Introduction to Probability Models

Introduction to Probability Models

Twelfth Edition

Sheldon M. Ross
University of Southern California
Los Angeles, CA, United States of America

ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-814346-9

Publisher: Katey Birtcher
Acquisition Editor: Katey Birtcher
Editorial Project Manager: Susan Ikeda
Production Project Manager: Divya Krishna Kumar
Designer: Matthew Limbert

Typeset by VTeX

Preface

This text is intended as an introduction to elementary probability theory and stochastic processes. It is particularly well suited for those wanting to see how probability theory can be applied to the study of phenomena in fields such as engineering, computer science, management science, the physical and social sciences, and operations research.

It is generally felt that there are two approaches to the study of probability theory. One approach is heuristic and nonrigorous and attempts to develop in the student an intuitive feel for the subject that enables him or her to "think probabilistically." The other approach attempts a rigorous development of probability by using the tools of measure theory. It is the first approach that is employed in this text. However, because it is extremely important in both understanding and applying probability theory to be able to "think probabilistically," this text should also be useful to students interested primarily in the second approach.

New to This Edition

The twelfth edition includes new text material, examples, and exercises in almost every chapter. Newly added Sections begin in Chapter 1 with Section 1.7, where it is shown that probability is a continuous function of events. The new Section 2.8 proves the Borel–Cantelli lemma and uses it as the basis of a proof of the strong law of large numbers. Subsection 5.2.5 introduces the Dirichlet distribution and details its relationship to exponential random variables. Notable also in Chapter 5 is a new approach for obtaining results for both stationary and non-stationary Poisson processes. The biggest change in the current edition, though, is the addition of Chapter 12 on coupling methods. Its usefulness in analyzing stochastic systems is indicated throughout this chapter.

Course

Ideally, this text would be used in a one-year course in probability models. Other possible courses would be a one-semester course in introductory probability theory (involving Chapters 1–3 and parts of others) or a course in elementary stochastic processes. The textbook is designed to be flexible enough to be used in a variety of

possible courses. For example, I have used Chapters 5 and 8, with smatterings from Chapters 4 and 6, as the basis of an introductory course in queueing theory.

Examples and Exercises

Many examples are worked out throughout the text, and there are also a large number of exercises to be solved by students. More than 100 of these exercises have been starred and their solutions provided at the end of the text. These starred problems can be used for independent study and test preparation. An Instructor's Manual, containing solutions to all exercises, is available free to instructors who adopt the book for class.

Organization

Chapters 1 and 2 deal with basic ideas of probability theory. In Chapter 1 an axiomatic framework is presented, while in Chapter 2 the important concept of a random variable is introduced. Section 2.6.1 gives a simple derivation of the joint distribution of the sample mean and sample variance of a normal data sample. Section 2.8 gives a proof of the strong law of large numbers, with the proof assuming that both the expected value and variance of the random variables under consideration are finite.

Chapter 3 is concerned with the subject matter of conditional probability and conditional expectation. "Conditioning" is one of the key tools of probability theory, and it is stressed throughout the book. When properly used, conditioning often enables us to easily solve problems that at first glance seem quite difficult. The final section of this chapter presents applications to (1) a computer list problem, (2) a random graph, and (3) the Polya urn model and its relation to the Bose–Einstein distribution. Section 3.6.5 presents k-record values and the surprising Ignatov's theorem.

In Chapter 4 we come into contact with our first random, or stochastic, process, known as a Markov chain, which is widely applicable to the study of many real-world phenomena. Applications to genetics and production processes are presented. The concept of time reversibility is introduced and its usefulness illustrated. Section 4.5.3 presents an analysis, based on random walk theory, of a probabilistic algorithm for the satisfiability problem. Section 4.6 deals with the mean times spent in transient states by a Markov chain. Section 4.9 introduces Markov chain Monte Carlo methods. In the final section we consider a model for optimally making decisions known as a Markovian decision process.

In Chapter 5 we are concerned with a type of stochastic process known as a counting process. In particular, we study a kind of counting process known as a Poisson process. The intimate relationship between this process and the exponential distribution is discussed. New derivations for the Poisson and nonhomogeneous Poisson processes are discussed. Examples relating to analyzing greedy algorithms, minimizing highway encounters, collecting coupons, and tracking the AIDS virus, as well as

material on compound Poisson processes, are included in this chapter. Section 5.2.4 gives a simple derivation of the convolution of exponential random variables.

Chapter 6 considers Markov chains in continuous time with an emphasis on birth and death models. Time reversibility is shown to be a useful concept, as it is in the study of discrete-time Markov chains. Section 6.8 presents the computationally important technique of uniformization.

Chapter 7, the renewal theory chapter, is concerned with a type of counting process more general than the Poisson. By making use of renewal reward processes, limiting results are obtained and applied to various fields. Section 7.9 presents new results concerning the distribution of time until a certain pattern occurs when a sequence of independent and identically distributed random variables is observed. In Section 7.9.1, we show how renewal theory can be used to derive both the mean and the variance of the length of time until a specified pattern appears, as well as the mean time until one of a finite number of specified patterns appears. In Section 7.9.2, we suppose that the random variables are equally likely to take on any of m possible values, and compute an expression for the mean time until a run of m distinct values occurs. In Section 7.9.3, we suppose the random variables are continuous and derive an expression for the mean time until a run of m consecutive increasing values occurs.

Chapter 8 deals with queueing, or waiting line, theory. After some preliminaries dealing with basic cost identities and types of limiting probabilities, we consider exponential queueing models and show how such models can be analyzed. Included in the models we study is the important class known as a network of queues. We then study models in which some of the distributions are allowed to be arbitrary. Included are Section 8.6.3 dealing with an optimization problem concerning a single server, general service time queue, and Section 8.8, concerned with a single server, general service time queue in which the arrival source is a finite number of potential users.

Chapter 9 is concerned with reliability theory. This chapter will probably be of greatest interest to the engineer and operations researcher. Section 9.6.1 illustrates a method for determining an upper bound for the expected life of a parallel system of not necessarily independent components and Section 9.7.1 analyzes a series structure reliability model in which components enter a state of suspended animation when one of their cohorts fails.

Chapter 10 is concerned with Brownian motion and its applications. The theory of options pricing is discussed. Also, the arbitrage theorem is presented and its relationship to the duality theorem of linear programming is indicated. We show how the arbitrage theorem leads to the Black–Scholes option pricing formula.

Chapter 11 deals with simulation, a powerful tool for analyzing stochastic models that are analytically intractable. Methods for generating the values of arbitrarily distributed random variables are discussed, as are variance reduction methods for increasing the efficiency of the simulation. Section 11.6.4 introduces the valuable simulation technique of importance sampling, and indicates the usefulness of tilted distributions when applying this method.

Chapter 12 introduces the concept of coupling and shows how it can be effectively employed in analyzing stochastic systems. Its use in showing stochastic order relations between random variables and processes—such as showing that a birth and death pro-

cess is stochastically increasing in its initial state—is illustrated. It is also shown how coupling can be of use in bounding the distance between distributions, in obtaining stochastic optimization results, in bounding the error of Poisson approximations, and in other areas of applied probability.

Acknowledgments

We would like to acknowledge with thanks the helpful suggestions made by the many reviewers of the text. These comments have been essential in our attempt to continue to improve the book and we owe these reviewers, and others who wish to remain anonymous, many thanks:

Mark Brown, City University of New York
Yang Cao, University of Southern California
Zhiqin Ginny Chen, University of Southern California
Tapas Das, University of South Florida
Israel David, Ben-Gurion University
Jay Devore, California Polytechnic Institute
Eugene Feinberg, State University of New York, Stony Brook
Rodrigo Gaitan, University of California, Riverside
Ramesh Gupta, University of Maine
Babak Haji, Sharif University
Marianne Huebner, Michigan State University
Garth Isaak, Lehigh University
Jonathan Kane, University of Wisconsin Whitewater
Amarjot Kaur, Pennsylvania State University
Zohel Khalil, Concordia University
Eric Kolaczyk, Boston University
Melvin Lax, California State University, Long Beach
Jean Lemaire, University of Pennsylvania
Xianxu Li
Andrew Lim, University of California, Berkeley
George Michailidis, University of Michigan
Donald Minassian, Butler University
Joseph Mitchell, State University of New York, Stony Brook
Krzysztof Osfaszewski, University of Illinois
Erol Pekoz, Boston University
Evgeny Poletsky, Syracuse University
James Propp, University of Massachusetts, Lowell
Anthony Quas, University of Victoria
Charles H. Roumeliotis, Proofreader
David Scollnik, University of Calgary
Mary Shepherd, Northwest Missouri State University
Galen Shorack, University of Washington, Seattle

John Shortle, George Mason University
Marcus Sommereder, Vienna University of Technology
Osnat Stramer, University of Iowa
Gabor Szekeley, Bowling Green State University
Marlin Thomas, Purdue University
Henk Tijms, Vrije University
Zhenyuan Wang, University of Binghamton
Ward Whitt, Columbia University
Bo Xhang, Georgia University of Technology
Zhengyu Zhang, University of Southern California
Jiang Zhiqiang
Julie Zhou, University of Victoria
Zheng Zuo, Stanford University

Introduction to Probability Theory 1

1.1 Introduction

Any realistic model of a real-world phenomenon must take into account the possibility of randomness. That is, more often than not, the quantities we are interested in will not be predictable in advance but, rather, will exhibit an inherent variation that should be taken into account by the model. This is usually accomplished by allowing the model to be probabilistic in nature. Such a model is, naturally enough, referred to as a probability model.

The majority of the chapters of this book will be concerned with different probability models of natural phenomena. Clearly, in order to master both the "model building" and the subsequent analysis of these models, we must have a certain knowledge of basic probability theory. The remainder of this chapter, as well as the next two chapters, will be concerned with a study of this subject.

1.2 Sample Space and Events

Suppose that we are about to perform an experiment whose outcome is not predictable in advance. However, while the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by S.

Some examples are the following.

1. If the experiment consists of the flipping of a coin, then

$$S = \{H, T\}$$

where H means that the outcome of the toss is a head and T that it is a tail.

2. If the experiment consists of rolling a die, then the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

where the outcome i means that i appeared on the die, $i = 1, 2, 3, 4, 5, 6$.

3. If the experiment consists of flipping two coins, then the sample space consists of the following four points:

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

The outcome will be (H, H) if both coins come up heads; it will be (H, T) if the first coin comes up heads and the second comes up tails; it will be (T, H) if the

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00006-8

first comes up tails and the second heads; and it will be (T, T) if both coins come up tails.

4. If the experiment consists of rolling two dice, then the sample space consists of the following 36 points:

$$S = \begin{Bmatrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6) \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{Bmatrix}$$

where the outcome (i, j) is said to occur if i appears on the first die and j on the second die.

5. If the experiment consists of measuring the lifetime of a car, then the sample space consists of all nonnegative real numbers. That is,[1]

$$S = [0, \infty)$$ ■

Any subset E of the sample space S is known as an *event*. Some examples of events are the following.

1′. In Example (1), if $E = \{H\}$, then E is the event that a head appears on the flip of the coin. Similarly, if $E = \{T\}$, then E would be the event that a tail appears.

2′. In Example (2), if $E = \{1\}$, then E is the event that one appears on the roll of the die. If $E = \{2, 4, 6\}$, then E would be the event that an even number appears on the roll.

3′. In Example (3), if $E = \{(H, H), (H, T)\}$, then E is the event that a head appears on the first coin.

4′. In Example (4), if $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then E is the event that the sum of the dice equals seven.

5′. In Example (5), if $E = (2, 6)$, then E is the event that the car lasts between two and six years. ■

We say that the event E occurs when the outcome of the experiment lies in E. For any two events E and F of a sample space S we define the new event $E \cup F$ to consist of all outcomes that are either in E or in F or in both E and F. That is, the event $E \cup F$ will occur if *either* E or F occurs. For example, in (1) if $E = \{H\}$ and $F = \{T\}$, then

$$E \cup F = \{H, T\}$$

That is, $E \cup F$ would be the whole sample space S. In (2) if $E = \{1, 3, 5\}$ and $F = \{1, 2, 3\}$, then

$$E \cup F = \{1, 2, 3, 5\}$$

[1] The set (a, b) is defined to consist of all points x such that $a < x < b$. The set $[a, b]$ is defined to consist of all points x such that $a \leqslant x \leqslant b$. The sets $(a, b]$ and $[a, b)$ are defined, respectively, to consist of all points x such that $a < x \leqslant b$ and all points x such that $a \leqslant x < b$.

and thus $E \cup F$ would occur if the outcome of the die is 1 or 2 or 3 or 5. The event $E \cup F$ is often referred to as the *union* of the event E and the event F.

For any two events E and F, we may also define the new event EF, sometimes written $E \cap F$, and referred to as the *intersection* of E and F, as follows. EF consists of all outcomes which are *both* in E and in F. That is, the event EF will occur only if both E and F occur. For example, in (2) if $E = \{1, 3, 5\}$ and $F = \{1, 2, 3\}$, then

$$EF = \{1, 3\}$$

and thus EF would occur if the outcome of the die is either 1 or 3. In Example (1) if $E = \{H\}$ and $F = \{T\}$, then the event EF would not consist of any outcomes and hence could not occur. To give such an event a name, we shall refer to it as the null event and denote it by Ø. (That is, Ø refers to the event consisting of no outcomes.) If $EF = Ø$, then E and F are said to be *mutually exclusive*.

We also define unions and intersections of more than two events in a similar manner. If $E_1, E_2, \ldots$ are events, then the union of these events, denoted by $\bigcup_{n=1}^{\infty} E_n$, is defined to be the event that consists of all outcomes that are in E_n for at least one value of $n = 1, 2, \ldots$. Similarly, the intersection of the events E_n, denoted by $\bigcap_{n=1}^{\infty} E_n$, is defined to be the event consisting of those outcomes that are in all of the events $E_n, n = 1, 2, \ldots$.

Finally, for any event E we define the new event E^c, referred to as the *complement* of E, to consist of all outcomes in the sample space S that are not in E. That is, E^c will occur if and only if E does not occur. In Example (4) if $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then E^c will occur if the sum of the dice does not equal seven. Also note that since the experiment must result in some outcome, it follows that $S^c = Ø$.

1.3 Probabilities Defined on Events

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following three conditions:

(i) $0 \leqslant P(E) \leqslant 1$.

(ii) $P(S) = 1$.

(iii) For any sequence of events $E_1, E_2, \ldots$ that are mutually exclusive, that is, events for which $E_n E_m = Ø$ when $n \neq m$, then

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

We refer to $P(E)$ as the probability of the event E.

Example 1.1. In the coin tossing example, if we assume that a head is equally likely to appear as a tail, then we would have

$$P(\{H\}) = P(\{T\}) = \tfrac{1}{2}$$

On the other hand, if we had a biased coin and felt that a head was twice as likely to appear as a tail, then we would have

$$P(\{H\}) = \tfrac{2}{3}, \qquad P(\{T\}) = \tfrac{1}{3}$$ ■

Example 1.2. In the die tossing example, if we supposed that all six numbers were equally likely to appear, then we would have

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \tfrac{1}{6}$$

From (iii) it would follow that the probability of getting an even number would equal

$$\begin{aligned} P(\{2,4,6\}) &= P(\{2\}) + P(\{4\}) + P(\{6\}) \\ &= \tfrac{1}{2} \end{aligned}$$ ■

Remark. We have chosen to give a rather formal definition of probabilities as being functions defined on the events of a sample space. However, it turns out that these probabilities have a nice intuitive property. Namely, if our experiment is repeated over and over again then (with probability 1) the proportion of time that event E occurs will just be $P(E)$.

Since the events E and E^c are always mutually exclusive and since $E \cup E^c = S$ we have by (ii) and (iii) that

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

or

$$P(E^c) = 1 - P(E) \tag{1.1}$$

In words, Eq. (1.1) states that the probability that an event does not occur is one minus the probability that it does occur.

We shall now derive a formula for $P(E \cup F)$, the probability of all outcomes either in E or in F. To do so, consider $P(E) + P(F)$, which is the probability of all outcomes in E plus the probability of all points in F. Since any outcome that is in both E and F will be counted twice in $P(E) + P(F)$ and only once in $P(E \cup F)$, we must have

$$P(E) + P(F) = P(E \cup F) + P(EF)$$

or equivalently

$$P(E \cup F) = P(E) + P(F) - P(EF) \tag{1.2}$$

Note that when E and F are mutually exclusive (that is, when $EF = \emptyset$), then Eq. (1.2) states that

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(\emptyset) \\ &= P(E) + P(F) \end{aligned}$$

a result which also follows from condition (iii). (Why is $P(\emptyset) = 0$?)

Example 1.3. Suppose that we toss two coins, and suppose that we assume that each of the four outcomes in the sample space

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

is equally likely and hence has probability $\frac{1}{4}$. Let

$$E = \{(H, H), (H, T)\} \quad \text{and} \quad F = \{(H, H), (T, H)\}$$

That is, E is the event that the first coin falls heads, and F is the event that the second coin falls heads.

By Eq. (1.2) we have that $P(E \cup F)$, the probability that either the first or the second coin falls heads, is given by

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(EF) \\ &= \tfrac{1}{2} + \tfrac{1}{2} - P(\{H, H\}) \\ &= 1 - \tfrac{1}{4} = \tfrac{3}{4} \end{aligned}$$

This probability could, of course, have been computed directly since

$$P(E \cup F) = P(\{(H, H), (H, T), (T, H)\}) = \tfrac{3}{4}$$ ■

We may also calculate the probability that any one of the three events E or F or G occurs. This is done as follows:

$$P(E \cup F \cup G) = P((E \cup F) \cup G)$$

which by Eq. (1.2) equals

$$P(E \cup F) + P(G) - P((E \cup F)G)$$

Now we leave it for you to show that the events $(E \cup F)G$ and $EG \cup FG$ are equivalent, and hence the preceding equals

$$\begin{aligned} &P(E \cup F \cup G) \\ &\quad = P(E) + P(F) - P(EF) + P(G) - P(EG \cup FG) \\ &\quad = P(E) + P(F) - P(EF) + P(G) - P(EG) - P(FG) + P(EGFG) \\ &\quad = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG) \end{aligned} \tag{1.3}$$

In fact, it can be shown by induction that, for any n events $E_1, E_2, E_3, \ldots, E_n$,

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_i P(E_i) - \sum_{i<j} P(E_i E_j) + \sum_{i<j<k} P(E_i E_j E_k) - \sum_{i<j<k<l} P(E_i E_j E_k E_l) + \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n) \tag{1.4}$$

In words, Eq. (1.4), known as the *inclusion–exclusion identity*, states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time minus the sum of the probabilities of these events taken two at a time plus the sum of the probabilities of these events taken three at a time, and so on.

1.4 Conditional Probabilities

Suppose that we toss two dice and that each of the 36 possible outcomes is equally likely to occur and hence has probability $\frac{1}{36}$. Suppose that we observe that the first die is a four. Then, given this information, what is the probability that the sum of the two dice equals six? To calculate this probability we reason as follows: Given that the initial die is a four, it follows that there can be at most six possible outcomes of our experiment, namely, $(4, 1)$, $(4, 2)$, $(4, 3)$, $(4, 4)$, $(4, 5)$, and $(4, 6)$. Since each of these outcomes originally had the same probability of occurring, they should still have equal probabilities. That is, given that the first die is a four, then the (conditional) probability of each of the outcomes $(4, 1)$, $(4, 2)$, $(4, 3)$, $(4, 4)$, $(4, 5)$, $(4, 6)$ is $\frac{1}{6}$ while the (conditional) probability of the other 30 points in the sample space is 0. Hence, the desired probability will be $\frac{1}{6}$.

If we let E and F denote, respectively, the event that the sum of the dice is six and the event that the first die is a four, then the probability just obtained is called the conditional probability that E occurs given that F has occurred and is denoted by

$$P(E|F)$$

A general formula for $P(E|F)$ that is valid for all events E and F is derived in the same manner as the preceding. Namely, if the event F occurs, then in order for E to occur it is necessary for the actual occurrence to be a point in both E and in F, that is, it must be in EF. Now, because we know that F has occurred, it follows that F becomes our new sample space and hence the probability that the event EF occurs will equal the probability of EF relative to the probability of F. That is,

$$P(E|F) = \frac{P(EF)}{P(F)} \tag{1.5}$$

Note that Eq. (1.5) is only well defined when $P(F) > 0$ and hence $P(E|F)$ is only defined when $P(F) > 0$.

Example 1.4. Suppose cards numbered one through ten are placed in a hat, mixed up, and then one of the cards is drawn. If we are told that the number on the drawn card is at least five, then what is the conditional probability that it is ten?

Solution: Let E denote the event that the number of the drawn card is ten, and let F be the event that it is at least five. The desired probability is $P(E|F)$. Now, from Eq. (1.5)

$$P(E|F) = \frac{P(EF)}{P(F)}$$

However, $EF = E$ since the number of the card will be both ten and at least five if and only if it is number ten. Hence,

$$P(E|F) = \frac{\frac{1}{10}}{\frac{6}{10}} = \frac{1}{6}$$ ■

Example 1.5. A family has two children. What is the conditional probability that both are boys given that at least one of them is a boy? Assume that the sample space S is given by $S = \{(b,b), (b,g), (g,b), (g,g)\}$, and all outcomes are equally likely. $((b,g)$ means, for instance, that the older child is a boy and the younger child a girl.)

Solution: Letting B denote the event that both children are boys, and A the event that at least one of them is a boy, then the desired probability is given by

$$\begin{aligned} P(B|A) &= \frac{P(BA)}{P(A)} \\ &= \frac{P(\{(b,b)\})}{P(\{(b,b),(b,g),(g,b)\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \end{aligned}$$ ■

Example 1.6. Bev can either take a course in computers or in chemistry. If Bev takes the computer course, then she will receive an A grade with probability $\frac{1}{2}$; if she takes the chemistry course then she will receive an A grade with probability $\frac{1}{3}$. Bev decides to base her decision on the flip of a fair coin. What is the probability that Bev will get an A in chemistry?

Solution: If we let C be the event that Bev takes chemistry and A denote the event that she receives an A in whatever course she takes, then the desired probability is $P(AC)$. This is calculated by using Eq. (1.5) as follows:

$$\begin{aligned} P(AC) &= P(C)P(A|C) \\ &= \tfrac{1}{2}\tfrac{1}{3} = \tfrac{1}{6} \end{aligned}$$ ■

Example 1.7. Suppose an urn contains seven black balls and five white balls. We draw two balls from the urn without replacement. Assuming that each ball in the urn is equally likely to be drawn, what is the probability that both drawn balls are black?

Solution: Let F and E denote, respectively, the events that the first and second balls drawn are black. Now, given that the first ball selected is black, there are six remaining black balls and five white balls, and so $P(E|F) = \frac{6}{11}$. As $P(F)$ is clearly $\frac{7}{12}$, our desired probability is

$$\begin{aligned} P(EF) &= P(F)P(E|F) \\ &= \tfrac{7}{12}\tfrac{6}{11} = \tfrac{42}{132} \end{aligned}$$

■

Example 1.8. Suppose that each of three men at a party throws his hat into the center of the room. The hats are first mixed up and then each man randomly selects a hat. What is the probability that none of the three men selects his own hat?

Solution: We shall solve this by first calculating the complementary probability that at least one man selects his own hat. Let us denote by $E_i, i = 1, 2, 3$, the event that the ith man selects his own hat. To calculate the probability $P(E_1 \cup E_2 \cup E_3)$, we first note that

$$\begin{aligned} P(E_i) &= \tfrac{1}{3}, \qquad i = 1, 2, 3 \\ P(E_i E_j) &= \tfrac{1}{6}, \qquad i \neq j \\ P(E_1 E_2 E_3) &= \tfrac{1}{6} \end{aligned} \tag{1.6}$$

To see why Eq. (1.6) is correct, consider first

$$P(E_i E_j) = P(E_i)P(E_j|E_i)$$

Now $P(E_i)$, the probability that the ith man selects his own hat, is clearly $\frac{1}{3}$ since he is equally likely to select any of the three hats. On the other hand, given that the ith man has selected his own hat, then there remain two hats that the jth man may select, and as one of these two is his own hat, it follows that with probability $\frac{1}{2}$ he will select it. That is, $P(E_j|E_i) = \frac{1}{2}$ and so

$$P(E_i E_j) = P(E_i)P(E_j|E_i) = \tfrac{1}{3}\tfrac{1}{2} = \tfrac{1}{6}$$

To calculate $P(E_1 E_2 E_3)$ we write

$$\begin{aligned} P(E_1 E_2 E_3) &= P(E_1 E_2)P(E_3|E_1 E_2) \\ &= \tfrac{1}{6}P(E_3|E_1 E_2) \end{aligned}$$

However, given that the first two men get their own hats it follows that the third man must also get his own hat (since there are no other hats left). That is, $P(E_3|E_1 E_2) = 1$ and so

$$P(E_1 E_2 E_3) = \tfrac{1}{6}$$

Now, from Eq. (1.4) we have that

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) - P(E_1E_2) \\ &\quad - P(E_1E_3) - P(E_2E_3) + P(E_1E_2E_3) \\ &= 1 - \tfrac{1}{2} + \tfrac{1}{6} \\ &= \tfrac{2}{3} \end{aligned}$$

Hence, the probability that none of the men selects his own hat is $1 - \frac{2}{3} = \frac{1}{3}$. ■

1.5 Independent Events

Two events E and F are said to be *independent* if

$$P(EF) = P(E)P(F)$$

By Eq. (1.5) this implies that E and F are independent if

$$P(E|F) = P(E)$$

(which also implies that $P(F|E) = P(F)$). That is, E and F are independent if knowledge that F has occurred does not affect the probability that E occurs. That is, the occurrence of E is independent of whether or not F occurs.

Two events E and F that are not independent are said to be *dependent*.

Example 1.9. Suppose we toss two fair dice. Let E_1 denote the event that the sum of the dice is six and F denote the event that the first die equals four. Then

$$P(E_1F) = P(\{4, 2\}) = \tfrac{1}{36}$$

while

$$P(E_1)P(F) = \tfrac{5}{36}\tfrac{1}{6} = \tfrac{5}{216}$$

and hence E_1 and F are not independent. Intuitively, the reason for this is clear for if we are interested in the possibility of throwing a six (with two dice), then we will be quite happy if the first die lands four (or any of the numbers 1, 2, 3, 4, 5) because then we still have a possibility of getting a total of six. On the other hand, if the first die landed six, then we would be unhappy as we would no longer have a chance of getting a total of six. In other words, our chance of getting a total of six depends on the outcome of the first die and hence E_1 and F cannot be independent.

Let E_2 be the event that the sum of the dice equals seven. Is E_2 independent of F? The answer is yes since

$$P(E_2F) = P(\{(4, 3)\}) = \tfrac{1}{36}$$

while

$$P(E_2)P(F) = \tfrac{1}{6}\tfrac{1}{6} = \tfrac{1}{36}$$

We leave it for you to present the intuitive argument why the event that the sum of the dice equals seven is independent of the outcome on the first die. ■

The definition of independence can be extended to more than two events. The events $E_1, E_2, \ldots, E_n$ are said to be independent if for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leqslant n$, of these events

$$P(E_{1'}E_{2'}\cdots E_{r'}) = P(E_{1'})P(E_{2'})\cdots P(E_{r'})$$

Intuitively, the events $E_1, E_2, \ldots, E_n$ are independent if knowledge of the occurrence of any of these events has no effect on the probability of any other event.

Example 1.10 (Pairwise Independent Events That Are Not Independent)**.** Let a ball be drawn from an urn containing four balls, numbered 1, 2, 3, 4. Let $E = \{1, 2\}$, $F = \{1, 3\}$, $G = \{1, 4\}$. If all four outcomes are assumed equally likely, then

$$P(EF) = P(E)P(F) = \tfrac{1}{4},$$
$$P(EG) = P(E)P(G) = \tfrac{1}{4},$$
$$P(FG) = P(F)P(G) = \tfrac{1}{4}$$

However,

$$\tfrac{1}{4} = P(EFG) \neq P(E)P(F)P(G)$$

Hence, even though the events E, F, G are pairwise independent, they are not jointly independent. ■

Example 1.11. There are r players, with player i initially having n_i units, $n_i > 0$, $i = 1, \ldots, r$. At each stage, two of the players are chosen to play a game, with the winner of the game receiving 1 unit from the loser. Any player whose fortune drops to 0 is eliminated, and this continues until a single player has all $n \equiv \sum_{i=1}^{r} n_i$ units, with that player designated as the victor. Assuming that the results of successive games are independent, and that each game is equally likely to be won by either of its two players, find the probability that player i is the victor.

Solution: To begin, suppose that there are n players, with each player initially having 1 unit. Consider player i. Each stage she plays will be equally likely to result in her either winning or losing 1 unit, with the results from each stage being independent. In addition, she will continue to play stages until her fortune becomes either 0 or n. Because this is the same for all players, it follows that each player has the same chance of being the victor. Consequently, each player has player probability $1/n$ of being the victor. Now, suppose these n players are divided into r teams, with team i containing n_i players, $i = 1, \ldots, r$. That is, suppose players $1, \ldots, n_1$ constitute team 1, players $n_1 + 1, \ldots, n_1 + n_2$ constitute team 2 and so on. Then the probability that the victor is a member of team i is n_i/n. But because team i initially has a total fortune of n_i units, $i = 1, \ldots, r$, and each game played by members of different teams results in the fortune of the winner's team increasing

by 1 and that of the loser's team decreasing by 1, it is easy to see that the probability that the victor is from team i is exactly the desired probability. Moreover, our argument also shows that the result is true no matter how the choices of the players in each stage are made. ■

Suppose that a sequence of experiments, each of which results in either a "success" or a "failure," is to be performed. Let E_i, $i \geqslant 1$, denote the event that the ith experiment results in a success. If, for all $i_1, i_2, \ldots, i_n$,

$$P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \prod_{j=1}^{n} P(E_{i_j})$$

we say that the sequence of experiments consists of *independent trials*.

1.6 Bayes' Formula

Let E and F be events. We may express E as

$$E = EF \cup EF^c$$

because in order for a point to be in E, it must either be in both E and F, or it must be in E and not in F. Since EF and EF^c are mutually exclusive, we have that

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \end{aligned} \tag{1.7}$$

Eq. (1.7) states that the probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred, each conditional probability being given as much weight as the event on which it is conditioned has of occurring.

Example 1.12. Consider two urns. The first contains two white and seven black balls, and the second contains five white and six black balls. We flip a fair coin and then draw a ball from the first urn or the second urn depending on whether the outcome was heads or tails. What is the conditional probability that the outcome of the toss was heads given that a white ball was selected?

Solution: Let W be the event that a white ball is drawn, and let H be the event that the coin comes up heads. The desired probability $P(H|W)$ may be calculated as follows:

$$P(H|W) = \frac{P(HW)}{P(W)} = \frac{P(W|H)P(H)}{P(W)}$$

$$= \frac{P(W|H)P(H)}{P(W|H)P(H) + P(W|H^c)P(H^c)}$$
$$= \frac{\frac{2}{9}\frac{1}{2}}{\frac{2}{9}\frac{1}{2} + \frac{5}{11}\frac{1}{2}} = \frac{22}{67}$$ ■

Example 1.13. In answering a question on a multiple-choice test a student either knows the answer or guesses. Let p be the probability that she knows the answer and $1 - p$ the probability that she guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that she answered it correctly?

Solution: Let C and K denote respectively the event that the student answers the question correctly and the event that she actually knows the answer.
Now

$$P(K|C) = \frac{P(KC)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)}$$
$$= \frac{p}{p + (1/m)(1-p)}$$
$$= \frac{mp}{1 + (m-1)p}$$

Thus, for example, if $m = 5$, $p = \frac{1}{2}$, then the probability that a student knew the answer to a question she correctly answered is $\frac{5}{6}$. ■

Example 1.14. A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply he has the disease.) If 0.5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

Solution: Let D be the event that the tested person has the disease, and E the event that his test result is positive. The desired probability $P(D|E)$ is obtained by

$$P(D|E) = \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)}$$
$$= \frac{(0.95)(0.005)}{(0.95)(0.005) + (0.01)(0.995)}$$
$$= \frac{95}{294} \approx 0.323$$

Thus, only 32 percent of those persons whose test results are positive actually have the disease. ■

Eq. (1.7) may be generalized in the following manner. Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$. In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ will occur. By writing

$$E = \bigcup_{i=1}^{n} EF_i$$

and using the fact that the events $EF_i, i = 1, \ldots, n$, are mutually exclusive, we obtain that

$$\begin{aligned} P(E) &= \sum_{i=1}^{n} P(EF_i) \\ &= \sum_{i=1}^{n} P(E|F_i)P(F_i) \end{aligned} \tag{1.8}$$

Thus, Eq. (1.8) shows how, for given events $F_1, F_2, \ldots, F_n$ of which one and only one must occur, we can compute $P(E)$ by first "conditioning" upon which one of the F_i occurs. That is, it states that $P(E)$ is equal to a weighted average of $P(E|F_i)$, each term being weighted by the probability of the event on which it is conditioned.

Suppose now that E has occurred and we are interested in determining which one of the F_j also occurred. By Eq. (1.8) we have that

$$\begin{aligned} P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \end{aligned} \tag{1.9}$$

Eq. (1.9) is known as *Bayes' formula*.

Example 1.15. You know that a certain letter is equally likely to be in any one of three different folders. Let α_i be the probability that you will find your letter upon making a quick examination of folder i if the letter is, in fact, in folder $i, i = 1, 2, 3$. (We may have $\alpha_i < 1$.) Suppose you look in folder 1 and do not find the letter. What is the probability that the letter is in folder 1?

Solution: Let $F_i, i = 1, 2, 3$ be the event that the letter is in folder i; and let E be the event that a search of folder 1 does not come up with the letter. We desire $P(F_1|E)$. From Bayes' formula we obtain

$$\begin{aligned} P(F_1|E) &= \frac{P(E|F_1)P(F_1)}{\sum_{i=1}^{3} P(E|F_i)P(F_i)} \\ &= \frac{(1-\alpha_1)\frac{1}{3}}{(1-\alpha_1)\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{1-\alpha_1}{3-\alpha_1} \end{aligned}$$

■

1.7 Probability Is a Continuous Event Function

We say that the sequence of events $A_1, A_2, \ldots$ is an *increasing sequence* if $A_n \subset A_{n+1}$ for all $n \geq 1$. If $A_n, n \geq 1$ is an increasing sequence of events, we define its limit by

$$\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i$$

Similarly, we say that $A_n, n \geq 1$ is a decreasing sequence of events if $A_{n+1} \subset A_n$ for all $n \geq 1$, and define its limit by

$$\lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i$$

We now show that probability is a continuous event function.

Proposition 1.1. *If $A_n, n \geq 1$ is either an increasing or a decreasing sequence of events, then*

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

Proof. We will prove this when $A_n, n \geq 1$ is an increasing sequence of events, and leave the proof in the decreasing case as an exercise. So, suppose that $A_n, n \geq 1$ is an increasing sequence of events. Now, define the events $B_n, n \geq 1$, by letting B_n be the set of points that are in A_n but were not in any of the events $A_1, \ldots, A_{n-1}$. That is, we let $B_1 = A_1$, and for $n > 1$ let

$$\begin{aligned} B_n &= A_n \cap (\cup_{i=1}^{n-1} A_i)^c \\ &= A_n A_{n-1}^c \end{aligned}$$

where the final equality used that $A_1, A_2, \ldots$ being increasing implies that $\cup_{i=1}^{n-1} A_i = A_{n-1}$. It is easy to see that the events $B_n, n \geq 1$ are mutually exclusive, and are such that

$$\cup_{i=1}^{n} B_i = \cup_{i=1}^{n} A_i = A_n\,,\ n \geq 1$$

and

$$\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$$

Hence,

$$\begin{aligned} P(\lim_{n\to\infty} A_n) &= P(\cup_{i=1}^{\infty} A_i) \\ &= P(\cup_{i=1}^{\infty} B_i) \\ &= \sum_{i=1}^{\infty} P(B_i) \qquad \text{since the } B_i \text{ are mutually exclusive} \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) \\
&= \lim_{n\to\infty} P(\cup_{i=1}^{n} B_i) \\
&= \lim_{n\to\infty} P(\cup_{i=1}^{n} A_i) \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}
$$

■

Example 1.16. Consider a population of individuals and let all individuals initially present constitute the first generation. Let the second generation consist of all offspring of the first generation, and in general let the $(n+1)$st generation consist of all the offspring of individuals of the nth generation. Let A_n denote the event that there are no individuals in the nth generation. Because $A_n \subset A_{n+1}$ it follows that $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i$. Because $\cup_{i=1}^{\infty} A_i$ is the event that the population eventually dies out, it follows from the continuity property of probability that

$$\lim_{n\to\infty} P(A_n) = P(\text{population dies out})$$

■

Exercises

1. A box contains three marbles: one red, one green, and one blue. Consider an experiment that consists of taking one marble from the box then replacing it in the box and drawing a second marble from the box. What is the sample space? If, at all times, each marble in the box is equally likely to be selected, what is the probability of each point in the sample space?

***2.** Repeat Exercise 1 when the second marble is drawn without replacing the first marble.

3. A coin is to be tossed until a head appears twice in a row. What is the sample space for this experiment? If the coin is fair, what is the probability that it will be tossed exactly four times?

4. Let E, F, G be three events. Find expressions for the events that of E, F, G

- **(a)** only F occurs,
- **(b)** both E and F but not G occur,
- **(c)** at least one event occurs,
- **(d)** at least two events occur,
- **(e)** all three events occur,
- **(f)** none occurs,
- **(g)** at most one occurs,
- **(h)** at most two occur.

***5.** An individual uses the following gambling system at Las Vegas. He bets \$1 that the roulette wheel will come up red. If he wins, he quits. If he loses then he makes the same bet a second time only this time he bets \$2; and then regardless of the outcome, quits. Assuming that he has a probability of $\frac{1}{2}$ of winning each

bet, what is the probability that he goes home a winner? Why is this system not used by everyone?

6. Show that $E(F \cup G) = EF \cup EG$.

7. Show that $(E \cup F)^c = E^c F^c$.

8. If $P(E) = 0.9$ and $P(F) = 0.8$, show that $P(EF) \geqslant 0.7$. In general, show that

$$P(EF) \geqslant P(E) + P(F) - 1$$

This is known as Bonferroni's inequality.

***9.** We say that $E \subset F$ if every point in E is also in F. Show that if $E \subset F$, then

$$P(F) = P(E) + P(FE^c) \geqslant P(E)$$

10. Show that

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{i=1}^{n} P(E_i)$$

This is known as Boole's inequality.

Hint: Either use Eq. (1.2) and mathematical induction, or else show that $\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{n} F_i$, where $F_1 = E_1$, $F_i = E_i \bigcap_{j=1}^{i-1} E_j^c$, and use property (iii) of a probability.

11. If two fair dice are tossed, what is the probability that the sum is $i, i = 2, 3, \ldots, 12$?

12. Let E and F be mutually exclusive events in the sample space of an experiment. Suppose that the experiment is repeated until either event E or event F occurs. What does the sample space of this new super experiment look like? Show that the probability that event E occurs before event F is $P(E)/[P(E) + P(F)]$.

Hint: Argue that the probability that the original experiment is performed n times and E appears on the nth time is $P(E) \times (1 - p)^{n-1}$, $n = 1, 2, \ldots$, where $p = P(E) + P(F)$. Add these probabilities to get the desired answer.

13. The dice game craps is played as follows. The player throws two dice, and if the sum is seven or eleven, then she wins. If the sum is two, three, or twelve, then she loses. If the sum is anything else, then she continues throwing until she either throws that number again (in which case she wins) or she throws a seven (in which case she loses). Calculate the probability that the player wins.

14. The probability of winning on a single toss of the dice is p. A starts, and if he fails, he passes the dice to B, who then attempts to win on her toss. They continue tossing the dice back and forth until one of them wins. What are their respective probabilities of winning?

15. Argue that $E = EF \cup EF^c$, $E \cup F = E \cup FE^c$.

16. Use Exercise 15 to show that $P(E \cup F) = P(E) + P(F) - P(EF)$.

***17.** Suppose each of three persons tosses a coin. If the outcome of one of the tosses differs from the other outcomes, then the game ends. If not, then the persons start over and retoss their coins. Assuming fair coins, what is the probability that

the game will end with the first round of tosses? If all three coins are biased and have probability $\frac{1}{4}$ of landing heads, what is the probability that the game will end at the first round?

18. Assume that each child who is born is equally likely to be a boy or a girl. If a family has two children, what is the probability that both are girls given that (a) the eldest is a girl, (b) at least one is a girl?

***19.** Two dice are rolled. What is the probability that at least one is a six? If the two faces are different, what is the probability that at least one is a six?

20. Three dice are thrown. What is the probability the same number appears on exactly two of the three dice?

21. Suppose that 5 percent of men and 0.25 percent of women are colorblind. A randomly chosen person is colorblind. What is the probability of this person being male? Assume that there are an equal number of males and females.

22. A and B play until one has 2 more points than the other. Assuming that each point is independently won by A with probability p, what is the probability they will play a total of $2n$ points? What is the probability that A will win?

23. For events $E_1, E_2, \ldots, E_n$ show that

$$P(E_1E_2\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

24. In an election, candidate A receives n votes and candidate B receives m votes, where $n > m$. Assume that in the count of the votes all possible orderings of the $n+m$ votes are equally likely. Let $P_{n,m}$ denote the probability that from the first vote on A is always in the lead. Find

(a) $P_{2,1}$ **(b)** $P_{3,1}$ **(c)** $P_{n,1}$ **(d)** $P_{3,2}$ **(e)** $P_{4,2}$
(f) $P_{n,2}$ **(g)** $P_{4,3}$ **(h)** $P_{5,3}$ **(i)** $P_{5,4}$
(j) Make a conjecture as to the value of $P_{n,m}$.

***25.** Two cards are randomly selected from a deck of 52 playing cards.
(a) What is the probability they constitute a pair (that is, that they are of the same denomination)?
(b) What is the conditional probability they constitute a pair given that they are of different suits?

26. A deck of 52 playing cards, containing all 4 aces, is randomly divided into 4 piles of 13 cards each. Define events E_1, E_2, E_3, and E_4 as follows:

$E_1 = \{\text{the first pile has exactly 1 ace}\}$,
$E_2 = \{\text{the second pile has exactly 1 ace}\}$,
$E_3 = \{\text{the third pile has exactly 1 ace}\}$,
$E_4 = \{\text{the fourth pile has exactly 1 ace}\}$

Use Exercise 23 to find $P(E_1E_2E_3E_4)$, the probability that each pile has an ace.

***27.** Suppose in Exercise 26 we had defined the events $E_i, i = 1, 2, 3, 4$, by

$$
\begin{aligned}
E_1 &= \{\text{one of the piles contains the ace of spades}\},\\
E_2 &= \{\text{the ace of spades and the ace of hearts are in different piles}\},\\
E_3 &= \{\text{the ace of spades, the ace of hearts,}\\
&\qquad \text{and the ace of diamonds are in different piles}\},\\
E_4 &= \{\text{all 4 aces are in different piles}\}
\end{aligned}
$$

Now use Exercise 23 to find $P(E_1E_2E_3E_4)$, the probability that each pile has an ace. Compare your answer with the one you obtained in Exercise 26.

28. If the occurrence of B makes A more likely, does the occurrence of A make B more likely?

29. Suppose that $P(E) = 0.6$. What can you say about $P(E|F)$ when
- **(a)** E and F are mutually exclusive?
- **(b)** $E \subset F$?
- **(c)** $F \subset E$?

***30.** Bill and George go target shooting together. Both shoot at a target at the same time. Suppose Bill hits the target with probability 0.7, whereas George, independently, hits the target with probability 0.4.
- **(a)** Given that exactly one shot hit the target, what is the probability that it was George's shot?
- **(b)** Given that the target is hit, what is the probability that George hit it?

31. What is the conditional probability that the first die is six given that the sum of the dice is seven?

***32.** Suppose all n men at a party throw their hats in the center of the room. Each man then randomly selects a hat. Show that the probability that none of the n men selects his own hat is

$$\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - + \cdots \frac{(-1)^n}{n!}$$

Note that as $n \to \infty$ this converges to e^{-1}. Is this surprising?

33. The winner of a tennis match is the first player to win 2 sets. A golden set occurs when one of the players wins all 24 points of a set. Supposing that the results of successive points are independent and that each point is equally likely to be won by either player, find the probability that at least one of the sets of a match is golden.

34. There is a 40 percent chance that A can fix her busted computer. If A cannot, then there is a 20 percent chance that her friend B can fix it. Find the probability it will be fixed by either A or B.

35. A fair coin is continually flipped. What is the probability that the first four flips are
- **(a)** H, H, H, H?
- **(b)** T, H, H, H?

(c) What is the probability that the pattern T, H, H, H occurs before the pattern H, H, H, H?

36. Consider two boxes, one containing one black and one white marble, the other, two black and one white marble. A box is selected at random and a marble is drawn at random from the selected box. What is the probability that the marble is black?

37. In Exercise 36, what is the probability that the first box was the one selected given that the marble is white?

38. Urn 1 contains two white balls and one black ball, while urn 2 contains one white ball and five black balls. One ball is drawn at random from urn 1 and placed in urn 2. A ball is then drawn from urn 2. It happens to be white. What is the probability that the transferred ball was white?

39. Stores A, B, and C have 50, 75, and 100 employees, and, respectively, 50, 60, and 70 percent of these are women. Resignations are equally likely among all employees, regardless of sex. One employee resigns and this is a woman. What is the probability that she works in store C?

***40.** **(a)** A gambler has in his pocket a fair coin and a two-headed coin. He selects one of the coins at random, and when he flips it, it shows heads. What is the probability that it is the fair coin?

(b) Suppose that he flips the same coin a second time and again it shows heads. Now what is the probability that it is the fair coin?

(c) Suppose that he flips the same coin a third time and it shows tails. Now what is the probability that it is the fair coin?

41. In a certain species of rats, black dominates over brown. Suppose that a black rat with two black parents has a brown sibling.

(a) What is the probability that this rat is a pure black rat (as opposed to being a hybrid with one black and one brown gene)?

(b) Suppose that when the black rat is mated with a brown rat, all five of their offspring are black. Now, what is the probability that the rat is a pure black rat?

42. There are three coins in a box. One is a two-headed coin, another is a fair coin, and the third is a biased coin that comes up heads 75 percent of the time. When one of the three coins is selected at random and flipped, it shows heads. What is the probability that it was the two-headed coin?

43. The blue-eyed gene for eye color is recessive, meaning that both the eye genes of an individual must be blue for that individual to be blue eyed. Jo (F) and Joe (M) are both brown-eyed individuals whose mothers had blue eyes. Their daughter Flo, who has brown eyes, is expecting a child conceived with a blue-eyed man. What is the probability that this child will be blue eyed?

44. Urn 1 has five white and seven black balls. Urn 2 has three white and twelve black balls. We flip a fair coin. If the outcome is heads, then a ball from urn 1 is selected, while if the outcome is tails, then a ball from urn 2 is selected. Suppose that a white ball is selected. What is the probability that the coin landed tails?

***45.** An urn contains b black balls and r red balls. One of the balls is drawn at random, but when it is put back in the urn c additional balls of the same color are put in with it. Now suppose that we draw another ball. Show that the probability that the first ball drawn was black given that the second ball drawn was red is $b/(b+r+c)$.

46. Three prisoners are informed by their jailer that one of them has been chosen at random to be executed, and the other two are to be freed. Prisoner A asks the jailer to tell him privately which of his fellow prisoners will be set free, claiming that there would be no harm in divulging this information, since he already knows that at least one will go free. The jailer refuses to answer this question, pointing out that if A knew which of his fellows were to be set free, then his own probability of being executed would rise from $\frac{1}{3}$ to $\frac{1}{2}$, since he would then be one of two prisoners. What do you think of the jailer's reasoning?

47. For a fixed event B, show that the collection $P(A|B)$, defined for all events A, satisfies the three conditions for a probability. Conclude from this that

$$P(A|B) = P(A|BC)P(C|B) + P(A|BC^c)P(C^c|B)$$

Then directly verify the preceding equation.

***48.** Sixty percent of the families in a certain community own their own car, thirty percent own their own home, and twenty percent own both their own car and their own home. If a family is randomly chosen, what is the probability that this family owns a car or a house but not both?

49. Prove Proposition 1.1 for a sequence of decreasing events.

50. If $A_1, A_2, \ldots$ is a sequence of events then $\limsup_{n\to\infty} A_n$ is defined as the set of points that are in an infinite number of the events $A_n, n \geq 1$; and $\liminf_{n\to\infty} A_n$ is defined as the set of points that are in all but a finite number of the events $A_n, n \geq 1$.

(a) If $A_n, n \geq 1$ is an increasing sequence of events, show that

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i.$$

(b) If $A_n, n \geq 1$ is a decreasing sequence of events, show that

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i.$$

51. There is a 40 percent chance of rain on Monday; a 30 percent chance of rain on Tuesday; and a 20 percent chance of rain on both days. It did not rain on Monday. What is the probability it will rain on Tuesday.

References

Reference [2] provides a colorful introduction to some of the earliest developments in probability theory. References [3,4], and [7] are all excellent introductory texts in modern probability theory. Reference [5] is the definitive work that established the axiomatic foundation of modern mathematical probability theory. Reference [6] is a nonmathematical introduction to probability theory and its applications, written by one of the greatest mathematicians of the eighteenth century.

[1] L. Breiman, Probability, Addison-Wesley, Reading, Massachusetts, 1968.
[2] F.N. David, Games, Gods, and Gambling, Hafner, New York, 1962.
[3] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, John Wiley, New York, 1957.
[4] B.V. Gnedenko, Theory of Probability, Chelsea, New York, 1962.
[5] A.N. Kolmogorov, Foundations of the Theory of Probability, Chelsea, New York, 1956.
[6] Marquis de Laplace, A Philosophical Essay on Probabilities, 1825 (English Translation), Dover, New York, 1951.
[7] S. Ross, A First Course in Probability, Tenth Edition, Prentice Hall, New Jersey, 2018.

Random Variables

2

2.1 Random Variables

It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1). These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as *random variables*.

Since the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example 2.1. Letting X denote the random variable that is defined as the sum of two fair dice; then

$$\begin{aligned}
P\{X=2\} &= P\{(1,1)\} = \tfrac{1}{36},\\
P\{X=3\} &= P\{(1,2),(2,1)\} = \tfrac{2}{36},\\
P\{X=4\} &= P\{(1,3),(2,2),(3,1)\} = \tfrac{3}{36},\\
P\{X=5\} &= P\{(1,4),(2,3),(3,2),(4,1)\} = \tfrac{4}{36},\\
P\{X=6\} &= P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = \tfrac{5}{36},\\
P\{X=7\} &= P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \tfrac{6}{36},\\
P\{X=8\} &= P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \tfrac{5}{36},\\
P\{X=9\} &= P\{(3,6),(4,5),(5,4),(6,3)\} = \tfrac{4}{36},\\
P\{X=10\} &= P\{(4,6),(5,5),(6,4)\} = \tfrac{3}{36},\\
P\{X=11\} &= P\{(5,6),(6,5)\} = \tfrac{2}{36},\\
P\{X=12\} &= P\{(6,6)\} = \tfrac{1}{36}
\end{aligned} \tag{2.1}$$

In other words, the random variable X can take on any integral value between two and twelve, and the probability that it takes on each value is given by Eq. (2.1). Since X must take on one of the values two through twelve, we must have

$$1 = P\left\{\bigcup_{n=2}^{12}\{X=n\}\right\} = \sum_{n=2}^{12} P\{X=n\}$$

which may be checked from Eq. (2.1). ■

Example 2.2. For a second example, suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00007-X

variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{Y=0\} = P\{(T,T)\} = \tfrac{1}{4},$$
$$P\{Y=1\} = P\{(T,H),(H,T)\} = \tfrac{2}{4},$$
$$P\{Y=2\} = P\{(H,H)\} = \tfrac{1}{4}$$

Of course, $P\{Y=0\} + P\{Y=1\} + P\{Y=2\} = 1$. ■

Example 2.3. Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \ldots$, with respective probabilities

$$P\{N=1\} = P\{H\} = p,$$
$$P\{N=2\} = P\{(T,H)\} = (1-p)p,$$
$$P\{N=3\} = P\{(T,T,H)\} = (1-p)^2 p,$$
$$\vdots$$
$$P\{N=n\} = P\{(\underbrace{T,T,\ldots,T}_{n-1},H)\} = (1-p)^{n-1}p, \qquad n \geq 1$$

As a check, note that

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty}\{N=n\}\right) &= \sum_{n=1}^{\infty} P\{N=n\} \\ &= p\sum_{n=1}^{\infty}(1-p)^{n-1} \\ &= \frac{p}{1-(1-p)} \\ &= 1 \end{aligned}$$

■

Example 2.4. Suppose that our experiment consists of seeing how long a battery can operate before wearing down. Suppose also that we are not primarily interested in the actual lifetime of the battery but are concerned only about whether or not the battery lasts at least two years. In this case, we may define the random variable I by

$$I = \begin{cases} 1, & \text{if the lifetime of battery is two or more years} \\ 0, & \text{otherwise} \end{cases}$$

If E denotes the event that the battery lasts two or more years, then the random variable I is known as the *indicator* random variable for event E. (Note that I equals 1 or 0 depending on whether or not E occurs.) ■

Example 2.5. Suppose that independent trials, each of which results in any of m possible outcomes with respective probabilities $p_1, \ldots, p_m, \sum_{i=1}^{m} p_i = 1$, are continually performed. Let X denote the number of trials needed until each outcome has occurred at least once.

Rather than directly considering $P\{X = n\}$ we will first determine $P\{X > n\}$, the probability that at least one of the outcomes has not yet occurred after n trials. Letting A_i denote the event that outcome i has not yet occurred after the first n trials, $i = 1, \ldots, m$, then

$$\begin{aligned} P\{X > n\} &= P\left(\bigcup_{i=1}^{m} A_i\right) \\ &= \sum_{i=1}^{m} P(A_i) - \sum\sum_{i<j} P(A_i A_j) \\ &\quad + \sum\sum\sum_{i<j<k} P(A_i A_j A_k) - \cdots + (-1)^{m+1} P(A_1 \cdots A_m) \end{aligned}$$

Now, $P(A_i)$ is the probability that each of the first n trials results in a non-i outcome, and so by independence

$$P(A_i) = (1 - p_i)^n$$

Similarly, $P(A_i A_j)$ is the probability that the first n trials all result in a non-i and non-j outcome, and so

$$P(A_i A_j) = (1 - p_i - p_j)^n$$

As all of the other probabilities are similar, we see that

$$\begin{aligned} P\{X > n\} &= \sum_{i=1}^{m} (1 - p_i)^n - \sum\sum_{i<j} (1 - p_i - p_j)^n \\ &\quad + \sum\sum\sum_{i<j<k} (1 - p_i - p_j - p_k)^n - \cdots \end{aligned}$$

Since $P\{X = n\} = P\{X > n - 1\} - P\{X > n\}$, we see, upon using the algebraic identity $(1 - a)^{n-1} - (1 - a)^n = a(1 - a)^{n-1}$, that

$$\begin{aligned} P\{X = n\} &= \sum_{i=1}^{m} p_i (1 - p_i)^{n-1} - \sum\sum_{i<j} (p_i + p_j)(1 - p_i - p_j)^{n-1} \\ &\quad + \sum\sum\sum_{i<j<k} (p_i + p_j + p_k)(1 - p_i - p_j - p_k)^{n-1} - \cdots \end{aligned}$$

■

In all of the preceding examples, the random variables of interest took on either a finite or a countable number of possible values.[1] Such random variables are called *discrete*. However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The *cumulative distribution function* (cdf) (or more simply the *distribution function*) $F(\cdot)$ of the random variable X is defined for any real number b, $-\infty < b < \infty$, by

$$F(b) = P\{X \leq b\}$$

In words, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b. Some properties of the cdf F are

(i) $F(b)$ is a nondecreasing function of b,
(ii) $\lim_{b\to\infty} F(b) = F(\infty) = 1$,
(iii) $\lim_{b\to-\infty} F(b) = F(-\infty) = 0$.

