OPERATIONS RESEARCH CALCULATIONS HANDBOOK

$$E[Y] = E[N]$$

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

$$Var[Y] = E[N]Var[X] + (E[X])^2 Var[N]$$

Dennis Blumenfeld

OPERATIONS RESEARCH CALCULATIONS HANDBOOK

Dennis Blumenfeld

CRC Press
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Blumenfeld, Dennis.
Operations research calculations handbook / by Dennis Blumenfeld.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-2127-1 (alk. paper)
1. Operations research—Handbooks, manuals, etc. 2. Mathematical.
analysis—Handbooks, manuals, etc. I. Title.

T57.6 B57 2001
658.4′034—dc21 2001025425

International Standard Book Number 0-8493-2127-1
Library of Congress Card Number 2001025425
Printed in the United States of America 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Preface

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 for frequently used results to be readily available in one reference source.

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 a 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 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 shared their expertise. In particular, I wish to thank David Kim for his valuable contributions and discussions on the content and organization of this handbook. My thanks also go to Debra Elkins, Bill Jordan, and Jonathan Owen for their helpful reviews of initial drafts. I gratefully acknowledge Cindy Carelli and Judith Kamin for their careful and professional editorial work throughout the publication process. I thank my wife Sharon for her patience and encouragement. She helped me keep to my deadline, with her occasional calls of "Author! Author!"

The Author

Dennis E. Blumenfeld is a staff research scientist at the General Motors Research and Development Center. He previously held faculty positions at Princeton University and University College London. He earned a B.Sc. in mathematics, M.Sc. in statistics and operations research from Imperial College, London, and Ph.D. 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*.

Table of Contents

1 Introduction

The objective of this handbook is to provide a concise collection of analytical results and formulas that arise in operations research applications. 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 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 three chapters cover 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 useful in logistics and spatial analysis.

Chapter 11 includes standard linear programming formulations. The subject of linear programming, and mathematical programming in general, involves the development of algorithms and methodologies in optimization. In keeping with the intent of this handbook to focus on analytical results and formulas, this chapter presents the mathematical formulations of basic linear programming problems and gives references for the solution methods.

Chapters 12–17 contain basic results in mathematics that are relevant to operations research. Chapter 12 lists some common mathematical functions that arise in applications. Chapter 13 presents useful results from elementary and more advanced calculus. Chapter 14 lists the standard properties of matrices and Chapter 15 gives the standard formulas for combinatorial calculations. Chapter 16 lists some common summation results. Chapter 17 gives basic interest formulas important in investment calculations.

For each result or formula in this handbook, 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 N taking integer values (N = …–2, –1, 0, 1, 2, …), the mean of N is given by

$$E[N] = \sum_{n=-\infty}^{\infty} n \cdot \Pr\{N = n\} \tag{2.1}$$

where

$E[N]$ denotes the mean (expected value) of N

and

$\Pr\{N = n\}$ denotes the probability that N takes the value n.

If N takes non-negative integer values only (N = 0, 1, 2, …), then the mean of N is given by

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

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

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
$f(x)$ denotes 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 takes non-negative values only ($0 \le X < \infty$), then the mean of X is given by

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

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

(Çinlar, 1975; Mood, Graybill, and Boes, 1974).

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

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

(Feller, 1964; Binmore, 1983; Çinlar, 1975; Mood, Graybill, and Boes, 1974; Ross, 1989).

2.2 COVARIANCE AND CORRELATION COEFFICIENT

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

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

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

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

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

(Feller, 1964; Mood, Graybill, and Boes, 1974; Ross, 1989).
The correlation coefficient is dimensionless and satisfies $-1 \le Corr[X, Y] \le 1$.

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

2.3 MEAN AND VARIANCE OF A 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 whether or not the random variables are 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 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.

These results 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 variance of the linear combination $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ are given by

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E\left[X_i\right] \tag{2.20}$$

and

$$Var\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 Var\left[X_i\right] + \sum_{i}\sum_{\neq j} a_i a_j Cov\left[X_i, X_j\right] \tag{2.21}$$

respectively (Bolch et al., 1998; Feller, 1964; Mood, Graybill, and Boes, 1974; Ross, 1989).

2.4 MEAN AND VARIANCE OF A PRODUCT OF TWO RANDOM VARIABLES

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

$$E[XY] = E[X]\,E[Y] \tag{2.22}$$

and

$$Var[XY] = (E[Y])^2\,Var[X] + (E[X])^2\,Var[Y] + Var[X]Var[Y] \tag{2.23}$$

respectively.

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

$$E[XY] = E[X]\,E[Y] + Cov[X, Y] \tag{2.24}$$

and

$$\begin{aligned} Var[XY] = {} & \\ & (E[Y])^2\,Var[X] + (E[X])^2\,Var[Y] + 2E[X]E[Y]\,Cov[X, Y] - (Cov[X, Y])^2 + \\ & E\{(X - E[X])^2\,(Y - E[Y])^2\} + 2E[Y]E\{(X - E[X])^2\,(Y - E[Y])\} + \\ & 2E[X]E\{(X - E[X])(Y - E[Y])^2\} \end{aligned} \tag{2.25}$$

respectively (Mood, Graybill, and Boes, 1974).

2.5 MEAN AND VARIANCE OF A QUOTIENT OF TWO RANDOM VARIABLES

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

$$E\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)\left(1 + \frac{Var[Y]}{(E[Y])^2}\right) \tag{2.26}$$

and

$$Var\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)^2 \left(\frac{Var[X]}{\left(E[X]\right)^2} + \frac{Var[Y]}{\left(E[Y]\right)^2}\right) \tag{2.27}$$

respectively.

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

$$E\left[\frac{X}{Y}\right] \cong \left(\frac{E[X]}{E[Y]}\right)\left(1 + \frac{Var[Y]}{\left(E[Y]\right)^2}\right) - \frac{1}{\left(E[Y]\right)^2} Cov[X,Y] \tag{2.28}$$

and

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

respectively.

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

2.6 CONDITIONAL MEAN AND VARIANCE FOR JOINTLY DISTRIBUTED RANDOM VARIABLES

For jointly distributed random variables X and Y,

$$E[Y] = E_X\{E[Y|X]\} \tag{2.30}$$

$$Var[Y] = E_X\{Var[Y|X]\} + Var_X\{E[Y|X]\} \tag{2.31}$$

where

$E[Y]$ and $Var[Y]$ denote the unconditional mean and variance of Y,
$E[Y|X]$ and $Var[Y|X]$ denote the conditional mean and variance of Y given a value of X, and
$E_X[\cdot]$ and $Var_X[\cdot]$ denote the mean and variance over the distribution of X,

respectively (Mood, Graybill, and Boes, 1974; Ross, 1988; Wolff, 1989).

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

$$E[X \mid X > a] = \frac{\int_a^\infty xf(x)dx}{\Pr\{X > a\}} \tag{2.32}$$

$$= \frac{\int_a^\infty xf(x)dx}{\int_a^\infty f(x)dx} \tag{2.33}$$

$$= \frac{\int_a^\infty xf(x)dx}{1 - F(a)} \tag{2.34}$$

where

$f(x)$ denotes 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

$$E[X \mid a < X < b] = \frac{\int_a^b xf(x)dx}{\Pr\{a < X < b\}} \tag{2.35}$$

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

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

(Stirzaker, 1994).

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 non-negative integer random variable (independent of $X_1, X_2, \ldots, X_N$),

and let

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

Then the sum

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

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

$$E[Y] = E[N]\, E[X] \tag{2.38}$$

$$Var[Y] = E[N]\, Var[X] + (E[X])^2\, Var[N] \tag{2.39}$$

(Benjamin and Cornell,1970; Drake, 1967; Mood, Graybill, and Boes, 1974; Ross, 1983; Ross, 1989; Wald, 1947).

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 mean $E[g(X)]$ given by

$$E\big[g(X)\big] = \int_{-\infty}^{\infty} g(x) f(x) dx = \int_{-\infty}^{\infty} g(x) dF(x) \tag{2.40}$$

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

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)] \tag{2.41}$$

(Çinlar, 1975; Mood, Graybill, and Boes, 1974; Ross, 1989).

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

$$E\left[g(X)\right] \cong g(\mu) + \tfrac{1}{2}\sigma^2 g''(\mu) \tag{2.42}$$

$$Var[g(X)] \cong \sigma^2[g'(\mu)]^2 \tag{2.43}$$

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

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

and

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

(Benjamin and Cornell, 1970; Papoulis, 1984).

2.11 MEAN AND VARIANCE OF THE MAXIMUM OF EXPONENTIALLY DISTRIBUTED RANDOM VARIABLES

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

$$E\left[\max\left(X_1, X_2, ..., X_n\right)\right] = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n}\right) \tag{2.44}$$

and

$$Var\left[\max\left(X_1, X_2, ..., X_n\right)\right] = \frac{1}{\lambda^2}\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + ... + \frac{1}{n^2}\right) \tag{2.45}$$

respectively (Balakrishnan and Sinha, 1995; Cox and Hinkley, 1974; Nahmias, 1989).

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,
$\sigma_1^2 = Var[X_1]$ and $\sigma_2^2 = Var[X_2]$ 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 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

$$E[Z] = \mu_1\Phi(\alpha) + \mu_2\Phi(-\alpha) + \beta\phi(\alpha) \tag{2.46}$$

and

$$E\left[Z^2\right] = \left(\mu_1^2 + \sigma_1^2\right)\Phi(\alpha) + \left(\mu_2^2 + \sigma_2^2\right)\Phi(-\alpha) + \left(\mu_1 + \mu_2\right)\beta\phi(\alpha) \tag{2.47}$$

respectively, and the variance of Z is given by

$$Var[Z] = E[Z^2] - (E[Z])^2 \tag{2.48}$$

(Clark, 1961).

These exact formulas for two normally distributed random variables can be used to obtain approximate expressions for 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

and

$\rho_2 = Corr[X_2, Y]$ be the correlation coefficient of X_1 and Y

The correlation coefficient of Y and Z is given by

$$Corr[Y, Z] = Corr\left[Y, \max(X_1, X_2)\right] = \frac{\sigma_1 \rho_1 \Phi(\alpha) + \sigma_2 \rho_2 \Phi(-\alpha)}{\sqrt{Var[Z]}} \tag{2.49}$$

The mean and 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

$$\max(X_1, X_2, Y) = \max[Y, \max(X_1, X_2)] \tag{2.50}$$

and applying the above formulas for the mean and 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).

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

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

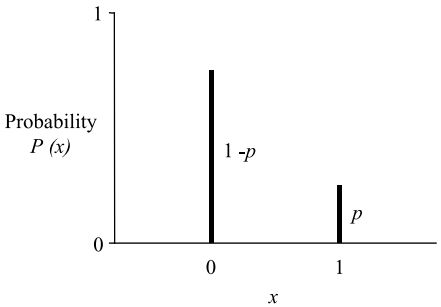

FIGURE 3.1 Example of the Bernoulli distribution.

referred to as "success" ($x = 1$) and "failure" ($x = 0$), with probabilities p and $1\text{-}p$, respectively.

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

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 $X = x$ ($x = 0, 1, 2, ..., N$)

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.

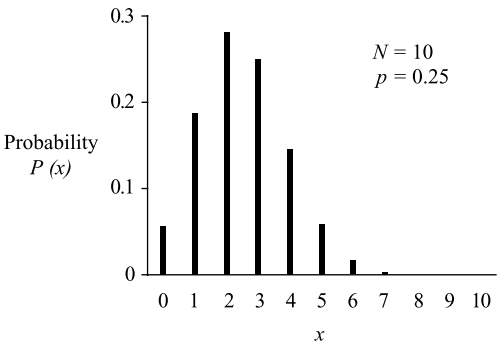

FIGURE 3.2 Example of the binomial distribution.

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

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

and

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

respectively (Ayyub and McCuen, 1997; Feller, 1964; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

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

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

The probabilities $P(x)$ for each x (x = 0, 1, 2, …, 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

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

Figure 3.3 shows an example of the geometric distribution.

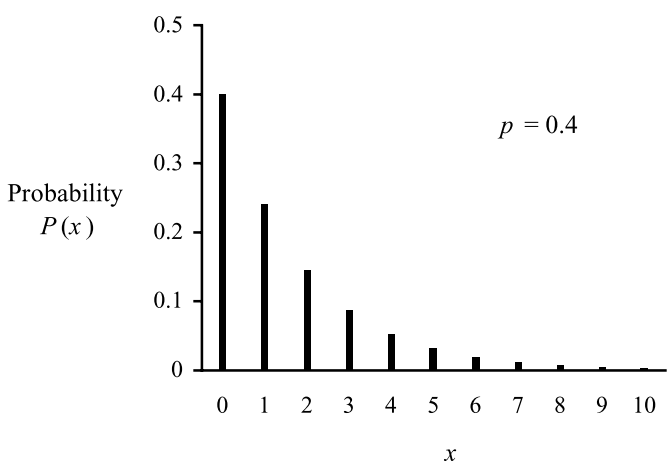

FIGURE 3.3 Example of the geometric distribution.

The mean $E[X]$ and 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; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

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 probability p of success and 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 ($x = 0, 1, 2, \ldots, N$), given by Equation 3.8, are the successive terms in the geometric series

$$p + p(1 - p) + p(1 - p)^2 + p(1 - p)^3 + \ldots$$

The sum of this series is

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

i.e.,

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

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 non-negative integer k is given by

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

The probability that X is greater than k is given by

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

The geometric distribution has the property that, for non-negative 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), i.e.,

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

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, …)

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

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

for $x = 1, 2, \ldots$, and are equal to 1 for $x = 0$. Figure 3.4 shows an example of the negative binomial distribution.

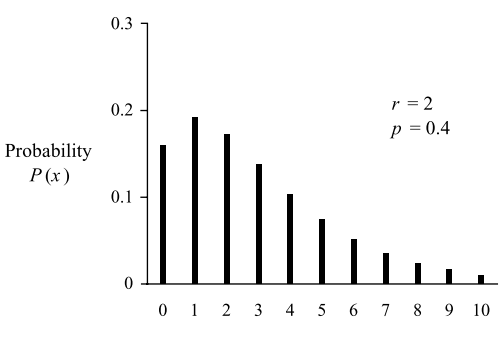

FIGURE 3.4 Example of the negative binomial distribution.

The mean $E[X]$ and 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; Feller, 1964; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

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

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

In the special case $r = 1$, the negative binomial distribution reduces to the geometric distribution. For general positive integer r, the sum of r independent and identically distributed geometric random variables has a negative binomial distribution. Thus, if $X_1, X_2, \ldots, X_r$ are r independent random variables that 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

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

The probabilities $P(x)$ for each x ($x = 0, 1, 2, \ldots$), given by Equations 3.13 or 3.14, 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 $\sum_{x=0}^{\infty} P(x) = 1$, as 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, …)

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.

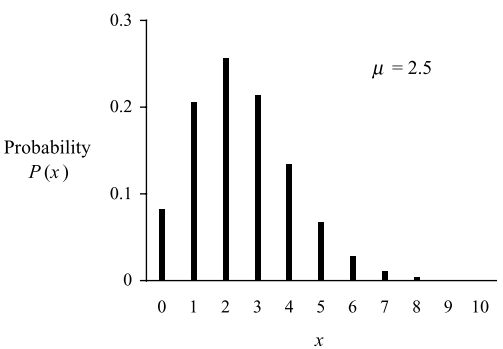

FIGURE 3.5 Example of the Poisson distribution.

The mean $E[X]$ and 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; Feller, 1964; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

Successive values of the Poisson distribution $P(x)$ (x = 0, 1, 2, …) 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).

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; Hoel, Port, and Stone, 1971), 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 Chapter 5, 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, ..., 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 number 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.

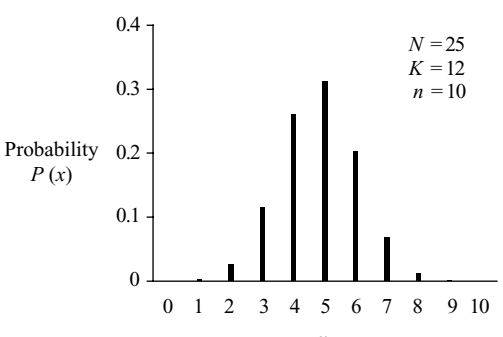

FIGURE 3.6 Example of the hypergeometric distribution.

The mean $E[X]$ and 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; Freund, 1992; Hoel, Port, and Ston, 1971; Mood, Graybill, and Boes, 1974).

The hypergeometric distribution arises in sampling from a finite population. Consider a population of N objects in total, of which K objects ($K \leq N$) 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 ($n \leq N$). The hypergeometric distribution $P(x)$ (x = 0, 1, 2, …, 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 compared to the sample size n, then there is little difference between sampling with and without replacement; 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

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

where $x_1 + x_2 + \cdots + x_k = N$ and $p_1 + p_2 + \cdots + p_k = 1$ (DeGroot, 1986; Freund, 1992; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

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 the outcomes are labeled $1, 2, \ldots, k$, 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.,

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

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

$$\frac{N!}{x_1! x_2!} p_1^{x_1} p_2^{x_2}$$

where $x_1 + x_2 = N$ and $p_1 + p_2 = 1$. Writing $x_1 = x$ and $x_2 = N - x$, with $p_1 = p$ and $p_2 = 1 - p$, the probability becomes

$$\frac{N!}{x!(N-x)!} p^x (1-p)^{N-x}$$

the standard form for the binomial distribution.

