reasoning. For any event B, we can write B as the union of the part of B in A and the part of B in A'. That is,

$$B = (A \cap B) \cup (A' \cap B)$$

This result is shown in the Venn diagram in Fig. 2-14. Because A and A' are mutually exclusive, $A \cap B$ and $A' \cap B$ are mutually exclusive. Therefore, from the probability of the union of mutually exclusive events in Equation 2-2 and the Multiplication Rule in Equation 2-6, the following **total probability rule** is obtained.

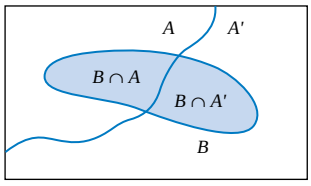
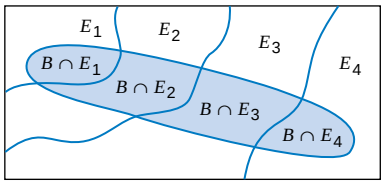


Figure 2-14 Partitioning an event into two mutually exclusive subsets.



$B = (B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3) \cup (B \cap E_4)$

Figure 2-15 Partitioning an event into several mutually exclusive subsets.

Total Probability Rule (two events)

For any events A and B,

$$P(B) = P(B \cap A) + P(B \cap A') = P(B \mid A)P(A) + P(B \mid A')P(A') \quad (2\text{-}7)$$

EXAMPLE 2-21

Consider the contamination discussion at the start of this section. Let F denote the event that the product fails, and let H denote the event that the chip is exposed to high levels of contamination. The requested probability is $P(F)$, and the information provided can be represented as

$$P(F \mid H) = 0.10 \quad \text{and} \quad P(F \mid H') = 0.005$$
$$P(H) = 0.20 \quad \text{and} \quad P(H') = 0.80$$

From Equation 2-7,

$$P(F) = 0.10(0.20) + 0.005(0.80) = 0.0235$$

which can be interpreted as just the weighted average of the two probabilities of failure.

The reasoning used to develop Equation 2-7 can be applied more generally. In the development of Equation 2-7, we only used the two mutually exclusive A and A'. However, the fact that $A \cup A' = S$, the entire sample space, was important. In general, a collection of sets $E_1, E_2, \ldots, E_k$ such that $E_1 \cup E_2 \cup \cdots \cup E_k = S$ is said to be **exhaustive.** A graphical display of partitioning an event B among a collection of mutually exclusive and exhaustive events is shown in Fig. 2-15 on page 43.

Total Probability Rule (multiple events)

Assume $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive sets. Then

$$P(B) = P(B \cap E_1) + P(B \cap E_2) + \cdots + P(B \cap E_k)$$
$$= P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k) \quad (2\text{-}8)$$

EXAMPLE 2-22

Continuing with the semiconductor manufacturing example, assume the following probabilities for product failure subject to levels of contamination in manufacturing:

Probability of Failure	Level of Contamination
0.10	High
0.01	Medium
0.001	Low

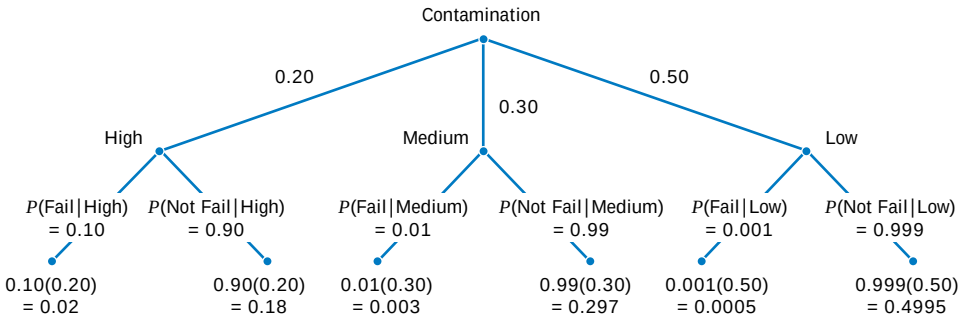


Figure 2-16 Tree diagram for Example 2-22.

In a particular production run, 20% of the chips are subjected to high levels of contamination, 30% to medium levels of contamination, and 50% to low levels of contamination. What is the probability that a product using one of these chips fails? Let

H denote the event that a chip is exposed to high levels of contamination

M denote the event that a chip is exposed to medium levels of contamination

L denote the event that a chip is exposed to low levels of contamination

Then,

$$P(F) = P(F|H)P(H) + P(F|M)P(M) + P(F|L)P(L)$$
$$= 0.10(0.20) + 0.01(0.30) + 0.001(0.50) = 0.0235$$

This problem is also conveniently solved using the tree diagram in Fig. 2-16.

EXERCISES FOR SECTION 2-5

2-70. Suppose that $P(A|B) = 0.4$ and $P(B) = 0.5$. Determine the following:
(a) $P(A \cap B)$
(b) $P(A' \cap B)$

2-71. Suppose that $P(A|B) = 0.2$, $P(A|B') = 0.3$, and $P(B) = 0.8$. What is $P(A)$?

2-72. The probability is 1% that an electrical connector that is kept dry fails during the warranty period of a portable computer. If the connector is ever wet, the probability of a failure during the warranty period is 5%. If 90% of the connectors are kept dry and 10% are wet, what proportion of connectors fail during the warranty period?

2-73. Suppose 2% of cotton fabric rolls and 3% of nylon fabric rolls contain flaws. Of the rolls used by a manufacturer, 70% are cotton and 30% are nylon. What is the probability that a randomly selected roll used by the manufacturer contains flaws?

2-74. In the manufacturing of a chemical adhesive, 3% of all batches have raw materials from two different lots. This occurs when holding tanks are replenished and the remaining portion of a lot is insufficient to fill the tanks.

Only 5% of batches with material from a single lot require reprocessing. However, the viscosity of batches consisting of two or more lots of material is more difficult to control, and 40% of such batches require additional processing to achieve the required viscosity.

Let *A* denote the event that a batch is formed from two different lots, and let *B* denote the event that a lot requires additional processing. Determine the following probabilities:
(a) $P(A)$ (b) $P(A')$
(c) $P(B|A)$ (d) $P(B|A')$
(e) $P(A \cap B)$ (f) $P(A \cap B')$
(g) $P(B)$

2-75. The edge roughness of slit paper products increases as knife blades wear. Only 1% of products slit with new blades have rough edges, 3% of products slit with blades of average sharpness exhibit roughness, and 5% of products slit with worn blades exhibit roughness. If 25% of the blades in manufacturing are new, 60% are of average sharpness, and 15% are worn, what is the proportion of products that exhibit edge roughness?

2-76. Samples of laboratory glass are in small, light packaging or heavy, large packaging. Suppose that 2 and 1% of the sample shipped in small and large packages, respectively, break during transit. If 60% of the samples are shipped in large packages and 40% are shipped in small packages, what proportion of samples break during shipment?

2-77. Incoming calls to a customer service center are classified as complaints (75% of call) or requests for information (25% of calls). Of the complaints, 40% deal with computer equipment that does not respond and 57% deal with incomplete software installation; and in the remaining 3% of complaints the user has improperly followed the installation instructions. The requests for information are evenly divided on technical questions (50%) and requests to purchase more products (50%).

(a) What is the probability that an incoming call to the customer service center will be from a customer who has not followed installation instructions properly?

(b) Find the probability that an incoming call is a request for purchasing more products.

2-78. Computer keyboard failures are due to faulty electrical connects (12%) or mechanical defects (88%). Mechanical defects are related to loose keys (27%) or improper assembly (73%). Electrical connect defects are caused by defective wires (35%), improper connections (13%), or poorly welded wires (52%).

(a) Find the probability that a failure is due to loose keys.

(b) Find the probability that a failure is due to improperly connected or poorly welded wires.

2-79. A batch of 25 injection-molded parts contains 5 that have suffered excessive shrinkage.

(a) If two parts are selected at random, and without replacement, what is the probability that the second part selected is one with excessive shrinkage?

(b) If three parts are selected at random, and without replacement, what is the probability that the third part selected is one with excessive shrinkage?

2-80. A lot of 100 semiconductor chips contains 20 that are defective.

(a) Two are selected, at random, without replacement, from the lot. Determine the probability that the second chip selected is defective.

(b) Three are selected, at random, without replacement, from the lot. Determine the probability that all are defective.

2-6 INDEPENDENCE

In some cases, the conditional probability of $P(B|A)$ might equal $P(B)$. In this special case, knowledge that the outcome of the experiment is in event A does not affect the probability that the outcome is in event B.

EXAMPLE 2-23

Suppose a day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Suppose two parts are selected from the batch, but the first part is replaced before the second part is selected. What is the probability that the second part is defective (denoted as B) given that the first part is defective (denoted as A)? The probability needed can be expressed as $P(B|A)$.

Because the first part is replaced prior to selecting the second part, the batch still contains 850 parts, of which 50 are defective. Therefore, the probability of B does not depend on whether or not the first part was defective. That is,

$$P(B|A) = 50/850$$

Also, the probability that both parts are defective is

$$P(A \cap B) = P(B|A)P(A) = \left(\frac{50}{850}\right)\left(\frac{50}{850}\right) = 0.0035$$

Table 2-4 Parts Classified

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	2	18	20
	No	38	342	380
	Total	40	360	400

EXAMPLE 2-24

The information in Table 2-3 related surface flaws to functionally defective parts. In that case, we determined that $P(D \mid F) = 10/40 = 0.25$ and $P(D) = 28/400 = 0.07$. Suppose that the situation is different and follows Table 2-4. Then,

$$P(D \mid F) = 2/40 = 0.05 \quad \text{and} \quad P(D) = 20/400 = 0.05$$

That is, the probability that the part is defective does not depend on whether it has surface flaws. Also,

$$P(F \mid D) = 2/20 = 0.10 \quad \text{and} \quad P(F) = 40/400 = 0.10$$

so the probability of a surface flaw does not depend on whether the part is defective. Furthermore, the definition of conditional probability implies that

$$P(F \cap D) = P(D \mid F)P(F)$$

but in the special case of this problem

$$P(F \cap D) = P(D)P(F) = \frac{2}{40} \cdot \frac{2}{20} = \frac{1}{200}$$

The preceding example illustrates the following conclusions. In the special case that $P(B \mid A) = P(B)$, we obtain

$$P(A \cap B) = P(B \mid A)P(A) = P(B)P(A)$$

and

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

These conclusions lead to an important definition.

Definition

Two events are **independent** if any one of the following equivalent statements is true:

(1) $P(A \mid B) = P(A)$

(2) $P(B \mid A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$ (2-9)

It is left as a mind-expanding exercise to show that independence implies related results such as

$$P(A_1 \cap B_1) = P(A_1)P(B_1).$$

The concept of independence is an important relationship between events and is used throughout this text. A mutually exclusive relationship between two events is based only on the outcomes that comprise the events. However, an independence relationship depends on the probability model used for the random experiment. Often, independence is assumed to be part of the random experiment that describes the physical system under study.

EXAMPLE 2-25

A day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Two parts are selected at random, without replacement, from the batch. Let A denote the event that the first part is defective, and let B denote the event that the second part is defective.

We suspect that these two events are not independent because knowledge that the first part is defective suggests that it is less likely that the second part selected is defective. Indeed, $P(B \mid A) = 49/849$. Now, what is $P(B)$? Finding the unconditional $P(B)$ is somewhat difficult because the possible values of the first selection need to be considered:

$$\begin{aligned} P(B) &= P(B \mid A)P(A) + P(B \mid A_1')P(A_1') \\ &= (49/849)(50/850) + (50/849)(800/850) \\ &= 50/850 \end{aligned}$$

Interestingly, $P(B)$, the unconditional probability that the second part selected is defective, without any knowledge of the first part, is the same as the probability that the first part selected is defective. Yet, our goal is to assess independence. Because $P(B \mid A)$ does not equal $P(B)$, the two events are not independent, as we suspected.

When considering three or more events, we can extend the definition of independence with the following general result.

Definition

The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \times P(E_{i_2}) \times \cdots \times P(E_{i_k}) \qquad (2\text{-}10)$$

This definition is typically used to calculate the probability that several events occur assuming that they are independent and the individual event probabilities are known. The knowledge that the events are independent usually comes from a fundamental understanding of the random experiment.

EXAMPLE 2-26

Assume that the probability that a wafer contains a large particle of contamination is 0.01 and that the wafers are independent; that is, the probability that a wafer contains a large particle is

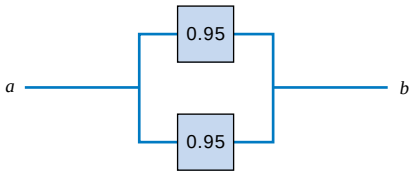
not dependent on the characteristics of any of the other wafers. If 15 wafers are analyzed, what is the probability that no large particles are found?

Let E_i denote the event that the ith wafer contains no large particles, $i = 1, 2, \ldots, 15$. Then, $P(E_i) = 0.99$. The probability requested can be represented as $P(E_1 \cap E_2 \cap \cdots \cap E_{15})$. From the independence assumption and Equation 2-10,

$$P(E_1 \cap E_2 \cap \cdots \cap E_{15}) = P(E_1)P(E_2)\cdots P(E_{15}) = 0.99^{15} = 0.86$$

EXAMPLE 2-27

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



Let T and B denote the events that the top and bottom devices operate, respectively. There is a path if at least one device operates. The probability that the circuit operates is

$$P(T \text{ or } B) = 1 - P[(T \text{ or } B)'] = 1 - P(T' \text{ and } B')$$

a simple formula for the solution can be derived from the complements T' and B'. From the independence assumption,

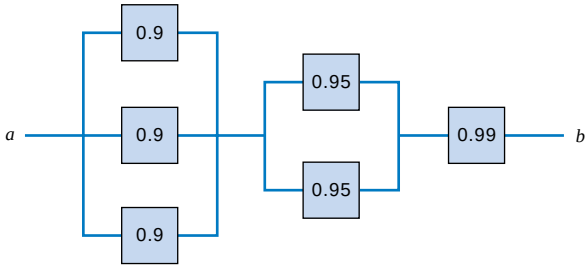
$$P(T' \text{ and } B') = P(T')P(B') = (1 - 0.95)^2 = 0.05^2$$

so

$$P(T \text{ or } B) = 1 - 0.05^2 = 0.9975$$

EXAMPLE 2-28

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



The solution can be obtained from a partition of the graph into three columns. The probability that there is a path of functional devices only through the three units on the left can be determined from the independence in a manner similar to the previous example. It is

$$1 - 0.1^3$$

Similarly, the probability that there is a path of functional devices only through the two units in the middle is

$$1 - 0.05^2$$

The probability that there is a path of functional devices only through the one unit on the right is simply the probability that the device functions, namely, 0.99. Therefore, with the independence assumption used again, the solution is

$$(1 - 0.1^3)(1 - 0.05^2)(0.99) = 0.987$$

EXERCISES FOR SECTION 2-6

2-81. If $P(A \mid B) = 0.4$, $P(B) = 0.8$, and $P(A) = 0.5$, are the events A and B independent?

2-82. If $P(A \mid B) = 0.3$, $P(B) = 0.8$, and $P(A) = 0.3$, are the events B and the complement of A independent?

2-83. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		shock resistance	
		high	low
scratch resistance	high	70	9
	low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Are events A and B independent?

2-84. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		length	
		excellent	good
surface finish	excellent	80	2
	good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent length. Are events A and B independent?

2-85. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		conforms	
		yes	no
supplier	1	22	8
	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications.
(a) Are events A and B independent?
(b) Determine $P(B \mid A)$.

2-86. If $P(A) = 0.2$, $P(B) = 0.2$, and A and B are mutually exclusive, are they independent?

2-87. The probability that a lab specimen contains high levels of contamination is 0.10. Five samples are checked, and the samples are independent.
(a) What is the probability that none contains high levels of contamination?
(b) What is the probability that exactly one contains high levels of contamination?
(c) What is the probability that at least one contains high levels of contamination?

2-88. In a test of a printed circuit board using a random test pattern, an array of 10 bits is equally likely to be 0 or 1. Assume the bits are independent.
(a) What is the probability that all bits are 1s?
(b) What is the probability that all bits are 0s?
(c) What is the probability that exactly five bits are 1s and five bits are 0s?

2-89. Eight cavities in an injection-molding tool produce plastic connectors that fall into a common stream. A sample is

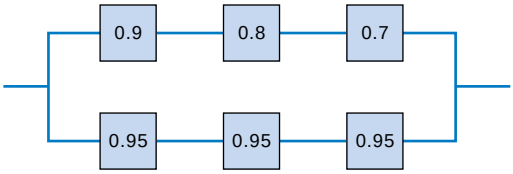
chosen every several minutes. Assume that the samples are independent.

(a) What is the probability that five successive samples were all produced in cavity one of the mold?

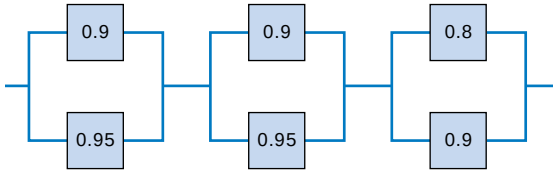
(b) What is the probability that five successive samples were all produced in the same cavity of the mold?

(c) What is the probability that four out of five successive samples were produced in cavity one of the mold?

2-90. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



2-91. The following circuit operates if and only if there is a path of functional devices from left to right. The probability each device functions is as shown. Assume that the probability that a device functions does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



2-92. An optical storage device uses an error recovery procedure that requires an immediate satisfactory readback of any written data. If the readback is not successful after three writing operations, that sector of the disk is eliminated as unacceptable for data storage. On an acceptable portion of the disk, the probability of a satisfactory readback is 0.98. Assume the readbacks are independent. What is the probability that an acceptable portion of the disk is eliminated as unacceptable for data storage?

2-93. A batch of 500 containers for frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement, from the batch. Let A and B denote the events that the first and second container selected is defective, respectively.

(a) Are A and B independent events?

(b) If the sampling were done with replacement, would A and B be independent?

2-7 BAYES' THEOREM

In some examples, we do not have a complete table of information such as the parts in Table 2-3. We might know one conditional probability but would like to calculate a different one. In the semiconductor contamination problem in Example 2-22, we might ask the following: If the semiconductor chip in the product fails, what is the probability that the chip was exposed to high levels of contamination?

From the definition of conditional probability,

$$P(A \cap B) = P(A|B)P(B) = P(B \cap A) = P(B|A)P(A)$$

Now considering the second and last terms in the expression above, we can write

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{for} \quad P(B) > 0 \tag{2-11}$$

This is a useful result that enables us to solve for $P(A|B)$ in terms of $P(B|A)$.

EXAMPLE 2-29

We can answer the question posed at the start of this section as follows: The probability requested can be expressed as $P(H \mid F)$. Then,

$$P(H \mid F) = \frac{P(F \mid H)P(H)}{P(F)} = \frac{0.10(0.20)}{0.0235} = 0.85$$

The value of $P(F)$ in the denominator of our solution was found in Example 2-20.

In general, if $P(B)$ in the denominator of Equation 2-11 is written using the Total Probability Rule in Equation 2-8, we obtain the following general result, which is known as **Bayes' Theorem.**

Bayes' Theorem

If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1 \mid B) = \frac{P(B \mid E_1)P(E_1)}{P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k)} \quad (2\text{-}12)$$

for $P(B) \neq 0$

EXAMPLE 2-30

Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive is 0.99, and the probability that the test correctly identifies someone without the illness as negative is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the illness?

Let D denote the event that you have the illness, and let S denote the event that the test signals positive. The probability requested can be denoted as $P(D \mid S)$. The probability that the test correctly signals someone without the illness as negative is 0.95. Consequently, the probability of a positive test without the illness is

$$P(S \mid D') = 0.05$$

From Bayes' Theorem,

$$\begin{aligned} P(D \mid S) &= P(S \mid D)P(D) / [P(S \mid D)P(D) + P(S \mid D')P(D')] \\ &= 0.99(0.0001) / [0.99(0.0001) + 0.05(1) + 0.0001)] \\ &= 1/506 = 0.002 \end{aligned}$$

Surprisingly, even though the test is effective, in the sense that $P(S \mid D)$ is high and $P(S \mid D')$ is low, because the incidence of the illness in the general population is low, the chances are quite small that you actually have the disease even if the test is positive.

EXERCISES FOR SECTION 2-7

2-94. Suppose that $P(A \mid B') = 0.7$, $P(A') = 0.5$, and $P(B') = 0.2$. Determine $P(B \mid A')$.

2-95. Software to detect fraud in consumer phone cards tracks the number of metropolitan areas where calls originate each day. It is found that 1% of the legitimate users originate calls from two or more metropolitan areas in a single day. However, 30% of fraudulent users originate calls from two or more metropolitan areas in a single day. The proportion of fraudulent users is 0.01%. If the same user originates calls from two or more metropolitan areas in a single day, what is the probability that the user is fraudulent?

2-96. Semiconductor lasers used in optical storage products require higher power levels for write operations than for read operations. High-power-level operations lower the useful life of the laser.

Lasers in products used for backup of higher speed magnetic disks primarily write, and the probability that the useful life exceeds five years is 0.95. Lasers that are in products that are used for main storage spend approximately an equal amount of time reading and writing, and the probability that the useful life exceeds five years is 0.995. Now, 25% of the products from a manufacturer are used for backup and 75% of the products are used for main storage.

Let A denote the event that a laser's useful life exceeds five years, and let B denote the event that a laser is in a product that is used for backup.

Use a tree diagram to determine the following:

(a) $P(B')$ (b) $P(A \mid B')$
(c) $P(A \mid B'^{c})$ (d) $P(A \cup B')$
(e) $P(A \cup B'^{c})$ (f) $P(A')$
(g) What is the probability that the useful life of a laser exceeds five years?
(h) What is the probability that a laser that failed before five years came from a product used for backup?

2-97. Customers are used to evaluate preliminary product designs. In the past, 95% of highly successful products received good reviews, 60% of moderately successful products received good reviews, and 10% of poor products received good reviews. In addition, 40% of products have been highly successful, 35% have been moderately successful, and 25% have been poor products.

(a) What is the probability that a product attains a good review?
(b) If a new design attains a good review, what is the probability that it will be a highly successful product?
(c) If a product does not attain a good review, what is the probability that it will be a highly successful product?

2-98. An inspector working for a manufacturing company has a 99% chance of correctly identifying defective items and a 0.5% chance of incorrectly classifying a good item as defective. The company has evidence that its line produces 0.9% of nonconforming items.

(a) What is the probability that an item selected for inspection is classified as defective?
(b) If an item selected at random is classified as nondefective, what is the probability that it is indeed good?

2-99. A new analytical method to detect pollutants in water is being tested. This new method of chemical analysis is important because, if adopted, it could be used to detect three different contaminants—organic pollutants, volatile solvents, and chlorinated compounds—instead of having to use a single test for each pollutant. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated compounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds.

A test sample is selected randomly.

(a) What is the probability that the test will signal?
(b) If the test signals, what is the probability that chlorinated compounds are present?

2-8 RANDOM VARIABLES

We often summarize the outcome from a random experiment by a simple number. In many of the examples of random experiments that we have considered, the sample space has been a description of possible outcomes. In some cases, descriptions of outcomes are sufficient, but in other cases, it is useful to associate a number with each outcome in the sample space. Because the particular outcome of the experiment is not known in advance, the resulting value of our variable is not known in advance. For this reason, the variable that associates a number with the outcome of a random experiment is referred to as a **random variable.**

Definition

A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

A random variable is denoted by an uppercase letter such as X. After an experiment is conducted, the measured value of the random variable is denoted by a lowercase letter such as x = 70 milliamperes.

Sometimes a measurement (such as current in a copper wire or length of a machined part) can assume any value in an interval of real numbers (at least theoretically). Then arbitrary precision in the measurement is possible. Of course, in practice, we might round off to the nearest tenth or hundredth of a unit. The random variable that represents this measurement is said to be a **continuous** random variable. The range of the random variable includes all values in an interval of real numbers; that is, the range can be thought of as a continuum.

In other experiments, we might record a count such as the number of transmitted bits that are received in error. Then the measurement is limited to integers. Or we might record that a proportion such as 0.0042 of the 10,000 transmitted bits were received in error. Then the measurement is fractional, but it is still limited to discrete points on the real line. Whenever the measurement is limited to discrete points on the real line, the random variable is said to be a **discrete** random variable.

Definition

A **discrete** random variable is a random variable with a finite (or countably infinite) range.

A **continuous** random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

In some cases, the random variable X is actually discrete but, because the range of possible values is so large, it might be more convenient to analyze X as a continuous random variable. For example, suppose that current measurements are read from a digital instrument that displays the current to the nearest one-hundredth of a milliampere. Because the possible measurements are limited, the random variable is discrete. However, it might be a more convenient, simple approximation to assume that the current measurements are values of a continuous random variable.

Examples of Random Variables

Examples of **continuous** random variables:

electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:

number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error.

EXERCISES FOR SECTION 2-8

2-100. Decide whether a discrete or continuous random variable is the best model for each of the following variables:

(a) The time until a projectile returns to earth.

(b) The number of times a transistor in a computer memory changes state in one operation.

(c) The volume of gasoline that is lost to evaporation during the filling of a gas tank.

(d) The outside diameter of a machined shaft.
(e) The number of cracks exceeding one-half inch in 10 miles of an interstate highway.
(f) The weight of an injection-molded plastic part.
(g) The number of molecules in a sample of gas.
(h) The concentration of output from a reactor.
(i) The current in an electronic circuit.

Supplemental Exercises

2-101. In circuit testing of printed circuit boards, each board either fails or does not fail the test. A board that fails the test is then checked further to determine which one of five defect types is the primary failure mode. Represent the sample space for this experiment.

2-102. The data from 200 machined parts are summarized as follows:

	depth of bore	
edge condition	above target	below target
coarse	15	10
moderate	25	20
smooth	50	80

(a) What is the probability that a part selected has a moderate edge condition and a below-target bore depth?
(b) What is the probability that a part selected has a moderate edge condition or a below-target bore depth?
(c) What is the probability that a part selected does not have a moderate edge condition or does not have a below-target bore depth?
(d) Construct a Venn diagram representation of the events in this sample space.

2-103. Computers in a shipment of 100 units contain a portable hard drive, CD RW drive, or both according to the following table:

	portable hard drive	
	yes	no
CD RW		
yes	15	80
no	4	1

Let A denote the events that a computer has a portable hard drive and let B denote the event that a computer has a CD RW drive. If one computer is selected randomly, compute

(a) $P(A')$ (b) $P(A \cup B)$
(c) $P(A \cap B)$ (d) $P(A' \cup B)$
(e) $P(A \mid B)$

2-104. The probability that a customer's order is not shipped on time is 0.05. A particular customer places three orders, and the orders are placed far enough apart in time that they can be considered to be independent events.

(a) What is the probability that all are shipped on time?
(b) What is the probability that exactly one is not shipped on time?
(c) What is the probability that two or more orders are not shipped on time?

2-105. Let E_1, E_2, and E_3 denote the samples that conform to a percentage of solids specification, a molecular weight specification, and a color specification, respectively. A total of 240 samples are classified by the E_1, E_2, and E_3 specifications, where *yes* indicates that the sample conforms.

E_3 yes

		E_2		
		yes	no	Total
E_1	yes	200	1	201
	no	5	4	9
Total		205	5	210

E_3 no

		E_2		
		yes	no	
E_1	yes	20	4	24
	no	6	0	6
Total		26	4	30

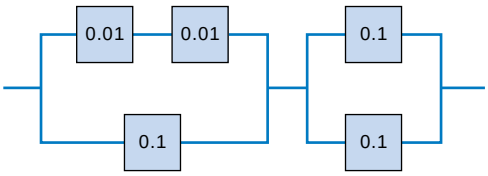
(a) Are E_1, E_2, and E_3 mutually exclusive events?
(b) Are E_1', E_2', and E_3' mutually exclusive events?
(c) What is $P(E_1'$ or E_2' or $E_3')$?
(d) What is the probability that a sample conforms to all three specifications?
(e) What is the probability that a sample conforms to the E_1 or E_3 specification?
(f) What is the probability that a sample conforms to the E_1 or E_2 or E_3 specification?

2-106. Transactions to a computer database are either new items or changes to previous items. The addition of an item can be completed less than 100 milliseconds 90% of the time, but only 20% of changes to a previous item can be completed in less than this time. If 30% of transactions are changes, what is the probability that a transaction can be completed in less than 100 milliseconds?

2-107. A steel plate contains 20 bolts. Assume that 5 bolts are not torqued to the proper limit. Four bolts are selected at random, without replacement, to be checked for torque.
(a) What is the probability that all four of the selected bolts are torqued to the proper limit?
(b) What is the probability that at least one of the selected bolts is not torqued to the proper limit?

2-108. The following circuit operates if and only if there is a path of functional devices from left to right. Assume devices fail independently and that the probability of *failure* of each

device is as shown. What is the probability that the circuit operates?



2-109. The probability of getting through by telephone to buy concert tickets is 0.92. For the same event, the probability of accessing the vendor's Web site is 0.95. Assume that these two ways to buy tickets are independent. What is the probability that someone who tries to buy tickets through the Internet and by phone will obtain tickets?

2-110. The British government has stepped up its information campaign regarding foot and mouth disease by mailing brochures to farmers around the country. It is estimated that 99% of Scottish farmers who receive the brochure possess enough information to deal with an outbreak of the disease, but only 90% of those without the brochure can deal with an outbreak. After the first three months of mailing, 95% of the farmers in Scotland received the informative brochure. Compute the probability that a randomly selected farmer will have enough information to deal effectively with an outbreak of the disease.

2-111. In an automated filling operation, the probability of an incorrect fill when the process is operated at a low speed is 0.001. When the process is operated at a high speed, the probability of an incorrect fill is 0.01. Assume that 30% of the containers are filled when the process is operated at a high speed and the remainder are filled when the process is operated at a low speed.

(a) What is the probability of an incorrectly filled container?
(b) If an incorrectly filled container is found, what is the probability that it was filled during the high-speed operation?

2-112. An encryption-decryption system consists of three elements: encode, transmit, and decode. A faulty encode occurs in 0.5% of the messages processed, transmission errors occur in 1% of the messages, and a decode error occurs in 0.1% of the messages. Assume the errors are independent.

(a) What is the probability of a completely defect-free message?
(b) What is the probability of a message that has either an encode or a decode error?

2-113. It is known that two defective copies of a commercial software program were erroneously sent to a shipping lot that has now a total of 75 copies of the program. A sample of copies will be selected from the lot without replacement.

(a) If three copies of the software are inspected, determine the probability that exactly one of the defective copies will be found.
(b) If three copies of the software are inspected, determine the probability that both defective copies will be found.
(c) If 73 copies are inspected, determine the probability that both copies will be found. Hint: Work with the copies that remain in the lot.

2-114. A robotic insertion tool contains 10 primary components. The probability that any component fails during the warranty period is 0.01. Assume that the components fail independently and that the tool fails if any component fails. What is the probability that the tool fails during the warranty period?

2-115. An e-mail message can travel through one of two server routes. The probability of transmission error in each of the servers and the proportion of messages that travel each route are shown in the following table. Assume that the servers are independent.

		probability of error			
	percentage of messages	server 1	server 2	server 3	server 4
route 1	30	0.01	0.015		
route 2	70			0.02	0.003

(a) What is the probability that a message will arrive without error?
(b) If a message arrives in error, what is the probability it was sent through route 1?

2-116. A machine tool is idle 15% of the time. You request immediate use of the tool on five different occasions during the year. Assume that your requests represent independent events.

(a) What is the probability that the tool is idle at the time of all of your requests?
(b) What is the probability that the machine is idle at the time of exactly four of your requests?
(c) What is the probability that the tool is idle at the time of at least three of your requests?

2-117. A lot of 50 spacing washers contains 30 washers that are thicker than the target dimension. Suppose that three washers are selected at random, without replacement, from the lot.

(a) What is the probability that all three washers are thicker than the target?
(b) What is the probability that the third washer selected is thicker than the target if the first two washers selected are thinner than the target?
(c) What is the probability that the third washer selected is thicker than the target?

2-118. Continuation of Exercise 2-117. Washers are selected from the lot at random, without replacement.

(a) What is the minimum number of washers that need to be selected so that the probability that all the washers are thinner than the target is less than 0.10?
(b) What is the minimum number of washers that need to be selected so that the probability that one or more washers are thicker than the target is at least 0.90?

2-119. The following table lists the history of 940 orders for features in an entry-level computer product.

		extra memory	
		no	yes
optional high-speed processor	no	514	68
	yes	112	246

Let A be the event that an order requests the optional high-speed processor, and let B be the event that an order requests extra memory. Determine the following probabilities:

(a) $P(A \cap B)$ (b) $P(A \cup B)$

(c) $P(A' \cap B)$ (d) $P(A' \cup B')$

(e) What is the probability that an order requests an optional high-speed processor given that the order requests extra memory?

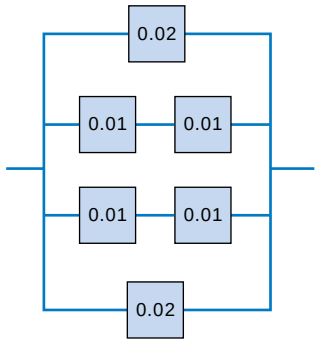
(f) What is the probability that an order requests extra memory given that the order requests an optional high-speed processor?

2-120. The alignment between the magnetic tape and head in a magnetic tape storage system affects the performance of the system. Suppose that 10% of the read operations are degraded by skewed alignments, 5% of the read operations are degraded by off-center alignments, and the remaining read operations are properly aligned. The probability of a read error is 0.01 from a skewed alignment, 0.02 from an off-center alignment, and 0.001 from a proper alignment.

(a) What is the probability of a read error?

(b) If a read error occurs, what is the probability that it is due to a skewed alignment?

2-121. The following circuit operates if and only if there is a path of functional devices from left to right. Assume that devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit does not operate?



2-122. A company that tracks the use of its web site determined that the more pages a visitor views, the more likely the visitor is to provide contact information. Use the following tables to answer the questions:

Number of pages viewed:	1	2	3	4 or more
Percentage of visitors:	40	30	20	10
Percentage of visitors in each page-view catgory that provide contact information:	10	10	20	40

(a) What is the probability that a visitor to the web site provides contact information?

(b) If a visitor provides contact information, what is the probability that the visitor viewed four or more pages?

MIND-EXPANDING EXERCISES

2-123. The alignment between the magnetic tape and head in a magnetic tape storage system affects the performance of the system. Suppose that 10% of the read operations are degraded by skewed alignments, 5% by off-center alignments, 1% by both skewness and off-center, and the remaining read operations are properly aligned. The probability of a read error is 0.01 from a skewed alignment, 0.02 from an off-center alignment, 0.06 from both conditions, and 0.001 from a proper alignment. What is the probability of a read error.

2-124. Suppose that a lot of washers is large enough that it can be assumed that the sampling is done with replacement. Assume that 60% of the washers exceed the target thickness.

(a) What is the minimum number of washers that need to be selected so that the probability that all the washers are thinner than the target is less than 0.10?

(b) What is the minimum number of washers that need to be selected so that the probability that one or more washers are thicker than the target is at least 0.90?

2-125. A biotechnology manufacturing firm can produce diagnostic test kits at a cost of $20. Each kit for which there is a demand in the week of production can be sold for $100. However, the half-life of components in the kit requires the kit to be scrapped if it is not sold in the week of production. The cost of scrapping the kit is $5. The weekly demand is summarized as follows:

	weekly demand			
Number of units	0	50	100	200
Probability of demand	0.05	0.4	0.3	0.25

How many kits should be produced each week to maximize the mean earnings of the firm?

2-126. Assume the following characteristics of the inspection process in Exercise 2-107. If an operator checks a bolt, the probability that an incorrectly torqued bolt is identified is 0.95. If a checked bolt is correctly torqued, the operator's conclusion is always correct. What is the probability that at least one bolt in the sample of four is identified as being incorrectly torqued?

2-127. If the events A and B are independent, show that $A_¿$ and $B_¿$ are independent.

2-128. Suppose that a table of part counts is generalized as follows:

		conforms	
		yes	no
supplier	1	ka	kb
	2	a	b

where a, b, and k are positive integers. Let A denote the event that a part is from supplier 1 and let B denote the event that a part conforms to specifications. Show that A and B are independent events.

This exercise illustrates the result that whenever the rows of a table (with r rows and c columns) are proportional, an event defined by a row category and an event defined by a column category are independent.

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Addition rule
Axioms of probability
Bayes' theorem
Conditional probability
Equally likely outcomes
Event
Independence
Multiplication rule
Mutually exclusive events
Outcome
Random experiment
Random variables discrete and continuous
Sample spaces—discrete and continuous
Total probability rule
With or without replacement

CD MATERIAL

Permutation
Combination

2-1.4 Counting Techniques (CD Only)

In many of the examples in Chapter 2, it is easy to determine the number of outcomes in each event. In more complicated examples, determining the number of outcomes that comprise the sample space (or an event) becomes more difficult. To associate probabilities with events, it is important to know the number of outcomes both in an event and in the sample space. Some simple rules can be used to simplify the calculations.

In Example 2-4, an automobile manufacturer provides vehicles equipped with selected options. Each vehicle is ordered

With or without an automatic transmission

With or without air conditioning

With one of three choices of a stereo system

With one of four exterior colors

The tree diagram in Fig. 2-6 describes the sample space of all possible vehicle types. The size of the sample space equals the number of branches in the last level of the tree and this quantity equals $2 \times 2 \times 3 \times 4 = 48$. This leads to the following useful result.

Multiplication Rule (for counting techniques)

If an operation can be described as a sequence of k steps, and

if the number of ways of completing step 1 is n_1, and

if the number of ways of completing step 2 is n_2 for each way of completing step 1, and

if the number of ways of completing step 3 is n_3 for each way of completing step 2, and so forth,

the total number of ways of completing the operation is

$$n_1 \times n_2 \times \cdots \times n_k$$

EXAMPLE S2-1 In the design of a casing for a gear housing, we can use four different types of fasteners, three different bolt lengths, and three different bolt locations. From the multiplication rule, $4 \times 3 \times 3 \times 36$ different designs are possible.

Permutations

Another useful calculation is the number of ordered sequences of the elements of a set. Consider a set of elements, such as $S \times \{a, b, c\}$. A **permutation** of the elements is an ordered sequence of the elements. For example, *abc*, *acb*, *bac*, *bca*, *cab*, and *cba* are all of the permutations of the elements of S.

The number of **permutations** of n different elements is $n!$ where

$$n! \times n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 \qquad \text{(S2-1)}$$

This result follows from the multiplication rule. A permutation can be constructed by selecting the element to be placed in the first position of the sequence from the n elements, then selecting the element for the second position from the $n - 1$ remaining elements, then selecting the element for the third position from the remaining $n - 2$ elements, and so forth. Permutations such as these are sometimes referred to as linear permutations.

In some situations, we are interested in the number of arrangements of only some of the elements of a set. The following result also follows from the multiplication rule.

The number of permutations of a subset of r elements selected from a set of n different elements is

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} \qquad \text{(S2-2)}$$

EXAMPLE S2-2

A printed circuit board has eight different locations in which a component can be placed. If four different components are to be placed on the board, how many different designs are possible?

Each design consists of selecting a location from the eight locations for the first component, a location from the remaining seven for the second component, a location from the remaining six for the third component, and a location from the remaining five for the fourth component. Therefore,

$$P_4^8 = 8 \cdot 7 \cdot 6 \cdot 5 = \frac{8!}{4!} = 1680 \text{ different designs are possible.}$$

Sometimes we are interested in counting the number of ordered sequences for objects that are not all different. The following result is a useful, general calculation.

The number of permutations of $n = n_1 + n_2 + \cdots + n_r$ objects of which n_1 are of one type, n_2 are of a second type, $\cdots$, and n_r are of an rth type is

$$\frac{n!}{n_1!\, n_2!\, n_3! \cdots n_r!} \qquad \text{(S2-3)}$$

EXAMPLE S2-3

Consider a machining operation in which a piece of sheet metal needs two identical diameter holes drilled and two identical size notches cut. We denote a drilling operation as d and a notching operation as n. In determining a schedule for a machine shop, we might be interested in the number of different possible sequences of the four operations. The number of possible sequences for two drilling operations and two notching operations is

$$\frac{4!}{2!\,2!} = 6$$

The six sequences are easily summarized: *ddnn*, *dndn*, *dnnd*, *nddn*, *ndnd*, *nndd*.

EXAMPLE S2-4

A part is labeled by printing with four thick lines, three medium lines, and two thin lines. If each ordering of the nine lines represents a different label, how many different labels can be generated by using this scheme?

From Equation S2-3, the number of possible part labels is

$$\frac{9!}{4!\,3!\,2!} = 2520$$

Combinations

Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements. Here, order is not important. Every subset of r elements can be indicated by listing the elements in the set and marking each element with a "*" if it is to be included in the subset. Therefore, each permutation of r *'s and $n - r$ blanks indicate a different subset and the number of these are obtained from Equation S2-3.

For example, if the set is $S = \{a, b, c, d\}$ the subset $\{a, c\}$ can be indicated as

a	b	c	d
*		*	

The number of subsets of size r that can be selected from a set of n elements is denoted as $\binom{n}{r}$ or C_r^n and

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{S2-4}$$

EXAMPLE S2-5

A printed circuit board has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many different designs are possible?

Each design is a subset of the eight locations that are to contain the components. From Equation S2-4, the number of possible designs is

$$\frac{8!}{5!\,3!} = 56$$

The following example uses the multiplication rule in combination with Equation S2-4 to answer a more difficult, but common, question.

EXAMPLE S2-6

A bin of 50 manufactured parts contains three defective parts and 47 nondefective parts. A sample of six parts is selected from the 50 parts. Selected parts are not replaced. That is, each part can only be selected once and the sample is a subset of the 50 parts. How many different samples are there of size six that contain exactly two defective parts?

A subset containing exactly two defective parts can be formed by first choosing the two defective parts from the three defective parts. Using Equation S2-4, this step can be completed in

$$\binom{3}{2} = \frac{3!}{2!\,1!} = 3 \text{ different ways}$$

Then, the second step is to select the remaining four parts from the 47 acceptable parts in the bin. The second step can be completed in

$$\binom{47}{4} = \frac{47!}{4!\,43!} = 178{,}365 \text{ different ways}$$

Therefore, from the multiplication rule, the number of subsets of size six that contain exactly two defective items is

$$3 \times 178{,}365 = 535{,}095$$

As an additional computation, the total number of different subsets of size six is found to be

$$\binom{50}{6} = \frac{50!}{6!\,44!} = 15{,}890{,}700$$

When probability is discussed in this chapter, the probability of an event is determined as the ratio of the number of outcomes in the event to the number of outcomes in the sample space (for equally likely outcomes). Therefore, the probability that a sample contains exactly two defective parts is

$$\frac{535{,}095}{15{,}890{,}700} = 0.034$$

Note that Example S2-7 illustrates the hypergeometric distribution.

EXERCISES FOR SECTION 2-1.4

S2-1. An order for a personal digital assistant can specify any one of five memory sizes, any one of three types of displays, any one of four sizes of a hard disk, and can either include or not include a pen tablet. How many different systems can be ordered?

S2-2. In a manufacturing operation, a part is produced by machining, polishing, and painting. If there are three machine tools, four polishing tools, and three painting tools, how many different routings (consisting of machining, followed by polishing, and followed by painting) for a part are possible?

S2-3. New designs for a wastewater treatment tank have proposed three possible shapes, four possible sizes, three locations for input valves, and four locations for output valves. How many different product designs are possible?

S2-4. A manufacturing process consists of 10 operations that can be completed in any order. How many different production sequences are possible?

S2-5. A manufacturing operations consists of 10 operations. However, five machining operations must be completed before any of the remaining five assembly operations can begin. Within each set of five, operations can be completed in any order. How many different production sequences are possible?

S2-6. In a sheet metal operation, three notches and four bends are required. If the operations can be done in any order, how many different ways of completing the manufacturing are possible?

S2-7. A lot of 140 semiconductor chips is inspected by choosing a sample of five chips. Assume 10 of the chips do not conform to customer requirements.

(a) How many different samples are possible?

(b) How many samples of five contain exactly one nonconforming chip?

(c) How many samples of five contain at least one nonconforming chip?

S2-8. In the layout of a printed circuit board for an electronic product, there are 12 different locations that can accommodate chips.

(a) If five different types of chips are to be placed on the board, how many different layouts are possible?

(b) If the five chips that are placed on the board are of the same type, how many different layouts are possible?

S2-9. In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed two times each day to check the calibration of the laboratory instruments.

(a) How many different sequences of process and control samples are possible each day? Assume that the five process samples are considered identical and that the two control samples are considered identical.
(b) How many different sequences of process and control samples are possible if we consider the five process samples to be different and the two control samples to be identical.
(c) For the same situation as part (b), how many sequences are possible if the first test of each day must be a control sample?

S2-10. In the design of an electromechanical product, seven different components are to be stacked into a cylindrical casing that holds 12 components in a manner that minimizes the impact of shocks. One end of the casing is designated as the bottom and the other end is the top.

(a) How many different designs are possible?
(b) If the seven components are all identical, how many different designs are possible?
(c) If the seven components consist of three of one type of component and four of another type, how many different designs are possible? (more difficult)

S2-11. The design of a communication system considered the following questions:

(a) How many three-digit phone prefixes that are used to represent a particular geographic area (such as an area code) can be created from the digits 0 through 9?
(b) As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit?
(c) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

S2-12. A byte is a sequence of eight bits and each bit is either 0 or 1.

(a) How many different bytes are possible?
(b) If the first bit of a byte is a parity check, that is, the first byte is determined from the other seven bits, how many different bytes are possible?

S2-13. In a chemical plant, 24 holding tanks are used for final product storage. Four tanks are selected at random and without replacement. Suppose that six of the tanks contain material in which the viscosity exceeds the customer requirements.

(a) What is the probability that exactly one tank in the sample contains high viscosity material?
(b) What is the probability that at least one tank in the sample contains high viscosity material?
(c) In addition to the six tanks with high viscosity levels, four different tanks contain material with high impurities. What is the probability that exactly one tank in the sample contains high viscosity material and exactly one tank in the sample contains material with high impurities?

S2-14. Plastic parts produced by an injection-molding operation are checked for conformance to specifications. Each tool contains 12 cavities in which parts are produced, and these parts fall into a conveyor when the press opens. An inspector chooses 3 parts from among the 12 at random. Two cavities are affected by a temperature malfunction that results in parts that do not conform to specifications.

(a) What is the probability that the inspector finds exactly one nonconforming part?
(b) What is the probability that the inspector finds at least one nonconforming part?

S2-15. A bin of 50 parts contains five that are defective. A sample of two is selected at random, without replacement.

(a) Determine the probability that both parts in the sample are defective by computing a conditional probability.
(b) Determine the answer to part (a) by using the subset approach that was described in this section.

3 Discrete Random Variables and Probability Distributions

CHAPTER OUTLINE

3-1 DISCRETE RANDOM VARIABLES

3-2 PROBABILITY DISTRIBUTIONS AND PROBABILITY MASS FUNCTIONS

3-3 CUMULATIVE DISTRIBUTION FUNCTIONS

3-4 MEAN AND VARIANCE OF A DISCRETE RANDOM VARIABLE

3-5 DISCRETE UNIFORM DISTRIBUTION

3-6 BINOMIAL DISTRIBUTION

3-7 GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

3-7.1 Geometric Distribution

3-7.2 Negative Binomial Distribution

3-8 HYPERGEOMETRIC DISTRIBUTION

3-9 POISSON DISTRIBUTION

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Determine probabilities from probability mass functions and the reverse
2. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability mass functions, and the reverse
3. Calculate means and variances for discrete random variables
4. Understand the assumptions for each of the discrete probability distributions presented
5. Select an appropriate discrete probability distribution to calculate probabilities in specific applications
6. Calculate probabilities, determine means and variances for each of the discrete probability distributions presented

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

3-1 DISCRETE RANDOM VARIABLES

Many physical systems can be modeled by the same or similar random experiments and random variables. The distribution of the random variables involved in each of these common systems can be analyzed, and the results of that analysis can be used in different applications and examples. In this chapter, we present the analysis of several random experiments and **discrete random variables** that frequently arise in applications. We often omit a discussion of the underlying sample space of the random experiment and directly describe the distribution of a particular random variable.

EXAMPLE 3-1

A voice communication system for a business contains 48 external lines. At a particular time, the system is observed, and some of the lines are being used. Let the random variable X denote the number of lines in use. Then, X can assume any of the integer values 0 through 48. When the system is observed, if 10 lines are in use, $x = 10$.

EXAMPLE 3-2

In a semiconductor manufacturing process, two wafers from a lot are tested. Each wafer is classified as *pass* or *fail*. Assume that the probability that a wafer passes the test is 0.8 and that wafers are independent. The sample space for the experiment and associated probabilities are shown in Table 3-1. For example, because of the independence, the probability of the outcome that the first wafer tested passes and the second wafer tested fails, denoted as *pf*, is

$$P(pf) = 0.8(0.2) = 0.16$$

The random variable X is defined to be equal to the number of wafers that pass. The last column of the table shows the values of X that are assigned to each outcome in the experiment.

EXAMPLE 3-3

Define the random variable X to be the number of contamination particles on a wafer in semiconductor manufacturing. Although wafers possess a number of characteristics, the random variable X summarizes the wafer only in terms of the number of particles.

The possible values of X are integers from zero up to some large value that represents the maximum number of particles that can be found on one of the wafers. If this maximum number is very large, we might simply assume that the range of X is the set of integers from zero to infinity.

Note that more than one random variable can be defined on a sample space. In Example 3-3, we might define the random variable Y to be the number of chips from a wafer that fail the final test.

Table 3-1 Wafer Tests

Outcome			
Wafer 1	Wafer 2	Probability	x
Pass	Pass	0.64	2
Fail	Pass	0.16	1
Pass	Fail	0.16	1
Fail	Fail	0.04	0

EXERCISES FOR SECTION 3-1

For each of the following exercises, determine the range (possible values) of the random variable.

3-1. The random variable is the number of nonconforming solder connections on a printed circuit board with 1000 connections.

3-2. In a voice communication system with 50 lines, the random variable is the number of lines in use at a particular time.

3-3. An electronic scale that displays weights to the nearest pound is used to weigh packages. The display shows only five digits. Any weight greater than the display can indicate is shown as 99999. The random variable is the displayed weight.

3-4. A batch of 500 machined parts contains 10 that do not conform to customer requirements. The random variable is the number of parts in a sample of 5 parts that do not conform to customer requirements.

3-5. A batch of 500 machined parts contains 10 that do not conform to customer requirements. Parts are selected successively, without replacement, until a nonconforming part is obtained. The random variable is the number of parts selected.

3-6. The random variable is the moisture content of a lot of raw material, measured to the nearest percentage point.

3-7. The random variable is the number of surface flaws in a large coil of galvanized steel.

3-8. The random variable is the number of computer clock cycles required to complete a selected arithmetic calculation.

3-9. An order for an automobile can select the base model or add any number of 15 options. The random variable is the number of options selected in an order.

3-10. Wood paneling can be ordered in thicknesses of 1□8, 1□4, or 3□8 inch. The random variable is the total thickness of paneling in two orders.

3-11. A group of 10,000 people are tested for a gene called Ifi202 that has been found to increase the risk for lupus. The random variable is the number of people who carry the gene.

3-12. A software program has 5000 lines of code. The random variable is the number of lines with a fatal error.

3-2 PROBABILITY DISTRIBUTIONS AND PROBABILITY MASS FUNCTIONS

Random variables are so important in random experiments that sometimes we essentially ignore the original sample space of the experiment and focus on the probability distribution of the random variable. For example, in Example 3-1, our analysis might focus exclusively on the integers $\{0, 1, \ldots, 48\}$ in the range of X. In Example 3-2, we might summarize the random experiment in terms of the three possible values of X, namely $\{0, 1, 2\}$. In this manner, a random variable can simplify the description and analysis of a random experiment.

The **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X. For a discrete random variable, the distribution is often specified by just a list of the possible values along with the probability of each. In some cases, it is convenient to express the probability in terms of a formula.

EXAMPLE 3-4

There is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. Based on a model for the errors that is presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are

$$P(X = 0) = 0.6561 \qquad P(X = 1) = 0.2916 \qquad P(X = 2) = 0.0486$$
$$P(X = 3) = 0.0036 \qquad P(X = 4) = 0.0001$$

The probability distribution of X is specified by the possible values along with the probability of each. A graphical description of the probability distribution of X is shown in Fig. 3-1.

Suppose a loading on a long, thin beam places mass only at discrete points. See Fig. 3-2. The loading can be described by a function that specifies the mass at each of the discrete points. Similarly, for a discrete random variable X, its distribution can be described by a function that specifies the probability at each of the possible discrete values for X.

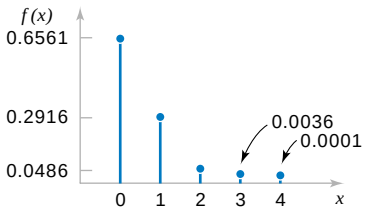


Figure 3-1 Probability distribution for bits in error.

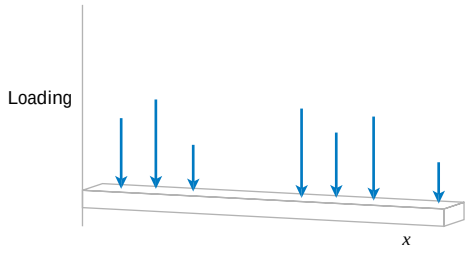


Figure 3-2 Loadings at discrete points on a long, thin beam.

Definition

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

$$(1)\quad f(x_i) \geq 0$$

$$(2)\quad \sum_{i=1}^{n} f(x_i) = 1$$

$$(3)\quad f(x_i) = P(X = x_i) \qquad (3\text{-}1)$$

For example, in Example 3-4, $f(0) = 0.6561$, $f(1) = 0.2916$, $f(2) = 0.0486$, $f(3) = 0.0036$, and $f(4) = 0.0001$. Check that the sum of the probabilities in Example 3-4 is 1.

EXAMPLE 3-5

Let the random variable X denote the number of semiconductor wafers that need to be analyzed in order to detect a large particle of contamination. Assume that the probability that a wafer contains a large particle is 0.01 and that the wafers are independent. Determine the probability distribution of X.

Let p denote a wafer in which a large particle is present, and let a denote a wafer in which it is absent. The sample space of the experiment is infinite, and it can be represented as all possible sequences that start with a string of a's and end with p. That is,

$$s = \{p, ap, aap, aaap, aaaap, aaaaap, \text{ and so forth}\}$$

Consider a few special cases. We have $P(X = 1) = P(p) = 0.01$. Also, using the independence assumption

$$P(X = 2) = P(ap) = 0.99(0.01) = 0.0099$$

A general formula is

$$P(X = x) = P(\underbrace{aa \ldots a}_{x - 1 \text{ a's}}\, p) = 0.99^{x-1}(0.01), \qquad \text{for } x = 1, 2, 3, \ldots$$

Describing the probabilities associated with X in terms of this formula is the simplest method of describing the distribution of X in this example. Clearly $f(x) \geq 0$. The fact that the sum of the probabilities is 1 is left as an exercise. This is an example of a geometric random variable, and details are provided later in this chapter.

EXERCISES FOR SECTION 3-2

3-13. The sample space of a random experiment is $\{a, b, c, d, e, f\}$, and each outcome is equally likely. A random variable is defined as follows:

outcome	a	b	c	d	e	f
x	0	0	1.5	1.5	2	3

Determine the probability mass function of X.

3-14. Use the probability mass function in Exercise 3-11 to determine the following probabilities:

(a) $P(X \le 1.5)$ (b) $P(0.5 < X \le 2.7)$
(c) $P(X \ge 3)$ (d) $P(0 < X < 2)$
(e) $P(X < 0 \text{ or } X > 2)$

Verify that the following functions are probability mass functions, and determine the requested probabilities.

3-15.

x	-2	-1	0	1	2
$f(x)$	$1/8$	$2/8$	$2/8$	$2/8$	$1/8$

(a) $P(X \le 2)$ (b) $P(X < -2)$
(c) $P(-1 < X \le 1)$ (d) $P(X < -1 \text{ or } X > 2)$

3-16. $f(x) = (3/7)(1/2)^x$, $x = 1, 2, 3$
(a) $P(X \le 1)$ (b) $P(X > 1)$
(c) $P(2 \le X \le 6)$ (d) $P(X < 1 \text{ or } X > 1)$

3-17. $f(x) = \dfrac{2x - 1}{25}$, $x = 0, 1, 2, 3, 4$
(a) $P(X \le 4)$ (b) $P(X > 1)$
(c) $P(2 \le X < 4)$ (d) $P(X < -10)$

3-18. $f(x) = (3/4)(1/4)^x$, $x = 0, 1, 2, \ldots$
(a) $P(X \le 2)$ (b) $P(X > 2)$
(c) $P(X \ge 2)$ (d) $P(X < 1)$

3-19. Marketing estimates that a new instrument for the analysis of soil samples will be very successful, moderately successful, or unsuccessful, with probabilities 0.3, 0.6, and 0.1, respectively. The yearly revenue associated with a very successful, moderately successful, or unsuccessful product is \$10 million, \$5 million, and \$1 million, respectively. Let the random variable X denote the yearly revenue of the product. Determine the probability mass function of X.

3-20. A disk drive manufacturer estimates that in five years a storage device with 1 terabyte of capacity will sell with probability 0.5, a storage device with 500 gigabytes capacity will sell with a probability 0.3, and a storage device with 100 gigabytes capacity will sell with probability 0.2. The revenue associated with the sales in that year are estimated to be \$50 million, \$25 million, and \$10 million, respectively. Let X be the revenue of storage devices during that year. Determine the probability mass function of X.

3-21. An optical inspection system is to distinguish among different part types. The probability of a correct classification of any part is 0.98. Suppose that three parts are inspected and that the classifications are independent. Let the random variable X denote the number of parts that are correctly classified. Determine the probability mass function of X.

3-22. In a semiconductor manufacturing process, three wafers from a lot are tested. Each wafer is classified as *pass* or *fail*. Assume that the probability that a wafer passes the test is 0.8 and that wafers are independent. Determine the probability mass function of the number of wafers from a lot that pass the test.

3-23. The distributor of a machine for cytogenics has developed a new model. The company estimates that when it is introduced into the market, it will be very successful with a probability 0.6, moderately successful with a probability 0.3, and not successful with probability 0.1. The estimated yearly profit associated with the model being very successful is \$15 million and being moderately successful is \$5 million; not successful would result in a loss of \$500,000. Let X be the yearly profit of the new model. Determine the probability mass function of X.

3-24. An assembly consists of two mechanical components. Suppose that the probabilities that the first and second components meet specifications are 0.95 and 0.98. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

3-25. An assembly consists of three mechanical components. Suppose that the probabilities that the first, second, and third components meet specifications are 0.95, 0.98, and 0.99. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

3-3 CUMULATIVE DISTRIBUTION FUNCTIONS

EXAMPLE 3-6 In Example 3-4, we might be interested in the probability of three or fewer bits being in error. This question can be expressed as $P(X \le 3)$.

The event that $X \le 3$ is the union of the events $X = 0$, $X = 1$, $X = 2$, and

$\{X \le 3\}$. Clearly, these three events are mutually exclusive. Therefore,

$$P(X \le 3) = P(X \le 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= 0.6561 + 0.2916 + 0.0486 + 0.0036 = 0.9999$$

This approach can also be used to determine

$$P(X \le 3) = P(X \le 3) - P(X = 2) = 0.0036$$

Example 3-6 shows that it is sometimes useful to be able to provide **cumulative probabilities** such as $P(X \le x)$ and that such probabilities can be used to find the probability mass function of a random variable. Therefore, using cumulative probabilities is an alternate method of describing the probability distribution of a random variable.

In general, for any discrete random variable with possible values $x_1, x_2, \ldots, x_n$, the events $\{X \le x_1\}, \{X \le x_2\}, \ldots, \{X \le x_n\}$ are mutually exclusive. Therefore, $P(X \le x) = \sum_{x_i \le x} f(x_i)$.

Definition

The **cumulative distribution function** of a discrete random variable X, denoted as $F(x)$, is

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

For a discrete random variable X, $F(x)$ satisfies the following properties.

(1) $F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$

(2) $0 \le F(x) \le 1$

(3) If $x \le y$, then $F(x) \le F(y)$ (3-2)

Like a probability mass function, a cumulative distribution function provides probabilities. Notice that even if the random variable X can only assume integer values, the cumulative distribution function can be defined at noninteger values. In Example 3-6, $F(1.5) = P(X \le 1.5) = P\{X \le 0\} = P(X \le 1) = 0.6561 + 0.2916 = 0.9477$. Properties (1) and (2) of a cumulative distribution function follow from the definition. Property (3) follows from the fact that if $x \le y$, the event that $\{X \le x\}$ is contained in the event $\{X \le y\}$.

The next example shows how the cumulative distribution function can be used to determine the probability mass function of a discrete random variable.

EXAMPLE 3-7

Determine the probability mass function of X from the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \le x < 0 \\ 0.7 & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

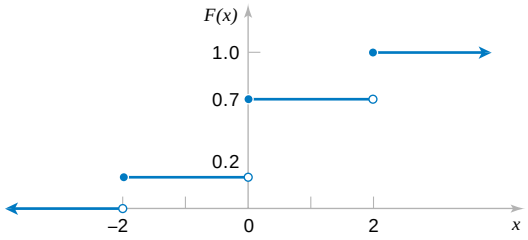


Figure 3-3 Cumulative distribution function for Example 3-7.

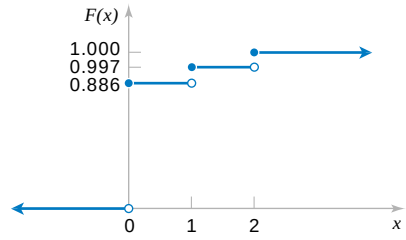


Figure 3-4 Cumulative distribution function for Example 3-8.

Figure 3-3 displays a plot of $F(x)$. From the plot, the only points that receive nonzero probability are −2, 0, and 2. The probability mass function at each point is the change in the cumulative distribution function at the point. Therefore,

$$f(-2) = 0.2 - 0 = 0.2 \qquad f(0) = 0.7 - 0.2 = 0.5 \qquad f(2) = 1.0 - 0.7 = 0.3$$

EXAMPLE 3-8

Suppose that a day's production of 850 manufactured parts contains 50 parts that do not conform to customer requirements. Two parts are selected at random, without replacement, from the batch. Let the random variable X equal the number of nonconforming parts in the sample. What is the cumulative distribution function of X?

The question can be answered by first finding the probability mass function of X.

$$P(X = 0) = \frac{800}{850} \cdot \frac{799}{849} = 0.886$$

$$P(X = 1) = 2 \cdot \frac{800}{850} \cdot \frac{50}{849} = 0.111$$

$$P(X = 2) = \frac{50}{850} \cdot \frac{49}{849} = 0.003$$

Therefore,

$$F(0) = P(X = 0) = 0.886$$

$$F(1) = P(X = 1) = 0.886 + 0.111 = 0.997$$

$$F(2) = P(X = 2) = 1$$

The cumulative distribution function for this example is graphed in Fig. 3-4. Note that $F(x)$ is defined for all x from $-\infty < x < \infty$ and not only for 0, 1, and 2.

EXERCISES FOR SECTION 3-3

3-26. Determine the cumulative distribution function of the random variable in Exercise 3-13.

3-27. Determine the cumulative distribution function for the random variable in Exercise 3-15; also determine the following probabilities:

(a) $P(X < 1.25)$ (b) $P(X \le 2.2)$ (c) $P(-1.1 < X < 1)$ (d) $P(X > 0)$

3-28. Determine the cumulative distribution function for the random variable in Exercise 3-17; also determine the following probabilities:

(a) $P(X < 1.5)$ (b) $P(X \ge 3)$ (c) $P(X > 2)$ (d) $P(1 < X < 2)$

3-29. Determine the cumulative distribution function for the random variable in Exercise 3-19.

3-30. Determine the cumulative distribution function for the random variable in Exercise 3-20.

3-31. Determine the cumulative distribution function for the random variable in Exercise 3-22.

3-32. Determine the cumulative distribution function for the variable in Exercise 3-23.

Verify that the following functions are cumulative distribution functions, and determine the probability mass function and the requested probabilities.

3-33.

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.5 & 1 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

(a) $P(X < 3)$ (b) $P(X \le 2)$
(c) $P(1 < X < 2)$ (d) $P(X > 2)$

3-34. Errors in an experimental transmission channel are found when the transmission is checked by a certifier that detects missing pulses. The number of errors found in an eight-bit byte is a random variable with the following distribution:

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.7 & 1 \le x < 4 \\ 0.9 & 4 \le x < 7 \\ 1 & 7 \le x \end{cases}$$

Determine each of the following probabilities:
(a) $P(X < 4)$ (b) $P(X < 7)$
(c) $P(X \le 5)$ (d) $P(X \ge 4)$
(e) $P(X = 2)$

3-35.

$$F(x) = \begin{cases} 0 & x < -10 \\ 0.25 & -10 \le x < 30 \\ 0.75 & 30 \le x < 50 \\ 1 & 50 \le x \end{cases}$$

(a) $P(X < 50)$ (b) $P(X \le 40)$
(c) $P(40 < X < 60)$ (d) $P(X > 0)$
(e) $P(10 < X < 10)$ (f) $P(-10 < X < 10)$

3-36. The thickness of wood paneling (in inches) that a customer orders is a random variable with the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 1/8 \\ 0.2 & 1/8 \le x < 1/4 \\ 0.9 & 1/4 \le x < 3/8 \\ 1 & 3/8 \le x \end{cases}$$

Determine the following probabilities:
(a) $P(X < 1/18)$ (b) $P(X < 1/4)$
(c) $P(X < 5/16)$ (d) $P(X > 1/4)$
(e) $P(X < 1/2)$

3-4 MEAN AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Two numbers are often used to summarize a probability distribution for a random variable X. The mean is a measure of the center or middle of the probability distribution, and the variance is a measure of the dispersion, or variability in the distribution. These two measures do not uniquely identify a probability distribution. That is, two different distributions can have the same mean and variance. Still, these measures are simple, useful summaries of the probability distribution of X.

Definition

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x f(x) \tag{3-3}$$

The **variance** of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_x (x - \mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

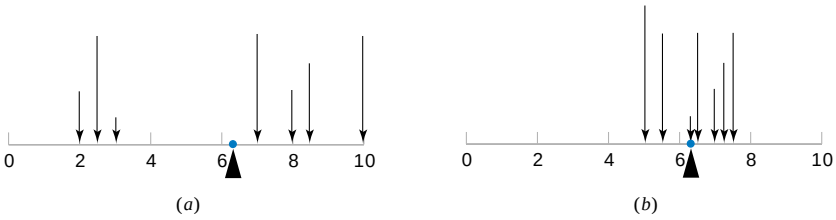


Figure 3-5 A probability distribution can be viewed as a loading with the mean equal to the balance point. Parts (a) and (b) illustrate equal means, but Part (a) illustrates a larger variance.

The mean of a discrete random variable X is a weighted average of the possible values of X, with weights equal to the probabilities. If $f(x)$ is the probability mass function of a loading on a long, thin beam, $E(X)$ is the point at which the beam balances. Consequently, $E(X)$ describes the "center" of the distribution of X in a manner similar to the balance point of a loading. See Fig. 3-5.

The variance of a random variable X is a measure of dispersion or scatter in the possible values for X. The variance of X uses weight $f(x)$ as the multiplier of each possible squared deviation $(x - \mu)^2$. Figure 3-5 illustrates probability distributions with equal means but different variances. Properties of summations and the definition of μ can be used to show the equality of the formulas for variance.

$$V(X) = \sum_x (x - \mu)^2 f(x) = \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu^2 + \mu^2 = \sum_x x^2 f(x) - \mu^2$$

Either formula for $V(X)$ can be used. Figure 3-6 illustrates that two probability distributions can differ even though they have identical means and variances.

EXAMPLE 3-9

In Example 3-4, there is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. Based on a model for the errors that is presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are

$$P(X = 0) = 0.6561 \qquad P(X = 2) = 0.0486 \qquad P(X = 4) = 0.0001$$
$$P(X = 1) = 0.2916 \qquad P(X = 3) = 0.0036$$

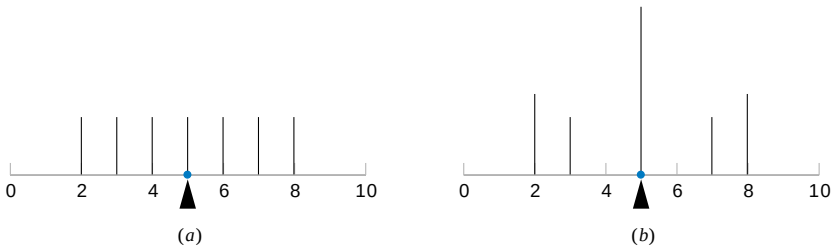


Figure 3-6 The probability distributions illustrated in Parts (a) and (b) differ even though they have equal means and equal variances.

Now

$$\mu = E(X) = 0f(0) + 1f(1) + 2f(2) + 3f(3) + 4f(4)$$
$$= 0(0.6561) + 1(0.2916) + 2(0.0486) + 3(0.0036) + 4(0.0001)$$
$$= 0.4$$

Although X never assumes the value 0.4, the weighted average of the possible values is 0.4.

To calculate $V(X)$, a table is convenient.

x	$x - 0.4$	$(x - 0.4)^2$	$f(x)$	$f(x)(x - 0.4)^2$
0	−0.4	0.16	0.6561	0.104976
1	0.6	0.36	0.2916	0.104976
2	1.6	2.56	0.0486	0.124416
3	2.6	6.76	0.0036	0.024336
4	3.6	12.96	0.0001	0.001296

$$V(X) = \sigma^2 = \sum_{i=1}^{5} f(x_i)(x_i - 0.4)^2 = 0.36$$

The alternative formula for variance could also be used to obtain the same result.

EXAMPLE 3-10

Two new product designs are to be compared on the basis of revenue potential. Marketing feels that the revenue from design A can be predicted quite accurately to be \$3 million. The revenue potential of design B is more difficult to assess. Marketing concludes that there is a probability of 0.3 that the revenue from design B will be \$7 million, but there is a 0.7 probability that the revenue will be only \$2 million. Which design do you prefer?

Let X denote the revenue from design A. Because there is no uncertainty in the revenue from design A, we can model the distribution of the random variable X as \$3 million with probability 1. Therefore, $E(X) = \$3$ million.

Let Y denote the revenue from design B. The expected value of Y in millions of dollars is

$$E(Y) = \$7(0.3) + \$2(0.7) = \$3.5$$

Because $E(Y)$ exceeds $E(X)$, we might prefer design B. However, the variability of the result from design B is larger. That is,

$$\sigma^2 = 1 - 3.5)^2(0.3) + (2 - 3.5)^2(0.7)$$
$$= 5.25 \text{ millions of dollars squared}$$

Because the units of the variables in this example are millions of dollars, and because the variance of a random variable squares the deviations from the mean, the units of σ^2 are millions of dollars squared. These units make interpretation difficult.

Because the units of standard deviation are the same as the units of the random variable, the standard deviation σ is easier to interpret. In this example, we can summarize our results as "the average deviation of Y from its mean is \$2.29 million."