Property (i) follows since for $a < b$ the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$, and so it must have a smaller probability. Properties (ii) and (iii) follow since X must take on some finite value.

All probability questions about X can be answered in terms of the cdf $F(\cdot)$. For example,

$$P\{a < X \leq b\} = F(b) - F(a) \qquad \text{for all } a < b$$

This follows since we may calculate $P\{a < X \leq b\}$ by first computing the probability that $X \leq b$ (that is, $F(b)$) and then subtracting from this the probability that $X \leq a$ (that is, $F(a)$).

If we desire the probability that X is strictly smaller than b, we may calculate this probability by

$$\begin{aligned} P\{X < b\} &= \lim_{h\to 0^+} P\{X \leq b - h\} \\ &= \lim_{h\to 0^+} F(b - h) \end{aligned}$$

where $\lim_{h\to 0^+}$ means that we are taking the limit as h decreases to 0. Note that $P\{X < b\}$ does not necessarily equal $F(b)$ since $F(b)$ also includes the probability that X equals b.

[1] A set is countable if its elements can be put in a one-to-one correspondence with the sequence of positive integers.

2.2 Discrete Random Variables

As was previously mentioned, a random variable that can take on at most a countable number of possible values is said to be *discrete*. For a discrete random variable X, we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$\begin{aligned} p(x_i) &> 0, \qquad i = 1, 2, \ldots \\ p(x) &= 0, \qquad \text{all other values of } x \end{aligned}$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x_i \le a} p(x_i)$$

For instance, suppose X has a probability mass function given by

$$p(1) = \tfrac{1}{2}, \qquad p(2) = \tfrac{1}{3}, \qquad p(3) = \tfrac{1}{6}$$

then, the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \le a < 2 \\ \frac{5}{6}, & 2 \le a < 3 \\ 1, & 3 \le a \end{cases}$$

This is graphically presented in Fig. 2.1.

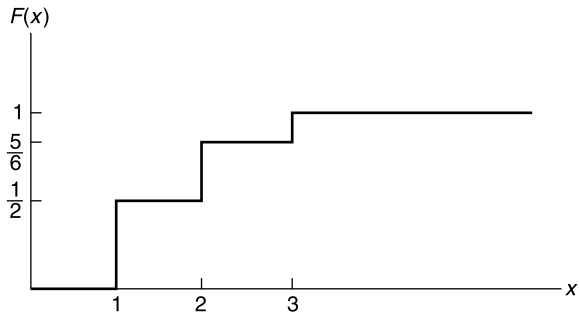

Figure 2.1 Graph of $F(x)$.

Discrete random variables are often classified according to their probability mass functions. We now consider some of these random variables.

2.2.1 The Bernoulli Random Variable

Suppose that a trial, or an experiment, whose outcome can be classified as either a "success" or as a "failure" is performed. If we let X equal 1 if the outcome is a success and 0 if it is a failure, then the probability mass function of X is given by

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p, \\ p(1) &= P\{X = 1\} = p \end{aligned} \tag{2.2}$$

where $p, 0 \le p \le 1$, is the probability that the trial is a "success."

A random variable X is said to be a *Bernoulli* random variable if its probability mass function is given by Eq. (2.2) for some $p \in (0, 1)$.

2.2.2 The Binomial Random Variable

Suppose that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p).

The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, \qquad i = 0, 1, \ldots, n \tag{2.3}$$

where

$$\binom{n}{i} = \frac{n!}{(n - i)!\, i!}$$

equals the number of different groups of i objects that can be chosen from a set of n objects. The validity of Eq. (2.3) may be verified by first noting that the probability of any particular sequence of the n outcomes containing i successes and $n - i$ failures is, by the assumed independence of trials, $p^i (1 - p)^{n-i}$. Eq. (2.3) then follows since there are $\binom{n}{i}$ different sequences of the n outcomes leading to i successes and $n - i$ failures. For instance, if $n = 3, i = 2$, then there are $\binom{3}{2} = 3$ ways in which the three trials can result in two successes. Namely, any one of the three outcomes $(s, s, f), (s, f, s), (f, s, s)$, where the outcome (s, s, f) means that the first two trials are successes and the third a failure. Since each of the three outcomes $(s, s, f), (s, f, s), (f, s, s)$ has a probability $p^2(1 - p)$ of occurring the desired probability is thus $\binom{3}{2} p^2 (1 - p)$.

Note that, by the binomial theorem, the probabilities sum to one, that is,

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1$$

Example 2.6. Four fair coins are flipped. If the outcomes are assumed independent, what is the probability that two heads and two tails are obtained?

Solution: Letting X equal the number of heads ("successes") that appear, then X is a binomial random variable with parameters $(n = 4, p = \frac{1}{2})$. Hence, by Eq. (2.3),

$$P\{X = 2\} = \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$$ ■

Example 2.7. It is known that any item produced by a certain machine will be defective with probability 0.1, independently of any other item. What is the probability that in a sample of three items, at most one will be defective?

Solution: If X is the number of defective items in the sample, then X is a binomial random variable with parameters (3, 0.1). Hence, the desired probability is given by

$$P\{X = 0\} + P\{X = 1\} = \binom{3}{0} (0.1)^0 (0.9)^3 + \binom{3}{1} (0.1)^1 (0.9)^2 = 0.972$$ ■

Example 2.8. Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine; suppose that the airplane will make a successful flight if at least 50 percent of its engines remain operative. For what values of p is a four-engine plane preferable to a two-engine plane?

Solution: Because each engine is assumed to fail or function independently of what happens with the other engines, it follows that the number of engines remaining operative is a binomial random variable. Hence, the probability that a four-engine plane makes a successful flight is

$$\begin{aligned} &\binom{4}{2} p^2 (1-p)^2 + \binom{4}{3} p^3 (1-p) + \binom{4}{4} p^4 (1-p)^0 \\ &\quad = 6p^2(1-p)^2 + 4p^3(1-p) + p^4 \end{aligned}$$

whereas the corresponding probability for a two-engine plane is

$$\binom{2}{1} p(1-p) + \binom{2}{2} p^2 = 2p(1-p) + p^2$$

Hence the four-engine plane is safer if

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 \geq 2p(1-p) + p^2$$

or equivalently if

$$6p(1-p)^2 + 4p^2(1-p) + p^3 \geq 2 - p$$

which simplifies to

$$3p^3 - 8p^2 + 7p - 2 \geq 0 \quad \text{or} \quad (p-1)^2(3p-2) \geq 0$$

which is equivalent to

$$3p - 2 \geq 0 \quad \text{or} \quad p \geq \tfrac{2}{3}$$

Hence, the four-engine plane is safer when the engine success probability is at least as large as $\frac{2}{3}$, whereas the two-engine plane is safer when this probability falls below $\frac{2}{3}$. ■

Example 2.9. Suppose that a particular trait of a person (such as eye color or left handedness) is classified on the basis of one pair of genes and suppose that d represents a dominant gene and r a recessive gene. Thus a person with dd genes is pure dominance, one with rr is pure recessive, and one with rd is hybrid. The pure dominance and the hybrid are alike in appearance. Children receive one gene from each parent. If, with respect to a particular trait, two hybrid parents have a total of four children, what is the probability that exactly three of the four children have the outward appearance of the dominant gene?

Solution: If we assume that each child is equally likely to inherit either of two genes from each parent, the probabilities that the child of two hybrid parents will have dd, rr, or rd pairs of genes are, respectively, $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$. Hence, because an offspring will have the outward appearance of the dominant gene if its gene pair is either dd or rd, it follows that the number of such children is binomially distributed with parameters $(4, \frac{3}{4})$. Thus the desired probability is

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 = \frac{27}{64}$$ ■

Remark on Terminology. If X is a binomial random variable with parameters (n, p), then we say that X has a binomial distribution with parameters (n, p).

2.2.3 The Geometric Random Variable

Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. If we let X be the number of trials required until the first success, then X is said to be a *geometric* random variable with parameter p. Its probability mass function is given by

$$p(n) = P\{X = n\} = (1-p)^{n-1}p, \qquad n = 1, 2, \ldots \tag{2.4}$$

Eq. (2.4) follows since in order for X to equal n it is necessary and sufficient that the first $n-1$ trials be failures and the nth trial a success. Eq. (2.4) follows since the outcomes of the successive trials are assumed to be independent.

To check that $p(n)$ is a probability mass function, we note that

$$\sum_{n=1}^{\infty} p(n) = p\sum_{n=1}^{\infty}(1-p)^{n-1} = 1$$

2.2.4 The Poisson Random Variable

A random variable X, taking on one of the values $0, 1, 2, \ldots$, is said to be a *Poisson* random variable with parameter λ, if for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 0, 1, \ldots \tag{2.5}$$

Eq. (2.5) defines a probability mass function since

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1$$

The Poisson random variable has a wide range of applications in a diverse number of areas, as will be seen in Chapter 5.

An important property of the Poisson random variable is that it may be used to approximate a binomial random variable when the binomial parameter n is large and p is small. To see this, suppose that X is a binomial random variable with parameters (n, p), and let $\lambda = np$. Then

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)!\,i!}p^i(1-p)^{n-i} \\ &= \frac{n!}{(n-i)!\,i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

Now, for n large and p small

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \qquad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1, \qquad \left(1-\frac{\lambda}{n}\right)^i \approx 1$$

Hence, for n large and p small,

$$P\{X = i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}$$

Example 2.10. Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one error on this page.

Solution:

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1} \approx 0.632$$ ■

Example 2.11. If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur today?

Solution:

$$P\{X = 0\} = e^{-3} \approx 0.05$$ ■

Example 2.12. Consider an experiment that consists of counting the number of α-particles given off in a one-second interval by one gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than two α-particles will appear?

Solution: If we think of the gram of radioactive material as consisting of a large number n of atoms each of which has probability $3.2/n$ of disintegrating and sending off an α-particle during the second considered, then we see that, to a very close approximation, the number of α-particles given off will be a Poisson random variable with parameter $\lambda = 3.2$. Hence the desired probability is

$$P\{X \leq 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \approx 0.380$$ ■

2.3 Continuous Random Variables

In this section, we shall concern ourselves with random variables whose set of possible values is uncountable. Let X be such a random variable. We say that X is a *continuous* random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)\,dx \tag{2.6}$$

The function $f(x)$ is called the *probability density function* of the random variable X.

In words, Eq. (2.6) states that the probability that X will be in B may be obtained by integrating the probability density function over the set B. Since X must assume some value, $f(x)$ must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx$$

All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain from Eq. (2.6) that

$$P\{a \le X \le b\} = \int_a^b f(x)\,dx \tag{2.7}$$

If we let $a = b$ in the preceding, then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any *particular* value is zero.

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x)\,dx$$

Differentiating both sides of the preceding yields

$$\frac{d}{da}F(a) = f(a)$$

That is, the density is the derivative of the cumulative distribution function. A somewhat more intuitive interpretation of the density function may be obtained from Eq. (2.7) as follows:

$$P\left\{a - \frac{\varepsilon}{2} \le X \le a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\,dx \approx \varepsilon f(a)$$

when ε is small. In other words, the probability that X will be contained in an interval of length ε around the point a is approximately $\varepsilon f(a)$. From this, we see that $f(a)$ is a measure of how likely it is that the random variable will be near a.

There are several important continuous random variables that appear frequently in probability theory. The remainder of this section is devoted to a study of certain of these random variables.

2.3.1 The Uniform Random Variable

A random variable is said to be *uniformly distributed* over the interval $(0, 1)$ if its probability density function is given by

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Note that the preceding is a density function since $f(x) \geq 0$ and

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 dx = 1$$

Since $f(x) > 0$ only when $x \in (0, 1)$, it follows that X must assume a value in $(0, 1)$. Also, since $f(x)$ is constant for $x \in (0, 1)$, X is just as likely to be "near" any value in $(0, 1)$ as any other value. To check this, note that, for any $0 < a < b < 1$,

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,dx = b - a$$

In other words, the probability that X is in any particular subinterval of $(0, 1)$ equals the length of that subinterval.

In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

Example 2.13. Calculate the cumulative distribution function of a random variable uniformly distributed over (α, β).

Solution: Since $F(a) = \int_{-\infty}^{a} f(x)\,dx$, we obtain from Eq. (2.8) that

$$F(a) = \begin{cases} 0, & a \leq \alpha \\ \dfrac{a - \alpha}{\beta - \alpha}, & \alpha < a < \beta \\ 1, & a \geq \beta \end{cases}$$

■

Example 2.14. If X is uniformly distributed over $(0, 10)$, calculate the probability that **(a)** $X < 3$, **(b)** $X > 7$, **(c)** $1 < X < 6$.

Solution:

$$P\{X < 3\} = \frac{\int_0^3 dx}{10} = \frac{3}{10},$$

$$P\{X > 7\} = \frac{\int_7^{10} dx}{10} = \frac{3}{10},$$

$$P\{1 < X < 6\} = \frac{\int_1^6 dx}{10} = \frac{1}{2}$$

■

2.3.2 Exponential Random Variables

A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

is said to be an *exponential random variable* with parameter λ. These random variables will be extensively studied in Chapter 5, so we will content ourselves here with just calculating the cumulative distribution function F:

$$F(a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}, \qquad a \geq 0$$

Note that $F(\infty) = \int_0^\infty \lambda e^{-\lambda x} dx = 1$, as, of course, it must.

2.3.3 Gamma Random Variables

A continuous random variable whose density is given by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x} (\lambda x)^{\alpha - 1}}{\Gamma(\alpha)}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

for some $\lambda > 0, \alpha > 0$ is said to be a *gamma random variable* with parameters α, λ. The quantity $\Gamma(\alpha)$ is called the gamma function and is defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha - 1} \, dx$$

It is easy to show by induction that for integral α, say, $\alpha = n$,

$$\Gamma(n) = (n - 1)!$$

2.3.4 Normal Random Variables

We say that X is a *normal random variable* (or simply that X is normally distributed) with parameters μ and σ^2 if the density of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\, \sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

This density function is a bell-shaped curve that is symmetric around μ (see Fig. 2.2).

An important fact about normal random variables is that if X is normally distributed with parameters μ and σ^2 then $Y = \alpha X + \beta$ is normally distributed with parameters

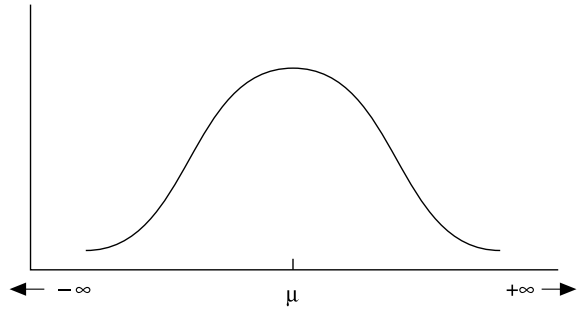

Figure 2.2 Normal density function.

$\alpha\mu + \beta$ and $\alpha^2\sigma^2$. To prove this, suppose first that $\alpha > 0$ and note that $F_Y(\cdot)$,[2] the cumulative distribution function of the random variable Y, is given by

$$\begin{aligned} F_Y(a) &= P\{Y \le a\} \\ &= P\{\alpha X + \beta \le a\} \\ &= P\left\{X \le \frac{a-\beta}{\alpha}\right\} \\ &= F_X\left(\frac{a-\beta}{\alpha}\right) \\ &= \int_{-\infty}^{(a-\beta)/\alpha} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}\,dx \\ &= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\,\alpha\sigma} \exp\left\{\frac{-(v-(\alpha\mu+\beta))^2}{2\alpha^2\sigma^2}\right\} dv \end{aligned} \tag{2.9}$$

where the last equality is obtained by the change in variables $v = \alpha x + \beta$. However, since $F_Y(a) = \int_{-\infty}^{a} f_Y(v)\,dv$, it follows from Eq. (2.9) that the probability density function $f_Y(\cdot)$ is given by

$$f_Y(v) = \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp\left\{\frac{-(v-(\alpha\mu+\beta))^2}{2(\alpha\sigma)^2}\right\}, \qquad -\infty < v < \infty$$

Hence, Y is normally distributed with parameters $\alpha\mu + \beta$ and $(\alpha\sigma)^2$. A similar result is also true when $\alpha < 0$.

One implication of the preceding result is that if X is normally distributed with parameters μ and σ^2 then $Y = (X - \mu)/\sigma$ is normally distributed with parameters 0 and 1. Such a random variable Y is said to have the *standard* or *unit* normal distribution.

[2] When there is more than one random variable under consideration, we shall denote the cumulative distribution function of a random variable Z by $F_z(\cdot)$. Similarly, we shall denote the density of Z by $f_z(\cdot)$.

2.4 Expectation of a Random Variable

2.4.1 The Discrete Case

If X is a discrete random variable having a probability mass function $p(x)$, then the *expected value* of X is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

In other words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes that value. For example, if the probability mass function of X is given by

$$p(1) = \tfrac{1}{2} = p(2)$$

then

$$E[X] = 1\left(\tfrac{1}{2}\right) + 2\left(\tfrac{1}{2}\right) = \tfrac{3}{2}$$

is just an ordinary average of the two possible values 1 and 2 that X can assume. On the other hand, if

$$p(1) = \tfrac{1}{3}, \qquad p(2) = \tfrac{2}{3}$$

then

$$E[X] = 1\left(\tfrac{1}{3}\right) + 2\left(\tfrac{2}{3}\right) = \tfrac{5}{3}$$

is a weighted average of the two possible values 1 and 2 where the value 2 is given twice as much weight as the value 1 since $p(2) = 2p(1)$.

Example 2.15. Find $E[X]$ where X is the outcome when we roll a fair die.

Solution: Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain

$$E[X] = 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right) = \tfrac{7}{2}$$ ■

Example 2.16 (Expectation of a Bernoulli Random Variable)**.** Calculate $E[X]$ when X is a Bernoulli random variable with parameter p.

Solution: Since $p(0) = 1 - p$, $p(1) = p$, we have

$$E[X] = 0(1 - p) + 1(p) = p$$

Thus, the expected number of successes in a single trial is just the probability that the trial will be a success. ■

Example 2.17 (Expectation of a Binomial Random Variable)**.** Calculate $E[X]$ when X is binomially distributed with parameters n and p.

Solution:

$$\begin{aligned}
E[X] &= \sum_{i=0}^{n} ip(i) \\
&= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{in!}{(n-i)!\,i!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n!}{(n-i)!(i-1)!} p^i (1-p)^{n-i} \\
&= np \sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!(i-1)!} p^{i-1} (1-p)^{n-i} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&= np[p + (1-p)]^{n-1} \\
&= np
\end{aligned}$$

where the third from the last equality follows by letting $k = i - 1$. Thus, the expected number of successes in n independent trials is n multiplied by the probability that a trial results in a success. ■

Example 2.18 (Expectation of a Geometric Random Variable)**.** Calculate the expectation of a geometric random variable having parameter p.

Solution: By Eq. (2.4), we have

$$\begin{aligned}
E[X] &= \sum_{n=1}^{\infty} np(1-p)^{n-1} \\
&= p \sum_{n=1}^{\infty} nq^{n-1}
\end{aligned}$$

where $q = 1 - p$,

$$E[X] = p \sum_{n=1}^{\infty} \frac{d}{dq}(q^n)$$

$$
\begin{aligned}
&= p\frac{d}{dq}\left(\sum_{n=1}^{\infty} q^n\right) \\
&= p\frac{d}{dq}\left(\frac{q}{1-q}\right) \\
&= \frac{p}{(1-q)^2} \\
&= \frac{1}{p}
\end{aligned}
$$

In words, the expected number of independent trials we need to perform until we attain our first success equals the reciprocal of the probability that any one trial results in a success. ■

Example 2.19 (Expectation of a Poisson Random Variable)**.** Calculate $E[X]$ if X is a Poisson random variable with parameter λ.

Solution: From Eq. (2.5), we have

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{\infty} \frac{i e^{-\lambda}\lambda^i}{i!} \\
&= \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

where we have used the identity $\sum_{k=0}^{\infty} \lambda^k/k! = e^{\lambda}$. ■

2.4.2 The Continuous Case

We may also define the expected value of a continuous random variable. This is done as follows. If X is a continuous random variable having a probability density function $f(x)$, then the expected value of X is defined by

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

Example 2.20 (Expectation of a Uniform Random Variable)**.** Calculate the expectation of a random variable uniformly distributed over (α, β).

Solution: From Eq. (2.8) we have

$$\begin{aligned} E[X] &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}\, dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

In other words, the expected value of a random variable uniformly distributed over the interval (α, β) is just the midpoint of the interval. ■

Example 2.21 (Expectation of an Exponential Random Variable)**.** Let X be exponentially distributed with parameter λ. Calculate $E[X]$.

Solution:

$$E[X] = \int_0^{\infty} x\lambda e^{-\lambda x}\, dx$$

Integrating by parts ($dv = \lambda e^{-\lambda x}\, dx, u = x$) yields

$$\begin{aligned} E[X] &= -xe^{-\lambda x}\big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx \\ &= 0 - \frac{e^{-\lambda x}}{\lambda}\bigg|_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

■

Example 2.22 (Expectation of a Normal Random Variable)**.** Calculate $E[X]$ when X is normally distributed with parameters μ and σ^2.

Solution:

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} xe^{-(x-\mu)^2/2\sigma^2}\, dx$$

Writing x as $(x - \mu) + \mu$ yields

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)e^{-(x-\mu)^2/2\sigma^2}\, dx + \mu \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

Letting $y = x - \mu$ leads to

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} ye^{-y^2/2\sigma^2} dy + \mu \int_{-\infty}^{\infty} f(x)\, dx$$

where $f(x)$ is the normal density. By symmetry, the first integral must be 0, and so

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

■

2.4.3 Expectation of a Function of a Random Variable

Suppose now that we are given a random variable X and its probability distribution (that is, its probability mass function in the discrete case or its probability density function in the continuous case). Suppose also that we are interested in calculating not the expected value of X, but the expected value of some function of X, say, $g(X)$. How do we go about doing this? One way is as follows. Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of X. Once we have obtained the distribution of $g(X)$, we can then compute $E[g(X)]$ by the definition of the expectation.

Example 2.23. Suppose X has the following probability mass function:

$$p(0) = 0.2, \qquad p(1) = 0.5, \qquad p(2) = 0.3$$

Calculate $E[X^2]$.

Solution: Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values $0^2, 1^2, 2^2$ with respective probabilities

$$\begin{aligned} p_Y(0) &= P\{Y = 0^2\} = 0.2, \\ p_Y(1) &= P\{Y = 1^2\} = 0.5, \\ p_Y(4) &= P\{Y = 2^2\} = 0.3 \end{aligned}$$

Hence,

$$E[X^2] = E[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$$

Note that

$$1.7 = E[X^2] \neq (E[X])^2 = 1.21$$ ■

Example 2.24. Let X be uniformly distributed over $(0, 1)$. Calculate $E[X^3]$.

Solution: Letting $Y = X^3$, we calculate the distribution of Y as follows. For $0 \leq a \leq 1$,

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{X^3 \leq a\} \\ &= P\{X \leq a^{1/3}\} \\ &= a^{1/3} \end{aligned}$$

where the last equality follows since X is uniformly distributed over $(0, 1)$. By differentiating $F_Y(a)$, we obtain the density of Y, namely,

$$f_Y(a) = \tfrac{1}{3}a^{-2/3}, \qquad 0 \leq a \leq 1$$

Hence,

$$\begin{aligned} E[X^3] = E[Y] &= \int_{-\infty}^{\infty} a f_Y(a)\, da \\ &= \int_0^1 a \tfrac{1}{3} a^{-2/3}\, da \\ &= \tfrac{1}{3} \int_0^1 a^{1/3}\, da \\ &= \tfrac{1}{3} \tfrac{3}{4} a^{4/3} \Big|_0^1 \\ &= \tfrac{1}{4} \end{aligned}$$

■

While the foregoing procedure will, in theory, always enable us to compute the expectation of any function of X from a knowledge of the distribution of X, there is, fortunately, an easier way to do this. The following proposition shows how we can calculate the expectation of $g(X)$ without first determining its distribution.

Proposition 2.1. **(a)** *If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,*

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

(b) *If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,*

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

■

Example 2.25. Applying the proposition to Example 2.23 yields

$$E[X^2] = 0^2(0.2) + (1^2)(0.5) + (2^2)(0.3) = 1.7$$

which, of course, checks with the result derived in Example 2.23.

Applying the proposition to Example 2.24 yields

$$\begin{aligned} E[X^3] &= \int_0^1 x^3\, dx \qquad (\text{since } f(x) = 1, 0 < x < 1) \\ &= \tfrac{1}{4} \end{aligned}$$

■

A simple corollary of Proposition 2.1 is the following.

Corollary 2.2. *If a and b are constants, then*

$$E[aX + b] = aE[X] + b$$

Proof. In the discrete case,

$$\begin{aligned} E[aX+b] &= \sum_{x:p(x)>0} (ax+b)p(x) \\ &= a\sum_{x:p(x)>0} xp(x) + b\sum_{x:p(x)>0} p(x) \\ &= aE[X]+b \end{aligned}$$

In the continuous case,

$$\begin{aligned} E[aX+b] &= \int_{-\infty}^{\infty} (ax+b)f(x)\,dx \\ &= a\int_{-\infty}^{\infty} xf(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx \\ &= aE[X]+b \end{aligned}$$

■

The expected value of a random variable X, $E[X]$, is also referred to as the *mean* or the first *moment* of X. The quantity $E[X^n]$, $n \geq 1$, is called the nth moment of X. By Proposition 2.1, we note that

$$E[X^n] = \begin{cases} \displaystyle\sum_{x:p(x)>0} x^n p(x), & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} x^n f(x)\,dx, & \text{if } X \text{ is continuous} \end{cases}$$

Another quantity of interest is the variance of a random variable X, denoted by $\mathrm{Var}(X)$, which is defined by

$$\mathrm{Var}(X) = E\Big[(X - E[X])^2\Big]$$

Thus, the variance of X measures the expected square of the deviation of X from its expected value.

Example 2.26 (Variance of the Normal Random Variable)**.** Let X be normally distributed with parameters μ and σ^2. Find $\mathrm{Var}(X)$.

Solution: Recalling (see Example 2.22) that $E[X] = \mu$, we have that

$$\begin{aligned} \mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}\,dx \end{aligned}$$

Substituting $y = (x-\mu)/\sigma$ yields

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}\,dy$$

Integrating by parts ($u = y, dv = ye^{-y^2/2}dy$) gives

$$\begin{aligned}\text{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}}\left(-ye^{-y^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right)\\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\,dy\\ &= \sigma^2\end{aligned}$$

Another derivation of $\text{Var}(X)$ will be given in Example 2.42. ■

Suppose that X is continuous with density f, and let $E[X] = \mu$. Then,

$$\begin{aligned}\text{Var}(X) &= E[(X-\mu)^2]\\ &= E[X^2 - 2\mu X + \mu^2]\\ &= \int_{-\infty}^{\infty}(x^2 - 2\mu x + \mu^2)f(x)\,dx\\ &= \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2\mu\int_{-\infty}^{\infty} xf(x)\,dx + \mu^2\int_{-\infty}^{\infty} f(x)\,dx\\ &= E[X^2] - 2\mu\mu + \mu^2\\ &= E[X^2] - \mu^2\end{aligned}$$

A similar proof holds in the discrete case, and so we obtain the useful identity

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

Example 2.27. Calculate $\text{Var}(X)$ when X represents the outcome when a fair die is rolled.

Solution: As previously noted in Example 2.15, $E[X] = \frac{7}{2}$. Also,

$$E[X^2] = 1\left(\tfrac{1}{6}\right) + 2^2\left(\tfrac{1}{6}\right) + 3^2\left(\tfrac{1}{6}\right) + 4^2\left(\tfrac{1}{6}\right) + 5^2\left(\tfrac{1}{6}\right) + 6^2\left(\tfrac{1}{6}\right) = \left(\tfrac{1}{6}\right)(91)$$

Hence,

$$\text{Var}(X) = \tfrac{91}{6} - \left(\tfrac{7}{2}\right)^2 = \tfrac{35}{12}$$ ■

2.5 Jointly Distributed Random Variables

2.5.1 Joint Distribution Functions

Thus far, we have concerned ourselves with the probability distribution of a single random variable. However, we are often interested in probability statements concerning two or more random variables. To deal with such probabilities, we define, for any

two random variables X and Y, the *joint cumulative probability distribution function* of X and Y by

$$F(a, b) = P\{X \leq a, Y \leq b\}, \qquad -\infty < a, b < \infty$$

The distribution of X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y < \infty\} \\ &= F(a, \infty) \end{aligned}$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(b) = P\{Y \leq b\} = F(\infty, b)$$

In the case where X and Y are both discrete random variables, it is convenient to define the *joint probability mass function* of X and Y by

$$p(x, y) = P\{X = x, Y = y\}$$

The probability mass function of X may be obtained from $p(x, y)$ by

$$p_X(x) = \sum_{y:p(x,y)>0} p(x, y)$$

Similarly,

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x, y)$$

We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$, defined for all real x and y, having the property that for all sets A and B of real numbers

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. The probability density of X can be obtained from a knowledge of $f(x, y)$ by the following reasoning:

$$\begin{aligned} P\{X \in A\} &= P\{X \in A, Y \in (-\infty, \infty)\} \\ &= \int_{-\infty}^{\infty} \int_A f(x, y)\, dx\, dy \\ &= \int_A f_X(x)\, dx \end{aligned}$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

is thus the probability density function of X. Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

Because

$$F(a, b) = P(X \le a, Y \le b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y)\, dy\, dx$$

differentiation yields

$$\frac{d^2}{da\, db} F(a, b) = f(a, b)$$

Thus, as in the single variable case, differentiating the probability distribution function gives the probability density function.

A variation of Proposition 2.1 states that if X and Y are random variables and g is a function of two variables, then

$$\begin{aligned} E[g(X, Y)] &= \sum_y \sum_x g(x, y) p(x, y) && \text{in the discrete case} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy && \text{in the continuous case} \end{aligned}$$

For example, if $g(X, Y) = X + Y$, then, in the continuous case,

$$\begin{aligned} E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y)\, dx\, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y)\, dx\, dy \\ &= E[X] + E[Y] \end{aligned}$$

where the first integral is evaluated by using the variation of Proposition 2.1 with $g(x, y) = x$, and the second with $g(x, y) = y$.

The same result holds in the discrete case and, combined with the corollary in Section 2.4.3, yields that for any constants a, b

$$E[aX + bY] = aE[X] + bE[Y] \tag{2.10}$$

Joint probability distributions may also be defined for n random variables. The details are exactly the same as when $n = 2$ and are left as an exercise. The corresponding

result to Eq. (2.10) states that if $X_1, X_2, \ldots, X_n$ are n random variables, then for any n constants $a_1, a_2, \ldots, a_n$,

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n] \quad (2.11)$$

Example 2.28. Calculate the expected sum obtained when three fair dice are rolled.

Solution: Let X denote the sum obtained. Then $X = X_1 + X_2 + X_3$ where X_i represents the value of the ith die. Thus,

$$E[X] = E[X_1] + E[X_2] + E[X_3] = 3\left(\frac{7}{2}\right) = \frac{21}{2}$$ ■

Example 2.29. As another example of the usefulness of Eq. (2.11), let us use it to obtain the expectation of a binomial random variable having parameters n and p. Recalling that such a random variable X represents the number of successes in n trials when each trial has probability p of being a success, we have

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

where

$$X_i = \begin{cases} 1, & \text{if the } i\text{th trial is a success} \\ 0, & \text{if the } i\text{th trial is a failure} \end{cases}$$

Hence, X_i is a Bernoulli random variable having expectation $E[X_i] = 1(p) + 0(1 - p) = p$. Thus,

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = np$$

This derivation should be compared with the one presented in Example 2.17. ■

Example 2.30. At a party N men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Find the expected number of men who select their own hats.

Solution: Letting X denote the number of men that select their own hats, we can best compute $E[X]$ by noting that

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

where

$$X_i = \begin{cases} 1, & \text{if the } i\text{th man selects his own hat} \\ 0, & \text{otherwise} \end{cases}$$

Now, because the ith man is equally likely to select any of the N hats, it follows that

$$P\{X_i = 1\} = P\{i\text{th man selects his own hat}\} = \frac{1}{N}$$

and so

$$E[X_i] = 1P\{X_i = 1\} + 0P\{X_i = 0\} = \frac{1}{N}$$

Hence, from Eq. (2.11) we obtain

$$E[X] = E[X_1] + \cdots + E[X_N] = \left(\frac{1}{N}\right) N = 1$$

Hence, no matter how many people are at the party, on the average exactly one of the men will select his own hat. ■

Example 2.31. Suppose there are 25 different types of coupons and suppose that each time one obtains a coupon, it is equally likely to be any one of the 25 types. Compute the expected number of different types that are contained in a set of 10 coupons.

Solution: Let X denote the number of different types in the set of 10 coupons. We compute $E[X]$ by using the representation

$$X = X_1 + \cdots + X_{25}$$

where

$$X_i = \begin{cases} 1, & \text{if at least one type } i \text{ coupon is in the set of 10} \\ 0, & \text{otherwise} \end{cases}$$

Now,

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= P\{\text{at least one type } i \text{ coupon is in the set of 10}\} \\ &= 1 - P\{\text{no type } i \text{ coupons are in the set of 10}\} \\ &= 1 - \left(\tfrac{24}{25}\right)^{10} \end{aligned}$$

when the last equality follows since each of the 10 coupons will (independently) not be a type i with probability $\frac{24}{25}$. Hence,

$$E[X] = E[X_1] + \cdots + E[X_{25}] = 25\left[1 - \left(\tfrac{24}{25}\right)^{10}\right]$$ ■

Example 2.32. Let $R_1, \ldots, R_{n+m}$ be a random permutation of $1, \ldots, n+m$. (That is, $R_1, \ldots, R_{n+m}$ is equally likely to be any of the $(n+m)!$ permutations of $1, \ldots, n+m$.) For a given $i \le n$, let X be the ith smallest of the values $R_1, \ldots, R_n$. Find $E[X]$.

Solution: If we let N be the number of the values $R_{n+1}, \ldots, R_{n+m}$ that are smaller than X, then X is the $(i+N)$th smallest of all the values $R_1, \ldots, R_{n+m}$. Because $R_1, \ldots, R_{n+m}$ consists of all numbers $1, \ldots, n+m$, it follows that $X = i + N$. Consequently,

$$E[X] = i + E[N]$$

To compute $E[N]$, for $k = 1, \ldots, m$ let I_{n+k} equal 1 if $R_{n+k} < X$ and let it equal 0 otherwise. Using that

$$N = \sum_{k=1}^{m} I_{n+k}$$

we obtain that

$$E[X] = i + \sum_{k=1}^{m} E[I_{n+k}]$$

Now,

$$\begin{aligned} E[I_{n+k}] &= P(R_{n+k} < X) \\ &= P(R_{n+k} < i\text{th smallest of } R_1, \ldots, R_n) \\ &= P(R_{n+k} \text{ is one of the } i \text{ smallest of the values } R_1, \ldots, R_n, R_{n+k}) \\ &= \frac{i}{n+1} \end{aligned}$$

where the final equality used that R_{n+k} is equally likely to be either the smallest, the second smallest, ..., or the $(n+1)$st smallest of the values $R_1, \ldots, R_n, R_{n+k}$. Hence,

$$E[X] = i + m\frac{i}{n+1}$$ ■

2.5.2 Independent Random Variables

The random variables X and Y are said to be *independent* if, for all a, b,

$$P\{X \le a, Y \le b\} = P\{X \le a\}P\{Y \le b\} \tag{2.12}$$

In other words, X and Y are independent if, for all a and b, the events $E_a = \{X \le a\}$ and $F_b = \{Y \le b\}$ are independent.

In terms of the joint distribution function F of X and Y, we have that X and Y are independent if

$$F(a, b) = F_X(a)F_Y(b) \qquad \text{for all } a, b$$

When X and Y are discrete, the condition of independence reduces to

$$p(x, y) = p_X(x)p_Y(y) \tag{2.13}$$

while if X and Y are jointly continuous, independence reduces to

$$f(x, y) = f_X(x)f_Y(y) \tag{2.14}$$

To prove this statement, consider first the discrete version, and suppose that the joint probability mass function $p(x, y)$ satisfies Eq. (2.13). Then

$$\begin{aligned} P\{X \le a, Y \le b\} &= \sum_{y \le b} \sum_{x \le a} p(x, y) \\ &= \sum_{y \le b} \sum_{x \le a} p_X(x) p_Y(y) \\ &= \sum_{y \le b} p_Y(y) \sum_{x \le a} p_X(x) \\ &= P\{Y \le b\} P\{X \le a\} \end{aligned}$$

and so X and Y are independent. That Eq. (2.14) implies independence in the continuous case is proven in the same manner and is left as an exercise.

An important result concerning independence is the following.

Proposition 2.3. *If X and Y are independent, then for any functions h and g*

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Proof. Suppose that X and Y are jointly continuous. Then

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y) f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y) f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} h(y) f_Y(y)\, dy \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \\ &= E[h(Y)]E[g(X)] \end{aligned}$$

The proof in the discrete case is similar. ■

2.5.3 Covariance and Variance of Sums of Random Variables

The covariance of any two random variables X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

$$\begin{aligned} \mathrm{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\ &= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Note that if X and Y are independent, then by Proposition 2.3 it follows that $\mathrm{Cov}(X, Y) = 0$.

Let us consider now the special case where X and Y are indicator variables for whether or not the events A and B occur. That is, for events A and B, define

$$X = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise,} \end{cases} \qquad Y = \begin{cases} 1, & \text{if } B \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

Then,

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

and, because XY will equal 1 or 0 depending on whether or not both X and Y equal 1, we see that

$$\text{Cov}(X, Y) = P\{X = 1, Y = 1\} - P\{X = 1\}P\{Y = 1\}$$

From this we see that

$$\begin{aligned} \text{Cov}(X, Y) > 0 &\Leftrightarrow P\{X = 1, Y = 1\} > P\{X = 1\}P\{Y = 1\} \\ &\Leftrightarrow \frac{P\{X = 1, Y = 1\}}{P\{X = 1\}} > P\{Y = 1\} \\ &\Leftrightarrow P\{Y = 1|X = 1\} > P\{Y = 1\} \end{aligned}$$

That is, the covariance of X and Y is positive if the outcome $X = 1$ makes it more likely that $Y = 1$ (which, as is easily seen by symmetry, also implies the reverse).

In general it can be shown that a positive value of $\text{Cov}(X, Y)$ is an indication that Y tends to increase as X does, whereas a negative value indicates that Y tends to decrease as X increases.

Example 2.33. The joint density function of X, Y is

$$f(x, y) = \frac{1}{y} e^{-(y+x/y)}, \quad 0 < x, y < \infty$$

(a) Verify that the preceding is a joint density function.
(b) Find Cov (X, Y).

Solution: To show that $f(x, y)$ is a joint density function we need to show it is nonnegative, which is immediate, and that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1$. We prove the latter as follows:

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx &= \int_0^{\infty} \int_0^{\infty} \frac{1}{y} e^{-(y+x/y)} dy\, dx \\ &= \int_0^{\infty} e^{-y} \int_0^{\infty} \frac{1}{y} e^{-x/y} dx\, dy \\ &= \int_0^{\infty} e^{-y} dy \\ &= 1 \end{aligned}$$

To obtain $\text{Cov}(X, Y)$, note that the density function of Y is

$$f_Y(y) = e^{-y} \int_0^\infty \frac{1}{y} e^{-x/y} dx = e^{-y}$$

Thus, Y is an exponential random variable with parameter 1, showing (see Example 2.21) that

$$E[Y] = 1$$

We compute $E[X]$ and $E[XY]$ as follows:

$$\begin{aligned} E[X] &= \int_{-\infty}^\infty \int_{-\infty}^\infty x f(x, y)\, dy\, dx \\ &= \int_0^\infty e^{-y} \int_0^\infty \frac{x}{y} e^{-x/y} dx\, dy \end{aligned}$$

Now, $\int_0^\infty \frac{x}{y} e^{-x/y} dx$ is the expected value of an exponential random variable with parameter $1/y$, and thus is equal to y. Consequently,

$$E[X] = \int_0^\infty y e^{-y} dy = 1$$

Also

$$\begin{aligned} E[XY] &= \int_{-\infty}^\infty \int_{-\infty}^\infty xy\, f(x, y)\, dy\, dx \\ &= \int_0^\infty y e^{-y} \int_0^\infty \frac{x}{y} e^{-x/y} dx\, dy \\ &= \int_0^\infty y^2 e^{-y} dy \end{aligned}$$

Integration by parts ($dv = e^{-y} dy, u = y^2$) gives

$$E[XY] = \int_0^\infty y^2 e^{-y} dy = -y^2 e^{-y}\big|_0^\infty + \int_0^\infty 2y e^{-y} dy = 2E[Y] = 2$$

Consequently,

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 1$$ ■

The following are important properties of covariance.

Properties of Covariance

For any random variables X, Y, Z and constant c,

1. $\text{Cov}(X, X) = \text{Var}(X)$,

2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
3. $\text{Cov}(cX, Y) = c\,\text{Cov}(X, Y)$,
4. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

Whereas the first three properties are immediate, the final one is easily proven as follows:

$$\begin{aligned}\text{Cov}(X, Y + Z) &= E[X(Y + Z)] - E[X]E[Y + Z] \\ &= E[XY] - E[X]E[Y] + E[XZ] - E[X]E[Z] \\ &= \text{Cov}(X, Y) + \text{Cov}(X, Z)\end{aligned}$$

The fourth property listed easily generalizes to give the following result:

$$\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, Y_j) \tag{2.15}$$

A useful expression for the variance of the sum of random variables can be obtained from Eq. (2.15) as follows:

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^{n} X_i\right) &= \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \text{Cov}(X_i, X_i) + \sum_{i=1}^{n}\sum_{j\neq i} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j<i} \text{Cov}(X_i, X_j)\end{aligned} \tag{2.16}$$

If $X_i, i = 1, \ldots, n$ are independent random variables, then Eq. (2.16) reduces to

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

Definition 2.1. If $X_1, \ldots, X_n$ are independent and identically distributed, then the random variable $\bar{X} = \sum_{i=1}^{n} X_i/n$ is called the *sample mean*.

The following proposition shows that the covariance between the sample mean and a deviation from that sample mean is zero. It will be needed in Section 2.6.1.

Proposition 2.4. *Suppose that $X_1, \ldots, X_n$ are independent and identically distributed with expected value μ and variance σ^2. Then,*

(a) $E[\bar{X}] = \mu$.

(b) $\text{Var}(\bar{X}) = \sigma^2/n$.
(c) $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0, i = 1, \ldots, n$.

Proof. Parts (a) and (b) are easily established as follows:

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu,$$

$$\text{Var}(\bar{X}) = \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n}$$

To establish part (c) we reason as follows:

$$\begin{aligned}
\text{Cov}(\bar{X}, X_i - \bar{X}) &= \text{Cov}(\bar{X}, X_i) - \text{Cov}(\bar{X}, \bar{X}) \\
&= \frac{1}{n}\text{Cov}\left(X_i + \sum_{j \neq i} X_j, X_i\right) - \text{Var}(\bar{X}) \\
&= \frac{1}{n}\text{Cov}(X_i, X_i) + \frac{1}{n}\text{Cov}\left(\sum_{j \neq i} X_j, X_i\right) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0
\end{aligned}$$

where the next to the last equality used the fact that X_i and $\sum_{j \neq i} X_j$ are independent and thus have covariance 0. ■

Eq. (2.16) is often useful when computing variances.

Example 2.34 (Variance of a Binomial Random Variable)**.** Compute the variance of a binomial random variable X with parameters n and p.

Solution: Since such a random variable represents the number of successes in n independent trials when each trial has a common probability p of being a success, we may write

$$X = X_1 + \cdots + X_n$$

where the X_i are independent Bernoulli random variables such that

$$X_i = \begin{cases} 1, & \text{if the } i\text{th trial is a success} \\ 0, & \text{otherwise} \end{cases}$$

Hence, from Eq. (2.16) we obtain

$$\text{Var}(X) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

But

$$\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2$$

$$= E[X_i] - (E[X_i])^2 \qquad \text{since } X_i^2 = X_i$$
$$= p - p^2$$

and thus

$$\text{Var}(X) = np(1-p)$$ ■

Example 2.35 (Sampling from a Finite Population: The Hypergeometric)**.** Consider a population of N individuals, some of whom are in favor of a certain proposition. In particular suppose that Np of them are in favor and $N - Np$ are opposed, where p is assumed to be unknown. We are interested in estimating p, the fraction of the population that is for the proposition, by randomly choosing and then determining the positions of n members of the population.

In such situations as described in the preceding, it is common to use the fraction of the sampled population that is in favor of the proposition as an estimator of p. Hence, if we let

$$X_i = \begin{cases} 1, & \text{if the } i\text{th person chosen is in favor} \\ 0, & \text{otherwise} \end{cases}$$

then the usual estimator of p is $\sum_{i=1}^{n} X_i/n$. Let us now compute its mean and variance. Now,

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

where the final equality follows since the ith person chosen is equally likely to be any of the N individuals in the population and so has probability Np/N of being in favor.

$$\text{Var}\left(\sum_{1}^{n} X_i\right) = \sum_{1}^{n} \text{Var}(X_i) + 2\sum\sum_{i<j} \text{Cov}(X_i, X_j)$$

Now, since X_i is a Bernoulli random variable with mean p, it follows that

$$\text{Var}(X_i) = p(1-p)$$

Also, for $i \neq j$,

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$
$$= P\{X_i = 1, X_j = 1\} - p^2$$
$$= P\{X_i = 1\}P\{X_j = 1 | X_i = 1\} - p^2$$
$$= \frac{Np}{N}\frac{(Np-1)}{N-1} - p^2$$

where the last equality follows since if the ith person to be chosen is in favor, then the jth person chosen is equally likely to be any of the other $N-1$ of which $Np-1$ are in favor. Thus, we see that

$$\begin{aligned}\text{Var}\left(\sum_1^n X_i\right) &= np(1-p) + 2\binom{n}{2}\left[\frac{p(Np-1)}{N-1} - p^2\right] \\ &= np(1-p) - \frac{n(n-1)p(1-p)}{N-1}\end{aligned}$$

and so the mean and variance of our estimator are given by

$$\begin{aligned}E\left[\sum_1^n \frac{X_i}{n}\right] &= p, \\ \text{Var}\left[\sum_1^n \frac{X_i}{n}\right] &= \frac{p(1-p)}{n} - \frac{(n-1)p(1-p)}{n(N-1)}\end{aligned}$$

Some remarks are in order: As the mean of the estimator is the unknown value p, we would like its variance to be as small as possible (why is this?), and we see by the preceding that, as a function of the population size N, the variance increases as N increases. The limiting value, as $N \to \infty$, of the variance is $p(1-p)/n$, which is not surprising since for N large each of the X_i will be (approximately) independent random variables, and thus $\sum_1^n X_i$ will have an (approximately) binomial distribution with parameters n and p.

The random variable $\sum_1^n X_i$ can be thought of as representing the number of white balls obtained when n balls are randomly selected from a population consisting of Np white and $N - Np$ black balls. (Identify a person who favors the proposition with a white ball and one against with a black ball.) Such a random variable is called *hypergeometric* and has a probability mass function given by

$$P\left\{\sum_1^n X_i = k\right\} = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}} \qquad \blacksquare$$

It is often important to be able to calculate the distribution of $X + Y$ from the distributions of X and Y when X and Y are independent. Suppose first that X and Y are continuous, X having probability density f and Y having probability density g. Then, letting $F_{X+Y}(a)$ be the cumulative distribution function of $X + Y$, we have

$$\begin{aligned}F_{X+Y}(a) &= P\{X+Y \le a\} \\ &= \iint_{x+y\le a} f(x)g(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f(x)g(y)\,dx\,dy\end{aligned}$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{a-y} f(x)\,dx \right) g(y)\,dy$$
$$= \int_{-\infty}^{\infty} F_X(a-y)g(y)\,dy \tag{2.17}$$

The cumulative distribution function F_{X+Y} is called the *convolution* of the distributions F_X and F_Y (the cumulative distribution functions of X and Y, respectively).

By differentiating Eq. (2.17), we obtain that the probability density function $f_{X+Y}(a)$ of $X+Y$ is given by

$$f_{X+Y}(a) = \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y)g(y)\,dy$$
$$= \int_{-\infty}^{\infty} \frac{d}{da}(F_X(a-y))g(y)\,dy$$
$$= \int_{-\infty}^{\infty} f(a-y)g(y)\,dy \tag{2.18}$$

Example 2.36 (Sum of Two Independent Uniform Random Variables)**.** If X and Y are independent random variables both uniformly distributed on (0, 1), then calculate the probability density of $X+Y$.

Solution: From Eq. (2.18), since

$$f(a) = g(a) = \begin{cases} 1, & 0 < a < 1 \\ 0, & \text{otherwise} \end{cases}$$

we obtain

$$f_{X+Y}(a) = \int_0^1 f(a-y)\,dy$$

For $0 \le a \le 1$, this yields

$$f_{X+Y}(a) = \int_0^a dy = a$$

For $1 < a < 2$, we get

$$f_{X+Y}(a) = \int_{a-1}^1 dy = 2-a$$

Hence,

$$f_{X+Y}(a) = \begin{cases} a, & 0 \le a \le 1 \\ 2-a, & 1 < a < 2 \\ 0, & \text{otherwise} \end{cases}$$

■

Rather than deriving a general expression for the distribution of $X+Y$ in the discrete case, we shall consider an example.

Example 2.37 (Sums of Independent Poisson Random Variables)**.** Let X and Y be independent Poisson random variables with respective means λ_1 and λ_2. Calculate the distribution of $X + Y$.

Solution: Since the event $\{X+Y=n\}$ may be written as the union of the disjoint events $\{X=k, Y=n-k\}, 0 \le k \le n$, we have

$$\begin{aligned}
P\{X+Y=n\} &= \sum_{k=0}^{n} P\{X=k, Y=n-k\} \\
&= \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\} \\
&= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} \\
&= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n
\end{aligned}$$

In words, $X + Y$ has a Poisson distribution with mean $\lambda_1 + \lambda_2$. ■

The concept of independence may, of course, be defined for more than two random variables. In general, the n random variables $X_1, X_2, \ldots, X_n$ are said to be independent if, for all values $a_1, a_2, \ldots, a_n$,

$$\begin{aligned}
&P\{X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n\} \\
&\quad = P\{X_1 \le a_1\}P\{X_2 \le a_2\}\cdots P\{X_n \le a_n\}
\end{aligned}$$

Example 2.38. Let $X_1, \ldots, X_n$ be independent and identically distributed continuous random variables with probability distribution F and density function $F' = f$. If we let $X_{(i)}$ denote the ith smallest of these random variables, then $X_{(1)}, \ldots, X_{(n)}$ are called the *order statistics*. To obtain the distribution of $X_{(i)}$, note that $X_{(i)}$ will be less than or equal to x if and only if at least i of the n random variables $X_1, \ldots, X_n$ are less than or equal to x. Hence,

$$P\{X_{(i)} \le x\} = \sum_{k=i}^{n}\binom{n}{k}(F(x))^k(1-F(x))^{n-k}$$

Differentiation yields that the density function of $X_{(i)}$ is as follows:

$$f_{X_{(i)}}(x) = f(x)\sum_{k=i}^{n}\binom{n}{k}k(F(x))^{k-1}(1-F(x))^{n-k}$$

$$
\begin{aligned}
&- f(x) \sum_{k=i}^{n} \binom{n}{k} (n-k)(F(x))^k (1-F(x))^{n-k-1} \\
&= f(x) \sum_{k=i}^{n} \frac{n!}{(n-k)!(k-1)!} (F(x))^{k-1} (1-F(x))^{n-k} \\
&\quad - f(x) \sum_{k=i}^{n-1} \frac{n!}{(n-k-1)!k!} (F(x))^k (1-F(x))^{n-k-1} \\
&= f(x) \sum_{k=i}^{n} \frac{n!}{(n-k)!(k-1)!} (F(x))^{k-1} (1-F(x))^{n-k} \\
&\quad - f(x) \sum_{j=i+1}^{n} \frac{n!}{(n-j)!(j-1)!} (F(x))^{j-1} (1-F(x))^{n-j} \\
&= \frac{n!}{(n-i)!(i-1)!} f(x) (F(x))^{i-1} (1-F(x))^{n-i}
\end{aligned}
$$

The preceding density is quite intuitive, since in order for $X_{(i)}$ to equal x, $i-1$ of the n values $X_1, \ldots, X_n$ must be less than x; $n-i$ of them must be greater than x; and one must be equal to x. Now, the probability density that every member of a specified set of $i-1$ of the X_j is less than x, every member of another specified set of $n-i$ is greater than x, and the remaining value is equal to x is $(F(x))^{i-1}(1-F(x))^{n-i} f(x)$. Therefore, since there are $n!/[(i-1)!(n-i)!]$ different partitions of the n random variables into the three groups, we obtain the preceding density function. ■

2.5.4 Joint Probability Distribution of Functions of Random Variables

Let X_1 and X_2 be jointly continuous random variables with joint probability density function $f(x_1, x_2)$. It is sometimes necessary to obtain the joint distribution of the random variables Y_1 and Y_2 that arise as functions of X_1 and X_2. Specifically, suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2.

Assume that the functions g_1 and g_2 satisfy the following conditions:

1. The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say, $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$.
2. The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that the following 2×2 determinant

$$
J(x_1, x_2) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} \end{vmatrix} \equiv \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0
$$

at all points (x_1, x_2).

Under these two conditions it can be shown that the random variables Y_1 and Y_2 are jointly continuous with joint density function given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1} \tag{2.19}$$

where $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$.