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 variance of each X_i are 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 illustrates 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 variance $Var[X]$ for the uniform distribution are given by

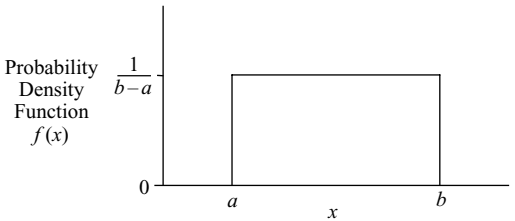

FIGURE 4.1 Example of the uniform distribution.

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

and

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

respectively (Allen, 1978; Freund, 1992; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

The uniform distribution is also known as the *rectangular distribution*. In the special case $a = 0$ and $b = 1$, the probability density function is simply $f(x) = 1$ $(0 \le x \le 1)$.

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

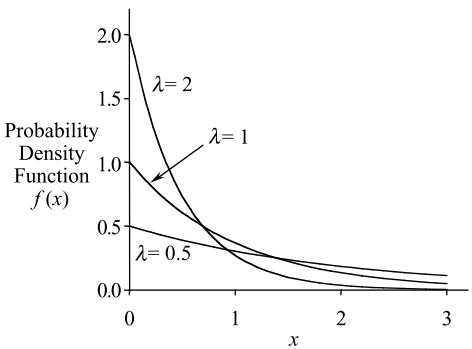

FIGURE 4.2 Examples of the exponential distribution.

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 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; DeGroot, 1986; Freund, 1992; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

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 $X > t$ (Allen, 1978; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974), i.e.,

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

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

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

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

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

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

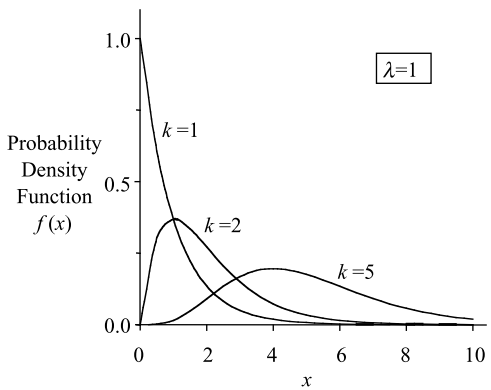

FIGURE 4.3 Examples of the Erlang distribution.

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

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

The mean $E[X]$ and 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; Law and Kelton, 1991; Tijms, 1986).

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 noninteger shape parameter), described in the next section.

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 general positive integer k, the sum of k independent and identically distributed exponential random variables has an Erlang distribution. Thus, if $X_1, X_2, \ldots, X_k$ are k independent random variables that each has an exponential distribution with mean $1/\lambda$, i.e., probability density function $\lambda e^{-\lambda x_i}$ ($0 \le x_i < \infty$, $i = 1, 2, \ldots, k$), then the sum $X = X_1 + X_2 + \ldots + X_k$ has an Erlang distribution with probability density function

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

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

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

with 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 $\int_0^x t^{\alpha-1} e^{-t} dt$ is the incomplete gamma function.

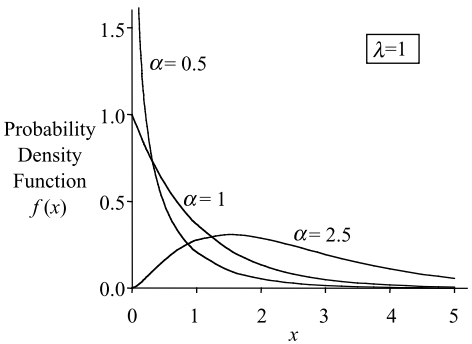

FIGURE 4.4 Examples of the gamma distribution.

The mean $E[X]$ and 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; Freund, 1992; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974; Tijms, 1986).

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 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, 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.

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; Mood, Graybill, and Boes, 1974). Thus, if Z_1, Z_2, …, Z_ν are ν independent random variables that each has a standard normal distribution, i.e., probability density function

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

then the sum $X = Z_1^2 + Z_2^2 + \ldots + Z_\nu^2$ has a χ^2 distribution with ν degrees of freedom, i.e., probability density function

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

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

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.

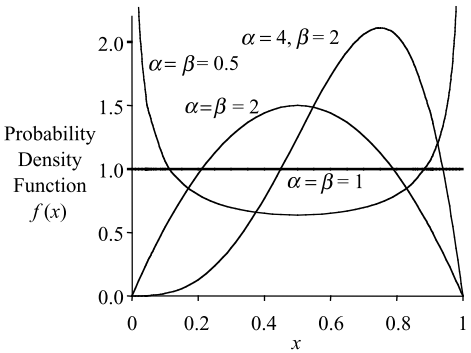

FIGURE 4.5 Examples of the beta distribution.

The mean $E[X]$ and 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; Freund, 1992; Mood, Graybill, and Boes, 1974).

If α and β are integers, the beta function

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

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, with probability density function $f(x) = 1$ $(0 < x < 1)$.

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 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; Freund, 1992; Mood, Graybill, and Boes, 1974).

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

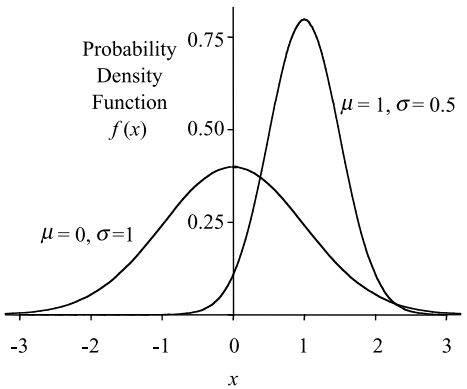

FIGURE 4.6 Examples of the normal distribution.

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

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

$$a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

has a normal distribution

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

(DeGroot, 1986; Mood, Graybill, and Boes, 1974).

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 probability density function denoted by $\phi(x)$ and 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; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

4.6.3 Partial Moments for the Normal Distribution

Let

X have 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

$$\int_{c}^{\infty} x f(x)dx$$

and

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

respectively, are given by

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

and

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

where $\Phi(x)$ is the cumulative distribution function for the standard normal distribution, given by Equation 4.36 (Hadley and Whitin, 1963; Winkler, Roodman, and Britney, 1972). Note: partial moments arise in the conditional mean and variance of a random variable X, given X is greater than a constant. For example, the conditional mean $E[X|X > c]$ is given by

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

(see Chapter 2, Equation 2.34).

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

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

where

$$a_1 = 0.4361836,\ a_2 = -0.1201676,\ a_3 = 0.9372980,$$

and

$$b = 0.33267.$$

The absolute error in this approximation is less than 1×10^{-5} (Abramowitz and Stegun, 1968; Hastings, 1955; Johnson, Kotz, and Balakrishnan, 1994). Note that this approximation can also be used for $\Phi(x)$ when $x \le 0$, 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

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

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

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

and

$$c_0 = 2.515517,\ c_1 = 0.802853,\ c_2 = 0.010328,$$

$$d_1 = 1.432788,\ d_2 = 0.189269,\ d_3 = 0.001308.$$

The absolute error in this approximation is less than 4.5×10^{-4} (Abramowitz and Stegun, 1968; Hastings, 1955). For $0 < p < 0.5$, 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.

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.

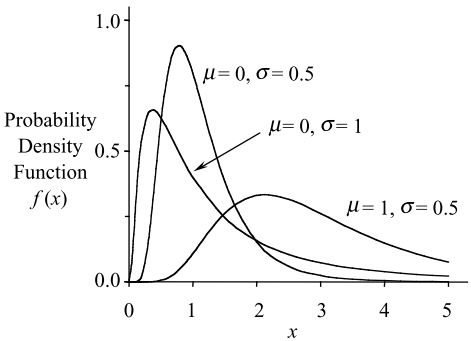

FIGURE 4.7 Examples of the lognormal distribution.

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

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

and

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

respectively (Bolch et al., 1998; Devore, 1987; Johnson, Kotz, and Balakrishnan, 1994; Mood, Graybill, and Boes, 1974; Tijms, 1986).

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 μ and variance σ^2, then X has a lognormal distribution with 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

and

- $\dfrac{\ln(X)-\mu}{\sigma}$ has a standard normal distribution.

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

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

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

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

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.

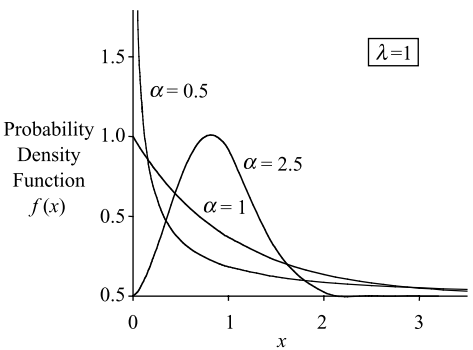

FIGURE 4.8 Examples of the Weibull distribution.

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

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

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

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

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 (Bolch et al., 1998; Devore, 1987; Mood, Graybill, and Boes, 1974).

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.

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.

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 variance $Var[X]$ for the logistic distribution are given by

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

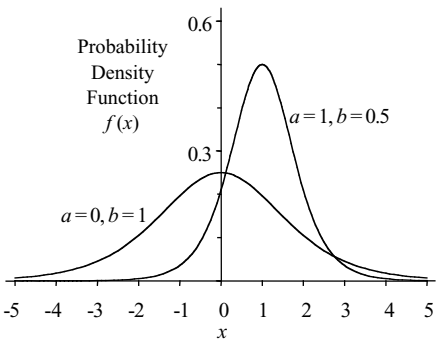

FIGURE 4.9 Examples of the logistic distribution.

and

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

respectively (Evans, Hastings, and Peacock, 2000; Mood, Graybill, and Boes, 1974).

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

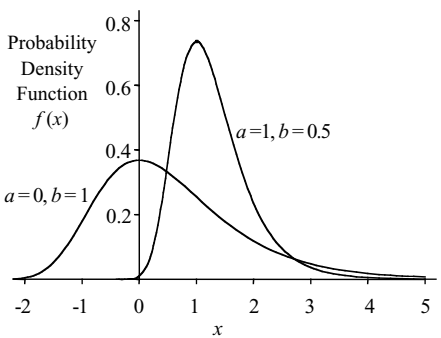

FIGURE 4.10 Examples of the Gumbel (extreme value) distribution.

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 variance $Var[X]$ for the Gumbel distribution are given by

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

where γ is Euler's constant, approximate value $\gamma \cong 0.577216$, and

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

respectively (Evans, Hastings, and Peacock, 2000; Mood, Graybill, and Boes, 1974).

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

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

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

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

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 scale parameter λ and 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.

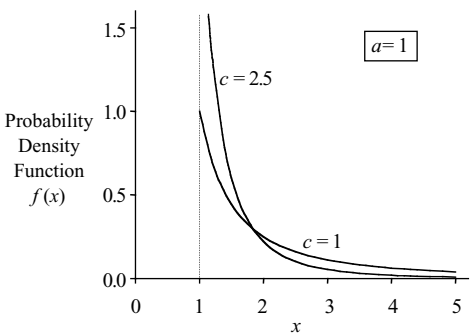

FIGURE 4.11 Examples of the Pareto distribution.

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 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 (Evans, Hastings, and Peacock, 2000; Mood, Graybill, and Boes, 1974).

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 variance $Var[X]$ for the triangular distribution are given by

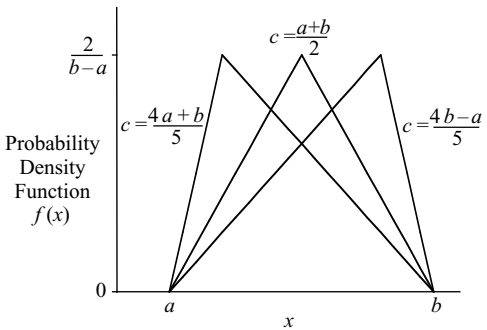

FIGURE 4.12 Examples of the triangular distribution.

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

and

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

respectively (Evans, Hastings, and Peacock, 2000; Law and Kelton, 1991).

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 mean and 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 that each has a uniform distribution over the range [0, 1], i.e., 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 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 $\dfrac{X_1 + X_2}{2}$ has a triangular distribution over the range [0, 1] with 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 $c = {}^1/_2$

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

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

(Bolch et al., 1998; Haight, 1981; Hoel, Port, and Stone, 1971; Hillier and Lieberman, 1980; Ross, 1989; Wolff, 1989). The distribution $h(z)$ is known as the *convolution* of $f(x)$ and $g(y)$, and is sometimes written as $h(z) = f(x)*g(y)$.

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

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 has a uniform distribution in the range [0, 1], then

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

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

$$E[Y] = \frac{n}{n+1} \quad \text{and} \quad E[Z] = \frac{1}{n+1}$$

respectively.

5.2.2 Example for the Exponential Distribution

If the random variables $X_1, X_2, \ldots, X_n$ are independent and each has an exponential distribution with 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

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

and

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

respectively, and have probability density functions

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

and

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

respectively. Note that the minimum Z in this example has an exponential distribution with 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 Chapter 2, Equation 2.44).

5.3 CHANGE OF VARIABLE IN A PROBABILITY DISTRIBUTION

Let

X be a continuous random variable
$f(x)$ be the probability density function of X
$Y = \psi(X)$ be a continuous, strictly increasing (or strictly decreasing) 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; Freund, 1992; Haight, 1981; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974).

In the cases where

$$\frac{dx}{dy}$$

is positive (x a strictly increasing function of y), this result can be written simply as

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

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

$$\int_{-\infty}^{x} f(t)dt$$

and

$$\int_{-\infty}^{y} g(u)du$$

and corresponds to the result for a change of variable in an integral (see Chapter 13, Section 13.9). As in the case of double integrals (Chapter 13, Section 13.10), this 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
$f(x_1, x_2)$ be the joint probability density function of X_1 and X_2
$Y_1 = \psi_1(X_1, X_2)$ 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
$g(y_1, y_2)$ 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; Freund, 1992; Haight, 1981; Mood, Graybill, and Boes, 1974).

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

Then

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

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

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

and

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

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

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

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

then

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

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

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

and

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

(Stirzaker, 1994). The corresponding results for the conditional means are given in Chapter 2, 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 β 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

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

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; McFadden, 1972; Mood, Graybill, and Boes, 1974). 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 and exhaustive events F_1, F_2, …, F_n, and any other event E, the conditional probabilities are related by

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

where $\Pr\{F_j|E\}$ denotes the conditional probability of event F_j given the event E, etc. (Feller, 1964; Ross, 1989; Clarke and Disney, 1985; Devore, 1982).

5.7 CENTRAL LIMIT THEOREM

Let

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

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

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, …, X_n (Allen, 1978; DeGroot, 1986; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974; Ross, 1989).

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

- the sum $S_n = X_1 + X_2 + \ldots + X_n$ is approximately normally distributed with mean $n\mu$ and variance $n\sigma^2$
- the sample mean $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ 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
- $\dfrac{\left(\overline{X} - \mu\right)\sqrt{n}}{\sigma}$ is approximately distributed with a standard normal distribution

5.8 PROBABILITY GENERATING FUNCTION (*z*-TRANSFORM)

Let

N be a discrete random variable taking non-negative 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

$$G'(0) = \frac{d}{dz} G(z)\bigg|_{z=0}, \; G''(0) = \frac{d^2}{dz^2} G(z)\bigg|_{z=0}, \text{ and } G^{(n)}(0) = \frac{d^n}{dz^n} G(z)\bigg|_{z=0}$$

The mean $E[N]$ and 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

$$G'(1) = \frac{d}{dz} G(z)\bigg|_{z=1} \quad \text{and} \quad G''(1) = \frac{d^2}{dz^2} G(z)\bigg|_{z=1}$$

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 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 probability generating function given by the product of the probability generating functions of the random variables.

(Allen, 1978; Bolch et al., 1998; Haight, 1981; Papadopoulos, Heavy, and Browne, 1993; Stirzaker, 1994).

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 $t = 0$:

$$E[X] = M'(0) \tag{5.51}$$

$$E[X^2] = M''(0) \tag{5.52}$$

and, in general,

$$E[X^n] = M^{(n)}(0) \tag{5.53}$$

where

$$M'(0) = \frac{d}{dt}M(t)\bigg|_{t=0}, \; M''(0) = \frac{d^2}{dt^2}M(t)\bigg|_{t=0}, \text{ and } M^{(n)}(0) = \frac{d^n}{dt^n}M(t)\bigg|_{t=0}$$

The variance of X is given by

$$Var[X] = E[X^2] - (E[X])^2 \tag{5.54}$$

$$= M''(0) - \{M'(0)\}^2 \tag{5.55}$$

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 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 moment generating function given by the product of the moment generating functions of the random variables.