EXAMPLE 3-11

The number of messages sent per hour over a computer network has the following distribution:

x = number of messages	10	11	12	13	14	15
$f(x)$	0.08	0.15	0.30	0.20	0.20	0.07

Determine the mean and standard deviation of the number of messages sent per hour.

$$E(X) = 10(0.08) + 11(0.15) + \cdots + 15(0.07) = 12.5$$

$$V(X) = 10^2(0.08) + 11^2(0.15) + \cdots + 15^2(0.07) - 12.5^2 = 1.85$$

$$\sigma = \sqrt{V(X)} = \sqrt{1.85} = 1.36$$

The variance of a random variable X can be considered to be the expected value of a specific function of X, namely, $h(X) = (X - \mu)^2$. In general, the expected value of any function $h(X)$ of a discrete random variable is defined in a similar manner.

Expected Value of a Function of a Discrete Random Variable

If X is a discrete random variable with probability mass function $f(x)$,

$$E[h(X)] = \sum_x h(x) f(x) \tag{3-4}$$

EXAMPLE 3-12

In Example 3-9, X is the number of bits in error in the next four bits transmitted. What is the expected value of the square of the number of bits in error? Now, $h(X) = X^2$. Therefore,

$$E[h(X)] = 0^2(0.6561) + 1^2(0.2916) + 2^2(0.0486) + 3^2(0.0036) + 4^2(0.0001) = 0.52$$

In the previous example, the expected value of X^2 does not equal $E(X)$ squared. However, in the special case that $h(X) = aX + b$ for any constants a and b, $E[h(X)] = aE(X) + b$. This can be shown from the properties of sums in the definition in Equation 3-4.

EXERCISES FOR SECTION 3-4

3-37. If the range of X is the set $\{0, 1, 2, 3, 4\}$ and $P(X = x) = 0.2$ determine the mean and variance of the random variable.

3-38. Determine the mean and variance of the random variable in Exercise 3-13.

3-39. Determine the mean and variance of the random variable in Exercise 3-15.

3-40. Determine the mean and variance of the random variable in Exercise 3-17.

3-41. Determine the mean and variance of the random variable in Exercise 3-19.

3-42. Determine the mean and variance of the random variable in Exercise 3-20.

3-43. Determine the mean and variance of the random variable in Exercise 3-22.

3-44. Determine the mean and variance of the random variable in Exercise 3-23.

3-45. The range of the random variable X is $\{0, 1, 2, 3, x\}$, where x is unknown. If each value is equally likely and the mean of X is 6, determine x.

3-5 DISCRETE UNIFORM DISTRIBUTION

The simplest discrete random variable is one that assumes only a finite number of possible values, each with equal probability. A random variable X that assumes each of the values $x_1, x_2, \ldots, x_n$, with equal probability $1/n$, is frequently of interest.

Definition

A random variable X has a **discrete uniform distribution** if each of the n values in its range, say, $x_1, x_2, \ldots, x_n$, has equal probability. Then,

$$f(x_i) = 1/n \tag{3-5}$$

EXAMPLE 3-13

The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \ldots, 9\}$. That is,

$$f(x) = 0.1$$

for each value in R. The probability mass function of X is shown in Fig. 3-7.

Suppose the range of the discrete random variable X is the consecutive integers $a, a + 1, a + 2, \ldots, b$, for $a < b$. The range of X contains $b - a + 1$ values each with probability $1/(b - a + 1)$. Now,

$$\mu = \sum_{k=a}^{b} k\left(\frac{1}{b - a + 1}\right)$$

The algebraic identity $\sum_{k=a}^{b} k = \dfrac{b(b+1) - (a-1)a}{2}$ can be used to simplify the result to $\mu = (b + a)/2$. The derivation of the variance is left as an exercise.

Suppose X is a discrete uniform random variable on the consecutive integers $a, a + 1, a + 2, \ldots, b$, for $a < b$. The mean of X is

$$\mu = E(X) = \frac{b + a}{2}$$

The variance of X is

$$\sigma^2 = \frac{(b - a + 1)^2 - 1}{12} \tag{3-6}$$

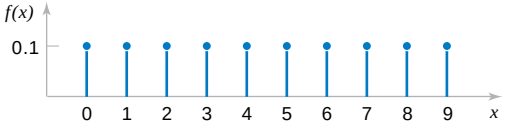


Figure 3-7 Probability mass function for a discrete uniform random variable.

EXAMPLE 3-14

As in Example 3-1, let the random variable X denote the number of the 48 voice lines that are in use at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Then,

$$E(X) = (48 + 0)/2 = 24$$

and

$$\sigma = \sqrt{(48 - 0 + 1)^2 - 1}/12 = 14.14$$

Equation 3-6 is more useful than it might first appear. If all the values in the range of a random variable X are multiplied by a constant (without changing any probabilities), the mean and standard deviation of X are multiplied by the constant. You are asked to verify this result in an exercise. Because the variance of a random variable is the square of the standard deviation, the variance of X is multiplied by the constant squared. More general results of this type are discussed in Chapter 5.

EXAMPLE 3-15

Let the random variable Y denote the proportion of the 48 voice lines that are in use at a particular time, and X denotes the number of lines that are in use at a particular time. Then, $Y = X/48$. Therefore,

$$E(Y) = E(X)/48 = 0.5$$

and

$$V(Y) = V(X)/48^2 = 0.087$$

EXERCISES FOR SECTION 3-5

3-46. Let the random variable X have a discrete uniform distribution on the integers $0 \le x \le 100$. Determine the mean and variance of X.

3-47. Let the random variable X have a discrete uniform distribution on the integers $1 \le x \le 3$. Determine the mean and variance of X.

3-48. Let the random variable X be equally likely to assume any of the values 1/8, 1/4, or 3/8. Determine the mean and variance of X.

3-49. Thickness measurements of a coating process are made to the nearest hundredth of a millimeter. The thickness measurements are uniformly distributed with values 0.15, 0.16, 0.17, 0.18, and 0.19. Determine the mean and variance of the coating thickness for this process.

3-50. Product codes of 2, 3, or 4 letters are equally likely. What is the mean and standard deviation of the number of letters in 100 codes?

3-51. The lengths of plate glass parts are measured to the nearest tenth of a millimeter. The lengths are uniformly distributed, with values at every tenth of a millimeter starting at 590.0 and continuing through 590.9. Determine the mean and variance of lengths.

3-52. Suppose that X has a discrete uniform distribution on the integers 0 through 9. Determine the mean, variance, and standard deviation of the random variable $Y = 5X$ and compare to the corresponding results for X.

3-53. Show that for a discrete uniform random variable X, if each of the values in the range of X is multiplied by the constant c, the effect is to multiply the mean of X by c and the variance of X by c^2. That is, show that $E(cX) = cE(X)$ and $V(cX) = c^2V(X)$.

3-54. The probability of an operator entering alphanumeric data incorrectly into a field in a database is equally likely. The random variable X is the number of fields on a data entry form with an error. The data entry form has 28 fields. Is X a discrete uniform random variable? Why or why not.

3-6 BINOMIAL DISTRIBUTION

Consider the following random experiments and random variables:

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A worn machine tool produces 1% defective parts. Let X = number of defective parts in the next 25 parts produced.
3. Each sample of air has a 10% chance of containing a particular rare molecule. Let X = the number of air samples that contain the rare molecule in the next 18 samples analyzed.
4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X = the number of bits in error in the next five bits transmitted.
5. A multiple choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
6. In the next 20 births at a hospital, let X = the number of female births.
7. Of all patients suffering a particular illness, 35% experience improvement from a particular medication. In the next 100 patients administered the medication, let X = the number of patients who experience improvement.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful.

Each of these random experiments can be thought of as consisting of a series of repeated, random trials: 10 flips of the coin in experiment 1, the production of 25 parts in experiment 2, and so forth. The random variable in each case is a count of the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion that X counts or it does not; consequently, each trial can be summarized as resulting in either a success or a failure. For example, in the multiple choice experiment, for each question, only the choice that is correct is considered a success. Choosing any one of the three incorrect choices results in the trial being summarized as a failure.

The terms *success* and *failure* are just labels. We can just as well use A and B or 0 or 1. Unfortunately, the usual labels can sometimes be misleading. In experiment 2, because X counts defective parts, the production of a defective part is called a success.

A trial with only two possible outcomes is used so frequently as a building block of a random experiment that it is called a **Bernoulli trial.** It is usually assumed that the trials that constitute the random experiment are **independent.** This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the **probability of a success in each trial is constant.** In the multiple choice experiment, if the test taker has no knowledge of the material and just guesses at each question, we might assume that the probability of a correct answer is 1/4 *for each question.*

Factorial notation is used in this section. Recall that $n!$ denotes the product of the integers less than or equal to n:

$$n! = n(n-1)(n-2) \cdots (2)(1)$$

For example,

$$5! = (5)(4)(3)(2)(1) = 120 \qquad 1! = 1$$

and by definition $0! = 1$. We also use the combinatorial notation

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

For example,

$$\binom{5}{2} = \frac{5!}{2!\,3!} = \frac{120}{2 \cdot 6} = 10$$

See Section 2-1.4, CD material for Chapter 2, for further comments.

EXAMPLE 3-16

The chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the transmission trials are independent. Let X = the number of bits in error in the next four bits transmitted. Determine $P(X = 2)$.

Let the letter E denote a bit in error, and let the letter O denote that the bit is okay, that is, received without error. We can represent the outcomes of this experiment as a list of four letters that indicate the bits that are in error and those that are okay. For example, the outcome $OEOE$ indicates that the second and fourth bits are in error and the other two bits are okay. The corresponding values for x are

Outcome	x	Outcome	x
OOOO	0	*EOOO*	1
OOOE	1	*EOOE*	2
OOEO	1	*EOEO*	2
OOEE	2	*EOEE*	3
OEOO	1	*EEOO*	2
OEOE	2	*EEOE*	3
OEEO	2	*EEEO*	3
OEEE	3	*EEEE*	4

The event that $X = 2$ consists of the six outcomes:

$$\{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE\}$$

Using the assumption that the trials are independent, the probability of $\{EEOO\}$ is

$$P(EEOO) = P(E)P(E)P(O)P(O) = (0.1)^2(0.9)^2 = 0.0081$$

Also, any one of the six mutually exclusive outcomes for which $X = 2$ has the same probability of occurring. Therefore,

$$P(X = 2) = 6(0.0081) = 0.0486$$

In general,

$$P(X = x) = (\text{number of outcomes that result in } x \text{ errors}) \text{ times } (0.1)^x(0.9)^{4-x}$$

To complete a general probability formula, only an expression for the number of outcomes that contain x errors is needed. An outcome that contains x errors can be constructed by partitioning the four trials (letters) in the outcome into two groups. One group is of size x and contains the errors, and the other group is of size $n - x$ and consists of the trials that are okay. The number of ways of partitioning four objects into two groups, one of which is of size x, is $\binom{4}{x} = \frac{4!}{x!(4-x)!}$. Therefore, in this example

$$P(X = x) = \binom{4}{x}(0.1)^x(0.9)^{4-x}$$

Notice that $\binom{4}{2} = 4!/(2!\,2!) = 6$, as found above. The probability mass function of X was shown in Example 3-4 and Fig. 3-1.

The previous example motivates the following result.

Definition

A random experiment consists of n Bernoulli trials such that

(1) The trials are independent

(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"

(3) The probability of a success in each trial, denoted as p, remains constant

The random variable X that equals the number of trials that result in a success has a **binomial random variable** with parameters $0 < p < 1$ and $n = 1, 2, \ldots$. The probability mass function of X is

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x} \qquad x = 0, 1, \ldots, n \tag{3-7}$$

As in Example 3-16, $\binom{n}{x}$ equals the total number of different sequences of trials that contain x successes and $n - x$ failures. The total number of different sequences that contain x successes and $n - x$ failures times the probability of each sequence equals $P(X = x)$.

The probability expression above is a very useful formula that can be applied in a number of examples. The name of the distribution is obtained from the *binomial expansion*. For constants a and b, the binomial expansion is

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

Let p denote the probability of success on a single trial. Then, by using the binomial expansion with $a = p$ and $b = 1 - p$, we see that the sum of the probabilities for a binomial random variable is 1. Furthermore, because each trial in the experiment is classified into two outcomes, {success, failure}, the distribution is called a "bi"-nomial. A more

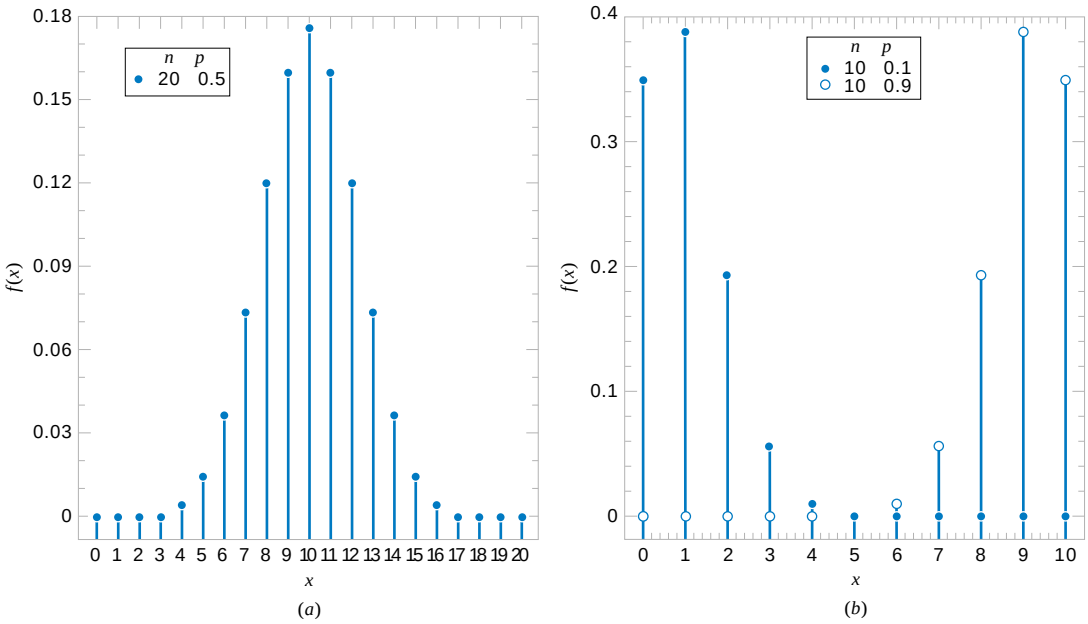


Figure 3-8 Binomial distributions for selected values of n and p.

general distribution, which includes the binomial as a special case, is the multinomial distribution.

Examples of binomial distributions are shown in Fig. 3-8. For a fixed n, the distribution becomes more symmetric as p increases from 0 to 0.5 or decreases from 1 to 0.5. For a fixed p, the distribution becomes more symmetric as n increases.

EXAMPLE 3-17 Several examples using the binomial coefficient $\binom{n}{x}$ follow.

$$\binom{10}{3} = \frac{10!}{3!\,7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2} = 120$$

$$\binom{15}{10} = \frac{15!}{10!\,5!} = \frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11}{5 \cdot 4 \cdot 3 \cdot 2} = 3003$$

$$\binom{100}{4} = \frac{100!}{4!\,96!} = \frac{100 \cdot 99 \cdot 98 \cdot 97}{4 \cdot 3 \cdot 2} = 3{,}921{,}225$$

EXAMPLE 3-18 Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that in the next 18 samples, exactly 2 contain the pollutant.

Let X = the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with $p = 0.1$ and $n = 18$.
Therefore,

$$P(X = 2) = \binom{18}{2} (0.1)^2 (0.9)^{16}$$

Now $\binom{18}{2} = 18!/(2!\,16!) = (18)(17)/2 = 153$. Therefore,

$$P(X = 2) = 153(0.1)^2(0.9)^{16} = 0.284$$

Determine the probability that at least four samples contain the pollutant. The requested probability is

$$P(X \geq 4) = \sum_{x=4}^{18} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$

However, it is easier to use the complementary event,

$$P(X \geq 4) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{3} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$
$$= 1 - (0.150 + 0.300 + 0.284 + 0.168) = 0.098$$

Determine the probability that $3 \leq X \leq 7$. Now

$$P(3 \leq X \leq 7) = \sum_{x=3}^{6} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$
$$= 0.168 + 0.070 + 0.022 + 0.005$$
$$= 0.265$$

The mean and variance of a binomial random variable depend only on the parameters p and n. Formulas can be developed from moment generating functions, and details are provided in Section 5-8, part of the CD material for Chapter 5. The results are simply stated here.

Definition

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1 - p) \tag{3-8}$$

EXAMPLE 3-19 For the number of transmitted bits received in error in Example 3-16, $n = 4$ and $p = 0.1$, so

$$E(X) = 4(0.1) = 0.4 \quad \text{and} \quad V(X) = 4(0.1)(0.9) = 0.36$$

and these results match those obtained from a direct calculation in Example 3-9.

EXERCISES FOR SECTION 3-6

3-55. For each scenario described below, state whether or not the binomial distribution is a reasonable model for the random variable and why. State any assumptions you make.

(a) A production process produces thousands of temperature transducers. Let X denote the number of nonconforming transducers in a sample of size 30 selected at random from the process.

(b) From a batch of 50 temperature transducers, a sample of size 30 is selected without replacement. Let X denote the number of nonconforming transducers in the sample.

(c) Four identical electronic components are wired to a controller that can switch from a failed component to one of the remaining spares. Let X denote the number of components that have failed after a specified period of operation.
(d) Let X denote the number of accidents that occur along the federal highways in Arizona during a one-month period.
(e) Let X denote the number of correct answers by a student taking a multiple choice exam in which a student can eliminate some of the choices as being incorrect in some questions and all of the incorrect choices in other questions.
(f) Defects occur randomly over the surface of a semiconductor chip. However, only 80% of defects can be found by testing. A sample of 40 chips with one defect each is tested. Let X denote the number of chips in which the test finds a defect.
(g) Reconsider the situation in part (f). Now, suppose the sample of 40 chips consists of chips with 1 and with 0 defects.
(h) A filling operation attempts to fill detergent packages to the advertised weight. Let X denote the number of detergent packages that are underfilled.
(i) Errors in a digital communication channel occur in bursts that affect several consecutive bits. Let X denote the number of bits in error in a transmission of 100,000 bits.
(j) Let X denote the number of surface flaws in a large coil of galvanized steel.

3-56. The random variable X has a binomial distribution with $n = 10$ and $p = 0.5$. Sketch the probability mass function of X.
(a) What value of X is most likely?
(b) What value(s) of X is(are) least likely?

3-57. The random variable X has a binomial distribution with $n = 10$ and $p = 0.5$. Determine the following probabilities:
(a) $P(X = 5)$ (b) $P(X = 2)$
(c) $P(X = 9)$ (d) $P(3 \le X \le 5)$

3-58. Sketch the probability mass function of a binomial distribution with $n = 10$ and $p = 0.01$ and comment on the shape of the distribution.
(a) What value of X is most likely?
(b) What value of X is least likely?

3-59. The random variable X has a binomial distribution with $n = 10$ and $p = 0.01$. Determine the following probabilities.
(a) $P(X = 5)$ (b) $P(X = 2)$
(c) $P(X = 9)$ (d) $P(3 \le X \le 5)$

3-60. Determine the cumulative distribution function of a binomial random variable with $n = 3$ and $p = 1/2$.

3-61. Determine the cumulative distribution function of a binomial random variable with $n = 3$ and $p = 1/4$.

3-62. An electronic product contains 40 integrated circuits. The probability that any integrated circuit is defective is 0.01, and the integrated circuits are independent. The product operates only if there are no defective integrated circuits. What is the probability that the product operates?

3-63. Let X denote the number of bits received in error in a digital communication channel, and assume that X is a binomial random variable with $p = 0.001$. If 1000 bits are transmitted, determine the following:
(a) $P(X = 1)$ (b) $P(X = 1)$
(c) $P(X = 2)$ (d) mean and variance of X

3-64. The phone lines to an airline reservation system are occupied 40% of the time. Assume that the events that the lines are occupied on successive calls are independent. Assume that 10 calls are placed to the airline.
(a) What is the probability that for exactly three calls the lines are occupied?
(b) What is the probability that for at least one call the lines are not occupied?
(c) What is the expected number of calls in which the lines are all occupied?

3-65. Batches that consist of 50 coil springs from a production process are checked for conformance to customer requirements. The mean number of nonconforming coil springs in a batch is 5. Assume that the number of nonconforming springs in a batch, denoted as X, is a binomial random variable.
(a) What are n and p?
(b) What is $P(X = 2)$?
(c) What is $P(X = 49)$?

3-66. **A statistical process control chart example.** Samples of 20 parts from a metal punching process are selected every hour. Typically, 1% of the parts require rework. Let X denote the number of parts in the sample of 20 that require rework. A process problem is suspected if X exceeds its mean by more than three standard deviations.
(a) If the percentage of parts that require rework remains at 1%, what is the probability that X exceeds its mean by more than three standard deviations?
(b) If the rework percentage increases to 4%, what is the probability that X exceeds 1?
(c) If the rework percentage increases to 4%, what is the probability that X exceeds 1 in at least one of the next five hours of samples?

3-67. Because not all airline passengers show up for their reserved seat, an airline sells 125 tickets for a flight that holds only 120 passengers. The probability that a passenger does not show up is 0.10, and the passengers behave independently.
(a) What is the probability that every passenger who shows up can take the flight?
(b) What is the probability that the flight departs with empty seats?

3-68. This exercise illustrates that poor quality can affect schedules and costs. A manufacturing process has 100 customer orders to fill. Each order requires one component part that is purchased from a supplier. However, typically, 2% of the components are identified as defective, and the components can be assumed to be independent.
(a) If the manufacturer stocks 100 components, what is the probability that the 100 orders can be filled without reordering components?

(b) If the manufacturer stocks 102 components, what is the probability that the 100 orders can be filled without reordering components?
(c) If the manufacturer stocks 105 components, what is the probability that the 100 orders can be filled without reordering components?

3-69. A multiple choice test contains 25 questions, each with four answers. Assume a student just guesses on each question.
(a) What is the probability that the student answers more than 20 questions correctly?
(b) What is the probability the student answers less than 5 questions correctly?

3-70. A particularly long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial.
(a) Over five mornings, what is the probability that the light is green on exactly one day?
(b) Over 20 mornings, what is the probability that the light is green on exactly four days?
(c) Over 20 mornings, what is the probability that the light is green on more than four days?

3-7 GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

3-7.1 Geometric Distribution

Consider a random experiment that is closely related to the one used in the definition of a binomial distribution. Again, assume a series of Bernoulli trials (independent trials with constant probability p of a success on each trial). However, instead of a fixed number of trials, trials are conducted until a success is obtained. Let the random variable X denote the number of trials until the first success. In Example 3-5, successive wafers are analyzed until a large particle is detected. Then, X is the number of wafers analyzed. In the transmission of bits, X might be the number of bits transmitted until an error occurs.

EXAMPLE 3-20

The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume the transmissions are independent events, and let the random variable X denote the number of bits transmitted *until* the first error.

Then, $P(X = 5)$ is the probability that the first four bits are transmitted correctly and the fifth bit is in error. This event can be denoted as $\{OOOOE\}$, where O denotes an okay bit. Because the trials are independent and the probability of a correct transmission is 0.9,

$$P(X = 5) = P(OOOOE) = 0.9^4 0.1 = 0.066$$

Note that there is some probability that X will equal any integer value. Also, if the first trial is a success, $X = 1$. Therefore, the range of X is $1, 2, 3, \ldots$, that is, all positive integers.

Definition

In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the first success. Then X is a **geometric random variable** with parameter $0 < p < 1$ and

$$f(x) = (1 - p)^{x-1} p \qquad x = 1, 2, \ldots \tag{3-9}$$

Examples of the probability mass functions for geometric random variables are shown in Fig. 3-9. Note that the height of the line at x is $(1 - p)$ times the height of the line at $x - 1$. That is, the probabilities decrease in a geometric progression. The distribution acquires its name from this result.

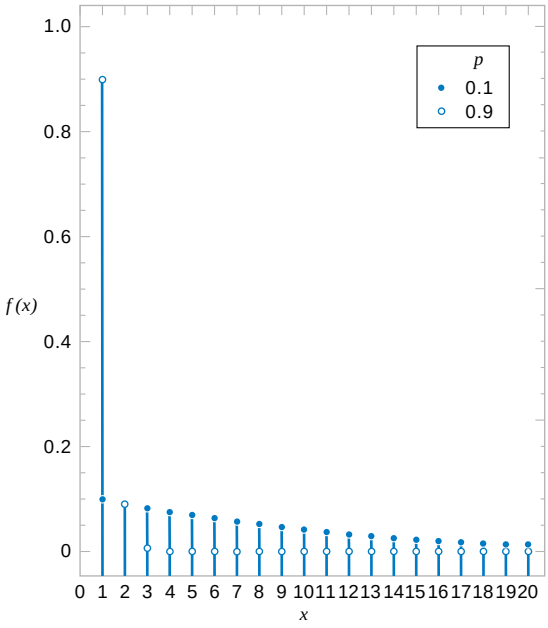


Figure 3-9 Geometric distributions for selected values of the parameter p.

EXAMPLE 3-21

The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, what is the probability that exactly 125 wafers need to be analyzed before a large particle is detected?

Let X denote the number of samples analyzed until a large particle is detected. Then X is a geometric random variable with $p = 0.01$. The requested probability is

$$P(X = 125) = (0.99)^{124}0.01 = 0.0029$$

The derivation of the mean and variance of a geometric random variable is left as an exercise. Note that $\sum_{k=1}^{\infty} k(1-p)^{k-1}p$ can be shown to equal $1/p$. The results are as follows.

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \quad \text{and} \quad \sigma^2 = V(X) = (1-p)/p^2 \tag{3-10}$$

EXAMPLE 3-22

Consider the transmission of bits in Example 3-20. Here, $p = 0.1$. The mean number of transmissions until the first error is $1/0.1 = 10$. The standard deviation of the number of transmissions before the first error is

$$\sigma = \sqrt{(0.9)/(0.1)^2} = 9.49$$

Lack of Memory Property

A geometric random variable has been defined as the number of trials until the first success. However, because the trials are independent, the count of the number of trials until the next

success can be started at any trial without changing the probability distribution of the random variable. For example, in the transmission of bits, if 100 bits are transmitted, the probability that the first error, after bit 100, occurs on bit 106 is the probability that the next six outcomes are *OOOOOE*. This probability is $(0.9)^5(0.1) = 0.059$, which is identical to the probability that the initial error occurs on bit 6.

The implication of using a geometric model is that the system presumably will not wear out. The probability of an error remains constant for all transmissions. In this sense, the geometric distribution is said to lack any memory. The **lack of memory property** will be discussed again in the context of an exponential random variable in Chapter 4.

EXAMPLE 3-23

In Example 3-20, the probability that a bit is transmitted in error is equal to 0.1. Suppose 50 bits have been transmitted. The mean number of bits until the next error is $1/0.1 = 10$—the same result as the mean number of bits until the first error.

3-7.2 Negative Binomial Distribution

A generalization of a geometric distribution in which the random variable is the number of Bernoulli trials required to obtain r successes results in the **negative binomial distribution**.

EXAMPLE 3-24

As in Example 3-20, suppose the probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume the transmissions are independent events, and let the random variable X denote the number of bits transmitted until the *fourth* error.

Then, X has a negative binomial distribution with $r = 4$. Probabilities involving X can be found as follows. The $P(X = 10)$ is the probability that exactly three errors occur in the first nine trials and then trial 10 results in the fourth error. The probability that exactly three errors occur in the first nine trials is determined from the binomial distribution to be

$$\binom{9}{3}(0.1)^3(0.9)^6$$

Because the trials are independent, the probability that exactly three errors occur in the first 9 trials and trial 10 results in the fourth error is the product of the probabilities of these two events, namely,

$$\binom{9}{3}(0.1)^3(0.9)^6(0.1) = \binom{9}{3}(0.1)^4(0.9)^6$$

The previous result can be generalized as follows.

Definition

In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until r successes occur. Then X is a **negative binomial random variable** with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$, and

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r \qquad x = r, r+1, r+2, \ldots \qquad (3\text{-}11)$$

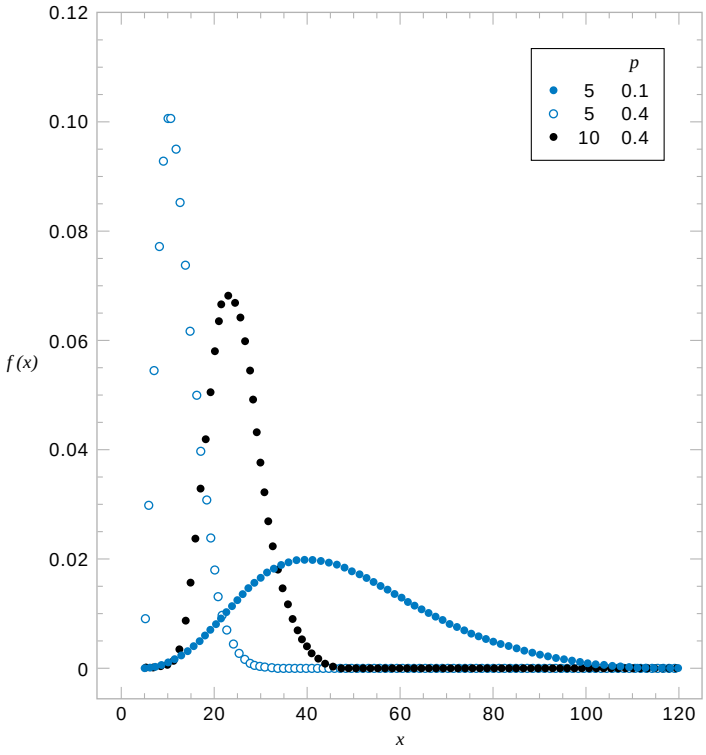


Figure 3-10 Negative binomial distributions for selected values of the parameters r and p.

Because at least r trials are required to obtain r successes, the range of X is from r to □. In the special case that r □ 1, a negative binomial random variable is a geometric random variable. Selected negative binomial distributions are illustrated in Fig. 3-10.

The lack of memory property of a geometric random variable implies the following. Let X denote the total number of trials required to obtain r successes. Let X_1 denote the number of trials required to obtain the first success, let X_2 denote the number of extra trials required to obtain the second success, let X_3 denote the number of extra trials to obtain the third success, and so forth. Then, the total number of trials required to obtain r successes is X □ X_1 □ X_2 □ p □ X_r. Because of the lack of memory property, each of the random variables X_1, X_2, p , X_r has a geometric distribution with the same value of p. Consequently, a negative binomial random variable can be interpreted as the sum of r geometric random variables. This concept is illustrated in Fig. 3-11.

Recall that a binomial random variable is a count of the number of successes in n Bernoulli trials. That is, the number of trials is predetermined, and the number of successes is random. A negative binomial random variable is a count of the number of trials required to

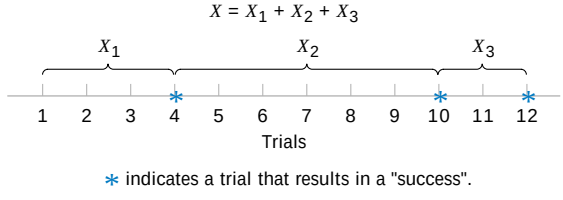


Figure 3-11 Negative binomial random variable represented as a sum of geometric random variables.

obtain r successes. That is, the number of successes is predetermined, and the number of trials is random. In this sense, a negative binomial random variable can be considered the opposite, or negative, of a binomial random variable.

The description of a negative binomial random variable as a sum of geometric random variables leads to the following results for the mean and variance. Sums of random variables are studied in Chapter 5.

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = r/p \quad \text{and} \quad \sigma^2 = V(X) = r(1-p)/p^2 \tag{3-12}$$

EXAMPLE 3-25

A Web site contains three identical computer servers. Only one is used to operate the site, and the other two are spares that can be activated in case the primary system fails. The probability of a failure in the primary computer (or any activated spare system) from a request for service is 0.0005. Assuming that each request represents an independent trial, what is the mean number of requests until failure of all three servers?

Let X denote the number of requests until all three servers fail, and let X_1, X_2, and X_3 denote the number of requests before a failure of the first, second, and third servers used, respectively. Now, $X = X_1 + X_2 + X_3$. Also, the requests are assumed to comprise independent trials with constant probability of failure $p = 0.0005$. Furthermore, a spare server is not affected by the number of requests before it is activated. Therefore, X has a negative binomial distribution with $p = 0.0005$ and $r = 3$. Consequently,

$$E(X) = 3/0.0005 = 6000 \text{ requests}$$

What is the probability that all three servers fail within five requests? The probability is $P(X = 5)$ and

$$\begin{aligned} P(X = 5) &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= 0.0005^3 + \binom{3}{2} 0.0005^3(0.9995)^2 + \binom{4}{2} 0.0005^3(0.9995)^2 \\ &= 1.25 \times 10^{-10} + 3.75 \times 10^{-10} + 7.49 \times 10^{-10} \\ &= 1.249 \times 10^{-9} \end{aligned}$$

EXERCISES FOR SECTION 3-7

3-71. Suppose the random variable X has a geometric distribution with $p = 0.5$. Determine the following probabilities:

(a) $P(X = 1)$ (b) $P(X = 4)$
(c) $P(X = 8)$ (d) $P(X = 2)$
(e) $P(X = 2)$

3-72. Suppose the random variable X has a geometric distribution with a mean of 2.5. Determine the following probabilities:

(a) $P(X = 1)$ (b) $P(X = 4)$
(c) $P(X = 5)$ (d) $P(X = 3)$
(e) $P(X = 3)$

3-73. The probability of a successful optical alignment in the assembly of an optical data storage product is 0.8. Assume the trials are independent.
(a) What is the probability that the first successful alignment requires exactly four trials?
(b) What is the probability that the first successful alignment requires at most four trials?
(c) What is the probability that the first successful alignment requires at least four trials?

3-74. In a clinical study, volunteers are tested for a gene that has been found to increase the risk for a disease. The probability that a person carries the gene is 0.1.
(a) What is the probability 4 or more people will have to be tested before 2 with the gene are detected?
(b) How many people are expected to be tested before 2 with the gene are detected?

3-75. Assume that each of your calls to a popular radio station has a probability of 0.02 of connecting, that is, of not obtaining a busy signal. Assume that your calls are independent.
(a) What is the probability that your first call that connects is your tenth call?
(b) What is the probability that it requires more than five calls for you to connect?
(c) What is the mean number of calls needed to connect?

3-76. In Exercise 3-70, recall that a particularly long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial.
(a) What is the probability that the first morning that the light is green is the fourth morning that you approach it?
(b) What is the probability that the light is not green for 10 consecutive mornings?

3-77. A trading company has eight computers that it uses to trade on the New York Stock Exchange (NYSE). The probability of a computer failing in a day is 0.005, and the computers fail independently. Computers are repaired in the evening and each day is an independent trial.
(a) What is the probability that all eight computers fail in a day?
(b) What is the mean number of days until a specific computer fails?
(c) What is the mean number of days until all eight computers fail in the same day?

3-78. In Exercise 3-66, recall that 20 parts are checked each hour and that X denotes the number of parts in the sample of 20 that require rework.
(a) If the percentage of parts that require rework remains at 1%, what is the probability that hour 10 is the first sample at which X exceeds 1?
(b) If the rework percentage increases to 4%, what is the probability that hour 10 is the first sample at which X exceeds 1?
(c) If the rework percentage increases to 4%, what is the expected number of hours until X exceeds 1?

3-79. Consider a sequence of independent Bernoulli trials with $p = 0.2$.
(a) What is the expected number of trials to obtain the first success?
(b) After the eighth success occurs, what is the expected number of trials to obtain the ninth success?

3-80. Show that the probability density function of a negative binomial random variable equals the probability density function of a geometric random variable when $r = 1$. Show that the formulas for the mean and variance of a negative binomial random variable equal the corresponding results for geometric random variable when $r = 1$.

3-81. Suppose that X is a negative binomial random variable with $p = 0.2$ and $r = 4$. Determine the following:
(a) $E(X)$ (b) $P(X = 20)$
(c) $P(X = 19)$ (d) $P(X = 21)$
(e) The most likely value for X

3-82. The probability is 0.6 that a calibration of a transducer in an electronic instrument conforms to specifications for the measurement system. Assume the calibration attempts are independent. What is the probability that at most three calibration attempts are required to meet the specifications for the measurement system?

3-83. An electronic scale in an automated filling operation stops the manufacturing line after three underweight packages are detected. Suppose that the probability of an underweight package is 0.001 and each fill is independent.
(a) What is the mean number of fills before the line is stopped?
(b) What is the standard deviation of the number of fills before the line is stopped?

3-84. A fault-tolerant system that processes transactions for a financial services firm uses three separate computers. If the operating computer fails, one of the two spares can be immediately switched online. After the second computer fails, the last computer can be immediately switched online. Assume that the probability of a failure during any transaction is 10^{-8} and that the transactions can be considered to be independent events.
(a) What is the mean number of transactions before all computers have failed?
(b) What is the variance of the number of transactions before all computers have failed?

3-85. Derive the expressions for the mean and variance of a geometric random variable with parameter p. (Formulas for infinite series are required.)

3-8 HYPERGEOMETRIC DISTRIBUTION

In Example 3-8, a day's production of 850 manufactured parts contains 50 parts that do not conform to customer requirements. Two parts are selected at random, without replacement from the day's production. That is, selected units are not replaced before the next selection is made. Let A and B denote the events that the first and second parts are nonconforming, respectively. In Chapter 2, we found $P(B \mid A) = 49/849$ and $P(A) = 50/850$. Consequently, knowledge that the first part is nonconforming suggests that it is less likely that the second part selected is nonconforming.

This experiment is fundamentally different from the examples based on the binomial distribution. In this experiment, the trials are not independent. Note that, in the unusual case that each unit selected is replaced before the next selection, the trials are independent and there is a constant probability of a nonconforming part on each trial. Then, the number of nonconforming parts in the sample is a binomial random variable.

Let X equal the number of nonconforming parts in the sample. Then

$$P(X = 0) = P(\text{both parts conform}) = (1800/850)(1799/849) = 0.886$$

$$P(X = 1) = P(\text{first part selected conforms and the second part selected does not, or the first part selected does not and the second part selected conforms})$$

$$= (1800/850)(150/849) + (150/850)(1800/849) = 0.111$$

$$P(X = 2) = P(\text{both parts do not conform}) = (150/850)(149/849) = 0.003$$

As in this example, samples are often selected without replacement. Although probabilities can be determined by the reasoning used in the example above, a general formula for computing probabilities when samples are selected without replacement is quite useful. The counting rules presented in Section 2-1.4, part of the CD material for Chapter 2, can be used to justify the formula given below.

Definition

A set of N objects contains

K objects classified as successes

$N - K$ objects classified as failures

A sample of size n objects is selected randomly (without replacement) from the N objects, where $K \le N$ and $n \le N$.

Let the random variable X denote the number of successes in the sample. Then X is a **hypergeometric random variable** and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max(0, n - (N - K)) \text{ to } \min(K, n) \qquad (3\text{-}13)$$

The expression $\min(K, n)$ is used in the definition of the range of X because the maximum number of successes that can occur in the sample is the smaller of the sample size, n,

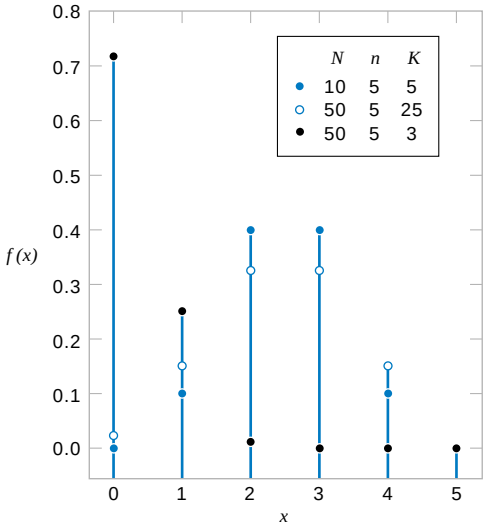


Figure 3-12 Hypergeometric distributions for selected values of parameters N, K, and n.

and the number of successes available, K. Also, if $n \square K \square N$, at least $n \square K \square N$ successes must occur in the sample. Selected hypergeometric distributions are illustrated in Fig. 3-12.

EXAMPLE 3-26

The example at the start of this section can be reanalyzed by using the general expression in the definition of a hypergeometric random variable. That is,

$$P(X \square 0) \square \frac{\binom{50}{0}\binom{800}{2}}{\binom{850}{2}} \square \frac{319600}{360825} \square 0.886$$

$$P(X \square 1) \square \frac{\binom{50}{1}\binom{800}{1}}{\binom{850}{2}} \square \frac{40000}{360825} \square 0.111$$

$$P(X \square 2) \square \frac{\binom{50}{2}\binom{800}{0}}{\binom{850}{2}} \square \frac{1225}{360825} \square 0.003$$

EXAMPLE 3-27

A batch of parts contains 100 parts from a local supplier of tubing and 200 parts from a supplier of tubing in the next state. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

Let X equal the number of parts in the sample from the local supplier. Then, X has a hypergeometric distribution and the requested probability is $P(X \leq 4)$. Consequently,

$$P(X \leq 4) = \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = 0.0119$$

What is the probability that two or more parts in the sample are from the local supplier?

$$P(X \geq 2) = \frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$
$$= 0.298 + 0.098 + 0.0119 = 0.408$$

What is the probability that at least one part in the sample is from the local supplier?

$$P(X \geq 1) = 1 - P(X \leq 0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.804$$

The mean and variance of a hypergeometric random variable can be determined from the trials that comprise the experiment. However, the trials are not independent, and so the calculations are more difficult than for a binomial distribution. The results are stated as follows.

If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1 - p)\left(\frac{N - n}{N - 1}\right) \tag{3-14}$$

where $p = K/N$.

Here p is interpreted as the proportion of successes in the set of N objects.

EXAMPLE 3-28 In the previous example, the sample size is 4. The random variable X is the number of parts in the sample from the local supplier. Then, $p = 100/300 = 1/3$. Therefore,

$$E(X) = 4(100/300) = 1.33$$

and

$$V(X) = \frac{4(11)(32)(32)}{1300} = \frac{42}{2994} = 0.88$$

For a hypergeometric random variable, $E(X)$ is similar to the mean a binomial random variable. Also, $V(X)$ differs from the result for a binomial random variable only by the term shown below.

Finite Population Correction Factor

The term in the variance of a hypergeometric random variable

$$\frac{N - n}{N - 1}$$

is called the finite population correction factor.

Sampling with replacement is equivalent to sampling from an infinite set because the proportion of success remains constant for every trial in the experiment. As mentioned previously, if sampling were done with replacement, X would be a binomial random variable and its variance would be $np(1 - p)$. Consequently, the finite population correction represents the correction to the binomial variance that results because the sampling is without replacement from the finite set of size N.

If n is small relative to N, the correction is small and the hypergeometric distribution is similar to the binomial. In this case, a binomial distribution can effectively approximate the distribution of the number of units of a specified type in the sample. A case is illustrated in Fig. 3-13.

EXAMPLE 3-29 A listing of customer accounts at a large corporation contains 1000 customers. Of these, 700 have purchased at least one of the corporation's products in the last three months. To evaluate a new product design, 50 customers are sampled at random from the corporate listing. What is

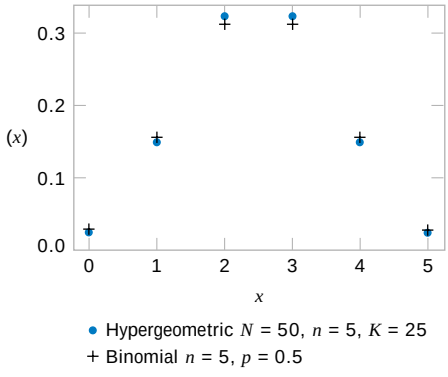


Figure 3-13 Comparison of hypergeometric and binomial distributions.

	0	1	2	3	4	5
Hypergeometric probability	0.025	0.149	0.326	0.326	0.149	0.025
Binomial probability	0.031	0.156	0.321	0.312	0.156	0.031

the probability that more than 45 of the sampled customers have purchased from the corporation in the last three months?

The sampling is without replacement. However, because the sample size of 50 is small relative to the number of customer accounts, 1000, the probability of selecting a customer who has purchased from the corporation in the last three months remains approximately constant as the customers are chosen.

For example, let A denote the event that the first customer selected has purchased from the corporation in the last three months, and let B denote the event that the second customer selected has purchased from the corporation in the last three months. Then, $P(A) = 700/1000 = 0.7$ and $P(B|A) = 699/999 = 0.6997$. That is, the trials are approximately independent.

Let X denote the number of customers in the sample who have purchased from the corporation in the last three months. Then, X is a hypergeometric random variable with $N = 1000$, $n = 50$, and $K = 700$. Consequently, $p = K/N = 0.7$. The requested probability is $P(X \le 45)$. Because the sample size is small relative to the batch size, the distribution of X can be approximated as binomial with $n = 50$ and $p = 0.7$. Using the binomial approximation to the distribution of X results in

$$P(X \le 45) = \sum_{x=46}^{50} \binom{50}{x} 0.7^x (1 - 0.7)^{50-x} = 0.00017$$

The probability from the hypergeometric distribution is 0.000166, but this requires computer software. The result agrees well with the binomial approximation.

EXERCISES FOR SECTION 3-8

3-86. Suppose X has a hypergeometric distribution with $N = 100$, $n = 4$, and $K = 20$. Determine the following:
(a) $P(X = 1)$ (b) $P(X = 6)$
(c) $P(X = 4)$ (d) Determine the mean and variance of X.

3-87. Suppose X has a hypergeometric distribution with $N = 20$, $n = 4$, and $K = 4$. Determine the following:
(a) $P(X = 1)$ (b) $P(X = 4)$
(c) $P(X = 2)$ (d) Determine the mean and variance of X.

3-88. Suppose X has a hypergeometric distribution with $N = 10$, $n = 3$, and $K = 4$. Sketch the probability mass function of X.

3-89. Determine the cumulative distribution function for X in Exercise 3-88.

3-90. A lot of 75 washers contains 5 in which the variability in thickness around the circumference of the washer is unacceptable. A sample of 10 washers is selected at random, without replacement.
(a) What is the probability that none of the unacceptable washers is in the sample?
(b) What is the probability that at least one unacceptable washer is in the sample?
(c) What is the probability that exactly one unacceptable washer is in the sample?
(d) What is the mean number of unacceptable washers in the sample?

3-91. A company employs 800 men under the age of 55. Suppose that 30% carry a marker on the male chromosome that indicates an increased risk for high blood pressure.
(a) If 10 men in the company are tested for the marker in this chromosome, what is the probability that exactly 1 man has the marker?
(b) If 10 men in the company are tested for the marker in this chromosome, what is the probability that more than 1 has the marker?

3-92. Printed circuit cards are placed in a functional test after being populated with semiconductor chips. A lot contains 140 cards, and 20 are selected without replacement for functional testing.
(a) If 20 cards are defective, what is the probability that at least 1 defective card is in the sample?
(b) If 5 cards are defective, what is the probability that at least 1 defective card appears in the sample?

3-93. Magnetic tape is slit into half-inch widths that are wound into cartridges. A slitter assembly contains 48 blades. Five blades are selected at random and evaluated each day for

sharpness. If any dull blade is found, the assembly is replaced with a newly sharpened set of blades.

(a) If 10 of the blades in an assembly are dull, what is the probability that the assembly is replaced the first day it is evaluated?

(b) If 10 of the blades in an assembly are dull, what is the probability that the assembly is not replaced until the third day of evaluation? [*Hint:* Assume the daily decisions are independent, and use the geometric distribution.]

(c) Suppose on the first day of evaluation, two of the blades are dull, on the second day of evaluation six are dull, and on the third day of evaluation, ten are dull. What is the probability that the assembly is not replaced until the third day of evaluation? [*Hint:* Assume the daily decisions are independent. However, the probability of replacement changes every day.]

3-94. A state runs a lottery in which 6 numbers are randomly selected from 40, without replacement. A player chooses 6 numbers before the state's sample is selected.

(a) What is the probability that the 6 numbers chosen by a player match all 6 numbers in the state's sample?

(b) What is the probability that 5 of the 6 numbers chosen by a player appear in the state's sample?

(c) What is the probability that 4 of the 6 numbers chosen by a player appear in the state's sample?

(d) If a player enters one lottery each week, what is the expected number of weeks until a player matches all 6 numbers in the state's sample?

3-95. Continuation of Exercises 3-86 and 3-87.

(a) Calculate the finite population corrections for Exercises 3-86 and 3-87. For which exercise should the binomial approximation to the distribution of X be better?

(b) For Exercise 3-86, calculate $P(X \le 1)$ and $P(X \le 4)$ assuming that X has a binomial distribution and compare these results to results derived from the hypergeometric distribution.

(c) For Exercise 3-87, calculate $P(X \le 1)$ and $P(X \le 4)$ assuming that X has a binomial distribution and compare these results to the results derived from the hypergeometric distribution.

3-96. Use the binomial approximation to the hypergeometric distribution to approximate the probabilities in Exercise 3-92. What is the finite population correction in this exercise?

3-9 POISSON DISTRIBUTION

We introduce the Poisson distribution with an example.

EXAMPLE 3-30

Consider the transmission of n bits over a digital communication channel. Let the random variable X equal the number of bits in error. When the probability that a bit is in error is constant and the transmissions are independent, X has a binomial distribution. Let p denote the probability that a bit is in error. Let $\lambda = pn$. Then, $E(X) = pn = \lambda$ and

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Now, suppose that the number of bits transmitted increases and the probability of an error decreases exactly enough that pn remains equal to a constant. That is, n increases and p decreases accordingly, such that $E(X) = \lambda$ remains constant. Then, with some work, it can be shown that

$$\lim_{n \to \infty} P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

Also, because the number of bits transmitted tends to infinity, the number of errors can equal any nonnegative integer. Therefore, the range of X is the integers from zero to infinity.

The distribution obtained as the limit in the above example is more useful than the derivation above implies. The following example illustrates the broader applicability.

EXAMPLE 3-31

Flaws occur at random along the length of a thin copper wire. Let X denote the random variable that counts the number of flaws in a length of L millimeters of wire and suppose that the average number of flaws in L millimeters is .

The probability distribution of X can be found by reasoning in a manner similar to the previous example. Partition the length of wire into n subintervals of small length, say, 1 micrometer each. If the subinterval chosen is small enough, the probability that more than one flaw occurs in the subinterval is negligible. Furthermore, we can interpret the assumption that flaws occur at random to imply that every subinterval has the same probability of containing a flaw, say, p. Finally, if we assume that the probability that a subinterval contains a flaw is independent of other subintervals, we can model the distribution of X as approximately a binomial random variable. Because

$$E(X) = np$$

we obtain

$$p = /n$$

That is, the probability that a subinterval contains a flaw is $/n$. With small enough subintervals, n is very large and p is very small. Therefore, the distribution of X is obtained as in the previous example.

Example 3-31 can be generalized to include a broad array of random experiments. The interval that was partitioned was a length of wire. However, the same reasoning can be applied to any interval, including an interval of time, an area, or a volume. For example, counts of (1) particles of contamination in semiconductor manufacturing, (2) flaws in rolls of textiles, (3) calls to a telephone exchange, (4) power outages, and (5) atomic particles emitted from a specimen have all been successfully modeled by the probability mass function in the following definition.

Definition

Given an interval of real numbers, assume counts occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

(1) the probability of more than one count in a subinterval is zero,

(2) the probability of one count in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and

(3) the count in each subinterval is independent of other subintervals, the random experiment is called a **Poisson process.**

The random variable X that equals the number of counts in the interval is a **Poisson random variable** with parameter $0 <$, and the probability mass function of X is

$$f(x) = \frac{e^{-} \quad x}{x!} \quad x = 0, 1, 2, \ldots \tag{3-15}$$

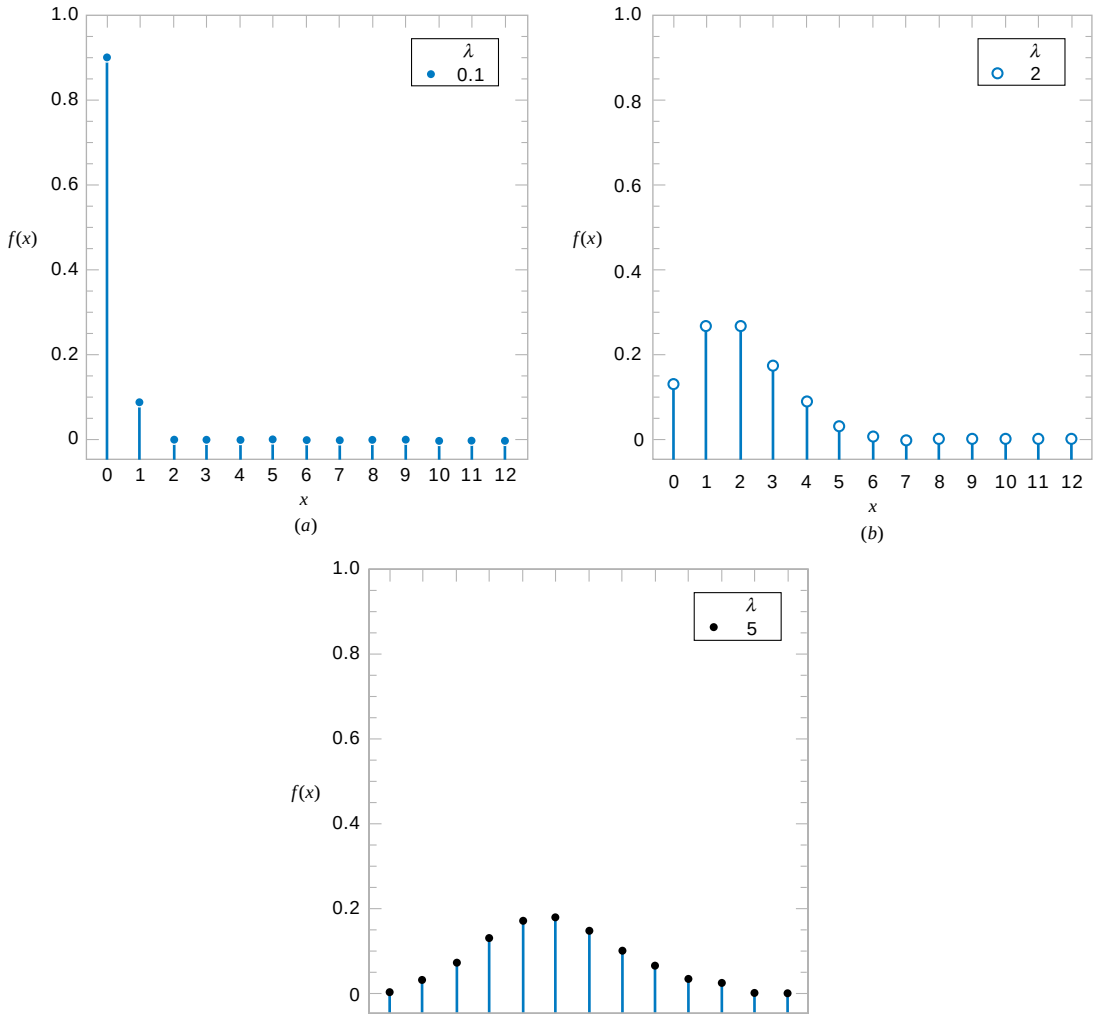


Figure 3-14 Poisson distributions for selected values of the parameters.

Historically, the term *process* has been used to suggest the observation of a system over time. In our example with the copper wire, we showed that the Poisson distribution could also apply to intervals such as lengths. Figure 3-14 provides graphs of selected Poisson distributions.

It is important to **use consistent units** in the calculation of probabilities, means, and variances involving Poisson random variables. The following example illustrates unit conversions. For example, if the

average number of flaws per millimeter of wire is 3.4, then the
average number of flaws in 10 millimeters of wire is 34, and the
average number of flaws in 100 millimeters of wire is 340.

If a Poisson random variable represents the number of counts in some interval, the mean of the random variable must equal the expected number of counts in the same length of interval.

EXAMPLE 3-32

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter. Determine the probability of exactly 2 flaws in 1 millimeter of wire.

Let X denote the number of flaws in 1 millimeter of wire. Then, $E(X) = 2.3$ flaws and

$$P(X = 2) = \frac{e^{-2.3}2.3^2}{2!} = 0.265$$

Determine the probability of 10 flaws in 5 millimeters of wire. Let X denote the number of flaws in 5 millimeters of wire. Then, X has a Poisson distribution with

$$E(X) = 5 \text{ mm} \times 2.3 \text{ flaws/mm} = 11.5 \text{ flaws}$$

Therefore,

$$P(X = 10) = e^{-11.5}\frac{11.5^{10}}{10!} = 0.113$$

Determine the probability of at least 1 flaw in 2 millimeters of wire. Let X denote the number of flaws in 2 millimeters of wire. Then, X has a Poisson distribution with

$$E(X) = 2 \text{ mm} \times 2.3 \text{ flaws/mm} = 4.6 \text{ flaws}$$

Therefore,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4.6} = 0.9899$$

EXAMPLE 3-33

Contamination is a problem in the manufacture of optical storage disks. The number of particles of contamination that occur on an optical disk has a Poisson distribution, and the average number of particles per centimeter squared of media surface is 0.1. The area of a disk under study is 100 squared centimeters. Find the probability that 12 particles occur in the area of a disk under study.

Let X denote the number of particles in the area of a disk under study. Because the mean number of particles is 0.1 particles per cm^2

$$E(X) = 100 \text{ cm}^2 \times 0.1 \text{ particles/cm}^2 = 10 \text{ particles}$$

Therefore,

$$P(X = 12) = \frac{e^{-10}10^{12}}{12!} = 0.095$$

The probability that zero particles occur in the area of the disk under study is

$$P(X = 0) = e^{-10} = 4.54 \times 10^{-5}$$

Determine the probability that 12 or fewer particles occur in the area of the disk under study. The probability is

$$P(X \leq 12) = P(X = 0) + P(X = 1) + \cdots + P(X = 12) = \sum_{i=0}^{12} \frac{e^{-10}10^i}{i!}$$

Because this sum is tedious to compute, many computer programs calculate cumulative Poisson probabilities. From one such program, $P(X \leq 12) = 0.791$.

The derivation of the mean and variance of a Poisson random variable is left as an exercise. The results are as follows.

If X is a Poisson random variable with parameter λ, then

$$\mu = E(X) = \lambda \quad \text{and} \quad \sigma^2 = V(X) = \lambda \tag{3-16}$$

The mean and variance of a Poisson random variable are equal. For example, if particle counts follow a Poisson distribution with a mean of 25 particles per square centimeter, the variance is also 25 and the standard deviation of the counts is 5 per square centimeter. Consequently, information on the variability is very easily obtained. Conversely, if the variance of count data is much greater than the mean of the same data, the Poisson distribution is not a good model for the distribution of the random variable.

EXERCISES FOR SECTION 3-9

3-97. Suppose X has a Poisson distribution with a mean of 4. Determine the following probabilities:
(a) $P(X = 0)$ (b) $P(X = 2)$
(c) $P(X = 4)$ (d) $P(X = 8)$

3-98. Suppose X has a Poisson distribution with a mean of 0.4. Determine the following probabilities:
(a) $P(X = 0)$ (b) $P(X = 2)$
(c) $P(X = 4)$ (d) $P(X = 8)$

3-99. Suppose that the number of customers that enter a bank in an hour is a Poisson random variable, and suppose that $P(X = 0) = 0.05$. Determine the mean and variance of X.

3-100. The number of telephone calls that arrive at a phone exchange is often modeled as a Poisson random variable. Assume that on the average there are 10 calls per hour.
(a) What is the probability that there are exactly 5 calls in one hour?
(b) What is the probability that there are 3 or less calls in one hour?
(c) What is the probability that there are exactly 15 calls in two hours?
(d) What is the probability that there are exactly 5 calls in 30 minutes?

3-101. The number of flaws in bolts of cloth in textile manufacturing is assumed to be Poisson distributed with a mean of 0.1 flaw per square meter.
(a) What is the probability that there are two flaws in 1 square meter of cloth?
(b) What is the probability that there is one flaw in 10 square meters of cloth?
(c) What is the probability that there are no flaws in 20 square meters of cloth?
(d) What is the probability that there are at least two flaws in 10 square meters of cloth?

3-102. When a computer disk manufacturer tests a disk, it writes to the disk and then tests it using a certifier. The certifier counts the number of missing pulses or errors. The number of errors on a test area on a disk has a Poisson distribution with $\lambda = 0.2$.
(a) What is the expected number of errors per test area?
(b) What percentage of test areas have two or fewer errors?

3-103. The number of cracks in a section of interstate highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile.
(a) What is the probability that there are no cracks that require repair in 5 miles of highway?
(b) What is the probability that at least one crack requires repair in $\frac{1}{2}$ mile of highway?
(c) If the number of cracks is related to the vehicle load on the highway and some sections of the highway have a heavy load of vehicles whereas other sections carry a light load, how do you feel about the assumption of a Poisson distribution for the number of cracks that require repair?

3-104. The number of failures for a cytogenics machine from contamination in biological samples is a Poisson random variable with a mean of 0.01 per 100 samples.
(a) If the lab usually processes 500 samples per day, what is the expected number of failures per day?

(b) What is the probability that the machine will not fail during a study that includes 500 participants? (Assume one sample per participant.)

3-105. The number of surface flaws in plastic panels used in the interior of automobiles has a Poisson distribution with a mean of 0.05 flaw per square foot of plastic panel. Assume an automobile interior contains 10 square feet of plastic panel.

(a) What is the probability that there are no surface flaws in an auto's interior?

(b) If 10 cars are sold to a rental company, what is the probability that none of the 10 cars has any surface flaws?

(c) If 10 cars are sold to a rental company, what is the probability that at most one car has any surface flaws?

3-106. The number of failures of a testing instrument from contamination particles on the product is a Poisson random variable with a mean of 0.02 failure per hour.

(a) What is the probability that the instrument does not fail in an 8-hour shift?

(b) What is the probability of at least one failure in a 24-hour day?

Supplemental Exercises

3-107. A shipment of chemicals arrives in 15 totes. Three of the totes are selected at random, without replacement, for an inspection of purity. If two of the totes do not conform to purity requirements, what is the probability that at least one of the nonconforming totes is selected in the sample?

3-108. The probability that your call to a service line is answered in less than 30 seconds is 0.75. Assume that your calls are independent.

(a) If you call 10 times, what is the probability that exactly 9 of your calls are answered within 30 seconds?

(b) If you call 20 times, what is the probability that at least 16 calls are answered in less than 30 seconds?

(c) If you call 20 times, what is the mean number of calls that are answered in less than 30 seconds?

3-109. Continuation of Exercise 3-108.

(a) What is the probability that you must call four times to obtain the first answer in less than 30 seconds?

(b) What is the mean number of calls until you are answered in less than 30 seconds?

3-110. Continuation of Exercise 3-109.

(a) What is the probability that you must call six times in order for two of your calls to be answered in less than 30 seconds?

(b) What is the mean number of calls to obtain two answers in less than 30 seconds?

3-111. The number of messages sent to a computer bulletin board is a Poisson random variable with a mean of 5 messages per hour.

(a) What is the probability that 5 messages are received in 1 hour?

(b) What is the probability that 10 messages are received in 1.5 hours?

(c) What is the probability that less than two messages are received in one-half hour?

3-112. A Web site is operated by four identical computer servers. Only one is used to operate the site; the others are spares that can be activated in case the active server fails. The probability that a request to the Web site generates a failure in the active server is 0.0001. Assume that each request is an independent trial. What is the mean time until failure of all four computers?

3-113. The number of errors in a textbook follow a Poisson distribution with a mean of 0.01 error per page. What is the probability that there are three or less errors in 100 pages?

3-114. The probability that an individual recovers from an illness in a one-week time period without treatment is 0.1. Suppose that 20 independent individuals suffering from this illness are treated with a drug and 4 recover in a one-week time period. If the drug has no effect, what is the probability that 4 or more people recover in a one-week time period?

3-115. Patient response to a generic drug to control pain is scored on a 5-point scale, where a 5 indicates complete relief. Historically the distribution of scores is

1	2	3	4	5
0.05	0.1	0.2	0.25	0.4

Two patients, assumed to be independent, are each scored.

(a) What is the probability mass function of the total score?

(b) What is the probability mass function of the average score?

3-116. In a manufacturing process that laminates several ceramic layers, 1% of the assemblies are defective. Assume that the assemblies are independent.

(a) What is the mean number of assemblies that need to be checked to obtain five defective assemblies?

(b) What is the standard deviation of the number of assemblies that need to be checked to obtain five defective assemblies?

3-117. Continuation of Exercise 3-116. Determine the minimum number of assemblies that need to be checked so that the probability of at least one defective assembly exceeds 0.95.

3-118. Determine the constant c so that the following function is a probability mass function: $f(x) = cx$ for $x = 1, 2, 3, 4$.

3-119. A manufacturer of a consumer electronics product expects 2% of units to fail during the warranty period. A sample of 500 independent units is tracked for warranty performance.

(a) What is the probability that none fails during the warranty period?

(b) What is the expected number of failures during the warranty period?

(c) What is the probability that more than two units fail during the warranty period?

3-120. Messages that arrive at a service center for an information systems manufacturer have been classified on the basis

of the number of keywords (used to help route messages) and the type of message, either email or voice. Also, 70% of the messages arrive via email and the rest are voice.

number of keywords	0	1	2	3	4
email	0.1	0.1	0.2	0.4	0.2
voice	0.3	0.4	0.2	0.1	0

Determine the probability mass function of the number of keywords in a message.

3-121. The random variable X has the following probability distribution:

x	2	3	5	8
probability	0.2	0.4	0.3	0.1

Determine the following:

(a) $P(X \leq 3)$ (b) $P(X \leq 2.5)$

(c) $P(2.7 \leq X \leq 5.1)$ (d) $E(X)$

(e) $V(X)$

3-122. Determine the probability mass function for the random variable with the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 2 \\ 0.2 & 2 \leq x < 5.7 \\ 0.5 & 5.7 \leq x < 6.5 \\ 0.8 & 6.5 \leq x < 8.5 \\ 1 & 8.5 \leq x \end{cases}$$

3-123. Each main bearing cap in an engine contains four bolts. The bolts are selected at random, without replacement, from a parts bin that contains 30 bolts from one supplier and 70 bolts from another.

(a) What is the probability that a main bearing cap contains all bolts from the same supplier?

(b) What is the probability that exactly three bolts are from the same supplier?

3-124. Assume the number of errors along a magnetic recording surface is a Poisson random variable with a mean of one error every 10^5 bits. A sector of data consists of 4096 eight-bit bytes.

(a) What is the probability of more than one error in a sector?

(b) What is the mean number of sectors until an error is found?

3-125. An installation technician for a specialized communication system is dispatched to a city only when three or more orders have been placed. Suppose orders follow a Poisson distribution with a mean of 0.25 per week for a city with a population of 100,000 and suppose your city contains a population of 800,000.

(a) What is the probability that a technician is required after a one-week period?

(b) If you are the first one in the city to place an order, what is the probability that you have to wait more than two weeks from the time you place your order until a technician is dispatched?

3-126. From 500 customers, a major appliance manufacturer will randomly select a sample without replacement. The company estimates that 25% of the customers will provide useful data. If this estimate is correct, what is the probability mass function of the number of customers that will provide useful data?

(a) Assume that the company samples 5 customers.

(b) Assume that the company samples 10 customers.

3-127. It is suspected that some of the totes containing chemicals purchased from a supplier exceed the moisture content target. Samples from 30 totes are to be tested for moisture content. Assume that the totes are independent. Determine the proportion of totes from the supplier that must exceed the moisture content target so that the probability is 0.90 that at least one tote in the sample of 30 fails the test.

3-128. Messages arrive to a computer server according to a Poisson distribution with a mean rate of 10 per hour. Determine the length of an interval of time such that the probability that no messages arrive during this interval is 0.90.

3-129. Flaws occur in the interior of plastic used for automobiles according to a Poisson distribution with a mean of 0.02 flaw per panel.

(a) If 50 panels are inspected, what is the probability that there are no flaws?

(b) What is the expected number of panels that need to be inspected before a flaw is found?

(c) If 50 panels are inspected, what is the probability that the number of panels that have one or more flaws is less than or equal to 2?

MIND-EXPANDING EXERCISES

3-130. Derive the mean and variance of a hypergeometric random variable (difficult exercise).

3-131. Show that the function $f(x)$ in Example 3-5 satisfies the properties of a probability mass function by summing the infinite series.

3-132. Derive the formula for the mean and standard deviation of a discrete uniform random variable over the range of integers $a, a \square 1, \mathsf{p}, b$.

3-133. A company performs inspection on shipments from suppliers in order to defect nonconforming products. Assume a lot contains 1000 items and 1% are nonconforming. What sample size is needed so that the probability of choosing at least one nonconforming item in the sample is at least 0.90? Assume the binomial approximation to the hypergeometric distribution is adequate.

3-134. A company performs inspection on shipments from suppliers in order to detect nonconforming products. The company's policy is to use a sample size that is always 10% of the lot size. Comment on the effectiveness of this policy as a general rule for all sizes of lots.

3-135. Surface flaws in automobile exterior panels follow a Poisson distribution with a mean of 0.1 flaw per panel. If 100 panels are checked, what is the probability that fewer than five panels have any flaws?

3-136. A large bakery can produce rolls in lots of either 0, 1000, 2000, or 3000 per day. The production cost per item is \$0.10. The demand varies randomly according to the following distribution:

demand for rolls	0	1000	2000	3000
probability of demand	0.3	0.2	0.3	0.2

Every roll for which there is a demand is sold for \$0.30. Every roll for which there is no demand is sold in a secondary market for \$0.05. How many rolls should the bakery produce each day to maximize the mean profit?

3-137. A manufacturer stocks components obtained from a supplier. Suppose that 2% of the components are defective and that the defective components occur independently. How many components must the manufacturer have in stock so that the probability that 100 orders can be completed without reordering components is at least 0.95?

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Bernoulli trial
Binomial distribution
Cumulative probability distribution function-discrete random variable
Discrete uniform distribution
Expected value of a function of a random variable
Finite population correction factor
Geometric distribution
Hypergeometric distribution
Lack of memory property-discrete random variable
Mean-discrete random variable
Mean-function of a discrete random variable
Negative binomial distribution
Poisson distribution
Poisson process
Probability distribution-discrete random variable
Probability mass function
Standard deviation-discrete random variable
Variance-discrete random variable

4 Continuous Random Variables and Probability Distributions

CHAPTER OUTLINE

4-1 CONTINUOUS RANDOM VARIABLES

4-2 PROBABILITY DISTRIBUTIONS AND PROBABILITY DENSITY FUNCTIONS

4-3 CUMULATIVE DISTRIBUTION FUNCTIONS

4-4 MEAN AND VARIANCE OF A CONTINUOUS RANDOM VARIABLE

4-5 CONTINUOUS UNIFORM DISTRIBUTION

4-6 NORMAL DISTRIBUTION

4-7 NORMAL APPROXIMATION TO THE BINOMIAL AND POISSON DISTRIBUTIONS

4-8 CONTINUITY CORRECTION TO IMPROVE THE APPROXIMATION (CD ONLY)

4-9 EXPONENTIAL DISTRIBUTION

4-10 ERLANG AND GAMMA DISTRIBUTIONS

4-10.1 Erlang Distribution

4-10.2 Gamma Distribution

4-11 WEIBULL DISTRIBUTION

4-12 LOGNORMAL DISTRIBUTION

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Determine probabilities from probability density functions.
2. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability density functions, and the reverse.
3. Calculate means and variances for continuous random variables.
4. Understand the assumptions for each of the continuous probability distributions presented.
5. Select an appropriate continuous probability distribution to calculate probabilities in specific applications.
6. Calculate probabilities, determine means and variances for each of the continuous probability distributions presented.
7. Standardize normal random variables.