A proof of Eq. (2.19) would proceed along the following lines:

$$P\{Y_1 \le y_1, Y_2 \le y_2\} = \iint\limits_{\substack{(x_1,x_2):\\ g_1(x_1,x_2)\le y_1\\ g_2(x_1,x_2)\le y_2}} f_{X_1,X_2}(x_1, x_2)\, dx_1\, dx_2 \tag{2.20}$$

The joint density function can now be obtained by differentiating Eq. (2.20) with respect to y_1 and y_2. That the result of this differentiation will be equal to the right-hand side of Eq. (2.19) is an exercise in advanced calculus whose proof will not be presented in the present text.

Example 2.39. If X and Y are independent gamma random variables with parameters (α, λ) and (β, λ), respectively, compute the joint density of $U = X + Y$ and $V = X/(X + Y)$.

Solution: The joint density of X and Y is given by

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \frac{\lambda e^{-\lambda y}(\lambda y)^{\beta-1}}{\Gamma(\beta)} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda(x+y)} x^{\alpha-1} y^{\beta-1} \end{aligned}$$

Now, if $g_1(x, y) = x + y$, $g_2(x, y) = x/(x + y)$, then

$$\frac{\partial g_1}{\partial x} = \frac{\partial g_1}{\partial y} = 1, \qquad \frac{\partial g_2}{\partial x} = \frac{y}{(x+y)^2}, \qquad \frac{\partial g_2}{\partial y} = -\frac{x}{(x+y)^2}$$

and so

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ \dfrac{y}{(x+y)^2} & \dfrac{-x}{(x+y)^2} \end{vmatrix} = -\frac{1}{x+y}$$

Finally, because the equations $u = x + y, v = x/(x + y)$ have as their solutions $x = uv, y = u(1 - v)$, we see that

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}[uv, u(1 - v)]u \\ &= \frac{\lambda e^{-\lambda u}(\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \frac{v^{\alpha-1}(1-v)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \end{aligned}$$

Hence $X + Y$ and $X/(X + Y)$ are independent, with $X + Y$ having a gamma distribution with parameters $(\alpha + \beta, \lambda)$ and $X/(X + Y)$ having density function

$$f_V(v) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} v^{\alpha-1}(1 - v)^{\beta-1}, \qquad 0 < v < 1$$

This is called the beta density with parameters (α, β).

This result is quite interesting. For suppose there are $n + m$ jobs to be performed, with each (independently) taking an exponential amount of time with rate λ for performance, and suppose that we have two workers to perform these jobs. Worker I will do jobs $1, 2, \ldots, n$, and worker II will do the remaining m jobs. If we let X and Y denote the total working times of workers I and II, respectively, then upon using the preceding result it follows that X and Y will be independent gamma random variables having parameters (n, λ) and (m, λ), respectively. Then the preceding result yields that independently of the working time needed to complete all $n + m$ jobs (that is, of $X + Y$), the proportion of this work that will be performed by worker I has a beta distribution with parameters (n, m). ■

When the joint density function of the n random variables $X_1, X_2, \ldots, X_n$ is given and we want to compute the joint density function of $Y_1, Y_2, \ldots, Y_n$, where

$$Y_1 = g_1(X_1, \ldots, X_n), \qquad Y_2 = g_2(X_1, \ldots, X_n), \qquad \ldots,$$
$$Y_n = g_n(X_1, \ldots, X_n)$$

the approach is the same. Namely, we assume that the functions g_i have continuous partial derivatives and that the Jacobian determinant $J(x_1, \ldots, x_n) \neq 0$ at all points $(x_1, \ldots, x_n)$, where

$$J(x_1, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

Furthermore, we suppose that the equations $y_1 = g_1(x_1, \ldots, x_n)$, $y_2 = g_2(x_1, \ldots, x_n)$, $\ldots, y_n = g_n(x_1, \ldots, x_n)$ have a unique solution, say, $x_1 = h_1(y_1, \ldots, y_n), \ldots, x_n = h_n(y_1, \ldots, y_n)$. Under these assumptions the joint density function of the random variables Y_i is given by

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)|J(x_1, \ldots, x_n)|^{-1}$$

where $x_i = h_i(y_1, \ldots, y_n), i = 1, 2, \ldots, n$.

2.6 Moment Generating Functions

The *moment generating function* $\phi(t)$ of the random variable X is defined for all values t by

$$\begin{aligned}\phi(t) &= E[e^{tX}] \\ &= \begin{cases} \sum_x e^{tx} p(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx, & \text{if } X \text{ is continuous} \end{cases}\end{aligned}$$

We call $\phi(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\begin{aligned}\phi'(t) &= \frac{d}{dt} E[e^{tX}] \\ &= E\left[\frac{d}{dt}(e^{tX})\right] \\ &= E[Xe^{tX}]\end{aligned}$$

Hence,

$$\phi'(0) = E[X]$$

Similarly,

$$\begin{aligned}\phi''(t) &= \frac{d}{dt}\phi'(t) \\ &= \frac{d}{dt} E[Xe^{tX}] \\ &= E\left[\frac{d}{dt}(Xe^{tX})\right] \\ &= E[X^2 e^{tX}]\end{aligned}$$

and so

$$\phi''(0) = E[X^2]$$

In general, the nth derivative of $\phi(t)$ evaluated at $t = 0$ equals $E[X^n]$, that is,

$$\phi^n(0) = E[X^n], \qquad n \geq 1$$

We now compute $\phi(t)$ for some common distributions.

Example 2.40 (The Binomial Distribution with Parameters n and p)**.**

$$\phi(t) = E[e^{tX}]$$

$$
\begin{aligned}
&= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

Hence,

$$\phi'(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

and so

$$E[X] = \phi'(0) = np$$

which checks with the result obtained in Example 2.17. Differentiating a second time yields

$$\phi''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

and so

$$E[X^2] = \phi''(0) = n(n-1)p^2 + np$$

Thus, the variance of X is given by

$$
\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= n(n-1)p^2 + np - n^2 p^2 \\
&= np(1-p)
\end{aligned}
$$

■

Example 2.41 (The Poisson Distribution with Mean λ)**.**

$$
\begin{aligned}
\phi(t) &= E[e^{tX}] \\
&= \sum_{n=0}^{\infty} \frac{e^{tn} e^{-\lambda} \lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= \exp\{\lambda(e^t - 1)\}
\end{aligned}
$$

Differentiation yields

$$
\begin{aligned}
&\phi'(t) = \lambda e^t \exp\{\lambda(e^t - 1)\}, \\
&\phi''(t) = (\lambda e^t)^2 \exp\{\lambda(e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\}
\end{aligned}
$$

and so

$$\begin{aligned} E[X] &= \phi'(0) = \lambda, \\ E[X^2] &= \phi''(0) = \lambda^2 + \lambda, \\ \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \lambda \end{aligned}$$

Thus, both the mean and the variance of the Poisson equal λ. ■

Example 2.42 (The Exponential Distribution with Parameter λ)**.**

$$\begin{aligned} \phi(t) &= E[e^{tX}] \\ &= \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^\infty e^{-(\lambda - t)x} dx \\ &= \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda \end{aligned}$$

We note by the preceding derivation that, for the exponential distribution, $\phi(t)$ is only defined for values of t less than λ. Differentiation of $\phi(t)$ yields

$$\phi'(t) = \frac{\lambda}{(\lambda - t)^2}, \qquad \phi''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

Hence,

$$E[X] = \phi'(0) = \frac{1}{\lambda}, \qquad E[X^2] = \phi''(0) = \frac{2}{\lambda^2}$$

The variance of X is thus given by

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{\lambda^2}$$ ■

Example 2.43 (The Normal Distribution with Parameters μ and σ^2)**.** The moment generating function of a standard normal random variable Z is obtained as follows.

$$\begin{aligned} E[e^{tZ}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{tx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(x^2 - 2tx)/2} dx \\ &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(x-t)^2/2} dx \\ &= e^{t^2/2} \end{aligned}$$

Table 2.1

Discrete probability distribution	Probability mass function, $p(x)$	Moment generating function, $\phi(t)$	Mean	Variance
Binomial with parameters n, p, $0 \le p \le 1$	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$(pe^t + (1-p))^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \le p \le 1$	$p(1-p)^{x-1}$, $x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

If Z is a standard normal, then $X = \sigma Z + \mu$ is normal with parameters μ and σ^2; therefore,

$$\phi(t) = E[e^{tX}] = E[e^{t(\sigma Z+\mu)}] = e^{t\mu} E[e^{t\sigma Z}] = \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

By differentiating we obtain

$$\phi'(t) = (\mu + t\sigma^2) \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\},$$
$$\phi''(t) = (\mu + t\sigma^2)^2 \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\} + \sigma^2 \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

and so

$$E[X] = \phi'(0) = \mu,$$
$$E[X^2] = \phi''(0) = \mu^2 + \sigma^2$$

implying that

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] - E([X])^2 \\ &= \sigma^2\end{aligned}$$

■

Tables 2.1 and 2.2 give the moment generating function for some common distributions.

An important property of moment generating functions is that the *moment generating function of the sum of independent random variables is just the product of the individual moment generating functions*. To see this, suppose that X and Y are independent and have moment generating functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. Then

Table 2.2

Continuous probability distribution	Probability density function, $f(x)$	Moment generating function, $\phi(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters (n, λ), $\lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^n$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \times \exp\{-(x-\mu)^2/2\sigma^2\}, \quad -\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

$\phi_{X+Y}(t)$, the moment generating function of $X + Y$, is given by

$$\begin{aligned} \phi_{X+Y}(t) &= E[e^{t(X+Y)}] \\ &= E[e^{tX} e^{tY}] \\ &= E[e^{tX}]E[e^{tY}] \\ &= \phi_X(t)\phi_Y(t) \end{aligned}$$

where the next to the last equality follows from Proposition 2.3 since X and Y are independent.

Another important result is that the *moment generating function uniquely determines the distribution*. That is, there exists a one-to-one correspondence between the moment generating function and the distribution function of a random variable.

Example 2.44 (Sums of Independent Binomial Random Variables)**.** If X and Y are independent binomial random variables with parameters (n, p) and (m, p), respectively, then what is the distribution of $X + Y$?

Solution: The moment generating function of $X + Y$ is given by

$$\begin{aligned} \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) &= (pe^t + 1 - p)^n (pe^t + 1 - p)^m \\ &= (pe^t + 1 - p)^{m+n} \end{aligned}$$

But $(pe^t + (1-p))^{m+n}$ is just the moment generating function of a binomial random variable having parameters $m+n$ and p. Thus, this must be the distribution of $X+Y$. ■

Example 2.45 (Sums of Independent Poisson Random Variables)**.** Calculate the distribution of $X+Y$ when X and Y are independent Poisson random variables with means λ_1 and λ_2, respectively.

Solution:

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t)\\ &= e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)}\\ &= e^{(\lambda_1+\lambda_2)(e^t-1)}\end{aligned}$$

Hence, $X+Y$ is Poisson distributed with mean $\lambda_1+\lambda_2$, verifying the result given in Example 2.37. ■

Example 2.46 (Sums of Independent Normal Random Variables)**.** Show that if X and Y are independent normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, then $X+Y$ is normal with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$.

Solution:

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t)\\ &= \exp\left\{\frac{\sigma_1^2t^2}{2}+\mu_1t\right\}\exp\left\{\frac{\sigma_2^2t^2}{2}+\mu_2t\right\}\\ &= \exp\left\{\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}+(\mu_1+\mu_2)t\right\}\end{aligned}$$

which is the moment generating function of a normal random variable with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$. Hence, the result follows since the moment generating function uniquely determines the distribution. ■

Example 2.47 (The Poisson Paradigm)**.** We showed in Section 2.2.4 that the number of successes that occur in n independent trials, each of which results in a success with probability p is, when n is large and p small, approximately a Poisson random variable with parameter $\lambda = np$. This result, however, can be substantially strengthened. First it is not necessary that the trials have the same success probability, only that all the success probabilities are small. To see that this is the case, suppose that the trials are independent, with trial i resulting in a success with probability p_i, where all the $p_i, i = 1, \ldots, n$ are small. Letting X_i equal 1 if trial i is a success, and 0 otherwise, it follows that the number of successes, call it X, can be expressed as

$$X = \sum_{i=1}^{n} X_i$$

Using that X_i is a Bernoulli (or binary) random variable, its moment generating function is

$$E[e^{tX_i}] = p_i e^t + 1 - p_i = 1 + p_i(e^t - 1)$$

Now, using the result that, for $|x|$ small,

$$e^x \approx 1 + x$$

it follows, because $p_i(e^t - 1)$ is small when p_i is small, that

$$E[e^{tX_i}] = 1 + p_i(e^t - 1) \approx \exp\{p_i(e^t - 1)\}$$

Because the moment generating function of a sum of independent random variables is the product of their moment generating functions, the preceding implies that

$$E[e^{tX}] \approx \prod_{i=1}^{n} \exp\{p_i(e^t - 1)\} = \exp\left\{\sum_i p_i(e^t - 1)\right\}$$

But the right side of the preceding is the moment generating function of a Poisson random variable with mean $\sum_i p_i$, thus arguing that this is approximately the distribution of X.

Not only is it not necessary for the trials to have the same success probability for the number of successes to approximately have a Poisson distribution, they need not even be independent, provided that their dependence is *weak*. For instance, recall the matching problem (Example 2.30) where n people randomly select hats from a set consisting of one hat from each person. By regarding the random selections of hats as constituting n trials, where we say that trial i is a success if person i chooses his or her own hat, it follows that, with A_i being the event that trial i is a success,

$$P(A_i) = \frac{1}{n} \quad \text{and} \quad P(A_i|A_j) = \frac{1}{n-1}, \qquad j \neq i$$

Hence, whereas the trials are not independent, their dependence appears, for large n, to be weak. Because of this weak dependence, and the small trial success probabilities, it would seem that the number of matches should approximately have a Poisson distribution with mean 1 when n is large, and this is shown to be the case in Example 3.27.

The statement that "the number of successes in n trials that are either independent or at most weakly dependent is, when the trial success probabilities are all small, approximately a Poisson random variable" is known as the *Poisson paradigm*. ■

Remark. For a nonnegative random variable X, it is often convenient to define its *Laplace transform* $g(t), t \geq 0$, by

$$g(t) = \phi(-t) = E[e^{-tX}]$$

That is, the Laplace transform evaluated at t is just the moment generating function evaluated at $-t$. The advantage of dealing with the Laplace transform, rather than the

moment generating function, when the random variable is nonnegative is that if $X \geq 0$ and $t \geq 0$, then

$$0 \leq e^{-tX} \leq 1$$

That is, the Laplace transform is always between 0 and 1. As in the case of moment generating functions, it remains true that nonnegative random variables that have the same Laplace transform must also have the same distribution. ■

It is also possible to define the joint moment generating function of two or more random variables. This is done as follows. For any n random variables $X_1, \ldots, X_n$, the joint moment generating function, $\phi(t_1, \ldots, t_n)$, is defined for all real values of $t_1, \ldots, t_n$ by

$$\phi(t_1, \ldots, t_n) = E[e^{(t_1 X_1 + \cdots + t_n X_n)}]$$

It can be shown that $\phi(t_1, \ldots, t_n)$ uniquely determines the joint distribution of $X_1, \ldots, X_n$.

Example 2.48 (The Multivariate Normal Distribution)**.** Let $Z_1, \ldots, Z_n$ be a set of n independent standard normal random variables. If, for some constants a_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, and μ_i, $1 \leq i \leq m$,

$$\begin{aligned}
X_1 &= a_{11} Z_1 + \cdots + a_{1n} Z_n + \mu_1, \\
X_2 &= a_{21} Z_1 + \cdots + a_{2n} Z_n + \mu_2, \\
&\vdots \\
X_i &= a_{i1} Z_1 + \cdots + a_{in} Z_n + \mu_i, \\
&\vdots \\
X_m &= a_{m1} Z_1 + \cdots + a_{mn} Z_n + \mu_m
\end{aligned}$$

then the random variables $X_1, \ldots, X_m$ are said to have a multivariate normal distribution.

It follows from the fact that the sum of independent normal random variables is itself a normal random variable that each X_i is a normal random variable with mean and variance given by

$$\begin{aligned}
E[X_i] &= \mu_i, \\
\mathrm{Var}(X_i) &= \sum_{j=1}^{n} a_{ij}^2
\end{aligned}$$

Let us now determine

$$\phi(t_1, \ldots, t_m) = E[\exp\{t_1 X_1 + \cdots + t_m X_m\}]$$

the joint moment generating function of $X_1, \ldots, X_m$. The first thing to note is that since $\sum_{i=1}^m t_i X_i$ is itself a linear combination of the independent normal random variables $Z_1, \ldots, Z_n$, it is also normally distributed. Its mean and variance are respectively

$$E\left[\sum_{i=1}^m t_i X_i\right] = \sum_{i=1}^m t_i \mu_i$$

and

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^m t_i X_i\right) &= \text{Cov}\left(\sum_{i=1}^m t_i X_i, \sum_{j=1}^m t_j X_j\right) \\ &= \sum_{i=1}^m \sum_{j=1}^m t_i t_j \text{Cov}(X_i, X_j)\end{aligned}$$

Now, if Y is a normal random variable with mean μ and variance σ^2, then

$$E[e^Y] = \phi_Y(t)|_{t=1} = e^{\mu+\sigma^2/2}$$

Thus, we see that

$$\phi(t_1, \ldots, t_m) = \exp\left\{\sum_{i=1}^m t_i \mu_i + \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m t_i t_j \text{Cov}(X_i, X_j)\right\}$$

which shows that the joint distribution of $X_1, \ldots, X_m$ is completely determined from a knowledge of the values of $E[X_i]$ and $\text{Cov}(X_i, X_j)$, $i, j = 1, \ldots, m$. ■

2.6.1 The Joint Distribution of the Sample Mean and Sample Variance from a Normal Population

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, each with mean μ and variance σ^2. The random variable S^2 defined by

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$$

is called the *sample variance* of these data. To compute $E[S^2]$ we use the identity

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \tag{2.21}$$

which is proven as follows:

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\mu - \bar{X})^2 + 2(\mu - \bar{X}) \sum_{i=1}^{n} (X_i - \mu) \\
&= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\mu - \bar{X})^2 + 2(\mu - \bar{X})(n\bar{X} - n\mu) \\
&= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\mu - \bar{X})^2 - 2n(\mu - \bar{X})^2
\end{aligned}
$$

and Identity (2.21) follows.

Using Identity (2.21) gives

$$
\begin{aligned}
E[(n-1)S^2] &= \sum_{i=1}^{n} E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] \\
&= n\sigma^2 - n \operatorname{Var}(\bar{X}) \\
&= (n-1)\sigma^2 \qquad \text{from Proposition 2.4(b)}
\end{aligned}
$$

Thus, we obtain from the preceding that

$$E[S^2] = \sigma^2$$

We will now determine the joint distribution of the sample mean $\bar{X} = \sum_{i=1}^{n} X_i/n$ and the sample variance S^2 when the X_i have a normal distribution. To begin we need the concept of a chi-squared random variable.

Definition 2.2. If $Z_1, \ldots, Z_n$ are independent standard normal random variables, then the random variable $\sum_{i=1}^{n} Z_i^2$ is said to be a *chi-squared random variable* with n *degrees of freedom*.

We shall now compute the moment generating function of $\sum_{i=1}^{n} Z_i^2$. To begin, note that

$$
\begin{aligned}
E[\exp\{tZ_i^2\}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-x^2/2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\, dx \qquad \text{where } \sigma^2 = (1-2t)^{-1} \\
&= \sigma \\
&= (1-2t)^{-1/2}
\end{aligned}
$$

Hence,

$$E\left[\exp\left\{t \sum_{i=1}^{n} Z_i^2\right\}\right] = \prod_{i=1}^{n} E[\exp\{tZ_i^2\}] = (1-2t)^{-n/2}$$

Now, let $X_1, \ldots, X_n$ be independent normal random variables, each with mean μ and variance σ^2, and let $\bar{X} = \sum_{i=1}^{n} X_i/n$ and S^2 denote their sample mean and sample

variance. Since the sum of independent normal random variables is also a normal random variable, it follows that $\bar{X}$ is a normal random variable with expected value μ and variance σ^2/n. In addition, from Proposition 2.4,

$$\text{Cov}(\bar{X}, X_i - \bar{X}) = 0, \qquad i = 1, \ldots, n \tag{2.22}$$

Also, since $\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$ are all linear combinations of the independent standard normal random variables $(X_i - \mu)/\sigma, i = 1, \ldots, n$, it follows that the random variables $\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$ have a joint distribution that is multivariate normal. However, if we let Y be a normal random variable with mean μ and variance σ^2/n that is independent of $X_1, \ldots, X_n$, then the random variables $Y, X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$ also have a multivariate normal distribution, and by Eq. (2.22), they have the same expected values and covariances as the random variables $\bar{X}, X_i - \bar{X}, i = 1, \ldots, n$. Thus, since a multivariate normal distribution is completely determined by its expected values and covariances, we can conclude that the random vectors $Y, X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$ and $\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$ have the same joint distribution; thus showing that $\bar{X}$ is independent of the sequence of deviations $X_i - \bar{X}, i = 1, \ldots, n$.

Since $\bar{X}$ is independent of the sequence of deviations $X_i - \bar{X}, i = 1, \ldots, n$, it follows that it is also independent of the sample variance

$$S^2 \equiv \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

To determine the distribution of S^2, use Identity (2.21) to obtain

$$(n-1)S^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

Dividing both sides of this equation by σ^2 yields

$$\frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \tag{2.23}$$

Now, $\sum_{i=1}^{n} (X_i - \mu)^2/\sigma^2$ is the sum of the squares of n independent standard normal random variables, and so is a chi-squared random variable with n degrees of freedom; it thus has moment generating function $(1 - 2t)^{-n/2}$. Also $[(\bar{X} - \mu)/(\sigma/\sqrt{n})]^2$ is the square of a standard normal random variable and so is a chi-squared random variable with one degree of freedom; it thus has moment generating function $(1 - 2t)^{-1/2}$. In addition, we have previously seen that the two random variables on the left side of Eq. (2.23) are independent. Therefore, because the moment generating function of the sum of independent random variables is equal to the product of their individual moment generating functions, we obtain that

$$E[e^{t(n-1)S^2/\sigma^2}](1 - 2t)^{-1/2} = (1 - 2t)^{-n/2}$$

or

$$E[e^{t(n-1)S^2/\sigma^2}] = (1-2t)^{-(n-1)/2}$$

But because $(1-2t)^{-(n-1)/2}$ is the moment generating function of a chi-squared random variable with $n-1$ degrees of freedom, we can conclude, since the moment generating function uniquely determines the distribution of the random variable, that this is the distribution of $(n-1)S^2/\sigma^2$.

Summing up, we have shown the following.

Proposition 2.5. *If $X_1, \ldots, X_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2, then the sample mean $\bar{X}$ and the sample variance S^2 are independent. $\bar{X}$ is a normal random variable with mean μ and variance σ^2/n; $(n-1)S^2/\sigma^2$ is a chi-squared random variable with $n-1$ degrees of freedom.*

2.7 Limit Theorems

We start this section by proving a result known as Markov's inequality.

Proposition 2.6 (Markov's Inequality)**.** *If X is a random variable that takes only nonnegative values, then for any value $a > 0$*

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof. We give a proof for the case where X is continuous with density f.

$$\begin{aligned}
E[X] &= \int_0^\infty x f(x)\,dx \\
&= \int_0^a x f(x)\,dx + \int_a^\infty x f(x)\,dx \\
&\geq \int_a^\infty x f(x)\,dx \\
&\geq \int_a^\infty a f(x)\,dx \\
&= a \int_a^\infty f(x)\,dx \\
&= a P\{X \geq a\}
\end{aligned}$$

and the result is proven. ■

As a corollary, we obtain the following.

Proposition 2.7 (Chebyshev's Inequality)**.** *If X is a random variable with mean μ and variance σ^2, then, for any value $k > 0$,*

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof. Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2}$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, the preceding is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and the proof is complete. ■

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed, and we would not need to resort to bounds.

Example 2.49. Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500.

(a) What can be said about the probability that this week's production will be at least 1000?

(b) If the variance of a week's production is known to equal 100, then what can be said about the probability that this week's production will be between 400 and 600?

Solution: Let X be the number of items that will be produced in a week.

(a) By Markov's inequality,

$$P\{X \geq 1000\} \leq \frac{E[X]}{1000} = \frac{500}{1000} = \frac{1}{2}$$

(b) By Chebyshev's inequality,

$$P\{|X - 500| \geq 100\} \leq \frac{\sigma^2}{(100)^2} = \frac{1}{100}$$

Hence,

$$P\{|X - 500| < 100\} \geq 1 - \frac{1}{100} = \frac{99}{100}$$

and so the probability that this week's production will be between 400 and 600 is at least 0.99. ■

The following theorem, known as the *strong law of large numbers*, is probably the most well-known result in probability theory. It states that the average of a sequence of independent random variables having the same distribution will, with probability 1, converge to the mean of that distribution.

Theorem 2.1 (Strong Law of Large Numbers). *Let $X_1, X_2, \ldots$ be a sequence of independent random variables having a common distribution, and let $E[X_i] = \mu$. Then, with probability 1,*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \qquad \text{as } n \to \infty$$

As an example of the preceding, suppose that a sequence of independent trials is performed. Let E be a fixed event and denote by $P(E)$ the probability that E occurs on any particular trial. Letting

$$X_i = \begin{cases} 1, & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0, & \text{if } E \text{ does not occur on the } i\text{th trial} \end{cases}$$

we have by the strong law of large numbers that, with probability 1,

$$\frac{X_1 + \cdots + X_n}{n} \to E[X] = P(E) \tag{2.24}$$

Since $X_1 + \cdots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret Eq. (2.24) as stating that, with probability 1, the limiting proportion of time that the event E occurs is just $P(E)$.

Running neck and neck with the strong law of large numbers for the honor of being probability theory's number one result is the *central limit theorem*. Besides its theoretical interest and importance, this theorem provides a simple method for computing approximate probabilities for sums of independent random variables. It also explains the remarkable fact that the empirical frequencies of so many natural "populations" exhibit a bell-shaped (that is, normal) curve.

Theorem 2.2 (Central Limit Theorem). *Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2. Then the distribution of*

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is,

$$P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$

as $n \to \infty$.

Note that like the other results of this section, this theorem holds for *any* distribution of the X_is; herein lies its power.

If X is binomially distributed with parameters n and p, then X has the same distribution as the sum of n independent Bernoulli random variables, each with parameter p. (Recall that the Bernoulli random variable is just a binomial random variable whose parameter n equals 1.) Hence, the distribution of

$$\frac{X - E[X]}{\sqrt{\text{Var}(X)}} = \frac{X - np}{\sqrt{np(1-p)}}$$

approaches the standard normal distribution as n approaches ∞. The normal approximation will, in general, be quite good for values of n satisfying $np(1-p) \geq 10$.

Example 2.50 (Normal Approximation to the Binomial)**.** Let X be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $X = 20$. Use the normal approximation and then compare it to the exact solution.

Solution: Since the binomial is a discrete random variable, and the normal a continuous random variable, it leads to a better approximation to write the desired probability as

$$\begin{aligned} P\{X = 20\} &= P\{19.5 < X < 20.5\} \\ &= P\left\{\frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right\} \\ &= P\left\{-0.16 < \frac{X - 20}{\sqrt{10}} < 0.16\right\} \\ &\approx \Phi(0.16) - \Phi(-0.16) \end{aligned}$$

where $\Phi(x)$, the probability that the standard normal is less than x is given by

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

By the symmetry of the standard normal distribution

$$\Phi(-0.16) = P\{N(0,1) > 0.16\} = 1 - \Phi(0.16)$$

where $N(0,1)$ is a standard normal random variable. Hence, the desired probability is approximated by

$$P\{X = 20\} \approx 2\Phi(0.16) - 1$$

Using Table 2.3, we obtain

$$P\{X = 20\} \approx 0.1272$$

Table 2.3 Area $\Phi(x)$ under the Standard Normal Curve to the Left of x.

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5597	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8557	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

The exact result is

$$P\{X = 20\} = \binom{40}{20}\left(\frac{1}{2}\right)^{40}$$

which can be shown to equal 0.1254. ■

Example 2.51. Let $X_i, i = 1, 2, \ldots, 10$ be independent random variables, each being uniformly distributed over (0, 1). Estimate $P\{\sum_1^{10} X_i > 7\}$.

Solution: Since $E[X_i] = \frac{1}{2}$, $\text{Var}(X_i) = \frac{1}{12}$ we have by the central limit theorem that

$$\begin{aligned} P\left\{\sum_1^{10} X_i > 7\right\} &= P\left\{\frac{\sum_1^{10} X_i - 5}{\sqrt{10\left(\frac{1}{12}\right)}} > \frac{7-5}{\sqrt{10\left(\frac{1}{12}\right)}}\right\} \\ &\approx 1 - \Phi(2.19) \\ &= 0.0143 \end{aligned}$$

■

Example 2.52. The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, at which point it is replaced by a new one. Assuming a stockpile of 25 such batteries, the lifetimes of which are independent, approximate the probability that over 1100 hours of use can be obtained.

Solution: If we let X_i denote the lifetime of the ith battery to be put in use, then we desire $p = P\{X_1 + \cdots + X_{25} > 1100\}$, which is approximated as follows:

$$\begin{aligned} p &= P\left\{\frac{X_1 + \cdots + X_{25} - 1000}{20\sqrt{25}} > \frac{1100 - 1000}{20\sqrt{25}}\right\} \\ &\approx P\{N(0, 1) > 1\} \\ &= 1 - \Phi(1) \\ &\approx 0.1587 \end{aligned}$$

■

We now present a heuristic proof of the central limit theorem. Suppose first that the X_i have mean 0 and variance 1, and let $E[e^{tX}]$ denote their common moment generating function. Then, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ is

$$\begin{aligned} E\left[\exp\left\{t\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right\}\right] &= E[e^{tX_1/\sqrt{n}} e^{tX_2/\sqrt{n}} \cdots e^{tX_n/\sqrt{n}}] \\ &= (E[e^{tX/\sqrt{n}}])^n \qquad \text{by independence} \end{aligned}$$

Now, for n large, we obtain from the Taylor series expansion of e^y that

$$e^{tX/\sqrt{n}} \approx 1 + \frac{tX}{\sqrt{n}} + \frac{t^2X^2}{2n}$$

Taking expectations shows that when n is large

$$\begin{aligned} E[e^{tX/\sqrt{n}}] &\approx 1 + \frac{tE[X]}{\sqrt{n}} + \frac{t^2E[X^2]}{2n} \\ &= 1 + \frac{t^2}{2n} \qquad \text{because } E[X] = 0, E[X^2] = 1 \end{aligned}$$

Therefore, we obtain that when n is large

$$E\left[\exp\left\{t\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right\}\right] \approx \left(1 + \frac{t^2}{2n}\right)^n$$

When n goes to ∞ the approximation can be shown to become exact and we have

$$\lim_{n\to\infty} E\left[\exp\left\{t\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right\}\right] = e^{t^2/2}$$

Thus, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the moment generating function of a (standard) normal random variable with mean 0 and variance 1. Using this, it can be proven that the distribution function of the random variable $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the standard normal distribution function Φ.

When the X_i have mean μ and variance σ^2, the random variables $\frac{X_i-\mu}{\sigma}$ have mean 0 and variance 1. Thus, the preceding shows that

$$P\left\{\frac{X_1 - \mu + X_2 - \mu + \cdots + X_n - \mu}{\sigma\sqrt{n}} \le a\right\} \to \Phi(a)$$

which proves the central limit theorem.

2.8 Proof of the Strong Law of Large Numbers

In this section we give a proof of the strong law of large numbers. Our proof of the strong law makes use of the *Borel–Cantelli lemma*.

Borel–Cantelli Lemma. *For a sequence of events A_i, $i \ge 1$, let N denote the number of these events that occur. If $\sum_{i=1}^{\infty} P(A_i) < \infty$, then $P(N = \infty) = 0$.*

Proof. Suppose that $\sum_{i=1}^{\infty} P(A_i) < \infty$. Now, if $N = \infty$, then for every $n < \infty$ at least one of the events $A_n, A_{n+1}, \ldots$ will occur. That is, $N = \infty$ implies that $\cup_{i=n}^{\infty} A_i$ occurs for every n. Thus, for every n

$$\begin{aligned} P(N = \infty) &\le P(\cup_{i=n}^{\infty} A_i) \\ &\le \sum_{i=n}^{\infty} P(A_i) \end{aligned}$$

where the final inequality follows from Boole's inequality. Because $\sum_{i=1}^{\infty} P(A_i) < \infty$ implies that $\sum_{i=n}^{\infty} P(A_i) \to 0$ as $n \to \infty$, we obtain from the preceding upon letting $n \to \infty$ that $P(N = \infty) = 0$, which proves the result. ■

Remark. The Borel–Cantelli lemma is actually quite intuitive, for if we define the indicator variable I_i to equal 1 if A_i occurs and to equal 0 otherwise, then $N = \sum_{i=1}^{\infty} I_i$, implying that

$$E[N] = \sum_{i=1}^{\infty} E[I_i] = \sum_{i=1}^{\infty} P(A_i)$$

Consequently, the Borel–Cantelli theorem states that if the expected number of events that occur is finite then the probability that an infinite number of them occur is 0, which is intuitive because if there were a positive probability that an infinite number of events could occur then $E[N]$ would be infinite. ■

Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean μ, and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ be the average of the first n of them. The strong law of large numbers states that $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$. That is, with probability 1, $\bar{X}_n$ converges to μ as $n \to \infty$. We will give a proof of this result under the assumption that σ^2, the variance of X_i, is finite (which is equivalent to assuming that $E[X_i^2] < \infty$). Because proving the strong law requires showing, for any $\epsilon > 0$, that $|\bar{X}_n - \mu| > \epsilon$ for only a finite number of values of n, it is natural to attempt to prove it by utilizing the Borel–Cantelli lemma. That is, the result would follow if we could show that $\sum_{n=1}^{\infty} P(|\bar{X}_n - \mu| > \epsilon) < \infty$. However, because $E[\bar{X}_n] = \mu$, $\text{Var}(\bar{X}_n) = \sigma^2/n$, attempting to show this by using Chebyshev's inequality yields

$$\sum_{n=1}^{\infty} P(|\bar{X}_n - \mu| > \epsilon) \le \sum_{n=1}^{\infty} \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Thus, a straightforward use of Borel–Cantelli does not work. However, as we now show, a tweaking of the argument, where we first consider a subsequence of $\bar{X}_n, n \ge 1$, allows us to prove the strong law.

Theorem (The Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with $E[X_i] = \mu$, and $Var(X_i) = \sigma^2 < \infty$. Then, with $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$,*

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$$

Proof. Suppose first that the X_i are nonnegative random variables. Fix $\alpha > 1$, and let n_j be the smallest integer greater than or equal to α^j, $j \ge 1$. From Chebyshev's inequality we see that

$$P(|\bar{X}_{n_j} - \mu| > \epsilon) \le \frac{\text{Var}(\bar{X}_{n_j})}{\epsilon^2} = \frac{\sigma^2}{n_j \epsilon^2}$$

Consequently,

$$\sum_{j=1}^{\infty} P(|\bar{X}_{n_j} - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \sum_{j=1}^{\infty} \frac{1}{n_j} \leq \frac{\sigma^2}{\epsilon^2} \sum_{j=1}^{\infty} (1/\alpha)^j < \infty.$$

Therefore, by the Borel–Cantelli lemma, it follows that, with probability 1, $|\bar{X}_{n_j} - \mu| > \epsilon$ for only a finite number of j. As this is true for any $\epsilon > 0$, we see that, with probability 1,

$$\lim_{j \to \infty} \bar{X}_{n_j} = \mu \tag{2.25}$$

Because $n_j \uparrow \infty$ as $j \uparrow \infty$, it follows that for any $m > \alpha$, there is an integer $j(m)$ such that $n_{j(m)} \leq m < n_{j(m)+1}$. The nonnegativity of the X_i yields that

$$\sum_{i=1}^{n_{j(m)}} X_i \leq \sum_{i=1}^{m} X_i \leq \sum_{i=1}^{n_{j(m)+1}} X_i$$

Dividing by m shows that

$$\frac{n_{j(m)}}{m} \bar{X}_{n_{j(m)}} \leq \bar{X}_m \leq \frac{n_{j(m)+1}}{m} \bar{X}_{n_{j(m)+1}}$$

Because $\frac{1}{n_{j(m)+1}} < \frac{1}{m} \leq \frac{1}{n_{j(m)}}$, this yields that

$$\frac{n_{j(m)}}{n_{j(m)+1}} \bar{X}_{n_{j(m)}} \leq \bar{X}_m \leq \frac{n_{j(m)+1}}{n_{j(m)}} \bar{X}_{n_{j(m)+1}}$$

Because $\lim_{m \to \infty} j(m) = \infty$ and $\lim_{j \to \infty} \frac{n_{j+1}}{n_j} = \alpha$, it follows, for any $\epsilon > 0$, that $\frac{n_{j(m)+1}}{n_{j(m)}} < \alpha + \epsilon$ for all but a finite number of m. Consequently, from (2.25) and the preceding, it follows, with probability 1, that $\frac{\mu}{\alpha+\epsilon} < \bar{X}_m < (\alpha + \epsilon)\mu$ for all but a finite number of values of m. As this is true for any $\epsilon > 0$, $\alpha > 1$, it follows that with probability 1

$$\lim_{m \to \infty} \bar{X}_m = \mu$$

Thus the result is proven when the X_i are nonnegative. In the general case, let

$$X_i^+ = \begin{cases} X_i, & \text{if } X_i \geq 0 \\ 0, & \text{if } X_i < 0 \end{cases}$$

and let

$$X_i^- = \begin{cases} 0, & \text{if } X_i \geq 0 \\ -X_i, & \text{if } X_i < 0 \end{cases}$$

X_i^+ and X_i^- are called, respectively, the positive and negative parts of X_i. Noting that

$$X_i = X_i^+ - X_i^-$$

it follows (since $X_i^+ X_i^- = 0$) that

$$X_i^2 = (X_i^+)^2 + (X_i^-)^2$$

Hence, the assumption that $E[X_i^2] < \infty$ implies that $E[(X_i^+)^2]$ and $E[(X_i^-)^2]$ are also both finite. Letting $\mu^+ = E[X_i^+]$ and $\mu^- = E[X_i^-]$, and using that X_i^+ and X_i^- are both nonnegative, it follows from the previous result for nonnegative random variables that, with probability 1,

$$\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} X_i^+ = \mu^+, \quad \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} X_i^- = \mu^-$$

Consequently, with probability 1,

$$\begin{aligned}\lim_{m\to\infty} \bar{X}_m &= \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m}(X_i^+ - X_i^-)\\ &= \mu^+ - \mu^-\\ &= \mu\end{aligned}$$

■

There is a partial converse of the Borel–Cantelli lemma that holds when the events are independent.

Converse to Borel–Cantelli Lemma. *If $\sum_{i=1}^{\infty} P(A_i) = \infty$, and the events A_i, $i \geq 1$ are independent, then*

$$P(\textit{an infinite number of the events } A_i,\ i \geq 1 \textit{ occur}) = 1$$

Proof. For any n, let $B_n = \cap_{i=n}^{\infty} A_i^c$ be the event that none the events $A_n, A_{n+1}, \ldots$ occur. Then

$$\begin{aligned}P(B_n) &= P(\cap_{i=n}^{\infty} A_i^c)\\ &= \prod_{i=n}^{\infty} P(A_i^c) \qquad \text{by independence}\\ &= \prod_{i=n}^{\infty}[1 - P(A_i)]\\ &\leq \prod_{i=n}^{\infty} e^{-P(A_i)} \qquad \text{by the inequality } e^{-x} \geq 1 - x\\ &= e^{-\sum_{i=n}^{\infty} P(A_i)}\\ &= 0\end{aligned}$$

Because B_n, $n \geq 1$ are increasing events, $\lim_{n\to\infty} B_n = \cup_{n=1}^{\infty} B_n$. Consequently, it follows from the continuity property of probabilities that

$$\begin{aligned} P(\cup_{n=1}^{\infty} B_n) &= P(\lim_{n\to\infty} B_n) \\ &= \lim_{n\to\infty} P(B_n) \\ &= 0 \end{aligned}$$

As $\cup_{n=1}^{\infty} B_n$ is the event that only a finite number of the events A_i occur, the result follows. ■

Example 2.53. Suppose that in each time period we must choose one of n drugs to use, with drug i resulting in a cure with unknown probability p_i, $i = 1, \ldots, n$. Assume that the result of a drug choice (either a cure or not) is immediately learned. Say that a drug is optimal if its cure probability is equal to $\max_i p_i$, and say that it is non-optimal otherwise. Suppose that our objective is to find a policy for deciding which drug to prescribe at each period that has the property that its use results in the long run proportion of time that a non-optimal drug is used being equal to 0. The following policy accomplishes this goal.

Suppose at time k that previous uses of drug i have resulted in $s_i(k)$ cures and $f_i(k)$ failures, for $i = 1, \ldots, n$, where $\sum_i (s_i(k) + f_i(k)) = k - 1$. Let the next choice be a "random choice" with probability $1/k$, or a "non-random choice" with probability $1 - 1/k$. If it is a random choice, let the drug used in period k be equally likely to be any of the n drugs; if it is a non-random choice, let the drug used in period k be any of the drugs with maximal value of $\frac{s_i(k)}{s_i(k)+f_i(k)}$.

To show that the use of the preceding procedure results in the long run proportion of time that a non-optimal drug is chosen being equal to 0, first note that the converse to the Borel–Cantelli lemma shows that, with probability 1, the number of random choices is infinite. As each such choice is equally likely to be any of the n drugs, it thus follows that each drug is, with probability 1, chosen infinitely often. Thus, by the strong law of large numbers, it follows that with probability 1

$$\lim_{k\to\infty} \frac{s_i(k)}{s_i(k) + f_i(k)} = p_i, \quad i = 1, \ldots, n$$

Hence, after some finite time no non-optimal drug will ever be selected by a non-random choice.

To complete the argument we now show that, with probability 1, the long run fraction of choices that are random is equal to 0. Suppose that these choices are determined by letting $U_k, k \geq 1$ be independent uniform $(0, 1)$ random variables, and then letting the choice at time k be random if $U_k \leq 1/k$. Then, with $I\{A\}$ being the indicator variable for the event A, equal to 1 if A occurs and to 0 otherwise, we have that for any m

$$\begin{aligned}\text{proportion of choices that are random} &= \lim_{r\to\infty} \frac{\sum_{k=1}^{r} I\{U_k \le 1/k\}}{r}\\ &= \lim_{r\to\infty} \frac{\sum_{k=m}^{m+r-1} I\{U_k \le 1/k\}}{r}\\ &\le \lim_{r\to\infty} \frac{\sum_{k=m}^{m+r-1} I\{U_k \le 1/m\}}{r}\\ &= 1/m\end{aligned}$$

where the next to last equality follows because if $k \ge m$ then $U_k \le 1/k \Rightarrow U_k \le 1/m$, and the final equality follows from the strong law of large numbers because $I\{U_k \le 1/m\}$, $k \ge m$, are independent and identically distributed Bernoulli random variables with mean $1/m$. As the preceding is true for all m, it follows that the proportion of choices that are random is equal to 0. It now follows from the earlier result that the long run proportion of non-random choices that are optimal is 1, that the long run proportion of choices that are optimal is equal to 1. ■

2.9 Stochastic Processes

A *stochastic process* $\{X(t), t \in T\}$ is a collection of random variables. That is, for each $t \in T$, $X(t)$ is a random variable. The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the *state* of the process at time t. For example, $X(t)$ might equal the total number of customers that have entered a supermarket by time t; or the number of customers in the supermarket at time t; or the total amount of sales that have been recorded in the market by time t; etc.

The set T is called the *index* set of the process. When T is a countable set the stochastic process is said to be a *discrete-time* process. If T is an interval of the real line, the stochastic process is said to be a *continuous-time* process. For instance, $\{X_n, n = 0, 1, \ldots\}$ is a discrete-time stochastic process indexed by the nonnegative integers; while $\{X(t), t \ge 0\}$ is a continuous-time stochastic process indexed by the nonnegative real numbers.

The *state space* of a stochastic process is defined as the set of all possible values that the random variables $X(t)$ can assume.

Thus, a stochastic process is a family of random variables that describes the evolution through time of some (physical) process. We shall see much of stochastic processes in the following chapters of this text.

Example 2.54. Consider a particle that moves along a set of $m + 1$ nodes, labeled $0, 1, \ldots, m$, that are arranged around a circle (see Fig. 2.3). At each step the particle is equally likely to move one position in either the clockwise or counterclockwise direction. That is, if X_n is the position of the particle after its nth step then

$$P\{X_{n+1} = i + 1 | X_n = i\} = P\{X_{n+1} = i - 1 | X_n = i\} = \tfrac{1}{2}$$

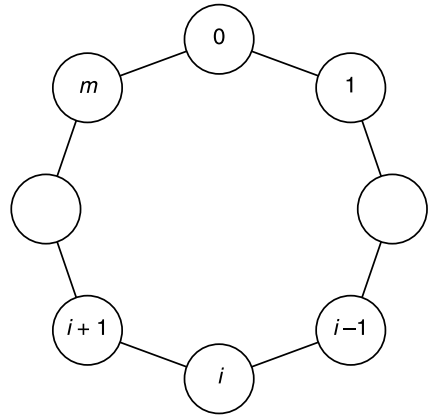

Figure 2.3 Particle moving around a circle.

where $i + 1 \equiv 0$ when $i = m$, and $i - 1 \equiv m$ when $i = 0$. Suppose now that the particle starts at 0 and continues to move around according to the preceding rules until all the nodes $1, 2, \ldots, m$ have been visited. What is the probability that node i, $i = 1, \ldots, m$, is the last one visited?

Solution: Surprisingly enough, the probability that node i is the last node visited can be determined without any computations. To do so, consider the first time that the particle is at one of the two neighbors of node i, that is, the first time that the particle is at one of the nodes $i - 1$ or $i + 1$ (with $m + 1 \equiv 0$). Suppose it is at node $i - 1$ (the argument in the alternative situation is identical). Since neither node i nor $i + 1$ has yet been visited, it follows that i will be the last node visited if and only if $i + 1$ is visited before i. This is so because in order to visit $i + 1$ before i the particle will have to visit all the nodes on the counterclockwise path from $i - 1$ to $i + 1$ before it visits i. But the probability that a particle at node $i - 1$ will visit $i + 1$ before i is just the probability that a particle will progress $m - 1$ steps in a specified direction before progressing one step in the other direction. That is, it is equal to the probability that a gambler who starts with one unit, and wins one when a fair coin turns up heads and loses one when it turns up tails, will have his fortune go up by $m - 1$ before he goes broke. Hence, because the preceding implies that the probability that node i is the last node visited is the same for all i, and because these probabilities must sum to 1, we obtain

$$P\{i \text{ is the last node visited}\} = 1/m, \qquad i = 1, \ldots, m$$ ■

Remark. The argument used in Example 2.54 also shows that a gambler who is equally likely to either win or lose one unit on each gamble will be down n before being up 1 with probability $1/(n + 1)$; or equivalently,

$$P\{\text{gambler is up 1 before being down } n\} = \frac{n}{n+1}$$

Suppose now we want the probability that the gambler is up 2 before being down n. Upon conditioning on whether he reaches up 1 before down n, we obtain that

$$\begin{aligned} &P\{\text{gambler is up 2 before being down } n\} \\ &\quad = P\{\text{up 2 before down } n | \text{up 1 before down } n\}\frac{n}{n+1} \\ &\quad = P\{\text{up 1 before down } n+1\}\frac{n}{n+1} \\ &\quad = \frac{n+1}{n+2}\frac{n}{n+1} = \frac{n}{n+2} \end{aligned}$$

Repeating this argument yields that

$$P\{\text{gambler is up } k \text{ before being down } n\} = \frac{n}{n+k}$$

Exercises

1. An urn contains five red, three orange, and two blue balls. Two balls are randomly selected. What is the sample space of this experiment? Let X represent the number of orange balls selected. What are the possible values of X? Calculate $P\{X = 0\}$.

2. Let X represent the difference between the number of heads and the number of tails obtained when a coin is tossed n times. What are the possible values of X?

3. In Exercise 2, if the coin is assumed fair, then, for $n = 2$, what are the probabilities associated with the values that X can take on?

***4.** Suppose a die is rolled twice. What are the possible values that the following random variables can take on?

(a) The maximum value to appear in the two rolls.

(b) The minimum value to appear in the two rolls.

(c) The sum of the two rolls.

(d) The value of the first roll minus the value of the second roll.

5. If the die in Exercise 4 is assumed fair, calculate the probabilities associated with the random variables in (a)–(d).

6. Suppose five fair coins are tossed. Let E be the event that all coins land heads. Define the random variable I_E

$$I_E = \begin{cases} 1, & \text{if } E \text{ occurs} \\ 0, & \text{if } E^c \text{ occurs} \end{cases}$$

For what outcomes in the original sample space does I_E equal 1? What is $P\{I_E = 1\}$?

7. Suppose a coin having probability 0.7 of coming up heads is tossed three times. Let X denote the number of heads that appear in the three tosses. Determine the probability mass function of X.

8. Suppose the distribution function of X is given by

$$F(b) = \begin{cases} 0, & b < 0 \\ \frac{1}{2}, & 0 \le b < 1 \\ 1, & 1 \le b < \infty \end{cases}$$

What is the probability mass function of X?

9. If the distribution function of F is given by

$$F(b) = \begin{cases} 0, & b < 0 \\ \frac{1}{2}, & 0 \le b < 1 \\ \frac{3}{5}, & 1 \le b < 2 \\ \frac{4}{5}, & 2 \le b < 3 \\ \frac{9}{10}, & 3 \le b < 3.5 \\ 1, & b \ge 3.5 \end{cases}$$

calculate the probability mass function of X.

10. Suppose three fair dice are rolled. What is the probability at most one six appears?

***11.** A ball is drawn from an urn containing three white and three black balls. After the ball is drawn, it is then replaced and another ball is drawn. This goes on indefinitely. What is the probability that of the first four balls drawn, exactly two are white?

12. On a multiple-choice exam with three possible answers for each of the five questions, what is the probability that a student would get four or more correct answers just by guessing?

13. An individual claims to have extrasensory perception (ESP). As a test, a fair coin is flipped ten times, and he is asked to predict in advance the outcome. Our individual gets seven out of ten correct. What is the probability he would have done at least this well if he had no ESP? (Explain why the relevant probability is $P\{X \ge 7\}$ and not $P\{X = 7\}$.)

14. Suppose X has a binomial distribution with parameters 6 and $\frac{1}{2}$. Show that $X = 3$ is the most likely outcome.

15. Let X be binomially distributed with parameters n and p. Show that as k goes from 0 to n, $P(X = k)$ increases monotonically, then decreases monotonically, reaching its largest value

(a) in the case that $(n+1)p$ is an integer, when k equals either $(n+1)p - 1$ or $(n+1)p$,

(b) in the case that $(n+1)p$ is not an integer, when k satisfies $(n+1)p - 1 < k < (n+1)p$.

Hint: Consider $P\{X = k\}/P\{X = k-1\}$ and see for what values of k it is greater or less than 1.

***16.** An airline knows that 5 percent of the people making reservations on a certain flight will not show up. Consequently, their policy is to sell 52 tickets for a

flight that can hold only 50 passengers. What is the probability that there will be a seat available for every passenger who shows up?

17. Suppose that an experiment can result in one of r possible outcomes, the ith outcome having probability p_i, $i = 1, \ldots, r$, $\sum_{i=1}^{r} p_i = 1$. If n of these experiments are performed, and if the outcome of any one of the n does not affect the outcome of the other $n - 1$ experiments, then show that the probability that the first outcome appears x_1 times, the second x_2 times, and the rth x_r times is

$$\frac{n!}{x_1!x_2!\ldots x_r!} p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r} \qquad \text{when } x_1 + x_2 + \cdots + x_r = n$$

This is known as the *multinomial* distribution.

18. In Exercise 17, let X_i denote the number of times that the ith type outcome occurs, $i = 1, \ldots, r$.

(a) For $0 \le j \le n$, use the definition of conditional probability to find $P(X_i = x_i, i = 1, \ldots, r-1 | X_r = j)$.

(b) What can you conclude about the conditional distribution of $X_1, \ldots, X_{r-1}$ given that $X_r = j$?

(c) Give an intuitive explanation for your answer to part (b).

19. In Exercise 17, let X_i denote the number of times the ith outcome appears, $i = 1, \ldots, r$. What is the probability mass function of $X_1 + X_2 + \ldots + X_k$?

20. In this problem we employ the multinomial distribution to solve an extension of the birthday problem. Assuming that each of n individuals is, independently of others, equally likely to have their birthday be any of the 365 days of the year, we want to derive an expression for the probability that the collection of n individuals will contain at least 3 having the same birthday.

(a) For a given partition of the 365 days of the year into a first set of size i, a second set of size $n - 2i$ and a third of size $365 - n + i$, find the probability that every day in the first set is the birthday of exactly 2 of the n individuals, every day in the second set is the birthday of exactly 1 of the n individuals, and every day in the third set is the birthday of none of the n individuals.

(b) For a given value i, determine the number of different partitions of the 365 days of the year into a first set of size i, a second set of size $n - 2i$ and a third set of size $365 - n + i$.

(c) Give an expression for the probability that a collection of n individuals does not contain at least 3 having the same birthday.

Remark. A computation gives that

$$1 - \sum_{i=0}^{44} \frac{365!}{i!(88-2i)!(365-88+i)!} \frac{88!}{2^i} \left(\frac{1}{365}\right)^{88} \approx .504.$$

21. Let X_1 and X_2 be independent binomial random variables, with X_i having parameters (n_i, p_i), $i = 1, 2$.

(a) Find $P(X_1 X_2 = 0)$.

(b) Find $P(X_1 + X_2 = 1)$.
(c) Find $P(X_1 + X_2 = 2)$.

22. If a fair coin is successively flipped, find the probability that a head first appears on the fifth trial.

***23.** A coin having probability p of coming up heads is successively flipped until the rth head appears. Argue that X, the number of flips required, will be $n, n \geq r$, with probability

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \qquad n \geq r$$

This is known as the negative binomial distribution.

Hint: How many successes must there be in the first $n-1$ trials?

24. The probability mass function of X is given by

$$p(k) = \binom{r+k-1}{r-1} p^r (1-p)^k, \qquad k = 0, 1, \ldots$$

Give a possible interpretation of the random variable X.

Hint: See Exercise 23.

In Exercises 25 and 26, suppose that two teams are playing a series of games, each of which is independently won by team A with probability p and by team B with probability $1 - p$. The winner of the series is the first team to win i games.

25. If $i = 4$, find the probability that a total of 7 games are played. Also show that this probability is maximized when $p = 1/2$.