(Allen, 1978; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974; Ross, 1989).

5.10 CHARACTERISTIC FUNCTION

Let

X be a random variable
$f(x)$ be the probability density function of X, if X is continuous
$p_x = \Pr\{X = x\}$ 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

$$\phi(t) = E[e^{itX}] \tag{5.57}$$

$$= \begin{cases} \displaystyle\int_{-\infty}^{\infty} e^{itx} f(x)dx & \text{if } X \text{ is continuous} \\ \displaystyle\sum_{x} e^{itx} p_x & \text{if } X \text{ is discrete} \end{cases} \tag{5.58}$$

where $i = \sqrt{-1}$. Like the moment generating function, the characteristic function generates the moments of X.

The characteristic function is a form of Fourier transform, and converges for real value of t. (The integral or sum defining a moment generating function does not necessarily converge.) The characteristic function $\phi(t)$ also has property that the probability distribution can be obtained from $\phi(t)$ by an inverse transform (Poularikas and Seely, 1985).

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,

$$E[X^n] = (-1)^n i^n \phi^{(n)}(0) \tag{5.62}$$

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 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 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

$$K(t) = \ln\{\phi(t)\} \tag{5.64}$$

$$= \kappa_0 + \kappa_1(it) + \kappa_2\frac{(it)^2}{2!} + \cdots + \kappa_n\frac{(it)^n}{n!} + \cdots \tag{5.65}$$

The coefficients κ_0, κ_1, κ_2, … in this series are known as the cumulants.

(Binmore, 1983; Hoel, Port, and Stone, 1971; Zwillinger, 1996).

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

$$\begin{aligned} L(s) &= E\left[e^{-sX}\right] \\ &= \int_0^\infty e^{-sx} f(x)dx \end{aligned} \tag{5.66}$$

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; Poularikas and Seely, 1985).

As in the cases of 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 $s = 0$:

$$E[X] = -L'(0) \tag{5.67}$$

$$E[X^2] = L''(0) \tag{5.68}$$

and, in general,

$$E[X^n] = (-1)^n L^n(0) \tag{5.69}$$

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 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 Laplace transform given by the product of the Laplace transforms of the random variables.

(Allen, 1978; Bolch et al., 1998; Haight, 1981).

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 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 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 mean $E[X] = \dfrac{1}{\lambda}$ and variance $Var[X] = \dfrac{1}{\lambda^2}$

- The number $N(t)$ of arrivals in time t is Poisson distributed, with 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 mean $E[N(t)] = \lambda t$ and variance $Var[N(t)] = \lambda t$.

(Cox and Miller, 1965; Ross, 1989; Wolff, 1989).

6.1.2 "Lack of Memory" Property of the Exponential Distribution

For any constants $s \geq 0$ and $t \geq 0$,

$$\Pr\{X > s + t \mid X > s\} = \Pr\{X > t\} \tag{6.4}$$

$$\Pr\{X > s + t\} = \Pr\{X > s\} \cdot \Pr\{X > t\} \tag{6.5}$$

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 "memoryless" property of the exponential distribution (see also Chapter 4, Section 4.2). (Allen, 1978; Heyman and Sobel, 1982; Hoel, Port, and Stone, 1971; Mood, Graybill, and Boes, 1974; Ross, 1989; Tijms, 1986)

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

$$\Pr\left\{X_i = \min\left(X_1, X_2, \ldots, X_n\right)\right\} = \frac{\lambda_i}{\sum_{k=1}^{n} \lambda_k} \tag{6.6}$$

(Ross, 1989; Tijms, 1986).

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 et al., 1998; Çinlar, 1975; Cox, 1962; Wolff, 1989).

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 ($i = 1, 2, ..., n$), where $\sum_{i=1}^{n} p_i = 1$, then arrivals of each type

i are independent Poisson processes with rates $\lambda_i = \lambda p_i$ (Bolch et al., 1998; Ross, 1989; Çinlar, 1975; Wolff, 1989).

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, 1989; Wolff, 1989).

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 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,

$$\frac{Var\left[N(t)\right]}{E\left[N(t)\right]} \cong \frac{Var[X]}{\left(E[X]\right)^2} \tag{6.7}$$

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

$$E\left[N(t)\right] = \frac{t}{E[X]} \tag{6.8}$$

$$Var\left[N(t)\right] = \frac{tVar[X]}{\left(E[X]\right)^3} \tag{6.9}$$

(Cox, 1962; Cox and Miller, 1965; Ross, 1989).

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 of $f_1(x)$ 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 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
$\sigma^2 = E[(X - \mu)^2]$ = variance of X
$\mu_3 = E[(X - \mu)^3]$ = 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 mean μ.

(Cox, 1962; Cox and Miller, 1965; Wagner, 1969).

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,

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \tag{6.14}$$

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,

$$\sum_{j=1}^{n} \pi_j = 1$$

and

$\mathbf{P} = \{P_{ij}\}$ is the matrix of transition probabilities

(Bolch et al., 1998; Çinlar, 1975; Gross and Harris, 1985; Kleinrock, 1975, 1976; Hillier and Lieberman, 1980; Ross, 1989).

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 ($i \neq j$)

$Q_{jj} = -\sum_{i,i\neq j} Q_{ij}$ (j = 1, 2, ... n)

π_j be the limiting probability for state j (j = 1, 2, ... n)

The limiting state probabilities are given by

$$0 = \pi_j Q_{jj} + \sum_{i,i\neq j} \pi_i Q_{ij} \quad (j = 1, 2, \ldots n) \tag{6.15}$$

or, in matrix notation,

$$\mathbf{0} = \boldsymbol{\pi}\mathbf{Q} \tag{6.16}$$

where

$\mathbf{0}$ = (0, 0, …, 0),
$\boldsymbol{\pi}$ = (π_1, π_2, …, π_n) is the row vector of limiting state probabilities,
$0 \leq \pi_j \leq 1$ for all j,

$$\sum_{j=1}^{n} \pi_j = 1$$

and

$\mathbf{Q}$ = $\{Q_{ij}\}$ is the transition rate matrix (also known as the intensity matrix, generator matrix, or infinitesimal generator matrix).

(Bolch et al., 1998; Gross and Harris, 1985; Hillier and Lieberman, 1980; Kleinrock, 1975, 1976; Ross, 1989).

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 first-come, first-served queueing discipline. Each system is defined by

A/B/m

where

A denotes the distribution of interarrival 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* and *G*. The symbol *M* is used for the exponential distribution to denote its Markovian ("memoryless") property. The symbol *G* is used for a general distribution of independent and identically distributed random variables. For example, an *M*/*G*/1 queue has exponentially distributed interarrival times, a general distribution for service times, and one server. A *G*/*G*/*m* queue has general distributions for interarrival times and service times, and *m* servers in parallel.

(Gross and Harris, 1985; Kendall, 1953; Wolff, 1989).

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 interarrival time
μ = service rate at one server (number of customers served per unit time); $1/\mu$ = mean service time

$\rho = \frac{\lambda}{m\mu}$ = Traffic intensity ($\rho < 1$)

σ_a^2 = variance of interarrival times
σ_s^2 = variance of service times
$C_a^2 = \lambda^2\sigma_a^2$ = Squared coefficient of variation of interarrival times
$C_s^2 = \mu^2\sigma_s^2$ = Squared coefficient of variation of service times.

7.3 LITTLE'S LAW AND GENERAL QUEUEING SYSTEM RELATIONSHIPS

$$L_q = \lambda W_q \text{ (Little's Law)} \tag{7.1}$$

$$L = \lambda W \text{ (Little's Law)} \tag{7.2}$$

$$L = L_q + \frac{\lambda}{\mu} \tag{7.3}$$

$$W = W_q + \frac{1}{\mu} \tag{7.4}$$

(Allen, 1978; Gillett, 1976; Gross and Harris, 1985; Hillier and Lieberman, 1980; Hopp and Spearman, 1996; Little, 1961; Medhi, 1991). Note: Equations 7.1–7.4 hold for all queue types defined in Section 7.1.

7.4 EXTENSION OF LITTLE'S LAW

For the $M/G/1$ queue, Little's Law $L = \lambda W$ can be extended to higher moments. For the k th moment:

$$E[N(N-1)(N-2)\ldots(N-\mathrm{k}+1)] = \lambda^k E[T^k] \tag{7.5}$$

where

N = Number of customers in system
T = Time customer spends in system

Special cases:

$$k = 1: \qquad E[N] = \lambda E[T] \quad \text{(i.e., } L = \lambda W\text{)} \tag{7.6}$$

$$k = 2: \qquad E[N(N-1)] = \lambda^2 E[T^2] \tag{7.7}$$

Hence

$$Var[N] = \lambda E[T] + \lambda^2 Var[T] \tag{7.8}$$

(Cox and Smith, 1961; Gross and Harris, 1985).

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)	$\dfrac{(m\rho)^m \rho}{m!(1-\rho)}\left(\dfrac{1}{\dfrac{(m\rho)^m}{m!}+(1-\rho)\sum_{k=0}^{m-1}\dfrac{(m\rho)^k}{k!}}\right)$ (7.12) $\dfrac{\rho^{\sqrt{2(m+1)}}}{1-\rho}$ approx. (7.13)
$M/G/m$	$\dfrac{\rho^2}{1-\rho}\left(\dfrac{1+C_s^2}{2}\right)$ (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 interarrival times)

$C_s^2 = \mu^2\sigma_s^2$ = (Service rate of one server)^2 × (Variance of service times)

References for L_q and W_q results are given in the table in Section 7.7. For the case $m = 1$, Equations 7.9 and 7.10 are exact, Equation 7.11 is approximate. For general m, Equation 7.12 is exact, Equations 7.13, 7.14, and 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)	$\left\{ \begin{array}{l} \frac{(m\rho)^m}{\mu m m!(1-\rho)}\left(\frac{1}{\frac{(m\rho)^m}{m!}+(1-\rho)\sum_{k=0}^{m-1}\frac{(m\rho)^k}{k!}}\right) \\ \frac{\rho^{\sqrt{2(m+1)}-1}}{\mu m(1-\rho)} \quad \text{approx.} \end{array} \right.$
$M/G/m$	$\frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)\left(\frac{1+C_s^2}{2}\right)$ (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 interarrival times)

$C_s^2 = \mu^2\sigma_s^2$ = (Service rate of one server)^2 × (Variance of service times)

References for L_q and W_q results are given in the table in Section 7.7. For the case $m = 1$, Equations 7.16 and 7.17 are exact, Equation 7.18 is approximate. For general m, Equation 7.19 is exact, Equations 7.20, 7.21, and 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 given in Equation 7.19.

7.7 REFERENCES FOR FORMULAS FOR AVERAGE QUEUE LENGTH AND TIME IN QUEUE

(given in Equations (7.9)–(7.15) and Equations (7.16)–(7.22))

Queue Type	$m = 1$	General m
$M/M/m$	Allen, 1978 Cohen, 1985 Gillett, 1976 Gross and Harris, 1985 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, 1985 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, 1985 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) and Khintchine (1932). 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 an $G/G/1$ 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 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, 1965; Kleinrock, 1976; Larson and Odoni, 1981; Medhi, 1991; Tanner, 1995).

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 interarrival 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):

$$G/G/1: \quad W_q \cong \frac{1}{\mu}\left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2 + C_s^2}{2}\right) g\left(\rho, C_a^2, C_s^2\right) \tag{7.24}$$

where

$$g\left(\rho, C_a^2, C_s^2\right) = \begin{cases} \exp\left[-\dfrac{2(1-\rho)}{3\rho}\dfrac{\left(1-C_a^2\right)}{\left(C_a^2 + C_s^2\right)}\right] & \left(C_a^2 < 1\right) \\ \exp\left[-\dfrac{(1-\rho)\left(C_a^2 - 1\right)}{\left(C_a^2 + 4C_s^2\right)}\right] & \left(C_a^2 \geq 1\right) \end{cases}$$

For analyzing networks of queues, Whitt 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).

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, 1962, 1965; Kleinrock, 1976; Larson and Odoni, 1981; Medhi, 1991).

This heavy traffic approximation can also be applied to a queue with multiple severs in parallel (i.e., a $G/G/m$ queue). When traffic intensity $\rho = \dfrac{\lambda}{m\mu}$ 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; Kleinrock, 1976; Larson and Odoni, 1981; Medhi, 1991; Whitt, 1993).

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

$$C_d^2 \cong \rho^2 C_s^2 + \left(1 - \rho^2\right) C_a^2 \tag{7.28}$$

(Hopp and Spearman, 1996; Whitt, 1983).

7.12 DISTRIBUTION RESULTS FOR 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 $\Pr\{\text{number} = n\}$	$1-\rho^2 \quad (n=0)$ $(1-\rho)\rho^{n+1} \quad (n \geq 1)$	(7.31)	$(1-\rho)\rho^n$	(7.35)
Cumulative distribution $\Pr\{\text{number} \leq n\}$	$1-\rho^{n+2}$	(7.32)	$1-\rho^{n+1}$	(7.36)

where

$$\rho = \frac{\lambda}{\mu} = \frac{\text{Arrival rate}}{\text{Service rate}} = \text{Traffic intensity } (\rho < 1)$$

(Allen, 1978; Cohen, 1985; Cox and Smith, 1961; Hillier and Lieberman, 1980; Morse, 1958; Wolff, 1989). Note: the distribution of the number of customers in system, given by Equations 7.33–7.36, is a geometric distribution with parameter $1 - \rho$ (see Chapter 3, 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 $E[t]$	$\dfrac{\rho}{\mu(1-\rho)}$	(7.37)	$\dfrac{1}{\mu(1-\rho)}$	(7.41)
Variance $Var[t]$	$\dfrac{(2-\rho)\rho}{\mu^2(1-\rho)^2}$	(7.38)	$\dfrac{1}{\mu^2(1-\rho)^2}$	(7.42)
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

$$\rho = \frac{\lambda}{\mu} = \frac{\text{Arrival rate}}{\text{Service rate}} = \text{Traffic intensity } (\rho < 1)$$

and

$u_0(t)$ = is the unit impulse function centered at $t = 0$, given by

$$u_0(t) = \begin{cases} \infty & (t = 0) \\ 0 & (t \neq 0) \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} u_0(t)dt = 1$$

(Allen, 1978; Cohen, 1985; Cox and Smith, 1961; Hillier and Lieberman, 1980; Kleinrock, 1975; Papadopoulos, Heavy, and Browne, 1993). Note: the distribution of the time in system, given by Equations 7.41–7.44, is an exponential distribution with parameter $\mu(1 - \rho)$ (see Chapter 4, 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 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:

$$\text{Cycle time } c = \frac{1}{S} \tag{8.1}$$

$$MTBF = c \times MCBF = \frac{MCBF}{S} \tag{8.2}$$

$$MTBF = \frac{1}{\lambda} \tag{8.3}$$

$$MTTR = \frac{1}{\mu} \tag{8.4}$$

$$\text{Availability} = \frac{MTBF}{MTBF + MTTR} = \frac{1}{1 + \frac{\lambda}{\mu}} \tag{8.5}$$

$$\text{Throughput} = (\text{Availability} \times S) \tag{8.6}$$

(Buzacott, 1968; Choong and Gershwin, 1987; Cohen, 1985; Goldman and Slattery, 1964; Hopp and Spearman, 1996; Nahmias, 1989).

Note: *MCBF* is an average count of the number of jobs produced between failures, which is generally much easier to measure in practice than *MTBF*. From Equation 8.2, *MTBF* can be obtained without the need to measure the time 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 mean $1/\lambda$)
- repair times are independent and exponentially distributed (with mean $1/\mu$)

the mean and 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

$$Var[T_n] = \frac{2n\lambda}{S\mu^2} \tag{8.8}$$

respectively (Kim and Alden, 1997).

From Equations 8.7 and 8.8, the mean and variance of the time T_1 to produce one job ($n = 1$) are

$$E[T_1] = \frac{1}{S}\left(1 + \frac{\lambda}{\mu}\right)$$

and

$$Var[T_1] = \frac{2\lambda}{S\mu^2}$$

respectively. In the more general case where processing time is a random variable, with mean $1/S$ and 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

$$Var[T_1] = \frac{2\lambda}{S\mu^2} + \sigma^2\left(1 + \frac{\lambda}{\mu}\right)^2 \tag{8.9}$$

(Hopp and Spearman, 1996).

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
$u_0(t)$ is the unit impulse function centered at $t = 0$, given by

$$u_0(t) = \begin{cases} \infty & (t = 0) \\ 0 & (t \neq 0) \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} u_0(t)dt = 1$$

(Kim and Alden, 1997).

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*).

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; Gershwin, 1994).