8. Use the table for the cumulative distribution function of a standard normal distribution to calculate probabilities.
9. Approximate probabilities for some binomial and Poisson distributions.

CD MATERIAL

10. Use continuity corrections to improve the normal approximation to those binomial and Poisson distributions.

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

4-1 CONTINUOUS RANDOM VARIABLES

Previously, we discussed the measurement of the current in a thin copper wire. We noted that the results might differ slightly in day-to-day replications because of small variations in variables that are not controlled in our experiment—changes in ambient temperatures, small impurities in the chemical composition of the wire, current source drifts, and so forth.

Another example is the selection of one part from a day's production and very accurately measuring a dimensional length. In practice, there can be small variations in the actual measured lengths due to many causes, such as vibrations, temperature fluctuations, operator differences, calibrations, cutting tool wear, bearing wear, and raw material changes. Even the measurement procedure can produce variations in the final results.

In these types of experiments, the measurement of interest—current in a copper wire experiment, length of a machined part—can be represented by a random variable. It is reasonable to model the range of possible values of the random variable by an interval (finite or infinite) of real numbers. For example, for the length of a machined part, our model enables the measurement from the experiment to result in any value within an interval of real numbers. Because the range is any value in an interval, the model provides for any precision in length measurements. However, because the number of possible values of the random variable X is uncountably infinite, X has a distinctly different distribution from the discrete random variables studied previously. The range of X includes all values in an interval of real numbers; that is, the range of X can be thought of as a continuum.

A number of continuous distributions frequently arise in applications. These distributions are described, and example computations of probabilities, means, and variances are provided in the remaining sections of this chapter.

4-2 PROBABILITY DISTRIBUTIONS AND PROBABILITY DENSITY FUNCTIONS

Density functions are commonly used in engineering to describe physical systems. For example, consider the density of a loading on a long, thin beam as shown in Fig. 4-1. For any point x along the beam, the density can be described by a function (in grams/cm). Intervals with large loadings correspond to large values for the function. The total loading between points a and b is determined as the integral of the density function from a to b. This integral is the area

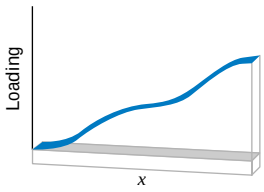


Figure 4-1 Density function of a loading on a long, thin beam.

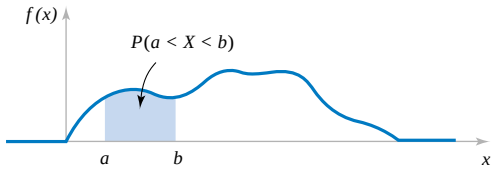


Figure 4-2 Probability determined from the area under $f(x)$.

under the density function over this interval, and it can be loosely interpreted as the sum of all the loadings over this interval.

Similarly, a **probability density function** $f(x)$ can be used to describe the probability distribution of a **continuous random variable** X. If an interval is likely to contain a value for X, its probability is large and it corresponds to large values for $f(x)$. The probability that X is between a and b is determined as the integral of $f(x)$ from a to b. See Fig. 4-2.

Definition

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx =$ area under $f(x)$ from a to b

for any a and b (4-1)

A probability density function provides a simple description of the probabilities associated with a random variable. As long as $f(x)$ is nonnegative and $\int_{-\infty}^{\infty} f(x)\,dx = 1$, $0 \leq P(a \leq X \leq b) \leq 1$ so that the probabilities are properly restricted. A probability density function is zero for x values that cannot occur and it is assumed to be zero wherever it is not specifically defined.

A **histogram** is an approximation to a probability density function. See Fig. 4-3. For each interval of the histogram, the area of the bar equals the relative frequency (proportion) of the measurements in the interval. The relative frequency is an estimate of the probability that a measurement falls in the interval. Similarly, the area under $f(x)$ over any interval equals the true probability that a measurement falls in the interval.

The important point is that ***f*(*x*) is used to calculate an area** that represents the probability that X assumes a value in $[a, b]$. For the current measurement example, the probability that X results in [14 mA, 15 mA] is the integral of the probability density function of X over this interval. The probability that X results in [14.5 mA, 14.6 mA] is the integral of

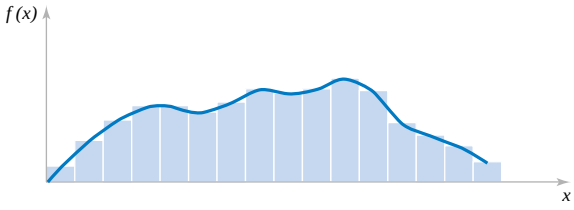


Figure 4-3 Histogram approximates a probability density function.

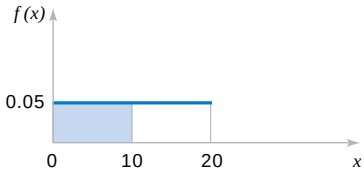


Figure 4-4 Probability density function for Example 4-1.

the same function, $f(x)$, over the smaller interval. By appropriate choice of the shape of $f(x)$, we can represent the probabilities associated with any continuous random variable X. The shape of $f(x)$ determines how the probability that X assumes a value in [14.5 mA, 14.6 mA] compares to the probability of any other interval of equal or different length.

For the density function of a loading on a long thin beam, because every point has zero width, the loading at any point is zero. Similarly, for a continuous random variable X and *any* value x.

$$P(X = x) = 0$$

Based on this result, it might appear that our model of a continuous random variable is useless. However, in practice, when a particular current measurement is observed, such as 14.47 milliamperes, this result can be interpreted as the rounded value of a current measurement that is actually in a range such as $14.465 \le x \le 14.475$. Therefore, the probability that the rounded value 14.47 is observed as the value for X is the probability that X assumes a value in the interval [14.465, 14.475], which is not zero. Similarly, because each point has zero probability, one need not distinguish between inequalities such as $<$ or $\le$ for continuous random variables.

If X is a **continuous random variable,** for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X \le x_2) = P(x_1 < X < x_2) \quad (4\text{-}2)$$

EXAMPLE 4-1

Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the range of X is [0, 20 mA], and assume that the probability density function of X is $f(x) = 0.05$ for $0 \le x \le 20$. What is the probability that a current measurement is less than 10 milliamperes?

The probability density function is shown in Fig. 4-4. It is assumed that $f(x) = 0$ wherever it is not specifically defined. The probability requested is indicated by the shaded area in Fig. 4-4.

$$P(X < 10) = \int_0^{10} f(x)\,dx = \int_0^{10} 0.05\,dx = 0.5$$

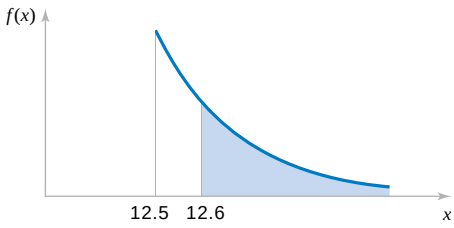


Figure 4-5 Probability density function for Example 4-2.

As another example,

$$P(5 \le X \le 20) = \int_5^{20} f(x)\,dx = 0.75$$

EXAMPLE 4-2

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimeters. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of X can be modeled by a probability density function $f(x) = 20e^{-20(x-12.5)}$, $x \ge 12.5$.

If a part with a diameter larger than 12.60 millimeters is scrapped, what proportion of parts is scrapped? The density function and the requested probability are shown in Fig. 4-5. A part is scrapped if $X \ge 12.60$. Now,

$$P(X \ge 12.60) = \int_{12.6}^{\infty} f(x)\,dx = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}\,dx = -e^{-20(x-12.5)}\Big|_{12.6}^{\infty} = 0.135$$

What proportion of parts is between 12.5 and 12.6 millimeters? Now,

$$P(12.5 \le X \le 12.6) = \int_{12.5}^{12.6} f(x)\,dx = -e^{-20(x-12.5)}\Big|_{12.5}^{12.6} = 0.865$$

Because the total area under $f(x)$ equals 1, we can also calculate $P(12.5 \le X \le 12.6) = 1 - P(X \le 12.6) - 1 - 0.135 = 0.865$.

EXERCISES FOR SECTION 4-2

4-1. Suppose that $f(x) = e^{-x}$ for $0 \le x$. Determine the following probabilities:

(a) $P(1 \le X)$ (b) $P(1 \le X \le 2.5)$
(c) $P(X \le 3)$ (d) $P(X \ge 4)$
(e) $P(3 \le X)$

4-2. Suppose that $f(x) = e^{-x}$ for $0 \le x$.

(a) Determine x such that $P(x \le X) = 0.10$.
(b) Determine x such that $P(X \le x) = 0.10$.

4-3. Suppose that $f(x) = x/8$ for $3 \le x \le 5$. Determine the following probabilities:

(a) $P(X \le 4)$ (b) $P(X \ge 3.5)$
(c) $P(4 \le X \le 5)$ (d) $P(X \ge 4.5)$
(e) $P(X \le 3.5 \text{ or } X \ge 4.5)$

4-4. Suppose that $f(x) = e^{-(x-4)}$ for $4 \le x$. Determine the following probabilities:

(a) $P(1 \le X)$ (b) $P(2 \le X \le 5)$

(c) $P(5 \le X)$ (d) $P(8 \le X \le 12)$
(e) Determine x such that $P(X \le x) = 0.90$.

4-5. Suppose that $f(x) = 1.5x^2$ for $-1 \le x \le 1$. Determine the following probabilities:
(a) $P(0 \le X)$ (b) $P(0.5 \le X)$
(c) $P(-0.5 \le X \le 0.5)$ (d) $P(X \le -2)$
(e) $P(X \le 0$ or $X \ge -0.5)$
(f) Determine x such that $P(x \le X) = 0.05$.

4-6. The probability density function of the time to failure of an electronic component in a copier (in hours) is $f(x) = \frac{e^{-x/1000}}{1000}$ for $x \ge 0$. Determine the probability that
(a) A component lasts more than 3000 hours before failure.
(b) A component fails in the interval from 1000 to 2000 hours.
(c) A component fails before 1000 hours.
(d) Determine the number of hours at which 10% of all components have failed.

4-7. The probability density function of the net weight in pounds of a packaged chemical herbicide is $f(x) = 2.0$ for $49.75 \le x \le 50.25$ pounds.
(a) Determine the probability that a package weighs more than 50 pounds.
(b) How much chemical is contained in 90% of all packages?

4-8. The probability density function of the length of a hinge for fastening a door is $f(x) = 1.25$ for $74.6 \le x \le 75.4$ millimeters. Determine the following:
(a) $P(X \le 74.8)$
(b) $P(X \le 74.8$ or $X \ge 75.2)$
(c) If the specifications for this process are from 74.7 to 75.3 millimeters, what proportion of hinges meets specifications?

4-9. The probability density function of the length of a metal rod is $f(x) = 2$ for $2.3 \le x \le 2.8$ meters.
(a) If the specifications for this process are from 2.25 to 2.75 meters, what proportion of the bars fail to meet the specifications?
(b) Assume that the probability density function is $f(x) = 2$ for an interval of length 0.5 meters. Over what value should the density be centered to achieve the greatest proportion of bars within specifications?

4-10. If X is a continuous random variable, argue that $P(x_1 \le X \le x_2) \le P(x_1 \le X \le x_2) \le P(x_1 \le X \le x_2) \le P(x_1 \le X \le x_2)$.

4-3 CUMULATIVE DISTRIBUTION FUNCTIONS

An alternative method to describe the distribution of a discrete random variable can also be used for continuous random variables.

Definition

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du \qquad (4\text{-}3)$$

for $-\infty < x < \infty$.

Extending the definition of $f(x)$ to the entire real line enables us to define the cumulative distribution function for all real numbers. The following example illustrates the definition.

EXAMPLE 4-3 For the copper current measurement in Example 4-1, the cumulative distribution function of the random variable X consists of three expressions. If $x \le 0$, $f(x) = 0$. Therefore,

$$F(x) = 0, \quad \text{for} \quad x \le 0$$

and

$$F(x) = \int_0^x f(u)\,du = 0.05x, \quad \text{for} \quad 0 \le x \le 20$$

Finally,

$$F(x) = \int_0^x f(u)\,du = 1, \quad \text{for} \quad 20 \le x$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.05x & 0 \le x < 20 \\ 1 & 20 \le x \end{cases}$$

The plot of $F(x)$ is shown in Fig. 4-6.

Notice that in the definition of $F(x)$ any $<$ can be changed to $\le$ and vice versa. That is, $F(x)$ can be defined as either $0.05x$ or 0 at the end-point $x = 0$, and $F(x)$ can be defined as either $0.05x$ or 1 at the end-point $x = 20$. In other words, $F(x)$ is a continuous function. For a discrete random variable, $F(x)$ is not a continuous function. Sometimes, a continuous random variable is defined as one that has a continuous cumulative distribution function.

EXAMPLE 4-4 For the drilling operation in Example 4-2, $F(x)$ consists of two expressions.

$$F(x) = 0 \quad \text{for} \quad x < 12.5$$

and for $12.5 \le x$

$$F(x) = \int_{12.5}^x 20e^{-20(u-12.5)}\,du$$
$$= 1 - e^{-20(x-12.5)}$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - e^{-20(x-12.5)} & 12.5 \le x \end{cases}$$

Figure 4-7 displays a graph of $F(x)$.

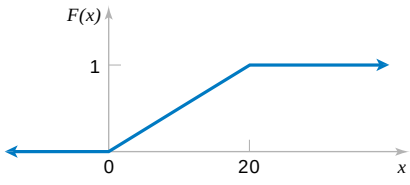


Figure 4-6 Cumulative distribution function for Example 4-3.

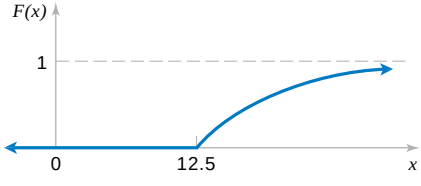


Figure 4-7 Cumulative distribution function for Example 4-4.

The probability density function of a continuous random variable can be determined from the cumulative distribution function by differentiating. Recall that the fundamental theorem of calculus states that

$$\frac{d}{dx}\int_{-\infty}^{x} f(u)\,du = f(x)$$

Then, given $F(x)$

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

EXAMPLE 4-5

The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-0.01x} & 0 \le x \end{cases}$$

Determine the probability density function of X. What proportion of reactions is complete within 200 milliseconds? Using the result that the probability density function is the derivative of the $F(x)$, we obtain

$$f(x) = \begin{cases} 0 & x < 0 \\ 0.01e^{-0.01x} & 0 \le x \end{cases}$$

The probability that a reaction completes within 200 milliseconds is

$$P(X < 200) = F(200) = 1 - e^{-2} = 0.8647.$$

EXERCISES FOR SECTION 4-3

4-11. Suppose the cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.2x & 0 \le x < 5 \\ 1 & 5 \le x \end{cases}$$

Determine the following:
(a) $P(X < 2.8)$ (b) $P(X < 1.5)$
(c) $P(X < -2)$ (d) $P(X < 6)$

4-12. Suppose the cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.25x - 0.5 & -2 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Determine the following:
(a) $P(X < 1.8)$ (b) $P(X < -1.5)$
(c) $P(X < -2)$ (d) $P(-1 < X < 1)$

4-13. Determine the cumulative distribution function for the distribution in Exercise 4-1.

4-14. Determine the cumulative distribution function for the distribution in Exercise 4-3.

4-15. Determine the cumulative distribution function for the distribution in Exercise 4-4.

4-16. Determine the cumulative distribution function for the distribution in Exercise 4-6. Use the cumulative distribution function to determine the probability that a component lasts more than 3000 hours before failure.

4-17. Determine the cumulative distribution function for the distribution in Exercise 4-8. Use the cumulative distribution function to determine the probability that a length exceeds 75 millimeters.

Determine the probability density function for each of the following cumulative distribution functions.

4-18. $F(x) = 1 - e^{-2x} \quad x \geq 0$

4-19.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 0.2x & 0 \leq x \leq 4 \\ 0.04x + 0.64 & 4 \leq x \leq 9 \\ 1 & 9 \leq x \end{cases}$$

4-20.

$$F(x) = \begin{cases} 0 & x \leq -2 \\ 0.25x + 0.5 & -2 \leq x \leq 1 \\ 0.5x + 0.25 & 1 \leq x \leq 1.5 \\ 1 & 1.5 \leq x \end{cases}$$

4-21. The gap width is an important property of a magnetic recording head. In coded units, if the width is a continuous random variable over the range from $0 \leq x \leq 2$ with $f(x) = 0.5x$, determine the cumulative distribution function of the gap width.

4-4 MEAN AND VARIANCE OF A CONTINUOUS RANDOM VARIABLE

The mean and variance of a continuous random variable are defined similarly to a discrete random variable. Integration replaces summation in the definitions. If a probability density function is viewed as a loading on a beam as in Fig. 4-1, the mean is the balance point.

Definition

Suppose X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \tag{4-4}$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

The equivalence of the two formulas for variance can be derived as one, as was done for discrete random variables.

EXAMPLE 4-6 For the copper current measurement in Example 4-1, the mean of X is

$$E(X) = \mu = \int_{0}^{20} x f(x)\, dx = 0.05 x^2 / 2 \Big|_{0}^{20} = 10$$

The variance of X is

$$V(X) = \sigma^2 = \int_{0}^{20} (x - 10)^2 f(x)\, dx = 0.05 (x - 10)^3 / 3 \Big|_{0}^{20} = 33.33$$

The expected value of a function $h(X)$ of a continuous random variable is defined similarly to a function of a discrete random variable.

Expected Value of a Function of a Continuous Random Variable

If X is a continuous random variable with probability density function $f(x)$,

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\, dx \tag{4-5}$$

EXAMPLE 4-7

In Example 4-1, X is the current measured in milliamperes. What is the expected value of the squared current? Now, $h(X) = X^2$. Therefore,

$$E[h(X)] = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_{0}^{20} 0.05x^2\, dx = 0.05 \frac{x^3}{3}\Big|_0^{20} = 133.33$$

In the previous example, the expected value of X^2 does not equal $E(X)$ squared. However, in the special case that $h(X) = aX + b$ for any constants a and b, $E[h(X)] = aE(X) + b$. This can be shown from the properties of integrals.

EXAMPLE 4-8

For the drilling operation in Example 4-2, the mean of X is

$$E(X) = \int_{12.5}^{\infty} x f(x)\, dx = \int_{12.5}^{\infty} x\, 20 e^{-20(x-12.5)}\, dx$$

Integration by parts can be used to show that

$$E(X) = -xe^{-20(x-12.5)} - \frac{e^{-20(x-12.5)}}{20}\Big|_{12.5}^{\infty} = 12.5 + 0.05 = 12.55$$

The variance of X is

$$V(X) = \int_{12.5}^{\infty} (x - 12.55)^2 f(x)\, dx$$

Although more difficult, integration by parts can be used two times to show that $V(X) = 0.0025$.

EXERCISES FOR SECTION 4-4

4-22. Suppose $f(x) = 0.25$ for $0 \le x \le 4$. Determine the mean and variance of X.

4-23. Suppose $f(x) = 0.125x$ for $0 \le x \le 4$. Determine the mean and variance of X.

4-24. Suppose $f(x) = 1.5x^2$ for $-1 \le x \le 1$. Determine the mean and variance of X.

4-25. Suppose that $f(x) = x/8$ for $3 \le x \le 5$. Determine the mean and variance for x.

4-26. Determine the mean and variance of the weight of packages in Exercise 4.7.

4-27. The thickness of a conductive coating in micrometers has a density function of $600x^{-2}$ for $100\ \mu\text{m} \le x \le 120\ \mu\text{m}$.

(a) Determine the mean and variance of the coating thickness.
(b) If the coating costs \$0.50 per micrometer of thickness on each part, what is the average cost of the coating per part?

4-28. Suppose that contamination particle size (in micrometers) can be modeled as $f(x) = 2x^{-3}$ for $1 < x$. Determine the mean of X.

4-29. Integration by parts is required. The probability density function for the diameter of a drilled hole in millimeters is $10e^{-10(x-5)}$ for $x > 5$ mm. Although the target diameter is 5 millimeters, vibrations, tool wear, and other nuisances produce diameters larger than 5 millimeters.

(a) Determine the mean and variance of the diameter of the holes.
(b) Determine the probability that a diameter exceeds 5.1 millimeters.

4-30. Suppose the probability density function of the length of computer cables is $f(x) = 0.1$ from 1200 to 1210 millimeters.
(a) Determine the mean and standard deviation of the cable length.
(b) If the length specifications are $1195 < x < 1205$ millimeters, what proportion of cables are within specifications?

4-5 CONTINUOUS UNIFORM DISTRIBUTION

The simplest continuous distribution is analogous to its discrete counterpart.

Definition

A continuous random variable X with probability density function

$$f(x) = 1/(b - a), \qquad a \le x \le b \tag{4-6}$$

is a **continuous uniform random variable.**

The probability density function of a continuous uniform random variable is shown in Fig. 4-8. The mean of the continuous uniform random variable X is

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{0.5x^2}{b-a}\Big|_a^b = \frac{(a+b)}{2}$$

The variance of X is

$$V(X) = \int_a^b \frac{\left(x - \left(\frac{a+b}{2}\right)\right)^2}{b-a}\,dx = \frac{\left(x - \frac{a+b}{2}\right)^3}{3(b-a)}\Big|_a^b = \frac{(b-a)^2}{12}$$

These results are summarized as follows.

If X is a continuous uniform random variable over $a \le x \le b$,

$$\mu = E(X) = \frac{(a+b)}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12} \tag{4-7}$$

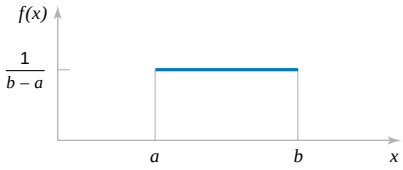


Figure 4-8 Continuous uniform probability density function.

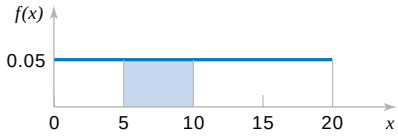


Figure 4-9 Probability for Example 4-9.

EXAMPLE 4-9

Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the range of X is [0, 20 mA], and assume that the probability density function of X is $f(x) = 0.05$, $0 \le x \le 20$.

What is the probability that a measurement of current is between 5 and 10 milliamperes? The requested probability is shown as the shaded area in Fig. 4-9.

$$P(5 \le X \le 10) = \int_5^{10} f(x)\,dx$$

$$= 5(0.05) = 0.25$$

The mean and variance formulas can be applied with $a = 0$ and $b = 20$. Therefore,

$$E(X) = 10 \text{ mA} \quad \text{and} \quad V(X) = 20^2/12 = 33.33 \text{ mA}^2$$

Consequently, the standard deviation of X is 5.77 mA.

The cumulative distribution function of a continuous uniform random variable is obtained by integration. If $a \le x \le b$,

$$F(x) = \int_a^x 1/(b-a)\,du = x/(b-a) - a/(b-a) = (x-a)/(b-a)$$

Therefore, the complete description of the cumulative distribution function of a continuous uniform random variable is

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

An example of $F(x)$ for a continuous uniform random variable is shown in Fig. 4-6.

EXERCISES FOR SECTION 4-5

4-31. Suppose X has a continuous uniform distribution over the interval [1.5, 5.5].
(a) Determine the mean, variance, and standard deviation of X.
(b) What is $P(X < 2.5)$?

4-32. Suppose X has a continuous uniform distribution over the interval $(-1, 1)$.
(a) Determine the mean, variance, and standard deviation of X.
(b) Determine the value for x such that $P(-x < X < x) = 0.90$.

4-33. The net weight in pounds of a packaged chemical herbicide is uniform for $49.75 \le x \le 50.25$ pounds.
(a) Determine the mean and variance of the weight of packages.

(b) Determine the cumulative distribution function of the weight of packages.
(c) Determine $P(X \leq 50.1)$.

4-34. The thickness of a flange on an aircraft component is uniformly distributed between 0.95 and 1.05 millimeters.
(a) Determine the cumulative distribution function of flange thickness.
(b) Determine the proportion of flanges that exceeds 1.02 millimeters.
(c) What thickness is exceeded by 90% of the flanges?
(d) Determine the mean and variance of flange thickness.

4-35. Suppose the time it takes a data collection operator to fill out an electronic form for a database is uniformly between 1.5 and 2.2 minutes.
(a) What is the mean and variance of the time it takes an operator to fill out the form?
(b) What is the probability that it will take less than two minutes to fill out the form?
(c) Determine the cumulative distribution function of the time it takes to fill out the form.

4-36. The probability density function of the time it takes a hematology cell counter to complete a test on a blood sample is $f(x) = 0.2$ for $50 \leq x \leq 75$ seconds.
(a) What percentage of tests require more than 70 seconds to complete.
(b) What percentage of tests require less than one minute to complete.
(c) Determine the mean and variance of the time to complete a test on a sample.

4-37. The thickness of photoresist applied to wafers in semiconductor manufacturing at a particular location on the wafer is uniformly distributed between 0.2050 and 0.2150 micrometers.
(a) Determine the cumulative distribution function of photoresist thickness.
(b) Determine the proportion of wafers that exceeds 0.2125 micrometers in photoresist thickness.
(c) What thickness is exceeded by 10% of the wafers?
(d) Determine the mean and variance of photoresist thickness.

4-38. The probability density function of the time required to complete an assembly operation is $f(x) = 0.1$ for $30 \leq x \leq 40$ seconds.
(a) Determine the proportion of assemblies that requires more than 35 seconds to complete.
(b) What time is exceeded by 90% of the assemblies?
(c) Determine the mean and variance of time of assembly.

4-6 NORMAL DISTRIBUTION

Undoubtedly, the most widely used model for the distribution of a random variable is a **normal distribution.** Whenever a random experiment is replicated, the random variable that equals the average (or total) result over the replicates tends to have a normal distribution as the number of replicates becomes large. De Moivre presented this fundamental result, known as the **central limit theorem,** in 1733. Unfortunately, his work was lost for some time, and Gauss independently developed a normal distribution nearly 100 years later. Although De Moivre was later credited with the derivation, a normal distribution is also referred to as a **Gaussian** distribution.

When do we average (or total) results? Almost always. For example, an automotive engineer may plan a study to average pull-off force measurements from several connectors. If we assume that each measurement results from a replicate of a random experiment, the normal distribution can be used to make approximate conclusions about this average. These conclusions are the primary topics in the subsequent chapters of this book.

Furthermore, sometimes the central limit theorem is less obvious. For example, assume that the deviation (or error) in the length of a machined part is the sum of a large number of infinitesimal effects, such as temperature and humidity drifts, vibrations, cutting angle variations, cutting tool wear, bearing wear, rotational speed variations, mounting and fixturing variations, variations in numerous raw material characteristics, and variation in levels of contamination. If the component errors are independent and equally likely to be positive or negative, the total error can be shown to have an approximate normal distribution. Furthermore, the normal distribution arises in the study of numerous basic physical phenomena. For example, the physicist Maxwell developed a normal distribution from simple assumptions regarding the velocities of molecules.

The theoretical basis of a normal distribution is mentioned to justify the somewhat complex form of the probability density function. Our objective now is to calculate probabilities for a normal random variable. The central limit theorem will be stated more carefully later.

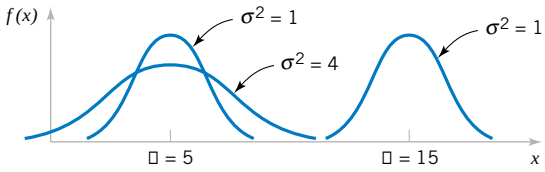


Figure 4-10 Normal probability density functions for selected values of the parameters □ and □².

Random variables with different means and variances can be modeled by normal probability density functions with appropriate choices of the center and width of the curve. The value of $E(X) = \mu$ determines the center of the probability density function and the value of $V(X) = \sigma^2$ determines the width. Figure 4-10 illustrates several normal probability density functions with selected values of μ and σ^2. Each has the characteristic symmetric bell-shaped curve, but the centers and dispersions differ. The following definition provides the formula for normal probability density functions.

Definition

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4-8}$$

is a **normal random variable** with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \tag{4-9}$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution. The mean and variance of X are shown to equal μ and σ^2, respectively, at the end of this Section 5-6.

EXAMPLE 4-10 Assume that the current measurements in a strip of wire follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement exceeds 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. This probability is shown as the shaded area under the normal probability density function in Fig. 4-11. Unfortunately, there is no closed-form expression for the integral of a normal probability density function, and probabilities based on the normal distribution are typically found numerically or from a table (that we will later introduce).

Some useful results concerning a normal distribution are summarized below and in Fig. 4-12. For any normal random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

Also, from the symmetry of $f(x)$, $P(X < \mu) = P(X > \mu) = 0.5$. Because $f(x)$ is positive for all x, this model assigns some probability to each interval of the real line. However, the

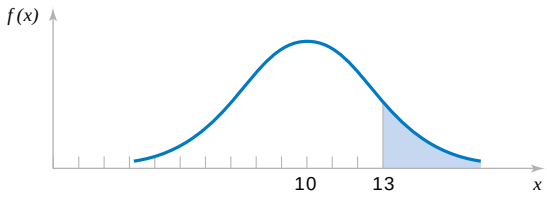


Figure 4-11 Probability that X □ 13 for a normal random variable with □ □ 10 and □^{2} □ 4.

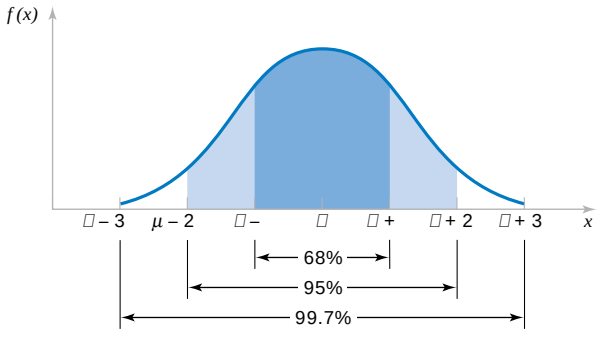


Figure 4-12 Probabilities associated with a normal distribution.

probability density function decreases as x moves farther from □. Consequently, the probability that a measurement falls far from □ is small, and at some distance from □ the probability of an interval can be approximated as zero.

The area under a normal probability density function beyond 3□ from the mean is quite small. This fact is convenient for quick, rough sketches of a normal probability density function. The sketches help us determine probabilities. Because more than 0.9973 of the probability of a normal distribution is within the interval 1□ □ 3□, □ □ 3□2, 6□ is often referred to as the **width** of a normal distribution. Advanced integration methods can be used to show that the area under the normal probability density function from □□ □ x □ □ is 1.

Definition

A normal random variable with

$$□ □ 0 \quad \text{and} \quad □^2 □ 1$$

is called a **standard normal random variable** and is denoted as Z.

The cumulative distribution function of a standard normal random variable is denoted as

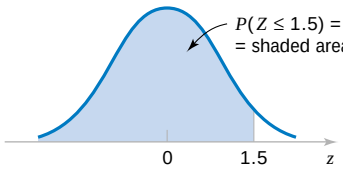
$$□1z2 □ P1Z □ z2$$

Appendix Table II provides cumulative probability values for □1z2, for a standard normal random variable. Cumulative distribution functions for normal random variables are also widely available in computer packages. They can be used in the same manner as Appendix Table II to obtain probabilities for these random variables. The use of Table II is illustrated by the following example.

EXAMPLE 4-11

Assume Z is a standard normal random variable. Appendix Table II provides probabilities of the form $P1Z □ z2$. The use of Table II to find $P1Z □ 1.52$ is illustrated in Fig. 4-13. Read down the z column to the row that equals 1.5. The probability is read from the adjacent column, labeled 0.00, to be 0.93319.

The column headings refer to the hundredth's digit of the value of z in $P1Z □ z2$. For example, $P1Z □ 1.532$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Figure 4-13 Standard normal probability density function.

Probabilities that are not of the form $P(Z \leq z)$ are found by using the basic rules of probability and the symmetry of the normal distribution along with Appendix Table II. The following examples illustrate the method.

EXAMPLE 4-12

The following calculations are shown pictorially in Fig. 4-14. In practice, a probability is often rounded to one or two significant digits.

(1) $P(Z \leq 1.26) = 1 - P(Z \leq 1.26) = 1 - 0.89616 = 0.10384$

(2) $P(Z \leq -0.86) = 0.19490$.

(3) $P(Z \leq -1.37) = P(Z \geq 1.37) = 0.91465$

(4) $P(-1.25 \leq Z \leq 0.37)$. This probability can be found from the difference of two areas, $P(Z \leq 0.37) - P(Z \leq -1.25)$. Now,

$$P(Z \leq 0.37) = 0.64431 \quad \text{and} \quad P(Z \leq -1.25) = 0.10565$$

Therefore,

$$P(-1.25 \leq Z \leq 0.37) = 0.64431 - 0.10565 = 0.53866$$

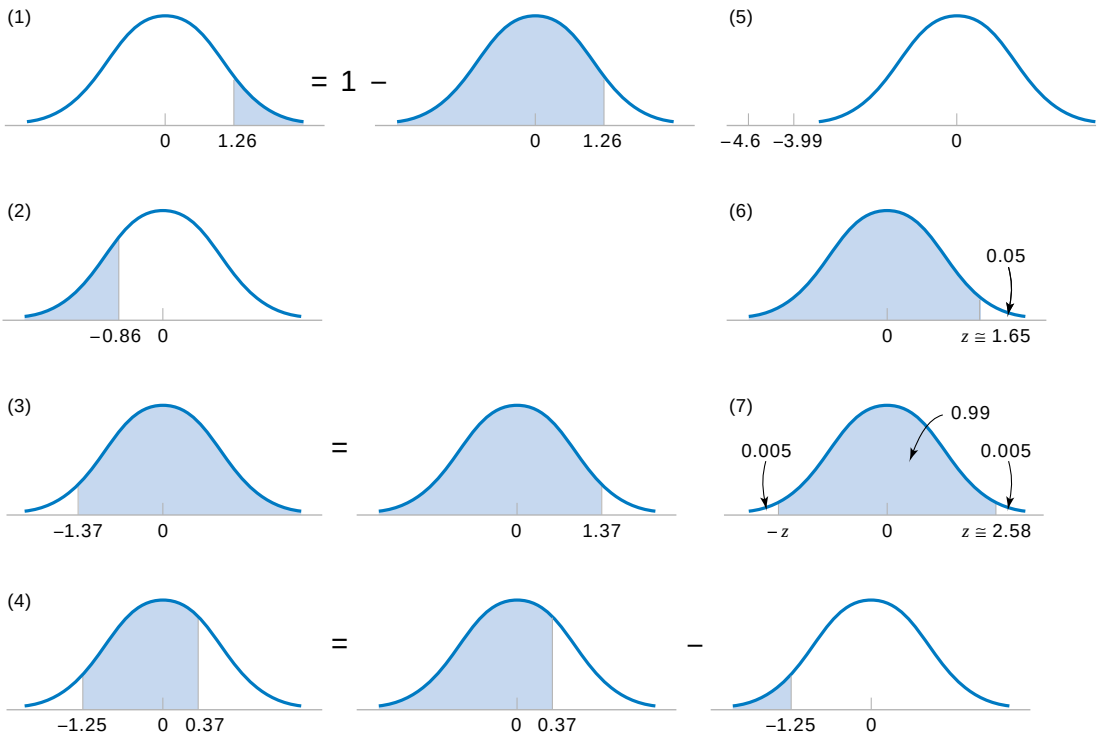


Figure 4-14 Graphical displays for standard normal distributions.

(5) $P(Z \leq -4.6)$ cannot be found exactly from Appendix Table II. However, the last entry in the table can be used to find that $P(Z \leq -3.99) = 0.00003$. Because $P(Z \leq -4.6) \leq P(Z \leq -3.99)$, $P(Z \leq -4.6)$ is nearly zero.

(6) Find the value z such that $P(Z \leq z) = 0.05$. This probability expression can be written as $P(Z \leq z) = 0.95$. Now, Table II is used in reverse. We search through the probabilities to find the value that corresponds to 0.95. The solution is illustrated in Fig. 4-14. We do not find 0.95 exactly; the nearest value is 0.95053, corresponding to $z = 1.65$.

(7) Find the value of z such that $P(-z \leq Z \leq z) = 0.99$. Because of the symmetry of the normal distribution, if the area of the shaded region in Fig. 4-14(7) is to equal 0.99, the area in each tail of the distribution must equal 0.005. Therefore, the value for z corresponds to a probability of 0.995 in Table II. The nearest probability in Table II is 0.99506, when $z = 2.58$.

The preceding examples show how to calculate probabilities for standard normal random variables. To use the same approach for an arbitrary normal random variable would require a separate table for every possible pair of values for μ and σ. Fortunately, all normal probability distributions are related algebraically, and Appendix Table II can be used to find the probabilities associated with an arbitrary normal random variable by first using a simple transformation.

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{4-10}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

Creating a new random variable by this transformation is referred to as **standardizing.** The random variable Z represents the distance of X from its mean in terms of standard deviations. It is the key step to calculate a probability for an arbitrary normal random variable.

EXAMPLE 4-13

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement will exceed 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. Let $Z = (X - 10)/2$. The relationship between the several values of X and the transformed values of Z are shown in Fig. 4-15. We note that $X = 13$ corresponds to $Z = 1.5$. Therefore, from Appendix Table II,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.93319 = 0.06681$$

Rather than using Fig. 4-15, the probability can be found from the inequality $X > 13$. That is,

$$P(X > 13) = P\left(\frac{X - 10}{2} > \frac{13 - 10}{2}\right) = P(Z > 1.5) = 0.06681$$

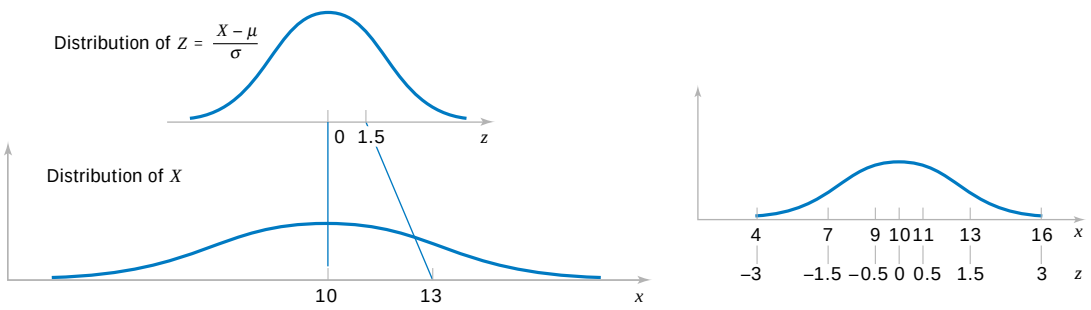


Figure 4-15 Standardizing a normal random variable.

In the preceding example, the value 13 is transformed to 1.5 by standardizing, and 1.5 is often referred to as the ***z*-value** associated with a probability. The following summarizes the calculation of probabilities derived from normal random variables.

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z) \tag{4-11}$$

where Z is a **standard normal random variable,** and $z = \frac{(x-\mu)}{\sigma}$ is the ***z*-value** obtained by **standardizing** X.

The probability is obtained by entering Appendix Table II with $z = (x-\mu)/\sigma$.

EXAMPLE 4-14

Continuing the previous example, what is the probability that a current measurement is between 9 and 11 milliamperes? From Fig. 4-15, or by proceeding algebraically, we have

$$\begin{aligned} P(9 \le X \le 11) &= P(9-10/2 \le X-10/2 \le 11-10/2) \\ &= P(-0.5 \le Z \le 0.5) = P(Z \le 0.5) - P(Z \le -0.5) \\ &= 0.69146 - 0.30854 = 0.38292 \end{aligned}$$

Determine the value for which the probability that a current measurement is below this value is 0.98. The requested value is shown graphically in Fig. 4-16. We need the value of x such that $P(X \le x) = 0.98$. By standardizing, this probability expression can be written as

$$\begin{aligned} P(X \le x) &= P(X-10/2 \le x-10/2) \\ &= P(Z \le x-10/2) \\ &= 0.98 \end{aligned}$$

Appendix Table II is used to find the z-value such that $P(Z \le z) = 0.98$. The nearest probability from Table II results in

$$P(Z \le 2.05) = 0.97982$$

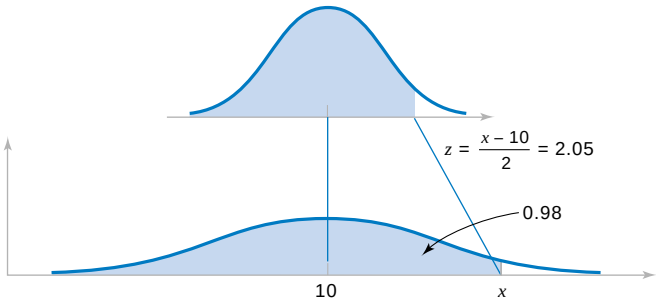


Figure 4-16 Determining the value of x to meet a specified probability.

Therefore, $(x - 10)/2 = 2.05$, and the standardizing transformation is used in reverse to solve for x. The result is

$$x = 2(2.05) + 10 = 14.1 \text{ milliamperes}$$

EXAMPLE 4-15

Assume that in the detection of a digital signal the background noise follows a normal distribution with a mean of 0 volt and standard deviation of 0.45 volt. The system assumes a digital 1 has been transmitted when the voltage exceeds 0.9. What is the probability of detecting a digital 1 when none was sent?

Let the random variable N denote the voltage of noise. The requested probability is

$$P(N > 0.9) = P\left(\frac{N}{0.45} > \frac{0.9}{0.45}\right) = P(Z > 2) = 1 - 0.97725 = 0.02275$$

This probability can be described as the probability of a false detection.

Determine symmetric bounds about 0 that include 99% of all noise readings. The question requires us to find x such that $P(-x < N < x) = 0.99$. A graph is shown in Fig. 4-17. Now,

$$P(-x < N < x) = P(-x/0.45 < N/0.45 < x/0.45)$$
$$= P(-x/0.45 < Z < x/0.45) = 0.99$$

From Appendix Table II

$$P(-2.58 < Z < 2.58) = 0.99$$

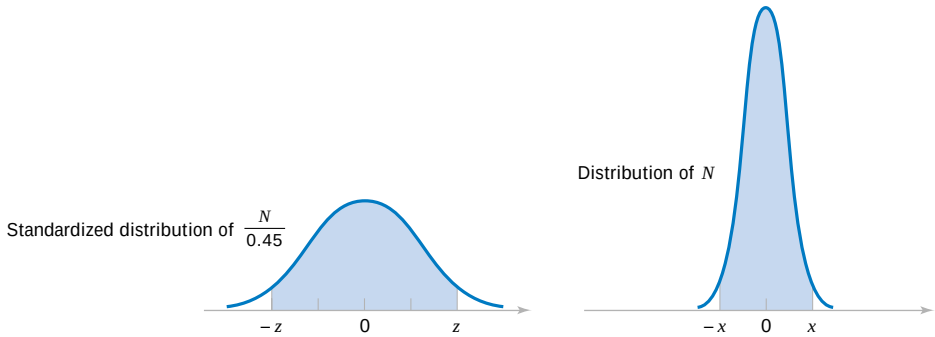


Figure 4-17 Determining the value of x to meet a specified probability.

Therefore,

$$x = 0.45 + 2.58$$

and

$$x + 2.58(0.45) = 1.16$$

Suppose a digital 1 is represented as a shift in the mean of the noise distribution to 1.8 volts. What is the probability that a digital 1 is not detected? Let the random variable S denote the voltage when a digital 1 is transmitted. Then,

$$P(S < 0.9) = P\left(\frac{S - 1.8}{0.45} < \frac{0.9 - 1.8}{0.45}\right) = P(Z < -2) = 0.02275$$

This probability can be interpreted as the probability of a missed signal.

EXAMPLE 4-16

The diameter of a shaft in an optical storage drive is normally distributed with mean 0.2508 inch and standard deviation 0.0005 inch. The specifications on the shaft are 0.2500 ± 0.0015 inch. What proportion of shafts conforms to specifications?

Let X denote the shaft diameter in inches. The requested probability is shown in Fig. 4-18 and

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485 - 0.2508}{0.0005} < Z < \frac{0.2515 - 0.2508}{0.0005}\right)$$
$$= P(-4.6 < Z < 1.4) = P(Z < 1.4) - P(Z < -4.6)$$
$$= 0.91924 - 0.0000 = 0.91924$$

Most of the nonconforming shafts are too large, because the process mean is located very near to the upper specification limit. If the process is centered so that the process mean is equal to the target value of 0.2500,

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485 - 0.2500}{0.0005} < Z < \frac{0.2515 - 0.2500}{0.0005}\right)$$
$$= P(-3 < Z < 3)$$
$$= P(Z < 3) - P(Z < -3)$$
$$= 0.99865 - 0.00135$$
$$= 0.9973$$

By recentering the process, the yield is increased to approximately 99.73%.

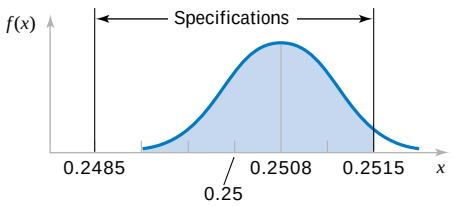


Figure 4-18 Distribution for Example 4-16.

Mean and Variance of the Normal Distribution (CD Only)

EXERCISES FOR SECTION 4-6

4-39. Use Appendix Table II to determine the following probabilities for the standard normal random variable Z:
(a) $P(Z \leq 1.32)$ (b) $P(Z \leq 3.0)$
(c) $P(Z \leq 1.45)$ (d) $P(Z \leq -2.15)$
(e) $P(-2.34 \leq Z \leq 1.76)$

4-40. Use Appendix Table II to determine the following probabilities for the standard normal random variable Z:
(a) $P(-1 \leq Z \leq 1)$ (b) $P(-2 \leq Z \leq 2)$
(c) $P(-3 \leq Z \leq 3)$ (d) $P(Z \leq 3)$
(e) $P(0 \leq Z \leq 1)$

4-41. Assume Z has a standard normal distribution. Use Appendix Table II to determine the value for z that solves each of the following:
(a) $P(Z \leq z) = 0.9$ (b) $P(Z \leq z) = 0.5$
(c) $P(Z \leq z) = 0.1$ (d) $P(Z \leq z) = 0.9$
(e) $P(-1.24 \leq Z \leq z) = 0.8$

4-42. Assume Z has a standard normal distribution. Use Appendix Table II to determine the value for z that solves each of the following:
(a) $P(-z \leq Z \leq z) = 0.95$ (b) $P(-z \leq Z \leq z) = 0.99$
(c) $P(-z \leq Z \leq z) = 0.68$ (d) $P(-z \leq Z \leq z) = 0.9973$

4-43. Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the following:
(a) $P(X \leq 13)$ (b) $P(X \leq 9)$
(c) $P(6 \leq X \leq 14)$ (d) $P(2 \leq X \leq 4)$
(e) $P(-2 \leq X \leq 8)$

4-44. Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the value for x that solves each of the following:
(a) $P(X \leq x) = 0.5$
(b) $P(X \leq x) = 0.95$
(c) $P(x \leq X \leq 10) = 0.2$
(d) $P(-x \leq X \leq 10 - x) = 0.95$
(e) $P(-x \leq X \leq 10 - x) = 0.99$

4-45. Assume X is normally distributed with a mean of 5 and a standard deviation of 4. Determine the following:
(a) $P(X \leq 11)$ (b) $P(X \leq 0)$
(c) $P(3 \leq X \leq 7)$ (d) $P(-2 \leq X \leq 9)$
(e) $P(2 \leq X \leq 8)$

4-46. Assume X is normally distributed with a mean of 5 and a standard deviation of 4. Determine the value for x that solves each of the following:
(a) $P(X \leq x) = 0.5$ (b) $P(X \leq x) = 0.95$
(c) $P(x \leq X \leq 9) = 0.2$ (d) $P(3 \leq X \leq x) = 0.95$
(e) $P(-x \leq X \leq x) = 0.99$

4-47. The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.
(a) What is the probability that a sample's strength is less than 6250 Kg/cm^2?
(b) What is the probability that a sample's strength is between 5800 and 5900 Kg/cm^2?
(c) What strength is exceeded by 95% of the samples?

4-48. The tensile strength of paper is modeled by a normal distribution with a mean of 35 pounds per square inch and a standard deviation of 2 pounds per square inch.
(a) What is the probability that the strength of a sample is less than 40 lb/in^2?
(b) If the specifications require the tensile strength to exceed 30 lb/in^2, what proportion of the samples is scrapped?

4-49. The line width of for semiconductor manufacturing is assumed to be normally distributed with a mean of 0.5 micrometer and a standard deviation of 0.05 micrometer.
(a) What is the probability that a line width is greater than 0.62 micrometer?
(b) What is the probability that a line width is between 0.47 and 0.63 micrometer?
(c) The line width of 90% of samples is below what value?

4-50. The fill volume of an automated filling machine used for filling cans of carbonated beverage is normally distributed with a mean of 12.4 fluid ounces and a standard deviation of 0.1 fluid ounce.
(a) What is the probability a fill volume is less than 12 fluid ounces?
(b) If all cans less than 12.1 or greater than 12.6 ounces are scrapped, what proportion of cans is scrapped?
(c) Determine specifications that are symmetric about the mean that include 99% of all cans.

4-51. The time it takes a cell to divide (called mitosis) is normally distributed with an average time of one hour and a standard deviation of 5 minutes.
(a) What is the probability that a cell divides in less than 45 minutes?
(b) What is the probability that it takes a cell more than 65 minutes to divide?
(c) What is the time that it takes approximately 99% of all cells to complete mitosis?

4-52. In the previous exercise, suppose that the mean of the filling operation can be adjusted easily, but the standard deviation remains at 0.1 ounce.
(a) At what value should the mean be set so that 99.9% of all cans exceed 12 ounces?
(b) At what value should the mean be set so that 99.9% of all cans exceed 12 ounces if the standard deviation can be reduced to 0.05 fluid ounce?

4-53. The reaction time of a driver to visual stimulus is normally distributed with a mean of 0.4 seconds and a standard deviation of 0.05 seconds.
(a) What is the probability that a reaction requires more than 0.5 seconds?
(b) What is the probability that a reaction requires between 0.4 and 0.5 seconds?
(c) What is the reaction time that is exceeded 90% of the time?

4-54. The speed of a file transfer from a server on campus to a personal computer at a student's home on a weekday evening is normally distributed with a mean of 60 kilobits per second and a standard deviation of 4 kilobits per second.
(a) What is the probability that the file will transfer at a speed of 70 kilobits per second or more?
(b) What is the probability that the file will transfer at a speed of less than 58 kilobits per second?
(c) If the file is 1 megabyte, what is the average time it will take to transfer the file? (Assume eight bits per byte.)

4-55. The length of an injection-molded plastic case that holds magnetic tape is normally distributed with a length of 90.2 millimeters and a standard deviation of 0.1 millimeter.
(a) What is the probability that a part is longer than 90.3 millimeters or shorter than 89.7 millimeters?
(b) What should the process mean be set at to obtain the greatest number of parts between 89.7 and 90.3 millimeters?
(c) If parts that are not between 89.7 and 90.3 millimeters are scrapped, what is the yield for the process mean that you selected in part (b)?

4-56. In the previous exercise assume that the process is centered so that the mean is 90 millimeters and the standard deviation is 0.1 millimeter. Suppose that 10 cases are measured, and they are assumed to be independent.
(a) What is the probability that all 10 cases are between 89.7 and 90.3 millimeters?
(b) What is the expected number of the 10 cases that are between 89.7 and 90.3 millimeters?

4-57. The sick-leave time of employees in a firm in a month is normally distributed with a mean of 100 hours and a standard deviation of 20 hours.
(a) What is the probability that the sick-leave time for next month will be between 50 and 80 hours?
(b) How much time should be budgeted for sick leave if the budgeted amount should be exceeded with a probability of only 10%?

4-58. The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.
(a) What is the probability that a laser fails before 5000 hours?
(b) What is the life in hours that 95% of the lasers exceed?
(c) If three lasers are used in a product and they are assumed to fail independently, what is the probability that all three are still operating after 7000 hours?

4-59. The diameter of the dot produced by a printer is normally distributed with a mean diameter of 0.002 inch and a standard deviation of 0.0004 inch.
(a) What is the probability that the diameter of a dot exceeds 0.0026 inch?
(b) What is the probability that a diameter is between 0.0014 and 0.0026 inch?
(c) What standard deviation of diameters is needed so that the probability in part (b) is 0.995?

4-60. The weight of a sophisticated running shoe is normally distributed with a mean of 12 ounces and a standard deviation of 0.5 ounce.
(a) What is the probability that a shoe weighs more than 13 ounces?
(b) What must the standard deviation of weight be in order for the company to state that 99.9% of its shoes are less than 13 ounces?
(c) If the standard deviation remains at 0.5 ounce, what must the mean weight be in order for the company to state that 99.9% of its shoes are less than 13 ounces?

4-7 NORMAL APPROXIMATION TO THE BINOMIAL AND POISSON DISTRIBUTIONS

We began our section on the normal distribution with the central limit theorem and the normal distribution as an approximation to a random variable with a large number of trials. Consequently, it should not be a surprise to learn that the normal distribution can be used to approximate binomial probabilities for cases in which n is large. The following example illustrates that for many physical systems the binomial model is appropriate with an extremely large value for n. In these cases, it is difficult to calculate probabilities by using the binomial distribution. Fortunately, the normal approximation is most effective in these cases. An illustration is provided in Fig. 4-19. The area of each bar equals the binomial probability of x. Notice that the area of bars can be approximated by areas under the normal density function.

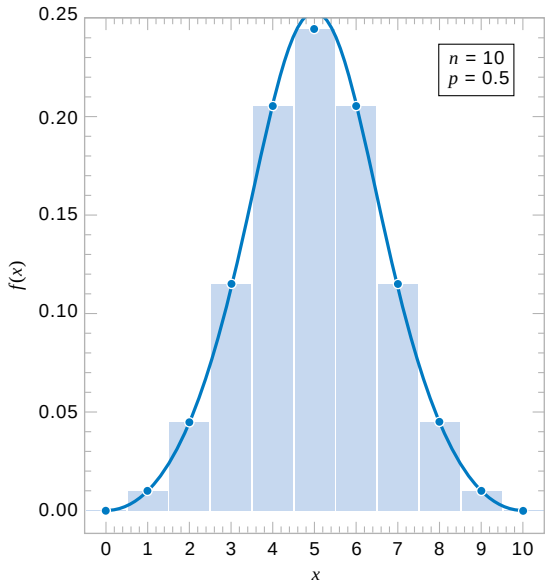


Figure 4-19 Normal approximation to the binomial distribution.

EXAMPLE 4-17

In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit is received in error is $1 \square 10^{\square 5}$. If 16 million bits are transmitted, what is the probability that more than 150 errors occur?

Let the random variable X denote the number of errors. Then X is a binomial random variable and

$$P(X \square 150) \square 1 \square P(x \square 150) \square 1 \square \sum_{x \square 0}^{150} \binom{16{,}000{,}000}{x} (10^{\square 5})^{x} (1 \square 10^{\square 5})^{16{,}000{,}000 \square x}$$

Clearly, the probability in Example 4-17 is difficult to compute. Fortunately, the normal distribution can be used to provide an excellent approximation in this example.

Normal Approximation to the Binomial Distribution

If X is a binomial random variable,

$$Z \square \frac{X \square np}{\sqrt{np(1 \square p)}} \tag{4-12}$$

is approximately a standard normal random variable. The approximation is good for

$$np \square 5 \quad \text{and} \quad n(1 \square p) \square 5$$

Recall that for a binomial variable X, $E(X) \square np$ and $V(X) \square np(1 \square p)$. Consequently, the expression in Equation 4-12 is nothing more than the formula for standardizing the random variable X. Probabilities involving X can be approximated by using a standard normal distribution. The approximation is good when n is large relative to p.

EXAMPLE 4-18

The digital communication problem in the previous example is solved as follows:

$$P(X > 150) = P\left(\frac{X - 160}{\sqrt{160(1 - 10^{-5})}} > \frac{150 - 160}{\sqrt{160(1 - 10^{-5})}}\right)$$

$$\approx P(Z > -0.79) = P(Z < 0.79) = 0.785$$

Because $np = 160 \times 10^{6}(1 - 10^{-5}) = 160$ and $n(1 - p)$ is much larger, the approximation is expected to work well in this case.

EXAMPLE 4-19

Again consider the transmission of bits in Example 4-18. To judge how well the normal approximation works, assume only $n = 50$ bits are to be transmitted and that the probability of an error is $p = 0.1$. The exact probability that 2 or less errors occur is

$$P(X \le 2) = \binom{50}{0} 0.9^{50} + \binom{50}{1} 0.1(0.9^{49}) + \binom{50}{2} 0.1^{2}(0.9^{48}) = 0.112$$

Based on the normal approximation

$$P(X \le 2) \approx P\left(\frac{X - 5}{2.12} < \frac{2 - 5}{2.12}\right) = P(Z < -1.42) = 0.08$$

Even for a sample as small as 50 bits, the normal approximation is reasonable.

If np or $n(1 - p)$ is small, the binomial distribution is quite skewed and the symmetric normal distribution is not a good approximation. Two cases are illustrated in Fig. 4-20. However, a correction factor can be used that will further improve the approximation. This factor is called a **continuity correction** and it is discussed in Section 4-8 on the CD.

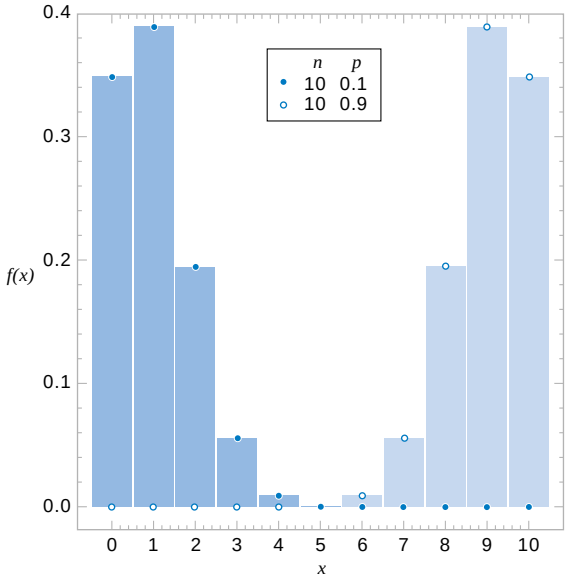


Figure 4-20 Binomial distribution is not symmetrical if p is near 0 or 1.

hypergometric distribution	$\frac{n}{N} \le 0.1$	binomial distribution	$np \ge 5$ $n(1 - p) \ge 5$	normal distribution

Figure 4-21 Conditions for approximating hypergeometric and binomial probabilities.

Recall that the binomial distribution is a satisfactory approximation to the hypergeometric distribution when n, the sample size, is small relative to N, the size of the population from which the sample is selected. A rule of thumb is that the binomial approximation is effective if $n/N \le 0.1$. Recall that for a hypergeometric distribution p is defined as $p = K/N$. That is, p is interpreted as the number of successes in the population. Therefore, the normal distribution can provide an effective approximation of hypergeometric probabilities when $n/N \le 0.1$, $np \ge 5$ and $n(1 - p) \ge 5$. Figure 4-21 provides a summary of these guidelines.

Recall that the Poisson distribution was developed as the limit of a binomial distribution as the number of trials increased to infinity. Consequently, it should not be surprising to find that the normal distribution can also be used to approximate probabilities of a Poisson random variable.

Normal Approximation to the Poisson Distribution

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \tag{4-13}$$

is approximately a standard normal random variable. The approximation is good for

$$\lambda \ge 5$$

EXAMPLE 4-20 Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that less than 950 particles are found?

This probability can be expressed exactly as

$$P(X < 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^x}{x!}$$

The computational difficulty is clear. The probability can be approximated as

$$P(X < x) = P\left(Z < \frac{950 - 1000}{\sqrt{1000}}\right) = P(Z < -1.58) = 0.057$$

EXERCISES FOR SECTION 4-7

4-61. Suppose that X is a binomial random variable with $n = 200$ and $p = 0.4$.

(a) Approximate the probability that X is less than or equal to 70.

(b) Approximate the probability that X is greater than 70 and less than 90.

4-62. Suppose that X is a binomial random variable with $n = 100$ and $p = 0.1$.

(a) Compute the exact probability that X is less than 4.

(b) Approximate the probability that X is less than 4 and compare to the result in part (a).

(c) Approximate the probability that $8 < X < 12$.

4-63. The manufacturing of semiconductor chips produces 2% defective chips. Assume the chips are independent and that a lot contains 1000 chips.
(a) Approximate the probability that more than 25 chips are defective.
(b) Approximate the probability that between 20 and 30 chips are defective.

4-64. A supplier ships a lot of 1000 electrical connectors. A sample of 25 is selected at random, without replacement. Assume the lot contains 100 defective connectors.
(a) Using a binomial approximation, what is the probability that there are no defective connectors in the sample?
(b) Use the normal approximation to answer the result in part (a). Is the approximation satisfactory?
(c) Redo parts (a) and (b) assuming the lot size is 500. Is the normal approximation to the probability that there are no defective connectors in the sample satisfactory in this case?

4-65. An electronic office product contains 5000 electronic components. Assume that the probability that each component operates without failure during the useful life of the product is 0.999, and assume that the components fail independently. Approximate the probability that 10 or more of the original 5000 components fail during the useful life of the product.

4-66. Suppose that the number of asbestos particles in a sample of 1 squared centimeter of dust is a Poisson random variable with a mean of 1000. What is the probability that 10 squared centimeters of dust contains more than 10,000 particles?

4-67. A corporate Web site contains errors on 50 of 1000 pages. If 100 pages are sampled randomly, without replacement, approximate the probability that at least 1 of the pages in error are in the sample.

4-68. Hits to a high-volume Web site are assumed to follow a Poisson distribution with a mean of 10,000 per day. Approximate each of the following:
(a) The probability of more than 20,000 hits in a day
(b) The probability of less than 9900 hits in a day
(c) The value such that the probability that the number of hits in a day exceed the value is 0.01

4-69. Continuation of Exercise 4-68.
(a) Approximate the expected number of days in a year (365 days) that exceed 10,200 hits.
(b) Approximate the probability that over a year (365 days) more than 15 days each have more than 10,200 hits.

4-70. The percentage of people exposed to a bacteria who become ill is 20%. Assume that people are independent. Assume that 1000 people are exposed to the bacteria. Approximate each of the following:
(a) The probability that more than 225 become ill
(b) The probability that between 175 and 225 become ill
(c) The value such that the probability that the number of people that become ill exceeds the value is 0.01

4-71. A high-volume printer produces minor print-quality errors on a test pattern of 1000 pages of text according to a Poisson distribution with a mean of 0.4 per page.
(a) Why are the number of errors on each page independent random variables?
(b) What is the mean number of pages with errors (one or more)?
(c) Approximate the probability that more than 350 pages contain errors (one or more).

4-8 CONTINUITY CORRECTION TO IMPROVE THE APPROXIMATION (CD ONLY)

4-9 EXPONENTIAL DISTRIBUTION

The discussion of the Poisson distribution defined a random variable to be the number of flaws along a length of copper wire. The distance between flaws is another random variable that is often of interest. Let the random variable X denote the length from any starting point on the wire until a flaw is detected.

As you might expect, the distribution of X can be obtained from knowledge of the distribution of the number of flaws. The key to the relationship is the following concept. The distance to the first flaw exceeds 3 millimeters if and only if there are no flaws within a length of 3 millimeters—simple, but sufficient for an analysis of the distribution of X.

In general, let the random variable N denote the number of flaws in x millimeters of wire. If the mean number of flaws is per millimeter, N has a Poisson distribution with mean x. We assume that the wire is longer than the value of x. Now,

$$P(X > x) = P(N = 0) = \frac{e^{-x}(x)^0}{0!} = e^{-x}$$

Therefore,

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

is the cumulative distribution function of X. By differentiating $F(x)$, the probability density function of X is calculated to be

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

The derivation of the distribution of X depends only on the assumption that the flaws in the wire follow a **Poisson process.** Also, the starting point for measuring X doesn't matter because the probability of the number of flaws in an interval of a Poisson process depends only on the length of the interval, not on the location. For any Poisson process, the following general result applies.

Definition

The random variable X that equals the distance between successive counts of a Poisson process with mean $\lambda > 0$ is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x < \infty \tag{4-14}$$

The exponential distribution obtains its name from the exponential function in the probability density function. Plots of the exponential distribution for selected values of λ are shown in Fig. 4-22. For any value of λ, the exponential distribution is quite skewed. The following results are easily obtained and are left as an exercise.

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \tag{4-15}$$

It is important to **use consistent units** in the calculation of probabilities, means, and variances involving exponential random variables. The following example illustrates unit conversions.

EXAMPLE 4-21

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in an interval of 6 minutes?

Let X denote the time in hours from the start of the interval until the first log-on. Then, X has an exponential distribution with $\lambda = 25$ log-ons per hour. We are interested in the probability that X exceeds 6 minutes. Because λ is given in log-ons per hour, we express all time units in hours. That is, 6 minutes $= 0.1$ hour. The probability requested is shown as the shaded area under the probability density function in Fig. 4-23. Therefore,

$$P(X > 0.1) = \int_{0.1}^{\infty} 25e^{-25x}\,dx = e^{-25(0.1)} = 0.082$$

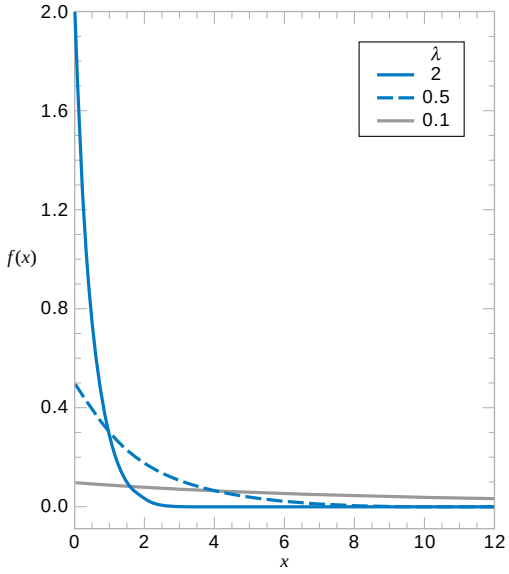


Figure 4-22 Probability density function of exponential random variables for selected values of .

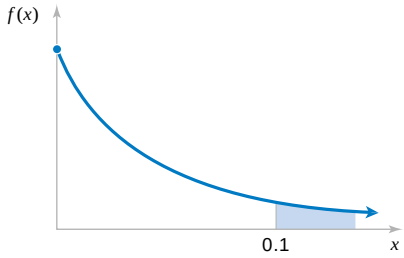


Figure 4-23 Probability for the exponential distribution in Example 4-21.

Also, the cumulative distribution function can be used to obtain the same result as follows:

$$P(X > 0.1) = 1 - F(0.1) = e^{-25(0.1)}$$

An identical answer is obtained by expressing the mean number of log-ons as 0.417 log-ons per minute and computing the probability that the time until the next log-on exceeds 6 minutes. Try it.

What is the probability that the time until the next log-on is between 2 and 3 minutes? Upon converting all units to hours,

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} 25e^{-25x}\,dx = -e^{-25x}\Big|_{0.033}^{0.05} = 0.152$$

An alternative solution is

$$P(0.033 < X < 0.05) = F(0.05) - F(0.033) = 0.152$$

Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90. The question asks for the length of time x such that $P(X > x) = 0.90$. Now,

$$P(X > x) = e^{-25x} = 0.90$$

Take the (natural) log of both sides to obtain $-25x = \ln(0.90) = -0.1054$. Therefore,

$$x = 0.00421 \text{ hour} = 0.25 \text{ minute}$$

Furthermore, the mean time until the next log-on is

$$\square\ \square\ 1\square 25\ \square\ 0.04 \text{ hour } \square\ 2.4 \text{ minutes}$$

The standard deviation of the time until the next log-on is

$$\square\ \square\ 1\square 25 \text{ hours } \square\ 2.4 \text{ minutes}$$

In the previous example, the probability that there are no log-ons in a 6-minute interval is 0.082 regardless of the starting time of the interval. A Poisson process assumes that events occur uniformly throughout the interval of observation; that is, there is no clustering of events. If the log-ons are well modeled by a Poisson process, the probability that the first log-on after noon occurs after 12:06 P.M. is the same as the probability that the first log-on after 3:00 P.M. occurs after 3:06 P.M. And if someone logs on at 2:22 P.M., the probability the next log-on occurs after 2:28 P.M. is still 0.082.

Our starting point for observing the system does not matter. However, if there are high-use periods during the day, such as right after 8:00 A.M., followed by a period of low use, a Poisson process is not an appropriate model for log-ons and the distribution is not appropriate for computing probabilities. It might be reasonable to model each of the high- and low-use periods by a separate Poisson process, employing a larger value for during the high-use periods and a smaller value otherwise. Then, an exponential distribution with the corresponding value of can be used to calculate log-on probabilities for the high- and low-use periods.

Lack of Memory Property

An even more interesting property of an exponential random variable is concerned with conditional probabilities.

EXAMPLE 4-22

Let X denote the time between detections of a particle with a geiger counter and assume that X has an exponential distribution with $\square$ 1.4 minutes. The probability that we detect a particle within 30 seconds of starting the counter is

$$P1X \square 0.5 \text{ minute}2 \square F10.52 \square 1 \square e^{\square 0.5\square 1.4} \square 0.30$$

In this calculation, all units are converted to minutes. Now, suppose we turn on the geiger counter and wait 3 minutes without detecting a particle. What is the probability that a particle is detected in the next 30 seconds?

Because we have already been waiting for 3 minutes, we feel that we are "due." That is, the probability of a detection in the next 30 seconds should be greater than 0.3. However, for an exponential distribution, this is not true. The requested probability can be expressed as the conditional probability that $P1X \square 3.5 \mid X \square 32$. From the definition of conditional probability,

$$P1X \square 3.5 \mid X \square 32 \square P13 \square X \square 3.52\square P1X \square 32$$

where

$$P13 \square X \square 3.52 \square F13.52 \square F132 \square 31 \square e^{\square 3.5\square 1.4}4 \square 31 \square e^{\square 3\square 1.4}4 \square 0.0035$$

and

$$P(X > 3) = 1 - F(3) = e^{-3/1.4} = 0.117$$

Therefore,

$$P(X > 3.5 \mid X > 3) = 0.035/0.117 = 0.30$$

After waiting for 3 minutes without a detection, the probability of a detection in the next 30 seconds is the same as the probability of a detection in the 30 seconds immediately after starting the counter. The fact that you have waited 3 minutes without a detection does not change the probability of a detection in the next 30 seconds.

Example 4-22 illustrates the **lack of memory property** of an exponential random variable and a general statement of the property follows. In fact, the exponential distribution is the only continuous distribution with this property.

Lack of Memory Property

For an exponential random variable X,

$$P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2) \tag{4-16}$$

Figure 4-24 graphically illustrates the lack of memory property. The area of region A divided by the total area under the probability density function $(A + B + C + D + 1)$ equals $P(X > t_2)$. The area of region C divided by the area $C + D$ equals $P(X > t_1 + t_2 \mid X > t_1)$. The lack of memory property implies that the proportion of the total area that is in A equals the proportion of the area in C and D that is in C. The mathematical verification of the lack of memory property is left as a mind-expanding exercise.

The lack of memory property is not that surprising when you consider the development of a Poisson process. In that development, we assumed that an interval could be partitioned into small intervals that were independent. These subintervals are similar to independent Bernoulli trials that comprise a binomial process; knowledge of previous results does not affect the probabilities of events in future subintervals. An exponential random variable is the continuous analog of a geometric random variable, and they share a similar lack of memory property.