26. Find the expected number of games that are played when
(a) $i = 2$;
(b) $i = 3$.
In both cases, show that this number is maximized when $p = 1/2$.

***27.** A fair coin is independently flipped n times, k times by A and $n - k$ times by B. Show that the probability that A and B flip the same number of heads is equal to the probability that there are a total of k heads.

28. Suppose that we want to generate a random variable X that is equally likely to be either 0 or 1, and that all we have at our disposal is a biased coin that, when flipped, lands on heads with some (unknown) probability p. Consider the following procedure:

1. Flip the coin, and let 0_1, either heads or tails, be the result.
2. Flip the coin again, and let 0_2 be the result.
3. If 0_1 and 0_2 are the same, return to step 1.
4. If 0_2 is heads, set $X = 0$, otherwise set $X = 1$.

(a) Show that the random variable X generated by this procedure is equally likely to be either 0 or 1.
(b) Could we use a simpler procedure that continues to flip the coin until the last two flips are different, and then sets $X = 0$ if the final flip is a head, and sets $X = 1$ if it is a tail?

29. Consider n independent flips of a coin having probability p of landing heads. Say a changeover occurs whenever an outcome differs from the one preceding it. For instance, if the results of the flips are $H\ H\ T\ H\ T\ H\ H\ T$, then there are a total of five changeovers. If $p = 1/2$, what is the probability there are k changeovers?

30. Let X be a Poisson random variable with parameter λ. Show that $P\{X = i\}$ increases monotonically and then decreases monotonically as i increases, reaching its maximum when i is the largest integer not exceeding λ.

Hint: Consider $P\{X = i\}/P\{X = i - 1\}$.

31. Compare the Poisson approximation with the correct binomial probability for the following cases:

(a) $P\{X = 2\}$ when $n = 8$, $p = 0.1$.
(b) $P\{X = 9\}$ when $n = 10$, $p = 0.95$.
(c) $P\{X = 0\}$ when $n = 10$, $p = 0.1$.
(d) $P\{X = 4\}$ when $n = 9$, $p = 0.2$.

32. If you buy a lottery ticket in 50 lotteries, in each of which your chance of winning a prize is $\frac{1}{100}$, what is the (approximate) probability that you will win a prize (a) at least once, (b) exactly once, (c) at least twice?

33. Let X be a random variable with probability density

$$f(x) = \begin{cases} c(1 - x^2), & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the value of c?
(b) What is the cumulative distribution function of X?

34. Let the probability density of X be given by

$$f(x) = \begin{cases} c(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the value of c?
(b) $P\left\{\frac{1}{2} < X < \frac{3}{2}\right\} = ?$

35. The density of X is given by

$$f(x) = \begin{cases} 10/x^2, & \text{for } x > 10 \\ 0, & \text{for } x \le 10 \end{cases}$$

What is the distribution function of X? Find $P\{X > 20\}$.

36. A point is uniformly distributed within the disk of radius 1. That is, its density is

$$f(x, y) = C, \qquad 0 \le x^2 + y^2 \le 1$$

Find the probability that its distance from the origin is less than $x, 0 \le x \le 1$.

37. Let $X_1, X_2, \ldots, X_n$ be independent random variables, each having a uniform distribution over $(0, 1)$. Let $M = \text{maximum } (X_1, X_2, \ldots, X_n)$. Show that the distribution function of M, $F_M(\cdot)$, is given by

$$F_M(x) = x^n, \qquad 0 \le x \le 1$$

What is the probability density function of M?

38. Let $X_1, \ldots, X_{10}$ be independent and identically distributed continuous random variables with distribution function F, and mean $\mu = E[X_i]$. Let $X_{(1)} < X_{(2)} < \cdots < X_{(10)}$ be the values arranged in increasing order. That is, for $i = 1, \ldots, 10$, $X_{(i)}$ is the ith smallest of $X_1, \ldots, X_{10}$.

(a) Find $E[\sum_{i=1}^{10} X_{(i)}]$.

(b) Let $N = \max\{i : X_{(i)} < x\}$. What is the distribution of N.

(c) If m is the median of the distribution (that is, if $F(m) = .5$), find $P(X_{(2)} < m < X_{(8)})$.

39. An urn has 8 red and 12 blue balls. Suppose that balls are chosen at random and removed from the urn, with the process stopping when all the red balls have been removed. Let X be the number of balls that have been removed when the process stops.

(a) Find $P(X = 14)$.

(b) Find the probability that a specified blue ball remains in the urn.

(c) Find $E[X]$.

40. Suppose that two teams are playing a series of games, each of which is independently won by team A with probability p and by team B with probability $1 - p$. The winner of the series is the first team to win four games. Find the expected number of games that are played, and evaluate this quantity when $p = 1/2$.

41. Consider the case of arbitrary p in Exercise 29. Compute the expected number of changeovers.

42. Suppose that each coupon obtained is, independent of what has been previously obtained, equally likely to be any of m different types. Find the expected number of coupons one needs to obtain in order to have at least one of each type.

Hint: Let X be the number needed. It is useful to represent X by

$$X = \sum_{i=1}^{m} X_i$$

where each X_i is a geometric random variable.

43. An urn contains $n + m$ balls, of which n are red and m are black. They are withdrawn from the urn, one at a time and without replacement. Let X be the number of red balls removed before the first black ball is chosen. We are interested in determining $E[X]$. To obtain this quantity, number the red balls from 1 to n. Now define the random variables $X_i, i = 1, \ldots, n$, by

$$X_i = \begin{cases} 1, & \text{if red ball } i \text{ is taken before any black ball is chosen} \\ 0, & \text{otherwise} \end{cases}$$

(a) Express X in terms of the X_i.
(b) Find $E[X]$.

44. In Exercise 43, let Y denote the number of red balls chosen after the first but before the second black ball has been chosen.
(a) Express Y as the sum of n random variables, each of which is equal to either 0 or 1.
(b) Find $E[Y]$.
(c) Compare $E[Y]$ to $E[X]$ obtained in Exercise 43.
(d) Can you explain the result obtained in part (c)?

45. A total of r keys are to be put, one at a time, in k boxes, with each key independently being put in box i with probability $p_i, \sum_{i=1}^{k} p_i = 1$. Each time a key is put in a nonempty box, we say that a collision occurs. Find the expected number of collisions.

46. If X is a nonnegative integer valued random variable, show that

(a) $$E[X] = \sum_{n=1}^{\infty} P\{X \geq n\} = \sum_{n=0}^{\infty} P\{X > n\}$$

Hint: Define the sequence of random variables $I_n, n \geq 1$, by

$$I_n = \begin{cases} 1, & \text{if } n \leq X \\ 0, & \text{if } n > X \end{cases}$$

Now express X in terms of the I_n.

(b) If X and Y are both nonnegative integer valued random variables, show that

$$E[XY] = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} P(X \geq n, Y \geq m)$$

***47.** Consider three trials, each of which is either a success or not. Let X denote the number of successes. Suppose that $E[X] = 1.8$.
(a) What is the largest possible value of $P\{X = 3\}$?
(b) What is the smallest possible value of $P\{X = 3\}$?
In both cases, construct a probability scenario that results in $P\{X = 3\}$ having the desired value.

48. For any event A, we define the random variable $I\{A\}$, called the *indicator variable* for A, by letting it equal 1 when A occurs and 0 when A does not. Now, if $X(t)$ is a nonnegative random variable for all $t \geq 0$, then it follows from a result in real analysis called Fubini's theorem that

$$E\left[\int_0^{\infty} X(t)\, dt\right] = \int_0^{\infty} E[X(t)]\, dt$$

Suppose that X is a nonnegative random variable and that g is a differentiable function such that $g(0) = 0$.

(a) Show that

$$g(X) = \int_0^\infty I\{t < X\} g'(t)\, dt$$

(b) Show that

$$E[g(X)] = \int_0^\infty \bar{F}(t)\, g'(t)\, dt$$

where $\bar{F}(t) = 1 - F(t) = P(X > t)$.

***49.** Prove that $E[X^2] \geq (E[X])^2$. When do we have equality?

50. Let c be a constant. Show that

(a) $\text{Var}(cX) = c^2\text{Var}(X)$;

(b) $\text{Var}(c + X) = \text{Var}(X)$.

51. A coin, having probability p of landing heads, is flipped until a head appears for the rth time. Let N denote the number of flips required. Calculate $E[N]$.

Hint: There is an easy way of doing this. It involves writing N as the sum of r geometric random variables.

52. **(a)** Calculate $E[X]$ for the maximum random variable of Exercise 37.

(b) Calculate $E[X]$ for X as in Exercise 33.

(c) Calculate $E[X]$ for X as in Exercise 34.

53. If X is uniform over (0, 1), calculate $E[X^n]$ and $\text{Var}(X^n)$.

54. Each member of a population is either type 1 with probability p_1 or type 2 with probability $p_2 = 1 - p_1$. Independent of other pairs, two individuals of the same type will be friends with probability α, whereas two individuals of different types will be friends with probability β. Let P_i be the probability that a type i person will be friends with a randomly chosen other person.

(a) Find P_1 and P_2.

Let $F_{k,r}$ be the event that persons k and r are friends.

(b) Find $P(F_{1,2})$.

(c) Show that $P(F_{1,2}|F_{1,3}) \geq P(F_{1,2})$.

Hint for (c): It might be useful to let X be such that $P(X = P_i) = p_i$, $i = 1, 2$.

55. Suppose that the joint probability mass function of X and Y is

$$P(X = i, Y = j) = \binom{j}{i} e^{-2\lambda}\lambda^j/j!, \quad 0 \leq i \leq j$$

(a) Find the probability mass function of Y.

(b) Find the probability mass function of X.

(c) Find the probability mass function of $Y - X$.

56. There are n types of coupons. Each newly obtained coupon is, independently, type i with probability p_i, $i = 1, \ldots, n$. Find the expected number and the variance of the number of distinct types obtained in a collection of k coupons.

57. Suppose that X and Y are independent binomial random variables with parameters (n, p) and (m, p). Argue probabilistically (no computations necessary) that $X + Y$ is binomial with parameters $(n + m, p)$.

58. An urn contains $2n$ balls, of which r are red. The balls are randomly removed in n successive pairs. Let X denote the number of pairs in which both balls are red.

(a) Find $E[X]$.

(b) Find $\text{Var}(X)$.

59. Let X_1, X_2, X_3, and X_4 be independent continuous random variables with a common distribution function F and let

$$p = P\{X_1 < X_2 > X_3 < X_4\}$$

(a) Argue that the value of p is the same for all continuous distribution functions F.

(b) Find p by integrating the joint density function over the appropriate region.

(c) Find p by using the fact that all 4! possible orderings of $X_1, \ldots, X_4$ are equally likely.

60. Let X and Y be independent random variables with means μ_x and μ_y and variances σ_x^2 and σ_y^2. Show that

$$\text{Var}(XY) = \sigma_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \mu_x^2\sigma_y^2$$

61. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed continuous random variables. We say that a record occurs at time n if $X_n > \max(X_1, \ldots, X_{n-1})$. That is, X_n is a record if it is larger than each of $X_1, \ldots, X_{n-1}$. Show

(a) $P\{\text{a record occurs at time } n\} = 1/n$;

(b) $E[\text{number of records by time } n] = \sum_{i=1}^{n} 1/i$;

(c) $\text{Var}(\text{number of records by time } n) = \sum_{i=1}^{n}(i-1)/i^2$;

(d) Let $N = \min\{n: n > 1 \text{ and a record occurs at time } n\}$. Show $E[N] = \infty$.

Hint: For (b) and (c) represent the number of records as the sum of indicator (that is, Bernoulli) random variables.

62. Let $a_1 < a_2 < \cdots < a_n$ denote a set of n numbers, and consider any permutation of these numbers. We say that there is an inversion of a_i and a_j in the permutation if $i < j$ and a_j precedes a_i. For instance the permutation 4, 2, 1, 5, 3 has 5 inversions—(4, 2), (4, 1), (4, 3), (2, 1), (5, 3). Consider now a random permutation of $a_1, a_2, \ldots, a_n$—in the sense that each of the $n!$ permutations is equally likely to be chosen—and let N denote the number of inversions in this permutation. Also, let

$$N_i = \text{number of } k: k < i,\ a_i \text{ precedes } a_k \text{ in the permutation}$$

and note that $N = \sum_{i=1}^{n} N_i$.

(a) Show that $N_1, \ldots, N_n$ are independent random variables.

(b) What is the distribution of N_i?
(c) Compute $E[N]$ and $\text{Var}(N)$.

63. Let X denote the number of white balls selected when k balls are chosen at random from an urn containing n white and m black balls.
(a) Compute $P\{X = i\}$.
(b) Let, for $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$,

$$X_i = \begin{cases} 1, & \text{if the } i\text{th ball selected is white} \\ 0, & \text{otherwise} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{if white ball } j \text{ is selected} \\ 0, & \text{otherwise} \end{cases}$$

Compute $E[X]$ in two ways by expressing X first as a function of the X_is and then of the Y_js.

***64.** Show that $\text{Var}(X) = 1$ when X is the number of men who select their own hats in Example 2.30.

65. The number of traffic accidents on successive days are independent Poisson random variables with mean 2.
(a) Find the probability that 3 of the next 5 days have two accidents.
(b) Find the probability that there are a total of six accidents over the next 2 days.
(c) If each accident is independently a "major accident" with probability p, what is the probability there are no major accidents tomorrow?

***66.** Show that the random variables $X_1, \ldots, X_n$ are independent if for each $i = 2, \ldots, n$, X_i is independent of $X_1, \ldots, X_{i-1}$.

Hint: $X_1, \ldots, X_n$ are independent if for any sets $A_1, \ldots, A_n$

$$P(X_j \in A_j, j = 1, \ldots, n) = \prod_{j=1}^{n} P(X_j \in A_j)$$

On the other hand X_i is independent of $X_1, \ldots, X_{i-1}$ if for any sets $A_1, \ldots, A_i$

$$P(X_i \in A_i | X_j \in A_j, j = 1, \ldots, i-1) = P(X_i \in A_i)$$

67. Calculate the moment generating function of the uniform distribution on $(0, 1)$. Obtain $E[X]$ and $\text{Var}[X]$ by differentiating.

68. Let X and W be the working and subsequent repair times of a certain machine. Let $Y = X + W$ and suppose that the joint probability density of X and Y is

$$f_{X,Y}(x, y) = \lambda^2 e^{-\lambda y}, \quad 0 < x < y < \infty$$

(a) Find the density of X.
(b) Find the density of Y.
(c) Find the joint density of X and W.
(d) Find the density of W.

69. In deciding upon the appropriate premium to charge, insurance companies sometimes use the exponential principle, defined as follows. With X as the random amount that it will have to pay in claims, the premium charged by the insurance company is

$$P = \frac{1}{a} \ln(E[e^{aX}])$$

where a is some specified positive constant. Find P when X is an exponential random variable with parameter λ, and $a = \alpha\lambda$, where $0 < \alpha < 1$.

70. Calculate the moment generating function of a geometric random variable.

***71.** Show that the sum of independent identically distributed exponential random variables has a gamma distribution.

72. Successive monthly sales are independent normal random variables with mean 100 and variance 100.

(a) Find the probability that at least one of the next 5 months has sales above 115.

(b) Find the probability that the total number of sales over the next 5 months exceeds 530.

73. Consider n people and suppose that each of them has a birthday that is equally likely to be any of the 365 days of the year. Furthermore, assume that their birthdays are independent, and let A be the event that no two of them share the same birthday. Define a "trial" for each of the $\binom{n}{2}$ pairs of people and say that trial (i, j), $i \neq j$, is a success if persons i and j have the same birthday. Let $S_{i,j}$ be the event that trial (i, j) is a success.

(a) Find $P(S_{i,j})$, $i \neq j$.

(b) Are $S_{i,j}$ and $S_{k,r}$ independent when i, j, k, r are all distinct?

(c) Are $S_{i,j}$ and $S_{k,j}$ independent when i, j, k are all distinct?

(d) Are $S_{1,2}$, $S_{1,3}$, $S_{2,3}$independent?

(e) Employ the Poisson paradigm to approximate $P(A)$.

(f) Show that this approximation yields that $P(A) \approx .5$ when $n = 23$.

(g) Let B be the event that no three people have the same birthday. Approximate the value of n that makes $P(B) \approx .5$. (Whereas a simple combinatorial argument explicitly determines $P(A)$, the exact determination of $P(B)$ is very complicated.)

Hint: Define a trial for each triplet of people.

***74.** If X is Poisson with parameter λ, show that its Laplace transform is given by

$$g(u) = E[e^{-uX}] = e^{\lambda(e^{-u}-1)}$$

75. Consider Example 2.48. Find $\text{Cov}(X_i, X_j)$ in terms of the a_{rs}.

76. Use Chebyshev's inequality to prove the *weak law of large numbers*. Namely, if $X_1, X_2, \ldots$ are independent and identically distributed with mean μ and vari-

ance σ^2 then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \qquad \text{as } n \to \infty$$

77. Suppose that X is a random variable with mean 10 and variance 15. What can we say about $P\{5 < X < 15\}$?

78. Let $X_1, X_2, \ldots, X_{10}$ be independent Poisson random variables with mean 1.
(a) Use the Markov inequality to get a bound on $P\{X_1 + \cdots + X_{10} \geq 15\}$.
(b) Use the central limit theorem to approximate $P\{X_1 + \cdots + X_{10} \geq 15\}$.

79. If X is normally distributed with mean 1 and variance 4, use the tables to find $P\{2 < X < 3\}$.

***80.** Show that

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}$$

Hint: Let X_n be Poisson with mean n. Use the central limit theorem to show that $P\{X_n \leq n\} \to \frac{1}{2}$.

81. Let X and Y be independent normal random variables, each having parameters μ and σ^2. Show that $X + Y$ is independent of $X - Y$.

Hint: Find their joint moment generating function.

82. Let $\phi(t_1, \ldots, t_n)$ denote the joint moment generating function of $X_1, \ldots, X_n$.
(a) Explain how the moment generating function of X_i, $\phi_{X_i}(t_i)$, can be obtained from $\phi(t_1, \ldots, t_n)$.
(b) Show that $X_1, \ldots, X_n$ are independent if and only if

$$\phi(t_1, \ldots, t_n) = \phi_{X_1}(t_1) \cdots \phi_{X_n}(t_n)$$

83. With $K(t) = \log(E\left[e^{tX}\right])$, show that

$$K'(0) = E[X], \quad K''(0) = \text{Var}(X)$$

84. Teams 1, 2, 3, 4 are all scheduled to play each of the other teams 10 times. Whenever team i plays team j, team i is the winner with probability $P_{i,j}$, where

$$P_{1,2} = .6, \quad P_{1,3} = .7, \quad P_{1,4} = .75$$
$$P_{2,3} = .6, \quad P_{2,4} = .70, \quad P_{3,4} = .5$$

(a) Approximate the probability that team 1 wins at least 20 games.
Suppose now that we want to approximate the probability that team 2 wins at least as many games as does team 1. To do so, let X be the number of games that team 2 wins against team 1, let Y be the total number of games that team 2 wins against teams 3 and 4, and let Z be the total number of games that team 1 wins against teams 3 and 4.
(b) Are X, Y, Z independent.

(c) Express the event that team 2 wins at least as many games as does team 1 in terms of the random variables X, Y, Z.

(d) Approximate the probability that team 2 wins at least as many games as does team 1.

85. The *standard deviation* of a random variable is the positive square root of its variance. Letting σ_X and σ_Y denote the standard deviations of the random variables X and Y, we define the *correlation* of X and Y by

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

(a) Starting with the inequality $\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \geq 0$, show that $-1 \leq \text{Corr}(X, Y)$.

(b) Prove the inequality

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

(c) If σ_{X+Y} is the standard deviation of $X + Y$, show that

$$\sigma_{X+Y} \leq \sigma_X + \sigma_Y$$

***86.** Each new book donated to a library must be processed. Suppose that the time it takes a librarian to process a book has mean 10 minutes and standard deviation 3 minutes. If a librarian has 40 books that must be processed one at a time,

(a) approximate the probability that it will take more than 420 minutes to process all these books;

(b) approximate the probability that at least 25 books will be processed in the first 240 minutes.

87. Recall that X is said to be a gamma random variable with parameters (α, λ) if its density is

$$f(x) = \lambda e^{-\lambda x} (\lambda x)^{\alpha - 1} / \Gamma(\alpha), \quad x > 0$$

(a) If Z is a standard normal random variable, show that Z^2 is a gamma random variable with parameters $(1/2, 1/2)$.

(b) If $Z_1, \ldots, Z_n$ are independent standard normal random variables, then $\sum_{i=1}^{n} Z_i^2$ is said to be a *chi-squared* random variable with n degrees of freedom. Explain how you can use results from Example 2.39 to show that the density function of $\sum_{i=1}^{n} Z_i^2$ is

$$f(x) = \frac{e^{-x/2} x^{n/2-1}}{2^{n/2} \Gamma(n/2)}, \quad x > 0$$

References

[1] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, John Wiley, New York, 1957.
[2] M. Fisz, Probability Theory and Mathematical Statistics, John Wiley, New York, 1963.
[3] E. Parzen, Modern Probability Theory and Its Applications, John Wiley, New York, 1960.
[4] S. Ross, A First Course in Probability, Tenth Edition, Prentice Hall, New Jersey, 2018.

Conditional Probability and Conditional Expectation

3

3.1 Introduction

One of the most useful concepts in probability theory is that of conditional probability and conditional expectation. The reason is twofold. First, in practice, we are often interested in calculating probabilities and expectations when some partial information is available; hence, the desired probabilities and expectations are conditional ones. Secondly, in calculating a desired probability or expectation it is often extremely useful to first "condition" on some appropriate random variable.

3.2 The Discrete Case

Recall that for any two events E and F, the conditional probability of E given F is defined, as long as $P(F) > 0$, by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Hence, if X and Y are discrete random variables, then it is natural to define the *conditional probability mass function* of X given that $Y = y$, by

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for all values of y such that $P\{Y = y\} > 0$. Similarly, the conditional probability distribution function of X given that $Y = y$ is defined, for all y such that $P\{Y = y\} > 0$, by

$$\begin{aligned} F_{X|Y}(x|y) &= P\{X \le x|Y = y\} \\ &= \sum_{a \le x} p_{X|Y}(a|y) \end{aligned}$$

Finally, the conditional expectation of X given that $Y = y$ is defined by

$$E[X|Y = y] = \sum_x x P\{X = x|Y = y\}$$

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00008-1

$$= \sum_x x p_{X|Y}(x|y)$$

In other words, the definitions are exactly as before with the exception that everything is now conditional on the event that $Y = y$. If X is independent of Y, then the conditional mass function, distribution, and expectation are the same as the unconditional ones. This follows, since if X is independent of Y, then

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\ &= P\{X = x\} \end{aligned}$$

Example 3.1. Suppose that $p(x, y)$, the joint probability mass function of X and Y, is given by

$$p(1, 1) = 0.5, \quad p(1, 2) = 0.1, \quad p(2, 1) = 0.1, \quad p(2, 2) = 0.3$$

Calculate the conditional probability mass function of X given that $Y = 1$.

Solution: We first note that

$$p_Y(1) = \sum_x p(x, 1) = p(1, 1) + p(2, 1) = 0.6$$

Hence,

$$\begin{aligned} p_{X|Y}(1|1) &= P\{X = 1|Y = 1\} \\ &= \frac{P\{X = 1, Y = 1\}}{P\{Y = 1\}} \\ &= \frac{p(1, 1)}{p_Y(1)} \\ &= \frac{5}{6} \end{aligned}$$

Similarly,

$$p_{X|Y}(2|1) = \frac{p(2, 1)}{p_Y(1)} = \frac{1}{6}$$ ■

Example 3.2. If X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p), calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

Solution: With $q = 1 - p$,

$$\begin{aligned} P\{X_1 = k|X_1 + X_2 = m\} &= \frac{P\{X_1 = k, X_1 + X_2 = m\}}{P\{X_1 + X_2 = m\}} \\ &= \frac{P\{X_1 = k, X_2 = m - k\}}{P\{X_1 + X_2 = m\}} \end{aligned}$$

$$= \frac{P\{X_1 = k\}P\{X_2 = m-k\}}{P\{X_1 + X_2 = m\}}$$
$$= \frac{\binom{n_1}{k} p^k q^{n_1-k} \binom{n_2}{m-k} p^{m-k} q^{n_2-m+k}}{\binom{n_1+n_2}{m} p^m q^{n_1+n_2-m}}$$

where we have used that $X_1 + X_2$ is a binomial random variable with parameters $(n_1 + n_2, p)$ (see Example 2.44). Thus, the conditional probability mass function of X_1, given that $X_1 + X_2 = m$, is

$$P\{X_1 = k | X_1 + X_2 = m\} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}} \tag{3.1}$$

The distribution given by Eq. (3.1), first seen in Example 2.35, is known as the *hypergeometric* distribution. It is the distribution of the number of blue balls that are chosen when a sample of m balls is randomly chosen from an urn that contains n_1 blue and n_2 red balls. (To intuitively see why the conditional distribution is hypergeometric, consider $n_1 + n_2$ independent trials that each result in a success with probability p; let X_1 represent the number of successes in the first n_1 trials and let X_2 represent the number of successes in the final n_2 trials. Because all trials have the same probability of being a success, each of the $\binom{n_1+n_2}{m}$ subsets of m trials is equally likely to be the success trials; thus, the number of the m success trials that are among the first n_1 trials is a hypergeometric random variable.) ■

Example 3.3. If X and Y are independent Poisson random variables with respective means λ_1 and λ_2, calculate the conditional expected value of X given that $X + Y = n$.

Solution: Let us first calculate the conditional probability mass function of X given that $X + Y = n$. We obtain

$$P\{X = k | X + Y = n\} = \frac{P\{X = k, X + Y = n\}}{P\{X + Y = n\}}$$
$$= \frac{P\{X = k, Y = n - k\}}{P\{X + Y = n\}}$$
$$= \frac{P\{X = k\}P\{Y = n - k\}}{P\{X + Y = n\}}$$

where the last equality follows from the assumed independence of X and Y. Recalling (see Example 2.37) that $X + Y$ has a Poisson distribution with mean $\lambda_1 + \lambda_2$, the preceding equation equals

$$P\{X = k | X + Y = n\} = \frac{e^{-\lambda_1}\lambda_1^k}{k!} \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1}$$

$$= \frac{n!}{(n-k)!k!} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}$$
$$= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

In other words, the conditional distribution of X given that $X+Y=n$ is the binomial distribution with parameters n and $\lambda_1/(\lambda_1+\lambda_2)$. Hence,

$$E\{X|X+Y=n\} = n\frac{\lambda_1}{\lambda_1+\lambda_2}$$ ■

Conditional expectations possess all of the properties of ordinary expectations. For example such identities such as

$$E\left[\sum_{i=1}^{n} X_i | Y=y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$
$$E[h(X)|Y=y] = \sum_x h(x)P(X=x|Y=y)$$

remain valid.

Example 3.4. There are n components. On a rainy day, component i will function with probability p_i; on a nonrainy day, component i will function with probability q_i, for $i=1,\ldots,n$. It will rain tomorrow with probability α. Calculate the conditional expected number of components that function tomorrow, given that it rains.

Solution: Let

$$X_i = \begin{cases} 1, & \text{if component } i \text{ functions tomorrow} \\ 0, & \text{otherwise} \end{cases}$$

Then, with Y defined to equal 1 if it rains tomorrow, and 0 otherwise, the desired conditional expectation is obtained as follows.

$$E\left[\sum_{i=1}^{n} X_i | Y=1\right] = \sum_{i=1}^{n} E[X_i|Y=1]$$
$$= \sum_{i=1}^{n} p_i$$ ■

3.3 The Continuous Case

If X and Y have a joint probability density function $f(x,y)$, then the *conditional probability density function* of X, given that $Y=y$, is defined for all values of y such

that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

To motivate this definition, multiply the left side by dx and the right side by $(dx\, dy)/dy$ to get

$$\begin{aligned} f_{X|Y}(x|y)\, dx &= \frac{f(x, y)\, dx\, dy}{f_Y(y)\, dy} \\ &\approx \frac{P\{x \le X \le x + dx, y \le Y \le y + dy\}}{P\{y \le Y \le y + dy\}} \\ &= P\{x \le X \le x + dx | y \le Y \le y + dy\} \end{aligned}$$

In other words, for small values dx and dy, $f_{X|Y}(x|y)\, dx$ is approximately the conditional probability that X is between x and $x + dx$ given that Y is between y and $y + dy$.

The *conditional expectation* of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

Example 3.5. Suppose the joint density of X and Y is given by

$$f(x, y) = \begin{cases} 6xy(2 - x - y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$.

Solution: We first compute the conditional density

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ &= \frac{6xy(2 - x - y)}{\int_0^1 6xy(2 - x - y)\, dx} \\ &= \frac{6xy(2 - x - y)}{y(4 - 3y)} \\ &= \frac{6x(2 - x - y)}{4 - 3y} \end{aligned}$$

Hence,

$$\begin{aligned} E[X|Y = y] &= \int_0^1 \frac{6x^2(2 - x - y)\, dx}{4 - 3y} \\ &= \frac{(2 - y)2 - \frac{6}{4}}{4 - 3y} \end{aligned}$$

$$= \frac{5-4y}{8-6y} \qquad \blacksquare$$

Example 3.6 (The t-Distribution)**.** If Z and Y are independent, with Z having a standard normal distribution and Y having a chi-squared distribution with n degrees of freedom, then the random variable T defined by

$$T = \frac{Z}{\sqrt{Y/n}} = \sqrt{n}\frac{Z}{\sqrt{Y}}$$

is said to be a *t-random variable* with n degrees of freedom. To compute its density function, we first derive the conditional density of T given that $Y = y$. Because Z and Y are independent, the conditional distribution of T given that $Y = y$ is the distribution of $\sqrt{n/y}Z$, which is normal with mean 0 and variance n/y. Hence, the conditional density function of T given that $Y = y$ is

$$f_{T|Y}(t|y) = \frac{1}{\sqrt{2\pi n/y}} e^{-t^2 y/2n} = \frac{y^{1/2}}{\sqrt{2\pi n}} e^{-t^2 y/2n}, \quad -\infty < t < \infty$$

Using the preceding, along with the following formula for the chi-squared density derived in Exercise 87 of Chapter 2,

$$f_Y(y) = \frac{e^{-y/2} y^{n/2-1}}{2^{n/2}\Gamma(n/2)}, \quad y > 0$$

we obtain the density function of T:

$$f_T(t) = \int_0^\infty f_{T,Y}(t, y)\, dy = \int_0^\infty f_{T|Y}(t|y) f_Y(y)\, dy$$

With

$$K = \frac{1}{\sqrt{\pi n}\, 2^{(n+1)/2}\Gamma(n/2)}, \quad c = \frac{t^2+n}{2n} = \frac{1}{2}\left(1 + \frac{t^2}{n}\right)$$

this yields

$$\begin{aligned} f_T(t) &= \frac{1}{K}\int_0^\infty e^{-cy} y^{(n-1)/2}\, dy \\ &= \frac{c^{-(n+1)/2}}{K}\int_0^\infty e^{-x} x^{(n-1)/2}\, dx \quad \text{(by letting } x = cy\text{)} \\ &= \frac{c^{-(n+1)/2}}{K}\Gamma\left(\frac{n+1}{2}\right) \\ &= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < \infty \end{aligned} \qquad \blacksquare$$

Example 3.7. The joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{2}ye^{-xy}, & 0 < x < \infty,\ 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

What is $E[e^{X/2}|Y=1]$?

Solution: The conditional density of X, given that $Y=1$, is given by

$$\begin{aligned} f_{X|Y}(x|1) &= \frac{f(x,1)}{f_Y(1)} \\ &= \frac{\frac{1}{2}e^{-x}}{\int_0^\infty \frac{1}{2}e^{-x}\,dx} = e^{-x} \end{aligned}$$

Hence, by Proposition 2.1,

$$\begin{aligned} E\left[e^{X/2}|Y=1\right] &= \int_0^\infty e^{x/2} f_{X|Y}(x|1)\,dx \\ &= \int_0^\infty e^{x/2}e^{-x}\,dx \\ &= 2 \end{aligned}$$

■

Example 3.8. Let X_1 and X_2 be independent exponential random variables with rates μ_1 and μ_2. Find the conditional density of X_1 given that $X_1 + X_2 = t$.

Solution: To begin, let us first note that if $f(x,y)$ is the joint density of X, Y, then the joint density of X and $X+Y$ is

$$f_{X,X+Y}(x,t) = f(x,t-x)$$

which is easily seen by noting that the Jacobian of the transformation

$$g_1(x,y) = x, \quad g_2(x,y) = x + y$$

is equal to 1.
Applying the preceding to our example yields

$$\begin{aligned} f_{X_1|X_1+X_2}(x|t) &= \frac{f_{X_1,X_1+X_2}(x,t)}{f_{X_1+X_2}(t)} \\ &= \frac{\mu_1 e^{-\mu_1 x}\mu_2 e^{-\mu_2(t-x)}}{f_{X_1+X_2}(t)}, \quad 0 \le x \le t \\ &= Ce^{-(\mu_1-\mu_2)x}, \quad 0 \le x \le t \end{aligned}$$

where

$$C = \frac{\mu_1\mu_2 e^{-\mu_2 t}}{f_{X_1+X_2}(t)}$$

Now, if $\mu_1 = \mu_2$, then

$$f_{X_1|X_1+X_2}(x|t) = C, \quad 0 \le x \le t$$

yielding that $C = 1/t$ and that X_1 given $X_1 + X_2 = t$ is uniformly distributed on $(0, t)$. On the other hand, if $\mu_1 \neq \mu_2$, then we use

$$1 = \int_0^t f_{X_1|X_1+X_2}(x|t)dx = \frac{C}{\mu_1 - \mu_2}\left(1 - e^{-(\mu_1-\mu_2)t}\right)$$

to obtain

$$C = \frac{\mu_1 - \mu_2}{1 - e^{-(\mu_1-\mu_2)t}}$$

thus yielding the result:

$$f_{X_1|X_1+X_2}(x|t) = \frac{(\mu_1 - \mu_2)e^{-(\mu_1-\mu_2)x}}{1 - e^{-(\mu_1-\mu_2)t}}$$

An interesting byproduct of our analysis is that

$$f_{X_1+X_2}(t) = \frac{\mu_1\mu_2 e^{-\mu_2 t}}{C} = \begin{cases} \mu^2 t e^{-\mu t}, & \text{if } \mu_1 = \mu_2 = \mu \\ \frac{\mu_1\mu_2(e^{-\mu_2 t} - e^{-\mu_1 t})}{\mu_1 - \mu_2}, & \text{if } \mu_1 \neq \mu_2 \end{cases}$$

■

3.4 Computing Expectations by Conditioning

Let us denote by $E[X|Y]$ that function of the random variable Y whose value at $Y = y$ is $E[X|Y = y]$. Note that $E[X|Y]$ is itself a random variable. An extremely important property of conditional expectation is that for all random variables X and Y

$$E[X] = E\big[E[X|Y]\big] \tag{3.2}$$

If Y is a discrete random variable, then Eq. (3.2) states that

$$E[X] = \sum_y E[X|Y = y]P\{Y = y\} \tag{3.2a}$$

while if Y is continuous with density $f_Y(y)$, then Eq. (3.2) says that

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y]f_Y(y)\,dy \tag{3.2b}$$

We now give a proof of Eq. (3.2) in the case where X and Y are both discrete random variables.

Proof of Eq. (3.2) When X and Y Are Discrete. We must show that

$$E[X] = \sum_y E[X|Y=y]P\{Y=y\} \tag{3.3}$$

Now, the right side of the preceding can be written

$$\begin{aligned}
\sum_y E[X|Y=y]P\{Y=y\} &= \sum_y \sum_x x P\{X=x|Y=y\}P\{Y=y\} \\
&= \sum_y \sum_x x \frac{P\{X=x, Y=y\}}{P\{Y=y\}} P\{Y=y\} \\
&= \sum_y \sum_x x P\{X=x, Y=y\} \\
&= \sum_x x \sum_y P\{X=x, Y=y\} \\
&= \sum_x x P\{X=x\} \\
&= E[X]
\end{aligned}$$

and the result is obtained. ■

One way to understand Eq. (3.3) is to interpret it as follows. It states that to calculate $E[X]$ we may take a weighted average of the conditional expected value of X given that $Y = y$, each of the terms $E[X|Y=y]$ being weighted by the probability of the event on which it is conditioned.

The following examples will indicate the usefulness of Eq. (3.2).

Example 3.9. Sam will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability book is Poisson distributed with mean 2 and if the number of misprints in his history chapter is Poisson distributed with mean 5, then assuming Sam is equally likely to choose either book, what is the expected number of misprints that Sam will come across?

Solution: Let X be the number of misprints. Because it would be easy to compute $E[X]$ if we know which book Sam chooses, let

$$Y = \begin{cases} 1, & \text{if Sam chooses his history book} \\ 2, & \text{if chooses his probability book} \end{cases}$$

Conditioning on Y yields

$$\begin{aligned}
E[X] &= E[X|Y=1]P\{Y=1\} + E[X|Y=2]P\{Y=2\} \\
&= 5\left(\tfrac{1}{2}\right) + 2\left(\tfrac{1}{2}\right) \\
&= \tfrac{7}{2}
\end{aligned}$$

■

Example 3.10 (The Expectation of the Sum of a Random Number of Random Variables)**.** Suppose that the expected number of accidents per week at an industrial plant is four. Suppose also that the numbers of workers injured in each accident are independent random variables with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week?

Solution: Letting N denote the number of accidents and X_i the number injured in the ith accident, $i = 1, 2, \ldots$, then the total number of injuries can be expressed as $\sum_{i=1}^{N} X_i$. Hence, we need to compute the expected value of the sum of a random number of random variables. Because it is easy to compute the expected value of the sum of a fixed number of random variables, let us try conditioning on N. This yields

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

But

$$\begin{aligned} E\left[\sum_{1}^{N} X_i | N = n\right] &= E\left[\sum_{1}^{n} X_i | N = n\right] \\ &= E\left[\sum_{1}^{n} X_i\right] \quad \text{by the independence of } X_i \text{ and } N \\ &= nE[X] \end{aligned}$$

which yields

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

and thus

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

Therefore, in our example, the expected number of injuries during a week equals $4 \times 2 = 8$. ■

The random variable $\sum_{i=1}^{N} X_i$, equal to the sum of a random number N of independent and identically distributed random variables that are also independent of N, is called a *compound random variable*. As just shown in Example 3.10, the expected value of a compound random variable is $E[X]E[N]$. Its variance will be derived in Example 3.20.

If there is some random variable Y such that it would be easy to compute $E[X]$ if we knew the value of Y, then conditioning on Y is likely to be a good strategy for determining $E[X]$. When there is no obvious random variable to condition on, it often turns out to be useful to condition on the first thing that occurs. This is illustrated in the following two examples.

Example 3.11 (The Mean of a Geometric Distribution)**.** A coin, having probability p of coming up heads, is to be successively flipped until the first head appears. What is the expected number of flips required?

Solution: Let N be the number of flips required, and let

$$Y = \begin{cases} 1, & \text{if the first flip results in a head} \\ 0, & \text{if the first flip results in a tail} \end{cases}$$

Now,

$$\begin{aligned} E[N] &= E[N|Y=1]P\{Y=1\} + E[N|Y=0]P\{Y=0\} \\ &= pE[N|Y=1] + (1-p)E[N|Y=0] \end{aligned} \tag{3.4}$$

However,

$$E[N|Y=1] = 1, \quad E[N|Y=0] = 1 + E[N] \tag{3.5}$$

To see why Eq. (3.5) is true, consider $E[N|Y=1]$. Since $Y=1$, we know that the first flip resulted in heads and so, clearly, the expected number of flips required is 1. On the other hand if $Y=0$, then the first flip resulted in tails. However, since the successive flips are assumed independent, it follows that, after the first tail, the expected additional number of flips until the first head is just $E[N]$. Hence $E[N|Y=0] = 1 + E[N]$. Substituting Eq. (3.5) into Eq. (3.4) yields

$$E[N] = p + (1-p)(1 + E[N])$$

or

$$E[N] = 1/p$$ ■

Because the random variable N is a geometric random variable with probability mass function $p(n) = p(1-p)^{n-1}$, its expectation could easily have been computed from $E[N] = \sum_1^\infty np(n)$ without recourse to conditional expectation. However, if you attempt to obtain the solution to our next example without using conditional expectation, you will quickly learn what a useful technique "conditioning" can be.

Example 3.12. A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to his mine after five hours. Assuming that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until the miner reaches safety?

Solution: Let X denote the time until the miner reaches safety, and let Y denote the door he initially chooses. Now,

$$\begin{aligned} E[X] &= E[X|Y=1]P\{Y=1\} + E[X|Y=2]P\{Y=2\} \\ &\quad + E[X|Y=3]P\{Y=3\} \\ &= \tfrac{1}{3}\big(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]\big) \end{aligned}$$

However,

$$\begin{aligned} &E[X|Y=1] = 2, \\ &E[X|Y=2] = 3 + E[X], \\ &E[X|Y=3] = 5 + E[X], \end{aligned} \tag{3.6}$$

To understand why this is correct consider, for instance, $E[X|Y=2]$, and reason as follows. If the miner chooses the second door, then he spends three hours in the tunnel and then returns to the mine. But once he returns to the mine the problem is as before, and hence his expected additional time until safety is just $E[X]$. Hence $E[X|Y=2] = 3 + E[X]$. The argument behind the other equalities in Eq. (3.6) is similar. Hence,

$$E[X] = \tfrac{1}{3}(2 + 3 + E[X] + 5 + E[X]) \quad \text{or} \quad E[X] = 10 \qquad \blacksquare$$

Example 3.13 (Multinomial Covariances)**.** Consider n independent trials, each of which results in one of the outcomes $1, \ldots, r$, with respective probabilities $p_1, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$. If we let N_i denote the number of trials that result in outcome i, then $(N_1, \ldots, N_r)$ is said to have a *multinomial distribution*. For $i \neq j$, let us compute

$$\text{Cov}(N_i, N_j) = E[N_i N_j] - E[N_i]E[N_j]$$

Because each trial independently results in outcome i with probability p_i, it follows that N_i is binomial with parameters (n, p_i), giving that $E[N_i]E[N_j] = n^2 p_i p_j$. To compute $E[N_i N_j]$, condition on N_i to obtain

$$\begin{aligned} E[N_i N_j] &= \sum_{k=0}^{n} E[N_i N_j | N_i = k] P(N_i = k) \\ &= \sum_{k=0}^{n} k E[N_j | N_i = k] P(N_i = k) \end{aligned}$$

Now, given that k of the n trials result in outcome i, each of the other $n - k$ trials independently results in outcome j with probability

$$P(j|\text{not } i) = \frac{p_j}{1 - p_i}$$

thus showing that the conditional distribution of N_j, given that $N_i = k$, is binomial with parameters $(n-k, \frac{p_j}{1-p_i})$. Using this yields

$$\begin{aligned} E[N_i N_j] &= \sum_{k=0}^{n} k(n-k)\frac{p_j}{1-p_i} P(N_i = k) \\ &= \frac{p_j}{1-p_i}\left(n\sum_{k=0}^{n} kP(N_i = k) - \sum_{k=0}^{n} k^2 P(N_i = k)\right) \\ &= \frac{p_j}{1-p_i}(nE[N_i] - E[N_i^2]) \end{aligned}$$

Because N_i is binomial with parameters (n, p_i)

$$E[N_i^2] = \mathrm{Var}(N_i) + (E[N_i])^2 = np_i(1-p_i) + (np_i)^2$$

Hence,

$$\begin{aligned} E[N_i N_j] &= \frac{p_j}{1-p_i}[n^2 p_i - np_i(1-p_i) - n^2 p_i^2] \\ &= \frac{np_i p_j}{1-p_i}[n - np_i - (1-p_i)] \\ &= n(n-1)p_i p_j \end{aligned}$$

which yields the result

$$\mathrm{Cov}(N_i, N_j) = n(n-1)p_i p_j - n^2 p_i p_j = -np_i p_j$$ ■

Example 3.14 (The Matching Rounds Problem)**.** Suppose in Example 2.30 that those choosing their own hats depart, while the others (those without a match) put their selected hats in the center of the room, mix them up, and then reselect. Also, suppose that this process continues until each individual has his own hat.

(a) Find $E[R_n]$ where R_n is the number of rounds that are necessary when n individuals are initially present.

(b) Find $E[S_n]$ where S_n is the total number of selections made by the n individuals, $n \geq 2$.

(c) Find the expected number of false selections made by one of the n people, $n \geq 2$.

Solution: (a) It follows from the results of Example 2.30 that no matter how many people remain there will, on average, be one match per round. Hence, one might suggest that $E[R_n] = n$. This turns out to be true, and an induction proof will now be given. Because it is obvious that $E[R_1] = 1$, assume that $E[R_k] = k$ for $k = 1, \ldots, n-1$. To compute $E[R_n]$, start by conditioning on X_n, the number of matches that occur in the first round. This gives

$$E[R_n] = \sum_{i=0}^{n} E[R_n | X_n = i]P\{X_n = i\}$$

Now, given a total of i matches in the initial round, the number of rounds needed will equal 1 plus the number of rounds that are required when $n - i$ persons are to be matched with their hats. Therefore,

$$\begin{aligned}
E[R_n] &= \sum_{i=0}^{n} (1 + E[R_{n-i}]) P\{X_n = i\} \\
&= 1 + E[R_n] P\{X_n = 0\} + \sum_{i=1}^{n} E[R_{n-i}] P\{X_n = i\} \\
&= 1 + E[R_n] P\{X_n = 0\} + \sum_{i=1}^{n} (n-i) P\{X_n = i\} \\
&\qquad \text{by the induction hypothesis} \\
&= 1 + E[R_n] P\{X_n = 0\} + n(1 - P\{X_n = 0\}) - E[X_n] \\
&= E[R_n] P\{X_n = 0\} + n(1 - P\{X_n = 0\})
\end{aligned}$$

where the final equality used the result, established in Example 2.30, that $E[X_n] = 1$. Since the preceding equation implies that $E[R_n] = n$, the result is proven.

(b) For $n \geq 2$, conditioning on X_n, the number of matches in round 1, gives

$$\begin{aligned}
E[S_n] &= \sum_{i=0}^{n} E[S_n | X_n = i] P\{X_n = i\} \\
&= \sum_{i=0}^{n} (n + E[S_{n-i}]) P\{X_n = i\} \\
&= n + \sum_{i=0}^{n} E[S_{n-i}] P\{X_n = i\}
\end{aligned}$$

where $E[S_0] = 0$. To solve the preceding equation, rewrite it as

$$E[S_n] = n + E[S_{n-X_n}]$$

Now, if there were *exactly* one match in each round, then it would take a total of $1 + 2 + \cdots + n = n(n+1)/2$ selections. Thus, let us try a solution of the form $E[S_n] = an + bn^2$. For the preceding equation to be satisfied by a solution of this type, for $n \geq 2$, we need

$$an + bn^2 = n + E\left[a(n - X_n) + b(n - X_n)^2\right]$$

or, equivalently,

$$an + bn^2 = n + a(n - E[X_n]) + b\left(n^2 - 2nE[X_n] + E\left[X_n^2\right]\right)$$

Now, using the results of Example 2.30 and Exercise 72 of Chapter 2 that $E[X_n] = \text{Var}(X_n) = 1$, the preceding will be satisfied if

$$an + bn^2 = n + an - a + bn^2 - 2nb + 2b$$

and this will be valid provided that $b = 1/2, a = 1$. That is,

$$E[S_n] = n + n^2/2$$

satisfies the recursive equation for $E[S_n]$.

The formal proof that $E[S_n] = n + n^2/2, n \geq 2$, is obtained by induction on n. It is true when $n = 2$ (since, in this case, the number of selections is twice the number of rounds and the number of rounds is a geometric random variable with parameter $p = 1/2$). Now, the recursion gives

$$E[S_n] = n + E[S_n]P\{X_n = 0\} + \sum_{i=1}^{n} E[S_{n-i}]P\{X_n = i\}$$

Hence, upon assuming that $E[S_0] = E[S_1] = 0$, $E[S_k] = k + k^2/2$, for $k = 2, \ldots, n-1$ and using that $P\{X_n = n-1\} = 0$, we see that

$$\begin{aligned} E[S_n] &= n + E[S_n]P\{X_n = 0\} + \sum_{i=1}^{n}[n - i + (n-i)^2/2]P\{X_n = i\} \\ &= n + E[S_n]P\{X_n = 0\} + (n + n^2/2)(1 - P\{X_n = 0\}) \\ &\quad - (n+1)E[X_n] + E[X_n^2]/2 \end{aligned}$$

Substituting the identities $E[X_n] = 1$, $E[X_n^2] = 2$ in the preceding shows that

$$E[S_n] = n + n^2/2$$

and the induction proof is complete.
(c) If we let C_j denote the number of hats chosen by person j, $j = 1, \ldots, n$ then

$$\sum_{j=1}^{n} C_j = S_n$$

Taking expectations, and using the fact that each C_j has the same mean, yields the result

$$E[C_j] = E[S_n]/n = 1 + n/2$$

Hence, the expected number of false selections by person j is

$$E[C_j - 1] = n/2.$$

■

Example 3.15. Independent trials, each of which is a success with probability p, are performed until there are k consecutive successes. What is the mean number of necessary trials?

Solution: Let N_k denote the number of necessary trials to obtain k consecutive successes, and let $M_k = E[N_k]$. We will determine M_k by deriving and then solving a recursive equation that it satisfies. To begin, write

$$N_k = N_{k-1} + A_{k-1,k}$$

where N_{k-1} is the number of trials needed for $k-1$ consecutive successes, and $A_{k-1,k}$ is the number of additional trials needed to go from having $k-1$ successes in a row to having k in a row. Taking expectations gives that,

$$M_k = M_{k-1} + E[A_{k-1,k}]$$

To determine $E[A_{k-1,k}]$, condition on the next trial after there have been $k-1$ successes in a row. If it is a success then that gives k in a row and no additional trials after that are needed; if it is a failure then at that point we are starting all over and so the expected additional number from then on would be $E[N_k]$. Thus,

$$E[A_{k-1,k}] = 1 \cdot p + (1 + M_k)(1 - p) = 1 + (1 - p)M_k$$

giving that

$$M_k = M_{k-1} + 1 + (1 - p)M_k$$

or

$$M_k = \frac{1}{p} + \frac{M_{k-1}}{p}$$

Since N_1, the time of the first success, is geometric with parameter p, we see that

$$M_1 = \frac{1}{p}$$

and, recursively

$$M_2 = \frac{1}{p} + \frac{1}{p^2},$$
$$M_3 = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3}$$

and, in general,

$$M_k = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k}$$

■

Example 3.16. Consider a gambler who in each game is equally likely to either win or lose 1, independent of the results from earlier games. Starting with a fortune of i, find m_i, the mean number of games until the gambler's fortune is either 0 or n, where $0 \le i \le n$.

Solution: Let N denote the number of games until the gambler's fortune is either 0 or n, and let S_i denote the event that the gambler starts with a fortune of i. To obtain an expression for $m_i = E[N|S_i]$, condition on the result of the first game. With W being the event that the first game is a win, and L that it is a loss, this yields that for $i = 1, \ldots, n-1$

$$\begin{aligned} m_i &= E[N|S_i] \\ &= E[N|S_iW]\,P(W|S_i) + E[N|S_iL]\,P(L|S_i) \\ &= (1+m_{i+1})\frac{1}{2} + (1+m_{i-1})\frac{1}{2} \\ &= 1 + \frac{1}{2}m_{i-1} + \frac{1}{2}m_{i+1}, \quad i = 1, \ldots, n-1 \end{aligned}$$

Using that $m_0 = 0$, the preceding can be rewritten as

$$\begin{aligned} m_2 &= 2(m_1 - 1) \\ m_{i+1} &= 2(m_i - 1) - m_{i-1}, \quad i = 2, \ldots, n-1 \end{aligned}$$

Letting $i = 2$ in the preceding yields that

$$\begin{aligned} m_3 &= 2m_2 - 2 - m_1 \\ &= 4m_1 - 4 - 2 - m_1 \\ &= 3(m_1 - 2) \end{aligned}$$

A check of m_4 shows a similar result, and indeed it is easily shown by induction that

$$m_i = i(m_1 - i + 1), \quad i = 2, \ldots, n$$

Using that $m_n = 0$, the preceding yields that $0 = n(m_1 - n + 1)$. Thus, $m_1 = n - 1$, and

$$m_i = i(n-i), \qquad i = 1, \ldots, n-1$$ ■

Example 3.17 (Analyzing the Quick-Sort Algorithm)**.** Suppose we are given a set of n distinct values—$x_1, \ldots, x_n$—and we desire to put these values in increasing order or, as it is commonly called, to *sort* them. An efficient procedure for accomplishing this is the quick-sort algorithm, which is defined recursively as follows: When $n = 2$ the algorithm compares the two values and puts them in the appropriate order. When $n > 2$ it starts by choosing at random one of the n values—say, x_i—and then compares each of the other $n - 1$ values with x_i, noting which are smaller and which are larger

than x_i. Letting S_i denote the set of elements smaller than x_i, and $\bar{S}_i$ the set of elements greater than x_i, the algorithm now sorts the set S_i and the set $\bar{S}_i$. The final ordering, therefore, consists of the ordered set of the elements in S_i, then x_i, and then the ordered set of the elements in $\bar{S}_i$. For instance, suppose that the set of elements is 10, 5, 8, 2, 1, 4, 7. We start by choosing one of these values at random (that is, each of the 7 values has probability of $\frac{1}{7}$ of being chosen). Suppose, for instance, that the value 4 is chosen. We then compare 4 with each of the other six values to obtain

$$\{2, 1\}, 4, \{10, 5, 8, 7\}$$

We now sort the set $\{2, 1\}$ to obtain

$$1, 2, 4, \{10, 5, 8, 7\}$$

Next we choose a value at random from $\{10, 5, 8, 7\}$—say 7 is chosen—and compare each of the other three values with 7 to obtain

$$1, 2, 4, 5, 7, \{10, 8\}$$

Finally, we sort $\{10, 8\}$ to end up with

$$1, 2, 4, 5, 7, 8, 10$$

One measure of the effectiveness of this algorithm is the expected number of comparisons that it makes. Let us denote by M_n the expected number of comparisons needed by the quick-sort algorithm to sort a set of n distinct values. To obtain a recursion for M_n we condition on the rank of the initial value selected to obtain

$$M_n = \sum_{j=1}^{n} E[\text{number of comparisons}|\text{value selected is } j\text{th smallest}]\frac{1}{n}$$

Now, if the initial value selected is the jth smallest, then the set of values smaller than it is of size $j-1$, and the set of values greater than it is of size $n-j$. Hence, as $n-1$ comparisons with the initial value chosen must be made, we see that

$$\begin{aligned} M_n &= \sum_{j=1}^{n}(n-1+M_{j-1}+M_{n-j})\frac{1}{n} \\ &= n-1+\frac{2}{n}\sum_{k=1}^{n-1}M_k \quad (\text{since } M_0=0) \end{aligned}$$

or, equivalently,

$$nM_n = n(n-1)+2\sum_{k=1}^{n-1}M_k$$

To solve the preceding, note that upon replacing n by $n+1$ we obtain

$$(n+1)M_{n+1} = (n+1)n + 2\sum_{k=1}^{n} M_k$$

Hence, upon subtraction,

$$(n+1)M_{n+1} - nM_n = 2n + 2M_n$$

or

$$(n+1)M_{n+1} = (n+2)M_n + 2n$$

Therefore,

$$\frac{M_{n+1}}{n+2} = \frac{2n}{(n+1)(n+2)} + \frac{M_n}{n+1}$$

Iterating this gives

$$\begin{aligned}\frac{M_{n+1}}{n+2} &= \frac{2n}{(n+1)(n+2)} + \frac{2(n-1)}{n(n+1)} + \frac{M_{n-1}}{n} \\ &= \cdots \\ &= 2\sum_{k=0}^{n-1} \frac{n-k}{(n+1-k)(n+2-k)} \quad \text{since } M_1 = 0\end{aligned}$$

Hence,

$$\begin{aligned}M_{n+1} &= 2(n+2)\sum_{k=0}^{n-1} \frac{n-k}{(n+1-k)(n+2-k)} \\ &= 2(n+2)\sum_{i=1}^{n} \frac{i}{(i+1)(i+2)}, \quad n \geq 1\end{aligned}$$

Using the identity $i/(i+1)(i+2) = 2/(i+2) - 1/(i+1)$, we can approximate M_{n+1} for large n as follows:

$$\begin{aligned}M_{n+1} &= 2(n+2)\left[\sum_{i=1}^{n} \frac{2}{i+2} - \sum_{i=1}^{n} \frac{1}{i+1}\right] \\ &\sim 2(n+2)\left[\int_3^{n+2} \frac{2}{x}\,dx - \int_2^{n+1} \frac{1}{x}\,dx\right] \\ &= 2(n+2)[2\log(n+2) - \log(n+1) + \log 2 - 2\log 3] \\ &= 2(n+2)\left[\log(n+2) + \log\frac{n+2}{n+1} + \log 2 - 2\log 3\right] \\ &\sim 2(n+2)\log(n+2)\end{aligned}$$

■

Although we usually employ the conditional expectation identity to more easily enable us to compute an unconditional expectation, in our next example we show how it can sometimes be used to obtain the conditional expectation.

