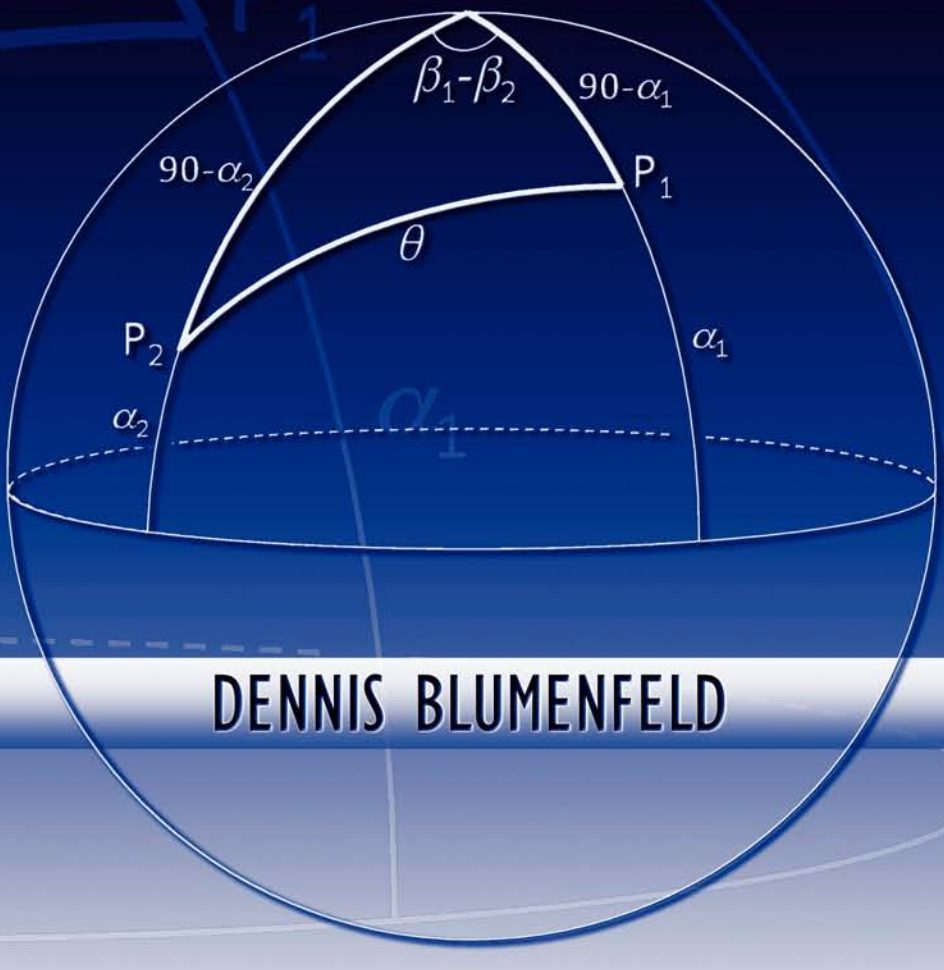
SECOND EDITION

OPERATIONS RESEARCH CALCULATIONS HANDBOOK



DENNIS BLUMENFELD

CRC Press
Taylor & Francis Group

SECOND EDITION

OPERATIONS RESEARCH CALCULATIONS HANDBOOK

The Operations Research Series

Series Editor: A. Ravi Ravindran
Dept. of Industrial & Manufacturing Engineering
The Pennsylvania State University, USA

Integer Programming: Theory and Practice
John K. Karlof

Operations Research Applications
A. Ravi Ravindran

Operations Research: A Practical Approach
Michael W. Carter and Camille C. Price

Operations Research Calculations Handbook, Second Edition
Dennis Blumenfeld

Operations Research and Management Science Handbook
A. Ravi Ravindran

Forthcoming Titles

Applied Nonlinear Optimization in Modeling Environments
Janos D. Pinter

Probability Models in Operations Research
Richard C. Cassady and Joel A. Nachlas

SECOND EDITION

OPERATIONS RESEARCH CALCULATIONS HANDBOOK

DENNIS BLUMENFELD

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2009 by Taylor and Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4200-5240-4 (Paperback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface to the Second Edition xi
Preface to the First Edition xiii
Acknowledgments xv
Author xvii

1. Introduction 1

2. Means and Variances 3
2.1 Mean (Expectation) and Variance of a Random Variable 3
2.2 Covariance and Correlation Coefficient 5
2.3 Mean and Variance of the Sum of Random Variables 5
2.4 Mean and Variance of the Product of Two Random Variables 7
2.5 Mean and Variance of the Quotient of Two Random Variables 8
2.6 Conditional Mean and Variance for Jointly Distributed Random Variables 9
2.7 Conditional Mean of a Constrained Random Variable 9
2.8 Mean and Variance of the Sum of a Random Number of Random Variables 11
2.9 Mean of a Function of a Random Variable 11
2.10 Approximations for the Mean and Variance of a Function of a Random Variable 12
2.11 Mean and Variance of the Maximum of Exponentially Distributed Random Variables 13
2.12 Mean and Variance of the Maximum of Normally Distributed Random Variables 13

3. Discrete Probability Distributions 17
3.1 Bernoulli Distribution 17
3.2 Binomial Distribution 18
3.3 Geometric Distribution 20
3.4 Negative Binomial Distribution 22
3.5 Poisson Distribution 24
3.6 Hypergeometric Distribution 26
3.7 Multinomial Distribution 29

4. Continuous Probability Distributions 31
4.1 Uniform Distribution 31
4.2 Exponential Distribution 32
4.3 Erlang Distribution 34
4.4 Gamma Distribution 37
4.5 Beta Distribution 41
4.6 Normal Distribution 43
4.6.1 Sum of Normally Distributed Random Variables 44
4.6.2 Standard Normal Distribution 45
4.6.3 Partial Moments for the Normal Distribution 46
4.6.4 Approximations for the Cumulative Normal Distribution Function 47
4.7 Lognormal Distribution 48
4.8 Weibull Distribution 51
4.9 Logistic Distribution 53
4.10 Gumbel (Extreme Value) Distribution 54
4.11 Pareto Distribution 56
4.12 Triangular Distribution 58
4.13 Hyper-Exponential and Hypo-Exponential Distributions 60
4.13.1 Hyper-Exponential Distribution 60
4.13.2 Hypo-Exponential Distribution 63
4.13.3 Additional Comments on Hyper- and Hypo-Exponential Distributions 66

5. Probability Relationships 69
5.1 Distribution of the Sum of Independent Random Variables 69
5.2 Distribution of the Maximum and Minimum of Random Variables 69
5.2.1 Example for the Uniform Distribution 70
5.2.2 Example for the Exponential Distribution 71
5.3 Change of Variable in a Probability Distribution 72
5.4 Conditional Probability Distribution for a Constrained Random Variable 73
5.5 Combination of Poisson and Gamma Distributions 75
5.6 Bayes' Formula 76
5.7 Central Limit Theorem 77
5.8 Probability Generating Function (z-Transform) 78
5.9 Moment Generating Function 79
5.10 Characteristic Function 81
5.11 Laplace Transform 83

6. Stochastic Processes 85
6.1 Poisson Process and Exponential Distribution 85
6.1.1 Properties of the Poisson Process 85
6.1.2 "Lack of Memory" Property of the Exponential Distribution 85
6.1.3 Competing Exponentials 86
6.1.4 Superposition of Independent Poisson Processes 86
6.1.5 Splitting of a Poisson Process 86
6.1.6 Arrivals from a Poisson Process in a Fixed Interval 87
6.2 Renewal Process Results 87
6.2.1 Mean and Variance of the Number of Arrivals in a Renewal Process 87
6.2.2 Distribution of First Interval in a Renewal Process 87
6.3 Markov Chain Results 88
6.3.1 Discrete-Time Markov Chains 89
6.3.2 Continuous-Time Markov Chains 89

7. Queueing Theory Results 91
7.1 Notation for Queue Types 91
7.2 Definitions of Queueing System Variables 91
7.3 Little's Law and General Queueing System Relationships 92
7.4 Extension of Little's Law 93
7.5 Formulas for Average Queue Length, L_q 93
7.6 Formulas for Average Time in Queue, W_q 94
7.7 References for the Formulas for Average Queue Length and Time in Queue 95
7.8 Pollaczek–Khintchine Formula for Average Time in Queue, W_q 96
7.9 Additional Formulas for Average Time in Queue, W_q 96
7.10 Heavy Traffic Approximation for Distribution of Time in Queue 97
7.11 Queue Departure Process 98
7.12 Distribution Results for the Number of Customers in $M/M/1$ Queue 99
7.13 Distribution Results for Time in $M/M/1$ Queue 99
7.14 Other Formulas in Queueing Theory 100

8. Production Systems Modeling 101
8.1 Definitions and Notation for Workstations 101
8.2 Basic Relationships between Workstation Parameters 101
8.3 Distribution of the Time to Produce a Fixed Lot Size at a Workstation 102

8.4 Throughput of a Serial Production Line with Failures 104
8.4.1 Line without Buffers 104
8.4.2 Line with Buffers 105
8.5 Throughput of a Two-Station Serial Production Line with Variable Processing Times 106
8.5.1 Two Stations without a Buffer 106
8.5.2 Two Stations with a Buffer 107
8.6 Throughput of an *N*-Station Serial Production Line with Variable Processing Times 107

9. Inventory Control 109
9.1 Economic Order Quantity 109
9.2 Economic Production Quantity 111
9.3 "Newsboy Problem": Optimal Inventory to Meet Uncertain Demand in a Single Period 114
9.4 Inventory Replenishment Policies 116
9.5 (*s*, *Q*) Policy: Estimates of Reorder Point (*s*) and Order Quantity (*Q*) 119
9.6 (*s*, *S*) Policy: Estimates of Reorder Point (*s*) and Order-Up-To Level (*S*) 122
9.7 (*T*, *S*) Policy: Estimates of Review Period (*T*) and Order-Up-To Level (*S*) 124
9.8 (*T*, *s*, *S*) Policy: Estimates of Review Period (*T*), Reorder Point (*s*), and Order-Up-To Level (*S*) 126
9.9 Summary of Results for Inventory Policies 128
9.10 Inventory in a Production/Distribution System 129
9.11 Note on Cumulative Plots 131

10. Distance Formulas for Logistics Analysis 133
10.1 Distance Norms 133
10.2 "Traveling Salesman Problem" Tour Distance: Shortest Path through a Set of Points in a Region 136
10.3 Distribution of Distance between Two Random Points in a Circle 136
10.4 Average Rectangular Grid Distance between Two Random Points in a Circle 139
10.5 Great Circle Distance 139

11. Traffic Flow and Delay 143
11.1 Traffic Flow Parameters 143
11.2 Traffic Speeds 143
11.3 Delay to Vehicle Merging with Traffic Stream 147

11.4 Critical Flow on Minor Road 148
11.5 Delay to Traffic Queue on Minor Road Waiting to Merge 149
11.6 Delay to Vehicle at Traffic Signal 150

12. Linear Programming Formulations 153
12.1 General Formulation 153
12.2 Terminology 154
12.3 Example of a Feasible Region 155
12.4 Alternative Formulations 156
12.4.1 Minimization vs. Maximization 156
12.4.2 Equality Constraints 156
12.4.3 Reversed Inequality Constraints 157
12.5 Diet Problem 157
12.6 Duality 158
12.7 Special Cases of Linear Programming Problems 160
12.7.1 Transportation Problem 160
12.7.2 Transshipment Problem 163
12.7.3 Assignment Problem 164
12.8 Integer Linear Programming Formulations 165
12.8.1 Knapsack Problem 166
12.8.2 Traveling Salesman Problem 167
12.9 Solution Methods 169
12.9.1 Simplex Method 169
12.9.2 Interior-Point Methods 170
12.9.3 Network Flow Methods 170
12.9.4 Cutting Planes 170
12.9.5 Branch and Bound 171

13. Heuristic Search Methods 173
13.1 Overview of Heuristics 173
13.2 Local Search Methods 175
13.3 Simulated Annealing 176
13.4 Tabu Search 178
13.5 Genetic Algorithms 179
13.6 Other Heuristics 180

14. Order Statistics 181
14.1 General Distribution Order Statistics 181
14.2 Uniform Distribution Order Statistics 182
14.3 Exponential Distribution Order Statistics 184

15. Mathematical Functions 187
15.1 Gamma Function 187
15.2 Incomplete Gamma Function 187

15.3 Beta Function 188
15.4 Incomplete Beta Function 188
15.5 Unit Impulse Function 189
15.6 Modified Bessel Functions 189
15.7 Stirling's Formula 190

16. Calculus Results 191
16.1 Basic Rules for Differentiation 191
16.2 Integration by Parts 192
16.3 Fundamental Theorem of Calculus 193
16.4 Taylor Series 193
16.5 Maclaurin Series 194
16.6 L'Hôpital's Rule 195
16.7 Lagrange Multipliers 196
16.8 Differentiation under the Integral Sign (Leibnitz's Rule) 197
16.9 Change of a Variable in an Integral 198
16.10 Change of Variables in a Double Integral 199
16.11 Changing the Order of Integration in a Double Integral 201
16.12 Changing the Order of Summation in a Double Sum 203
16.13 Numerical Integration 204

17. Matrices 209
17.1 Rules for Matrix Calculations 209
17.2 Inverses of Matrices 210
17.2.1 Inverse of 2×2 Matrix 210
17.2.2 Inverse of 3×3 Matrix 210
17.3 Series of Matrices 211
17.4 Derivatives of Matrices 211

18. Combinatorics 213

19. Summations 215
19.1 Finite Sums 215
19.2 Infinite Sums 215

20. Interest Formulas 217

References 219

Index 231

Preface to the Second Edition

The field of operations research encompasses a growing number of technical areas. The scope of the second edition has been expanded to cover several additional topics. These include new chapters on order statistics, heuristic search methods, and traffic flow and delay. Some chapters have also been updated with new material, and many new references have been added. As before, the focus is on presenting handy analytical results and formulas that allow quick calculations and provide the understanding of system models.

Dennis E. Blumenfeld

Preface to the First Edition

Operations research uses analyses and techniques from a variety of branches of mathematics, statistics, and other scientific disciplines. Certain analytical results arise repeatedly in applications of operations research to industrial and service operations. These results are scattered among many different textbooks and journal articles, sometimes in the midst of extensive derivations. The idea for a handbook of operations research results came from a need to have frequently used results to be readily available in one source of reference.

This handbook is a compilation of analytical results and formulas that have been found useful in various applications. The objective is to provide students, researchers, and practitioners with convenient access to wide range of operations research results in a concise format.

Given the extensive variety of applications of operations research, a collection of results cannot be exhaustive. The selection of results included in this handbook is based on experience in the manufacturing industry. Many of the results are basic to system modeling, and are likely to carry over to applications in other areas of operations research and management science.

This handbook focuses on areas of operations research that yield explicit analytical results and formulas. With the widespread availability of computer software for simulations and algorithms, many analyses can be easily performed numerically without knowledge of explicit formulas. However, formulas continue to play a significant role in system modeling. While software packages are useful for obtaining numerical results for given values of input parameters, formulas allow general conclusions to be drawn about system behavior as parameter values vary. Analytical results and formulas also help to provide an intuitive understanding of the underlying models for system performance. Such understanding is important in the implementation of operations research models as it allows analysts and decision makers to use models with confidence.

Dennis E. Blumenfeld

Happy is the man that findeth wisdom, and the man that getteth understanding.

—Proverbs 3:13

Acknowledgments

It is a pleasure to thank colleagues who have given me suggestions and ideas, and have shared their expertise. In particular, I wish to thank David Kim for his valuable contributions and discussions on the basic content and organization. My thanks also go to Jeffrey Alden, Robert Bordley, Debra Elkins, Randolph Hall, Ningjian Huang, William Jordan, Jonathan Owen, and Brian Tomlin for their willingness to review earlier versions and provide many helpful and constructive comments, and to Cindy Carelli and her colleagues at CRC Press for their careful and professional editorial work. I thank my wife, Sharon, for her patience and encouragement. She helped me to adhere to a deadline, with her repeated calls of "Author! Author!"

Author

Dennis E. Blumenfeld teaches on the adjunct faculty at the University of Michigan, Ann Arbor. He has spent most of his career as a staff research scientist at the General Motors Research and Development Center, Warren, Michigan. Before joining General Motors, he held faculty positions at Princeton University, New Jersey, and at University College London. He received his BSc in mathematics and his MSc in statistics and operations research from Imperial College London, and his PhD in civil engineering from University College London. He is a member of the Institute for Operations Research and the Management Sciences, and a fellow of the Royal Statistical Society. He has published articles on transportation models, traffic flow and queueing, logistics, inventory control, and production systems, and has served on the editorial advisory board of *Transportation Research*.

1

Introduction

Operations research can be considered as the science of decision making. It encompasses many scientific disciplines, such as mathematics, statistics, computer science, physics, engineering, economics, and social sciences, and has been successful in providing a systematic approach to complex decisions in manufacturing, service, military, and financial operations.

One of the reasons for the appeal and success of operations research is that it draws on basic mathematical principles and uses them in clever and novel ways to solve all kinds of real-world problems. Many of the applications make use of handy analytical results and formulas derived from system models, and can reveal how system performance varies with model parameters.

Often, these analytical results and formulas offer insight that numerical methods do not provide. Even though numerical solutions can now be easily obtained with the greatly increased speed and power of computers in recent years, there is still a need for analytical results to highlight trade-offs between the different parameters in a model, and to make the mathematical relationships between the parameters readily apparent.

Analytical results and formulas often require minimal data and allow quick "back-of-the-envelope" calculations that are very useful for initial analyses. This is important when an approximate estimate is all that is needed, or all there is time for, in the many real-world situations where decisions must be made quickly. In situations where there is more time, and many alternatives are to be evaluated, such initial analyses can provide focus as to where more detailed numerical analyses are warranted. Since formulas are not limited to any particular programming language, computer operating system, or user interface, they can be readily used on their own for system analyses or be included as components of comprehensive decision-making tools.

The objective of this handbook is to provide a concise collection of analytical results and formulas that arise in operations research applications. The material is organized into chapters based on the following topics.

The first few chapters are devoted to results on the stochastic modeling aspects of operations research. Chapter 2 covers a range of formulas that involve the mean and the variance of random variables. Chapters 3 and 4 list the main properties of widely used discrete and continuous probability distributions. Chapter 5 contains a collection of other analytical results

that frequently arise in probability. Chapters 6 and 7 present formulas that arise in stochastic processes and queueing theory.

The next four chapters cover specific applications of operations research in the areas of stochastic modeling. Chapter 8 presents some results in production systems modeling and Chapter 9 covers the basic formulas in inventory control. Chapter 10 gives distance formulas that are useful in logistics and spatial analyses. Chapter 11 presents basic results in traffic flow and delay.

Chapters 12 and 13 cover the standard linear programming formulations and heuristic search methods. These subjects deal with the development of algorithms and methodologies in optimization. In keeping with the intent of this handbook, which is to focus on analytical results and formulas, these two chapters present the mathematical formulations and basic concepts, and give references for the solution methods.

The remaining chapters contain basic mathematical results that are relevant to operations research. Chapter 14 covers key results in order statistics, Chapter 15 lists some common mathematical functions that arise in applications, Chapter 16 presents useful results from elementary and more advanced calculus, Chapter 17 lists the standard properties of matrices, Chapter 18 gives the standard formulas for combinatorial calculations, Chapter 19 lists some common results for finite and infinite sums, and, finally, Chapter 20 gives basic interest formulas that are important in economic analysis.

To supplement the various results and formulas, references are given for derivations and additional details.

2

Means and Variances

2.1 Mean (Expectation) and Variance of a Random Variable

For a discrete random variable X that takes the values x_0, x_1, x_2, …, the mean of X is given by

$$E[X] = \sum_{i=0}^{\infty} x_i \cdot \Pr\{X = x_i\} \tag{2.1}$$

where
$E[X]$ denotes the mean (expected value or expectation) of X
$\Pr\{X = x_i\}$ denotes the probability that X takes the value x_i ($i = 0, 1, 2, \ldots$)

If X takes nonnegative integer values only ($X = 0, 1, 2, \ldots$), then the mean of X is given by

$$E[X] = \sum_{n=0}^{\infty} n \cdot \Pr\{X = n\} \tag{2.2}$$

$$= \sum_{n=0}^{\infty} \Pr\{X > n\} \tag{2.3}$$

For a continuous random variable X ($-\infty < X < \infty$), the mean of X is given by

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \tag{2.4}$$

$$= \int_{0}^{\infty} [1 - F(x)] dx - \int_{-\infty}^{0} F(x) dx \tag{2.5}$$

where
$E[X]$ denotes the mean (expected value) of X
$f(x)$ is the probability density function of X

and

$$F(x) = \Pr\{X \le x\} = \int_{-\infty}^{x} f(t)dt$$

denotes the cumulative distribution function of X.

If X is continuous and takes nonnegative values only ($0 \le X < \infty$), then the mean of X is given by

$$E[X] = \int_{0}^{\infty} xf(x)dx \tag{2.6}$$

$$= \int_{0}^{\infty} \left[1 - F(x)\right]dx \tag{2.7}$$

Çinlar (1975, pp. 22, 25–26); Lefebvre (2006, pp. 96–97); Mood, Graybill, and Boes (1974, pp. 64–65).

For any random variable X, the variance is given by

$$Var[X] = E\left\{\left(X - E[X]\right)^2\right\} \tag{2.8}$$

$$= E\left[X^2\right] - \left(E[X]\right)^2 \tag{2.9}$$

where $Var[X]$ denotes the variance of X
and

$$E\left[X^2\right] = \begin{cases} \sum_{x} x^2 \cdot \Pr\{X = x\} & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^2 f(x)dx & \text{if } X \text{ is continuous} \end{cases} \tag{2.10}$$

The standard deviation of X, *St Dev*[X], is given by

$$St\,Dev[X] = \sqrt{Var[X]} \tag{2.11}$$

Binmore (1983, pp. 268–269); Çinlar (1975, p. 31); Feller (1964, p. 213); Lefebvre (2006, pp. 99–100); Mood, Graybill, and Boes (1974, pp. 68, 70); Ross (2003, pp. 46–47).

2.2 Covariance and Correlation Coefficient

For any random variables X and Y, the covariance *Cov*[X, Y] is given by

$$Cov[X,Y] = E\left\{\left(X - E[X]\right)\left(Y - E[Y]\right)\right\} \tag{2.12}$$

$$= E[XY] - E[X]E[Y] \tag{2.13}$$

and the correlation coefficient *Corr*[X, Y] is given by

$$Corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} \tag{2.14}$$

The correlation coefficient is dimensionless and satisfies the condition $-1 \le Corr[X, Y] \le 1$.

If X and Y are independent, then the covariance, *Cov*[X, Y], and correlation coefficient, *Corr*[X, Y], are zero.

Feller (1964, pp. 215, 221); Mood, Graybill, and Boes (1974, pp. 155–156, 161); Ross (2003, p. 53).

2.3 Mean and Variance of the Sum of Random Variables

For any random variables X and Y, the mean of the sum $X+Y$ is given by

$$E[X+Y] = E[X] + E[Y] \tag{2.15}$$

This result for the mean of a sum holds even if the random variables are not independent.

If the random variables X and Y are independent, then the variance of the sum $X+Y$ is given by

$$Var[X+Y]=Var[X]+Var[Y] \tag{2.16}$$

If the random variables X and Y are not independent, then the variance of the sum $X+Y$ is given by

$$Var[X+Y]=Var[X]+Var[Y]+2Cov[X,Y] \tag{2.17}$$

where $Cov[X, Y]$ is the covariance of X and Y given by Equation 2.12.

For any random variables X and Y, and any constants a and b, the mean and the variance of the linear combination $aX+bY$ are given by

$$E[aX+bY]=aE[X]+bE[Y] \tag{2.18}$$

and

$$Var[aX+bY]=a^2Var[X]+b^2Var[Y]+2abCov[X,Y] \tag{2.19}$$

respectively.

In the special case $a=1$ and $b=-1$, Equations 2.18 and 2.19 give the mean and the variance of the *difference* between the two random variables. Thus, for any random variables X and Y, the mean of the difference $X-Y$ is given by

$$E[X-Y]=E[X]-E[Y] \tag{2.20}$$

and the variance of the difference $X-Y$ is given by

$$Var[X-Y]=Var[X]+Var[Y]-2Cov[X,Y] \tag{2.21}$$

If the random variables X and Y are independent, then $Cov[X, Y]=0$ and the variance of the difference $X-Y$ is given by

$$Var[X-Y]=Var[X]+Var[Y] \tag{2.22}$$

Equation 2.20 for the mean of the difference $X-Y$ holds even if the random variables are not independent. Note that the mean of the difference is simply the difference of the means (Equation 2.20), while the variance of the difference (for the case of independent random variables) is the *sum*

of the variances (Equation 2.22), i.e., the same variance as for the sum $X+Y$ (Equation 2.16).

The results in Equations 2.18 and 2.19 for a linear combination can be generalized to n random variables. For any random variables $X_1, X_2, \ldots, X_n$ and any constants $a_1, a_2, \ldots, a_n$, the mean and the variance of the linear combination $a_1X_1+a_2X_2+\cdots+a_nX_n$ are given by

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E[X_i] \tag{2.23}$$

and

$$Var\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2\, Var[X_i] + \sum_{i \neq j}\sum a_i a_j Cov[X_i, X_j] \tag{2.24}$$

respectively.

Bolch, Greiner, de Meer, and Trivedi (1998, pp. 23–24); Feller (1964, pp. 208, 214, 216); Mood, Graybill, and Boes (1974, pp. 178–179); Ross (2003, pp. 49, 53–54).

2.4 Mean and Variance of the Product of Two Random Variables

If X and Y are independent random variables, then the mean and the variance of the product XY are given by

$$E[XY] = E[X]E[Y] \tag{2.25}$$

and

$$Var[XY] = (E[Y])^2\, Var[X] + (E[X])^2\, Var[Y] + Var[X]Var[Y] \tag{2.26}$$

respectively.

If the random variables X and Y are not independent, then the mean and the variance of the product XY are given by

$$E[XY] = E[X]E[Y] + Cov[X, Y] \tag{2.27}$$

and

$$
\begin{aligned}
&Var[XY] = \\
&(E[Y])^2 Var[X] + (E[X])^2 Var[Y] + 2E[X]E[Y]Cov[X,Y] - (Cov[X,Y])^2 \\
&+ E\left\{(X - E[X])^2 (Y - E[Y])^2\right\} + 2E[Y]E\left\{(X - E[X])^2 (Y - E[Y])\right\} \\
&+ 2E[X]E\left\{(X - E[X])(Y - E[Y])^2\right\}
\end{aligned} \tag{2.28}
$$

respectively.

Mood, Graybill, and Boes (1974, p. 180).

2.5 Mean and Variance of the Quotient of Two Random Variables

If X and Y are independent random variables, then the approximate expressions for the mean and the variance of the quotient X/Y are given by

$$E\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)\left(1 + \frac{Var[Y]}{(E[Y])^2}\right) \tag{2.29}$$

and

$$Var\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)^2\left(\frac{Var[X]}{(E[X])^2} + \frac{Var[Y]}{(E[Y])^2}\right) \tag{2.30}$$

respectively.

If the random variables X and Y are not independent, then the approximate expressions for the mean and the variance of the quotient X/Y are given by

$$E\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)\left(1 + \frac{Var[Y]}{(E[Y])^2}\right) - \frac{1}{(E[Y])^2} Cov[X,Y] \tag{2.31}$$

and

$$Var\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)^2 \left(\frac{Var[X]}{(E[X])^2} + \frac{Var[Y]}{(E[Y])^2} - \frac{2Cov[X,Y]}{E[X]E[Y]}\right) \tag{2.32}$$

respectively.

These approximations for a quotient are obtained from the Taylor series expansions about the means $E[X]$ and $E[Y]$ up to second-order terms. Mood, Graybill, and Boes (1974, p. 181).

2.6 Conditional Mean and Variance for Jointly Distributed Random Variables

For jointly distributed random variables X and Y,

$$E[Y] = E_X\left\{E\left[Y|X\right]\right\} \tag{2.33}$$

$$Var[Y] = E_X\left\{Var\left[Y|X\right]\right\} + Var_X\left\{E\left[Y|X\right]\right\} \tag{2.34}$$

where

$E[Y]$ and $Var[Y]$ denote the unconditional mean and the variance of Y

$E[Y|X]$ and $Var[Y|X]$ denote the conditional mean and the variance of Y, given a value of X

$E_X[\cdot]$ and $Var_X[\cdot]$ denote the mean and the variance over the distribution of X, respectively

Mood, Graybill, and Boes (1974, pp. 158–159); Ross (1988, pp. 285, 292); Wolff (1989, pp. 32, 34).

2.7 Conditional Mean of a Constrained Random Variable

For a continuous random variable $X(-\infty < X < \infty)$ and any constant a, the conditional mean of X, given that X is greater than a, is given by

$$E\left[X|X > a\right] = \frac{\int_a^\infty xf(x)dx}{\Pr\{X > a\}} \tag{2.35}$$

$$= \frac{\int_a^{\infty} xf(x)dx}{\int_a^{\infty} f(x)dx} \tag{2.36}$$

$$= \frac{\int_a^{\infty} xf(x)dx}{1-F(a)} \tag{2.37}$$

where
$f(x)$ is the probability density function of X

and

$$F(x) = \Pr\{X \le x\} = \int_{-\infty}^{x} f(t)dt$$

denotes the cumulative distribution function of X.

More generally, for any constants a and b where $a<b$, the conditional mean of X, given that X lies between a and b, is given by

$$E[X \mid a < X < b] = \frac{\int_a^b xf(x)dx}{\Pr\{a < X < b\}} \tag{2.38}$$

$$= \frac{\int_a^b xf(x)dx}{\int_a^b f(x)dx} \tag{2.39}$$

$$= \frac{\int_a^b xf(x)dx}{F(b)-F(a)} \tag{2.40}$$

Cassady and Nachlas (2009, p. 69); Stirzaker (1994, p. 243).

2.8 Mean and Variance of the Sum of a Random Number of Random Variables

Let

$X_1, X_2, \ldots, X_N$ be N independent and identically distributed random variables,

where

N is a nonnegative integer random variable (independent of $X_1, X_2, \ldots, X_N$), and let

$E[X]$ and $Var[X]$ be the mean and the variance of X_i ($i = 1, 2, \ldots, N$)

$E[N]$ and $Var[N]$ be the mean and the variance of N, respectively

Then the sum

$$Y = X_1 + X_2 + \cdots + X_N$$

has a mean $E[Y]$ and a variance $Var[Y]$ given by

$$E[Y] = E[N]E[X] \tag{2.41}$$

$$Var[Y] = E[N]Var[X] + (E[X])^2 Var[N] \tag{2.42}$$

Benjamin and Cornell (1970, p. 179); Drake (1967, pp. 111–112); Mood, Graybill, and Boes (1974, p. 197); Ross (1983, p. 16); Ross (2003, pp. 107, 119); Wald (1947, p. 53).

2.9 Mean of a Function of a Random Variable

Let

X be a continuous random variable ($-\infty < X < \infty$)

$g(X)$ be a function of X

$f(x)$ be the probability density function of X

$F(x)$ be the cumulative distribution function of X

The function $g(X)$ is a random variable with a mean $E[g(x)]$ given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx = \int_{-\infty}^{\infty} g(x)\, dF(x) \tag{2.43}$$

If X and Y are independent random variables, then for any functions $g(\cdot)$ and $h(\cdot)$,

$$E\left[g(X)h(Y)\right] = E\left[g(X)\right]E\left[h(Y)\right] \tag{2.44}$$

Çinlar (1975, pp. 29–30); Lefebvre (2006, pp. 97–98); Mood, Graybill, and Boes (1974, p. 160); Ross (2003, pp. 45, 52).

2.10 Approximations for the Mean and Variance of a Function of a Random Variable

Let
X be a random variable $(-\infty < X < \infty)$
$g(X)$ be a function of X
$\mu = E[X]$ be the mean of X
$\sigma^2 = Var[X]$ be the variance of X

The mean and the variance of the function $g(X)$ are given in terms of the mean and the variance of X by the following approximations:

$$E\left[g(X)\right] \cong g(\mu) + \tfrac{1}{2}\sigma^2 g''(\mu) \tag{2.45}$$

$$Var\left[g(X)\right] \cong \sigma^2\left[g'(\mu)\right]^2 \tag{2.46}$$

where $g'(\mu)$ and $g''(\mu)$ denote the first and second derivatives of $g(x)$, respectively, evaluated at $x = \mu$, i.e.,

$$g'(\mu) = \frac{d}{dx}g(x)\bigg|_{x=\mu}$$

and

$$g''(\mu) = \frac{d^2}{dx^2}g(x)\bigg|_{x=\mu}$$

Benjamin and Cornell (1970, pp. 180–181); Papoulis (1984, p. 113).

2.11 Mean and Variance of the Maximum of Exponentially Distributed Random Variables

Let X_1, X_2, …, X_n be n independent and identically distributed random variables, each having an exponential distribution with a mean $1/\lambda$, i.e., a probability density function $f(x_i) = \lambda e^{-\lambda x_i}$ ($i = 1, 2, \ldots, n$). The mean and the variance of the maximum of the n random variables are given by

$$E\left[\max\left(X_1, X_2, \ldots, X_n\right)\right] = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \tag{2.47}$$

and

$$Var\left[\max\left(X_1, X_2, \ldots, X_n\right)\right] = \frac{1}{\lambda^2}\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}\right) \tag{2.48}$$

respectively.

Balakrishnan and Sinha (1995, p. 19); Cox and Hinkley (1974, p. 468); Nahmias (1989, p. 553).

Note that the ***minimum*** of n independent exponentially distributed random variables, each with a mean $1/\lambda$, has simply an exponential distribution with a mean $1/(n\lambda)$ and a variance $1/(n\lambda)^2$ (see Equation 5.15).

2.12 Mean and Variance of the Maximum of Normally Distributed Random Variables

Let

X_1 and X_2 be jointly normally distributed random variables, and let
$\mu_1 = E[X_1]$ and $\mu_2 = E[X_2]$ be the means of X_1 and X_2, respectively
$\sigma_1^2 = Var\left[X_1\right]$ and $\sigma_2^2 = Var\left[X_2\right]$ be the variances of X_1 and X_2, respectively
$\rho = Corr[X_1, X_2]$ be the correlation coefficient of X_1 and X_2

Assume $\sigma_1 \neq \sigma_2$ and $\rho \neq 1$, and let the parameters α and β be defined as

$$\beta^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho$$

and

$$\alpha = \frac{\mu_1 - \mu_2}{\beta}$$

Let the functions $\phi(x)$ and $\Phi(x)$ denote the probability density function and the cumulative distribution function, respectively, for the standard normal distribution given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt$$

Let $Z = \max(X_1, X_2)$ be the maximum of X_1 and X_2.

The means of Z and Z^2 are given by

$$E[Z] = \mu_1\Phi(\alpha) + \mu_2\Phi(-\alpha) + \beta\phi(\alpha) \tag{2.49}$$

and

$$E[Z^2] = (\mu_1^2 + \sigma_1^2)\Phi(\alpha) + (\mu_2^2 + \sigma_2^2)\Phi(-\alpha) + (\mu_1 + \mu_2)\beta\phi(\alpha) \tag{2.50}$$

respectively, and the variance of Z is given by

$$Var[Z] = E[Z^2] - (E[Z])^2 \tag{2.51}$$

Clark (1961, pp. 146–147).

Since $\phi(x)$ is symmetric about $x = 0$, the probabilities $\Phi(\alpha)$ and $\Phi(-\alpha)$ are related by

$$\Phi(-\alpha) = 1 - \Phi(\alpha)$$

and Clark's results (Equations 2.49 and 2.50) can be written as

$$E[Z] = \mu_1\Phi(\alpha) + \mu_2\{1 - \Phi(\alpha)\} + \beta\phi(\alpha) \tag{2.52}$$

and

$$E[Z^2] = (\mu_1^2 + \sigma_1^2)\Phi(\alpha) + (\mu_2^2 + \sigma_2^2)\{1 - \Phi(\alpha)\} + (\mu_1 + \mu_2)\beta\,\phi(\alpha) \tag{2.53}$$

Equation 2.52 shows that the mean of the maximum Z is given by a probability-weighted average of the means μ_1 and μ_2 of the normal

variables X_1 and X_2 separately, plus an adjustment term. Similarly, Equation 2.53 shows that the mean of Z^2 is given by a probability-weighted average of the means of X_1^2 and X_2^2 separately, plus an adjustment term.

The above exact formulas for the moments of the maximum of two normally distributed random variables can be used to obtain approximate expressions for the case of more than two normal random variables, as follows.

Let X_1, X_2, and Y be jointly normally distributed random variables, and let
$\rho_1 = Corr[X_1, Y]$ be the correlation coefficient of X_1 and Y
$\rho_2 = Corr[X_2, Y]$ be the correlation coefficient of X_2 and Y

The correlation coefficient of Y and Z is given by

$$Corr\left[Y, Z\right] = Corr\left[Y, \max\left(X_1, X_2\right)\right] = \frac{\sigma_1 \rho_1 \Phi(\alpha) + \sigma_2 \rho_2 \Phi(-\alpha)}{\sqrt{Var\left[Z\right]}} \tag{2.54}$$

$$= \frac{\sigma_1 \rho_1 \Phi(\alpha) + \sigma_2 \rho_2 \left\{1 - \Phi(\alpha)\right\}}{\sqrt{Var\left[Z\right]}} \tag{2.55}$$

The mean and the variance for the maximum of the three normal random variables, X_1, X_2, and Y, are obtained by expressing $\max(X_1, X_2, Y)$ as

$$\max\left(X_1, X_2, Y\right) = \max\left[Y,\ \max\left(X_1, X_2\right)\right] \tag{2.56}$$

and applying the above formulas for the mean and the variance in the two-variable case and the correlation of Y and $\max(X_1, X_2)$. The results for the three-variable case are approximate since $\max(X_1, X_2)$ is not normally distributed. This procedure for approximate results can be extended to any finite number of normal random variables.

Clark (1961, pp. 147–148).

3

Discrete Probability Distributions

3.1 Bernoulli Distribution

Let
p be a constant, where $0<p<1$
X be a random variable that can only take the values 0 or 1
$P(x)$ be the probability that $X=x$ $(x=0, 1)$

The random variable, X, has a Bernoulli distribution if $P(x)$ is given by

$$P(x)=\begin{cases} p & \text{for } x=1 \\ 1-p & \text{for } x=0 \end{cases} \tag{3.1}$$

Figure 3.1 shows an example of the Bernoulli distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the Bernoulli distribution are given by

$$E[X]=p \tag{3.2}$$

and

$$Var[X]=p(1-p) \tag{3.3}$$

respectively.

Ayyub and McCuen (1997, p. 91); Hoel, Port, and Stone (1971, pp. 66, 83); Mood, Graybill, and Boes (1974, p. 87).

Note that the Bernoulli distribution $P(x)$ $(x=0, 1)$ is used to characterize a random experiment with two possible outcomes. The outcomes are generally referred to as "success" $(x=1)$ and "failure" $(x=0)$, with probabilities p and $1-p$, respectively.

Bernoulli trials. Repeated random experiments that are independent and have two possible outcomes with constant probabilities are called *Bernoulli trials*.

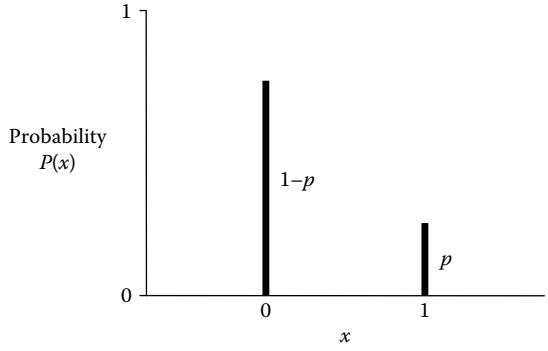


FIGURE 3.1
Example of the Bernoulli distribution.

3.2 Binomial Distribution

Let
N be a positive integer
p be a constant, where $0<p<1$
X be a random variable that can take the values 0, 1, 2, …, N
$P(x)$ be the probability that $X=x$ ($x=0, 1, 2, \ldots, N$)

The random variable X has a binomial distribution if $P(x)$ is given by

$$P(x)=\binom{N}{x}p^x(1-p)^{N-x} \quad (x=0, 1, 2, \ldots, N) \tag{3.4}$$

The term $\binom{N}{x}$ denotes the number of combinations of x objects selected from a total of N objects, and is given by

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

Figure 3.2 shows an example of the binomial distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the binomial distribution are given by

$$E[X]=Np \tag{3.5}$$

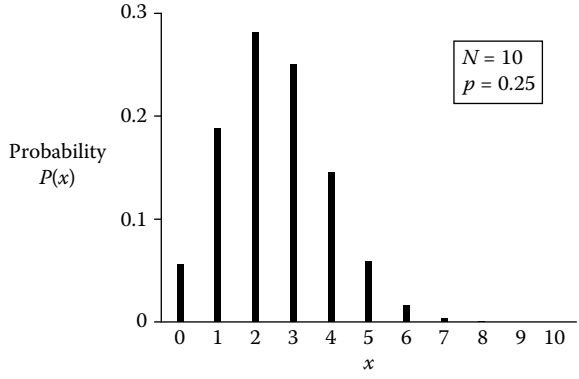


FIGURE 3.2
Example of the binomial distribution.

and

$$Var[X] = Np(1-p) \tag{3.6}$$

respectively.

Ayyub and McCuen (1997, p. 92); Feller (1964, pp. 137, 209, 214); Hoel, Port, and Stone (1971, pp. 51, 83, 97–98); Mood, Graybill, and Boes (1974, pp. 88–89).

The binomial distribution $P(x)$ ($x=0, 1, 2, \ldots, N$) gives the probability of x successes out of N Bernoulli trials, where each trial has a probability p of success and a probability $(1-p)$ of failure.

In the special case $N=1$, the binomial distribution reduces to the Bernoulli distribution. For the general positive integer N, the sum of N Bernoulli random variables (i.e., the sum of random variables that take the values 0 or 1 in N Bernoulli trials) has a binomial distribution.

The probabilities, $P(x)$, for each x ($x=0, 1, 2, \ldots, N$), given by Equation 3.4, are the successive terms in the binomial expansion of $[(1-p)+p]^N$. Since $[(1-p)+p]^N=1$ for any p and N, the sum of the terms in the expansion is equal to 1 (i.e., $\sum_{x=0}^{N} P(x)=1$), as required for $P(x)$ to be a probability distribution.

For any N, the ratio of the variance to the mean for the binomial distribution is

$$\frac{Var[X]}{E[X]} = (1-p) < 1 \tag{3.7}$$

3.3 Geometric Distribution

Let
p be a constant, where $0<p<1$
X be a random variable that can take the values 0, 1, 2, …
$P(x)$ be the probability that $X=x$ ($x=0, 1, 2, \ldots$)

The random variable X has a geometric distribution if $P(x)$ is given by

$$P(x) = p(1-p)^x \quad (x = 0, 1, 2, \ldots) \tag{3.8}$$

Figure 3.3 shows an example of the geometric distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the geometric distribution are given by

$$E[X] = \frac{1-p}{p} \tag{3.9}$$

and

$$Var[X] = \frac{1-p}{p^2} \tag{3.10}$$

respectively.

DeGroot (1986, pp. 260–261); Hoel, Port, and Stone (1971, pp. 55, 84–85, 96); Mood, Graybill, and Boes (1974, pp. 99–100).

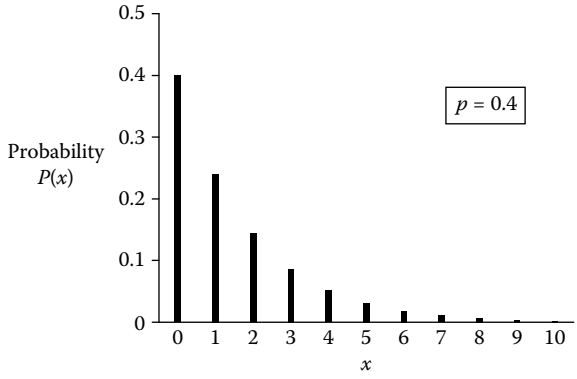


FIGURE 3.3
Example of the geometric distribution.

The geometric distribution $P(x)$ ($x=0, 1, 2, \ldots$) gives the probability of x trials (or failures) occurring before the first success in an unlimited sequence of Bernoulli trials, where each trial has a probability p of success and a probability $(1-p)$ of failure.

Note that the geometric random variable X is sometimes defined as the number of trials needed to achieve the first success (rather than the number of trials *before* the first success) in an unlimited sequence of Bernoulli trials. Under this definition, X can take the values 1, 2, … (but not 0), and the distribution for $P(x)$ is given by $P(x)=p(1-p)^{x-1}$ ($x=1, 2, \ldots$). The mean in this case is $E[X]=1/p$, while the variance remains the same as before, $Var[X]=(1-p)/p^2$.

The probabilities $P(x)$ for each x ($x=0, 1, 2, \ldots$), given by Equation 3.8, are the successive terms in the infinite geometric series

$$p+p\left(1-p\right)+p\left(1-p\right)^2+p\left(1-p\right)^3+\cdots$$

The sum of this series is

$$\frac{p}{\left[1-\left(1-p\right)\right]}=1$$

i.e.,

$$\sum_{x=0}^{\infty}P\left(x\right)=1$$

as required for $P(x)$ to be a probability distribution.

The probability that the geometric random variable X is less than or equal to a nonnegative integer k is given by

$$\Pr\left\{X\le k\right\}=\sum_{x=0}^{k}P\left(x\right)=1-\left(1-p\right)^{k+1}$$

The probability that X is greater than k is given by

$$\Pr\left\{X>k\right\}=\left(1-p\right)^{k+1}$$

The geometric distribution has the property that, for the nonnegative integers k and m, the conditional probability that $X>k+m$, given that $X>k$, is

equal to the unconditional probability that $X>m$ (Hoel, Port, and Stone, 1971, p. 59), i.e.,

$$\Pr\{X > k+m \mid X > k\} = \Pr\{X > m\} \tag{3.11}$$

This is the "lack of memory" property (also known as the "memoryless" property). The geometric distribution is the discrete counterpart to the continuous exponential distribution, which also has the lack of memory property.

3.4 Negative Binomial Distribution

Let
r be a constant, where $0<r<\infty$
p be a constant, where $0<p<1$
X be a random variable that can take the values 0, 1, 2, …
$P(x)$ be the probability that $X=x$ ($x=0, 1, 2, \ldots$)

The random variable X has a negative binomial distribution if $P(x)$ is given by

$$P(x) = \binom{r+x-1}{x} p^r (1-p)^x \quad (x = 0, 1, 2, \ldots) \tag{3.12}$$

or, in its alternative form,

$$P(x) = \binom{-r}{x} (-1)^x p^r (1-p)^x \quad (x = 0, 1, 2, \ldots) \tag{3.13}$$

The terms

$$\binom{r+x-1}{x}$$

and

$$\binom{-r}{x}(-1)^x$$

are given by

$$\binom{r+x-1}{x} = \binom{-r}{x}(-1)^x = \frac{r(r+1)\cdots(r+x-1)}{x!} \tag{3.14}$$

for $x=1, 2, \ldots$, and are equal to 1 for $x=0$.

Figure 3.4 shows an example of the negative binomial distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the negative binomial distribution are given by

$$E[X] = \frac{r(1-p)}{p} \tag{3.15}$$

and

$$Var[X] = \frac{r(1-p)}{p^2} \tag{3.16}$$

respectively.

DeGroot (1986, pp. 259, 261); Feller (1964, pp. 155, 210, 253); Hoel, Port, and Stone (1971, pp. 55–56, 95–96); Mood, Graybill, and Boes (1974, pp. 99, 102).

The negative binomial distribution, $P(x)$, is defined only for nonnegative integer values of x ($x=0, 1, 2, \ldots$). The constant, r, may be any positive number, not necessarily an integer.

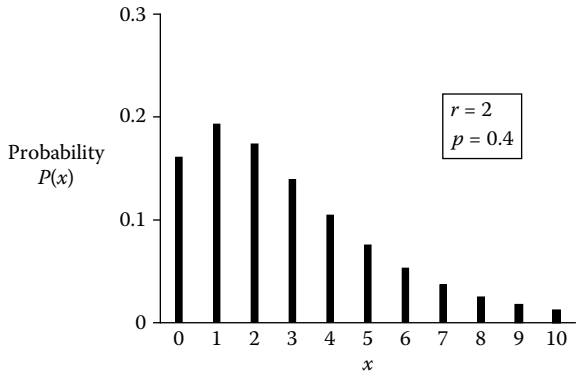


FIGURE 3.4
Example of the negative binomial distribution.

If r is an integer, the negative binomial distribution $P(x)$ gives the probability of x failures occurring before the rth success in an unlimited sequence of Bernoulli trials, where each trial has a probability p of success and a probability $(1-p)$ of failure. The negative binomial distribution with r as integer is sometimes called the Pascal distribution.

In the special case $r=1$, the negative binomial distribution reduces to the geometric distribution. For the general positive integer, r, the sum of r independent and identically distributed geometric random variables has a negative binomial distribution. Thus, if X_1, X_2, …, X_r are r independent random variables where each has a geometric distribution $P(x_i) = p(1-p)^{x_i}$ $(i=1, 2, \ldots, r)$, then the sum $X=X_1+X_2+\cdots+X_r$ has a negative binomial distribution, given by

$$P(x)=\binom{r+x-1}{x}p^r(1-p)^x$$

The probabilities, $P(x)$, for each x $(x=0, 1, 2, \ldots)$, given by Equation 3.12 or Equation 3.13, are equal to p^r multiplied by the successive terms in the binomial expansion of $[1-(1-p)]^{-r}$. Since $p^r[1-(1-p)]^{-r}=1$ for any p and r, the sum $\sum_{x=0}^{\infty} P(x)=1$, is required for $P(x)$ to be a probability distribution.