The exponential distribution is often used in reliability studies as the model for the time until failure of a device. For example, the lifetime of a semiconductor chip might be modeled as an exponential random variable with a mean of 40,000 hours. The lack of

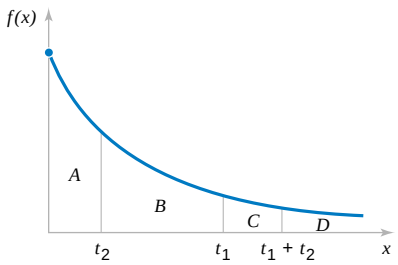


Figure 4-24 Lack of memory property of an exponential distribution.

memory property of the exponential distribution implies that the device does not wear out. That is, regardless of how long the device has been operating, the probability of a failure in the next 1000 hours is the same as the probability of a failure in the first 1000 hours of operation. The lifetime L of a device with failures caused by random shocks might be appropriately modeled as an exponential random variable. However, the lifetime L of a device that suffers slow mechanical wear, such as bearing wear, is better modeled by a distribution such that $P(L \le t + \Delta t \mid L \ge t)$ increases with t. Distributions such as the Weibull distribution are often used, in practice, to model the failure time of this type of device. The Weibull distribution is presented in a later section.

EXERCISES FOR SECTION 4-9

4-72. Suppose X has an exponential distribution with $\lambda = 2$. Determine the following:
(a) $P(X \le 0)$ (b) $P(X \ge 2)$
(c) $P(X \le 1)$ (d) $P(1 \le X \le 2)$
(e) Find the value of x such that $P(X \le x) = 0.05$.

4-73. Suppose X has an exponential distribution with mean equal to 10. Determine the following:
(a) $P(X \le 10)$
(b) $P(X \ge 20)$
(c) $P(X \le 30)$
(d) Find the value of x such that $P(X \le x) = 0.95$.

4-74. Suppose the counts recorded by a geiger counter follow a Poisson process with an average of two counts per minute.
(a) What is the probability that there are no counts in a 30-second interval?
(b) What is the probability that the first count occurs in less than 10 seconds?
(c) What is the probability that the first count occurs between 1 and 2 minutes after start-up?

4-75. Suppose that the log-ons to a computer network follow a Poisson process with an average of 3 counts per minute.
(a) What is the mean time between counts?
(b) What is the standard deviation of the time between counts?
(c) Determine x such that the probability that at least one count occurs before time x minutes is 0.95.

4-76. The time to failure (in hours) for a laser in a cytometry machine is modeled by an exponential distribution with $\lambda = 0.00004$.
(a) What is the probability that the laser will last at least 20,000 hours?
(b) What is the probability that the laser will last at most 30,000 hours?
(c) What is the probability that the laser will last between 20,000 and 30,000 hours?

4-77. The time between calls to a plumbing supply business is exponentially distributed with a mean time between calls of 15 minutes.
(a) What is the probability that there are no calls within a 30-minute interval?
(b) What is the probability that at least one call arrives within a 10-minute interval?
(c) What is the probability that the first call arrives within 5 and 10 minutes after opening?
(d) Determine the length of an interval of time such that the probability of at least one call in the interval is 0.90.

4-78. The life of automobile voltage regulators has an exponential distribution with a mean life of six years. You purchase an automobile that is six years old, with a working voltage regulator, and plan to own it for six years.
(a) What is the probability that the voltage regulator fails during your ownership?
(b) If your regulator fails after you own the automobile three years and it is replaced, what is the mean time until the next failure?

4-79. The time to failure (in hours) of fans in a personal computer can be modeled by an exponential distribution with $\lambda = 0.0003$.
(a) What proportion of the fans will last at least 10,000 hours?
(b) What proportion of the fans will last at most 7000 hours?

4-80. The time between the arrival of electronic messages at your computer is exponentially distributed with a mean of two hours.
(a) What is the probability that you do not receive a message during a two-hour period?
(b) If you have not had a message in the last four hours, what is the probability that you do not receive a message in the next two hours?
(c) What is the expected time between your fifth and sixth messages?

4-81. The time between arrivals of taxis at a busy intersection is exponentially distributed with a mean of 10 minutes.
(a) What is the probability that you wait longer than one hour for a taxi?
(b) Suppose you have already been waiting for one hour for a taxi, what is the probability that one arrives within the next 10 minutes?

4-82. Continuation of Exercise 4-81.
(a) Determine x such that the probability that you wait more than x minutes is 0.10.

(b) Determine x such that the probability that you wait less than x minutes is 0.90.
(c) Determine x such that the probability that you wait less than x minutes is 0.50.

4-83. The distance between major cracks in a highway follows an exponential distribution with a mean of 5 miles.
(a) What is the probability that there are no major cracks in a 10-mile stretch of the highway?
(b) What is the probability that there are two major cracks in a 10-mile stretch of the highway?
(c) What is the standard deviation of the distance between major cracks?

4-84. Continuation of Exercise 4-83.
(a) What is the probability that the first major crack occurs between 12 and 15 miles of the start of inspection?
(b) What is the probability that there are no major cracks in two separate 5-mile stretches of the highway?
(c) Given that there are no cracks in the first 5 miles inspected, what is the probability that there are no major cracks in the next 10 miles inspected?

4-85. The lifetime of a mechanical assembly in a vibration test is exponentially distributed with a mean of 400 hours.
(a) What is the probability that an assembly on test fails in less than 100 hours?
(b) What is the probability that an assembly operates for more than 500 hours before failure?
(c) If an assembly has been on test for 400 hours without a failure, what is the probability of a failure in the next 100 hours?

4-86. Continuation of Exercise 4-85.
(a) If 10 assemblies are tested, what is the probability that at least one fails in less than 100 hours? Assume that the assemblies fail independently.
(b) If 10 assemblies are tested, what is the probability that all have failed by 800 hours? Assume the assemblies fail independently.

4-87. When a bus service reduces fares, a particular trip from New York City to Albany, New York, is very popular. A small bus can carry four passengers. The time between calls for tickets is exponentially distributed with a mean of 30 minutes. Assume that each call orders one ticket. What is the probability that the bus is filled in less than 3 hours from the time of the fare reduction?

4-88. The time between arrivals of small aircraft at a county airport is exponentially distributed with a mean of one hour. What is the probability that more than three aircraft arrive within an hour?

4-89. Continuation of Exercise 4-88.
(a) If 30 separate one-hour intervals are chosen, what is the probability that no interval contains more than three arrivals?
(b) Determine the length of an interval of time (in hours) such that the probability that no arrivals occur during the interval is 0.10.

4-90. The time between calls to a corporate office is exponentially distributed with a mean of 10 minutes.
(a) What is the probability that there are more than three calls in one-half hour?
(b) What is the probability that there are no calls within one-half hour?
(c) Determine x such that the probability that there are no calls within x hours is 0.01.

4-91. Continuation of Exercise 4-90.
(a) What is the probability that there are no calls within a two-hour interval?
(b) If four nonoverlapping one-half hour intervals are selected, what is the probability that none of these intervals contains any call?
(c) Explain the relationship between the results in part (a) and (b).

4-92. If the random variable X has an exponential distribution with mean θ, determine the following:
(a) $P(X < \theta)$ (b) $P(X > 2\theta)$
(c) $P(X > 3\theta)$
(d) How do the results depend on θ?

4-93. Assume that the flaws along a magnetic tape follow a Poisson distribution with a mean of 0.2 flaw per meter. Let X denote the distance between two successive flaws.
(a) What is the mean of X?
(b) What is the probability that there are no flaws in 10 consecutive meters of tape?
(c) Does your answer to part (b) change if the 10 meters are not consecutive?
(d) How many meters of tape need to be inspected so that the probability that at least one flaw is found is 90%?

4-94. Continuation of Exercise 4-93. (More difficult questions.)
(a) What is the probability that the first time the distance between two flaws exceeds 8 meters is at the fifth flaw?
(b) What is the mean number of flaws before a distance between two flaws exceeds 8 meters?

4-95. Derive the formula for the mean and variance of an exponential random variable.

4-10 ERLANG AND GAMMA DISTRIBUTIONS

4-10.1 Erlang Distribution

An exponential random variable describes the length until the first count is obtained in a Poisson process. A generalization of the exponential distribution is the length until r counts

occur in a Poisson process. The random variable that equals the interval length until r counts occur in a Poisson process has an **Erlang random variable.**

EXAMPLE 4-23

The failures of the central processor units of large computer systems are often modeled as a Poisson process. Typically, failures are not caused by components wearing out, but by more random failures of the large number of semiconductor circuits in the units. Assume that the units that fail are immediately repaired, and assume that the mean number of failures per hour is 0.0001. Let X denote the time until four failures occur in a system. Determine the probability that X exceeds 40,000 hours.

Let the random variable N denote the number of failures in 40,000 hours of operation. The time until four failures occur exceeds 40,000 hours if and only if the number of failures in 40,000 hours is three or less. Therefore,

$$P(X > 40{,}000) = P(N \leq 3)$$

The assumption that the failures follow a Poisson process implies that N has a Poisson distribution with

$$E(N) = 40{,}000(0.0001) = 4 \text{ failures per } 40{,}000 \text{ hours}$$

Therefore,

$$P(X > 40{,}000) = P(N \leq 3) = \sum_{k=0}^{3} \frac{e^{-4}4^k}{k!} = 0.433$$

The cumulative distribution function of a general Erlang random variable X can be obtained from $P(X \leq x) = 1 - P(X > x)$, and $P(X > x)$ can be determined as in the previous example. Then, the probability density function of X can be obtained by differentiating the cumulative distribution function and using a great deal of algebraic simplification. The details are left as an exercise. In general, we can obtain the following result.

Definition

The random variable X that equals the interval length until r counts occur in a Poisson process with mean > 0 has an **Erlang random variable** with parameters and r. The probability density function of X is

$$f(x) = \frac{^{r}x^{r-1}e^{-x}}{(r-1)!}, \quad \text{for } x \geq 0 \text{ and } r = 1, 2, \ldots \tag{4-17}$$

Sketches of the Erlang probability density function for several values of r and are shown in Fig. 4-25. Clearly, an Erlang random variable with $r = 1$ is an exponential random variable. Probabilities involving Erlang random variables are often determined by computing a summation of Poisson random variables as in Example 4-23. The probability density function of an Erlang random variable can be used to determine probabilities; however, integrating by parts is often necessary. As was the case for the exponential distribution, one must be careful to define the random variable and the parameter in consistent units.

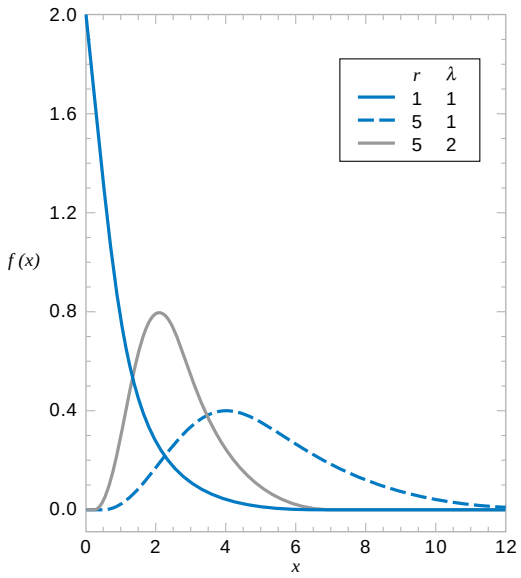


Figure 4-25 Erlang probability density functions for selected values of r and .

EXAMPLE 4-24

An alternative approach to computing the probability requested in Example 4-24 is to integrate the probability density function of X. That is,

$$P(X > 40{,}000) = \int_{40{,}000}^{\infty} f(x)\,dx = \int_{40{,}000}^{\infty} \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}\,dx$$

where $r = 4$ and $= 0.0001$. Integration by parts can be used to verify the result obtained previously.

An Erlang random variable can be thought of as the continuous analog of a negative binomial random variable. A negative binomial random variable can be expressed as the sum of r geometric random variables. Similarly, an Erlang random variable can be represented as the sum of r exponential random variables. Using this conclusion, we can obtain the following plausible result. Sums of random variables are studied in Chapter 5.

If X is an Erlang random variable with parameters and r,

$$= E(X) = r/ \quad \text{and} \quad \sigma^2 = V(X) = r/^2 \tag{4-18}$$

4-10.2 Gamma Distribution

The Erlang distribution is a special case of the **gamma distribution.** If the parameter r of an Erlang random variable is not an integer, but $r > 0$, the random variable has a gamma distribution. However, in the Erlang density function, the parameter r appears as r factorial.

Therefore, to define a gamma random variable, we require a generalization of the factorial function.

Definition

The **gamma function** is

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\, dx, \quad \text{for } r > 0 \tag{4-19}$$

It can be shown that the integral in the definition of $\Gamma(r)$ is finite. Furthermore, by using integration by parts it can be shown that

$$\Gamma(r) = (r-1)\Gamma(r-1)$$

This result is left as an exercise. Therefore, if r is a positive integer (as in the Erlang distribution),

$$\Gamma(r) = (r-1)!$$

Also, $\Gamma(1) = 0! = 1$ and it can be shown that $\Gamma(1/2) = \sqrt{\pi}$. The gamma function can be interpreted as a generalization to noninteger values of r of the term $(r-1)!$ that is used in the Erlang probability density function.

Now the gamma probability density function can be stated.

Definition

The random variable X with probability density function

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \quad \text{for } x > 0 \tag{4-20}$$

has a **gamma random variable** with parameters $\lambda > 0$ and $r > 0$. If r is an integer, X has an Erlang distribution.

Sketches of the gamma distribution for several values of λ and r are shown in Fig. 4-26. It can be shown that $f(x)$ satisfies the properties of a probability density function, and the following result can be obtained. Repeated integration by parts can be used, but the details are lengthy.

If X is a **gamma random variable** with parameters λ and r,

$$\mu = E(X) = r/\lambda \quad \text{and} \quad \sigma^2 = V(X) = r/\lambda^2 \tag{4-21}$$

Although the gamma distribution is not frequently used as a model for a physical system, the special case of the Erlang distribution is very useful for modeling random experiments. The exercises provide illustrations. Furthermore, the **chi-squared distribution** is a special case of

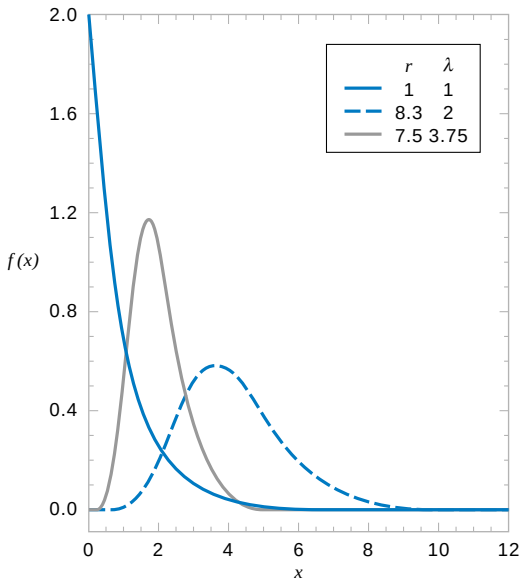


Figure 4-26 Gamma probability density functions for selected values of and *r*.

the gamma distribution in which □ 1□2 and *r* equals one of the values 1□2, 1, 3□2, 2, p . This distribution is used extensively in interval estimation and tests of hypotheses that are discussed in subsequent chapters.

EXERCISES FOR SECTION 4-10

4-96. Calls to a telephone system follow a Poisson distribution with a mean of five calls per minute.
(a) What is the name applied to the distribution and parameter values of the time until the tenth call?
(b) What is the mean time until the tenth call?
(c) What is the mean time between the ninth and tenth calls?

4-97. Continuation of Exercise 4-96.
(a) What is the probability that exactly four calls occur within one minute?
(b) If 10 separate one-minute intervals are chosen, what is the probability that all intervals contain more than two calls?

4-98. Raw materials are studied for contamination. Suppose that the number of particles of contamination per pound of material is a Poisson random variable with a mean of 0.01 particle per pound.
(a) What is the expected number of pounds of material required to obtain 15 particles of contamination?
(b) What is the standard deviation of the pounds of materials required to obtain 15 particles of contamination?

4-99. The time between failures of a laser in a cytogenics machine is exponentially distributed with a mean of 25,000 hours.
(a) What is the expected time until the second failure?
(b) What is the probability that the time until the third failure exceeds 50,000 hours?

4-100. In a data communication system, several messages that arrive at a node are bundled into a packet before they are transmitted over the network. Assume the messages arrive at the node according to a Poisson process with □ □ 30 messages per minute. Five messages are used to form a packet.
(a) What is the mean time until a packet is formed, that is, until five messages arrived at the node?
(b) What is the standard deviation of the time until a packet is formed?
(c) What is the probability that a packet is formed in less than 10 seconds?
(d) What is the probability that a packet is formed in less than 5 seconds?

4-101. Errors caused by contamination on optical disks occur at the rate of one error every 10^5 bits. Assume the errors follow a Poisson distribution.
(a) What is the mean number of bits until five errors occur?
(b) What is the standard deviation of the number of bits until five errors occur?

(c) The error-correcting code might be ineffective if there are three or more errors within 10^5 bits. What is the probability of this event?

4-102. Calls to the help line of a large computer distributor follow a Possion distribution with a mean of 20 calls per minute.
(a) What is the mean time until the one-hundredth call?
(b) What is the mean time between call numbers 50 and 80?
(c) What is the probability that three or more calls occur within 15 seconds?

4-103. The time between arrivals of customers at an automatic teller machine is an exponential random variable with a mean of 5 minutes.
(a) What is the probability that more than three customers arrive in 10 minutes?
(b) What is the probability that the time until the fifth customer arrives is less than 15 minutes?

4-104. The time between process problems in a manufacturing line is exponentially distributed with a mean of 30 days.
(a) What is the expected time until the fourth problem?
(b) What is the probability that the time until the fourth problem exceeds 120 days?

4-105. Use the properties of the gamma function to evaluate the following:
(a) $\Gamma(6)$ (b) $\Gamma(5/2)$
(c) $\Gamma(9/2)$

4-106. Use integration by parts to show that $\Gamma(r) = (r-1)\Gamma(r-1)$.

4-107. Show that the gamma density function $f(x, \lambda, r)$ integrates to 1.

4-108. Use the result for the gamma distribution to determine the mean and variance of a chi-square distribution with $r = 7$.

4-11 WEIBULL DISTRIBUTION

As mentioned previously, the Weibull distribution is often used to model the time until failure of many different physical systems. The parameters in the distribution provide a great deal of flexibility to model systems in which the number of failures increases with time (bearing wear), decreases with time (some semiconductors), or remains constant with time (failures caused by external shocks to the system).

Definition

The random variable X with probability density function

$$f(x) = \frac{\delta}{\beta}\left(\frac{x}{\beta}\right)^{\delta-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\delta}\right], \quad \text{for } x \geq 0 \tag{4-22}$$

is a **Weibull random variable** with scale parameter $\beta > 0$ and shape parameter $\delta > 0$.

The flexibility of the Weibull distribution is illustrated by the graphs of selected probability density functions in Fig. 4-27. By inspecting the probability density function, it is seen that when $\delta = 1$, the Weibull distribution is identical to the exponential distribution.

The cumulative distribution function is often used to compute probabilities. The following result can be obtained.

If X has a Weibull distribution with parameters δ and β, then the cumulative distribution function of X is

$$F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^{\delta}} \tag{4-23}$$

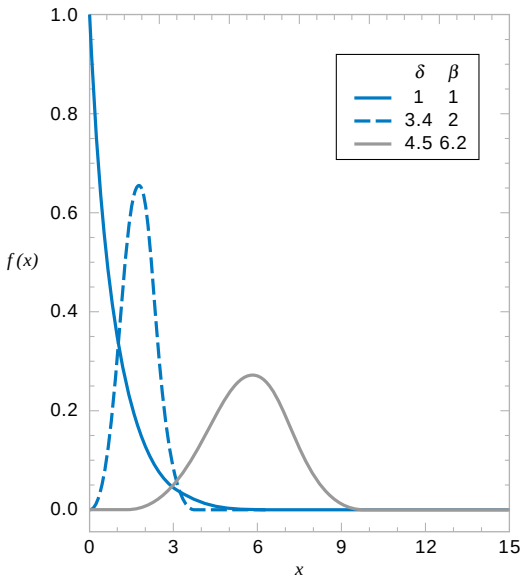


Figure 4-27 Weibull probability density functions for selected values of δ and β.

Also, the following result can be obtained.

If X has a Weibull distribution with parameters δ and β,

$$\mu = E(X) = \delta\Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{and} \quad \sigma^2 = V(X) = \delta^2\Gamma\left(1 + \frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2 \tag{4-24}$$

EXAMPLE 4-25

The time to failure (in hours) of a bearing in a mechanical shaft is satisfactorily modeled as a Weibull random variable with $\beta = 1/2$, and $\delta = 5000$ hours. Determine the mean time until failure.

From the expression for the mean,

$$E(X) = 5000\Gamma(1 + 1/0.5) = 5000\Gamma(3) = 5000 \times 2! = 10{,}000 \text{ hours}$$

Determine the probability that a bearing lasts at least 6000 hours. Now

$$P(x > 6000) = 1 - F(6000) = \exp\left[-\left(\frac{6000}{5000}\right)^{1/2}\right] = e^{-1.095} = 0.334$$

Consequently, only 33.4% of all bearings last at least 6000 hours.

EXERCISES FOR SECTION 4-11

4-109. Suppose that X has a Weibull distribution with $\beta = 0.2$ and $\delta = 100$ hours. Determine the mean and variance of X.

4-110. Suppose that X has a Weibull distribution $\beta = 0.2$ and $\delta = 100$ hours. Determine the following:
(a) $P(X > 10{,}000)$ (b) $P(X > 5000)$

4-111. Assume that the life of a roller bearing follows a Weibull distribution with parameters $\beta = 2$ and $\delta = 10{,}000$ hours.
(a) Determine the probability that a bearing lasts at least 8000 hours.
(b) Determine the mean time until failure of a bearing.
(c) If 10 bearings are in use and failures occur independently, what is the probability that all 10 bearings last at least 8000 hours?

4-112. The life (in hours) of a computer processing unit (CPU) is modeled by a Weibull distribution with parameters $\beta = 3$ and $\delta = 900$ hours.
(a) Determine the mean life of the CPU.
(b) Determine the variance of the life of the CPU.
(c) What is the probability that the CPU fails before 500 hours?

4-113. Assume the life of a packaged magnetic disk exposed to corrosive gases has a Weibull distribution with $\beta = 0.5$ and the mean life is 600 hours.
(a) Determine the probability that a packaged disk lasts at least 500 hours.
(b) Determine the probability that a packaged disk fails before 400 hours.

4-114. The life of a recirculating pump follows a Weibull distribution with parameters $\beta = 2$, and $\delta = 700$ hours.
(a) Determine the mean life of a pump.
(b) Determine the variance of the life of a pump.
(c) What is the probability that a pump will last longer than its mean?

4-115. The life (in hours) of a magnetic resonance imaging machine (MRI) is modeled by a Weibull distribution with parameters $\beta = 2$ and $\delta = 500$ hours.
(a) Determine the mean life of the MRI.
(b) Determine the variance of the life of the MRI.
(c) What is the probability that the MRI fails before 250 hours?

4-116. If X is a Weibull random variable with $\beta = 1$, and $\delta = 1000$, what is another name for the distribution of X and what is the mean of X?

4-12 LOGNORMAL DISTRIBUTION

Variables in a system sometimes follow an exponential relationship as $x = exp(w)$. If the exponent is a random variable, say W, $X = \exp(W)$ is a random variable and the distribution of X is of interest. An important special case occurs when W has a normal distribution. In that case, the distribution of X is called a **lognormal distribution.** The name follows from the transformation $\ln(X) = W$. That is, the natural logarithm of X is normally distributed.

Probabilities for X are obtained from the transformation to W, but we need to recognize that the range of X is $(0, \infty)$. Suppose that W is normally distributed with mean θ and variance ω^2; then the cumulative distribution function for X is

$$F(x) = P[X \le x] = P[\exp(W) \le x] = P[W \le \ln(x)]$$
$$= P\left(Z \le \frac{\ln(x) - \theta}{\omega}\right) = \Phi\left(\frac{\ln(x) - \theta}{\omega}\right)$$

for $x > 0$, where Z is a standard normal random variable. Therefore, Appendix Table II can be used to determine the probability. Also, $F(x) = 0$, for $x \le 0$.

The probability density function of X can be obtained from the derivative of $F(x)$. This derivative is applied to the last term in the expression for $F(x)$, the integral of the standard normal density function. Furthermore, from the probability density function, the mean and variance of X can be derived. The details are omitted, but a summary of results follows.

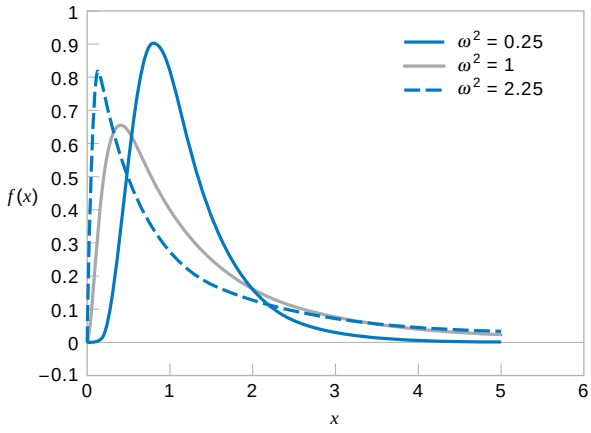


Figure 4-28 Lognormal probability density functions with $\theta = 0$ for selected values of ω^2.

Let W have a normal distribution mean θ and variance ω^2; then $X = \exp(W)$ is a **lognormal random variable** with probability density function

$$f(x) = \frac{1}{x\sqrt{2\pi\omega^2}} \exp\left[-\frac{(\ln x - \theta)^2}{2\omega^2}\right] \qquad 0 < x < \infty$$

The mean and variance of X are

$$E(X) = e^{\theta + \omega^2/2} \quad \text{and} \quad V(X) = e^{2\theta + \omega^2}(e^{\omega^2} - 1) \tag{4-25}$$

The parameters of a lognormal distribution are θ and ω^2, but care is needed to interpret that these are the mean and variance of the normal random variable W. The mean and variance of X are the functions of these parameters shown in (4-25). Figure 4-28 illustrates lognormal distributions for selected values of the parameters.

The lifetime of a product that degrades over time is often modeled by a lognormal random variable. For example, this is a common distribution for the lifetime of a semiconductor laser. A Weibull distribution can also be used in this type of application, and with an appropriate choice for parameters, it can approximate a selected lognormal distribution. However, a lognormal distribution is derived from a simple exponential function of a normal random variable, so it is easy to understand and easy to evaluate probabilities.

EXAMPLE 4-26 The lifetime of a semiconductor laser has a lognormal distribution with $\theta = 10$ hours and $\omega = 1.5$ hours. What is the probability the lifetime exceeds 10,000 hours?

From the cumulative distribution function for X

$$P(X > 10{,}000) = 1 - P[\exp(W) > 10{,}000] = 1 - P[W > \ln(10{,}000)]$$
$$= \Phi\left(\frac{\ln(10{,}000) - 10}{1.5}\right) = 1 - \Phi(0.52) = 1 - 0.30 = 0.70$$

What lifetime is exceeded by 99% of lasers? The question is to determine x such that $P(X > x) = 0.99$. Therefore,

$$P(X > x) = P(\exp(W) > x) = P(W > \ln(x)) = 1 - \Phi\left(\frac{\ln(x) - 10}{1.5}\right) = 0.99$$

From Appendix Table II, $1 - \Phi(z) = 0.99$ when $z = -2.33$. Therefore,

$$\frac{\ln(x) - 10}{1.5} = -2.33 \quad \text{and} \quad x = \exp(6.505) = 668.48 \text{ hours.}$$

Determine the mean and standard deviation of lifetime. Now,

$$E(X) = e^{\theta + \sigma^2/2} = \exp(10 + 1.125) = 67{,}846.3$$
$$V(X) = e^{2\theta + \sigma^2}(e^{\sigma^2} - 1) = \exp(20 + 2.25)(\exp(2.25) - 1) = 39{,}070{,}059{,}886.6$$

so the standard deviation of X is 197,661.5 hours. Notice that the standard deviation of lifetime is large relative to the mean.

EXERCISES FOR SECTION 4-12

4-117. Suppose that X has a lognormal distribution with parameters $\theta = 5$ and $\sigma^2 = 9$. Determine the following:
(a) $P(X < 13{,}300)$
(b) The value for x such that $P(X > x) = 0.95$
(c) The mean and variance of X

4-118. Suppose that X has a lognormal distribution with parameters $\theta = -2$ and $\sigma^2 = 9$. Determine the following:
(a) $P(500 < X < 1000)$
(b) The value for x such that $P(X > x) = 0.1$
(c) The mean and variance of X

4-119. Suppose that X has a lognormal distribution with parameters $\theta = 2$ and $\sigma^2 = 4$. Determine the following:
(a) $P(X < 500)$
(b) The conditional probability that $X < 1500$ given that $X < 1000$
(c) What does the difference between the probabilities in parts (a) and (b) imply about lifetimes of lognormal random variables?

4-120. The length of time (in seconds) that a user views a page on a Web site before moving to another page is a lognormal random variable with parameters $\theta = 0.5$ and $\sigma^2 = 1$.
(a) What is the probability that a page is viewed for more than 10 seconds?
(b) What is the length of time that 50% of users view the page?
(c) What is the mean and standard deviation of the time until a user moves from the page?

4-121. Suppose that X has a lognormal distribution and that the mean and variance of X are 100 and 85,000, respectively. Determine the parameters θ and σ^2 of the lognormal distribution. (*Hint:* define $x = \exp(\theta)$ and $y = \exp(\sigma^2)$ and write two equations in terms of x and y.)

4-122. The lifetime of a semiconductor laser has a lognormal distribution, and it is known that the mean and standard deviation of lifetime are 10,000 and 20,000, respectively.
(a) Calculate the parameters of the lognormal distribution
(b) Determine the probability that a lifetime exceeds 10,000 hours
(c) Determine the lifetime that is exceeded by 90% of lasers

4-123. Derive the probability density function of a lognormal random variable from the derivative of the cumulative distribution function.

Supplemental Exercises

4-124. Suppose that $f(x) = 0.5x - 1$ for $2 < x < 4$. Determine the following:
(a) $P(X < 2.5)$
(b) $P(X > 3)$
(c) $P(2.5 < X < 3.5)$

4-125. Continuation of Exercise 4-124. Determine the cumulative distribution function of the random variable.

4-126. Continuation of Exercise 4-124. Determine the mean and variance of the random variable.

4-127. The time between calls is exponentially distributed with a mean time between calls of 10 minutes.
(a) What is the probability that the time until the first call is less than 5 minutes?
(b) What is the probability that the time until the first call is between 5 and 15 minutes?
(c) Determine the length of an interval of time such that the probability of at least one call in the interval is 0.90.

4-128. Continuation of Exercise 4-127.
(a) If there has not been a call in 10 minutes, what is the probability that the time until the next call is less than 5 minutes?
(b) What is the probability that there are no calls in the intervals from 10:00 to 10:05, from 11:30 to 11:35, and from 2:00 to 2:05?

4-129. Continuation of Exercise 4-127.
(a) What is the probability that the time until the third call is greater than 30 minutes?
(b) What is the mean time until the fifth call?

4-130. The CPU of a personal computer has a lifetime that is exponentially distributed with a mean lifetime of six years. You have owned this CPU for three years. What is the probability that the CPU fails in the next three years?

4-131. Continuation of Exercise 4-130. Assume that your corporation has owned 10 CPUs for three years, and assume that the CPUs fail independently. What is the probability that at least one fails within the next three years?

4-132. Suppose that X has a lognormal distribution with parameters $\theta = 0$ and $\omega^2 = 4$. Determine the following:
(a) $P(10 \le X \le 50)$
(b) The value for x such that $P(X \le x) = 0.05$
(c) The mean and variance of X

4-133. Suppose that X has a lognormal distribution and that the mean and variance of X are 50 and 4000, respectively. Determine the following:
(a) The parameters θ and ω^2 of the lognormal distribution
(b) The probability that X is less than 150

4-134. Asbestos fibers in a dust sample are identified by an electron microscope after sample preparation. Suppose that the number of fibers is a Poisson random variable and the mean number of fibers per squared centimeter of surface dust is 100. A sample of 800 square centimeters of dust is analyzed. Assume a particular grid cell under the microscope represents 1/160,000 of the sample.
(a) What is the probability that at least one fiber is visible in the grid cell?
(b) What is the mean of the number of grid cells that need to be viewed to observe 10 that contain fibers?
(c) What is the standard deviation of the number of grid cells that need to be viewed to observe 10 that contain fibers?

4-135. Without an automated irrigation system, the height of plants two weeks after germination is normally distributed with a mean of 2.5 centimeters and a standard deviation of 0.5 centimeters.
(a) What is the probability that a plant's height is greater than 2.25 centimeters?
(b) What is the probability that a plant's height is between 2.0 and 3.0 centimeters?
(c) What height is exceeded by 90% of the plants?

4-136. Continuation of Exercise 4-135. With an automated irrigation system, a plant grows to a height of 3.5 centimeters two weeks after germination.
(a) What is the probability of obtaining a plant of this height or greater from the distribution of heights in Exercise 4-135.
(b) Do you think the automated irrigation system increases the plant height at two weeks after germination?

4-137. The thickness of a laminated covering for a wood surface is normally distributed with a mean of 5 millimeters and a standard deviation of 0.2 millimeter.
(a) What is the probability that a covering thickness is greater than 5.5 millimeters?
(b) If the specifications require the thickness to be between 4.5 and 5.5 millimeters, what proportion of coverings do not meet specifications?
(c) The covering thickness of 95% of samples is below what value?

4-138. The diameter of the dot produced by a printer is normally distributed with a mean diameter of 0.002 inch. Suppose that the specifications require the dot diameter to be between 0.0014 and 0.0026 inch. If the probability that a dot meets specifications is to be 0.9973, what standard deviation is needed?

4-139. Continuation of Exercise 4-138. Assume that the standard deviation of the size of a dot is 0.0004 inch. If the probability that a dot meets specifications is to be 0.9973, what specifications are needed? Assume that the specifications are to be chosen symmetrically around the mean of 0.002.

4-140. The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.
(a) What is the probability that a laser fails before 5,800 hours?
(b) What is the life in hours that 90% of the lasers exceed?

4-141. Continuation of Exercise 4-140. What should the mean life equal in order for 99% of the lasers to exceed 10,000 hours before failure?

4-142. Continuation of Exercise 4-140. A product contains three lasers, and the product fails if any of the lasers fails. Assume the lasers fail independently. What should the mean life equal in order for 99% of the products to exceed 10,000 hours before failure?

4-143. Continuation of Exercise 140. Rework parts (a) and (b). Assume that the lifetime is an exponential random variable with the same mean.

4-144. Continuation of Exercise 4-140. Rework parts (a) and (b). Assume that the lifetime is a lognormal random variable with the same mean and standard deviation.

4-145. A square inch of carpeting contains 50 carpet fibers. The probability of a damaged fiber is 0.0001. Assume the damaged fibers occur independently.
(a) Approximate the probability of one or more damaged fibers in 1 square yard of carpeting.
(b) Approximate the probability of four or more damaged fibers in 1 square yard of carpeting.

4-146. An airline makes 200 reservations for a flight that holds 185 passengers. The probability that a passenger arrives for the flight is 0.9 and the passengers are assumed to be independent.
(a) Approximate the probability that all the passengers that arrive can be seated.
(b) Approximate the probability that there are empty seats.
(c) Approximate the number of reservations that the airline should make so that the probability that everyone who arrives can be seated is 0.95. [*Hint*: Successively try values for the number of reservations.]

MIND-EXPANDING EXERCISES

4-147. The steps in this exercise lead to the probability density function of an Erlang random variable X with parameters and r, $f(x) = r x^{r-1} e^{-x}/(r-1)!$, $x > 0$, $r = 1, 2, \ldots$.
(a) Use the Poisson distribution to express $P(X > x)$.
(b) Use the result from part (a) to determine the cumulative distribution function of X.
(c) Differentiate the cumulative distribution function in part (b) and simplify to obtain the probability density function of X.

4-148. A bearing assembly contains 10 bearings. The bearing diameters are assumed to be independent and normally distributed with a mean of 1.5 millimeters and a standard deviation of 0.025 millimeter. What is the probability that the maximum diameter bearing in the assembly exceeds 1.6 millimeters?

4-149. Let the random variable X denote a measurement from a manufactured product. Suppose the target value for the measurement is m. For example, X could denote a dimensional length, and the target might be 10 millimeters. The **quality loss** of the process producing the product is defined to be the expected value of $\$k(X - m)^2$, where k is a constant that relates a deviation from target to a loss measured in dollars.
(a) Suppose X is a continuous random variable with $E(X) = m$ and $V(X) = \sigma^2$. What is the quality loss of the process?
(b) Suppose X is a continuous random variable with $E(X) = \mu$ and $V(X) = \sigma^2$. What is the quality loss of the process?

4-150. The lifetime of an electronic amplifier is modeled as an exponential random variable. If 10% of the amplifiers have a mean of 20,000 hours and the remaining amplifiers have a mean of 50,000 hours, what proportion of the amplifiers fail before 60,000 hours?

4-151. **Lack of Memory Property.** Show that for an exponential random variable X, $P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2)$

4-152. A process is said to be of **six-sigma quality** if the process mean is at least six standard deviations from the nearest specification. Assume a normally distributed measurement.
(a) If a process mean is centered between the upper and lower specifications at a distance of six standard deviations from each, what is the probability that a product does not meet specifications? Using the result that 0.000001 equals one part per million, express the answer in parts per million.
(b) Because it is difficult to maintain a process mean centered between the specifications, the probability of a product not meeting specifications is often calculated after assuming the process shifts. If the process mean positioned as in part (a) shifts upward by 1.5 standard deviations, what is the probability that a product does not meet specifications? Express the answer in parts per million.
(c) Rework part (a). Assume that the process mean is at a distance of three standard deviations.
(d) Rework part (b). Assume that the process mean is at a distance of three standard deviations and then shifts upward by 1.5 standard deviations.
(e) Compare the results in parts (b) and (d) and comment.

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Chi-squared distribution
Continuous uniform distribution
Cumulative probability distribution function-continuous random variable
Erlang distribution
Exponential distribution
Gamma distribution
Lack of memory property-continuous random variable
Lognormal distribution
Mean-continuous random variable
Mean-function of a continuous random variable
Normal approximation to binomial and Poisson probabilities
Normal distribution
Probability density function
Probability distribution-continuous random variable
Standard deviation-continuous random variable
Standard normal distribution
Standardizing
Variance-continuous random variable
Weibull distribution

CD MATERIAL

Continuity correction

Mean and Variance of the Normal Distribution (CD Only)

In the derivations below, the mean and variance of a normal random variable are shown to be μ and σ^2, respectively. The mean of x is

$$E(X) = \int_{-\infty}^{\infty} x \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}\, dx$$

By making the change of variable $y = (x - \mu)/\sigma$, the integral becomes

$$E(X) = \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy = \int_{-\infty}^{\infty} y \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy$$

The first integral in the expression above equals 1 because $\frac{e^{-y^2/2}}{\sqrt{2\pi}}$ is a probability density function and the second integral is found to be 0 by either formally making the change of variable $u = -y^2/2$ or noticing the symmetry of the integrand about $y = 0$. Therefore, $E(X) = \mu$.

The variance of X is

$$V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}\, dx$$

By making the change of variable $y = (x - \mu)/\sigma$, the integral becomes

$$V(X) = \sigma^2 \int_{-\infty}^{\infty} y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy$$

Upon integrating by parts with $u = y$ and $dv = y \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy$, $V(X)$ is found to be σ^2.

4-8 CONTINUITY CORRECTIONS TO IMPROVE THE APPROXIMATION

From Fig. 4-19 it can be seen that a probability such as $P(3 \le X \le 7)$ is better approximated by the area under the normal curve from 2.5 to 7.5. This observation provides a method to improve the approximation of binomial probabilities. Because a continuous normal distribution is used to approximate a discrete binomial distribution, the modification is referred to as a **continuity correction.**

If X is a binomial random variable with parameters n and p, and if $x = 0, 1, 2, \ldots, n$, the **continuity correction** to improve approximations obtained from the normal distribution is

$$P(X \le x) = P(X \le x + 0.5) \approx P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) \approx P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right)$$

A way to remember the approximation is to write the probability in terms of $\leq$ or $<$ and then add or subtract the 0.5 correction factor to make the probability greater.

EXAMPLE S4-1 Consider the situation in Example 4-20 with $n = 50$ and $p = 0.1$. The probability $P(X \leq 2)$ is better approximated as

$$P(X \leq 2) \approx P(X \leq 2.5) \approx P\left(Z \leq \frac{2 + 0.5 - 5}{2.12}\right) \approx P(Z \leq -1.18) \approx 0.119$$

and this result is closer to the exact probability of 0.112 than the previous result of 0.08.

As another example, $P(8 \leq X) \approx P(9 \leq X)$ and this is better approximated as

$$P(9 \leq X) \approx P(8.5 \leq X) \approx P\left(\frac{9 - 0.5 - 5}{2.12} \leq Z\right) \approx P(1.65 \leq Z) \approx 0.05$$

We can even approximate $P(X = 5) \approx P(5 \leq X \leq 5)$ as

$$P(5 \leq X \leq 5) \approx P\left(\frac{5 - 0.5 - 5}{2.12} \leq Z \leq \frac{5 + 0.5 - 5}{2.12}\right) \approx P(-0.24 \leq Z \leq 0.24) \approx 0.19$$

and this compares well with the exact answer of 0.1849.

EXERCISES FOR SECTION 4-8

S4-1. **Continuity correction.** The normal approximation of a binomial probability is sometimes modified by a correction factor of 0.5 that improves the approximation. Suppose that X is binomial with $n = 50$ and $p = 0.1$. Because X is a discrete random variable, $P(X \leq 2) \approx P(X \leq 2.5)$. However, the normal approximation to $P(X \leq 2)$ can be improved by applying the approximation to $P(X \leq 2.5)$.

(a) Approximate $P(X \leq 2)$ by computing the z-value corresponding to $x = 2.5$.

(b) Approximate $P(X \leq 2)$ by computing the z-value corresponding to $x = 2$.

(c) Compare the results in parts (a) and (b) to the exact value of $P(X \leq 2)$ to evaluate the effectiveness of the continuity correction.

(d) Use the continuity correction to approximate $P(X \leq 10)$.

S4-2. **Continuity correction.** Suppose that X is binomial with $n = 50$ and $p = 0.1$. Because X is a discrete random variable, $P(X \leq 2) \approx P(X < 1.5)$. However, the normal approximation to $P(X \leq 2)$ can be improved by applying the approximation to $P(X < 1.5)$. The continuity correction of 0.5 is either added or subtracted. The easy rule to remember is that the continuity correction is always applied to make the approximating normal probability greatest.

(a) Approximate $P(X \leq 2)$ by computing the z-value corresponding to 1.5.

(b) Approximate $P(X \leq 2)$ by computing the z-value corresponding to 2.

(c) Compare the results in parts (a) and (b) to the exact value of $P(X \leq 2)$ to evaluate the effectiveness of the continuity correction.

(d) Use the continuity correction to approximate $P(X \leq 6)$.

S4-3. **Continuity correction.** Suppose that X is binomial with $n = 50$ and $p = 0.1$. Because X is a discrete random variable, $P(2 \leq X \leq 5) \approx P(1.5 \leq X \leq 5.5)$. However, the normal approximation to $P(2 \leq X \leq 5)$ can be improved by applying the approximation to $P(1.5 \leq X \leq 5.5)$.

(a) Approximate $P(2 \leq X \leq 5)$ by computing the z-values corresponding to 1.5 and 5.5.

(b) Approximate $P(2 \leq X \leq 5)$ by computing the z-values corresponding to 2 and 5.

S4-4. **Continuity correction.** Suppose that X is binomial with $n = 50$ and $p = 0.1$. Then, $P(X \leq 10) \approx P(10 \leq X \leq 10)$. Using the results for the continuity corrections, we can approximate $P(10 \leq X \leq 10)$ by applying the normal standardization to $P(9.5 \leq X \leq 10.5)$.

(a) Approximate $P(X \leq 10)$ by computing the z-values corresponding to 9.5 and 10.5.

(b) Approximate $P(X = 5)$.

S4-5. **Continuity correction.** The manufacturing of semiconductor chips produces 2% defective chips. Assume that the chips are independent and that a lot contains 1000 chips.

(a) Use the continuity correction to approximate the probability that 20 to 30 chips in the lot are defective.

(b) Use the continuity correction to approximate the probability that exactly 20 chips are defective.

(c) Determine the number of defective chips, x, such that the normal approximation for the probability of obtaining x defective chips is greatest.

5 Joint Probability Distributions

CHAPTER OUTLINE

5-1 TWO DISCRETE RANDOM VARIABLES
- 5-1.1 Joint Probability Distributions
- 5-1.2 Marginal Probability Distributions
- 5-1.3 Conditional Probability Distributions
- 5-1.4 Independence

5-2 MULTIPLE DISCRETE RANDOM VARIABLES
- 5-2.1 Joint Probability Distributions
- 5-2.2 Multinomial Probability Distribution

5-3 TWO CONTINUOUS RANDOM VARIABLES
- 5-3.1 Joint Probability Distributions
- 5-3.2 Marginal Probability Distributions
- 5-3.3 Conditional Probability Distributions
- 5-3.4 Independence

5-4 MULTIPLE CONTINUOUS RANDOM VARIABLES

5-5 COVARIANCE AND CORRELATION

5-6 BIVARIATE NORMAL DISTRIBUTION

5-7 LINEAR COMBINATIONS OF RANDOM VARIABLES

5-8 FUNCTIONS OF RANDOM VARIABLES (CD ONLY)

5-9 MOMENT GENERATING FUNCTIONS (CD ONLY)

5-10 CHEBYSHEV'S INEQUALITY (CD ONLY)

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Use joint probability mass functions and joint probability density functions to calculate probabilities
2. Calculate marginal and conditional probability distributions from joint probability distributions
3. Use the multinomial distribution to determine probabilities
4. Interpret and calculate covariances and correlations between random variables
5. Understand properties of a bivariate normal distribution and be able to draw contour plots for the probability density function
6. Calculate means and variance for linear combinations of random variables and calculate probabilities for linear combinations of normally distributed random variables

CD MATERIAL

7. Determine the distribution of a function of one or more random variables
8. Calculate moment generating functions and use them to determine moments for random variables and use the uniqueness property to determine the distribution of a random variable
9. Provide bounds on probabilities for arbitrary distributions based on Chebyshev's inequality

---

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

In Chapters 3 and 4 we studied probability distributions for a single random variable. However, it is often useful to have more than one random variable defined in a random experiment. For example, in the classification of transmitted and received signals, each signal can be classified as high, medium, or low quality. We might define the random variable X to be the number of high-quality signals received and the random variable Y to be the number of low-quality signals received. In another example, the continuous random variable X can denote the length of one dimension of an injection-molded part, and the continuous random variable Y might denote the length of another dimension. We might be interested in probabilities that can be expressed in terms of both X and Y. For example, if the specifications for X and Y are (2.95 to 3.05) and (7.60 to 7.80) millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 \leq X \leq 3.05 \text{ and } 7.60 \leq Y \leq 7.80)$.

In general, if X and Y are two random variables, the probability distribution that defines their simultaneous behavior is called a **joint probability distribution.** In this chapter, we investigate some important properties of these joint distributions.

5-1 TWO DISCRETE RANDOM VARIABLES

5-1.1 Joint Probability Distributions

For simplicity, we begin by considering random experiments in which only two random variables are studied. In later sections, we generalize the presentation to the joint probability distribution of more than two random variables.

EXAMPLE 5-1

In the development of a new receiver for the transmission of digital information, each received bit is rated as *acceptable*, *suspect*, or *unacceptable*, depending on the quality of the received signal, with probabilities 0.9, 0.08, and 0.02, respectively. Assume that the ratings of each bit are independent.

In the first four bits transmitted, let

X denote the number of acceptable bits

Y denote the number of suspect bits

Then, the distribution of X is binomial with $n = 4$ and $p = 0.9$, and the distribution of Y is binomial with $n = 4$ and $p = 0.08$. However, because only four bits are being rated, the possible values of X and Y are restricted to the points shown in the graph in Fig. 5-1. Although the possible values of X are 0, 1, 2, 3, or 4, if $y = 3$, $x = 0$ or 1. By specifying the probability of each of the points in Fig. 5-1, we specify the joint probability distribution of X and Y. Similarly to an individual random variable, we define the range of the random variables (X, Y) to be the set of points (x, y) in two-dimensional space for which the probability that $X = x$ and $Y = y$ is positive.

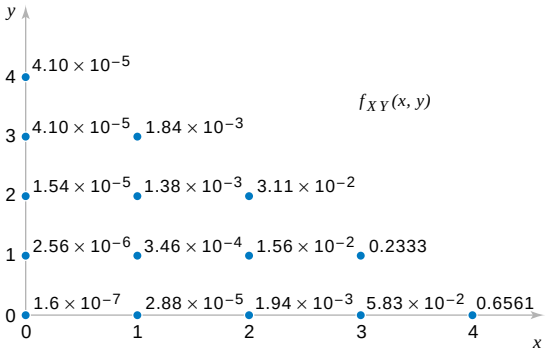


Figure 5-1 Joint probability distribution of X and Y in Example 5-1.

If X and Y are discrete random variables, the joint probability distribution of X and Y is a description of the set of points (x, y) in the range of (X, Y) along with the probability of each point. The joint probability distribution of two random variables is sometimes referred to as the **bivariate probability distribution** or **bivariate distribution** of the random variables. One way to describe the joint probability distribution of two discrete random variables is through a joint probability mass function. Also, $P(X \leq x \text{ and } Y \leq y)$ is usually written as $P(X \leq x, Y \leq y)$.

Definition

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

$$(1)\quad f_{XY}(x, y) \geq 0$$

$$(2)\quad \sum_x \sum_y f_{XY}(x, y) = 1$$

$$(3)\quad f_{XY}(x, y) = P(X = x, Y = y) \qquad (5\text{-}1)$$

Subscripts are used to indicate the random variables in the bivariate probability distribution. Just as the probability mass function of a single random variable X is assumed to be zero at all values outside the range of X, so the joint probability mass function of X and Y is assumed to be zero at values for which a probability is not specified.

EXAMPLE 5-2

Probabilities for each point in Fig. 5-1 are determined as follows. For example, $P(X \leq 2, Y = 1)$ is the probability that exactly two acceptable bits and exactly one suspect bit are received among the four bits transferred. Let a, s, and u denote acceptable, suspect, and unacceptable bits, respectively. By the assumption of independence,

$$P(aasu) = 0.9(0.9)(0.08)(0.02) = 0.0013$$

The number of possible sequences consisting of two a's, one s, and one u is shown in the CD material for Chapter 2:

$$\frac{4!}{2!1!1!} = 12$$

Therefore,

$$P(aasu) = 12(0.0013) = 0.0156$$

and

$$f_{XY}(2, 12) = P(X = 2, Y = 12) = 0.0156$$

The probabilities for all points in Fig. 5-1 are shown next to the point and the figure describes the joint probability distribution of X and Y.

5-1.2 Marginal Probability Distributions

If more than one random variable is defined in a random experiment, it is important to distinguish between the joint probability distribution of X and Y and the probability distribution of each variable individually. The individual probability distribution of a random variable is referred to as its **marginal probability distribution.** In Example 5-1, we mentioned that the marginal probability distribution of X is binomial with $n = 4$ and $p = 0.9$ and the marginal probability distribution of Y is binomial with $n = 4$ and $p = 0.08$.

In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, to determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

EXAMPLE 5-3

The joint probability distribution of X and Y in Fig. 5-1 can be used to find the marginal probability distribution of X. For example,

$$P(X = 3) = P(X = 3, Y = 0) = P(X = 3, Y = 1) = 0.0583 = 0.2333 = 0.292$$

As expected, this probability matches the result obtained from the binomial probability distribution for X; that is, $P(X = 3) = \binom{4}{3}0.9^3 0.1^1 = 0.292$. The marginal probability distribution for X is found by summing the probabilities in each column, whereas the marginal probability distribution for Y is found by summing the probabilities in each row. The results are shown in Fig. 5-2.

Although the marginal probability distribution of X in the previous example can be determined directly from the description of the experiment, in some problems the marginal probability distribution is determined from the joint probability distribution.

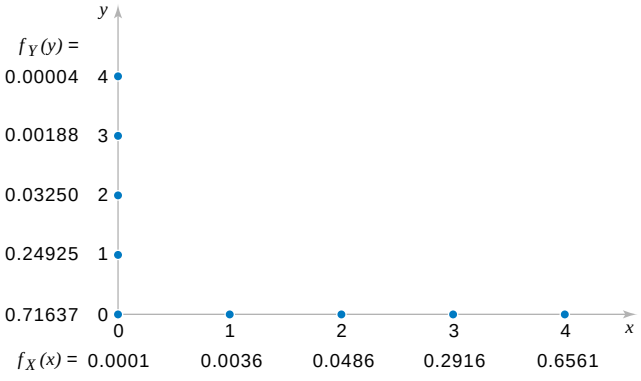


Figure 5-2 Marginal probability distributions of X and Y from Fig. 5-1.

Definition

If X and Y are discrete random variables with joint probability mass function $f_{XY}(x, y)$, then the **marginal probability mass functions** of X and Y are

$$f_X(x) = P(X = x) = \sum_{R_x} f_{XY}(x, y) \quad \text{and} \quad f_Y(y) = P(Y = y) = \sum_{R_y} f_{XY}(x, y) \tag{5-2}$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R_y denotes the set of all points in the range of (X, Y) for which $Y = y$

Given a joint probability mass function for random variables X and Y, $E(X)$ and $V(X)$ can be obtained directly from the joint probability distribution of X and Y or by first calculating the marginal probability distribution of X and then determining $E(X)$ and $V(X)$ by the usual method. This is shown in the following equation.

Mean and Variance from Joint Distribution

If the marginal probability distribution of X has the probability mass function $f_X(x)$, then

$$E(X) = \mu_X = \sum_x x f_X(x) = \sum_x x \left(\sum_{R_x} f_{XY}(x, y) \right) = \sum_x \sum_{R_x} x f_{XY}(x, y)$$
$$= \sum_R x f_{XY}(x, y) \tag{5-3}$$

and

$$V(X) = \sigma_X^2 = \sum_x (x - \mu_X)^2 f_X(x) = \sum_x (x - \mu_X)^2 \sum_{R_x} f_{XY}(x, y)$$
$$= \sum_x \sum_{R_x} (x - \mu_X)^2 f_{XY}(x, y) = \sum_R (x - \mu_X)^2 f_{XY}(x, y)$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R denotes the set of all points in the range of (X, Y)

EXAMPLE 5-4

In Example 5-1, $E(X)$ can be found as

$$\begin{aligned} E(X) &= 0[f_{XY}(0, 0) + f_{XY}(0, 1) + f_{XY}(0, 2) + f_{XY}(0, 3) + f_{XY}(0, 4)] \\ &\quad + 1[f_{XY}(1, 0) + f_{XY}(1, 1) + f_{XY}(1, 2) + f_{XY}(1, 3)] \\ &\quad + 2[f_{XY}(2, 0) + f_{XY}(2, 1) + f_{XY}(2, 2)] \\ &\quad + 3[f_{XY}(3, 0) + f_{XY}(3, 1)] \\ &\quad + 4[f_{XY}(4, 0)] \\ &= 0(0.0001) + 1(0.0036) + 2(0.0486) + 3(0.02916) + 4(0.6561) = 3.6 \end{aligned}$$

Alternatively, because the marginal probability distribution of X is binomial,

$$E(X) = np = 4(0.9) = 3.6$$

The calculation using the joint probability distribution can be used to determine $E(X)$ even in cases in which the marginal probability distribution of X is not known. As practice, you can use the joint probability distribution to verify that $E(Y) = 0.32$ in Example 5-1.

Also,

$$V(X) = np(1 - p) = 4(0.9)(1 - 0.9) = 0.36$$

Verify that the same result can be obtained from the joint probability distribution of X and Y.

5-1.3 Conditional Probability Distributions

When two random variables are defined in a random experiment, knowledge of one can change the probabilities that we associate with the values of the other. Recall that in Example 5-1, X denotes the number of acceptable bits and Y denotes the number of suspect bits received by a receiver. Because only four bits are transmitted, if $X = 4$, Y must equal 0. Using the notation for conditional probabilities from Chapter 2, we can write this result as $P(Y = 0 | X = 4) = 1$. If $X = 3$, Y can only equal 0 or 1. Consequently, the random variables X and Y can be considered to be dependent. Knowledge of the value obtained for X changes the probabilities associated with the values of Y.

Recall that the definition of conditional probability for events A and B is $P(B|A) = P(A \cap B)/P(A)$. This definition can be applied with the event A defined to be $X = x$ and event B defined to be $Y = y$.

EXAMPLE 5-5

For Example 5-1, X and Y denote the number of acceptable and suspect bits received, respectively. The remaining bits are unacceptable.

$$P(Y = 0 | X = 3) = P(X = 3, Y = 0)/P(X = 3)$$
$$= f_{XY}(3, 0)/f_X(3) = 0.05832/0.2916 = 0.200$$

The probability that $Y = 1$ given that $X = 3$ is

$$P(Y = 1 | X = 3) = P(X = 3, Y = 1)/P(X = 3)$$
$$= f_{XY}(3, 1)/f_X(3) = 0.2333/0.2916 = 0.800$$

Given that $X = 3$, the only possible values for Y are 0 and 1. Notice that $P(Y = 0 | X = 3) + P(Y = 1 | X = 3) = 1$. The values 0 and 1 for Y along with the probabilities 0.200 and 0.800 define the conditional probability distribution of Y given that $X = 3$.

Example 5-5 illustrates that the conditional probabilities that $Y = y$ given that $X = x$ can be thought of as a new probability distribution. The following definition generalizes these ideas.

Definition

Given discrete random variables X and Y with joint probability mass function $f_{XY}(x, y)$ the **conditional probability mass function** of Y given $X = x$ is

$$f_{Y|x}(y) = f_{XY}(x, y)/f_X(x) \quad \text{for } f_X(x) > 0 \tag{5-4}$$

The function $f_{Y|x}(y)$ is used to find the probabilities of the possible values for Y given that $X = x$. That is, it is the probability mass function for the possible values of Y given that $X = x$. More precisely, let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The conditional probability mass function provides the conditional probabilities for the values of Y in the set R_x.

Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

$$(1)\quad f_{Y|x}(y) \geq 0$$

$$(2)\quad \sum_{R_x} f_{Y|x}(y) = 1$$

$$(3)\quad P(Y = y \mid X = x) = f_{Y|x}(y) \tag{5-5}$$

EXAMPLE 5-6

For the joint probability distribution in Fig. 5-1, $f_{Y|x}(y)$ is found by dividing each $f_{XY}(x, y)$ by $f_X(x)$. Here, $f_X(x)$ is simply the sum of the probabilities in each column of Fig. 5-1. The function $f_{Y|x}(y)$ is shown in Fig. 5-3. In Fig. 5-3, each column sums to one because it is a probability distribution.

Properties of random variables can be extended to a conditional probability distribution of Y given $X = x$. The usual formulas for mean and variance can be applied to a conditional probability mass function.

Definition

Let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \sum_{R_x} y f_{Y|x}(y) \tag{5-6}$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \sum_{R_x} (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_{R_x} y^2 f_{Y|x}(y) - \sigma^2_{Y|x}$$

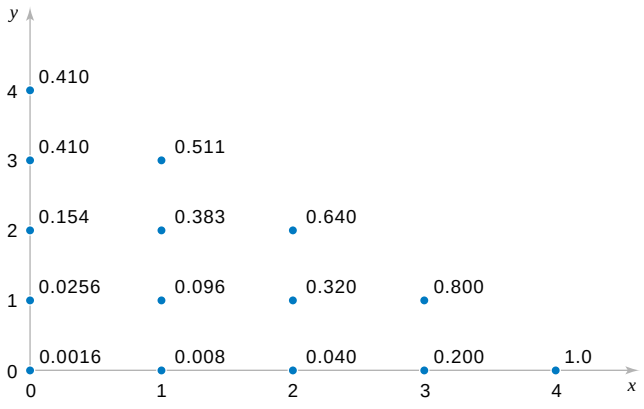


Figure 5-3
Conditional probability distributions of Y given $X = x$, $f_{Y|x}(y)$ in Example 5-6.

EXAMPLE 5-7

For the random variables in Example 5-1, the conditional mean of Y given $X = 2$ is obtained from the conditional distribution in Fig. 5-3:

$$E(Y|2) = \mu_{Y|2} = 0(0.040) + 1(0.320) + 2(0.640) = 1.6$$

The conditional mean is interpreted as the expected number of acceptable bits given that two of the four bits transmitted are suspect. The conditional variance of Y given $X = 2$ is

$$V(Y|2) = (0 - \mu_{Y|2})^2(0.040) + (1 - \mu_{Y|2})^2(0.320) + (2 - \mu_{Y|2})^2(0.640) = 0.32$$

5-1.4 Independence

In some random experiments, knowledge of the values of X does not change any of the probabilities associated with the values for Y.

EXAMPLE 5-8

In a plastic molding operation, each part is classified as to whether it conforms to color and length specifications. Define the random variable X and Y as

$$X = \begin{cases} 1 & \text{if the part conforms to color specifications} \\ 0 & \text{otherwise} \end{cases}$$

$$Y = \begin{cases} 1 & \text{if the part conforms to length specifications} \\ 0 & \text{otherwise} \end{cases}$$

Assume the joint probability distribution of X and Y is defined by $f_{XY}(x, y)$ in Fig. 5-4(a). The marginal probability distributions of X and Y are also shown in Fig. 5-4(a). Note that $f_{XY}(x, y) = f_X(x)\, f_Y(y)$. The conditional probability mass function $f_{Y|x}(y)$ is shown in Fig. 5-4(b). Notice that for any x, $f_{Y|x}(y) = f_Y(y)$. That is, knowledge of whether or not the part meets color specifications does not change the probability that it meets length specifications.

By analogy with independent events, we define two random variables to be **independent** whenever $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y. Notice that independence implies that $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for *all* x and y. If we find one pair of x and y in which the equality fails, X and Y are not independent. If two random variables are independent, then

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{f_X(x)\, f_Y(y)}{f_X(x)} = f_Y(y)$$

With similar calculations, the following equivalent statements can be shown.

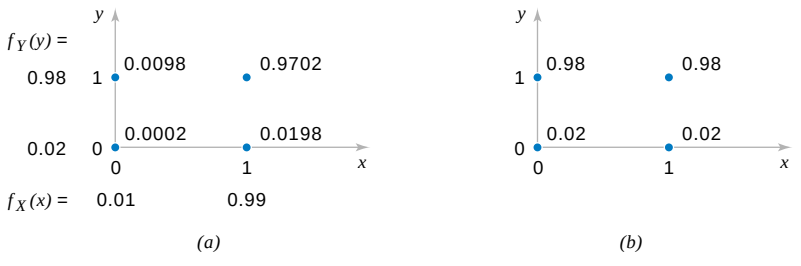


Figure 5-4 (a) Joint and marginal probability distributions of X and Y in Example 5-8. (b) Conditional probability distribution of Y given $X = x$ in Example 5-8.

For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-7)

Rectangular Range for (X, Y)!

If the set of points in two-dimensional space that receive positive probability under $f_{XY}(x, y)$ does not form a rectangle, X and Y are not independent because knowledge of X can restrict the range of values of Y that receive positive probability. In Example 5-1 knowledge that $X = 3$ implies that Y can equal only 0 or 1. Consequently, the marginal probability distribution of Y does not equal the conditional probability distribution $f_{Y|3}(y)$ for $X = 3$. Using this idea, we know immediately that the random variables X and Y with joint probability mass function in Fig. 5-1 are not independent. If the set of points in two-dimensional space that receives positive probability under $f_{XY}(x, y)$ forms a rectangle, independence is possible but not demonstrated. One of the conditions in Equation 5-7 must still be verified.

Rather than verifying independence from a joint probability distribution, knowledge of the random experiment is often used to assume that two random variables are independent. Then, the joint probability mass function of X and Y is computed from the product of the marginal probability mass functions.

EXAMPLE 5-9

In a large shipment of parts, 1% of the parts do not conform to specifications. The supplier inspects a random sample of 30 parts, and the random variable X denotes the number of parts in the sample that do not conform to specifications. The purchaser inspects another random sample of 20 parts, and the random variable Y denotes the number of parts in this sample that do not conform to specifications. What is the probability that $X = 1$ and $Y = 1$?

Although the samples are typically selected without replacement, if the shipment is large, relative to the sample sizes being used, approximate probabilities can be computed by assuming the sampling is with replacement and that X and Y are independent. With this assumption, the marginal probability distribution of X is binomial with $n = 30$ and $p = 0.01$, and the marginal probability distribution of Y is binomial with $n = 20$ and $p = 0.01$.

If independence between X and Y were not assumed, the solution would have to proceed as follows:

$$\begin{aligned} P(X = 1, Y = 1) &= P(X = 0, Y = 0) = P(X = 1, Y = 0) \\ &\quad = P(X = 0, Y = 1) = P(X = 1, Y = 1) \\ &= f_{XY}(0, 0) = f_{XY}(1, 0) = f_{XY}(0, 1) = f_{XY}(1, 1) \end{aligned}$$

However, with independence, property (4) of Equation 5-7 can be used as

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1)$$

and the binomial distributions for X and Y can be used to determine these probabilities as $P(X \le 1) \le 0.9639$ and $P(Y \le 1) \le 0.9831$. Therefore, $P(X \le 1, Y \le 1) \le 0.948$.

Consequently, the probability that the shipment is accepted for use in manufacturing is 0.948 even if 1% of the parts do not conform to specifications. If the supplier and the purchaser change their policies so that the shipment is acceptable only if zero nonconforming parts are found in the sample, the probability that the shipment is accepted for production is still quite high. That is,

$$P(X \le 0, Y \le 0) \le P(X \le 0)P(Y \le 0) \le 0.605$$

This example shows that inspection is not an effective means of achieving quality.

EXERCISES FOR SECTION 5-1

5-1. Show that the following function satisfies the properties of a joint probability mass function.

x	y	$f_{XY}(x, y)$
1	1	1/4
1.5	2	1/8
1.5	3	1/4
2.5	4	1/4
3	5	1/8

5-2. Continuation of Exercise 5-1. Determine the following probabilities:

(a) $P(X \le 2.5, Y \le 3)$ (b) $P(X \le 2.5)$

(c) $P(Y \le 3)$ (d) $P(X \le 1.8, Y \le 4.7)$

5-3. Continuation of Exercise 5-1. Determine $E(X)$ and $E(Y)$.

5-4. Continuation of Exercise 5-1. Determine

(a) The marginal probability distribution of the random variable X.

(b) The conditional probability distribution of Y given that $X \le 1.5$.

(c) The conditional probability distribution of X given that $Y \le 2$.

(d) $E(Y \mid X \le 1.5)$

(e) Are X and Y independent?

5-5. Determine the value of c that makes the function $f(x, y) \le c(x \le y)$ a joint probability mass function over the nine points with $x \le 1, 2, 3$ and $y \le 1, 2, 3$.

5-6. Continuation of Exercise 5-5. Determine the following probabilities:

(a) $P(X \le 1, Y \le 4)$ (b) $P(X \le 1)$

(c) $P(Y \le 2)$ (d) $P(X \le 2, Y \le 2)$

5-7. Continuation of Exercise 5-5. Determine $E(X)$, $E(Y)$, $V(X)$, and $V(Y)$.

5-8. Continuation of Exercise 5-5. Determine

(a) The marginal probability distribution of the random variable X.

(b) The conditional probability distribution of Y given that $X \le 1$.

(c) The conditional probability distribution of X given that $Y \le 2$.

(d) $E(Y \mid X \le 1)$

(e) Are X and Y independent?

5-9. Show that the following function satisfies the properties of a joint probability mass function.

x	y	$f_{XY}(x, y)$
−1	−2	1/8
−0.5	−1	1/4
0.5	1	1/2
1	2	1/8

5-10. Continuation of Exercise 5-9. Determine the following probabilities:

(a) $P(X \le 0.5, Y \le 1.5)$ (b) $P(X \le 0.5)$

(c) $P(Y \le 1.5)$ (d) $P(X \le 0.25, Y \le 4.5)$

5-11. Continuation of Exercise 5-9. Determine $E(X)$ and $E(Y)$.

5-12. Continuation of Exercise 5-9. Determine

(a) The marginal probability distribution of the random variable X.

(b) The conditional probability distribution of Y given that $X \le 1$.

(c) The conditional probability distribution of X given that $Y \le 1$.

(d) $E(X \mid y \le 1)$

(e) Are X and Y independent?

5-13. Four electronic printers are selected from a large lot of damaged printers. Each printer is inspected and classified as containing either a major or a minor defect. Let the random variables X and Y denote the number of printers with major and minor defects, respectively. Determine the range of the joint probability distribution of X and Y.

5-14. In the transmission of digital information, the probability that a bit has high, moderate, and low distortion is 0.01, 0.10, and 0.95, respectively. Suppose that three bits are transmitted and that the amount of distortion of each bit is assumed to be independent. Let X and Y denote the number of bits with high and moderate distortion out of the three, respectively. Determine

(a) $f_{XY}(x, y)$ (b) $f_X(x)$
(c) $E(XY)$ (d) $f_{Y|1}(y)$
(e) $E(Y|X = 1)$ (f) Are X and Y independent?

5-15. A small-business Web site contains 100 pages and 60%, 30%, and 10% of the pages contain low, moderate, and high graphic content, respectively. A sample of four pages is selected without replacement, and X and Y denote the number of pages with moderate and high graphics output in the sample. Determine

(a) $f_{XY}(x, y)$ (b) $f_X(x)$
(c) $E(XY)$ (d) $f_{Y|3}(y)$
(e) $E(Y|X = 3)$ (f) $V(Y|X = 3)$
(g) Are X and Y independent?

5-16. A manufacturing company employs two inspecting devices to sample a fraction of their output for quality control purposes. The first inspection monitor is able to accurately detect 99.3% of the defective items it receives, whereas the second is able to do so in 99.7% of the cases. Assume that four defective items are produced and sent out for inspection. Let X and Y denote the number of items that will be identified as defective by inspecting devices 1 and 2, respectively. Assume the devices are independent. Determine

(a) $f_{XY}(x, y)$ (b) $f_X(x)$
(c) $E(XY)$ (d) $f_{Y|2}(y)$
(e) $E(Y|X = 2)$ (f) $V(Y|X = 2)$
(g) Are X and Y independent?

5-2 MULTIPLE DISCRETE RANDOM VARIABLES

5-2.1 Joint Probability Distributions

EXAMPLE 5-10

In some cases, more than two random variables are defined in a random experiment, and the concepts presented earlier in the chapter can easily be extended. The notation can be cumbersome and if doubts arise, it is helpful to refer to the equivalent concept for two random variables. Suppose that the quality of each bit received in Example 5-1 is categorized even more finely into one of the four classes, excellent, good, fair, or poor, denoted by E, G, F, and P, respectively. Also, let the random variables X_1, X_2, X_3, and X_4 denote the number of bits that are E, G, F, and P, respectively, in a transmission of 20 bits. In this example, we are interested in the joint probability distribution of four random variables. Because each of the 20 bits is categorized into one of the four classes, only values for x_1, x_2, x_3, and x_4 such that $x_1 + x_2 + x_3 + x_4 = 20$ receive positive probability in the probability distribution.