Example 3.18. In the match problem of Example 2.30 involving $n, n > 1$, individuals, find the conditional expected number of matches given that the first person did not have a match.

Solution: Let X denote the number of matches, and let X_1 equal 1 if the first person has a match and 0 otherwise. Then,

$$\begin{aligned} E[X] &= E[X|X_1 = 0]P\{X_1 = 0\} + E[X|X_1 = 1]P\{X_1 = 1\} \\ &= E[X|X_1 = 0]\frac{n-1}{n} + E[X|X_1 = 1]\frac{1}{n} \end{aligned}$$

But, from Example 2.30

$$E[X] = 1$$

Moreover, given that the first person has a match, the expected number of matches is equal to 1 plus the expected number of matches when $n - 1$ people select among their own $n - 1$ hats, showing that

$$E[X|X_1 = 1] = 2$$

Therefore, we obtain the result

$$E[X|X_1 = 0] = \frac{n-2}{n-1}$$ ■

3.4.1 Computing Variances by Conditioning

Conditional expectations can also be used to compute the variance of a random variable. Specifically, we can use

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

and then use conditioning to obtain both $E[X]$ and $E[X^2]$. We illustrate this technique by determining the variance of a geometric random variable.

Example 3.19 (Variance of the Geometric Random Variable)**.** Independent trials, each resulting in a success with probability p, are performed in sequence. Let N be the trial number of the first success. Find $\text{Var}(N)$.

Solution: Let $Y = 1$ if the first trial results in a success, and $Y = 0$ otherwise.

$$\text{Var}(N) = E[N^2] - (E[N])^2$$

To calculate $E[N^2]$ and $E[N]$ we condition on Y. For instance,

$$E[N^2] = E\big[E[N^2|Y]\big]$$

However,

$$E[N^2|Y=1] = 1,$$
$$E[N^2|Y=0] = E[(1+N)^2]$$

These two equations are true since if the first trial results in a success, then clearly $N = 1$ and so $N^2 = 1$. On the other hand, if the first trial results in a failure, then the total number of trials necessary for the first success will equal one (the first trial that results in failure) plus the necessary number of additional trials. Since this latter quantity has the same distribution as N, we get that $E[N^2|Y=0] = E[(1+N)^2]$. Hence, we see that

$$\begin{aligned} E[N^2] &= E[N^2|Y=1]P\{Y=1\} + E[N^2|Y=0]P\{Y=0\} \\ &= p + E[(1+N)^2](1-p) \\ &= 1 + (1-p)E[2N + N^2] \end{aligned}$$

Since, as was shown in Example 3.11, $E[N] = 1/p$, this yields

$$E[N^2] = 1 + \frac{2(1-p)}{p} + (1-p)E[N^2]$$

or

$$E[N^2] = \frac{2-p}{p^2}$$

Therefore,

$$\begin{aligned} \mathrm{Var}(N) &= E[N^2] - (E[N])^2 \\ &= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 \\ &= \frac{1-p}{p^2} \end{aligned}$$

■

Another way to use conditioning to obtain the variance of a random variable is to apply the conditional variance formula. The conditional variance of X given that $Y = y$ is defined by

$$\mathrm{Var}(X|Y=y) = E\big[(X - E[X|Y=y])^2|Y=y\big]$$

That is, the conditional variance is defined in exactly the same manner as the ordinary variance with the exception that all probabilities are determined conditional on the event that $Y = y$. Expanding the right side of the preceding and taking expectation term by term yields

$$\mathrm{Var}(X|Y=y) = E[X^2|Y=y] - (E[X|Y=y])^2$$

Letting $\text{Var}(X|Y)$ denote that function of Y whose value when $Y = y$ is $\text{Var}(X|Y = y)$, we have the following result.

Proposition 3.1 (The Conditional Variance Formula)**.**

$$\text{Var}(X) = E\big[\text{Var}(X|Y)\big] + \text{Var}\big(E[X|Y]\big) \tag{3.7}$$

Proof.

$$\begin{aligned} E\big[\text{Var}(X|Y)\big] &= E[E[X^2|Y] - (E[X|Y])^2] \\ &= E\big[E[X^2|Y]\big] - E\big[(E[X|Y])^2\big] \\ &= E[X^2] - E\big[(E[X|Y])^2\big] \end{aligned}$$

and

$$\begin{aligned} \text{Var}(E[X|Y]) &= E\big[(E[X|Y])^2\big] - \big(E\big[E[X|Y]\big]\big)^2 \\ &= E\big[(E[X|Y])^2\big] - (E[X])^2 \end{aligned}$$

Therefore,

$$E\big[\text{Var}(X|Y)\big] + \text{Var}\big(E[X|Y]\big) = E[X^2] - (E[X])^2$$

which completes the proof. ■

Example 3.20 (The Variance of a Compound Random Variable)**.** Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with distribution F having mean μ and variance σ^2, and assume that they are independent of the nonnegative integer valued random variable N. As noted in Example 3.10, where its expected value was determined, the random variable $S = \sum_{i=1}^{N} X_i$ is called a compound random variable. Find its variance.

Solution: Whereas we could obtain $E[S^2]$ by conditioning on N, let us instead use the conditional variance formula. Now,

$$\begin{aligned} \text{Var}(S|N = n) &= \text{Var}\left(\sum_{i=1}^{N} X_i | N = n\right) \\ &= \text{Var}\left(\sum_{i=1}^{n} X_i | N = n\right) \\ &= \text{Var}\left(\sum_{i=1}^{n} X_i\right) \\ &= n\sigma^2 \end{aligned}$$

By the same reasoning,

$$E[S|N = n] = n\mu$$

Therefore,

$$\text{Var}(S|N) = N\sigma^2, \quad E[S|N] = N\mu$$

and the conditional variance formula gives

$$\text{Var}(S) = E[N\sigma^2] + \text{Var}(N\mu) = \sigma^2 E[N] + \mu^2 \text{Var}(N)$$

If N is a Poisson random variable, then $S = \sum_{i=1}^{N} X_i$ is called a *compound Poisson* random variable. Because the variance of a Poisson random variable is equal to its mean, it follows that for a compound Poisson random variable having $E[N] = \lambda$

$$\text{Var}(S) = \lambda\sigma^2 + \lambda\mu^2 = \lambda E[X^2]$$

where X has the distribution F. ■

Example 3.21 (The Variance in the Matching Rounds Problem)**.** Consider the matching rounds problem of Example 3.14, and let $V_n = \text{Var}(R_n)$ denote the variance of the number of rounds needed when there are initially n people. Using the conditional variance formula, we will show that

$$V_n = n, \quad n \geq 2$$

The proof of the preceding is by induction on n. To begin, note that when $n = 2$ the number of rounds needed is geometric with parameter $p = 1/2$ and so

$$V_2 = \frac{1-p}{p^2} = 2$$

So assume the induction hypothesis that

$$V_j = j, \quad 2 \leq j < n$$

and now consider the case when there are n individuals. If X is the number of matches in the first round then, conditional on X, the number of rounds R_n is distributed as 1 plus the number of rounds needed when there are initially $n - X$ individuals. Consequently,

$$\begin{aligned} E[R_n|X] &= 1 + E[R_{n-X}] \\ &= 1 + n - X \qquad \text{by Example 3.14} \end{aligned}$$

Also, with $V_0 = 0$,

$$\text{Var}(R_n|X) = \text{Var}(R_{n-X}) = V_{n-X}$$

Hence, by the conditional variance formula

$$V_n = E[\text{Var}(R_n|X)] + \text{Var}(E[R_n|X])$$

$$
\begin{aligned}
&= E[V_{n-X}] + \mathrm{Var}(X) \\
&= \sum_{j=0}^{n} V_{n-j} P(X = j) + \mathrm{Var}(X) \\
&= V_n P(X = 0) + \sum_{j=1}^{n} V_{n-j} P(X = j) + \mathrm{Var}(X)
\end{aligned}
$$

Because $P(X = n-1) = 0$, it follows from the preceding and the induction hypothesis that

$$
\begin{aligned}
V_n &= V_n P(X = 0) + \sum_{j=1}^{n} (n - j) P(X = j) + \mathrm{Var}(X) \\
&= V_n P(X = 0) + n(1 - P(X = 0)) - E[X] + \mathrm{Var}(X)
\end{aligned}
$$

As it is easily shown (see Example 2.30 and Exercise 64 of Chapter 2) that $E[X] = \mathrm{Var}(X) = 1$, the preceding gives

$$
V_n = V_n P(X = 0) + n(1 - P(X = 0))
$$

thus proving the result. ■

3.5 Computing Probabilities by Conditioning

Not only can we obtain expectations by first conditioning on an appropriate random variable, but we may also use this approach to compute probabilities. To see this, let E denote an arbitrary event and define the indicator random variable X by

$$
X = \begin{cases} 1, & \text{if } E \text{ occurs} \\ 0, & \text{if } E \text{ does not occur} \end{cases}
$$

It follows from the definition of X that

$$
\begin{aligned}
E[X] &= P(E), \\
E[X|Y = y] &= P(E|Y = y), \quad \text{for any random variable } Y
\end{aligned}
$$

Therefore, from Eqs. (3.2a) and (3.2b) we obtain

$$
\begin{aligned}
P(E) &= \sum_{y} P(E|Y = y) P(Y = y), \quad \text{if } Y \text{ is discrete} \\
&= \int_{-\infty}^{\infty} P(E|Y = y) f_Y(y)\, dy, \quad \text{if } Y \text{ is continuous}
\end{aligned}
$$

Example 3.22. Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute $P\{X < Y\}$.

Solution: Conditioning on the value of Y yields

$$\begin{aligned} P\{X < Y\} &= \int_{-\infty}^{\infty} P\{X < Y | Y = y\} f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} P\{X < y | Y = y\} f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} P\{X < y\} f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y)\, dy \end{aligned}$$

where

$$F_X(y) = \int_{-\infty}^{y} f_X(x)\, dx$$ ■

Example 3.23. An insurance company supposes that the number of accidents that each of its policyholders will have in a year is Poisson distributed, with the mean of the Poisson depending on the policyholder. If the Poisson mean of a randomly chosen policyholder has a gamma distribution with density function

$$g(\lambda) = \lambda e^{-\lambda}, \quad \lambda \geq 0$$

what is the probability that a randomly chosen policyholder has exactly n accidents next year?

Solution: Let X denote the number of accidents that a randomly chosen policyholder has next year. Letting Y be the Poisson mean number of accidents for this policyholder, then conditioning on Y yields

$$\begin{aligned} P\{X = n\} &= \int_0^{\infty} P\{X = n | Y = \lambda\} g(\lambda)\, d\lambda \\ &= \int_0^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \lambda e^{-\lambda}\, d\lambda \\ &= \frac{1}{n!} \int_0^{\infty} \lambda^{n+1} e^{-2\lambda}\, d\lambda \end{aligned}$$

However, because

$$h(\lambda) = \frac{2e^{-2\lambda}(2\lambda)^{n+1}}{(n+1)!}, \quad \lambda > 0$$

is the density function of a gamma $(n+2, 2)$ random variable, its integral is 1. Therefore,

$$1 = \int_0^{\infty} \frac{2e^{-2\lambda}(2\lambda)^{n+1}}{(n+1)!}\, d\lambda = \frac{2^{n+2}}{(n+1)!} \int_0^{\infty} \lambda^{n+1} e^{-2\lambda}\, d\lambda$$

showing that

$$P\{X = n\} = \frac{n+1}{2^{n+2}} \qquad \blacksquare$$

Example 3.24. Suppose that the number of people who visit a yoga studio each day is a Poisson random variable with mean λ. Suppose further that each person who visits is, independently, female with probability p or male with probability $1 - p$. Find the joint probability that exactly n women and m men visit the academy today.

Solution: Let N_1 denote the number of women and N_2 the number of men who visit the academy today. Also, let $N = N_1 + N_2$ be the total number of people who visit. Conditioning on N gives

$$P\{N_1 = n, N_2 = m\} = \sum_{i=0}^{\infty} P\{N_1 = n, N_2 = m | N = i\} P\{N = i\}$$

Because $P\{N_1 = n, N_2 = m | N = i\} = 0$ when $i \neq n + m$, the preceding equation yields

$$P\{N_1 = n, N_2 = m\} = P\{N_1 = n, N_2 = m | N = n + m\} e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!}$$

Given that $n + m$ people visit it follows, because each of these $n + m$ is independently a woman with probability p, that the conditional probability that n of them are women (and m are men) is just the binomial probability of n successes in $n+m$ trials. Therefore,

$$\begin{aligned} P\{N_1 = n, N_2 = m\} &= \binom{n+m}{n} p^n (1-p)^m e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} \\ &= \frac{(n+m)!}{n!m!} p^n (1-p)^m e^{-\lambda p} e^{-\lambda(1-p)} \frac{\lambda^n \lambda^m}{(n+m)!} \\ &= e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} \end{aligned}$$

Because the preceding joint probability mass function factors into two products, one of which depends only on n and the other only on m, it follows that N_1 and N_2 are independent. Moreover, because

$$\begin{aligned} P\{N_1 = n\} &= \sum_{m=0}^{\infty} P\{N_1 = n, N_2 = m\} \\ &= e^{-\lambda p} \frac{(\lambda p)^n}{n!} \sum_{m=0}^{\infty} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} = e^{-\lambda p} \frac{(\lambda p)^n}{n!} \end{aligned}$$

and, similarly,

$$P\{N_2 = m\} = e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!}$$

we can conclude that N_1 and N_2 are independent Poisson random variables with respective means λp and $\lambda(1-p)$. Therefore, this example establishes the important result that when each of a Poisson number of events is independently classified either as being type 1 with probability p or type 2 with probability $1-p$, then the numbers of type 1 and type 2 events are independent Poisson random variables. ■

The result of Example 3.24 generalizes to the case where each of a Poisson distributed number of events, N, with mean λ is independently classified as being one of k types, with the probability that it is type i being $p_i, i = 1, \ldots, k, \sum_{i=1}^{k} p_i = 1$. If N_i is the number that are classified as type i, then $N_1, \ldots, N_k$ are independent Poisson random variables with respective means $\lambda p_1, \ldots, \lambda p_k$. This follows, since for $n = \sum_{i=1}^{k} n_i$

$$\begin{aligned} P(N_1 = n_1, \ldots, N_k = n_k) &= P(N_1 = n_1, \ldots, N_k = n_k | N = n) P(N = n) \\ &= \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k} e^{-\lambda} \lambda^n / n! \\ &= \prod_{i=1}^{k} e^{-\lambda p_i} (\lambda p_i)^{n_i} / n_i! \end{aligned}$$

where the second equality used that, given a total of n events, the numbers of each type has a multinomial distribution with parameters $(n, p_1, \ldots, p_k)$.

Example 3.25 (The Distribution of the Sum of Independent Bernoulli Random Variables)**.** Let $X_1, \ldots, X_n$ be independent Bernoulli random variables, with X_i having parameter $p_i, i = 1, \ldots, n$. That is, $P\{X_i = 1\} = p_i, P\{X_i = 0\} = q_i = 1 - p_i$. Suppose we want to compute the probability mass function of their sum, $X_1 + \cdots + X_n$. To do so, we will recursively obtain the probability mass function of $X_1 + \cdots + X_k$, first for $k = 1$, then $k = 2$, and on up to $k = n$. To begin, let

$$P_k(j) = P\{X_1 + \cdots + X_k = j\}$$

and note that

$$P_k(k) = \prod_{i=1}^{k} p_i, \qquad P_k(0) = \prod_{i=1}^{k} q_i$$

For $0 < j < k$, conditioning on X_k yields the recursion

$$\begin{aligned} P_k(j) &= P\{X_1 + \cdots + X_k = j | X_k = 1\} p_k + P\{X_1 + \cdots + X_k = j | X_k = 0\} q_k \\ &= P\{X_1 + \cdots + X_{k-1} = j - 1 | X_k = 1\} p_k \\ &\quad + P\{X_1 + \cdots + X_{k-1} = j | X_k = 0\} q_k \\ &= P\{X_1 + \cdots + X_{k-1} = j - 1\} p_k + P\{X_1 + \cdots + X_{k-1} = j\} q_k \\ &= p_k P_{k-1}(j-1) + q_k P_{k-1}(j) \end{aligned}$$

Starting with $P_1(1) = p_1$, $P_1(0) = q_1$, the preceding equations can be recursively solved to obtain the functions $P_2(j)$, $P_3(j)$, up to $P_n(j)$. ■

Example 3.26 (The Best Prize Problem)**.** Suppose that we are to be presented with n distinct prizes in sequence. After being presented with a prize we must immediately decide whether to accept it or reject it and consider the next prize. The only information we are given when deciding whether to accept a prize is the relative rank of that prize compared to ones already seen. That is, for instance, when the fifth prize is presented we learn how it compares with the first four prizes already seen. Suppose that once a prize is rejected it is lost, and that our objective is to maximize the probability of obtaining the best prize. Assuming that all $n!$ orderings of the prizes are equally likely, how well can we do?

Solution: Rather surprisingly, we can do quite well. To see this, fix a value k, $0 \leq k < n$, and consider the strategy that rejects the first k prizes and then accepts the first one that is better than all of those first k. Let P_k (best) denote the probability that the best prize is selected when this strategy is employed. To compute this probability, condition on X, the position of the best prize. This gives

$$\begin{aligned} P_k(\text{best}) &= \sum_{i=1}^{n} P_k(\text{best}|X=i)P(X=i) \\ &= \frac{1}{n}\sum_{i=1}^{n} P_k(\text{best}|X=i) \end{aligned}$$

Now, if the overall best prize is among the first k, then no prize is ever selected under the strategy considered. On the other hand, if the best prize is in position i, where $i > k$, then the best prize will be selected if the best of the first k prizes is also the best of the first $i-1$ prizes (for then none of the prizes in positions $k+1, k+2, \ldots, i-1$ would be selected). Hence, we see that

$$\begin{aligned} P_k(\text{best}|X=i) &= 0, \quad \text{if } i \leq k \\ P_k(\text{best}|X=i) &= P\{\text{best of first } i-1 \text{ is among the first } k\} \\ &= k/(i-1), \quad \text{if } i > k \end{aligned}$$

From the preceding, we obtain

$$\begin{aligned} P_k(\text{best}) &= \frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1} \\ &\approx \frac{k}{n}\int_{k}^{n-1}\frac{1}{x}\,dx \\ &= \frac{k}{n}\log\left(\frac{n-1}{k}\right) \\ &\approx \frac{k}{n}\log\left(\frac{n}{k}\right) \end{aligned}$$

Now, if we consider the function

$$g(x) = \frac{x}{n} \log\left(\frac{n}{x}\right)$$

then

$$g'(x) = \frac{1}{n} \log\left(\frac{n}{x}\right) - \frac{1}{n}$$

and so

$$g'(x) = 0 \Rightarrow \log(n/x) = 1 \Rightarrow x = n/e$$

Thus, since $P_k(\text{best}) \approx g(k)$, we see that the best strategy of the type considered is to let the first n/e prizes go by and then accept the first one to appear that is better than all of those. In addition, since $g(n/e) = 1/e$, the probability that this strategy selects the best prize is approximately $1/e \approx 0.36788$.

Remark. Most students are quite surprised by the size of the probability of obtaining the best prize, thinking that this probability would be close to 0 when n is large. However, even without going through the calculations, a little thought reveals that the probability of obtaining the best prize can be made to be reasonably large. Consider the strategy of letting half of the prizes go by, and then selecting the first one to appear that is better than all of those. The probability that a prize is actually selected is the probability that the overall best is among the second half and this is $1/2$. In addition, given that a prize is selected, at the time of selection that prize would have been the best of more than $n/2$ prizes to have appeared, and would thus have probability of at least $1/2$ of being the overall best. Hence, the strategy of letting the first half of all prizes go by and then accepting the first one that is better than all of those prizes results in a probability greater than $1/4$ of obtaining the best prize. ■

Example 3.27. At a party n men take off their hats. The hats are then mixed up and each man randomly selects one. We say that a match occurs if a man selects his own hat. What is the probability of no matches? What is the probability of exactly k matches?

Solution: Let E denote the event that no matches occur, and to make explicit the dependence on n, write $P_n = P(E)$. We start by conditioning on whether or not the first man selects his own hat—call these events M and M^c. Then

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$$

Clearly, $P(E|M) = 0$, and so

$$P_n = P(E|M^c)\frac{n-1}{n} \tag{3.8}$$

Now, $P(E|M^c)$ is the probability of no matches when $n-1$ men select from a set of $n-1$ hats that does not contain the hat of one of these men. This can happen

in either of two mutually exclusive ways. Either there are no matches and the extra man does not select the extra hat (this being the hat of the man that chose first), or there are no matches and the extra man does select the extra hat. The probability of the first of these events is just P_{n-1}, which is seen by regarding the extra hat as "belonging" to the extra man. Because the second event has probability $[1/(n-1)]P_{n-2}$, we have

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}$$

and thus, from Eq. (3.8),

$$P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2}$$

or, equivalently,

$$P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2}) \tag{3.9}$$

However, because P_n is the probability of no matches when n men select among their own hats, we have

$$P_1 = 0, \quad P_2 = \tfrac{1}{2}$$

and so, from Eq. (3.9),

$$P_3 - P_2 = -\frac{(P_2 - P_1)}{3} = -\frac{1}{3!} \quad \text{or} \quad P_3 = \frac{1}{2!} - \frac{1}{3!},$$
$$P_4 - P_3 = -\frac{(P_3 - P_2)}{4} = \frac{1}{4!} \quad \text{or} \quad P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

and, in general, we see that

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}$$

To obtain the probability of exactly k matches, we consider any fixed group of k men. The probability that they, and only they, select their own hats is

$$\frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k} = \frac{(n-k)!}{n!}P_{n-k}$$

where P_{n-k} is the conditional probability that the other $n-k$ men, selecting among their own hats, have no matches. Because there are $\binom{n}{k}$ choices of a set of k men, the desired probability of exactly k matches is

$$\frac{P_{n-k}}{k!} = \frac{\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-k}}{(n-k)!}}{k!}$$

which, for n large, is approximately equal to $e^{-1}/k!$.

Remark. The recursive equation, Eq. (3.9), could also have been obtained by using the concept of a cycle, where we say that the sequence of distinct individuals $i_1, i_2, \ldots, i_k$ constitutes a *cycle* if i_1 chooses i_2's hat, i_2 chooses i_3's hat, $\ldots$, i_{k-1} chooses i_k's hat, and i_k chooses i_1's hat. Note that every individual is part of a cycle, and that a cycle of size $k = 1$ occurs when someone chooses his or her own hat. With E being, as before, the event that no matches occur, it follows upon conditioning on the size of the cycle containing a specified person, say person 1, that

$$P_n = P(E) = \sum_{k=1}^{n} P(E|C = k)P(C = k) \tag{3.10}$$

where C is the size of the cycle that contains person 1. Now, call person 1 the first person, and note that $C = k$ if the first person does not choose 1's hat; the person whose hat was chosen by the first person—call this person the second person—does not choose 1's hat; the person whose hat was chosen by the second person—call this person the third person—does not choose 1's hat; $\ldots$, the person whose hat was chosen by the $(k-1)$st person does choose 1's hat. Consequently,

$$P(C = k) = \frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\frac{1}{n-k+1} = \frac{1}{n} \tag{3.11}$$

That is, the size of the cycle that contains a specified person is equally likely to be any of the values $1, 2, \ldots, n$. Moreover, since $C = 1$ means that 1 chooses his or her own hat, it follows that

$$P(E|C = 1) = 0$$

On the other hand, if $C = k$, then the set of hats chosen by the k individuals in this cycle is exactly the set of hats of these individuals. Hence, conditional on $C = k$, the problem reduces to determining the probability of no matches when $n - k$ people randomly choose among their own $n - k$ hats. Therefore, for $k > 1$

$$P(E|C = k) = P_{n-k} \tag{3.12}$$

Substituting (3.11)–(3.13) back into Eq. (3.10) gives

$$P_n = \frac{1}{n}\sum_{k=2}^{n} P_{n-k} \tag{3.13}$$

which is easily shown to be equivalent to Eq. (3.9). ■

Example 3.28 (The Ballot Problem)**.** In an election, candidate A receives n votes, and candidate B receives m votes where $n > m$. Assuming that all orderings are equally likely, show that the probability that A is always ahead in the count of votes is $(n - m)/(n + m)$.

Solution: Let $P_{n,m}$ denote the desired probability. By conditioning on which candidate receives the last vote counted we have

$$\begin{aligned} P_{n,m} &= P\{A \text{ always ahead} | A \text{ receives last vote}\}\frac{n}{n+m} \\ &\quad + P\{A \text{ always ahead} | B \text{ receives last vote}\}\frac{m}{n+m} \end{aligned}$$

Now, given that A receives the last vote, we can see that the probability that A is always ahead is the same as if A had received a total of $n - 1$ and B a total of m votes. Because a similar result is true when we are given that B receives the last vote, we see from the preceding that

$$P_{n,m} = \frac{n}{n+m}P_{n-1,m} + \frac{m}{m+n}P_{n,m-1} \tag{3.14}$$

We can now prove that $P_{n,m} = (n - m)/(n + m)$ by induction on $n + m$. As it is true when $n + m = 1$, that is, $P_{1,0} = 1$, assume it whenever $n + m = k$. Then when $n + m = k + 1$, we have by Eq. (3.14) and the induction hypothesis that

$$\begin{aligned} P_{n,m} &= \frac{n}{n+m}\frac{n-1-m}{n-1+m} + \frac{m}{m+n}\frac{n-m+1}{n+m-1} \\ &= \frac{n-m}{n+m} \end{aligned}$$

and the result is proven. ■

The ballot problem has some interesting applications. For example, consider successive flips of a coin that always land on "heads" with probability p, and let us determine the probability distribution of the first time, after beginning, that the total number of heads is equal to the total number of tails. The probability that the first time this occurs is at time $2n$ can be obtained by first conditioning on the total number of heads in the first $2n$ trials. This yields

$$\begin{aligned} &P\{\text{first time equal} = 2n\} \\ &\quad = P\{\text{first time equal} = 2n | n \text{ heads in first } 2n\}\binom{2n}{n}p^n(1-p)^n \end{aligned}$$

Now, given a total of n heads in the first $2n$ flips we can see that all possible orderings of the n heads and n tails are equally likely, and thus the preceding conditional probability is equivalent to the probability that in an election, in which each candidate receives n votes, one of the candidates is always ahead in the counting until the last

vote (which ties them). But by conditioning on whomever receives the last vote, we see that this is just the probability in the ballot problem when $m = n - 1$. Hence,

$$P\{\text{first time equal} = 2n\} = P_{n,n-1}\binom{2n}{n}p^n(1-p)^n$$
$$= \frac{\binom{2n}{n}p^n(1-p)^n}{2n-1}$$

Suppose now that we wanted to determine the probability that the first time there are i more heads than tails occurs after the $(2n+i)$th flip. Now, in order for this to be the case, the following two events must occur:

(a) The first $2n+i$ tosses result in $n+i$ heads and n tails; and
(b) The order in which the $n+i$ heads and n tails occur is such that the number of heads is never i more than the number of tails until after the final flip.

Now, it is easy to see that event (b) will occur if and only if the order of appearance of the $n+i$ heads and n tails is such that starting from the final flip and working backwards heads is always in the lead. For instance, if there are 4 heads and 2 tails $(n = 2, i = 2)$, then the outcome $\_\_\_\_TH$ would not suffice because there would have been 2 more heads than tails sometime before the sixth flip (since the first 4 flips resulted in 2 more heads than tails).

Now, the probability of the event specified in (a) is just the binomial probability of getting $n+i$ heads and n tails in $2n+i$ flips of the coin.

We must now determine the conditional probability of the event specified in (b) given that there are $n+i$ heads and n tails in the first $2n+i$ flips. To do so, note first that given that there are a total of $n+i$ heads and n tails in the first $2n+i$ flips, all possible orderings of these flips are equally likely. As a result, the conditional probability of (b) given (a) is just the probability that a random ordering of $n+i$ heads and n tails will, when counted in reverse order, always have more heads than tails. Since all reverse orderings are also equally likely, it follows from the ballot problem that this conditional probability is $i/(2n+i)$.

That is, we have shown that

$$P\{a\} = \binom{2n+i}{n}p^{n+i}(1-p)^n,$$
$$P\{b|a\} = \frac{i}{2n+i}$$

and so

$$P\{\text{first time heads leads by } i \text{ is after flip } 2n+i\} = \binom{2n+i}{n}p^{n+i}(1-p)^n\frac{i}{2n+i}$$

Example 3.29. Let $U_1, U_2, \ldots$ be a sequence of independent uniform (0, 1) random variables, and let

$$N = \min\{n \geq 2: U_n > U_{n-1}\}$$

and

$$M = \min\{n \geq 1: U_1 + \cdots + U_n > 1\}$$

That is, N is the index of the first uniform random variable that is larger than its immediate predecessor, and M is the number of uniform random variables we need sum to exceed 1. Surprisingly, N and M have the same probability distribution, and their common mean is e!

Solution: It is easy to find the distribution of N. Since all $n!$ possible orderings of $U_1, \ldots, U_n$ are equally likely, we have

$$P\{N > n\} = P\{U_1 > U_2 > \cdots > U_n\} = 1/n!$$

To show that $P\{M > n\} = 1/n!$, we will use mathematical induction. However, to give ourselves a stronger result to use as the induction hypothesis, we will prove the stronger result that for $0 < x \leq 1$, $P\{M(x) > n\} = x^n/n!, n \geq 1$, where

$$M(x) = \min\{n \geq 1: U_1 + \cdots + U_n > x\}$$

is the minimum number of uniforms that need be summed to exceed x. To prove that $P\{M(x) > n\} = x^n/n!$, note first that it is true for $n = 1$ since

$$P\{M(x) > 1\} = P\{U_1 \leq x\} = x$$

So assume that for all $0 < x \leq 1$, $P\{M(x) > n\} = x^n/n!$. To determine $P\{M(x) > n + 1\}$, condition on U_1 to obtain

$$\begin{aligned}
P\{M(x) > n+1\} &= \int_0^1 P\{M(x) > n+1 | U_1 = y\}\, dy \\
&= \int_0^x P\{M(x) > n+1 | U_1 = y\}\, dy \\
&= \int_0^x P\{M(x-y) > n\}\, dy \\
&= \int_0^x \frac{(x-y)^n}{n!}\, dy \quad \text{by the induction hypothesis} \\
&= \int_0^x \frac{u^n}{n!}\, du \\
&= \frac{x^{n+1}}{(n+1)!}
\end{aligned}$$

where the third equality of the preceding follows from the fact that given $U_1 = y$, $M(x)$ is distributed as 1 plus the number of uniforms that need be summed to exceed $x - y$. Thus, the induction is complete and we have shown that for $0 < x \leq 1, n \geq 1$,

$$P\{M(x) > n\} = x^n/n!$$

Letting $x = 1$ shows that N and M have the same distribution. Finally, we have

$$E[M] = E[N] = \sum_{n=0}^{\infty} P\{N > n\} = \sum_{n=0}^{\infty} 1/n! = e \qquad \blacksquare$$

Example 3.30. Let $X_1, X_2, \ldots$ be independent continuous random variables with a common distribution function F and density $f = F'$, and suppose that they are to be observed one at a time in sequence. Let

$$N = \min\{n \geqslant 2\text{: } X_n = \text{second largest of } X_1, \ldots, X_n\}$$

and let

$$M = \min\{n \geq 2\text{: } X_n = \text{second smallest of } X_1, \ldots, X_n\}$$

Which random variable—X_N, the first random variable which when observed is the second largest of those that have been seen, or X_M, the first one that on observation is the second smallest to have been seen—tends to be larger?

Solution: To calculate the probability density function of X_N, it is natural to condition on the value of N; so let us start by determining its probability mass function. Now, if we let

$$A_i = \{X_i \neq \text{second largest of } X_1, \ldots, X_i\}, \quad i \geq 2$$

then, for $n \geq 2$,

$$P\{N = n\} = P\left(A_2 A_3 \cdots A_{n-1} A_n^c\right)$$

Since the X_i are independent and identically distributed it follows that, for any $m \geq 1$, knowing the rank ordering of the variables $X_1, \ldots, X_m$ yields no information about the set of m values $\{X_1, \ldots, X_m\}$. That is, for instance, knowing that $X_1 < X_2$ gives us no information about the values of $\min(X_1, X_2)$ or $\max(X_1, X_2)$. It follows from this that the events $A_i, i \geq 2$ are independent. Also, since X_i is equally likely to be the largest, or the second largest, $\ldots$, or the ith largest of $X_1, \ldots, X_i$ it follows that $P(A_i) = (i-1)/i, i \geq 2$. Therefore, we see that

$$P\{N = n\} = \frac{1}{2}\frac{2}{3}\frac{3}{4} \cdots \frac{n-2}{n-1}\frac{1}{n} = \frac{1}{n(n-1)}$$

Hence, conditioning on N yields that the probability density function of X_N is as follows:

$$f_{X_N}(x) = \sum_{n=2}^{\infty} \frac{1}{n(n-1)} f_{X_N|N}(x|n)$$

Now, since the ordering of the variables $X_1, \ldots, X_n$ is independent of the set of values $\{X_1, \ldots, X_n\}$, it follows that the event $\{N = n\}$ is independent of

$\{X_1, \ldots, X_n\}$. From this, it follows that the conditional distribution of X_N given that $N = n$ is equal to the distribution of the second largest from a set of n random variables having distribution F. Thus, using the results of Example 2.38 concerning the density function of such a random variable, we obtain

$$\begin{aligned} f_{X_N}(x) &= \sum_{n=2}^{\infty} \frac{1}{n(n-1)} \frac{n!}{(n-2)!1!} (F(x))^{n-2} f(x)(1 - F(x)) \\ &= f(x)(1 - F(x)) \sum_{i=0}^{\infty} (F(x))^i \\ &= f(x) \end{aligned}$$

Thus, rather surprisingly, X_N has the same distribution as X_1, namely, F. Also, if we now let $W_i = -X_i$, $i \geq 1$, then W_M will be the value of the first W_i, which on observation is the second largest of all those that have been seen. Hence, by the preceding, it follows that W_M has the same distribution as W_1. That is, $-X_M$ has the same distribution as $-X_1$, and so X_M also has distribution F! In other words, whether we stop at the first random variable that is the second largest of all those presently observed, or we stop at the first one that is the second smallest of all those presently observed, we will end up with a random variable having distribution F.

Whereas the preceding result is quite surprising, it is a special case of a general result known as *Ignatov's theorem*, which yields even more surprises. For instance, for $k \geqslant 1$, let

$$N_k = \min\{n \geqslant k\colon X_n = k\text{th largest of } X_1, \ldots, X_n\}$$

Therefore, N_2 is what we previously called N, and X_{N_k} is the first random variable that upon observation is the kth largest of all those observed up to this point. It can then be shown by a similar argument as used in the preceding that X_{N_k} has distribution function F for all k (see Exercise 82 at the end of this chapter). In addition, it can be shown that the random variables $X_{N_k}, k \geqslant 1$ are independent. (A statement and proof of Ignatov's theorem in the case of discrete random variables are given in Section 3.6.6.) ■

Example 3.31. A population consists of m families. Let X_j denote the size of family j, and suppose that $X_1, \ldots, X_m$ are independent random variables having the common probability mass function

$$p_k = P(X_j = k), \quad \sum_{k=1}^{\infty} p_k = 1$$

with mean $\mu = \Sigma_k k p_k$. Suppose a member of the population is randomly chosen, in that the selection is equally likely to be any of the members of the population, and let

S_i be the event that the selected individual is from a family of size i. Argue that

$$P(S_i) \to \frac{ip_i}{\mu} \text{ as } \quad m \to \infty$$

Solution: A heuristic argument for the preceding formula is that because each family is of size i with probability p_i, it follows that there are approximately mp_i families of size i when m is large. Thus, imp_i members of the population come from a family of size i, implying that the probability that the selected individual is from a family of size i is approximately $\frac{imp_i}{\sum_j jmp_j} = \frac{ip_i}{\mu}$.

For a more formal argument, let N_i denote the number of families that are of size i. That is,

$$N_i = \text{number } \{k : k = 1, \ldots, m : X_k = i\}$$

Then, conditional on $\mathbf{X} = (X_1, \ldots, X_m)$

$$P(S_i|\mathbf{X}) = \frac{iN_i}{\sum_{k=1}^m X_k}$$

Hence,

$$\begin{aligned} P(S_i) &= E[P(S_i|X)] \\ &= E\left[\frac{iN_i}{\sum_{k=1}^m X_k}\right] \\ &= E\left[\frac{iN_i/m}{\sum_{k=1}^m X_k/m}\right] \end{aligned}$$

Because each family is independently of size i with probability p_i, it follows by the strong law of large numbers that N_i/m, the fraction of families that are of size i, converges to p_i as $m \to \infty$. Also by the strong law of large numbers, $\sum_{k=1}^m X_k/m \to E[X] = \mu$ as $m \to \infty$. Consequently, with probability 1,

$$\frac{iN_i/m}{\sum_{k=1}^m X_k/m} \to \frac{ip_i}{\mu} \quad \text{as } m \to \infty$$

Because the random variable $\frac{iN_i}{\sum_{k=1}^m X_k}$ converges to $\frac{ip_i}{\mu}$ so does its expectation, which proves the result. (While it is not always the case that $\lim_{m\to\infty} Y_m = c$ implies that $\lim_{m\to\infty} E[Y_m] = c$, the implication is true when the Y_m are uniformly bounded random variables, and the random variables $\frac{iNi}{\sum_{k=1}^m X_k}$ are all between 0 and 1.) ■

The use of conditioning can also result in a more computationally efficient solution than a direct calculation. This is illustrated by our next example.

Example 3.32. Consider n independent trials in which each trial results in one of the outcomes $1, \ldots, k$ with respective probabilities $p_1, \ldots, p_k$, $\sum_{i=1}^k p_i = 1$. Suppose

further that $n > k$, and that we are interested in determining the probability that each outcome occurs at least once. If we let A_i denote the event that outcome i does not occur in any of the n trials, then the desired probability is $1 - P(\bigcup_{i=1}^{k} A_i)$, and it can be obtained by using the inclusion–exclusion theorem as follows:

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i) - \sum_{i}\sum_{j>i} P(A_i A_j) + \sum_{i}\sum_{j>i}\sum_{k>j} P(A_i A_j A_k) - \cdots + (-1)^{k+1} P(A_1 \cdots A_k)$$

where

$$\begin{aligned} P(A_i) &= (1 - p_i)^n \\ P(A_i A_j) &= (1 - p_i - p_j)^n, \quad i < j \\ P(A_i A_j A_k) &= (1 - p_i - p_j - p_k)^n, \quad i < j < k \end{aligned}$$

The difficulty with the preceding solution is that its computation requires the calculation of $2^k - 1$ terms, each of which is a quantity raised to the power n. The preceding solution is thus computationally inefficient when k is large. Let us now see how to make use of conditioning to obtain an efficient solution.

To begin, note that if we start by conditioning on N_k (the number of times that outcome k occurs) then when $N_k > 0$ the resulting conditional probability will equal the probability that all of the outcomes $1, \ldots, k-1$ occur at least once when $n - N_k$ trials are performed, and each results in outcome i with probability $p_i / \sum_{j=1}^{k-1} p_j$, $i = 1, \ldots, k-1$. We could then use a similar conditioning step on these terms.

To follow through on the preceding idea, let $A_{m,r}$, for $m \leqslant n, r \leqslant k$, denote the event that each of the outcomes $1, \ldots, r$ occurs at least once when m independent trials are performed, where each trial results in one of the outcomes $1, \ldots, r$ with respective probabilities $p_1/P_r, \ldots, p_r/P_r$, where $P_r = \sum_{j=1}^{r} p_j$. Let $P(m, r) = P(A_{m,r})$ and note that $P(n, k)$ is the desired probability. To obtain an expression for $P(m, r)$, condition on the number of times that outcome r occurs. This gives

$$\begin{aligned} P(m, r) &= \sum_{j=0}^{m} P\{A_{m,r} | r \text{ occurs } j \text{ times}\} \binom{m}{j} \left(\frac{p_r}{P_r}\right)^j \left(1 - \frac{p_r}{P_r}\right)^{m-j} \\ &= \sum_{j=1}^{m-r+1} P(m - j, r - 1) \binom{m}{j} \left(\frac{p_r}{P_r}\right)^j \left(1 - \frac{p_r}{P_r}\right)^{m-j} \end{aligned}$$

Starting with

$$\begin{aligned} P(m, 1) &= 1, \quad \text{if } m \geqslant 1 \\ P(m, 1) &= 0, \quad \text{if } m = 0 \end{aligned}$$

we can use the preceding recursion to obtain the quantities $P(m, 2), m = 2, \ldots, n - (k - 2)$, and then the quantities $P(m, 3), m = 3, \ldots, n - (k - 3)$, and so on, up to $P(m, k - 1)$, $m = k - 1, \ldots, n - 1$. At this point we can then use the recursion to compute $P(n, k)$. It is not difficult to check that the amount of computation needed is a polynomial function of k, which will be much smaller than 2^k when k is large. ■

Our next example is concerned with final score probabilities in serve and rally games such as table tennis, squash, paddle ball, volleyball, and others.

Example 3.33 (Serve and Rally Competitions)**.** Consider a serve and rally competition involving players A and B. Suppose that each rally that begins with a serve by player A is won by player A with probability p_a and is won by player B with probability $q_a = 1 - p_a$. Furthermore, suppose that each rally that begins with a serve by player B is won by player A with probability p_b and is won by player B with probability $q_b = 1 - p_b$. Suppose that the winner of the rally earns a point and becomes the server of the next rally. The competition is decided either when A has won a total of N points or when B has won a total of M. Given that A serves first, we are interested in determining the final score probabilities.

The format of this example is used in a variety of serve and rally games, including international volleyball and American squash, both of which changed from their original format which gave service to the winner of the previous rally but only awarded a point if the winner of a rally was the server. (See Exercise 84 for an analysis of this latter format.)

Let F denote the final score, with $F = (i, j)$ meaning that A won a total of i points and B a total of j points. Clearly

$$P(F = (N, 0)) = p_a^N, \quad P(F = (0, M)) = q_a q_b^{M-1}$$

To determine the other final score probabilities, imagine that A and B continue to play even after the competition is decided. Define the concept of a "round" by letting the initial serve of A start the first round and letting a new round begin each time A serves. Let B_i denote the number of points won by B in round i. Note that if the first point of a round is won by A, then that round ends with B winning 0 points in it. On the other hand, if B wins the first point in a round then B will continue serving until A wins a point, showing that the number of points won by B in a round is equal to the number of times that B serves. Because the number of consecutive serves of B before A wins a point is geometric with parameter p_b, we see that

$$B_i = \begin{cases} 0, & \text{with probability } p_a \\ \text{Geometric}(p_b), & \text{with probability } q_a \end{cases}$$

That is,

$$P(B_i = 0) = p_a$$
$$P(B_i = k | B_i > 0) = q_b^{k-1} p_b, k > 0$$

Because a new round begins each time A wins a point, it follows that B_i is the number of points that B wins between the time that A has won $i-1$ points until A has won i points. Consequently, $B(n) \equiv \sum_{i=1}^{n} B_i$ is the number of points that B has won at the moment that A wins its nth point. Noting that the final score will be $(N, m), m < M$, if $B(N) = m$, let us determine $P(B(n) = m)$ for $m > 0$. To do so, we condition on the number of $B_1, \ldots, B_n$ that are positive. Calling this number Y, that is,

$$Y = \text{number of } i \leq n \text{ such that } B_i > 0$$

we obtain

$$\begin{aligned} P(B(n) = m) &= \sum_{r=0}^{n} P(B(n) = m | Y = r) P(Y = r) \\ &= \sum_{r=1}^{n} P(B(n) = m | Y = r) P(Y = r) \end{aligned}$$

where the last equality followed since $m > 0$ and so $P(B(n) = m | Y = 0) = 0$. Because each of $B_1, \ldots, B_n$ is independently positive with probability q_a, it follows that Y, the number of them that are positive, is binomial with parameters n, q_a. Consequently,

$$P(B(n) = m) = \sum_{r=1}^{n} P(B(n) = m | Y = r) \binom{n}{r} q_a^r p_a^{n-r}$$

Now, if r of the variables $B_1, \ldots, B_n$ are positive, then $B(n)$ is distributed as the sum of r independent geometric random variables with parameter p_b, which is the negative binomial distribution of the number of trials until there have been r successes when each trial is independently a success with probability p_b. Hence,

$$P(B(n) = m | Y = r) = \binom{m-1}{r-1} p_b^r q_b^{m-r}$$

where we are using the convention that $\binom{a}{b} = 0$ if $b > a$. This gives

$$\begin{aligned} P(B(n) = m) &= \sum_{r=1}^{n} \binom{m-1}{r-1} p_b^r q_b^{m-r} \binom{n}{r} q_a^r p_a^{n-r} \\ &= q_b^m p_a^n \sum_{r=1}^{n} \binom{m-1}{r-1} \binom{n}{r} \left(\frac{p_b q_a}{q_b p_a} \right)^r \end{aligned}$$

Thus, we have shown that

$$P(F = (N, m)) = P(B(N) = m)$$

$$= q_b^m p_a^N \sum_{r=1}^{N} \binom{m-1}{r-1} \binom{N}{r} \left(\frac{p_b q_a}{q_b p_a}\right)^r, \quad 0 < m < M$$

To determine the probability that the final score will be $(n, M), 0 < n < N$, we condition on the number of wins that B has at the moment that A wins its nth game to obtain

$$\begin{aligned} P(F = (n, M)) &= \sum_{m=0}^{\infty} P(F = (n, M) | B(n) = m) P(B(n) = m) \\ &= \sum_{m=0}^{M-1} P(F = (n, M) | B(n) = m) P(B(n) = m) \end{aligned}$$

Now, given that B has $m < M$ points at the moment that A wins its nth point, in order for the final score to be (n, M) B must win the next point with A serving and must then win the final $M - m - 1$ points on its serve. Hence, $P(F = (n, M) | B(n) = m) = q_a q_b^{M-m-1}$, giving that

$$\begin{aligned} P(F = (n, M)) &= \sum_{m=0}^{M-1} q_a q_b^{M-m-1} P(B(n) = m) \\ &= q_a q_b^{M-1} p_a^n + \sum_{m=1}^{M-1} q_a q_b^{M-m-1} P(B(n) = m) \\ &= q_a q_b^{M-1} p_a^n \left[1 + \sum_{m=1}^{M-1} \sum_{r=1}^{n} \binom{m-1}{r-1} \binom{n}{r} \left(\frac{p_b q_a}{q_b p_a}\right)^r \right], \\ & \qquad\qquad 0 < n < N \end{aligned}$$

■

As noted previously, conditional expectations given that $Y = y$ are exactly the same as ordinary expectations except that all probabilities are computed conditional on the event that $Y = y$. As such, conditional expectations satisfy all the properties of ordinary expectations. For instance, the analog of

$$E[X] = \begin{cases} \sum_w E[X | W = w] P\{W = w\}, & \text{if } W \text{ is discrete} \\ \int_w E[X | W = w] f_W(w)\, dw, & \text{if } W \text{ is continuous} \end{cases}$$

is

$$E[X | Y = y] = \begin{cases} \sum_w E[X | W = w, Y = y] P\{W = w | Y = y\}, & \text{if } W \text{ is discrete} \\ \int_w E[X | W = w, Y = y] f_{W|Y}(w|y)\, dw, & \text{if } W \text{ is continuous} \end{cases}$$

If $E[X|Y, W]$ is defined to be that function of Y and W that, when $Y = y$, and $W = w$, is equal to $E[X|Y = y, W = w]$, then the preceding can be written as

$$E[X|Y] = E\big[E[X|Y, W]|Y\big]$$

Example 3.34. An automobile insurance company classifies each of its policyholders as being of one of the types $i = 1, \ldots, k$. It supposes that the numbers of accidents that a type i policyholder has in successive years are independent Poisson random variables with mean λ_i, $i = 1, \ldots, k$. The probability that a newly insured policyholder is type i is p_i, $\sum_{i=1}^{k} p_i = 1$. Given that a policyholder had n accidents in her first year, what is the expected number that she has in her second year? What is the conditional probability that she has m accidents in her second year?

Solution: Let N_i denote the number of accidents the policyholder has in year $i, i = 1, 2$. To obtain $E[N_2|N_1 = n]$, condition on her risk type T.

$$\begin{aligned}
E[N_2|N_1 = n] &= \sum_{j=1}^{k} E[N_2|T = j, N_1 = n]P\{T = j|N_1 = n\} \\
&= \sum_{j=1}^{k} E[N_2|T = j]P\{T = j|N_1 = n\} \\
&= \sum_{j=1}^{k} \lambda_j P\{T = j|N_1 = n\} \\
&= \frac{\sum_{j=1}^{k} e^{-\lambda_j}\lambda_j^{n+1} p_j}{\sum_{j=1}^{k} e^{-\lambda_j}\lambda_j^{n} p_j}
\end{aligned}$$

where the final equality used that

$$\begin{aligned}
P\{T = j|N_1 = n\} &= \frac{P\{T = j, N_1 = n\}}{P\{N_1 = n\}} \\
&= \frac{P\{N_1 = n|T = j\}P\{T = j\}}{\sum_{j=1}^{k} P\{N_1 = n|T = j\}P\{T = j\}} \\
&= \frac{p_j e^{-\lambda_j}\lambda_j^n/n!}{\sum_{j=1}^{k} p_j e^{-\lambda_j}\lambda_j^n/n!}
\end{aligned}$$

The conditional probability that the policyholder has m accidents in year 2 given that she had n in year 1 can also be obtained by conditioning on her type.

$$\begin{aligned}
P\{N_2 = m|N_1 = n\} &= \sum_{j=1}^{k} P\{N_2 = m|T = j, N_1 = n\}P\{T = j|N_1 = n\} \\
&= \sum_{j=1}^{k} e^{-\lambda_j}\frac{\lambda_j^m}{m!} P\{T = j|N_1 = n\}
\end{aligned}$$

$$= \frac{\sum_{j=1}^{k} e^{-2\lambda_j} \lambda_j^{m+n} p_j}{m! \sum_{j=1}^{k} e^{-\lambda_j} \lambda_j^n p_j}$$

Another way to calculate $P\{N_2 = m | N_1 = n\}$ is first to write

$$P\{N_2 = m | N_1 = n\} = \frac{P\{N_2 = m, N_1 = n\}}{P\{N_1 = n\}}$$

and then determine both the numerator and denominator by conditioning on T. This yields

$$\begin{aligned} P\{N_2 = m | N_1 = n\} &= \frac{\sum_{j=1}^{k} P\{N_2 = m, N_1 = n | T = j\} p_j}{\sum_{j=1}^{k} P\{N_1 = n | T = j\} p_j} \\ &= \frac{\sum_{j=1}^{k} e^{-\lambda_j} \frac{\lambda_j^m}{m!} e^{-\lambda_j} \frac{\lambda_j^n}{n!} p_j}{\sum_{j=1}^{k} e^{-\lambda_j} \frac{\lambda_j^n}{n!} p_j} \\ &= \frac{\sum_{j=1}^{k} e^{-2\lambda_j} \lambda_j^{m+n} p_j}{m! \sum_{j=1}^{k} e^{-\lambda_j} \lambda_j^n p_j} \end{aligned}$$

■

3.6 Some Applications

3.6.1 A List Model

Consider n elements—$e_1, e_2, \ldots, e_n$—that are initially arranged in some ordered list. At each unit of time a request is made for one of these elements—e_i being requested, independently of the past, with probability P_i. After being requested the element is then moved to the front of the list. That is, for instance, if the present ordering is e_1, e_2, e_3, e_4 and e_3 is requested, then the next ordering is e_3, e_1, e_2, e_4.

We are interested in determining the expected position of the element requested after this process has been in operation for a long time. However, before computing this expectation, let us note two possible applications of this model. In the first we have a stack of reference books. At each unit of time a book is randomly selected and is then returned to the top of the stack. In the second application we have a computer receiving requests for elements stored in its memory. The request probabilities for the elements may not be known, so to reduce the average time it takes the computer to locate the element requested (which is proportional to the position of the requested element if the computer locates the element by starting at the beginning and then going down the list), the computer is programmed to replace the requested element at the beginning of the list.