For any r, the ratio of the variance to the mean for the negative binomial distribution is

$$\frac{Var[X]}{E[X]}=\frac{1}{p}>1 \tag{3.17}$$

3.5 Poisson Distribution

Let

μ be a constant, where $0<\mu<\infty$

X be a random variable that can take the values 0, 1, 2, …

$P(x)$ be the probability that $X=x$ $(x=0, 1, 2, \ldots)$

The random variable X has a Poisson distribution if $P(x)$ is given by

$$P(x)=\frac{e^{-\mu}\mu^x}{x!} \quad (x=0, 1, 2, \ldots) \tag{3.18}$$

Figure 3.5 shows an example of the Poisson distribution.

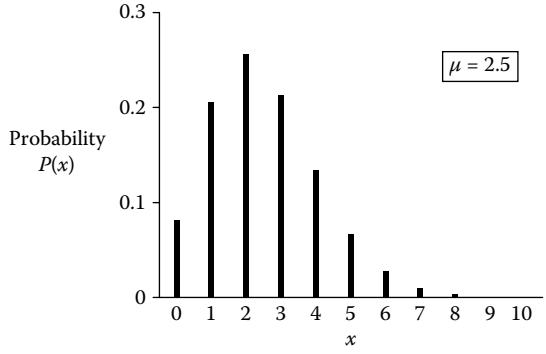


FIGURE 3.5
Example of the Poisson distribution.

Note that, for the Poisson distribution, the cumulative distribution function (i.e., the probability that $x \le n$, for any integer n) can be expressed as

$$\sum_{x=0}^{n} P(x) = e^{-\mu} \sum_{x=0}^{n} \frac{\mu^x}{x!} = 1 - \frac{\gamma(n+1, \mu)}{n!}$$

where $\gamma(\cdot,\cdot)$ denotes the incomplete gamma function (see Chapter 15).

The mean, $E[X]$, and the variance, $Var[X]$, for the Poisson distribution are given by

$$E[X] = \mu \tag{3.19}$$

and

$$Var[X] = \mu \tag{3.20}$$

respectively.

DeGroot (1986, pp. 252–253); Feller (1964, pp. 146, 209, 210, 214); Hoel, Port, and Stone (1971, pp. 56, 84, 96); Mood, Graybill, and Boes (1974, pp. 93–94).

Successive values of the Poisson distribution $P(x)$ ($x = 0, 1, 2, \ldots$) can be conveniently computed from the relationships.

$$\left.\begin{aligned} P(0) &= e^{-\mu} \\ P(x+1) &= \frac{\mu P(x)}{x+1} \end{aligned}\right\} \tag{3.21}$$

Evans, Hastings, and Peacock (2000, p. 157).

The relationships given in Equation 3.21 help to avoid overflow or underflow problems that can occur in computing $P(x)$ directly from Equation 3.18 for large values of x.

For any μ, the ratio of the variance to the mean for the Poisson distribution is

$$\frac{Var[X]}{E[X]} = 1 \tag{3.22}$$

The Poisson distribution with parameter μ is the limiting form of the binomial distribution with parameters N and p, as N becomes large and p becomes small in such a way that the product Np remains fixed and equal to μ (DeGroot, 1986, pp. 256–257; Hoel, Port, and Stone, 1971, p. 69), i.e., for $p = \mu/N$,

$$\lim_{N\to\infty} \binom{N}{x} p^x (1-p)^{N-x} = \frac{e^{-\mu}\mu^x}{x!} \tag{3.23}$$

In the case where the parameter μ in a Poisson distribution is a continuous random variable rather than a constant, the combination of the Poisson distribution with a gamma distribution for μ results in a negative binomial distribution (see Section 5.5).

3.6 Hypergeometric Distribution

Let

N be a positive integer
K be a positive integer, where $K \le N$
n be a positive integer, where $n \le N$
X be a random variable that can take the values 0, 1, 2, …, n
$P(x)$ be the probability that $X = x$ ($x = 0, 1, 2, \ldots, n$)

The random variable, X, has a hypergeometric distribution if $P(x)$ is given by

$$P(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \quad (x = 0, 1, 2, \ldots, n) \tag{3.24}$$

Terms of the form $\binom{a}{b}$ denote the numbers of combinations of b objects selected from a total of a objects, and are given by

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}$$

Figure 3.6 shows an example of the hypergeometric distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the hypergeometric distribution are given by

$$E[X] = \frac{nK}{N} \tag{3.25}$$

and

$$Var[X] = \left(\frac{nK}{N}\right)\left(1 - \frac{K}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{3.26}$$

respectively.

DeGroot (1986, pp. 247–250); Freund (1992, pp. 199–202); Hoel, Port, and Stone (1971, pp. 52, 90, 98); Mood, Graybill, and Boes (1974, pp. 91–92).

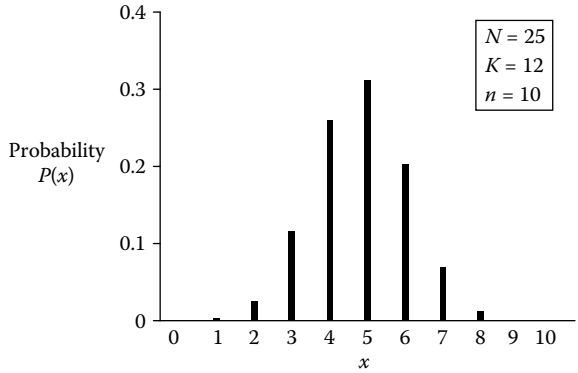


FIGURE 3.6
Example of the hypergeometric distribution.

The hypergeometric distribution arises in sampling from a finite population. Consider a population of N objects in total, of which K objects ($K \le N$) are of a specific type (referred to as "successes"), and suppose that a random sample of size n is selected without replacement from the N objects in the population ($n \le N$). The hypergeometric distribution $P(x)$ ($x = 0, 1, 2, \ldots, n$) gives the probability of x successes out of the n objects in the sample.

The number of combinations of x successes from the total of K successes and $(n-x)$ objects from the remaining $(N-K)$ objects is $\binom{K}{x}\binom{N-K}{n-x}$. The number of combinations of any n objects from the total of N objects is $\binom{N}{n}$. The ratio of these numbers gives the probability of x successes in the sample of size n, i.e.,

$$\binom{K}{x}\binom{N-K}{n-x} \Big/ \binom{N}{n}$$

as given by Equation 3.24.

If the objects in the sample were selected with replacement, rather than without replacement, then the probability of selecting a success would be a constant p, given by $p = K/N$, and the probability of x successes in the sample of size n would be given by the binomial distribution with parameters n and p, i.e.,

$$\binom{n}{x} p^x (1-p)^{n-x}$$

If the population size N is large when compared to the sample size n, then there is little difference between sampling with and without replacement, and the hypergeometric distribution with parameters n, N, and K, can be approximated in this case by the binomial distribution with parameters n and $p = K/N$. In general, the hypergeometric distribution has the same mean as the binomial distribution (i.e., np), but a smaller variance. The variance for the hypergeometric distribution is

$$np(1-p)\left(\frac{N-n}{N-1}\right)$$

while the variance for the binomial distribution is $np(1-p)$. As N becomes large, the factor $\left(\frac{N-n}{N-1}\right)$ approaches 1, and the variance for the hypergeometric distribution becomes approximately equal to the variance for the binomial distribution.

3.7 Multinomial Distribution

Let

N be a positive integer

k be a positive integer

$p_1, p_2, \ldots, p_k$ be constants, where $0<p_i<1$ ($i=1, 2, \ldots, k$) and $p_1+p_2+\cdots+p_k=1$

$X_1, X_2, \ldots, X_k$ be random variables that can take the values 0, 1, 2, …, N, subject to the constraint $X_1+X_2+\cdots+X_k=N$

$P(x_1, x_2, \ldots, x_k)$ be the joint probability $\Pr(X_1=x_1, X_2=x_2, \ldots, X_k=x_k)$

The random variables $X_1, X_2, \ldots, X_k$ have a multinomial distribution if $P(x_1, x_2, \ldots, x_k)$ is given by

$$P(x_1, x_2, \ldots, x_k) = \frac{N!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \quad (x_i = 0, 1, 2, \ldots, N; \quad i = 1, 2, \ldots, k) \tag{3.27}$$

where $x_1+x_2+\cdots+x_k=N$ and $p_1+p_2+\cdots+p_k=1$.

DeGroot (1986, pp. 297–298); Freund (1992, pp. 216–217); Hoel, Port, and Stone (1971, pp. 68–69); Mood, Graybill, and Boes (1974, pp. 137–138).

The multinomial distribution is a multivariate generalization of the binomial distribution. It arises in repeated independent random experiments, where each experiment has k possible outcomes. Suppose that the outcomes are labeled 1, 2, …, k, and occur with probabilities $p_1, p_2, \ldots, p_k$, respectively, where $p_1+p_2+\cdots+p_k=1$. The multinomial distribution $P(x_1, x_2, \ldots, x_k)$ gives the probability that, out of a total of N experiments, x_1 are of outcome 1, x_2 are of outcome 2, …, and x_k are of outcome k.

The probabilities $P(x_1, x_2, \ldots, x_k)$ for $x_i=0, 1, \ldots, N$ ($i=1, 2, \ldots, k$), given by Equation 3.27, are the terms in the expansion of $(p_1+p_2+\cdots+p_k)^N$. Since $p_1+p_2+\cdots+p_k=1$, the sum of the terms in the expansion is equal to 1, i.e.,

$$\sum_{\substack{x_1, x_2, \ldots, x_k \\ x_1+x_2+\cdots+x_k=N}} P(x_1, x_2, \ldots, x_k) = 1$$

as required for $P(x_1, x_2, \ldots, x_k)$ to be a probability distribution.

In the special case $k=2$, the multinomial distribution reduces to the binomial distribution. The multinomial distribution probability $\Pr(X_1=x_1, X_2=x_2)$ in this case is given by

$$P(x_1, x_2) = \frac{N!}{x_1! x_2!} p_1^{x_1} p_2^{x_2}$$

where $x_1+x_2=N$ and $p_1+p_2=1$. Introducing a single random variable x and a single parameter p, and writing $x_1=x$ and $x_2=N-x$, and $p_1=p$ and $p_2=1-p$, this probability becomes

$$P(x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} \quad (x = 0, 1, 2, \ldots, N)$$

which is the standard form for the binomial distribution, as given by Equation 3.4.

The marginal distribution of each random variable X_i ($i=1, 2, \ldots, k$) in the multinomial distribution is a binomial distribution with parameters N and p_i. The mean and the variance of each X_i is given by

$$E[X_i] = Np_i \tag{3.28}$$

and

$$Var[X_i] = Np_i(1-p_i) \tag{3.29}$$

respectively.

4

Continuous Probability Distributions

4.1 Uniform Distribution

Let

a and b be constants, where $b>a$

X be a random variable that can take any value in the range $[a, b]$

$f(x)$ be the probability density function of X $(a \le x \le b)$

$F(x)$ be the cumulative distribution function of X $(a \le x \le b)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_a^x f(t)dt \quad (a \le x \le b)$$

The random variable, X, has a uniform distribution if $f(x)$ is given by

$$f(x) = \frac{1}{b-a} \quad (a \le x \le b) \tag{4.1}$$

Figure 4.1 shows the probability density function, $f(x)$, for the uniform distribution.

The cumulative distribution function, $F(x)$, for the uniform distribution is given by

$$F(x) = \frac{x-a}{b-a} \quad (a \le x \le b) \tag{4.2}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the uniform distribution are given by

$$E[X] = \frac{a+b}{2} \tag{4.3}$$

and

$$Var[X] = \frac{(b-a)^2}{12} \tag{4.4}$$

respectively.

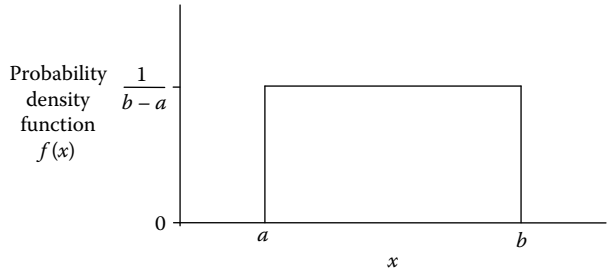


FIGURE 4.1
Example of the uniform distribution.

Allen (1978, pp. 80–81); Freund (1992, p. 223); Hoel, Port, and Stone (1971, pp. 118, 173); Mood, Graybill, and Boes (1974, pp. 105–106).

The uniform distribution is also known as the *rectangular distribution*. In the special case $a=0$ and $b=1$, the probability density function and the cumulative distribution function are simply

$$f(x)=1 \quad (0 \le x \le 1)$$

$$F(x)=x \quad (0 \le x \le 1)$$

respectively.

4.2 Exponential Distribution

Let

λ be a constant, where $\lambda>0$
X be a random variable that can take any value in the range $[0, \infty)$
$f(x)$ be the probability density function of X $(0 \le x < \infty)$
$F(x)$ be the cumulative distribution function of X $(0 \le x < \infty)$, i.e.,

$$F(x)=\Pr\{X \le x\}=\int_0^x f(t)dt$$

The random variable, X, has an exponential distribution if $f(x)$ is given by

$$f(x)=\lambda e^{-\lambda x} \quad (0 \le x < \infty) \tag{4.5}$$

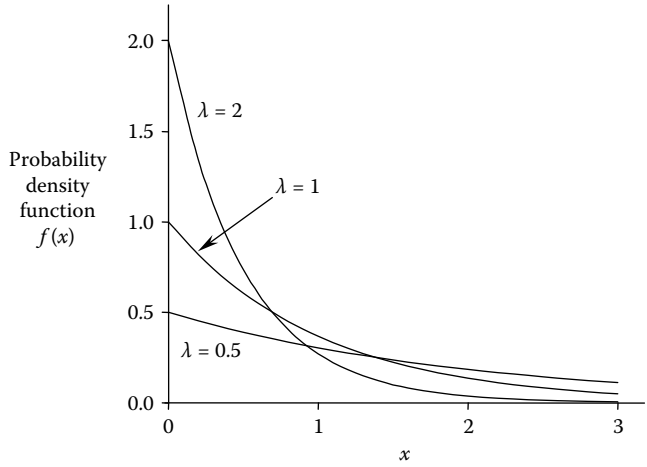


FIGURE 4.2
Examples of the exponential distribution.

Figure 4.2 shows examples of the probability density function, $f(x)$, for the exponential distribution.

The cumulative distribution function, $F(x)$, for the exponential distribution is given by

$$F(x) = 1 - e^{-\lambda x} \quad (0 \le x < \infty) \tag{4.6}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the exponential distribution are given by

$$E[X] = \frac{1}{\lambda} \tag{4.7}$$

and

$$Var[X] = \frac{1}{\lambda^2} \tag{4.8}$$

respectively.

Allen (1978, pp. 82, 85); DeGroot (1986, pp. 289–290); Freund (1992, pp. 225, 228); Hoel, Port, and Stone (1971, pp. 126, 174, 177); Mood, Graybill, and Boes (1974, p. 112).

The probability that the exponential random variable X is greater than x is given by

$$\Pr\{X > x\} = 1 - F(x) = e^{-\lambda x}$$

The exponential distribution has the property that, for any $s \geq 0$ and $t \geq 0$, the conditional probability that $X > s+t$, given that $X > s$, is equal to the unconditional probability that $X > t$, i.e.,

$$\Pr\{X > s+t \,|\, X > s\} = \Pr\{X > t\} \tag{4.9}$$

Allen (1978, pp. 82–83); Hoel, Port, and Stone (1971, p. 127); Krishnan (2006, p. 81); Mood, Graybill, and Boes (1974, p. 114). This is the "lack of memory" property (or "memoryless" property). The geometric distribution (the discrete counterpart to the exponential distribution) has the same property.

The standard deviation, *St Dev* [X], for the exponential distribution is

$$St\,Dev[X] = \sqrt{Var[X]} = \frac{1}{\lambda} \tag{4.10}$$

and the coefficient of variation for the exponential distribution is

$$Coeff\ of\ Var = \frac{St\,Dev[X]}{E[X]} = 1 \tag{4.11}$$

4.3 Erlang Distribution

Let

λ be a constant, where $\lambda > 0$

k be a positive integer

X be a random variable that can take any value in the range $[0, \infty)$

$f(x)$ be the probability density function of X $(0 \leq x < \infty)$

$F(x)$ be the cumulative distribution function of X $(0 < x < \infty)$, i.e.,

$$F(x) = \Pr\{X \leq x\} = \int_0^x f(t)dt$$

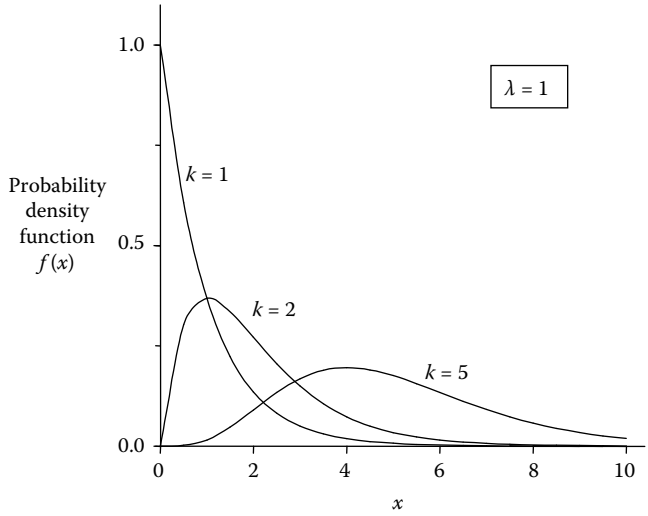


FIGURE 4.3
Examples of the Erlang distribution.

The random variable, X, has an Erlang distribution if $f(x)$ is given by

$$f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \quad (0 \le x < \infty) \tag{4.12}$$

Figure 4.3 shows examples of the probability density function $f(x)$ for the Erlang distribution.

The cumulative distribution function, $F(x)$, for the Erlang distribution is given by

$$F(x) = 1 - e^{-\lambda x} \sum_{i=0}^{k-1} \frac{(\lambda x)^i}{i!} \quad (0 \le x < \infty) \tag{4.13}$$

Note that Equation 4.13 can be expressed as

$$F(x) = \frac{\gamma(k, \lambda x)}{(k-1)!}$$

where $\gamma(\cdot,\cdot)$ denotes the incomplete gamma function (see Chapter 15).

The mean, $E[X]$, and the variance, $Var[X]$, for the Erlang distribution are given by

$$E[X] = \frac{k}{\lambda} \tag{4.14}$$

and

$$Var[X] = \frac{k}{\lambda^2} \tag{4.15}$$

respectively.

Çinlar (1975, pp. 81, 83); Law and Kelton (1991, p. 332); Tijms (1986, p. 395).

The constant, λ, is the scale parameter, and the integer, k, is the shape parameter. The Erlang distribution with shape parameter, k, is sometimes denoted by Erlang-k or E_k.

The Erlang distribution is a special case of the gamma distribution (which can have a non-integer shape parameter). The gamma distribution is described in Section 4.4.

The standard deviation, $St\ Dev[X]$, for the Erlang distribution is

$$St\,Dev[X] = \sqrt{Var[X]} = \frac{\sqrt{k}}{\lambda} \tag{4.16}$$

and the coefficient of variation for the Erlang distribution is

$$Coeff\ of\ Var = \frac{St\,Dev[X]}{E[X]} = \frac{1}{\sqrt{k}} \tag{4.17}$$

In the special case $k=1$, the Erlang distribution reduces to the exponential distribution. For the general positive integer k, the sum of k independent and identically distributed exponential random variables has an Erlang distribution. Thus, if X_1, X_2, …, X_k are k independent random variables where each has an exponential distribution with a mean $1/\lambda$, i.e., a probability density function $\lambda e^{-\lambda x_i}$ $(0 \le x_i < \infty,\ i = 1, 2, \ldots, k)$, then the sum $X = X_1 + X_2 + \cdots + X_k$ has an Erlang distribution with a probability density function, $f(x)$, given by

$$f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \quad (0 \le x < \infty)$$

as in Equation 4.12.

Likewise, the sum of m independent random variables that have Erlang distributions, with different shape parameters $k_1, k_2, \ldots, k_m$ and common scale parameter λ, also has an Erlang distribution. The resulting shape parameter, $k_1+k_2+\cdots+k_m$, since this case is the same as the sum of $k_1+k_2+\cdots+k_m$ independent and identically distributed exponential random variables. Thus, if $X_1, X_2, \ldots, X_m$ are m independent random variables where each has an Erlang distribution, with parameters λ and k_i ($i=1, 2, \ldots, m$), then the sum $S=X_1+X_2+\cdots+X_m$ has an Erlang distribution with a probability density function, $f(s)$, given by

$$f(s)=\frac{\lambda^{(k_1+k_2+\cdots+k_m)}}{(k_1+k_2+\cdots+k_m-1)!}s^{(k_1+k_2+\cdots+k_m-1)}e^{-\lambda s} \quad (0\le s<\infty)$$

The probability density function, $f(x)$, for the Erlang distribution is sometimes expressed as

$$f(x)=\frac{(\theta k)^k}{(k-1)!}x^{k-1}e^{-\theta kx}$$

with a scale parameter, θ, rather than λ, where $\theta=\lambda/k$. With the distribution expressed in terms of these parameters, the mean is given by $E[X]=\frac{1}{\theta}$ and is thus the same for any value of the shape parameter k. The variance in this case is given by $Var[X]=\frac{1}{k\theta^2}$.

4.4 Gamma Distribution

Let

λ and α be constants, where $\lambda>0$ and $\alpha>0$

X be a random variable that can take any value in the range $(0, \infty)$

$f(x)$ be the probability density function of X $(0<x<\infty)$

$F(x)$ be the cumulative distribution function of X $(0<x<\infty)$, i.e.,

$$F(x)=\Pr\{X\le x\}=\int_0^x f(t)dt$$

The random variable X has a gamma distribution if $f(x)$ is given by

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad (0 < x < \infty) \tag{4.18}$$

where $\Gamma(\alpha)$ is a gamma function, given by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

Figure 4.4 shows examples of the probability density function $f(x)$ for the gamma distribution.

The cumulative distribution function $F(x)$ for the gamma distribution is given by

$$F(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\lambda x} t^{\alpha-1} e^{-t} dt \quad (0 < x < \infty) \tag{4.19}$$

where the integral $\int_0^{\lambda x} t^{\alpha-1} e^{-t}\, dt$ is an incomplete gamma function $\gamma(\alpha, \lambda x)$ (see Chapter 15).

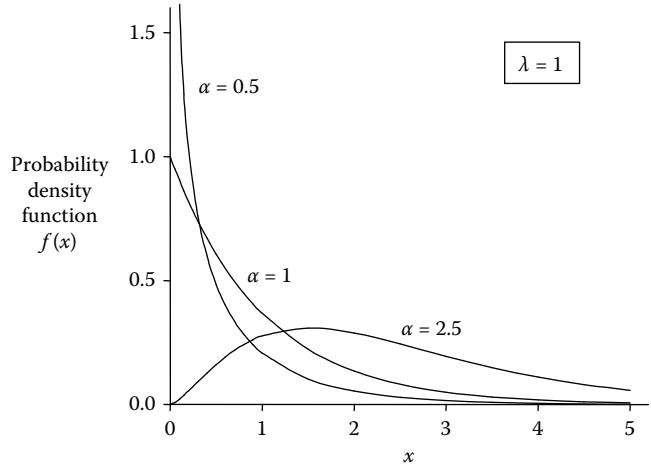


FIGURE 4.4
Examples of the gamma distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the gamma distribution are given by

$$E[X] = \frac{\alpha}{\lambda} \tag{4.20}$$

and

$$Var[X] = \frac{\alpha}{\lambda^2} \tag{4.21}$$

respectively.

DeGroot (1986, pp. 288–289); Freund (1992, pp. 225, 228); Hoel, Port, and Stone (1971, pp. 129, 174, 177); Mood, Graybill, and Boes (1974, pp. 112–113); Tijms (1986, pp. 394–395).

The constant, λ, is the scale parameter, and the constant, α, is the shape parameter. The standard deviation, $St\ Dev[X]$, for the gamma distribution is

$$St\,Dev[X] = \sqrt{Var[X]} = \frac{\sqrt{\alpha}}{\lambda} \tag{4.22}$$

and the coefficient of variation for the gamma distribution is

$$Coeff\ of\ Var = \frac{St\,Dev[X]}{E[X]} = \frac{1}{\sqrt{\alpha}} \tag{4.23}$$

From Equations 4.20 and 4.21, the parameters λ and α can be expressed in terms of the mean and the variance, and are given by

$$\lambda = \frac{E[X]}{Var[X]} \tag{4.24}$$

and

$$\alpha = \frac{(E[X])^2}{Var[X]} \tag{4.25}$$

respectively.

In the special case $\alpha=k$, where k is an integer, the gamma distribution is known as the Erlang distribution, as described in the previous section. The cumulative distribution function, $F(x)$, for this special case is given by Equation 4.13.

In the special case $\alpha=1$, the gamma distribution reduces to the exponential distribution, which is defined in Section 4.2.

In the special case $\lambda=1/2$ and $\alpha=\nu/2$, where ν is an integer, the gamma distribution is known as the χ^2 (chi-squared) distribution with ν degrees of freedom. The χ^2 distribution arises in statistical inference. It is the distribution of the sum of the squares of ν independent standard normal random variables.

Allen (1978, pp. 97–98); Mood, Graybill, and Boes, (1974, pp. 241–243).

Thus, if $Z_1, Z_2, \ldots, Z_\nu$ are ν independent random variables where each has a standard normal distribution, i.e., a probability density function $\phi(z_i)$ given by

$$\phi(z_i)=\frac{1}{\sqrt{2\pi}}e^{-z_i^2/2} \quad (-\infty < z_i < \infty, \quad i=1,2,\ldots,\nu)$$

then the sum $X=Z_1^2+Z_2^2+\cdots+Z_v^2$ has a χ^2 distribution with v degrees of freedom, i.e., a probability density function, $f(x)$, given by

$$f(x)=\frac{(1/2)^{v/2}}{\Gamma(v/2)}x^{(v/2)-1}e^{-x/2} \quad (0 \le x < \infty)$$

From Equations 4.20 and 4.21, the mean and the variance for the χ^2 distribution are given by $E[X]=\nu$ and $Var[X]=2\nu$, respectively.

The sum of m independent random variables that have gamma distributions, with different shape parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$ and a common scale parameter λ, has a gamma distribution with a shape parameter $\alpha_1+\alpha_2+\cdots+\alpha_m$ and a scale parameter λ. Thus, if $X_1, X_2, \ldots, X_m$ are m independent random variables where each has a gamma distribution, with parameters λ and α_i ($i=1, 2, \ldots, m$), then the sum $S=X_1+X_2+\cdots+X_m$ has a gamma distribution with a probability density function, $f(s)$, given by

$$f(s)=\frac{\lambda^{(\alpha_1+\alpha_2+\cdots+\alpha_m)}}{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_m)}s^{(\alpha_1+\alpha_2+\cdots+\alpha_m-1)}e^{-\lambda s} \quad (0 \le s < \infty)$$

4.5 Beta Distribution

Let
α and β be constants, where $\alpha>0$ and $\beta>0$
X be a random variable that can take any value in the range (0, 1)
$f(x)$ be the probability density function of X $(0<x<1)$
$F(x)$ be the cumulative distribution function of X $(0<x<1)$, i.e.,

$$F(x)=\Pr\{X\le x\}=\int_0^x f(t)dt$$

The random variable X has a beta distribution if $f(x)$ is given by

$$f(x)=\frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}\quad(0<x<1) \tag{4.26}$$

where $\mathrm{B}(\alpha, \beta)$ is a beta function, given by

$$\mathrm{B}(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

The beta function is related to the gamma function by

$$\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Figure 4.5 shows examples of the probability density function, $f(x)$, for the beta distribution.

The cumulative distribution function, $F(x)$, for the beta distribution is given by

$$F(x)=\frac{1}{\mathrm{B}(\alpha,\beta)}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt\quad(0<x<1) \tag{4.27}$$

where the integral $\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$ is the incomplete beta function (see Chapter 15).

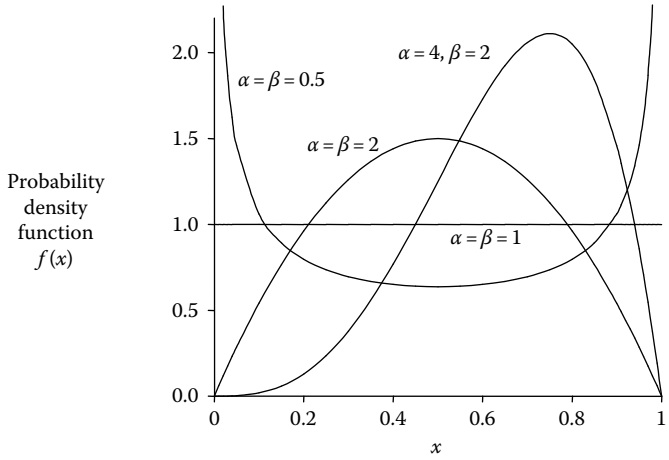


FIGURE 4.5
Examples of the beta distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the beta distribution are given by

$$E[X] = \frac{\alpha}{\alpha + \beta} \tag{4.28}$$

and

$$Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4.29}$$

respectively.

DeGroot (1986, pp. 294–296); Freund (1992, pp. 229–231); Mood, Graybill, and Boes (1974, pp. 115–116).

If α and β are integers, then the beta function, $B(\alpha, \beta)$, is given by

$$B(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$$

and the probability density function, $f(x)$, for the beta distribution becomes

$$f(x) = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} x^{\alpha-1}(1-x)^{\beta-1} \tag{4.30}$$

In the special case $\alpha=1$ and $\beta=1$, the beta distribution reduces to a uniform distribution (see Section 4.1), with a probability density function $f(x)=1$ $(0\le x\le 1)$.

4.6 Normal Distribution

Let

μ be any constant

σ be a constant, where $\sigma>0$

X be a random variable that can take any value in the range $(-\infty, \infty)$

$f(x)$ be the probability density function of X $(-\infty<x<\infty)$

$F(x)$ be the cumulative distribution function of X $(-\infty<x<\infty)$, i.e.,

$$F(x)=\Pr\{X\le x\}=\int_{-\infty}^{x} f(t)dt$$

The random variable X, has a normal distribution if $f(x)$ is given by

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (-\infty<x<\infty) \tag{4.31}$$

Figure 4.6 shows examples of the probability density function, $f(x)$, for the normal distribution.

The cumulative distribution function, $F(x)$, for the normal distribution is given by

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x-\mu)/\sigma} e^{-t^2/2}dt \quad (-\infty<x<\infty) \tag{4.32}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the normal distribution are given by

$$E[X]=\mu \tag{4.33}$$

and

$$Var[X]=\sigma^2 \tag{4.34}$$

respectively.

DeGroot (1986, pp. 264–266); Freund (1992, pp. 236–238); Mood, Graybill, and Boes (1974, pp. 107–109).

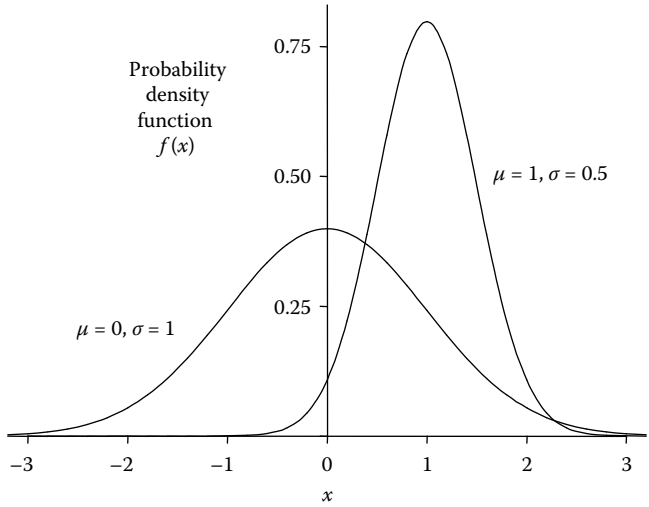


FIGURE 4.6
Examples of the normal distribution.

The notation $N(\mu, \sigma^2)$ is generally used to represent a normal distribution with a mean, μ, and a variance, σ^2. The normal distribution has the following properties.

4.6.1 Sum of Normally Distributed Random Variables

If X_1 and X_2 are independent random variables that have normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, then

- The sum $X_1 + X_2$ has a normal distribution $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- The difference $X_1 - X_2$ has a normal distribution $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

In general, if X_1, X_2, …, X_n are n independent random variables that have normal distributions $N(\mu_i, \sigma_i^2)$ $(i = 1, 2, \ldots, n)$, and a_1, a_2, …, a_n are any constants, then the sum

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

has a normal distribution

$$N(a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2)$$

DeGroot (1986, p. 270); Mood, Graybill, and Boes (1974, pp. 193–194).

4.6.2 Standard Normal Distribution

In the special case $\mu=0$ and $\sigma^2=1$, the normal distribution is called the *standard normal distribution*, with the probability density function denoted by $\phi(x)$ and the cumulative distribution function denoted by $\Phi(x)$, where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (-\infty < x < \infty) \tag{4.35}$$

and

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt \quad (-\infty < x < \infty) \tag{4.36}$$

The standard normal distribution is symmetrical about $x=0$, and hence

$$\phi(-x) = \phi(x) \tag{4.37}$$

and

$$\Phi(-x) = 1 - \Phi(x) \tag{4.38}$$

If X has a normal distribution with mean μ and variance σ^2, then $\frac{X-\mu}{\sigma}$ has a standard normal distribution.

The cumulative distribution function, $F(x)$, for the normal distribution is related to the corresponding function, $\Phi(x)$, for the standard normal distribution by

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad (-\infty < x < \infty) \tag{4.39}$$

DeGroot (1986, pp. 267–269); Hoel, Port, and Stone (1971, p. 125); Mood, Graybill, and Boes (1974, pp. 108–111).

4.6.3 Partial Moments for the Normal Distribution

Let

X have a normal distribution with mean μ and variance σ^2

$f(x)$ be the probability density function of X, given by Equation 4.31

c be any constant

The first and second partial moments of X,

$$\int_c^\infty xf(x)dx$$

and

$$\int_c^\infty x^2 f(x)dx$$

respectively, are given by

$$\int_c^\infty xf(x)dx = \mu\left\{1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right\}+\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2\right\} \tag{4.40}$$

and

$$\int_c^\infty x^2 f(x)dx = \left(\mu^2+\sigma^2\right)\left\{1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right\}+\frac{\sigma(c+\mu)}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2\right\} \tag{4.41}$$

where $\Phi(x)$ is the cumulative distribution function for the standard normal distribution, given by Equation 4.36.

Hadley and Whitin (1963, pp. 144, 167); Winkler, Roodman, and Britney (1972, p. 292).

Note that partial moments arise in the conditional mean and variance of a random variable, X, given that X is greater than a constant. For example, the conditional mean, $E[X|X>c]$, is given by

$$E\left[X|X>c\right]=\frac{\int_c^\infty xf(x)dx}{1-F(c)}$$

(see Equation 2.37), so that in the case of the normal distribution for X, the conditional mean is given by

$$E\left[X|X>c\right]=\frac{\mu\left\{1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right\}+\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2\right\}}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)}$$

i.e.,

$$E\left[X|X>c\right]=\mu+\frac{\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2\right\}}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)}$$

4.6.4 Approximations for the Cumulative Normal Distribution Function

For $x\geq 0$, the cumulative distribution function, $\Phi(x)$, for the standard normal distribution can be approximated by

$$\Phi\left(x\right)\cong 1-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\left\{\frac{a_1}{\left(1+bx\right)}+\frac{a_2}{\left(1+bx\right)^2}+\frac{a_3}{\left(1+bx\right)^3}\right\}\quad(0\leq x<\infty)\qquad(4.42)$$

where

$a_1=0.4361836$
$a_2=-0.1201676$
$a_3=0.9372980$
$b=0.33267$

The absolute error in this approximation is less than 1×10^{-5}.

Abramowitz and Stegun (1968, p. 932); Hastings (1955, p. 167); Johnson, Kotz, and Balakrishnan (1994, p. 13).

Note: This approximation can also be used for $\Phi(x)$ when $x\leq 0$, by evaluating $\Phi(-x)$ and using $\Phi(x)=1-\Phi(-x)$ from Equation 4.38.

The following is an approximation for the *inverse* of the cumulative distribution function for the standard normal distribution. Let

$$p=\Phi\left(x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt\quad(0\leq p\leq 1)$$

The value of x for a given probability, p, is given by the inverse function $x=\Phi^{-1}(p)$. For values of p where $0.5 \le p < 1$, the inverse of the cumulative distribution function can be approximated by

$$x = \Phi^{-1}(p) \cong u - \frac{c_0 + c_1 u + c_2 u^2}{1 + d_1 u + d_2 u^2 + d_3 u^3} \quad (0.5 \le p < 1) \tag{4.43}$$

where

$$u = \sqrt{\ln\left[\frac{1}{(1-p)^2}\right]}$$

$c_0=2.515517$, $c_1=0.802853$, $c_2=0.010328$
$d_1=1.432788$, $d_2=0.189269$, $d_3=0.001308$

The absolute error in this approximation is less than 4.5×10^{-4}.
Abramowitz and Stegun (1968, p. 933); Hastings (1955, p. 192).

For values of p where $0<p<0.5$, the inverse function $x=\Phi^{-1}(p)$ can first be rewritten as $x=-\Phi^{-1}(1-p)$ from Equation 4.38 and then evaluated using the same approximation.

4.7 Lognormal Distribution

Let
μ be any constant
σ be a constant, where $\sigma>0$
X be a random variable that can take any value in the range $(0, \infty)$
$f(x)$ be the probability density function of X $(0<x<\infty)$
$F(x)$ be the cumulative distribution function of X $(0<x<\infty)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_0^x f(t)dt$$

The random variable, X, has a lognormal distribution if $f(x)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma\,x} \exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2\right\} \quad (0 < x < \infty) \tag{4.44}$$

Figure 4.7 shows examples of the probability density function, $f(x)$, for the lognormal distribution.

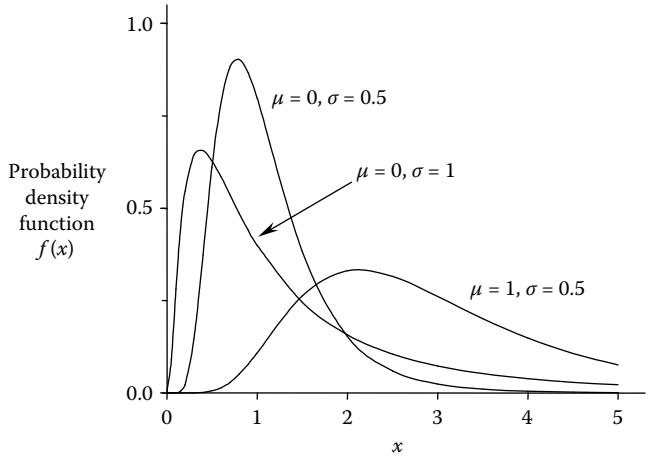


FIGURE 4.7
Examples of the lognormal distribution.

The cumulative distribution function, $F(x)$, for the lognormal distribution is given by

$$F(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \quad (0 < x < \infty) \tag{4.45}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

is the cumulative distribution function for the standard normal distribution.

The mean, $E[X]$, and the variance, $Var[X]$, for the lognormal distribution are given by

$$E[X] = e^{\mu + \sigma^2/2} \tag{4.46}$$

and

$$Var[X] = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) \tag{4.47}$$

respectively.

Bolch, Greiner, de Meer, and Trivedi (1998, p. 19); Devore (2008, pp. 166–167); Johnson, Kotz, and Balakrishnan (1994, pp. 208, 212); Mood, Graybill, and Boes (1974, p. 117); Tijms (1986, p. 395).

The lognormal distribution is derived from the normal distribution by a logarithmic transformation. If X and Y are random variables related by $Y=\ln(X)$, and Y has a normal distribution with mean μ and variance σ^2, then X has a lognormal distribution with a probability density function given by Equation 4.44.

Note that the lognormal random variable, X, is not the log of a normal random variable. Rather, it is the normal random variable that is the log of X (i.e., X is the exponential of a normal random variable). Thus, if X has a lognormal distribution with parameters μ and σ, as given by Equation 4.44, then

- $\ln(X)$ has a normal distribution with mean μ and variance σ^2
- $\dfrac{\ln(X)-\mu}{\sigma}$ has a standard normal distribution

The standard deviation, *St Dev* [X], for the lognormal distribution is

$$St\,Dev\left[X\right]=\sqrt{Var\left[X\right]}=e^{\mu+\sigma^2/2}\sqrt{e^{\sigma^2}-1} \tag{4.48}$$

and the coefficient of variation for the lognormal distribution is

$$Coeff\ of\ Var=\frac{St\,Dev[X]}{E[X]}=\sqrt{e^{\sigma^2}-1} \tag{4.49}$$

From Equations 4.46 and 4.47, σ^2 is given by

$$\sigma^2=\ln\left(1+\frac{Var\left[X\right]}{\left(E\left[X\right]\right)^2}\right)$$

and from Equation 4.46, μ is given by

$$\mu=\ln\left(E\left[X\right]\right)-\frac{1}{2}\sigma^2$$

Thus, the parameters μ and σ can be expressed in terms of the mean and the variance of the lognormal distribution, and are given by

$$\mu = \ln\left(E[X]\right) - \frac{1}{2}\ln\left(1 + \frac{Var[X]}{\left(E[X]\right)^2}\right) \tag{4.50}$$

and

$$\sigma = \sqrt{\ln\left(1 + \frac{Var[X]}{\left(E[X]\right)^2}\right)} \tag{4.51}$$

respectively.

4.8 Weibull Distribution

Let

λ and α be constants, where $\lambda > 0$ and $\alpha > 0$

X be a random variable that can take any value in the range $(0, \infty)$

$f(x)$ be the probability density function of X $(0 < x < \infty)$

$F(x)$ be the cumulative distribution function of X $(0 < x < \infty)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_0^x f(t)dt$$

The random variable, X, has a Weibull distribution if $f(x)$ is given by

$$f(x) = \alpha\lambda^{\alpha} x^{\alpha-1} \exp\left\{-(\lambda x)^{\alpha}\right\} \quad (0 < x < \infty) \tag{4.52}$$

Figure 4.8 shows examples of the probability density function, $f(x)$, for the Weibull distribution.

The cumulative distribution function, $F(x)$, for the Weibull distribution is given by

$$F(x) = 1 - \exp\left\{-(\lambda x)^{\alpha}\right\} \quad (0 < x < \infty) \tag{4.53}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the Weibull distribution are given by

$$E[X] = \frac{1}{\lambda}\Gamma\left(\frac{\alpha+1}{\alpha}\right) \tag{4.54}$$

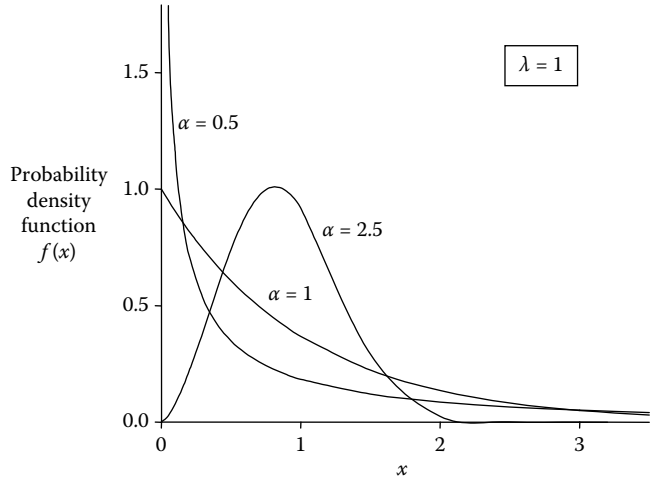


FIGURE 4.8
Examples of the Weibull distribution.

and

$$Var[X] = \frac{1}{\lambda^2}\left\{\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \left[\Gamma\left(\frac{\alpha+1}{\alpha}\right)\right]^2\right\} \tag{4.55}$$

respectively, where $\Gamma(u)$ is a gamma function, given by

$$\Gamma(u) = \int_0^\infty t^{u-1}e^{-t}dt$$

Bolch, Greiner, de Meer, and Trivedi (1998, p. 19); Devore (2008, pp. 163–166); Mood, Graybill, and Boes (1974, pp. 117–118).

The Weibull distribution, defined for $0 < x < \infty$, is a two-parameter distribution that has a closed form expression for the cumulative distribution function $F(x)$. The constant, λ, is the scale parameter, and the constant, α, is the shape parameter. In the special case $\alpha = 1$, the Weibull distribution reduces to the exponential distribution, i.e.,

$$f(x) = \lambda e^{-\lambda x}$$

as defined in Section 4.2.

4.9 Logistic Distribution

Let
a be any constant
b be a constant, where $b>0$
X be a random variable that can take any value in the range $(-\infty, \infty)$
$f(x)$ be the probability density function of X $(-\infty<x<\infty)$
$F(x)$ be the cumulative distribution function of X $(-\infty<x<\infty)$, i.e.,

$$F(x)=\Pr\{X\le x\}=\int_{-\infty}^{x} f(t)dt$$

The random variable, X, has a logistic distribution if $f(x)$ is given by

$$f(x)=\frac{e^{-(x-a)/b}}{b\left\{1+e^{-(x-a)/b}\right\}^2} \quad (-\infty<x<\infty) \tag{4.56}$$

Figure 4.9 shows examples of the probability density function, $f(x)$, for the logistic distribution.

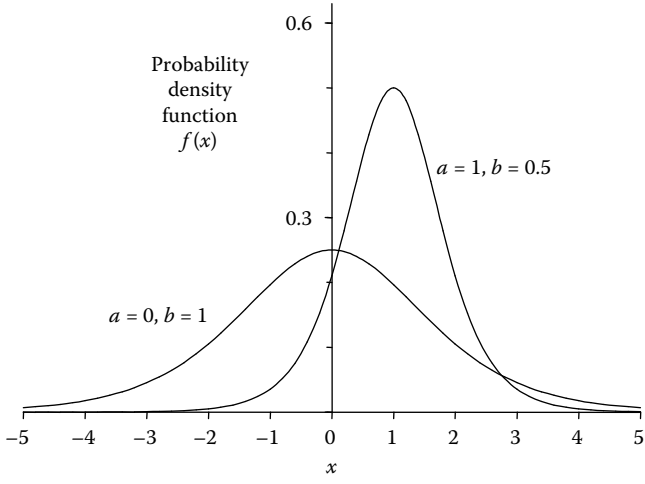


FIGURE 4.9
Examples of the logistic distribution.

The cumulative distribution function, $F(x)$, for the logistic distribution is given by

$$F(x) = \frac{1}{1 + e^{-(x-a)/b}} \quad (-\infty < x < \infty) \tag{4.57}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the logistic distribution are given by

$$E[X] = a \tag{4.58}$$

and

$$Var[X] = \frac{\pi^2 b^2}{3} \tag{4.59}$$

respectively.

Evans, Hastings, and Peacock (2000, pp. 124–126); Mood, Graybill, and Boes (1974, p. 118).

The logistic distribution, defined for $-\infty < x < \infty$, is a two-parameter distribution that has a closed form expression for the cumulative distribution function $F(x)$. The constant, a, is the location parameter, and the constant, b, is the scale parameter.

4.10 Gumbel (Extreme Value) Distribution

Let

a be a constant

b be a constant, where $b > 0$

X be a random variable that can take any value in the range $(-\infty, \infty)$

$f(x)$ be the probability density function of X $(-\infty < x < \infty)$

$F(x)$ be the cumulative distribution function of X $(-\infty < x < \infty)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_{-\infty}^{x} f(t)dt$$

The random variable, X, has a Gumbel distribution if $f(x)$ is given by

$$f(x) = \frac{1}{b} e^{-(x-a)/b} \exp\left\{-e^{-(x-a)/b}\right\} \tag{4.60}$$

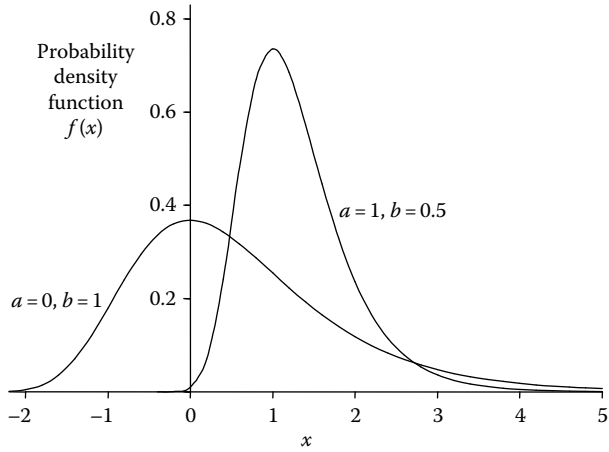


FIGURE 4.10
Examples of the Gumbel (extreme value) distribution.

Figure 4.10 shows examples of the probability density function, $f(x)$, for the Gumbel distribution.