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 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 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

$$P = \frac{1}{\frac{1}{S_1} + \frac{1}{S_2} - \frac{1}{S_1 + S_2}} \tag{8.12}$$

i.e.,

$$P = \frac{S_1 S_2 (S_1 + S_2)}{S_1^2 + S_1 S_2 + S_2^2} \tag{8.13}$$

In the special case of identical stations ($S_1 = S_2 = S$), the throughput result reduces to

$$P = \frac{2}{3} S \tag{8.14}$$

(Baker, 1992; Makino, 1964; Hunt, 1956).

8.5.2 Two Stations with 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 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

$$P = S_1 S_2 \left(\frac{S_1^{B+2} - S_2^{B+2}}{S_1^{B+3} - S_2^{B+3}} \right) \tag{8.15}$$

In the special case of identical stations ($S_1 = S_2 = S$), this throughput result reduces to

$$P = \left(\frac{B+2}{B+3} \right) S \tag{8.16}$$

(Hillier and Boling, 1966; Hunt, 1956).

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
$\overline{T} = 1/S$ be the mean processing time of a station

σ^2 be the variance of processing times
$C = \sigma/\bar{T}$ 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

an approximate estimate of the throughput of the line (average number of jobs per unit time) is given by Muth:

$$P \cong \frac{S}{1+\dfrac{1.67(N-1)C}{1+N+0.31C}} \tag{8.17}$$

(Baker, 1992; Blumenfeld, 1990; Muth, 1987).

For a line with buffers between the stations, where the buffer sizes are equal, an extension to Muth's approximation is given by

$$P \cong \frac{S}{1+\dfrac{1.67(N-1)C}{1+N+0.31C+1.67NB/(2C)}} \tag{8.18}$$

where B is the buffer size (Askin and Standridge, 1993; Blumenfeld, 1990).

9 Inventory Control

9.1 ECONOMIC ORDER QUANTITY (EOQ)

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 height between these two curves at any point in time is the inventory level. Figure 9.2 plots this inventory level, which displays the classic sawtooth pattern over time. 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.

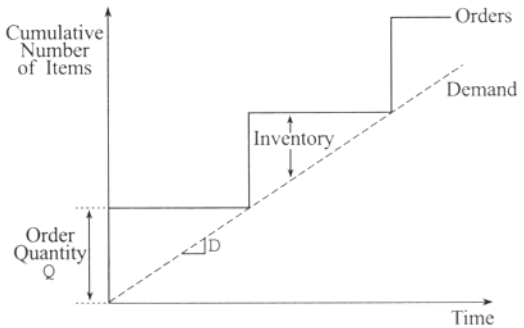

FIGURE 9.1 Cumulative orders and demand over time in EOQ model.

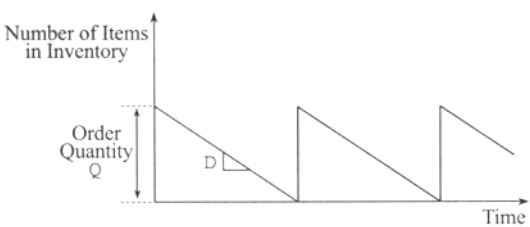

FIGURE 9.2 Inventory level over time in EOQ model.

Total cost per unit time $C(Q)$ is given by

$$\begin{aligned} C(Q) &= \text{Inventory Cost} + \text{Ordering Cost} \\ &= \frac{HQ}{2} + \frac{AD}{Q} \end{aligned} \tag{9.1}$$

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

$$Q^* \sqrt{\frac{2AD}{H}} \tag{9.2}$$

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; Cohen, 1985; Harris, 1913; Hax and Candea, 1984; Hopp and Spearman, 1996; Nahmias, 1989; Stevenson, 1986; Tersine, 1985; Wilson, 1934; Woolsey and Swanson, 1975).

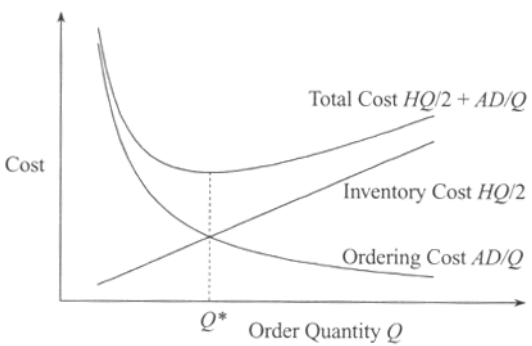

FIGURE 9.3 Trade-off between inventory and ordering costs in EOQ model.

9.2 ECONOMIC PRODUCTION QUANTITY (EPQ)

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 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 set-up)
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 height 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.

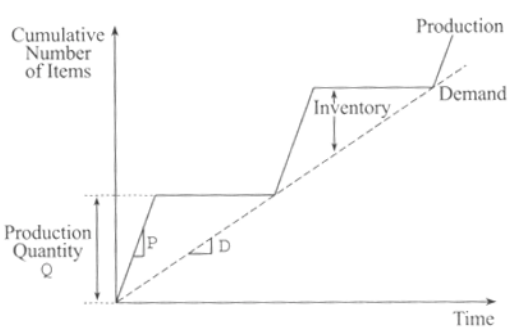

FIGURE 9.4 Cumulative production and demand over time in EPQ model.

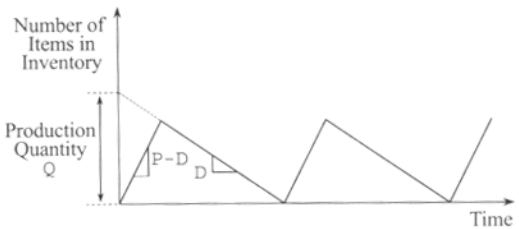

FIGURE 9.5 Inventory level over time in EPQ model.

Total cost per unit time $C(Q)$ is given by

$$\begin{aligned} C(Q) &= \text{Inventory Cost} + \text{Production Set-Up Cost} \\ &= \frac{H(1-D/P)Q}{2} + \frac{AD}{Q} \end{aligned} \tag{9.3}$$

The optimal quantity Q^* to produce (i.e., the production quantity that minimizes total cost) is given by

$$\frac{d}{dQ}C(Q) = 0$$

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; Hopp and Spearman, 1996; Nahmias, 1989; Stevenson, 1986; Taft, 1918; Tersine, 1985).

9.3 "NEWSBOY PROBLEM": OPTIMAL INVENTORY TO MEET UNCERTAIN DEMAND IN A SINGLE PERIOD

The optimal number of items to hold in inventory to meet uncertain demand in a single period is given by the trade-off between

- cost of holding too many items and
- cost of 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

$$\begin{aligned} C(Q) &= c_0 E[\text{number of items over}] + c_s E[\text{number of items short}] \\ &= c_0 \int_0^Q (Q-x) f(x)dx + C_s \int_Q^\infty (x-Q) f(x)dx \end{aligned} \tag{9.5}$$

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 13.8, Equation 13.23),

$$\frac{d}{dQ}C(Q) = c_0\int_0^Q \frac{\partial}{\partial Q}\{(Q-x)f(x)\}dx + c_s\int_Q^\infty \frac{\partial}{\partial Q}\{(x-Q)f(x)\}dx$$

$$= c_0\int_0^Q f(x)dx - c_s\int_Q^\infty f(x)dx$$

$$= c_0F(Q) - c_s\{1-F(Q)\}$$

$$= (c_0 + c_s)F(Q) - c_s$$

Hence, setting $\frac{d}{dQ}C(Q) = 0$, the optimal quantity Q^* is given by

$$F(Q^*) = \frac{c_s}{c_s + c_0} \tag{9.6}$$

Equation 9.6 is the solution to the classic "Newsboy 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

$$F(Q^*) = \frac{a + p - c}{a + p - v} \tag{9.7}$$

(Hannsmann, 1962; Hopp and Spearman, 1996; Nahmias, 1989; Ravindran, Phillips, and Solberg, 1987).

9.4 INVENTORY REPLENISHMENT POLICIES

Figures 9.6, 9.7, 9.8, and 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*.
  (*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 inventory position is at level *s* or below, an order is placed for a sufficient quantity to bring inventory position up to a given level *S*; if 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; Hadley and Whitin, 1963; Hax and Candea, 1984; Johnson and Montgomery, 1974; Silver, Pyke, and Peterson, 1998).

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)

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 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; in that case, the review of inventory can be based simply on the inventory on hand. (Evans et al., 1984; Johnson and Montgomery, 1974.)

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; Silver, Pyke and Peterson, 1998).

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; Silver, Pyke, and Peterson, 1998).

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.

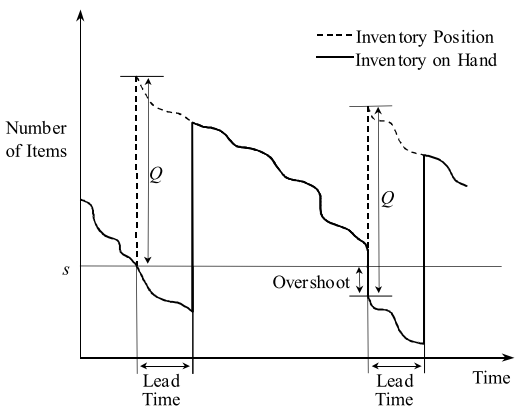

FIGURE 9.6 Inventory pattern over time in (s, Q) policy.

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; McClain and Thomas, 1985; Silver, Pyke and Peterson, 1998; Sipper and Bulfin, 1997; Stevenson, 1986).

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 mean DL and standard deviation $\sqrt{L\sigma_D^2 + D^2\sigma_L^2}$ (Hadley and Whitin, 1963).

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). 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 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 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

$$\beta = 1 - \frac{\sigma_l}{Q}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}k^2\right) - k\left[1 - \Phi(k)\right]\right\} \tag{9.15}$$

where $\Phi(k)$ is the cumulative distribution function of the standard normal distribution, i.e.,

$$\Phi(k) = \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right) dx \tag{9.16}$$

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; Nahmias, 1989; Schneider, 1981; Sipper and Bulfin, 1997).

To ensure a high value of the probability α, the service level factor k is typically set in the range 2–3. From Equation 9.16, when $k = 2$, $\alpha = 97.7\%$, and when $k = 3$, $\alpha = 99.9\%$ (Lewis, 1970; Mood, Graybill, and Boes, 1974; Sipper and Bulfin, 1997).

The relationship between the fill rate β and 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 given shortage cost per item, s and Q can also be optimized jointly (Hadley and Whitin, 1963; Nahmias, 1989; Sipper and Bulfin, 1997).

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.

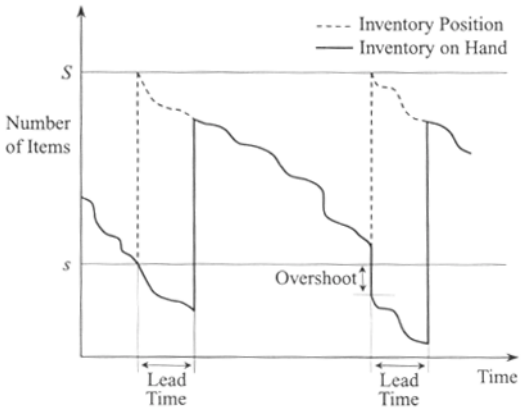

FIGURE 9.7 Inventory pattern over time in (*s*, *S*) policy.

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 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 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; Silver, Pyke, and Peterson, 1998).

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 mean DL and standard deviation $\sqrt{L\sigma_D^2 + D^2\sigma_L^2}$ (Hadley and Whitin, 1963).

The service level factor k in 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 a heuristic estimate based simply on the reorder point plus the EOQ, 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.

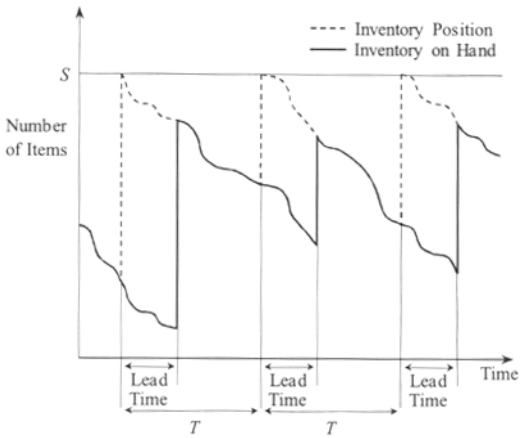

FIGURE 9.8 Inventory pattern over time in (*T*, *S*) policy.

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 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 the inventory position in each review period. 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 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; Lewis, 1970; McClain and Thomas, 1985; Silver, Pyke, and Peterson, 1998; Sipper and Bulfin, 1997).

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

$$S = D(L+T) + k\sigma_D\sqrt{(L+T)} \tag{9.23}$$

The review period T, given by Equation 9.21, is determined from the EOQ, given in Section 9.1. For 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, $k\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2}$, is the additional inventory (i.e., safety stock) needed to avoid stocking out due to 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 demand during the lead time plus review period (rather than just the lead time, as in the (s, Q) policy). The demand during the lead time plus review period has mean $D(L + T)$ and standard deviation $k\sqrt{(L+T)\sigma_D^2 + D^2\sigma_L^2}$ (Tijms and Groenevelt, 1984).

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).

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 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.*

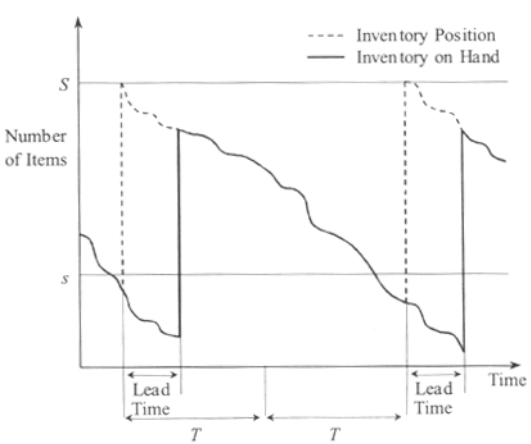

FIGURE 9.9 Inventory pattern over time in (T, s, S) policy.

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 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 the inventory position in each review period. 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 $\sigma_D^2\, t$.

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; Tijms and Groenevelt, 1984).

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 demand during the lead time plus review period. This expression for s is based on the result that demand during the lead time plus review period has mean $D(L + T)$ and 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 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).

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 are illustrated here for a single link between one origin and one destination.

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

$\bar{I}_1$, $\bar{I}_2$, $\bar{I}_3$, and $\bar{I}_4$

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

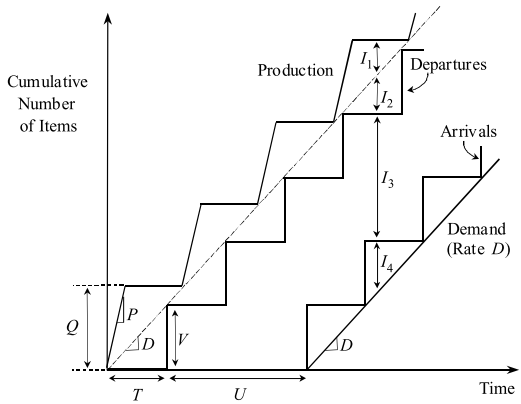

FIGURE 9.10 Cumulative production, shipments, and demand over time (Figure modified from Blumenfeld et al., 1985, p. 370).

rate is P (where $P > D$ to ensure 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.)

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 $\bar{I}_1$, $\bar{I}_2$, $\bar{I}_3$, and $\bar{I}_4$ are given by

$$\bar{I}_1 = \frac{Q}{2}\left(1 - \frac{D}{P}\right) \tag{9.41}$$

$$\bar{I}_2 = \frac{V}{2} \tag{9.42}$$

$$\bar{I}_3 = DU \tag{9.43}$$

$$\bar{I}_4 = \frac{V}{2} \tag{9.44}$$

(Blumenfeld et al., 1985; Hall, 1996).

Equation 9.41 for the average production cycle inventory $\bar{I}_1$ is the same as the expression for average inventory in the EPQ model, given in Section 9.2. The total number of items in 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 9.42–9.44 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 A 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; Daganzo, 1991; Love, 1979).

Cumulative plots also have applications in areas closely related to inventory control, such as queueing theory (Newell, 1982; Medhi, 1991) and transportation and traffic flow analysis (Daganzo, 1997; Newell, 1993). Cumulative plots have long been used in hydraulic engineering for determining reservoir capacity (Linsley and Franzini, 1955).

10 Distance Formulas for Logistics Analysis

10.1 "TRAVELING SALESMAN PROBLEM" TOUR DISTANCE: SHORTEST PATH THROUGH A SET OF POINTS IN A REGION

The average tour distance d on the shortest closed path connecting n points randomly distributed within a region of area A (see Figures 10.1 and 10.2) is given approximately by

$$d \cong K\sqrt{nA} \tag{10.1}$$

where K is a constant. Based on simulation experiments, the value for the constant K is generally taken as $K = 0.75$ (Beardwood, Halton, and Hammersley, 1959; Eilon, Watson-Gandy, and Christofides, 1971; Larson and Odoni, 1981; Stein, 1978).