To compute the expected position of the element requested, we start by conditioning on which element is selected. This yields

$$
\begin{aligned}
&E[\text{ position of element requested }] \\
&= \sum_{i=1}^{n} E[\text{ position}|e_i \text{ is selected }]P_i \\
&= \sum_{i=1}^{n} E[\text{ position of } e_i|e_i \text{ is selected }]P_i \\
&= \sum_{i=1}^{n} E[\text{ position of } e_i \text{ }]P_i \qquad (3.15)
\end{aligned}
$$

where the final equality used that the position of e_i and the event that e_i is selected are independent because, regardless of its position, e_i is selected with probability P_i.

Now,

$$\text{position of } e_i = 1 + \sum_{j \neq i} I_j$$

where

$$I_j = \begin{cases} 1, & \text{if } e_j \text{ precedes } e_i \\ 0, & \text{otherwise} \end{cases}$$

and so,

$$
\begin{aligned}
E[\text{ position of } e_i \text{ }] &= 1 + \sum_{j \neq i} E[I_j] \\
&= 1 + \sum_{j \neq i} P\{e_j \text{ precedes } e_i\} \qquad (3.16)
\end{aligned}
$$

To compute $P\{e_j \text{ precedes } e_i\}$, note that e_j will precede e_i if the most recent request for either of them was for e_j. But given that a request is for either e_i or e_j, the probability that it is for e_j is

$$P\{e_j|e_i \text{ or } e_j\} = \frac{P_j}{P_i + P_j}$$

and, thus,

$$P\{e_j \text{ precedes } e_i\} = \frac{P_j}{P_i + P_j}$$

Hence, from Eqs. (3.15) and (3.16) we see that

$$E\{\text{position of element requested}\} = 1 + \sum_{i=1}^{n} P_i \sum_{j \neq i} \frac{P_j}{P_i + P_j}$$

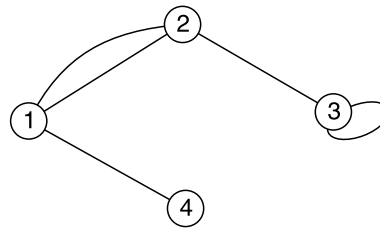

Figure 3.1 A graph.

This list model will be further analyzed in Section 4.8, where we will assume a different reordering rule—namely, that the element requested is moved one closer to the front of the list as opposed to being moved to the front of the list as assumed here. We will show there that the average position of the requested element is less under the one-closer rule than it is under the front-of-the-line rule.

3.6.2 A Random Graph

A graph consists of a set V of elements called nodes and a set A of pairs of elements of V called arcs. A graph can be represented graphically by drawing circles for nodes and drawing lines between nodes i and j whenever (i, j) is an arc. For instance if $V = \{1, 2, 3, 4\}$ and $A = \{(1, 2), (1, 4), (2, 3), (1, 2), (3, 3)\}$, then we can represent this graph as shown in Fig. 3.1. Note that the arcs have no direction (a graph in which the arcs are ordered pairs of nodes is called a directed graph); and that in the figure there are multiple arcs connecting nodes 1 and 2, and a self-arc (called a self-loop) from node 3 to itself.

We say that there exists a path from node i to node j, $i \neq j$, if there exists a sequence of nodes $i, i_1, \ldots, i_k, j$ such that $(i, i_1), (i_1, i_2), \ldots, (i_k, j)$ are all arcs. If there is a path between each of the $\binom{n}{2}$ distinct pair of nodes we say that the graph is *connected*. The graph in Fig. 3.1 is connected but the graph in Fig. 3.2 is not. Consider now the following graph where $V = \{1, 2, \ldots, n\}$ and $A = \{(i, X(i)), i = 1, \ldots, n\}$ where the $X(i)$ are independent random variables such that

$$P\{X(i) = j\} = \frac{1}{n}, \quad j = 1, 2, \ldots, n$$

In other words from each node i we select at random one of the n nodes (including possibly the node i itself) and then join node i and the selected node with an arc. Such a graph is commonly referred to as a *random graph*.

We are interested in determining the probability that the random graph so obtained is connected. As a prelude, starting at some node—say, node 1—let us follow the sequence of nodes, $1, X(1), X^2(1), \ldots$, where $X^n(1) = X(X^{n-1}(1))$; and define N to equal the first k such that $X^k(1)$ is not a new node. In other words,

$$N = 1\text{st } k \text{ such that } X^k(1) \in \{1, X(1), \ldots, X^{k-1}(1)\}$$

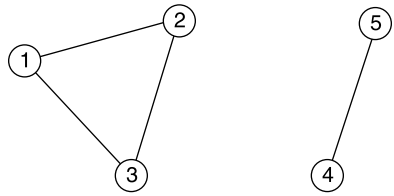

Figure 3.2 A disconnected graph.

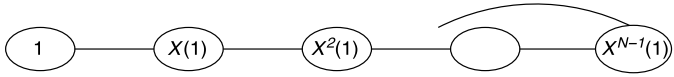

Figure 3.3

We can represent this as shown in Fig. 3.3 where the arc from $X^{N-1}(1)$ goes back to a node previously visited.

To obtain the probability that the graph is connected we first condition on N to obtain

$$P\{\text{graph is connected}\} = \sum_{k=1}^{n} P\{\text{connected}|N=k\}P\{N=k\} \tag{3.17}$$

Now, given that $N = k$, the k nodes $1, X(1), \ldots, X^{k-1}(1)$ are connected to each other, and there are no other arcs emanating out of these nodes. In other words, if we regard these k nodes as being one supernode, the situation is the same as if we had one supernode and $n-k$ ordinary nodes with arcs emanating from the ordinary nodes—each arc going into the supernode with probability k/n. The solution in this situation is obtained from Lemma 3.1 by taking $r = n-k$.

Lemma 3.1. *Given a random graph consisting of nodes* $0, 1, \ldots, r$ *and* r *arcs—namely,* $(i, Y_i), i = 1, \ldots, r$, *where*

$$Y_i = \begin{cases} j & \text{with probability } \dfrac{1}{r+k}, \quad j = 1, \ldots, r \\ 0 & \text{with probability } \dfrac{k}{r+k} \end{cases}$$

then

$$P\{\text{graph is connected}\} = \frac{k}{r+k}$$

(In other words, for the preceding graph there are $r+1$ nodes—r ordinary nodes and one supernode. Out of each ordinary node an arc is chosen. The arc goes to the supernode with probability $k/(r+k)$ and to each of the ordinary ones with probability $1/(r+k)$. There is no arc emanating out of the supernode.)

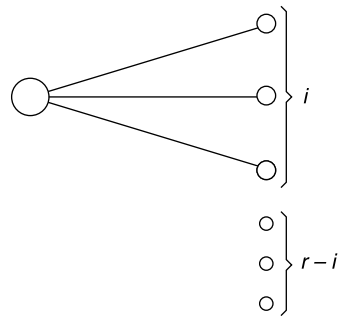

Figure 3.4 The situation given that i of the r arcs are into the supernode.

Proof. The proof is by induction on r. As it is true when $r = 1$ for any k, assume it true for all values less than r. Now, in the case under consideration, let us first condition on the number of arcs (j, Y_j) for which $Y_j = 0$. This yields

$$P\{\text{connected}\} = \sum_{i=0}^{r} P\{\text{connected}|i \text{ of the } Y_j = 0\} \binom{r}{i} \left(\frac{k}{r+k}\right)^i \left(\frac{r}{r+k}\right)^{r-i} \tag{3.18}$$

Now, given that exactly i of the arcs are into the supernode (see Fig. 3.4), the situation for the remaining $r - i$ arcs which do not go into the supernode is the same as if we had $r - i$ ordinary nodes and one supernode with an arc going out of each of the ordinary nodes—into the supernode with probability i/r and into each ordinary node with probability $1/r$. But by the induction hypothesis the probability that this would lead to a connected graph is i/r.

Hence,

$$P\{\text{connected}|i \text{ of the } Y_j = 0\} = \frac{i}{r}$$

and from Eq. (3.18)

$$\begin{aligned} P\{\text{connected}\} &= \sum_{i=0}^{r} \frac{i}{r} \binom{r}{i} \left(\frac{k}{r+k}\right)^i \left(\frac{r}{r+k}\right)^{r-i} \\ &= \frac{1}{r} E\left[\text{binomial}\left(r, \frac{k}{r+k}\right)\right] \\ &= \frac{k}{r+k} \end{aligned}$$

which completes the proof of the lemma. ■

Hence, as the situation given $N = k$ is exactly as described by Lemma 3.1 when $r = n - k$, we see that, for the original graph,

$$P\{\text{graph is connected}|N=k\}=\frac{k}{n}$$

and, from Eq. (3.17),

$$P\{\text{graph is connected}\}=\frac{E(N)}{n} \tag{3.19}$$

To compute $E(N)$ we use the identity

$$E(N)=\sum_{i=1}^{\infty}P\{N\geqslant i\}$$

which can be proved by defining indicator variables I_i, $i\geqslant 1$, by

$$I_i=\begin{cases}1, & \text{if } i\leqslant N\\ 0, & \text{if } i>N\end{cases}$$

Hence,

$$N=\sum_{i=1}^{\infty}I_i$$

and so

$$\begin{aligned}E(N)&=E\left[\sum_{i=1}^{\infty}I_i\right]\\&=\sum_{i=1}^{\infty}E[I_i]\\&=\sum_{i=1}^{\infty}P\{N\geqslant i\}\end{aligned} \tag{3.20}$$

Now, the event $\{N\geqslant i\}$ occurs if the nodes $1, X(1), \ldots, X^{i-1}(1)$ are all distinct. Hence,

$$\begin{aligned}P\{N\geq i\}&=\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-i+1)}{n}\\&=\frac{(n-1)!}{(n-i)!n^{i-1}}\end{aligned}$$

and so, from Eqs. (3.19) and (3.20),

$$\begin{aligned}P\{\text{graph is connected}\}&=(n-1)!\sum_{i=1}^{n}\frac{1}{(n-i)!n^i}\\&=\frac{(n-1)!}{n^n}\sum_{j=0}^{n-1}\frac{n^j}{j!}\quad(\text{by } j=n-i)\end{aligned} \tag{3.21}$$

We can also use Eq. (3.21) to obtain a simple approximate expression for the probability that the graph is connected when n is large. To do so, we first note that if X is a Poisson random variable with mean n, then

$$P\{X < n\} = e^{-n} \sum_{j=0}^{n-1} \frac{n^j}{j!}$$

Since a Poisson random variable with mean n can be regarded as being the sum of n independent Poisson random variables each with mean 1, it follows from the central limit theorem that for n large such a random variable has approximately a normal distribution and as such has probability $\frac{1}{2}$ of being less than its mean. That is, for n large,

$$P\{X < n\} \approx \tfrac{1}{2}$$

and so for n large,

$$\sum_{j=0}^{n-1} \frac{n^j}{j!} \approx \frac{e^n}{2}$$

Hence, from Eq. (3.21), for n large,

$$P\{\text{graph is connected}\} \approx \frac{e^n (n-1)!}{2n^n}$$

By employing an approximation due to Stirling that states that for n large,

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}$$

we see that, for n large,

$$P\{\text{graph is connected}\} \approx \sqrt{\frac{\pi}{2(n-1)}} e \left(\frac{n-1}{n}\right)^n$$

and as

$$\lim_{n\to\infty} \left(\frac{n-1}{n}\right)^n = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = e^{-1}$$

we see that, for n large,

$$P\{\text{graph is connected}\} \approx \sqrt{\frac{\pi}{2(n-1)}}$$

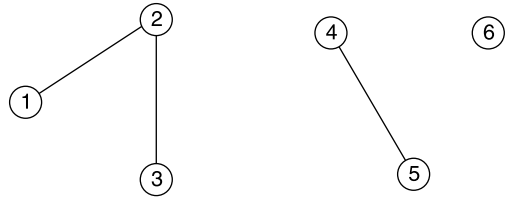

Figure 3.5 A graph having three connected components.

Now a graph is said to consist of r connected components if its nodes can be partitioned into r subsets so that each of the subsets is connected and there are no arcs between nodes in different subsets. For instance, the graph in Fig. 3.5 consists of three connected components—namely, $\{1, 2, 3\}$, $\{4, 5\}$, and $\{6\}$. Let C denote the number of connected components of our random graph and let

$$P_n(i) = P\{C = i\}$$

where we use the notation $P_n(i)$ to make explicit the dependence on n, the number of nodes. Since a connected graph is by definition a graph consisting of exactly one component, from Eq. (3.21) we have

$$\begin{aligned} P_n(1) &= P\{C = 1\} \\ &= \frac{(n-1)!}{n^n} \sum_{j=0}^{n-1} \frac{n^j}{j!} \end{aligned} \tag{3.22}$$

To obtain $P_n(2)$, the probability of exactly two components, let us first fix attention on some particular node—say, node 1. In order that a given set of $k-1$ other nodes—say, nodes $2, \ldots, k$—will along with node 1 constitute one connected component, and the remaining $n-k$ a second connected component, we must have

(i) $X(i) \in \{1, 2, \ldots, k\}$, for all $i = 1, \ldots, k$.
(ii) $X(i) \in \{k+1, \ldots, n\}$, for all $i = k+1, \ldots, n$.
(iii) The nodes $1, 2, \ldots, k$ form a connected subgraph.
(iv) The nodes $k+1, \ldots, n$ form a connected subgraph.

The probability of the preceding occurring is clearly

$$\left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} P_k(1) P_{n-k}(1)$$

and because there are $\binom{n-1}{k-1}$ ways of choosing a set of $k-1$ nodes from the nodes 2 through n, we have

$$P_n(2) = \sum_{k=1}^{n-1} \binom{n-1}{k-1} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} P_k(1) P_{n-k}(1)$$

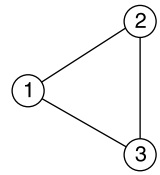

Figure 3.6 A cycle.

and so $P_n(2)$ can be computed from Eq. (3.22). In general, the recursive formula for $P_n(i)$ is given by

$$P_n(i) = \sum_{k=1}^{n-i+1} \binom{n-1}{k-1} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} P_k(1) P_{n-k}(i-1)$$

To compute $E[C]$, the expected number of connected components, first note that every connected component of our random graph must contain exactly one cycle (a cycle is a set of arcs of the form $(i, i_1), (i_1, i_2), \ldots, (i_{k-1}, i_k), (i_k, i)$ for distinct nodes $i, i_1, \ldots, i_k$). For example, Fig. 3.6 depicts a cycle.

The fact that every connected component of our random graph must contain exactly one cycle is most easily proved by noting that if the connected component consists of r nodes, then it must also have r arcs and, hence, must contain exactly one cycle (why?). Thus, we see that

$$\begin{aligned} E[C] &= E[\text{number of cycles}] \\ &= E\left[\sum_S I(S)\right] \\ &= \sum_S E[I(S)] \end{aligned}$$

where the sum is over all subsets $S \subset \{1, 2, \ldots, n\}$ and

$$I(S) = \begin{cases} 1, & \text{if the nodes in } S \text{ are all the nodes of a cycle} \\ 0, & \text{otherwise} \end{cases}$$

Now, if S consists of k nodes, say $1, \ldots, k$, then

$$\begin{aligned} E[I(S)] &= P\{1, X(1), \ldots, X^{k-1}(1) \text{ are all distinct and contained in} \\ &\qquad 1, \ldots, k \text{ and } X^k(1) = 1\} \\ &= \frac{k-1}{n}\frac{k-2}{n} \cdots \frac{1}{n}\frac{1}{n} = \frac{(k-1)!}{n^k} \end{aligned}$$

Hence, because there are $\binom{n}{k}$ subsets of size k we see that

$$E[C] = \sum_{k=1}^{n} \binom{n}{k} \frac{(k-1)!}{n^k}$$

3.6.3 Uniform Priors, Polya's Urn Model, and Bose–Einstein Statistics

Suppose that n independent trials, each of which is a success with probability p, are performed. If we let X denote the total number of successes, then X is a binomial random variable such that

$$P\{X = k|p\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

However, let us now suppose that whereas the trials all have the same success probability p, its value is not predetermined but is chosen according to a uniform distribution on (0, 1). (For instance, a coin may be chosen at random from a huge bin of coins representing a uniform spread over all possible values of p, the coin's probability of coming up heads. The chosen coin is then flipped n times.) In this case, by conditioning on the actual value of p, we have

$$\begin{aligned} P\{X = k\} &= \int_0^1 P\{X = k|p\} f(p)\, dp \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\, dp \end{aligned}$$

Now, it can be shown that

$$\int_0^1 p^k (1-p)^{n-k} dp = \frac{k!(n-k)!}{(n+1)!} \tag{3.23}$$

and thus

$$\begin{aligned} P\{X = k\} &= \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} \\ &= \frac{1}{n+1}, \quad k = 0, 1, \ldots, n \end{aligned} \tag{3.24}$$

In other words, each of the $n + 1$ possible values of X is equally likely.

As an alternate way of describing the preceding experiment, let us compute the conditional probability that the $(r + 1)$st trial will result in a success given a total of k successes (and $r - k$ failures) in the first r trials.

$$
\begin{aligned}
&P\{(r+1)\text{st trial is a success}|k \text{ successes in first } r\} \\
&\quad = \frac{P\{(r+1)\text{st is a success, } k \text{ successes in first } r \text{ trials}\}}{P\{k \text{ successes in first } r \text{ trials}\}} \\
&\quad = \frac{\int_0^1 P\{(r+1)\text{st is a success, } k \text{ in first } r|p\}\, dp}{1/(r+1)} \\
&\quad = (r+1)\int_0^1 \binom{r}{k} p^{k+1}(1-p)^{r-k}\, dp \\
&\quad = (r+1)\binom{r}{k}\frac{(k+1)!(r-k)!}{(r+2)!} \quad \text{by Eq. (3.23)} \\
&\quad = \frac{k+1}{r+2} \qquad (3.25)
\end{aligned}
$$

That is, if the first r trials result in k successes, then the next trial will be a success with probability $(k+1)/(r+2)$.

It follows from Eq. (3.25) that an alternative description of the stochastic process of the successive outcomes of the trials can be described as follows: There is an urn that initially contains one white and one black ball. At each stage a ball is randomly drawn and is then replaced along with another ball of the same color. Thus, for instance, if of the first r balls drawn, k were white, then the urn at the time of the $(r+1)$th draw would consist of $k+1$ white and $r-k+1$ black, and thus the next ball would be white with probability $(k+1)/(r+2)$. If we identify the drawing of a white ball with a successful trial, then we see that this yields an alternate description of the original model. This latter urn model is called *Polya's urn model*.

Remarks. **(i)** In the special case when $k = r$, Eq. (3.25) is sometimes called Laplace's rule of succession, after the French mathematician Pierre de Laplace. In Laplace's era, this "rule" provoked much controversy, for people attempted to employ it in diverse situations where its validity was questionable. For instance, it was used to justify such propositions as "If you have dined twice at a restaurant and both meals were good, then the next meal also will be good with probability $\frac{3}{4}$," and "Since the sun has risen the past 1,826,213 days, so will it rise tomorrow with probability 1,826,214/1,826,215." The trouble with such claims resides in the fact that it is not at all clear the situation they are describing can be modeled as consisting of independent trials having a common probability of success that is itself uniformly chosen.

(ii) In the original description of the experiment, we referred to the successive trials as being independent, and in fact they are independent when the success probability is known. However, when p is regarded as a random variable, the successive outcomes are no longer independent because knowing whether an outcome is a success or not gives us some information about p, which in turn yields information about the other outcomes.

The preceding can be generalized to situations in which each trial has more than two possible outcomes. Suppose that n independent trials, each resulting in one of m possible outcomes $1, \ldots, m$, with respective probabilities $p_1, \ldots, p_m$ are performed. If

we let X_i denote the number of type i outcomes that result in the n trials, $i = 1, \ldots, m$, then the vector $X_1, \ldots, X_m$ will have the multinomial distribution given by

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m | \mathbf{p}\} = \frac{n!}{x_1! \cdots x_m!} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}$$

where $x_1, \ldots, x_m$ is any vector of nonnegative integers that sum to n. Now let us suppose that the vector $\mathbf{p} = (p_1, \ldots, p_m)$ is not specified, but instead is chosen by a "uniform" distribution. Such a distribution would be of the form

$$f(p_1, \ldots, p_m) = \begin{cases} c, & 0 \leqslant p_i \leqslant 1, i = 1, \ldots, m, \ \sum_1^m p_i = 1 \\ 0, & \text{otherwise} \end{cases}$$

The preceding multivariate distribution is a special case of what is known as the *Dirichlet distribution*, and it is not difficult to show, using the fact that the distribution must integrate to 1, that $c = (m-1)!$.

The unconditional distribution of the vector $\mathbf{X}$ is given by

$$\begin{aligned} P\{X_1 = x_1, \ldots, X_m = x_m\} &= \iint \cdots \int P\{X_1 = x_1, \ldots, X_m = x_m | p_1, \ldots, p_m\} \\ &\times f(p_1, \ldots, p_m)\, dp_1 \cdots dp_m = \frac{(m-1)!n!}{x_1! \cdots x_m!} \iint \cdots \int_{\substack{0 \leqslant p_i \leqslant 1 \\ \sum_1^m p_i = 1}} p_1^{x_1} \cdots p_m^{x_m} dp_1 \cdots dp_m \end{aligned}$$

Now it can be shown that

$$\iint \cdots \int_{\substack{0 \leqslant p_i \leqslant 1 \\ \sum_1^m p_i = 1}} p_1^{x_1} \cdots p_m^{x_m} dp_1 \cdots dp_m = \frac{x_1! \cdots x_m!}{\left(\sum_1^m x_i + m - 1\right)!} \tag{3.26}$$

and thus, using the fact that $\sum_1^m x_i = n$, we see that

$$\begin{aligned} P\{X_1 = x_1, \ldots, X_m = x_m\} &= \frac{n!(m-1)!}{(n+m-1)!} \\ &= \binom{n+m-1}{m-1}^{-1} \end{aligned} \tag{3.27}$$

Hence, all of the $\binom{n+m-1}{m-1}$ possible outcomes (there are $\binom{n+m-1}{m-1}$ possible nonnegative integer valued solutions of $x_1 + \cdots + x_m = n$) of the vector $(X_1, \ldots, X_m)$ are equally likely. The probability distribution given by Eq. (3.27) is sometimes called the *Bose–Einstein distribution*.

To obtain an alternative description of the foregoing, let us compute the conditional probability that the $(n+1)$st outcome is of type j if the first n trials have resulted in x_i type i outcomes, $i = 1, \ldots, m$, $\sum_1^m x_i = n$. This is given by

$$P\{(n+1)\text{st is } j | x_i \text{ type } i \text{ in first } n, i = 1, \ldots, m\}$$

$$= \frac{P\{(n+1)\text{st is } j, x_i \text{ type } i \text{ in first } n, i = 1, \ldots, m\}}{P\{x_i \text{ type } i \text{ in first } n, i = 1, \ldots, m\}}$$

$$= \frac{\dfrac{n!(m-1)!}{x_1! \cdots x_m!} \iint \cdots \int p_1^{x_1} \cdots p_j^{x_j+1} \cdots p_m^{x_m} dp_1 \cdots dp_m}{\binom{n+m-1}{m-1}^{-1}}$$

where the numerator is obtained by conditioning on the **p** vector and the denominator is obtained by using Eq. (3.27). By Eq. (3.26), we have

$$\begin{aligned} &P\{(n+1)\text{st is } j | x_i \text{ type } i \text{ in first } n, i = 1, \ldots, m\} \\ &= \frac{\dfrac{(x_j+1)n!(m-1)!}{(n+m)!}}{\dfrac{(m-1)!n!}{(n+m-1)!}} \\ &= \frac{x_j + 1}{n+m} \end{aligned} \tag{3.28}$$

Using Eq. (3.28), we can now present an urn model description of the stochastic process of successive outcomes. Namely, consider an urn that initially contains one of each of m types of balls. Balls are then randomly drawn and are replaced along with another of the same type. Hence, if in the first n drawings there have been a total of x_j type j balls drawn, then the urn immediately before the $(n+1)$st draw will contain $x_j + 1$ type j balls out of a total of $m + n$, and so the probability of a type j on the $(n+1)$st draw will be given by Eq. (3.28).

Remark. Consider a situation where n particles are to be distributed at random among m possible regions; and suppose that the regions appear, at least before the experiment, to have the same physical characteristics. It would thus seem that the most likely distribution for the number of particles that fall into each of the regions is the multinomial distribution with $p_i \equiv 1/m$. (This, of course, would correspond to each particle, independent of the others, being equally likely to fall in any of the m regions.) Physicists studying how particles distribute themselves observed the behavior of such particles as photons and atoms containing an even number of elementary particles. However, when they studied the resulting data, they were amazed to discover that the observed frequencies did not follow the multinomial distribution but rather seemed to follow the Bose–Einstein distribution. They were amazed because they could not imagine a physical model for the distribution of particles that would result in all possible outcomes being equally likely. (For instance, if 10 particles are to distribute themselves between two regions, it hardly seems reasonable that it is just as likely that both regions will contain 5 particles as it is that all 10 will fall in region 1 or that all 10 will fall in region 2.)

However, from the results of this section we now have a better understanding of the cause of the physicists' dilemma. In fact, two possible hypotheses present themselves.

First, it may be that the data gathered by the physicists were actually obtained under a variety of different situations, each having its own characteristic **p** vector that gave rise to a uniform spread over all possible **p** vectors. A second possibility (suggested by the urn model interpretation) is that the particles select their regions sequentially and a given particle's probability of falling in a region is roughly proportional to the fraction of the landed particles that are in that region. (In other words, the particles presently in a region provide an "attractive" force on elements that have not yet landed.)

3.6.4 Mean Time for Patterns

Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a sequence of independent and identically distributed discrete random variables such that

$$p_i = P\{X_j = i\}$$

For a given subsequence, or *pattern*, $i_1, \ldots, i_n$ let $T = T(i_1, \ldots, i_n)$ denote the number of random variables that we need to observe until the pattern appears. For instance, if the subsequence of interest is $3, 5, 1$ and the sequence is $\mathbf{X} = (5, 3, 1, 3, 5, 3, 5, 1, 6, 2, \ldots)$ then $T = 8$. We want to determine $E[T]$.

To begin, let us consider whether the pattern has an overlap, where we say that the pattern $i_1, i_2, \ldots, i_n$ has an overlap if for some k, $1 \le k < n$, the sequence of its final k elements is the same as that of its first k elements. That is, it has an overlap if for some $1 \le k < n$,

$$(i_{n-k+1}, \ldots, i_n) = (i_1, \ldots, i_k), \quad k < n$$

For instance, the pattern $3, 5, 1$ has no overlaps, whereas the pattern $3, 3, 3$ does.

Case 1. The pattern has no overlaps.
In this case we will argue that T will equal $j + n$ if and only if the pattern does not occur within the first j values, and the next n values are $i_1, \ldots, i_n$.
That is,

$$T = j + n \Leftrightarrow \{T > j, (X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)\} \tag{3.29}$$

To verify (3.29), note first that $T = j + n$ clearly implies both that $T > j$ and that $(X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)$. On the other hand, suppose that

$$T > j \quad \text{and} \quad (X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n) \tag{3.30}$$

Let $k < n$. Because $(i_1, \ldots, i_k) \ne (i_{n-k+1}, \ldots, i_n)$, it follows that $T \ne j + k$. But (3.30) implies that $T \le j + n$, so we can conclude that $T = j + n$. Thus we have verified (3.29).

Using (3.29), we see that

$$P\{T = j + n\} = P\{T > j, (X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)\}$$

However, whether $T > j$ is determined by the values $X_1, \ldots, X_j$, and is thus independent of $X_{j+1}, \ldots, X_{j+n}$. Consequently,

$$\begin{aligned} P\{T = j+n\} &= P\{T > j\} P\{(X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)\} \\ &= P\{T > j\} p \end{aligned}$$

where

$$p = p_{i_1} p_{i_2} \cdots p_{i_n}$$

Summing both sides of the preceding over all j yields

$$1 = \sum_{j=0}^{\infty} P\{T = j+n\} = p \sum_{j=0}^{\infty} P\{T > j\} = pE[T]$$

or

$$E[T] = \frac{1}{p}$$

Case 2. The pattern has overlaps.
For patterns having overlaps there is a simple trick that will enable us to obtain $E[T]$ by making use of the result for nonoverlapping patterns. To make the analysis more transparent, consider a specific pattern, say $\mathbf{P} = (3, 5, 1, 3, 5)$. Let x be a value that does not appear in the pattern, and let T_x denote the time until the pattern $\mathbf{P}_x = (3, 5, 1, 3, 5, x)$ appears. That is, T_x is the time of occurrence of the new pattern that puts x at the end of the original pattern. Because x did not appear in the original pattern it follows that the new pattern has no overlaps; thus,

$$E[T_x] = \frac{1}{p_x p}$$

where $p = \prod_{j=1}^{n} p_{i_j} = p_3^2 p_5^2 p_1$. Because the new pattern can occur only after the original one, write

$$T_x = T + A$$

where T is the time at which the pattern $\mathbf{P} = (3, 5, 1, 3, 5)$ occurs, and A is the additional time after the occurrence of the pattern $\mathbf{P}$ until $\mathbf{P}_x$ occurs. Also, let $E[T_x | i_1, \ldots i_r]$ denote the expected additional time after time r until the pattern $\mathbf{P}_x$ appears given that the first r data values are $i_1, \ldots, i_r$. Conditioning on X, the next data value after the occurrence of the pattern $(3, 5, 1, 3, 5)$, gives

$$E[A|X = i] = \begin{cases} 1 + E[T_x|3, 5, 1], & \text{if } i = 1 \\ 1 + E[T_x|3], & \text{if } i = 3 \\ 1, & \text{if } i = x \\ 1 + E[T_x], & \text{if } i \neq 1, 3, x \end{cases}$$

Therefore,

$$\begin{aligned} E[T_x] &= E[T] + E[A] \\ &= E[T] + 1 + E[T_x|3,5,1]p_1 + E[T_x|3]p_3 \\ &\quad + E[T_x](1 - p_1 - p_3 - p_x) \end{aligned} \tag{3.31}$$

But

$$E[T_x] = E[T(3,5,1)] + E[T_x|3,5,1]$$

giving

$$E[T_x|3,5,1] = E[T_x] - E[T(3,5,1)]$$

Similarly,

$$E[T_x|3] = E[T_x] - E[T(3)]$$

Substituting back into Eq. (3.31) gives

$$p_x E[T_x] = E[T] + 1 - p_1 E[T(3,5,1)] - p_3 E[T(3)]$$

But, by the result in the nonoverlapping case,

$$E[T(3,5,1)] = \frac{1}{p_3 p_5 p_1}, \quad E[T(3)] = \frac{1}{p_3}$$

yielding the result

$$E[T] = p_x E[T_x] + \frac{1}{p_3 p_5} = \frac{1}{p} + \frac{1}{p_3 p_5}$$

For another illustration of the technique, let us reconsider Example 3.15, which is concerned with finding the expected time until n consecutive successes occur in independent Bernoulli trials. That is, we want $E[T]$, when the pattern is $\mathbf{P} = (1, 1, \ldots, 1)$. Then, with $x \neq 1$ we consider the nonoverlapping pattern $\mathbf{P}_x = (1, \ldots, 1, x)$, and let T_x be its occurrence time. With A and X as previously defined, we have

$$E[A|X = i] = \begin{cases} 1 + E[A], & \text{if } i = 1 \\ 1, & \text{if } i = x \\ 1 + E[T_x], & \text{if } i \neq 1, x \end{cases}$$

Therefore,

$$E[A] = 1 + E[A]p_1 + E[T_x](1 - p_1 - p_x)$$

or

$$E[A] = \frac{1}{1 - p_1} + E[T_x]\frac{1 - p_1 - p_x}{1 - p_1}$$

Consequently,

$$\begin{aligned} E[T] &= E[T_x] - E[A] \\ &= \frac{p_x E[T_x] - 1}{1 - p_1} \\ &= \frac{(1/p_1)^n - 1}{1 - p_1} \end{aligned}$$

where the final equality used that $E[T_x] = \frac{1}{p_1^n p_x}$.

The mean occurrence time of any overlapping pattern $\mathbf{P} = (i_1, \ldots, i_n)$ can be obtained by the preceding method. Namely, let T_x be the time until the nonoverlapping pattern $\mathbf{P}_x = (i_1, \ldots, i_n, x)$ occurs; then use the identity

$$E[T_x] = E[T] + E[A]$$

to relate $E[T]$ and $E[T_x] = \frac{1}{p\, p_x}$; then condition on the next data value after $\mathbf{P}$ occurs to obtain an expression for $E[A]$ in terms of quantities of the form

$$E[T_x | i_1, \ldots, i_r] = E[T_x] - E[T(i_1, \ldots, i_r)]$$

If $(i_1, \ldots, i_r)$ is nonoverlapping, use the nonoverlapping result to obtain $E[T(i_1, \ldots, i_r)]$; otherwise, repeat the process on the subpattern $(i_1, \ldots, i_r)$.

Remark. We can utilize the preceding technique even when the pattern $i_1, \ldots, i_n$ includes all the distinct data values. For instance, in coin tossing the pattern of interest might be h, t, h. Even in such cases, we should let x be a data value that is not in the pattern and use the preceding technique (even though $p_x = 0$). Because p_x will appear only in the final answer in the expression $p_x E[T_x] = \frac{p_x}{p_x p}$, by interpreting this fraction as $1/p$ we obtain the correct answer. (A rigorous approach, yielding the same result, would be to reduce one of the positive p_i by ϵ, take $p_x = \epsilon$, solve for $E[T]$, and then let ϵ go to 0.) ■

3.6.5 The k-Record Values of Discrete Random Variables

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables whose set of possible values is the positive integers, and let $P\{X = j\}$, $j \geqslant 1$, denote their common probability mass function. Suppose that these random variables are observed in sequence, and say that X_n is a *k-record value* if

$$X_i \geqslant X_n \quad \text{for exactly } k \text{ of the values } i, i = 1, \ldots, n$$

That is, the nth value in the sequence is a k-record value if exactly k of the first n values (including X_n) are at least as large as it. Let $\mathbf{R}_k$ denote the ordered set of k-record values.

It is a rather surprising result that not only do the sequences of k-record values have *the same probability distributions* for all k, these sequences are also *independent* of each other. This result is known as Ignatov's theorem.

Theorem 3.1 (Ignatov's Theorem). $\mathbf{R}_k, k \geqslant 1$, *are independent and identically distributed random vectors.*

Proof. Define a series of subsequences of the data sequence $X_1, X_2, \ldots$ by letting the ith subsequence consist of all data values that are at least as large as $i, i \geqslant 1$. For instance, if the data sequence is

$$2, 5, 1, 6, 9, 8, 3, 4, 1, 5, 7, 8, 2, 1, 3, 4, 2, 5, 6, 1, \ldots$$

then the subsequences are as follows:

$$\begin{aligned}
&\geqslant 1: \quad 2, 5, 1, 6, 9, 8, 3, 4, 1, 5, 7, 8, 2, 1, 3, 4, 2, 5, 6, 1, \ldots \\
&\geqslant 2: \quad 2, 5, 6, 9, 8, 3, 4, 5, 7, 8, 2, 3, 4, 2, 5, 6, \ldots \\
&\geqslant 3: \quad 5, 6, 9, 8, 3, 4, 5, 7, 8, 3, 4, 5, 6, \ldots
\end{aligned}$$

and so on.

Let X_j^i be the jth element of subsequence i. That is, X_j^i is the jth data value that is at least as large as i. An important observation is that i is a k-record value if and only if $X_k^i = i$. That is, i will be a k-record value if and only if the kth value to be at least as large as i is equal to i. (For instance, for the preceding data, since the fifth value to be at least as large as 3 is equal to 3 it follows that 3 is a five-record value.) Now, it is not difficult to see that, independent of which values in the first subsequence are equal to 1, the values in the second subsequence are independent and identically distributed according to the mass function

$$P\{\text{value in second subsequence} = j\} = P\{X = j | X \geqslant 2\}, \quad j \geqslant 2$$

Similarly, independent of which values in the first subsequence are equal to 1 and which values in the second subsequence are equal to 2, the values in the third subsequence are independent and identically distributed according to the mass function

$$P\{\text{value in third subsequence} = j\} = P\{X = j | X \geqslant 3\}, \quad j \geqslant 3$$

and so on. It therefore follows that the events $\{X_j^i = i\}, i \geqslant 1, j \geqslant 1$, are independent and

$$P\{i \text{ is a } k\text{-record value}\} = P\{X_k^i = i\} = P\{X = i | X \geqslant i\}$$

It now follows from the independence of the events $\{X_k^i = i\}, i \geqslant 1$, and the fact that $P\{i \text{ is a } k\text{-record value}\}$ does not depend on k, that $\mathbf{R}_k$ has the same distribution for all $k \geqslant 1$. In addition, it follows from the independence of the events $\{X_k^i = 1\}$, that the random vectors $\mathbf{R}_k, k \geqslant 1$, are also independent. ■

Suppose now that the $X_i, i \geqslant 1$ are independent finite-valued random variables with probability mass function

$$p_i = P\{X = i\}, \quad i = 1, \ldots, m$$

and let

$$T = \min\{n : X_i \geqslant X_n \text{ for exactly } k \text{ of the values } i, i = 1, \ldots, n\}$$

denote the first *k-record index*. We will now determine its mean.

Proposition 3.2. *Let* $\lambda_i = p_i / \sum_{j=i}^{m} p_j,\ i = 1, \ldots, m$. *Then*

$$E[T] = k + (k-1) \sum_{i=1}^{m-1} \lambda_i$$

Proof. To begin, suppose that the observed random variables $X_1, X_2, \ldots$ take on one of the values $i, i+1, \ldots, m$ with respective probabilities

$$P\{X = j\} = \frac{p_j}{p_i + \cdots + p_m}, \quad j = i, \ldots, m$$

Let T_i denote the first k-record index when the observed data have the preceding mass function, and note that since the each data value is at least i it follows that the k-record value will equal i, and T_i will equal k, if $X_k = i$. As a result,

$$E[T_i | X_k = i] = k$$

On the other hand, if $X_k > i$ then the k-record value will exceed i, and so all data values equal to i can be disregarded when searching for the k-record value. In addition, since each data value greater than i will have probability mass function

$$P\{X = j | X > i\} = \frac{p_j}{p_{i+1} + \cdots + p_m}, \quad j = i+1, \ldots, m$$

it follows that the total number of data values greater than i that need be observed until a k-record value appears has the same distribution as T_{i+1}. Hence,

$$E[T_i | X_k > i] = E[T_{i+1} + N_i | X_k > i]$$

where T_{i+1} is the total number of variables greater than i that we need observe to obtain a k-record, and N_i is the number of values equal to i that are observed in that time. Now, given that $X_k > i$ and that $T_{i+1} = n$ $(n \geqslant k)$ it follows that the time to observe T_{i+1} values greater than i has the same distribution as the number of trials to obtain n successes given that trial k is a success and that each trial is independently a success with probability $1 - p_i / \sum_{j \geqslant i} p_j = 1 - \lambda_i$. Thus, since the number of trials needed to obtain a success is a geometric random variable with mean $1/(1-\lambda_i)$, we see that

$$E[T_i | T_{i+1}, X_k > i] = 1 + \frac{T_{i+1} - 1}{1 - \lambda_i} = \frac{T_{i+1} - \lambda_i}{1 - \lambda_i}$$

Taking expectations gives

$$E[T_i|X_k > i] = E\left[\frac{T_{i+1} - \lambda_i}{1 - \lambda_i}\middle| X_k > i\right] = \frac{E[T_{i+1}] - \lambda_i}{1 - \lambda_i}$$

Thus, upon conditioning on whether $X_k = i$, we obtain

$$\begin{aligned} E[T_i] &= E[T_i|X_k = i]\lambda_i + E[T_i|X_k > i](1 - \lambda_i) \\ &= (k-1)\lambda_i + E[T_{i+1}] \end{aligned}$$

Starting with $E[T_m] = k$, the preceding gives

$$\begin{aligned} E[T_{m-1}] &= (k-1)\lambda_{m-1} + k \\ E[T_{m-2}] &= (k-1)\lambda_{m-2} + (k-1)\lambda_{m-1} + k \\ &= (k-1)\sum_{j=m-2}^{m-1} \lambda_j + k \\ E[T_{m-3}] &= (k-1)\lambda_{m-3} + (k-1)\sum_{j=m-2}^{m-1} \lambda_j + k \\ &= (k-1)\sum_{j=m-3}^{m-1} \lambda_j + k \end{aligned}$$

In general,

$$E[T_i] = (k-1)\sum_{j=i}^{m-1} \lambda_j + k$$

and the result follows since $T = T_1$. ■

3.6.6 Left Skip Free Random Walks

Let $X_i, i \geq 1$ be independent and identically distributed random variables. Let $P_j = P(X_i = j)$ and suppose that

$$\sum_{j=-1}^{\infty} P_j = 1$$

That is, the possible values of the X_i are $-1, 0, 1, \ldots$. If we take

$$S_0 = 0, \quad S_n = \sum_{i=1}^{n} X_i$$

then the sequence of random variables $S_n, n \geq 0$ is called a *left skip free random walk*. (It is called left skip free because S_n can decrease from S_{n-1} by at most 1.)

For an application consider a gambler who makes a sequence of identical bets, for which he can lose at most 1 on each bet. Then if X_i represents the gambler's winnings on bet i, then S_n would represent his total winnings after the first n bets.

Suppose that the gambler is playing in an unfair game, in the sense that $E[X_i] < 0$, and let $v = -E[X_i]$. Also, let $T_0 = 0$, and for $k > 0$, let T_{-k} denote the number of bets until the gambler is losing k. That is,

$$T_{-k} = \min\{n : S_n = -k\}$$

It should be noted that $T_{-k} < \infty$; that is, the random walk will eventually hit $-k$. This is so because, by the strong law of large numbers, $S_n/n \to E[X_i] < 0$, which implies that $S_n \to -\infty$. We are interested in determining $E[T_{-k}]$ and $\text{Var}(T_{-k})$. (It can be shown that both are finite when $E[X_i] < 0$.)

The key to the analysis is to note that the number of bets until one's fortune decreases by k can be expressed as the number of bets until it decreases by 1 (namely, T_{-1}), plus the additional number of bets after the decrease is 1 until the total decrease is 2 (namely, $T_{-2} - T_{-1}$), plus the additional number of bets after the decrease is 2 until it is 3 (namely, $T_{-3} - T_{-2}$), and so on. That is,

$$T_{-k} = T_{-1} + \sum_{j=2}^{k} \left(T_{-j} - T_{-(j-1)}\right)$$

However, because the results of all bets are independent and identically distributed, it follows that $T_{-1}, T_{-2} - T_{-1}, T_{-3} - T_{-2}, \ldots, T_{-k} - T_{-(k-1)}$ are all independent and identically distributed. (That is, starting at any instant, the number of additional bets until the gambler's fortune is one less than it is at that instant is independent of prior results and has the same distribution as T_{-1}.) Consequently, the mean and variance of T_{-k}, the sum of these k random variables, are

$$E[T_{-k}] = kE[T_{-1}]$$

and

$$\text{Var}(T_{-k}) = k\text{Var}(T_{-1})$$

We now compute the mean and variance of T_{-1} by conditioning on X_1, the result of the initial bet. Now, given X_1, T_{-1} is equal to 1 plus the number of bets it takes until the gambler's fortune decreases by $X_1 + 1$ from what it is after the initial bet. Consequently, given X_1, T_{-1} has the same distribution as $1 + T_{-(X_1+1)}$. Hence,

$$E[T_{-1}|X_1] = 1 + E[T_{-(X_1+1)}] = 1 + (X_1 + 1)E[T_{-1}]$$
$$\text{Var}(T_{-1}|X_1) = \text{Var}(T_{-(X_1+1)}) = (X_1 + 1)\text{Var}(T_{-1})$$

Consequently,

$$E[T_{-1}] = E\left[E[T_{-1}|X_1]\right] = 1 + (-v+1)E[T_{-1}]$$

or

$$E[T_{-1}] = \frac{1}{v}$$

which shows that

$$E[T_{-k}] = \frac{k}{v} \tag{3.32}$$

Similarly, with $\sigma^2 = \text{Var}(X_1)$, the conditional variance formula yields

$$\begin{aligned}\text{Var}(T_{-1}) &= E[(X_1+1)\text{Var}(T_{-1})] + \text{Var}(X_1 E[T_{-1}]) \\ &= (1-v)\text{Var}(T_{-1}) + (E[T_{-1}])^2\sigma^2 \\ &= (1-v)\text{Var}(T_{-1}) + \frac{\sigma^2}{v^2}\end{aligned}$$

thus showing that

$$\text{Var}(T_{-1}) = \frac{\sigma^2}{v^3}$$

and yielding the result

$$\text{Var}(T_{-k}) = \frac{k\sigma^2}{v^3} \tag{3.33}$$

There are many interesting results about skip free random walks. For instance, the *hitting time theorem*.

Proposition 3.3 (The Hitting Time Theorem)**.**

$$P(T_{-k} = n) = \frac{k}{n} P(S_n = -k), \quad n \geq 1$$

Proof. The proof is by induction on n. Now, when $n = 1$ we must prove

$$P(T_{-k} = 1) = k P(S_1 = -k)$$

However, the preceding is true when $k = 1$ because

$$P(T_{-1} = 1) = P(S_1 = -1) = P_{-1}$$

and it is true when $k > 1$ because

$$P(T_{-k} = 1) = 0 = P(S_1 = -k), \quad k > 1$$

Thus the result is true when $n = 1$. So assume that for a fixed value $n > 1$ and all $k > 0$

$$P(T_{-k} = n - 1) = \frac{k}{n-1} P(S_{n-1} = -k) \tag{3.34}$$

Now consider $P(T_{-k} = n)$. Conditioning on X_1 yields

$$P(T_{-k} = n) = \sum_{j=-1}^{\infty} P(T_{-k} = n | X_1 = j) P_j$$

Now, if the gambler wins j on his initial bet, then the first time that he is down k will occur after bet n if the first time that his cumulative losses after the initial gamble is $k + j$ occurs after an additional $n - 1$ bets. That is,

$$P(T_{-k} = n | X_1 = j) = P(T_{-(k+j)} = n - 1)$$

Consequently,

$$\begin{aligned} P(T_{-k} = n) &= \sum_{j=-1}^{\infty} P(T_{-k} = n | X_1 = j) P_j \\ &= \sum_{j=-1}^{\infty} P(T_{-(k+j)} = n - 1) P_j \\ &= \sum_{j=-1}^{\infty} \frac{k+j}{n-1} P\{S_{n-1} = -(k+j)\} P_j \end{aligned}$$

where the last equality follows by Induction Hypothesis (3.34). Using that

$$P(S_n = -k | X_1 = j) = P\{S_{n-1} = -(k+j)\}$$

the preceding yields

$$\begin{aligned} P(T_{-k} = n) &= \sum_{j=-1}^{\infty} \frac{k+j}{n-1} P(S_n = -k | X_1 = j) P_j \\ &= \sum_{j=-1}^{\infty} \frac{k+j}{n-1} P(S_n = -k, X_1 = j) \\ &= \sum_{j=-1}^{\infty} \frac{k+j}{n-1} P(X_1 = j | S_n = -k) P(S_n = -k) \\ &= P(S_n = -k) \left\{ \frac{k}{n-1} \sum_{j=-1}^{\infty} P(X_1 = j | S_n = -k) \right. \end{aligned}$$

$$+\frac{1}{n-1}\sum_{j=-1}^{\infty} jP(X_1=j|S_n=-k)\Bigg\}$$
$$=P(S_n=-k)\left\{\frac{k}{n-1}+\frac{1}{n-1}E[X_1|S_n=-k]\right\} \tag{3.35}$$

However,

$$\begin{aligned}-k&=E[S_n|S_n=-k]\\&=E[X_1+\cdots+X_n|S_n=-k]\\&=\sum_{i=1}^{n}E[X_i|S_n=-k]\\&=nE[X_1|S_n=-k]\end{aligned}$$

where the final equation follows because $X_1,\ldots,X_n$ are independent and identically distributed and thus the distribution of X_i given that $X_1+\cdots+X_n=-k$ is the same for all i. Hence,

$$E[X_1|S_n=-k]=-\frac{k}{n}$$

Substituting the preceding into (3.35) gives

$$P(T_{-k}=n)=P(S_n=-k)\left(\frac{k}{n-1}-\frac{1}{n-1}\frac{k}{n}\right)=\frac{k}{n}P(S_n=-k)$$

and completes the proof. ■

Suppose that after n bets the gambler is down k. Then the conditional probability that this is the first time he has ever been down k is

$$\begin{aligned}P(T_{-k}=n|S_n=-k)&=\frac{P(T_{-k}=n,S_n=-k)}{P(S_n=-k)}\\&=\frac{P(T_{-k}=n)}{P(S_n=-k)}\\&=\frac{k}{n}\quad\text{(by the hitting time theorem)}\end{aligned}$$

Let us suppose for the remainder of this section that $-v=E[X]<0$. Combining the hitting time theorem with our previously derived result about $E[T_{-k}]$ gives the following:

$$\begin{aligned}\frac{k}{v}&=E[T_{-k}]\\&=\sum_{n=1}^{\infty}nP(T_{-k}=n)\end{aligned}$$

$$= \sum_{n=1}^{\infty} k P(S_n = -k)$$

where the final equality used the hitting time theorem. Hence,

$$\sum_{n=1}^{\infty} P(S_n = -k) = \frac{1}{v}$$

Let I_n be an indicator variable for the event that $S_n = -k$. That is, let

$$I_n = \begin{cases} 1, & \text{if } S_n = -k \\ 0, & \text{if } S_n \neq -k \end{cases}$$

and note that

$$\text{total time gambler's fortune is } -k = \sum_{n=1}^{\infty} I_n$$

Taking expectations gives

$$E[\text{total time gambler's fortune is } -k] = \sum_{n=1}^{\infty} P(S_n = -k) = \frac{1}{v} \tag{3.36}$$

Now, let α be the probability that the random walk is always negative after the initial movement. That is,

$$\alpha = P(S_n < 0 \text{ for all } n \geq 1)$$

To determine α note that each time the gambler's fortune is $-k$ the probability that it will never again hit $-k$ (because all cumulative winnings starting at that time are negative) is α. Hence, the number of times that the gambler's fortune is $-k$ is a geometric random variable with parameter α, and thus has mean $1/\alpha$. Consequently, from (3.36)

$$\alpha = v$$

Let us now define L_{-k} to equal the last time that the random walk hits $-k$. Because L_{-k} will equal n if $S_n = -k$ and the sequence of cumulative winnings from time n onwards is always negative, we see that

$$P(L_{-k} = n) = P(S_n = -k)\alpha = P(S_n = -k)v$$

Hence,

$$E[L_{-k}] = \sum_{n=0}^{\infty} n P(L_{-k} = n)$$

$$
\begin{aligned}
&= v \sum_{n=0}^{\infty} n P(S_n = -k) \\
&= v \sum_{n=0}^{\infty} n \frac{n}{k} P(T_{-k} = n) \quad \text{by the hitting time theorem} \\
&= \frac{v}{k} \sum_{n=0}^{\infty} n^2 P(T_{-k} = n) \\
&= \frac{v}{k} E[T_{-k}^2] \\
&= \frac{v}{k} \{E^2[T_{-k}] + \text{Var}(T_{-k})\} \\
&= \frac{k}{v} + \frac{\sigma^2}{v^2}
\end{aligned}
$$

3.7 An Identity for Compound Random Variables

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, and let $S_n = \sum_{i=1}^{n} X_i$ be the sum of the first n of them, $n \geq 0$, where $S_0 = 0$. Recall that if N is a nonnegative integer valued random variable that is independent of the sequence $X_1, X_2, \ldots$ then

$$S_N = \sum_{i=1}^{N} X_i$$

is said to be a *compound random variable*, with the distribution of N called the *compounding distribution*. In this subsection we will first derive an identity involving such random variables. We will then specialize to where the X_i are positive integer valued random variables, prove a corollary of the identity, and then use this corollary to develop a recursive formula for the probability mass function of S_N, for a variety of common compounding distributions.

To begin, let M be a random variable that is independent of the sequence $X_1, X_2, \ldots$, and which is such that

$$P\{M = n\} = \frac{n P\{N = n\}}{E[N]}, \quad n = 1, 2, \ldots$$

Proposition 3.4 (The Compound Random Variable Identity)**.** *For any function h*

$$E[S_N h(S_N)] = E[N] E[X_1 h(S_M)]$$

Proof.

$$
\begin{aligned}
E[S_N h(S_N)] &= E\left[\sum_{i=1}^{N} X_i h(S_N)\right] \\
&= \sum_{n=0}^{\infty} E\left[\sum_{i=1}^{N} X_i h(S_N) | N = n\right] P\{N = n\} \\
&\qquad \text{(by conditioning on } N\text{)} \\
&= \sum_{n=0}^{\infty} E\left[\sum_{i=1}^{n} X_i h(S_n) | N = n\right] P\{N = n\} \\
&= \sum_{n=0}^{\infty} E\left[\sum_{i=1}^{n} X_i h(S_n)\right] P\{N = n\} \\
&\qquad \text{(by independence of } N \text{ and } X_1, \ldots, X_n\text{)} \\
&= \sum_{n=0}^{\infty} \sum_{i=1}^{n} E\big[X_i h(S_n)\big] P\{N = n\}
\end{aligned}
$$

Now, because $X_1, \ldots, X_n$ are independent and identically distributed, and $h(S_n) = h(X_1 + \cdots + X_n)$ is a symmetric function of $X_1, \ldots, X_n$, it follows that the distribution of $X_i h(S_n)$ is the same for all $i = 1, \ldots, n$. Therefore, continuing the preceding string of equalities yields

$$
\begin{aligned}
E[S_N h(S_N)] &= \sum_{n=0}^{\infty} n E[X_1 h(S_n)] P\{N = n\} \\
&= E[N] \sum_{n=0}^{\infty} E[X_1 h(S_n)] P\{M = n\} \quad \text{(definition of } M\text{)} \\
&= E[N] \sum_{n=0}^{\infty} E[X_1 h(S_n) | M = n] P\{M = n\} \\
&\qquad \text{(independence of } M \text{ and } X_1, \ldots, X_n\text{)} \\
&= E[N] \sum_{n=0}^{\infty} E[X_1 h(S_M) | M = n] P\{M = n\} \\
&= E[N] E[X_1 h(S_M)]
\end{aligned}
$$

which proves the proposition. ■

Suppose now that the X_i are positive integer valued random variables, and let

$$\alpha_j = P\{X_1 = j\}, \quad j > 0$$

The successive values of $P\{S_N = k\}$ can often be obtained from the following corollary to Proposition 3.4.