The cumulative distribution function, $F(x)$, for the Gumbel distribution is given by

$$F(x) = \exp\left\{-e^{-(x-a)/b}\right\} \tag{4.61}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the Gumbel distribution are given by

$$E[X] = a + b\gamma \tag{4.62}$$

where γ is Euler's constant, of approximate value $\gamma \cong 0.577216$ (see Equation 19.11), and

$$Var[X] = \frac{\pi^2 b^2}{6} \tag{4.63}$$

respectively.

Evans, Hastings, and Peacock (2000, pp. 85–89); Mood, Graybill, and Boes (1974, pp. 118, 542).

The Gumbel distribution, defined for $-\infty < x < \infty$, is a two-parameter distribution that has a closed-form expression for the cumulative distribution function, $F(x)$ (as well as for the probability density function, $f(x)$).

The constant, a, is the location parameter, and the constant, b, is the scale parameter. In the special case $a=0$ and $b=1$, the probability density function, $f(x)$, becomes

$$f(x) = e^{-x} \exp\left\{-e^{-x}\right\} \tag{4.64}$$

and the cumulative distribution function, $F(x)$, becomes

$$F(x) = \exp\left\{-e^{-x}\right\} \tag{4.65}$$

The distribution in this special case is known as the *standard Gumbel distribution*.

The log of a Weibull-distributed random variable has a Gumbel distribution. If Z has a Weibull distribution with a scale parameter, λ, and a shape parameter, α, then $X=-\alpha \ln(\lambda Z)$ has a standard Gumbel distribution.

The Gumbel distribution is also known as the *extreme value distribution*. It is the limiting distribution for the largest (or smallest) value of a large number of identically distributed random variables. The Gumbel distribution given in Equations 4.60 and 4.61 is for the case of the largest value. For the case of the smallest value, the distribution has the sign reversed in the exponent, so that the probability density function, $f(x)$, is given by

$$f(x) = \frac{1}{b} e^{(x-a)/b} \exp\left\{-e^{(x-a)/b}\right\} \tag{4.66}$$

and the cumulative distribution function, $F(x)$, in this case is given by

$$F(x) = 1 - \exp\left\{-e^{(x-a)/b}\right\} \tag{4.67}$$

4.11 Pareto Distribution

Let

a and c be constants, where $a>0$ and $c>0$

X be a random variable that can take any value in the range $[a, \infty)$

$f(x)$ be the probability density function of X $(a \le x < \infty)$

$F(x)$ be the cumulative distribution function of X $(a \le x < \infty)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_a^x f(t)dt$$

The random variable, X, has a Pareto distribution if $f(x)$ is given by

$$f(x) = \frac{ca^c}{x^{c+1}} \tag{4.68}$$

Figure 4.11 shows examples of the probability density function, $f(x)$, for the Pareto distribution.

The cumulative distribution function, $F(x)$, for the Pareto distribution is given by

$$F(x) = 1 - \left(\frac{a}{x}\right)^c \tag{4.69}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the Pareto distribution are given by

$$E[X] = \frac{ca}{c-1} \quad (c > 1) \tag{4.70}$$

and

$$Var[X] = \frac{ca^2}{(c-1)^2(c-2)} \quad (c > 2) \tag{4.71}$$

respectively.

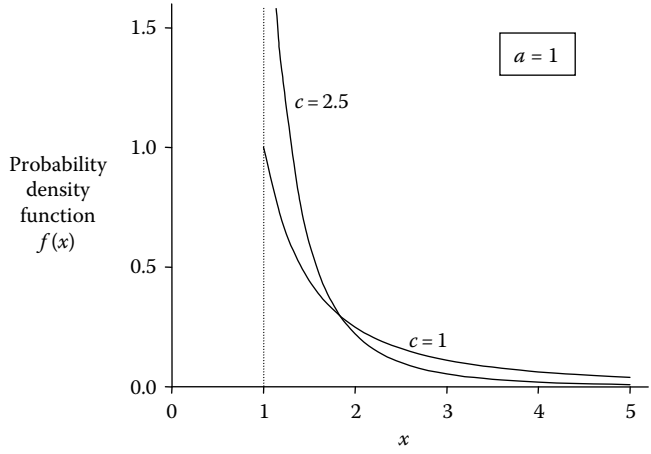


FIGURE 4.11
Examples of the Pareto distribution.

Evans, Hastings, and Peacock (2000, pp. 151–152); Mood, Graybill, and Boes (1974, p. 118).

The constant, a, is the location parameter, and the constant, c, is the shape parameter. For finite mean and variance, c must be greater than 2.

4.12 Triangular Distribution

Let

a, b, and c be constants, where $a < c < b$

X be a random variable that can take any value in the range $[a, b]$

$f(x)$ be the probability density function of X $(a \le x \le b)$

$F(x)$ be the cumulative distribution function of X $(a \le x \le b)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_a^x f(t)dt$$

The random variable, X, has a triangular distribution if $f(x)$ is given by

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & (a \le x \le c) \\ \dfrac{2(b-x)}{(b-a)(b-c)} & (c < x \le b) \end{cases} \tag{4.72}$$

Figure 4.12 shows examples of the probability density function, $f(x)$, for the triangular distribution.

The cumulative distribution function, $F(x)$, for the triangular distribution is given by

$$F(x) = \begin{cases} \dfrac{(x-a)^2}{(b-a)(c-a)} & (a \le x \le c) \\ 1 - \dfrac{(b-x)^2}{(b-a)(b-c)} & (c < x \le b) \end{cases} \tag{4.73}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the triangular distribution are given by

$$E[X] = \frac{a+b+c}{3} \tag{4.74}$$

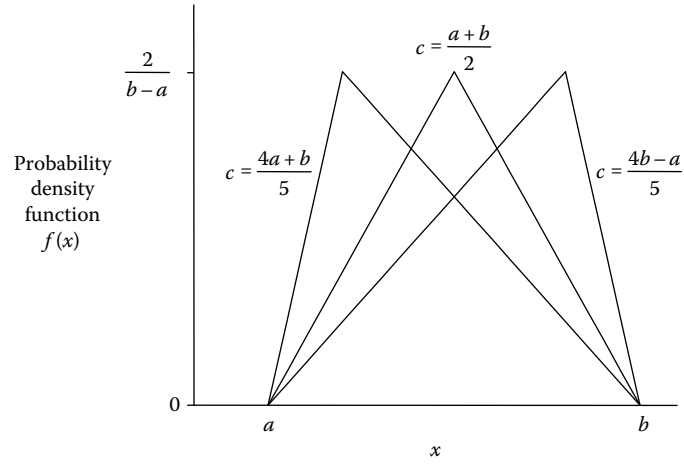


FIGURE 4.12
Examples of the triangular distribution.

and

$$Var[X] = \frac{a^2 + b^2 + c^2 - ab - ac - bc}{18} \tag{4.75}$$

respectively.

Evans, Hastings, and Peacock (2000, pp. 187–188); Law and Kelton (1991, pp. 341–342).

The constants a and b are the location parameters, and the constant c is the shape parameter. The mode of the triangular distribution occurs at $x = c$.

In the special case $c = \frac{a+b}{2}$, the triangular distribution is symmetrical about the mode, with the mean and the variance given by

$$E[X] = \frac{a+b}{2}$$

and

$$Var[X] = \frac{(b-a)^2}{24}$$

respectively.

The sum of two independent and identically distributed uniform random variables has a triangular distribution. If X_1 and X_2 are independent random variables where each has a uniform distribution over the range [0, 1], i.e., the probability density function $f(x_i)=1$ $(0\le x_i\le 1,\ i=1,\ 2)$, then the sum $X=X_1+X_2$ has a triangular distribution over the range [0, 2] with a probability density function, $f(x)$, given by

$$f(x)=\begin{cases} x & (0\le x\le 1) \\ (2-x) & (1<x\le 2) \end{cases} \tag{4.76}$$

i.e., with parameters $a=0$, $b=2$, and $c=1$. Similarly, the mean $\frac{X_1+X_2}{2}$ has a triangular distribution over the range [0, 1] with a probability density function, $f(x)$, given by

$$f(x)=\begin{cases} 4x & \left(0\le x\le \frac{1}{2}\right) \\ 4(1-x) & \left(\frac{1}{2}<x\le 1\right) \end{cases} \tag{4.77}$$

i.e., with parameters $a=0$, $b=1$, and $c=½$.

Hoel, Port, and Stone, (1971); Mood, Graybill, and Boes, (1974).

4.13 Hyper-Exponential and Hypo-Exponential Distributions

The following two distributions are extensions of the basic exponential distribution:

4.13.1 Hyper-Exponential Distribution

Let

$\lambda_1, \lambda_2, \ldots, \lambda_n$ be n constants, where $\lambda_i>0$ $(i=1, 2, \ldots, n)$

$p_1, p_2, \ldots, p_n$ be n constants, where $0<p_i<1$ $(i=1, 2, \ldots, n)$ and $p_1+p_2+\cdots+p_n=1$

X be a random variable that can take any value in the range $[0, \infty)$

$f(x)$ be the probability density function of X $(0\le x<\infty)$

$F(x)$ be the cumulative distribution function of X $(0\le x<\infty)$, i.e.,

$$F(x)=\Pr\{X\le x\}=\int_0^x f(t)dt$$

The random variable, X, has a hyper-exponential distribution if $f(x)$ is given by

$$f(x) = p_1\lambda_1 e^{-\lambda_1 x} + p_2\lambda_2 e^{-\lambda_2 x} + \cdots + p_n\lambda_n e^{-\lambda_n x} \quad (0 \le x < \infty) \tag{4.78}$$

Figure 4.13 shows examples of the probability density function, $f(x)$, for the hyper-exponential distribution.

The cumulative distribution function, $F(x)$, for the hyper-exponential distribution is given by

$$F(x) = p_1(1 - e^{-\lambda_1 x}) + p_2(1 - e^{-\lambda_2 x}) + \cdots + p_n(1 - e^{-\lambda_n x}) \quad (0 \le x < \infty) \tag{4.79}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the hyper-exponential distribution are given by

$$E[X] = \frac{p_1}{\lambda_1} + \frac{p_2}{\lambda_2} + \cdots + \frac{p_n}{\lambda_n} \tag{4.80}$$

and

$$Var[X] = \left(\frac{2p_1}{\lambda_1^2} + \frac{2p_2}{\lambda_2^2} + \cdots + \frac{2p_n}{\lambda_n^2}\right) - \left(\frac{p_1}{\lambda_1} + \frac{p_2}{\lambda_2} + \cdots + \frac{p_n}{\lambda_n}\right)^2 \tag{4.81}$$

respectively.

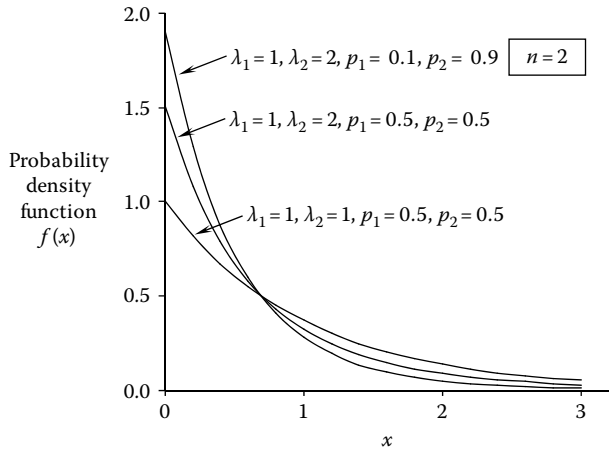


FIGURE 4.13
Examples of the hyper-exponential distribution.

The standard deviation, *St Dev*[*X*], for the hyper-exponential distribution is

$$St\,Dev[X] = \sqrt{Var[X]} = \sqrt{\left(\frac{2p_1}{\lambda_1^2} + \frac{2p_2}{\lambda_2^2} + \cdots + \frac{2p_n}{\lambda_n^2}\right) - \left(\frac{p_1}{\lambda_1} + \frac{p_2}{\lambda_2} + \cdots + \frac{p_n}{\lambda_n}\right)^2} \tag{4.82}$$

and the coefficient of variation for the hyper-exponential distribution is

$$Coeff\ of\ Var = \frac{St\,Dev[X]}{E[X]} = \sqrt{\frac{\left(\frac{2p_1}{\lambda_1^2} + \frac{2p_2}{\lambda_2^2} + \cdots + \frac{2p_n}{\lambda_n^2}\right)}{\left(\frac{p_1}{\lambda_1} + \frac{p_2}{\lambda_2} + \cdots + \frac{p_n}{\lambda_n}\right)^2} - 1} \tag{4.83}$$

For any values of the different parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $p_1, p_2, \ldots, p_n$, the coefficient of variation for the hyper-exponential distribution, given by Equation 4.83, is greater than 1. This can be seen by writing the coefficient of variation as

$$Coeff\ of\ Var = \sqrt{\frac{2\sum_{i=1}^{n} p_i\mu_i^2}{\left(\sum_{i=1}^{n} p_i\mu_i\right)^2} - 1} \quad \text{where } \mu_i = 1/\lambda_i \quad i = 1, 2, \ldots, n \tag{4.84}$$

and noting that, for $\sum_{i=1}^{n} p_i = 1$,

$$\sum_{i=1}^{n} p_i\mu_i^2 - \left(\sum_{i=1}^{n} p_i\mu_i\right)^2 = \sum_{i=1}^{n} p_i\left(\mu_i - \sum_{i=1}^{n} p_i\mu_i\right)^2$$

$$> 0$$

since the terms $\left(\mu_i - \sum_{i=1}^{n} p_i\mu_i\right)^2$ are positive,
i.e.,

$$\frac{\sum_{i=1}^{n} p_i\mu_i^2}{\left(\sum_{i=1}^{n} p_i\mu_i\right)^2} > 1$$

or

$$\frac{2\sum_{i=1}^{n} p_i \mu_i^2}{\left(\sum_{i=1}^{n} p_i \mu_i\right)^2} - 1 > 1.$$

Thus, for the hyper-exponential distribution, *Coeff of Var*>1. (Note that, in the case of the basic exponential distribution, *Coeff of Var*=1, as in Equation 4.11.)

The hyper-exponential distribution is a weighted sum of exponential distributions with different means. This mixture distribution arises in situations where a random variable takes on values from different exponential distributions, occurring with different probabilities. For example, in a queueing system with n servers in parallel, where customer arrivals to each server occur with probabilities $p_1, p_2, \ldots, p_n$ and service times at each server are exponentially distributed with means $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$, the service times for the system as a whole have a hyper-exponential distribution given by Equation 4.78. Or in a manufacturing facility that processes n different products occurring with probabilities $p_1, p_2, \ldots, p_n$, where processing times for each product are exponentially distributed with means $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$, the processing times for all products combined have a hyper-exponential distribution given by Equation 4.78.

Biernacki (2006, pp. 203–204); Bolch, Greiner, de Meer, and Trivedi (1998, pp. 12–13); Gross and Harris (1998, pp. 52, 152); Sokhan-Sanj, Gaxiola, Mackulak, and Malmgren (1999, p. 776); Trivedi (2002, pp. 133–134, 224–225).

4.13.2 Hypo-Exponential Distribution

Let

$\lambda_1, \lambda_2, \ldots, \lambda_n$ be n constants, where $\lambda_i > 0$ $(i = 1, 2, \ldots, n)$

$X_1, X_2, \ldots, X_n$ be n independent random variables that have exponential distributions with probability density functions $\lambda_i e^{-\lambda_i x_i} \quad 0 \le x_i < \infty, \quad i = 1, 2, \ldots, n$

X be the sum of the random variables $X = X_1 + X_2 + \cdots + X_n$

$f(x)$ be the probability density function of X $(0 \le x < \infty)$

$F(x)$ be the cumulative distribution function of X $(0 \le x < \infty)$, i.e.,

$$F(x) = \Pr\{X \le x\} = \int_0^x f(t)dt$$

Then the random variable X has a hypo-exponential distribution, given by

$$f(x) = a_1\lambda_1 e^{-\lambda_1 x} + a_2\lambda_2 e^{-\lambda_2 x} + \cdots + a_n\lambda_n e^{-\lambda_n x} \quad (0 \le x < \infty) \tag{4.85}$$

where the constants $a_1, a_2, \ldots, a_n$ are given by

$$a_i = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{\lambda_j}{\lambda_j - \lambda_i} \quad (i = 1, 2, \ldots n) \tag{4.86}$$

Figure 4.14 shows examples of the probability density function, $f(x)$, for the hypo-exponential distribution.

The cumulative distribution function, $F(x)$, for the hypo-exponential distribution is given by

$$F(x) = a_1(1 - e^{-\lambda_1 x}) + a_2(1 - e^{-\lambda_2 x}) + \cdots + a_n(1 - e^{-\lambda_n x}) \quad (0 \le x < \infty) \tag{4.87}$$

The mean, $E[X]$, and the variance, $Var[X]$, for the hypo-exponential distribution are given by

$$E[X] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_n} \tag{4.88}$$

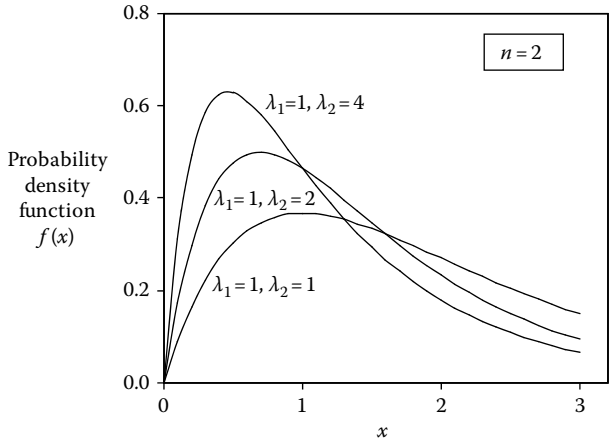


FIGURE 4.14
Examples of the hypo-exponential distribution.

and

$$Var[X] = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} + \cdots + \frac{1}{\lambda_n^2} \tag{4.89}$$

respectively.

The standard deviation, *St Dev*[*X*], for the hypo-exponential distribution is

$$St\,Dev[X] = \sqrt{Var[X]} = \sqrt{\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} + \cdots + \frac{1}{\lambda_n^2}} \tag{4.90}$$

and the coefficient of variation for the hypo-exponential distribution is

$$Coeff\ of\ Var = \frac{St\,Dev[X]}{E[X]} = \sqrt{\frac{\left(\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} + \cdots + \frac{1}{\lambda_n^2}\right)}{\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_n}\right)^2}} \tag{4.91}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} \mu_i^2}{\left(\sum_{i=1}^{n} \mu_i\right)^2}} \quad \text{where } \mu_i = 1/\lambda_i \quad (i = 1, 2, \ldots n) \tag{4.92}$$

Note that

$$\frac{\sum_{i=1}^{n} \mu_i^2}{\left(\sum_{i=1}^{n} \mu_i\right)^2} = \frac{\sum_{i=1}^{n} \mu_i^2}{\sum_{i=1}^{n} \mu_i^2 + \sum_{i \neq j} \mu_i \mu_j} < 1$$

since the parameters μ_1, μ_2, …, μ_n are positive. Thus, for the hypo-exponential distribution, *Coeff of Var* < 1. (Note that, in the case of the basic exponential distribution, *Coeff of Var* = 1, as in Equation 4.11.)

The hypo-exponential distribution arises in sequential processes, where the overall performance measure is the sum of independent random variables where each has an exponential distribution with a different mean. For example, in a manufacturing process that requires a series of *n* separate operations, where the processing times at each operation are

independent and exponentially distributed with means $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$, the total processing time has a hypo-exponential distribution, given by Equation 4.85.

In the special case $n=2$, the constants a_i in Equation 4.86 simplify to

$$a_1 = \frac{\lambda_2}{\lambda_2 - \lambda_1} \quad \text{and} \quad a_2 = \frac{\lambda_1}{\lambda_1 - \lambda_2}$$

and the probability density function, $f(x)$, in Equation 4.85 simplifies to

$$f(x) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1}\left(e^{-\lambda_1 x} - e^{-\lambda_2 x}\right) \quad (0 \le x < \infty) \tag{4.93}$$

In the special case where the parameters λ_i are the same for all i ($i=1, 2, \ldots, n$), i.e., where

$\lambda_1 = \lambda_2 = \ldots = \lambda_n = \lambda$, the hypo-exponential distribution given by Equation 4.85 reduces the Erlang distribution, given by

$$f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \quad (0 \le x < \infty) \tag{4.94}$$

(see Equation 4.12).

Bolch, Greiner, de Meer, and Trivedi (1998, pp. 14–15); Gross and Harris (1998, p. 282); Ross (2003, pp. 284–286); Trivedi (2002, pp. 129–130, 174–175, 223).

4.13.3 Additional Comments on Hyper- and Hypo-Exponential Distributions

The hyper-exponential and hypo-exponential distributions both arise in processes with n phases (or stages), where each phase has a different exponential distribution. In the case of the hyper-exponential distribution, these n phases are arranged in parallel, while in the case of the hypo-exponential distribution, these n phases are arranged in series.

Properties of the coefficient of variation for these distributions, shown earlier, can be summarized as follows:

$$\textit{Coeff of Var} \begin{cases} < 1 \text{ (Hypo-exponential distribution)} \\ = 1 \text{ (Exponential distribution)} \\ > 1 \text{ (Hyper-exponential distribution)} \end{cases}$$

Bolch, Greiner, de Meer, and Trivedi (1998, pp. 12–14); Trivedi (2002, pp. 223, 225, 256).

In the hyper-exponential distribution, the probabilities p_1, p_2, …, p_n are separate parameters, while in the hypo-exponential distribution, the constants a_1, a_2, …, a_n are related to the parameters λ_1, λ_2, …, λ_n (Equation 4.86). As in any distribution defined over the range $0 \leq x < \infty$, the probability density function $f(x)$ for the hypo-exponential distribution must satisfy the integral $\int_0^\infty f(x)dx = 1$. Hence, from Equation 4.85, the constants a_i ($i = 1$, 2, …., n) satisfy

$$a_1 + a_2 + \cdots + a_n = 1$$

However, unlike the probabilities p_1, p_2, …, p_n in the hyper-exponential distribution, the constants $a_1, a_2, \ldots, a_n$ in the hypo-exponential distribution are not probabilities, as some of them will be negative, given the different values for the parameters λ_1, λ_2, …, λ_n Ross (2003, p. 286).

5

Probability Relationships

5.1 Distribution of the Sum of Independent Random Variables

If X and Y are independent continuous random variables, then the distribution of the sum $Z=X+Y$ is given by

$$h(z)=\int_{-\infty}^{\infty} f(x)g(z-x)dx \tag{5.1}$$

where
$f(x)$ is the probability density function of X
$g(y)$ is the probability density function of Y
$h(z)$ is the probability density function of Z

The distribution $h(z)$ is known as the *convolution* of $f(x)$ and $g(y)$, and is sometimes written as $h(z)=f(x)*g(y)$.

Bolch, Greiner, de Meer, and Trivedi (1998, p. 32); Haight (1981, p. 148); Hoel, Port, and Stone (1971, p. 146); Hillier and Lieberman (1980, p. 369); Ross (2003, p. 58); Wolff (1989, p. 25).

5.2 Distribution of the Maximum and Minimum of Random Variables

If $X_1, X_2, \ldots, X_n$ are n independent and identically distributed random variables, each with cumulative distribution function $F(x)$, then

- $Y=\max(X_1, X_2, \ldots, X_n)$ has cumulative distribution function, $G(y)$, given by

$$G(y)=\{F(y)\}^n \tag{5.2}$$

- $Z=\min(X_1, X_2, \ldots, X_n)$ has cumulative distribution function, $H(z)$, given by

$$H(z)=1-\{1-F(z)\}^n \tag{5.3}$$

The corresponding probability density functions of Y and Z are given by

$$g(y)=G'(y)=n\{F(y)\}^{n-1}f(y) \tag{5.4}$$

and

$$h(z)=H'(z)=n\{1-F(z)\}^{n-1}f(z) \tag{5.5}$$

respectively, where $f(x)=F'(x)$ is the probability density function for each of the random variables $X_1, X_2, \ldots, X_n$.

DeGroot (1986, p. 160); Hoel, Port, and Stone (1971, p. 161); Mood, Graybill, and Boes (1974, pp. 183–184).

The following are examples of these results for the uniform and exponential distributions.

5.2.1 Example for the Uniform Distribution

If the random variables $X_1, X_2, \ldots, X_n$ are independent and each have a uniform distribution in the range [0, 1], then

$$F(x)=x \quad (0\le x\le 1) \tag{5.6}$$

and $Y=\max(X_1, X_2, \ldots, X_n)$ and $Z=\min(X_1, X_2, \ldots, X_n)$ have cumulative distribution functions

$$G(y)=y^n \quad (0\le y\le 1) \tag{5.7}$$

and

$$H(z)=1-(1-z)^n \quad (0\le z\le 1) \tag{5.8}$$

respectively, and have probability density functions

$$g(y)=ny^{n-1} \quad (0\le y\le 1) \tag{5.9}$$

and

$$h(z) = n(1-z)^{n-1} \quad (0 \le z \le 1) \tag{5.10}$$

respectively. The means of Y and Z in this example are $E[Y] = \frac{n}{n+1}$ and $E[Z] = \frac{1}{n+1}$, respectively.

5.2.2 Example for the Exponential Distribution

If the random variables $X_1, X_2, \ldots, X_n$ are independent and each have an exponential distribution with a mean $1/\lambda$, then

$$F(x) = 1 - e^{-\lambda x} \quad (0 \le x < \infty) \tag{5.11}$$

and $Y = \max(X_1, X_2, \ldots, X_n)$ and $Z = \min(X_1, X_2, \ldots, X_n)$ have cumulative distribution functions

$$G(y) = \left(1 - e^{-\lambda y}\right)^n \quad (0 \le y < \infty) \tag{5.12}$$

and

$$H(z) = 1 - e^{-n\lambda z} \quad (0 \le z < \infty) \tag{5.13}$$

respectively, and have probability density functions

$$g(y) = n\lambda\left(1 - e^{-\lambda y}\right)^{n-1} e^{-\lambda y} \quad (0 \le y < \infty) \tag{5.14}$$

and

$$h(z) = n\lambda e^{-n\lambda z} \quad (0 \le z < \infty) \tag{5.15}$$

respectively. Note that Z in this example has an exponential distribution with a mean $E[Z] = \frac{1}{n\lambda}$. The mean of Y in this example is

$$E[Y] = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i}$$

(see Equation 2.44).

5.3 Change of Variable in a Probability Distribution

Let

X be a continuous random variable

$Y = \psi(X)$ be a continuous, strictly increasing (or strictly decreasing) function of X

$f(x)$ be the probability density function of X

$g(y)$ be the probability density function of Y

If the inverse function $X = \psi^{-1}(Y)$ is a continuous and differentiable function of Y, then

$$g(y) = f(x)\left|\frac{dx}{dy}\right| \tag{5.16}$$

where $\left|\frac{dx}{dy}\right|$ denotes the absolute value of the derivative of x with respect to y.

DeGroot (1986, p. 153); Freund (1992, p. 266); Haight (1981, p. 146); Hoel, Port, and Stone (1971, p. 119); Mood, Graybill, and Boes (1974, p. 200).

In the cases where $\frac{dx}{dy}$ is positive (x a strictly increasing function of y), this result for a change of variable can be written as

$$g(y)dy = f(x)dx \tag{5.17}$$

Note: The result for a change of variable is derived by considering the cumulative distribution functions

$$\int_{-\infty}^{x} f(t)dt \quad \text{and} \quad \int_{-\infty}^{y} g(u)du$$

and corresponds to the result for a change of variable in an integral (see Section 16.9).

As in the case of double integrals (see Section 16.10), the above result can be extended to two or more variables. For the two-variable case, let

X_1 and X_2 be continuous jointly distributed random variables

$Y_1 = \psi_1(X_1, X_2)$ and $Y_2 = \psi_2(X_1, X_2)$ be continuous functions that define a one-to-one transformation of X_1 and X_2 to Y_1 and Y_2

$f(x_1, x_2)$ be the joint probability density function of X_1 and X_2

$g(y_1, y_2)$ be the joint probability density function of Y_1 and Y_2

From the functions ψ_1 and ψ_2, let

$X_1 = \varphi_1(Y_1, Y_2)$ and $X_2 = \varphi_2(Y_1, Y_2)$ be the corresponding functions for X_1 and X_2 in terms of Y_1 and Y_2.

If the functions φ_1 and φ_2 are continuous and differentiable, then

$$g(y_1, y_2) = f(x_1, x_2)|J| \tag{5.18}$$

where J is the Jacobian, given by the determinant of partial derivatives

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \tag{5.19}$$

i.e.,

$$J = \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1} \tag{5.20}$$

DeGroot (1986, p. 162); Freund (1992, p. 275); Haight (1981, p. 147); Mood, Graybill, and Boes (1974, p. 205).

5.4 Conditional Probability Distribution for a Constrained Random Variable

Let

X be a continuous random variable $(-\infty < X < \infty)$

$f(x)$ be the probability density function of X

$F(x) = \Pr\{X \le x\} = \int_{-\infty}^{x} f(t)\,dt$ be the cumulative distribution function of X, a be any constant

$f(x|x>a)$ be the conditional probability density function of X, given that X is greater than a

$F(x|x>a) = \Pr(X \le x|x>a)$ be the conditional cumulative distribution function of X, given that X is greater than a

Then

$$f(x|x > a) = \frac{f(x)}{\Pr\{X > a\}} \tag{5.21}$$

$$= \frac{f(x)}{\int_a^{\infty} f(x)dx} \tag{5.22}$$

$$= \frac{f(x)}{1-F(a)} \quad (a < x < \infty) \tag{5.23}$$

and

$$F(x|x > a) = \frac{F(x)}{1-F(a)} \quad (a < x < \infty) \tag{5.24}$$

More generally, for any constants a and b where $a<b$, let

$f(x|a<x<b)$ be the conditional probability density function of X, given that X lies between a and b,

$F(x|a<x<b) = \Pr(X<x|a<x<b)$ be the conditional cumulative distribution function of X given that X lies between a and b

Then

$$f(x|a < x < b) = \frac{f(x)}{\Pr\{a < X < b\}} \tag{5.25}$$

$$= \frac{f(x)}{\int_a^b f(x)dx} \tag{5.26}$$

$$= \frac{f(x)}{F(b)-F(a)} \quad (a < x < b) \tag{5.27}$$

and

$$F(x|a < x < b) = \frac{F(x)-F(a)}{F(b)-F(a)} \quad (a < x < b) \tag{5.28}$$

Stirzaker (1994, p. 243).

The corresponding results for the conditional means are given in Section 2.7.

5.5 Combination of Poisson and Gamma Distributions

Let X be a discrete random variable with a Poisson distribution with parameter μ, given by

$$P(x;\mu)=\frac{e^{-\mu}\mu^{x}}{x!} \quad (x=0,\,1,\,2,\,\ldots) \tag{5.29}$$

and let the parameter μ be a random variable with a gamma distribution, given by

$$f(\mu)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\mu^{\alpha-1}e^{-\lambda\mu} \quad (0<\mu<\infty) \tag{5.30}$$

where

$P(x;\mu)$ is the probability that $X=x$ for given μ
$f(\mu)$ is the probability density function of μ
$\Gamma(\alpha)=\int_0^{\infty} t^{\alpha-1}e^{-t}dt$ is the gamma function
α and λ are positive constants

Let

$P(x)=\Pr\{X=x\}$ be the probability that $X=x$ (for all values of μ)

The probability $P(x)$ is the probability $P(x;\mu)$ averaged over the distribution of μ, i.e.,

$$P(x)=\int_0^{\infty} P(x;\mu)f(\mu)d\mu \tag{5.31}$$

$$=\int_0^{\infty}\frac{e^{-\mu}\mu^{x}}{x!}\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\mu^{\alpha-1}e^{-\lambda\mu}d\mu \tag{5.32}$$

$$=\frac{\Gamma(\alpha+x)}{\Gamma(\alpha)}\left(\frac{\lambda}{1+\lambda}\right)^{\alpha}\left(\frac{1}{1+\lambda}\right)^{x} \tag{5.33}$$

Since x is an integer, the gamma functions $\Gamma(\alpha+x)$ and $\Gamma(\alpha)$ are related by

$$\Gamma(\alpha+x)=(\alpha+x-1)(\alpha+x-2)\cdots(\alpha+1)\alpha\Gamma(\alpha) \tag{5.34}$$

so that $P(x)$ can be written as

$$P(x)=\binom{\alpha+x-1}{x}\left(\frac{\lambda}{1+\lambda}\right)^{\alpha}\left(\frac{1}{1+\lambda}\right)^{x} \quad (x=0,\,1,\,2,\,\ldots) \tag{5.35}$$

The probability $P(x)$ is therefore a negative binomial distribution, given in standard form by

$$P(x)=\binom{\alpha+x-1}{x}p^{\alpha}(1-p)^{x} \quad (x=0,\,1,\,2,\,\ldots) \tag{5.36}$$

where

$$p=\frac{\lambda}{1+\lambda} \tag{5.37}$$

Haight (1981, p. 155); McFadden (1972, pp. 114–116); Mood, Graybill, and Boes (1974, p. 123).

Note that the parameters α and λ can take any positive values. In particular, α need not be an integer.

5.6 Bayes' Formula

For mutually exclusive (i.e., disjoint) and exhaustive events A_1, A_2, …, A_n, and any other event E, the conditional probabilities are related by

$$\Pr\{A_j|E\}=\frac{\Pr\{E|A_j\}\Pr\{A_j\}}{\sum_{i=1}^{n}\Pr\{E|A_i\}\Pr\{A_i\}} \tag{5.38}$$

where

$\Pr\{A_j|E\}$ denotes the conditional probability of event A_j given the event E

$\Pr\{E|A_j\}$ denotes the conditional probability of event E given the event A_j ($j=1, 2, \ldots, n$)

Ayyub and McCuen (1997); Clarke and Disney (1985, pp. 31–32); Devore (2008, pp. 72–73); Feller (1964, p. 114); Krishnan (2006, pp. 20, 131); Ross (2003, p. 14).

5.7 Central Limit Theorem

Let

$X_1, X_2, \ldots, X_n$ be n independent and identically distributed random variables, each with a mean μ and a variance σ^2, and let

$$S_n = X_1 + X_2 + \cdots + X_n$$

Then

$$\lim_{n\to\infty} \Pr\left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \le x \right\} = \Phi(x) \tag{5.39}$$

where

Pr{·} denotes probability

and

$\Phi(x)$ is the cumulative distribution function for the standard normal distribution, given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \quad (-\infty < x < \infty) \tag{5.40}$$

This theorem holds for any distribution of the random variables $X_1, X_2, \ldots, X_n$.

Allen (1978, p. 104); DeGroot (1986, p. 275); Hoel, Port, and Stone (1971, p. 185); Mood, Graybill, and Boes (1974, pp. 195, 233); Ross (2003, pp. 79–80).

The result from the central limit theorem above can be stated in the following ways. For large n,

- The sum $S_n = X_1 + X_2 + \cdots + X_n$ is approximately normally distributed with mean $n\mu$ and variance $n\sigma^2$
- The sample mean $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ is approximately normally distributed with mean μ and variance $\dfrac{\sigma^2}{n}$
- $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ is approximately distributed with a standard normal distribution
- $\dfrac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$ is approximately distributed with a standard normal distribution

5.8 Probability Generating Function (*z*-Transform)

Let

N be a discrete random variable taking nonnegative integer values ($N=0, 1, 2, \ldots$)

$p_n=\Pr\{N=n\}$ be the probability that N takes the value n

The probability generating function $G(z)$ of N (also known as the z-transform of N) is given by

$$G(z)=E\left[z^N\right] \tag{5.41}$$

$$=\sum_{n=0}^{\infty} p_n z^n \tag{5.42}$$

$$=p_0+p_1 z+p_2 z^2+p_3 z^3+\cdots \tag{5.43}$$

The probability generating function characterizes a discrete probability distribution $\{p_n\}$ ($n=0, 1, 2, \ldots$) by a single function $G(z)$.

A given probability distribution has a unique probability generating function. The converse also holds. A given probability generating function corresponds to a unique probability distribution.

The probability generating function, $G(z)$, generates the individual probabilities. The probabilities are obtained from $G(z)$ by evaluating the function and its successive derivatives at $z=0$, so that $p_0=G(0)$, $p_1=G'(0)$, $p_2=G''(0)/2!$, and in general

$$p_n=\frac{G^{(n)}(0)}{n!} \tag{5.44}$$

where

$$G'(0)=\frac{d}{dz}G(z)\bigg|_{z=0}$$

$$G''(0)=\frac{d^2}{dz^2}G(z)\bigg|_{z=0}$$

$$G^{(n)}(0)=\frac{d^n}{dz^n}G(z)\bigg|_{z=0}$$

The mean, $E[N]$, and the variance, $Var[N]$, of the random variable, N, are obtained from the first two derivatives of $G(z)$, evaluated at $z=1$:

$$E[N] = G'(1) \tag{5.45}$$

and

$$Var[N] = G''(1) + G'(1) - \{G'(1)\}^2 \tag{5.46}$$

where

$$G'(1) = \left.\frac{d}{dz} G(z)\right|_{z=1}$$

$$G''(1) = \left.\frac{d^2}{dz^2} G(z)\right|_{z=1}$$

If X and Y are independent discrete random variables with probability generating functions $G_X(z)$ and $G_Y(z)$, respectively, then the sum $S=X+Y$ has a probability generating function, $G_S(z)$, given by

$$G_S(z) = G_X(z) \cdot G_Y(z) \tag{5.47}$$

In general, the sum of any number of independent random variables has a probability generating function given by the product of the probability generating functions of the random variables.

Allen (1978, p. 57); Bolch, Greiner, de Meer, and Trivedi (1998, pp. 25–26); Haight (1981, pp. 33–34); Krishnan (2006, p. 162); Lefebvre (2006, p. 106); Papadopoulos, Heavy, and Browne (1993, pp. 361–362); Stirzaker (1994, pp. 179–180, 183).

5.9 Moment Generating Function

Let

X be a random variable

$f(x)$ be the probability density function of X, if X is continuous

$p_x = \Pr\{X=x\}$ be the probability that X takes the value x, if X is discrete

The moment generating function $M(t)$ of X is given by

$$M(t) = E\left[e^{tX}\right] \tag{5.48}$$

$$= \begin{cases} \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x} e^{tx} p_x & \text{if } X \text{ is discrete} \end{cases} \tag{5.49}$$

A given probability distribution has a unique moment generating function. The converse also holds. A given moment generating function corresponds to a unique probability distribution.

The moment generating function, $M(t)$, generates the individual moments of X, and can be written as

$$M(t) = 1 + tE[X] + \frac{t^2}{2!} E\left[X^2\right] + \cdots + \frac{t^n}{n!} E\left[X^n\right] + \cdots \tag{5.50}$$

The moments are obtained from successive derivatives of $M(t)$, evaluated at $t = 0$:

$$E[X] = M'(0) \tag{5.51}$$

$$E\left[X^2\right] = M''(0) \tag{5.52}$$

and in general

$$E\left[X^n\right] = M^{(n)}(0) \tag{5.53}$$

where

$$M'(0) = \frac{d}{dt} M(t)\Big|_{t=0}$$

$$M''(0) = \frac{d^2}{dt^2} M(t)\Big|_{t=0}$$

$$M^{(n)}(0) = \frac{d^n}{dt^n} M(t)\Big|_{t=0}$$

The variance of X is given by

$$Var[X] = E\left[X^2\right] - \left(E[X]\right)^2 \tag{5.54}$$

$$= M''(0) - \left\{M'(0)\right\}^2 \tag{5.55}$$

If X and Y are independent random variables and have moment generating functions $M_X(t)$ and $M_Y(t)$, respectively, then the sum $Z = X + Y$ has a moment generating function $M_Z(t)$ given by

$$M_Z(t) = M_X(t) \cdot M_Y(t) \tag{5.56}$$

In general, the sum of any number of independent random variables has a moment generating function given by the product of the moment generating functions of the random variables.

Allen (1978, pp. 53–54); Hoel, Port, and Stone (1971, pp. 197–199); Krishnan (2006, p. 164); Lefebvre (2006, pp. 107–108); Mood, Graybill, and Boes (1974, pp. 78–79, 192); Ross (2003, pp. 64, 68).

5.10 Characteristic Function

Let

X be a random variable

$f(x)$ be the probability density function of X, if X is continuous

$p_x = \Pr\{N = x\}$ be the probability that X takes the value x, if X is discrete

The characteristic function $\phi(t)$ of the random variable X is given by

$$\phi(t) = E\left[e^{itX}\right] \tag{5.57}$$

$$= \begin{cases} \int_{-\infty}^{\infty} e^{itx} f(x)\,dx & \text{if } X \text{ is continuous} \\ \sum_x e^{itx} p_x & \text{if } X \text{ is discrete} \end{cases} \tag{5.58}$$

where $i = \sqrt{-1}$. Like the moment generating function, the characteristic function generates the moments of X.

The characteristic function in Equation 5.58 is a form of the Fourier transform, and converges for real values of t (whereas the integral or sum defining the moment generating function in Equation 5.49 does not necessarily converge). The characteristic function, $\phi(t)$, also has the property that the probability distribution can be obtained from $\phi(t)$ by an inverse transform (Poularikas and Seely, 1985, pp. 163, 290).

As in the case of the moment generating function, a given probability distribution has a unique characteristic function, and, conversely, a given characteristic function corresponds to a unique probability distribution.

The characteristic function, $\phi(t)$, can be written as

$$\phi(t) = 1 + itE[X] + \frac{(it)^2}{2!}E[X^2] + \cdots + \frac{(it)^n}{n!}E[X^n] + \cdots \tag{5.59}$$

and the moments are obtained from successive derivatives of $\phi(t)$, evaluated at $t=0$:

$$E[X] = -i\phi'(0) \tag{5.60}$$

$$E[X^2] = -\phi''(0) \tag{5.61}$$

and in general

$$E[X^n] = (-1)^n i^n \phi^{(n)}(0) \tag{5.62}$$

where

$$\phi(0) = \frac{d}{dt}\phi(t)\bigg|_{t=0}$$

$$\phi''(0) = \frac{d^2}{dt^2}\phi(t)\bigg|_{t=0}$$

$$\phi^{(n)}(0) = \frac{d^n}{dt^n}\phi(t)\bigg|_{t=0}$$

If X and Y are independent random variables and have characteristic functions $\phi_X(t)$ and $\phi_Y(t)$, respectively, then the sum $Z = X + Y$ has a characteristic function $\phi_Z(t)$ given by

$$\phi_Z(t) = \phi_X(t) \cdot \phi_Y(t) \tag{5.63}$$

In general, the sum of any number of independent random variables has a characteristic function given by the product of the characteristic functions of the random variables.

The logarithm of the characteristic function is the *cumulant generating function*. A Taylor series expansion of the cumulant generating function, $K(t)$, is given by

$$K(t) = \ln\left\{\phi(t)\right\} \tag{5.64}$$

$$= \kappa_0 + \kappa_1 (it) + \kappa_2 \frac{(it)^2}{2!} + \cdots + \kappa_n \frac{(it)^n}{n!} + \cdots \tag{5.65}$$

The coefficients κ_0, κ_1, κ_2, … in this series are known as the cumulants.

Binmore (1983, pp. 358–359); Hoel, Port, and Stone (1971, pp. 200–202); Krishnan (2006, pp. 151, 167); Zwillinger (1996, p. 574).

5.11 Laplace Transform

For a continuous random variable, X, defined in the range $0 \le X < \infty$, the Laplace transform $L(s)$ of X is given by

$$L(s) = E\left[e^{-sX}\right]$$

$$= \int_0^\infty e^{-sx} f(x)\, dx \tag{5.66}$$

As in the case of the characteristic function, the probability distribution can be obtained from the Laplace transform by an inverse transform (Papadopoulos, Heavy, and Browne, 1993, pp. 360–361; Poularikas and Seely, 1985, pp. 291–292).

As in the cases of both the moment generating function and the characteristic function, a given probability distribution has a unique Laplace transform, and, conversely, a given Laplace transform corresponds to a unique probability distribution.

Moments are obtained from successive derivatives of $L(s)$, evaluated at $s = 0$:

$$E[X] = -L'(0) \tag{5.67}$$

$$E\left[X^2\right] = L''(0) \tag{5.68}$$

and in general

$$E\left[X^n\right] = (-1)^n L^{(n)}(0) \tag{5.69}$$

If X and Y are independent random variables and have Laplace transforms $L_X(t)$ and $L_Y(t)$, respectively, then the sum $Z = X + Y$ has a Laplace transform $L_Z(t)$ given by

$$L_Z(t) = L_X(t) \cdot L_Y(t) \tag{5.70}$$

In general, the sum of any number of independent random variables has a Laplace transform given by the product of the Laplace transforms of the random variables.

Allen (1978, pp. 59–60); Bolch, Greiner, de Meer, and Trivedi (1998, pp. 26–27); Haight (1981, pp. 165–169).

6

Stochastic Processes

6.1 Poisson Process and Exponential Distribution

6.1.1 Properties of the Poisson Process

For a Poisson process with rate λ arrivals (or events) per unit time:

- The time, X, between successive arrivals is exponentially distributed, with a cumulative distribution function, $F(x)$, given by

$$F(x) = \Pr\{X \le x\} = 1 - e^{-\lambda x} \quad (0 \le x < \infty) \tag{6.1}$$

  and a probability density function, $f(x)$, given by

$$f(x) = F'(x) = \lambda e^{-\lambda x} \quad (0 \le x < \infty) \tag{6.2}$$

  and with a mean, $E[X] = \frac{1}{\lambda}$, and a variance, $Var[X] = \frac{1}{\lambda^2}$.

- The number, $N(t)$, of arrivals in time, t, is Poisson distributed, with a probability distribution given by

$$\Pr\{N(t) = n\} = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad (n = 0, 1, 2, \ldots) \tag{6.3}$$

  and with a mean, $E[N(t)] = \lambda t$, and a variance, $Var[N(t)] = \lambda t$.

Cox and Miller (1965, pp. 6, 147, 150); Ross (2003, pp. 270–271, 289); Wolff (1989, pp. 70–71).

6.1.2 "Lack of Memory" Property of the Exponential Distribution

For an exponentially distributed random variable, X, and any constants, $s \ge 0$ and $t \ge 0$,

$$\Pr\{X > s + t \mid X > s\} = \Pr\{X > t\} \tag{6.4}$$

$$\Pr\{X>s+t\}=\Pr\{X>s\}\cdot\Pr\{X>t\} \tag{6.5}$$

From Equation 6.4, the conditional probability that $X>s+t$, given that $X>s$, is equal to the unconditional probability that $X>t$. This is the "lack of memory" or the "memoryless" property of the exponential distribution (see also Section 4.2).

Allen (1978, pp. 82–83); Heyman and Sobel (1982, p. 511); Hoel, Port, and Stone (1971, p. 127); Mood, Graybill, and Boes (1974, p. 114); Krishnan (2006, p. 80); Pal, Jin, and Lim (2006, pp. 157–159) Ross (2003, p. 272); Tijms (1986, p. 9).

6.1.3 Competing Exponentials

If $X_1, X_2, \ldots, X_n$ are independent and exponentially distributed random variables, with probability density functions $f_i(x_i)=\lambda_i\exp(-\lambda_i x_i)$ $(i=1, 2, \ldots, n)$, then the probability that X_i has the smallest value of the n random variables is given by

$$\Pr\{X_i=\min(X_1, X_2, \ldots, X_n)\}=\frac{\lambda_i}{\sum_{k=1}^{n}\lambda_k} \tag{6.6}$$

Example: If the above random variables $X_1, X_2, \ldots, X_n$ are the times-to-next-failure of different machines 1, 2, …, n, then the probability that the ith machine fails first is given by Equation 6.6.

Ross (2003, pp. 279–280); Tijms (1986, p. 20); Trivedi (2002, pp. 252–253).

6.1.4 Superposition of Independent Poisson Processes

If customers of type i arrive according to a Poisson process with rate λ_i, then the arrival process of all customer types is Poisson with rate $\sum_i \lambda_i$.

Bolch, Greiner, de Meer, and Trivedi (1998, p. 11); Çinlar (1975, p. 87); Cox (1962, p. 73); Wolff (1989, p. 75); Trivedi (2002, p. 308).

6.1.5 Splitting of a Poisson Process

If arrivals from a Poisson process with rate λ are independently classified as type i with probability p_i $(i=1, 2, \ldots, n)$, where $\sum_{i=1}^{n} p_i=1$, then the arrivals of each type i are independent Poisson processes with rates $\lambda_i=\lambda p_i$.

Bolch, Greiner, de Meer, and Trivedi (1998, pp. 11–12); Ross (2003, p. 296); Çinlar (1975, p. 89); Trivedi (2002, p. 309); Wolff (1989, pp. 74–75).

6.1.6 Arrivals from a Poisson Process in a Fixed Interval

Given n Poisson arrivals in some fixed interval, the times of the n arrivals are uniformly distributed in that interval.

Ross (2003, pp. 301–302); Trivedi (2002, p. 311); Wolff (1989, p. 73).

6.2 Renewal Process Results

For a renewal process, the times between successive arrivals (or events) are independent and identically distributed with an arbitrary distribution.

6.2.1 Mean and Variance of the Number of Arrivals in a Renewal Process

If

$N(t)$ = number of arrivals in time t

X = time between successive arrivals

Then, for large t,

$$\frac{Var\left[N\left(t\right)\right]}{E\left[N\left(t\right)\right]} \cong \frac{Var[X]}{\left(E\left[X\right]\right)^2} \tag{6.7}$$

Note that, in the special case of a Poisson process, X is exponentially distributed, $N(t)$ is Poisson distributed, and the above ratio is equal to 1.