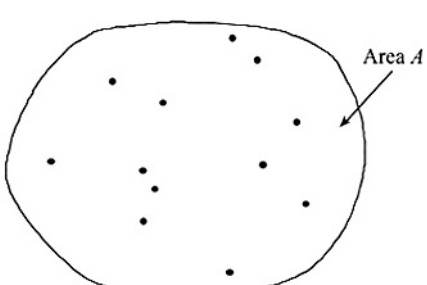

FIGURE 10.1 Points randomly distributed within a region of area A.

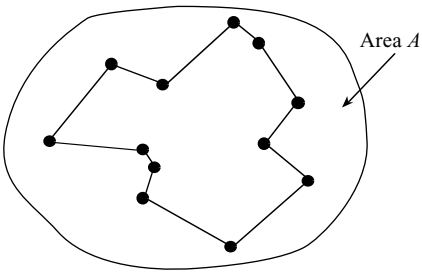

FIGURE 10.2 Shortest closed path through points in Figure 10.1.

This formula is useful in logistics for developing delivery strategies, planning transportation service requirements, and evaluating "traveling salesman problem" algorithms (Burns et al., 1985; Daganzo, 1991; Larson and Odoni, 1981; Stein, 1978). The approximation holds well for regions of various shapes (Christofides and Eilon, 1969; Eilon, Watson-Gandy, and Christofides, 1971).

The formula is derived for large n, but also provides an approximation when n is small (Daganzo, 1984). 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 10.1 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 10.4).

10.2 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 10.3

$f(x)$ = probability density function of x

$F(x) = \int_0^x f(u)du$ = cumulative distribution function of x

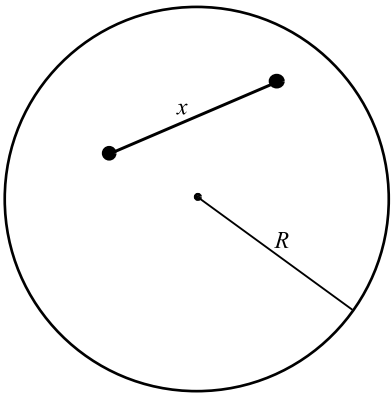

FIGURE 10.3 Distance x between two random points in a circle of radius R.

$$E(x) = \int_0^{2R} x f(x)\,dx \;= \text{average distance}$$

$$= \sqrt{\int_0^{2R} x^2 f(x)\,dx - \{E(x)\}^2} \;= \text{standard deviation of } x.$$

The probability density function $f(x)$ is given by

$$f(x) = \frac{2x}{\pi R^2}\;2\cos^{-1}\frac{x}{2R} - \frac{x}{R}\sqrt{1-\frac{x^2}{4R^2}} \qquad (0 \le x \le 2R) \qquad (10.2)$$

(Garwood, 1947; Garwood and Tanner, 1958; Fairthorne, 1965; Kendall and Moran, 1963; Vaughan, 1987). The probability density function $f(x)$ is plotted in Figure 10.4.

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

$$F(x) = 1 + \frac{2}{\pi}\left(\frac{x^2}{R^2} - 1\right)\cos^{-1}\left(\frac{x}{2R}\right) - \frac{x}{\pi R}\left(1+\frac{x^2}{2R^2}\right)\sqrt{1-\frac{x^2}{4R^2}} \quad (0 \le x \le 2R) \qquad (10.3)$$

(Borel, 1925; Garwood, 1947).

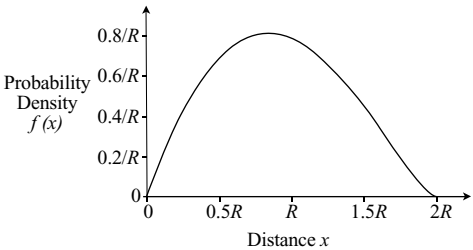

FIGURE 10.4 Probability density function $f(x)$ for distance x between two random points in a circle of radius R.

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 10.5.

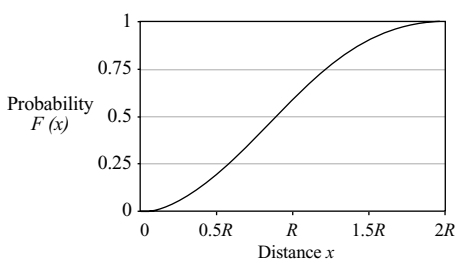

FIGURE 10.5 Cumulative distribution function $F(x)$ for distance x between two random points in a circle of radius R.

The average distance $E(x)$ is given by

$$E(x) = \frac{128R}{45\pi} \quad (\cong 0.9054R) \tag{10.4}$$

(Apsimon, 1958; Eilon, Watson-Gandy and Christofides, 1971; Garwood and Tanner, 1958; Fairthorne, 1965; Smeed, 1971; Vaughan, 1987).

The standard deviation is given by

$$\sigma = \sqrt{R^2 - \left(\frac{128R}{45\pi}\right)^2} \quad (\cong 0.4245R) \tag{10.5}$$

(Garwood and Tanner, 1958)

Note: 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; Larson and Odoni, 1981; Smeed, 1967).

10.3 AVERAGE RECTANGULAR GRID DISTANCE BETWEEN TWO RANDOM POINTS IN A CIRCLE

Average distance d_{grid} on a rectangular grid between two random points in a circle of radius R (see Figure 10.6) is given by

$$d_{grid} = \frac{4}{\pi} d_{direct} \tag{10.6}$$

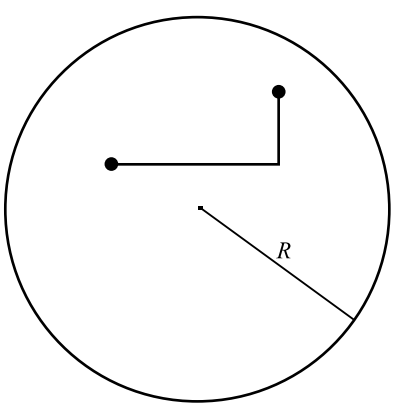

FIGURE 10.6 Rectangular grid distance between two random points in a circle of radius R.

where d_{direct} is the average direct (Euclidean) distance between the two points. From Equation 10.4

$$d_{direct} = \frac{128R}{45\pi}$$

Hence

$$d_{grid} = \frac{512R}{45\pi^2} \quad (\cong 1.1528R) \tag{10.7}$$

(Eilon, Watson-Gandy, and Christofides, 1971; Fairthorne, 1955; Vaughan, 1987).

10.4 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

$$D = R\theta \tag{10.8}$$

where

$$R = \text{mean radius of the earth } (\cong 3960 \text{ miles or } 6370 \text{ km}) \tag{10.9}$$

and

$$\theta = \cos^{-1}\{\sin\alpha_1 \sin\alpha_2 + \cos\alpha_1 \cos\alpha_2 \cos(\beta_1 - \beta_2)\}, \tag{10.10}$$

with angle θ expressed in radians (1 degree $= \pi/180$ radians). (Jennings, 1994; Melzak, 1983).

Note that 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 10.10 is based on the *law of cosines for sides* for a spherical triangle. Under the assumption that the earth is a sphere, the 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 vertex angle at the North Pole given by $\beta_1 - \beta_2$, as shown in Figure 10.7. The law of cosines for sides for this spherical triangle gives

$$\cos\theta = \cos(90 - \alpha_1)\cos(90 - \alpha_2) + \sin(90 - \alpha_1)\sin(90 - \alpha_2)\cos(\beta_1 - \beta_2)$$

(Jennings, 1994; Melzak, 1983; Thurston, 1997). Since $\cos(90 - \alpha_1) = \sin\alpha_1$, $\sin(90 - \alpha_1) = \cos\alpha_1$, etc., this law yields the result given by Equation 10.10.

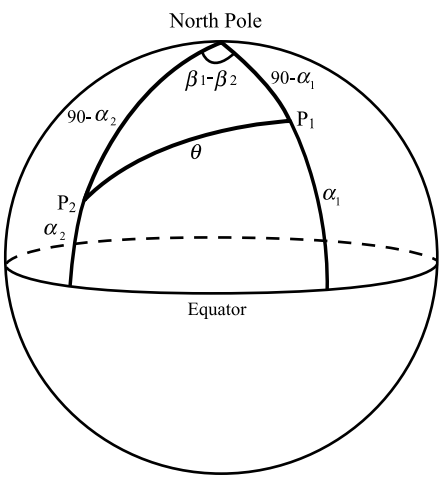

FIGURE 10.7 Spherical triangle formed by points P_1, P_2, and the North Pole (with arc angles $90\text{-}\alpha_1$, $90\text{-}\alpha_2$, and θ, and vertex angle $\beta_1\text{-}\beta_2$ at the North Pole).

The formula for θ given by Equation 10.10 can also be written as

$$\theta = 2\sin^{-1}\sqrt{\sin^2\frac{\alpha_1 - \alpha_2}{2} + \cos\alpha_1\cos\alpha_2\sin^2\frac{\beta_1 - \beta_2}{2}} \qquad (10.11)$$

(Simmons, 1945; Sinnott, 1984).

This version of the formula is mathematically the same as Equation 10.10, 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* (Simmons,1945):

$$\text{hav}(\theta) = \text{hav}(\alpha_1 - \alpha_2) + \cos(\alpha_1)\cos(\alpha_2)\,\text{hav}(\beta_1 - \beta_2)$$

where, for any angle a, hav(a) is the haversine of a and is given by

$$\text{hav}(a) = \tfrac{1}{2}\left[1 - \cos(a)\right] = \sin^2\left(\tfrac{1}{2}a\right)$$

11 Linear Programming Formulations

11.1 GENERAL FORMULATION

Let $x_1, x_2, \ldots, x_N$ be N variables in a linear programming problem. The problem is to find the values of the variables $x_1, x_2, \ldots, x_N$ 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

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{11.1}$$

subject to the constraints

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &\le b_2 \\ &\vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &\le b_M \end{aligned} \tag{11.2}$$

and

$$x_1 \ge 0, x_2 \ge 0, \ldots, x_N \ge 0 \tag{11.3}$$

where a_{ij}, b_i, c_j ($i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, N$) are constants.

(Chvátal, 1983; Gass, 1964; Hillier and Lieberman, 1980; Ignizio, 1982; Munakata, 1979; Ozan, 1986; Vanderbei, 1997; Wagner, 1969).

In matrix notation, Equations 11.1–11.3 for the general formulation are written as:

$$\text{Maximize } Z = \mathbf{c}^{\mathbf{T}}\mathbf{x} \tag{11.4}$$

subject to the constraints

$$\mathbf{Ax} \leq \mathbf{b} \tag{11.5}$$

and

$$\mathbf{x} \geq \mathbf{0} \tag{11.6}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \tag{11.7}$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{11.8}$$

and where $\mathbf{c}^{\mathbf{T}}$ denotes the transpose of the vector $\mathbf{c}$.

11.2 TERMINOLOGY

The following terms are commonly used in linear programming:

- *Decision variables*: variables $x_1, x_2, \ldots, x_N$ in Equation 11.1
- *Objective function*: function Z given by Equation 11.1
- *Objective function coefficients*: constants $c_1, c_2, \ldots, c_N$ in Equation 11.1
- *Constraint coefficients*: constants a_{ij} in Equation 11.2
- *Non-negativity constraints*: constraints given by Equation 11.3

- *Feasible solution*: set of $x_1, x_2, \ldots, x_N$ values 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 11.1).

11.3 EXAMPLE OF FEASIBLE REGION

The feasible region can be shown graphically in the case of two decision variables ($N = 2$). Figure 11.1 illustrates the feasible region for the following example:

$$\text{Maximize } Z = 2x_1 + x_2 \tag{11.9}$$

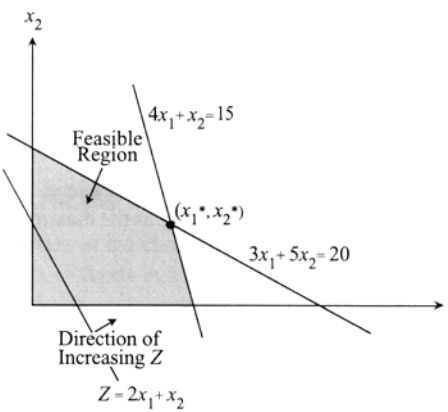

FIGURE 11.1 Feasible region for linear programming example.

subject to the constraints

$$\begin{aligned} 3x_1 + 5x_2 &\le 20 \\ 4x_1 + x_2 &\le 15 \end{aligned} \tag{11.10}$$

and

$$x_1 \ge 0, \; x_2 \ge 0 \tag{11.11}$$

The optimal solution Z^* (i.e., maximum Z) occurs at the point $\left(x_1^*, x_2^*\right)$ in Figure 11.1. For this example, the point $\left(x_1^*, x_2^*\right) = \left(\frac{55}{17}, \frac{35}{17}\right)$, giving $Z^* = \frac{145}{17}$.

11.4 ALTERNATIVE FORMULATIONS

Linear programming problems in which

(a) the objective function is to be minimized (rather than maximized) or
(b) the constraints contain equalities (= rather than ≤) or
(c) the constraints contain inequalities with the sign reversed (≥ rather than ≤)

can be reformulated in terms of the general formulation given by Equations 11.1–11.3.

11.4.1 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; Hillier and Lieberman, 1980; Vanderbei, 1997; Wagner, 1969).

11.4.2 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); Hillier and Lieberman, 1980; Vanderbei, 1997; Wagner, 1969).

11.4.3 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; Munakata, 1979; Wagner, 1969).

11.5 DIET PROBLEM

The *diet problem* arises in the choice of foods for a healthy diet. The problem is to determine the mix of foods in a diet that minimizes 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

$$\text{Minimize } C = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{11.12}$$

subject to the constraints

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &\geq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &\geq b_2 \\ &\vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &\geq b_M \end{aligned} \tag{11.13}$$

and

$$x_1 \geq 0,\ x_2 \geq,\ \ldots,\ x_N \geq 0 \tag{11.14}$$

where a_{ij}, b_i, c_j (i = 1, 2, …, M; j = 1, 2, …, N) are constants (Gass, 1964; Luenberger, 1984; Spivey and Thrall, 1970; Vanderbei, 1997).

11.6 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 11.1–11.3 above, the corresponding dual problem is to find the values of the M variables y_1, y_2, …, y_M to solve the following:

$$\text{Minimize } V = b_1y_1 + b_2y_2 + \cdots + b_My_M \tag{11.15}$$

subject to the constraints

$$\left.\begin{aligned} a_{11}y_1 + a_{21}y_2 + \cdots + a_{M1}y_M &\geq c_1 \\ a_{12}y_1 + a_{22}y_2 + \cdots + a_{M2}y_M &\geq c_2 \\ &\vdots \\ a_{1N}y_1 + a_{2N}y_2 + \cdots + a_{MN}y_M &\geq c_N \end{aligned}\right\} \tag{11.16}$$

and

$$y_1 \geq 0,\ y_2 \geq 0,\ \ldots,\ y_M \geq 0 \tag{11.17}$$

In matrix notation, the primal and dual problems are formulated as:

Primal			***Dual***		
Maximize	$Z = \mathbf{c}^{\mathbf{T}}\mathbf{x}$		Minimize	$V = \boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{y}$	
subject to	$\mathbf{A}\mathbf{x} \le \mathbf{b}$	(11.18)	subject to	$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{y} \ge \boldsymbol{c}$	(11.19)
and	$\mathbf{x} \ge \mathbf{0}$		and	$\boldsymbol{y} \ge \boldsymbol{0}$	

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix}, \tag{11.20}$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{11.21}$$

and where $\mathbf{c}^{\mathbf{T}}$ 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; Dantzig, 1963; Gass, 1964; Hillier and Lieberman, 1980; Ignizio, 1982; Munakata, 1979; Vanderbei, 1997; Wagner, 1969).

11.7 SPECIAL CASES OF LINEAR PROGRAMMING PROBLEMS

11.7.1 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, ..., m)
d_j = number of units of demand at destination j (j = 1, 2, ..., 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 this 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, …, m; j = 1, 2, …, n) to minimize the total shipping cost, C.

The linear programming formulation for the transportation problem is

$$\text{Minimize } C = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{11.22}$$

subject to the constraints

(11.23)

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} &= s_i \quad (i = 1,2,...,m) \\ \sum_{i=1}^{m} x_{ij} &= d_j \quad (j = 1,2,...,n) \end{aligned}\right\}$$

and

$$x_{ij} \geq 0 \ (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n) \tag{11.24}$$

where s_i and d_j ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) are constants (Chvátal, 1983; Dantzig, 1951b; Gass, 1964; Hillier and Lieberman, 1980; Ignizio, 1982; Luenberger, 1984; Ozan, 1986; Spivey and Thrall, 1970; Vanderbei, 1997; Wagner, 1969).

The above formulation requires that total supply is equal to total demand, i.e.,

$$\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j \tag{11.25}$$

In the more general case, total supply may be greater total demand, i.e.,

$$\sum_{i=1}^{m} s_i \geq \sum_{j=1}^{n} d_j$$

For this case, the problem can be reformulated in terms of an equality as in Equation 11.25 by adding a fictitious destination with demand equal to the difference between total supply and 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; Wagner, 1969).