Corollary 3.5.

$$P\{S_N = 0\} = P\{N = 0\}$$

$$P\{S_N = k\} = \frac{1}{k} E[N] \sum_{j=1}^{k} j\alpha_j P\{S_{M-1} = k - j\}, \quad k > 0$$

Proof. For k fixed, let

$$h(x) = \begin{cases} 1, & \text{if } x = k \\ 0, & \text{if } x \neq k \end{cases}$$

and note that $S_N h(S_N)$ is either equal to k if $S_N = k$ or is equal to 0 otherwise. Therefore,

$$E[S_N h(S_N)] = k P\{S_N = k\}$$

and the compound identity yields

$$\begin{aligned} k P\{S_N = k\} &= E[N] E[X_1 h(S_M)] \\ &= E[N] \sum_{j=1}^{\infty} E[X_1 h(S_M))|X_1 = j]\alpha_j \\ &= E[N] \sum_{j=1}^{\infty} j E[h(S_M)|X_1 = j]\alpha_j \\ &= E[N] \sum_{j=1}^{\infty} j P\{S_M = k|X_1 = j\}\alpha_j \end{aligned} \tag{3.37}$$

Now,

$$\begin{aligned} P\{S_M = k|X_1 = j\} &= P\left\{\sum_{i=1}^{M} X_i = k \,|X_1 = j\right\} \\ &= P\left\{j + \sum_{i=2}^{M} X_i = k \,|X_1 = j\right\} \\ &= P\left\{j + \sum_{i=2}^{M} X_i = k\right\} \\ &= P\left\{j + \sum_{i=1}^{M-1} X_i = k\right\} \end{aligned}$$

$$= P\{S_{M-1} = k - j\}$$

The next to last equality followed because $X_2, \ldots, X_M$ and $X_1, \ldots, X_{M-1}$ have the same joint distribution; namely that of $M - 1$ independent random variables that all have the distribution of X_1, where $M - 1$ is independent of these random variables. Thus the proof follows from Eq. (3.37). ■

When the distributions of $M - 1$ and N are related, the preceding corollary can be a useful recursion for computing the probability mass function of S_N, as is illustrated in the following subsections.

3.7.1 Poisson Compounding Distribution

If N is the Poisson distribution with mean λ, then

$$\begin{aligned} P\{M - 1 = n\} &= P\{M = n + 1\} \\ &= \frac{(n+1)P\{N = n+1\}}{E[N]} \\ &= \frac{1}{\lambda}(n+1)e^{-\lambda}\frac{\lambda^{n+1}}{(n+1)!} \\ &= e^{-\lambda}\frac{\lambda^n}{n!} \end{aligned}$$

Consequently, $M - 1$ is also Poisson with mean λ. Thus, with

$$P_n = P\{S_N = n\}$$

the recursion given by Corollary 3.5 can be written

$$P_0 = e^{-\lambda}$$

$$P_k = \frac{\lambda}{k}\sum_{j=1}^{k} j\alpha_j P_{k-j}, \quad k > 0$$

Remark. When the X_i are identically 1, the preceding recursion reduces to the well-known identity for a Poisson random variable having mean λ:

$$P\{N = 0\} = e^{-\lambda}$$

$$P\{N = n\} = \frac{\lambda}{n}P\{N = n - 1\}, \quad n \geq 1$$

Example 3.35. Let S be a compound Poisson random variable with $\lambda = 4$ and

$$P\{X_i = i\} = 1/4, \quad i = 1, 2, 3, 4$$

Let us use the recursion given by Corollary 3.5 to determine $P\{S = 5\}$. It gives

$$P_0 = e^{-\lambda} = e^{-4}$$
$$P_1 = \lambda\alpha_1 P_0 = e^{-4}$$
$$P_2 = \frac{\lambda}{2}(\alpha_1 P_1 + 2\alpha_2 P_0) = \frac{3}{2}e^{-4}$$
$$P_3 = \frac{\lambda}{3}(\alpha_1 P_2 + 2\alpha_2 P_1 + 3\alpha_3 P_0) = \frac{13}{6}e^{-4}$$
$$P_4 = \frac{\lambda}{4}(\alpha_1 P_3 + 2\alpha_2 P_2 + 3\alpha_3 P_1 + 4\alpha_4 P_0) = \frac{73}{24}e^{-4}$$
$$P_5 = \frac{\lambda}{5}(\alpha_1 P_4 + 2\alpha_2 P_3 + 3\alpha_3 P_2 + 4\alpha_4 P_1 + 5\alpha_5 P_0) = \frac{381}{120}e^{-4}$$ ■

3.7.2 Binomial Compounding Distribution

Suppose that N is a binomial random variable with parameters r and p. Then,

$$\begin{aligned} P\{M-1=n\} &= \frac{(n+1)P\{N=n+1\}}{E[N]} \\ &= \frac{n+1}{rp}\binom{r}{n+1}p^{n+1}(1-p)^{r-n-1} \\ &= \frac{n+1}{rp}\frac{r!}{(r-1-n)!(n+1)!}p^{n+1}(1-p)^{r-1-n} \\ &= \frac{(r-1)!}{(r-1-n)!n!}p^n(1-p)^{r-1-n} \end{aligned}$$

Thus, $M-1$ is a binomial random variable with parameters $r-1$, p.

Fixing p, let $N(r)$ be a binomial random variable with parameters r and p, and let

$$P_r(k) = P\{S_{N(r)} = k\}$$

Then, Corollary 3.5 yields

$$P_r(0) = (1-p)^r$$
$$P_r(k) = \frac{rp}{k}\sum_{j=1}^{k} j\alpha_j P_{r-1}(k-j), \quad k > 0$$

For instance, letting k equal 1, then 2, and then 3 gives

$$P_r(1) = rp\alpha_1(1-p)^{r-1}$$
$$\begin{aligned} P_r(2) &= \frac{rp}{2}[\alpha_1 P_{r-1}(1) + 2\alpha_2 P_{r-1}(0)] \\ &= \frac{rp}{2}\left[(r-1)p\alpha_1^2(1-p)^{r-2} + 2\alpha_2(1-p)^{r-1}\right] \end{aligned}$$

$$
\begin{aligned}
P_r(3) &= \frac{rp}{3}\left[\alpha_1 P_{r-1}(2) + 2\alpha_2 P_{r-1}(1) + 3\alpha_3 P_{r-1}(0)\right] \\
&= \frac{\alpha_1 rp}{3}\frac{(r-1)p}{2}\left[(r-2)p\alpha_1^2(1-p)^{r-3} + 2\alpha_2(1-p)^{r-2}\right] \\
&\quad + \frac{2\alpha_2 rp}{3}(r-1)p\alpha_1(1-p)^{r-2} + \alpha_3 rp(1-p)^{r-1}
\end{aligned}
$$

3.7.3 A Compounding Distribution Related to the Negative Binomial

Suppose, for a fixed value of p, $0 < p < 1$, the compounding random variable N has a probability mass function

$$
P\{N = n\} = \binom{n+r-1}{r-1} p^r (1-p)^n, \quad n = 0, 1, \ldots
$$

Such a random variable can be thought of as being the number of failures that occur before a total of r successes have been amassed when each trial is independently a success with probability p. (There will be n such failures if the rth success occurs on trial $n + r$. Consequently, $N + r$ is a negative binomial random variable with parameters r and p.) Using that the mean of the negative binomial random variable $N + r$ is $E[N + r] = r/p$, we see that $E[N] = r\frac{1-p}{p}$.

Regard p as fixed, and call N an NB(r) random variable. The random variable $M - 1$ has probability mass function

$$
\begin{aligned}
P\{M - 1 = n\} &= \frac{(n+1)P\{N = n+1\}}{E[N]} \\
&= \frac{(n+1)p}{r(1-p)}\binom{n+r}{r-1} p^r (1-p)^{n+1} \\
&= \frac{(n+r)!}{r!n!} p^{r+1}(1-p)^n \\
&= \binom{n+r}{r} p^{r+1}(1-p)^n
\end{aligned}
$$

In other words, $M - 1$ is an NB($r + 1$) random variable.

Letting, for an NB(r) random variable N,

$$
P_r(k) = P\{S_N = k\}
$$

Corollary 3.5 yields

$$
\begin{aligned}
P_r(0) &= p^r \\
P_r(k) &= \frac{r(1-p)}{kp}\sum_{j=1}^{k} j\alpha_j P_{r+1}(k-j), \quad k > 0
\end{aligned}
$$

Thus,

$$
\begin{aligned}
P_r(1) &= \frac{r(1-p)}{p}\alpha_1 P_{r+1}(0) \\
&= rp^r(1-p)\alpha_1, \\
P_r(2) &= \frac{r(1-p)}{2p}\big[\alpha_1 P_{r+1}(1) + 2\alpha_2 P_{r+1}(0)\big] \\
&= \frac{r(1-p)}{2p}\big[\alpha_1^2(r+1)p^{r+1}(1-p) + 2\alpha_2 p^{r+1}\big], \\
P_r(3) &= \frac{r(1-p)}{3p}\big[\alpha_1 P_{r+1}(2) + 2\alpha_2 P_{r+1}(1) + 3\alpha_3 P_{r+1}(0)\big]
\end{aligned}
$$

and so on.

Exercises

1. If X and Y are both discrete, show that $\sum_x p_{X|Y}(x|y) = 1$ for all y such that $p_Y(y) > 0$.

***2.** Let X_1 and X_2 be independent geometric random variables having the same parameter p. Guess the value of

$$P\{X_1 = i | X_1 + X_2 = n\}$$

Hint: Suppose a coin having probability p of coming up heads is continually flipped. If the second head occurs on flip number n, what is the conditional probability that the first head was on flip number $i, i = 1, \ldots, n-1$? Verify your guess analytically.

3. The joint probability mass function of X and Y, $p(x, y)$, is given by

$$
\begin{aligned}
&p(1,1) = \tfrac{1}{9}, \quad p(2,1) = \tfrac{1}{3}, \quad p(3,1) = \tfrac{1}{9}, \\
&p(1,2) = \tfrac{1}{9}, \quad p(2,2) = 0, \quad p(3,2) = \tfrac{1}{18}, \\
&p(1,3) = 0, \quad p(2,3) = \tfrac{1}{6}, \quad p(3,3) = \tfrac{1}{9}
\end{aligned}
$$

Compute $E[X|Y = i]$ for $i = 1, 2, 3$.

4. In Exercise 3, are the random variables X and Y independent?

5. An urn contains three white, six red, and five black balls. Six of these balls are randomly selected from the urn. Let X and Y denote respectively the number of white and black balls selected. Compute the conditional probability mass function of X given that $Y = 3$. Also compute $E[X|Y = 1]$.

***6.** Repeat Exercise 5 but under the assumption that when a ball is selected its color is noted, and it is then replaced in the urn before the next selection is made.

7. Suppose $p(x, y, z)$, the joint probability mass function of the random variables X, Y, and Z, is given by

$$p(1,1,1) = \tfrac{1}{8}, \quad p(2,1,1) = \tfrac{1}{4},$$
$$p(1,1,2) = \tfrac{1}{8}, \quad p(2,1,2) = \tfrac{3}{16},$$
$$p(1,2,1) = \tfrac{1}{16}, \quad p(2,2,1) = 0,$$
$$p(1,2,2) = 0, \quad p(2,2,2) = \tfrac{1}{4}$$

What is $E[X|Y=2]$? What is $E[X|Y=2, Z=1]$?

8. An unbiased die is successively rolled. Let X and Y denote, respectively, the number of rolls necessary to obtain a six and a five. Find (a) $E[X]$, (b) $E[X|Y=1]$, (c) $E[X|Y=5]$.

9. If Z_1 and Z_2 are independent standard normal random variables, find the conditional density function of Z_1 given that $Z_1 + Z_2 = x$.

10. Let $X_1, \ldots, X_n$ be independent uniform $(0, 1)$ random variables. Find the conditional density of X_1 given that it is not the smallest of the n values.

11. The joint density of X and Y is

$$f(x,y) = \frac{(y^2 - x^2)}{8} e^{-y}, \quad 0 < y < \infty, \quad -y \leqslant x \leqslant y$$

Show that $E[X|Y=y] = 0$.

12. The joint density of X and Y is given by

$$f(x,y) = \frac{e^{-x/y} e^{-y}}{y}, \quad 0 < x < \infty, \quad 0 < y < \infty$$

Show $E[X|Y=y] = y$.

***13.** Let X be exponential with mean $1/\lambda$; that is,

$$f_X(x) = \lambda e^{-\lambda x}, \quad 0 < x < \infty$$

Find $E[X|X > 1]$.

14. Let X be uniform over $(0, 1)$. Find $E[X|X < \frac{1}{2}]$.

15. The joint density of X and Y is given by

$$f(x,y) = \frac{e^{-y}}{y}, \quad 0 < x < y, \quad 0 < y < \infty$$

Compute $E[X^2|Y=y]$.

16. The random variables X and Y are said to have a bivariate normal distribution if their joint density function is given by

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \times \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

for $-\infty < x < \infty, -\infty < y < \infty$, where $\sigma_x, \sigma_y, \mu_x, \mu_y$, and ρ are constants such that $-1 < \rho < 1, \sigma_x > 0, \sigma_y > 0, -\infty < \mu_x < \infty, -\infty < \mu_y < \infty$.

(a) Show that X is normally distributed with mean μ_x and variance σ_x^2, and Y is normally distributed with mean μ_y and variance σ_y^2.

(b) Show that the conditional density of X given that $Y = y$ is normal with mean $\mu_x + (\rho\sigma_x/\sigma_y)(y - \mu_y)$ and variance $\sigma_x^2(1-\rho^2)$.

The quantity ρ is called the correlation between X and Y. It can be shown that

$$\rho = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x\sigma_y} = \frac{\text{Cov}(X, Y)}{\sigma_x\sigma_y}$$

17. Let Y be a gamma random variable with parameters (s, α). That is, its density is

$$f_Y(y) = Ce^{-\alpha y}y^{s-1}, \quad y > 0$$

where C is a constant that does not depend on y. Suppose also that the conditional distribution of X given that $Y = y$ is Poisson with mean y. That is,

$$P\{X = i | Y = y\} = e^{-y}y^i/i!, \quad i \geqslant 0$$

Show that the conditional distribution of Y given that $X = i$ is the gamma distribution with parameters $(s + i, \alpha + 1)$.

18. Let $X_1, \ldots, X_n$ be independent random variables having a common distribution function that is specified up to an unknown parameter θ. Let $T = T(\mathbf{X})$ be a function of the data $\mathbf{X} = (X_1, \ldots, X_n)$. If the conditional distribution of $X_1, \ldots, X_n$ given $T(\mathbf{X})$ does not depend on θ then $T(\mathbf{X})$ is said to be a *sufficient statistic* for θ. In the following cases, show that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for θ.

(a) The X_i are normal with mean θ and variance 1.

(b) The density of X_i is $f(x) = \theta e^{-\theta x}, x > 0$.

(c) The mass function of X_i is $p(x) = \theta^x(1-\theta)^{1-x}, x = 0, 1, 0 < \theta < 1$.

(d) The X_i are Poisson random variables with mean θ.

***19.** Prove that if X and Y are jointly continuous, then

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y]f_Y(y)\,dy$$

20. There are 3 coins which when flipped come up heads, respectively, with probabilities 1/4, 1/2, 3/4. One of these coins is randomly chosen and continually flipped.

(a) Find the expected number of flips until the first head.
(b) Find the mean number of heads in the first 8 flips.

21. Consider Example 3.12, which refers to a miner trapped in a mine. Let N denote the total number of doors selected before the miner reaches safety. Also, let T_i denote the travel time corresponding to the ith choice, $i \geqslant 1$. Again let X denote the time when the miner reaches safety.
(a) Give an identity that relates X to N and the T_i.
(b) What is $E[N]$?
(c) What is $E[T_N]$?
(d) What is $E[\sum_{i=1}^{N} T_i | N = n]$?
(e) Using the preceding, what is $E[X]$?

22. Suppose that independent trials, each of which is equally likely to have any of m possible outcomes, are performed until the same outcome occurs k consecutive times. If N denotes the number of trials, show that

$$E[N] = \frac{m^k - 1}{m - 1}$$

Some people believe that the successive digits in the expansion of $\pi = 3.14159\ldots$ are "uniformly" distributed. That is, they believe that these digits have all the appearance of being independent choices from a distribution that is equally likely to be any of the digits from 0 through 9. Possible evidence against this hypothesis is the fact that starting with the 24,658,601st digit there is a run of nine successive 7s. Is this information consistent with the hypothesis of a uniform distribution?
To answer this, we note from the preceding that if the uniform hypothesis were correct, then the expected number of digits until a run of nine of the same value occurs is

$$(10^9 - 1)/9 = 111{,}111{,}111$$

Thus, the actual value of approximately 25 million is roughly 22 percent of the theoretical mean. However, it can be shown that under the uniformity assumption the standard deviation of N will be approximately equal to the mean. As a result, the observed value is approximately 0.78 standard deviations less than its theoretical mean and is thus quite consistent with the uniformity assumption.

***23.** A coin having probability p of coming up heads is successively flipped until two of the most recent three flips are heads. Let N denote the number of flips. (Note that if the first two flips are heads, then $N = 2$.) Find $E[N]$.

24. A coin, having probability p of landing heads, is continually flipped until at least one head and one tail have been flipped.
(a) Find the expected number of flips needed.
(b) Find the expected number of flips that land on heads.
(c) Find the expected number of flips that land on tails.

(d) Repeat part (a) in the case where flipping is continued until a total of at least two heads and one tail have been flipped.

25. Independent trials, resulting in one of the outcomes 1, 2, 3 with respective probabilities p_1, p_2, p_3, $\sum_{i=1}^{3} p_i = 1$, are performed.

(a) Let N denote the number of trials needed until the initial outcome has occurred exactly 3 times. For instance, if the trial results are 3, 2, 1, 2, 3, 2, 3 then $N = 7$. Find $E[N]$.

(b) Find the expected number of trials needed until both outcome 1 and outcome 2 have occurred.

26. You have two opponents with whom you alternate play. Whenever you play A, you win with probability p_A; whenever you play B, you win with probability p_B, where $p_B > p_A$. If your objective is to minimize the expected number of games you need to play to win two in a row, should you start with A or with B?

Hint: Let $E[N_i]$ denote the mean number of games needed if you initially play i. Derive an expression for $E[N_A]$ that involves $E[N_B]$; write down the equivalent expression for $E[N_B]$ and then subtract.

27. A coin that comes up heads with probability p is continually flipped until the pattern T, T, H appears. (That is, you stop flipping when the most recent flip lands heads, and the two immediately preceding it lands tails.) Let X denote the number of flips made, and find $E[X]$.

28. Polya's urn model supposes that an urn initially contains r red and b blue balls. At each stage a ball is randomly selected from the urn and is then returned along with m other balls of the same color. Let X_k be the number of red balls drawn in the first k selections.

(a) Find $E[X_1]$.

(b) Find $E[X_2]$.

(c) Find $E[X_3]$.

(d) Conjecture the value of $E[X_k]$, and then verify your conjecture by a conditioning argument.

(e) Give an intuitive proof for your conjecture.

Hint: Number the initial r red and b blue balls, so the urn contains one type i red ball, for each $i = 1, \ldots, r$; as well as one type j blue ball, for each $j = 1, \ldots, b$. Now suppose that whenever a red ball is chosen it is returned along with m others of the same type, and similarly whenever a blue ball is chosen it is returned along with m others of the same type. Now, use a symmetry argument to determine the probability that any given selection is red.

29. Two players take turns shooting at a target, with each shot by player i hitting the target with probability $p_i, i = 1, 2$. Shooting ends when two consecutive shots hit the target. Let μ_i denote the mean number of shots taken when player i shoots first, $i = 1, 2$.

(a) Find μ_1 and μ_2.

(b) Let h_i denote the mean number of times that the target is hit when player i shoots first, $i = 1, 2$. Find h_1 and h_2.

30. Let $X_i, i \geqslant 0$ be independent and identically distributed random variables with probability mass function

$$p(j) = P\{X_i = j\}, \quad j = 1, \ldots, m, \quad \sum_{j=1}^{m} P(j) = 1$$

Find $E[N]$, where $N = \min\{n > 0 : X_n = X_0\}$.

31. Each element in a sequence of binary data is either 1 with probability p or 0 with probability $1 - p$. A maximal subsequence of consecutive values having identical outcomes is called a run. For instance, if the outcome sequence is $1, 1, 0, 1, 1, 1, 0$, the first run is of length 2, the second is of length 1, and the third is of length 3.

(a) Find the expected length of the first run.

(b) Find the expected length of the second run.

32. Independent trials, each resulting in success with probability p, are performed.

(a) Find the expected number of trials needed for there to have been both at least n successes and at least m failures.

Hint: Is it useful to know the result of the first $n + m$ trials?

(b) Find the expected number of trials needed for there to have been either at least n successes or at least m failures.

Hint: Make use of the result from part (a).

33. If R_i denotes the random amount that is earned in period i, then $\sum_{i=1}^{\infty} \beta^{i-1} R_i$, where $0 < \beta < 1$ is a specified constant, is called the total discounted reward with discount factor β. Let T be a geometric random variable with parameter $1 - \beta$ that is independent of the R_i. Show that the expected total discounted reward is equal to the expected total (undiscounted) reward earned by time T. That is, show that

$$E\left[\sum_{i=1}^{\infty} \beta^{i-1} R_i\right] = E\left[\sum_{i=1}^{T} R_i\right]$$

34. A set of n dice is thrown. All those that land on six are put aside, and the others are again thrown. This is repeated until all the dice have landed on six. Let N denote the number of throws needed. (For instance, suppose that $n = 3$ and that on the initial throw exactly two of the dice land on six. Then the other die will be thrown, and if it lands on six, then $N = 2$.) Let $m_n = E[N]$.

(a) Derive a recursive formula for m_n and use it to calculate $m_i, i = 2, 3, 4$ and to show that $m_5 \approx 13.024$.

(b) Let X_i denote the number of dice rolled on the ith throw. Find $E[\sum_{i=1}^{N} X_i]$.

35. Consider n multinomial trials, where each trial independently results in outcome i with probability p_i, $\sum_{i=1}^{k} p_i = 1$. With X_i equal to the number of trials that result in outcome i, find $E[X_1 | X_2 > 0]$.

36. Let $p_0 = P\{X = 0\}$ and suppose that $0 < p_0 < 1$. Let $\mu = E[X]$ and $\sigma^2 = \text{Var}(X)$.
- **(a)** Find $E[X|X \neq 0]$.
- **(b)** Find $\text{Var}(X|X \neq 0)$.

37. A manuscript is sent to a typing firm consisting of typists A, B, and C. If it is typed by A, then the number of errors made is a Poisson random variable with mean 2.6; if typed by B, then the number of errors is a Poisson random variable with mean 3; and if typed by C, then it is a Poisson random variable with mean 3.4. Let X denote the number of errors in the typed manuscript. Assume that each typist is equally likely to do the work.
- **(a)** Find $E[X]$.
- **(b)** Find $\text{Var}(X)$.

38. Suppose Y is uniformly distributed on $(0, 1)$, and that the conditional distribution of X given that $Y = y$ is uniform on $(0, y)$. Find $E[X]$ and $\text{Var}(X)$.

39. A deck of n cards, numbered 1 through n, is randomly shuffled so that all $n!$ possible permutations are equally likely. The cards are then turned over one at a time until card number 1 appears. These upturned cards constitute the first cycle. We now determine (by looking at the upturned cards) the lowest numbered card that has not yet appeared, and we continue to turn the cards face up until that card appears. This new set of cards represents the second cycle. We again determine the lowest numbered of the remaining cards and turn the cards until it appears, and so on until all cards have been turned over. Let m_n denote the mean number of cycles.
- **(a)** Derive a recursive formula for m_n in terms of $m_k, k = 1, \ldots, n-1$.
- **(b)** Starting with $m_0 = 0$, use the recursion to find m_1, m_2, m_3, and m_4.
- **(c)** Conjecture a general formula for m_n.
- **(d)** Prove your formula by induction on n. That is, show it is valid for $n = 1$, then assume it is true for any of the values $1, \ldots, n-1$ and show that this implies it is true for n.
- **(e)** Let X_i equal 1 if one of the cycles ends with card i, and let it equal 0 otherwise, $i = 1, \ldots, n$. Express the number of cycles in terms of these X_i.
- **(f)** Use the representation in part (e) to determine m_n.
- **(g)** Are the random variables $X_1, \ldots, X_n$ independent? Explain.
- **(h)** Find the variance of the number of cycles.

40. A prisoner is trapped in a cell containing three doors. The first door leads to a tunnel that returns him to his cell after two days of travel. The second leads to a tunnel that returns him to his cell after three days of travel. The third door leads immediately to freedom.
- **(a)** Assuming that the prisoner will always select doors 1, 2, and 3 with probabilities 0.5, 0.3, 0.2, what is the expected number of days until he reaches freedom?
- **(b)** Assuming that the prisoner is always equally likely to choose among those doors that he has not used, what is the expected number of days until he reaches freedom? (In this version, for instance, if the prisoner

initially tries door 1, then when he returns to the cell, he will now select only from doors 2 and 3.)

(c) For parts (a) and (b) find the variance of the number of days until the prisoner reaches freedom.

41. Workers $1, \ldots, n$ are currently idle. Suppose that each worker, independently, has probability p of being eligible for a job, and that a job is equally likely to be assigned to any of the workers that are eligible for it (if none are eligible, the job is rejected). Find the probability that the next job is assigned to worker 1.

***42.** If $X_i, i = 1, \ldots, n$ are independent normal random variables, with X_i having mean μ_i and variance 1, then the random variable $\sum_{i=1}^{n} X_i^2$ is said to be a *noncentral chi-squared* random variable.

(a) if X is a normal random variable having mean μ and variance 1 show, for $|t| < 1/2$, that the moment generating function of X^2 is

$$(1-2t)^{-1/2} e^{\frac{t\mu^2}{1-2t}}$$

(b) Derive the moment generating function of the noncentral chi-squared random variable $\sum_{i=1}^{n} X_i^2$, and show that its distribution depends on the sequence of means $\mu_1, \ldots, \mu_n$ only through the sum of their squares. As a result, we say that $\sum_{i=1}^{n} X_i^2$ is a noncentral chi-squared random variable with parameters n and $\theta = \sum_{i=1}^{n} \mu_i^2$.

(c) If all $\mu_i = 0$, then $\sum_{i=1}^{n} X_i^2$ is called a chi-squared random variable with n degrees of freedom. Determine, by differentiating its moment generating function, its expected value and variance.

(d) Let K be a Poisson random variable with mean $\theta/2$, and suppose that conditional on $K = k$, the random variable W has a chi-squared distribution with $n + 2k$ degrees of freedom. Show, by computing its moment generating function, that W is a noncentral chi-squared random variable with parameters n and θ.

(e) Find the expected value and variance of a noncentral chi-squared random variable with parameters n and θ.

***43.** For $P(Y \in A) > 0$, show that

$$E[X|Y \in A] = \frac{E[XI\{Y \in A\}]}{P(Y \in A)}$$

where $I\{B\}$ is the indicator variable of the event B, equal to 1 if B occurs and to 0 otherwise.

44. The number of customers entering a store on a given day is Poisson distributed with mean $\lambda = 10$. The amount of money spent by a customer is uniformly distributed over (0, 100). Find the mean and variance of the amount of money that the store takes in on a given day.

45. An individual traveling on the real line is trying to reach the origin. However, the larger the desired step, the greater is the variance in the result of that step.

Specifically, whenever the person is at location x, he next moves to a location having mean 0 and variance βx^2. Let X_n denote the position of the individual after having taken n steps. Supposing that $X_0 = x_0$, find

(a) $E[X_n]$;

(b) $\text{Var}(X_n)$.

46. **(a)** Show that

$$\text{Cov}(X, Y) = \text{Cov}(X, E[Y|X])$$

(b) Suppose, that, for constants a and b,

$$E[Y|X] = a + bX$$

Show that

$$b = \text{Cov}(X, Y)/\text{Var}(X)$$

***47.** If $E[Y|X] = 1$, show that

$$\text{Var}(XY) \geqslant \text{Var}(X)$$

48. Suppose that we want to predict the value of a random variable X by using one of the predictors $Y_1, \ldots, Y_n$, each of which satisfies $E[Y_i|X] = X$. Show that the predictor Y_i that minimizes $E[(Y_i - X)^2]$ is the one whose variance is smallest.

Hint: Compute $\text{Var}(Y_i)$ by using the conditional variance formula.

49. A and B play a series of games with A winning each game with probability p. The overall winner is the first player to have won two more games than the other.

(a) Find the probability that A is the overall winner.

(b) Find the expected number of games played.

50. Suppose that N, the number of flips made of a coin that comes up heads with probability p, is a geometric random variable with parameter α, independent of the results of the flips. Let A be the event that all flips land heads.

(a) Find $P(A)$ by conditioning on N.

(b) Find $P(A)$ by conditioning on the result of the first flip.

51. If X is geometric with parameter p, find the probability that X is even.

52. Each applicant has a score. If there are a total of n applicants then each applicant whose score is above s_n is accepted, where $s_1 = .2$, $s_2 = .4$, $s_n = .5$, $n \geq 3$. Suppose the scores of the applicants are independent uniform $(0, 1)$ random variables and are independent of N, the number of applicants, which is Poisson distributed with mean 2. Let X denote the number of applicants that are accepted. Derive expressions for

(a) $P(X = 0)$.

(b) $E[X]$.

***53.** Suppose X is a Poisson random variable with mean λ. The parameter λ is itself a random variable whose distribution is exponential with mean 1. Show that $P\{X = n\} = (\frac{1}{2})^{n+1}$.

54. Independent trials, each resulting in a success with probability p, are performed until k consecutive successful trials have occurred. Let X be the total number of successes in these trial, and let $P_n = P(X = n)$.

(a) Find P_k.

(b) Derive a recursive equation for the $P_n, n \geq k$, by imagining that the trials continue forever and conditioning on the time of the first failure.

(c) Verify your answer in part (a) by solving the recursion for P_k.

(d) When $p = .6, k = 3$, find P_8.

55. In the preceding problem let $M_k = E[X]$. Derive a recursive equation for M_k and then solve.

Hint: Start by writing $X_k = X_{k-1} + A_{k-1,k}$, where X_i is the total number of successes attained up to the first time there have been i consecutive successes, and $A_{k-1,k}$ is the additional number of successes after there have been $k - 1$ successes in a row until there have been k successes in a row.

56. Data indicate that the number of traffic accidents in Berkeley on a rainy day is a Poisson random variable with mean 9, whereas on a dry day it is a Poisson random variable with mean 3. Let X denote the number of traffic accidents tomorrow. If it will rain tomorrow with probability 0.6, find

(a) $E[X]$;

(b) $P\{X = 0\}$;

(c) $\text{Var}(X)$.

57. The number of storms in the upcoming rainy season is Poisson distributed but with a parameter value that is uniformly distributed over $(0, 5)$. That is, Λ is uniformly distributed over $(0, 5)$, and given that $\Lambda = \lambda$, the number of storms is Poisson with mean λ. Find the probability there are at least three storms this season.

***58.** Suppose that the conditional distribution of N, given that $Y = y$, is Poisson with mean y. Further suppose that Y is a gamma random variable with parameters (r, λ), where r is a positive integer. That is, suppose that

$$P(N = n|Y = y) = e^{-y}\frac{y^n}{n!}$$

and

$$f_Y(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{r-1}}{(r-1)!}, \quad y > 0$$

(a) Find $E[N]$.

(b) Find $\text{Var}(N)$.

(c) Find $P(N = n)$

(d) Using (c), argue that N is distributed as the total number of failures before the rth success when each trial is independently a success with probability $p = \frac{\lambda}{1+\lambda}$.

59. Suppose each new coupon collected is, independent of the past, a type i coupon with probability p_i. A total of n coupons is to be collected. Let A_i be the event that there is at least one type i in this set. For $i \neq j$, compute $P(A_i A_j)$ by

(a) conditioning on N_i, the number of type i coupons in the set of n coupons;

(b) conditioning on F_i, the first time a type i coupon is collected;

(c) using the identity $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i A_j)$.

***60.** Two players alternate flipping a coin that comes up heads with probability p. The first one to obtain a head is declared the winner. We are interested in the probability that the first player to flip is the winner. Before determining this probability, which we will call $f(p)$, answer the following questions.

(a) Do you think that $f(p)$ is a monotone function of p? If so, is it increasing or decreasing?

(b) What do you think is the value of $\lim_{p\to 1} f(p)$?

(c) What do you think is the value of $\lim_{p\to 0} f(p)$?

(d) Find $f(p)$.

61. Suppose in Exercise 29 that the shooting ends when the target has been hit twice. Let m_i denote the mean number of shots needed for the first hit when player i shoots first, $i = 1, 2$. Also, let $P_i, i = 1, 2$, denote the probability that the first hit is by player 1, when player i shoots first.

(a) Find m_1 and m_2.

(b) Find P_1 and P_2.

For the remainder of the problem, assume that player 1 shoots first.

(c) Find the probability that the final hit was by 1.

(d) Find the probability that both hits were by 1.

(e) Find the probability that both hits were by 2.

(f) Find the mean number of shots taken.

62. A, B, and C are evenly matched tennis players. Initially A and B play a set, and the winner then plays C. This continues, with the winner always playing the waiting player, until one of the players has won two sets in a row. That player is then declared the overall winner. Find the probability that A is the overall winner.

63. Suppose there are n types of coupons, and that the type of each new coupon obtained is independent of past selections and is equally likely to be any of the n types. Suppose one continues collecting until a complete set of at least one of each type is obtained.

(a) Find the probability that there is exactly one type i coupon in the final collection.

Hint: Condition on T, the number of types that are collected before the first type i appears.

(b) Find the expected number of types that appear exactly once in the final collection.

64. A and B roll a pair of dice in turn, with A rolling first. A's objective is to obtain a sum of 6, and B's is to obtain a sum of 7. The game ends when either player reaches his or her objective, and that player is declared the winner.
(a) Find the probability that A is the winner.
(b) Find the expected number of rolls of the dice.
(c) Find the variance of the number of rolls of the dice.

65. The number of red balls in an urn that contains n balls is a random variable that is equally likely to be any of the values $0, 1, \ldots, n$. That is,

$$P\{i \text{ red}, n-i \text{ non-red}\} = \frac{1}{n+1}, \quad i = 0, \ldots, n$$

The n balls are then randomly removed one at a time. Let Y_k denote the number of red balls in the first k selections, $k = 1, \ldots, n$.
(a) Find $P\{Y_n = j\}$, $j = 0, \ldots, n$.
(b) Find $P\{Y_{n-1} = j\}$, $j = 0, \ldots, n$.
(c) What do you think is the value of $P\{Y_k = j\}$, $j = 0, \ldots, n$?
(d) Verify your answer to part (c) by a backwards induction argument. That is, check that your answer is correct when $k = n$, and then show that whenever it is true for k it is also true for $k-1, k = 1, \ldots, n$.

66. The number of goals that J scores in soccer games that her team wins is Poisson with mean 2, while the number she scores in games that her team loses is Poisson with mean 1. Assume that J's team wins each game they play with probability p.
(a) Find the expected number of goals that J scores in her next 3 games.
(b) Find the probability that she scores a total of n goals in her next 3 games.

***67.** A coin having probability p of coming up heads is continually flipped. Let $P_j(n)$ denote the probability that a run of j successive heads occurs within the first n flips.
(a) Argue that

$$P_j(n) = P_j(n-1) + p^j(1-p)[1 - P_j(n-j-1)]$$

(b) By conditioning on the first non-head to appear, derive another equation relating $P_j(n)$ to the quantities $P_j(n-k), k = 1, \ldots, j$.

68. If the level of infection of a tree is x, $0 \le x \le 1$, then each cure attempt will independently be successful with probability $1 - x$. Consider a tree whose infection level, call it L, is assumed to be the value of a uniform (0, 1) random variable.
(a) What is the probability that a single attempt will result in a cure.
(b) Find the probability that the first two cure attempts are unsuccessful.
(c) Find the conditional expected value of L given that it took 3 attempts to cure the tree.

69. In the match problem, say that $(i, j), i < j$, is a pair if i chooses j's hat and j chooses i's hat.
(a) Find the expected number of pairs.

(b) Let Q_n denote the probability that there are no pairs, and derive a recursive formula for Q_n in terms of Q_j, $j < n$.

Hint: Use the cycle concept.

(c) Use the recursion of part (b) to find Q_8.

70. Let N denote the number of cycles that result in the match problem.

(a) Let $M_n = E[N]$, and derive an equation for M_n in terms of $M_1, \ldots, M_{n-1}$.

(b) Let C_j denote the size of the cycle that contains person j. Argue that

$$N = \sum_{j=1}^{n} 1/C_j$$

and use the preceding to determine $E[N]$.

(c) Find the probability that persons $1, 2, \ldots, k$ are all in the same cycle.

(d) Find the probability that $1, 2, \ldots, k$ is a cycle.

71. Use Eq. (3.13) to obtain Eq. (3.9).

Hint: First multiply both sides of Eq. (3.13) by n, then write a new equation by replacing n with $n - 1$, and then subtract the former from the latter.

72. In Example 3.29 show that the conditional distribution of N given that $U_1 = y$ is the same as the conditional distribution of M given that $U_1 = 1 - y$. Also, show that

$$E[N|U_1 = y] = E[M|U_1 = 1 - y] = 1 + e^y$$

***73.** Suppose that we continually roll a die until the sum of all throws exceeds 100. What is the most likely value of this total when you stop?

74. There are five components. The components act independently, with component i working with probability p_i, $i = 1, 2, 3, 4, 5$. These components form a system as shown in Fig. 3.7.

The system is said to work if a signal originating at the left end of the diagram can reach the right end, where it can pass through a component only if that component is working. (For instance, if components 1 and 4 both work, then the system also works.) What is the probability that the system works?

75. This problem will present another proof of the ballot problem of Example 3.28.

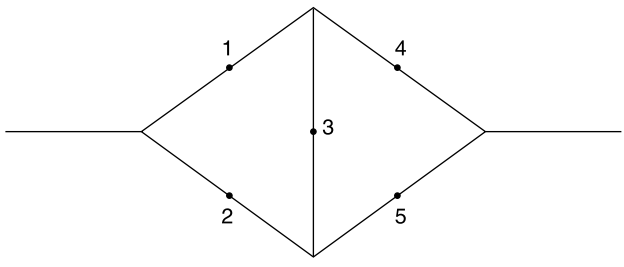

Figure 3.7

(a) Argue that

$$P_{n,m} = 1 - P\{A \text{ and } B \text{ are tied at some point}\}$$

(b) Explain why

$$P\{A \text{ receives first vote and they are eventually tied}\}$$
$$= P\{B \text{ receives first vote and they are eventually tied}\}$$

Hint: Any outcome in which they are eventually tied with A receiving the first vote corresponds to an outcome in which they are eventually tied with B receiving the first vote. Explain this correspondence.

(c) Argue that $P\{\text{eventually tied}\} = 2m/(n+m)$, and conclude that $P_{n,m} = (n-m)/(n+m)$.

76. Consider a gambler who on each bet either wins 1 with probability 18/38 or loses 1 with probability 20/38. (These are the probabilities if the bet is that a roulette wheel will land on a specified color.) The gambler will quit either when he or she is winning a total of 5 or after 100 plays. What is the probability he or she plays exactly 15 times?

77. Show that

(a) $E[XY|Y=y] = yE[X|Y=y]$

(b) $E[g(X,Y)|Y=y] = E[g(X,y)|Y=y]$

(c) $E[XY] = E[YE[X|Y]]$

78. In the ballot problem (Example 3.28), compute $P\{A \text{ is never behind}\}$.

79. An urn contains n white and m black balls that are removed one at a time. If $n > m$, show that the probability that there are always more white than black balls in the urn (until, of course, the urn is empty) equals $(n-m)/(n+m)$. Explain why this probability is equal to the probability that the set of withdrawn balls always contains more white than black balls. (This latter probability is $(n-m)/(n+m)$ by the ballot problem.)

80. A coin that comes up heads with probability p is flipped n consecutive times. What is the probability that starting with the first flip there are always more heads than tails that have appeared?

81. Let $X_i, i \geqslant 1$, be independent uniform (0, 1) random variables, and define N by

$$N = \min\{n: X_n < X_{n-1}\}$$

where $X_0 = x$. Let $f(x) = E[N]$.

(a) Derive an integral equation for $f(x)$ by conditioning on X_1.

(b) Differentiate both sides of the equation derived in part (a).

(c) Solve the resulting equation obtained in part (b).

(d) For a second approach to determining $f(x)$ argue that

$$P\{N \geqslant k\} = \frac{(1-x)^{k-1}}{(k-1)!}$$

(e) Use part (d) to obtain $f(x)$.

82. Let $X_1, X_2, \ldots$ be independent continuous random variables with a common distribution function F and density $f = F'$, and for $k \geqslant 1$ let

$$N_k = \min\{n \geqslant k: X_n = k\text{th largest of } X_1, \ldots, X_n\}$$

(a) Show that $P\{N_k = n\} = \frac{k-1}{n(n-1)}$, $n \geqslant k$.

(b) Argue that

$$f_{X_{N_k}}(x) = f(x)(\bar{F}(x))^{k-1} \sum_{i=0}^{\infty} \binom{i+k-2}{i} (F(x))^i$$

(c) Prove the following identity:

$$a^{1-k} = \sum_{i=0}^{\infty} \binom{i+k-2}{i} (1-a)^i, \quad 0 < a < 1, k \geqslant 2$$

Hint: Use induction. First prove it when $k = 2$, and then assume it for k. To prove it for $k+1$, use the fact that

$$\sum_{i=1}^{\infty} \binom{i+k-1}{i} (1-a)^i = \sum_{i=1}^{\infty} \binom{i+k-2}{i} (1-a)^i + \sum_{i=1}^{\infty} \binom{i+k-2}{i-1} (1-a)^i$$

where the preceding used the combinatorial identity

$$\binom{m}{i} = \binom{m-1}{i} + \binom{m-i}{i-1}$$

Now, use the induction hypothesis to evaluate the first term on the right side of the preceding equation.

(d) Conclude that X_{N_k} has distribution F.

83. An urn contains n balls, with ball i having weight $w_i, i = 1, \ldots, n$. The balls are withdrawn from the urn one at a time according to the following scheme: When S is the set of balls that remains, ball $i, i \in S$, is the next ball withdrawn with probability $w_i / \sum_{j \in S} w_j$. Find the expected number of balls that are withdrawn before ball $i, i = 1, \ldots, n$.

84. Suppose in Example 3.33 that a point is only won if the winner of the rally was the server of that rally.

(a) If A is currently serving, what is the probability that A wins the next point?

(b) Explain how to obtain the final score probabilities.

85. In the list problem, when the P_i are known, show that the best ordering (best in the sense of minimizing the expected position of the element requested) is to place the elements in decreasing order of their probabilities. That is, if $P_1 > P_2 > \cdots > P_n$, show that $1, 2, \ldots, n$ is the best ordering.

86. Consider the random graph of Section 3.6.2 when $n = 5$. Compute the probability distribution of the number of components and verify your solution by using it to compute $E[C]$ and then comparing your solution with

$$E[C] = \sum_{k=1}^{5} \binom{5}{k} \frac{(k-1)!}{5^k}$$

87. **(a)** From the results of Section 3.6.3 we can conclude that there are $\binom{n+m-1}{m-1}$ nonnegative integer valued solutions of the equation $x_1 + \cdots + x_m = n$. Prove this directly.

(b) How many positive integer valued solutions of $x_1 + \cdots + x_m = n$ are there?

Hint: Let $y_i = x_i - 1$.

(c) For the Bose–Einstein distribution, compute the probability that exactly k of the X_i are equal to 0.

88. In Section 3.6.3, we saw that if U is a random variable that is uniform on (0, 1) and if, conditional on $U = p$, X is binomial with parameters n and p, then

$$P\{X = i\} = \frac{1}{n+1}, \quad i = 0, 1, \ldots, n$$

For another way of showing this result, let $U, X_1, X_2, \ldots, X_n$ be independent uniform (0, 1) random variables. Define X by

$$X = \#i\colon X_i < U$$

That is, if the $n+1$ variables are ordered from smallest to largest, then U would be in position $X + 1$.

(a) What is $P\{X = i\}$?

(b) Explain how this proves the result of Section 3.6.3.

89. Let $I_1, \ldots, I_n$ be independent random variables, each of which is equally likely to be either 0 or 1. A well-known nonparametric statistical test (called the signed rank test) is concerned with determining $P_n(k)$ defined by

$$P_n(k) = P\left\{\sum_{j=1}^{n} j I_j \leqslant k\right\}$$

Justify the following formula:

$$P_n(k) = \tfrac{1}{2} P_{n-1}(k) + \tfrac{1}{2} P_{n-1}(k-n)$$

90. The number of accidents in each period is a Poisson random variable with mean 5. With $X_n, n \geq 1$, equal to the number of accidents in period n, find $E[N]$ when
(a) $N = \min(n: X_{n-2} = 2, X_{n-1} = 1, X_n = 0)$;
(b) $N = \min(n: X_{n-3} = 2, X_{n-2} = 1, X_{n-1} = 0, X_n = 2)$.

91. Find the expected number of flips of a coin, which comes up heads with probability p, that are necessary to obtain the pattern h, t, h, h, t, h, t, h.

92. The number of coins that Josh spots when walking to work is a Poisson random variable with mean 6. Each coin is equally likely to be a penny, a nickel, a dime, or a quarter. Josh ignores the pennies but picks up the other coins.
(a) Find the expected amount of money that Josh picks up on his way to work.
(b) Find the variance of the amount of money that Josh picks up on his way to work.
(c) Find the probability that Josh picks up exactly 25 cents on his way to work.

***93.** Consider a sequence of independent trials, each of which is equally likely to result in any of the outcomes $0, 1, \ldots, m$. Say that a round begins with the first trial, and that a new round begins each time outcome 0 occurs. Let N denote the number of trials that it takes until all of the outcomes $1, \ldots, m-1$ have occurred in the same round. Also, let T_j denote the number of trials that it takes until j distinct outcomes have occurred, and let I_j denote the jth distinct outcome to occur. (Therefore, outcome I_j first occurs at trial T_j.)
(a) Argue that the random vectors $(I_1, \ldots, I_m)$ and $(T_1, \ldots, T_m)$ are independent.
(b) Define X by letting $X = j$ if outcome 0 is the jth distinct outcome to occur. (Thus, $I_X = 0$.) Derive an equation for $E[N]$ in terms of $E[T_j], j = 1, \ldots, m-1$ by conditioning on X.
(c) Determine $E[T_j], j = 1, \ldots, m-1$.

Hint: See Exercise 42 of Chapter 2.
(d) Find $E[N]$.

94. Let N be a hypergeometric random variable having the distribution of the number of white balls in a random sample of size r from a set of w white and b blue balls. That is,

$$P\{N = n\} = \frac{\binom{w}{n}\binom{b}{r-n}}{\binom{w+b}{r}}$$

where we use the convention that $\binom{m}{j} = 0$ if either $j < 0$ or $j > m$. Now, consider a compound random variable $S_N = \sum_{i=1}^{N} X_i$, where the X_i are positive integer valued random variables with $\alpha_j = P\{X_i = j\}$.
(a) With M as defined as in Section 3.7, find the distribution of $M - 1$.

(b) Suppressing its dependence on b, let $P_{w,r}(k) = P\{S_N = k\}$, and derive a recursion equation for $P_{w,r}(k)$.

(c) Use the recursion of (b) to find $P_{w,r}(2)$.

95. For the left skip free random walk of Section 3.6.6 let $\beta = P(S_n \leq 0 \text{ for all } n)$ be the probability that the walk is never positive. Find β when $E[X_i] < 0$.

96. Consider a large population of families, and suppose that the number of children in the different families are independent Poisson random variables with mean λ. Show that the number of siblings of a randomly chosen child is also Poisson distributed with mean λ.

***97.** Use the conditional variance formula to find the variance of a geometric random variable.

98. For a compound random variable $S = \sum_{i=1}^{N} X_i$, find $\text{Cov}(N, S)$.

99. Let N be the number of trials until k consecutive successes have occurred, when each trial is independently a success with probability p.

(a) What is $P(N = k)$?

(b) Argue that

$$P(N = k + r) = P(N > r - 1)qp^k, \quad r > 0$$

(c) Show that

$$1 - p^k = qp^k E[N]$$

100. In the fair gambler's ruin problem of Example 3.16, let P_i denote the probability that, starting with a fortune of i, the gambler's fortune reaches n before 0. Find P_i, $0 \leq i \leq n$.

101. For the left skip free random walk of Section 3.6.6,

(a) Show, for $0 < k \leq n$, that $P(T_k = n | S_n = -k) = k/n$.

(b) Show that part (a) implies that $P(S_j < 0,\ j = 1, \ldots, n | S_n = -k) = k/n$.

(c) Explain why part (b) implies the ballot theorem.

Markov Chains

4

4.1 Introduction

Consider a process that has a value in each time period. Let X_n denote its value in time period n, and suppose we want to make a probability model for the sequence of successive values $X_0, X_1, X_2 \ldots$. The simplest model would probably be to assume that the X_n are independent random variables, but often such an assumption is clearly unjustified. For instance, starting at some time suppose that X_n represents the price of one share of some security, such as Google, at the end of n additional trading days. Then it certainly seems unreasonable to suppose that the price at the end of day $n+1$ is independent of the prices on days $n, n-1, n-2$ and so on down to day 0. However, it might be reasonable to suppose that the price at the end of trading day $n+1$ depends on the previous end-of-day prices only through the price at the end of day n. That is, it might be reasonable to assume that the conditional distribution of X_{n+1} given all the past end-of-day prices $X_n, X_{n-1}, \ldots, X_0$ depends on these past prices only through the price at the end of day n. Such an assumption defines a Markov chain, a type of stochastic process that will be studied in this chapter, and which we now formally define.

Let $\{X_n, n = 0, 1, 2, \ldots\}$ be a stochastic process that takes on a finite or countable number of possible values. Unless otherwise mentioned, this set of possible values of the process will be denoted by the set of nonnegative integers $\{0, 1, 2, \ldots\}$. If $X_n = i$, then the process is said to be in state i at time n. We suppose that whenever the process is in state i, there is a fixed probability P_{ij} that it will next be in state j. That is, we suppose that

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij} \tag{4.1}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geqslant 0$. Such a stochastic process is known as a *Markov chain*. Eq. (4.1) may be interpreted as stating that, for a Markov chain, the conditional distribution of any future state X_{n+1}, given the past states $X_0, X_1, \ldots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state.

The value P_{ij} represents the probability that the process will, when in state i, next make a transition into state j. Since probabilities are nonnegative and since the process must make a transition into some state, we have

$$P_{ij} \geqslant 0, \qquad i, j \geqslant 0; \qquad \sum_{j=0}^{\infty} P_{ij} = 1, \qquad i = 0, 1, \ldots$$

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00009-3

Let $\mathbf{P}$ denote the matrix of one-step transition probabilities P_{ij}, so that

$$\mathbf{P} = \begin{Vmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{Vmatrix}$$

Example 4.1 (Forecasting the Weather)**.** Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain tomorrow with probability β.

If we say that the process is in state 0 when it rains and state 1 when it does not rain, then the preceding is a two-state Markov chain whose transition probabilities are given by

$$\mathbf{P} = \begin{Vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{Vmatrix}$$ ■

Example 4.2 (A Communications System)**.** Consider a communications system that transmits the digits 0 and 1. Each digit transmitted must pass through several stages, at each of which there is a probability p that the digit entered will be unchanged when it leaves. Letting X_n denote the digit entering the nth stage, then $\{X_n, n = 0, 1, \ldots\}$ is a two-state Markov chain having a transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} p & 1-p \\ 1-p & p \end{Vmatrix}$$ ■

Example 4.3. On any given day Gary is either cheerful (C), so-so (S), or glum (G). If he is cheerful today, then he will be C, S, or G tomorrow with respective probabilities 0.5, 0.4, 0.1. If he is feeling so-so today, then he will be C, S, or G tomorrow with probabilities 0.3, 0.4, 0.3. If he is glum today, then he will be C, S, or G tomorrow with probabilities 0.2, 0.3, 0.5.

Letting X_n denote Gary's mood on the nth day, then $\{X_n, n \geqslant 0\}$ is a three-state Markov chain (state $0 = C$, state $1 = S$, state $2 = G$) with transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{Vmatrix}$$ ■

Example 4.4 (Transforming a Process into a Markov Chain)**.** Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain

tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

If we let the state at time n depend only on whether or not it is raining at time n, then the preceding model is not a Markov chain (why not?). However, we can transform this model into a Markov chain by saying that the state at any time is determined by the weather conditions during both that day and the previous day. In other words, we can say that the process is in

state 0 if it rained both today and yesterday,
state 1 if it rained today but not yesterday,
state 2 if it rained yesterday but not today,
state 3 if it did not rain either yesterday or today.

The preceding would then represent a four-state Markov chain having a transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix}$$

You should carefully check the matrix $\mathbf{P}$, and make sure you understand how it was obtained. ■

Example 4.5 (A Random Walk Model)**.** A Markov chain whose state space is given by the integers $i = 0, \pm 1, \pm 2, \ldots$ is said to be a *random walk* if, for some number $0 < p < 1$,

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 0, \pm 1, \ldots$$

The preceding Markov chain is called a random walk for we may think of it as being a model for an individual walking on a straight line who at each point of time either takes one step to the right with probability p or one step to the left with probability $1 - p$. ■

Example 4.6 (A Gambling Model)**.** Consider a gambler who, at each play of the game, either wins \$1 with probability p or loses \$1 with probability $1 - p$. If we suppose that our gambler quits playing either when he goes broke or he attains a fortune of \$$N$, then the gambler's fortune is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 1, 2, \ldots, N-1,$$
$$P_{00} = P_{NN} = 1$$

States 0 and N are called *absorbing* states since once entered they are never left. Note that the preceding is a finite state random walk with absorbing barriers (states 0 and N). ■

Example 4.7. In most of Europe and Asia annual automobile insurance premiums are determined by use of a Bonus Malus (Latin for Good-Bad) system. Each policyholder is given a positive integer valued state and the annual premium is a function of this state (along, of course, with the type of car being insured and the level of insurance). A policyholder's state changes from year to year in response to the number of claims made by that policyholder. Because lower numbered states correspond to lower annual premiums, a policyholder's state will usually decrease if he or she had no claims in the preceding year, and will generally increase if he or she had at least one claim. (Thus, no claims is good and typically results in a decreased premium, while claims are bad and typically result in a higher premium.)

For a given Bonus Malus system, let $s_i(k)$ denote the next state of a policyholder who was in state i in the previous year and who made a total of k claims in that year. If we suppose that the number of yearly claims made by a particular policyholder is a Poisson random variable with parameter λ, then the successive states of this policyholder will constitute a Markov chain with transition probabilities

$$P_{i,j} = \sum_{k:s_i(k)=j} e^{-\lambda}\frac{\lambda^k}{k!}, \quad j \geqslant 0$$

Whereas there are usually many states (20 or so is not atypical), the following table specifies a hypothetical Bonus Malus system having four states.