In general, given discrete random variables $X_1, X_2, X_3, \ldots, X_p$, the joint probability distribution of $X_1, X_2, X_3, \ldots, X_p$ is a description of the set of points $(x_1, x_2, x_3, \ldots, x_p)$ in the range of $X_1, X_2, X_3, \ldots, X_p$, along with the probability of each point. A joint probability mass function is a simple extension of a bivariate probability mass function.

Definition

The **joint probability mass function** of $X_1, X_2, \ldots, X_p$ is

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) \qquad (5\text{-}8)$$

for all points $(x_1, x_2, \ldots, x_p)$ in the range of $X_1, X_2, \ldots, X_p$.

A marginal probability distribution is a simple extension of the result for two random variables.

Definition

If $X_1, X_2, X_3, \ldots, X_p$ are discrete random variables with joint probability mass function $f_{X_1X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$, the **marginal probability mass function** of any X_i is

$$f_{X_i}(x_i) = P(X_i = x_i) = \sum_{R_{x_i}} f_{X_1X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) \tag{5-9}$$

where R_{x_i} denotes the set of points in the range of $(X_1, X_2, \ldots, X_p)$ for which $X_i = x_i$.

EXAMPLE 5-11

Points that have positive probability in the joint probability distribution of three random variables X_1, X_2, X_3 are shown in Fig. 5-5. The range is the nonnegative integers with $x_1 + x_2 + x_3 \leq 3$. The marginal probability distribution of X_2 is found as follows.

$$P(X_2 = 0) = f_{X_1X_2X_3}(3, 0, 0) + f_{X_1X_2X_3}(0, 0, 3) + f_{X_1X_2X_3}(1, 0, 2) + f_{X_1X_2X_3}(2, 0, 1)$$
$$P(X_2 = 1) = f_{X_1X_2X_3}(2, 1, 0) + f_{X_1X_2X_3}(0, 1, 2) + f_{X_1X_2X_3}(1, 1, 1)$$
$$P(X_2 = 2) = f_{X_1X_2X_3}(1, 2, 0) + f_{X_1X_2X_3}(0, 2, 1)$$
$$P(X_2 = 3) = f_{X_1X_2X_3}(0, 3, 0)$$

Furthermore, $E(X_i)$ and $V(X_i)$ for $i = 1, 2, \ldots, p$ can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

Mean and Variance from Joint Distribution

$$E(X_i) = \sum_{R} x_i f_{X_1X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$$

and

$$V(X_i) = \sum_{R} (x_i - \mu_{X_i})^2 f_{X_1X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) \tag{5-10}$$

where R is the set of all points in the range of $X_1, X_2, \ldots, X_p$.

With several random variables, we might be interested in the probability distribution of some subset of the collection of variables. The probability distribution of $X_1, X_2, \ldots, X_k, k < p$ can be obtained from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

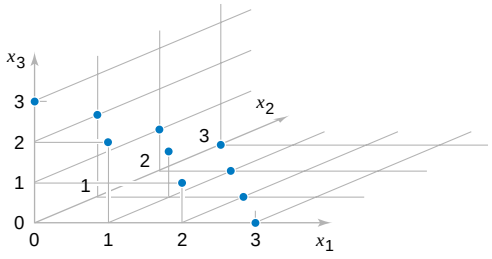


Figure 5-5 Joint probability distribution of X_1, X_2, and X_3.

Distribution of a Subset of Random Variables

If $X_1, X_2, X_3, \ldots, X_p$ are discrete random variables with joint probability mass function $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$, the joint **probability mass function** of $X_1, X_2, \ldots, X_k$, $k < p$, is

$$f_{X_1 X_2 \ldots X_k}(x_1, x_2, \ldots, x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \sum_{R_{x_1 x_2 \ldots x_k}} P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) \quad (5\text{-}11)$$

where $R_{x_1 x_2 \ldots x_k}$ denotes the set of all points in the range of $X_1, X_2, \ldots, X_p$ for which $X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k$.

That is, $P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$ is the sum of the probabilities over all points in the range of $X_1, X_2, X_3, \ldots, X_p$ for which $X_1 = x_1, X_2 = x_2, \ldots$, and $X_k = x_k$. An example is presented in the next section. Any k random variables can be used in the definition. The first k simplifies the notation.

Conditional Probability Distributions

Conditional probability distributions can be developed for multiple discrete random variables by an extension of the ideas used for two discrete random variables. For example, the conditional joint probability mass function of X_1, X_2, X_3 given X_4, X_5 is

$$f_{X_1 X_2 X_3 | x_4 x_5}(x_1, x_2, x_3) = \frac{f_{X_1 X_2 X_3 X_4 X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4 X_5}(x_4, x_5)}$$

for $f_{X_4 X_5}(x_4, x_5) > 0$. The conditional joint probability mass function of X_1, X_2, X_3 given X_4, X_5 provides the conditional probabilities at all points in the range of X_1, X_2, X_3, X_4, X_5 for which $X_4 = x_4$ and $X_5 = x_5$.

The concept of independence can be extended to multiple discrete random variables.

Definition

Discrete variables $X_1, X_2, \ldots, X_p$ are **independent** if and only if

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_p}(x_p) \quad (5\text{-}12)$$

for *all* $x_1, x_2, \ldots, x_p$.

Similar to the result for bivariate random variables, independence implies that Equation 5-12 holds for *all* $x_1, x_2, \ldots, x_p$. If we find one point for which the equality fails, $X_1, X_2, \ldots, X_p$ are not independent. It can be shown that if $X_1, X_2, \ldots, X_p$ are independent,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = P(X_1 \in A_1) P(X_2 \in A_2) \cdots P(X_p \in A_p)$$

for *any* sets $A_1, A_2, \ldots, A_p$.

5-2.2 Multinomial Probability Distribution

A joint probability distribution for multiple discrete random variables that is quite useful is an extension of the binomial. The random experiment that generates the probability distribution consists of a series of independent trials. However, the results from each trial can be categorized into one of k classes.

EXAMPLE 5-12

We might be interested in a probability such as the following. Of the 20 bits received, what is the probability that 14 are excellent, 3 are good, 2 are fair, and 1 is poor? Assume that the classifications of individual bits are independent events and that the probabilities of E, G, F, and P are 0.6, 0.3, 0.08, and 0.02, respectively. One sequence of 20 bits that produces the specified numbers of bits in each class can be represented as

$$EEEEEEEEEEEEEEGGGFFP$$

Using independence, we find that the probability of this sequence is

$$P(EEEEEEEEEEEEEEGGGFFP) = 0.6^{14}0.3^{3}0.08^{2}0.02^{1} = 2.708 \times 10^{-9}$$

Clearly, all sequences that consist of the same numbers of E's, G's, F's, and P's have the same probability. Consequently, the requested probability can be found by multiplying 2.708×10^{-9} by the number of sequences with 14 E's, three G's, two F's, and one P. The number of sequences is found from the CD material for Chapter 2 to be

$$\frac{20!}{14!3!2!1!} = 2325600$$

Therefore, the requested probability is

$$P(\text{14 } E\text{'s, three } G\text{'s, two } F\text{'s, and one } P) = 2325600(2.708 \times 10^{-9}) = 0.0063$$

Example 5-12 leads to the following generalization of a binomial experiment and a binomial distribution.

Multinomial Distribution

Suppose a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.

(2) The probability of a trial generating a result in class 1, class 2, ..., class k is constant over the trials and equal to $p_1, p_2, \ldots, p_k$, respectively.

(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, ..., class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \qquad (5\text{-}13)$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

The multinomial distribution is considered a multivariable extension of the binomial distribution.

EXAMPLE 5-13

In Example 5-12, let the random variables X_1, X_2, X_3, and X_4 denote the number of bits that are E, G, F, and P, respectively, in a transmission of 20 bits. The probability that 12 of the bits received are E, 6 are G, 2 are F, and 0 are P is

$$P(X_1 = 12, X_2 = 6, X_3 = 2, X_4 = 0) = \frac{20!}{12!6!2!0!}\, 0.6^{12}0.3^{6}0.08^{2}0.02^{0} = 0.0358$$

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with

$$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \tag{5-14}$$

EXAMPLE 5-14

In Example 5-13, the marginal probability distribution of X_2 is binomial with $n = 20$ and $p = 0.3$. Furthermore, the joint marginal probability distribution of X_2 and X_3 is found as follows. The $P(X_2 = x_2, X_3 = x_3)$ is the probability that exactly x_2 trials result in G and that x_3 result in F. The remaining $n - x_2 - x_3$ trials must result in either E or P. Consequently, we can consider each trial in the experiment to result in one of three classes, $\{G\}$, $\{F\}$, or $\{E, P\}$, with probabilities 0.3, 0.08, and $0.6 + 0.02 = 0.62$, respectively. With these new classes, we can consider the trials to comprise a new multinomial experiment. Therefore,

$$f_{X_2X_3}(x_2, x_3) = P(X_2 = x_2, X_3 = x_3) = \frac{n!}{x_2!x_3!(n - x_2 - x_3)!}\, 0.3^{x_2}0.08^{x_3}0.62^{n-x_2-x_3}$$

The joint probability distribution of other sets of variables can be found similarly.

EXERCISES FOR SECTION 5-2

5-17. Suppose the random variables X, Y, and Z have the following joint probability distribution

x	y	z	$f(x, y, z)$
1	1	1	0.05
1	1	2	0.10
1	2	1	0.15
1	2	2	0.20
2	1	1	0.20
2	1	2	0.15
2	2	1	0.10
2	2	2	0.05

Determine the following:

(a) $P(X = 2)$ (b) $P(X = 1, Y = 2)$

(c) $P(Z = 1.5)$ (d) $P(X = 1 \text{ or } Z = 2)$

(e) $E(X)$

5-18. Continuation of Exercise 5-17. Determine the following:

(a) $P(X = 1 \mid Y = 1)$ (b) $P(X = 1, Y = 1 \mid Z = 2)$

(c) $P(X = 1 \mid Y = 1, Z = 2)$

5-19. Continuation of Exercise 5-17. Determine the conditional probability distribution of X given that $Y = 1$ and $Z = 2$.

5-20. Based on the number of voids, a ferrite slab is classified as either high, medium, or low. Historically, 5% of the slabs are classified as high, 85% as medium, and 10% as low.

A sample of 20 slabs is selected for testing. Let X, Y, and Z denote the number of slabs that are independently classified as high, medium, and low, respectively.

(a) What is the name and the values of the parameters of the joint probability distribution of X, Y, and Z?

(b) What is the range of the joint probability distribution of X, Y, Z?

(c) What is the name and the values of the parameters of the marginal probability distribution of X?

(d) Determine $E(X)$ and $V(X)$.

5-21. Continuation of Exercise 5-20. Determine the following:

(a) $P(X = 1, Y = 17, Z = 3)$

(b) $P(X = 1, Y = 17, Z = 3)$

(c) $P(X = 1)$

(d) $E(X)$

5-22. Continuation of Exercise 5-20. Determine the following:

(a) $P(X = 2, Z = 3 \mid Y = 17)$ (b) $P(X = 2 \mid Y = 17)$

(c) $E(X \mid Y = 17)$

5-23. An order of 15 printers contains four with a graphics-enhancement feature, five with extra memory, and six with both features. Four printers are selected at random, without replacement, from this set. Let the random variables X, Y, and Z denote the number of printers in the sample with graphics enhancement only, extra memory only, and both, respectively.

(a) Describe the range of the joint probability distribution of X, Y, and Z.

(b) Is the probability distribution of X, Y, and Z multinomial? Why or why not?

5-24. Continuation of Exercise 5-23. Determine the conditional probability distribution of X given that $Y = 2$.

5-25. Continuation of Exercise 5-23. Determine the following:

(a) $P(X = 1, Y = 2, Z = 1)$ (b) $P(X = 1, Y = 1)$

(c) $E(X)$ and $V(X)$

5-26. Continuation of Exercise 5-23. Determine the following:

(a) $P(X = 1, Y = 2 \mid Z = 1)$ (b) $P(X = 2 \mid Y = 2)$

(c) The conditional probability distribution of X given that $Y = 0$ and $Z = 3$.

5-27. Four electronic ovens that were dropped during shipment are inspected and classified as containing either a major, a minor, or no defect. In the past, 60% of dropped ovens had a major defect, 30% had a minor defect, and 10% had no defect. Assume that the defects on the four ovens occur independently.

(a) Is the probability distribution of the count of ovens in each category multinomial? Why or why not?

(b) What is the probability that, of the four dropped ovens, two have a major defect and two have a minor defect?

(c) What is the probability that no oven has a defect?

5-28. Continuation of Exercise 5-27. Determine the following:

(a) The joint probability mass function of the number of ovens with a major defect and the number with a minor defect.

(b) The expected number of ovens with a major defect.

(c) The expected number of ovens with a minor defect.

5-29. Continuation of Exercise 5-27. Determine the following:

(a) The conditional probability that two ovens have major defects given that two ovens have minor defects

(b) The conditional probability that three ovens have major defects given that two ovens have minor defects

(c) The conditional probability distribution of the number of ovens with major defects given that two ovens have minor defects

(d) The conditional mean of the number of ovens with major defects given that two ovens have minor defects

5-30. In the transmission of digital information, the probability that a bit has high, moderate, or low distortion is 0.01, 0.04, and 0.95, respectively. Suppose that three bits are transmitted and that the amount of distortion of each bit is assumed to be independent.

(a) What is the probability that two bits have high distortion and one has moderate distortion?

(b) What is the probability that all three bits have low distortion?

5-31. Continuation of Exercise 5-30. Let X and Y denote the number of bits with high and moderate distortion out of the three transmitted, respectively. Determine the following:

(a) The probability distribution, mean and variance of X.

(b) The conditional probability distribution, conditional mean and conditional variance of X given that $Y = 2$.

5-32. A marketing company performed a risk analysis for a manufacturer of synthetic fibers and concluded that new competitors present no risk 13% of the time (due mostly to the diversity of fibers manufactured), moderate risk 72% of the time (some overlapping of products), and very high risk (competitor manufactures the exact same products) 15% of the time. It is known that 12 international companies are planning to open new facilities for the manufacture of synthetic fibers within the next three years. Assume the companies are independent. Let X, Y, and Z denote the number of new competitors that will pose no, moderate, and very high risk for the interested company, respectively.

(a) What is the range of the joint probability distribution of X, Y, and Z?

(b) Determine $P(X = 1, Y = 3, Z = 1)$

(c) Determine $P(Z = 2)$

5-33. Continuation of Exercise 5-32. Determine the following:

(a) $P(Z = 2 \mid Y = 1, X = 10)$ (b) $P(Z = 1 \mid X = 10)$

(c) $P(Y = 1, Z = 1 \mid X = 10)$ (d) $E(Z \mid X = 10)$

5-3 TWO CONTINUOUS RANDOM VARIABLES

5-3.1 Joint Probability Distributions

Our presentation of the joint probability distribution of two continuous random variables is similar to our discussion of two discrete random variables. As an example, let the continuous random variable X denote the length of one dimenson of an injection-molded part, and let the continuous random variable Y denote the length of another dimension. The sample space of the random experiment consists of points in two dimensions.

We can study each random variable separately. However, because the two random variables are measurements from the same part, small disturbances in the injection-molding process, such as pressure and temperature variations, might be more likely to generate values for X and Y in specific regions of two-dimensional space. For example, a small pressure increase might generate parts such that both X and Y are greater than their respective targets and a small pressure decrease might generate parts such that X and Y are both less than their respective targets. Therefore, based on pressure variations, we expect that the probability of a part with X much greater than its target and Y much less than its target is small. Knowledge of the joint probability distribution of X and Y provides information that is not obvious from the marginal probability distributions.

The joint probability distribution of two continuous random variables X and Y can be specified by providing a method for calculating the probability that X and Y assume a value in any region R of two-dimensional space. Analogous to the probability density function of a single continuous random variable, a **joint probability density function** can be defined over two-dimensional space. The double integral of $f_{XY}(x, y)$ over a region R provides the probability that (X, Y) assumes a value in R. This integral can be interpreted as the volume under the surface $f_{XY}(x, y)$ over the region R.

A joint probability density function for X and Y is shown in Fig. 5-6. The probability that (X, Y) assumes a value in the region R equals the volume of the shaded region in Fig. 5-6. In this manner, a joint probability density function is used to determine probabilities for X and Y.

Definition

A **joint probability density function** for the continuous random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies the following properties:

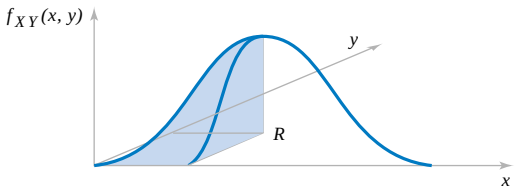
(1) $f_{XY}(x, y) \geq 0$ for all x, y

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1$

(3) For any region R of two-dimensional space

$$P((X, Y) \in R) = \iint_{R} f_{XY}(x, y)\, dx\, dy \tag{5-15}$$

Typically, $f_{XY}(x, y)$ is defined over all of two-dimensional space by assuming that $f_{XY}(x, y) = 0$ for all points for which $f_{XY}(x, y)$ is not specified.



Probability that (X, Y) is in the region R is determined by the volume of $f_{XY}(x, y)$ over the region R.

Figure 5-6 Joint probability density function for random variables X and Y.

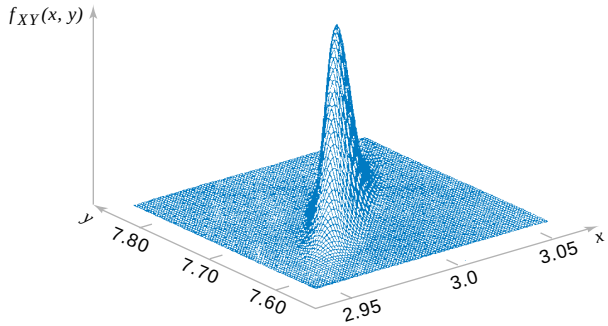


Figure 5-7 Joint probability density function for the lengths of different dimensions of an injection-molded part.

At the start of this chapter, the lengths of different dimensions of an injection-molded part were presented as an example of two random variables. Each length might be modeled by a normal distribution. However, because the measurements are from the same part, the random variables are typically not independent. A probability distribution for two normal random variables that are not independent is important in many applications and it is presented later in this chapter. If the specifications for X and Y are 2.95 to 3.05 and 7.60 to 7.80 millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 \le X \le 3.05, 7.60 \le Y \le 7.80)$. Suppose that $f_{XY}(x, y)$ is shown in Fig. 5-7. The required probability is the volume of $f_{XY}(x, y)$ within the specifications. Often a probability such as this must be determined from a numerical integration.

EXAMPLE 5-15

Let the random variable X denote the time until a computer server connects to your machine (in milliseconds), and let Y denote the time until the server authorizes you as a valid user (in milliseconds). Each of these random variables measures the wait from a common starting time and $X \le Y$. Assume that the joint probability density function for X and Y is

$$f_{XY}(x, y) = 6 \times 10^{-6} \exp(-0.001x - 0.002y) \quad \text{for } x \le y$$

Reasonable assumptions can be used to develop such a distribution, but for now, our focus is only on the joint probability density function.

The region with nonzero probability is shaded in Fig. 5-8. The property that this joint probability density function integrates to 1 can be verified by the integral of $f_{XY}(x, y)$ over this region as follows:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = \int_{0}^{\infty}\left[\int_{x}^{\infty} 6 \times 10^{-6} e^{-0.001x-0.002y}\, dy\right] dx$$

$$= 6 \times 10^{-6}\int_{0}^{\infty}\left[\int_{x}^{\infty} e^{-0.002y}\, dy\right] e^{-0.001x}\, dx$$

$$= 6 \times 10^{-6}\int_{0}^{\infty}\left[\frac{e^{-0.002x}}{0.002}\right] e^{-0.001x}\, dx$$

$$= 0.003\int_{0}^{\infty} e^{-0.003x}\, dx = 0.003\left(\frac{1}{0.003}\right) = 1$$

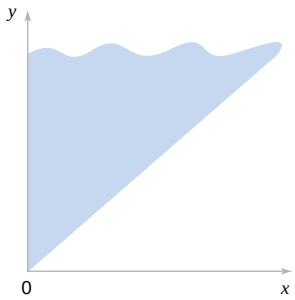


Figure 5-8 The joint probability density function of X and Y is nonzero over the shaded region.

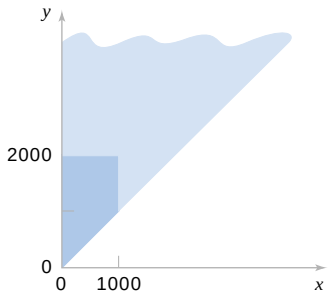


Figure 5-9 Region of integration for the probability that $X \leq 1000$ and $Y \leq 2000$ is darkly shaded.

The probability that $X \leq 1000$ and $Y \leq 2000$ is determined as the integral over the darkly shaded region in Fig. 5-9.

$$P(X \leq 1000, Y \leq 2000) = \int_0^{1000} \int_x^{2000} f_{XY}(x, y)\, dy\, dx$$

$$= 6 \times 10^{-6} \int_0^{1000} \left(\int_x^{2000} e^{-0.002y}\, dy \right) e^{-0.001x}\, dx$$

$$= 6 \times 10^{-6} \int_0^{1000} \left(\frac{e^{-0.002x} - e^{-4}}{0.002} \right) e^{-0.001x}\, dx$$

$$= 0.003 \int_0^{1000} e^{-0.003x} - e^{-4} e^{-0.001x}\, dx$$

$$= 0.003 \left[\left(\frac{1 - e^{-3}}{0.003} \right) - e^{-4} \left(\frac{1 - e^{-1}}{0.001} \right) \right]$$

$$= 0.003(316.738 - 11.578) = 0.915$$

5-3.2 Marginal Probability Distributions

Similar to joint discrete random variables, we can find the marginal probability distributions of X and Y from the joint probability distribution.

Definition

If the joint probability density function of continuous random variables X and Y is $f_{XY}(x, y)$, the **marginal probability density functions** of X and Y are

$$f_X(x) = \int_{R_x} f_{XY}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_{R_y} f_{XY}(x, y)\, dx \tag{5-16}$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R_y denotes the set of all points in the range of (X, Y) for which $Y = y$

A probability involving only one random variable, say, for example, $P(a \le X \le b)$, can be found from the marginal probability distribution of X or from the joint probability distribution of X and Y. For example, $P(a \le X \le b)$ equals $P(a \le X \le b, -\infty < Y < \infty)$. Therefore,

$$P(a \le X \le b) = \int_a^b \int_{R_x} f_{XY}(x, y)\, dy\, dx = \int_a^b \left[\int_{R_x} f_{XY}(x, y)\, dy \right] dx = \int_a^b f_X(x)\, dx$$

Similarly, $E(X)$ and $V(X)$ can be obtained directly from the joint probability distribution of X and Y or by first calculating the marginal probability distribution of X. The details, shown in the following equations, are similar to those used for discrete random variables.

Mean and Variance from Joint Distribution

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} x \left[\int_{R_x} f_{XY}(x, y)\, dy \right] dx$$

$$= \iint_R x f_{XY}(x, y)\, dx\, dy \qquad (5\text{-}17)$$

and

$$V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx = \int_{-\infty}^{\infty} (x - \mu_X)^2 \left[\int_{R_x} f_{XY}(x, y)\, dy \right] dx$$

$$= \iint_R (x - \mu_X)^2 f_{XY}(x, y)\, dx\, dy$$

where R_X denotes the set of all points in the range of (X, Y) for which $X = x$ and R_Y denotes the set of all points in the range of (X, Y)

EXAMPLE 5-16

For the random variables that denote times in Example 5-15, calculate the probability that Y exceeds 2000 milliseconds.

This probability is determined as the integral of $f_{XY}(x, y)$ over the darkly shaded region in Fig. 5-10. The region is partitioned into two parts and different limits of integration are determined for each part.

$$P(Y > 2000) = \int_0^{2000} \left[\int_{2000}^{\infty} 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy \right] dx$$

$$= \int_{2000}^{\infty} \left[\int_{x}^{\infty} 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy \right] dx$$

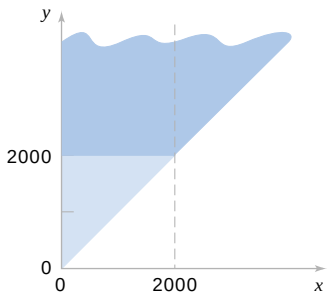


Figure 5-10 Region of integration for the probability that $Y \le 2000$ is darkly shaded and it is partitioned into two regions with $x \le 2000$ and $x \ge 2000$.

The first integral is

$$6 \times 10^{-6} \int_0^{2000} \left(\frac{e^{-0.002y}}{-0.002} \right)_{2000}^{\infty} e^{-0.001x}\, dx = \frac{6 \times 10^{-6}}{0.002} e^{-4} \int_0^{2000} e^{-0.001x}\, dx$$

$$= \frac{6 \times 10^{-6}}{0.002} e^{-4} \left(\frac{1 - e^{-2}}{0.001} \right) = 0.0475$$

The second integral is

$$6 \times 10^{-6} \int_{2000}^{\infty} \left(\frac{e^{-0.002y}}{-0.002} \right)_{x}^{\infty} e^{-0.001x}\, dx = \frac{6 \times 10^{-6}}{0.002} \int_{2000}^{\infty} e^{-0.003x}\, dx$$

$$= \frac{6 \times 10^{-6}}{0.002} \left(\frac{e^{-6}}{0.003} \right) = 0.0025$$

Therefore,

$$P(Y \le 2000) = 0.0475 + 0.0025 = 0.05.$$

Alternatively, the probability can be calculated from the marginal probability distribution of Y as follows. For $y \ge 0$

$$f_Y(y) = \int_0^y 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dx = 6 \times 10^{-6} e^{-0.002y} \int_0^y e^{-0.001x}\, dx$$

$$= 6 \times 10^{-6} e^{-0.002y} \left(\frac{e^{-0.001x}}{-0.001} \right)_0^y = 6 \times 10^{-6} e^{-0.002y} \left(\frac{1 - e^{-0.001y}}{0.001} \right)$$

$$= 6 \times 10^{-3} e^{-0.002y} (1 - e^{-0.001y}) \quad \text{for } y \ge 0$$

We have obtained the marginal probability density function of Y. Now,

$$P(Y > 2000) = 6 \times 10^{-3} \int_{2000}^{\infty} e^{-0.002y}(1 - e^{-0.001y})\,dy$$

$$= 6 \times 10^{-3} \left[\left(\frac{e^{-0.002y}}{-0.002} \right)_{2000}^{\infty} - \left(\frac{e^{-0.003y}}{-0.003} \right)_{2000}^{\infty} \right]$$

$$= 6 \times 10^{-3} \left(\frac{e^{-4}}{0.002} - \frac{e^{-6}}{0.003} \right) = 0.05$$

5-3.3 Conditional Probability Distributions

Analogous to discrete random variables, we can define the conditional probability distribution of Y given $X = x$.

Definition

Given continuous random variables X and Y with joint probability density function $f_{XY}(x, y)$, the **conditional probability density function** of Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{for} \quad f_X(x) > 0 \tag{5-18}$$

The function $f_{Y|x}(y)$ is used to find the probabilities of the possible values for Y given that $X = x$. Let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The conditional probability density function provides the conditional probabilities for the values of Y in the set R_x.

Because the conditional probability density function $f_{Y|x}(y)$ is a probability density function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\int_{R_x} f_{Y|x}(y)\,dy = 1$

(3) $P(Y \in B \mid X = x) = \int_{B} f_{Y|x}(y)\,dy$ for any set B in the range of Y

(5-19)

It is important to state the region in which a joint, marginal, or conditional probability density function is not zero. The following example illustrates this.

EXAMPLE 5-17 For the random variables that denote times in Example 5-15, determine the conditional probability density function for Y given that $X = x$.

First the marginal density function of x is determined. For $x > 0$

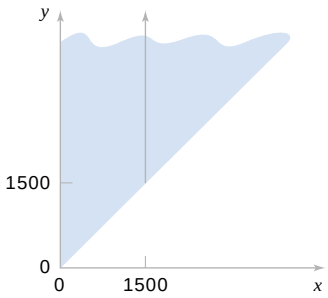


Figure 5-11 The conditional probability density function for Y, given that $x = 1500$, is nonzero over the solid line.

$$f_X(x) = \int_x^{\infty} 6 \times 10^{-6} e^{-0.001x-0.002y}\,dy = 6 \times 10^{-6} e^{-0.001x}\left(\frac{e^{-0.002y}}{-0.002}\right)\Big|_x^{\infty}$$

$$= 6 \times 10^{-6} e^{-0.001x}\left(\frac{e^{-0.002x}}{0.002}\right) = 0.003e^{-0.003x} \quad \text{for} \quad x > 0$$

This is an exponential distribution with $\theta = 0.003$. Now, for $0 < x$ and $x < y$ the conditional probability density function is

$$f_{Y|x}(y) = f_{XY}(x,y)/f_x(x) = \frac{6 \times 10^{-6} e^{-0.001x-0.002y}}{0.003e^{-0.003x}}$$

$$= 0.002e^{0.002x-0.002y} \quad \text{for } 0 < x \quad \text{and} \quad x < y$$

The conditional probability density function of Y, given that $x = 1500$, is nonzero on the solid line in Fig. 5-11.

Determine the probability that Y exceeds 2000, given that $x = 1500$. That is, determine $P(Y > 2000 \mid x = 1500)$. The conditional probability density function is integrated as follows:

$$P(Y > 2000 \mid x = 1500) = \int_{2000}^{\infty} f_{Y|1500}(y)\,dy = \int_{2000}^{\infty} 0.002e^{0.002(1500)-0.002y}\,dy$$

$$= 0.002e^{3}\left(\frac{e^{-0.002y}}{-0.002}\right)\Big|_{2000}^{\infty} = 0.002e^{3}\left(\frac{e^{-4}}{0.002}\right) = 0.368$$

Definition

Let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The **conditional mean** of Y given $X = x$, denoted as $E(Y \mid x)$ or $\mu_{Y|x}$, is

$$E(Y \mid x) = \int_{R_x} y f_{Y|x}(y)\,dy$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y \mid x)$ or $\sigma^2_{Y|x}$, is

$$V(Y \mid x) = \int_{R_x} (y - \mu_{Y|x})^2 f_{Y|x}(y)\,dy = \int_{R_x} y^2 f_{Y|x}(y)\,dy - \sigma^2_{Y|x} \qquad (5\text{-}20)$$

EXAMPLE 5-18

For the random variables that denote times in Example 5-15, determine the conditional mean for Y given that $x = 1500$.

The conditional probability density function for Y was determined in Example 5-17. Because $f_{Y|1500}(y)$ is nonzero for $y > 1500$,

$$E(Y \mid x = 1500) = \int_{1500}^{\infty} y(0.002e^{0.002(1500) - 0.002y})\,dy = 0.002e^{3} \int_{1500}^{\infty} ye^{-0.002y}\,dy$$

Integrate by parts as follows:

$$\int_{1500}^{\infty} ye^{-0.002y}\,dy = y\frac{e^{-0.002y}}{-0.002}\Big|_{1500}^{\infty} - \int_{1500}^{\infty} \left(\frac{e^{-0.002y}}{-0.002}\right) dy$$

$$= \frac{1500}{0.002}e^{-3} - \left(\frac{e^{-0.002y}}{(-0.002)(-0.002)}\right)\Big|_{1500}^{\infty}$$

$$= \frac{1500}{0.002}e^{-3} - \frac{e^{-3}}{(0.002)(0.002)} = \frac{e^{-3}}{0.002}(2000)$$

With the constant $0.002e^{3}$ reapplied

$$E(Y \mid x = 1500) = 2000$$

5-3.4 Independence

The definition of independence for continuous random variables is similar to the definition for discrete random variables. If $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y, X and Y are **independent.** Independence implies that $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for *all* x and y. If we find one pair of x and y in which the equality fails, X and Y are not independent.

Definition

For continuous random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are said to be **independent.**

(1) $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-21)

EXAMPLE 5-19

For the joint distribution of times in Example 5-15, the

- Marginal distribution of Y was determined in Example 5-16.
- Conditional distribution of Y given $X = x$ was determined in Example 5-17.

Because the marginal and conditional probability densities are not the same for all values of x, property (2) of Equation 5-20 implies that the random variables are not independent. The

fact that these variables are not independent can be determined quickly by noticing that the range of (X, Y), shown in Fig. 5-8, is not rectangular. Consequently, knowledge of X changes the interval of values for Y that receives nonzero probability.

EXAMPLE 5-20

Suppose that Example 5-15 is modified so that the joint probability density function of X and Y is $f_{XY}(x, y) = 2 \times 10^{-6} e^{-0.001x - 0.002y}$ for x $\geq$ 0 and $y \geq 0$. Show that X and Y are independent and determine $P(X \leq 1000, Y \leq 1000)$.

The marginal probability density function of X is

$$f_X(x) = \int_0^\infty 2 \times 10^{-6} e^{-0.001x - 0.002y}\, dy = 0.001 e^{-0.001x} \quad \text{for } x \geq 0$$

The marginal probability density function of y is

$$f_Y(y) = \int_0^\infty 2 \times 10^{-6} e^{-0.001x - 0.002y}\, dx = 0.002 e^{-0.002y} \quad \text{for } y \geq 0$$

Therefore, $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y and X and Y are independent.

To determine the probability requested, property (4) of Equation 5-21 and the fact that each random variable has an exponential distribution can be applied.

$$P(X \leq 1000, Y \leq 1000) = P(X \leq 1000)P(Y \leq 1000) = e^{-1}(1 - e^{-2}) = 0.318$$

Often, based on knowledge of the system under study, random variables are assumed to be independent. Then, probabilities involving both variables can be determined from the marginal probability distributions.

EXAMPLE 5-21

Let the random variables X and Y denote the lengths of two dimensions of a machined part, respectively. Assume that X and Y are independent random variables, and further assume that the distribution of X is normal with mean 10.5 millimeters and variance 0.0025 (millimeter)^2 and that the distribution of Y is normal with mean 3.2 millimeters and variance 0.0036 (millimeter)^2. Determine the probability that $10.4 \leq X \leq 10.6$ and $3.15 \leq Y \leq 3.25$.

Because X and Y are independent,

$$P(10.4 \leq X \leq 10.6, 3.15 \leq Y \leq 3.25) = P(10.4 \leq X \leq 10.6)P(3.15 \leq Y \leq 3.25)$$

$$= P\left(\frac{10.4 - 10.5}{0.05} \leq Z \leq \frac{10.6 - 10.5}{0.05}\right) P\left(\frac{3.15 - 3.2}{0.06} \leq Z \leq \frac{3.25 - 3.2}{0.06}\right)$$

$$= P(-2 \leq Z \leq 2)P(-0.833 \leq Z \leq 0.833) = 0.566$$

where Z denotes a standard normal random variable.

EXERCISES FOR SECTION 5-3

5-34. Determine the value of c such that the function $f(x, y) = cxy$ for $0 \leq x \leq 3$ and $0 \leq y \leq 3$ satisfies the properties of a joint probability density function.

5-35. Continuation of Exercise 5-34. Determine the following:

(a) $P(X \leq 2, Y \leq 3)$
(b) $P(X \leq 2.5)$
(c) $P(1 \leq Y \leq 2.5)$
(d) $P(X \leq 1.8, 1 \leq Y \leq 2.5)$
(e) $E(X)$
(f) $P(X \leq 0, Y \leq 4)$

5-36. Continuation of Exercise 5-34. Determine the following:
(a) Marginal probability distribution of the random variable X
(b) Conditional probability distribution of Y given that $X = 1.5$
(c) $E(Y|X = 1.5)$
(d) $P(Y = 2|X = 1.5)$
(e) Conditional probability distribution of X given that $Y = 2$

5-37. Determine the value of c that makes the function $f(x, y) = c(x + y)$ a joint probability density function over the range $0 \le x \le 3$ and $x \le y \le x + 2$.

5-38. Continuation of Exercise 5-37. Determine the following:
(a) $P(X = 1, Y = 2)$ (b) $P(1 \le X \le 2)$
(c) $P(Y = 1)$ (d) $P(X = 2, Y = 2)$
(e) $E(X)$

5-39. Continuation of Exercise 5-37. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given that $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y = 2|X = 1)$
(e) Conditional probability distribution of X given that $Y = 2$

5-40. Determine the value of c that makes the function $f(x, y) = cxy$ a joint probability density function over the range $0 \le x \le 3$ and $0 \le y \le x$.

5-41. Continuation of Exercise 5-40. Determine the following:
(a) $P(X = 1, Y = 2)$ (b) $P(1 \le X \le 2)$
(c) $P(Y = 1)$ (d) $P(X = 2, Y = 2)$
(e) $E(X)$ (f) $E(Y)$

5-42. Continuation of Exercise 5-40. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y = 2|X = 1)$
(e) Conditional probability distribution of X given $Y = 2$

5-43. Determine the value of c that makes the function $f(x, y) = ce^{-2x-3y}$ a joint probability density function over the range $0 \le x$ and $0 \le y \le x$.

5-44. Continuation of Exercise 5-43. Determine the following:
(a) $P(X = 1, Y = 2)$ (b) $P(1 \le X \le 2)$
(c) $P(Y = 3)$ (d) $P(X = 2, Y = 2)$
(e) $E(X)$ (f) $E(Y)$

5-45. Continuation of Exercise 5-43. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) Conditional probability distribution of X given $Y = 2$

5-46. Determine the value of c that makes the function $f(x, y) = ce^{-2x-3y}$ a joint probability density function over the range $0 \le x$ and $x \le y$.

5-47. Continuation of Exercise 5-46. Determine the following:
(a) $P(X = 1, Y = 2)$ (b) $P(1 \le X \le 2)$
(c) $P(Y = 3)$ (d) $P(X = 2, Y = 2)$
(e) $E(X)$ (f) $E(Y)$

5-48. Continuation of Exercise 5-46. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y = 2|X = 1)$
(e) Conditional probability distribution of X given $Y = 2$

5-49. Two methods of measuring surface smoothness are used to evaluate a paper product. The measurements are recorded as deviations from the nominal surface smoothness in coded units. The joint probability distribution of the two measurements is a uniform distribution over the region $0 \le x \le 4$, $0 \le y$, and $x - 1 \le y \le x + 1$. That is, $f_{XY}(x, y) = c$ for x and y in the region. Determine the value for c such that $f_{XY}(x, y)$ is a joint probability density function.

5-50. Continuation of Exercise 5-49. Determine the following:
(a) $P(X = 0.5, Y = 0.5)$ (b) $P(X = 0.5)$
(c) $E(X)$ (d) $E(Y)$

5-51. Continuation of Exercise 5-49. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y = 0.5|X = 1)$

5-52. The time between surface finish problems in a galvanizing process is exponentially distributed with a mean of 40 hours. A single plant operates three galvanizing lines that are assumed to operate independently.
(a) What is the probability that none of the lines experiences a surface finish problem in 40 hours of operation?
(b) What is the probability that all three lines experience a surface finish problem between 20 and 40 hours of operation?
(c) Why is the joint probability density function not needed to answer the previous questions?

5-53. A popular clothing manufacturer receives Internet orders via two different routing systems. The time between orders for each routing system in a typical day is known to be exponentially distributed with a mean of 3.2 minutes. Both systems operate independently.
(a) What is the probability that no orders will be received in a 5 minute period? In a 10 minute period?
(b) What is the probability that both systems receive two orders between 10 and 15 minutes after the site is officially open for business?

(c) Why is the joint probability distribution not needed to answer the previous questions?

5-54. The conditional probability distribution of Y given $X = x$ is $f_{Y|x}(y) = xe^{-xy}$ for $y > 0$ and the marginal probability distribution of X is a continuous uniform distribution over 0 to 10.

(a) Graph $f_{Y|X}(y) = xe^{-xy}$ for $y > 0$ for several values of x. Determine
(b) $P(Y > 2 \mid X = 2)$ (c) $E(Y \mid X = 2)$
(d) $E(Y \mid X = x)$ (e) $f_{XY}(x, y)$
(f) $f_Y(y)$

5-4 MULTIPLE CONTINUOUS RANDOM VARIABLES

As for discrete random variables, in some cases, more than two continuous random variables are defined in a random experiment.

EXAMPLE 5-22

Many dimensions of a machined part are routinely measured during production. Let the random variables, X_1, X_2, X_3, and X_4 denote the lengths of four dimensions of a part. Then, at least four random variables are of interest in this study.

The joint probability distribution of continuous random variables, $X_1, X_2, X_3, \ldots, X_p$ can be specified by providing a method of calculating the probability that $X_1, X_2, X_3, \ldots, X_p$ assume a value in a region R of p-dimensional space. A joint probability density function $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$ is used to determine the probability that $(X_1, X_2, X_3, \ldots, X_p)$ assume a value in a region R by the multiple integral of $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$ over the region R.

Definition

A **joint probability density function** for the continuous random variables $X_1, X_2, X_3, \ldots, X_p$, denoted as $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$, satisfies the following properties:

(1) $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) \geq 0$

(2) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \cdots dx_p = 1$

(3) For any region B of p-dimensional space

$$P[(X_1, X_2, \ldots, X_p) \in B] = \int \cdots \int_B f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \cdots dx_p \quad (5\text{-}22)$$

Typically, $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$ is defined over all of p-dimensional space by assuming that $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = 0$ for all points for which $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$ is not specified.

EXAMPLE 5-23

In an electronic assembly, let the random variables X_1, X_2, X_3, X_4 denote the lifetimes of four components in hours. Suppose that the joint probability density function of these variables is

$$f_{X_1 X_2 X_3 X_4}(x_1, x_2, x_3, x_4) = 9 \times 10^{-2} e^{-0.001x_1 - 0.002x_2 - 0.0015x_3 - 0.003x_4}$$
$$\text{for } x_1 > 0, x_2 > 0, x_3 > 0, x_4 > 0$$

What is the probability that the device operates for more than 1000 hours without any failures?

The requested probability is $P(X_1 \le 1000, X_2 \le 1000, X_3 \le 1000, X_4 \le 1000)$, which equals the multiple integral of $f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)$ over the region $x_1 \le 1000$, $x_2 \le 1000$, $x_3 \le 1000$, $x_4 \le 1000$. The joint probability density function can be written as a product of exponential functions, and each integral is the simple integral of an exponential function. Therefore,

$$P(X_1 \le 1000, X_2 \le 1000, X_3 \le 1000, X_4 \le 1000) = e^{-1-2-1.5-3} = 0.00055$$

Suppose that the joint probability density function of several continuous random variables is a constant, say c over a region R (and zero elsewhere). In this special case,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_p = c \times (\text{volume of region R}) = 1$$

by property (2) of Equation 5-22. Therefore, $c = 1/(\text{volume } R)$. Furthermore, by property (3) of Equation 5-22.

$$\begin{aligned} &P[(X_1, X_2, \ldots, X_p) \in B] \\ &= \int\cdots\int_{B} f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_p = c \times \text{volume}(B \cap R) \\ &= \frac{\text{volume}(B \cap R)}{\text{volume}(R)} \end{aligned}$$

When the joint probability density function is constant, the probability that the random variables assume a value in the region B is just the ratio of the volume of the region $B \cap R$ to the volume of the region R for which the probability is positive.

EXAMPLE 5-24

Suppose the joint probability density function of the continuous random variables X and Y is constant over the region $x^2 + y^2 \le 4$. Determine the probability that $X^2 + Y^2 \le 1$.

The region that receives positive probability is a circle of radius 2. Therefore, the area of this region is 4π. The area of the region $x^2 + y^2 \le 1$ is π. Consequently, the requested probability is $\pi/(4\pi) = 1/4$.

Definition

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1X_2\cdots X_p}(x_1, x_2 \ldots, x_p)$ the **marginal probability density function** of X_i is

$$f_{X_i}(x_i) = \int\cdots\int_{R_{x_i}} f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_{i-1}\,dx_{i+1}\cdots dx_p \qquad (5\text{-}23)$$

where R_{x_i} denotes the set of all points in the range of $X_1, X_2, \ldots, X_p$ for which $X_i = x_i$.

As for two random variables, a probability involving only one random variable, say, for example $P(a \le X_i \le b)$, can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$. That is,

$$P(a \le X_i \le b) = P(-\infty < X_1 < \infty, \ldots, -\infty < X_{i-1} < \infty, a \le X_i \le b, -\infty < X_{i+1} < \infty, \ldots, -\infty < X_p < \infty)$$

Furthermore, $E(X_i)$ and $V(X_i)$, for $i = 1, 2, \ldots, p$, can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

Mean and Variance from Joint Distribution

$$E(X_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{X_1 X_2 \cdots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \cdots dx_p$$

and (5-24)

$$V(X_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_1 X_2 \cdots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \cdots dx_p$$

The probability distribution of a subset of variables such as $X_1, X_2, \ldots, X_k, k < p$, can be obtained from the joint probability distribution of $X_1, X_2, X_3, \ldots, X_p$ as follows.

Distribution of a Subset of Random Variables

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1 X_2 \cdots X_p}(x_1, x_2, \ldots, x_p)$, the **probability density function** of $X_1, X_2, \ldots, X_k, k < p$, is

$$f_{X_1 X_2 \cdots X_k}(x_1, x_2, \ldots, x_k) = \int_{R_{x_1 x_2 \cdots x_k}} \int \cdots \int f_{X_1 X_2 \cdots X_p}(x_1, x_2, \ldots, x_p)\, dx_{k+1}\, dx_{k+2} \cdots dx_p \quad (5\text{-}25)$$

where $R_{x_1 x_2 \cdots x_k}$ denotes the set of all points in the range of $X_1, X_2, \ldots, X_k$ for which $X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k$.

Conditional Probability Distribution

Conditional probability distributions can be developed for multiple continuous random variables by an extension of the ideas used for two continuous random variables.

$$f_{X_1 X_2 X_3 | X_4 X_5}(x_1, x_2, x_3) = \frac{f_{X_1 X_2 X_3 X_4 X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4 X_5}(x_4, x_5)}$$

for $f_{X_4 X_5}(x_4, x_5) > 0$.

The concept of independence can be extended to multiple continuous random variables.

Definition

Continuous random variables $X_1, X_2, \ldots, X_p$ are **independent** if and only if

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_p}(x_p) \quad \text{for all } x_1, x_2, \ldots, x_p \qquad (5\text{-}26)$$

Similar to the result for only two random variables, independence implies that Equation 5-26 holds for *all* $x_1, x_2, \ldots, x_p$. If we find one point for which the equality fails, $X_1, X_2, \ldots, X_p$ are not independent. It is left as an exercise to show that if $X_1, X_2, \ldots, X_p$ are independent,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = P(X_1 \in A_1) P(X_2 \in A_2) \ldots P(X_p \in A_p)$$

for *any* regions $A_1, A_2, \ldots, A_p$ in the range of $X_1, X_2, \ldots, X_p$, respectively.

EXAMPLE 5-25

In Chapter 3, we showed that a negative binomial random variable with parameters p and r can be represented as a sum of r geometric random variables $X_1, X_2, \ldots, X_r$. Each geometric random variable represents the additional trials required to obtain the next success. Because the trials in a binomial experiment are independent, $X_1, X_2, \ldots, X_r$ are independent random variables.

EXAMPLE 5-26

Suppose X_1, X_2, and X_3 represent the thickness in micrometers of a substrate, an active layer, and a coating layer of a chemical product. Assume that X_1, X_2, and X_3 are independent and normally distributed with $\mu_1 = 10000$, $\mu_2 = 1000$, $\mu_3 = 80$, $\sigma_1 = 250$, $\sigma_2 = 20$, and $\sigma_3 = 4$, respectively. The specifications for the thickness of the substrate, active layer, and coating layer are $9200 \le x_1 \le 10800$, $950 \le x_2 \le 1050$, and $75 \le x_3 \le 85$, respectively. What proportion of chemical products meets all thickness specifications? Which one of the three thicknesses has the least probability of meeting specifications?

The requested probability is $P(9200 \le X_1 \le 10800, 950 \le X_2 \le 1050, 75 \le X_3 \le 85)$. Because the random variables are independent,

$$P(9200 \le X_1 \le 10800, 950 \le X_2 \le 1050, 75 \le X_3 \le 85)$$
$$= P(9200 \le X_1 \le 10800) P(950 \le X_2 \le 1050) P(75 \le X_3 \le 85)$$

After standardizing, the above equals

$$P(-3.2 \le Z \le 3.2) P(-2.5 \le Z \le 2.5) P(-1.25 \le Z \le 1.25)$$

where Z is a standard normal random variable. From the table of the standard normal distribution, the above equals

$$(0.9986)(0.9875)(0.7887) = 0.7778$$

The thickness of the coating layer has the least probability of meeting specifications. Consequently, a priority should be to reduce variability in this part of the process.

EXERCISES FOR SECTION 5–4

5-55. Suppose the random variables X, Y, and Z have the joint probability density function $f(x, y, z) = 8xyz$ for $0 \le x \le 1$, $0 \le y \le 1$, and $0 \le z \le 1$. Determine the following:
(a) $P(X \le 0.5)$ (b) $P(X \le 0.5, Y \le 0.5)$
(c) $P(Z \le 2)$ (d) $P(X \le 0.5 \text{ or } Z \le 2)$
(e) $E(X)$

5-56. Continuation of Exercise 5-55. Determine the following:
(a) $P(X \le 0.5 \mid Y \le 0.5)$
(b) $P(X \le 0.5, Y \le 0.5 \mid Z \le 0.8)$

5-57. Continuation of Exercise 5-55. Determine the following:
(a) Conditional probability distribution of X given that $Y \le 0.5$ and $Z \le 0.8$
(b) $P(X \le 0.5 \mid Y \le 0.5, Z \le 0.8)$

5-58. Suppose the random variables X, Y, and Z have the joint probability density function $f_{XYZ}(x, y, z) = c$ over the cylinder $x^2 + y^2 \le 4$ and $0 \le z \le 4$. Determine the following.
(a) The constant c so that $f_{XYZ}(x, y, z)$ is a probability density function
(b) $P(X^2 + Y^2 \le 2)$
(c) $P(Z \le 2)$
(d) $E(X)$

5-59. Continuation of Exercise 5-58. Determine the following:
(a) $P(X \le 1 \mid Y \le 1)$ (b) $P(X^2 + Y^2 \le 1 \mid Z \le 1)$

5-60. Continuation of Exercise 5-58. Determine the conditional probability distribution of Z given that $X \le 1$ and $Y \le 1$.

5-61. Determine the value of c that makes $f_{XYZ}(x, y, z) = c$ a joint probability density function over the region $x \ge 0$, $y \ge 0$, $z \ge 0$, and $x + y + z \le 1$.

5-62. Continuation of Exercise 5-61. Determine the following:
(a) $P(X \le 0.5, Y \le 0.5, Z \le 0.5)$
(b) $P(X \le 0.5, Y \le 0.5)$
(c) $P(X \le 0.5)$
(d) $E(X)$

5-63. Continuation of Exercise 5-61. Determine the following:
(a) Marginal distribution of X
(b) Joint distribution of X and Y
(c) Conditional probability distribution of X given that $Y \le 0.5$ and $Z \le 0.5$
(d) Conditional probability distribution of X given that $Y \le 0.5$

5-64. The yield in pounds from a day's production is normally distributed with a mean of 1500 pounds and standard deviation of 100 pounds. Assume that the yields on different days are independent random variables.
(a) What is the probability that the production yield exceeds 1400 pounds on each of five days next week?
(b) What is the probability that the production yield exceeds 1400 pounds on at least four of the five days next week?

5-65. The weights of adobe bricks used for construction are normally distributed with a mean of 3 pounds and a standard deviation of 0.25 pound. Assume that the weights of the bricks are independent and that a random sample of 20 bricks is selected.
(a) What is the probability that all the bricks in the sample exceed 2.75 pounds?
(b) What is the probability that the heaviest brick in the sample exceeds 3.75 pounds?

5-66. A manufacturer of electroluminescent lamps knows that the amount of luminescent ink deposited on one of its products is normally distributed with a mean of 1.2 grams and a standard deviation of 0.03 grams. Any lamp with less than 1.14 grams of luminescent ink will fail to meet customer's specifications. A random sample of 25 lamps is collected and the mass of luminescent ink on each is measured.
(a) What is the probability that at least 1 lamp fails to meet specifications?
(b) What is the probability that 5 lamps or fewer fail to meet specifications?
(c) What is the probability that all lamps conform to specifications?
(d) Why is the joint probability distribution of the 25 lamps not needed to answer the previous questions?

5-5 COVARIANCE AND CORRELATION

When two or more random variables are defined on a probability space, it is useful to describe how they vary together; that is, it is useful to measure the relationship between the variables. A common measure of the relationship between two random variables is the **covariance.** To define the covariance, we need to describe the expected value of a function of two random variables $h(X, Y)$. The definition simply extends that used for a function of a single random variable.

Definition

$$E[h(X, Y)] = \begin{cases} \sum_R h(x, y) f_{XY}(x, y) & X, Y \text{ discrete} \\ \iint_R h(x, y) f_{XY}(x, y)\, dx\, dy & X, Y \text{ continuous} \end{cases} \quad (5\text{-}27)$$

That is, $E[h(X, Y)]$ can be thought of as the weighted average of $h(x, y)$ for each point in the range of (X, Y). The value of $E[h(X, Y)]$ represents the average value of $h(X, Y)$ that is expected in a long sequence of repeated trials of the random experiment.

EXAMPLE 5-27

For the joint probability distribution of the two random variables in Fig. 5-12, calculate $E[(X - \mu_X)(Y - \mu_Y)]$.

The result is obtained by multiplying $x - \mu_X$ times $y - \mu_Y$, times $f_{XY}(x, y)$ for each point in the range of (X, Y). First, μ_X and μ_Y are determined from Equation 5-3 as

$$\mu_X = 1 \times 0.3 + 3 \times 0.7 = 2.4$$

and

$$\mu_Y = 1 \times 0.3 + 2 \times 0.4 + 3 \times 0.3 = 2.0$$

Therefore,

$$\begin{aligned} E[(X - \mu_X)(Y - \mu_Y)] &= (1 - 2.4)(1 - 2.0)0.1 \\ &\quad + (1 - 2.4)(2 - 2.0)0.2 + (3 - 2.4)(1 - 2.0)0.2 \\ &\quad + (3 - 2.4)(2 - 2.0)0.2 + (3 - 2.4)(3 - 2.0)0.3 = 0.2 \end{aligned}$$

The covariance is defined for both continuous and discrete random variables by the same formula.

Definition

The **covariance** between the random variables X and Y, denoted as $\text{cov}(X, Y)$ or σ_{XY}, is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y \quad (5\text{-}28)$$

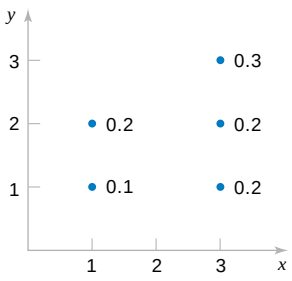


Figure 5-12 Joint distribution of X and Y for Example 5-27.

If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, σ_{XY} is positive (or negative). If the points tend to fall along a line of positive slope, X tends to be greater than μ_X when Y is greater than μ_Y. Therefore, the product of the two terms $x - \mu_X$ and $y - \mu_Y$ tends to be positive. However, if the points tend to fall along a line of negative slope, $x - \mu_X$ tends to be positive when $y - \mu_Y$ is negative, and vice versa. Therefore, the product of $x - \mu_X$ and $y - \mu_Y$ tends to be negative. In this sense, the covariance between X and Y describes the variation between the two random variables. Figure 5-13 shows examples of pairs of random variables with positive, negative, and zero covariance.

Covariance is a measure of **linear relationship** between the random variables. If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship. This is illustrated in Fig. 5-13(d). The only points with nonzero probability are the points on the circle. There is an identifiable relationship between the variables. Still, the covariance is zero.

The equality of the two expressions for covariance in Equation 5-28 is shown for continuous random variables as follows. By writing the expectations as integrals,

$$E[(Y - \mu_Y)(X - \mu_X)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy - \mu_X y - x\mu_Y + \mu_X\mu_Y) f_{XY}(x, y)\, dx\, dy$$

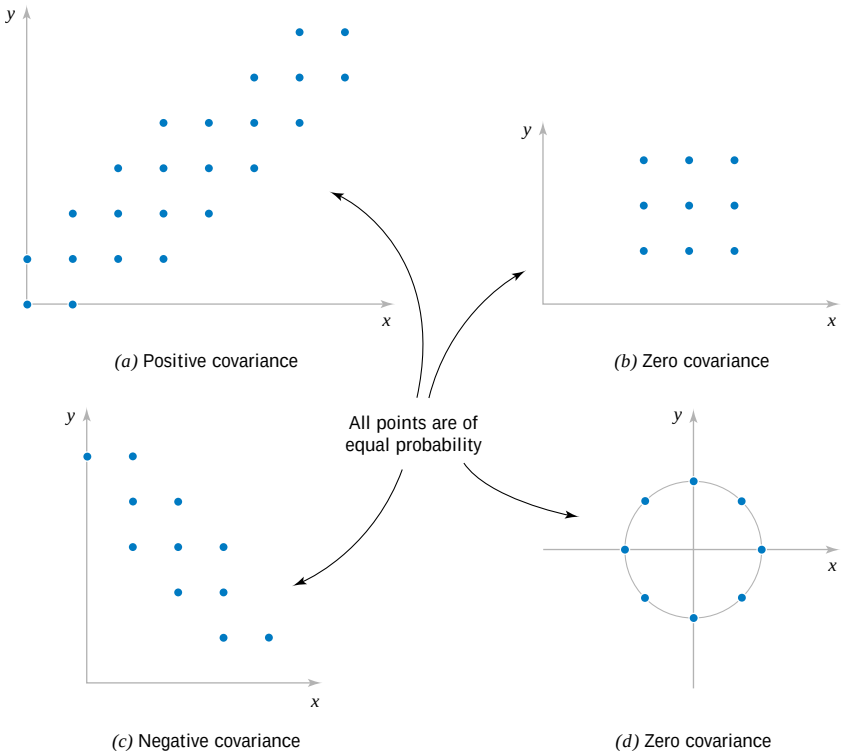


Figure 5-13 Joint probability distributions and the sign of covariance between X and Y.

Now

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)y\,f_{XY}(x,y)\,dx\,dy = \mu_X\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{XY}(x,y)\,dx\,dy\right] = \mu_X\mu_Y$$

Therefore,

$$E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{XY}(x,y)\,dx\,dy - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{XY}(x,y)\,dx\,dy - \mu_X\mu_Y = E(XY) - \mu_X\mu_Y$$

EXAMPLE 5-28

In Example 5-1, the random variables X and Y are the number of acceptable and suspect bits among four bits received during a digital communication, respectively. Is the covariance between X and Y positive or negative?

Because X and Y are the number of acceptable and suspect bits out of the four received, $X + Y = 4$. If X is near 4, Y must be near 0. Therefore, X and Y have a negative covariance. This can be verified from the joint probability distribution in Fig. 5-1.

There is another measure of the relationship between two random variables that is often easier to interpret than the covariance.

Definition

The **correlation** between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{5-29}$$

Because $\sigma_X > 0$ and $\sigma_Y > 0$, if the covariance between X and Y is positive, negative, or zero, the correlation between X and Y is positive, negative, or zero, respectively. The following result can be shown.

For any two random variables X and Y

$$-1 \le \rho_{XY} \le 1 \tag{5-30}$$

The correlation just scales the covariance by the standard deviation of each variable. Consequently, the correlation is a dimensionless quantity that can be used to compare the linear relationships between pairs of variables in different units.

If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, ρ_{XY} is near $+1$ (or -1). If ρ_{XY} equals $+1$ or -1, it can be shown that the points in the joint probability distribution that receive positive probability fall exactly along a straight line. Two random variables with nonzero correlation are said to be **correlated.** Similar to covariance, the correlation is a measure of the **linear relationship** between random variables.

EXAMPLE 5-29

For the discrete random variables X and Y with the joint distribution shown in Fig. 5-14, determine σ_{XY} and ρ_{XY}.

The calculations for $E(XY)$, $E(X)$, and $V(X)$ are as follows.

$$E(XY) = 0 \times 0 \times 0.2 + 1 \times 1 \times 0.1 + 1 \times 2 \times 0.1 + 2 \times 1 \times 0.1 + 2 \times 2 \times 0.1 + 3 \times 3 \times 0.4 = 4.5$$

$$E(X) = 0 \times 0.2 + 1 \times 0.2 + 2 \times 0.2 + 3 \times 0.4 = 1.8$$

$$V(X) = (0 - 1.8)^2 \times 0.2 + (1 - 1.8)^2 \times 0.2 + (2 - 1.8)^2 \times 0.2 + (3 - 1.8)^2 \times 0.4 = 1.36$$

Because the marginal probability distribution of Y is the same as for X, $E(Y) = 1.8$ and $V(Y) = 1.36$. Consequently,

$$\sigma_{XY} = E(XY) - E(X)E(Y) = 4.5 - (1.8)(1.8) = 1.26$$

Furthermore,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{1.26}{\sqrt{1.36}\sqrt{1.36}} = 0.926$$

EXAMPLE 5-30

Suppose that the random variable X has the following distribution: $P(X = 1) = 0.2$, $P(X = 2) = 0.6$, $P(X = 3) = 0.2$. Let $Y = 2X + 5$. That is, $P(Y = 7) = 0.2$, $P(Y = 9) = 0.6$, $P(Y = 11) = 0.2$. Determine the correlation between X and Y. Refer to Fig. 5-15.

Because X and Y are linearly related, $\rho = 1$. This can be verified by direct calculations: Try it.

For independent random variables, we do not expect any relationship in their joint probability distribution. The following result is left as an exercise.

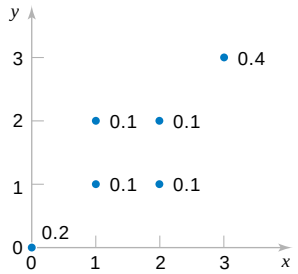


Figure 5-14 Joint distribution for Example 5-29.

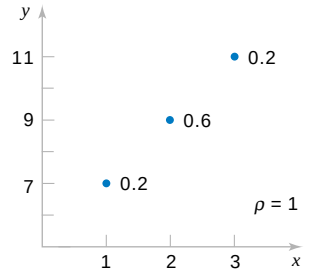


Figure 5-15 Joint distribution for Example 5-30.

If X and Y are independent random variables,

$$\sigma_{XY} = \rho_{XY} = 0 \tag{5-31}$$

EXAMPLE 5-31 For the two random variables in Fig. 5-16, show that $\sigma_{XY} = 0$.

The two random variables in this example are continuous random variables. In this case $E(XY)$ is defined as the double integral over the range of (X, Y). That is,

$$E(XY) = \int_0^4 \int_0^2 xy f_{XY}(x, y)\, dx\, dy = \frac{1}{16} \int_0^4 \left[\int_0^2 x^2 y^2\, dx \right] dy = \frac{1}{16} \int_0^4 y^2 \left[x^3/3 \right]_0^2 \Big]$$

$$= \frac{1}{16} \int_0^4 y^2 (8/3)\, dy = \frac{1}{6} \left[y^3/3 \right]_0^4 \Big] = \frac{1}{6}(64/3) = 32/9$$

Also,

$$E(X) = \int_0^4 \int_0^2 x f_{XY}(x, y)\, dx\, dy = \frac{1}{16} \int_0^4 \left[\int_0^2 x^2\, dx \right] dy = \frac{1}{16} \int_0^4 \left[x^3/3 \right]_0^2 \Big] dy$$

$$= \frac{1}{16} \left[y^2/2 \right]_0^4 \Big] (8/3) = \frac{1}{6}(16/2) = 4/3$$

$$E(Y) = \int_0^4 \int_0^2 y f_{XY}(x, y)\, dx\, dy = \frac{1}{16} \int_0^4 y^2 \left[\int_0^2 x\, dx \right] dy = \frac{1}{16} \int_0^4 y^2 \left[x^2/2 \right]_0^2 \Big] dy$$

$$= \frac{2}{16} \left[y^3/3 \right]_0^4 \Big] = \frac{1}{8}(64/3) = 8/3$$

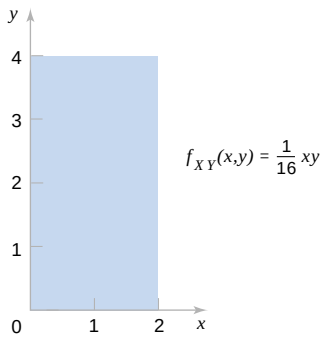


Figure 5-16 Random variables with zero covariance from Example 5-31.

Thus,

$$E(XY) = E(X)E(Y) = 3(3) = 14(3)18(3) = 0$$

It can be shown that these two random variables are independent. You can check that $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y.

However, if the correlation between two random variables is zero, we *cannot* immediately conclude that the random variables are independent. Figure 5-13(d) provides an example.

EXERCISES FOR SECTION 5-5

5-67. Determine the covariance and correlation for the following joint probability distribution:

x	1	1	2	4
y	3	4	5	6
$f_{XY}(x, y)$	1/8	1/4	1/2	1/8

5-68. Determine the covariance and correlation for the following joint probability distribution:

x	−1	−0.5	0.5	1
y	−2	−1	1	2
$f_{XY}(x, y)$	1/8	1/4	1/2	1/8

5-69. Determine the value for c and the covariance and correlation for the joint probability mass function $f_{XY}(x, y) = c(x + y)$ for $x = 1, 2, 3$ and $y = 1, 2, 3$.

5-70. Determine the covariance and correlation for the joint probability distribution shown in Fig. 5-4(a) and described in Example 5-8.

5-71. Determine the covariance and correlation for X_1 and X_2 in the joint distribution of the multinomial random variables X_1, X_2 and X_3 in with $p_1 = p_2 = p_3 = 1/3$ and $n = 3$. What can you conclude about the sign of the correlation between two random variables in a multinomial distribution?

5-72. Determine the value for c and the covariance and correlation for the joint probability density function $f_{XY}(x, y) = cxy$ over the range $0 < x < 3$ and $0 < y < x$.

5-73. Determine the value for c and the covariance and correlation for the joint probability density function $f_{XY}(x, y) = c$ over the range $0 < x < 5$, $0 < y$, and $x - 1 < y < x + 1$.

5-74. Determine the covariance and correlation for the joint probability density function $f_{XY}(x, y) = 6 \times 10^{-6} e^{-0.001x - 0.002y}$ over the range $0 < x$ and $x < y$ from Example 5-15.

5-75. Determine the covariance and correlation for the joint probability density function $f_{XY}(x, y) = e^{-x-y}$ over the range $0 < x$ and $0 < y$.

5-76. Suppose that the correlation between X and Y is ρ. For constants a, b, c, and d, what is the correlation between the random variables $U = aX + b$ and $V = cY + d$?

5-77. The joint probability distribution is

x	−1	0	0	1
y	0	−1	1	0
$f_{XY}(x, y)$	1/4	1/4	1/4	1/4

Show that the correlation between X and Y is zero, but X and Y are not independent.

5-78. Suppose X and Y are independent continuous random variables. Show that $\rho_{XY} = 0$.

5-6 BIVARIATE NORMAL DISTRIBUTION

An extension of a normal distribution to two random variables is an important bivariate probability distribution.

EXAMPLE 5-32

At the start of this chapter, the length of different dimensions of an injection-molded part was presented as an example of two random variables. Each length might be modeled by a normal distribution. However, because the measurements are from the same part, the random variables are typically not independent. A probability distribution for two normal random variables that are not independent is important in many applications. As stated at the start of the

chapter, if the specifications for X and Y are 2.95 to 3.05 and 7.60 to 7.80 millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 \le X \le 3.05, 7.60 \le Y \le 7.80)$.

Definition

The probability density function of a **bivariate normal distribution** is

$$f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \quad (5\text{-}32)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, with parameters $\sigma_X > 0$, $\sigma_Y > 0$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, and $-1 < \rho < 1$.

The result that $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$ integrates to 1 is left as an exercise. Also, the bivariate normal probability density function is positive over the entire plane of real numbers.

Two examples of bivariate normal distributions are illustrated in Fig. 5-17 along with corresponding contour plots. Each curve on the contour plots is a set of points for which the probability density function is constant. As seen in the contour plots, the bivariate normal probability density function is constant on ellipses in the (x, y) plane. (We can consider a circle to be a special case of an ellipse.) The center of each ellipse is at the point (μ_X, μ_Y). If $\rho > 0$ ($\rho < 0$), the major axis of each ellipse has positive (negative) slope, respectively. If $\rho = 0$, the major axis of the ellipse is aligned with either the x or y coordinate axis.

EXAMPLE 5-33 The joint probability density function $f_{XY}(x, y) = \frac{1}{2\pi} e^{-0.5(x^2+y^2)}$ is a special case of a bivariate normal distribution with $\mu_X = 1$, $\mu_Y = 1$, $\sigma_X = 0$, $\sigma_Y = 0$, and $\rho = 0$. This probability density function is illustrated in Fig. 5-18. Notice that the contour plot consists of concentric circles about the origin.

By completing the square in the exponent, the following results can be shown. The details are left as an exercise.

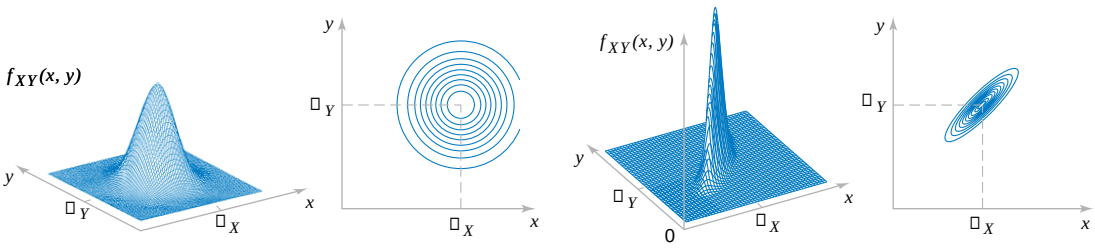


Figure 5-17 Examples of bivariate normal distributions.

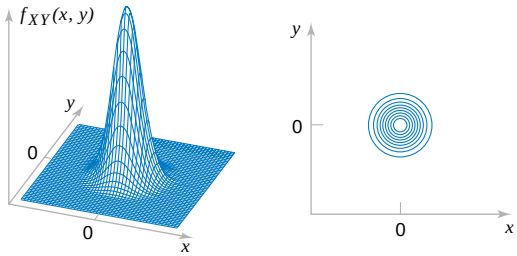


Figure 5-18 Bivariate normal probability density function with $\sigma_X = 1$, $\sigma_Y = 1$, $\rho = 0$, $\mu_X = 0$, and $\mu_Y = 0$.

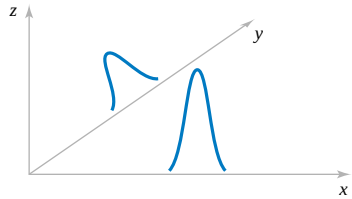


Figure 5-19 Marginal probability density functions of a bivariate normal distribution.

Marginal Distributions of Bivariate Normal Random Variables

If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, the **marginal probability distributions** of X and Y are normal with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. (5-33)

Figure 5-19 illustrates that the marginal probability distributions of X and Y are normal. Furthermore, as the notation suggests, ρ represents the correlation between X and Y. The following result is left as an exercise.

If X and Y have a bivariate normal distribution with joint probability density function $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, the correlation between X and Y is ρ. (5-34)

The contour plots in Fig. 5-17 illustrate that as ρ moves from zero (left graph) to 0.9 (right graph), the ellipses narrow around the major axis. The probability is more concentrated about a line in the (x, y) plane and graphically displays greater correlation between the variables. If $\rho = -1$ or -1, all the probability is concentrated on a line in the (x, y) plane. That is, the probability that X and Y assume a value that is not on the line is zero. In this case, the bivariate normal probability density is not defined.