In general, for large t, $N(t)$ is approximately normally distributed with mean and variance given by

$$E\left[N\left(t\right)\right] = \frac{t}{E\left[X\right]} \tag{6.8}$$

$$Var\left[N\left(t\right)\right] = \frac{t\,Var\left[X\right]}{\left(E\left[X\right]\right)^3} \tag{6.9}$$

Cox (1962, p. 40); Cox and Miller (1965, p. 343); Ross (2003, p. 416).

6.2.2 Distribution of First Interval in a Renewal Process

If

X = time between successive arrivals

X_1 = time between an arbitrary origin and first arrival after the origin in an equilibrium renewal process, then the probability density function, $f_1(x)$ of X_1, is given by

$$f_1(x) = \frac{1 - F(x)}{\mu} \tag{6.10}$$

where

$\mu = E[X]$ = mean of X

and

$F(x) = \Pr\{X \le x\}$ = cumulative distribution function of X.

The mean and the variance of X_1 are given by

$$E[X_1] = \frac{1}{2}\left(\mu + \frac{\sigma^2}{\mu}\right) \tag{6.11}$$

$$Var[X_1] = \frac{\mu_3}{3\mu} + \frac{\sigma^2}{2}\left(1 - \frac{\sigma^2}{2\mu^2}\right) + \frac{\mu^2}{12} \tag{6.12}$$

where

$\mu = E[X]$ = mean of X

$\sigma^2 = E\left[(X - \mu)^2\right]$ = variance of X

$\mu_3 = E\left[(X - \mu)^3\right]$ = third moment of X about the mean

Note that, in the special case of a Poisson process, X and X_1 are both exponentially distributed with a mean μ.

Cox (1962, pp. 64, 66); Cox and Miller (1965, pp. 347–348); Wagner (1969, p. 848).

6.3 Markov Chain Results

The following are limiting (steady-state) results for irreducible, aperiodic Markov chains with a finite number of states.

6.3.1 Discrete-Time Markov Chains

For a discrete-time Markov chain with n states, let
P_{ij} be the probability of a transition from state i to state j ($i, j=1, 2, \ldots, n$)
π_j be the limiting probability for state j ($j=1, 2, \ldots, n$)
The limiting state probabilities are given by

$$\pi_j = \sum_{i=1}^{n} \pi_i P_{ij} \quad (j=1, 2, \ldots, n) \tag{6.13}$$

or, in matrix notation,

$$\pi = \pi \mathbf{P} \tag{6.14}$$

where
$\boldsymbol{\pi}=(\pi_1, \pi_2, \ldots, \pi_n)$ is the row vector of limiting state probabilities
$0 \le \pi_j \le 1$ for all j
$\sum_{j=1}^{n} \pi_j = 1$
$\mathbf{P}=\{P_{ij}\}$ is the matrix of transition probabilities

Bolch, Greiner, de Meer, and Trivedi (1998, p. 41); Çinlar (1975, pp. 152–153); Gross and Harris (1998, p. 35); Kleinrock (1975, p. 31); Kleinrock (1976, p. 7); Hillier and Lieberman (1980, p. 381); Ross (2003, p. 201).

6.3.2 Continuous-Time Markov Chains

For a continuous-time Markov chain with n states, let
Q_{ij} be the transition rate from state i to state j ($i \ne j$)
$Q_{jj} = -\sum_{i, i \ne j} Q_{ij}$ ($j=1, 2, \ldots, n$)
π_j be the limiting probability for state j ($j=1, 2, \ldots, n$)
The limiting state probabilities are given by

$$0 = \pi_j Q_{jj} + \sum_{i, i \ne j} \pi_i Q_{ij} \quad (j = 1, 2, \ldots, n) \tag{6.15}$$

or, in matrix notation,

$$\mathbf{0} = \boldsymbol{\pi} \mathbf{Q} \tag{6.16}$$

where

$\mathbf{0} = (0, 0, \ldots, 0)$

$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ is the row vector of limiting state probabilities

$0 \le \pi_j \le 1$ for all j

$\sum_{j=1}^{n} \pi_j = 1$

$\mathbf{Q} = \{Q_{ij}\}$ is the transition rate matrix (also known as the intensity matrix, generator matrix, or infinitesimal generator matrix)

Bolch, Greiner, de Meer, and Trivedi (1998, p. 53); Gross and Harris (1998, p. 35); Hillier and Lieberman (1980, p. 389); Kleinrock (1975, p. 52); Kleinrock (1976, p. 8); Ross (2003, p. 369).

7

Queueing Theory Results

7.1 Notation for Queue Types

The following standard notation is used to characterize systems with a single queue in equilibrium, identical parallel servers, unlimited waiting room, and with first-come first-served (FCFS) queueing discipline. Each system is defined by

$$A/B/m$$

where

A denotes the distribution of inter-arrival times
B denotes the distribution of service times
m denotes the number of servers in parallel

Commonly used symbols for both the *A* and *B* positions in this notation are *M*, *D*, and *G*. The symbol *M* is used for the exponential distribution to denote its Markovian ("memoryless") property. The symbol *D* is used for deterministic times. The symbol *G* is used for a general distribution of independent and identically distributed random variables. Examples:

M/M/1 queue: Exponential distribution for inter-arrival times, exponential distribution for service times, and one server.

M/D/2 queue: Exponential distribution for inter-arrival times, deterministic service times, and two servers in parallel.

G/G/c queue: General distribution for inter-arrival times, general distribution for service times, and *c* servers in parallel.

Gross and Harris (1998, pp. 8–9); Kendall (1953, pp. 339–340); Wolff (1989, p. 245).

7.2 Definitions of Queueing System Variables

L_q = average queue length (average number of customers in queue)
L = average system length (average number of customers in system, including those being served)

W_q = average waiting time in queue (average time a customer spends in queue)
W = average time in system (average time a customer spends in queue plus service)
N = number of customers in system ($E[N] = L$)
T = time customer spends in system ($E[T] = W$)
m = number of servers
λ = arrival rate (number of customers arriving per unit time); $1/\lambda$ = mean inter-arrival time
μ = service rate at one server (number of customers served per unit time); $1/\mu$ = mean service time)
$\rho = \dfrac{\lambda}{m\mu}$ = traffic intensity ($\rho < 1$)
σ_a^2 = variance of inter-arrival times
σ_s^2 = variance of service times
$C_a^2 = \lambda^2\sigma_a^2$ = squared coefficient of variation of inter-arrival times
$C_s^2 = \mu^2\sigma_s^2$ = squared coefficient of variation of service times

For exponentially distributed service times, $C_s^2 = 1$, and for deterministic service times, $C_s^2 = 0$ (similarly for inter-arrival times).

7.3 Little's Law and General Queueing System Relationships

$$\left.\begin{aligned} L_q &= \lambda W_q \quad (7.1)\\ L &= \lambda W \quad (7.2)\end{aligned}\right\} \text{Little's Law}$$

$$L = L_q + \frac{\lambda}{\mu} \quad (7.3)$$

$$W = W_q + \frac{1}{\mu} \quad (7.4)$$

Allen (1978, p. 354); Gillett (1976, p. 462); Gross and Harris (1988, pp. 11–13); Hillier and Lieberman (1980, p. 406); Hopp and Spearman (1996, p. 273); Little (1961, p. 383); Medhi (1991, pp. 62–63).

Note: Equations 7.1 through 7.4 hold for all queue types listed in Section 7.1.

7.4 Extension of Little's Law

For the *M/G/*1 queue, Little's Law $L=\lambda W$ given earlier can be extended to higher moments. For the kth moment:

$$E\left[N(N-1)(N-2)\cdots(N-k+1)\right]=\lambda^k E\left[T^k\right] \tag{7.5}$$

where
N=number of customers in system
T=time customer spends in system

Special cases:

$$k=1:\quad E[N]=\lambda E[T] \quad \text{(i.e., } L=\lambda W\text{)} \tag{7.6}$$

$$k=2:\quad E\left[N(N-1)\right]=\lambda^2 E\left[T^2\right] \tag{7.7}$$

Hence

$$Var[N]=\lambda E[T]+\lambda^2 Var[T] \tag{7.8}$$

Cox and Smith (1971, pp. 52, 56); Gross and Harris (1988, p. 225).

7.5 Formulas for Average Queue Length, L_q

Queue Type	$m=1$	General m
M/M/m	$\dfrac{\rho^2}{1-\rho}$ (7.9)	$\left\{\begin{array}{l}\dfrac{(m\rho)^m\rho}{m!(1-\rho)}\left(\dfrac{1}{\dfrac{(m\rho)^m}{m!}+(1-\rho)\displaystyle\sum_{k=0}^{m-1}\dfrac{(m\rho)^k}{k!}}\right) \quad (7.12)\\ \dfrac{\rho^{\sqrt{2(m+1)}}}{1-\rho} \quad \text{approx.} \quad (7.13)\end{array}\right.$
M/G/m	$\dfrac{\rho^2}{1-\rho}\left(\dfrac{1+C_s^2}{2}\right)$ (7.10)	$\dfrac{\rho^{\sqrt{2(m+1)}}}{1-\rho}\left(\dfrac{1+C_s^2}{2}\right)$ (7.14)
G/G/m	$\dfrac{\rho^2}{1-\rho}\left(\dfrac{C_a^2+C_s^2}{2}\right)$ (7.11)	$\dfrac{\rho^{\sqrt{2(m+1)}}}{1-\rho}\left(\dfrac{C_a^2+C_s^2}{2}\right)$ (7.15)

where

m = number of servers

$$\rho = \frac{\lambda}{m\mu} = \frac{\text{arrival rate}}{\text{total service rate}} = \text{traffic intensity } (\rho < 1)$$

$C_a^2 = \lambda^2\sigma_a^2$ = (arrival rate)^2 × (variance of inter-arrival times)

$C_s^2 = \mu^2\sigma_s^2$ = (service rate of one server)^2 × (variance of service times)

References for these results are given in the table in Section 7.7. For the case $m=1$, Equations 7.9 and 7.10 are exact, and Equation 7.11 is approximate. For general m, Equation 7.12 is exact, and Equations 7.13 through 7.15 are approximate. The approximations for *M/G/m* and *G/G/m* with general m (Equations 7.14 and 7.15) can be improved slightly by replacing the first term, $\rho^{\sqrt{2(m+1)}}/(1-\rho)$, by the more complicated but exact expression given in Equation 7.12.

7.6 Formulas for Average Time in Queue, W_q

Queue Type	$m=1$	General m
M/M/m	$\frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)$ (7.16)	$\left\{\begin{array}{l}\dfrac{(m\rho)^m}{\mu m m!(1-\rho)}\left(\dfrac{1}{\dfrac{(m\rho)^m}{m!}+(1-\rho)\sum_{k=0}^{m-1}\dfrac{(m\rho)^k}{k!}}\right) \quad (7.19)\\ \dfrac{\rho^{\sqrt{2(m+1)}-1}}{\mu m(1-\rho)} \quad \text{approx.} \quad (7.20)\end{array}\right.$
M/G/m	$\frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)\left(\frac{1+C_s^2}{2}\right)$ (7.17)	$\frac{\rho^{\sqrt{2(m+1)}-1}}{\mu m(1-\rho)}\left(\frac{1+C_s^2}{2}\right)$ (7.21)
G/G/m	$\frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2+C_s^2}{2}\right)$ (7.18)	$\frac{\rho^{\sqrt{2(m+1)}-1}}{\mu m(1-\rho)}\left(\frac{C_a^2+C_s^2}{2}\right)$ (7.22)

where

m = number of servers

$$\rho = \frac{\lambda}{m\mu} = \frac{\text{arrival rate}}{\text{total service rate}} = \text{traffic intensity } (\rho < 1)$$

$C_a^2 = \lambda^2\sigma_a^2$ = (arrival rate)^2 × (variance of inter-arrival times)

$C_s^2 = \mu^2\sigma_s^2$ = (service rate of one server)^2 × (variance of service times)

References for these results are given in the table in Section 7.7. For the case $m=1$, Equations 7.16 and 7.17 are exact, and Equation 7.18 is approximate. For general m, Equation 7.19 is exact, and Equations 7.20 through 7.22 are approximate. The approximations for *M/G/m* and *G/G/m* with general m (Equations 7.21 and 7.22) can be improved slightly by replacing the first term, $\rho^{\sqrt{2(m+1)}-1}/\{\mu m(1-\rho)\}$, by the more complicated but exact expression in Equation 7.19.

7.7 References for the Formulas for Average Queue Length and Time in Queue

(Given in Equations 7.9 through 7.15 and Equations 7.16 through 7.22.)

Queue Type	$m=1$	**General m**
M/M/m	Allen (1978) Cohen (1985) Gillett (1976) Gross and Harris (1998) Hopp and Spearman (1996) Morse (1958) Wolff (1989)	*For exact Equations 7.12 and 7.19*: Allen (1978) Buzacott and Shanthikumar (1993) Gillett (1976) Gross and Harris (1998) Hall (1991) Hillier and Lieberman (1980) *For approximate Equations 7.13 and 7.20*: Hopp and Spearman (1996) Sakasegawa (1977) Whitt (1993)
M/G/m	Cox and Smith (1971) Gross and Harris (1998) Hall (1991) Kendall (1951) Kleinrock (1975) Sakasegawa (1977)	Sakasegawa (1977)
G/G/m	Hall (1991) Hopp and Spearman (1996) Sakasegawa (1977) Tanner (1995) Shanthikumar and Buzacott (1980)	Hall (1991) Hopp and Spearman (1996) Sakasegawa (1977) Tanner (1995) Allen (1978) Kimura (1986) Whitt (1993)

7.8 Pollaczek–Khintchine Formula for Average Time in Queue, W_q

Equation 7.17 for the average time in queue W_q in an $M/G/1$ queue is sometimes called the Pollaczek–Khintchine formula. It is an exact result, based on derivations by Pollaczek (1930, p. 77) and Khintchine (1932, p. 79). More recent references are given in Section 7.7.

7.9 Additional Formulas for Average Time in Queue, W_q

The average time in queue W_q in a $G/G/1$ queue is given by the approximate result in Equation 7.18. An earlier approximation for W_q in a $G/G/1$ queue, derived by Kingman (1961, p. 903, 1965, p. 139) for the case of heavy traffic (traffic intensity ρ close to 1), is given by

$$G/G/1: \quad W_q \cong \lambda\left(\frac{1}{1-\rho}\right)\left(\frac{\sigma_a^2+\sigma_s^2}{2}\right) \tag{7.23}$$

Kingman (1961, p. 903, 1965, p. 139); Kleinrock (1976, p. 31); Larson and Odoni (1981, p. 230); Medhi (1991, p. 376); Tanner (1995, p. 158).

In terms of coefficients of variation C_a and C_s, Kingman's result (Equation 7.23) becomes

$$W_q \cong \frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)\left(\frac{(C_a/\rho)^2+C_s^2}{2}\right)$$

For ρ close to 1, this expression and Equation 7.18 are in close agreement. Numerical comparisons for the performance of these formulas are given in Tanner (1995). Note that, when inter-arrival times are exponentially distributed ($C_a=1$), Equation 7.18 reduces to the exact result for W_q in an $M/G/1$ queue, given by Equation 7.17.

A refinement to Equation 7.18 for W_q in a $G/G/1$ queue is given in Krämer and Langenbach-Belz (1976, 1978):

$$G/G/1: \quad W_q \cong \frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2+C_s^2}{2}\right)g(\rho, C_a^2, C_s^2) \tag{7.24}$$

where

$$g\left(\rho, C_a^2, C_s^2\right) = \begin{cases} \exp\left[-\dfrac{2(1-\rho)}{3\rho}\dfrac{\left(1-C_a^2\right)^2}{\left(C_a^2+C_s^2\right)}\right] & \left(C_a^2 < 1\right) \\ \exp\left[-\dfrac{(1-\rho)\left(C_a^2-1\right)}{\left(C_a^2+4C_s^2\right)}\right] & \left(C_a^2 \geq 1\right) \end{cases}$$

For analyzing networks of queues, Whitt (1983) uses this refinement in the case $C_a^2 < 1$, and the simpler result (Equation 7.18) in the case $C_a^2 \geq 1$ (Whitt, 1983, p. 2802).

7.10 Heavy Traffic Approximation for Distribution of Time in Queue

For a *G*/*G*/1 queue, let

t be the time spent in the queue
$f(t)$ be the probability density function of t
$F(t)$ be the cumulative distribution function of t

where

$$F(t) = \int_0^t f(x)dx = \Pr\{\text{time in queue} \leq t\}$$

Under heavy traffic conditions (traffic intensity $\rho = \lambda/\mu$ close to 1), the distribution of time spent in a *G*/*G*/1 queue can be approximated by an exponential distribution, i.e.,

$$f(t) \cong \frac{1}{W_q} e^{-t/W_q} \tag{7.25}$$

and

$$F(t) \cong 1 - e^{-t/W_q} \tag{7.26}$$

where W_q is the average time in queue for a *G/G/1* queue.

Kingman (1961, pp. 902–903, 1962, pp. 383–384, 1965, p. 139), Kleinrock (1976, p. 31), Larson and Odoni (1981, pp. 229–230), Medhi (1991, pp. 373, 376).

This heavy traffic approximation can also be applied to a queue with multiple servers in parallel (i.e., a *G/G/m* queue). When traffic intensity $\rho = \frac{\lambda}{m\mu}$ is close to 1, the distribution of time spent in a *G/G/m* queue can be approximated by the exponential distribution given by Equations 7.25 and 7.26, where W_q in this case is the average time in queue for a *G/G/m* queue.

Kingman (1965, p. 153); Kleinrock (1976, p. 47); Larson and Odoni (1981, p. 231); Medhi (1991, p. 381); Whitt (1993, p. 122).

Approximations for the average time in queue W_q for *G/G/1* and *G/G/m* queues are given in Sections 7.6 and 7.9.

7.11 Queue Departure Process

For a *G/G/1* queue, let

ν = departure rate (number of customers departing after service per unit time); ($1/\nu$ = mean inter-departure time)

σ_d^2 = variance of inter-departure times

$C_d^2 = \nu^2\sigma_d^2$ = squared coefficient of variation of inter-departure times

For traffic intensity $\rho < 1$, the departure rate must equal the arrival rate, i.e.,

$$\nu = \lambda \tag{7.27}$$

The squared coefficient of variation of inter-departure times C_d^2 is given approximately by

$$C_d^2 \cong \rho^2 C_s^2 + (1 - \rho^2) C_a^2 \tag{7.28}$$

Hopp and Spearman (1996, pp. 268–269); Whitt (1983, p. 2799).

7.12 Distribution Results for the Number of Customers in *M*/*M*/1 Queue

	Number of Customers in Queue		Number of Customers in System (Including One Being Served)	
Mean $E[n]$	$\frac{\rho^2}{1-\rho}$	(7.29)	$\frac{\rho}{1-\rho}$	(7.33)
Variance $Var[n]$	$\frac{\rho^2\left(1+\rho-\rho^2\right)}{(1-\rho)^2}$	(7.30)	$\frac{\rho}{(1-\rho)^2}$	(7.34)
Probability distribution Pr{# in queue = n}	$\begin{cases} 1-\rho^2 & (n=0) \\ (1-\rho)\rho^{n+1} & (n \geq 1) \end{cases}$	(7.31)	$(1-\rho)\rho^n$	(7.35)
Cumulative distribution Pr{# in queue ≤ n}	$1-\rho^{n+2}$	(7.32)	$1-\rho^{n+1}$	(7.36)

where

$$\rho = \frac{\lambda}{\mu} = \frac{\text{arrival rate}}{\text{service rate}} = \text{traffic intensity} \quad (\rho < 1)$$

Allen (1978, pp. 161–162, 365); Cohen (1985, p. 173); Cox and Smith (1961, p. 41); Hillier and Lieberman (1980, pp. 418–419); Morse (1958, p. 22); Wolff (1989, p. 148).

Note that the distribution of the number of customers in a system, given by Equations 7.33 through 7.36, is a geometric distribution with parameter $1 - \rho$ (see Section 3.3).

7.13 Distribution Results for Time in *M*/*M*/1 Queue

	Time in Queue		Time in System (Time in Queue Plus Service Time)	
Mean $E[t]$	$\frac{\rho}{\mu(1-\rho)}$	(7.37)	$\frac{1}{\mu(1-\rho)}$	(7.41)
Variance $Var[t]$	$\frac{(2-\rho)\rho}{\mu^2(1-\rho)^2}$	(7.38)	$\frac{1}{\mu^2(1-\rho)^2}$	(7.42)

(*continued*)

	Time in Queue		Time in System (Time in Queue Plus Service Time)	
Probability density function $f(t)$ $(t \geq 0)$	$(1-\rho)u_0(t)+\rho\mu(1-\rho)e^{-\mu(1-\rho)t}$	(7.39)	$\mu(1-\rho)e^{-\mu(1-\rho)t}$	(7.43)
Cumulative distribution function $F(t)$ Pr{time in queue ≤ t}	$1-\rho e^{-\mu(1-\rho)t}$	(7.40)	$1-e^{-\mu(1-\rho)t}$	(7.44)

where

$$\rho = \frac{\lambda}{\mu} = \frac{\text{arrival rate}}{\text{service rate}} = \text{traffic intensity} \quad (\rho < 1)$$

$u_0(t)$ is the unit impulse function centered at $t=0$, given by

$$u_0(t) = \begin{cases} \infty & (t = 0) \\ 0 & (t \neq 0) \end{cases}$$

and

$$\int_{-\infty}^{\infty} u_0(t)\,dt = 1$$

Allen (1978, pp. 163, 365); Cohen (1985, p. 173); Cox and Smith (1961, pp. 57–58); Hillier and Lieberman (1980, p. 420); Kleinrock (1975, pp. 202–203); Papadopoulos, Heavy, and Browne (1993, pp. 363–364).

Note that the distribution of the time in system, given by Equations 7.41 through 7.44, is an exponential distribution with parameter $\mu(1-\rho)$ (see Section 4.2).

7.14 Other Formulas in Queueing Theory

A comprehensive summary of formulas for a variety of different types of queueing systems is given in Allen (1978). The formulas include means, variances, and probability distributions for queue length, waiting time, and other system measures. Summaries of basic queueing theory results are also given in Bolch, Greiner, de Meer, and Trivedi (1998) and in Papadopoulos, Heavy, and Browne (1993).

8

Production Systems Modeling

8.1 Definitions and Notation for Workstations

The number of jobs a station can produce per unit time is the *speed* (or *service rate*) of the station. The production of one job is a *cycle*. The time to produce one job, when there is no station failure, is the *cycle time* (or *service time* or *processing time*).

For a station that is subject to failures, the frequency of failures is determined by the *failure rate*, and the time taken on average for repair is determined by the *repair rate*. The ratio of the operating time to total time is the station's *availability* (or *stand-alone availability*). The number of jobs the station can produce per unit time, taking account of failures, is the station's *throughput* (or *stand-alone throughput*).

Station parameters:

S = speed of station (number of jobs per unit time)
c = cycle time of station
λ = failure rate (number of failures per unit time)
μ = repair rate (number of repairs per unit time)
$MCBF$ = mean number of cycles between failures
$MTBF$ = mean operating time between failures
$MTTR$ = mean time to repair (mean down time)

8.2 Basic Relationships between Workstation Parameters

For a single station:

$$\text{Cycle time } \ c = \frac{1}{S} \tag{8.1}$$

$$MTBF = c \times MCBF = \frac{MCBF}{S} \tag{8.2}$$

$$MTBF = \frac{1}{\lambda} \tag{8.3}$$

$$MTTR = \frac{1}{\mu} \tag{8.4}$$

$$\text{Availability} = \frac{MTBF}{MTBF + MTTR} = \frac{1}{1 + \dfrac{\lambda}{\mu}} \tag{8.5}$$

$$\text{Throughput} = (\text{Availability} \times S) \tag{8.6}$$

Buzacott (1968, pp. 176, 180); Choong and Gershwin (1987, p. 152); Cohen (1985, pp. 185, 191); Goldman and Slattery (1964, pp. 26, 41); Hopp and Spearman (1996, p. 261); Li and Meerkov (2009, pp. 53, 95, 361); Nahmias (1989, p. 557).

Note that the *MCBF* is an average count of the number of jobs produced between failures, which is generally much easier to measure in practice than the *MTBF*. From Equation 8.2, the *MTBF* can be obtained without the need to measure time between failures directly.

8.3 Distribution of the Time to Produce a Fixed Lot Size at a Workstation

For a single station with random failures and random repair times, let

S be the station speed (jobs per unit time)
λ be the failure rate for the station
μ be the repair rate for the station
n be the lot size (number of jobs)
T_n be the time to produce n jobs (T_n a random variable, n fixed)

Assuming

- Processing times (cycle times) are constant
- Operating times between failures are independent and exponentially distributed (with a mean $1/\lambda$)
- Repair times are independent and exponentially distributed (with a mean $1/\mu$)

the mean and the variance of the time, T_n, are given by

$$E[T_n] = \frac{n}{S}\left(1 + \frac{\lambda}{\mu}\right) \tag{8.7}$$

and

$$Var\left[T_n\right] = \frac{2n\lambda}{S\mu^2} \tag{8.8}$$

respectively.

Kim and Alden (1997, p. 3411).

From Equations 8.7 and 8.8, the mean and the variance of the time, T_1, to produce one job ($n=1$) are

$$E\left[T_1\right] = \frac{1}{S}\left(1+\frac{\lambda}{\mu}\right)$$

and

$$Var\left[T_1\right] = \frac{2\lambda}{S\mu^2}$$

respectively. In the more general case, where processing time is a random variable, with a mean $1/S$ and a variance σ^2, the mean of time T_1 remains the same, and the variance of time T_1 has an additional term and becomes

$$Var\left[T_1\right] = \frac{2\lambda}{S\mu^2} + \sigma^2\left(1+\frac{\lambda}{\mu}\right)^2 \tag{8.9}$$

Hopp and Spearman (1996, p. 262).

For constant processing times, the probability density function $f(t)$ of the time T_n to produce n jobs is given by

$$f(t) = \begin{cases} 0 & (t < n/S) \\ u_0(t-n/S)e^{-\lambda n/S} + \dfrac{(\lambda\mu n/S)I_1\left(2\sqrt{x}\right)e^{-\lambda n/S-\mu(t-n/S)}}{\sqrt{x}} & (t \geq n/S) \end{cases} \tag{8.10}$$

where

$$x = (\lambda\mu n/S)(t-n/S)$$

$I_1(x)$ is a modified Bessel function of order one (see Section 15.6)
$u_0(t)$ is the unit impulse function centered at $t=0$, given by

$$u_0(t) = \begin{cases} \infty & (t = 0) \\ 0 & (t \neq 0) \end{cases}$$

and

$$\int_{-\infty}^{\infty} u_0(t)dt = 1$$

Kim and Alden (1997, pp. 3407, 3409).

8.4 Throughput of a Serial Production Line with Failures

The average number of jobs per unit time that can flow through a production line is the line's *throughput* (or *production rate*).

8.4.1 Line without Buffers

For a production line with stations arranged in series, let
N be the number of stations
S be the speed (service rate) of a station (jobs per unit time)
λ_i be the failure rate of station i ($i=1, 2, \ldots, N$)
μ_i be the repair rate of station i ($i=1, 2, \ldots, N$)
P be the throughput of the line (jobs per unit time)

Assuming
- Stations have the same speed
- Processing times (cycle times) are constant
- Station failures are independent
- Repair times are independent
- A failure at one station stops the entire line
- There are no buffers between stations

the throughput of the line (average number of jobs per unit time) is given by

$$P = \frac{S}{1 + \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i}} \tag{8.11}$$

Buzacott (1968, p. 176); Gershwin (1994, p. 66).

8.4.2 Line with Buffers

For a serial production line with buffers of equal size between adjacent stations, where all stations are identical (i.e., have the same speed, failure rate, and repair rate), let

N be the number of stations
S be the speed (service rate) of a station (jobs per unit time)
λ be the failure rate of a station
μ be the repair rate of a station
B be the buffer size (number of jobs that can be held in the buffer)
P be the throughput of the line (jobs per unit time)

Assuming

- Stations are identical
- Processing times (cycle times) are constant
- Station failures are independent
- Repair times are independent

the throughput of the line (average number of jobs per unit time) for a two-station line ($N=2$) is given by

$$P = \frac{S}{1 + \frac{\lambda}{\mu} + \frac{\lambda/\mu}{1 + \frac{1}{2}\left(1 + 2\lambda/\mu\right)\left(B\mu/S\right)}} \tag{8.12}$$

and for an N-station line is approximately given by

$$P \cong \frac{S}{1 + \frac{\lambda}{\mu} + \frac{(N-1)\lambda/\mu}{1 + \frac{N}{4}\left(1 + 2\lambda/\mu\right)\left(B\mu/S\right)}} \tag{8.13}$$

Alden (2002, pp. 84, 107, 112); Blumenfeld and Li (2005, pp. 298, 300).

8.5 Throughput of a Two-Station Serial Production Line with Variable Processing Times

The following results are for a production line with no failures and with processing times at each station that are random variables.

8.5.1 Two Stations without a Buffer

For a production line with two stations arranged in series, let
S_i be the speed (service rate) of a station i ($i=1, 2$) (jobs per unit time)
P be the throughput of the line (jobs per unit time)

Assuming
- Processing times at station i are independent and exponentially distributed (with a mean $1/S_i$) ($i=1, 2$)
- Stations are not subject to failures
- There is no buffer between the two stations

the throughput of the line (average number of jobs per unit time) is given by

$$P = \frac{1}{\frac{1}{S_1} + \frac{1}{S_2} - \frac{1}{S_1 + S_2}} \tag{8.14}$$

i.e.,

$$P = \frac{S_1 S_2 (S_1 + S_2)}{S_1^2 + S_1 S_2 + S_2^2} \tag{8.15}$$

In the special case of identical stations ($S_1 = S_2 = S$), the throughput result reduces to

$$P = \frac{2}{3} S \tag{8.16}$$

Baker (1992, p. 387); Makino (1964, p. 21); Hunt (1956, pp. 678–679).

8.5.2 Two Stations with a Buffer

A buffer between stations holds jobs that have been processed at one station and are waiting to be processed at the next station. For a production line with two stations arranged in series, let

S_i be the speed (service rate) of a station i ($i=1, 2$) (jobs per unit time)
B be the buffer size (number of jobs that can be held in the buffer)
P be the throughput of the line (jobs per unit time)

Assuming

- Processing times at station i are independent and exponentially distributed (with a mean $1/S_i$) ($i=1, 2$)
- Stations are not subject to failures

the throughput of the line (average number of jobs per unit time) is given by

$$P = S_1 S_2 \left(\frac{S_1^{B+2} - S_2^{B+2}}{S_1^{B+3} - S_2^{B+3}} \right) \tag{8.17}$$

In the special case of identical stations ($S_1=S_2=S$), this throughput result reduces to

$$P = \left(\frac{B+2}{B+3} \right) S \tag{8.18}$$

Hillier and Boling (1966, p. 657); Hunt (1956, p. 680).

8.6 Throughput of an *N*-Station Serial Production Line with Variable Processing Times

The following is an approximate result for throughput of a serial production line with identical stations.

For a production line with stations arranged in series, let
N be the number of stations
S be the speed (service rate) of a station (jobs per unit time)
$\overline{T}=1/S$ be the mean processing time of a station
σ^2 be the variance of processing times
$C=\sigma/\overline{T}$ be the coefficient of variation of processing times
P be the throughput of the line (jobs per unit time)

Assuming

- Processing times are independent random variables (with an arbitrary distribution)
- Stations are identical (same mean and variance of processing times)
- Stations are not subject to failures
- There are no buffers between stations

the throughput of the line (average number of jobs per unit time) is given by Muth's approximate formula:

$$P \cong \frac{S}{1+\dfrac{1.67\,(N-1)C}{1+N+0.31C}} \tag{8.19}$$

Baker (1992, p. 388); Blumenfeld (1990, p. 1165); Muth (1987, p. 7).

For a line with buffers between the stations, where the buffer sizes are equal, an extension to Muth's approximation is given by

$$P \cong \frac{S}{1+\dfrac{1.67\,(N-1)C}{1+N+0.31C+1.67NB/(2C)}} \tag{8.20}$$

where B is the buffer size.

Askin and Standridge (1993, p. 87); Blumenfeld (1990, p. 1169).

9

Inventory Control

9.1 Economic Order Quantity

The economic order quantity (EOQ) is the optimal quantity to order to replenish inventory, based on a trade-off between inventory and ordering costs. The trade-off analysis assumes the following:

Demand for items from inventory is continuous and at a constant rate.
Orders are placed to replenish inventory at regular intervals.
Ordering cost is fixed (independent of quantity ordered).
Replenishment is instantaneous.

Let

D = demand (number of items per unit time)
A = ordering cost ($ per order)
c = cost of an item ($ per item)
r = inventory carrying charge (fraction per unit time)
$H = cr$ = holding cost of an item ($ per item per unit time)
Q = order quantity (number of items per order)

Figure 9.1 plots cumulative curves of orders and demand over time. The curve for orders increases in steps of size Q each time an order is placed, and the demand curve increases linearly with slope D. The vertical distance between these two curves at any point in time is the inventory level. Figure 9.2 plots this inventory level over time, which displays the classic saw-tooth pattern. The inventory increases by Q each time an order is placed, and decreases at rate D between orders. The average inventory level in Figure 9.2 is $Q/2$, which determines the inventory cost in the EOQ model.

Total cost per unit time $C(Q)$ is given by

$$\begin{aligned} C(Q) &= \text{Inventory cost} + \text{Ordering cost} \\ &= \frac{HQ}{2} + \frac{AD}{Q} \end{aligned} \tag{9.1}$$

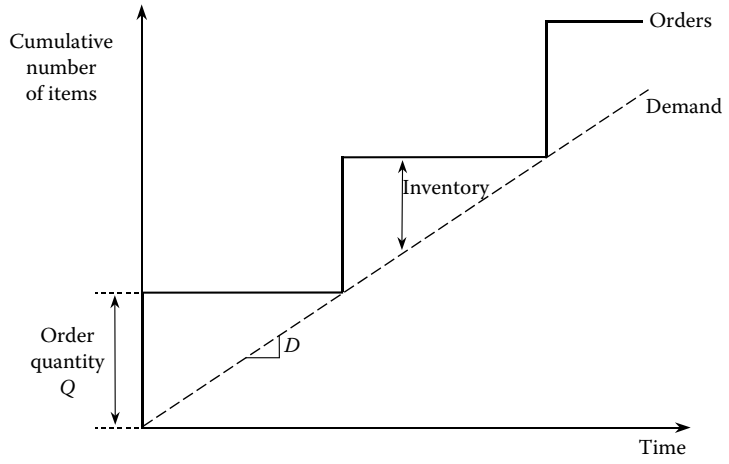


FIGURE 9.1
Cumulative orders and demand over time in an EOQ model.

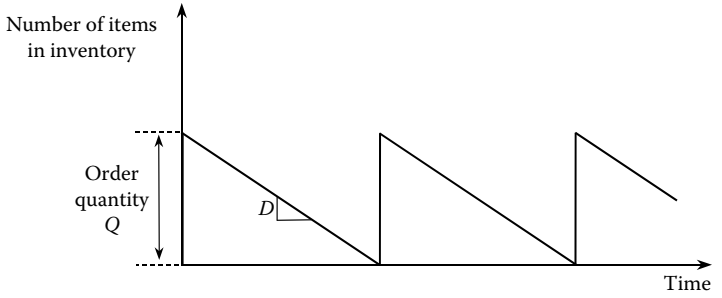


FIGURE 9.2
Inventory level over time in an EOQ model.

The optimal quantity Q^* to order (i.e., the order quantity that minimizes total cost) is given by

$$\frac{d}{dQ}C(Q)=0$$

Hence

$$Q^*=\sqrt{\frac{2AD}{H}} \tag{9.2}$$

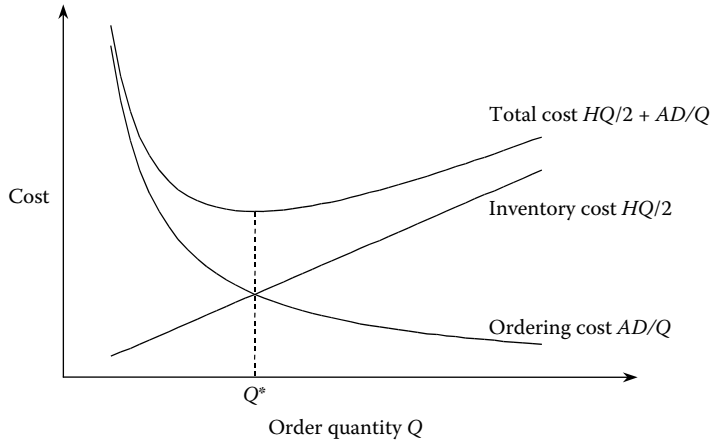


FIGURE 9.3
Trade-off between inventory and ordering costs in an EOQ model.

Equation 9.2 for Q^* is known as the EOQ formula. Figure 9.3 illustrates the trade-off between the inventory and ordering costs.

Arrow, Karlin, and Scarf (1958, p. 6); Cohen (1985, p. 144); Harris (1913, p. 136); Hax and Candea (1984, p. 134); Hopp and Spearman (1996, p. 58); Nahmias (1989, p. 148); Stevenson (1986, p. 480); Tersine (1985, p. 590); Wilson (1934, p. 122); Woolsey and Swanson (1975, p. 39).

9.2 Economic Production Quantity

The economic production quantity (EPQ) is the optimal quantity to produce to replenish inventory, based on a trade-off between inventory and production set-up costs. The trade-off analysis assumes the following:

- Demand for items from inventory is continuous and at a constant rate.
- Production runs to replenish inventory are made at regular intervals.
- During a production run, the production of items is continuous and at a constant rate.
- Production set-up cost is fixed (independent of the quantity produced).

The EPQ model is similar to that for the EOQ model. The difference is in the time to replenish inventory. The EOQ model assumes replenishment is instantaneous, while the EPQ model assumes replenishment is gradual, due to a finite production rate.

Let

D = demand (number of items per unit time)
P = production rate during a production run (number of items per unit time)
A = production set-up cost ($ per setup)
c = cost of an item ($ per item)
r = inventory carrying charge (fraction per unit time)
$H = cr$ = holding cost of an item ($ per item per unit time)
Q = production quantity (number of items per production run)

The EPQ model assumes $P > D$. Figure 9.4 plots cumulative curves of production and demand over time. The slope of the production curve during a production run is P. The slope of the demand curve is D. The vertical distance between these two curves at any point in time is the inventory level. Figure 9.5 plots this inventory level over time. Inventory increases at rate $P–D$ during a production run, and decreases at rate D between production runs. The average inventory level in Figure 9.5 is

$$\frac{(1 - D/P)Q}{2}$$

which determines the inventory cost in the EPQ model.

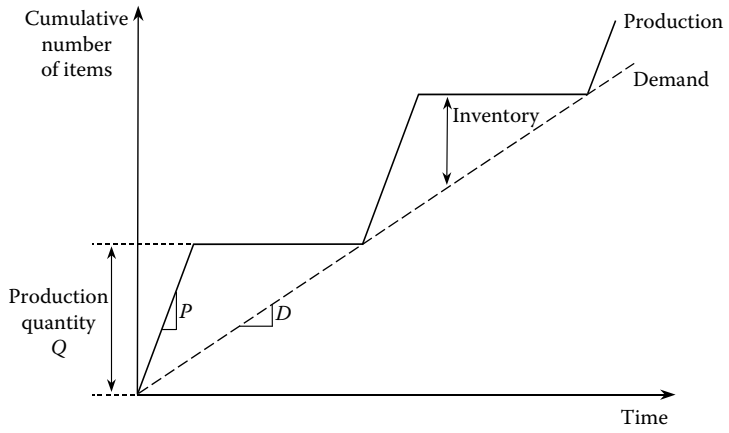


FIGURE 9.4
Cumulative production and demand over time in an EPQ model.

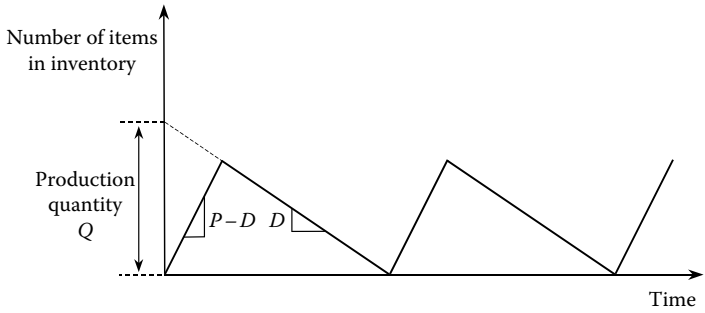


FIGURE 9.5
Inventory level over time in an EPQ model.

Total cost per unit time $C(Q)$ is given by

$$C(Q) = \text{Inventory cost} + \text{Production set-up cost}$$

$$= \frac{H(1-D/P)Q}{2} + \frac{AD}{Q} \tag{9.3}$$

The optimal quantity Q^* to produce (i.e., the production quantity that minimizes total cost) is given by

$$\frac{d}{dQ}C(Q) = 0$$

and hence

$$Q^* = \sqrt{\frac{2AD}{H(1-D/P)}} \tag{9.4}$$

Equation 9.4 for Q^* is known as the EPQ formula.

The trade-off between inventory and production set-up costs is the same as for the EOQ model illustrated in Figure 9.3, except for a different slope for the linear inventory cost curve.

Note: As P approaches infinity, replenishment becomes instantaneous, and the EPQ formula given by Equation 9.4 reduces to the EOQ formula given by Equation 9.2.

Hax and Candea (1984, p. 136); Hopp and Spearman (1996, p. 63); Nahmias (1989, p. 154); Stevenson (1986, p. 483); Taft (1918, p. 1411); Tersine (1985, p. 593).

9.3 "Newsboy Problem": Optimal Inventory to Meet Uncertain Demand in a Single Period

The classic *newsboy problem* (also called more recently the *newsvendor problem*) considers the inventory level needed under uncertain demand. The problem is to determine the optimal number of items to hold in inventory to meet uncertain demand in a single period. The optimum is given by the trade-off between cost of

- Holding too many items
- Not meeting demand

Let

c_o = cost per item of items left over after demand is met (overage cost per item)
c_s = cost per item of unmet demand (shortage cost per item)
x = demand in given period (number of items)
$f(x)$ = probability density function (pdf) of demand
$F(x) = \int_0^x f(u)du$ = cumulative distribution function of demand
Q = quantity held in inventory (number of items)

The optimal cost trade-off depends on the expected numbers of items over demand and short of demand.

Expected cost $C(Q)$ is given by

$$C(Q) = c_o \text{ E[number of items over]} + c_s \text{ E[number of items short]}$$

$$= c_o \int_0^Q (Q - x) f(x) dx + c_s \int_Q^\infty (x - Q) f(x) dx \tag{9.5}$$

The optimal quantity, Q^*, to hold in inventory (i.e., the quantity that minimizes expected cost) is given by

$$\frac{d}{dQ} C(Q) = 0$$

Applying Leibnitz's rule for differentiation under the integral sign (see Section 16.8, Equation 16.23), we get

$$\frac{d}{dQ}C(Q) = c_o\int_0^Q \frac{\partial}{\partial Q}\{(Q-x)f(x)\}dx + c_s\int_Q^\infty \frac{\partial}{\partial Q}\{(x-Q)f(x)\}dx$$

$$= c_o\int_0^Q f(x)dx - c_s\int_Q^\infty f(x)dx$$

$$= c_oF(Q) - c_s\{1-F(Q)\}$$

$$= (c_o + c_s)F(Q) - c_s$$

Hence, setting $\frac{d}{dQ}C(Q) = 0$, the optimal quantity, Q^*, is given by

$$F(Q^*) = \frac{c_s}{c_s + c_o} \tag{9.6}$$

Equation 9.6 is the solution to the classic *newsboy problem* or *newsvendor problem*.

The overage and shortage costs, c_o and c_s, can be expressed in terms of the following economic parameters. Let

c = cost per item
a = selling price per item
p = lost sales penalty per item
v = salvage value per item

The profit for each item sold is $a-c$. Hence, the lost profit per item for unmet demand is $a-c$. An additional cost of unmet demand is the lost sales penalty p, representing loss of some customers in future periods. Hence, the shortage cost, c_s, is

$$c_s = a - c + p$$

For unsold items left over after demand is met, the net cost per item is the cost minus the salvage value. Hence, the overage cost c_o is

$$c_o = c - v$$

From Equation 9.6, the optimal quantity Q^* given by

$$F(Q^*) = \frac{a+p-c}{a+p-v} \tag{9.7}$$

Hanssmann (1962, p. 47); Hopp and Spearman (1996, p. 75); Nahmias (1989, p. 233); Ravindran, Phillips, and Solberg (1987, p. 355).

9.4 Inventory Replenishment Policies

Figures 9.6 through 9.9 illustrate the following basic policies for replenishing inventory in continuous review and periodic review systems:

- **Continuous Review Systems**

(*s*, *Q*) Policy: Whenever the inventory position (items on hand plus items on order) drops to a given level, *s*, or below, an order is placed for a fixed quantity, *Q*.

(*s*, *S*) Policy: Whenever the inventory position (items on hand plus items on order) drops to a given level, *s*, or below, an order is placed for a sufficient quantity to bring the inventory position up to a given level, *S*.

- **Periodic Review Systems**

(*T*, *S*) Policy: Inventory position (items on hand plus items on order) is reviewed at regular instants, spaced at time intervals of length *T*. At each review, an order is placed for a sufficient quantity to bring the inventory position up to a given level, *S*.

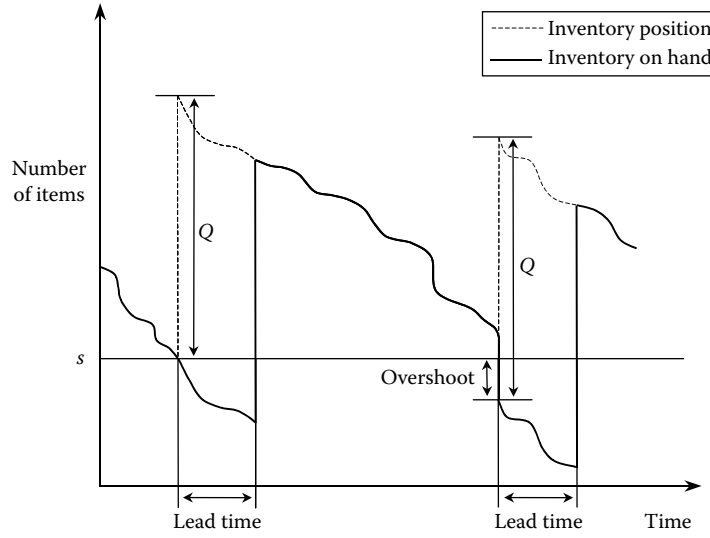


FIGURE 9.6
Inventory pattern over time in an (*s*, *Q*) policy.

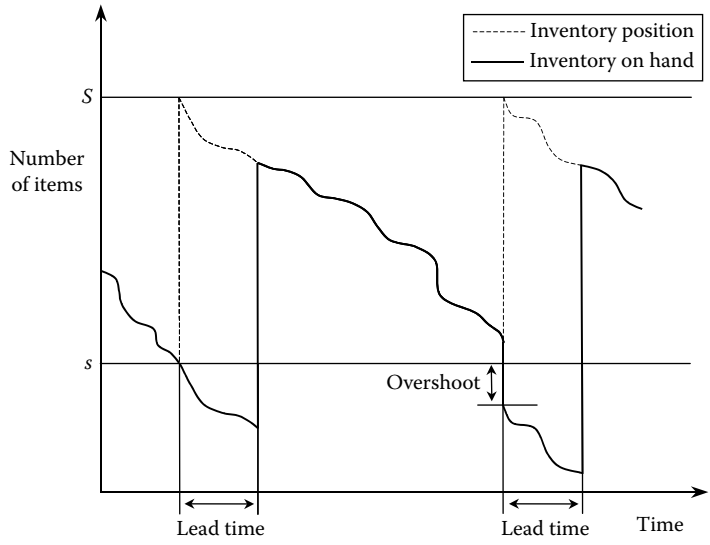


FIGURE 9.7
Inventory pattern over time in an (*s*, *S*) policy.

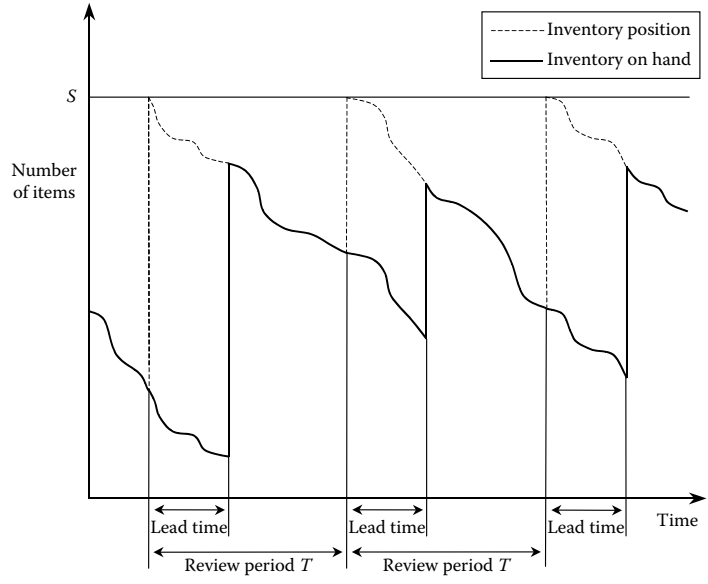


FIGURE 9.8
Inventory pattern over time in a (*T*, *S*) policy.