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; therefore, there is no need to specify an integer constraint on x_{ij} (Luenberger, 1984; Spivey and Thrall, 1970; Wagner, 1969). This result is known as the *integrality property* (Ahuja, Magnanti, and Orlin, 1993; Chvátal, 1983; Vanderbei, 1997).

The constraints given by Equation 11.23 are a special case of the general constraints given by Equation 11.2. 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:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \tag{11.26}$$

as shown in Equation 11.7. For the transportation problem with mn variables and $(m + n)$ constraints, the constraint coefficients are given by the special $(m + n) \times mn$ matrix:

$$\begin{array}{l} \textit{Coefficients of} \\ x_{11}\ x_{12}\ \cdots\ x_{1n}\ x_{21}\ x_{22}\ \cdots\ x_{2n}\ \cdots\ x_{m1}\ x_{m2}\ \cdots\ x_{mn} \\ A = \begin{pmatrix} 1 & 1 & \cdots & 1 & & & & & & & & \\ & & & & 1 & 1 & \cdots & 1 & & & & \\ & & & & & & & & \ddots & & & \\ & & & & & & & & 1 & 1 & \cdots & 1 \\ 1 & & & & 1 & & & & 1 & & & \\ & 1 & & & & 1 & & & & 1 & & \\ & & \ddots & & & & \ddots & & & & \ddots & \\ & & & 1 & & & & 1 & & & & 1 \end{pmatrix} \end{array} \tag{11.27}$$

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 shown in Equation 11.27. 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 11.1 shows the correspondence between the general linear programming formulation and the transportation problem formulation, based on the notation used in Equations 11.1–11.3 and 11.22–11.24.

TABLE 11.1

Correspondence between General and Transportation Problem Formulations

	General Linear Programming Problem Eqns. 11.1–11.3	Transportation Problem Eqns. 11.22–11.24
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} given by Equation 11.26	0 or 1 in pattern given by Equation 11.27

11.7.2 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 can be 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 by 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. (Hillier and Lieberman, 1980; Ignizio, 1982; Ozan, 1986; Wagner, 1969).

11.7.3 Assignment Problem

The *assignment problem* is a special case of the transportation problem, with specific values for the decision variables and 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 this 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, …, n; j = 1, 2, …, n) to minimize the total assignment cost, C.

The linear programming formulation for the assignment problem is:

$$\text{Minimize } C = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{11.28}$$

subject to the constraints

$$\left.\begin{aligned}\sum_{j=1}^{n} x_{ij} &= 1 \quad (i = 1,2,\ldots,n)\\ \sum_{i=1}^{n} x_{ij} &= 1 \quad (j = 1,2,\ldots,n)\end{aligned}\right\} \tag{11.29}$$

and

$$x_{ij} \geq 0 \ (i = 1, 2, \ldots, n; j = 1, 2, \ldots, n) \tag{11.30}$$

(Bradley, Hax, and Magnanti, 1977; Gass, 1964; Hillier and Lieberman, 1980; Ignizio, 1982; Luenberger, 1984; Ozan, 1986; Spivey and Thrall, 1970; Vanderbei, 1997; Wagner, 1969).

The assignment problem given by Equations 11.28–11.30 is a special case of the transportation problem given by Equations 11.22–11.24 in which

$$m = n$$

$$s_i = 1 \ (i = 1, 2, \ldots, n)$$

$$d_j = 1 \ (j = 1, 2, \ldots, n)$$

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 has an optimal solution in which each x_{ij} is 0 or 1 (Bradley, Hax, and Magnanti, 1977; Luenberger, 1984; Spivey and Thrall, 1970). This result is an example of the *integrality property* (Ahuja, Magnanti, and Orlin, 1993; Chvátal, 1983; Vanderbei, 1997).

11.8 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

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{11.31}$$

subject to the constraints

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &\le b_2 \\ &\vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N &\le b_M \end{aligned}\right\} \tag{11.32}$$

$$x_1 \ge 0,\ x_2 \ge 0,\ \ldots,\ x_N \ge 0 \tag{11.33}$$

and

$$x_j \text{ integer (for some or all } j = 1, 2, \ldots, N) \tag{11.34}$$

where a_{ij}, b_i, c_j ($i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, N$) are constants (Bradley, Hax, and Magnanti, 1977; Garfinkel and Nemhauser, 1972; Nemhauser and Wolsey, 1988).

An example of the general formulation for an integer linear programming problem is the *diet problem* given in Section 11.5, 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*.

11.8.1 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 $x_1, x_2, \ldots, x_N$ to maximize the total value Z of items included in knapsack.

The integer linear programming formulation for the knapsack problem is

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{11.35}$$

subject to the constraints

$$a_1x_1 + a_2x_2 + \cdots + a_Nx_N \le b \tag{11.36}$$

$$x_1 \ge 0,\ x_2 \ge 0,\ \ldots,\ x_N \ge 0 \tag{11.37}$$

and

$$x_1, x_2, \ldots, x_N \text{ integers} \tag{11.38}$$

where a_j, b, c_j (j = 1, 2, …, N) are constants.

The knapsack problem given by Equations 11.35–11.38 is a special case of the integer linear programming problem given by Equations 11.31–11.34, in which the set of constraints in Equation 11.32 reduces to only one constraint (i.e., M = 1), with $a_{ij} = a_j$, and $b_i = b$.

If no more than one of any item type may be included in the knapsack, then the integer constraint given by Equation 11.38 is replaced by the constraint

$$x_j = \begin{cases} 1 & \text{if item type } j \text{ is included in the knapsack} \\ 0 & \text{otherwise} \end{cases} \qquad (j = 1, 2, \ldots, N) \tag{11.39}$$

and the problem is called a *0–1 knapsack problem.*

(Bradley, Hax, and Magnanti, 1977; Garfinkel and Nemhauser, 1972; Murty, 1976; Ozan, 1986).

11.8.2 Traveling Salesman Problem

The *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 ($i, j = 1, 2, ..., n$)

$$x_{ij} = \begin{cases} 1 & \text{if tour leads directly from location } i \text{ to location } j \\ 0 & \text{otherwise} \end{cases} \qquad (i, j = 1, 2, ..., n)$$

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

$$\text{Minimize } C = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{11.40}$$

subject to the constraints

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1 \qquad (i = 1, 2, ..., n) \\ \sum_{i=1}^{n} x_{ij} &= 1 \qquad (j = 1, 2, ..., n) \end{aligned}\right\} \tag{11.41}$$

$$x_{ij} \geq 0 \ (i = 1, 2, \ldots, n; j = 1, 2, \ldots, n) \tag{11.42}$$

$$x_{ij} \text{ integers } (i = 1, 2, \ldots, n; j = 1, 2, \ldots, n) \tag{11.43}$$

and

$$\sum_{i \in S} \sum_{j \in S} x_{ij} \geq 1 \tag{11.44}$$

where S is any nonempty subset of the n locations (Bradley, Hax, and Magnanti, 1977; Garfinkel and Nemhauser, 1972).

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 constraints

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad (i = 1, 2, ..., n)$$

in Equation 11.41 ensure that the tour leaves each location exactly once. The constraints

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad (j = 1, 2, ..., n)$$

in Equation 11.41 ensure that the tour enters each location exactly once.

The set of constraints for subsets S, given by Equation 11.44, ensures that the tour connects all n locations. Without such constraints, the problem formulation would allow for disconnected subtours, which are not valid. Figure 11.2 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

$$u_i - u_j + nx_{ij} \leq n - 1 \ (i = 2, 3, \ldots, n; j = 2, 3, \ldots, n; i \neq j) \qquad (11.45)$$

where u_i ($i = 2, 3, \ldots, n$) are arbitrary real numbers (Gass, 1964; Murty, 1976; Wagner, 1969).

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).

11.9 SOLUTION METHODS

11.9.1 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

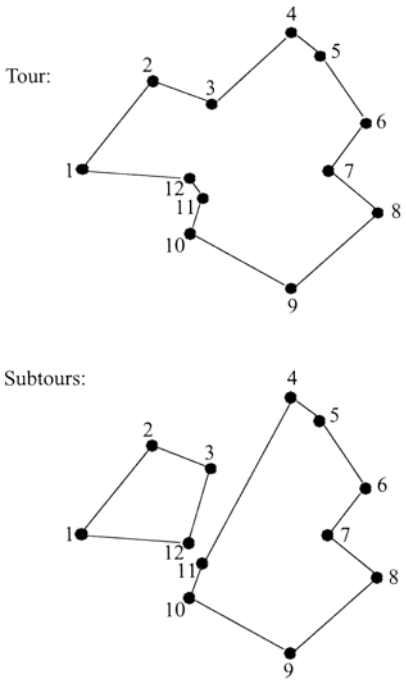

FIGURE 11.2 Tour and subtours in traveling salesman problem.

is reached (Chvátal, 1983; Cohen, 1985; Dantzig, 1951a, 1963; Gass, 1964; Hillier and Lieberman, 1980; Karloff, 1991; Munakata, 1979; Vanderbei, 1997; Wagner, 1969).

11.9.2 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; Karmarkar, 1984; Mehrotra, 1992; Wright, 1997).

11.9.3 Network Flow Methods

A technique for solving the above special cases of linear programming problems (transportation problem, transshipment problem, and 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; Bradley, Hax, and Magnanti, 1977; Ignizio, 1982).

11.9.4 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 noninteger solutions without eliminating feasible integer solutions (Garfinkel and Nemhauser, 1972; Nemhauser and Wolsey, 1988; Ozan, 1986).

11.9.5 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; Garfinkel and Nemhauser, 1972; Hillier and Lieberman, 1980; Murty, 1976; Nemhauser and Wolsey, 1988; Ozan, 1986).

12 Mathematical Functions

12.1 GAMMA FUNCTION

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx \tag{12.1}$$

$$\Gamma(n + 1) = n\Gamma(n) \tag{12.2}$$

For n a postive integer:

$$\Gamma(n + 1) = n! \tag{12.3}$$

(Abramowitz and Stegun, 1968; Binmore, 1983).

12.2 BETA FUNCTION

$$B(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \tag{12.4}$$

$$= \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} \tag{12.5}$$

For m and n positive integers:

$$B(m, n) = \frac{(m-1)!(n-1)!}{(m+n-1)!} \tag{12.6}$$

(Abramowitz and Stegun, 1968).

12.3 UNIT IMPULSE FUNCTION

$$u_0(t) = \begin{cases} \infty & (t = 0) \\ 0 & (t \neq 0) \end{cases} \tag{12.7}$$

and

$$\int_{-\infty}^{\infty} u_0(t)dt = 1 \tag{12.8}$$

The unit impulse function is also known as the Dirac delta function. (Kleinrock, 1975; Poularikas and Seely, 1985).

12.4 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:

$$I_0(z) = 1 + \frac{\frac{1}{4}z^2}{(1!)^2} + \frac{\left(\frac{1}{4}z^2\right)^2}{(2!)^2} + \frac{\left(\frac{1}{4}z^2\right)^3}{(3!)^2} + \frac{\left(\frac{1}{4}z^2\right)^4}{(4!)^2} + \cdots \tag{12.9}$$

$$I_1(z) = \frac{d}{dz} I_0(z) \tag{12.10}$$

$$I_1(z) = \frac{1}{2} z \left\{ 1 + \frac{\frac{1}{4}z^2}{1!2!} + \frac{\left(\frac{1}{4}z^2\right)^2}{2!3!} + \frac{\left(\frac{1}{4}z^2\right)^3}{3!4!} + \frac{\left(\frac{1}{4}z^2\right)^4}{4!5!} + \cdots \right\} \tag{12.11}$$

(Abramowitz and Stegun, 1968).

12.5 STIRLING'S FORMULA

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n} \tag{12.12}$$

(Abramowitz and Stegun, 1968; Binmore, 1977; Feller, 1964).

13 Calculus Results

13.1 BASIC RULES FOR DIFFERENTIATION

(a) Differentiation of a Sum of Functions

$$\frac{d}{dx}(u+v)=\frac{du}{dx}+\frac{dv}{dx} \tag{13.1}$$

where $u = u(x)$ and $v = v(x)$ are differentiable functions of x.

(b) Differentiation of a Function and Constant Multiple

$$\frac{d}{dx}(ku)=k\frac{du}{dx} \tag{13.2}$$

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)

$$\frac{d}{dx}(uv)=v\frac{du}{dx}+u\frac{dv}{dx} \tag{13.3}$$

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 differentiation of a product of three functions

$$\frac{d}{dx}(uvw)=vw\frac{du}{dx}+uw\frac{dv}{dx}+uv\frac{dw}{dx} \tag{13.4}$$

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)

$$\frac{d}{dx}\left(\frac{u}{v}\right)=\frac{v\dfrac{du}{dx}-u\dfrac{dv}{dx}}{v^2} \tag{13.5}$$

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

$$\frac{du}{dx}=\frac{du}{dy}\frac{dy}{dx} \tag{13.6}$$

Thus, if $u = f(y)$ and $y = g(x)$, this rule gives

$$\frac{d}{dx}f\big(g(x)\big)=f'\big(g(x)\big)g'(x) \tag{13.7}$$

where f' and g' are the derivatives of f and g, respectively, i.e., where

$$f'(y)=\frac{d}{dy}f(y) \quad \text{and} \quad g'(x)=\frac{d}{dx}g(x)$$

(f) Differentiation of an Inverse Function

If $y = f(x)$ is a differentiable function of x with nonzero derivative $f'(x)$, and $x = f^{-1}(y)$ is the inverse function, then

$$\frac{dx}{dy}=\frac{1}{\dfrac{dy}{dx}}=\frac{1}{f'(x)} \tag{13.8}$$

(Adams, 1999; Binmore, 1983; Granville, Smith, and Longley, 1957; Hardy, 1963).

13.2 INTEGRATION BY PARTS

$$\int u\,dv = uv - \int v\,du \tag{13.9}$$

where $u = u(x)$ and $v = v(x)$ are differentiable functions of x, $du = \frac{du}{dx}dx$, and $dv = \frac{dv}{dx}dx$.

(Adams, 1999; Binmore, 1983; Hardy, 1963; Stewart, 1995; Trim, 1983). Note: this result is derived from the product rule for differentiation, given by Equation 13.3.

13.3 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 $f(x)$:

1. 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

$$\frac{d}{dx}F(x) = f(x) \tag{13.10}$$

2. If $G(x)$ is a function such that its derivative $\frac{d}{dx}G(x) = f(x)$ for any x in the interval, then

$$\int_a^b f(x)dx = G(b) - G(a) \tag{13.11}$$

(Adams, 1999; Binmore, 1983; Sokolnikoff, 1939; Stewart, 1995; Trim, 1983).

13.4 TAYLOR SERIES

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f'''(a) + \cdots \quad (13.12)$$

where

$$f'(a) = \frac{d}{dx}f(x)\bigg|_{x=a},\ f''(a) = \frac{d^2}{dx^2}f(x)\bigg|_{x=a},\ f'''(a) = \frac{d^3}{dx^3}f(x)\bigg|_{x=a}, \text{ etc.}$$

Equation 13.12 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; Finney and Thomas, 1990; Gradshteyn and Ryzhik, 1965; Granville, Smith, and Longley, 1957; Hardy, 1963; Sokolnikoff, 1939; Stewart, 1995).

13.5 MACLAURIN SERIES

(Special case of Taylor series with $a = 0$)

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f'''(0) + \cdots \quad (13.13)$$

where

$$f'(0) = \frac{d}{dx}f(x)\bigg|_{x=0},\ f''(0) = \frac{d^2}{dx^2}f(x)\bigg|_{x=0},\ f'''(0) = \frac{d^3}{dx^3}f(x)\bigg|_{x=0}, \text{ etc.}$$

Equation 13.13 holds for values of x around the point 0 for which $f(x)$ has derivatives of all orders and the series satisfies convergence criteria (Adams, 1999; Finney and Thomas, 1990; Gradshteyn and Ryzhik, 1965; Granville, Smith, and Longley, 1957; Hardy, 1963; Sokolnikoff, 1939; Stewart, 1995).

13.6 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

$$f(a) = g(a) = 0 \tag{13.14}$$

and

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = L \tag{13.15}$$

then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = L \tag{13.16}$$

(Adams, 1999; Brand, 1955; Finney and Thomas, 1990; Sokolnikoff, 1939; Stewart, 1995; Trim, 1983). 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 13.14, i.e., if

$$\lim_{x \to a} f(x) = \pm \infty \tag{13.17}$$

and

$$\lim_{x \to a} g(x) = \pm \infty \tag{13.18}$$

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.

13.7 LAGRANGE MULTIPLIERS

The method of Lagrange multipliers 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

$$f = f(x_1, x_2, \ldots, x_n) \tag{13.19}$$

subject to the m constraints

$$\left.\begin{aligned} g_1(x_1, x_2, \ldots, x_n) &= 0 \\ g_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}\right\} \tag{13.20}$$

the method of Lagrange multipliers consists of the following steps:

1. Introduce m new variables, called Lagrange multipliers

   $\lambda_1, \lambda_2, \ldots, \lambda_m$ (one for each constraint equation).