In general, zero correlation does not imply independence. But in the special case that X and Y have a bivariate normal distribution, if $\rho = 0$, X and Y are independent. The details are left as an exercise.

If X and Y have a bivariate normal distribution with $\rho = 0$, X and Y are independent. (5-35)

An important use of the bivariate normal distribution is to calculate probabilities involving two correlated normal random variables.

EXAMPLE 5-34 Suppose that the X and Y dimensions of an injection-molded part have a bivariate normal distribution with $\mu_X = 3.00$, $\mu_Y = 7.70$, $\sigma_X = 0.04$, $\sigma_Y = 0.08$, and $\rho = 0.8$. Then, the probability that a part satisfies both specifications is

$$P(2.95 \le X \le 3.05, 7.60 \le Y \le 7.80)$$

This probability can be obtained by integrating $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$ over the region $2.95 \le x \le 3.05$ and $7.60 \le y \le 7.80$, as shown in Fig. 5-7. Unfortunately, there is often no closed-form solution to probabilities involving bivariate normal distributions. In this case, the integration must be done numerically.

EXERCISES FOR SECTION 5-6

5-79. Let X and Y represent concentration and viscosity of a chemical product. Suppose X and Y have a bivariate normal distribution with $\mu_X = 4$, $\mu_Y = 1$, $\sigma_X^2 = 2$, and $\sigma_Y^2 = 1$. Draw a rough contour plot of the joint probability density function for each of the following values for ρ:
(a) $\rho = 0$ (b) $\rho = 0.8$
(c) $\rho = -0.8$

5-80. Let X and Y represent two dimensions of an injection molded part. Suppose X and Y have a bivariate normal distribution with $\mu_X = 3.00$, $\sigma_X = 0.04$, $\mu_Y = 7.70$, and $\sigma_Y = 0.08$. Determine $P(2.95 \le X \le 3.05, 7.60 \le Y \le 7.80)$.

5-81. In the manufacture of electroluminescent lamps, several different layers of ink are deposited onto a plastic substrate. The thickness of these layers is critical if specifications regarding the final color and intensity of light of the lamp are to be met. Let X and Y denote the thickness of two different layers of ink. It is known that X is normally distributed with a mean of 0.1 millimeter and a standard deviation of 0.00031 millimeter, and Y is also normally distributed with a mean of 0.23 millimeter and a standard deviation of 0.00017 millimeter. The value of ρ for these variables is equal to zero. Specifications call for a lamp to have a thickness of the ink corresponding to X in the range of 0.099535 to 0.100465 millimeters and Y in the range of 0.22966 to 0.23034 millimeters. What is the probability that a randomly selected lamp will conform to specifications?

5-82. Suppose that X and Y have a bivariate normal distribution with joint probability density function $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$.
(a) Show that the conditional distribution of Y, given that $X = x$ is normal.
(b) Determine $E(Y \mid X = x)$.
(c) Determine $V(Y \mid X = x)$.

5-83. If X and Y have a bivariate normal distribution with $\rho = 0$, show that X and Y are independent.

5-84. Show that the probability density function $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$ of a bivariate normal distribution integrates to one. [*Hint:* Complete the square in the exponent and use the fact that the integral of a normal probability density function for a single variable is 1.]

5-85. If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, show that the marginal probability distribution of X is normal with mean μ_X and standard deviation σ_X. [*Hint:* Complete the square in the exponent and use the fact that the integral of a normal probability density function for a single variable is 1.]

5-86. If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, show that the correlation between X and Y is ρ. [*Hint:* Complete the square in the exponent].

5-7 LINEAR COMBINATIONS OF RANDOM VARIABLES

A random variable is sometimes defined as a function of one or more random variables. The CD material presents methods to determine the distributions of general functions of random variables. Furthermore, moment-generating functions are introduced on the CD

and used to determine the distribution of a sum of random variables. In this section, results for linear functions are highlighted because of their importance in the remainder of the book. References are made to the CD material as needed. For example, if the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part, $Y = 2X_1 + 2X_2$ is a random variable that represents the perimeter of the part. As another example, recall that the negative binomial random variable was represented as the sum of several geometric random variables.

In this section, we develop results for random variables that are linear combinations of random variables.

Definition

Given random variables $X_1, X_2, \ldots, X_p$ and constants $c_1, c_2, \ldots, c_p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p \tag{5-36}$$

is a **linear combination** of $X_1, X_2, \ldots, X_p$.

Now, $E(Y)$ can be found from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows. Assume $X_1, X_2, \ldots, X_p$ are continuous random variables. An analogous calculation can be used for discrete random variables.

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(c_1x_1 + c_2x_2 + \cdots + c_px_p)f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_p \\
&= c_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1 f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_p \\
&\quad + c_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_2 f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_p + \cdots + \\
&\quad + c_p\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_p f_{X_1X_2\cdots X_p}(x_1, x_2, \ldots, x_p)\,dx_1\,dx_2\cdots dx_p
\end{aligned}$$

By using Equation 5-24 for each of the terms in this expression, we obtain the following.

Mean of a Linear Combination

If $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p) \tag{5-37}$$

Furthermore, it is left as an exercise to show the following.

Variance of a Linear Combination

If $X_1, X_2, \ldots, X_p$ are random variables, and $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$, then in general

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_j \operatorname{cov}(X_i, X_j) \quad (5\text{-}38)$$

If $X_1, X_2, \ldots, X_p$ are **independent,**

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) \quad (5\text{-}39)$$

Note that the result for the variance in Equation 5-39 requires the random variables to be independent. To see why the independence is important, consider the following simple example. Let X_1 denote any random variable and define $X_2 = -X_1$. Clearly, X_1 and X_2 are not independent. In fact, $\rho_{XY} = -1$. Now, $Y = X_1 + X_2$ is 0 with probability 1. Therefore, $V(Y) = 0$, regardless of the variances of X_1 and X_2.

EXAMPLE 5-35

In Chapter 3, we found that if Y is a negative binomial random variable with parameters p and r, $Y = X_1 + X_2 + \cdots + X_r$, where each X_i is a geometric random variable with parameter p and they are independent. Therefore, $E(X_i) = 1/p$ and $E(X_i^2) = (1 + p)/p^2$. From Equation 5-37, $E(Y) = r/p$ and from Equation 5-39, $V(Y) = r(1 + p)/p^2$.

An approach similar to the one applied in the above example can be used to verify the formulas for the mean and variance of an Erlang random variable in Chapter 4.

EXAMPLE 5-36

Suppose the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume $E(X_1) = 2$ centimeters with standard deviation 0.1 centimeter and $E(X_2) = 5$ centimeters with standard deviation 0.2 centimeter. Also, assume that the covariance between X_1 and X_2 is -0.005. Then, $Y = 2X_1 + 2X_2$ is a random variable that represents the perimeter of the part. From Equation 5-36,

$$E(Y) = 2(2) + 2(5) = 14 \text{ centimeters}$$

and from Equation 5-38

$$V(Y) = 2^2(0.1^2) + 2^2(0.2^2) + 2 \times 2 \times 2(-0.005)$$
$$= 0.04 + 0.16 - 0.04 = 0.16 \text{ centimeters squared}$$

Therefore, the standard deviation of Y is $0.16^{1/2} = 0.4$ centimeters.

The particular linear combination that represents the average of p random variables, with identical means and variances, is used quite often in the subsequent chapters. We highlight the results for this special case.

Mean and Variance of an Average

If $\bar{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu$ for $i = 1, 2, \ldots, p$

$$E(\bar{X}) = \mu \tag{5-40a}$$

if $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2$ for $i = 1, 2, \ldots, p$,

$$V(\bar{X}) = \frac{\sigma^2}{p} \tag{5-40b}$$

The conclusion for $V(\bar{X})$ is obtained as follows. Using Equation 5-39, with $c_i = 1/p$ and $V(X_i) = \sigma^2$, yields

$$V(\bar{X}) = (1/p)^2\sigma^2 + \cdots + (1/p)^2\sigma^2 = p(1/p)^2\sigma^2 = \sigma^2/p$$

(p terms)

Another useful result concerning linear combinations of random variables is a **reproductive property** that holds for independent, normal random variables.

Reproductive Property of the Normal Distribution

If $X_1, X_2, \ldots, X_p$ are independent, normal random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, for $i = 1, 2, \ldots, p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

is a normal random variable with

$$E(Y) = c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

and

$$V(Y) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_p^2\sigma_p^2 \tag{5-41}$$

The mean and variance of Y follow from Equations 5-37 and 5-39. The fact that Y has a normal distribution can be obtained from moment-generating functions discussed in Section 5-9 in the CD material.

EXAMPLE 5-37

Let the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume that X_1 is normal with $E(X_1) = 2$ centimeters and standard deviation 0.1 centimeter and that X_2 is normal with $E(X_2) = 5$ centimeters and standard deviation 0.2 centimeter. Also, assume that X_1 and X_2 are independent. Determine the probability that the perimeter exceeds 14.5 centimeters.

Then, $Y = 2X_1 + 2X_2$ is a normal random variable that represents the perimeter of the part. We obtain, $E(Y) = 14$ centimeters and the variance of Y is

$$V(Y) = 4 \times 0.1^2 + 4 \times 0.2^2 = 0.0416$$

Now,

$$P(Y < 14.5) = P\left(\frac{Y - \mu_Y}{\sigma_Y} < \frac{14.5 - 14}{\sqrt{0.0416}}\right) = P(Z < 1.12) = 0.13$$

EXAMPLE 5-38

Soft-drink cans are filled by an automated filling machine. The mean fill volume is 12.1 fluid ounces, and the standard deviation is 0.1 fluid ounce. Assume that the fill volumes of the cans are independent, normal random variables. What is the probability that the average volume of 10 cans selected from this process is less than 12 fluid ounces?

Let $X_1, X_2, \ldots, X_{10}$ denote the fill volumes of the 10 cans. The average fill volume (denoted as $\bar{X}$) is a normal random variable with

$$E(\bar{X}) = 12.1 \quad \text{and} \quad V(\bar{X}) = \frac{0.1^2}{10} = 0.001$$

Consequently,

$$P(\bar{X} < 12) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{12 - 12.1}{\sqrt{0.001}}\right)$$

$$= P(Z < -3.16) = 0.00079$$

EXERCISES FOR SECTION 5-7

5-87. If X and Y are independent, normal random variables with $E(X) = 0$, $V(X) = 4$, $E(Y) = 10$, and $V(Y) = 9$. Determine the following:

(a) $E(2X - 3Y)$ (b) $V(2X - 3Y)$

(c) $P(2X - 3Y > 30)$ (d) $P(2X - 3Y > 40)$

5-88. Suppose that the random variable X represents the length of a punched part in centimeters. Let Y be the length of the part in millimeters. If $E(X) = 5$ and $V(X) = 0.25$, what are the mean and variance of Y?

5-89. A plastic casing for a magnetic disk is composed of two halves. The thickness of each half is normally distributed with a mean of 2 millimeters and a standard deviation of 0.1 millimeter and the halves are independent.

(a) Determine the mean and standard deviation of the total thickness of the two halves.

(b) What is the probability that the total thickness exceeds 4.3 millimeters?

5-90. In the manufacture of electroluminescent lamps, several different layers of ink are deposited onto a plastic substrate. The thickness of these layers is critical if specifications regarding the final color and intensity of light of the lamp are to be met. Let X and Y denote the thickness of two different layers of ink. It is known that X is normally distributed with a mean of 0.1 millimeter and a standard deviation of 0.00031 millimeter and Y is also normally distributed with a mean of 0.23 millimeter and a standard deviation of 0.00017 millimeter. Assume that these variables are independent.

(a) If a particular lamp is made up of these two inks only, what is the probability that the total ink thickness is less than 0.2337 millimeter?

(b) A lamp with a total ink thickness exceeding 0.2405 millimeters lacks the uniformity of color demanded by the customer. Find the probability that a randomly selected lamp fails to meet customer specifications.

5-91. The width of a casing for a door is normally distributed with a mean of 24 inches and a standard deviation of 1/8 inch. The width of a door is normally distributed with a mean of 23 and 7/8 inches and a standard deviation of 1/16 inch. Assume independence.

(a) Determine the mean and standard deviation of the difference between the width of the casing and the width of the door.

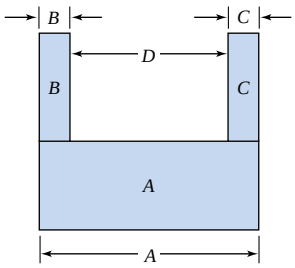


Figure 5-20 Figure for the U-shaped component.

(b) What is the probability that the width of the casing minus the width of the door exceeds 1/4 inch?
(c) What is the probability that the door does not fit in the casing?

5-92. A U-shaped component is to be formed from the three parts *A*, *B*, and *C*. The picture is shown in Fig. 5-20. The length of *A* is normally distributed with a mean of 10 millimeters and a standard deviation of 0.1 millimeter. The thickness of parts *B* and *C* is normally distributed with a mean of 2 millimeters and a standard deviation of 0.05 millimeter. Assume all dimensions are independent.
(a) Determine the mean and standard deviation of the length of the gap *D*.
(b) What is the probability that the gap *D* is less than 5.9 millimeters?

5-93. Soft-drink cans are filled by an automated filling machine and the standard deviation is 0.5 fluid ounce. Assume that the fill volumes of the cans are independent, normal random variables.
(a) What is the standard deviation of the average fill volume of 100 cans?
(b) If the mean fill volume is 12.1 ounces, what is the probability that the average fill volume of the 100 cans is below 12 fluid ounces?
(c) What should the mean fill volume equal so that the probability that the average of 100 cans is below 12 fluid ounces is 0.005?
(d) If the mean fill volume is 12.1 fluid ounces, what should the standard deviation of fill volume equal so that the probability that the average of 100 cans is below 12 fluid ounces is 0.005?
(e) Determine the number of cans that need to be measured such that the probability that the average fill volume is less than 12 fluid ounces is 0.01.

5-94. The photoresist thickness in semiconductor manufacturing has a mean of 10 micrometers and a standard deviation of 1 micrometer. Assume that the thickness is normally distributed and that the thicknesses of different wafers are independent.
(a) Determine the probability that the average thickness of 10 wafers is either greater than 11 or less than 9 micrometers.
(b) Determine the number of wafers that needs to be measured such that the probability that the average thickness exceeds 11 micrometers is 0.01.
(c) If the mean thickness is 10 micrometers, what should the standard deviation of thickness equal so that the probability that the average of 10 wafers is either greater than 11 or less than 9 micrometers is 0.001?

5-95. Assume that the weights of individuals are independent and normally distributed with a mean of 160 pounds and a standard deviation of 30 pounds. Suppose that 25 people squeeze into an elevator that is designed to hold 4300 pounds.
(a) What is the probability that the load (total weight) exceeds the design limit?
(b) What design limit is exceeded by 25 occupants with probability 0.0001?

5-8 FUNCTIONS OF RANDOM VARIABLES (CD ONLY)

5-9 MOMENT GENERATING FUNCTION (CD ONLY)

5-10 CHEBYSHEV'S INEQUALITY (CD ONLY)

Supplemental Exercises

5-96. Show that the following function satisfies the properties of a joint probability mass function:

x	y	$f(x, y)$
0	0	1/4
0	1	1/8
1	0	1/8
1	1	1/4
2	2	1/4

5-97. Continuation of Exercise 5-96. Determine the following probabilities:
(a) $P(X \le 0.5, Y \le 1.5)$ (b) $P(X \le 1)$
(c) $P(X \le 1.5)$ (d) $P(X \le 0.5, Y \le 1.5)$
(e) Determine $E(X)$, $E(Y)$, $V(X)$, and $V(Y)$.

5-98. Continuation of Exercise 5-96. Determine the following:
(a) Marginal probability distribution of the random variable X
(b) Conditional probability distribution of Y given that $X = 1$
(c) $E(Y \mid X = 1)$
(d) Are X and Y independent? Why or why not?
(e) Calculate the correlation between X and Y.

5-99. The percentage of people given an antirheumatoid medication who suffer severe, moderate, or minor side effects

are 10, 20, and 70%, respectively. Assume that people react independently and that 20 people are given the medication. Determine the following:

(a) The probability that 2, 4, and 14 people will suffer severe, moderate, or minor side effects, respectively
(b) The probability that no one will suffer severe side effects
(c) The mean and variance of the number of people that will suffer severe side effects
(d) What is the conditional probability distribution of the number of people who suffer severe side effects given that 19 suffer minor side effects?
(e) What is the conditional mean of the number of people who suffer severe side effects given that 19 suffer minor side effects?

5-100. The backoff torque required to remove bolts in a steel plate is rated as high, moderate, or low. Historically, the probability of a high, moderate, or low rating is 0.6, 0.3, or 0.1, respectively. Suppose that 20 bolts are evaluated and that the torque ratings are independent.

(a) What is the probability that 12, 6, and 2 bolts are rated as high, moderate, and low, respectively?
(b) What is the marginal distribution of the number of bolts rated low?
(c) What is the expected number of bolts rated low?
(d) What is the probability that the number of bolts rated low is greater than two?

5-101. Continuation of Exercise 5-100

(a) What is the conditional distribution of the number of bolts rated low given that 16 bolts are rated high?
(b) What is the conditional expected number of bolts rated low given that 16 bolts are rated high?
(c) Are the numbers of bolts rated high and low independent random variables?

5-102. To evaluate the technical support from a computer manufacturer, the number of rings before a call is answered by a service representative is tracked. Historically, 70% of the calls are answered in two rings or less, 25% are answered in three or four rings, and the remaining calls require five rings or more. Suppose you call this manufacturer 10 times and assume that the calls are independent.

(a) What is the probability that eight calls are answered in two rings or less, one call is answered in three or four rings, and one call requires five rings or more?
(b) What is the probability that all 10 calls are answered in four rings or less?
(c) What is the expected number of calls answered in four rings or less?

5-103. Continuation of Exercise 5-102

(a) What is the conditional distribution of the number of calls requiring five rings or more given that eight calls are answered in two rings or less?
(b) What is the conditional expected number of calls requiring five rings or more given that eight calls are answered in two rings or less?
(c) Are the number of calls answered in two rings or less and the number of calls requiring five rings or more independent random variables?

5-104. Determine the value of c such that the function $f(x, y) = cx^2y$ for $0 \le x \le 3$ and $0 \le y \le 2$ satisfies the properties of a joint probability density function.

5-105. Continuation of Exercise 5-104. Determine the following:

(a) $P(X = 1, Y \le 1)$ (b) $P(X \le 2.5)$
(c) $P(1 \le Y \le 2.5)$ (d) $P(X \le 2, 1 \le Y \le 1.5)$
(e) $E(X)$ (f) $E(Y)$

5-106. Continuation of Exercise 5-104.

(a) Determine the marginal probability distribution of the random variable X.
(b) Determine the conditional probability distribution of Y given that $X = 1$.
(c) Determine the conditional probability distribution of X given that $Y = 1$.

5-107. The joint distribution of the continuous random variables X, Y, and Z is constant over the region $x^2 \le y^2 \le 1$, $0 \le z \le 4$.

(a) Determine $P(X^2 \le Y^2 \le 0.5)$
(b) Determine $P(X^2 \le Y^2 \le 0.5, Z \le 2)$
(c) What is the joint conditional probability density function of X and Y given that $Z = 1$?
(d) What is the marginal probability density function of X?

5-108. Continuation of Exercise 5-107.

(a) Determine the conditional mean of Z given that $X = 0$ and $Y = 0$.
(b) In general, determine the conditional mean of Z given that $X = x$ and $Y = y$.

5-109. Suppose that X and Y are independent, continuous uniform random variables for $0 \le x \le 1$ and $0 \le y \le 1$. Use the joint probability density function to determine the probability that $|X - Y| \le 0.5$.

5-110. The lifetimes of six major components in a copier are independent exponential random variables with means of 8000, 10,000, 10,000, 20,000, 20,000, and 25,000 hours, respectively.

(a) What is the probability that the lifetimes of all the components exceed 5000 hours?
(b) What is the probability that at least one component lifetime exceeds 25,000 hours?

5-111. Contamination problems in semiconductor manufacturing can result in a functional defect, a minor defect, or no defect in the final product. Suppose that 20, 50, and 30% of the contamination problems result in functional, minor, and no defects, respectively. Assume that the effects of 10 contamination problems are independent.

(a) What is the probability that the 10 contamination problems result in two functional defects and five minor defects?
(b) What is the distribution of the number of contamination problems that result in no defects?

(c) What is the expected number of contamination problems that result in no defects?

5-112. The weight of adobe bricks for construction is normally distributed with a mean of 3 pounds and a standard deviation of 0.25 pound. Assume that the weights of the bricks are independent and that a random sample of 25 bricks is chosen.

(a) What is the probability that the mean weight of the sample is less than 2.95 pounds?

(b) What value will the mean weight exceed with probability 0.99?

5-113. The length and width of panels used for interior doors (in inches) are denoted as X and Y, respectively. Suppose that X and Y are independent, continuous uniform random variables for $17.75 \le x \le 18.25$ and $4.75 \le y \le 5.25$, respectively.

(a) By integrating the joint probability density function over the appropriate region, determine the probability that the area of a panel exceeds 90 squared inches.

(b) What is the probability that the perimeter of a panel exceeds 46 inches?

5-114. The weight of a small candy is normally distributed with a mean of 0.1 ounce and a standard deviation of 0.01 ounce. Suppose that 16 candies are placed in a package and that the weights are independent.

(a) What are the mean and variance of package net weight?

(b) What is the probability that the net weight of a package is less than 1.6 ounces?

(c) If 17 candies are placed in each package, what is the probability that the net weight of a package is less than 1.6 ounces?

5-115. The time for an automated system in a warehouse to locate a part is normally distributed with a mean of 45 seconds and a standard deviation of 30 seconds. Suppose that independent requests are made for 10 parts.

(a) What is the probability that the average time to locate 10 parts exceeds 60 seconds?

(b) What is the probability that the total time to locate 10 parts exceeds 600 seconds?

5-116. A mechanical assembly used in an automobile engine contains four major components. The weights of the components are independent and normally distributed with the following means and standard deviations (in ounces):

Component	Mean	Standard Deviation
Left case	4	0.4
Right case	5.5	0.5
Bearing assembly	10	0.2
Bolt assembly	8	0.5

(a) What is the probability that the weight of an assembly exceeds 29.5 ounces?

(b) What is the probability that the mean weight of eight independent assemblies exceeds 29 ounces?

5-117. Suppose X and Y have a bivariate normal distribution with $\mu_X = 4$, $\mu_Y = 1$, $\sigma_X = 4$, $\sigma_Y = 4$, and $\rho = -0.2$. Draw a rough contour plot of the joint probability density function.

5-118. If $f_{XY}(x, y) = \frac{1}{1.2\pi} \exp\left(\frac{-1}{0.72}\left[(x-1)^2 - 1.6(x-1)(y-2) + (y-2)^2\right]\right)$ determine $E(X)$, $E(Y)$, $V(X)$, $V(Y)$, and ρ by recorganizing the parameters in the joint probability density function.

5-119. The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of two integrated layers. The layers are normally distributed with means of 0.5 and 1 millimeters, respectively. The standard deviations of layer thickness are 0.1 and 0.2 millimeters, respectively. The correlation between layers is 0.7.

(a) Determine the mean and variance of the total thickness of the two layers.

(b) What is the probability that the total thickness is less than 1 millimeter?

(c) Let X_1 and X_2 denote the thickness of layers 1 and 2, respectively. A measure of performance of the membrane is a function $2X_1 + 3X_2$ of the thickness. Determine the mean and variance of this performance measure.

5-120. The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of three integrated layers. Layers 1, 2, and 3 are normally distributed with means of 0.5, 1, and 1.5 millimeters, respectively. The standard deviations of layer thickness are 0.1, 0.2, and 0.3, respectively. Also, the correlation between layers 1 and 2 is 0.7, between layers 2 and 3 is 0.5, and between layers 1 and 3 is 0.3.

(a) Determine the mean and variance of the total thickness of the three layers.

(b) What is the probability that the total thickness is less than 1.5 millimeters?

5-121. A small company is to decide what investments to use for cash generated from operations. Each investment has a mean and standard deviation associated with the percentage gain. The first security has a mean percentage gain of 5% with a standard deviation of 2%, and the second security provides the same mean of 5% with a standard deviation of 4%. The securities have a correlation of -0.5, so there is a negative correlation between the percentage returns. If the company invests two million dollars with half in each security, what is the mean and standard deviation of the percentage return? Compare the standard deviation of this strategy to one that invests the two million dollars into the first security only.

MIND-EXPANDING EXERCISES

5-122. Show that if $X_1, X_2, \ldots, X_p$ are independent, continuous random variables, $P(X_1 \le A_1, X_2 \le A_2, \ldots, X_p \le A_p) = P(X_1 \le A_1)P(X_2 \le A_2) \cdots P(X_p \le A_p)$ for any regions $A_1, A_2, \ldots, A_p$ in the range of $X_1, X_2, \ldots, X_p$ respectively.

5-123. Show that if $X_1, X_2, \ldots, X_p$ are independent random variables and $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p)$$

You can assume that the random variables are continuous.

5-124. Suppose that the joint probability function of the continuous random variables X and Y is constant on the rectangle $0 \le x \le a$, $0 \le y \le b$. Show that X and Y are independent.

5-125. Suppose that the range of the continuous variables X and Y is $0 \le x \le a$ and $0 \le y \le b$. Also suppose that the joint probability density function $f_{XY}(x, y) = g(x)h(y)$, where $g(x)$ is a function only of x and $h(y)$ is a function only of y. Show that X and Y are independent.

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Bivariate normal distribution
Conditional mean
Conditional probability density function
Conditional probability mass function
Conditional variance
Contour plots
Correlation
Covariance
Independence
Joint probability density function
Joint probability mass function
Linear combinations of random variables
Marginal probability distribution
Multinomial distribution
Reproductive property of the normal distribution

CD MATERIAL

Convolution
Functions of random variables
Jacobian of a transformation
Moment generating function
Uniqueness property of moment generating function
Chebyshev's inequality

5-8 FUNCTIONS OF RANDOM VARIABLES (CD ONLY)

In many situations in statistics, it is necessary to derive the probability distribution of a function of one or more random variables. In this section, we present some results that are helpful in solving this problem.

Suppose that X is a discrete random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ be a function of X that defines a one-to-one transformation between the values of X and Y, and we wish to find the probability distribution of Y. By a one-to-one transformation, we mean that each value x is related to one and only one value of $y = h(x)$ and that each value of y is related to one and only one value of x, say, $x = u(y)$, where $u(y)$ is found by solving $y = h(x)$ for x in terms of y.

Now, the random variable Y takes on the value y when X takes on the value $u(y)$. Therefore, the probability distribution of Y is

$$f_Y(y) = P(Y = y) = P[X = u(y)] = f_X[u(y)]$$

We may state this result as follows.

Suppose that X is a **discrete** random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = h(x)$ can be solved uniquely for x in terms of y. Let this solution be $x = u(y)$. Then the probability distribution of the random variable Y is

$$f_Y(y) = f_X[u(y)] \qquad \text{(S5-1)}$$

EXAMPLE S5-1

Let X be a geometric random variable with probability distribution

$$f_X(x) = p(1 - p)^{x-1}, \qquad x = 1, 2, \ldots$$

Find the probability distribution of $Y = X^2$.

Since $X > 0$, the transformation is one to one; that is, $y = x^2$ and $x = \sqrt{y}$. Therefore, Equation S5-1 indicates that the distribution of the random variable Y is

$$f_Y(y) = f(\sqrt{y}) = p(1 - p)^{\sqrt{y}-1}, \qquad y = 1, 4, 9, 16, \ldots$$

Now suppose that we have two discrete random variables X_1 and X_2 with joint probability distribution $f_{X_1X_2}(x_1, x_2)$ and we wish to find the joint probability distribution $f_{Y_1Y_2}(y_1, y_2)$ of two new random variables $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$. We assume that the functions h_1 and h_2 define a one-to-one transformation between (x_1, x_2) and (y_1, y_2). Solving the equations $y_1 = h_1(x_1, x_2)$ and $y_2 = h_2(x_1, x_2)$ simultaneously, we obtain the unique solution $x_1 = u_1(y_1, y_2)$ and $x_2 = u_2(y_1, y_2)$. Therefore, the random variables Y_1 and Y_2 take on the values y_1 and y_2 when X_1 takes on the value $u_1(y_1, y_2)$ and X_2 takes the value $u_2(y_1, y_2)$. The joint probability distribution of Y_1 and Y_2 is

$$\begin{aligned} f_{Y_1Y_2}(y_1, y_2) &= P(Y_1 = y_1, Y_2 = y_2) \\ &= P[X_1 = u_1(y_1, y_2), X_2 = u_2(y_1, y_2)] \\ &= f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)] \end{aligned}$$

We will also state this result as follows.

Suppose that X_1 and X_2 are **discrete** random variables with joint probability distribution $f_{X_1X_2}(x_1, x_2)$, and let $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ define one-to-one transformations between the points (x_1, x_2) and (y_1, y_2) so that the equations $y_1 = h_1(x_1, x_2)$ and $y_2 = h_2(x_1, x_2)$ can be solved uniquely for x_1 and x_2 in terms of y_1 and y_2. Let this solution be $x_1 = u_1(y_1, y_2)$ and $x_2 = u_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)] \qquad \text{(S5-2)}$$

A very important application of Equation S5-2 is in finding the distribution of a random variable Y_1 that is a function of two other random variables X_1 and X_2. That is, let $Y_1 = h_1(X_1, X_2)$ where X_1 and X_2 are discrete random variables with joint distribution $f_{X_1X_2}(x_1, x_2)$. We want to find the probability distribution of Y_1, say, $f_{Y_1}(y_1)$. To do this, we define a second function $Y_2 = h_2(X_1, X_2)$ so that the one-to-one correspondence between the points (x_1, x_2) and (y_1, y_2) is maintained, and we use the result in Equation S5-2 to find the joint probability distribution of Y_1 and Y_2. Then the distribution of Y_1 alone is found by summing over the y_2 values in this joint distribution. That is, $f_{Y_1}(y_1)$ is just the **marginal probability distribution** of Y_1, or

$$f_{Y_1}(y_1) = \sum_{y_2} f_{Y_1Y_2}(y_1, y_2)$$

EXAMPLE S5-2

Consider the case where X_1 and X_2 are independent Poisson random variables with parameters λ_1 and λ_2, respectively. We will find the distribution of the random variable $Y_1 = X_1 + X_2$.

The joint distribution of X_1 and X_2 is

$$\begin{aligned} f_{X_1X_2}(x_1, x_2) &= f_{X_1}(x_1)\, f_{X_2}(x_2) \\ &= \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\,\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1}\lambda_2^{x_2}}{x_1!\,x_2!}, \qquad x_1 = 0, 1, \ldots, \qquad x_2 = 0, 1, \ldots \end{aligned}$$

because X_1 and X_2 are independent. Now to use Equation S5-2 we need to define a second function $Y_2 = h_2(X_1, X_2)$. Let this function be $Y_2 = X_2$. Now the solutions for x_1 and x_2 are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus, from Equation S5-2 the joint probability distribution of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y_1-y_2}\lambda_2^{y_2}}{(y_1-y_2)!\,y_2!}, \qquad y_1 = 0, 1, 2, \ldots, \qquad y_2 = 0, 1, \ldots, y_1$$

Because $x_1 \geq 0$, the transformation $x_1 = y_1 - y_2$ requires that $x_2 = y_2$ must always be less than or equal to y_1. Thus, the values of y_2 are $0, 1, \ldots, y_1$, and the marginal probability distribution of Y_1 is obtained as follows:

$$f_{Y_1}(y_1) = \sum_{y_2=0}^{y_1} f_{Y_1Y_2}(y_1, y_2) = \sum_{y_2=0}^{y_1} \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y_1-y_2}\lambda_2^{y_2}}{(y_1-y_2)!\,y_2!}$$

$$\frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y_1}}{y_1!}\sum_{y_2=0}^{y_1}\frac{y_1!}{(y_1-y_2)!\,y_2!}\left(\frac{\lambda_1}{\lambda_2}\right)^{y_2}=\frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y_1}}{y_1!}\sum_{y_2=0}^{y_1}\binom{y_1}{y_2}\left(\frac{\lambda_1}{\lambda_2}\right)^{y_2}$$

The summation in this last expression is the binomial expansion of $(1+\lambda_1/\lambda_2)^{y_1}$, so

$$f_{Y_1}(y_1)=\frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y_1}(1+\lambda_1/\lambda_2)^{y_1}}{y_1!},\qquad y_1=0,1,\ldots$$

We recognize this as a Poisson distribution with parameter $\lambda_1+\lambda_2$. Therefore, we have shown that the sum of two independent Poisson random variables with parameters λ_1 and λ_2 has a Poisson distribution with parameter $\lambda_1+\lambda_2$.

We now consider the situation where the random variables are continuous. Let $Y=h(X)$, with X continuous and the transformation is one to one.

Suppose that X is a **continuous** random variable with probability distribution $f_X(x)$. The function $Y=h(X)$ is a one-to-one transformation between the values of Y and X so that the equation $y=h(x)$ can be uniquely solved for x in terms of y. Let this solution be $x=u(y)$. The probability distribution of Y is

$$f_Y(y)=f_X[u(y)]\,|J| \tag{S5-3}$$

where $J=u'(y)$ is called the **Jacobian** of the transformation and the absolute value of J is used.

Equation S5-3 is shown as follows. Let the function $y=h(x)$ be an increasing function of x. Now

$$P(Y\le a)=P[X\le u(a)]=\int_{-\infty}^{u(a)}f_X(x)\,dx$$

If we change the variable of integration from x to y by using $x=u(y)$, we obtain $dx=u'(y)\,dy$ and then

$$P(Y\le a)=\int_{-\infty}^{a}f_X[u(y)]u'(y)\,dy$$

Since the integral gives the probability that $Y\le a$ for all values of a contained in the feasible set of values for y, $f_X[u(y)]u'(y)$ must be the probability density of Y. Therefore, the probability distribution of Y is

$$f_Y(y)=f_X[u(y)]u'(y)=f_X[u(y)]J$$

If the function $y=h(x)$ is a decreasing function of x, a similar argument holds.

EXAMPLE S5-3

Let X be a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{8}, \quad 0 \le x \le 4$$

Find the probability distribution of $Y = h(X) = 2X + 4$.

Note that $y = h(x) = 2x + 4$ is an increasing function of x. The inverse solution is $x = u(y) = (y - 4)/2$, and from this we find the Jacobian to be $J = u'(y) = dx/dy = 1/2$. Therefore, from S5-3 the probability distribution of Y is

$$f_Y(y) = \frac{(y - 4)/2}{8}\left(\frac{1}{2}\right) = \frac{y - 4}{32}, \qquad 4 \le y \le 12$$

We now consider the case where X_1 and X_2 are continuous random variables and we wish to find the joint probability distribution of $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ where the transformation is one to one. The application of this will typically be in finding the probability distribution of $Y_1 = h_1(X_1, X_2)$, analogous to the discrete case discussed above. We will need the following result.

Suppose that X_1 and X_2 are **continuous** random variables with joint probability distribution $f_{X_1X_2}(x_1, x_2)$, and let $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2). Let the equations $y_1 = h_1(x_1, x_2)$ and $y_2 = h_2(x_1, x_2)$ be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 as $x_1 = u_1(y_1, y_2)$ and $x_2 = u_2(y_1, y_2)$. Then the joint probability of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)]\,|J| \tag{S5-4}$$

where J is the *Jacobian* and is given by the following determinant:

$$J = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix}$$

and the absolute value of the determinant is used.

This result can be used to find $f_{Y_1Y_2}(y_1, y_2)$, the joint probability distribution of Y_1 and Y_2. Then the probability distribution of Y_1 is

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1Y_2}(y_1, y_2)\,dy_2$$

That is, $f_{Y_1}(y_1)$ is the **marginal probability distribution** of Y_1.

EXAMPLE S5-4

Suppose that X_1 and X_2 are independent exponential random variables with $f_{X_1}(x_1) = 2e^{-2x_1}$ and $f_{X_2}(x_2) = 2e^{-2x_2}$. Find the probability distribution of $Y = X_1/X_2$.

The joint probability distribution of X_1 and X_2 is

$$f_{X_1X_2}(x_1, x_2) = 4e^{-2(x_1+x_2)}, \qquad x_1 > 0, \qquad x_2 > 0$$

because X_1 and X_2 are independent. Let $Y_1 = h_1(X_1, X_2) = X_1/X_2$ and $Y_2 = h_2(X_1, X_2) = X_1 + X_2$. The inverse solutions of $y_1 = x_1/x_2$ and $y_2 = x_1 + x_2$ are $x_1 = y_1y_2/(1 + y_1)$ and $x_2 = y_2/(1 + y_1)$, and it follows that

$$\frac{\partial x_1}{\partial y_1} = y_2\left(\frac{1}{(1+y_1)^2}\right), \qquad \frac{\partial x_1}{\partial y_2} = \left(\frac{y_1}{(1+y_1)}\right)$$

$$\frac{\partial x_2}{\partial y_1} = y_2\left(\frac{-1}{(1+y_1)^2}\right), \qquad \frac{\partial x_2}{\partial y_2} = \left(\frac{1}{(1+y_1)}\right)$$

Therefore

$$J = \begin{vmatrix} \dfrac{y_2}{(1+y_1)^2} & \dfrac{y_1}{(1+y_1)} \\ \dfrac{-y_2}{(1+y_1)^2} & \dfrac{1}{(1+y_1)} \end{vmatrix} = \frac{y_2}{(1+y_1)^2}$$

and from Equation S5-4 the joint probability distribution of Y_1 and Y_2 is

$$\begin{aligned} f_{Y_1Y_2}(y_1, y_2) &= f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)]\,|J| \\ &= 4e^{-2[y_1y_2/(1+y_1)+y_2/(1+y_1)]}\,\frac{y_2}{(1+y_1)^2} \\ &= 4e^{-2y_2}\,\frac{y_2}{(1+y_1)^2} \end{aligned}$$

for $y_1 > 0, y_2 > 0$. We need to find the distribution of $Y_1 = X_1/X_2$. This is the marginal probability distribution of Y_1, or

$$\begin{aligned} f_{Y_1}(y_1) &= \int_0^\infty f_{Y_1Y_2}(y_1, y_2)\,dy_2 \\ &= \int_0^\infty 4e^{-2y_2}\,[y_2/(1+y_1)^2]\,dy_2 \\ &= \frac{1}{(1+y_1)^2}, \qquad y_1 > 0 \end{aligned}$$

An important application of Equation S5-4 is to obtain the distribution of the sum of two *independent* random variables X_1 and X_2. Let $Y_1 = X_1 + X_2$ and let $Y_2 = X_2$. The inverse solutions are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Therefore,

$$\frac{\partial x_1}{\partial y_1} = 1 \qquad \frac{\partial x_1}{\partial y_2} = -1$$

$$\frac{\partial x_2}{\partial y_1} = 0 \qquad \frac{\partial x_2}{\partial y_2} = 1$$

and $|J| = 1$. From Equation S5-4, the joint probability density function of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1}(y_1 - y_2) f_{X_2}(y_2)$$

Therefore, the marginal probability density function of Y_1 is

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1}(y_1 - y_2) f_{X_2}(y_2)\, dy_2$$

The notation is simpler if the variable of integration y_2 is replaced with x and y_1 is replaced with y. Then the following result is obtained.

Convolution of X_1 and X_2

If X_1 and X_2 are independent random variables with probability density functions $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$, respectively, the probability density function of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x) f_{X_2}(x)\, dx \qquad \text{(S5-5)}$$

The probability density function of Y in Equation S5-5 is referred to as the **convolution** of the probability density functions for X_1 and X_2. This concept is commonly used for transformations (such as Fourier transformations) in mathematics. This integral may be evaluated numerically to obtain the probability density function of Y, even for complex probability density functions for X_1 and X_2. A similar result can be obtained for discrete random variables with the integral replaced with a sum.

In some problems involving transformations, we need to find the probability distribution of the random variable $Y = h(X)$ when X is a continuous random variable, but the transformation is not one to one. The following result is helpful.

Suppose that X is a **continuous** random variable with probability distribution $f_X(x)$, and $Y = h(X)$ is a transformation that is not one to one. If the interval over which X is defined can be partitioned into m mutually exclusive disjoint sets such that each of the inverse functions $x_1 = u_1(y)$, $x_2 = u_2(y)$, $\ldots$, $x_m = u_m(y)$ of $y = u(x)$ is one to one, the probability distribution of Y is

$$f_Y(y) = \sum_{i=1}^{m} f_X[u_i(y)] |J_i| \qquad \text{(S5-6)}$$

where $J_i = u_i'(y)$, $i = 1, 2, \ldots, m$ and the absolute values are used.

To illustrate how this equation is used, suppose that X is a normal random variable with mean μ and variance σ^2, and we wish to show that the distribution of $Y = (X - \mu)^2/\sigma^2$ is a

chi-squared distribution with one degree of freedom. Let $Z = (X - \mu)/\sigma$, and $Y = Z^2$. The probability distribution of Z is the standard normal; that is,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty$$

The inverse solutions of $y = z^2$ are $z = \pm\sqrt{y}$, so the transformation is not one to one. Define $z_1 = -\sqrt{y}$ and $z_2 = +\sqrt{y}$ so that $J_1 = -1/(2\sqrt{y})$ and $J_2 = 1/(2\sqrt{y})$. Then by Equation S5-6, the probability distribution of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} \left|\frac{-1}{2\sqrt{y}}\right| + \frac{1}{\sqrt{2\pi}} e^{-y/2} \left|\frac{1}{2\sqrt{y}}\right|$$
$$= \frac{1}{2^{1/2}\sqrt{\pi}} y^{1/2-1} e^{-y/2}, \qquad y > 0$$

Now it can be shown that $\sqrt{\pi} = \Gamma(1/2)$, so we may write $f(y)$ as

$$f_Y(y) = \frac{1}{2^{1/2}\,\Gamma\left(\frac{1}{2}\right)} y^{1/2-1} e^{-y/2}, \qquad y > 0$$

which is the chi-squared distribution with 1 degree of freedom.

EXERCISES FOR SECTION 5-8

S5-1. Suppose that X is a random variable with probability distribution

$$f_X(x) = 1/4, \qquad x = 1, 2, 3, 4$$

Find the probability distribution of the random $Y = 2X - 1$.

S5-2. Let X be a binomial random variable with $p = 0.25$ and $n = 3$. Find the probability distribution of the random variable $Y = X^2$.

S5-3. Suppose that X is a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{18}, \qquad 0 < x < 6$$

(a) Find the probability distribution of the random variable $Y = 2X - 10$.
(b) Find the expected value of Y.

S5-4. Suppose that X has a uniform probability distribution

$$f_X(x) = 1, \qquad 0 < x < 1$$

Show that the probability distribution of the random variable $Y = -2 \ln X$ is chi-squared with two degrees of freedom.

S5-5. A current of I amperes flows through a resistance of R ohms according to the probability distribution

$$f_I(i) = 2i, \qquad 0 < i < 1$$

Suppose that the resistance is also a random variable with probability distribution

$$f_R(r) = 1, \qquad 0 < r < 1$$

Assume that I and R are independent.
(a) Find the probability distribution for the power (in watts) $P = I^2R$.
(b) Find $E(P)$.

S5-6. A random variable X has the following probability distribution:

$$f_X(x) = e^{-x}, \qquad x > 0$$

(a) Find the probability distribution for $Y = X^2$.
(b) Find the probability distribution for $Y = X^{1/2}$.
(c) Find the probability distribution for $Y = \ln X$.

S5-7. The velocity of a particle in a gas is a random variable V with probability distribution

$$f_V(v) = av^2e^{-bv} \qquad v \geq 0$$

where b is a constant that depends on the temperature of the gas and the mass of the particle.
(a) Find the value of the constant a.
(b) The kinetic energy of the particle is $W = mV^2/2$. Find the probability distribution of W.

S5-8. Suppose that X has the probability distribution

$$f_X(x) = 1, \qquad 1 \leq x \leq 2$$

Find the probability distribution of the random variable $Y = e^X$.

S5-9. Prove that Equation S5-3 holds when $y = h(x)$ is a decreasing function of x.

S5-10. The random variable X has the probability distribution

$$f_X(x) = \frac{x}{8}, \qquad 0 \leq x \leq 4$$

Find the probability distribution of $Y = (X - 2)^2$.

S5-11. Consider a rectangle with sides of length S_1 and S_2, where S_1 and S_2 are independent random variables. The probability distributions of S_1 and S_2 are

$$f_{S_1}(s_1) = 2s_1, \qquad 0 \leq s_1 \leq 1$$

and

$$f_{S_2}(s_2) = \frac{s_2}{8}, \qquad 0 \leq s_2 \leq 4$$

(a) Find the joint distribution of the area of the rectangle $A = S_1 S_2$ and the random variable $Y = S_1$.
(b) Find the probability distribution of the area A of the rectangle.

S5-12. Suppose we have a simple electrical circuit in which Ohm's law $V = IR$ holds. We are interested in the probability distribution of the resistance R given that V and I are independent random variables with the following distributions:

$$f_V(v) = e^{-v}, \qquad v \geq 0$$

and

$$f_I(i) = 1, \qquad 1 \leq i \leq 2$$

Find the probability distribution of R.

5-9 MOMENT GENERATING FUNCTIONS (CD ONLY)

Suppose that X is a random variable with mean μ. Throughout this book we have used the idea of the expected value of the random variable X, and in fact $E(X) = \mu$. Now suppose that we are interested in the expected value of a particular function of X, say, $g(X) = X^r$. The expected value of this function, or $E[g(X)] = E(X^r)$, is called the rth moment about the origin of the random variable X, which we will denote by μ'_r.

Definition

The rth **moment about the origin** of the random variable X is

$$\mu'_r = E(X^r) = \begin{cases} \sum_x x^r f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^r f(x)\,dx, & X \text{ continuous} \end{cases} \tag{S5-7}$$

Notice that the first moment about the origin is just the mean, that is, $E(X) = \mu'_1 = \mu$. Furthermore, since the second moment about the origin is $E(X^2) = \mu'_2$, we can write the variance of a random variable in terms of origin moments as follows:

$$\sigma^2 = E(X^2) = [E(X)]^2 = \mu'_2 - \mu^2$$

The moments of a random variable can often be determined directly from the definition in Equation S5-7, but there is an alternative procedure that is frequently useful that makes use of a special function.

Definition

The **moment generating function** of the random variable X is the expected value of e^{tX} and is denoted by $M_X(t)$. That is,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, & X \text{ continuous} \end{cases} \tag{S5-8}$$

The moment generating function $M_X(t)$ will exist only if the sum or integral in the above definition converges. If the moment generating function of a random variable does exist, it can be used to obtain all the origin moments of the random variable.

Let X be a random variable with moment generating function $M_X(t)$. Then

$$\mu'_r = \left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} \tag{S5-9}$$

Assuming that we can differentiate inside the summation and integral signs,

$$\frac{d^r M_X(t)}{dt^r} = \begin{cases} \sum_x x^r e^{tx} f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^r e^{tx} f(x)\, dx, & X \text{ continuous} \end{cases}$$

Now if we set $t = 0$ in this expression, we find that

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = E(X^r)$$

EXAMPLE S5-5

Suppose that X has a binomial distribution, that is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

Determine the moment generating function and use it to verify that the mean and variance of the binomial random variable are $\mu = np$ and $\sigma^2 = np(1-p)$.

From the definition of a moment generating function, we have

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

This last summation is the binomial expansion of $[pe^t + (1 - p)]^n$, so

$$M_X(t) = [pe^t + (1 - p)]^n$$

Taking the first and second derivatives, we obtain

$$M'_X(t) = \frac{dM_X(t)}{dt} = npe^t[1 - p(1 - e^t)]^{n-1}$$

and

$$M''_X(t) = \frac{d^2M_X(t)}{dt^2} = npe^t(1 - p) + npe^t[(n - 1)pe^t(1 - p(1 - e^t)]^{n-2}$$

If we set $t = 0$ in $M'_X(t)$, we obtain

$$M'_X(t)|_{t=0} = \mu'_1 = np$$

which is the mean of the binomial random variable X. Now if we set $t = 0$ in $M''_X(t)$,

$$M''_X(t)|_{t=0} = \mu'_2 = np(1 - p) + np = np(1 - p) + (np)^2$$

Therefore, the variance of the binomial random variable is

$$\sigma^2 = \mu'_2 - \mu^2 = np(1 - p) + (np)^2 - (np)^2 = np(1 - p)$$

EXAMPLE S5-6

Find the moment generating function of the normal random variable and use it to show that the mean and variance of this random variable are μ and σ^2, respectively.

The moment generating function is

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}\,dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-(x^2 - 2\mu x - 2\sigma^2 tx + \mu^2)/2\sigma^2}\,dx$$

If we complete the square in the exponent, we have

$$x^2 - 2\mu x - 2\sigma^2 tx + \mu^2 = x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2 - 2\mu\sigma^2 t - \sigma^4 t^2$$

and then

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-[x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2]/2\sigma^2}e^{(2\mu\sigma^2 t + \sigma^4 t^2)/2\sigma^2}\,dx$$

$$= e^{\mu t + \sigma^2 t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-[x - (\mu + \sigma^2 t)]^2/2\sigma^2}dx$$

Let $u = (x - \mu - t\sigma^2)/\sigma$. Then $dx = \sigma\, du$, and this last expression above becomes

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

Now the integral is just the total area under a standard normal density, which is 1, so the moment generating function of a normal random variable is

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Differentiating this function twice with respect to t and setting $t = 0$ in the result, we find

$$\left.\frac{dM_X(t)}{dt}\right|_{t=0} = \mu \qquad \text{and} \qquad \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \mu^2 + \sigma^2$$

Therefore, the variance of the normal random variable is

$$\sigma^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

Moment generating functions have many important and useful properties. One of the most important of these is the **uniqueness property.** That is, the moment generating function of a random variable is unique when it exists, so if we have two random variables X and Y, say, with moment generating functions $M_X(t)$ and $M_Y(t)$, then if $M_X(t) = M_Y(t)$ for all values of t, both X and Y have the same probability distribution. Some of the other useful properties of the moment generating function are summarized as follows.

Properties of Moment Generating Functions

If X is a random variable and a is a constant, then

(1) $M_{X-a}(t) = e^{at} M_X(t)$

(2) $M_{aX}(t) = M_X(at)$

If $X_1, X_2, \ldots, X_n$ are independent random variables with moment generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively, and if $Y = X_1 + X_2 + \cdots + X_n$, then the moment generating function of Y is

(3) $M_Y(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots \cdot M_{X_n}(t)$ (S5-10)

Property (1) follows from $M_{X-a}(t) = E[e^{t(X-a)}] = e^{at}E[e^{tX}] = e^{at}M_X(t)$. Property (2) follows from $M_{aX}(t) = E[e^{t(aX)}] = E[e^{(at)X}] = M_X(at)$. Consider property (3) for the case where the X's are continuous random variables:

$$M_Y(t) = E[e^{tY}] = E[e^{t(X_1+X_2+\cdots+X_n)}]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{t(x_1+x_2+\cdots+x_n)} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n$$

Since the X's are independent,

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots \cdot f_{X_n}(x_n)$$

and one may write

$$M_Y(t) = \int e^{tx_1} f_{X_1}(x_1)\, dx_1 \int e^{tx_2} f_{X_2}(x_2)\, dx_2 \ldots \int e^{tx_n} f_{X_n}(x_n)\, dx_n$$
$$= M_{X_1}(t) \cdot M_{X_2}(t) \cdot \ldots \cdot M_{X_n}(t)$$

For the case when the X's are discrete, we would use the same approach replacing integrals with summations.

Equation S5-10 is particularly useful. In many situations we need to find the distribution of the sum of two or more independent random variables, and often this result makes the problem very easy. This is illustrated in the following example.

EXAMPLE S5-7

Suppose that X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, respectively. Find the probability distribution of $Y = X_1 + X_2$.

The moment generating function of a Poisson random variable with parameter λ is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

so the moment generating functions of X_1 and X_2 are $M_{X_1}(t) = e^{\lambda_1(e^t - 1)}$ and $M_{X_2}(t) = e^{\lambda_2(e^t - 1)}$, respectively. Using Equation S5-10, we find that the moment generating function of $Y = X_1 + X_2$ is

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) = e^{\lambda_1(e^t - 1)} e^{\lambda_2(e^t - 1)} = e^{(\lambda_1 + \lambda_2)(e^t - 1)}$$

which is recognized as the moment generating function of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Therefore, we have shown that the sum of two independent Poisson random variables with parameters λ_1 and λ_2 is a Poisson random variable with parameters equal to the sum of the two parameters $\lambda_1 + \lambda_2$.

EXERCISES FOR SECTION 5-9

S5-13. A random variable X has the discrete uniform distribution

$$f(x) = \frac{1}{m}, \qquad x = 1, 2, \ldots, m$$

(a) Show that the moment generating function is

$$M_X(t) = \frac{e^t(1 - e^{tm})}{m(1 - e^t)}$$

(b) Use $M_X(t)$ to find the mean and variance of X.

S5-14. A random variable X has the Poisson distribution

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, \ldots$$

(a) Show that the moment generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

(b) Use $M_X(t)$ to find the mean and variance of the Poisson random variable.

S5-15. The geometric random variable X has probability distribution

$$f(x) = (1 - p)^{x-1} p, \quad x = 1, 2, \ldots$$

(a) Show that the moment generating function is

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}$$

(b) Use $M_X(t)$ to find the mean and variance of X.

S5-16. The chi-squared random variable with k degrees of freedom has moment generating function $M_X(t) = (1 - 2t)^{-k/2}$. Suppose that X_1 and X_2 are independent chi-squared random variables with k_1 and k_2 degrees of freedom, respectively. What is the distribution of $Y = X_1 + X_2$?

S5-17. A continuous random variable X has the following probability distribution:

$$f(x) = 4xe^{-2x}, \qquad x > 0$$

(a) Find the moment generating function for X.
(b) Find the mean and variance of X.

S5-18. The continuous uniform random variable X has density function

$$f(x) = \frac{1}{\beta - \alpha}, \qquad \alpha < x < \beta$$

(a) Show that the moment generating function is

$$M_X(t) = \frac{e^{t\beta} - e^{t\alpha}}{t(\beta - \alpha)}$$

(b) Use $M_X(t)$ to find the mean and variance of X.

S5-19. A random variable X has the exponential distribution

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

(a) Show that the moment generating function of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

(b) Find the mean and variance of X.

S5-20. A random variable X has the gamma distribution

$$f(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}, \qquad x > 0$$

(a) Show that the moment generating function of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}$$

(b) Find the mean and variance of X.

S5-21. Let $X_1, X_2, \ldots, X_r$ be independent exponential random variables with parameter λ.
(a) Find the moment generating function of $Y = X_1 + X_2 + \cdots + X_r$.
(b) What is the distribution of the random variable Y?
[*Hint:* Use the results of Exercise S5-20].

S5-22. Suppose that X_i has a normal distribution with mean μ_i and variance $\sigma_i^2, i = 1, 2$. Let X_1 and X_2 be independent.
(a) Find the moment generating function of $Y = X_1 + X_2$.
(b) What is the distribution of the random variable Y?

S5-23. Show that the moment generating function of the chi-squared random variable with k degrees of freedom is $M_X(t) = (1 - 2t)^{-k/2}$. Show that the mean and variance of this random variable are k and $2k$, respectively.

S5-24. Continuation of Exercise S5-20.
(a) Show that by expanding e^{tX} in a power series and taking expectations term by term we may write the moment generating function as

$$M_X(t) = E(e^{tX})$$
$$= 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \cdots + \mu_r' \frac{t^r}{r!} + \cdots$$

Thus, the coefficient of $t^r/r!$ in this expansion is μ_r', the rth origin moment.
(b) Continuation of Exercise S5-20. Write the power series expansion for $M_X(t)$, the gamma random variable.
(c) Continuation of Exercise S5-20. Find μ_1' and μ_2' using the results of parts (a) and (b). Does this approach give the same answers that you found for the mean and variance of the gamma random variable in Exercise S5-20?

5-10 CHEBYSHEV'S INEQUALITY (CD ONLY)

In Chapter 3 we showed that if X is a normal random variable with mean μ and standard deviation σ, $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95$. This result relates the probability of a normal random variable to the magnitude of the standard deviation. An interesting, similar result that applies to any discrete or continuous random variable was developed by the mathematician Chebyshev in 1867.

Chebyshev's Inequality

For any random variable X with mean μ and variance σ^2,

$$P(|X - \mu| \geq c\sigma) \leq 1/c^2$$

for $c > 0$.

This result is interpreted as follows. The probability that a random variable differs from its mean by at least c standard deviations is less than or equal to $1/c^2$. Note that the rule is useful only for $c > 1$.

For example, using $c = 2$ implies that the probability that *any* random variable differs from its mean by at least two standard deviations is no greater than 1/4. We know that for a normal random variable, this probability is less than 0.05. Also, using $c = 3$ implies that the probability that any random variable differs from its mean by at least three standard deviations is no greater than 1/9. **Chebyshev's inequality** provides a relationship between the standard deviation and the dispersion of the probability distribution of any random variable. The proof is left as an exercise.

Table S5-1 compares probabilities computed by Chebyshev's rule to probabilities computed for a normal random variable.

EXAMPLE S5-8

The process of drilling holes in printed circuit boards produces diameters with a standard deviation of 0.01 millimeter. How many diameters must be measured so that the probability is at least 8/9 that the average of the measured diameters is within 0.005 of the process mean diameter μ?

Let $X_1, X_2, \ldots, X_n$ be the random variables that denote the diameters of n holes. The average measured diameter is $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$. Assume that the X's are independent random variables. From Equation 5-40, $E(\bar{X}) = \mu$ and $V(\bar{X}) = 0.01^2/n$. Consequently, the standard deviation of $\bar{X}$ is $(0.01^2/n)^{1/2}$. By applying Chebyshev's inequality to $\bar{X}$,

$$P(|\bar{X} - \mu| \geq c(0.01^2/n)^{1/2}) \leq 1/c^2$$

Let $c = 3$. Then,

$$P(|\bar{X} - \mu| \geq 3(0.01^2/n)^{1/2}) \leq 1/9$$

Therefore,

$$P(|\bar{X} - \mu| \geq 3(0.01^2/n)^{1/2}) \geq 8/9$$

Table S5-1 Percentage of Distribution Greater than c Standard Deviations from the Mean

c	Chebyshev's Rule for any Probability Distribution	Normal Distribution
1.5	less than 44.4%	13.4%
2	less than 25.0%	4.6%
3	less than 11.1%	0.27%
4	less than 6.3%	0.01%

Thus, the probability that $\overline{X}$ is within $3(0.01^2/n)^{1/2}$ of μ is at least 8/9. Finally, n is chosen such that $3(0.01^2/n)^{1/2} \leq 0.005$. That is,

$$n \geq 3^2\{0.01^2/0.005^2\} = 36$$

EXERCISES FOR SECTION 5-10

S5-25. The photoresist thickness in semiconductor manufacturing has a mean of 10 micrometers and a standard deviation of 1 micrometer. Bound the probability that the thickness is less than 6 or greater than 14 micrometers.

S5-26. Suppose X has a continuous uniform distribution with range $0 \leq x \leq 10$. Use Chebyshev's rule to bound the probability that X differs from its mean by more than two standard deviations and compare to the actual probability.

S5-27. Suppose X has an exponential distribution with mean 20. Use Chebyshev's rule to bound the probability that X differs from its mean by more than two standard deviations and by more than three standard deviations and compare to the actual probabilities.

S5-28. Suppose X has a Poisson distribution with mean $\mu = 4$. Use Chebyshev's rule to bound the probability that X differs from its mean by more than two standard deviations and by more than three standard deviations and compare to the actual probabilities.

S5-29. Consider the process of drilling holes in printed circuits boards. Assume that the standard deviation of the diameters is 0.01 and that the diameters are independent. Suppose that the average of 500 diameters is used to estimate the process mean.

(a) The probability is at least 15/16 that the measured average is within some bound of the process mean. What is the bound?

(b) If it is assumed that the diameters are normally distributed, determine the bound such that the probability is 15/16 that the measured average is closer to the process mean than the bound.

S5-30. Prove Chebyshev's rule from the following steps. Define the random variable Y as follows:

$$Y = \begin{cases} 1 & \text{if } |X - \mu| \geq c \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine $E(Y)$

(b) Show that $(X - \mu)^2 \geq (X - \mu)^2 Y \geq c^2 Y$

(c) Using part (b), show that $E\{(X - \mu)^2\} \geq c^2 E\{Y\}$

(d) Using part (c), complete the derivation of Chebyshev's inequality.

6 Random Sampling and Data Description

CHAPTER OUTLINE

6-1 DATA SUMMARY AND DISPLAY
6-2 RANDOM SAMPLING
6-3 STEM-AND-LEAF DIAGRAMS
6-4 FREQUENCY DISTRIBUTIONS AND HISTOGRAMS
6-5 BOX PLOTS
6-6 TIME SEQUENCE PLOTS
6-7 PROBABILITY PLOTS
6-8 MORE ABOUT PROBABILITY PLOTTING (CD ONLY)

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Compute and interpret the sample mean, sample variance, sample standard deviation, sample median, and sample range
2. Explain the concepts of sample mean, sample variance, population mean, and population variance
3. Construct and interpret visual data displays, including the stem-and-leaf display, the histogram, and the box plot
4. Explain the concept of random sampling
5. Construct and interpret normal probability plots
6. Explain how to use box plots and other data displays to visually compare two or more samples of data
7. Know how to use simple time series plots to visually display the important features of time-oriented data.

CD MATERIAL

8. Interpret probability plots for distributions other than normal.

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

6-1 DATA SUMMARY AND DISPLAY

Well-constructed data summaries and displays are essential to good statistical thinking, because they can focus the engineer on important features of the data or provide insight about the type of model that should be used in solving the problem. The computer has become an important tool in the presentation and analysis of data. While many statistical techniques require only a hand-held calculator, much time and effort may be required by this approach, and a computer will perform the tasks much more efficiently.

Most statistical analysis is done using a prewritten library of statistical programs. The user enters the data and then selects the types of analysis and output displays that are of interest. Statistical software packages are available for both mainframe machines and personal computers. We will present examples of output from Minitab (one of the most widely-used PC packages), throughout the book. We will not discuss the hands-on use of Minitab for entering and editing data or using commands. This information is found in the software documentation.

We often find it useful to describe data features **numerically.** For example, we can characterize the location or central tendency in the data by the ordinary arithmetic average or mean. Because we almost always think of our data as a sample, we will refer to the arithmetic mean as the **sample mean.**

Definition

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{6-1}$$

EXAMPLE 6-1

Let's consider the eight observations collected from the prototype engine connectors from Chapter 1. The eight observations are $x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. The sample mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{12.6 + 12.9 + \cdots + 13.1}{8}$$

$$= \frac{104}{8} = 13.0 \text{ pounds}$$

A physical interpretation of the sample mean as a measure of location is shown in the dot diagram of the pull-off force data. See Figure 6-1. Notice that the sample mean $\bar{x} = 13.0$ can be thought of as a "balance point." That is, if each observation represents 1 pound of mass placed at the point on the x-axis, a fulcrum located at $\bar{x}$ would exactly balance this system of weights.

The sample mean is the average value of all the observations in the data set. Usually, these data are a **sample** of observations that have been selected from some larger **population** of observations. Here the population might consist of all the connectors that will be manufactured and sold to customers. Recall that this type of population is called a **conceptual** or

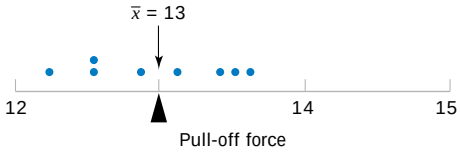


Figure 6-1 The sample mean as a balance point for a system of weights.

hypothetical population, because it does not physically exist. Sometimes there is an actual physical population, such as a lot of silicon wafers produced in a semiconductor factory.

In previous chapters we have introduced the mean of a probability distribution, denoted □. If we think of a probability distribution as a **model** for the population, one way to think of the mean is as the average of all the measurements in the population. For a finite population with N measurements, the mean is

$$\square \square \frac{\sum_{i\square 1}^{N} x_i}{N} \tag{6-2}$$

The sample mean, $\bar{x}$, is a reasonable estimate of the population mean, □. Therefore, the engineer designing the connector using a 3□32-inch wall thickness would conclude, on the basis of the data, that an estimate of the mean pull-off force is 13.0 pounds.

Although the sample mean is useful, it does not convey all of the information about a sample of data. The variability or scatter in the data may be described by the **sample variance** or the **sample standard deviation.**

Definition

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the **sample variance** is

$$s^2 \square \frac{\sum_{i\square 1}^{n} (x_i \square \bar{x})^2}{n \square 1} \tag{6-3}$$

The **sample standard deviation,** s, is the positive square root of the sample variance.

The units of measurements for the sample variance are the square of the original units of the variable. Thus, if x is measured in pounds, the units for the sample variance are $(\text{pounds})^2$. The standard deviation has the desirable property of measuring variability in the original units of the variable of interest, x.

How Does the Sample Variance Measure Variability?

To see how the sample variance measures dispersion or variability, refer to Fig. 6-2, which shows the deviations $x_i \square \bar{x}$ for the connector pull-off force data. The greater the amount of variability in the pull-off force data, the larger in absolute magnitude some of the deviations $x_i \square \bar{x}$ will be. Since the deviations $x_i \square \bar{x}$ always sum to zero, we must use a measure of variability that changes the negative deviations to nonnegative quantities. Squaring the deviations is the approach used in the sample variance. Consequently, if s^2 is small, there is relatively little variability in the data, but if s^2 is large, the variability is relatively large.

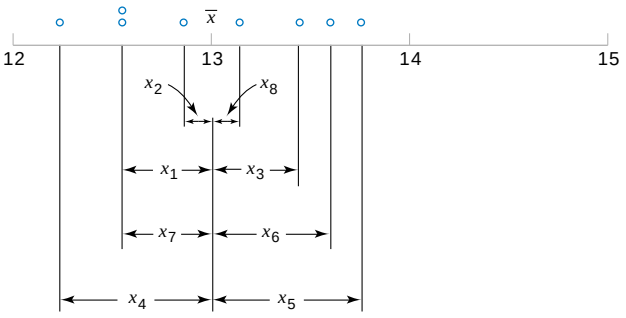


Figure 6-2 How the sample variance measures variability through the deviations $x_i - \bar{x}$.

EXAMPLE 6-2

Table 6-1 displays the quantities needed for calculating the sample variance and sample standard deviation for the pull-off force data. These data are plotted in Fig. 6-2. The numerator of s^2 is

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 1.60$$

so the sample variance is

$$s^2 = \frac{1.60}{8 - 1} = \frac{1.60}{7} = 0.2286 \text{ (pounds)}^2$$

and the sample standard deviation is

$$s = \sqrt{0.2286} = 0.48 \text{ pounds}$$

Computation of s^2

The computation of s^2 requires calculation of $\bar{x}$, n subtractions, and n squaring and adding operations. If the original observations or the deviations $x_i - \bar{x}$ are not integers, the deviations $x_i - \bar{x}$ may be tedious to work with, and several decimals may have to be carried to ensure

Table 6-1 Calculation of Terms for the Sample Variance and Sample Standard Deviation

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
	104.0	0.0	1.60

numerical accuracy. A more efficient computational formula for the sample variance is obtained as follows:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i^2 - \bar{x}^2 - 2\bar{x}x_i)}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n} x_i}{n-1}$$

and since $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$, this last equation reduces to

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \tag{6-4}$$

Note that Equation 6-4 requires squaring each individual x_i, then squaring the sum of the x_i, subtracting $(\sum x_i)^2/n$ from $\sum x_i^2$, and finally dividing by $n - 1$. Sometimes this is called the shortcut method for calculating s^2 (or s).

EXAMPLE 6-3

We will calculate the sample variance and standard deviation using the shortcut method, Equation 6-4. The formula gives

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{1353.6 - \frac{(104)^2}{8}}{7} = \frac{1.60}{7} = 0.2286 \text{ (pounds)}^2$$

and

$$s = \sqrt{0.2286} = 0.48 \text{ pounds}$$

These results agree exactly with those obtained previously.

Analogous to the sample variance s^2, the variability in the population is defined by the **population variance** (σ^2). As in earlier chapters, the positive square root of σ^2, or σ, will denote the **population standard deviation.** When the population is finite and consists of N values, we may define the population variance as

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \tag{6-5}$$

We observed previously that the sample mean could be used as an estimate of the population mean. Similarly, the sample variance is an estimate of the population variance. In Chapter 7, we will discuss **estimation of parameters** more formally.

Note that the divisor for the sample variance is the sample size minus one $(n - 1)$, while for the population variance it is the population size N. If we knew the true value of the population mean μ, we could find the *sample* variance as the average squared deviation of the sample observations about μ. In practice, the value of μ is almost never known, and so the sum of

the squared deviations about the sample average $\bar{x}$ must be used instead. However, the observations x_i tend to be closer to their average, $\bar{x}$, than to the population mean, μ. Therefore, to compensate for this we use $n - 1$ as the divisor rather than n. If we used n as the divisor in the sample variance, we would obtain a measure of variability that is, on the average, consistently smaller than the true population variance σ^2.

Another way to think about this is to consider the sample variance s^2 as being based on $n - 1$ **degrees of freedom.** The term *degrees of freedom* results from the fact that the n deviations $x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}$ always sum to zero, and so specifying the values of any $n - 1$ of these quantities automatically determines the remaining one. This was illustrated in Table 6-1. Thus, only $n - 1$ of the n deviations, $x_i - \bar{x}$, are freely determined.

In addition to the sample variance and sample standard deviation, the **sample range,** or the difference between the largest and smallest observations, is a useful measure of variability. The sample range is defined as follows.

Definition

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample range** is

$$r = \max(x_i) - \min(x_i) \tag{6-6}$$

For the pull-off force data, the sample range is $r = 13.6 - 12.3 = 1.3$. Generally, as the variability in sample data increases, the sample range increases.

The sample range is easy to calculate, but it ignores all of the information in the sample data between the largest and smallest values. For example, the two samples 1, 3, 5, 8, and 9 and 1, 5, 5, 5, and 9, both have the same range ($r = 8$). However, the standard deviation of the first sample is $s_1 = 3.35$, while the standard deviation of the second sample is $s_2 = 2.83$. The variability is actually less in the second sample.

Sometimes, when the sample size is small, say $n = 8$ or 10, the information loss associated with the range is not too serious. For example, the range is used widely in statistical quality control where sample sizes of 4 or 5 are fairly common. We will discuss some of these applications in Chapter 16.

EXERCISES FOR SECTIONS 6-1 AND 6-2

6-1. Eight measurements were made on the inside diameter of forged piston rings used in an automobile engine. The data (in millimeters) are 74.001, 74.003, 74.015, 74.000, 74.005, 74.002, 74.005, and 74.004. Calculate the sample mean and sample standard deviation, construct a dot diagram, and comment on the data.

6-2. In *Applied Life Data Analysis* (Wiley, 1982), Wayne Nelson presents the breakdown time of an insulating fluid between electrodes at 34 kV. The times, in minutes, are as follows: 0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, and 72.89. Calculate the sample mean and sample standard deviation.

6-3. The January 1990 issue of *Arizona Trend* contains a supplement describing the 12 "best" golf courses in the state. The yardages (lengths) of these courses are as follows: 6981, 7099, 6930, 6992, 7518, 7100, 6935, 7518, 7013, 6800, 7041, and 6890. Calculate the sample mean and sample standard deviation. Construct a dot diagram of the data.