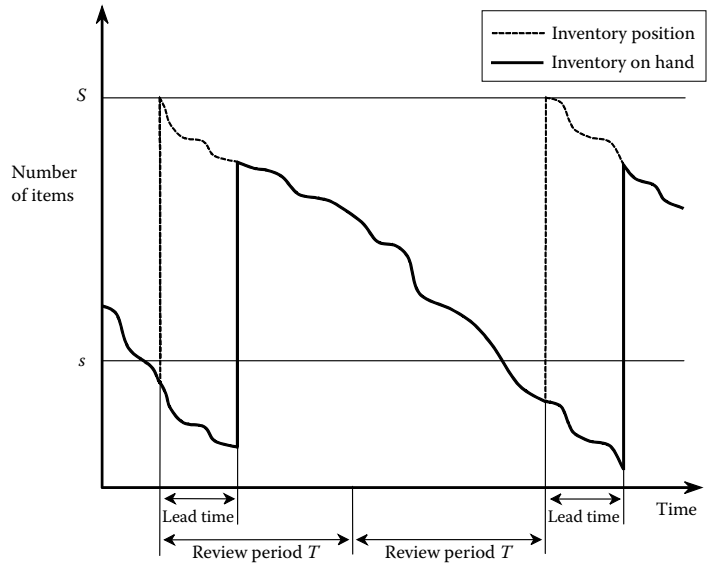


FIGURE 9.9
Inventory pattern over time in a (*T*, *s*, *S*) policy.

(*T*, *s*, *S*) Policy: Inventory position (items on hand plus items on order) is reviewed at regular instants, spaced at time intervals of length *T*. At each review, if the inventory position is at level *s* or below, an order is placed for a sufficient quantity to bring the inventory position up to a given level *S*. If the inventory position is above *s*, no order is placed. This policy is also known as a *periodic review* (*s*, *S*) *policy*.

Elsayed and Boucher (1985, pp. 58–60); Hadley and Whitin (1963, pp. 236–237); Hax and Candea (1984, p. 220); Johnson and Montgomery (1974, pp. 23–25); Silver, Pyke, and Peterson (1998, pp. 237–241).

The quantities *Q*, *s*, *S*, and *T* in these policies are defined as follows:
Q = order quantity
s = reorder point
S = order-up-to level
T = review period (time interval between reviews)

The notation for these quantities varies in the inventory literature. For example, some references denote the reorder point by *R*, while other references use *R* for the order-up-to level, and still others use *R* for the review period. The notation defined above is intended to avoid ambiguity, while being consistent with the notation frequently used in the literature.

Inventory position is the sum of inventory on hand (i.e., items immediately available to meet demand) and inventory on order (i.e., items

ordered but not yet arrived due to the lead time). The above policies for replenishment are based on inventory position, rather than simply inventory on hand, to account for cases where the lead time is longer than the time between replenishments. If the lead time is always shorter than the time between replenishments, then there will never be any items on order at the time an order is placed, and in that case the review of inventory can be based simply on the inventory on hand (Evans, Anderson, Sweeney, and Williams, 1984, p. 381; Johnson and Montgomery, 1974, pp. 24–25).

A note on the continuous review systems: If demand occurs one item at a time, then the (s, S) policy is the same as the (s, Q) policy. If, however, demand can occur in batches, so that the inventory position can drop from a level above s to a level below s instantaneously (i.e., an overshoot can occur), then the (s, Q) and (s, S) policies are different. A comparison of Figures 9.6 and 9.7 illustrates the difference. In the (s, Q) policy, the order quantity is fixed, and the inventory position just after a replenishment order is placed is variable from one replenishment cycle to another. In the (s, S) policy, the inventory position just after a replenishment order is placed is fixed, and the order quantity is variable (Hax and Candea, 1984, pp. 222–223; Silver, Pyke, and Peterson, 1998, p. 238).

The (s, S) policy is a special case of the (T, s, S) policy in which $T=0$. The (T, s, S) policy can thus be regarded as a periodic version of the (s, S) policy. The (T, S) policy represents a special case of the (T, s, S) policy in which $s=S$ (Johnson and Montgomery, 1974, p. 24; Silver, Pyke, and Peterson, 1998, p. 241).

9.5 (s, Q) Policy: Estimates of Reorder Point (s) and Order Quantity (Q)

Replenishment policy: Whenever the inventory position (items on hand plus items on order) drops to the reorder point s or below, an order is placed for a fixed quantity. Figure 9.6 illustrates the (s, Q) policy.

Assume:

- Demand for items is a random variable with fixed mean and variance.
- Demands in separate increments of time are independent.
- Lead time (i.e., time from when an order for replenishment is placed until the replenishment arrives) is a random variable with fixed mean and variance.
- Lead times are independent.

Let

s = reorder point (number of items)
Q = order quantity (number of items)
D = average demand (number of items per unit time)
σ_D^2 = variance of demand (items^2 per unit time)
L = average lead time (units of time)
σ_L^2 = variance of lead time (units of time^2)
k = service level factor
A = ordering cost ($ per order)
H = holding cost of an item ($ per item per unit time)

The demand variance, σ_D^2, is defined for demand in one time unit. Since the demands in each time unit are assumed to be independent, the variance of demand in a fixed time of t units is $\sigma_D^2 t$.

The reorder point, s, and order quantity, Q, in the (s, Q) policy are given approximately by

$$s = DL + k\sqrt{L\sigma_D^2 + D^2\sigma_L^2} \tag{9.8}$$

$$Q = \sqrt{\frac{2AD}{H}} \tag{9.9}$$

Lewis (1970, pp. 50, 60); McClain and Thomas (1985, pp. 293, 304); Silver, Pyke, and Peterson (1998, pp. 255, 258, 283); Sipper and Bulfin (1997, pp. 272, 276–277); Stevenson (1986, pp. 480, 501, 514).

In the special case of fixed lead times $\sigma_L^2 = 0$ and Equation 9.8 for the reorder point s reduces to

$$s = DL + k\sigma_D\sqrt{L} \tag{9.10}$$

The order quantity Q, given by Equation 9.9, is the EOQ, as given in Section 9.1.

The reorder point s, given by Equation 9.8, is the inventory level needed to cover demand during the lead time. The first term, DL, is the inventory needed on average. The second term, $k\sqrt{L\sigma_D^2 + D^2\sigma_L^2}$, is the additional inventory needed to avoid stocking-out due to random variability in demand and lead time. This additional inventory is the safety stock, i.e.,

$$\text{Safety Stock} = k\sqrt{L\sigma_D^2 + D^2\sigma_L^2} \tag{9.11}$$

The above two terms for s are based on the result that demand during the lead time has a mean, DL, and a standard deviation, $\sqrt{L\sigma_D^2 + D^2\sigma_L^2}$ (Hadley and Whitin, 1963, p. 153).

The service level factor, k, in Equations 9.8 and 9.11 is a dimensionless constant that represents the number of standard deviations beyond the mean, DL, needed to achieve a given service level (i.e., a given measure of performance for meeting demand from inventory). The service level is typically measured using one of the following two quantities—α and β:

α=probability of meeting demand from inventory

β=fraction of demand met from inventory (also known as "fill rate").

The probability, α, is the proportion of replenishment cycles in which no shortage occurs (regardless of the number of items short, when a shortage does occur). The fill rate β is the proportion of total items demanded that are filled from inventory (regardless of the number of replenishment cycles in which a shortage occurs).

If the demand during the lead time has a general distribution with probability density function (pdf) denoted by $f_l(x)$, then the quantities α and β are given by

$$\alpha = 1 - \int_{s}^{\infty} f_l(x)dx \tag{9.12}$$

and

$$\beta = 1 - \frac{1}{Q}\int_{s}^{\infty} (x - s) f_l(x)dx \tag{9.13}$$

where s is related to the service level factor, k, by Equation 9.8.

If the demand during the lead time is normally distributed, then the quantities α and β are related to the service level factor, k, by

$$\alpha = \Phi(k) \tag{9.14}$$

and

$$\beta = 1 - \frac{\sigma_l}{Q}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}k^2\right) - k\left[1 - \Phi(k)\right]\right\} \tag{9.15}$$

where $\Phi(k)$ is the cumulative distribution function of the standard normal distribution, i.e.,

$$\Phi(k) = \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)dx \tag{9.16}$$

and where σ_l is the standard deviation of demand during the lead time, i.e.,

$$\sigma_l = \sqrt{L\sigma_D^2 + D^2\sigma_L^2} \tag{9.17}$$

Fortuin (1980, pp. 939–940); Nahmias (1989, p. 202); Schneider (1981, pp. 620–621); Sipper and Bulfin (1997, pp. 275, 607–608).

To ensure a high value of the probability α, the service level factor k is typically set in the range of 2–3. From Equation 9.16, when $k=2$, $\alpha=97.7\%$, and when $k=3$, $\alpha=99.9\%$ (Lewis, 1970, p. 47; Mood, Graybill, and Boes, 1974, p. 552; Sipper and Bulfin, 1997, p. 607).

The relationship between the fill rate, β, and the service level factor, k, is more complex than that for α, since it depends also on the order quantity, Q. The EOQ given by Equation 9.9 provides a useful heuristic approximation for Q, which allows s to be estimated separately from Q. For a given shortage cost per item, s and Q can also be optimized jointly (Hadley and Whitin, 1963, p. 167; Nahmias, 1989, p. 204; Sipper and Bulfin, 1997, p. 284).

Note: If demand in each time unit is normally distributed, and the lead time is constant ($\sigma_L^2 = 0$), then the demand during the lead time is normally distributed. If, however, demand in each time unit is normally distributed, and the lead time is variable ($\sigma_L^2 > 0$), then in general the demand during the lead time is not normally distributed. In the case of variable lead time, therefore, the relationships between the safety level factor, k, and the quantities α and β, given by Equations 9.14 and 9.15, may not be sufficiently close approximations.

9.6 (*s*, *S*) Policy: Estimates of Reorder Point (*s*) and Order-Up-To Level (*S*)

Replenishment policy: Whenever the inventory position (items on hand plus items on order) drops to the reorder point, s, or below, an order is placed for a sufficient quantity to raise the inventory position to the order-up-to level S. Figure 9.7 illustrates the (s, S) policy.

Assume:

- Demand for items is a random variable with fixed mean and variance.
- Demands in separate increments of time are independent.

- Lead time (i.e., time from when an order for replenishment is placed until the replenishment arrives) is a random variable with a fixed mean and variance.
- Lead times are independent.

Let

s = reorder point (number of items)
S = order-up-to level (number of items)
D = average demand (number of items per unit time)
σ_D^2 = variance of demand (items^2 per unit time)
L = average lead time (units of time)
σ_L^2 = variance of lead time (units of time^2)
k = service level factor
A = ordering cost ($ per order)
H = holding cost of an item ($ per item per unit time)

The demand variance, σ_D^2, is defined for a demand in one time unit. Since the demands in each time unit are assumed to be independent, the variance of demand in a fixed time of t units is $\sigma_D^2 t$.

The reorder point, s, and order-up-to level, S, in the (s, S) policy are given approximately by

$$s = DL + k\sqrt{L\sigma_D^2 + D^2\sigma_L^2} \tag{9.18}$$

$$S = s + Q \tag{9.19}$$

where

$$Q = \sqrt{\frac{2AD}{H}} \tag{9.20}$$

Hax and Candea (1984, p. 223); Silver, Pyke, and Peterson (1998, pp. 255, 331–332).

The reorder point s, given by Equation 9.18, is the inventory level needed to cover demand during the lead time. This expression for s is based on the result that demand during the lead time has a mean, DL, and a standard deviation, $\sqrt{L\sigma_D^2 + D^2\sigma_L^2}$ (Hadley and Whitin, 1963, p. 153).

The service level factor k, given by Equation 9.18, is a dimensionless constant that represents the number of standard deviations demand beyond the mean, DL, needed to achieve a given service level (see the (s, Q) policy above).

The order-up-to level S, given by Equation 9.19, is an heuristic estimate based simply on the reorder point plus the EOQ, where the EOQ is given in Section 9.1.

9.7 (*T*, *S*) Policy: Estimates of Review Period (*T*) and Order-Up-To Level (*S*)

Replenishment policy: Inventory position (items on hand plus items on order) is reviewed at regular instants, spaced at time intervals of length T. At each review, an order is placed for a sufficient quantity to raise the inventory position to the order-up-to level S. Figure 9.8 illustrates the (T, S) policy.

Assume:

- Demand for items is a random variable with a fixed mean and variance.
- Demands in separate increments of time are independent.
- Lead time (i.e., time from when an order for replenishment is placed until the replenishment arrives) is a random variable with a fixed mean and variance.
- Lead times are independent.
- Review period (i.e., time interval between reviews) is a constant.

Let

T = review period (units of time)
S = order-up-to level (number of items)
D = average demand (number of items per unit time)
σ_D^2 = variance of demand (items^2 per unit time)
L = average lead time (units of time)
σ_L^2 = variance of lead time (units of time^2)
k = service level factor
A = ordering cost ($ per order)
H = holding cost of an item ($ per item per unit time)

The ordering cost, A, in this policy includes the cost, if any, of reviewing of the inventory position in each review period. The demand variance, σ_D^2, is defined for a demand in one time unit. Since the demands in each time unit are assumed to be independent, the variance of demand in a fixed time of t units is $\sigma_D^2 t$.

The review period, T, and order-up-to level, S, in the (T, S) policy are given approximately by

$$T = \sqrt{\frac{2A}{DH}} \tag{9.21}$$

$$S = D(L+T) + k\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2} \tag{9.22}$$

Hax and Candea (1984, pp. 227–228); Lewis (1970, p. 84); McClain and Thomas (1985, p. 309); Silver, Pyke, and Peterson (1998, pp. 276, 279); Sipper and Bulfin (1997, pp. 291–292).

In the special case of fixed lead times, $\sigma_L^2 = 0$ and Equation 9.22 for the order-up-to level S reduces to

$$S = D\left(L+T\right) + k\sigma_D\sqrt{\left(L+T\right)} \tag{9.23}$$

The review period T, given by Equation 9.21, is determined from the EOQ, given in Section 9.1. For a given EOQ denoted by Q, the optimal time between successive replenishments is Q/D. This provides the estimate for T. In practice, the review period, T, may be rounded to a whole number of days or weeks, or set at some other convenient interval of time.

The order-up-to level S, given by Equation 9.22, is the inventory needed to ensure a given service level (i.e., a given probability that demand is met). The first term, $D(L+T)$, is the inventory needed to meet demand on average. The second term, $k\sqrt{\left(L+T\right)\sigma_D^2 + D^2\sigma_L^2}$, is the additional inventory (i.e., safety stock) needed to avoid stocking out due to the random variability in the demand and the lead time.

Orders for replenishment in the (T, S) policy are placed every T time units, as shown in Figure 9.8. After an order is placed, it takes l time units for the replenishment to arrive, where l is a random variable (the lead time). Thus, the time from when an order for a replenishment is placed until the subsequent replenishment arrives (i.e., the time from ordering replenishment, i, to the arrival of replenishment $i+1$) is $l+T$. To avoid a shortage, therefore, the inventory in the (T, S) policy must be sufficient to meet the demand during the lead time plus the review period (rather than just the lead time, as in the (s, Q) policy). The demand during the lead time plus the review period has a mean $D(L+T)$ and a standard deviation $k\sqrt{\left(L+T\right)\sigma_D^2 + D^2\sigma_L^2}$ (Tijms and Groenevelt, 1984, p. 180).

The service level factor, k, in Equation 9.22 is a dimensionless constant that represents the number of standard deviations beyond the mean needed to ensure a given service level (see (s, Q) policy above).

9.8 (*T*, *s*, *S*) Policy: Estimates of Review Period (*T*), Reorder Point (*s*), and Order-Up-To Level (*S*)

Replenishment policy: Inventory position (items on hand plus items on order) is reviewed at regular instants, spaced at time intervals of length *T*. At each review, if the inventory position is at the reorder point, *s*, or below, an order is placed for a sufficient quantity to raise the inventory position to the order-up-to level *S*; if the inventory position is above the reorder point, *s*, no order is placed. Figure 9.9 illustrates the (*T*, *s*, *S*) policy. This policy is also known as a *periodic review* (*s*, *S*) *policy*.

Assume:

- Demand for items is a random variable with a fixed mean and variance.
- Demands in separate increments of time are independent.
- Lead time (i.e., time from when an order for replenishment is placed until the replenishment arrives) is a random variable with a fixed mean and variance.
- Lead times are independent.
- Review period (i.e., time interval between reviews) is a constant.

Let

T = review period (units of time)
s = reorder point (number of items)
S = order-up-to level (number of items)
D = average demand (number of items per unit time)
σ_D^2 = variance of demand (items^2 per unit time)
L = average lead time (units of time)
σ_L^2 = variance of lead time (units of time^2)
k = service level factor
A = ordering cost ($ per order)
H = holding cost of an item ($ per item per unit time)

The ordering cost, A, in this policy includes the cost, if any, of reviewing of the inventory position in each review period. The demand variance, σ_D^2, is defined for a demand in one time unit. Since the demands in each time unit are assumed to be independent, the variance of demand in a fixed time of t units is $\sigma_D^2 t$.

Joint optimization of the three parameters (*T*, *s*, and *S*) in this policy leads to complicated mathematics (Lewis, 1970; Silver, Pyke, and Peterson, 1998). Simple heuristic approximations are presented here instead.

The review period T, reorder point s, and order-up-to level S in the (T, s, S) policy are given approximately by

$$T = \sqrt{\frac{2A}{DH}} \tag{9.24}$$

$$s = D(L+T) + k\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2} \tag{9.25}$$

$$S = s + Q \tag{9.26}$$

where

$$Q = \sqrt{\frac{2AD}{H}} \tag{9.27}$$

Porteus (1985, p. 138); Tijms and Groenevelt (1984, pp. 180, 183).

In the special case of fixed lead times, $\sigma_L^2 = 0$ and Equation 9.25 for the reorder point s reduces to

$$s = D(L+T) + k\sigma_D\sqrt{(L+T)} \tag{9.28}$$

The review period T, given by Equation 9.24, is the same as for the (T, S) policy (i.e., it is obtained from $T=Q/D$). In practice, T may be rounded to a whole number of days or weeks, or set at some other convenient interval of time. The quantity Q, given by Equation 9.27, is the EOQ, given in Section 9.1.

The reorder point s, given by Equation 9.25, is the inventory level needed to cover the demand during the lead time plus the review period. This expression for s is based on the result that demand during the lead time plus the review period has a mean $D(L+T)$ and a standard deviation $\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2}$.

The order-up-to level S, given by Equation 9.26, is the reorder point plus the economic order quantity (EOQ), as in the (s, S) policy for a continuous review system.

The service level factor, k, in Equation 9.25 is a dimensionless constant that represents the number of standard deviations of lead time demand beyond the mean needed to achieve a given service level (see (s, Q) policy earlier).

9.9 Summary of Results for Inventory Policies

(Details given in preceding sections)

(*s*, *Q*) Policy:

$$s = DL + k\sqrt{L\sigma_D^2 + D^2\sigma_L^2} \tag{9.29}$$

$$Q = \sqrt{\frac{2AD}{H}} \tag{9.30}$$

(*s*, *S*) Policy:

$$s = DL + k\sqrt{L\sigma_D^2 + D^2\sigma_L^2} \tag{9.31}$$

$$S = s + Q \tag{9.32}$$

$$Q = \sqrt{\frac{2AD}{H}} \tag{9.33}$$

(*T*, *S*) Policy:

$$T = \sqrt{\frac{2A}{DH}} \tag{9.34}$$

$$S = D(L+T) + k\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2} \tag{9.35}$$

(*T*, *s*, *S*) Policy:

$$T = \sqrt{\frac{2A}{DH}} \tag{9.36}$$

$$s = D(L+T) + k\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2} \tag{9.37}$$

$$S = s + Q \tag{9.38}$$

$$Q = \sqrt{\frac{2AD}{H}} \tag{9.39}$$

9.10 Inventory in a Production/Distribution System

The components of inventory in a production/distribution system for a single link between one origin and one destination are illustrated here.

Assume:

- Demand for items at the destination is continuous and at a constant rate.
- The origin has a production cycle and makes production runs for the destination at regular intervals.
- During a production run, the production of items at the origin for the destination is continuous and at a constant rate.
- The origin ships the items directly to the destination at regular intervals.
- The production schedule and shipment schedule are independent.
- Transit time (i.e., time for a shipment to travel from the origin to the destination) is a constant.

Let

P = production rate at origin (number of items per unit time)
D = demand at destination (number of items per unit time)
Q = production lot size (number of items)
V = shipment size (number of items)
T = time interval between shipments (units of time)
U = transit time (units of time)
I_1 = inventory at origin due to production cycle schedule (number of items)
I_2 = inventory at origin due to shipment cycle schedule (number of items)
I_3 = in-transit inventory (number of items)
I_4 = inventory at destination due to shipment cycle schedule (number of items)

Let

$$\overline{I}_1, \overline{I}_2, \overline{I}_3, \text{ and } \overline{I}_4$$

denote the averages of I_1, I_2, I_3, and I_4 over time, respectively.

Figure 9.10 shows the cumulative production, shipments, and demand over time for the single link between one origin and one destination.

The cumulative production curve in Figure 9.10 represents production cycling at the origin. During a production run for the destination, the production rate is P (where $P > D$ to ensure the demand is met). During the remainder of the production cycle, the production rate is zero. (The origin may produce items for other destinations during this time.)

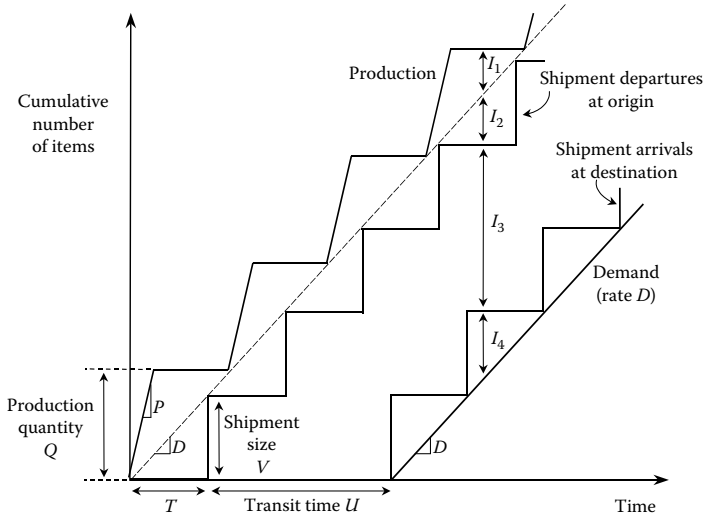


FIGURE 9.10
Cumulative production, shipments, and demand over time.

The cumulative shipment departure curve in Figure 9.10 represents shipments from the origin. Each step in the curve represents a shipment of size V. The cumulative shipment arrival curve represents these shipments when they arrive at the destination, U time units later. The cumulative demand curve represents the demand at the destination, with rate D items per unit time.

The average slope of each cumulative curve in Figure 9.10 must be D to match the demand. The shipment size and the time between shipments are related by

$$V = DT \tag{9.40}$$

The quantities I_1, I_2, I_3, and I_4 are the inventories at each stage in the production/distribution system. At any point in time, the vertical distances between the cumulative curves in Figure 9.10 represent these inventories.

The average inventories $\overline{I}_1$, $\overline{I}_2$, $\overline{I}_3$, and $\overline{I}_4$ are given by

$$\overline{I}_1 = \frac{Q}{2}\left(1 - \frac{D}{P}\right) \tag{9.41}$$

$$\overline{I}_2 = \frac{V}{2} \tag{9.42}$$

$$\bar{I}_3 = DU \tag{9.43}$$

$$\bar{I}_4 = \frac{V}{2} \tag{9.44}$$

Blumenfeld, Burns, Diltz, and Daganzo (1985, pp. 364, 370); Hall (1996, pp. 391–392).

Equation 9.41 for the average production cycle inventory $\bar{I}_1$ is the same as the expression for the average inventory in the EPQ model given in Section 9.2. The total number of items in the inventory at the origin is the sum of two separate inventories: production cycle inventory, I_1, and shipment cycle inventory, I_2, as shown in Figure 9.10.

Equations 9.42 through 9.44 for $\bar{I}_2$, $\bar{I}_3$, and $\bar{I}_4$ give the average components of the inventory associated with shipping. This inventory is, on average, made up of half a shipment at the origin, half a shipment at the destination, and DU items in transit. The total inventory associated with shipping is $V+DU=D(T+U)$ items.

9.11 Note on Cumulative Plots

The cumulative plots of production, shipments, and demand over time shown in Figure 9.10 provide a useful visual tool for representing the stages of a production/distribution system. A major benefit of cumulative plots is that they allow inventories at the various stages of the system to be conveniently displayed on one chart (Daganzo, 1991). For any point in time, the vertical distances between the cumulative curves represent the numbers of items in inventory at each stage, as indicated in Figure 9.10. If items pass through the system in a FIFO (first in, first out) sequence, the horizontal distances between the cumulative curves represent the times spent in inventory by an item at each stage.

Figures 9.1 and 9.4 show cumulative plots of orders and demand over time for the EOQ and EPQ models. For inventory systems in general, cumulative plots can be used to plan schedules for orders, analyze holding costs, and identify conditions for shortages (Brown, 1977, pp. 271–286; Daganzo, 1991, pp. 46–51; Love, 1979, pp. 42–47).

Cumulative plots also have applications in areas closely related to inventory control, such as in queueing theory (Newell, 1982, pp. 3–7; Medhi, 1991, pp. 62–63) and transportation and traffic flow analysis (Daganzo, 1997, pp. 25–29; Newell, 1993, pp. 282–283). Cumulative plots have long been used in hydraulic engineering for determining reservoir capacity (Linsley and Franzini, 1955, pp. 138–139).

10

Distance Formulas for Logistics Analysis

10.1 Distance Norms

A distance norm is a metric used to determine how close two points are to each other. The coordinates that define the locations of points in space can be represented by vectors, and a distance norm is expressed as a function of the coordinates in the vector space.

The location of a point in N-dimensional space with coordinates v_1, v_2, …, v_N is represented by the vector **v**, where

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}$$

The norm for vector **v** is denoted by $\|v\|$. A set of norms known as L_p norms provide various distance metrics in the vector space. For a given p, the L_p norm for vector **v** is given by

$$\|v\|_p = \left(\sum_{i=1}^{N} |v_i|^p \right)^{\frac{1}{p}} = \left(|v_1|^p + |v_2|^p + \cdots + |v_N|^p \right)^{\frac{1}{p}} \tag{10.1}$$

where $|v_i|$ denotes the absolute value of v_i (i=1, 2, …, N).

In a plane (two-dimensional space), the location of a point P with coordinates (x, y) is represented by the two-dimensional vector **v**, where

$$\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$$

In this case of two-dimensional space, $N=2$ and Equation 10.1 becomes

$$\|v\|_p = \left(|x|^p + |y|^p\right)^{\frac{1}{p}} \tag{10.2}$$

Let P_1 and P_2 be two points in a plane with coordinates (x_1, y_1) and (x_2, y_2), respectively. The L_p norm for a general distance D between points P_1 and P_2 is given by

$$D = \left(|x_1 - x_2|^p + |y_1 - y_2|^p\right)^{\frac{1}{p}} \tag{10.3}$$

Common values for p are 1, 2, and ∞. With these values for p in Equation 10.3, the norms for distance D are the L_1, L_2, and L_∞ norms, respectively, which are defined as follows:

$$L_1 \text{ norm: } D = |x_1 - x_2| + |y_1 - y_2| \tag{10.4}$$

$$L_2 \text{ norm: } D = \sqrt{|x_1 - x_2|^2 + |y_1 - y_2|^2} \tag{10.5}$$

$$\begin{aligned} L_\infty \text{ norm: } D &= \lim_{p \to \infty} \left(|x_1 - x_2|^p + |y_1 - y_2|^p\right)^{\frac{1}{p}} \\ &= \max\left(|x_1 - x_2|, |y_1 - y_2|\right) \end{aligned} \tag{10.6}$$

Fröberg (1965, pp. 62–63); Kumaresan (2005, pp. 3, 5, 8); Smith (2008); Van der Heijden, Duin, de Ridder, and Tax (2004, p. 356).

Figures 10.1 through 10.3 illustrate the L_1, L_2, and L_∞ norms, respectively. The rectilinear distance (L_1 norm), shown in Figure 10.1, is the rectangular

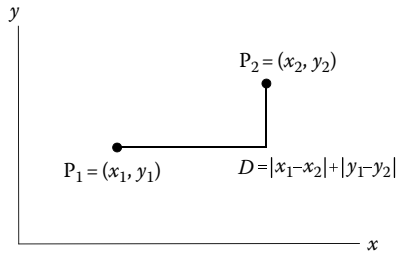


FIGURE 10.1
L_1 norm: rectangular grid (rectilinear) distance.

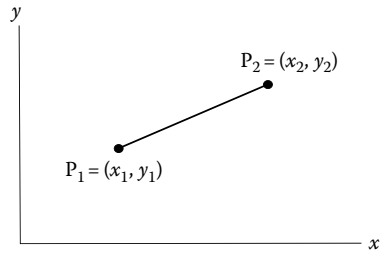


FIGURE 10.2
L_2 norm: Euclidean (straight Line) distance.

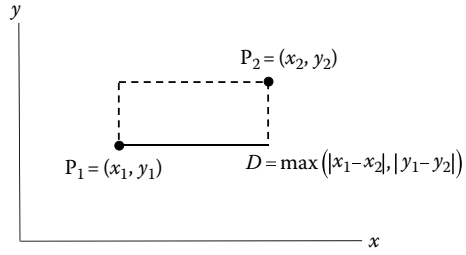


FIGURE 10.3
L_∞ norm: Chebyshev distance.

grid distance, also known as city block, Manhattan, or taxicab distance. The Euclidean distance (L_2 norm), shown in Figure 10.2, is the direct or straight line distance. The Chebyshev distance (L_∞ norm), shown in Figure 10.3, is a measure of the maximum separation in one direction.

The above distance norms are metrics for distance between two points on a plane surface, with the shortest distance given by the L_2 norm (Euclidean or straight line distance). Different metrics are needed for distances over a sphere. The shortest distance between two points on the surface of a sphere is given by the great circle distance (or orthodromic distance). A great circle is any circle around the surface of a sphere that divides the sphere into two equal halves. Great circles are the largest circles that can be drawn on a sphere. The shortest path between two points on a sphere is an arc of a great circle, known as a geodesic. Since the earth is approximately spherical, a great circle distance is the metric for determining shortest distances between points on the earth's surface (Jennings, 1994, p. 47; Simmons, 1945, p. 299).

The following sections present formulas based on Euclidean distance (Sections 10.2 and 10.3), rectilinear distance (Section 10.4), and great circle distance (Section 10.5).

10.2 "Traveling Salesman Problem" Tour Distance: Shortest Path through a Set of Points in a Region

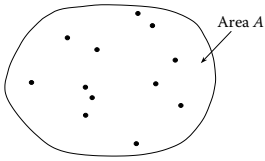


FIGURE 10.4
Points randomly distributed within a region of area A.

The average tour distance d on the shortest closed path connecting n points randomly distributed within a region of area A (see Figures 10.4 and 10.5) is given approximately by

$$d \cong K\sqrt{n\,A} \tag{10.7}$$

where K is a constant. Based on simulation experiments, the value for the constant K is generally taken as $K = 0.75$.

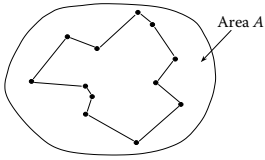


FIGURE 10.5
Shortest closed path through points in Figure 10.4.

Beardwood, Halton, and Hammersley (1959, p. 303); Eilon, Watson-Gandy, and Christofides (1971, p. 169); Larson and Odoni (1981, p. 408); Stein (1978, p. 90).

This formula is useful in logistics for developing delivery strategies, planning transportation service requirements, and evaluating "traveling salesman problem" algorithms (Burns, Hall, Blumenfeld, and Daganzo, 1985, pp. 474–477; Daganzo, 1991, pp. 122–125; Larson and Odoni, 1981, pp. 408–411; Stein, 1978, pp. 89–93). The approximation holds well for regions of various shapes (Christofides and Eilon, 1969, p. 439; Eilon, Watson-Gandy, and Christofides, 1971, pp. 170, 174).

The formula is derived for large n, but also provides an approximation when n is small (Daganzo, 1984, p. 135). In the extreme case of two random points ($n = 2$), the tour distance d is the distance from one point to the other and back (i.e., twice the distance between the two points). If the region is a circle of radius R, then $A = \pi R^2$, and d from Equation 10.7 is given by $d \cong K\sqrt{2\pi}\,R = 1.88R$. This estimate for d is close to $1.81R$, the theoretical result for twice the average distance between two random points in a circle (see Equation 10.10).

10.3 Distribution of Distance between Two Random Points in a Circle

Let

x = distance between two random points in a circle of radius R ($0 \le x \le 2R$), as illustrated in Figure 10.6

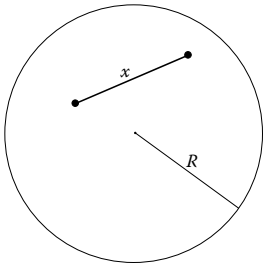


FIGURE 10.6
Distance x between two random points in a circle of radius R.

$f(x)$ = probability density function of x

$F(x) = \int_0^x f(u)\,du$ = cumulative distribution function of x

$E(x) = \int_0^{2R} x f(x)\,dx$ = average distance

$$\sigma = \sqrt{\int_0^{2R} x^2 f(x)\,dx - \{E(x)\}^2}$$

= standard deviation of x

The probability density function $f(x)$ is given by

$$f(x) = \frac{2x}{\pi R^2}\left\{2\cos^{-1}\left(\frac{x}{2R}\right) - \frac{x}{R}\sqrt{1-\frac{x^2}{4R^2}}\right\} \quad (0 \le x \le 2R) \tag{10.8}$$

Garwood (1947, p. 9); Garwood and Tanner (1958, p. 293); Fairthorne (1965, p. 396); Kendall and Moran (1963, p. 42); Vaughan (1987, p. 222); Grimmett and Stirzaker (2001, p. 136). The probability density function $f(x)$ is plotted in Figure 10.7.

The cumulative distribution function $F(x)$ is given by

$$F(x) = 1 + \frac{2}{\pi}\left(\frac{x^2}{R^2} - 1\right)\cos^{-1}\left(\frac{x}{2R}\right) - \frac{x}{\pi R}\left(1 + \frac{x^2}{2R^2}\right)\sqrt{1-\frac{x^2}{4R^2}} \quad (0 \le x \le 2R) \tag{10.9}$$

Borel (1925, Chapter 4, p. 78); Garwood (1947, p. 8).

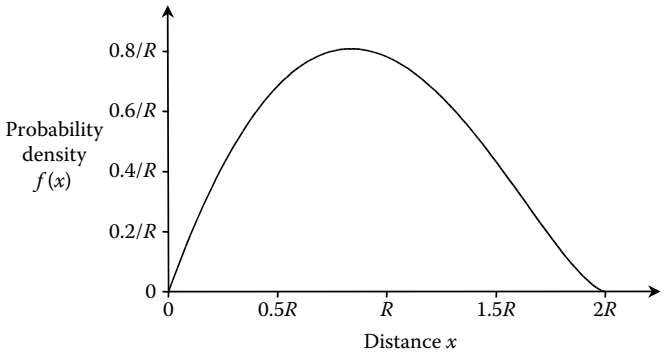


FIGURE 10.7
Probability density function $f(x)$ for distance x between two random points in a circle of radius R.

The cumulative distribution function $F(x)$ gives the probability that the distance between the two points is less than x. This function is plotted in Figure 10.8.

The average distance $E(x)$ is given by

$$E(x) = \frac{128R}{45\pi} \quad (\cong 0.9054R) \tag{10.10}$$

Apsimon (1958, p. 52); Eilon, Watson-Gandy, and Christofides (1971, p. 154); Garwood and Tanner (1958, pp. 292–293); Fairthorne (1965, p. 396); Smeed (1971, p. 15); Vaughan (1987, pp. 222, 242).

The standard deviation σ of distance x is given by

$$\sigma = \sqrt{R^2 - \left(\frac{128R}{45\pi}\right)^2} \quad (\cong 0.4245R) \tag{10.11}$$

Garwood and Tanner (1958, p. 293).

Note: The above results on average distances for a circle provide useful approximations for regions of general shape. Spatial analyses indicate that average distances within a region of a given area do not depend strongly on the shape of the region (Eilon, Watson-Gandy, and Christofides, 1971, p. 174; Larson and Odoni, 1981, pp. 135–136; Smeed, 1967, p. 23).

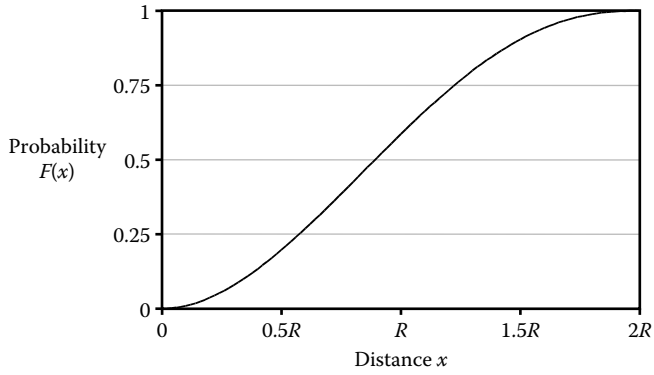


FIGURE 10.8
Cumulative distribution function $F(x)$ for distance x between two random points in a circle of radius R.

10.4 Average Rectangular Grid Distance between Two Random Points in a Circle

The average distance d_{grid} on a rectangular grid (i.e., average rectilinear distance) between two random points in a circle of radius R (see Figure 10.9) is given by

$$d_{grid} = \frac{4}{\pi} d_{direct} \qquad (10.12)$$

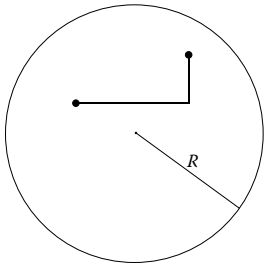


FIGURE 10.9
Rectangular grid distance between two random points in a circle of radius R.

where d_{direct} is the average direct (Euclidean) distance between the two points. From the result for d_{direct} in Equation 10.10, d_{grid} is given by

$$d_{grid} = \frac{512R}{45\pi^2} \quad (\cong 1.1528R) \qquad (10.13)$$

Eilon, Watson-Gandy, and Christofides (1971, pp. 162–163); Fairthorne (1965, p. 403); Vaughan (1987, p. 235).

10.5 Great Circle Distance

Let P_1 and P_2 be two points on the earth's surface, with positions (α_1, β_1) and (α_2, β_2) defined by

α_1 = latitude of P_1
β_1 = longitude of P_1
α_2 = latitude of P_2
β_2 = longitude of P_2

Assuming the earth is a sphere, the great circle distance D between points P_1 and P_2 is given by

$$D = R\theta \qquad (10.14)$$

where

$$R = \text{mean radius of the earth } (\cong 3960 \text{ miles or } 6370 \text{ km}) \qquad (10.15)$$

and

$$\theta = \cos^{-1}\left\{\sin\alpha_1 \sin\alpha_2 + \cos\alpha_1 \cos\alpha_2 \cos\left(\beta_1 - \beta_2\right)\right\} \tag{10.16}$$

with angle θ expressed in radians ($1° = \pi/180$ rad).

Jennings (1994, pp. 54, 64); Melzak (1983, pp. 132, 140–141).

Note: East and west longitudes can be distinguished by taking longitudes east of the prime meridian as positive, and longitudes west of the prime meridian as negative (or vice versa). Likewise, north and south latitudes can be distinguished by taking latitudes north of the equator as positive, and latitudes south of the equator as negative.

Equation 10.16 is based on the *law of cosines for sides* for a spherical triangle. Under the assumption that the earth is a sphere, the three points P_1, P_2, and the north pole form a spherical triangle with arc angles (in degrees) given by $90 - \alpha_1$, $90 - \alpha_2$, and θ, and with a vertex angle at the north pole given by $\beta_1 - \beta_2$, as shown in Figure 10.10. The law of cosines for sides for this spherical triangle gives

$$\cos\theta = \cos\left(90 - \alpha_1\right)\cos\left(90 - \alpha_2\right) + \sin\left(90 - \alpha_1\right)\sin\left(90 - \alpha_2\right)\cos\left(\beta_1 - \beta_2\right) \tag{10.17}$$

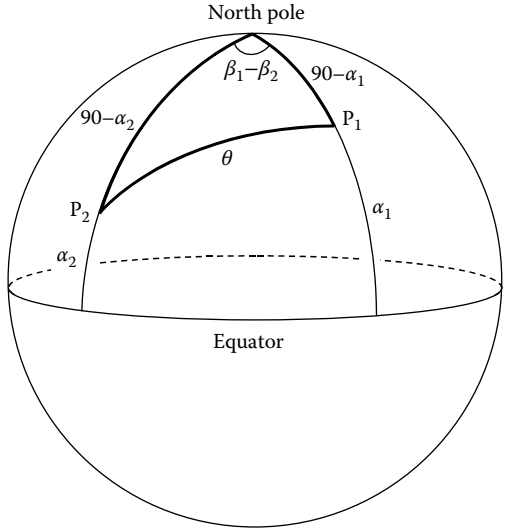


FIGURE 10.10
Spherical triangle formed by points P_1, P_2, and the north pole (with arc angles $90 - \alpha_1$, $90 - \alpha_2$, and θ, and vertex angle $\beta_1 - \beta_2$ at the north pole).

Ashton and Marsh (1902, p. 129); Chauvenet (1875, p. 179); Jennings (1994, pp. 49, 54); Kells, Kern, and Bland (1942, p. 70); Melzak (1983, p. 132); Thurston (1997, p. 75).

Since $\cos(90-\alpha_1)=\sin\alpha_1$, $\sin(90-\alpha_1)=\cos\alpha_1$, etc., this law (Equation 10.17) yields the result given by Equation 10.16.

The formula for the angle θ given by Equation 10.16 can also be written as

$$\theta = 2\sin^{-1}\left\{\sqrt{\sin^2\left(\frac{\alpha_1-\alpha_2}{2}\right)+\cos\alpha_1\cos\alpha_2\sin^2\left(\frac{\beta_1-\beta_2}{2}\right)}\right\} \tag{10.18}$$

Simmons (1945, pp. 292, 294, 300); Sinnott (1984, p. 159). This version of the formula for θ is mathematically the same as Equation 10.16, but is in a more suitable form for computations when the points P_1 and P_2 are close together. It is based on the *haversine formula*:

$$\text{hav}(\theta)=\text{hav}(\alpha_1-\alpha_2)+\cos(\alpha_1)\cos(\alpha_2)\text{hav}(\beta_1-\beta_2) \tag{10.19}$$

where, for any angle a, hav(a) is the haversine of a and is given by

$$\text{hav}(a)=\frac{1}{2}\left[1-\cos(a)\right]=\sin^2\left(\frac{1}{2}a\right) \tag{10.20}$$

Ayres (1954, p. 181); Kells, Kern, and Bland (1942, p. 97); Love and Morris (1972, p. 62); Simmons (1945, p. 294).

11

Traffic Flow and Delay

11.1 Traffic Flow Parameters

Basic parameters for a traffic stream of vehicles traveling along a roadway are

q: traffic flow (vehicles per unit time)
k: concentration (vehicles per unit distance of roadway)
v: speed (distance per unit time)

Traffic flow, q, is the number of vehicles passing a point per unit time, typically measured in vehicles per hour. Concentration (or density), k, is the number of vehicles per unit length of roadway, typically measured in vehicles per kilometer or vehicles per mile. Speed, v, denotes the instantaneous speed of a given vehicle (also known as spot speed), typically measured in kilometers per hour or miles per hour.

If the vehicles in a traffic stream are all traveling at the same speed v, where v is a constant, then the three parameters above are related by

$$q = kv \tag{11.1}$$

Wardrop (1952, p. 328); Haight (1963, p. 70).

11.2 Traffic Speeds

Consider a traffic stream composed of C independent subsidiary streams, where vehicles in the stream i travel at a constant speed v_i ($i = 1, 2, \ldots, C$).

Let

q_i be the traffic flow in stream i (vehicles per unit time)
k_i be the concentration of traffic in stream i (vehicles per unit distance)
Q be the total traffic flow in the C streams combined (vehicles per unit time)
K be the total concentration of traffic in the C streams combined (vehicles per unit distance)

Then the total flow Q and the total concentration K are given by

$$Q = \sum_{i=1}^{C} q_i \tag{11.2}$$

and

$$K = \sum_{i=1}^{C} k_i \tag{11.3}$$

respectively. Let

$$f_i^t = \frac{q_i}{Q} \quad (i = 1, 2, ..., C)$$

$$f_i^s = \frac{k_i}{Q} \quad (i = 1, 2, ..., C)$$

Then the quantities $f_1^t, f_2^t, \ldots f_C^t$ are the relative frequencies in *time* of vehicles with speeds $v_1, v_2, \ldots v_C$. Similarly, the quantities $f_1^s, f_2^s, \ldots f_C^s$ are the relative frequencies in *space* of vehicles with speeds $v_1, v_2, \ldots v_C$.

These two sets of relative frequencies define distributions of speeds of the vehicles in the traffic stream:

$$\{f_1^t, f_2^t, \ldots f_C^t\} = \text{time-distribution of speeds}$$

$$\{f_1^s, f_2^s, \ldots f_C^s\} = \text{space-distribution of speeds}$$

with $\sum_{i=1}^{C} f_i^t = 1$ and $\sum_{i=1}^{C} f_i^s = 1$. Corresponding to these distributions, there are two different measures for average speed: *time-mean speed* and *space-mean speed*. Let

$\bar{v}_t$ be the time-mean speed

$\bar{v}_s$ be the space-mean speed

Then the time-mean speed, $\bar{v}_t$, and the space-mean speed, $\bar{v}_s$, are given by

$$\overline{v}_t = \sum_{i=1}^{C} f_i^s v_i = \frac{1}{Q} \sum_{i=1}^{C} q_i v_i \tag{11.4}$$

and

$$\overline{v}_s = \sum_{i=1}^{C} f_i^s v_i = \frac{1}{K} \sum_{i=1}^{C} k_i v_i \tag{11.5}$$

respectively.

From Equation 11.1, the concentration, k_i, of traffic stream i is

$$k_i = \frac{q_i}{v_i} \quad (i = 1, 2, ..., C) \tag{11.6}$$

and Equation 11.5 for the space-mean speed therefore becomes

$$\overline{v}_s = \frac{1}{K} \sum_{i=1}^{C} q_i = \frac{1}{K} Q \tag{11.7}$$

Hence, total flow, Q, total concentration, K, and space-mean speed, $\bar{v}_s$, are related by

$$Q = K \overline{v}_s \tag{11.8}$$

Wardrop (1952, p. 330); Hall (1992, pp. 2-9–2-10).

Equation 11.8 is known as the fundamental relationship in traffic. It extends Equation 11.1 to the case of traffic with vehicles traveling at different speeds.

Note that $\bar{v}_t \geq \bar{v}_s$ in all cases, with $\bar{v}_t > \bar{v}_s$ when there are any differences in speed among the vehicles. The two mean speeds are related by

$$\overline{v}_t = \overline{v}_s + \frac{\sigma_s^2}{\overline{v}_s} \tag{11.9}$$

where σ_s^2 is the variance of the space-distribution of speeds, given by

$$\sigma_s^2 = \sum_{i=1}^{C} f_i^s (v_i - \overline{v}_s)^2 = \frac{1}{K} \sum_{i=1}^{C} k_i (v_i - \overline{v}_s)^2 \tag{11.10}$$

If c_s denotes the coefficient of variation of the space-distribution of speed, i.e.,

$$c_s = \frac{\sigma_s}{\overline{v}_s} \tag{11.11}$$

then Equation 11.9 can be expressed as

$$\bar{v}_t = \bar{v}_s \left(1 + c_s^2\right) \tag{11.12}$$

Wardrop (1952, pp. 330–331, 356); Hall (1992, pp. 2-6–2-7).

Calculations of time-mean and space-mean speeds in practice depend on how vehicle speeds are measured. If the speeds of N vehicles are observed as the vehicles pass a fixed point in the roadway over a period of time, where the observed speed of vehicle j is v_j (j=1, 2, …, N), then the average of the observed speeds is the time-mean speed $\bar{v}_t$, i.e.,

$$\bar{v}_t = \frac{1}{N}\sum_{j=1}^{N} v_j \tag{11.13}$$

Assuming these speeds are constant as the vehicles travel along a stretch of roadway of length, L, the travel time t_j of vehicle j is $t_j = L/v_j$ (j=1, 2, …, N). The average travel time $\bar{t}$ of the N vehicles is then given by

$$\bar{t} = \frac{1}{N}\sum_{j=1}^{N} t_j = \frac{1}{N}\sum_{j=1}^{N} \frac{L}{v_j} \tag{11.14}$$

The space-mean speed, $\bar{v}_s$, of these vehicles is the distance traveled divided by the average travel time, i.e.,

$$\bar{v}_s = \frac{L}{\bar{t}} \tag{11.15}$$

From Equation 11.14, the space-mean speed, $\bar{v}_s$, becomes

$$\bar{v}_s = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{v_i}} \tag{11.16}$$

Thus, the space-mean speed is the harmonic mean of the observed speeds, v_j, of vehicles passing a point in the roadway over a period of time, while the time-mean speed is the arithmetic mean.