2. Form the function L, called the Lagrangian and defined as

$$\begin{aligned} L = f(x_1, x_2, \ldots, x_n) + \lambda_1 g_1(x_1, x_2, \ldots, x_n) + \\ \lambda_2 g_2(x_1, x_2, \ldots, x_n) + \cdots + \lambda_m g_m(x_1, x_2, \ldots, x_n) \end{aligned} \tag{13.21}$$

3. Take the partial derivatives of the Lagrangian L with respect to each of the variables, and solve the equations

$$\left.\begin{aligned} \frac{\partial L}{\partial x_i} &= 0 \quad (i = 1, 2, \ldots, n) \\ \frac{\partial L}{\partial \lambda_j} &= 0 \quad (j = 1, 2, \ldots, m) \end{aligned}\right\} \tag{13.22}$$

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 13.20.

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; Binmore, 1983; Sokolnikoff, 1939).

13.8 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, and $f(x, t)$ and $\frac{\partial}{\partial t} f(x, t)$ are continuous functions in the region of integration. Then

$$\frac{d}{dt} I(t) = \int_{\alpha(t)}^{\beta(t)} \frac{\partial}{\partial t} f(x, t)dx + f(\beta(t), t)\frac{d\beta}{dt} - f(\alpha(t), t)\frac{d\alpha}{dt} \qquad (13.23)$$

Special cases:

$$\frac{d}{dt}\int_a^b f(x, t)dx = \int_a^b \frac{\partial}{\partial t} f(x, t)dx \qquad (13.24)$$

$$\frac{d}{dt}\int_a^t f(x)dx = f(t) \qquad (13.25)$$

$$\frac{d}{dt}\int_t^b f(x)dx = -f(t) \qquad (13.26)$$

where a and b are constants, and where the integrand $f(x)$ in Equations 13.25 and 13.26 is a function of x only (and not t).

(Adams, 1999; Binmore, 1983; Buck, 1956; Sokolnikoff, 1939; Trim, 1983). Note: Equation 13.23 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 13.23 is in the classic "Newsboy Problem" (see Section 9.3).

13.9 CHANGE OF 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

$$x = g(u) \tag{13.27}$$

be a continuous single-valued function with a continuous derivative $\frac{dx}{du} = g'(u)$ in the interval $c \le u \le d$, where c and d are given by

$$a = g(c) \tag{13.28}$$

and

$$b = g(d) \tag{13.29}$$

respectively. Then

$$\int_a^b f(x)dx = \int_c^d f\big(g(u)\big)g'(u)du \tag{13.30}$$

(Adams, 1999; Binmore, 1983; Brand, 1955; Sokolnikoff, 1939; Trim, 1983). Note that this transformation can substantially simplify the evaluation of an integral. The formula is derived from the chain rule for differentiation.

13.10 CHANGE OF VARIABLES IN A DOUBLE INTEGRAL

Given the double integral $\iint_R f(x, y)\, dxdy$ over the region R in the xy-plane, let

$$x = g(u, v) \tag{13.31}$$

and

$$y = h(u, v) \tag{13.32}$$

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

$$F(u, v) = f(g(u, v), h(u, v)) \tag{13.33}$$

In addition, let J be the Jacobian, given by the determinant

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \tag{13.34}$$

i.e.,

$$J = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \tag{13.35}$$

and let $|J|$ denote the absolute value of J. If the Jacobian J is nonzero, then

$$\iint_R f(x, y)dxdy = \iint_S F(u, v)|J|dudv \tag{13.36}$$

(Adams, 1999; Binmore, 1983; Brand, 1955; Sokolnikoff, 1939; Stewart, 1995).

Note: the determinants $\dfrac{\partial(x, y)}{\partial(u, v)}$ and $\dfrac{\partial(u, v)}{\partial(x, y)}$ are related by

$$\frac{\partial(x, y)}{\partial(u, v)} = \frac{1}{\dfrac{\partial(u, v)}{\partial(x, y)}} \tag{13.37}$$

where

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} \tag{13.38}$$

(Adams, 1999; Binmore, 1983; Brand, 1955). Equation 13.37 is useful for deriving J in 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 13.36 can be extended to triple and higher-order multiple integrals. Thus, for a triple integral $\iiint_R f(x, y, z)\, dxdydz$ over the region R in xyz-space, the formula becomes

$$\iiint_R f(x, y, z)dxdydz = \iiint_S F(u, v, w)|J|dudvdw \tag{13.39}$$

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

$$F(u, v, w) = f(x(u, v, w), y(u, v, w), z(u, v, w)) \tag{13.40}$$

and where the Jacobian J is given by the determinant

$$J = \frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} & \dfrac{\partial y}{\partial w} \\ \dfrac{\partial z}{\partial u} & \dfrac{\partial z}{\partial v} & \dfrac{\partial z}{\partial w} \end{vmatrix} \tag{13.41}$$

and is nonzero.

(Adams, 1999; Binmore, 1983; Brand, 1955; Sokolnikoff, 1939; Stewart, 1995).

13.11 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 13.1 and 13.2),

$$\int_{y=0}^{a}\int_{x=y}^{a} f(x,y)\,dx\,dy = \int_{x=0}^{a}\int_{y=0}^{x} f(x,y)\,dy\,dx \tag{13.42}$$

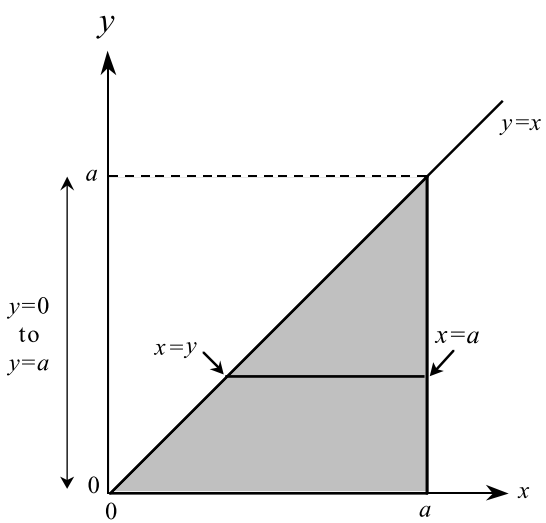

FIGURE 13.1 Limits of integration for left-hand-side of Equation (13.42) (integration in *x* direction first).

Extension for the case $a \to \infty$: If $f(x, y) \geq 0$ over the region of integration, then

$$\int_{y=0}^{\infty}\int_{x=y}^{\infty} f(x,y)\,dx\,dy = \int_{x=0}^{\infty}\int_{y=0}^{x} f(x,y)\,dy\,dx \tag{13.43}$$

provided the integrals converge.

(Adams, 1999; Binmore, 1983; Brand, 1955; Finney and Thomas, 1990; Trim, 1983). Note: changing the order of integration can substantially simplify

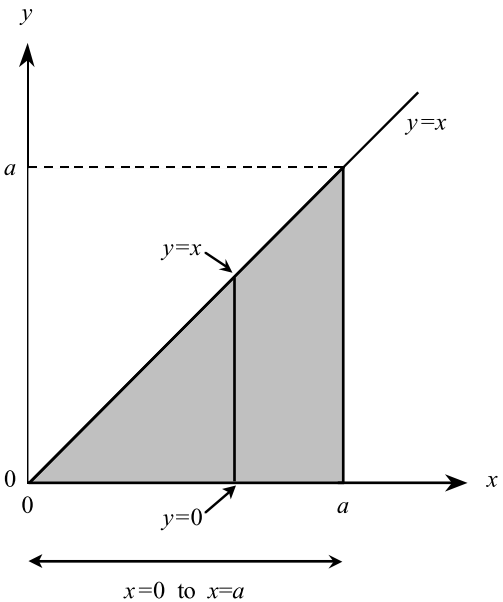

FIGURE 13.2 Limits of integration for right-hand-side of Equation (13.42) (integration in y direction first).

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.

13.12 CHANGING THE ORDER OF SUMMATION IN A DOUBLE SUM

If $a_{ij} \geq 0$ for all i and j (i = 0, 1, 2, …; j = 0, 1, 2, …), then

$$\sum_{j=0}^{\infty}\sum_{i=j}^{\infty} a_{ij} = \sum_{i=0}^{\infty}\sum_{j=0}^{i} a_{ij} \tag{13.44}$$

provided the sums converge (see Figures 13.3 and 13.4.). (Binmore, 1983; Clarke and Disney, 1985).

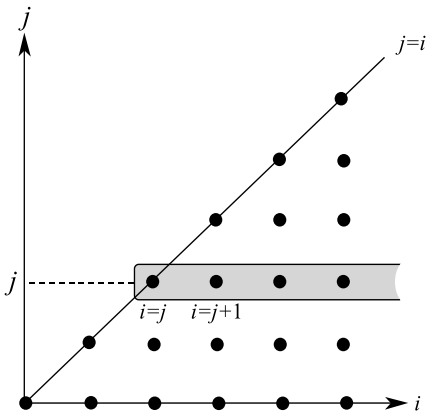

FIGURE 13.3 First summation range for left-hand side of Equation 13.44.

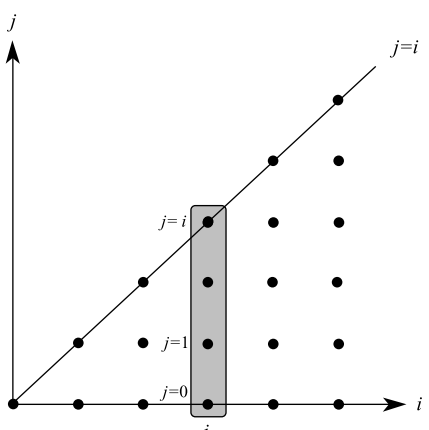

FIGURE 13.4 First summation range for right-hand side of Equation 13.44.

13.13 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

$$h = \frac{b-a}{n} \tag{13.45}$$

Let $x_0, x_1, x_2, \ldots, x_n$ be equally spaced points on the interval, given by

$$\left.\begin{aligned} x_0 &= A \\ x_1 &= a + h \\ x_2 &= a + 2h \\ &\vdots \\ x_{n-1} &= a + (n-1)h \\ x_n &= b \end{aligned}\right\} \tag{13.46}$$

In addition, for the midpoint rule, let $m_1, m_2, \ldots, m_n$ denote the midpoints of each subinterval, given by

$$\left.\begin{aligned} m_1 &= \frac{x_0 + x_1}{2} \\ m_2 &= \frac{x_1 + x_2}{2} \\ &\vdots \\ m_n &= \frac{x_{n-1} + x_n}{2} \end{aligned}\right\} \tag{13.47}$$

Figures 13.5, 13.6, and 13.7 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 13.5). The trapezoidal rule is based on linear approximations, and the area under the curve is thus approximated by trapezoids (Figure 13.6). 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 13.7).

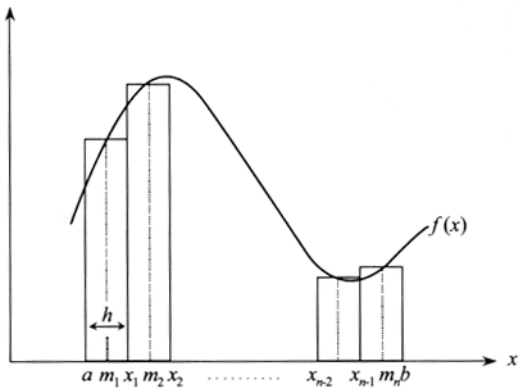

FIGURE 13.5 Midpoint rule approximation for numerical integration.

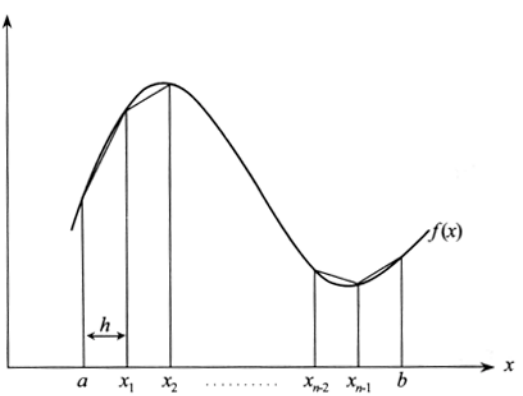

FIGURE 13.6 Trapezoidal rule approximation for numerical integration.

The three rules give the following approximate formulas for numerical evaluation of the integral $\int_a^b f(x)dx$:

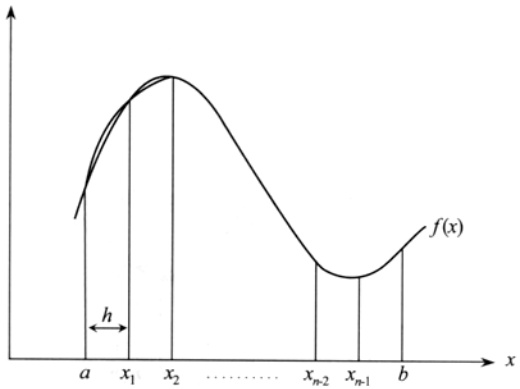

FIGURE 13.7 Simpson's rule approximation for numerical integration.

Midpoint Rule

$$\int_a^b f(x)dx \cong h\left[f(m_1) + f(m_2) + \ldots + f(m_n)\right] \tag{13.48}$$

Trapezoidal Rule

$$\int_a^b f(x)dx \cong \frac{h}{2}\left[f(a) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(b)\right] \tag{13.49}$$

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 an even number for Simpson's rule). (13.50)

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

$$\text{Midpoint Rule:} \qquad |\text{Error}| \le \frac{M(b-a)^3}{24n^2} \tag{13.51}$$

$$\text{Trapezoidal Rule:} \qquad |\text{Error}| \le \frac{M(b-a)^3}{12n^2} \tag{13.52}$$

$$\text{Simpson's Rule:} \qquad |\text{Error}| \le \frac{N(b-a)^5}{180n^4} \tag{13.53}$$

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 $a \le x \le b$,

and $|\ |$ denotes absolute value (Adams, 1999; Finney and Thomas, 1990; Press et al., 1992; Stewart, 1995; Trim, 1983).

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), etc.

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 the 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 et al., 1992).

14 Matrices

14.1 RULES FOR MATRIX CALCULATIONS

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{14.1}$$

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{14.2}$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \tag{14.3}$$

$$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \tag{14.4}$$

$$\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B} \tag{14.5}$$

$$(\mathbf{A}^T)^T = \mathbf{A} \tag{14.6}$$

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \tag{14.7}$$

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T \tag{14.8}$$

$$(\mathbf{ABC})^T = \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T \tag{14.9}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{14.10}$$

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{14.11}$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A} \tag{14.12}$$

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T \tag{14.13}$$

where

A, **B**, and **C** are matrices
λ is a scalar
$\mathbf{A}^T$ is the transpose of **A**
$\mathbf{A}^{-1}$ is the inverse of **A**

(Neter, Wasserman, and Kutner, 1985; Press, 1982).

14.2 INVERSES OF MATRICES

2×2 matrix: If $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{d}{D} & -\frac{b}{D} \\ -\frac{c}{D} & \frac{a}{D} \end{pmatrix} \tag{14.14}$$

where

$$D = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

(D = determinant of **A**)

3×3 matrix: If $\mathbf{B} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix}$ then

$$\mathbf{B}^{-1} = \begin{pmatrix} (ek - fh)/Z & -(bk - ch)/Z & (bf - ce)/Z \\ -(dk - fg)/Z & (ak - cg)/Z & -(af - cd)/Z \\ (dh - eg)/Z & -(ah - bg)/Z & (ae - bd)/Z \end{pmatrix} \tag{14.15}$$

where

$$Z = a(ek - fh) - b(dk - fg) + c(dh - eg).$$

(Z = determinant of **B**)

(Neter, Wasserman, and Kutner, 1985).

14.3 SERIES OF MATRICES

If **A** tends to the zero matrix as n tends to infinity, then (**I** – **A**) has an inverse, given by

$$(\mathbf{I}-\mathbf{A})^{-1} = \mathbf{I}+\mathbf{A}+\mathbf{A}^2+\mathbf{A}^3+\cdots = \sum_{k=0}^{\infty}\mathbf{A}^k \tag{14.16}$$

where **I** is the identity (unity) matrix (Kemeny and Snell, 1976).