6-4. An article in the *Journal of Structural Engineering* (Vol. 115, 1989) describes an experiment to test the yield strength of circular tubes with caps welded to the ends. The first yields (in kN) are 96, 96, 102, 102, 102, 104, 104, 108, 126, 126, 128, 128, 140, 156, 160, 160, 164, and 170. Calculate the sample mean and sample standard deviation. Construct a dot diagram of the data.

6-5. An article in *Human Factors* (June 1989) presented data on visual accommodation (a function of eye movement) when recognizing a speckle pattern on a high-resolution CRT screen. The data are as follows: 36.45, 67.90, 38.77, 42.18, 26.72, 50.77, 39.30, and 49.71. Calculate the sample mean and sample standard deviation. Construct a dot diagram of the data.

6-6. The following data are direct solar intensity measurements (watts/m^2) on different days at a location in southern Spain: 562, 869, 708, 775, 775, 704, 809, 856, 655, 806, 878, 909, 918, 558, 768, 870, 918, 940, 946, 661, 820, 898, 935, 952, 957, 693, 835, 905, 939, 955, 960, 498, 653, 730, and 753. Calculate the sample mean and sample standard deviation.

6-7. The April 22, 1991 issue of *Aviation Week and Space Technology* reports that during Operation Desert Storm, U.S. Air Force F-117A pilots flew 1270 combat sorties for a total of 6905 hours. What is the mean duration of an F-117A mission during this operation? Why is the parameter you have calculated a population mean?

6-8. Preventing fatigue crack propagation in aircraft structures is an important element of aircraft safety. An engineering study to investigate fatigue crack in $n = 9$ cyclically loaded wing boxes reported the following crack lengths (in mm): 2.13, 2.96, 3.02, 1.82, 1.15, 1.37, 2.04, 2.47, 2.60.

(a) Calculate the sample mean.
(b) Calculate the sample variance and sample standard deviation.
(c) Prepare a dot diagram of the data.

6-9. Consider the solar intensity data in Exercise 6-6. Prepare a dot diagram of this data. Indicate where the sample mean falls on this diagram. Give a practical interpretation of the sample mean.

6-10. Exercise 6-5 describes data from an article in *Human Factors* on visual accommodation from an experiment involving a high-resolution CRT screen.

(a) Construct a dot diagram of this data.
(b) Data from a second experiment using a low-resolution screen were also reported in the article. They are 8.85, 35.80, 26.53, 64.63, 9.00, 15.38, 8.14, and 8.24. Prepare a dot diagram for this second sample and compare it to the one for the first sample. What can you conclude about CRT resolution in this situation?

6-11. The pH of a solution is measured eight times by one operator using the same instrument. She obtains the following data: 7.15, 7.20, 7.18, 7.19, 7.21, 7.20, 7.16, and 7.18.

(a) Calculate the sample mean.
(b) Calculate the sample variance and sample standard deviation.
(c) What are the major sources of variability in this experiment?

6-12. An article in the *Journal of Aircraft* (1988) describes the computation of drag coefficients for the NASA 0012 airfoil. Different computational algorithms were used at $M_\infty = 0.7$ with the following results (drag coefficients are in units of drag counts; that is, one count is equivalent to a drag coefficient of 0.0001): 79, 100, 74, 83, 81, 85, 82, 80, and 84. Compute the sample mean, sample variance, and sample standard deviation, and construct a dot diagram.

6-13. The following data are the joint temperatures of the O-rings (°F) for each test firing or actual launch of the space shuttle rocket motor (from *Presidential Commission on the Space Shuttle Challenger Accident*, Vol. 1, pp. 129–131): 84, 49, 61, 40, 83, 67, 45, 66, 70, 69, 80, 58, 68, 60, 67, 72, 73, 70, 57, 63, 70, 78, 52, 67, 53, 67, 75, 61, 70, 81, 76, 79, 75, 76, 58, 31.

(a) Compute the sample mean and sample standard deviation.
(b) Construct a dot diagram of the temperature data.
(c) Set aside the smallest observation (31°F) and recompute the quantities in part (a). Comment on your findings. How "different" are the other temperatures from this last value?

6-2 RANDOM SAMPLING

In most statistics problems, we work with a sample of observations selected from the population that we are interested in studying. Figure 6-3 illustrates the relationship between the population and the sample. We have informally discussed these concepts before; however, we now give the formal definitions of some of these terms.

Definition

A **population** consists of the totality of the observations with which we are concerned.

In any particular problem, the population may be small, large but finite, or infinite. The number of observations in the population is called the **size** of the population. For example, the number of underfilled bottles produced on one day by a soft-drink company is a population of finite size. The observations obtained by measuring the carbon monoxide level every day is a population of infinite size. We often use a **probability distribution** as a **model** for a population. For example, a structural engineer might consider the population of tensile strengths of a

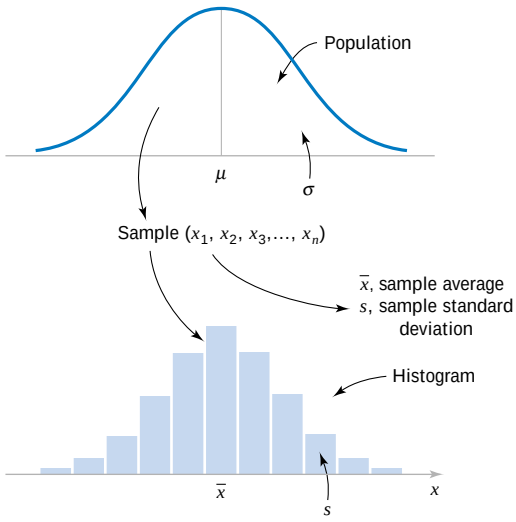


Figure 6-3 Relationship between a population and a sample.

chassis structural element to be normally distributed with mean □ and variance □^{2}. We could refer to this as a **normal population** or a normally distributed population.

In most situations, it is impossible or impractical to observe the entire population. For example, we could not test the tensile strength of all the chassis structural elements because it would be too time consuming and expensive. Furthermore, some (perhaps many) of these structural elements do not yet exist at the time a decision is to be made, so to a large extent, we must view the population as **conceptual.** Therefore, we depend on a subset of observations from the population to help make decisions about the population.

Definition

A **sample** is a subset of observations selected from a population.

For statistical methods to be valid, the sample must be representative of the population. It is often tempting to select the observations that are most convenient as the sample or to exercise judgment in sample selection. These procedures can frequently introduce **bias** into the sample, and as a result the parameter of interest will be consistently underestimated (or overestimated) by such a sample. Furthermore, the behavior of a judgment sample cannot be statistically described. To avoid these difficulties, it is desirable to select a **random sample** as the result of some chance mechanism. Consequently, the selection of a sample is a random experiment and each observation in the sample is the observed value of a random variable. The observations in the population determine the probability distribution of the random variable.

To define a random sample, let X be a random variable that represents the result of one selection of an observation from the population. Let $f(x)$ denote the probability density function of X. Suppose that each observation in the sample is obtained independently, under unchanging conditions. That is, the observations for the sample are obtained by observing X independently under unchanging conditions, say, n times. Let X_i denote the random variable that represents the ith replicate. Then, $X_1, X_2, \ldots, X_n$ is a random sample and the numerical values obtained are denoted as $x_1, x_2, \ldots, x_n$. The random variables in a random sample are independent with the same probability distribution $f(x)$ because of the identical conditions under which each observation is obtained. That is, the marginal probability density function of $X_1, X_2, \ldots, X_n$ is

$f(x_1), f(x_2), \ldots, f(x_n)$, respectively, and by independence the joint probability density function of the random sample is $f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$.

Definition

> The random variables $X_1, X_2, \ldots, X_n$ are a random sample of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

To illustrate this definition, suppose that we are investigating the effective service life of an electronic component used in a cardiac pacemaker and that component life is normally distributed. Then we would expect each of the observations on component life $X_1, X_2, \ldots, X_n$ in a random sample of n components to be independent random variables with exactly the same normal distribution. After the data are collected, the numerical values of the observed lifetimes are denoted as $x_1, x_2, \ldots, x_n$.

The primary purpose in taking a random sample is to obtain information about the unknown population parameters. Suppose, for example, that we wish to reach a conclusion about the proportion of people in the United States who prefer a particular brand of soft drink. Let p represent the unknown value of this proportion. It is impractical to question every individual in the population to determine the true value of p. In order to make an inference regarding the true proportion p, a more reasonable procedure would be to select a random sample (of an appropriate size) and use the observed proportion $\hat{p}$ of people in this sample favoring the brand of soft drink.

The sample proportion, $\hat{p}$ is computed by dividing the number of individuals in the sample who prefer the brand of soft drink by the total sample size n. Thus, $\hat{p}$ is a function of the observed values in the random sample. Since many random samples are possible from a population, the value of $\hat{p}$ will vary from sample to sample. That is, $\hat{p}$ is a random variable. Such a random variable is called a **statistic.**

Definition

> A **statistic** is any function of the observations in a random sample.

We have encountered statistics before. For example, if $X, X_2, \ldots, X_n$ is a random sample of size n, the **sample mean** $\bar{X}$, the **sample variance** S^2, and the **sample standard deviation** S are statistics.

Although numerical summary statistics are very useful, **graphical displays** of sample data are a very powerful and extremely useful way to visually examine the data. We now present a few of the techniques that are most relevant to engineering applications of probability and statistics.

6-3 STEM-AND-LEAF DIAGRAMS

The dot diagram is a useful data display for small samples, up to (say) about 20 observations. However, when the number of observations is moderately large, other graphical displays may be more useful.

For example, consider the data in Table 6-2. These data are the compressive strengths in pounds per square inch (psi) of 80 specimens of a new aluminum-lithium alloy undergoing evaluation as a possible material for aircraft structural elements. The data were recorded in the order of testing, and in this format they do not convey much information about compressive strength. Questions such as "What percent of the specimens fail below 120 psi?" are not easy to answer.

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Because there are many observations, constructing a dot diagram of these data would be relatively inefficient; more effective displays are available for large data sets.

A **stem-and-leaf diagram** is a good way to obtain an informative visual display of a data set $x_1, x_2, \ldots, x_n$, where each number x_i consists of at least two digits. To construct a stem-and-leaf diagram, use the following steps.

Steps for Constructing a Stem-and-Leaf Diagram

(1) Divide each number x_i into two parts: a **stem,** consisting of one or more of the leading digits and a **leaf,** consisting of the remaining digit.

(2) List the stem values in a vertical column.

(3) Record the leaf for each observation beside its stem.

(4) Write the units for stems and leaves on the display.

To illustrate, if the data consist of percent defective information between 0 and 100 on lots of semiconductor wafers, we can divide the value 76 into the stem 7 and the leaf 6. In general, we should choose relatively few stems in comparison with the number of observations. It is usually best to choose between 5 and 20 stems.

EXAMPLE 6-4 To illustrate the construction of a stem-and-leaf diagram, consider the alloy compressive strength data in Table 6-2. We will select as stem values the numbers 7, 8, 9, ..., 24. The resulting stem-and-leaf diagram is presented in Fig. 6-4. The last column in the diagram is a frequency count of the number of leaves associated with each stem. Inspection of this display immediately reveals that most of the compressive strengths lie between 110 and 200 psi and that a central value is somewhere between 150 and 160 psi. Furthermore, the strengths are distributed approximately symmetrically about the central value. The stem-and-leaf diagram enables us to determine quickly some important features of the data that were not immediately obvious in the original display in Table 6-2.

In some data sets, it may be desirable to provide more classes or stems. One way to do this would be to modify the original stems as follows: Divide the stem 5 (say) into two new stems, 5L and 5U. The stem 5L has leaves 0, 1, 2, 3, and 4, and stem 5U has leaves 5, 6, 7, 8, and 9. This will double the number of original stems. We could increase the number of original stems by four by defining five new stems: 5z with leaves 0 and 1, 5t (for twos and three) with leaves 2 and 3, 5f (for fours and fives) with leaves 4 and 5, 5s (for six and seven) with leaves 6 and 7, and 5e with leaves 8 and 9.

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem : Tens and hundreds digits (psi); Leaf: Ones digits (psi)

Figure 6-4 Stem-and-leaf diagram for the compressive strength data in Table 6-2.

EXAMPLE 6-5

Figure 6-5 illustrates the stem-and-leaf diagram for 25 observations on batch yields from a chemical process. In Fig. 6-5(a) we have used 6, 7, 8, and 9 as the stems. This results in too few stems, and the stem-and-leaf diagram does not provide much information about the data. In Fig. 6-5(b) we have divided each stem into two parts, resulting in a display that more

Stem	Leaf
6	1 3 4 5 5 6
7	0 1 1 3 5 7 8 8 9
8	1 3 4 4 7 8 8
9	2 3 5

(a)

Stem	Leaf
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

(b)

Stem	Leaf
6z	1
6t	3
6f	4 5 5
6s	6
6e	
7z	0 1 1
7t	3
7f	5
7s	7
7e	8 8 9
8z	1
8t	3
8f	4 4
8s	7
8e	8 8
9z	
9t	2 3
9f	5
9s	
9e	

(c)

Figure 6-5 Stem-and-leaf displays for Example 6-5. Stem: Tens digits. Leaf: Ones digits.

Character Stem-and-Leaf Display

Stem-and-leaf of Strength
N = 80 Leaf Unit = 1.0

1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

Figure 6-6 A stem-and-leaf diagram from Minitab.

adequately displays the data. Figure 6-5(c) illustrates a stem-and-leaf display with each stem divided into five parts. There are too many stems in this plot, resulting in a display that does not tell us much about the shape of the data.

Figure 6-6 shows a stem-and-leaf display of the compressive strength data in Table 6-2 produced by Minitab. The software uses the same stems as in Fig. 6-4. Note also that the computer orders the leaves from smallest to largest on each stem. This form of the plot is usually called an **ordered stem-and-leaf diagram.** This is not usually done when the plot is constructed manually because it can be time consuming. The computer adds a column to the left of the stems that provides a count of the observations at and above each stem in the upper half of the display and a count of the observations at and below each stem in the lower half of the display. At the middle stem of 16, the column indicates the number of observations at this stem.

The ordered stem-and-leaf display makes it relatively easy to find data features such as percentiles, quartiles, and the median. The sample **median** is a measure of central tendency that divides the data into two equal parts, half below the median and half above. If the number of observations is even, the median is halfway between the two central values. From Fig. 6-6 we find the 40th and 41st values of strength as 160 and 163, so the median is $(160 + 163)/2 = 161.5$. If the number of observations is odd, the median is the central value. The sample **mode** is the most frequently occurring data value. Figure 6-6 indicates that the mode is 158; this value occurs four times, and no other value occurs as frequently in the sample.

We can also divide data into more than two parts. When an ordered set of data is divided into four equal parts, the division points are called **quartiles.** The *first* or *lower quartile*, q_1, is a value that has approximately 25% of the observations below it and approximately 75% of the observations above. The *second quartile*, q_2, has approximately 50% of the observations below its value. The second quartile is exactly equal to the median. The *third* or *upper quartile*, q_3, has approximately 75% of the observations below its value. As in the case of the median, the quartiles may not be unique. The compressive strength data in Fig. 6-6 contains $n = 80$ observations. Minitab software calculates the first and third quartiles as the $(n + 1)/4$

Table 6-3 Summary Statistics for the Compressive Strength Data from Minitab

Variable	N	Mean	Median	StDev	SE Mean
	80	162.66	161.50	33.77	3.78
	Min	Max	Q1	Q3	
	76.00	245.00	143.50	181.00	

and $3(n + 1)/4$ ordered observations and interpolates as needed. For example, $(80 + 1)/4 = 20.25$ and $3(80 + 1)/4 = 60.75$. Therefore, Minitab interpolates between the 20th and 21st ordered observation to obtain $q_1 = 143.50$ and between the 60th and 61st observation to obtain $q_3 = 181.00$. In general, the $100k$th **percentile** is a data value such that approximately $100k\%$ of the observations are at or below this value and approximately $100(1 - k)\%$ of them are above it. Finally, we may use the **interquartile range,** defined as $IQR = q_3 - q_1$, as a measure of variability. The interquartile range is less sensitive to the extreme values in the sample than is the ordinary sample range.

Many statistics software packages provide data summaries that include these quantities. The output obtained for the compressive strength data in Table 6-2 from Minitab is shown in Table 6-3.

EXERCISES FOR SECTION 6-3

6-14. An article in *Technometrics* (Vol. 19, 1977, p. 425) presents the following data on the motor fuel octane ratings of several blends of gasoline:

88.5	98.8	89.6	92.2	92.7	88.4	87.5	90.9
94.7	88.3	90.4	83.4	87.9	92.6	87.8	89.9
84.3	90.4	91.6	91.0	93.0	93.7	88.3	91.8
90.1	91.2	90.7	88.2	94.4	96.5	89.2	89.7
89.0	90.6	88.6	88.5	90.4	84.3	92.3	92.2
89.8	92.2	88.3	93.3	91.2	93.2	88.9	
91.6	87.7	94.2	87.4	86.7	88.6	89.8	
90.3	91.1	85.3	91.1	94.2	88.7	92.7	
90.0	86.7	90.1	90.5	90.8	92.7	93.3	
91.5	93.4	89.3	100.3	90.1	89.3	86.7	
89.9	96.1	91.1	87.6	91.8	91.0	91.0	

Construct a stem-and-leaf display for these data.

6-15. The following data are the numbers of cycles to failure of aluminum test coupons subjected to repeated alternating stress at 21,000 psi, 18 cycles per second:

1115	865	1015	885	1594	1000	1416	1501
1310	2130	845	1223	2023	1820	1560	1238
1540	1421	1674	375	1315	1940	1055	990
1502	1109	1016	2265	1269	1120	1764	1468
1258	1481	1102	1910	1260	910	1330	1512
1315	1567	1605	1018	1888	1730	1608	1750
1085	1883	706	1452	1782	1102	1535	1642
798	1203	2215	1890	1522	1578	1781	
1020	1270	785	2100	1792	758	1750	

(a) Construct a stem-and-leaf display for these data.

(b) Does it appear likely that a coupon will "survive" beyond 2000 cycles? Justify your answer.

6-16. The percentage of cotton in material used to manufacture men's shirts follows. Construct a stem-and-leaf display for the data.

34.2	37.8	33.6	32.6	33.8	35.8	34.7	34.6
33.1	36.6	34.7	33.1	34.2	37.6	33.6	33.6
34.5	35.4	35.0	34.6	33.4	37.3	32.5	34.1
35.6	34.6	35.4	35.9	34.7	34.6	34.1	34.7
36.3	33.8	36.2	34.7	34.6	35.5	35.1	35.7
35.1	37.1	36.8	33.6	35.2	32.8	36.8	36.8
34.7	34.0	35.1	32.9	35.0	32.1	37.9	34.3
33.6	34.1	35.3	33.5	34.9	34.5	36.4	32.7

6-17. The following data represent the yield on 90 consecutive batches of ceramic substrate to which a metal coating has been applied by a vapor-deposition process. Construct a stem-and-leaf display for these data.

94.1	86.1	95.3	84.9	88.8	84.6	94.4	84.1
93.2	90.4	94.1	78.3	86.4	83.6	96.1	83.7
90.6	89.1	97.8	89.6	85.1	85.4	98.0	82.9
91.4	87.3	93.1	90.3	84.0	89.7	85.4	87.3
88.2	84.1	86.4	93.1	93.7	87.6	86.6	86.4
86.1	90.1	87.6	94.6	87.7	85.1	91.7	84.5
95.1	95.2	94.1	96.3	90.6	89.6	87.5	
90.0	86.1	92.1	94.7	89.4	90.0	84.2	
92.4	94.3	96.4	91.1	88.6	90.1	85.1	
87.3	93.2	88.2	92.4	84.1	94.3	90.5	
86.6	86.7	86.4	90.6	82.6	97.3	95.6	
91.2	83.0	85.0	89.1	83.1	96.8	88.3	

6-18. Find the median and the quartiles for the motor fuel octane data in Exercise 6-14.

6-19. Find the median and the quartiles for the failure data in Exercise 6-15.

6-20. Find the median, mode, and sample average of the data in Exercise 6-16. Explain how these three measures of location describe different features in the data.

6-21. Find the median and the quartiles for the yield data in Exercise 6-19.

6-22. The female students in an undergraduate engineering core course at ASU self-reported their heights to the nearest inch. The data are

62	64	66	67	65	68	61	65	67	65	64	63	67
68	64	66	68	69	65	67	62	66	68	67	66	65
69	65	70	65	67	68	65	63	64	67	67		

(a) Calculate the sample mean and standard deviation of height.
(b) Construct a stem-and-leaf diagram for the height data and comment on any important features that you notice.
(c) What is the median height of this group of female engineering students?

6-23. The shear strengths of 100 spot welds in a titanium alloy follow. Construct a stem-and-leaf diagram for the weld strength data and comment on any important features that you notice.

5408	5431	5475	5442	5376	5388	5459	5422	5416	5435
5420	5429	5401	5446	5487	5416	5382	5357	5388	5457
5407	5469	5416	5377	5454	5375	5409	5459	5445	5429
5463	5408	5481	5453	5422	5354	5421	5406	5444	5466
5399	5391	5477	5447	5329	5473	5423	5441	5412	5384
5445	5436	5454	5453	5428	5418	5465	5427	5421	5396
5381	5425	5388	5388	5378	5481	5387	5440	5482	5406
5401	5411	5399	5431	5440	5413	5406	5342	5452	5420
5458	5485	5431	5416	5431	5390	5399	5435	5387	5462
5383	5401	5407	5385	5440	5422	5448	5366	5430	5418

6-24. An important quality characteristic of water is the concentration of suspended solid material. Following are 60 measurements on suspended solids from a certain lake. Construct a stem-and-leaf diagram for this data and comment on any important features that you notice. Compute the sample mean, sample standard deviation, and the sample median.

42.4	65.7	29.8	58.7	52.1	55.8	57.0	68.7	67.3	67.3
54.3	54.0	73.1	81.3	59.9	56.9	62.2	69.9	66.9	59.0
56.3	43.3	57.4	45.3	80.1	49.7	42.8	42.4	59.6	65.8
61.4	64.0	64.2	72.6	72.5	46.1	53.1	56.1	67.2	70.7
42.6	77.4	54.7	57.1	77.3	39.3	76.4	59.3	51.1	73.8
61.4	73.1	77.3	48.5	89.8	50.7	52.0	59.6	66.1	31.6

6-25. The United States Golf Association tests golf balls to ensure that they conform to the rules of golf. Balls are tested for weight, diameter, roundness, and overall distance. The overall distance test is conducted by hitting balls with a driver swung by a mechanical device nicknamed "Iron Byron" after the legendary great Byron Nelson, whose swing the machine is said to emulate. Following are 100 distances (in yards) achieved by a particular brand of golf ball in the overall distance test. Construct a stem-and-leaf diagram for this data and comment on any important features that you notice. Compute the sample mean, sample standard deviation, and the sample median.

261.3	259.4	265.7	270.6	274.2	261.4	254.5	283.7
258.1	270.5	255.1	268.9	267.4	253.6	234.3	263.2
254.2	270.7	233.7	263.5	244.5	251.8	259.5	257.5
257.7	272.6	253.7	262.2	252.0	280.3	274.9	233.7
237.9	274.0	264.5	244.8	264.0	268.3	272.1	260.2
255.8	260.7	245.5	279.6	237.8	278.5	273.3	263.7
241.4	260.6	280.3	272.7	261.0	260.0	279.3	252.1
244.3	272.2	248.3	278.7	236.0	271.2	279.8	245.6
241.2	251.1	267.0	273.4	247.7	254.8	272.8	270.5
254.4	232.1	271.5	242.9	273.6	256.1	251.6	
256.8	273.0	240.8	276.6	264.5	264.5	226.8	
255.3	266.6	250.2	255.8	285.3	255.4	240.5	
255.0	273.2	251.4	276.1	277.8	266.8	268.5	

6-26. A semiconductor manufacturer produces devices used as central processing units in personal computers. The speed of the device (in megahertz) is important because it determines the price that the manufacturer can charge for the devices. The following table contains measurements on 120 devices. Construct a stem-and-leaf diagram for this data and comment on any important features that you notice. Compute the sample mean, sample standard deviation, and the sample median. What percentage of the devices has a speed exceeding 700 megahertz?

680	669	719	699	670	710	722	663	658	634	720	690
677	669	700	718	690	681	702	696	692	690	694	660
649	675	701	721	683	735	688	763	672	698	659	704
681	679	691	683	705	746	706	649	668	672	690	724
652	720	660	695	701	724	668	698	668	660	680	739
717	727	653	637	660	693	679	682	724	642	704	695
704	652	664	702	661	720	695	670	656	718	660	648
683	723	710	680	684	705	681	748	697	703	660	722
662	644	683	695	678	674	656	667	683	691	680	685
681	715	665	676	665	675	655	659	720	675	697	663

6-27. A group of wine enthusiasts taste-tested a pinot noir wine from Oregon. The evaluation was to grade the wine on a 0 to 100 point scale. The results follow:

94	90	92	91	91	86	89	91	91	90
90	93	87	90	91	92	89	86	89	90
88	95	91	88	89	92	87	89	95	92
85	91	85	89	88	84	85	90	90	83

(a) Construct a stem-and-leaf diagram for this data and comment on any important features that you notice.

(b) Compute the sample mean, sample standard deviation, and the sample median.
(c) A wine rated above 90 is considered truly exceptional. What proportion of the taste-tasters considered this particular pinot noir truly exceptional?

6-28. In their book *Introduction to Linear Regression Analysis* (3rd edition, Wiley, 2001) Montgomery, Peck, and Vining present measurements on $NbOCl_3$ concentration from a tube-flow reactor experiment. The data, in gram□mole per liter □ $10^{□3}$, are as follows:

450	450	473	507	457	452	453	1215	1256
1145	1085	1066	1111	1364	1254	1396	1575	1617
1733	2753	3186	3227	3469	1911	2588	2635	2725

(a) Construct a stem-and-leaf diagram for this data and comment on any important features that you notice.
(b) Compute the sample mean, sample standard deviation, and the sample median.

6-29. **A Comparative Stem-and-Leaf Diagram.** In Exercise 6-22, we presented height data that was self-reported by female undergraduate engineering students in a core course at ASU. In the same class, the male students self-reported their heights as follows:

69 67 69 70 65 68 69 70 71 69 66 67 69 75 68 67 68
69 70 71 72 68 69 69 70 71 68 72 69 69 68 69 73 70
73 68 69 71 67 68 65 68 68 69 70 74 71 69 70 69

(a) Construct a comparative stem-and-leaf diagram by listing the stems in the center of the display and then placing the female leaves on the left and the male leaves on the right.
(b) Comment on any important features that you notice in this display.

6-4 FREQUENCY DISTRIBUTIONS AND HISTOGRAMS

A **frequency distribution** is a more compact summary of data than a stem-and-leaf diagram. To construct a frequency distribution, we must divide the range of the data into intervals, which are usually called **class intervals, cells,** or **bins.** If possible, the bins should be of equal width in order to enhance the visual information in the frequency distribution. Some judgment must be used in selecting the number of bins so that a reasonable display can be developed. The number of bins depends on the number of observations and the amount of scatter or dispersion in the data. A frequency distribution that uses either too few or too many bins will not be informative. We usually find that between 5 and 20 bins is satisfactory in most cases and that the number of bins should increase with n. Choosing the **number of bins** approximately equal to the square root of the number of observations often works well in practice.

A frequency distribution for the comprehensive strength data in Table 6-2 is shown in Table 6-4. Since the data set contains 80 observations, and since $\sqrt{80} \approx 9$, we suspect that about eight to nine bins will provide a satisfactory frequency distribution. The largest and smallest data values are 245 and 76, respectively, so the bins must cover a range of at least 245 □ 76 □ 169 units on the psi scale. If we want the lower limit for the first bin to begin slightly below the smallest data value and the upper limit for the last bin to be slightly above the largest data value, we might start the frequency distribution at 70 and end it at 250. This is an interval or range of 180 psi units. Nine bins, each of width 20 psi, give a reasonable frequency distribution, so the frequency distribution in Table 6-4 is based on nine bins.

The second row of Table 6-4 contains a **relative frequency distribution.** The relative frequencies are found by dividing the observed frequency in each bin by the total number of

Table 6-4 Frequency Distribution for the Compressive Strength Data in Table 6-2

Class	70 □ x □ 90	90 □ x □ 110	110 □ x □ 130	130 □ x □ 150	150 □ x □ 170	170 □ x □ 190	190 □ x □ 210	210 □ x □ 230	230 □ x □ 250
Frequency	2	3	6	14	22	17	10	4	2
Relative frequency	0.0250	0.0375	0.0750	0.1750	0.2750	0.2125	0.1250	0.0500	0.0250
Cumulative relative frequency	0.0250	0.0625	0.1375	0.3125	0.5875	0.8000	0.9250	0.9750	1.0000

observations. The last row of Table 6-4 expresses the relative frequencies on a cumulative basis. Frequency distributions are often easier to interpret than tables of data. For example, from Table 6-4 it is very easy to see that most of the specimens have compressive strengths between 130 and 190 psi and that 97.5 percent of the specimens fail below 230 psi.

The **histogram** is a visual display of the frequency distribution. The stages for constructing a histogram follow.

Constructing a Histogram (Equal Bin Widths)

(1) Label the bin (class interval) boundaries on a horizontal scale.
(2) Mark and label the vertical scale with the frequencies or the relative frequencies.
(3) Above each bin, draw a rectangle where height is equal to the frequency (or relative frequency) corresponding to that bin.

Figure 6-7 is the histogram for the compression strength data. The histogram, like the **stem-and-leaf diagram,** provides a visual impression of the shape of the distribution of the measurements and information about the central tendency and scatter or dispersion in the data. Notice the symmetric, bell-shaped distribution of the strength measurements in Fig. 6-7. This display often gives insight about possible choices of probability distribution to use as a model for the population. For example, here we would likely conclude that the **normal distribution** is a reasonable model for the population of compression strength measurements.

Sometimes a histogram with **unequal bin widths** will be employed. For example, if the data have several extreme observations or outliers, using a few equal-width bins will result in nearly all observations falling in just of few of the bins. Using many equal-width bins will result in many bins with zero frequency. A better choice is to use shorter intervals in the region where most of the data falls and a few wide intervals near the extreme observations. When the bins are of unequal width, the rectangle's **area** (not its height) should be proportional to the bin frequency. This implies that the rectangle height should be

$$\text{Rectangle height} \propto \frac{\text{bin frequency}}{\text{bin width}}$$

In passing from either the original data or stem-and-leaf diagram to a frequency distribution or histogram, we have lost some information because we no longer have the individual observations. However, this information loss is often small compared with the conciseness and ease of interpretation gained in using the frequency distribution and histogram.

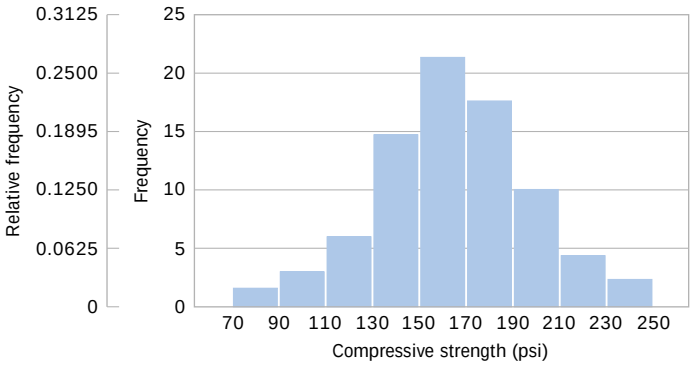


Figure 6-7 Histogram of compressive strength for 80 aluminum-lithium alloy specimens.

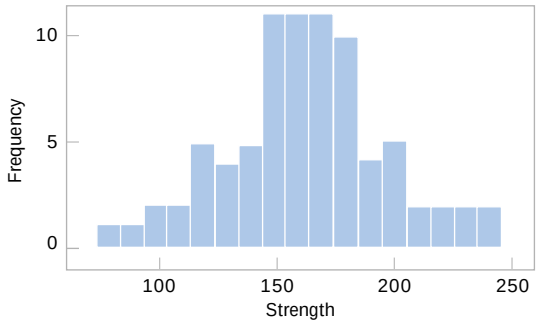


Figure 6-8 A histogram of the compressive strength data from Minitab with 17 bins.

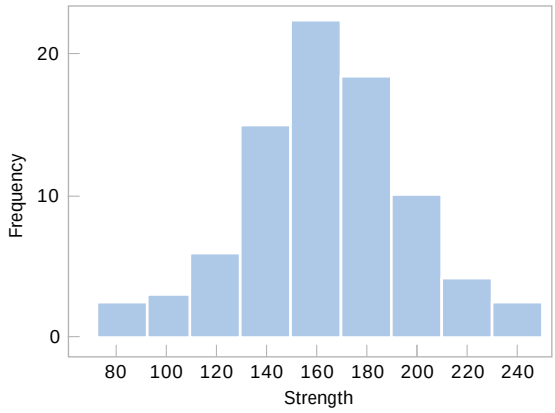


Figure 6-9 A histogram of the compressive strength data from Minitab with nine bins.

Figure 6-8 shows a histogram of the compressive strength data from Minitab. The "default" settings were used in this histogram, leading to 17 bins. We have noted that histograms may be relatively sensitive to the number of bins and their width. For small data sets, histograms may change dramatically in appearance if the number and/or width of the bins changes. Histograms are more stable for larger data sets, preferably of size 75 to 100 or more. Figure 6-9 shows the Minitab histogram for the compressive strength data with nine bins. This is similar to the original histogram shown in Fig. 6-7. Since the number of observations is moderately large (n □ 80), the choice of the number of bins is not especially important, and both Figs. 6-8 and 6-9 convey similar information.

Figure 6-10 shows a variation of the histogram available in Minitab, the **cumulative frequency plot.** In this plot, the height of each bar is the total number of observations that are less than or equal to the upper limit of the bin. Cumulative distributions are also useful in data interpretation; for example, we can read directly from Fig. 6-10 that there are approximately 70 observations less than or equal to 200 psi.

When the sample size is large, the histogram can provide a reasonably reliable indicator of the general **shape** of the distribution or population of measurements from which the sample was drawn. Figure 6-11 presents three cases. The median is denoted as $\tilde{x}$. Generally, if the data are symmetric, as in Fig. 6-11(*b*), the mean and median coincide. If, in addition, the data have only one mode (we say the data are *unimodal*), the mean, median, and mode all coincide. If the data are *skewed* (asymmetric, with a long tail to one side), as in Fig. 6-11(*a*) and (*c*), the mean, median, and mode do not coincide. Usually, we find that mode □ median □ mean if the

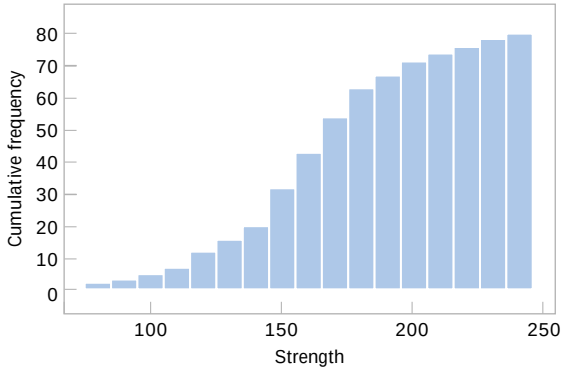


Figure 6-10 A cumulative distribution plot of the compressive strength data from Minitab.

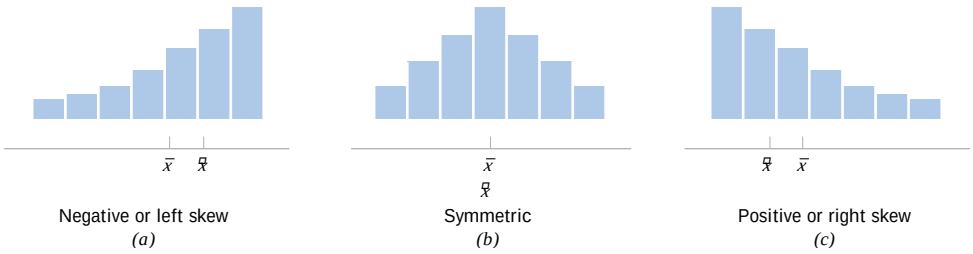


Figure 6-11
Histograms for symmetric and skewed distributions.

distribution is skewed to the right, whereas mode □ median □ mean if the distribution is skewed to the left.

Frequency distributions and histograms can also be used with qualitative or categorical data. In some applications there will be a natural ordering of the categories (such as freshman, sophomore, junior, and senior), whereas in others the order of the categories will be arbitrary (such as male and female). When using categorical data, the bins should have equal width.

EXAMPLE 6-6 Figure 6-12 presents the production of transport aircraft by the Boeing Company in 1985. Notice that the 737 was the most popular model, followed by the 757, 747, 767, and 707.

A chart of occurrences by category (in which the categories are ordered by the number of occurrences) is sometimes referred to as a **Pareto chart.** See Exercise 6-41.

In this section we have concentrated on descriptive methods for the situation in which each observation in a data set is a single number or belongs to one category. In many cases, we work with data in which each observation consists of several measurements. For example, in a gasoline mileage study, each observation might consist of a measurement of miles per gallon, the size of the engine in the vehicle, engine horsepower, vehicle weight, and vehicle length. This is an example of **multivariate data.** In later chapters, we will discuss analyzing this type of data.

EXERCISES FOR SECTION 6-4

6-30. Construct a frequency distribution and histogram for the motor fuel octane data from Exercise 6-14. Use eight bins.

6-31. Construct a frequency distribution and histogram using the failure data from Exercise 6-15.

6-32. Construct a frequency distribution and histogram for the cotton content data in Exercise 6-16.

6-33. Construct a frequency distribution and histogram for the yield data in Exercise 6-17.

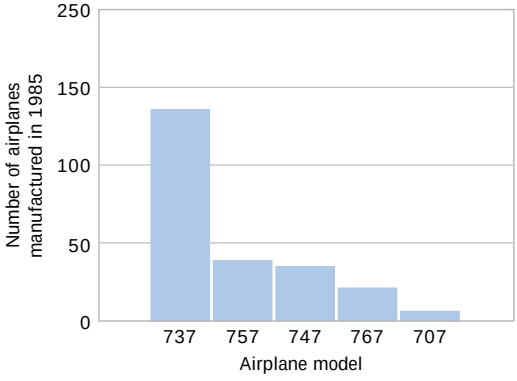


Figure 6-12
Airplane production in 1985. (*Source:* Boeing Company.)

6-34. Construct a frequency distribution and histogram with 16 bins for the motor fuel octane data in Exercise 6-14. Compare its shape with that of the histogram with eight bins from Exercise 6-30. Do both histograms display similar information?

6-35. Construct a histogram for the female student height data in Exercise 6-22.

6-36. Construct a histogram with 10 bins for the spot weld shear strength data in Exercise 6-23. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?

6-37. Construct a histogram for the water quality data in Exercise 6-24. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?

6-38. Construct a histogram with 10 bins for the overall distance data in Exercise 6-25. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?

6-39. Construct a histogram for the semiconductor speed data in Exercise 6-26. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?

6-40. Construct a histogram for the pinot noir wine rating data in Exercise 6-27. Comment on the shape of the histogram. Does it convey the same information as the stem-and-leaf display?

6-41. **The Pareto Chart.** An important variation of a histogram for categorical data is the Pareto chart. This chart is widely used in quality improvement efforts, and the categories usually represent different types of defects, failure modes, or product/process problems. The categories are ordered so that the category with the largest frequency is on the left, followed by the category with the second largest frequency and so forth. These charts are named after the Italian economist V. Pareto, and they usually exhibit "Pareto's law"; that is, most of the defects can be accounted for by only a few categories. Suppose that the following information on structural defects in automobile doors is obtained: dents, 4; pits, 4; parts assembled out of sequence, 6; parts undertrimmed, 21; missing holes/slots, 8; parts not lubricated, 5; parts out of contour, 30; and parts not deburred, 3. Construct and interpret a Pareto chart.

6-5 BOX PLOTS

The stem-and-leaf display and the histogram provide general visual impressions about a data set, while numerical quantities such as $\bar{x}$ or s provide information about only one feature of the data. The **box plot** is a graphical display that simultaneously describes several important features of a data set, such as center, spread, departure from symmetry, and identification of unusual observations or outliers.

A box plot displays the three quartiles, the minimum, and the maximum of the data on a rectangular box, aligned either horizontally or vertically. The box encloses the interquartile range with the left (or lower) edge at the first quartile, q_1, and the right (or upper) edge at the third quartile, q_3. A line is drawn through the box at the second quartile (which is the 50th percentile or the median), q_2 □ $\bar{x}$. A line, or **whisker,** extends from each end of the box. The lower whisker is a line from the first quartile to the smallest data point within 1.5 interquartile ranges from the first quartile. The upper whisker is a line from the third quartile to the largest data point within 1.5 interquartile ranges from the third quartile. Data farther from the box than the whiskers are plotted as individual points. A point beyond a whisker, but less than 3 interquartile ranges from the box edge, is called an **outlier.** A point more than 3 interquartile ranges from the box edge is called an **extreme outlier.** See Fig. 6-13. Occasionally, different symbols, such as open and filled circles, are used to identify the two types of outliers. Sometimes box plots are called *box-and-whisker plots*.

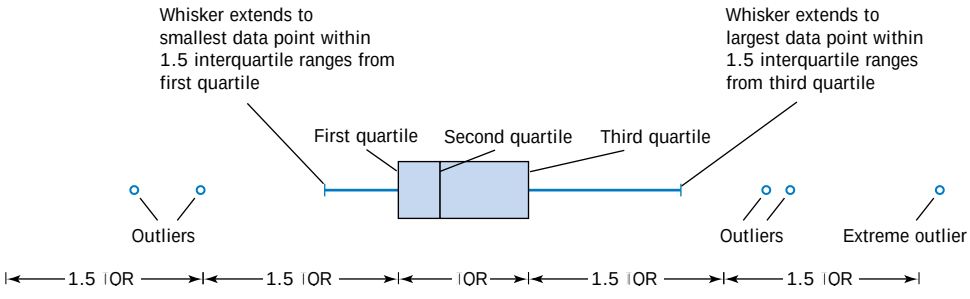


Figure 6-13 Description of a box plot.

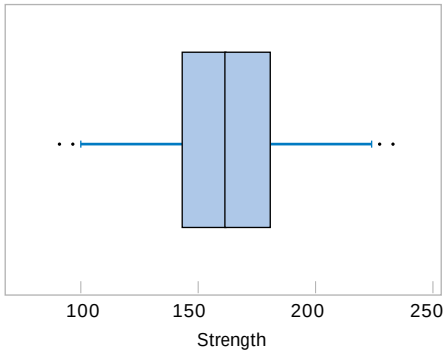


Figure 6-14 Box plot for compressive strength data in Table 6-2.

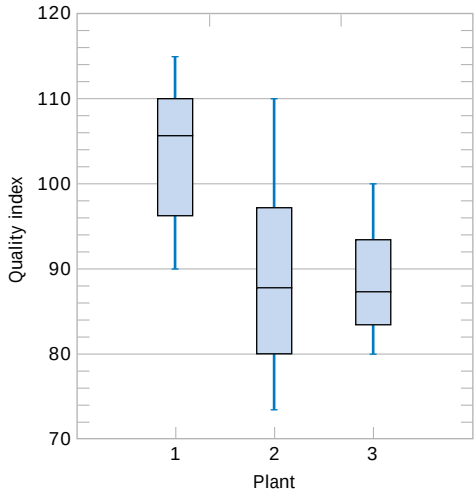


Figure 6-15 Comparative box plots of a quality index at three plants.

Figure 6-14 presents the box plot from Minitab for the alloy compressive strength data shown in Table 6-2. This box plot indicates that the distribution of compressive strengths is fairly symmetric around the central value, because the left and right whiskers and the lengths of the left and right boxes around the median are about the same. There are also two mild outliers on either end of the data.

Box plots are very useful in graphical comparisons among data sets, because they have high visual impact and are easy to understand. For example, Fig. 6-15 shows the comparative box plots for a manufacturing quality index on semiconductor devices at three manufacturing plants. Inspection of this display reveals that there is too much variability at plant 2 and that plants 2 and 3 need to raise their quality index performance.

EXERCISES FOR SECTION 6-5

6-42. Exercise 6-13 presented the joint temperatures of the O-rings (°F) for each test firing or actual launch of the space shuttle rocket motor. In that exercise you were asked to find the sample mean and sample standard deviation of temperature.

(a) Find the upper and lower quartiles of temperature.
(b) Find the median.
(c) Set aside the smallest observation (31°F) and recompute the quantities in parts (a) and (b). Comment on your findings. How "different" are the other temperatures from this smallest value?
(d) Construct a box plot of the data and comment on the possible presence of outliers.

6-43. An article in the *Transactions of the Institution of Chemical Engineers* (Vol. 34, 1956, pp. 280–293) reported data from an experiment investigating the effect of several process variables on the vapor phase oxidation of naphthalene. A sample of the percentage mole conversion of naphthalene to maleic anhydride follows: 4.2, 4.7, 4.7, 5.0, 3.8, 3.6, 3.0, 5.1, 3.1, 3.8, 4.8, 4.0, 5.2, 4.3, 2.8, 2.0, 2.8, 3.3, 4.8, 5.0.

(a) Calculate the sample mean.
(b) Calculate the sample variance and sample standard deviation.
(c) Construct a box plot of the data.

6-44. The "cold start ignition time" of an automobile engine is being investigated by a gasoline manufacturer. The following times (in seconds) were obtained for a test vehicle: 1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91.

(a) Calculate the sample mean and sample standard deviation.
(b) Construct a box plot of the data.

6-45. The nine measurements that follow are furnace temperatures recorded on successive batches in a semiconductor

manufacturing process (units are °F): 953, 950, 948, 955, 951, 949, 957, 954, 955.

(a) Calculate the sample mean, sample variance, and standard deviation.
(b) Find the median. How much could the largest temperature measurement increase without changing the median value?
(c) Construct a box plot of the data.

6-46. Exercise 6-12 presents drag coefficients for the NASA 0012 airfoil. You were asked to calculate the sample mean, sample variance, and sample standard deviation of those coefficients.

(a) Find the upper and lower quartiles of the drag coefficients.
(b) Construct a box plot of the data.
(c) Set aside the largest observation (100) and rework parts a and b. Comment on your findings.

6-47. The following data are the temperatures of effluent at discharge from a sewage treatment facility on consecutive days:

43	47	51	48	52	50	46	49
45	52	46	51	44	49	46	51
49	45	44	50	48	50	49	50

(a) Calculate the sample mean and median.
(b) Calculate the sample variance and sample standard deviation.
(c) Construct a box plot of the data and comment on the information in this display.

6-48. Reconsider the golf course yardage data in Exercise 6-3. Construct a box plot of the yardages and write an interpretation of the plot.

6-49. Reconsider the motor fuel octane rating data in Exercise 6-14. Construct a box plot of the yardages and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram in Exercise 6-14?

6-50. Reconsider the spot weld shear strength data in Exercise 6-23. Construct a box plot of the strengths and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram in Exercise 6-23?

6-51. Reconsider the female engineering student height data in Exercise 6-22. Construct a box plot of the heights and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram in Exercise 6-22?

6-52. Reconsider the water quality data in Exercise 6-24. Construct a box plot of the concentrations and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram in Exercise 6-24?

6-53. Reconsider the golf ball overall distance data in Exercise 6-25. Construct a box plot of the yardage distance and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram in Exercise 6-25?

6-54. Reconsider the wine rating data in Exercise 6-27. Construct a box plot of the wine ratings and write an interpretation of the plot. How does the box plot compare in interpretive value to the original stem-and-leaf diagram in Exercise 6-27?

6-55. Use the data on heights of female and male engineering students from Exercises 6-22 and 6-29 to construct comparative box plots. Write an interpretation of the information that you see in these plots.

6-56. In Exercise 6-44, data was presented on the cold start ignition time of a particular gasoline used in a test vehicle. A second formulation of the gasoline was tested in the same vehicle, with the following times (in seconds): 1.83, 1.99, 3.13, 3.29, 2.65, 2.87, 3.40, 2.46, 1.89, and 3.35. Use this new data along with the cold start times reported in Exercise 6-44 to construct comparative box plots. Write an interpretation of the information that you see in these plots.

6-6 TIME SEQUENCE PLOTS

The graphical displays that we have considered thus far such as histograms, stem-and-leaf plots, and box plots are very useful visual methods for showing the variability in data. However, we noted in Section 1-2.2 that time is an important factor that contributes to variability in data, and those graphical methods do not take this into account. A **time series** or **time sequence** is a data set in which the observations are recorded in the order in which they occur. A **time series plot** is a graph in which the vertical axis denotes the observed value of the variable (say x) and the horizontal axis denotes the time (which could be minutes, days, years, etc.) When measurements are plotted as a time series, we often see trends, cycles, or other broad features of the data that could not be seen otherwise.

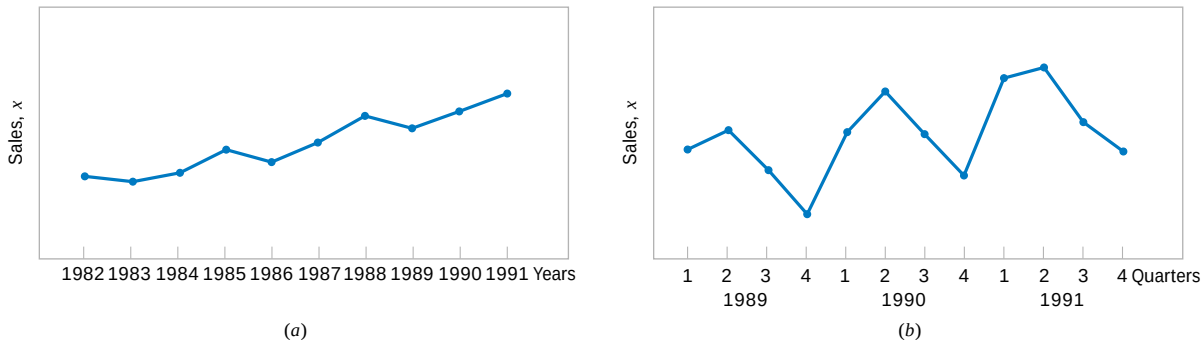


Figure 6-16 Company sales by year (*a*) and by quarter (*b*).

For example, consider Fig. 6-16(*a*), which presents a time series plot of the annual sales of a company for the last 10 years. The general impression from this display is that sales show an upward **trend.** There is some variability about this trend, with some years' sales increasing over those of the last year and some years' sales decreasing. Figure 6-16(*b*) shows the last three years of sales reported by quarter. This plot clearly shows that the annual sales in this business exhibit a **cyclic** variability by quarter, with the first- and second-quarter sales being generally greater than sales during the third and fourth quarters.

Sometimes it can be very helpful to combine a time series plot with some of the other graphical displays that we have considered previously. J. Stuart Hunter (*The American Statistician*, Vol. 42, 1988, p. 54) has suggested combining the stem-and-leaf plot with a time series plot to form a **digidot plot.**

Figure 6-17 shows a digidot plot for the observations on compressive strength from Table 6-2, assuming that these observations are recorded in the order in which they

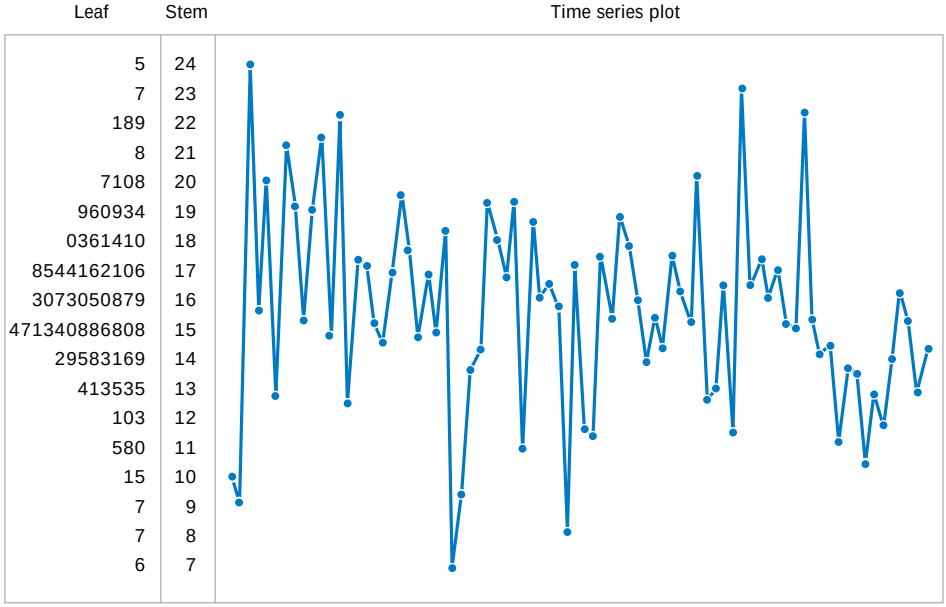


Figure 6-17 A digidot plot of the compressive strength data in Table 6-2.

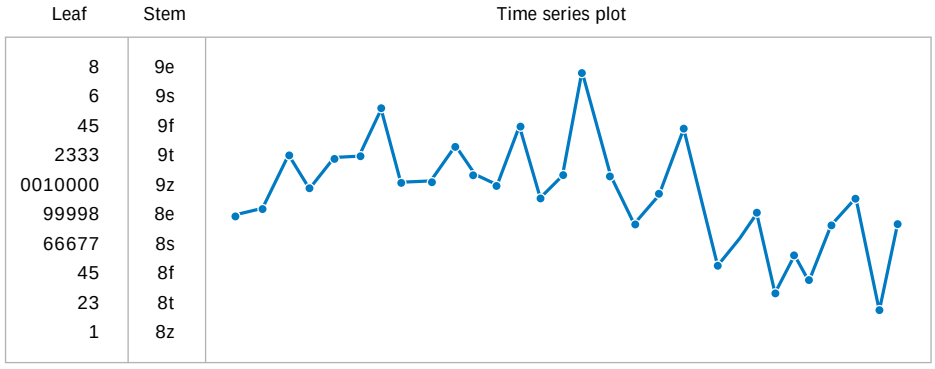


Figure 6-18 A digidot plot of chemical process concentration readings, observed hourly.

occurred. This plot effectively displays the overall variability in the compressive strength data and simultaneously shows the variability in these measurements over time. The general impression is that compressive strength varies around the mean value of 162.67, and there is no strong obvious pattern in this variability over time.

The digidot plot in Fig. 6-18 tells a different story. This plot summarizes 30 observations on concentration of the output product from a chemical process, where the observations are recorded at one-hour time intervals. This plot indicates that during the first 20 hours of operation this process produced concentrations generally above 85 grams per liter, but that following sample 20, something may have occurred in the process that results in lower concentrations. If this variability in output product concentration can be reduced, operation of this process can be improved.

EXERCISES FOR SECTION 6-6

6-57. The College of Engineering and Applied Science at Arizona State University had a VAX computer system. Response times for 20 consecutive jobs were recorded and are as follows: (read across)

5.3	10.1	5.9	12.2	11.2	12.4	9.2
5.0	5.8	7.2	8.5	7.3	3.9	10.5
9.5	6.2	10.0	4.7	6.4	8.1	

Construct and interpret a time series plot of these data.

6-58. The following data are the viscosity measurements for a chemical product observed hourly (read down, then left to right).

47.9	48.6	48.0	48.1	43.0	43.2
47.9	48.8	47.5	48.0	42.9	43.6
48.6	48.1	48.6	48.3	43.6	43.2
48.0	48.3	48.0	43.2	43.3	43.5
48.4	47.2	47.9	43.0	43.0	43.0
48.1	48.9	48.3	43.5	42.8	
48.0	48.6	48.5	43.1	43.1	

(a) Construct and interpret either a digidot plot or a separate stem-and-leaf and time series plot of these data.

(b) Specifications on product viscosity are at 48 □ 2. What conclusions can you make about process performance?

6-59. The pull-off force for a connector is measured in a laboratory test. Data for 40 test specimens follow (read down, then left to right).

241	203	201	251	236	190
258	195	195	238	245	175
237	249	255	210	209	178
210	220	245	198	212	175
194	194	235	199	185	190
225	245	220	183	187	
248	209	249	213	218	

(a) Construct a time series plot of the data.

(b) Construct and interpret either a digidot plot or a stem-and-leaf plot of the data.

6-60. In their book *Time Series Analysis, Forecasting, and Control* (Prentice Hall, 1994), G. E. P. Box, G. M. Jenkins, and G. C. Reinsel present chemical process concentration readings made every two hours. Some of these data follow (read down, then left to right).

Table 6-5 United Kingdom Passenger Airline Miles Flown

Month	1964	1965	1966	1967	1968	1969	1970
Jan.	7.269	8.350	8.186	8.334	8.639	9.491	10.840
Feb.	6.775	7.829	7.444	7.899	8.772	8.919	10.436
Mar.	7.819	8.829	8.484	9.994	10.894	11.607	13.589
Apr.	8.371	9.948	9.864	10.078	10.455	8.852	13.402
May	9.069	10.638	10.252	10.801	11.179	12.537	13.103
June	10.248	11.253	12.282	12.953	10.588	14.759	14.933
July	11.030	11.424	11.637	12.222	10.794	13.667	14.147
Aug.	10.882	11.391	11.577	12.246	12.770	13.731	14.057
Sept.	10.333	10.665	12.417	13.281	13.812	15.110	16.234
Oct.	9.109	9.396	9.637	10.366	10.857	12.185	12.389
Nov.	7.685	7.775	8.094	8.730	9.290	10.645	11.594
Dec.	7.682	7.933	9.280	9.614	10.925	12.161	12.772

17.0	16.7	17.1	17.5	17.6
16.6	17.4	17.4	18.1	17.5
16.3	17.2	17.4	17.5	16.5
16.1	17.4	17.5	17.4	17.8
17.1	17.4	17.4	17.4	17.3
16.9	17.0	17.6	17.1	17.3
16.8	17.3	17.4	17.6	17.1
17.4	17.2	17.3	17.7	17.4
17.1	17.4	17.0	17.4	16.9
17.0	16.8	17.8	17.8	17.3

6-61. Construct and interpret either a digidot plot or a stem-and-leaf plot of these data. The 100 annual Wolfer sunspot numbers from 1770 to 1869 follow. (For an interesting analysis and interpretation of these numbers, see the book by Box, Jenkins, and Reinsel referenced in Exercise 6-60. Their analysis requires some advanced knowledge of statistics and statistical model building.) (read down, then left to right)

(a) Construct a time series plot of these data.

(b) Construct and interpret either a digidot plot or a stem-and-leaf plot of these data.

101	31	154	38	83	90
82	7	125	23	132	67
66	20	85	10	131	60
35	92	68	24	118	47

41	10	16	8	62	94
21	8	7	13	98	96
16	2	4	57	124	77
6	0	2	122	96	59
4	1	8	138	66	44
7	5	17	103	64	47
14	12	36	86	54	30
34	14	50	63	39	16
45	35	62	37	21	7
43	46	67	24	7	37
48	41	71	11	4	74
42	30	48	15	23	
28	24	28	40	55	

6-62. In their book *Forecasting and Time Series Analysis*, 2nd edition (McGraw-Hill, 1990), D. C. Montgomery, L. A. Johnson, and J. S. Gardiner analyze the data in Table 6-5, which are the monthly total passenger airline miles flown in the United Kingdom, 1964–1970 (in millions of miles).

(a) Draw a time series plot of the data and comment on any features of the data that are apparent.

(b) Construct and interpret either a digidot plot or a stem-and-leaf plot of these data.

6-7 PROBABILITY PLOTS

How do we know if a particular probability distribution is a reasonable model for data? Sometimes, this is an important question because many of the statistical techniques presented in subsequent chapters are based on an assumption that the population distribution is of a specific type. Thus, we can think of determining whether data come from a specific

probability distribution as **verifying assumptions.** In other cases, the form of the distribution can give insight into the underlying physical mechanism generating the data. For example, in reliability engineering, verifying that time-to-failure data come from an exponential distribution identifies the **failure mechanism** in the sense that the failure rate is constant with respect to time.

Some of the visual displays we have used earlier, such as the histogram, can provide insight about the form of the underlying distribution. However, histograms are usually not really reliable indicators of the distribution form unless the sample size is very large. **Probability plotting** is a graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data. The general procedure is very simple and can be performed quickly. It is also more reliable than the histogram for small to moderate size samples. Probability plotting typically uses special graph paper, known as **probability paper,** that has been designed for the hypothesized distribution. Probability paper is widely available for the normal, lognormal, Weibull, and various chi-square and gamma distributions. We focus primarily on normal probability plots because many statistical techniques are appropriate only when the population is (at least approximately) normal.

To construct a probability plot, the observations in the sample are first ranked from smallest to largest. That is, the sample $x_1, x_2, \ldots, x_n$ is arranged as $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, where $x_{(1)}$ is the smallest observation, $x_{(2)}$ is the second smallest observation, and so forth, with $x_{(n)}$ the largest. The ordered observations $x_{(j)}$ are then plotted against their observed cumulative frequency $(j - 0.5)/n$ on the appropriate probability paper. If the hypothesized distribution adequately describes the data, the plotted points will fall approximately along a straight line; if the plotted points deviate significantly from a straight line, the hypothesized model is not appropriate. Usually, the determination of whether or not the data plot as a straight line is subjective. The procedure is illustrated in the following example.

EXAMPLE 6-7

Ten observations on the effective service life in minutes of batteries used in a portable personal computer are as follows: 176, 191, 214, 220, 205, 192, 201, 190, 183, 185. We hypothesize that battery life is adequately modeled by a normal distribution. To use probability plotting to investigate this hypothesis, first arrange the observations in ascending order and calculate their cumulative frequencies $(j - 0.5)/10$ as shown in Table 6-6.

The pairs of values $x_{(j)}$ and $(j - 0.5)/10$ are now plotted on normal probability paper. This plot is shown in Fig. 6-19. Most normal probability paper plots $100(j - 0.5)/n$ on the left vertical scale and $100[1 - (j - 0.5)/n]$ on the right vertical scale, with the variable value plotted on the horizontal scale. A straight line, chosen subjectively, has been drawn through the plotted points. In drawing the straight line, you should be influenced more by the points near the middle of the plot than by the extreme points. A good rule of thumb is to draw the line approximately between the 25th and 75th percentile points. This is how the line in Fig. 6-19 was determined. In assessing the "closeness" of the points to the straight line, imagine a "fat pencil" lying along the line. If all the points are covered by this imaginary pencil, a normal distribution adequately describes the data. Since the points in Fig. 6-19 would pass the "fat pencil" test, we conclude that the normal distribution is an appropriate model.

A **normal probability plot** can also be constructed on ordinary graph paper by plotting the standardized normal scores z_j against $x_{(j)}$, where the standardized normal scores satisfy

$$\frac{j - 0.5}{n} = P(Z \le z_j) = \Phi(z_j)$$

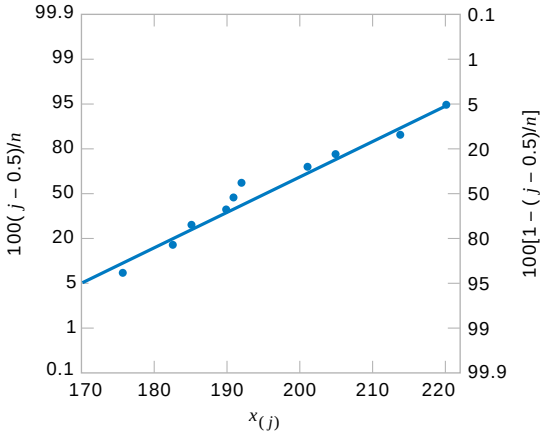


Figure 6-19 Normal probability plot for battery life.

Table 6-6 Calculation for Constracting a Normal Probability Plot

j	$x_{(j)}$	$(j - 0.5)/10$	z_j
1	176	0.05	−1.64
2	183	0.15	−1.04
3	185	0.25	−0.67
4	190	0.35	−0.39
5	191	0.45	−0.13
6	192	0.55	0.13
7	201	0.65	0.39
8	205	0.75	0.67
9	214	0.85	1.04
10	220	0.95	1.64

For example, if $(j - 0.5)/n = 0.05$, $(z_j) = 0.05$ implies that $z_j = -1.64$. To illustrate, consider the data from Example 6-4. In the last column of Table 6-6 we show the standarized normal scores. Figure 6-20 presents the plot of z_j versus $x_{(j)}$. This normal probability plot is equivalent to the one in Fig. 6-19.

We have constructed our probability plots with the probability scale (or the z-scale) on the vertical axis. Some computer packages "flip" the axis and put the probability scale on the horizontal axis.

The normal probability plot can be useful in identifying distributions that are symmetric but that have tails that are "heavier" or "lighter" than the normal. They can also be useful in identifying skewed distributions. When a sample is selected from a light-tailed distribution (such as the uniform distribution), the smallest and largest observations will not be as extreme as would be expected in a sample from a normal distribution. Thus if we consider the straight line drawn through the observations at the center of the normal probability plot, observations on the left side will tend to fall below the line, whereas observations on the right side will tend to fall above the line. This will produce an S-shaped normal probability plot such as shown in

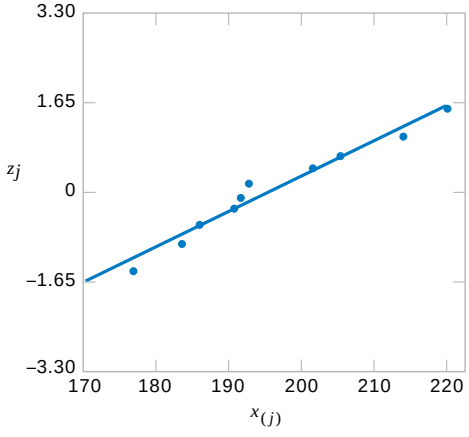


Figure 6-20 Normal probability plot obtained from standardized normal scores.

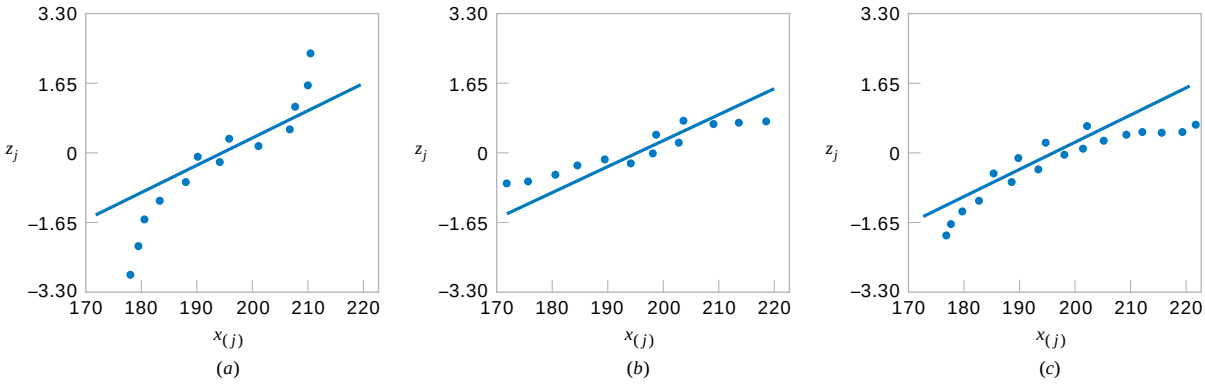


Figure 6-21 Normal probability plots indicating a nonnormal distrubution. (a) Light-tailed distribution. (b) Heavy-tailed distribution. (c) A distribution with positive (or right) skew.

Fig. 6-21(a). A heavy-tailed distribution will result in data that also produces an S-shaped normal probability plot, but now the observations on the left will be above the straight line and the observations on the right will lie below the line. See Fig. 6-19(b). A positively skewed distribution will tend to produce a pattern such as shown in Fig. 6-19(c), where points on both ends of the plot tend to fall below the line, giving a curved shape to the plot. This occurs because both the smallest and the largest observations from this type of distribution are larger than expected in a sample from a normal distribution.

Even when the underlying population is exactly normal, the sample data will not plot exactly on a straight line. Some judgment and experience are required to evaluate the plot. Generally, if the sample size is $n \square 30$, there can be a lot of deviation from linearity in normal plots, so in these cases only a very severe departure from linearity should be interpreted as a strong indication of nonnormality. As n increases, the linear pattern will tend to become stronger, and the normal probability plot will be easier to interpret and more reliable as an indicator of the form of the distribution.

EXERCISES FOR SECTION 6-7

6-63. Construct a normal probability plot of the piston ring diameter data in Exercise 6-1. Does it seem reasonable to assume that piston ring diameter is normally distributed?

6-64. Construct a normal probability plot of the insulating fluid breakdown time data in Exercise 6-2. Does it seem reasonable to assume that breakdown time is normally distributed?

6-65. Construct a normal probability plot of the visual accommodation data in Exercise 6-5. Does it seem reasonable to assume that visual accommodation is normally distributed?

6-66. Construct a normal probability plot of the O-ring joint temperature data in Exercise 6-13. Does it seem reasonable to assume that O-ring joint temperature is normally distributed? Discuss any interesting features that you see on the plot.

6-67. Construct a normal probability plot of the octane rating data in Exercise 6-14. Does it seem reasonable to assume that octane rating is normally distributed?

6-68. Construct a normal probability plot of the cycles to failure data in Exercise 6-15. Does it seem reasonable to assume that cycles to failure is normally distributed?

6-69. Construct a normal probability plot of the wine quality rating data in Exercise 6-27. Does it seem reasonable to assume that this variable is normally distributed?

6-70. Construct a normal probability plot of the suspended solids concentration data in Exercise 6-24. Does it seem reasonable to assume that the concentration of suspended solids in water from this particular lake is normally distributed?

6-71. Construct two normal probability plots for the height data in Exercises 6-22 and 6-29. Plot the data for female and male students on the same axes. Does height seem to be normally distributed for either group of students? If both populations have the same variance, the two normal probability plots should have identical slopes. What conclusions would you draw about the heights of the two groups of students from visual examination of the normal probability plots?

6-72. It is possible to obtain a "quick and dirty" estimate of the mean of a normal distribution from the fiftieth percentile value on a normal probability plot. Provide an argument why this is so. It is also possible to obtain an estimate of the standard deviation of a normal distribution by subtracting the sixty-fourth percentile value from the fiftieth percentile value. Provide an argument why this is so.

6-8 MORE ABOUT PROBABILITY PLOTTING (CD ONLY)

Supplemental Exercises

6-73. The concentration of a solution is measured six times by one operator using the same instrument. She obtains the following data: 63.2, 67.1, 65.8, 64.0, 65.1, and 65.3 (grams per liter).

(a) Calculate the sample mean. Suppose that the desirable value for this solution has been specified to be 65.0 grams per liter. Do you think that the sample mean value computed here is close enough to the target value to accept the solution as conforming to target? Explain your reasoning.

(b) Calculate the sample variance and sample standard deviation.

(c) Suppose that in measuring the concentration, the operator must set up an apparatus and use a reagent material. What do you think the major sources of variability are in this experiment? Why is it desirable to have a small variance of these measurements?

6-74. A sample of six resistors yielded the following resistances (ohms): $x_1 = 45$, $x_2 = 38$, $x_3 = 47$, $x_4 = 41$, $x_5 = 35$, and $x_6 = 43$.

(a) Compute the sample variance and sample standard deviation.

(b) Subtract 35 from each of the original resistance measurements and compute s^2 and s. Compare your results with those obtained in part (a) and explain your findings.

(c) If the resistances were 450, 380, 470, 410, 350, and 430 ohms, could you use the results of previous parts of this problem to find s^2 and s?

6-75. Consider the following two samples:

Sample 1: 10, 9, 8, 7, 8, 6, 10, 6

Sample 2: 10, 6, 10, 6, 8, 10, 8, 6

(a) Calculate the sample range for both samples. Would you conclude that both samples exhibit the same variability? Explain.