If, instead, the speeds v_j (j=1, 2, …, N) are the speeds of N vehicles observed on the roadway in a one-time snapshot (i.e., the speeds in space

at a given instant of time), then it is the space-mean speed that is the arithmetic mean, i.e.,

$$\bar{v}_s = \frac{1}{N}\sum_{j=1}^{N} v_j \tag{11.17}$$

and the variance, σ_s^2, in this case is given by

$$\sigma_s^2 = \frac{1}{N}\sum_{j=1}^{N}(v_j - \bar{v}_s)^2 = \frac{1}{N}\left(\sum_{j=1}^{N} v_j^2\right) - \bar{v}_s^2 \tag{11.18}$$

so that the time-mean speed, $\bar{v}_t$, from Equation 11.9 is given in this case by

$$\bar{v}_t = \frac{\frac{1}{N}\sum_{j=1}^{N} v_j^2}{\bar{v}_s} = \frac{\sum_{j=1}^{N} v_j^2}{\sum_{j=1}^{N} v_j} \tag{11.19}$$

Wardrop (1952, p. 330).

11.3 Delay to Vehicle Merging with Traffic Stream

Consider a vehicle on a minor road waiting to merge with the traffic stream on a main road. The delay to the merging vehicle is the time it must wait until there is a large enough time gap to accept in the main road traffic stream. The analytical results given below are from a basic model of merging under the following assumptions:

1. Traffic flows along the main road according to a Poisson process with rate Q vehicles per unit time (i.e., time gaps between successive vehicles are exponentially distributed, with mean $1/Q$).
2. The main road traffic has priority.
3. Lengths of vehicles can be neglected.
4. Driver of merging vehicle accepts a time gap if it is greater than (or equal to) a fixed time gap, called the driver's *critical gap*. Driver rejects gaps that are less than the critical gap.

Let

Q be the traffic flow on the main road (vehicles per unit time)
T_c be the critical gap of the driver of the merging vehicle (units of time)
D be the average delay to the merging vehicle (units of time)
σ_D^2 be the variance of delay to the merging vehicle (units of time^2)

Given assumption 1, the time gaps, t, between successive vehicles on the main road have an exponential distribution with probability density function, $f(t)$, given by

$$f(t) = Qe^{-Qt} \quad (0 \le t < \infty)$$

By the lack of memory property of the exponential distribution (Equation 4.9), the first gap the merging vehicle encounters has the same probability density function. The probability of no delay is

$$\Pr\{\text{no delay}\} = e^{-QT_c} \tag{11.20}$$

For this model, the average delay, D, to a merging vehicle is

$$D = \frac{1}{Q}\left(e^{QT_c} - QT_c - 1\right) \tag{11.21}$$

and the variance of delay, σ_D^2, is

$$\sigma_D^2 = \frac{1}{Q^2}\left(e^{2QT_c} - 2QT_c e^{QT_c} - 1\right) \tag{11.22}$$

Herman and Weiss (1961, p. 837); Weiss and Maradudin (1962, pp. 80–81, 92); Haight (1963, pp. 140–141); Blumenfeld and Weiss (1970a, p. 126); McNeil and Weiss (1974, p. 119).

11.4 Critical Flow on Minor Road

For the merging delay model given above, the critical flow on the minor road is the capacity at the merge point (i.e., maximum flow of traffic on the minor road to ensure a finite queue). Given a queue of vehicles on the minor road waiting to merge into the main toad traffic, the critical flow depends on the time for a vehicle to move up from second in the queue to the head of the queue (move-up time).

Let

Q be the traffic flow on the main road (vehicles per unit time)
T_c be the critical gap of each driver of a merging vehicle
T_m be the move-up time (time for a vehicle to move up from second to the head of the queue)
q_c be the critical flow of traffic on the minor road (vehicles per unit time)

For the model and assumptions as in Section 11.3, the critical flow is given q_c by

$$q_c = \frac{Qe^{-QT_c}}{1-e^{-QT_m}} \tag{11.23}$$

Evans, Herman, and Weiss (1964, p. 850); Ashworth (1969, p. 273); Blumenfeld and Weiss (1970a, p. 136).

11.5 Delay to Traffic Queue on Minor Road Waiting to Merge

For the merging delay model given earlier, the delay to a queue of vehicles on the minor road is the time from when a vehicle arrives to join the queue until it merges into the main road traffic. The model is based on the same assumptions as in Sections 11.3 and 11.4, and the following additional assumptions:

1. Vehicles arrive at the queue on the minor road according to a Poisson process with rate q vehicles per unit time (i.e., times between arrivals of vehicles are exponentially distributed, with mean $1/q$).
2. Traffic flow on the minor road is less than the critical flow (i.e., $q<q_c$).

Let

Q be the traffic flow on the main road (vehicles per unit time)
q be the traffic flow on the minor road (vehicles per unit time) (i.e., q=arrival rate to the queue on the minor road)
T_c be the critical gap of each driver of a merging vehicle
T_m be the move-up time (time for a vehicle to move up from second to the head of the queue)
D_q be the average delay to a vehicle in the queue on the minor road (units of time)

For this model, the average delay, D_q, to a vehicle in the queue on the minor road waiting to merge into the main road traffic is

$$D_q = \frac{Qe^{QT_m}\left(e^{QT_c} - QT_c - 1\right) + qe^{QT_c}\left(e^{QT_m} - QT_m - 1\right)}{Q\{Qe^{QT_m} - qe^{QT_c}(e^{QT_m} - 1)\}} \tag{11.24}$$

Tanner (1962, p. 160); Blumenfeld and Weiss (1970b, p. 142).

11.6 Delay to Vehicle at Traffic Signal

A traffic signal controls the flow of traffic on each approach to an intersection. It cycles through the red, green, and yellow phases for each approach. The time for one complete cycle is the *cycle time*. For a given approach to the signal, the maximum departure rate of vehicles during the green phase is called the *saturation flow*. The portion of the cycle time during which traffic travels through the signal at the saturation flow rate, as long as there are vehicles on the approach, is the *effective green time*. This time is effectively the time the signal is green, plus part of the subsequent yellow time, minus a short start-up time at the beginning of the green phase.

There are various types of traffic signals, such as fixed-time signals, vehicle-actuated signals, and signals with adaptive control. For a fixed-time signal, the cycle time is fixed, and the times of the phases on each approach are fixed. For a vehicle-actuated signal, the phase times vary according to traffic flows. Signals with adaptive control adjust phase times to optimize traffic flows over a network. There are many analyses on the different types of traffic signals (see, e.g., Allsop, 1972; Allsop, Ali, Crosta, and Heydecker, 2005; McNeil and Weiss, 1974; Newell, 1982, 1989; Webster and Cobbe, 1966).

The following formula gives a basic approximation of average delay for the simplest case of a fixed-time signal.

Let

c = cycle time

and for a given approach, let

g = effective green time

q = traffic flow (vehicles per unit time)

s = saturation flow (vehicles per unit time)

$\lambda = \frac{g}{c}$ = proportion of the cycle that is effectively green

$x = \frac{qc}{gs}$ = ratio of the average number of arrivals per cycle (qc) to a maximum number of departures per cycle (gs)

Then the average delay per vehicle, D, on the approach to the signal is given by Webster's approximate formula:

$$D = 0.9\left\{\frac{c(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)}\right\} \tag{11.25}$$

Webster (1958, pp. 4–5); Wardrop (1968, p. 532); Homburger (1982, p. 457).

The first term accounts for delay to uniform traffic (i.e., traffic arriving in a steady flow). The second term accounts for extra delay to randomness of arrivals, based on average queueing time in an $M/D/1$ queue (see Chapter 7). The factor 0.9 is a correction factor, based on simulations, to provide a close approximation for the overall average delay per vehicle.

12

Linear Programming Formulations

12.1 General Formulation

Let x_1, x_2, …, x_N, be N variables in a linear programming problem. The problem is to find the values of the variables x_1, x_2, …, x_N to maximize (or minimize) a given linear function of the variables, subject to a given set of constraints that are linear in the variables.

The general formulation for a linear programming problem is

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{12.1}$$

subject to the constraints

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &\le b_2 \\ &\vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &\le b_M \end{aligned}\right\} \tag{12.2}$$

and

$$x_1 \ge 0, x_2 \ge 0, \ldots, x_N \ge 0 \tag{12.3}$$

where a_{ij}, b_i, c_j ($i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, N$) are constants.

Chvátal (1983, p. 6); Gass (1964, p. 9); Hillier and Lieberman (1980, p. 22); Ignizio (1982, pp. 81–82); Munakata (1979, p. 251); Ozan (1986, pp. 5–6); Vanderbei (1997, p. 7); Wagner (1969, p. 81).

In matrix notation, Equations 12.1 through 12.3 for the general formulation are written as

$$\text{Maximize } Z = \mathbf{c}^{\mathrm{T}}\mathbf{x} \tag{12.4}$$

subject to the constraints

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{12.5}$$

and

$$\mathbf{x} \geq \mathbf{0} \tag{12.6}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \tag{12.7}$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12.8}$$

and where $\mathbf{c}^T$ denotes the transpose of the vector $\mathbf{c}$.

12.2 Terminology

The following terms are commonly used in linear programming:

- *Decision variables*: Variables $x_1, x_2, \ldots, x_N$ in Equation 12.1
- *Objective function*: Function Z given by Equation 12.1
- *Objective function coefficients*: Constants $c_1, c_2, \ldots, c_N$ in Equation 12.1
- *Constraint coefficients*: Constants a_{ij} in Equation 12.2
- *Nonnegativity constraints*: Constraints given by Equation 12.3
- *Feasible solution*: Set of values of $x_1, x_2, \ldots, x_N$ that satisfy all the constraints
- *Feasible region*: Collection of all feasible solutions
- *Optimal solution*: Feasible solution that gives an optimal value of the objective function (i.e., the maximum value of Z in Equation 12.1).

12.3 Example of a Feasible Region

The feasible region can be shown graphically in the case of two decision variables ($N=2$). Figure 12.1 illustrates the feasible region for the following example:

$$\text{Maximize } Z = 2x_1 + x_2 \tag{12.9}$$

subject to the constraints

$$\left.\begin{aligned} 3x_1 + 5x_2 &\le 20 \\ 4x_1 + x_2 &\le 15 \end{aligned}\right\} \tag{12.10}$$

and

$$x_1 \ge 0, x_2 \ge 0 \tag{12.11}$$

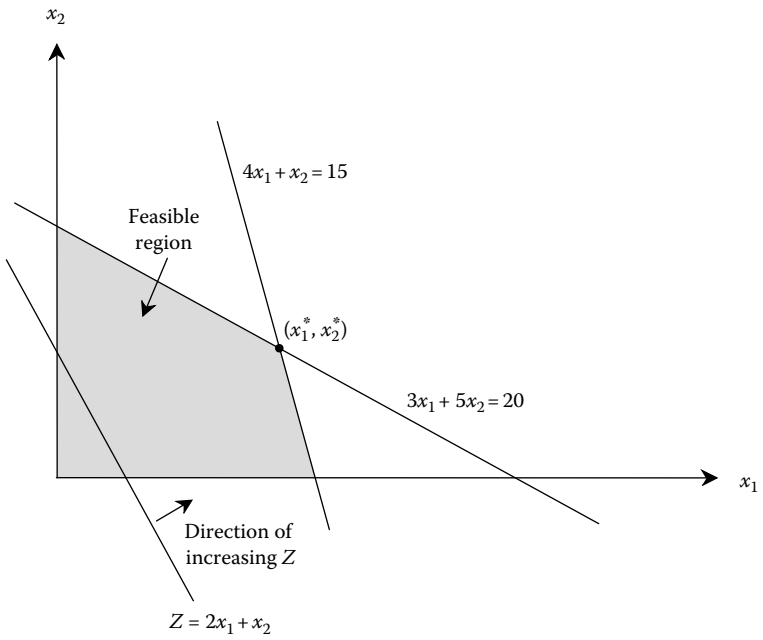


FIGURE 12.1
Feasible region for a linear programming example.

The optimal solution Z^* (i.e., maximum Z) occurs at the point (x_1^*, x_2^*) in Figure 12.1. For this example, the point $(x_1^*, x_2^*) = \left(\frac{55}{17}, \frac{35}{17}\right)$, giving $Z^* = \frac{145}{17}$.

12.4 Alternative Formulations

Linear programming problems in which

- The objective function is to be minimized (rather than maximized), or
- The constraints contain equalities (= rather than ≤), or
- The constraints contain inequalities with the sign reversed (≥ rather than ≤),

can be reformulated in terms of the general formulation given by Equations 12.1 through 12.3 in the following ways.

12.4.1 Minimization vs. Maximization

For cases where the objective function Z is to be minimized, the problem can be reformulated by expressing the minimization of Z as the maximization of $-Z$. Thus, the problem

$$\text{Minimize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N$$

is equivalent to

$$\text{Maximize } -Z = (-c_1)x_1 + (-c_2)x_2 + \cdots + (-c_N)x_N$$

Dantzig (1963, p. 61); Hillier and Lieberman (1980, p. 51); Vanderbei (1997, p. 53); Wagner (1969, p. 78).

12.4.2 Equality Constraints

For cases where some or all of the constraints contain equalities, the problem can be reformulated by expressing an equality as two inequalities with opposite signs. Thus, the constraint

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{iN}x_N = b_i$$

is equivalent to

$$\left.\begin{aligned} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{iN}x_N \le b_i \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{iN}x_N \ge b_i \end{aligned}\right\}$$

Dantzig (1963, p. 88); Daskin (1995, p. 22); Hillier and Lieberman (1980, p. 51); Vanderbei (1997, p. 7); Wagner (1969, p. 79).

12.4.3 Reversed Inequality Constraints

For cases where some or all of the constraints contain inequalities with the sign reversed (≥ rather than ≤), the ≥ signs can be converted to ≤ signs by multiplying both sides of the constraints by –1. Thus, the constraint

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{iN}x_N \ge b_i$$

is equivalent to

$$-a_{i1}x_1 - a_{i2}x_2 - \cdots - a_{iN}x_N \le -b_i$$

Hillier and Lieberman (1980, p. 51); Munakata (1979, p. 251); Wagner (1969, p. 78).

12.5 Diet Problem

The *diet problem* arises in optimizing the choice of foods for a healthy diet. The problem is to determine the mix of foods in a diet that minimizes the total cost per day, subject to constraints that ensure minimum daily nutritional requirements are met. The diet problem is an example of a general linear programming problem, in which the objective function is to be minimized and the constraints contain ≥ signs. Let

M=number of nutrients
N=number of types of food
a_{ij}=number of units of nutrient i in food j (i=1, 2, …, M; j=1, 2, …, N)
b_i=number of units of nutrient i required per day (i=1, 2, …, M)
c_j=cost per unit of food j (j=1, 2, …, N)
x_j=number of units of food j in the diet per day (j=1, 2, …, N)

The objective is to find the values of the N variables $x_1, x_2, \ldots, x_N$ to minimize the total cost per day, C.

The linear programming formulation for the diet problem is

$$\text{Minimize } C = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{12.12}$$

subject to the constraints

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &\geq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &\geq b_2 \\ &\vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &\geq b_M \end{aligned}\right\} \tag{12.13}$$

and

$$x_1 \geq 0, x_2 \geq 0, \ldots, x_N \geq 0 \tag{12.14}$$

where a_{ij}, b_i, c_j ($i=1, 2, \ldots, M$; $j=1, 2, \ldots, N$) are constants.

Gass (1964, pp. 9–10); Luenberger (1984, pp. 14–15); Spivey and Thrall (1970, pp. 39–40); Vanderbei (1997, pp. 78–79).

12.6 Duality

Each linear programming problem has a related linear programming problem called the *dual problem*. The original linear programming problem is called the *primal problem*. For the primal problem defined by Equations 12.1 through 12.3, the corresponding dual problem is to find the values of the M variables $y_1, y_2, \ldots, y_M$ to solve the following:

$$\text{Minimize } V = b_1y_1 + b_2y_2 + \cdots + b_My_M \tag{12.15}$$

subject to the constraints

$$\left.\begin{aligned} a_{11}y_1 + a_{21}y_2 + \cdots + a_{M1}y_M &\geq c_1 \\ a_{12}y_1 + a_{22}y_2 + \cdots + a_{M2}y_M &\geq c_2 \\ &\vdots \\ a_{1N}y_1 + a_{2N}y_2 + \cdots + a_{MN}y_M &\geq c_N \end{aligned}\right\} \tag{12.16}$$

and

$$y_1 \geq 0, y_2 \geq 0, \ldots, y_M \geq 0 \tag{12.17}$$

In matrix notation, the primal and dual problems are formulated as
Primal

$$\left.\begin{array}{ll} \text{Maximize} & Z = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ \text{and} & \mathbf{x} \geq \mathbf{0} \end{array}\right\} \tag{12.18}$$

Dual

$$\left.\begin{array}{ll} \text{Minimize} & V = \mathbf{b}^{\mathrm{T}}\mathbf{y} \\ \text{subject to} & \mathbf{A}^{\mathrm{T}}\mathbf{y} \geq \mathbf{c} \\ \text{and} & \mathbf{y} \geq \mathbf{0} \end{array}\right\} \tag{12.19}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1N} \\ a_{21} & a_{22} & \ldots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \ldots & a_{MN} \end{pmatrix} \tag{12.20}$$

and

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12.21}$$

and where $\mathbf{c}^{\mathrm{T}}$ denotes the transpose of $\mathbf{c}$, etc.

Note that the dual of a dual is the primal.

Duality Theorem: If the primal problem has an optimal solution, then the dual problem also has an optimal solution, and the optimal values of their objective functions are equal (i.e., Max(Z)=Min(V)).

Chvátal (1983, pp. 56, 58); Dantzig (1963, pp. 124–125); Gass (1964, pp. 84, 90); Hillier and Lieberman (1980, pp. 93, 102); Ignizio (1982, p. 174); Munakata (1979, pp. 384, 386); Vanderbei (1997, pp. 53, 56); Wagner (1969, pp. 134–135).

12.7 Special Cases of Linear Programming Problems

12.7.1 Transportation Problem

The *transportation problem* arises in the distribution of material between different locations. The problem is to determine the minimum cost of shipping material from a set of sources to a set of destinations, given constraints on the supply at each source and the demand at each destination. Let

m = number of sources
n = number of destinations
s_i = number of units of supply at source i ($i = 1, 2, \ldots, m$)
d_j = number of units of demand at destination j ($j = 1, 2, \ldots, n$)
c_{ij} = cost per unit of shipping from source i to destination j
x_{ij} = number of units to be shipped from source i to destination j

The shipments from each source to each destination are displayed in the following table:

		Destination				
		1	2	⋯	n	Supply
Source	1	x_{11}	x_{12}	…	x_{1n}	s_1
	2	x_{21}	x_{22}	…	x_{2n}	s_2
	⋮	⋮	⋮		⋮	⋮
	m	x_{m1}	x_{m2}	…	x_{mn}	s_m
	Demand	d_1	d_2	…	d_n	

The decision variables x_{ij} in the above table must be chosen so that the row totals are equal to the supply quantities s_i and the column totals are equal to the demand quantities d_j. This ensures that the total number of units shipped from each source matches the supply at that source, and the total number of units shipped to each destination matches the demand at that destination.

The objective is to find the values of the mn variables x_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) to minimize the total shipping cost, C.

The linear programming formulation for the transportation problem is

$$\text{Minimize } C = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{12.22}$$

subject to the constraints

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} &= s_i \quad (i = 1, 2, \ldots, m) \\ \sum_{i=1}^{m} x_{ij} &= d_j \quad (j = 1, 2, \ldots, n) \end{aligned}\right\} \tag{12.23}$$

and

$$x_{ij} \geq 0 \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n) \tag{12.24}$$

where s_i and d_j ($i=1, 2, \ldots, m$; $j=1, 2, \ldots, n$) are constants.

Chvátal (1983, p. 345); Dantzig (1951b, pp. 359–360); Gass (1964, pp. 8, 194); Hillier and Lieberman (1980, p. 122); Ignizio (1982, p. 283); Luenberger (1984, p. 118); Ozan (1986, p. 208); Spivey and Thrall (1970, p. 32); Vanderbei (1997, p. 225); Wagner (1969, p. 168).

The above formulation requires that the total supply is equal to the total demand, i.e.,

$$\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j \tag{12.25}$$

In the more general case, the total supply may be greater than the total demand, i.e.,

$$\sum_{i=1}^{m} s_i \geq \sum_{j=1}^{n} d_j$$

and the problem can then be reformulated in terms of an equality as in Equation 12.25 by adding a fictitious destination, with the demand equal to the difference between the total supply and the total demand

$$\sum_{i=1}^{m} s_i - \sum_{j=1}^{n} d_j$$

and with zero shipping costs from each source (Hillier and Lieberman, 1980, pp. 123, 126; Wagner, 1969, p. 168).

The quantities s_i, d_j, and x_{ij} may often be restricted to integers. Note that, if s_i and d_j are integers, the transportation problem as formulated above has an optimal solution in which each x_{ij} is an integer, and there is no need therefore to specify an integer constraint on x_{ij} (Luenberger, 1984, p. 126; Spivey and Thrall, 1970, p. 32; Wagner, 1969, pp. 167–168). This result is known as the *integrality property* (Ahuja, Magnanti, and Orlin, 1993, p. 318; Chvátal, 1983, p. 327; Vanderbei, 1997, p. 218).

The constraints given by Equation 12.23 are a special case of the general constraints given by Equation 12.2. In matrix notation, the constraint coefficients a_{ij} for the general linear programming problem with N variables and M constraints are given by the general $M \times N$ matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \tag{12.26}$$

as shown in Equation 12.7. For the transportation problem with mn variables and $(m+n)$ constraints, the constraint coefficients are given by the special $(m+n) \times mn$ matrix:

Coefficients of :

$$\begin{matrix} x_{11} & x_{12} & \cdots & x_{1n} & x_{21} & x_{22} & \cdots & x_{2n} & \cdots & x_{m1} & x_{m2} & \cdots & x_{mn} \end{matrix}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 & & & & & & & & & \\ & & & & 1 & 1 & \cdots & 1 & & & & & \\ & & & & & & & & \ddots & & & & \\ & & & & & & & & & 1 & 1 & \cdots & 1 \\ 1 & & & & 1 & & & & & 1 & & & \\ & 1 & & & & 1 & & & & & 1 & & \\ & & \ddots & & & & \ddots & & & & & \ddots & \\ & & & 1 & & & & 1 & & & & & 1 \end{pmatrix} \tag{12.27}$$

where all the empty elements in this matrix are zero. The first m rows are the coefficients for the supply constraints. The remaining n rows are the coefficients for the demand constraints.

Thus, the transportation problem is a special case of the general linear programming problem, in which the constraint coefficients a_{ij} are 0 or 1 in the particular pattern as shown in Equation 12.27. Any linear programming problem that can be formulated with this special structure is called a transportation problem, even if it does not involve the physical transportation of material.

Table 12.1 shows the correspondence between the general linear programming formulation and the transportation problem formulation, based on the notation used in Equation 12.1 through 12.3 and 12.22 through 12.24.

12.7.2 Transshipment Problem

In the transportation problem, shipments only occur from sources to destinations. The *transshipment problem* is an extension of the transportation problem, in which shipments can occur between any two locations (i.e., from source to source, destination to destination, and destination to source, as well as from source to destination). This extension allows sources and destinations to serve as intermediate transfer points (known as transshipment points), so that there are alternative routings in the distribution of material to meet demand.

The transshipment problem can be reformulated as a transportation problem, by treating all locations as both potential sources and destinations, and considering movements of material between all pairs of locations. For the problem of m sources supplying material to meet demand at n destinations, the $m \times n$ transshipment problem becomes an $(m+n) \times (m+n)$ transportation problem.

TABLE 12.1

Correspondence between General and Transportation Problem Formulations

	General Linear Programming Problem (Equations 12.1 through 12.3)	**Transportation Problem (Equations 12.22 through 12.24)**
Number of decision variables	N	mn
Number of constraints	M	$m+n$
Constraint conditions	Inequalities ($\leq$)	Equalities ($=$)
Decision variables	$x_1, x_2, \ldots, x_N$	$x_{11}, \ldots, x_{1n}, \ldots, x_{m1}, \ldots, x_{mn}$
Objective function coefficients	$c_1, c_2, \ldots, c_N$	$c_{11}, \ldots, c_{1n}, \ldots, c_{m1}, \ldots, c_{mn}$
Constraint constants	$b_1, b_2, \ldots, b_M$	$s_1, \ldots, s_m, d_1, \ldots, d_n$
Constraint coefficients	a_{ij} (as in Equation 12.26)	0 and 1 (as in Equation 12.27)

Hillier and Lieberman (1980, pp. 146–151); Ignizio (1982, pp. 305–312); Ozan (1986, pp. 259–262); Wagner (1969, pp. 171–176).

12.7.3 Assignment Problem

The *assignment problem* is a special case of the transportation problem, with specific values for the decision variables and the constraint constants. It arises in the decision on how to allocate a group of individuals to a group of tasks, under the following rules:

- Each individual is assigned to one task only.
- Each task is performed by one individual only.

The number of tasks is equal to the number of individuals. The problem is to determine the assignment of individuals to tasks to minimize total cost (or time), given the costs (or times) of each individual for each task. Let

n = number of individuals = number of tasks

c_{ij} = cost of assigning individual i to task j

$$x_{ij} = \begin{cases} 1 & \text{if individual } i \text{ is assigned to task } j \\ 0 & \text{otherwise} \end{cases}$$

The assignments of individuals to tasks are displayed in the following table:

		Task				
		1	2	⋯	n	Total
Individual	1	x_{11}	x_{12}	…	x_{1n}	1
	2	x_{21}	x_{22}	…	x_{2n}	1
	⋮	⋮	⋮		⋮	⋮
	n	x_{n1}	x_{n2}	…	x_{nn}	1
	Total	1	1	⋯	1	n

The decision variables x_{ij} in the above table must be chosen so that each row and each column adds up to 1. This ensures that there is exactly one task per individual, and exactly one individual per task.

The objective is to find the values of the n^2 variables x_{ij} ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$) to minimize the total assignment cost, C.

The linear programming formulation for the assignment problem is

$$\text{Minimize } C = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{12.28}$$

subject to the constraints

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1 \quad (i = 1, 2, \ldots, n) \\ \sum_{i=1}^{n} x_{ij} &= 1 \quad (j = 1, 2, \ldots, n) \end{aligned}\right\} \tag{12.29}$$

and

$$x_{ij} \geq 0 \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n) \tag{12.30}$$

Bradley, Hax, and Magnanti (1977, p. 316); Gass (1964, p. 210); Hillier and Lieberman (1980, pp. 151–153); Ignizio (1982, p. 323); Luenberger (1984, p. 133); Ozan (1986, p. 271); Spivey and Thrall (1970, pp. 217–218); Vanderbei (1997, p. 226); Wagner (1969, p. 177).

The assignment problem given by Equations 12.28 through 12.30 is a special case of the transportation problem given by Equations 12.22 through 12.24 in which

$$m = n$$

$$s_i = 1 \quad (i = 1, 2, \ldots, n)$$

$$d_j = 1 \quad (j = 1, 2, \ldots, n)$$

and in which the decision variables x_{ij} are constrained to be integers 0 or 1.

Note that the integer constraint on x_{ij} need not be specified in the assignment problem formulation. The standard constraint $x_{ij} \geq 0$ is sufficient (even though it allows fractional values for x_{ij}), since the assignment problem as formulated above has an optimal solution in which each x_{ij} is 0 or 1 (Bradley, Hax, and Magnanti, 1977, p. 316; Luenberger, 1984, p. 134; Spivey and Thrall, 1970, p. 218). This result is an example of the *integrality property* (Ahuja, Magnanti, and Orlin, 1993, p. 318; Chvátal, 1983, p. 327; Vanderbei, 1997, p. 218).

12.8 Integer Linear Programming Formulations

Integer linear programming problems are linear programming problems in which some or all of the decision variables are constrained to be

integers. For the N variables $x_1, x_2, \ldots, x_N$, the general formulation for an integer linear programming problem is

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{12.31}$$

subject to the constraints

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &\le b_2 \\ &\vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &\le b_M \end{aligned}\right\} \tag{12.32}$$

$$x_1 \ge 0,\ x_2 \ge 0,\ \ldots,\ x_N \ge 0 \tag{12.33}$$

and

$$x_j \text{ integer (for some or all } j = 1, 2, \ldots, N) \tag{12.34}$$

where a_{ij}, b_i, c_j (i=1, 2, …, M; j=1, 2, …, N) are constants.

Bradley, Hax, and Magnanti (1977, p. 366); Garfinkel and Nemhauser (1972, p. 5); Nemhauser and Wolsey (1988, p. 27).

An example of the general formulation for an integer linear programming problem is the *diet problem* given in Section 12.5, with the added constraint that some or all of the numbers of units of food per day in the diet are restricted to integers. Examples of special cases of integer linear programming problems are the *knapsack problem* and the *traveling salesman problem*.

12.8.1 Knapsack Problem

The *knapsack problem* arises in the selection of items to include in a knapsack, given a limit on how much can be carried. For items of different values and different weights, the problem is to determine the optimal (i.e., most valuable) selection of items, subject to a weight constraint on the total number of items. Let

N=number of types of items
c_j=value of item type j (j=1, 2, …, N)
a_j=weight of item type j (j=1, 2, …, N)
b=limit on total weight of items
x_j=number of items of type j included in knapsack

The objective is to find the values of the N variables x_1, x_2, …, x_N to maximize the total value Z of items included in knapsack.

The integer linear programming formulation for the knapsack problem is

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{12.35}$$

subject to the constraints

$$a_1x_1 + a_2x_2 + \cdots + a_Nx_N \le b \tag{12.36}$$

$$x_1 \ge 0,\ x_2 \ge 0,\ \ldots,\ x_N \ge 0 \tag{12.37}$$

and

$$x_1, x_2, \ldots, x_N \text{ integers} \tag{12.38}$$

where a_j, b, c_j (j=1, 2, …, N) are constants.

The knapsack problem given by Equations 12.35 through 12.38 is a special case of the integer linear programming problem given by Equations 12.31 through 12.34, in which the set of constraints in Equation 12.32 reduces to only one constraint (i.e., number of constraints M=1) with a_{ij}=a_j, and b_i=b, as given by Equation 12.36.

If no more than one of any item type may be included in the knapsack, then the integer constraint given by Equation 12.38 is replaced by the constraint

$$x_j = \begin{cases} 1 & \text{if item type } j \text{ is included in the knapsack} \\ 0 & \text{otherwise} \end{cases} \qquad (j = 1, 2, \ldots, N) \tag{12.39}$$

and the problem is called a *0-1 knapsack problem*.

Bradley, Hax, and Magnanti (1977, p. 368); Garfinkel and Nemhauser (1972, p. 215); Murty (1976, p. 404); Ozan (1986, p. 362).

12.8.2 Traveling Salesman Problem

The classic *traveling salesman problem* arises in the choice of route to visit a set of locations. The problem is to determine the shortest tour connecting all locations, when each location is to be visited exactly once. Let

n=number of locations to be visited

c_{ij}=travel distance from location i to location j (i, j=1, 2, …, n)

$$x_{ij} = \begin{cases} 1 & \text{if tour leads directly from location } i \text{ to location } j \\ 0 & \text{otherwise} \end{cases} \qquad (i, j = 1, 2, \ldots, n)$$

The objective is to find the values of the n^2 variables x_{ij} ($i=1, 2, \ldots, n$; $j=1, 2, \ldots, n$) to minimize the total tour distance, C. Note that c_{ij} and C may represent travel time or cost, rather than distance.

The integer linear programming formulation for the traveling salesman problem is

$$\text{Minimize } C = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{12.40}$$

subject to the constraints

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} = 1 \quad (i = 1, 2, \ldots, n) \\ \sum_{i=1}^{n} x_{ij} = 1 \quad (j = 1, 2, \ldots, n) \end{aligned}\right\} \tag{12.41}$$

$$x_{ij} \geq 0 \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n) \tag{12.42}$$

$$x_{ij} \text{ integers} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n) \tag{12.43}$$

and

$$\sum_{i \in S} \sum_{j \notin S} x_{ij} \geq 1 \tag{12.44}$$

where S is any nonempty subset of the n locations.

Bradley, Hax, and Magnanti (1977, pp. 371–372); Garfinkel and Nemhauser (1972, p. 356).

To ensure that a tour does not contain any link going from and to the same location, the solution must have $x_{ii}=0$ for all i. This may be achieved by setting c_{ii} to a very large number for all i.

The first set of constraints in Equation 12.41, i.e., $\sum_{j=1}^{n} x_{ij} = 1 (i = 1, 2, \ldots, n)$, ensures that the tour leaves each location exactly once. The

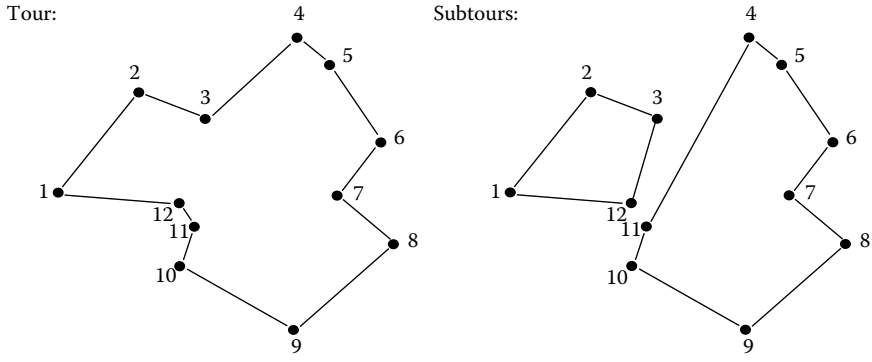


FIGURE 12.2
Tour and subtours in traveling salesman problem.

second set of constraints in Equation 12.41, i.e., $\sum_{i=1}^{n} x_{ij} = 1 (j = 1, 2, \ldots, n)$, ensures that the tour enters each location exactly once.

The set of constraints for subsets S, given by Equation 12.44, ensures that the tour connects all n locations. Without such constraints, the problem formulation would allow for disconnected subtours, which are not valid. Figure 12.2 illustrates the cases of a tour and subtours. The requirement that subtours be excluded can also be formulated in other ways. An alternative formulation for excluding subtours is given by the constraints

$$u_i - u_j + n x_{ij} \leq n - 1 \quad (i = 2, 3, \ldots, n;\ j = 2, 3, \ldots, n;\ i \neq j) \tag{12.45}$$

where u_i (i=2, 3, …, n) are arbitrary real numbers (Gass, 1964, pp. 165–166; Murty, 1976, p. 411; Wagner, 1969, pp. 455–456).

Note that, without a set of constraints to exclude subtours, the formulation would be the same as for the assignment problem. Thus, the solution to the assignment problem provides a lower bound on the cost C in the traveling salesman problem (Garfinkel and Nemhauser, 1972, p. 356).

12.9 Solution Methods

12.9.1 Simplex Method

The standard technique for solving general linear programming problems is the *simplex method*. It can be used to solve maximization or minimization problems with any of the standard constraints. The simplex method

systematically searches for solutions on the boundary of the feasible region, improving on the value of the objective function with each iteration until the optimum is reached.

Chvátal (1983, pp. 13–134); Cohen (1985, pp. 22–25); Cormen, Leiserson, Rivest, and Stein, (2001); Dantzig (1951a, pp. 339–347); Dantzig (1963, pp. 94–111); Gass (1964, pp. 59–80, 96–113); Hillier and Lieberman (1980, pp. 33–46, 68–91); Karloff (1991, pp. 23–47); Munakata (1979, pp. 323–348); Vanderbei (1997, pp. 11–110); Wagner (1969, pp. 96–122).

12.9.2 Interior-Point Methods

A different technique for solving general linear programming problems is the class of methods known as *interior-point methods*. These methods include *Karmarkar's algorithm* and *primal-dual interior-point methods*. Interior-point methods search for solutions in the interior of the feasible region, improving on the value of the objective function with each iteration. At the final step, the search procedure jumps to the boundary of the feasible region for the optimal solution.

Karloff (1991, pp. 103–130); Karmarkar (1984, pp. 373–395); Mehrotra (1992, pp. 575–601); Wright (1997, pp. 4–46).

12.9.3 Network Flow Methods

A technique for solving the above special cases of linear programming problems (transportation problem, transshipment problem, assignment problem) is the class of methods known as *network flow methods*. These methods take advantage of the special structure of such problems, representing each problem in terms of flows on a network of nodes connected by arcs.

Ahuja, Magnanti, and Orlin (1993, pp. 294–344); Bradley, Hax, and Magnanti (1977, pp. 310–319); Cormen, Leiserson, Rivest, and Stein (2001, pp. 643–663); Ignizio (1982, pp. 344–368).

12.9.4 Cutting Planes

A technique for solving integer linear programming problems is the method of *cutting planes*. Cutting planes are additional linear constraints that are introduced into the linear programming formulation. By the addition of appropriate constraints, the cutting planes method systematically cuts away part of the feasible region to eliminate non-integer solutions without eliminating feasible integer solutions.

Garfinkel and Nemhauser (1972, pp. 155–207); Nemhauser and Wolsey (1988, pp. 367–378); Ozan (1986, pp. 306–323).

12.9.5 Branch and Bound

Another technique for solving integer linear programming problems is the method known as *branch and bound*. Since the decision variables in integer programming problems are discrete, the number of feasible solutions is finite. If the number is small, the solutions can be enumerated. In general, however, there are too many feasible solutions to allow a complete enumeration. Branch and bound provides a systematic enumeration procedure that considers bounds on the objective function for different subsets of solutions and eliminates the subsets of nonoptimal solutions.

Bradley, Hax, and Magnanti (1977, pp. 387–395); Garfinkel and Nemhauser (1972, pp. 111–122); Hillier and Lieberman (1980, pp. 716–732); Murty (1976, pp. 437–478); Nemhauser and Wolsey (1988, pp. 355–367); Ozan (1986, pp. 324–351).

13

Heuristic Search Methods

13.1 Overview of Heuristics

Let $x_1, x_2, \ldots, x_N$ be N decision variables in an optimization problem. In the general case, the problem is to search for values of the variables, $x_1, x_2, \ldots, x_N$, to maximize (or minimize) a given general function of the variables, subject to a given set of general constraints on the variables. The general formulation for the maximization problem is

$$\text{Maximize } f\left(x_1, x_2, \ldots, x_N\right) \tag{13.1}$$

subject to the M constraints

$$\left.\begin{aligned} g_1\left(x_1, x_2, \ldots, x_N\right) &\le b_1 \\ g_2\left(x_1, x_2, \ldots, x_N\right) &\le b_2 \\ &\vdots \\ g_M\left(x_1, x_2, \ldots, x_N\right) &\le b_M \end{aligned}\right\} \tag{13.2}$$

where
$f(.)$ and $g_i(.)$ $(i=1, 2, \ldots, M)$ are general functions
b_i $(i=1, 2, \ldots, M)$ are constants

Expressing the decision variables in vector notation:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \tag{13.3}$$

the general formulation for the maximization problem is written as

$$\text{Maximize } f(\mathbf{x}) \tag{13.4}$$

subject to the constraints

$$g_i(\mathbf{x}) \leq b_i \quad (i = 1, 2, \ldots, M) \tag{13.5}$$

For the minimization problem, the corresponding general formulation is

$$\text{Minimize } f(\mathbf{x}) \tag{13.6}$$

subject to the constraints

$$g_i(\mathbf{x}) \geq b_i \quad (i = 1, 2, \ldots, M) \tag{13.7}$$

A set of values for $\mathbf{x}$ that satisfies all the constraints in a given problem is a feasible solution. The collection of all feasible solutions is the feasible search space.

In cases where $f(\mathbf{x})$ and $g_i(\mathbf{x})$ take on particular functional forms, the optimization problem can be solved using specific techniques or algorithms. For example, in the special case where $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are linear functions of the decision variables, the problem reduces to a linear programming problem with standard solution methods (see Chapter 12). Also, for certain types of functions $f(\mathbf{x})$ and $g_i(\mathbf{x})$, the problem can be solved using Lagrange multipliers (see Chapter 16). In many cases, however, there is no known method for finding the optimal solution, and heuristics are often employed to obtain an acceptable solution.

Heuristics can be defined as methods that seek near-optimal solutions, using a reasonable amount of computational effort. They typically involve systematic but non-rigorous methods of exploration. Unlike formal mathematical optimization methods, heuristics do not guarantee optimal solutions, and may not even guarantee feasible solutions, but can generally yield attainable solutions for use in practice.

If the variables $x_1, x_2, \ldots, x_N$ take discrete values, the problem becomes a combinatorial optimization problem, with a finite number of feasible solutions. However, in general the number would be too large to enumerate them all and find the optimal solution. Various heuristics have been developed to reduce the number of computations needed in any optimization problem. Examples of commonly used heuristics are given in the following sections.

Heuristics typically start with a feasible solution and use an iterative procedure to search for improved solutions. For the minimization problem (Equation 13.6) with feasible search space $\mathcal{F}$, an heuristic searches for a practical solution close to the optimal solution $\mathbf{x}^*$ where, for any $\mathbf{x} \in \mathcal{F}$,

$$f(\mathbf{x}^*) < f(\mathbf{x})$$

Gilli and Winker (2008); Michalewicz and Fogel (2004); Rayward-Smith, Osman, Reeves, and Smith (1996); Reeves (1993); Silver (2004); Winkler and Gilli (2004).

13.2 Local Search Methods

Local search (or neighborhood search) methods are a class of heuristics in which the iterative procedure starts with a feasible solution, and then at each iteration finds an improvement on the current solution by searching the neighborhood of the current solution. This neighborhood is a set of feasible solutions where the values of the decision variables are close to those of the current solution. Each time a new solution in the neighborhood is an improvement, it is used to update the current solution. The iterative procedure ends based on pre-specified stopping criteria, such as when no further improvement is found or when the total number of iterations reaches a given limit.

Let

$\mathcal{F}$ be the feasible search space

$\mathcal{N}(\mathbf{x})$ be the set of solutions in the neighborhood of $\mathbf{x}$, where $\mathcal{N}(\mathbf{x}) \subset \mathcal{F}$

In terms of the minimization problem given by Equation 13.6, the algorithm for a local search heuristic consists of the following steps:

1. Select a solution $\mathbf{x}^{current} \in \mathcal{F}$ as the initial current solution
2. Select a new solution $\mathbf{x}^{new} \in \mathcal{N}(\mathbf{x}^{current})$
3. If $f(\mathbf{x}^{new}) < f(\mathbf{x}^{current})$ then set $\mathbf{x}^{new} = \mathbf{x}^{current}$
4. Repeat steps 2 and 3 until the stopping criteria is met (e.g., limit on number of iterations)

Gilli and Winker (2008); Michalewicz and Fogel (2004); Michiels, Aarts, and Korst (2007); Rayward-Smith, Osman, Reeves, and Smith (1996); Silver (2004); Winkler and Gilli (2004).

Note that a local search heuristic has the drawback that it might yield a local optimum rather than the global optimum. The performance of the

heuristic depends on the choice of size for a neighborhood $\mathcal{N}(\mathbf{x})$. If the size of the neighborhood is small, the algorithm needs only to search for a few solutions at each iteration, but is likely to become trapped in a local optimum. If the size of the neighborhood is large, the number of computations is likely to become prohibitively large. The trade-off between neighborhood size and efficiency of the search is a limitation of the local search method.

13.3 Simulated Annealing

Simulated annealing is an iterative heuristic that extends the local search method, described above, to allow for a new solution at some iterations to be worse than the current solution, rather than an improvement. This extension helps to avoid getting trapped in a local optimum. By accepting inferior solutions in some neighborhoods, the heuristic searches more widely within the feasible search space, so that it is more likely to escape a local optimum and move to the global optimum.

Like the local search method, the simulated annealing heuristic searches for a new solution $\mathbf{x}^{new}$ at each iteration in the neighborhood of the current solution $\mathbf{x}^{current}$. If the new solution is an improvement, it is accepted as the update to the current solution, just as in the local search method. In addition, if the new solution is inferior to the current solution, the new solution is sometimes accepted, with a given probability that depends on the difference between the values of $f(\mathbf{x})$ for the new and current solutions. The bigger this difference, the smaller the probability that the new (inferior) solution is accepted as the update to the current solution. The acceptance probability is determined by whether a random number u generated between 0 and 1 is less than or greater than the function $e^{-\Delta/T}$, where Δ is the difference between $f(\mathbf{x}^{new})$ and $f(\mathbf{x}^{current})$, and T is a specified control parameter. When a new solution is inferior, it is still accepted if u satisfies the condition

$$u < e^{-\Delta/T}$$

Figure 13.1 illustrates this acceptance criterion. When incorporated into the heuristic, it provides a stochastic approach for seeking solutions away from a local optimum.

The simulated annealing heuristic is based on an analogy to the physical process of annealing that occurs in thermodynamics, when a heated material cools down and changes its structure under a controlled temperature-lowering schedule. The function $e^{-\Delta/T}$ is the equivalent of the Boltzmann

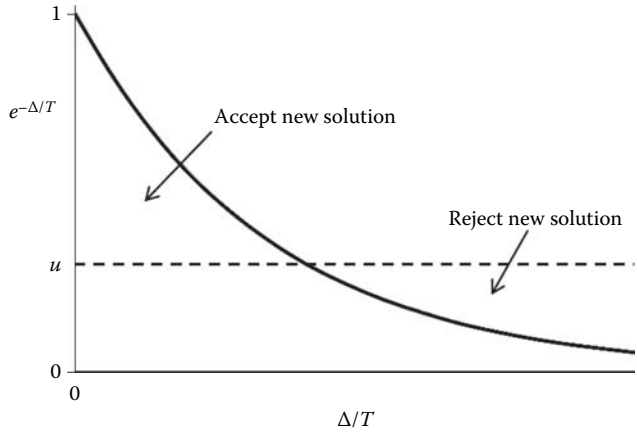


FIGURE 13.1
Acceptance criterion for the simulated annealing heuristic.

factor used in the physical annealing process to express the probability of the state of a system in terms of energy and temperature (Kittel and Kroemer, 1980; Landsberg, 1978; Sturge, 2003). Simulated annealing is a method that simulates this process in order to calculate properties of the changing states of a material during annealing (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller, 1953). Analogous to investigating states in a physical system, the same method has been applied to calculating solutions for optimization problems (Černý, 1985; Kirkpatrick, Gelatt, and Vecchi, 1983; Schwarzschild, 1982).

In applying this analogy for the simulated annealing heuristic in optimization, the control parameter T in the function $e^{-\Delta/T}$ is referred to as the "temperature." The "temperature" T is initially set at a high value, in order to accept inferior solutions frequently, and is then gradually lowered as the iterative procedure progresses to allow fewer and fewer inferior solutions.

Let

$\mathcal{F}$ be the feasible search space

$\mathcal{N}(\mathbf{x})$ be the set of solutions in the neighborhood of $\mathbf{x}$, where $\mathcal{N}(\mathbf{x}) \subset \mathcal{F}$

In terms of the minimization problem given by Equation 13.6, the algorithm for a simulated annealing heuristic consists of the following steps:

1. Select a solution $\mathbf{x}^{current} \in \mathcal{F}$ as the initial current solution
2. Set value for the control parameter T
3. Select a new solution $\mathbf{x}^{new} \in \mathcal{N}(\mathbf{x}^{current})$
4. Compute $\Delta = f(\mathbf{x}^{new}) - f(\mathbf{x}^{current})$

5. Generate a random number u from the uniform distribution in the range [0, 1]
6. If ($\Delta < 0$), or if ($\Delta > 0$ and $u < e^{-\Delta/T}$), then set $\mathbf{x}^{new} = \mathbf{x}^{current}$
7. Repeat steps 2 and 6 until the stopping criteria is met (e.g., limit reached on number of iterations)

Dowsland (1993); Duque-Antón (1997); Gilli and Winker (2008); Michalewicz and Fogel (2004); Monticelli, Romero, and Asada (2008); Rayward-Smith, Osman, Reeves, and Smith (1996); Silver (2004); Winkler and Gilli (2004).

13.4 Tabu Search

Tabu search is an iterative heuristic that, like simulated annealing, is based on the local search method and includes a strategy to avoid a local optimum. Unlike simulated annealing, however, the basic form of tabu search uses a deterministic rather than a stochastic approach for avoiding a local optimum. The tabu search heuristic is a search method with memory. The memory is structured to record previously visited solutions and other solutions that are not wanted. The heuristic uses the memory to ensure such solutions are forbidden (tabu) in the search for a new solution. This helps to avoid cycling (i.e., visiting the same solution more than once) and to broaden the search within the feasible search space.

A tabu list contains the forbidden solutions. This list is updated as the heuristic goes through more iterations. Criteria vary in different problems for determining which solutions are added to the tabu list and for how many iterations they remain on the list.

Let
$\mathcal{F}$ be the feasible search space
$\mathcal{N}(\mathbf{x})$ be the set of solutions in the neighborhood of $\mathbf{x}$, where $\mathcal{N}(\mathbf{x}) \subset \mathcal{F}$
$\mathcal{T}$ be the tabu list (i.e., the set of forbidden solutions)

In terms of the minimization problem given by Equation 13.6, the algorithm for a tabu search heuristic consists of the following steps:

1. Select a solution $\mathbf{x}^{current} \in \mathcal{F}$ as the initial current solution
2. Select a new solution $\mathbf{x}^{new} \in \mathcal{N}(\mathbf{x}^{current})$
3. If $f(\mathbf{x}^{new}) < f(\mathbf{x}^{current})$ then set $\mathbf{x}^{new} = \mathbf{x}^{current}$
4. Add $\mathbf{x}^{new}$ to the tabu list $\mathcal{T}$
5. Update memory (e.g., delete older solutions from $\mathcal{T}$)
6. Repeat steps 2 to 5 until the stopping criteria is met (e.g., limit reached on number of iterations)

Gilli and Winker (2008); Glover (1986); Glover and Laguna (1993); Glover, Tallard, and de Werra (1993); Ibidapo-Obe and Asaolu (2006); Michalewicz and Fogel (2004); Rayward-Smith, Osman, Reeves, and Smith (1996); Silver (2004); Winkler and Gilli (2004).