15 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:

$$P_N^M = \frac{M!}{(M-N)!} \quad (N \le M) \tag{15.1}$$

2. Permutations with replacement:

$$P_N^M = M^N \tag{15.2}$$

3. Combinations without replacement:

$$C_N^M = \binom{M}{N} = \frac{M!}{N!(M-N)!} \quad (N \le M) \tag{15.3}$$

4. Combinations with replacement:

$$C_N^M = \binom{M+N-1}{N} = \frac{(M+N-1)!}{N!(M-1)!} \tag{15.4}$$

5. Circular permutations (permutations of objects forming a circle) without replacement:

$$P_N^M = \frac{M!}{N(M-N)!} \quad (N \le M) \tag{15.5}$$

6. Combinations with replacement that contain at least one of each object type:

$$C_N^M = \binom{N-1}{M-1} = \frac{(N-1)!}{(N-M)!(M-1)!} \quad (N \ge M) \tag{15.6}$$

7. Permutations of M objects, consisting of M_1 objects of type 1, M_2 objects of type 2, up to k types, where $M_1 + M_2 + \cdots + M_k = M$

$$P_N^M = \frac{M!}{M_1!M_2!\cdots M_k!} \tag{15.7}$$

(multinomial coefficients).

(Brualdi, 1977; Feller, 1964).

16 Summations

16.1 FINITE SUMS

$$1+2+3+\cdots+n=\frac{n(n+1)}{2} \tag{16.1}$$

$$1^2+2^2+3^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6} \tag{16.2}$$

$$1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2 \tag{16.3}$$

$$1+3+5+\cdots+(2n-1)=n^2 \tag{16.4}$$

$$a+ar+ar^2+ar^3+\cdots+ar^n=\frac{a\left(1-r^{n+1}\right)}{1-r} \quad (r\neq 1) \tag{16.5}$$

16.2 INFINITE SUMS

$$a+ar+ar^2+ar^3+\cdots=\frac{a}{1-r} \quad \left(|r|<1\right) \tag{16.6}$$

$$1ar+2ar^2+3ar^3+\cdots=\frac{ar}{(1-r)^2} \quad \left(|r|<1\right) \tag{16.7}$$

$$1^2ar+2^2ar^2+3^2ar^3+\cdots=\frac{ar(1+r)}{(1-r)^3} \quad \left(|r|<1\right) \tag{16.8}$$

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots=\frac{\pi^2}{6} \tag{16.9}$$

$$\frac{1}{1^2}-\frac{1}{2^2}+\frac{1}{3^2}-\cdots=\frac{\pi^2}{12} \tag{16.10}$$

$$\lim_{n\to\infty}\left\{1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\ln(n)\right\}=\gamma \tag{16.11}$$

where $\gamma\cong 0.577216$ is Euler's constant.

(Abramowitz and Stegun, 1968; Gradshteyn and Ryzhik, 1965; Zwillinger, 1996).

17 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 (future worth)
- A = uniform end-of-period payment continuing for n periods that, in total, is equivalent to P at interest rate i (annuity)

then

$$F = P(1 + i)^n \tag{17.1}$$

$$A = P\left(\frac{i(1+i)^n}{(1+i)^n - 1}\right) \tag{17.2}$$

Let

- E_k = portion of A in period k paid against principal (equity payment)
- I_k = portion of A in period k paid as interest (interest payment)

then

$$E_k = A\left(\frac{1}{(1+i)^{n-k+1}}\right) \tag{17.3}$$

$$I_k = A\left(1 - \frac{1}{(1+i)^{n-k+1}}\right) \tag{17.4}$$

(Grant, Ireson, and Leavenworth, 1982; White, Agee, and Case, 1977).

References

Abramowitz, M. and Stegun, I. A. (1968). *Handbook of Mathematical Functions*. Dover, New York.

Adams, R. A. (1999). *Calculus — A Complete Course* (4th ed.). Addison-Wesley, Reading, MA.

Ahuja, R. K., Magnanti, T. L., and Orlin, J., B. (1993). *Network Flows*. Prentice-Hall, Englewood Cliffs, NJ.

Allen, A. O. (1978). *Probability, Statistics, and Queueing Theory — With Computer Science Applications*. Academic Press, Orlando, FL.

Apsimon, H. G. (1958). Mathematical note 2754 — A repeated integral. *Math. Gaz.*, 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.

Askin, R. G. and Standridge, C. R. (1993). *Modeling and Analysis of Manufacturing Systems*. John Wiley & Sons, 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. *J. Manuf. Syst.*, 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 and Breach, Amsterdam, Chapter 3.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. *Proc. Cambridge Philos. Soc.*, 55(4), 299–327.

Benjamin, J. R. and Cornell, C. A. (1970). *Probability, Statistics, and Decision for Civil Engineers*. McGraw-Hill, New York.

Binmore, K. G. (1977). *Mathematical Analysis*. Cambridge University Press, Cambridge, U.K.

Binmore, K. G. (1983). *Calculus*. Cambridge University Press, Cambridge, U.K.

Blumenfeld, D. E. (1990). A simple formula for estimating throughput of serial production lines with variable processing times and limited buffer capacity. *Int. J. Prod. Res.*, 28(6), 1163–1182.

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. *Transport. Res.*, 19B(5), 361–380.

Bolch, G., Greiner, S., de Meer, H., and Trivedi, K. S. (1998). *Queueing Networks and Markov Chains*. John Wiley & Sons, 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.

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*. John Wiley & Sons, New York.

Brown, R. G. (1977). *Materials Management Systems*. John Wiley & Sons, 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. *Oper. Res.*, 33(3), 469–490.

Buzacott, J. A. (1968). Prediction of the Efficiency of Production Systems Without Internal Storage. *Int. J. Prod. Res.*, 6(3), 173–188.

Buzacott, J. A. and Shanthikumar, J. G. (1993). *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Englewood Cliffs, NJ.

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 Trans.*, 19(2), 150–159.

Christofides, N. and Eilon, S. (1969). Expected distances in distribution problems. *Operational Res. Q.*, 20(4), 437–443.

Chvátal, V. (1983). *Linear Programming*. W. H. Freeman, New York.

Çinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ.

Clark, C. E. (1961). The Greatest of a Finite Set of Random Variables. *Oper. Res.*, 9(2), 145–162.

Clarke, A. B. and Disney, R. L. (1985). *Probability and Random Processes: A First Course with Applications*. John Wiley & Sons, New York.

Cohen, S. S. (1985). *Operational Research*. Edward Arnold, London.

Cox, D. R. (1962). *Renewal Theory*. Methuen, London.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. Methuen, London.

Cox, D. R. and Smith, W. L. (1971). *Queues*. Chapman & Hall, London.

Daganzo, C. F. (1984). The length of tours in zones of different shapes. *Transp. Res.*, 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., John Wiley & Sons, 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., John Wiley & Sons, New York, Chapter XXIII, pp. 359–373.

Dantzig, G. B. (1963). *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ.

DeGroot, M. H. (1986). *Probability and Statistics* (2nd ed.). Addison-Wesley, Reading, MA.

Devore, J. L. (1987). *Probability and Statistics for Engineering and the Sciences* (2nd ed.). Brooks/Cole, Monterey, CA.

Drake, A. W. (1967). *Fundamentals of Applied Probability Theory.* McGraw-Hill, New York.

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, J. R., Anderson, D. R., Sweeney, D. J., and Williams, T. A. (1984). *Applied Production and Operations Management* (2nd ed.). West Publishing Company, St. Paul, MN.

Evans, M., Hastings, N., and Peacock, B. (2000). *Statistical Distributions* (3rd ed.). John Wiley & Sons, New York.

Fairthorne, D. (1965). The distances between pairs of points in towns of simple geometric shapes. *Proceedings of the 2nd Symposium on the Theory of Traffic Flow* (London, 1963). Almond, J., Ed., Organization for Economic Cooperation and Development (OECD), Paris, 1965, pp. 391–406.

Feller, W. (1964). *An Introduction to Probability Theory and Its Applications — Volume 1*. John Wiley & Sons, New York.

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. *J. Oper. Res. Soc.*, 31(10), 937–942.

Freund, J. E. (1992). *Mathematical Statistics* (5th Edition). Prentice-Hall, Englewood Cliffs, NJ.

Garfinkel, R. S. and Nemhauser, G. L. (1972). *Integer Programming*. John Wiley & Sons, 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. *Math. Gaz.*, 42(342), 292–293.

Gass, S. I. (1964). *Linear Programming—- Methods and Applications* (2nd ed.). 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.

Goldman, A. S. and Slattery, T. B. (1964). *Maintainability: A Major Element of System Effectiveness*. John Wiley & Sons, 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.* John Wiley & Sons, New York.

Granville, W. A., Smith, P. F., and Longley, W. R. (1957). *Elements of Differential and Integral Calculus*. Ginn and Company, Boston.

Gross, D. and Harris, C. M. (1985). *Fundamentals of Queueing Theory.* John Wiley & Sons, New York.

Hadley, G. and Whitin, T. M. (1963). *Analysis of Inventory Systems.* Prentice-Hall, Englewood Cliffs, NJ.

Haight, F. A. (1981). *Applied Probability.* Plenum Press, New York.

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. *Transport. Res.*, 30B(5), 387–403.

Hanssmann, F. (1962). *Operations Research in Production and Inventory Control*. John Wiley & Sons, New York.

Hardy, G. H. (1963). *A Course of Pure Mathematics* (10th ed.). Cambridge University Press, Cambridge, U.K.

Harris, F. W. (1913). How many parts to make at once. *Factory*: *Mag. Manage.*, 10(2), 135–136, 152. (Reprinted 1990 in *Oper. Res.*, 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.

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. *J. Ind. Eng.*, 17(12), 651–658.

Hillier, F. S. and Lieberman, G. J. (1980). *Introduction to Operations Research* (3rd ed.). Holden-Day, San Francisco.

Hoel, P. G., Port, S. C., and Stone, C. J. (1971). *Introduction to Probability Theory.* Houghton Mifflin, Boston

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. *Oper. Res.*, 4(6), 674–683.

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*. John Wiley & Sons, New York.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1* (2nd ed.). John Wiley & Sons, New York.

Karloff, H. (1991). *Linear Programming*. Birkhäuser, Boston.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, 4, 373–395.

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. *J. R. Stat. Soc.*, 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. *Ann. Math. Stat.*, 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. *Int. J. Prod. Res.*, 35(12), 3405–3414.

Kimura, T. (1986). A two-moment approximation for the mean waiting time in the GI/G/s queue. *Manage. Sci.*, 32(6), 751–763.

Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Proc. Cambridge Philosophical Soc.*, 57(1), 902–904.

Kingman, J. F. C. (1962). On queues in heavy traffic. *J. R. Stat. Soc., 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, NC.

Kleinrock, L. (1975). *Queueing Systems — Volume I*: *Theory.* John Wiley & Sons, New York.

Kleinrock, L. (1976). *Queueing Systems — Volume II*: *Computer Applications*. John Wiley & Sons, New York.

Krämer, W. and Langenbach-Belz, M. (1976). Approximate formulae for the delay in the queueing system GI/G/1. *Congressbook, 8th International Teletraffic Congress (ITC)*, Melbourne, Australia, pp. 235-1 to 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).

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 ed.). McGraw-Hill, New York.

Lewis, C. D. (1970). *Scientific Inventory Control*. Elsevier, 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: $L = \lambda W$. *Oper. Res.*, 9(3), 383–387.

Love, S. F. (1979). *Inventory Control*. McGraw-Hill, New York.

Luenberger, D. G. (1984). *Linear and Non-Linear Programming* (2nd ed.). Addison-Wesley, Reading, MA.

Makino, T. (1964). On the mean passage time concerning some queuing problems of the tandem type. *J. Oper. Res. Soc, Jpn*, 7(1), 17–47.
McClain, J. O. and Thomas, L. J. (1985). *Operations Management*: *Production of Goods and Services* (2nd ed.). Prentice-Hall, Englewood Cliffs, NJ.
McFadden, F. R. (1972). On lead time demand distributions. *Decision Sci.*, 3, 106–126.
Medhi, J. (1991). *Stochastic Models in Queueing Theory.* Academic Press, Boston.
Mehrotra, S. (1992). On the implementation of a primal-dual interior point method. *SIAM J. Optimization*, 2(4), 575–601.
Melzak, Z. A. (1983). *Invitation to Geometry.* John Wiley & Sons, New York.
Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd ed.). McGraw-Hill, New York.
Morse, P. M. (1958). *Queues, Inventories and Maintenance*. John Wiley & Sons, New York.
Munakata, T. (1979). *Matrices and Linear Programming with Applications.* Holden-Day, San Francisco.
Murty, K. G. (1976). *Linear and Combinatorial Programming*. John Wiley & Sons, 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, Gainesville, FL.
Nahmias, S. (1989). *Production and Operations Analysis*. Richard D. Irwin, Homewood, IL.
Nemhauser, G. L. and Wolsey, L. A. (1988). *Integer and Combinatorial Optimization.* John Wiley & Sons, 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 ed.). Chapman & Hall, London.
Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part 1: general theory. *Transport. Res.*, 27B(4), 281–287.
Ozan, T. M. (1986). *Applied Mathematical Programming for Engineering and Production Management*. Prentice-Hall, Englewood Cliffs, NJ.
Papadopoulos, H. T., Heavy, C., and Browne, J. (1993). *Queueing Theory in Manufacturing Systems Analysis and Design*. Chapman & Hall, London.
Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes* (2nd ed.). 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. *Oper. Res.*, 33(1), 134–152.
Poularikas, A. D. and Seely, S. (1985). *Signals and Systems*. PWS Engineering, Boston.
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. (1992). *Numerical Recipes in FORTRAN — The Art of Scientific Computing* (2nd ed.), Cambridge University Press, Cambridge, U.K.

Ravindran, A., Phillips, D. T., and Solberg, J. J. (1987). *Operations Research*: *Principles and Practice*. John Wiley & Sons, New York.
Ross, S. M. (1983). *Stochastic Processes*. John Wiley & Sons, New York.
Ross, S. M. (1988). *A First Course in Probability* (3rd ed.). Macmillan, New York.
Ross, S. M. (1989). *Introduction to Probability Models* (4th ed.). Academic Press, New York.
Sakasegawa, H. (1977). An approximation formula $L_q \cong \alpha \cdot \rho^{\beta}/(1-\rho)$. *Ann. Inst. Stat. Math*,, 29, Part A, 67–75.
Schneider, H. (1981). Effect of service-levels on order-points or order-levels in inventory models. *Inter. J. Prod. Res.*, 19(6), 615–631.
Shanthikumar, J. G. and Buzacott, J. A. (1980). On the approximations to the single server queue. *Inter. J. Prod. Res.*, 18(6), 761–773.
Silver, E. A., Pyke, D, F., and Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling* (3rd ed.). John Wiley & Sons, New York.
Simmons. H. A. (1945). *Plane and Spherical Trigonometry.* John Wiley & Sons, 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 Res. Rec.*, 169, Transportation Research Board, Washington, D.C.
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.
Sokolnikoff, I. S. (1939). *Advanced Calculus*. McGraw-Hill, New York.
Spivey, W. A. and Thrall, R. M. (1970). *Linear Optimization*. Holt, Rinehart and Winston, New York.
Stein, D. M. (1978). An asymptotic, probabilistic analysis of a routing problem. *Math. Oper. Res.*, 3(2), 89–101.
Stevenson, W. J. (1986). *Production/Operations Management*. Richard D. Irwin, Homewood, IL.
Stewart, J. (1995). *Calculus* (3rd ed.). Brooks/Cole, Pacific Grove, CA.
Stirzaker, D. (1994). *Elementary Probability.* Cambridge University Press, Cambridge, U.K.
Taft, E. W. (1918). The most economical production lot. *The Iron Age*, 101, 1410–1412.
Tanner, M. (1995). *Practical Queueing Analysis*. McGraw-Hill, New York.
Tersine, R. J. (1985). *Production/Operations Management*: *Concepts, Structure, and Analysis* (2nd ed.). 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*. John Wiley & Sons, 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. *Eur. J. Oper. Res.*, 17, 175–190.

Trim, D. W. (1983). *Calculus and Analytic Geometry*. Addison-Wesley, Reading, MA.

Vanderbei, R. J. (1997). *Linear Programming*: *Foundations and Extensions*. Kluwer, Boston.

Vaughan, R. J. (1987). *Urban Spatial Traffic Patterns*. Pion, London.

Wagner, H. M. (1969). *Principles of Operations Research — with Applications to Managerial Decisions*. Prentice-Hall, Englewood Cliffs, NJ.

Wald, A. (1947). *Sequential Analysis*. John Wiley & Sons, New York.

White, J., A., Agee, M. A., and Case, K. E. (1977). *Principles of Engineering Economic Analysis*. John Wiley & Sons, New York.

Whitt, W. (1983). The queueing network analyzer. *Bell Syst. Tech. J.*, 62(9), 2779–2815.

Whitt, W. (1993). Approximations for the GI/G/m queue. *Prod. Oper. Manage.*, 2(2), 114-161.

Wilson, R. H. (1934). A scientific routine for stock control. *Harv. Bus. Rev.*, 13(1), 116–128.

Winkler, R. L., Roodman, G. M., and Britney, R. R. (1972). The determination of partial moments. *Manage. Sci.*, 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 and Row, New York.

Wright, S. J. (1997). *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, Philadelphia.

Zwillinger, D. (1996). *Standard Mathematical Tables and Formulae* (30th ed.). CRC Press, Boca Raton, FL.