(b) Calculate the sample standard deviations for both samples. Do these quantities indicate that both samples have the same variability? Explain.

(c) Write a short statement contrasting the sample range versus the sample standard deviation as a measure of variability.

6-76. An article in *Quality Engineering* (Vol. 4, 1992, pp. 487–495) presents viscosity data from a batch chemical process. A sample of these data follows:

13.3	14.3	14.9	15.2	15.8	14.2	16.0	14.0
14.5	16.1	13.7	15.2	13.7	16.9	14.9	14.4
15.3	13.1	15.2	15.9	15.1	14.9	13.6	13.7
15.3	15.5	14.5	16.5	13.4	15.2	15.3	13.8
14.3	12.6	15.3	14.8	14.1	14.4	14.3	15.6
14.8	14.6	15.6	15.1	14.8	15.2	15.6	14.5
15.2	14.3	15.8	17.0	14.3	14.6	16.1	12.8
14.5	15.4	13.3	14.9	14.3	16.4	13.9	16.1
14.6	15.2	14.1	14.8	16.4	14.2	15.2	16.6
14.1	16.8	15.4	14.0	16.9	15.7	14.4	15.6

(a) Reading down and left to right, draw a time series plot of all the data and comment on any features of the data that are revealed by this plot.

(b) Consider the notion that the first 40 observations were generated from a specific process, whereas the last 40 observations were generated from a different process. Does the plot indicate that the two processes generate similar results?

(c) Compute the sample mean and sample variance of the first 40 observations; then compute these values for the second 40 observations. Do these quantities indicate that both processes yield the same mean level? The same variability? Explain.

6-77. Reconsider the data from Exercise 6-76. Prepare comparative box plots for two groups of observations: the first 40 and the last 40. Comment on the information in the box plots.

6-78. The data shown in Table 6-7 are monthly champagne sales in France (1962-1969) in thousands of bottles.

(a) Construct a time series plot of the data and comment on any features of the data that are revealed by this plot.

(b) Speculate on how you would use a graphical procedure to forecast monthly champagne sales for the year 1970.

6-79. A manufacturer of coil springs is interested in implementing a quality control system to monitor his production process. As part of this quality system, it is decided to record the number of nonconforming coil springs in each production

Table 6-7 Champagne Sales in France

Month	1962	1963	1964	1965	1966	1967	1968	1969
Jan.	2.851	2.541	3.113	5.375	3.633	4.016	2.639	3.934
Feb.	2.672	2.475	3.006	3.088	4.292	3.957	2.899	3.162
Mar.	2.755	3.031	4.047	3.718	4.154	4.510	3.370	4.286
Apr.	2.721	3.266	3.523	4.514	4.121	4.276	3.740	4.676
May	2.946	3.776	3.937	4.520	4.647	4.968	2.927	5.010
June	3.036	3.230	3.986	4.539	4.753	4.677	3.986	4.874
July	2.282	3.028	3.260	3.663	3.965	3.523	4.217	4.633
Aug.	2.212	1.759	1.573	1.643	1.723	1.821	1.738	1.659
Sept.	2.922	3.595	3.528	4.739	5.048	5.222	5.221	5.591
Oct.	4.301	4.474	5.211	5.428	6.922	6.873	6.424	6.981
Nov.	5.764	6.838	7.614	8.314	9.858	10.803	9.842	9.851
Dec.	7.132	8.357	9.254	10.651	11.331	13.916	13.076	12.670

batch of size 50. During 40 days of production, 40 batches of data were collected as follows:

Read data across.

9	12	6	9	7	14	12	4	6	7
8	5	9	7	8	11	3	6	7	7
11	4	4	8	7	5	6	4	5	8
19	19	18	12	11	17	15	17	13	13

(a) Construct a stem-and-leaf plot of the data.
(b) Find the sample average and standard deviation.
(c) Construct a time series plot of the data. Is there evidence that there was an increase or decrease in the average number of nonconforming springs made during the 40 days? Explain.

6-80. A communication channel is being monitored by recording the number of errors in a string of 1000 bits. Data for 20 of these strings follow:

Read data across.

3	1	0	1	3	2	4	1	3	1
1	1	2	3	3	2	0	2	0	1

(a) Construct a stem-and-leaf plot of the data.
(b) Find the sample average and standard deviation.
(c) Construct a time series plot of the data. Is there evidence that there was an increase or decrease in the number of errors in a string? Explain.

6-81. Reconsider the data in Exercise 6-76. Construct normal probability plots for two groups of the data: the first 40 and the last 40 observations. Construct both plots on the same axes. What tentative conclusions can you draw?

6-82. Construct a normal probability plot of the effluent discharge temperature data from Exercise 6-47. Based on the plot, what tentative conclusions can you draw?

6-83. Construct normal probability plots of the cold start ignition time data presented in Exercises 6-44 and 6-56. Construct a separate plot for each gasoline formulation, but arrange the plots on the same axes. What tentative conclusions can you draw?

6-84. **Transformations.** In some data sets, a transformation by some mathematical function applied to the original data, such as $1\bar{y}$ or log y, can result in data that are simpler to work with statistically than the original data. To illustrate the effect of a transformation, consider the following data, which represent cycles to failure for a yarn product: 675, 3650, 175, 1150, 290, 2000, 100, 375.

(a) Construct a normal probability plot and comment on the shape of the data distribution.
(b) Transform the data using logarithms; that is, let y^* (new value) = log y (old value). Construct a normal probability plot of the transformed data and comment on the effect of the transformation.

6-85. In 1879, A. A. Michelson made 100 determinations of the velocity of light in air using a modification of a method proposed by the French physicist Foucault. He made the measurements in five trials of 20 measurements each. The observations (in kilometers per second) follow. Each value has 299,000 substracted from it.

Trial 1

850	900	930	950	980
1000	930	760	1000	960
740	1070	850	980	880
980	650	810	1000	960

Trial 2

960	960	880	850	900
830	810	880	800	760
940	940	800	880	840
790	880	830	790	800

Trial 3

880	880	720	620	970
880	850	840	850	840
880	860	720	860	950
910	870	840	840	840

Trial 4

890	810	800	760	750
910	890	880	840	850
810	820	770	740	760
920	860	720	850	780

Trial 5

890	780	760	790	820
870	810	810	950	810
840	810	810	810	850
870	740	940	800	870

The currently accepted true velocity of light in a vacuum is 299, 792.5 kilometers per second. Stigler (1977, *The Annals of Statistics*) reports that the "true" value for comparison to these measurements is 734.5. Construct comparative box plots of these measurements. Does it seem that all five trials are consistent with respect to the variability of the measurements? Are all five trials centered on the same value? How does each group of trials compare to the true value? Could there have been "startup" effects in the experiment that Michelson performed? Could there have been bias in the measuring instrument?

6-86. In 1789, Henry Cavendish estimated the density of the earth by using a torsion balance. His 29 measurements follow, expressed as a multiple of the density of water.

5.50	5.30	5.47	5.10	5.29	5.65
5.55	5.61	5.75	5.63	5.27	5.44
5.57	5.36	4.88	5.86	5.34	5.39
5.34	5.53	5.29	4.07	5.85	5.46
5.42	5.79	5.62	5.58	5.26	



(a) Calculate the sample mean, sample standard deviation, and median of the Cavendish density data.

(b) Construct a normal probability plot of the data. Comment on the plot. Does there seem to be a "low" outlier in the data?

(c) Would the sample median be a better estimate of the density of the earth than the sample mean? Why?

MIND-EXPANDING EXERCISES

6-87. Consider the airfoil data in Exercise 6-12. Subtract 30 from each value and then multiply the resulting quantities by 10. Now compute s^2 for the new data. How is this quantity related to s^2 for the *original* data? Explain why.

6-88. Consider the quantity $\sum_{i=1}^{n}(x_i - a)^2$. For what value of a is this quantity minimized?

6-89 Using the results of Exercise 6-87, which of the two quantities $\sum_{i=1}^{n}(x_i - \bar{x})^2$ and $\sum_{i=1}^{n}(x_i - 0)^2$ will be smaller, provided that $\bar{x} \neq 0$?

6-90. **Coding the Data.** Let $y_i = a + bx_i$, $i = 1, 2, \ldots, n$, where a and b are nonzero constants. Find the relationship between $\bar{x}$ and $\bar{y}$, and between s_x and s_y.

6-91. A sample of temperature measurements in a furnace yielded a sample average (°F) of 835.00 and a sample standard deviation of 10.5. Using the results from Exercise 6-90, what are the sample average and sample standard deviations expressed in °C?

6-92. Consider the sample $x_1, x_2, \ldots, x_n$ with sample mean $\bar{x}$ and sample standard deviation s. Let $z_i = (x_i - \bar{x})/s$, $i = 1, 2, \ldots, n$. What are the values of the sample mean and sample standard deviation of the z_i?

6-93. An experiment to investigate the survival time in hours of an electronic component consists of placing the parts in a test cell and running them for 100 hours under elevated temperature conditions. (This is called an "accelerated" life test.) Eight components were tested with the following resulting failure times:

75, 63, 100+, 36, 51, 45, 80, 90

The observation 100+ indicates that the unit still functioned at 100 hours. Is there any meaningful measure of location that can be calculated for these data? What is its numerical value?

6-94. Suppose that we have a sample $x_1, x_2, \ldots, x_n$ and we have calculated $\bar{x}_n$ and s_n^2 for the sample. Now an $(n+1)$st observation becomes available. Let $\bar{x}_{n+1}$ and s_{n+1}^2 be the sample mean and sample variance for the sample using all $n+1$ observations.

(a) Show how $\bar{x}_{n+1}$ can be computed using $\bar{x}_n$ and x_{n+1}.

(b) Show that $ns_{n+1}^2 + (n+1)s_n^2 = \dfrac{n(x_{n+1} - \bar{x}_n)^2}{n+1}$

(c) Use the results of parts (a) and (b) to calculate the new sample average and standard deviation for the data of Exercise 6-22, when the new observation is $x_{38} = 64$.

6-95. **The Trimmed Mean.** Suppose that the data are arranged in increasing order, T% of the observations are removed from each end and the sample mean of the remaining numbers is calculated. The resulting quantity is called a *trimmed mean*. The trimmed mean generally lies between the sample mean $\bar{x}$ and the sample median $\tilde{x}$. Why?

(a) Calculate the 10% trimmed mean for the yield data in Exercise 6-17.

(b) Calculate the 20% trimmed mean for the yield data in Exercise 6-17 and compare it with the quantity found in part (a).

(c) Compare the values calculated in parts (a) and (b) with the sample mean and median for the yield data. Is there much difference in these quantities? Why?

6-96. **The Trimmed Mean.** Suppose that the sample size n is such that the quantity $nT/100$ is not an integer. Develop a procedure for obtaining a trimmed mean in this case.

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Box plot
Frequency distribution and histogram
Median, quartiles and percentiles
Normal probability plot
Population mean
Population standard deviation
Population variance
Random sample
Sample mean
Sample standard deviation
Sample variance
Stem-and-leaf diagram
Time series plots

CD MATERIAL

Exponential probability plot
Goodness of fit
Weibull probability plot

6-8 MORE ABOUT PROBABILITY PLOTTING (CD ONLY)

Probability plots are extremely useful and are often the first technique used in an effort to determine which probability distribution is likely to provide a reasonable model for the data.

We give a simple illustration of how a normal probability plot can be useful in distinguishing between normal and nonnormal data. Table S6-1 contains 50 observations generated at random from an exponential distribution with mean 20 (or □ 0.05). These data were generated using the random number generation capability in Minitab. Figure S6-1 presents a normal probability plot of these data, constructed using Minitab. The observations do not even approximately lie along a straight line, giving a clear indication that the data do not follow a normal distribution. The strong curvature at both ends of the plot suggests that the data come from a distribution with right or positive skew. Compare Fig. S6-1 with Fig. 6-19c.

Minitab also provides estimates of the mean and standard deviation of the distribution using the **method of maximum likelihood** (abbreviated ML on the graph in Figure S6-1). We will discuss maximum likelihood estimation in Chapter 7. For the normal distribution, this is the familiar sample mean and sample standard deviation that we first presented in Chapter 1. Minitab also presents a quantitative measure of how well the data are described by a normal distribution. This goodness-of-fit measure is called the Anderson-Darling statistic (abbreviated AD on the Minitab probability plot). The Anderson-Darling statistic is based on the probability integral transformation

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

that can be used to convert the data to a uniform distribution if the hypothesized distribution is correct. Thus, if $x_1, x_2, \ldots, x_n$ are independent and identically distributed random variables whose cumulative distribution function is $F(x)$, then $F(x_1), F(x_2), \ldots, F(x_n)$ are independent uniform (0, 1) random variables. The Anderson-Darling statistic essentially compares how close the $F(x_1), F(x_2), \ldots, F(x_n)$ values are to values from a uniform (0, 1) distribution. For

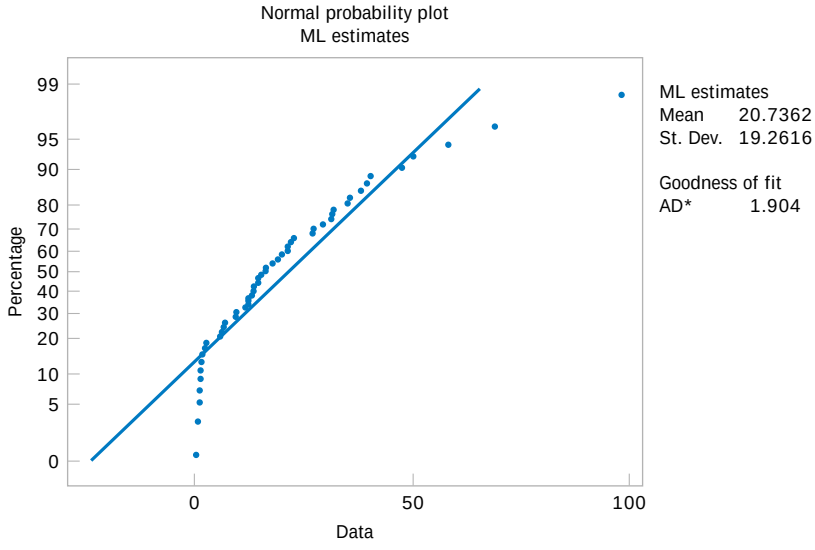


Figure S6-1. Normal probability plot (from Minitab) of the data from Table S6-1.

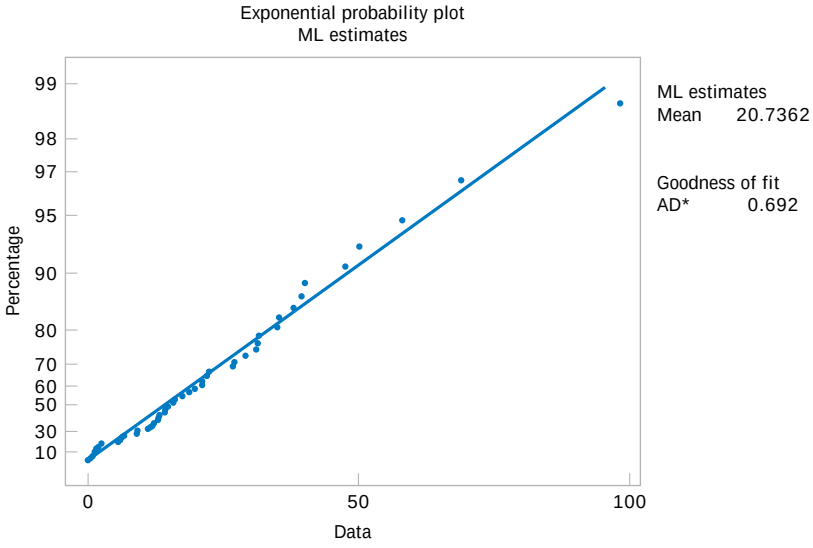


Figure S6-2. Exponential probability plot (from Minitab) of the data from Table S6-1.

this reason, the Anderson-Darling test is sometimes called a "distance" test. The test is upper-tailed; that is, if the computed value exceeds a critical value, the hypothesis of normality is rejected. The 5% critical value of the Anderson-Darling statistic is 0.752 and the 1% value is 1.035. Because the Anderson-Darling statistic in Figure S6-1 is 1.904, and this exceeds the 1% critical value, we conclude that the assumption of normality would be inappropriate.

Minitab can construct several other types of probability plots. An exponential probability plot of the data in Table S6-1 is shown in Figure S6-2. Notice that the data lies very close to the straight line in this plot, implying that the exponential is a good model for the data. Minitab also provides an estimate of the mean of the exponential distribution. This estimate is just the sample mean.

Figure S6-3 is a Weibull probability plot of the data from Table S6-1, constructed using Minitab. The data lies approximately along a straight line, suggesting that the Weibull distribution is also a reasonable model for the data. Notice that Minitab provides maximum

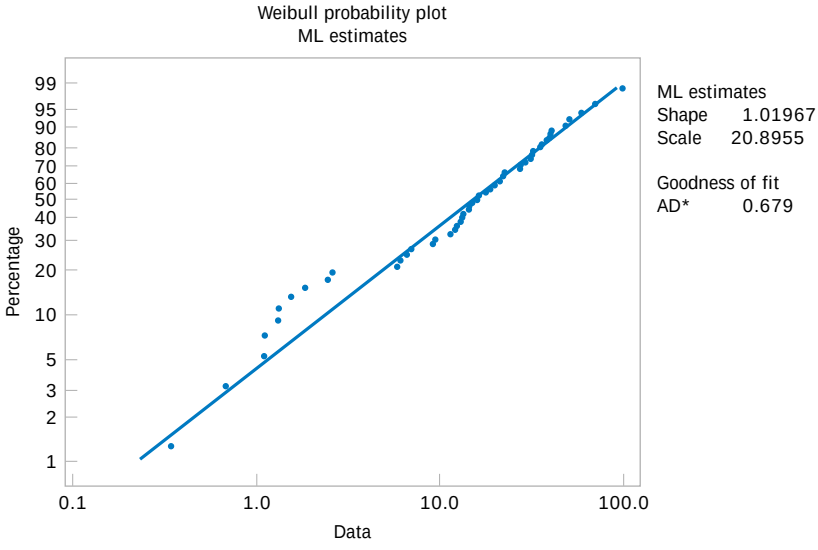


Figure S6-3. Weibull probability plot (from Minitab) of the data from Table S6-1.

Table S6-1 50 Observations Drawn at Random from an Exponential Distribution with Mean 20 1 □ 0.052

1.2934	14.1968	13.2798	16.1154	14.8891	31.6489
2.4330	13.1818	5.7511	39.9558	97.8874	21.1057
15.8770	9.0942	31.3655	1.3104	12.0008	11.2846
21.9606	21.1336	26.8364	9.3134	31.0346	29.0222
47.4481	0.3389	1.0999	19.8350	1.5191	2.5623
34.8720	39.2494	12.1621	18.8295	35.3307	
27.0908	0.6731	14.2467	49.9397	6.0479	
17.5253	68.7876	57.8919	1.0882	22.4244	
6.8290	37.9023	1.8219	6.4967	12.8239	

likelihood estimates of the shape parameter □ and the scale parameter □. The shape parameter estimate is very close to unity, and we know that a Weibull distribution with □ □ 1 is the exponential distribution.

7 Point Estimation of Parameters

CHAPTER OUTLINE

7-1 INTRODUCTION

7-2 GENERAL CONCEPTS OF POINT ESTIMATION
- 7-2.1 Unbiased Estimators
- 7-2.2 Proof that S is a Biased Estimator of □ (CD Only)
- 7-2.3 Variance of a Point Estimator
- 7-2.4 Standard Error: Reporting a Point Estimate
- 7-2.5 Bootstrap Estimate of the Standard Error (CD Only)
- 7-2.6 Mean Square Error of an Estimator

7-3 METHODS OF POINT ESTIMATION
- 7-3.1 Method of Moments
- 7-3.2 Method of Maximum Likelihood
- 7-3.3 Bayesian Estimation of Parameters (CD Only)

7-4 SAMPLING DISTRIBUTIONS

7-5 SAMPLING DISTRIBUTIONS OF MEANS

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Explain the general concepts of estimating the parameters of a population or a probability distribution
2. Explain important properties of point estimators, including bias, variance, and mean square error
3. Know how to construct point estimators using the method of moments and the method of maximum likelihood
4. Know how to compute and explain the precision with which a parameter is estimated
5. Understand the central limit theorem
6. Explain the important role of the normal distribution as a sampling distribution

CD MATERIAL

7. Use bootstrapping to find the standard error of a point estimate
8. Know how to construct a point estimator using the Bayesian approach

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

7-1 INTRODUCTION

The field of statistical inference consists of those methods used to make decisions or to draw conclusions about a **population.** These methods utilize the information contained in a **sample** from the population in drawing conclusions. This chapter begins our study of the statistical methods used for inference and decision making.

Statistical inference may be divided into two major areas: **parameter estimation** and **hypothesis testing.** As an example of a parameter estimation problem, suppose that a structural engineer is analyzing the tensile strength of a component used in an automobile chassis. Since variability in tensile strength is naturally present between the individual components because of differences in raw material batches, manufacturing processes, and measurement procedures (for example), the engineer is interested in estimating the mean tensile strength of the components. In practice, the engineer will use sample data to compute a number that is in some sense a reasonable value (or guess) of the true mean. This number is called a **point estimate.** We will see that it is possible to establish the precision of the estimate.

Now consider a situation in which two different reaction temperatures can be used in a chemical process, say t_1 and t_2. The engineer conjectures that t_1 results in higher yields than does t_2. Statistical hypothesis testing is a framework for solving problems of this type. In this case, the hypothesis would be that the mean yield using temperature t_1 is greater than the mean yield using temperature t_2. Notice that there is no emphasis on estimating yields; instead, the focus is on drawing conclusions about a stated hypothesis.

Suppose that we want to obtain a point estimate of a population parameter. We know that before the data is collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$. Therefore, any function of the observation, or any **statistic,** is also a random variable. For example, the sample mean $\overline{X}$ and the sample variance S^2 are statistics and they are also random variables.

Since a statistic is a random variable, it has a probability distribution. We call the probability distribution of a statistic a **sampling distribution.** The notion of a sampling distribution is very important and will be discussed and illustrated later in the chapter.

When discussing inference problems, it is convenient to have a general symbol to represent the parameter of interest. We will use the Greek symbol θ (theta) to represent the parameter. The objective of point estimation is to select a single number, based on sample data, that is the most plausible value for θ. A numerical value of a sample statistic will be used as the point estimate.

In general, if X is a random variable with probability distribution $f(x)$, characterized by the unknown parameter θ, and if $X_1, X_2, \ldots, X_n$ is a random sample of size n from X, the statistic $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ is called a **point estimator** of θ. Note that $\hat{\Theta}$ is a random variable because it is a function of random variables. After the sample has been selected, $\hat{\Theta}$ takes on a particular numerical value $\hat{\theta}$ called the **point estimate** of θ.

Definition

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

As an example, suppose that the random variable X is normally distributed with an unknown mean μ. The sample mean is a point estimator of the unknown population mean μ. That is, $\hat{\mu} = \bar{X}$. After the sample has been selected, the numerical value $\bar{x}$ is the point estimate of μ. Thus, if $x_1 = 25$, $x_2 = 30$, $x_3 = 29$, and $x_4 = 31$, the point estimate of μ is

$$\bar{x} = \frac{25 + 30 + 29 + 31}{4} = 28.75$$

Similarly, if the population variance σ^2 is also unknown, a point estimator for σ^2 is the sample variance S^2, and the numerical value $s^2 = 6.9$ calculated from the sample data is called the point estimate of σ^2.

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \bar{x}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

We may have several different choices for the point estimator of a parameter. For example, if we wish to estimate the mean of a population, we might consider the sample mean, the sample median, or perhaps the average of the smallest and largest observations in the sample as point estimators. In order to decide which point estimator of a particular parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators.

7-2 GENERAL CONCEPTS OF POINT ESTIMATION

7-2.1 Unbiased Estimators

An estimator should be "close" in some sense to the true value of the unknown parameter. Formally, we say that $\hat{\Theta}$ is an unbiased estimator of θ if the expected value of $\hat{\Theta}$ is equal to θ. This is equivalent to saying that the mean of the probability distribution of $\hat{\Theta}$ (or the mean of the sampling distribution of $\hat{\Theta}$) is equal to θ.

Definition

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \tag{7-1}$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \tag{7-2}$$

is called the **bias** of the estimator $\hat{\Theta}$.

When an estimator is unbiased, the bias is zero; that is, $E(\hat{\Theta}) - \theta = 0$.

EXAMPLE 7-1

Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\bar{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

First consider the sample mean. In Equation 5.40a in Chapter 5, we showed that $E(\bar{X}) = \mu$. Therefore, the sample mean $\bar{X}$ is an unbiased estimator of the population mean μ.

Now consider the sample variance. We have

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i^2 - \bar{X}^2 - 2\bar{X}X_i)\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2) - nE(\bar{X}^2)\right]$$

The last equality follows from Equation 5-37 in Chapter 5. However, since $E(X_i^2) = \sigma^2 + \mu^2$ and $E(\bar{X}^2) = \sigma^2 + \mu^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n(\sigma^2 + \mu^2/n)\right]$$

$$= \frac{1}{n-1}(n\sigma^2 + n\mu^2 - n\sigma^2 - \mu^2)$$

$$= \sigma^2$$

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

Although S^2 is unbiased for σ^2, S is a biased estimator of σ. For large samples, the bias is very small. However, there are good reasons for using S as an estimator of σ in samples from normal distributions, as we will see in the next three chapters when are discuss confidence intervals and hypothesis testing.

Sometimes there are several unbiased estimators of the sample population parameter. For example, suppose we take a random sample of size $n = 10$ from a normal population and obtain the data $x_1 = 12.8$, $x_2 = 9.4$, $x_3 = 8.7$, $x_4 = 11.6$, $x_5 = 13.1$, $x_6 = 9.8$, $x_7 = 14.1$, $x_8 = 8.5$, $x_9 = 12.1$, $x_{10} = 10.3$. Now the sample mean is

$$\bar{x} = \frac{12.8 + 9.4 + 8.7 + 11.6 + 13.1 + 9.8 + 14.1 + 8.5 + 12.1 + 10.3}{10} = 11.04$$

the sample median is

$$\tilde{x} = \frac{10.3 + 11.6}{2} = 10.95$$

and a 10% trimmed mean (obtained by discarding the smallest and largest 10% of the sample before averaging) is

$$\bar{x}_{tr10} = \frac{8.7 + 9.4 + 9.8 + 10.3 + 11.6 + 12.1 + 12.8 + 13.1}{8} = 10.98$$

We can show that all of these are unbiased estimates of μ. Since there is not a unique unbiased estimator, we cannot rely on the property of unbiasedness alone to select our estimator. We need a method to select among unbiased estimators. We suggest a method in Section 7-2.3.

7-2.2 Proof That S is a Biased Estimator of σ (CD Only)

7-2.3 Variance of a Point Estimator

Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of θ. This indicates that the distribution of each estimator is centered at the true value of θ. However, the variance of these distributions may be different. Figure 7-1 illustrates the situation. Since $\hat{\Theta}_1$ has a smaller variance than $\hat{\Theta}_2$, the estimator $\hat{\Theta}_1$ is more likely to produce an estimate close to the true value θ. A logical principle of estimation, when selecting among several estimators, is to choose the estimator that has minimum variance.

Definition

If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

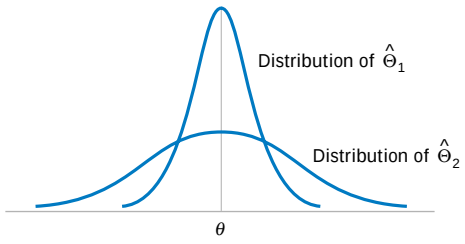


Figure 7-1 The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

In a sense, the MVUE is most likely among all unbiased estimators to produce an estimate $\hat{\theta}$ that is close to the true value of θ. It has been possible to develop methodology to identify the MVUE in many practical situations. While this methodology is beyond the scope of this book, we give one very important result concerning the normal distribution.

Theorem 7-1

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\bar{X}$ is the MVUE for μ.

In situations in which we do not know whether an MVUE exists, we could still use a minimum variance principle to choose among competing estimators. Suppose, for example, we wish to estimate the mean of a population (not necessarily a *normal* population). We have a random sample of n observations $X_1, X_2, \ldots, X_n$ and we wish to compare two possible estimators for μ: the sample mean $\bar{X}$ and a single observation from the sample, say, X_i. Note that both $\bar{X}$ and X_i are unbiased estimators of μ; for the sample mean, we have $V(\bar{X}) = \sigma^2/n$ from Equation 5-40b and the variance of any observation is $V(X_i) = \sigma^2$. Since $V(\bar{X}) < V(X_i)$ for sample sizes $n > 2$, we would conclude that the sample mean is a better estimator of μ than a single observation X_i.

7-2.4 Standard Error: Reporting a Point Estimate

When the numerical value or point estimate of a parameter is reported, it is usually desirable to give some idea of the precision of estimation. The measure of precision usually employed is the standard error of the estimator that has been used.

Definition

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error,** denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Sometimes the estimated standard error is denoted by $s_{\hat{\Theta}}$ or $se(\hat{\Theta})$.

Suppose we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\bar{X}$ is normal with mean μ and variance σ^2/n, so the standard error of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the above equation, the estimated standard error of $\bar{X}$ would be

$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}}$$

When the estimator follows a normal distribution, as in the above situation, we can be reasonably confident that the true value of the parameter lies within two standard errors of the

estimate. Since many point estimators are normally distributed (or approximately so) for large n, this is a very useful result. Even in cases in which the point estimator is not normally distributed, we can state that so long as the estimator is unbiased, the estimate of the parameter will deviate from the true value by as much as four standard errors at most 6 percent of the time. Thus a very conservative statement is that the true value of the parameter differs from the point estimate by at most four standard errors. See Chebyshev's inequality in the CD only material.

EXAMPLE 7-2

An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

41.60, 41.48, 42.34, 41.95, 41.86,
42.18, 41.72, 42.26, 41.81, 42.04

A point estimate of the mean thermal conductivity at 100°F and 550 watts is the sample mean or

$$\bar{x} = 41.924 \text{ Btu/hr-ft-°F}$$

The standard error of the sample mean is $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, and since σ is unknown, we may replace it by the sample standard deviation $s = 0.284$ to obtain the estimated standard error of $\bar{X}$ as

$$\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

Notice that the standard error is about 0.2 percent of the sample mean, implying that we have obtained a relatively precise point estimate of thermal conductivity. If we can assume that thermal conductivity is normally distributed, 2 times the standard error is $2\hat{\sigma}_{\bar{X}} = 2(0.0898) = 0.1796$, and we are highly confident that the true mean thermal conductivity is with the interval 41.924 ± 0.1756, or between 41.744 and 42.104.

7-2.5 Bootstrap Estimate of the Standard Error (CD Only)

7-2.6 Mean Square Error of an Estimator

Sometimes it is necessary to use a biased estimator. In such cases, the mean square error of the estimator can be important. The **mean square error** of an estimator $\hat{\Theta}$ is the expected squared difference between $\hat{\Theta}$ and Θ.

Definition

The **mean square error** of an estimator $\hat{\Theta}$ of the parameter Θ is defined as

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2 \tag{7-3}$$

The mean square error can be rewritten as follows:

$$MSE(\hat{\Theta}) = E[\hat{\Theta} - E(\hat{\Theta})]^2 + [E(\hat{\Theta}) - \Theta]^2$$
$$= V(\hat{\Theta}) + (\text{bias})^2$$

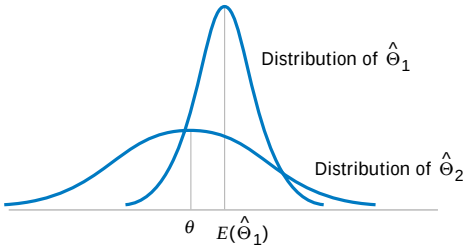


Figure 7-2 A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.

That is, the mean square error of $\hat{\Theta}$ is equal to the variance of the estimator plus the squared bias. If $\hat{\Theta}$ is an unbiased estimator of θ, the mean square error of $\hat{\Theta}$ is equal to the variance of $\hat{\Theta}$.

The mean square error is an important criterion for comparing two estimators. Let $\hat{\Theta}_1$ and $\hat{\Theta}_2$ be two estimators of the parameter θ, and let MSE $(\hat{\Theta}_1)$ and MSE $(\hat{\Theta}_2)$ be the mean square errors of $\hat{\Theta}_1$ and $\hat{\Theta}_2$. Then the **relative efficiency** of $\hat{\Theta}_2$ to $\hat{\Theta}_1$ is defined as

$$\frac{\text{MSE}(\hat{\Theta}_1)}{\text{MSE}(\hat{\Theta}_2)} \tag{7-4}$$

If this relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$, in the sense that it has a smaller mean square error.

Sometimes we find that biased estimators are preferable to unbiased estimators because they have smaller mean square error. That is, we may be able to reduce the variance of the estimator considerably by introducing a relatively small amount of bias. As long as the reduction in variance is greater than the squared bias, an improved estimator from a mean square error viewpoint will result. For example, Fig. 7-2 shows the probability distribution of a biased estimator $\hat{\Theta}_1$ that has a smaller variance than the unbiased estimator $\hat{\Theta}_2$. An estimate based on $\hat{\Theta}_1$ would more likely be close to the true value of θ than would an estimate based on $\hat{\Theta}_2$. Linear regression analysis (Chapters 11 and 12) is an area in which biased estimators are occasionally used.

An estimator $\hat{\Theta}$ that has a mean square error that is less than or equal to the mean square error of any other estimator, for all values of the parameter θ, is called an **optimal** estimator of θ. Optimal estimators rarely exist.

EXERCISES FOR SECTION 7-2

7-1. Suppose we have a random sample of size $2n$ from a population denoted by X, and $E(X) = \mu$ and $V(X) = \sigma^2$. Let

$$\bar{X}_1 = \frac{1}{2n}\sum_{i=1}^{2n} X_i \quad \text{and} \quad \bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

be two estimators of μ. Which is the better estimator of μ? Explain your choice.

7-2. Let $X_1, X_2, \ldots, X_7$ denote a random sample from a population having mean μ and variance σ^2. Consider the following estimators of μ:

$$\hat{\Theta}_1 = \frac{X_1 + X_2 + \cdots + X_7}{7}$$

$$\hat{\Theta}_2 = \frac{2X_1 + X_6 + X_4}{2}$$

(a) Is either estimator unbiased?
(b) Which estimator is best? In what sense is it best?

7-3. Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of the parameter θ. We know that $V(\hat{\Theta}_1) = 10$ and $V(\hat{\Theta}_2) = 4$. Which estimator is best and in what sense is it best?

7-4. Calculate the relative efficiency of the two estimators in Exercise 7-2.

7-5. Calculate the relative efficiency of the two estimators in Exercise 7-3.

7-6. Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are estimators of the parameter θ. We know that $E(\hat{\Theta}_1) = \theta$, $E(\hat{\Theta}_2) = \theta + 2$, $V(\hat{\Theta}_1) = 10$, $V(\hat{\Theta}_2) = 4$. Which estimator is best? In what sense is it best?

7-7. Suppose that $\hat{\Theta}_1$, $\hat{\Theta}_2$, and $\hat{\Theta}_3$ are estimators of θ. We know that $E(\hat{\Theta}_1) = E(\hat{\Theta}_2) = \theta$, $E(\hat{\Theta}_3) = \theta$, $V(\hat{\Theta}_1) = 12$, $V(\hat{\Theta}_2) = 10$, and $E(\hat{\Theta}_3 - \theta)^2 = 6$. Compare these three estimators. Which do you prefer? Why?

7-8. Let three random samples of sizes $n_1 = 20$, $n_2 = 10$, and $n_3 = 8$ be taken from a population with mean μ and variance σ^2. Let S_1^2, S_2^2, and S_3^2 be the sample variances. Show that $S^2 = (20S_1^2 + 10S_2^2 + 8S_3^2)/38$ is an unbiased estimator of σ^2.

7-9. (a) Show that $\sum_{i=1}^{n}(X_i - \bar{X})^2/n$ is a biased estimator of σ^2.
(b) Find the amount of bias in the estimator.
(c) What happens to the bias as the sample size n increases?

7-10. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population with mean μ and variance σ^2.
(a) Show that $\bar{X}^2$ is a biased estimator for σ^2.
(b) Find the amount of bias in this estimator.
(c) What happens to the bias as the sample size n increases?

7-11. Data on pull-off force (pounds) for connectors used in an automobile engine application are as follows: 79.3, 75.1, 78.2, 74.1, 73.9, 75.0, 77.6, 77.3, 73.8, 74.6, 75.5, 74.0, 74.7, 75.9, 72.9, 73.8, 74.2, 78.1, 75.4, 76.3, 75.3, 76.2, 74.9, 78.0, 75.1, 76.8.
(a) Calculate a point estimate of the mean pull-off force of all connectors in the population. State which estimator you used and why.
(b) Calculate a point estimate of the pull-off force value that separates the weakest 50% of the connectors in the population from the strongest 50%.
(c) Calculate point estimates of the population variance and the population standard deviation.
(d) Calculate the standard error of the point estimate found in part (a). Provide an interpretation of the standard error.
(e) Calculate a point estimate of the proportion of all connectors in the population whose pull-off force is less than 73 pounds.

7-12. Data on oxide thickness of semiconductors are as follows: 425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413, 416.
(a) Calculate a point estimate of the mean oxide thickness for all wafers in the population.
(b) Calculate a point estimate of the standard deviation of oxide thickness for all wafers in the population.
(c) Calculate the standard error of the point estimate from part (a).
(d) Calculate a point estimate of the median oxide thickness for all wafers in the population.
(e) Calculate a point estimate of the proportion of wafers in the population that have oxide thickness greater than 430 angstrom.

7-13. $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2. Let $X_{\min}$ and $X_{\max}$ be the smallest and largest observations in the sample.
(a) Is $(X_{\min} + X_{\max})/2$ an unbiased estimate for μ?
(b) What is the standard error of this estimate?
(c) Would this estimate be preferable to the sample mean $\bar{X}$?

7-14. Suppose that X is the number of observed "successes" in a sample of n observations where p is the probability of success on each observation.
(a) Show that $\hat{P} = X/n$ is an unbiased estimator of p.
(b) Show that the standard error of $\hat{P}$ is $\sqrt{p(1-p)/n}$. How would you estimate the standard error?

7-15. $\bar{X}_1$ and S_1^2 are the sample mean and sample variance from a population with mean μ and variance σ_1^2. Similarly, $\bar{X}_2$ and S_2^2 are the sample mean and sample variance from a second independent population with mean μ_1 and variance σ_2^2. The sample sizes are n_1 and n_2, respectively.
(a) Show that $\bar{X}_1 - \bar{X}_2$ is an unbiased estimator of $\mu_1 - \mu_2$.
(b) Find the standard error of $\bar{X}_1 - \bar{X}_2$. How could you estimate the standard error?

7-16. Continuation of Exercise 7-15. Suppose that both populations have the same variance; that is, $\sigma_1^2 = \sigma_2^2 = \sigma_2$. Show that

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

7-17. Two different plasma etchers in a semiconductor factory have the same mean etch rate μ. However, machine 1 is newer than machine 2 and consequently has smaller variability in etch rate. We know that the variance of etch rate for machine 1 is σ_1^2 and for machine 2 it is $\sigma_2^2 = a\sigma_1^2$. Suppose that we have n_1 independent observations on etch rate from machine 1 and n_2 independent observations on etch rate from machine 2.
(a) Show that $\hat{\mu} = \omega\bar{X}_1 + (1 - \omega)\bar{X}_2$ is an unbiased estimator of μ for any value of ω between 0 and 1.
(b) Find the standard error of the point estimate of ω in part (a).
(c) What value of ω would minimize the standard error of the point estimate of μ?
(d) Suppose that $a = 4$ and $n_1 = 2n_2$. What value of ω would you select to minimize the standard error of the point estimate of μ. How "bad" would it be to arbitrarily choose $\omega = 0.5$ in this case?

7-18. Of n_1 randomly selected engineering students at ASU, X_1 owned an HP calculator, and of n_2 randomly selected engineering students at Virginia Tech X_2 owned an HP calculator. Let p_1 and p_2 be the probability that randomly selected ASU and Va. Tech engineering students, respectively, own HP calculators.
(a) Show that an unbiased estimate for $p_1 - p_2$ is $(X_1/n_1) - (X_2/n_2)$.
(b) What is the standard error of the point estimate in part (a)?
(c) How would you compute an estimate of the standard error found in part (b)?
(d) Suppose that $n_1 = 200$, $X_1 = 150$, $n_2 = 250$, and $X_2 = 185$. Use the results of part (a) to compute an estimate of $p_1 - p_2$.
(e) Use the results in parts (b) through (d) to compute an estimate of the standard error of the estimate.

7-3 METHODS OF POINT ESTIMATION

The definitions of unbiasness and other properties of estimators do not provide any guidance about how good estimators can be obtained. In this section, we discuss two methods for obtaining point estimators: the method of moments and the method of maximum likelihood. Maximum likelihood estimates are generally preferable to moment estimators because they have better efficiency properties. However, moment estimators are sometimes easier to compute. Both methods can produce unbiased point estimators.

7-3.1 Method of Moments

The general idea behind the method of moments is to equate **population moments,** which are defined in terms of expected values, to the corresponding **sample moments.** The population moments will be functions of the unknown parameters. Then these equations are solved to yield estimators of the unknown parameters.

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

To illustrate, the first population moment is $E(X) = \mu$, and the first sample moment is $(1/n)\sum_{i=1}^{n} X_i = \bar{X}$. Thus by equating the population and sample moments, we find that $\hat{\mu} = \bar{X}$. That is, the sample mean is the **moment estimator** of the population mean. In the general case, the population moments will be functions of the unknown parameters of the distribution, say, $\theta_1, \theta_2, \ldots, \theta_m$.

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

EXAMPLE 7-3

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from an exponential distribution with parameter λ. Now there is only one parameter to estimate, so we must equate $E(X)$ to $\bar{X}$. For the exponential, $E(X) = 1/\lambda$. Therefore $E(X) = \bar{X}$ results in $1/\lambda = \bar{X}$, so $\hat{\lambda} = 1/\bar{X}$ is the moment estimator of λ.

As an example, suppose that the time to failure of an electronic module used in an automobile engine controller is tested at an elevated temperature to accelerate the failure mechanism.

The time to failure is exponentially distributed. Eight units are randomly selected and tested, resulting in the following failure time (in hours): $x_1 = 11.96$, $x_2 = 5.03$, $x_3 = 67.40$, $x_4 = 16.07$, $x_5 = 31.50$, $x_6 = 7.73$, $x_7 = 11.10$, and $x_8 = 22.38$. Because $\bar{x} = 21.65$, the moment estimate of λ is $\hat{\lambda} = 1/\bar{x} = 1/21.65 = 0.0462$.

EXAMPLE 7-4

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with parameters μ and σ^2. For the normal distribution $E(X) = \mu$ and $E(X^2) = \mu^2 + \sigma^2$. Equating $E(X)$ to $\bar{X}$ and $E(X^2)$ to $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ gives

$$\mu = \bar{X}, \qquad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solving these equations gives the moment estimators

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

Notice that the moment estimator of σ^2 is not an unbiased estimator.

EXAMPLE 7-5

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a gamma distribution with parameters r and λ. For the gamma distribution $E(X) = r/\lambda$ and $E(X^2) = r(r+1)/\lambda^2$. The moment estimators are found by solving

$$r/\lambda = \bar{X}, \quad r(r+1)/\lambda^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

The resulting estimators are

$$\hat{r} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2} \qquad \hat{\lambda} = \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

To illustrate, consider the time to failure data introduced following Example 7-3. For this data, $\bar{x} = 21.65$ and $\sum_{i=1}^{8} x_i^2 = 6639.40$, so the moment estimates are

$$\hat{r} = \frac{(21.65)^2}{(1/8)(6639.40) - (21.65)^2} = 1.30, \qquad \hat{\lambda} = \frac{21.65}{(1/8)(6639.40) - (21.65)^2} = 0.0599$$

When $r = 1$, the gamma reduces to the exponential distribution. Because $\hat{r}$ slightly exceeds unity, it is quite possible that either the gamma or the exponential distribution would provide a reasonable model for the data.

7-3.2 Method of Maximum Likelihood

One of the best methods of obtaining a point estimator of a parameter is the method of maximum likelihood. This technique was developed in the 1920s by a famous British statistician, Sir R. A. Fisher. As the name implies, the estimator will be the value of the parameter that maximizes the **likelihood function.**

Definition

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \tag{7-5}$$

Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

In the case of a discrete random variable, the interpretation of the likelihood function is clear. The likelihood function of the sample $L(\theta)$ is just the probability

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

That is, $L(\theta)$ is just the probability of obtaining the sample values $x_1, x_2, \ldots, x_n$. Therefore, in the discrete case, the maximum likelihood estimator is an estimator that maximizes the probability of occurrence of the sample values.

EXAMPLE 7-6

Let X be a Bernoulli random variable. The probability mass function is

$$f(x; p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$L(p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

We observe that if $\hat{p}$ maximizes $L(p)$, $\hat{p}$ also maximizes $\ln L(p)$. Therefore,

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

Now

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

Equating this to zero and solving for p yields $\hat{p} = (1/n)\sum_{i=1}^{n} x_i$. Therefore, the maximum likelihood estimator of p is

$$\hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Suppose that this estimator was applied to the following situation: n items are selected at random from a production line, and each item is judged as either defective (in which case we set $x_i = 1$) or nondefective (in which case we set $x_i = 0$). Then $\sum_{i=1}^{n} x_i$ is the number of defective units in the sample, and $\hat{p}$ is the **sample proportion defective.** The parameter p is the **population proportion defective;** and it seems intuitively quite reasonable to use $\hat{p}$ as an estimate of p.

Although the interpretation of the likelihood function given above is confined to the discrete random variable case, the method of maximum likelihood can easily be extended to a continuous distribution. We now give two examples of maximum likelihood estimation for continuous distributions.

EXAMPLE 7-7

Let X be normally distributed with unknown μ and known variance σ^2. The likelihood function of a random sample of size n, say $X_1, X_2, \ldots, X_n$, is

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2} = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)/2\sigma^2}$$

Now

$$\ln L(\mu) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

and

$$\frac{d\ln L(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Equating this last result to zero and solving for μ yields

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Thus the sample mean is the maximum likelihood estimator of μ. Notice that this is identical to the moment estimator.

EXAMPLE 7-8

Let X be exponentially distributed with parameter λ. The likelihood function of a random sample of size n, say $X_1, X_2, \ldots, X_n$, is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum_{i=1}^{n} x_i}$$

The log likelihood is

$$\ln L(\lambda) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i$$

Now

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

and upon equating this last result to zero we obtain

$$\hat{\lambda} = n \Big/ \sum_{i=1}^{n} X_i = 1/\bar{X}$$

Thus the maximum likelihood estimator of λ is the reciprocal of the sample mean. Notice that this is the same as the moment estimator.

It is easy to illustrate graphically just how the method of maximum likelihood works. Figure 7-3(a) plots the log of the likelihood function for the exponential parameter from Example 7-8, using the $n = 8$ observations on failure time given following Example 7-3. We found that the estimate of λ was $\hat{\lambda} = 0.0462$. From Example 7-8, we know that this is a maximum likelihood estimate. Figure 7-3(a) shows clearly that the log likelihood function is maximized at a value of λ that is approximately equal to 0.0462. Notice that the log likelihood function is relatively flat in the region of the maximum. This implies that the parameter is not estimated very precisely. If the parameter were estimated precisely, the log likelihood function would be very peaked at the maximum value. The sample size here is relatively small, and this has led to the imprecision in estimation. This is illustrated in Fig. 7-3(b) where we have plotted the difference in log likelihoods for the maximum value, assuming that the sample sizes were $n = 8$, 20, and 40 but that the sample average time to failure remained constant at $\bar{x} = 21.65$. Notice how much steeper the log likelihood is for $n = 20$ in comparsion to $n = 8$, and for $n = 40$ in comparison to both smaller sample sizes.

The method of maximum likelihood can be used in situations where there are several unknown parameters, say, $\theta_1, \theta_2, \ldots, \theta_k$ to estimate. In such cases, the likelihood function is a function of the k unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$, and the maximum likelihood estimators $\{\hat{\Theta}_i\}$ would be found by equating the k partial derivatives $\partial L(\theta_1, \theta_2, \ldots, \theta_k)/\partial\theta_i$, $i = 1, 2, \ldots, k$ to zero and solving the resulting system of equations.

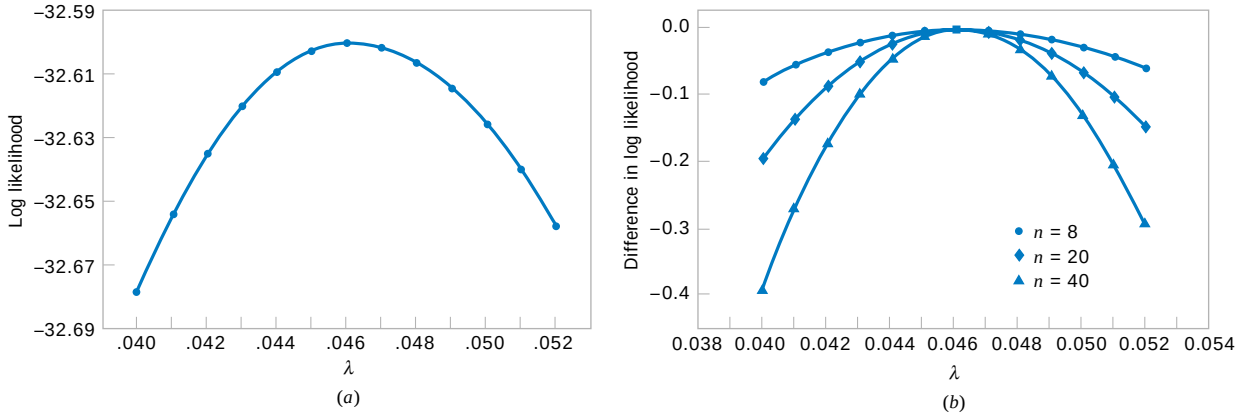


Figure 7-3 Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Log likelihood if $n = 8$, 20, and 40.

EXAMPLE 7-9

Let X be normally distributed with mean μ and variance σ^2, where both μ and σ^2 are unknown. The likelihood function for a random sample of size n is

$$L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\left[\sum_{i=1}^{n}(x_i-\mu)^2\right]/2\sigma^2}$$

and

$$\ln L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Now

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \mu, \sigma^2)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

The solutions to the above equation yield the maximum likelihood estimators

$$\hat{\mu} = \bar{X} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Once again, the maximum likelihood estimators are equal to the moment estimators.

Properties of the Maximum Likelihood Estimator

The method of maximum likelihood is often the estimation method that mathematical statisticians prefer, because it is usually easy to use and produces estimators with good statistical properties. We summarize these properties as follows.

Properties of the Maximum Likelihood Estimator

Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ $[E(\hat{\Theta}) = \theta]$,

(2) the variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator, and

(3) $\hat{\Theta}$ has an approximate normal distribution.

Properties 1 and 2 essentially state that the maximum likelihood estimator is approximately an MVUE. This is a very desirable result and, coupled with the fact that it is fairly easy to obtain in many situations and has an asymptotic normal distribution (the "asymptotic" means "when n is large"), explains why the maximum likelihood estimation technique is widely used. To use maximum likelihood estimation, remember that the distribution of the population must be either known or assumed.

To illustrate the "large-sample" or asymptotic nature of the above properties, consider the maximum likelihood estimator for σ^2, the variance of the normal distribution, in Example 7-9. It is easy to show that

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

The bias is

$$E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$

Because the bias is negative, $\hat{\sigma}^2$ tends to underestimate the true variance σ^2. Note that the bias approaches zero as n increases. Therefore, $\hat{\sigma}^2$ is an asymptotically unbiased estimator for σ^2.

We now give another very important and useful property of maximum likelihood estimators.

The Invariance Property

> Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$.

EXAMPLE 7-10

In the normal distribution case, the maximum likelihood estimators of μ and σ^2 were $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$. To obtain the maximum likelihood estimator of the function $h(\mu, \sigma^2) = 2\sigma^2 - \mu$, substitute the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ into the function h, which yields

$$\hat{\mu} - 2\hat{\sigma}^2 = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^{1/2}$$

Thus, the maximum likelihood estimator of the standard deviation σ is *not* the sample standard deviation S.

Complications in Using Maximum Likelihood Estimation

While the method of maximum likelihood is an excellent technique, sometimes complications arise in its use. For example, it is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\theta)/d\theta = 0$ may be difficult to solve. Furthermore, it may not always be possible to use calculus methods directly to determine the maximum of $L(\theta)$. These points are illustrated in the following two examples.

EXAMPLE 7-11

Let X be uniformly distributed on the interval 0 to a. Since the density function is $f(x) = 1/a$ for $0 \le x \le a$ and zero otherwise, the likelihood function of a random sample of size n is

$$L(a) = \prod_{i=1}^{n}\frac{1}{a} = \frac{1}{a^n}$$

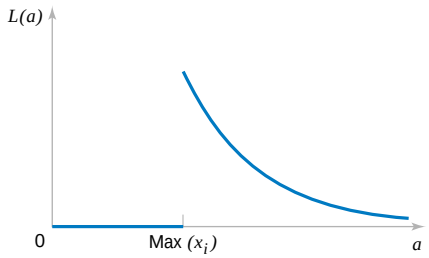


Figure 7-4 The likelihood function for the uniform distribution in Example 7-10.

if $0 \le x_1 \le a$, $0 \le x_2 \le a$, $\ldots$, $0 \le x_n \le a$. Note that the slope of this function is not zero anywhere. That is, as long as $\max(x_i) \le a$, the likelihood is $1/a^n$, which is positive, but when $a < \max(x_i)$, the likelihood goes to zero, as illustrated in Fig. 7-4. Therefore, calculus methods cannot be used directly because the maximum value of the likelihood function occurs at a point of discontinuity. However, since $d/da(1/a^n) = -n/a^{n+1}$ is less than zero for all values of $a > 0$, a^{-n} is a decreasing function of a. This implies that the maximum of the likelihood function $L(a)$ occurs at the lower boundary point. The figure clearly shows that we could maximize $L(a)$ by setting $\hat{a}$ equal to the smallest value that it could logically take on, which is $\max(x_i)$. Clearly, a cannot be smaller than the largest sample observation, so setting $\hat{a}$ equal to the largest sample value is reasonable.

EXAMPLE 7-12 Let $X_1, X_2, \ldots, X_n$ be a random sample from the gamma distribution. The log of the likelihood function is

$$\ln L(r, \lambda) = \ln \left(\prod_{i=1}^{n} \frac{\lambda^r x_i^{r-1} e^{-\lambda x_i}}{\Gamma(r)} \right)$$

$$= nr \ln(\lambda) - n \ln[\Gamma(r)] + (r-1) \sum_{i=1}^{n} \ln(x_i) - \lambda \sum_{i=1}^{n} x_i$$

The derivatives of the log likelihood are

$$\frac{\partial \ln L(r, \lambda)}{\partial r} = n \ln(\lambda) - n \frac{\Gamma'(r)}{\Gamma(r)} + \sum_{i=1}^{n} \ln(x_i)$$

$$\frac{\partial \ln L(r, \lambda)}{\partial \lambda} = \frac{nr}{\lambda} - \sum_{i=1}^{n} x_i$$

When the derivatives are equated to zero, we obtain the equations that must be solved to find the maximum likelihood estimators of r and λ:

$$\hat{\lambda} = \frac{\hat{r}}{\bar{x}}$$

$$n \ln(\hat{\lambda}) - n \frac{\Gamma'(\hat{r})}{\Gamma(\hat{r})} + \sum_{i=1}^{n} \ln(x_i) = 0$$

There is no closed form solution to these equations.

Figure 7-5 shows a graph of the log likelihood for the gamma distribution using the $n = 8$ observations on failure time introduced previously. Figure 7-5(a) shows the **log likelihood**

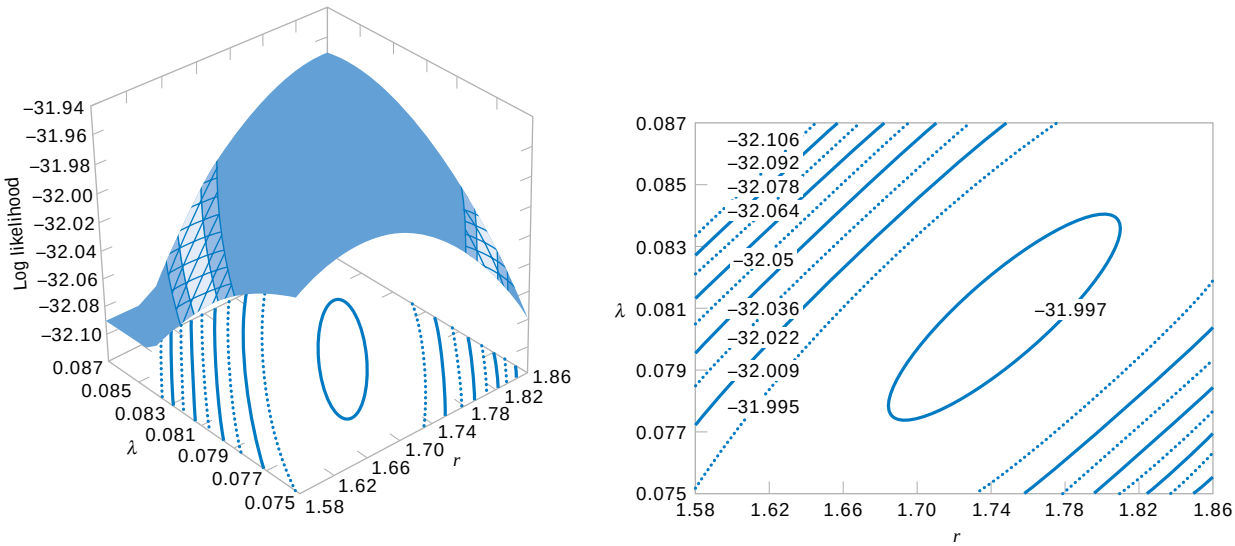


Figure 7-5 Log likelihood for the gamma distribution using the failure time data. (a) Log likelihood surface. (b) Contour plot.

surface as a function of r and λ, and Figure 7-5(b) is a **contour plot.** These plots reveal that the log likelihood is maximized at approximately $\hat{r} = 1.75$ and $\hat{\lambda} = 0.08$. Many statistics computer programs use numerical techniques to solve for the maximum likelihood estimates when no simple solution exists.

7-3.3 Bayesian Estimation of Parameters (CD Only)

EXERCISES FOR SECTION 7-3

7-19. Consider the Poisson distribution

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0, 1, 2, \ldots$$

Find the maximum likelihood estimator of λ, based on a random sample of size n.

7-20. Consider the shifted exponential distribution

$$f(x) = e^{-(x-\theta)}, \qquad x \geq \theta$$

When $\theta = 0$, this density reduces to the usual exponential distribution. When $\theta > 0$, there is only positive probability to the right of θ.

(a) Find the maximum likelihood estimator of θ and λ, based on a random sample of size n.

(b) Describe a practical situation in which one would suspect that the shifted exponential distribution is a plausible model.

7-21. Let X be a geometric random variable with parameter p. Find the maximum likelihood estimator of p, based on a random sample of size n.

7-22. Let X be a random variable with the following probability distribution:

$$f(x) = \begin{cases} (\theta+1)x^{\theta}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimator of θ, based on a random sample of size n.

7-23. Consider the Weibull distribution

$$f(x) = \begin{cases} \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}, & 0 \leq x \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the likelihood function based on a random sample of size n. Find the log likelihood.

(b) Show that the log likelihood is maximized by solving the equations

$$\frac{\sum_{i=1}^{n} x_i^{\theta} \ln(x_i)}{\sum_{i=1}^{n} x_i^{\theta}} - \frac{\sum_{i=1}^{n} \ln(x_i)}{n} = \frac{1}{\theta}$$

$$\theta = \left[\frac{\sum_{i=1}^{n} x_i^{\theta}}{n}\right]^{1/\theta}$$

(c) What complications are involved in solving the two equations in part (b)?

7-24. Consider the probability distribution in Exercise 7-22. Find the moment estimator of θ.

7-25. Let $X_1, X_2, \ldots, X_n$ be uniformly distributed on the interval 0 to a. Show that the moment estimator of a is $\hat{a} = 2\bar{X}$. Is this an unbiased estimator? Discuss the reasonableness of this estimator.

7-26. Let $X_1, X_2, \ldots, X_n$ be uniformly distributed on the interval 0 to a. Recall that the maximum likelihood estimator of a is $\hat{a} = \max(X_i)$.

(a) Argue intuitively why $\hat{a}$ cannot be an unbiased estimator for a.

(b) Suppose that $E(\hat{a}) = na/(n+1)$. Is it reasonable that $\hat{a}$ consistently underestimates a? Show that the bias in the estimator approaches zero as n gets large.

(c) Propose an unbiased estimator for a.

(d) Let $Y = \max(X_i)$. Use the fact that $Y \le y$ if and only if each $X_i \le y$ to derive the cumulative distribution function of Y. Then show that the probability density function of Y is

$$f(y) = \begin{cases} \dfrac{ny^{n-1}}{a^n}, & 0 \le y \le a \\ 0, & \text{otherwise} \end{cases}$$

Use this result to show that the maximum likelihood estimator for a is biased.

7-27. For the continuous distribution of the interval 0 to a, we have two unbiased estimators for a: the moment estimator $\hat{a}_1 = 2\bar{X}$ and $\hat{a}_2 = [(n+1)/n] \max(X_i)$, where $\max(X_i)$ is the largest observation in a random sample of size n (see Exercise 7-26). It can be shown that $V(\hat{a}_1) = a^2/(3n)$ and that $V(\hat{a}_2) = a^2/[n(n+2)]$. Show that if $n > 1$, $\hat{a}_2$ is a better estimator than $\hat{a}$. In what sense is it a better estimator of a?

7-28. Consider the probability density function

$$f(x) = \frac{1}{\theta^2} x e^{-x/\theta}, \qquad 0 \le x \le \infty, \quad 0 \le \theta \le \infty$$

Find the maximum likelihood estimator for θ.

7-29. The Rayleigh distribution has probability density function

$$f(x) = \frac{x}{\theta} e^{-x^2/2\theta}, \qquad x \ge 0, \qquad 0 \le \theta \le \infty$$

(a) It can be shown that $E(X^2) = 2\theta$. Use this information to construct an unbiased estimator for θ.

(b) Find the maximum likelihood estimator of θ. Compare your answer to part (a).

(c) Use the invariance property of the maximum likelihood estimator to find the maximum likelihood estimator of the median of the Raleigh distribution.

7-30. Consider the probability density function

$$f(x) = c(1 - x), \quad -1 \le x \le 1$$

(a) Find the value of the constant c.

(b) What is the moment estimator for θ?

(c) Show that $\hat{\theta} = 3\bar{X}$ is an unbiased estimator for θ.

(d) Find the maximum likelihood estimator for θ.

7-31. Reconsider the oxide thickness data in Exercise 7-12 and suppose that it is reasonable to assume that oxide thickness is normally distributed.

(a) Use the results of Example 7-9 to compute the maximum likelihood estimates of μ and σ^2.

(b) Graph the likelihood function in the vicinity of $\hat{\mu}$ and $\hat{\sigma}^2$, the maximum likelihood estimates, and comment on its shape.

7-32. Continuation of Exercise 7-31. Suppose that for the situation of Exercise 7-12, the sample size was larger ($n = 40$) but the maximum likelihood estimates were numerically equal to the values obtained in Exercise 7-31. Graph the likelihood function for $n = 40$, compare it to the one from Exercise 7-31 (b), and comment on the effect of the larger sample size.

7-4 SAMPLING DISTRIBUTIONS

Statistical inference is concerned with making **decisions** about a population based on the information contained in a random sample from that population. For instance, we may be interested in the mean fill volume of a can of soft drink. The mean fill volume in the

population is required to be 300 milliliters. An engineer takes a random sample of 25 cans and computes the sample average fill volume to be $\bar{x}$ □ 298 milliliters. The engineer will probably decide that the population mean is □ □ 300 milliliters, even though the sample mean was 298 milliliters because he or she knows that the sample mean is a reasonable estimate of □ and that a sample mean of 298 milliliters is very likely to occur, even if the true population mean is □ □ 300 milliliters. In fact, if the true mean is 300 milliliters, tests of 25 cans made repeatedly, perhaps every five minutes, would produce values of $\bar{x}$ that vary both above and below □ □ 300 milliliters.

The sample mean is a statistic; that is, it is a random variable that depends on the results obtained in each particular sample. Since a statistic is a random variable, it has a probability distribution.

Definition

The probability distribution of a statistic is called a **sampling distribution.**

For example, the probability distribution of $\overline{X}$ is called the **sampling distribution of the mean.**

The sampling distribution of a statistic depends on the distribution of the population, the size of the sample, and the method of sample selection. The next section presents perhaps the most important sampling distribution. Other sampling distributions and their applications will be illustrated extensively in the following two chapters.

7-5 SAMPLING DISTRIBUTIONS OF MEANS

Consider determining the sampling distribution of the sample mean $\overline{X}$. Suppose that a random sample of size n is taken from a normal population with mean □ and variance □^{2}. Now each observation in this sample, say, X_1, X_2, p , X_n, is a normally and independently distributed random variable with mean □ and variance □^{2}. Then by the reproductive property of the normal distribution, Equation 5-41 in Chapter 5, we conclude that the sample mean

$$\overline{X} \square \frac{X_1 \square X_2 \square \text{p} \square X_n}{n}$$

has a normal distribution with mean

$$\square_{\overline{X}} \square \frac{\square \square \square \square \text{p} \square \square}{n} \square \square$$

and variance

$$\square^2_{\overline{X}} \square \frac{\square^2 \square \square^2 \square \text{p} \square \square^2}{n^2} \square \frac{\square^2}{n}$$

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean □ and

variance σ^2/n, if the sample size n is large. This is one of the most useful theorems in statistics, called the **central limit theorem.** The statement is as follows:

Theorem 7-2: The Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{7-6}$$

as $n \to \infty$, is the standard normal distribution.

The normal approximation for $\bar{X}$ depends on the sample size n. Figure 7-6(a) shows the distribution obtained for throws of a single, six-sided true die. The probabilities are equal (1/6) for all the values obtained, 1, 2, 3, 4, 5, or 6. Figure 7-6(b) shows the distribution of the average score obtained when tossing two dice, and Fig. 7-6(c), 7-6(d), and 7-6(e) show the distributions of average scores obtained when tossing three, five, and ten dice, respectively. Notice that, while the population (one die) is relatively far from normal, the distribution of averages is approximated reasonably well by the normal distribution for sample sizes as small as five. (The dice throw distributions are discrete, however, while the normal is continuous). Although the central limit theorem will work well for small samples ($n = 4, 5$) in most cases, particularly where the population is continuous, unimodal, and symmetric, larger samples will be required in other situations, depending on the shape of the population. In many cases of practical interest, if $n > 30$, the normal approximation will be satisfactory regardless of the

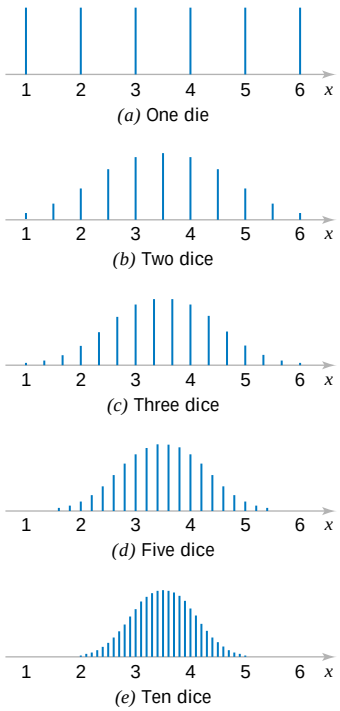


Figure 7-6 Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

shape of the population. If $n \geq 30$, the central limit theorem will work if the distribution of the population is not severely nonnormal.

EXAMPLE 7-13

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\bar{X}$ is normal, with mean $\mu_{\bar{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-7. Standardizing the point $\bar{X} = 95$ in Fig. 7-7, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$P(\bar{X} \leq 95) = P(Z \leq -2.5) = 0.0062$$

The following example makes use of the central limit theorem.

EXAMPLE 7-14

Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2, & 4 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.

The mean and variance of X are $\mu = 5$ and $\sigma^2 = (6-4)^2/12 = 1/3$. The central limit theorem indicates that the distribution of $\bar{X}$ is approximately normal with mean $\mu_{\bar{X}} = 5$ and variance $\sigma^2_{\bar{X}} = \sigma^2/n = 1/(3)(40) = 1/120$. The distributions of X and $\bar{X}$ are shown in Fig. 7-8.

Now consider the case in which we have two independent populations. Let the first population have mean μ_1 and variance σ_1^2 and the second population have mean μ_2 and variance σ_2^2. Suppose that both populations are normally distributed. Then, using the fact that linear

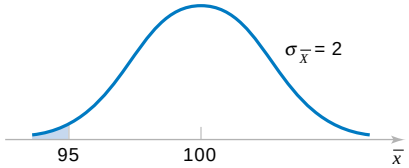


Figure 7-7 Probability for Example 7-13.

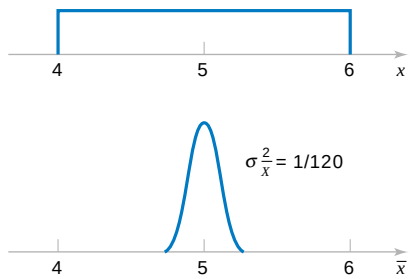


Figure 7-8 The distributions of X and $\bar{X}$ for Example 7-14.

combinations of independent normal random variables follow a normal distribution (see Equation 5-41), we can say that the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is normal with mean

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2 \tag{7-7}$$

and variance

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \tag{7-8}$$

If the two populations are not normally distributed and if both sample sizes n_1 and n_2 are greater than 30, we may use the central limit theorem and assume that $\bar{X}_1$ and $\bar{X}_2$ follow approximately independent normal distributions. Therefore, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is approximately normal with mean and variance given by Equations 7-7 and 7-8, respectively. If either n_1 or n_2 is less than 30, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ will still be approximately normal with mean and variance given by Equations 7-7 and 7-8, provided that the population from which the small sample is taken is not dramatically different from the normal. We may summarize this with the following definition.

Definition

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\bar{X}_1$ and $\bar{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{7-9}$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

EXAMPLE 7-15

The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the "old" process and a random sample of $n_2 = 25$ components is selected from the "improved" process. What is the probability that the difference in the two sample means $\bar{X}_2 - \bar{X}_1$ is at least 25 hours? Assume that the old and improved processes can be regarded as independent populations.

To solve this problem, we first note that the distribution of $\bar{X}_1$ is normal with mean $\mu_1 = 5000$ hours and standard deviation $\sigma_1/\sqrt{n_1} = 40/\sqrt{16} = 10$ hours, and the distribution of $\bar{X}_2$ is normal with mean $\mu_2 = 5050$ hours and standard deviation $\sigma_2/\sqrt{n_2} = 30/\sqrt{25} = 6$ hours. Now the distribution of $\bar{X}_2 - \bar{X}_1$ is normal with mean $\mu_2 - \mu_1 = 5050 - 5000 = 50$ hours and variance $\sigma_2^2/n_2 + \sigma_1^2/n_1 = 6^2 + 10^2 = 136$ hours^2. This sampling distribution is shown in Fig. 7-9. The probability that $\bar{X}_2 - \bar{X}_1 = 25$ is the shaded portion of the normal distribution in this figure.