13.5 Genetic Algorithms

Genetic algorithms are iterative heuristics that use a group (or population) of current solutions, rather than one current solution at a time as in simulated annealing or tabu search methods, in order to find new improved solutions.

At each iteration, the heuristic for a genetic algorithm evaluates the individual solutions in the current population, and uses the evaluations to define probabilities for selecting solutions. For each solution $\mathbf{x}$, the evaluations may be the values of the function $f(\mathbf{x})$ to be optimized, or some other related measure. The probabilities determine the likelihood of each solution being selected, and are used to randomly select pairs of solutions. For each pair, the heuristic combines part of one solution with part of the other, so that the components from the individual solutions are recombined to construct a new solution. Occasionally, some components of solutions at random are also altered to widen the exploration of solutions throughout the search space. The set of new solutions becomes the new population for the next iteration. The algorithm ends according to pre-specified stopping criteria, such as when an acceptable solution is found or when the total number of iterations reaches a given limit.

The heuristic is based on an analogy to the generation of successive populations of organisms, where individuals of each new population inherit genetic characteristics of pairs of individuals from the previous population. In this analogy for a mathematical optimization problem, solutions represent the individuals in a population.

Let
$\boldsymbol{P}$ be a population of solutions
$\mathbf{x}$ be a solution in $\boldsymbol{P}$

A genetic algorithm consists of the following steps:

1. Generate a population $\boldsymbol{P}$ of solutions as the initial population
2. Evaluate each solution $\mathbf{x}$ in $\boldsymbol{P}$
3. Define probabilities $p(\mathbf{x})$ for selecting solutions based on evaluations
4. Use probabilities $p(\mathbf{x})$ to generate pairs of solutions $\mathbf{x}^a$ and $\mathbf{x}^b$ from $\boldsymbol{P}$
5. Recombine components from each pair $\mathbf{x}^a$ and $\mathbf{x}^b$ to form new solutions $\mathbf{x}^c$

6. Alter some components at random with a pre-specified low rate
7. Use new solutions $\mathbf{x}^c$ as new population $\boldsymbol{P}$
8. Repeat steps 2 to 7 until the stopping criteria is met (e.g., acceptable solution found or limit reached on number of iterations)

Chu and Beasley (1997); Gilli and Winker (2008); Michalewicz and Fogel (2004); Rayward-Smith, Osman, Reeves, and Smith (1996); Silver (2004); Winkler and Gilli (2004).

13.6 Other Heuristics

Various other heuristics have been developed to solve optimization problems (Lee and El-Sharkawi, 2008; Michalewicz and Fogel, 2004; Reeves, 1993; Silver, 2004). An heuristic known as the ant colony optimization uses the analogy of ants searching for food, and mimics their collective behavior in establishing the shortest routes between two points (Blum and Dorigo, 2004; Dorigo, Maniezzo, and Colorni, 1996; Song, Lu, Lee, and Yo, 2008).

An enhancement of local search methods (Section 13.2) is the variable neighborhood search heuristic, which allows for a systematic change of neighborhoods to avoid a local optimum (Hansen and Mladenović, 2001). One of the simplest heuristics for solving general optimization problems is the greedy algorithm, which sequentially finds the best value for each of the decision variables (Cormen, Leiserson, Rivest, and Stein, 2001; Michalewicz and Fogel, 2004). The greedy algorithm uses a convenient but myopic approach that may yield acceptable but suboptimal solutions. To improve exploration of the search space, heuristics have been developed that combine greedy algorithms with randomized search techniques (Consoli, Darby-Dowman, Mladenović, and Pérez, 2009; Feo and Resende, 1995).

For some optimization problems, various techniques have been constructed that integrate two of more basic heuristic procedures to form hybrid heuristics (Clark, 2003; Lee and Lee, 2005; Michalewicz and Fogel, 2004; Silver, 2004; Winkler and Gilli, 2004).

14

Order Statistics

14.1 General Distribution Order Statistics

Let X_1, X_2, …, X_n be n independent and identically distributed random variables, each with a probability density function $f(x)$ and a cumulative distribution function $F(x)$, i.e.,

$$\Pr\{X_i \le x\} = F(x) = \int_{-\infty}^{x} f(t)\,dt$$

For given X_1, X_2, …, X_n, the order statistics $X_{(1)}$, $X_{(2)}$, …, $X_{(n)}$ are random variables with their values arranged in ascending order, i.e.,

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

In particular, the random variables $X_{(1)}$ and $X_{(n)}$ are the smallest and largest order statistics, respectively, i.e.,

$$X_{(1)} = \min\left(X_1, X_2, \ldots, X_n\right)$$

$$X_{(n)} = \max\left(X_1, X_2, \ldots, X_n\right)$$

In general, the random variable $X_{(k)}$ is known as the kth-order statistic.
Let

$f_k(x)$ be the probability density function of $X_{(k)}$

$F_k(x)$ be the cumulative distribution function of $X_{(k)}$

Then $f_k(x)$ is given by

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!}\{F(x)\}^{k-1}\{1-F(x)\}^{n-k} f(x) \tag{14.1}$$

and $F_k(x)$ is given by

$$F_k(x) = \sum_{i=k}^{n} \binom{n}{i} \{F(x)\}^i \{1 - F(x)\}^{n-i} \tag{14.2}$$

where

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

Balakrishnan and Sinha (1995, p. 18); Bolch, Greiner, de Meer, and Trivedi (1998, p. 31); Cox and Hinkley (1974, p. 466); Ross (2003, pp. 60–61).

When $k=1$, Equations 14.1 and 14.2 reduce to the probability density function and the cumulative distribution function for the minimum of the n random variables, $\min(X_1, X_2, \ldots, X_n)$:

$$f_1(x) = n\{1 - F(x)\}^{n-1} f(x) \tag{14.3}$$

$$F_1(x) = 1 - \{1 - F(x)\}^n \tag{14.4}$$

(see Equations 5.5 and 5.3).

Similarly, when $k=n$, Equations 14.1 and 14.2 reduce to the probability density function and the cumulative distribution function for the maximum of the n random variables, $\max(X_1, X_2, \ldots, X_n)$:

$$f_n(x) = n\{F(x)\}^{n-1} f(x) \tag{14.5}$$

$$F_n(x) = \{F(x)\}^n \tag{14.6}$$

(see Equations 5.4 and 5.2).

14.2 Uniform Distribution Order Statistics

For the special case where $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are order statistics of a uniform distribution over the range [0,1], i.e.,

$$f(x)=1 \quad (0\le x\le 1)$$

$$F(x)=x \quad (0\le x\le 1)$$

the probability density function $f_k(x)$ for the kth-order statistic $X_{(k)}$ ($k=1, 2, \ldots, n$) is given by

$$f_k(x)=\frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k} \quad (0\le x\le 1) \tag{14.7}$$

David (1970, pp. 8, 11); Ebrahimi, Soofi, and Zahedi (2004, p. 177).

Equation 14.7 is the probability density function for a beta distribution (see Equation 4.30). Thus, the kth-order statistic of a uniform distribution has a beta distribution.

The mean and the variance of the kth-order statistic, $X_{(k)}$, in this case are given by

$$E[X_{(k)}]=\frac{k}{n+1} \tag{14.8}$$

and

$$Var[X_{(k)}]=\frac{k(n-k+1)}{(n+1)^2(n+2)} \tag{14.9}$$

respectively (Ahsanullah and Nevzorov, 2005, pp. 104–105).

When $k=1$, Equations 14.8 and 14.9 reduce to the mean and the variance of the minimum of n random variables from a uniform distribution:

$$E[X_{(1)}]=\frac{1}{n+1} \tag{14.10}$$

and

$$Var[X_{(1)}]=\frac{n}{(n+1)^2(n+2)} \tag{14.11}$$

respectively.

Similarly, when $k=n$, Equations 14.8 and 14.9 reduce to the mean and the variance of the maximum of n random variables from a uniform distribution:

$$E[X_{(n)}]=\frac{n}{n+1} \tag{14.12}$$

and

$$Var[X_{(n)}] = \frac{n}{(n+1)^2(n+2)} \tag{14.13}$$

respectively.

Note that the variances of the minimum and maximum, given by Equations 14.11 and 14.13, respectively, are the same, i.e., $Var[X_{(1)}] = Var[X_{(n)}]$ when the order statistics are from a uniform distribution.

14.3 Exponential Distribution Order Statistics

For the special case where $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are order statistics of an exponential distribution with parameter λ, i.e.,

$$f(x) = \lambda e^{-\lambda x} \quad (0 \le x < \infty)$$

$$F(x) = 1 - e^{-\lambda x} \quad (0 \le x < \infty)$$

the probability density function $f_k(x)$ for the kth-order statistic $X_{(k)}$ (k=1, 2, …, n) is given by

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} \lambda \left(1 - e^{-\lambda x}\right)^{k-1} e^{-(n-k+1)\lambda x} \quad (0 \le x < \infty) \tag{14.14}$$

Balakrishnan and Sinha (1995, p. 18).

The mean and the variance of the kth-order statistic $X_{(k)}$ in this case are given by

$$E[X_{(k)}] = \sum_{i=1}^{k} \frac{1}{n-i+1} = \frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-k+1} \tag{14.15}$$

and

$$Var[X_{(k)}] = \sum_{i=1}^{k} \frac{1}{(n-i+1)^2} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2} + \cdots + \frac{1}{(n-k+1)^2} \tag{14.16}$$

respectively Ahsanullah and Nevzorov (2005, p. 112); Balakrishnan and Sinha (1995, p. 19); Cox and Hinkley (1974, p. 468).

When $k=1$, Equation 14.14 reduces to

$$f_1(x) = n\,\lambda e^{-n\lambda x} \quad (0 \le x < \infty) \tag{14.17}$$

Thus the minimum of n random variables from an exponential distribution with parameter λ has an exponential distribution with parameter $n\lambda$. The mean and the variance for the case $k=1$ (i.e., for the minimum of n random variables from an exponential distribution) are given by

$$E[X_{(1)}] = \frac{1}{n\lambda} \tag{14.18}$$

$$Var[X_{(1)}] = \frac{1}{(n\lambda)^2} \tag{14.19}$$

When $k=n$, Equation 14.14 becomes

$$f_n(x) = n\,\lambda\left(1 - e^{-\lambda x}\right)^{n-1} e^{-\lambda x} \quad (0 \le x < \infty) \tag{14.20}$$

and from Equations 14.15 and 14.16 the mean and the variance for the maximum of n random variables from an exponential distribution are given by Equations 2.47 and 2.48, i.e.,

$$E\left[X_{(n)}\right] = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \tag{14.21}$$

and

$$Var\left[X_{(n)}\right] = \frac{1}{\lambda^2}\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}\right) \tag{14.22}$$

respectively Balakrishnan and Sinha (1995, p. 19); Cox and Hinkley (1974, p. 468); Nahmias (1989, p. 553).

15

Mathematical Functions

15.1 Gamma Function

The gamma function, denoted by $\Gamma(n)$, is defined by

$$\Gamma(n)=\int_0^\infty t^{n-1}e^{-t}dt \tag{15.1}$$

For any $n>0$:

$$\Gamma(n+1)=n\Gamma(n) \tag{15.2}$$

For n a positive integer:

$$\Gamma(n+1)=n! \tag{15.3}$$

Abramowitz and Stegun (1968, pp. 255–256); Binmore (1983, pp. 258–259); Landsberg (1978, p. 345); Zwillinger (1996, pp. 494–495).

15.2 Incomplete Gamma Function

The incomplete gamma function, denoted by $\gamma(n, x)$, is defined by

$$\gamma(n, x)=\int_0^x t^{n-1}e^{-t}dt \tag{15.4}$$

For any n, limit as $x \to \infty$:

$$\gamma(n, \infty)=\Gamma(n) \tag{15.5}$$

where $\Gamma(n)$ is the gamma function, given by Equation 15.1.

For n a positive integer:

$$\gamma(n, x) = (n-1)!\left\{1 - e^{-x}\sum_{i=0}^{n-1}\frac{x^i}{i!}\right\} \tag{15.6}$$

Abramowitz and Stegun (1968, pp. 260, 262); Gradshteyn and Ryzhik (1965, p. 940); Press, Teukolsky, Vetterling, and Flannery (2007, p. 259).

Note that the summation in Equation 15.6 appears in the cumulative distribution functions for the Poisson probability distribution (Equation 3.18) and the Erlang probability distribution (Equations 4.12 and 4.13).

15.3 Beta Function

The beta function, denoted by $B(m, n)$, is defined by

$$B(m,n) = \int_0^1 t^{m-1}(1-t)^{n-1}\,dt \tag{15.7}$$

$$= \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} \tag{15.8}$$

where $\Gamma(m)$ and $\Gamma(n)$ are gamma functions, given by Equation 15.1.

For m and n positive integers:

$$B(m,n) = \frac{(m-1)!(n-1)!}{(m+n-1)!} \tag{15.9}$$

Abramowitz and Stegun (1968, p. 258); Zwillinger (1996, pp. 497–498).

15.4 Incomplete Beta Function

The incomplete beta function, denoted by $B(x; m, n)$, is defined by

$$B(x;m,n) = \int_0^x t^{m-1}(1-t)^{n-1}\,dt \tag{15.10}$$

For special case $x=1$:

$$B(1;m,n)=B(m,n) \tag{15.11}$$

where $B(m, n)$ is the beta function, given by Equation 15.7.
Abramowitz and Stegun (1968, p. 258).

15.5 Unit Impulse Function

$$u_0(t)=\begin{cases}\infty & (t=0)\\ 0 & (t\neq 0)\end{cases} \tag{15.12}$$

and

$$\int_{-\infty}^{\infty} u_0(t)dt=1 \tag{15.13}$$

The unit impulse function is also known as the Dirac delta function.
Kleinrock (1975, pp. 341–342); Poularikas and Seely (1985, pp. 689–692).

15.6 Modified Bessel Functions

Let
$I_0(z)$ = modified Bessel function of order zero
$I_1(z)$ = modified Bessel function of order one

These functions can be expressed as

$$I_0(z)=1+\frac{\frac{1}{4}z^2}{(1!)^2}+\frac{\left(\frac{1}{4}z^2\right)^2}{(2!)^2}+\frac{\left(\frac{1}{4}z^2\right)^3}{(3!)^2}+\frac{\left(\frac{1}{4}z^2\right)^4}{(4!)^2}+\cdots \tag{15.14}$$

$$I_1(z)=\frac{d}{dz}I_0(z) \tag{15.15}$$

$$I_1(z) = \frac{1}{2} z \left\{ 1 + \frac{\frac{1}{4}z^2}{1!2!} + \frac{\left(\frac{1}{4}z^2\right)^2}{2!3!} + \frac{\left(\frac{1}{4}z^2\right)^3}{3!4!} + \frac{\left(\frac{1}{4}z^2\right)^4}{4!5!} + \cdots \right\} \tag{15.16}$$

Abramowitz and Stegun (1968, pp. 375–376).

15.7 Stirling's Formula

Factorial approximation. For n a positive integer:

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n} \tag{15.17}$$

Abramowitz and Stegun (1968, p. 257); Binmore (1977, p. 156); Feller (1964, p. 50); Landsberg (1978, p. 435).

16

Calculus Results

16.1 Basic Rules for Differentiation

(a) Differentiation of a Sum of Functions

$$\frac{d}{dx}(u+v)=\frac{du}{dx}+\frac{dv}{dx} \tag{16.1}$$

where $u=u(x)$ and $v=v(x)$ are differentiable functions of x.

(b) Differentiation of a Function and a Constant Multiple

$$\frac{d}{dx}(ku)=k\frac{du}{dx} \tag{16.2}$$

where k is a constant, and $u=u(x)$ is a differentiable function of x.

(c) Differentiation of a Product of Two Functions (Product Rule)

$$\frac{d}{dx}(uv)=v\frac{du}{dx}+u\frac{dv}{dx} \tag{16.3}$$

where $u=u(x)$ and $v=v(x)$ are differentiable functions of x.

Note: This rule can be extended to the product of three or more functions, by successive applications of the rule for two functions. Thus, for the differentiation of a product of three functions,

$$\frac{d}{dx}(uvw)=vw\frac{du}{dx}+uw\frac{dv}{dx}+uv\frac{dw}{dx} \tag{16.4}$$

where $u=u(x)$, $v=v(x)$, and $w=w(x)$ are differentiable functions of x.

(d) Differentiation of a Quotient of Two Functions (Quotient Rule)

$$\frac{d}{dx}\left(\frac{u}{v}\right)=\frac{v\dfrac{du}{dx}-u\dfrac{dv}{dx}}{v^2} \tag{16.5}$$

where $u=u(x)$ and $v=v(x)$ are differentiable functions of x.

(e) Differentiation of a Function of a Function (Chain Rule)
If y is a function of x, and u is a function of y, where both functions are differentiable, then

$$\frac{du}{dx} = \frac{du}{dy}\frac{dy}{dx} \tag{16.6}$$

Thus, if $u=f(y)$ and $y=g(x)$, this rule gives

$$\frac{d}{dx} f\left(g(x)\right) = f'\left(g(x)\right)g'(x) \tag{16.7}$$

where f' and g' are the derivatives of f and g, respectively, i.e., where $f'(y) = \frac{d}{dy} f(y)$ and $g'(x) = \frac{d}{dx} g(x)$.

(f) Differentiation of an Inverse Function
If $y=f(x)$ is a differentiable function of x with a nonzero derivative $f'(x)$, and $x=f^{-1}(y)$ is the inverse function, then

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{f'(x)} \tag{16.8}$$

Adams (1999, pp. 110–122); Binmore (1983, pp. 43–44); Granville, Smith, and Longley (1957, pp. 28–39); Hardy (1963, pp. 216–220).

16.2 Integration by Parts

$$\int u\,dv = uv - \int v\,du \tag{16.9}$$

where
$u=u(x)$ and $v=v(x)$ are differentiable functions of x

$$du = \frac{du}{dx}dx$$

$$dv = \frac{dv}{dx}dx$$

Note: This result is derived from the product rule for differentiation, given by Equation 16.3.

Adams (1999, p. 345); Binmore (1983, p. 242); Hardy (1963, p. 258); Stewart (1995, p. 437); Trim (1983, p. 360).

16.3 Fundamental Theorem of Calculus

Let $f(x)$ be a continuous function in the interval $a \le x \le b$. The fundamental theorem of calculus consists of the following two statements concerning $f(x)$:

If $F(x)$ is a function defined by the integral $F(x) = \int_a^x f(t)dt$ for any x in the interval, then

$$\frac{d}{dx}F(x) = f(x) \tag{16.10}$$

If $G(x)$ is a function such that its derivative $\frac{d}{dx}G(x) = f(x)$ for any x in the interval, then

$$\int_a^b f(x)dx = G(b) - G(a) \tag{16.11}$$

Adams (1999, p. 324); Binmore (1983, pp. 230–231); Sokolnikoff (1939, p. 120); Stewart (1995, p. 291); Trim (1983, p. 202).

16.4 Taylor Series

The Taylor series for the function $f(x)$ is a series expansion about a given point a, given by

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f'''(a) + \cdots \tag{16.12}$$

where

$$f'(a) = \frac{d}{dx} f(x)\Big|_{x=a}$$

$$f''(a) = \frac{d^2}{dx^2} f(x)\Big|_{x=a}$$

$$f'''(a) = \frac{d^3}{dx^3} f(x)\Big|_{x=a}, \text{ etc.}$$

Equation 16.12 holds for values of x around the point a for which $f(x)$ has derivatives of all orders and the series satisfies convergence criteria.

Adams (1999, p. 566); Finney and Thomas (1990, p. 623); Gradshteyn and Ryzhik (1965, p. 15); Granville, Smith, and Longley (1957, p. 369); Hardy (1963, p. 291); Sokolnikoff (1939, p. 296); Stewart (1995, p. 654).

16.5 Maclaurin Series

The Maclaurin series for the function $f(x)$ is a special case of the above Taylor series with $a = 0$:

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f'''(0) + \cdots \tag{16.13}$$

where

$$f'(0) = \frac{d}{dx} f(x)\Big|_{x=0}$$

$$f''(0) = \frac{d^2}{dx^2} f(x)\Big|_{x=0}$$

$$f'''(0) = \frac{d^3}{dx^3} f(x)\Big|_{x=0}, \text{ etc.}$$

Equation 16.13 holds for values of x around the point $x=0$ for which $f(x)$ has derivatives of all orders and the series satisfies convergence criteria.

Adams (1999, p. 566); Finney and Thomas (1990, p. 623); Gradshteyn and Ryzhik (1965, p. 15); Granville, Smith, and Longley (1957, p. 357); Hardy (1963, p. 291); Sokolnikoff (1939, p. 296); Stewart (1995, p. 654).

16.6 L'Hôpital's Rule

Let $f(x)$ and $g(x)$ be functions that are differentiable in an open interval containing the point a, and have derivatives $f'(x)=\frac{d}{dx}f(x)$ and $g'(x)=\frac{d}{dx}g(x)$.

If

$$f(a)=g(a)=0 \tag{16.14}$$

and

$$\lim_{x\to a}\frac{f'(x)}{g'(x)}=L \tag{16.15}$$

then

$$\lim_{x\to a}\frac{f(x)}{g(x)}=L \tag{16.16}$$

Adams (1999, pp. 290, 292); Brand (1955, p. 119); Finney and Thomas (1990, p. 483); Sokolnikoff (1939, p. 54); Stewart (1995, p. 420); Trim (1983, pp. 157, 159).

Note: L'Hôpital's rule holds when the limit L is finite, or ∞, or $-\infty$, and when a is finite, or ∞, or $-\infty$. It also holds if $f(x)$ and $g(x)$ at $x=a$ are ∞ or $-\infty$, instead of 0 as given by Equation 16.14, i.e., if

$$\lim_{x\to a}f(x)=\pm\infty \tag{16.17}$$

and

$$\lim_{x \to a} g(x) = \pm\infty \tag{16.18}$$

Thus, l'Hôpital's rule applies when the limit of a ratio has an indeterminate form of the type $\frac{0}{0}$ or $\frac{\infty}{\infty}$. The rule states that the limit for these indeterminate forms may be found by taking the derivatives of the numerator and denominator (separately), and evaluating the limit of the resulting ratio. If the ratio of the derivatives approaches a limit, the original ratio approaches the same limit.

16.7 Lagrange Multipliers

The method of Lagrange multipliers is a calculus-based optimization technique that finds the stationary points (maxima, minima, etc.) of a function of several variables, when the variables are subject to constraints.

To find the stationary points of a function f of n variables given by

$$f = f(x_1, x_2, \ldots, x_n) \tag{16.19}$$

subject to the m constraints

$$\left.\begin{aligned} g_1(x_1, x_2, \ldots, x_n) &= 0 \\ g_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}\right\} \tag{16.20}$$

where the functions f and g_i ($i = 1, 2, \ldots, m$) have certain differentiability properties, the method of Lagrange multipliers consists of the following steps:

1. Introduce m new variables, called Lagrange multipliers

$$\lambda_1, \lambda_2, \ldots, \lambda_m \text{ (one for each constraint equation)}$$

2. Form the function L, called the Lagrangian and defined as

$$L = f(x_1, x_2, \ldots, x_n) + \lambda_1 g_1(x_1, x_2, \ldots, x_n) + \lambda_2 g_2(x_1, x_2, \ldots, x_n) + \cdots + \lambda_m g_m(x_1, x_2, \ldots, x_n) \tag{16.21}$$

3. Take the partial derivatives of the Lagrangian L with respect to each of the variables and solve the equations

$$\left.\begin{aligned} \frac{\partial L}{\partial x_i} &= 0 \qquad (i = 1, 2, \ldots, n) \\ \frac{\partial L}{\partial \lambda_j} &= 0 \qquad (j = 1, 2, \ldots, m) \end{aligned}\right\} \tag{16.22}$$

for $x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m$ ($n+m$ equations for $n+m$ unknowns).

Note: Since $\lambda_1, \lambda_2, \ldots, \lambda_m$ appear in L only as multipliers of $g_1, g_2, \ldots, g_m$, the m equations $\frac{\partial L}{\partial \lambda_j} = 0$ ($j=1, 2, \ldots, m$) are just the constraint equations given by Equation 16.20.

The solutions obtained for $x_1, x_2, \ldots, x_n$ from step 3 are the values of these variables at the stationary points of the function f.

Adams (1999, pp. 792–793); Binmore (1983, pp. 85–86); Sokolnikoff (1939, pp. 331–333).

16.8 Differentiation under the Integral Sign (Leibnitz's Rule)

Let

$$I(t) = \int_{\alpha(t)}^{\beta(t)} f(x, t)\, dx$$

where

$\alpha(t)$ and $\beta(t)$ are differentiable functions of t

$f(x, t)$ and $\frac{\partial}{\partial t} f(x, t)$ are continuous functions in the region of integration

Then

$$\frac{d}{dt}I(t)=\int_{\alpha(t)}^{\beta(t)}\frac{\partial}{\partial t}f(x,t)\,dx+f(\beta(t),t)\frac{d\beta}{dt}-f(\alpha(t),t)\frac{d\alpha}{dt} \tag{16.23}$$

Special cases:

$$\frac{d}{dt}\int_a^b f(x,t)dx=\int_a^b\frac{\partial}{\partial t}f(x,t)dx \tag{16.24}$$

$$\frac{d}{dt}\int_a^t f(x)dx=f(t) \tag{16.25}$$

$$\frac{d}{dt}\int_t^b f(x)dx=-f(t), \tag{16.26}$$

where a and b are constants, and where the integrand $f(x)$ in Equations 16.25 and 16.26 is a function of x only (and not t).

Adams (1999, p. 805); Binmore (1983, p. 255); Buck (1956, p. 73); Sokolnikoff (1939, p. 121); Trim (1983, p. 723).

Note: Equation 16.23 is a useful result in optimization, where the derivative with respect to a parameter is needed and the objective function may be an integral that is difficult or impossible to evaluate analytically. Such an integral can arise, for example, in the expected value of a variable over a probability distribution. An example of the use of Equation 16.23 is in the classic *newsboy problem* (see Section 9.3).

16.9 Change of a Variable in an Integral

Let $f(x)$ be a continuous function of x in the interval $a\le x\le b$. Given the integral $\int_a^b f(x)dx$, let

$$x=g(u) \tag{16.27}$$

be a continuous single-valued function with a continuous derivative $\frac{dx}{du} = g'(u)$ in the interval $c \le t \le d$, where c and d are given by

$$a = g(c) \tag{16.28}$$

and

$$b = g(d) \tag{16.29}$$

respectively. Then

$$\int_a^b f(x)\,dx = \int_c^d f(g(u))g'(u)\,du \tag{16.30}$$

Adams (1999, pp. 332, 352); Binmore (1983, pp. 239–240); Brand (1955, p. 270); Sokolnikoff (1939, p. 125); Trim (1983, p. 206).

Note: This transformation can substantially simplify the evaluation of an integral. The formula is derived from the chain rule for differentiation.

16.10 Change of Variables in a Double Integral

Given the integral $\iint_R f(x, y)\,dx\,dy$ over the region R in the xy-plane, let

$$x = g(u, v) \tag{16.31}$$

and

$$y = h(u, v) \tag{16.32}$$

be continuously differentiable functions that define a one-to-one transformation of the region R in the xy-plane to the region S in the uv-plane, and let

$$F(u, v) = f(g(u, v), h(u, v)) \tag{16.33}$$

In addition, let J be the Jacobian, given by the determinant

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \tag{16.34}$$

i.e.,

$$J = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \tag{16.35}$$

and let $|J|$ denote the absolute value of J. If the Jacobian J is nonzero, then

$$\iint_R f(x, y)dx\,dy = \iint_S F(u, v)|J|du\,dv \tag{16.36}$$

Adams (1999, p. 844); Binmore (1983, p. 293); Brand (1955, p. 364); Sokolnikoff (1939, p. 151); Stewart (1995, p. 883).

Note: The determinants $\frac{\partial(x, y)}{\partial(u, v)}$ and $\frac{\partial(u, v)}{\partial(x, y)}$ are related by

$$\frac{\partial(x, y)}{\partial(u, v)} = \frac{1}{\frac{\partial(u, v)}{\partial(x, y)}} \tag{16.37}$$

where

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} \tag{16.38}$$

Adams (1999, pp. 758, 843); Binmore (1983, p. 293); Brand (1955, p. 176).

Equation 16.37 is useful for deriving J in the cases where it is easier to evaluate the partial derivatives of u and v with respect to x and y than the partial derivatives of x and y with respect to u and v.

The change of variables formula given by Equation 16.36 can be extended to triple and higher-order multiple integrals. Thus, for a triple integral $\iiint_R f(x, y, z)dx\,dy\,dz$ over the region R in xyz-space, the formula becomes

$$\iiint_R f(x, y, z)dx\,dy\,dz = \iiint_S F(u, v, w)|J|\,du\,dv\,dw \tag{16.39}$$

where $x=x(u, v, w)$, $y=y(u, v, w)$, and $z=z(u, v, w)$ are continuously differentiable functions that define a one-to-one transformation of the region R in xyz-space to the region S in uvw-space, where the function $F(u, v, w)$ is given by

$$F(u,v,w) = f(x(u,v,w),\ y(u,v,w),\ z(u,v,w)) \tag{16.40}$$

and where the Jacobian J is given by the determinant

$$J = \frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} \tag{16.41}$$

and is nonzero.

Adams (1999, p. 855); Binmore (1983, p. 293); Brand (1955, p. 383); Sokolnikoff (1939, p. 157); Stewart (1995, p. 886).

16.11 Changing the Order of Integration in a Double Integral

For a double integral over a triangular region in the xy-plane bounded by the x-axis, line $y=x$, and line $x=a$ (see Figures 16.1 and 16.2),

$$\int_{y=0}^{a}\int_{x=y}^{a} f(x,y)dx\,dy = \int_{x=0}^{a}\int_{y=0}^{x} f(x,y)dy\,dx \tag{16.42}$$

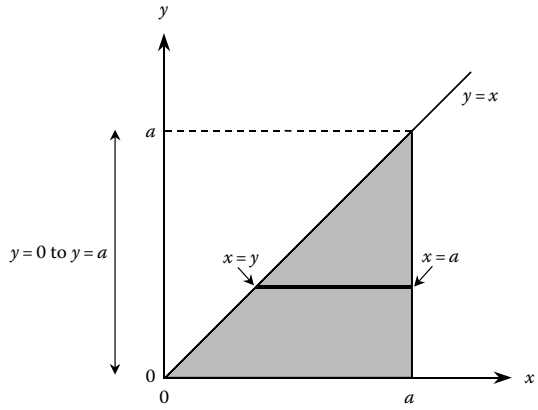


FIGURE 16.1
Limits of integration for the left-hand side of Equation 16.42 (integration in x-direction first).

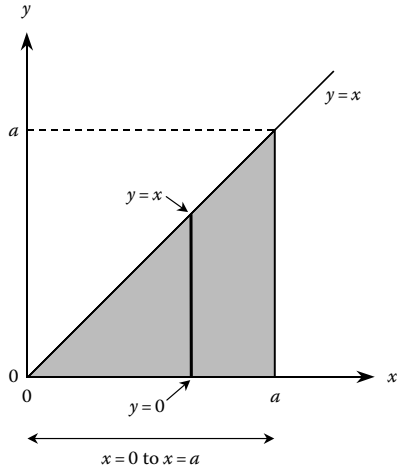


FIGURE 16.2
Limits of integration for the right-hand side of Equation 16.42 (integration in y-direction first).

Extension for the case $a \to \infty$: if $f(x, y) \geq 0$ over the region of integration, then

$$\int_{y=0}^{\infty} \int_{x=y}^{\infty} f(x, y)\,dx\,dy = \int_{x=0}^{\infty} \int_{y=0}^{x} f(x, y)\,dy\,dx \tag{16.43}$$

provided the integrals converge.

Adams (1999, p. 828); Binmore (1983, pp. 282, 298); Brand (1955, p. 354); Finney and Thomas (1990, p. 899); Trim (1983, p. 672).

Note: Changing the order of integration can substantially simplify the evaluation of a double integral. The limits of integration depend in general on the shape of the region of integration. The above results are given for the case of a triangular region, since it occurs commonly in double integrals and the limits of integration are easily specified.

16.12 Changing the Order of Summation in a Double Sum

If $a_{ij} \geq 0$ for all i and j ($i = 0, 1, 2, \ldots\ldots; j = 0, 1, 2, \ldots\ldots$), then

$$\sum_{j=0}^{\infty}\sum_{i=j}^{\infty} a_{ij} = \sum_{i=0}^{\infty}\sum_{j=0}^{i} a_{ij} \tag{16.44}$$

provided the sums converge (see Figures 16.3 and 16.4).

Binmore (1983, p. 300); Clarke and Disney (1985, p. 312).

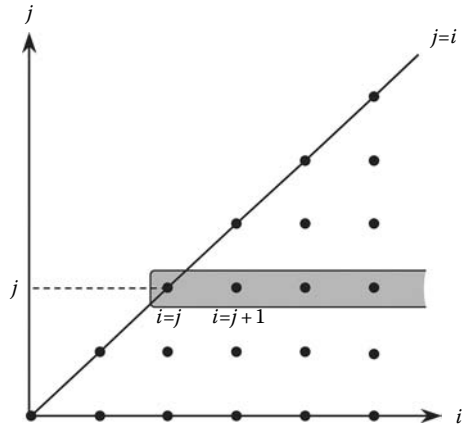


FIGURE 16.3
First summation range for the left-hand side of Equation 16.44.

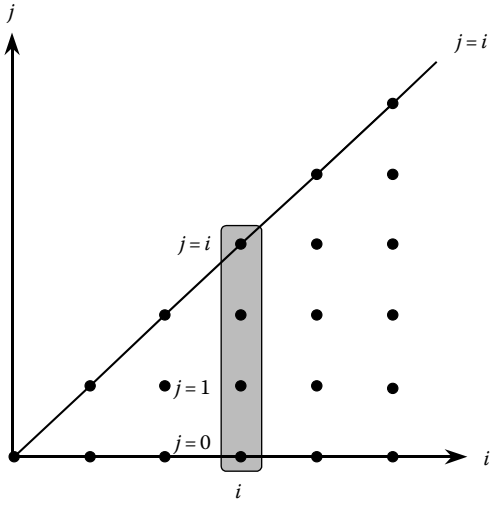


FIGURE 16.4
First summation range for the right-hand side of Equation 16.44.

16.13 Numerical Integration

The integral $\int_a^b f(x)dx$ can be evaluated numerically using several different methods. The following rules provide three standard methods of numerical approximation: midpoint rule, trapezoidal rule, and Simpson's rule.

For each rule, let the interval $a \le x \le b$ be divided into n equal subintervals. Let h be the width of each subinterval, given by

$$h = \frac{b-a}{n} \tag{16.45}$$

Let $x_0, x_1, x_2, \ldots, x_n$ be equally spaced points on the interval, given by

$$\left.\begin{aligned} x_0 &= a \\ x_1 &= a + h \\ x_2 &= a + 2h \\ &\vdots \\ x_{n-1} &= a + (n-1)h \\ x_n &= b \end{aligned}\right\} \tag{16.46}$$

In addition, for the midpoint rule, let $m_1, m_2, \ldots, m_n$ denote the midpoints of each subinterval, given by

$$\left.\begin{aligned} m_1 &= \frac{x_0 + x_1}{2} \\ m_2 &= \frac{x_1 + x_2}{2} \\ &\vdots \\ m_n &= \frac{x_{n-1} + x_n}{2} \end{aligned}\right\} \tag{16.47}$$

Figures 16.5 through 16.7 illustrate the three methods of approximation. The midpoint rule is based on step function approximations to the curve $f(x)$ on the subintervals, and the area under the curve is thus approximated by rectangles (Figure 16.5). The trapezoidal rule is based on linear approximations, and the area under the curve is thus approximated by trapezoids (Figure 16.6). Simpson's rule is based on quadratic function approximations, and the area under the curve is thus approximated by the areas under segments of parabolas (Figure 16.7).

The three rules give the following approximate formulas for numerical evaluation of the integral $\int_a^b f(x)dx$:

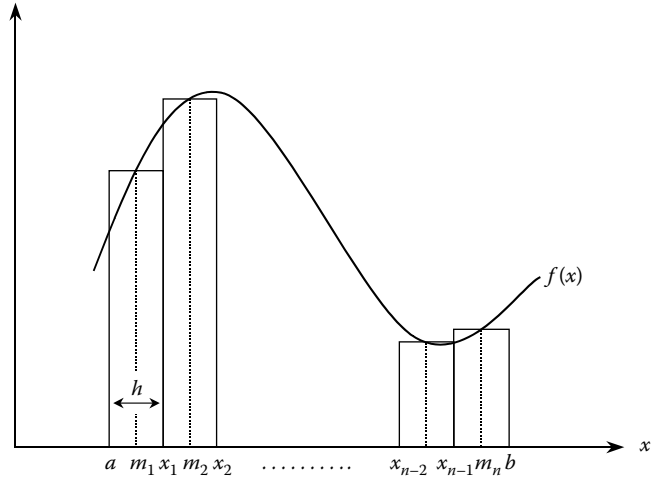


FIGURE 16.5
Midpoint rule approximation for numerical integration.

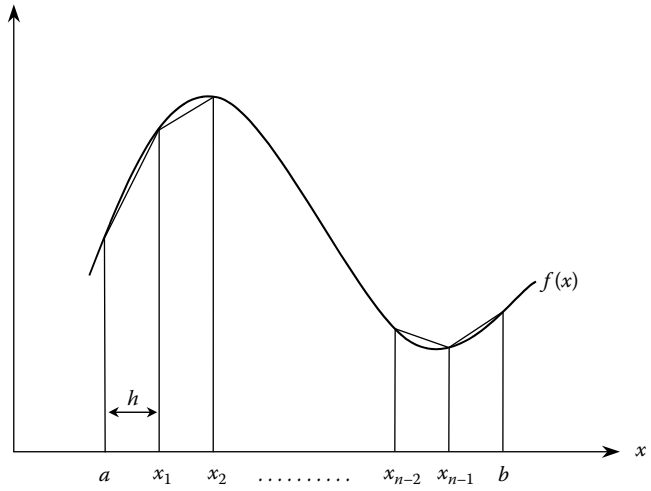


FIGURE 16.6
Trapezoidal rule approximation for numerical integration.

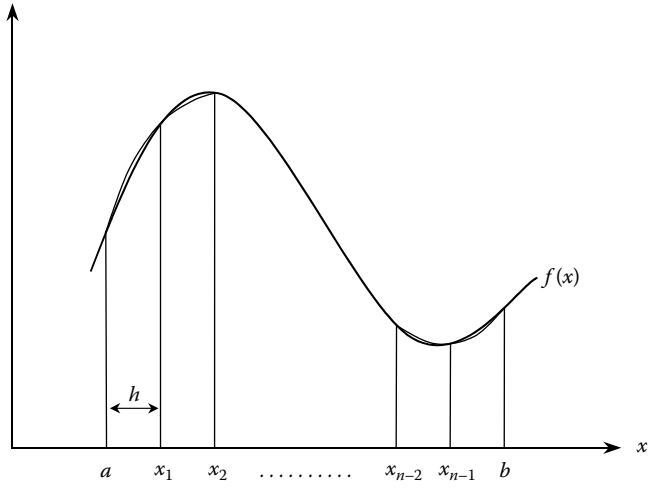


FIGURE 16.7
Simpson's rule approximation for numerical integration.

Midpoint rule

$$\int_a^b f(x)dx \cong h\left[f(m_1)+f(m_2)+\cdots+f(m_n)\right] \tag{16.48}$$

Trapezoidal rule

$$\int_a^b f(x)dx \cong \frac{h}{2}\left[f(a)+2f(x_1)+2f(x_2)+\cdots+2f(x_{n-1})+f(b)\right] \tag{16.49}$$

Simpson's rule

$$\int_a^b f(x)dx \cong \frac{h}{3}\left[f(a)+4f(x_1)+2f(x_2)+4f(x_3)+\cdots+2f(x_{n-2})+4f(x_{n-1})+f(b)\right]$$

(n is an even number for Simpson's rule)

(16.50)

For each rule, the error (difference between the numerical approximation and the exact value of the integral) can be estimated within given bounds. The error bounds are given for the midpoint and trapezoidal rules when $f(x)$ has a continuous second derivative $f''(x)$ on the interval $a \le x \le b$, and for Simpson's rule when $f(x)$ has a continuous fourth derivative $f^{(4)}(x)$ on the interval $a \le x \le b$. Estimates of the error bounds for the three rules are given by

$$\text{Midpoint rule: } |\text{Error}| \le \frac{M(b-a)^3}{24n^2} \tag{16.51}$$

$$\text{Trapezoidal rule: } |\text{Error}| \le \frac{M(b-a)^3}{12n^2} \tag{16.52}$$

$$\text{Simpson's rule: } |\text{Error}| \le \frac{N(b-a)^5}{180n^4} \tag{16.53}$$

where

M = maximum value of $|f''(x)|$ on the interval $a \le x \le b$
N = maximum value of $|f^{(4)}(x)|$ on the interval $a \le x \le b$
|| denotes absolute value

Adams (1999, pp. 384–392); Finney and Thomas (1990, pp. 337–343); Press, Teukolsky, Vetterling, and Flannery (2007, p. 156); Stewart (1995, pp. 478–484); Trim (1983, pp. 381–385).

The formulas for the three rules are derived for a single interval (double interval for Simpson's rule), and are extended to the case of n subintervals. The term *extended* (or *composite*) is therefore sometimes used to describe

these rules for the case of general n. Thus, Simpson's rule for general n, as given above, is sometimes referred to as the extended Simpson's rule (or composite Simpson's rule), and similarly for the other rules.

For the midpoint and trapezoidal rules, n may be an odd or even number. For Simpson's rule, n must be an even number. The error bounds for each rule give estimates of the theoretical errors as a function of n. Note that, if n is large, there may also be round-off errors that accumulate in the numerical computations.

The formulas for trapezoidal rule and Simpson's rule are examples of a *closed* formula, since they use values of $f(x)$ in the closed interval $[a, b]$, i.e., including the values at the endpoints a and b. The formula for the midpoint rule is an example of an *open* formula, since it only uses values of $f(x)$ in the open interval (a, b), i.e., not at the endpoints a and b. An open formula is useful in cases where $f(x)$ cannot be readily evaluated at the endpoints Press, Teukolsky, Vetterling, and Flannery (2007, p. 157).

17

Matrices

17.1 Rules for Matrix Calculations

Let
A, **B**, and **C** be matrices
λ be a scalar
$\mathbf{A}^{\mathrm{T}}$ be the transpose of **A**
$\mathbf{A}^{-1}$ be the inverse of **A**

Then the following identities hold:

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \tag{17.1}$$

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C}) \tag{17.2}$$

$$(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC}) \tag{17.3}$$

$$\mathbf{C}(\mathbf{A}+\mathbf{B})=\mathbf{CA}+\mathbf{CB} \tag{17.4}$$

$$\lambda(\mathbf{A}+\mathbf{B})=\lambda\mathbf{A}+\lambda\mathbf{B} \tag{17.5}$$

$$(\mathbf{A}^{\mathrm{T}})^{\mathrm{T}}=\mathbf{A} \tag{17.6}$$

$$(\mathbf{A}+\mathbf{B})^{\mathrm{T}}=\mathbf{A}^{\mathrm{T}}+\mathbf{B}^{\mathrm{T}} \tag{17.7}$$

$$(\mathbf{AB})^{\mathrm{T}}=\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} \tag{17.8}$$

$$(\mathbf{ABC})^{\mathrm{T}}=\mathbf{C}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} \tag{17.9}$$

$$(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{17.10}$$

$$(\mathbf{ABC})^{-1}=\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{17.11}$$

$$(\mathbf{A}^{-1})^{-1}=\mathbf{A} \tag{17.12}$$

$$(\mathbf{A}^{\mathrm{T}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{T}} \tag{17.13}$$

Neter, Wasserman, and Kutner (1985, pp. 204–205); Press (1982, pp. 18, 22).

17.2 Inverses of Matrices

17.2.1 Inverse of 2×2 Matrix

If $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$\mathbf{A}^{-1} = \begin{pmatrix} \dfrac{d}{D} & -\dfrac{b}{D} \\ -\dfrac{c}{D} & \dfrac{a}{D} \end{pmatrix} \tag{17.14}$$

where D = determinant of $\mathbf{A}$, given by

$$D = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

17.2.2 Inverse of 3×3 Matrix

If $\mathbf{B} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix}$ then

$$\mathbf{B}^{-1} = \begin{pmatrix} (ek - fh)/Z & -(bk - ch)/Z & (bf - ce)/Z \\ -(dk - fg)/Z & (ak - cg)/Z & -(af - cd)/Z \\ (dh - eg)/Z & -(ah - bg)/Z & (ae - bd)/Z \end{pmatrix} \tag{17.15}$$

where Z = determinant of $\mathbf{B}$, given by

$$Z = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & k \end{vmatrix} = a(ek - fh) - b(dk - fg) + c(dh - eg)$$

Neter, Wasserman, and Kutner (1985, p. 202).

17.3 Series of Matrices

If $\mathbf{A}^n$ tends to the zero matrix as n tends to infinity, then $(\mathbf{I}-\mathbf{A})$ has an inverse, given by

$$(\mathbf{I}-\mathbf{A})^{-1} = \mathbf{I}+\mathbf{A}+\mathbf{A}^2+\mathbf{A}^3+\cdots = \sum_{k=0}^{\infty}\mathbf{A}^k \tag{17.16}$$

where $\mathbf{I}$ is the identity (unity) matrix.

Kemeny and Snell (1976, p. 22).

17.4 Derivatives of Matrices

If $\mathbf{A}$ is a matrix with elements that are differentiable functions of the scalar variable x, and has an inverse matrix $\mathbf{A}^{-1}$ for any x, then the derivative of the inverse matrix is given by

$$\frac{d\mathbf{A}^{-1}}{dx} = -\mathbf{A}^{-1}\frac{d\mathbf{A}}{dx}\mathbf{A}^{-1} \tag{17.17}$$

Equation 17.17 is obtained by differentiating the identify $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Fröberg (1965, p. 71).

18

Combinatorics

Terminology:

Permutations: Selections when order matters.

Combinations: Selections when order does not matter.

Without replacement: Once selected, an object cannot be selected again.

With replacement: Object can be selected any number of times.

Let

P_N^M be the number of permutations of N objects selected from a total of M objects

C_N^M be the number of combinations of N objects selected from a total of M objects

1. Permutations without replacement:

$$P_N^M = \frac{M!}{(M-N)!} \quad (N \le M) \tag{18.1}$$

2. Permutations with replacement:

$$P_N^M = M^N \tag{18.2}$$

3. Combinations without replacement:

$$C_N^M = \binom{M}{N} = \frac{M!}{N!(M-N)!} \quad (N \le M) \tag{18.3}$$

4. Combinations with replacement:

$$C_N^M = \binom{M+N-1}{N} = \frac{(M+N-1)!}{N!(M-1)!} \tag{18.4}$$

5. Circular permutations (permutations of objects forming a circle) without replacement:

$$P_N^M = \frac{M!}{N(M-N)!} \quad (N \le M) \tag{18.5}$$

6. Combinations with replacement that contain at least one of each object type:

$$C_N^M = \binom{N-1}{M-1} = \frac{(N-1)!}{(N-M)!(M-1)!} \quad (N \ge M) \tag{18.6}$$

7. Permutations of M objects, consisting of M_1 objects of type 1, M_2 objects of type 2, up to k types, where $M_1+M_2+\cdots+M_k = M$

$$P_{\{M_k\}}^M = \frac{M!}{M_1!M_2!\cdots M_k!} \tag{18.7}$$

(Equation 18.7 gives multinomial coefficients—see Section 3.7.)

Brualdi (1977, pp. 29, 31–32, 34–35, 37–38); Feller (1964, pp. 28, 33, 35–37).

19

Summations

19.1 Finite Sums

$$1+2+3+\cdots+n=\frac{n(n+1)}{2} \tag{19.1}$$

$$1^2+2^2+3^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6} \tag{19.2}$$

$$1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2=\frac{n^2(n+1)^2}{4} \tag{19.3}$$

$$1+3+5+\cdots+(2n-1)=n^2 \tag{19.4}$$

$$a+ar+ar^2+ar^3+\cdots+ar^n=\frac{a(1-r^{n+1})}{1-r} \quad (r\neq 1) \tag{19.5}$$

19.2 Infinite Sums

$$a+ar+ar^2+ar^3+\cdots=\frac{a}{1-r} \quad (|r|<1) \tag{19.6}$$

$$1ar+2ar^2+3ar^3+\cdots=\frac{ar}{(1-r)^2} \quad (|r|<1) \tag{19.7}$$

$$1^2ar+2^2ar^2+3^2ar^3+\cdots=\frac{ar(1+r)}{(1-r)^3} \quad (|r|<1) \tag{19.8}$$

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots=\frac{\pi^2}{6} \tag{19.9}$$

$$\frac{1}{1^2}-\frac{1}{2^2}+\frac{1}{3^2}-\cdots=\frac{\pi^2}{12} \tag{19.10}$$

$$\lim_{n\to\infty}\left\{1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\ln(n)\right\}=\gamma \tag{19.11}$$

where $\gamma\cong 0.577216$ is Euler's constant.

Abramowitz and Stegun (1968, pp. 255, 807–808); Gradshteyn and Ryzhik (1965, pp. 1, 7); Zwillinger (1996, pp. 15, 21, 23, 80).

20

Interest Formulas

Let
i=interest rate per period
n=number of interest (payment) periods
P=present sum of money (present worth)
F=sum of money at the end of n periods from the present date that is equivalent to P at interest rate i (F=*future worth*)
A=uniform end-of-period payment continuing for n periods that in total is equivalent to P at interest rate i (A=*annuity*)

Then

$$F = P(1+i)^n \tag{20.1}$$

$$A = P\left(\frac{i(1+i)^n}{(1+i)^n - 1}\right) \tag{20.2}$$

Let
E_k=portion of A in period k paid against principal (E_k=*equity payment*)
I_k=portion of A in period k paid as interest (I_k=*interest payment*)

Then

$$E_k = A\left(\frac{1}{(1+i)^{n-k+1}}\right) \tag{20.3}$$

$$I_k = A\left(1 - \frac{1}{(1+i)^{n-k+1}}\right) \tag{20.4}$$

Grant, Ireson, and Leavenworth (1982, p. 33); White, Agee, and Case (1977, pp. 65, 92, 93).

References

Abramowitz, M. and Stegun, I. A. (1968). *Handbook of Mathematical Functions*. Dover, New York.

Adams, R. A. (1999). *Calculus—A Complete Course* (4th edition). Addison-Wesley, Reading, MA.

Ahsanullah, M. and Nevzorov, V. B. (2005). *Order Statistics: Examples and Exercises*. Nova Science Publishers, New York.

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows*. Prentice-Hall, Englewood Cliffs, NJ.

Alden, J. M. (2002). Estimating performance of two workstations in series with downtime and unequal speeds. General Motors R&D Center, Research Publication R&D 9434.

Allen, A. O. (1978). *Probability, Statistics, and Queueing Theory—With Computer Science Applications*. Academic Press, Orlando, FL.

Allsop, R. E. (1972). Delay at a fixed time traffic signal—I: Theoretical analysis. *Transportation Science*, **6**(3), 260–285.

Allsop, R. E., Ali, A. T., Crosta, D. A., and Heydecker, B. G. (2005). Video vehicle detection at signalised junctions—A simulation-based study. *Proceedings of TRL Conference MOVA and Adaptive Traffic Signal Control*, Birmingham, U.K.

Apsimon, H. G. (1958). Mathematical note 2754—A repeated integral. *The Mathematical Gazette*, **42**(339), 52.

Arrow, K. J., Karlin, S., and Scarf, H. (1958). *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA.

Ashton, C. H. and Marsh, W. R. (1902). *Plane and Spherical Trigonometry*. Charles Scribner's Sons, New York.

Askin, R. G. and Standridge, C. R. (1993). *Modeling and Analysis of Manufacturing Systems*. Wiley, New York.

Ashworth, R. (1969). The capacity of priority-type intersections with a non-uniform distribution of critical gaps. *Transportation Research*, **3**(2), 272–279.

Ayres, F. (1954). *Schaum's Outline of Theory and Problems of Plane and Spherical Trigonometry*. McGraw-Hill, New York.

Ayyub, B. M. and McCuen, R. H. (1997). *Probability, Statistics, and Reliability for Engineers*. CRC Press, Boca Raton, FL.

Baker, K. R. (1992). Tightly-coupled production systems: Models, analysis, and insights. *Journal of Manufacturing Systems*, **11**(6), 385–400.

Balakrishnan, N. and Sinha, B. K. (1995). Order statistics and their properties. In *The Exponential Distribution: Theory, Methods and Applications*, Balakrishnan, N. and Basu, A. P., Eds. Gordon & Breach, Amsterdam, the Netherlands, Chapter 3.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. *Proceedings of the Cambridge Philosophical Society*, **55**(4), 299–327.

Benjamin, J. R. and Cornell, C. A. (1970). *Probability, Statistics, and Decision for Civil Engineers*. McGraw-Hill, New York.

Biernacki, A. (2006). VoIP source model based on the hyperexponential distribution. In *Proceedings of the World Academy of Science, Engineering and Technology*, Helsinki, Finland, February 2006, Vol. **11**, pp. 202–206.

Binmore, K. G. (1977). *Mathematical Analysis*. Cambridge University Press, Cambridge, U.K.

Binmore, K. G. (1983). *Calculus*. Cambridge University Press, Cambridge, U.K.

Blum, C. and Dorigo, M. (2004). The hyper-cube framework for ant colony optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, **34**(2), 1161–1172.

Blumenfeld, D. E. (1990). A simple formula for estimating throughput of serial production lines with variable processing times and limited buffer capacity. *International Journal of Production Research*, **28**(6), 1163–1182.

Blumenfeld, D. E. and Li, J. (2005). An analytical formula for throughput of a production line with identical stations and random failures. *Mathematical Problems in Engineering*, **2005**(3), 293–308.

Blumenfeld, D. E., Burns, L. D., Diltz, J. D., and Daganzo, C. F. (1985). Analyzing trade-offs between transportation, inventory and production costs on freight networks. *Transportation Research*, **19B**(5), 361–380.

Blumenfeld, D. E. and Weiss, G. H. (1970a). On the robustness of certain assumptions in the merging delay problem. *Transportation Research*, **4**(2), 125–139.

Blumenfeld, D. E. and Weiss, G. H. (1970b). On queue splitting to reduce waiting times. *Transportation Research*, **4**(2), 141–144.

Bolch, G., Greiner, S., de Meer, H., and Trivedi, K. S. (1998). *Queueing Networks and Markov Chains*. Wiley, New York.

Borel, E. (1925). *Traité du Calcul des Probabilitiés et de ses Applications*. Vol. 1, Part 1: *Principes et Formules Classiques du Calcul des Probabilitiés*. Gauthier-Villars, Paris, France.

Bradley, S. P., Hax, A. C., and Magnanti, T. L. (1977). *Applied Mathematical Programming*. Addison-Wesley, Reading, MA.

Brand, L. (1955). *Advanced Calculus—An Introduction to Classical Analysis*. Wiley, New York.

Brown, R. G. (1977). *Materials Management Systems*. Wiley, New York.

Brualdi, R. A. (1977). *Introductory Combinatorics*. North-Holland, New York.

Buck, R. C. (1956). *Advanced Calculus*. McGraw-Hill, New York.

Burns, L. D., Hall, R. W., Blumenfeld, D. E., and Daganzo, C. F. (1985). Distribution strategies that minimize transportation and inventory costs. *Operations Research*, **33**(3), 469–490.

Buzacott, J. A. (1968). Prediction of the efficiency of production systems without internal storage. *International Journal of Production Research*, **6**(3), 173–188.

Buzacott, J. A. and Shanthikumar, J. G. (1993). *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Cassady, C. R. and Nachlas, J. A. (2009). *Probability Models in Operations Research.* CRC Press, Boca Raton, FL.

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, **45**(1), 41–51.

Chauvenet, W. (1875). *A Treatise on Plane and Spherical Trigonometry* (9th edition). Lippincott, Philadelphia, PA.

Choong, Y. F. and Gershwin, S. B. (1987). A decomposition method for the approximate evaluation of capacitated transfer lines with unreliable machines and random processing times. *IIE Transactions*, **19**(2), 150–159 (Institute of Industrial Engineers).

Christofides, N. and Eilon, S. (1969). Expected distances in distribution problems. *Operational Research Quarterly*, **20**(4), 437–443.

Chu, P. C. and Beasley, J. E. (1997). A genetic algorithm for the generalised assignment problem. *Computers and Operations Research*, **24**(1), 17–23.

Chvátal, V. (1983). *Linear Programming*. W. H. Freeman, New York.

Çinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ.

Clark, A. R. (2003). Hybrid heuristics for planning lot setups and sizes. *Computers & Industrial Engineering*, **45**(4), 545–562.

Clark, C. E. (1961). The greatest of a finite set of random variables. *Operations Research*, **9**(2), 145–162.

Clarke, A. B. and Disney, R. L. (1985). *Probability and Random Processes: A First Course with Applications*. Wiley, New York.

Cohen, S. S. (1985). *Operational Research*. Edward Arnold, London, U.K.

Consoli, S., Darby-Dowman, K., Mladenović, N., and Pérez, J. A. (2009). Greedy randomized adaptive search and variable neighbourhood search for the minimum labelling spanning tree problem. *European Journal of Operational Research*, **196**(2), 440–449.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001). *Introduction to Algorithms* (2nd edition). MIT Press, Cambridge, MA.

Cox, D. R. (1962). *Renewal Theory*. Methuen, London, U.K.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London, U.K.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. Methuen, London, U.K.

Cox, D. R. and Smith, W. L. (1971). *Queues*. Chapman & Hall, London, U.K.

Daganzo, C. F. (1984). The length of tours in zones of different shapes. *Transportation Research*, **18B**(2), 135–145.

Daganzo, C. F. (1991). Logistics systems analysis. *Lecture Notes in Economics and Mathematical Systems*, Vol. 361. Springer-Verlag, New York.

Daganzo, C. F. (1997). *Fundamentals of Transportation and Traffic Operations*. Elsevier Science, New York.

Dantzig, G. B. (1951a). Maximization of a linear function of variables subject to linear inequalities. In *Activity Analysis of Production and Allocation*, Koopmans, T. C., Ed. Wiley, New York, Chapter XXI, pp. 339–347.

Dantzig, G. B. (1951b). Application of the simplex method to a transportation problem. In *Activity Analysis of Production and Allocation*, Koopmans, T. C., Ed. Wiley, New York, Chapter XXIII, pp. 359–373.

Dantzig, G. B. (1963). *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ.

David, H. A. (1970). *Order Statistics*. Wiley, New York.

Daskin, M. S. (1995). *Network and Discrete Location: Models, Algorithms, and Applications*. Wiley, New York.

DeGroot, M. H. (1986). *Probability and Statistics* (2nd edition). Addison-Wesley, Reading, MA.

Devore, J. L. (2008). *Probability and Statistics for Engineering and the Sciences* (7th edition). Thomson Brooks/Cole, Belmont, CA.

Dorigo, M., Maniezzo, V., and Colorni, A. (1996). The ant system: Optimization by a colony of cooperating ants. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, **26**(1), 29–41.

Dowsland, K. A. (1993). Simulated annealing. In *Modern Heuristic Techniques for Combinatorial Problems*, Reeves, C. R., Ed. Wiley, New York, Chapter 2.

Drake, A. W. (1967). *Fundamentals of Applied Probability Theory*. McGraw-Hill, New York.

Duque-Antón, M. (1997). Constructing efficient simulated annealing algorithms. *Discrete Applied Mathematics*, **77**(2), 139–159.

Ebrahimi, N., Soofi, E. S. and Zahedi, H. (2004). Information properties of order statistics and spacings. *IEEE Transactions on Information Theory*, **50**(1), 177–173.

Eilon, S., Watson-Gandy, C. D. T., and Christofides, N. (1971). *Distribution Management: Mathematical Modelling and Practical Analysis*. Hafner, New York.

Elsayed, E. A. and Boucher, T. O. (1985). *Analysis and Control of Production Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Evans, D. H., Herman, R., and Weiss, G. H. (1964). Comments on the highway crossing problem. *Operations Research*, **12**(6), 832–857.

Evans, J. R., Anderson, D. R., Sweeney, D. J., and Williams, T. A. (1984). *Applied Production and Operations Management* (2nd edition). West Publishing Company, St. Paul, MN.

Evans, M., Hastings, N., and Peacock, B. (2000). *Statistical Distributions* (3rd edition). Wiley, New York.

Fairthorne, D. (1965). The distances between pairs of points in towns of simple geometric shapes. *Proceedings of the Second Symposium on the Theory of Traffic Flow* (London, 1963), Almond, J., Ed., Organization for Economic Cooperation and Development (OECD), Paris, France, 1965, pp. 391–406.

Feller, W. (1964). *An Introduction to Probability Theory and its Applications—Volume 1*. Wiley, New York.

Feo, T. A. and Resende, M. G. C. (1995). Greedy Randomized Adaptive Search Procedures. *Journal of Global Optimization*, **6**(2), 109–133.

Finney, R. L. and Thomas, G. B. (1990). *Calculus*. Addison-Wesley, Reading, MA.

Fortuin, L. (1980). Five popular probability density functions: A comparison in the field of stock-control models. *Journal of the Operational Research Society*, **31**(10), 937–942.

Freund, J. E. (1992). *Mathematical Statistics* (5th edition). Prentice-Hall, Englewood Cliffs, NJ.

Fröberg, C.-E. (1965). *Introduction to Numerical Analysis*. Addison-Wesley, Reading, MA.

Garfinkel, R. S. and Nemhauser, G. L. (1972). *Integer Programming*. Wiley, New York.

Garwood, F. (1947). The variance of the overlap of geometrical figures with reference to a bombing problem. *Biometrika*, **34**(1&2), 1–17.

Garwood, F. and Tanner, J. C. (1958). Mathematical note 2800: On note 2754—A repeated integral. *The Mathematical Gazette*, **42**(342), 292–293.

Gass, S. I. (1964). *Linear Programming—Methods and Applications* (2nd edition). McGraw-Hill, New York.

Gershwin, S. B. (1994). *Manufacturing Systems Engineering*. PTR Prentice-Hall, Englewood Cliffs, NJ.

Gillett, B. E. (1976). *Introduction to Operations Research: A Computer-Oriented Algorithmic Approach*. McGraw-Hill, New York.

Gilli, M. and Winker, P. (2008). A review of heuristic optimization methods in econometrics. Swiss Finance Institute Research Paper No. 08-12.

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, **15**(5), 533–549.

Glover, F. and Laguna, M. (1993). Tabu search. In *Modern Heuristic Techniques for Combinatorial Problems*, Reeves, C. R., Ed. Wiley, New York, Chapter 3.

Glover, F., Tallard, E., and de Werra, D. (1993). A user's guide to tabu search. *Annals of Operations Research*, **41**(1), 3–28.

Goldman, A. S. and Slattery, T. B. (1964). *Maintainability: A Major Element of System Effectiveness*. Wiley, New York.

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Tables of Integrals, Series, and Products*. Academic Press, New York.

Grant, E. L., Ireson, W. G., and Leavenworth, R. S. (1982). *Principles of Engineering Economy*. Wiley, New York.

Granville, W. A., Smith, P. F., and Longley, W. R. (1957). *Elements of Differential and Integral Calculus*. Ginn & Company, Boston, MA.

Grimmett, G. R. and Stirzaker, D. R. (2001). *Probability and Random Processes*. Oxford University Press, Oxford, U.K.

Gross, D. and Harris, C. M. (1998). *Fundamentals of Queueing Theory* (3rd edition). Wiley, New York.

Hadley, G. and Whitin, T. M. (1963). *Analysis of Inventory Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Haight, F. A. (1963). *Mathematical Theories of Traffic Flow*. Academic Press, New York.

Haight, F. A. (1981). *Applied Probability*. Plenum Press, New York.

Hall, F. L. (1992). Traffic stream characteristics. In *Traffic Flow Theory: A State-of-the-Art Report* (revised monograph), Gartner, N. H., Rathi, A. J., and C.J. Messer, C. J., Eds. Turner-Fairbank Highway Research Center, Federal Highway Administration, U.S. Department of Transportation, McLean, VA, Chapter 2 (http://www.tfhrc.gov/its/tft/tft.htm).

Hall, R. W. (1991). *Queueing Methods—For Services and Manufacturing*. Prentice-Hall, Englewood Cliffs, NJ.

Hall, R. W. (1996). On the integration of production and distribution: Economic order and production quantity implications. *Transportation Research*, **30B**(5), 387–403.

Hansen, P. and Mladenović, N. (2001). Variable neighborhood search: Principles and applications. *European Journal of Operational Research*, **130**(3), 449–467.

Hanssmann, F. (1962). *Operations Research in Production and Inventory Control*. Wiley, New York.

Hardy, G. H. (1963). *A Course of Pure Mathematics* (10th edition). Cambridge University Press, Cambridge, U.K.

Harris, F. W. (1913). How many parts to make at once. *Factory: The Magazine of Management*, **10**(2), 135–136, 152. (Reprinted 1990 in *Operations Research*, **38**(6), 947–950.)

Hastings, C. (1955). *Approximations for Digital Computers*. Princeton University Press, Princeton, NJ.

Hax, C. and Candea, D. (1984). *Production and Inventory Management*. Prentice-Hall, Englewood Cliffs, NJ.

Herman, R. and Weiss, G. H. (1961). Comments on the highway crossing problem. *Operations Research*, **9**(6), 828–840.

Heyman, D. P. and Sobel, M. J. (1982). *Stochastic Models in Operations Research, Volume 1*: *Stochastic Processes and Operating Characteristics*. McGraw-Hill, New York.

Hillier, F. S. and Boling, R. W. (1966). The effect of some design factors on the efficiency of production lines with variable operation times. *Journal of Industrial Engineering*, **17**(12), 651–658.

Hillier, F. S. and Lieberman, G. J. (1980). *Introduction to Operations Research* (3rd edition). Holden-Day, San Francisco, CA.

Hoel, P. G., Port, S. C., and Stone, C. J. (1971). *Introduction to Probability Theory*. Houghton Mifflin, Boston, MA.

Homburger, S. H. (1982). *Transportation and Traffic Engineering Handbook* (2nd edition), Prentice-Hall, Englewood Cliffs, NJ (Institute of Transportation Engineers).

Hopp, W. J. and Spearman, M. L. (1996). *Factory Physics: Foundations of Manufacturing Management*. Irwin/McGraw-Hill, New York.

Hunt, G. C. (1956). Sequential arrays of waiting times. *Operations Research*, **4**(6), 674–683.

Ibidapo-Obe, O. and Asaolu, S. (2006). Optimization problems in applied sciences: From classical through stochastic to intelligent metaheuristic approaches. In *Handbook of Industrial and Systems Engineering*, Badiru, A., Ed. CRC Press, Boca Raton, FL, Chapter 22.

Ignizio, J. P. (1982). *Linear Programming in Single- and Multiple-Objective Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Jennings, G. A. (1994). *Modern Geometry with Applications*. Springer-Verlag, New York.

Johnson, L. A. and Montgomery, D. C. (1974). *Operations Research in Production Planning, Scheduling, and Inventory Control*. Wiley, New York.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1 (2nd edition). Wiley, New York.

Karloff, H. (1991). *Linear Programming*. Birkhäuser, Boston, MA.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, **4**, 373–395.

Kells, L. M., Kern, W. F., and Bland, J. R. (1942). *Spherical Trigonometry with Naval and Military Applications*. McGraw-Hill, New York.

Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov Chains*. Springer-Verlag, New York.

Kendall, D. G. (1951). Some problems in the theory of queues. *Journal of the Royal Statistical Society*, Series B, **13**(2), 151–173.

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Annals of Mathematical Statistics*, **24**, 338–354.

Kendall, M. G. and Moran, P. A. P. (1963). *Geometric Probability*. Hafner, New York.

Khintchine, A. Y. (1932). Mathematical theory of stationary queues. *Matem. Sbornik*, **39**(4), 73–84 (paper in Russian, abstract in German).

Kim, D. S. and Alden, J. M. (1997). Estimating the distribution and variance of time required to produce a fixed lot size given deterministic processing times and random downtimes. *International Journal of Production Research*, **35**(12), 3405–3414.

Kimura, T. (1986). A two-moment approximation for the mean waiting time in the GI/G/s queue. *Management Science*, **32**(6), 751–763.

Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Proceedings of the Cambridge Philosophical Society*, **57**(1), 902–904.

Kingman, J. F. C. (1962). On queues in heavy traffic. *Journal of the Royal Statistical Society*, Series B, **24**(2), 383–392.

Kingman, J. F. C. (1965). The heavy traffic approximation in the theory of queues. *Proceedings of the Symposium on Congestion Theory*, Smith, W. L. and Wilkinson, W. E., Eds. University of North Carolina Press, Chapel Hill, North Carolina.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, **220**, May 13, 1983, 761–679.

Kittel, C. and Kroemer, H. (1980). *Thermal Physics* (2nd edition). Freeman & Company, San Francisco, CA.

Kleinrock, L. (1975). *Queueing Systems—Volume I: Theory*. Wiley, New York.

Kleinrock, L. (1976). *Queueing Systems—Volume II: Computer Applications*. Wiley, New York.

Krämer, W. and Langenbach-Belz, M. (1976). Approximate formulae for the delay in the queueing system GI/G/1. *Congressbook, Eighth International Teletraffic Congress (ITC)*, Melbourne, Australia, pp. 235-1–235-8.

Krämer, W. and Langenbach-Belz, M. (1978). Approximate formulae for general single server systems with single and batch arrivals. *Angewandte Informatik*, 396–402. (Concise version of Krämer and Langenbach-Belz (1976), listed above).

Krishnan, V. (2006). *Probability and Random Process*. Wiley, Hoboken, NJ.

Kumaresan, S. (2005). *Topology of Metric Spaces*. Alpha Science International, Harrow, U.K.

Landsberg, P. T. (1978). *Thermodynamics and Statistical Mechanics*. Oxford University Press, Oxford, U.K.

Larson, R. C. and Odoni, A. R. (1981). *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, NJ.

Law, A. M. and Kelton, W. D. (1991). *Simulation Modeling and Analysis* (2nd edition). McGraw-Hill, New York.

Lee, K. Y. and El-Sharkawi, M. A., Eds. (2008). *Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems*. IEEE Press Series on Power Engineering, Wiley, Hoboken, NJ.

Lee, Z.-J. and Lee, C.-Y. (2005). A hybrid search algorithm with heuristics for resource allocation problem. *Information Sciences*, **173**(1–3), 155–167.

Lefebvre, M. (2006). *Applied Probability and Statistics*. Springer, New York.

Lewis, C. D. (1970). *Scientific Inventory Control*. Elsevier, New York.

Li, J. and Meerkov, S. M. (2009). *Production Systems Engineering*. Springer, New York.

Linsley, R. K. and Franzini, J. B. (1955). *Elements of Hydraulic Engineering*. McGraw-Hill, New York.

Little, J. D. C. (1961). A proof for the queueing formula: $L=\lambda W$. *Operations Research*, **9**(3), 383–387.

Love, R. F. and Morris, J. G. (1972). Modelling inter-city road distances by mathematical functions. *Operational Research Quarterly*, **23**(1), 61–71.

Love, S. F. (1979). *Inventory Control*. McGraw-Hill, New York.

Luenberger, D. G. (1984). *Linear and Non-Linear Programming* (2nd edition). Addison-Wesley, Reading, MA.

Makino, T. (1964). On the mean passage time concerning some queuing problems of the tandem type. *Journal of the Operations Research Society of Japan*, **7**(1), 17–47.

McClain, J. O. and Thomas, L. J. (1985). *Operations Management: Production of Goods and Services* (2nd edition). Prentice-Hall, Englewood Cliffs, NJ.

McFadden, F. R. (1972). On lead time demand distributions. *Decision Sciences*, **3**, 106–126.

McNeil, D. R. and Weiss, G. H. (1974). Delay problems for isolated intersections. In *Traffic Science*, Gazis, D. C., Ed. Wiley, New York.

Medhi, J. (1991). *Stochastic Models in Queueing Theory*. Academic Press, Boston, MA.

Mehrotra, S. (1992). On the implementation of a primal-dual interior point method. *SIAM Journal of Optimization*, **2**(4), 575–601.

Melzak, Z. A. (1983). *Invitation to Geometry*. Wiley, New York.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**(6), 1087–1092.

Michalewicz, Z. and Fogel, D. B. (2004). *How to Solve it: Modern Heuristics* (2nd edition). Springer, Berlin, Germany.

Michiels, W., Aarts, E., and Korst, J. (2007). *Theoretical Aspects of Local Search*. Springer, Berlin, Germany.

Monticelli, A. J., Romero, R., and Asada, E. N. (2008). Fundamentals of simulated annealing. In *Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems*, Lee, K. Y. and El-Sharkawi, M. A., Eds. IEEE Press Series on Power Engineering, Wiley, Hoboken, NJ, Chapter 7.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd edition). McGraw-Hill, New York.

Morse, P. M. (1958). *Queues, Inventories and Maintenance*. Wiley, New York.

Munakata, T. (1979). *Matrices and Linear Programming with Applications*. Holden-Day, San Francisco, CA.

Murty, K. G. (1976). *Linear and Combinatorial Programming*. Wiley, New York.

Muth, E. J. (1987). An update on analytical models of serial transfer lines. Research Report No. 87-15, Department of Industrial and Systems Engineering, University of Florida, Gainsville, FL.

Nahmias, S. (1989). *Production and Operations Analysis*. Irwin, Homewood, IL.

Nemhauser, G. L. and Wolsey, L. A. (1988). *Integer and Combinatorial Optimization*. Wiley, New York.

Neter, J., Wasserman, W., and Kutner, M. H. (1985). *Applied Linear Statistical Models*. Richard D. Irwin, Homewood, IL.

Newell, G. F. (1982). *Applications of Queueing Theory* (2nd edition). Chapman & Hall, London, U.K.

Newell, G. F. (1989). *Theory of Highway Traffic Signals*. Publication No. UCB-ITS-CN-89-1, Institute of Transportation Studies, University of California, Berkeley, CA.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, Part I: General theory. *Transportation Research*, **27B**(4), 281–287.

Ozan, T. M. (1986). *Applied Mathematical Programming for Engineering and Production Management*. Prentice-Hall, Englewood Cliffs, NJ.

Pal, N., Jin, C., and Lim, W. K. (2006). *Handbook of Exponential and Related Distributions for Engineers and Scientists*. Chapman and Hall/CRC, Boca Raton, FL.

Papadopoulos, H. T., Heavy, C., and Browne, J. (1993). *Queueing Theory in Manufacturing Systems Analysis and Design*. Chapman & Hall, London, U.K.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes* (2nd edition). McGraw-Hill, New York.

Pollaczek, F. (1930). Über eine Aufgabe der Wahrscheinlichkeitstheorie. *Mathematische Zeitschrift*, **32**, 64–100, 729–750.

Porteus, E. L. (1985). Numerical comparisons of inventory policies for periodic review systems. *Operations Research*, **33**(1), 134–152.

Poularikas, A. D. and Seely, S. (1985). *Signals and Systems*. PWS Engineering, Boston, MA.

Press, S. J. (1982). *Applied Multivariate Analysis*. Robert E. Krieger, Malabar, FL.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing* (3rd edition). Cambridge University Press, Cambridge, U.K.

Ravindran, A., Phillips, D. T., and Solberg, J. J. (1987). *Operations Research: Principles and Practice*. Wiley, New York.

Rayward-Smith, V. J., Osman, I. H., Reeves, C. R., and Smith, G. D. (1996). *Modern Heuristic Search Methods*. Wiley, New York.

Reeves, C. R. (1993). *Modern Heuristic Techniques for Combinatorial Problems*. Wiley, New York.

Ross, S. M. (1983). *Stochastic Processes*. Wiley, New York.

Ross, S. M. (1988). *A First Course in Probability* (3rd edition). Macmillan, New York.

Ross, S. M. (2003). *Introduction to Probability Models* (8th edition). Academic Press, New York.

Sakasegawa, H. (1977). An approximation formula $L_q \cong \alpha \cdot \rho^{\beta}/(1-\rho)$. *Annals of the Institute of Statistical Mathematics*, **29**, Part A, 67–75.

Schneider, H. (1981). Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research*, **19**(6), 615–631.

Schwarzschild, B. M. (1982). Statistical mechanics algorithm for Monte Carlo simulation. *Physics Today*, **35**(5), May 1982, 17–19.

Shanthikumar, J. G. and Buzacott, J. A. (1980). On the approximations to the single server queue. *International Journal of Production Research*, **18**(6), 761–773.

Silver, E. A. (2004). An overview of heuristic solution methods. *Journal of the Operational Research Society*, **55**(9), 936–956.

Silver, E. A., Pyke, D. F., and Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling* (3rd edition). Wiley, New York.

Simmons, H. A. (1945). *Plane and Spherical Trigonometry*. Wiley, New York.

Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope*, **68**(2), 159.

Sipper, D. and Bulfin, R. L. (1997). *Production: Planning, Control, and Integration*. McGraw-Hill, New York.

Smeed, R. J. (1967). The road capacity of city centers. *Highway Research Record*, **169**, Transportation Research Board, Washington, DC.

Smeed, R. J. (1971). The effect of the design of road network on the intensity of traffic movement in different parts of a town with special reference to the effects of ring roads. Technical Note 17, Construction Industry Research and Information Association, London, U.K.

Smith III, J. O. (2008). *Spectral Audio Signal Processing*. Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA.

Sokhan-Sanj, S., Gaxiola, G., Mackulak, G. T., and Malmgren, F. B. (1999). A comparison of the exponential and the hyperexponential distributions for modeling move requests in a semiconductor fab. *Proceedings of the 1999 Winter Simulation Conference*, Phoenix, AZ, December 1999; Farrington, P. A., Nembhard, H. B., Sturrock, D. T. and Evans, G. W., Eds., pp. 774–778.

Sokolnikoff, I. S. (1939). *Advanced Calculus*. McGraw-Hill, New York.

Song, Y.-H., Lu, H., Lee, K. Y. and Yo, I. K. (2008). Fundamentals of ant colony search algorithms. In *Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems*, Lee, K. Y. and El-Sharkawi, M. A., Eds. IEEE Press Series on Power Engineering, Wiley, Hoboken, NJ, Chapter 5.

Spivey, W. A. and Thrall, R. M. (1970). *Linear Optimization*. Holt, Rinehart & Winston, New York.

Stein, D. M. (1978). An asymptotic, probabilistic analysis of a routing problem. *Mathematics of Operations Research*, **3**(2), 89–101.

Stevenson, W. J. (1986). *Production/Operations Management*. Irwin, Homewood, IL.

Stewart, J. (1995). *Calculus* (3rd edition). Brooks/Cole, Pacific Grove, CA.

Stirzaker, D. (1994). *Elementary Probability*. Cambridge University Press, Cambridge, U.K.

Sturge, M. D. (2003). *Statistical and Thermal Physics: Fundamentals and Applications*. A. K. Peters, Natick, MA.

Taft, E. W. (1918). The most economical production lot. *The Iron Age*, **101**, 1410–1412.

Tanner, J. C. (1962). A theoretical analysis of delays at an uncontrolled intersection. *Biometrika*, **49**(1 & 2), 163–170.

Tanner, M. (1995). *Practical Queueing Analysis*. McGraw-Hill, New York.

Tersine, R. J. (1985). *Production/Operations Management: Concepts, Structure, and Analysis* (2nd edition). North-Holland, New York.

Thurston, W. P. (1997). *Three-Dimensional Geometry and Topology*, Volume 1. Princeton University Press, Princeton, NJ.

Tijms, H. C. (1986). *Stochastic Modelling and Analysis: A Computational Approach*. Wiley, New York.

Tijms, H. C. and Groenevelt, H. (1984). Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints. *European Journal of Operational Research*, **17**, 175–190.

Trim, D. W. (1983). *Calculus and Analytic Geometry*. Addison-Wesley, Reading, MA.

Trivedi, K. S. (2002). *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. Wiley, New York.

Vanderbei, R. J. (1997). *Linear Programming*: *Foundations and Extensions*. Kluwer, Boston, MA.

Van der Heijden, F., Duin, R., de Ridder, D., and Tax, D. (2004). *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*. Wiley, Hoboken, NJ.

Vaughan, R. J. (1987). *Urban Spatial Traffic Patterns*. Pion, London, U.K.

Wagner, H. M. (1969). *Principles of Operations Research—With Applications to Managerial Decisions*. Prentice-Hall, Englewood Cliffs, NJ.

Wald, A. (1947). *Sequential Analysis*. Wiley, New York.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, *U.K.*, Part II, **1**(2), 325–378 (Road Paper No. 36).

Wardrop, J. G. (1968). Journey speed and flow in central urban areas. *Traffic Engineering & Control (TEC)*, **9**(11), 528–539.

Webster, F. V. (1958). Traffic signal settings. Road Research Technical Paper No. 39, Department of Scientific and Industrial Research, Her Majesty's Stationery Office, London, U.K.

Webster, F. V. and Cobbe, B. M. (1966). Traffic signals. Road Research Technical Paper No. 56, Road Research Laboratory, Her Majesty's Stationery Office, London, U.K.

Weiss, G. H. and Maradudin, A. A. (1962). Some problems in traffic delay. *Operations Research*, **10**(1), 74–104.

White, J. A., Agee, M. A., and Case, K. E. (1977). *Principles of Engineering Economic Analysis*. Wiley, New York.

Whitt, W. (1983). The queueing network analyzer. *The Bell System Technical Journal*, **62**(9), 2779–2815.

Whitt, W. (1993). Approximations for the GI/G/m queue. *Production and Operations Management*, **2**(2), 114–161.

Wilson, R. H. (1934). A scientific routine for stock control. *Harvard Business Review*, **13**(1), 116–128.

Winkler, P. and Gilli, M. (2004). Applications of optimization heuristics to estimation and modelling problems. *Computational Statistics and Data Analysis*, **47**(2), 211–223.

Winkler, R. L., Roodman, G. M., and Britney, R. R. (1972). The determination of partial moments. *Management Science*, **19**(3), 290–296.

Wolff, R. W. (1989). *Stochastic Modeling and the Theory of Queues*. Prentice-Hall, Englewood Cliffs, NJ.

Woolsey, R. E. D. and Swanson, H. S. (1975). *Operations Research for Immediate Application: A Quick and Dirty Manual*. Harper & Row, New York.

Wright, S. J. (1997). *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Zwillinger, D. (1996). *Standard Mathematical Tables and Formulae* (30th edition). CRC Press, Boca Raton, FL.

Index

A

Absolute error, 47–48
Analytical results, 1
Assignment problem, 164–165
Average delay, 148, 150

B

"Back-of-the-envelope" calculations, 1
Bayes' formula, 76
Bernoulli trials, 17
Beta distribution
  beta function, 41–42
  cumulative distribution function, 41
  mean and variance, 42
  probability density function, 41–43
Beta function, 188
Binomial distribution
  mean, 18
  negative distribution, 22–24
  probability, 18–19
  variance, 19
Branch and bound technique, 171

C

Calculus
  change of variable
    double integral, 199–201
    integral, 198–199
  differentiation
    chain rule, 192
    function and constant multiple, 191
    integral sign (Leibnitz's rule), 197–198
    inverse function, 192
    sum of functions, 191
  fundamental theorem, 193
  integration
    numerical, 204–208
    order change, double integral, 201–203
    parts, 192–193
  Lagrange multipliers, 196–197
  L'Hôpital's rule, 195–196
  Maclaurin series, 194–195
  summation order change, 203–204
  Taylor series, 193–194
Central limit theorem, 77
Chebyshev distance, 134–135
Circular permutations, 214
Combinatorics, 213–214
Conditional cumulative distribution function, 73–74
Conditional mean, 47
Conditional probability density function, 73–74
Constrained random variable, 9–10
Continuous probability distributions
  beta distribution
    beta function, 41–42
    cumulative distribution function, 41
    mean and variance, 42
    probability density function, 41–43
  Erlang distribution
    coefficient of variation, 36
    cumulative distribution function, 35
    mean and variance, 36
    probability density function, 34–35, 37
    scale parameter, 36–37
    shape parameter, 37
    standard deviation, 36
  exponential distribution
    coefficient of variation, 34
    cumulative distribution function, 33

probability density function, 32–33
standard deviation, 34
variance and mean, 33
gamma distribution
chi-squared (χ^2) distribution, 40
gamma function, 38
mean and variance, 39
parameters, 39–40
probability density function, 38, 40
standard deviation, 39
Gumbel (extreme value) distribution
cumulative distribution function, 54–55
Euler's constant, 55
location and scale parameter, 56
probability density function, 55
standard Gumbel distribution, 56
variance and mean, 55
hyper-exponential distribution
coefficient of variation, 62
cumulative distribution function, 60–61
mean and variance, 61
queueing system, 63
standard deviation, 62
hypo-exponential distribution
coefficient of variation, 65
constants, 64
cumulative distribution function, 63–64
probability density function, 64
standard deviation, 65
variance, 64–65
logistic distribution, 53–54
lognormal distribution
coefficient of variation, 50
cumulative distribution function, 49
logarithmic transformation, 50
mean and variance, 49
μ and α parameters, 50–51
probability density function, 48–49
standard deviation, 50
normal distribution
approximation, cumulative function, 47–48
partial moments, 46–47
standard, 45
sum, random variables, 44–45
n phases, 66–67
Pareto distribution, 56–58
triangular distribution
cumulative distribution function, 58
mean and variance, 58–59
probability density function, 58–60
uniform distribution, 31–32
Weibull distribution, 51–52
Continuous-time Markov chain, 89–90
Covariance and correlation coefficient, 5
Cumulative distribution function, 4, 10
Cycle time, 101, 150

D

Differentiation
chain rule, 192
function and constant multiple, 191
integral sign (Leibnitz's rule), 197–198
inverse function, 192
sum of functions, 191
Dirac delta function, 189
Discrete probability distributions
Bernoulli distribution, 17–18
Binomial distribution
mean, 18
negative, 22–24
probability, 18–19
variance, 19
geometric distribution
Bernoulli trials, 21
mean and variance, 20
memoryless property, 22
nonnegative integer, 21
probability, 20–21

hypergeometric distribution
mean and variance, 27
numbers of combinations, object, 27–28
parameters, 28
probability, 26
multinomial distribution, 29–30
Poisson distribution
cumulative distribution function, 25
mean and variance, 25
μ parameter, 26
probability, 24–25
Discrete-time Markov chain, 89
Driver's critical gap, 147

E

Economic order quantity (EOQ)
cost per unit time, 109
cumulative orders and demand *vs.* time, 109–110
optimal quantity, 110
time *vs.* inventory level, 109–110
trade-off, inventory and ordering costs, 111
Economic production quantity (EPQ)
cost per unit time, 113
cumulative production and demand *vs.* time, 112
time *vs.* inventory level, 112–113
trade-off, inventory and production set-up cost, 111, 113
Effective green time, 150
Equity payment, 217
Erlang distribution
coefficient of variation, 36
cumulative distribution function, 35
mean and variance, 36
probability density function, 34–35, 37
scale parameter, 36–37
shape parameter, 37
standard deviation, 36
Euclidean (straight line) distance, 134–135
Exponential distribution, 13
Exponential distribution order statistics, 184–185

F

Finite sums, 215
First-come first-served (FCFS) queueing discipline, 91
Fixed-time signal, 150

G

Gamma distribution, 75–76
Gamma functions, 38, 52, 75, 187
General distribution order statistics, 181–182
Genetic algorithms, 179–180
Geometric distribution
Bernoulli trials, 21
mean and variance, 20
memoryless property, 22
nonnegative integer, 21
probability, 20–21
G/G/c queue, 91
*G/G/*1 queue, 97–98
Greedy algorithm, 180
Gumbel (extreme value) distribution
cumulative distribution function, 54–55
Euler's constant, 55
location and scale parameter, 56
probability density function, 55
standard Gumbel distribution, 56

H

Heuristic search methods
ant colony optimization, 180
decision variables, 173
definition, 174
general formulation, maximization problem, 173–174
genetic algorithms, 179–180
greedy algorithm, 180
local search methods, 175–176
minimization problem, 174–175
simulated annealing

acceptance criterion, 176–177
minimization problem, algorithm, 177–178
new solution search, iteration, 176
temperature, 176–177
tabu search, 178–179
Hypergeometric distribution
mean and variance, 27
numbers of combinations, object, 27–28
parameters, 28
probability, 26

I

Identity (unity) matrix, 211
Incomplete beta function, 188–189
Incomplete gamma function, 187–188
Infinite sums, 215–216
Integer linear programming
knapsack problem, 166–167
traveling salesman problem
constraints, 168–169
n^2 variables, 168
tour and subtours, 169
Integration
numerical
midpoint rule approximation, 205–206
Simpson's rule approximation, 205–208
trapezoidal rule approximation, 205–207
order change, double integral, 201–203
parts, 192–193
Interest payment, 217
Interior-point methods, 170
Inventory control
economic order quantity (EOQ)
cost per unit time, 109
cumulative orders and demand *vs.* time, 109–110
optimal quantity, 110
time *vs.* inventory level, 109–110
trade-off, inventory and ordering costs, 111
economic production quantity (EPQ)
cost per unit time, 113
cumulative production and demand *vs.* time, 112
time *vs.* inventory level, 112–113
trade-off, inventory and production set-up cost, 111, 113
newsboy problem, uncertain demand
expected cost, 114
optimal quantity, 114–115
overage and shortage costs, 115
(s, Q) policy, 128
cumulative distribution function, 121
demand variance, 120
fill rate, β, 121–122
inventory pattern *vs.* time, 116
lead time, 120–121
order quantity, 120
probability, α, 121–122
reorder point, 120
replenishment policy, 119
safety stock, 120
standard deviation, 120–122
(s, S) policy, 128
inventory pattern *vs.* time, 116–117
reorder point and order-up-to level, 123
replenishment policy, 122
service level factor, 123
(T, S) policy, 128
inventory pattern *vs.* time, 116–117, 125
ordering cost and demand variance, 124
review period and standard deviation, 125
(T, s, S) policy, 128
heuristic approximations, 126
inventory pattern *vs.* time, 116, 118, 126
ordering cost and demand variance, 126
review period, 127

production/distribution system
average inventories, 130–131
cumulative production, shipments, and demand *vs.* time, 129–130
inventory components, 129
Inventory position, 118–119
Inverse function, 72

K

Knapsack problem, 166–167
*k*th-order statistic, 181, 183

L

Lagrange multipliers, 196–197
Laplace transform, 83–84
L'Hôpital's rule, 195–196
Linear combination, 6–7
Linear programming formulations, 2
alternative formulations
equality constraints, 156–157
minimization *vs.* maximization, 156
reversed inequality constraints, 157
assignment problem, 164–165
branch and bound technique, 171
cutting planes, 170
diet problem, 157–158
duality
dual and primal problem, 158–159
theorem, 159
feasible region, 155–156
integer linear programming
knapsack problem, 166–167
traveling salesman problem, 167–169
interior-point methods, 170
Maximize *Z*, 153
network flow methods, 170
simplex method, 169–170
terminology, 154
transportation problem
constraint coefficients, 162–163
correspondence, general linear programming, 163
definition, 160
integrality property, 162
shipments, 160
supply and demand, 160–161
transshipment problem, 163–164
Logistics analysis
average rectangular grid distance, 139
distance norms
Chebyshev distance, L_∞ norm, 134–135
definition, 133
Euclidean (straight line) distance, L_2 norm, 134–135
geodesic, 135
L_p norm, 133–134
rectangular grid (rectilinear) distance, L_1 norm, 134
two-dimensional space, 133–134
vector, 133
great circle distance
haversine formula, 141
law of cosines, 140
mean radius, 139
spherical triangle, 140
random points distance, circle
average distance, 138
cumulative distribution function, 137–138
probability density function, 137
standard deviation, 138
traveling salesman problem tour distance, 136

M

Maclaurin series, 194–195
Markovian property, 91
Mathematical results, 2
Matrices
2×2 and 3×3 matrix inverse, 210
calculation rules, 209–210
derivatives, 211
identity (unity) matrix, 211
M/D/2 queue, 91

Means and variances
  Bernoulli distribution, 17
  constrained random variable, 9–10
  covariance and correlation coefficient, 5
  random number, 11
  random variable
    approximations, 12
    cumulative distribution function, 4
    discrete and continuous, 3
    exponential distribution, 13
    function, 11–12
    joint distribution, 9
    normal distribution, 13–15
    probability density function, 4
    product, 7–8
    quotient, 8–9
    standard deviation, 5
    sum, 5–7
M/*M*/1 queue, 91
M objects permutations, 214
Modified Bessel functions, 189–190
Multinomial distribution, 29–30
Muth's approximation, 108

N

Negative binomial distribution, 76
Network flow methods, 170
Normal distribution
  correlation coefficient, 13, 15
  cumulative distribution function, 14
  α and β parameters, 13
  probability-weighted average mean, 14–15
  variance, 14

P

Pareto distribution, 56–58
Pascal distribution, 24
Poisson distribution, 75–76
  cumulative distribution function, 25
  mean and variance, 25
  μ parameter, 26
  probability, 24–25
Poisson process
  arrivals, fixed interval, 87
  independent Poisson process, superposition, 86
  properties, 85
  splitting, 86
Pollaczek-Khintchine formula, 96
Probability density function, 4, 10
Probability generating function, 78–79
Probability relationships
  Bayes' formula, 76
  central limit theorem, 77
  characteristic function
    cumulant generating function, 83
    derivatives, 82
    Fourier transform, 81–82
    random variable, 81
    Taylor series expansion, 83
  independent random variables, 69
  Laplace transform, 83–84
  maximum and minimum distribution, random variable
    cumulative distribution function, 69–70
    exponential distribution, 71
    probability density functions, 70
    uniform distribution, 70–71
  moment generating function
    successive derivatives, 80
    variance, 81
  Poisson and gamma distributions, 75–76
  probability distribution
    change of variable, 72–73
    conditional, constrained random variable, 73–74
  probability generating function, 78–79
Production/distribution system
  average inventories, 130–131
  cumulative production, shipments, and demand *vs.* time, 129–130
  inventory components, 129
Production systems modeling
  serial production line throughput
    line with buffers, 105
    line without buffers, 104–105

N-station, 107–108
two stations with buffer, 107
two stations without buffer, 106
workstations
definitions and notation, 101
parameter relationship, 101–102
time distribution, fixed lot size, 102–104
Product rule, 191

Q

Queue departure process, 98
Queueing theory
average queue length formulas, 93–94
average time, 94–95
additional formulas, 96–97
Pollaczek-Khintchine formula, 96
heavy traffic approximation, time distribution, 97–98
Little's law
extension, 93
general system relationship, 92
M/*M*/1 queue, distribution results
number of customers, 99
time, 99–100
queue departure process, 98
queueing system variables, 91–92
queue types notation, 91
references, formula, 95
Quotient rule, 191

R

Random variables, 181–182
Real-world problems, 1
Rectangular distribution, *see* Uniform distribution
Rectangular grid (rectilinear) distance, 134

S

Saturation flow, 150
Serial production line throughput
line with buffers, 105
line without buffers, 104–105
N-station, 107–108
two stations with buffer, 107
two stations without buffer, 106
Simplex method, 169–170
Simulated annealing
acceptance criterion, 176–177
minimization problem, algorithm, 177–178
new solution search, iteration, 176
temperature, 176–177
Standard normal distribution, 45
Stirling's formula, 190
Stochastic processes, 1–2
competing exponentials, 86
continuous-time Markov chain, 89–90
discrete-time Markov chain, 89
"lack of memory" property, exponential distribution, 85–86
Poisson process
arrivals, fixed interval, 87
independent Poisson process, superposition, 86
properties, 85
splitting, 86
renewal process
first interval distribution, 87–88
mean and variance, number of arrivals, 87
Summations
finite sums, 215
infinite sums, 215–216

T

Tabu search, 178–179
Taylor series, 9, 83, 193–194
Traffic flow and delay
minor road
critical flow, 148–149
traffic queue, 149–150
parameters, 143
traffic speed
average travel time, 146
coefficient of variation, 145

relative frequencies, time and space, 144
space-mean speed, 144–146
time-mean speed, 144, 146–147
total flow and concentration, 144–145
variance, space-distribution of speeds, 145
vehicle merging, traffic stream, 147–148
vehicle, traffic signal, 150–151
Traffic intensity, 94, 98
z-Transform, *see* Probability generating function
Transportation problem
constraint coefficients, 162–163
correspondence, general linear programming, 163
definition, 160
integrality property, 162
shipments, 160
supply and demand, 160–161
Transshipment problem, 163–164
Traveling salesman problem tour distance, 136

U

Uniform distribution, 31–32
Uniform distribution order statistics
*k*th-order statistic, 183
mean and variance, 183–184
Unique probability distribution, 80
Unit impulse function, 189

V

Vehicle-actuated signal, 150

W

Weibull distribution, 51–52
Workstations, production systems modeling
definitions and notation, 101
parameter relationship, 101–102
time distribution, fixed lot size
mean and variance, 102–103
probability density function, 103
unit impulse function, 104

Manufacturing & Industrial Engineering

A handbook in the truest sense of the word, the first edition of the *Operations Research Calculations Handbook* quickly became an indispensible resource. While other books available tend to give detailed information about specific topics, this one contains comprehensive information and results useful for real-world problem solving. Reflecting the breadth and depth of growth in the field, the scope of the second edition has been expanded to cover several additional topics. And as with the first edition, it focuses on presenting analytical results and formulas that allow quick calculations and provide understanding of system models.

See what's in the *Second Edition*

- New chapters include *Order Statistics*, *Traffic Flow and Delay*, and *Heuristic Search Methods*
- New sections include *Distance Norms*, *Hyper-Exponential and Hypo-Exponential Distributions*
- Newly derived formulas and an expanded reference list

Like its predecessor, the new edition of this handbook presents the analytical results and formulas needed in the scientific applications of operations research and management. It continues to provide quick calculations and insight into system performance. Presenting practical results and formulas without derivations, the material is organized by topic and offered in a concise format that allows ready-access to a wide range of results in a single volume.

The field of operations research encompasses a growing number of technical areas and uses analyses and techniques from a variety of branches of mathematics, statistics, and other scientific disciplines. And as the field continues to grow, there is an even greater need for key results to be summarized and easily accessible in one reference volume. Yet many of the important results and formulas are widely scattered among different textbooks and journals and are often hard to find in the midst of mathematical derivations. This book provides a one-stop resource for many important results and formulas needed in operations research and management science applications.

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

52403
ISBN: 978-1-4200-5240-4
90000
9 781420 052404