State	Annual Premium	Next state if			
		0 claims	*1 claim*	*2 claims*	*$\geqslant$ 3 claims*
1	200	1	2	3	4
2	250	1	3	4	4
3	400	2	4	4	4
4	600	3	4	4	4

Thus, for instance, the table indicates that $s_2(0) = 1; s_2(1) = 3; s_2(k) = 4, k \geqslant 2$. Consider a policyholder whose annual number of claims is a Poisson random variable with parameter λ. If a_k is the probability that such a policyholder makes k claims in a year, then

$$a_k = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \geqslant 0$$

For the Bonus Malus system specified in the preceding table, the transition probability matrix of the successive states of this policyholder is

$$\mathbf{P} = \begin{Vmatrix} a_0 & a_1 & a_2 & 1-a_0-a_1-a_2 \\ a_0 & 0 & a_1 & 1-a_0-a_1 \\ 0 & a_0 & 0 & 1-a_0 \\ 0 & 0 & a_0 & 1-a_0 \end{Vmatrix}$$

■

4.2 Chapman–Kolmogorov Equations

We have already defined the one-step transition probabilities P_{ij}. We now define the n-step transition probabilities P_{ij}^n to be the probability that a process in state i will be in state j after n additional transitions. That is,

$$P_{ij}^n = P\{X_{n+k} = j | X_k = i\}, \quad n \geqslant 0,\ i, j \geqslant 0$$

Of course $P_{ij}^1 = P_{ij}$. The *Chapman–Kolmogorov equations* provide a method for computing these n-step transition probabilities. These equations are

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m \quad \text{for all } n,\ m \geqslant 0,\ \text{all } i, j \tag{4.2}$$

and are most easily understood by noting that $P_{ik}^n P_{kj}^m$ represents the probability that starting in i the process will go to state j in $n + m$ transitions through a path which takes it into state k at the nth transition. Hence, summing over all intermediate states k yields the probability that the process will be in state j after $n + m$ transitions. Formally, we have

$$\begin{aligned} P_{ij}^{n+m} &= P\{X_{n+m} = j | X_0 = i\} \\ &= \sum_{k=0}^{\infty} P\{X_{n+m} = j, X_n = k | X_0 = i\} \\ &= \sum_{k=0}^{\infty} P\{X_{n+m} = j | X_n = k, X_0 = i\} P\{X_n = k | X_0 = i\} \\ &= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n \end{aligned}$$

If we let $\mathbf{P}^{(n)}$ denote the matrix of n-step transition probabilities P_{ij}^n, then Eq. (4.2) asserts that

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$$

where the dot represents matrix multiplication.[1] Hence, in particular,

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P} \cdot \mathbf{P} = \mathbf{P}^2$$

and by induction

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1+1)} = \mathbf{P}^{n-1} \cdot \mathbf{P} = \mathbf{P}^n$$

[1] If $\mathbf{A}$ is an $N \times M$ matrix whose element in the ith row and jth column is a_{ij} and $\mathbf{B}$ is an $M \times K$ matrix whose element in the ith row and jth column is b_{ij}, then $\mathbf{A} \cdot \mathbf{B}$ is defined to be the $N \times K$ matrix whose element in the ith row and jth column is $\sum_{k=1}^{M} a_{ik} b_{kj}$.

That is, the n-step transition matrix may be obtained by multiplying the matrix $\mathbf{P}$ by itself n times.

Example 4.8. Consider Example 4.1 in which the weather is considered as a two-state Markov chain. If $\alpha = 0.7$ and $\beta = 0.4$, then calculate the probability that it will rain four days from today given that it is raining today.

Solution: The one-step transition probability matrix is given by

$$\mathbf{P} = \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix}$$

Hence,

$$\begin{aligned} \mathbf{P}^{(2)} = \mathbf{P}^2 &= \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix} \cdot \begin{Vmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Vmatrix} \\ &= \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix}, \\ \mathbf{P}^{(4)} = (\mathbf{P}^2)^2 &= \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix} \cdot \begin{Vmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{Vmatrix} \\ &= \begin{Vmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{Vmatrix} \end{aligned}$$

and the desired probability P_{00}^4 equals 0.5749. ■

Example 4.9. Consider Example 4.4. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution: The two-step transition matrix is given by

$$\begin{aligned} \mathbf{P}^{(2)} = \mathbf{P}^2 &= \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix} \cdot \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix} \\ &= \begin{Vmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{Vmatrix} \end{aligned}$$

Since rain on Thursday is equivalent to the process being in either state 0 or state 1 on Thursday, the desired probability is given by $P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61$. ■

Example 4.10. An urn always contains 2 balls. Ball colors are red and blue. At each stage a ball is randomly chosen and then replaced by a new ball, which with probability 0.8 is the same color, and with probability 0.2 is the opposite color, as the ball it replaces. If initially both balls are red, find the probability that the fifth ball selected is red.

Solution: To find the desired probability we first define an appropriate Markov chain. This can be accomplished by noting that the probability that a selection is red is determined by the composition of the urn at the time of the selection. So, let us define X_n to be the number of red balls in the urn after the nth selection and subsequent replacement. Then $X_n, n \geq 0$, is a Markov chain with states 0, 1, 2 and with transition probability matrix $\mathbf{P}$ given by

$$\begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

To understand the preceding, consider for instance $P_{1,0}$. Now, to go from 1 red ball in the urn to 0 red balls, the ball chosen must be red (which occurs with probability 0.5) and it must then be replaced by a ball of opposite color (which occurs with probability 0.2), showing that

$$P_{1,0} = (0.5)(0.2) = 0.1$$

To determine the probability that the fifth selection is red, condition on the number of red balls in the urn after the fourth selection. This yields

$$\begin{aligned} &P(\text{fifth selection is red}) \\ &\quad = \sum_{i=0}^{2} P(\text{fifth selection is red} | X_4 = i) P(X_4 = i | X_0 = 2) \\ &\quad = (0) P_{2,0}^4 + (0.5) P_{2,1}^4 + (1) P_{2,2}^4 \\ &\quad = 0.5 P_{2,1}^4 + P_{2,2}^4 \end{aligned}$$

To calculate the preceding we compute $\mathbf{P}^4$. Doing so yields

$$P_{2,1}^4 = 0.4352, \quad P_{2,2}^4 = 0.4872$$

giving the answer $P(\text{fifth selection is red}) = 0.7048$. ■

Example 4.11. Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 nonempty urns after 9 balls have been distributed?

Solution: If we let X_n be the number of nonempty urns after n balls have been distributed, then $X_n, n \geqslant 0$ is a Markov chain with states $0, 1, \ldots, 8$ and transition probabilities

$$P_{i,i} = i/8 = 1 - P_{i,i+1}, \quad i = 0, 1, \ldots, 8$$

The desired probability is $P_{0,3}^9 = P_{1,3}^8$, where the equality follows because $P_{0,1} = 1$. Now, starting with 1 occupied urn, if we had wanted to determine the entire probability distribution of the number of occupied urns after 8 additional balls

had been distributed we would need to consider the transition probability matrix with states $1, 2, \ldots, 8$. However, because we only require the probability, starting with a single occupied urn, that there are 3 occupied urns after an additional 8 balls have been distributed we can make use of the fact that the state of the Markov chain cannot decrease to collapse all states $4, 5, \ldots, 8$ into a single state 4 with the interpretation that the state is 4 whenever four or more of the urns are occupied. Consequently, we need only determine the eight-step transition probability $P_{1,3}^8$ of the Markov chain with states 1, 2, 3, 4 having transition probability matrix $\mathbf{P}$ given by

$$\begin{pmatrix} 1/8 & 7/8 & 0 & 0 \\ 0 & 2/8 & 6/8 & 0 \\ 0 & 0 & 3/8 & 5/8 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Raising the preceding matrix to the power 4 yields the matrix $\mathbf{P}^4$ given by

$$\begin{pmatrix} 0.0002 & 0.0256 & 0.2563 & 0.7178 \\ 0 & 0.0039 & 0.0952 & 0.9009 \\ 0 & 0 & 0.0198 & 0.9802 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Hence,

$$\begin{aligned} P_{1,3}^8 &= 0.0002 \times 0.2563 + 0.0256 \times 0.0952 + 0.2563 \times 0.0198 \\ &\quad + 0.7178 \times 0 = 0.00756 \end{aligned}$$

■

Consider a Markov chain with transition probabilities P_{ij}. Let $\mathscr{A}$ be a set of states, and suppose we are interested in the probability that the Markov chain ever enters any of the states in $\mathscr{A}$ by time m. That is, for a given state $i \notin \mathscr{A}$, we are interested in determining

$$\beta = P(X_k \in \mathscr{A} \text{ for some } k = 1, \ldots, m | X_0 = i)$$

To determine the preceding probability we will define a Markov chain $\{W_n, n \geqslant 0\}$ whose states are the states that are not in $\mathscr{A}$ plus an additional state, which we will call A in our general discussion (though in specific examples we will usually give it a different name). Once the $\{W_n\}$ Markov chain enters state A it remains there forever.

The new Markov chain is defined as follows. Letting X_n denote the state at time n of the Markov chain with transition probabilities $P_{i,j}$, define

$$N = \min\{n : X_n \in \mathscr{A}\}$$

and let $N = \infty$ if $X_n \notin \mathscr{A}$ for all n. In words, N is the first time the Markov chain enters the set of states $\mathscr{A}$. Now, define

$$W_n = \begin{cases} X_n, & \text{if } n < N \\ A, & \text{if } n \geqslant N \end{cases}$$

So the state of the $\{W_n\}$ process is equal to the state of the original Markov chain up to the point when the original Markov chain enters a state in $\mathscr{A}$. At that time the new process goes to state A and remains there forever. From this description it follows that $W_n, n \geqslant 0$ is a Markov chain with states $i, i \notin \mathscr{A}, A$ and with transition probabilities $Q_{i,j}$, given by

$$
\begin{aligned}
Q_{i,j} &= P_{i,j}, \quad \text{if} \quad i \notin \mathscr{A}, j \notin \mathscr{A} \\
Q_{i,A} &= \sum_{j \in \mathscr{A}} P_{i,j}, \quad \text{if} \quad i \notin \mathscr{A} \\
Q_{A,A} &= 1
\end{aligned}
$$

Because the original Markov chain will have entered a state in $\mathscr{A}$ by time m if and only if the state at time m of the new Markov chain is A, we see that

$$
\begin{aligned}
&P(X_k \in \mathscr{A} \text{ for some } k = 1, \ldots, m | X_0 = i) \\
&\quad = P(W_m = A | X_0 = i) = P(W_m = A | W_0 = i) = Q_{i,A}^m
\end{aligned}
$$

That is, the desired probability is equal to an m-step transition probability of the new chain.

Example 4.12. In a sequence of independent flips of a fair coin, let N denote the number of flips until there is a run of three consecutive heads. Find

(a) $P(N \leq 8)$ and
(b) $P(N = 8)$.

Solution: To determine $P(N \leq 8)$, define a Markov chain with states $0, 1, 2, 3$ where for $i < 3$ state i means that we currently are on a run of i consecutive heads, and where state 3 means that a run of three consecutive heads has already occurred. Thus, the transition probability matrix is

$$
\mathbf{P} = \begin{pmatrix}
1/2 & 1/2 & 0 & 0 \\
1/2 & 0 & 1/2 & 0 \\
1/2 & 0 & 0 & 1/2 \\
0 & 0 & 0 & 1
\end{pmatrix}
$$

where, for instance, the values for row 2 are obtained by noting that if we currently are on a run of size 1 then the next state will be 0 if the next flip is a tail, or 2 if it is a head. Hence, $P_{1,0} = P_{1,2} = 1/2$. Because there would be a run of three consecutive heads within the first eight flips if and only if $X_8 = 3$, the desired probability is $P_{0,3}^8$. Squaring $\mathbf{P}$ to obtain $\mathbf{P}^2$, then squaring the result to obtain $\mathbf{P}^4$, and then squaring that matrix gives the result

$$
\mathbf{P}^8 = \begin{pmatrix}
81/256 & 44/256 & 24/256 & 107/256 \\
68/256 & 37/256 & 20/256 & 131/256 \\
44/256 & 24/256 & 13/256 & 175/256 \\
0 & 0 & 0 & 1
\end{pmatrix}
$$

Hence, the probability that there will be a run of three consecutive heads within the first eight flips is $107/256 \approx .4180$.
(b) Noting that $N = 8$ if the pattern hasn't yet occurred in the first 7 transitions, the state after 7 transitions is 2, and the next flip lands heads, shows that

$$P(N = 8) = \frac{1}{2} P_{0,2}^7.$$ ■

We can also use Markov chains to determine probabilities concerning the time until a pattern appears when the data itself comes from a Markov chain. We illustrate this by an example.

Example 4.13. Let $\{X_n, n \geq 0\}$ be a Markov chain with states 0, 1, 2, 3 and transition probabilities $P_{i,j}$, $i, j = 0, 1, 2, 3$, and let N denote the number of transitions, starting in state 0, until the pattern 1, 2, 1, 2 appears. That is,

$$N = \min\{n \geq 4 : X_{n-3} = 1, X_{n-2} = 2, X_{n-1} = 1, X_n = 2\}.$$

Suppose we are interested in evaluating $P(N \leq k)$ for some specified value k. To do so, we define a new Markov chain $\{Y_n, n \geq 0\}$, that tracks the progress towards the pattern. The Y_n are defined as follows:

- If the pattern has appeared by the nth transition—that is, if $X_0, \ldots, X_n$ includes 1, 2, 1, 2—then $Y_n = 4$.
- If the pattern has not yet appeared by the nth transition

  $Y_n = 1$ if $X_n = 1$ and $(X_{n-2}, X_{n-1}) \neq (1, 2)$.
  $Y_n = 2$ if $X_{n-1} = 1$, $X_n = 2$.
  $Y_n = 3$ if $X_{n-2} = 1$, $X_{n-1} = 2$, $X_n = 1$.
  $Y_n = 5$ if $X_n = 2$, $X_{n-1} \neq 1$.
  $Y_n = 6$ if $X_n = 0$.
  $Y_n = 7$ if $X_n = 3$

Thus, for $i = 1, 2, 3, 4$, $Y_n = i$ signifies that we are i steps into the pattern (or in the case $i = 4$ that the pattern has appeared). $Y_n = 5$ (or 6 or 7) if there is no current progress with regards to the pattern and the current state is 2 (or 0 or 3). The desired probability $P(N \leq k)$ is equal to the probability that the number of transitions of the Markov chain $\{Y_n\}$ to go from state 6 to state 4 is less than or equal to k. Because state 4 is an absorbing state of this chain, this probability is $Q_{6,4}^k$ where $Q_{i,j}$ are the transition probabilities of the Markov chain $\{Y_n\}$. ■

Suppose now that we want to compute the probability that the $\{X_n, n \geqslant 0\}$ chain, starting in state i, enters state j at time m without ever entering any of the states in $\mathscr{A}$, where neither i nor j is in $\mathscr{A}$. That is, for $i, j \notin \mathscr{A}$, we are interested in

$$\alpha = P(X_m = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-1 | X_0 = i)$$

Noting that the event that $X_m = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-1$ is equivalent to the event that $W_m = j$, it follows that for $i, j \notin \mathscr{A}$,

$$P(X_m = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-1 | X_0 = i) = P(W_m = j | X_0 = i)$$
$$= P(W_m = j | W_0 = i) = Q^m_{i,j}.$$

Example 4.14. Consider a Markov chain with states 1, 2, 3, 4, 5, and suppose that we want to compute

$$P(X_4 = 2, X_3 \leqslant 2, X_2 \leqslant 2, X_1 \leqslant 2 | X_0 = 1)$$

That is, we want the probability that, starting in state 1, the chain is in state 2 at time 4 and has never entered any of the states in the set $\mathscr{A} = \{3, 4, 5\}$.

To compute this probability all we need to know are the transition probabilities $P_{11}, P_{12}, P_{21}, P_{22}$. So, suppose that

$$P_{11} = 0.3 \quad P_{12} = 0.3$$
$$P_{21} = 0.1 \quad P_{22} = 0.2$$

Then we consider the Markov chain having states 1, 2, 3 (we are giving state A the name 3), and having the transition probability matrix $\mathbf{Q}$ as follows:

$$\begin{pmatrix} 0.3 & 0.3 & 0.4 \\ 0.1 & 0.2 & 0.7 \\ 0 & 0 & 1 \end{pmatrix}$$

The desired probability is Q^4_{12}. Raising $\mathbf{Q}$ to the power 4 yields the matrix

$$\begin{pmatrix} 0.0219 & 0.0285 & 0.9496 \\ 0.0095 & 0.0124 & 0.9781 \\ 0 & 0 & 1 \end{pmatrix}$$

Hence, the desired probability is $\alpha = 0.0285$. ■

When $i \notin \mathscr{A}$ but $j \in \mathscr{A}$ we can determine the probability

$$\alpha = P(X_m = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-1 | X_0 = i)$$

as follows.

$$\begin{aligned} \alpha &= \sum_{r \notin \mathscr{A}} P(X_m = j, X_{m-1} = r, X_k \notin \mathscr{A}, k = 1, \ldots, m-2 | X_0 = i) \\ &= \sum_{r \notin \mathscr{A}} P(X_m = j | X_{m-1} = r, X_k \notin \mathscr{A}, k = 1, \ldots, m-2, X_0 = i) \\ &\quad \times P(X_{m-1} = r, X_k \notin \mathscr{A}, k = 1, \ldots, m-2 | X_0 = i) \\ &= \sum_{r \notin \mathscr{A}} P_{r,j} \, P(X_{m-1} = r, X_k \notin \mathscr{A}, k = 1, \ldots, m-2 | X_0 = i) \\ &= \sum_{r \notin \mathscr{A}} P_{r,j} \, Q^{m-1}_{i,r} \end{aligned}$$

Also, when $i \in \mathscr{A}$ we could determine

$$\alpha = P(X_m = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-1 | X_0 = i)$$

by conditioning on the first transition to obtain

$$\begin{aligned}\alpha &= \sum_{r \notin \mathscr{A}} P(X_m = j, X_k \notin \mathscr{A}, \\ &\qquad k = 1, \ldots, m-1 | X_0 = i, X_1 = r) P(X_1 = r | X_0 = i) \\ &= \sum_{r \notin \mathscr{A}} P(X_{m-1} = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-2 | X_0 = r) P_{i,r}\end{aligned}$$

For instance, if $i \in \mathscr{A}$, $j \notin \mathscr{A}$ then the preceding equation yields

$$P(X_m = j, X_k \notin \mathscr{A}, k = 1, \ldots, m-1 | X_0 = i) = \sum_{r \notin \mathscr{A}} Q_{r,j}^{m-1} P_{i,r}$$

We can also compute the conditional probability of X_n given that the chain starts in state i and has not entered any state in $\mathscr{A}$ by time n, as follows. For $i, j \notin \mathscr{A}$,

$$\begin{aligned}&P\{X_n = j | X_0 = i, X_k \notin \mathscr{A}, k = 1, \ldots, n\} \\ &\quad = \frac{P\{X_n = j, X_k \notin \mathscr{A}, k = 1, \ldots, n | X_0 = i\}}{P\{X_k \notin \mathscr{A}, k = 1, \ldots, n | X_0 = i\}} = \frac{Q_{i,j}^n}{\sum_{r \notin \mathscr{A}} Q_{i,r}^n}\end{aligned}$$

Remark. So far, all of the probabilities we have considered are conditional probabilities. For instance, P_{ij}^n is the probability that the state at time n is j given that the initial state at time 0 is i. If the unconditional distribution of the state at time n is desired, it is necessary to specify the probability distribution of the initial state. Let us denote this by

$$\alpha_i \equiv P\{X_0 = i\}, \quad i \geqslant 0 \left(\sum_{i=0}^{\infty} \alpha_i = 1 \right)$$

All unconditional probabilities may be computed by conditioning on the initial state. That is,

$$\begin{aligned}P\{X_n = j\} &= \sum_{i=0}^{\infty} P\{X_n = j | X_0 = i\} P\{X_0 = i\} \\ &= \sum_{i=0}^{\infty} P_{ij}^n \alpha_i\end{aligned}$$

For instance, if $\alpha_0 = 0.4, \alpha_1 = 0.6$, in Example 4.8, then the (unconditional) probability that it will rain four days after we begin keeping weather records is

$$P\{X_4 = 0\} = 0.4 P_{00}^4 + 0.6 P_{10}^4$$

$$= (0.4)(0.5749) + (0.6)(0.5668)$$
$$= 0.5700$$

4.3 Classification of States

State j is said to be *accessible* from state i if $P_{ij}^n > 0$ for some $n \geqslant 0$. Note that this implies that state j is accessible from state i if and only if, starting in i, it is possible that the process will ever enter state j. This is true since if j is not accessible from i, then

$$\begin{aligned} P\{\text{ever be in } j | \text{start in } i\} &= P\left\{ \bigcup_{n=0}^{\infty} \{X_n = j\} \,\middle|\, X_0 = i \right\} \\ &\leq \sum_{n=0}^{\infty} P\{X_n = j | X_0 = i\} \\ &= \sum_{n=0}^{\infty} P_{ij}^n \\ &= 0 \end{aligned}$$

Two states i and j that are accessible to each other are said to *communicate*, and we write $i \leftrightarrow j$.

Note that any state communicates with itself since, by definition,

$$P_{ii}^0 = P\{X_0 = i | X_0 = i\} = 1$$

The relation of communication satisfies the following three properties:

- **(i)** State i communicates with state i, all $i \geqslant 0$.
- **(ii)** If state i communicates with state j, then state j communicates with state i.
- **(iii)** If state i communicates with state j, and state j communicates with state k, then state i communicates with state k.

Properties (i) and (ii) follow immediately from the definition of communication. To prove (iii) suppose that i communicates with j, and j communicates with k. Thus, there exist integers n and m such that $P_{ij}^n > 0, P_{jk}^m > 0$. Now by the Chapman–Kolmogorov equations, we have

$$P_{ik}^{n+m} = \sum_{r=0}^{\infty} P_{ir}^n P_{rk}^m \geqslant P_{ij}^n P_{jk}^m > 0$$

Hence, state k is accessible from state i. Similarly, we can show that state i is accessible from state k. Hence, states i and k communicate.

Two states that communicate are said to be in the same *class*. It is an easy consequence of (i), (ii), and (iii) that any two classes of states are either identical or disjoint.

In other words, the concept of communication divides the state space up into a number of separate classes. The Markov chain is said to be *irreducible* if there is only one class, that is, if all states communicate with each other.

Example 4.15. Consider the Markov chain consisting of the three states 0, 1, 2 and having transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{Vmatrix}$$

It is easy to verify that this Markov chain is irreducible. For example, it is possible to go from state 0 to state 2 since

$$0 \to 1 \to 2$$

That is, one way of getting from state 0 to state 2 is to go from state 0 to state 1 (with probability $\frac{1}{2}$) and then go from state 1 to state 2 (with probability $\frac{1}{4}$). ■

Example 4.16. Consider a Markov chain consisting of the four states 0, 1, 2, 3 and having transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

The classes of this Markov chain are $\{0, 1\}$, $\{2\}$, and $\{3\}$. Note that while state 0 (or 1) is accessible from state 2, the reverse is not true. Since state 3 is an absorbing state, that is, $P_{33} = 1$, no other state is accessible from it. ■

For any state i we let f_i denote the probability that, starting in state i, the process will ever reenter state i. State i is said to be *recurrent* if $f_i = 1$ and *transient* if $f_i < 1$.

Suppose that the process starts in state i and i is recurrent. Hence, with probability 1, the process will eventually reenter state i. However, by the definition of a Markov chain, it follows that the process will be starting over again when it reenters state i and, therefore, state i will eventually be visited again. Continual repetition of this argument leads to the conclusion that *if state i is recurrent then, starting in state i, the process will reenter state i again and again and again—in fact, infinitely often.*

On the other hand, suppose that state i is transient. Hence, each time the process enters state i there will be a positive probability, namely, $1 - f_i$, that it will never again enter that state. Therefore, starting in state i, the probability that the process will be in state i for exactly n time periods equals $f_i^{n-1}(1 - f_i)$, $n \geqslant 1$. In other words, *if state i is transient then, starting in state i, the number of time periods that the process will be in state i has a geometric distribution with finite mean* $1/(1 - f_i)$.

From the preceding two paragraphs, it follows that *state i is recurrent if and only if, starting in state i, the expected number of time periods that the process is in state i is infinite*. But, letting

$$I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$$

we have that $\sum_{n=0}^{\infty} I_n$ represents the number of periods that the process is in state i. Also,

$$\begin{aligned} E\left[\sum_{n=0}^{\infty} I_n | X_0 = i\right] &= \sum_{n=0}^{\infty} E[I_n | X_0 = i] \\ &= \sum_{n=0}^{\infty} P\{X_n = i | X_0 = i\} \\ &= \sum_{n=0}^{\infty} P_{ii}^n \end{aligned}$$

We have thus proven the following.

Proposition 4.1. *State i is*

$$\text{recurrent if } \sum_{n=1}^{\infty} P_{ii}^n = \infty,$$

$$\text{transient if } \sum_{n=1}^{\infty} P_{ii}^n < \infty$$

The argument leading to the preceding proposition is doubly important because it also shows that a transient state will only be visited a finite number of times (hence the name transient). This leads to the conclusion that in a finite-state Markov chain not all states can be transient. To see this, suppose the states are $0, 1, \ldots, M$ and suppose that they are all transient. Then after a finite amount of time (say, after time T_0) state 0 will never be visited, and after a time (say, T_1) state 1 will never be visited, and after a time (say, T_2) state 2 will never be visited, and so on. Thus, after a finite time $T = \max\{T_0, T_1, \ldots, T_M\}$ no states will be visited. But as the process must be in some state after time T we arrive at a contradiction, which shows that at least one of the states must be recurrent.

Another use of Proposition 4.1 is that it enables us to show that recurrence is a class property.

Corollary 4.2. *If state i is recurrent, and state i communicates with state j, then state j is recurrent.*

Proof. To prove this we first note that, since state i communicates with state j, there exist integers k and m such that $P_{ij}^k > 0$, $P_{ji}^m > 0$. Now, for any integer n

$$P_{jj}^{m+n+k} \geqslant P_{ji}^m P_{ii}^n P_{ij}^k$$

This follows since the left side of the preceding is the probability of going from j to j in $m+n+k$ steps, while the right side is the probability of going from j to j in $m+n+k$ steps via a path that goes from j to i in m steps, then from i to i in an additional n steps, then from i to j in an additional k steps.

From the preceding we obtain, by summing over n, that

$$\sum_{n=1}^{\infty} P_{jj}^{m+n+k} \geqslant P_{ji}^m P_{ij}^k \sum_{n=1}^{\infty} P_{ii}^n = \infty$$

since $P_{ji}^m P_{ij}^k > 0$ and $\sum_{n=1}^{\infty} P_{ii}^n$ is infinite since state i is recurrent. Thus, by Proposition 4.1 it follows that state j is also recurrent. ■

Remarks. **(i)** Corollary 4.2 also implies that transience is a class property. For if state i is transient and communicates with state j, then state j must also be transient. For if j were recurrent then, by Corollary 4.2, i would also be recurrent and hence could not be transient.

(ii) Corollary 4.2 along with our previous result that not all states in a finite Markov chain can be transient leads to the conclusion that all states of a finite irreducible Markov chain are recurrent.

Example 4.17. Let the Markov chain consisting of the states 0, 1, 2, 3 have the transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{Vmatrix}$$

Determine which states are transient and which are recurrent.

Solution: It is a simple matter to check that all states communicate and, hence, since this is a finite chain, all states must be recurrent. ■

Example 4.18. Consider the Markov chain having states 0, 1, 2, 3, 4 and

$$\mathbf{P} = \begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{Vmatrix}$$

Determine the recurrent state.

Solution: This chain consists of the three classes $\{0, 1\}$, $\{2, 3\}$, and $\{4\}$. The first two classes are recurrent and the third transient. ■

Example 4.19 (A Random Walk)**.** Consider a Markov chain whose state space consists of the integers $i = 0, \pm 1, \pm 2, \ldots$, and has transition probabilities given by

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i = 0, \pm 1, \pm 2, \ldots$$

where $0 < p < 1$. In other words, on each transition the process either moves one step to the right (with probability p) or one step to the left (with probability $1 - p$). One colorful interpretation of this process is that it represents the wanderings of a drunken man as he walks along a straight line. Another is that it represents the winnings of a gambler who on each play of the game either wins or loses one dollar.

Since all states clearly communicate, it follows from Corollary 4.2 that they are either all transient or all recurrent. So let us consider state 0 and attempt to determine if $\sum_{n=1}^{\infty} P_{00}^n$ is finite or infinite.

Since it is impossible to be even (using the gambling model interpretation) after an odd number of plays we must, of course, have that

$$P_{00}^{2n-1} = 0, \quad n = 1, 2, \ldots$$

On the other hand, we would be even after $2n$ trials if and only if we won n of these and lost n of these. Because each play of the game results in a win with probability p and a loss with probability $1 - p$, the desired probability is thus the binomial probability

$$P_{00}^{2n} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!} (p(1-p))^n, \quad n = 1, 2, 3, \ldots$$

By using an approximation, due to Stirling, which asserts that

$$n! \sim n^{n+1/2} e^{-n} \sqrt{2\pi} \tag{4.3}$$

where we say that $a_n \sim b_n$ when $\lim_{n\to\infty} a_n/b_n = 1$, we obtain

$$\binom{2n}{n} \sim \frac{(2n)^{2n+1/2}\, e^{-2n}\, \sqrt{2\pi}}{n^{2n+1}\, e^{-2n}\, (2\pi)} = \frac{2^{2n}}{\sqrt{n\pi}}$$

Hence,

$$P_{00}^{2n} \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

Now it is easy to verify, for positive a_n, b_n, that if $a_n \sim b_n$, then $\sum_n a_n < \infty$ if and only if $\sum_n b_n < \infty$. Hence, $\sum_{n=1}^{\infty} P_{00}^n$ will converge if and only if

$$\sum_{n=1}^{\infty} \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

does. However, $4p(1-p) \leq 1$ with equality holding if and only if $p = \frac{1}{2}$. Hence, $\sum_{n=1}^{\infty} P_{00}^n = \infty$ if and only if $p = \frac{1}{2}$. Thus, the chain is recurrent when $p = \frac{1}{2}$ and transient if $p \neq \frac{1}{2}$.

When $p = \frac{1}{2}$, the preceding process is called a *symmetric random walk*. We could also look at symmetric random walks in more than one dimension. For instance, in the two-dimensional symmetric random walk the process would, at each transition, either take one step to the left, right, up, or down, each having probability $\frac{1}{4}$. That is, the state is the pair of integers (i, j) and the transition probabilities are given by

$$P_{(i,j),(i+1,j)} = P_{(i,j),(i-1,j)} = P_{(i,j),(i,j+1)} = P_{(i,j),(i,j-1)} = \tfrac{1}{4}$$

By using the same method as in the one-dimensional case, we now show that this Markov chain is also recurrent.

Since the preceding chain is irreducible, it follows that all states will be recurrent if state $\mathbf{0} = (0, 0)$ is recurrent. So consider $P_{\mathbf{00}}^{2n}$. Now after $2n$ steps, the chain will be back in its original location if for some $i, 0 \leqslant i \leqslant n$, the $2n$ steps consist of i steps to the left, i to the right, $n - i$ up, and $n - i$ down. Since each step will be either of these four types with probability $\frac{1}{4}$, it follows that the desired probability is a multinomial probability. That is,

$$\begin{aligned}
P_{\mathbf{00}}^{2n} &= \sum_{i=0}^{n} \frac{(2n)!}{i!i!(n-i)!(n-i)!} \left(\frac{1}{4}\right)^{2n} \\
&= \sum_{i=0}^{n} \frac{(2n)!}{n!n!} \frac{n!}{(n-i)!i!} \frac{n!}{(n-i)!i!} \left(\frac{1}{4}\right)^{2n} \\
&= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \sum_{i=0}^{n} \binom{n}{i} \binom{n}{n-i} \\
&= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \binom{2n}{n} \qquad (4.4)
\end{aligned}$$

where the last equality uses the combinatorial identity

$$\binom{2n}{n} = \sum_{i=0}^{n} \binom{n}{i} \binom{n}{n-i}$$

which follows upon noting that both sides represent the number of subgroups of size n one can select from a set of n white and n black objects. Now,

$$\begin{aligned}
\binom{2n}{n} &= \frac{(2n)!}{n!n!} \\
&\sim \frac{(2n)^{2n+1/2} e^{-2n} \sqrt{2\pi}}{n^{2n+1} e^{-2n} (2\pi)} \quad \text{by Stirling's approximation}
\end{aligned}$$

$$= \frac{4^n}{\sqrt{\pi n}}$$

Hence, from Eq. (4.4) we see that

$$P_{00}^{2n} \sim \frac{1}{\pi n}$$

which shows that $\sum_n P_{00}^{2n} = \infty$, and thus all states are recurrent.

Interestingly enough, whereas the symmetric random walks in one and two dimensions are both recurrent, all higher-dimensional symmetric random walks turn out to be transient. (For instance, the three-dimensional symmetric random walk is at each transition equally likely to move in any of six ways—either to the left, right, up, down, in, or out.) ■

Remark. For the one-dimensional random walk of Example 4.19 here is a direct argument for establishing recurrence in the symmetric case, and for determining the probability that it ever returns to state 0 in the nonsymmetric case. Let

$$\beta = P\{\text{ever return to } 0\}$$

To determine β, start by conditioning on the initial transition to obtain

$$\beta = P\{\text{ever return to } 0 | X_1 = 1\}p + P\{\text{ever return to } 0 | X_1 = -1\}(1 - p) \quad (4.5)$$

Now, let α denote the probability that the Markov chain will ever return to state 0 given that it is currently in state 1. Because the Markov chain will always increase by 1 with probability p or decrease by 1 with probability $1 - p$ no matter what its current state, note that α is also the probability that the Markov chain currently in state i will ever enter state $i - 1$, for any i. To obtain an equation for α, condition on the next transition to obtain

$$\begin{aligned}\alpha &= P\{\text{ever return} | X_1 = 1, X_2 = 0\}(1 - p) + P\{\text{ever return} | X_1 = 1, X_2 = 2\}p \\ &= 1 - p + P\{\text{ever return} | X_1 = 1, X_2 = 2\}p \\ &= 1 - p + p\alpha^2\end{aligned}$$

where the final equation follows by noting that in order for the chain to ever go from state 2 to state 0 it must first go to state 1—and the probability of that ever happening is α—and if it does eventually go to state 1 then it must still go to state 0—and the conditional probability of that ever happening is also α. Therefore,

$$\alpha = 1 - p + p\alpha^2$$

The two roots of this equation are $\alpha = 1$ and $\alpha = (1 - p)/p$. Consequently, in the case of the symmetric random walk where $p = 1/2$ we can conclude that $\alpha = 1$. By symmetry, the probability that the symmetric random walk will ever enter state 0 given

that it is currently in state -1 is also 1, proving that the symmetric random walk is recurrent.

Suppose now that $p > 1/2$. In this case, it can be shown (see Exercise 17 at the end of this chapter) that $P\{\text{ever return to } 0 | X_1 = -1\} = 1$. Consequently, Eq. (4.5) reduces to

$$\beta = \alpha p + 1 - p$$

Because the random walk is transient in this case we know that $\beta < 1$, showing that $\alpha \neq 1$. Therefore, $\alpha = (1 - p)/p$, yielding that

$$\beta = 2(1 - p), \quad p > 1/2$$

Similarly, when $p < 1/2$ we can show that $\beta = 2p$. Thus, in general

$$P\{\text{ever return to } 0\} = 2\min(p, 1 - p)$$ ■

In our next example we use the recurrence of the symmetric random walk to construct an example where $E[\sum_{n=1}^{\infty} X_n] \neq \sum_{n=1}^{\infty} E[X_n]$.

Example 4.20. Whereas it is true that $E[\sum_{n=1}^{\infty} X_n] = \sum_{n=1}^{\infty} E[X_n]$ when the random variables $X_n, n \geq 1$ are all nonnegative, it is not true in general. For an example where it does not hold, suppose that $Y_1, Y_2, \ldots$ are independent and identically distributed with $P(Y_n = 1) = P(Y_n = -1) = 1/2, n \geq 1$. Note that $E[Y_n] = 0$. Let

$$N = \min(k : Y_1 + \ldots + Y_k = 1)$$

and note that it follows from the fact that the symmetric random walk is recurrent that N will be finite with probability 1. Now, let

$$I_n = \begin{cases} 1, & \text{if } n \leq N \\ 0, & \text{if } n > N \end{cases}$$

Because $I_n = 1$ if $N > n - 1$, and it is equal to 0 otherwise, it follows that the value of I_n is determined by $Y_1, \ldots, Y_{n-1}$. Indeed, because N is defined to be the first time the sum of the Y's is equal to 1, it follows that

$$\{I_n = 1\} = \{N > n - 1\} = \{Y_1 \neq 1, Y_1 + Y_2 \neq 1, \ldots, Y_1 + \ldots + Y_{n-1} \neq 1\}$$

which shows that I_n and Y_n are independent. Now, define $X_n, n \geq 1$, by

$$X_n = Y_n I_n = \begin{cases} Y_n, & \text{if } n \leq N \\ 0, & \text{if } n > N \end{cases}$$

By the independence of Y_n and I_n,

$$E[X_n] = E[Y_n]\, E[I_n] = 0$$

showing that

$$\sum_{n=1}^{\infty} E[X_n] = 0$$

However,

$$\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} Y_n I_n = \sum_{n=1}^{N} Y_n = 1$$

and so

$$E[\sum_{n=1}^{\infty} X_n] = 1$$

Thus,

$$E[\sum_{n=1}^{\infty} X_n] = 1 \text{ and } \sum_{n=1}^{\infty} E[X_n] = 0$$ ■

Example 4.21 (On the Ultimate Instability of the Aloha Protocol)**.** Consider a communications facility in which the numbers of messages arriving during each of the time periods $n = 1, 2, \ldots$ are independent and identically distributed random variables. Let $a_i = P\{i \text{ arrivals}\}$, and suppose that $a_0 + a_1 < 1$. Each arriving message will transmit at the end of the period in which it arrives. If exactly one message is transmitted, then the transmission is successful and the message leaves the system. However, if at any time two or more messages simultaneously transmit, then a collision is deemed to occur and these messages remain in the system. Once a message is involved in a collision it will, independently of all else, transmit at the end of each additional period with probability p—the so-called Aloha protocol (because it was first instituted at the University of Hawaii). We will show that such a system is asymptotically unstable in the sense that the number of successful transmissions will, with probability 1, be finite.

To begin let X_n denote the number of messages in the facility at the beginning of the nth period, and note that $\{X_n, n \geqslant 0\}$ is a Markov chain. Now for $k \geqslant 0$ define the indicator variables I_k by

$$I_k = \begin{cases} 1, & \text{if the first time that the chain departs state } k \text{ it} \\ & \text{directly goes to state } k-1 \\ 0, & \text{otherwise} \end{cases}$$

and let it be 0 if the system is never in state $k, k \geqslant 0$. (For instance, if the successive states are $0, 1, 3, 4, \ldots$, then $I_3 = 0$ since when the chain first departs state 3 it goes to state 4; whereas, if they are $0, 3, 3, 2, \ldots$, then $I_3 = 1$ since this time it goes to state 2.)

Now,

$$E\left[\sum_{k=0}^{\infty} I_k\right] = \sum_{k=0}^{\infty} E[I_k]$$
$$= \sum_{k=0}^{\infty} P\{I_k = 1\}$$
$$\leqslant \sum_{k=0}^{\infty} P\{I_k = 1 | k \text{ is ever visited}\} \tag{4.6}$$

Now, $P\{I_k = 1|k \text{ is ever visited}\}$ is the probability that when state k is departed the next state is $k-1$. That is, it is the conditional probability that a transition from k is to $k-1$ given that it is not back into k, and so

$$P\{I_k = 1|k \text{ is ever visited}\} = \frac{P_{k,k-1}}{1 - P_{k,k}}$$

Because

$$P_{k,k-1} = a_0 kp(1-p)^{k-1},$$
$$P_{k,k} = a_0[1 - kp(1-p)^{k-1}] + a_1(1-p)^k$$

which is seen by noting that if there are k messages present on the beginning of a day, then (a) there will be $k-1$ at the beginning of the next day if there are no new messages that day and exactly one of the k messages transmits; and (b) there will be k at the beginning of the next day if either

(i) there are no new messages and it is not the case that exactly one of the existing k messages transmits, or

(ii) there is exactly one new message (which automatically transmits) and none of the other k messages transmits.

Substitution of the preceding into Eq. (4.6) yields

$$E\left[\sum_{k=0}^{\infty} I_k\right] \leqslant \sum_{k=0}^{\infty} \frac{a_0 kp(1-p)^{k-1}}{1 - a_0[1 - kp(1-p)^{k-1}] - a_1(1-p)^k}$$
$$< \infty$$

where the convergence follows by noting that when k is large the denominator of the expression in the preceding sum converges to $1 - a_0$ and so the convergence or divergence of the sum is determined by whether or not the sum of the terms in the numerator converge and $\sum_{k=0}^{\infty} k(1-p)^{k-1} < \infty$.

Hence, $E[\sum_{k=0}^{\infty} I_k] < \infty$, which implies that $\sum_{k=0}^{\infty} I_k < \infty$ with probability 1 (for if there was a positive probability that $\sum_{k=0}^{\infty} I_k$ could be ∞, then its mean would be ∞). Hence, with probability 1, there will be only a finite number of states that are

initially departed via a successful transmission; or equivalently, there will be some finite integer N such that whenever there are N or more messages in the system, there will never again be a successful transmission. From this (and the fact that such higher states will eventually be reached—why?) it follows that, with probability 1, there will only be a finite number of successful transmissions. ■

Remark. For a (slightly less than rigorous) probabilistic proof of Stirling's approximation, let $X_1, X_2, \ldots$ be independent Poisson random variables each having mean 1. Let $S_n = \sum_{i=1}^{n} X_i$, and note that both the mean and variance of S_n are equal to n. Now,

$$\begin{aligned} P\{S_n = n\} &= P\{n-1 < S_n \leqslant n\} \\ &= P\{-1/\sqrt{n} < (S_n - n)/\sqrt{n} \leqslant 0\} \\ &\approx \int_{-1/\sqrt{n}}^{0} (2\pi)^{-1/2} e^{-x^2/2}\, dx \qquad \text{when } n \text{ is large, by the central limit theorem} \\ &\approx (2\pi)^{-1/2}(1/\sqrt{n}) \\ &= (2\pi n)^{-1/2} \end{aligned}$$

But S_n is Poisson with mean n, and so

$$P\{S_n = n\} = \frac{e^{-n} n^n}{n!}$$

Hence, for n large

$$\frac{e^{-n} n^n}{n!} \approx (2\pi n)^{-1/2}$$

or, equivalently

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}$$

which is Stirling's approximation.

4.4 Long-Run Proportions and Limiting Probabilities

For pairs of states $i \neq j$, let $f_{i,j}$ denote the probability that the Markov chain, starting in state i, will ever make a transition into state j. That is,

$$f_{i,j} = P(X_n = j \text{ for some } n > 0 | X_0 = i)$$

We then have the following result.

Proposition 4.3. *If i is recurrent and i communicates with j, then $f_{i,j} = 1$.*

Proof. Because i and j communicate there is a value n such that $P_{i,j}^n > 0$. Let $X_0 = i$ and say that the first opportunity is a success if $X_n = j$, and note that the first opportunity is a success with probability $P_{i,j}^n > 0$. If the first opportunity is not a success then consider the next time (after time n) that the chain enters state i. (Because state i is recurrent we can be certain that it will eventually reenter state i after time n.) Say that the second opportunity is a success if n time periods later the Markov chain is in state j. If the second opportunity is not a success then wait until the next time the chain enters state i and say that the third opportunity is a success if n time periods later the Markov chain is in state j. Continuing in this manner, we can define an unlimited number of opportunities, each of which is a success with the same positive probability $P_{i,j}^n$. Because the number of opportunities until the first success occurs is geometric with parameter $P_{i,j}^n$, it follows that with probability 1 a success will eventually occur and so, with probability 1, state j will eventually be entered. ■

If state j is recurrent, let m_j denote the expected number of transitions that it takes the Markov chain when starting in state j to return to that state. That is, with

$$N_j = \min\{n > 0 : X_n = j\}$$

equal to the number of transitions until the Markov chain makes a transition into state j,

$$m_j = E[N_j | X_0 = j]$$

Definition. Say that the recurrent state j is *positive recurrent* if $m_j < \infty$ and say that it is *null recurrent* if $m_j = \infty$.

Now suppose that the Markov chain is irreducible and recurrent. In this case we now show that the long-run proportion of time that the chain spends in state j is equal to $1/m_j$. That is, letting π_j denote the long-run proportion of time that the Markov chain is in state j, we have the following proposition.

Proposition 4.4. *If the Markov chain is irreducible and recurrent, then for any initial state*

$$\pi_j = 1/m_j$$

Proof. Suppose that the Markov chain starts in state i, and let T_1 denote the number of transitions until the chain enters state j; then let T_2 denote the additional number of transitions from time T_1 until the Markov chain next enters state j; then let T_3 denote the additional number of transitions from time $T_1 + T_2$ until the Markov chain next enters state j, and so on. Note that T_1 is finite because Proposition 4.3 tells us that with probability 1 a transition into j will eventually occur. Also, for $n \geq 2$, because T_n is the number of transitions between the $(n-1)$th and the nth transition into state j, it follows from the Markovian property that $T_2, T_3, \ldots$ are independent and identically distributed with mean m_j. Because the nth transition into state j occurs at time $T_1 +$

$\ldots + T_n$ we obtain that π_j, the long-run proportion of time that the chain is in state j, is

$$\begin{aligned}\pi_j &= \lim_{n\to\infty} \frac{n}{\sum_{i=1}^{n} T_i} \\ &= \lim_{n\to\infty} \frac{1}{\frac{1}{n}\sum_{i=1}^{n} T_i} \\ &= \lim_{n\to\infty} \frac{1}{\frac{T_1}{n} + \frac{T_2+\ldots+T_n}{n}} \\ &= \frac{1}{m_j}\end{aligned}$$

where the last equality follows because $\lim_{n\to\infty} T_1/n = 0$ and, from the strong law of large numbers, $\lim_{n\to\infty} \frac{T_2+\ldots+T_n}{n} = \lim_{n\to\infty} \frac{T_2+\ldots+T_n}{n-1}\,\frac{n-1}{n} = m_j$. ■

Because $m_j < \infty$ is equivalent to $1/m_j > 0$, it follows from the preceding that state j is positive recurrent if and only if $\pi_j > 0$. We now exploit this to show that positive recurrence is a class property.

Proposition 4.5. *If i is positive recurrent and $i \leftrightarrow j$ then j is positive recurrent.*

Proof. Suppose that i is positive recurrent and that $i \leftrightarrow j$. Now, let n be such that $P_{i,j}^n > 0$. Because π_i is the long-run proportion of time that the chain is in state i, and $P_{i,j}^n$ is the long-run proportion of time when the Markov chain is in state i that it will be in state j after n transitions

$$\begin{aligned}\pi_i P_{i,j}^n &= \text{long-run proportion of time the chain is in } i \\ &\qquad \text{and will be in } j \text{ after } n \text{ transitions} \\ &= \text{long-run proportion of time the chain is in } j \\ &\qquad \text{and was in } i \text{ } n \text{ transitions ago} \\ &\le \text{long-run proportion of time the chain is in } j\end{aligned}$$

Hence, $\pi_j \ge \pi_i P_{i,j}^n > 0$, showing that j is positive recurrent. ■

Remarks. **(i)** It follows from the preceding result that null recurrence is also a class property. For suppose that i is null recurrent and $i \leftrightarrow j$. Because i is recurrent and $i \leftrightarrow j$ we can conclude that j is recurrent. But if j were positive recurrent then by the preceding proposition i would also be positive recurrent. Because i is not positive recurrent, neither is j.

(ii) An irreducible finite state Markov chain must be positive recurrent. For we know that such a chain must be recurrent; hence, all its states are either positive recurrent or null recurrent. If they were null recurrent then all the long run proportions would equal 0, which is impossible when there are only a finite number of states. Consequently, we can conclude that the chain is positive recurrent.

(iii) The classical example of a null recurrent Markov chain is the one dimensional symmetric random walk of Example 4.18. One way to show it is null recurrent is to argue that the mean time to return to a state is infinite. (For another argument, see Exercise 39.) To show this, let $m_{i,j}$ denote the mean number of transitions, starting in state i, until a transition into state j occurs. Now, in Example 3.16, it was shown that the mean number of transitions to go from state 1 to either 0 or n is $n-1$, implying that

$$m_{1,0} \geq n-1.$$

As the preceding is true for all n, we obtain upon letting $n \to \infty$ that

$$m_{1,0} = \infty.$$

Conditioning on the result of the first bet gives

$$m_{0,0} = 1 + m_{1,0}\frac{1}{2} + m_{-1,0}\frac{1}{2}.$$

Hence, $m_{0,0} = \infty$, establishing that the symmetric random walk is null recurrent. ■

To determine the long-run proportions $\{\pi_j, j \geq 1\}$, note, because π_i is the long-run proportion of transitions that come from state i, that

$$\pi_i P_{i,j} = \text{long-run proportion of transitions that go from state } i \text{ to state } j$$

Summing the preceding over all i now yields that

$$\pi_j = \sum_i \pi_i P_{i,j}$$

Indeed, the following important theorem can be proven.

Theorem 4.1. *Consider an irreducible Markov chain. If the chain is positive recurrent then the long-run proportions are the unique solution of the equations*

$$\begin{aligned} \pi_j &= \sum_i \pi_i P_{i,j}, \quad j \geq 1 \\ \sum_j \pi_j &= 1 \end{aligned} \tag{4.7}$$

Moreover, if there is no solution of the preceding linear equations, then the Markov chain is either transient or null recurrent and all $\pi_j = 0$.

Example 4.22. Consider Example 4.1, in which we assume that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain

tomorrow with probability β. If we say that the state is 0 when it rains and 1 when it does not rain, then by Theorem 4.1 the long-run proportions π_0 and π_1 are given by

$$\begin{aligned}\pi_0 &= \alpha\pi_0 + \beta\pi_1,\\ \pi_1 &= (1-\alpha)\pi_0 + (1-\beta)\pi_1,\\ \pi_0 + \pi_1 &= 1\end{aligned}$$

which yields that

$$\pi_0 = \frac{\beta}{1+\beta-\alpha}, \quad \pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

For example if $\alpha = 0.7$ and $\beta = 0.4$, then the long-run proportion of rain is $\pi_0 = \frac{4}{7} = 0.571$. ■

Example 4.23. Consider Example 4.3 in which the mood of an individual is considered as a three-state Markov chain having a transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{Vmatrix}$$

In the long run, what proportion of time is the process in each of the three states?

Solution: The long run proportions $\pi_i, i = 0, 1, 2$, are obtained by solving the set of equations in Eq. (4.7). In this case these equations are

$$\begin{aligned}\pi_0 &= 0.5\pi_0 + 0.3\pi_1 + 0.2\pi_2,\\ \pi_1 &= 0.4\pi_0 + 0.4\pi_1 + 0.3\pi_2,\\ \pi_2 &= 0.1\pi_0 + 0.3\pi_1 + 0.5\pi_2,\\ \pi_0 + \pi_1 + \pi_2 &= 1\end{aligned}$$

Solving yields

$$\pi_0 = \tfrac{21}{62}, \quad \pi_1 = \tfrac{23}{62}, \quad \pi_2 = \tfrac{18}{62}$$ ■

Example 4.24 (A Model of Class Mobility)**.** A problem of interest to sociologists is to determine the proportion of society that has an upper- or lower-class occupation. One possible mathematical model would be to assume that transitions between social classes of the successive generations in a family can be regarded as transitions of a Markov chain. That is, we assume that the occupation of a child depends only on his or her parent's occupation. Let us suppose that such a model is appropriate and that the transition probability matrix is given by

$$\mathbf{P} = \begin{Vmatrix} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.70 & 0.25 \\ 0.01 & 0.50 & 0.49 \end{Vmatrix} \tag{4.8}$$

That is, for instance, we suppose that the child of a middle-class worker will attain an upper-, middle-, or lower-class occupation with respective probabilities 0.05, 0.70, 0.25.

The long-run proportions π_i thus satisfy

$$\begin{aligned}\pi_0 &= 0.45\pi_0 + 0.05\pi_1 + 0.01\pi_2,\\ \pi_1 &= 0.48\pi_0 + 0.70\pi_1 + 0.50\pi_2,\\ \pi_2 &= 0.07\pi_0 + 0.25\pi_1 + 0.49\pi_2,\\ \pi_0 + \pi_1 + \pi_2 &= 1\end{aligned}$$

Hence,

$$\pi_0 = 0.07, \quad \pi_1 = 0.62, \quad \pi_2 = 0.31$$

In other words, a society in which social mobility between classes can be described by a Markov chain with transition probability matrix given by Eq. (4.8) has, in the long run, 7 percent of its people in upper-class jobs, 62 percent of its people in middle-class jobs, and 31 percent in lower-class jobs.

Example 4.25 (The Hardy–Weinberg Law and a Markov Chain in Genetics)**.** Consider a large population of individuals, each of whom possesses a particular pair of genes, of which each individual gene is classified as being of type A or type a. Assume that the proportions of individuals whose gene pairs are AA, aa, or Aa are, respectively, p_0, q_0, and r_0 ($p_0 + q_0 + r_0 = 1$). When two individuals mate, each contributes one of his or her genes, chosen at random, to the resultant offspring. Assuming that the mating occurs at random, in that each individual is equally likely to mate with any other individual, we are interested in determining the proportions of individuals in the next generation whose genes are AA, aa, or Aa. Calling these proportions p, q, and r, they are easily obtained by focusing attention on an individual of the next generation and then determining the probabilities for the gene pair of that individual.

To begin, note that randomly choosing a parent and then randomly choosing one of its genes is equivalent to just randomly choosing a gene from the total gene population. By conditioning on the gene pair of the parent, we see that a randomly chosen gene will be type A with probability

$$\begin{aligned}P\{A\} &= P\{A|AA\}p_0 + P\{A|aa\}q_0 + P\{A|Aa\}r_0\\ &= p_0 + r_0/2\end{aligned}$$

Similarly, it will be type a with probability

$$P\{a\} = q_0 + r_0/2$$

Thus, under random mating a randomly chosen member of the next generation will be type AA with probability p, where

$$p = P\{A\}P\{A\} = (p_0 + r_0/2)^2$$

Similarly, the randomly chosen member will be type aa with probability

$$q = P\{a\}P\{a\} = (q_0 + r_0/2)^2$$

and will be type Aa with probability

$$r = 2P\{A\}P\{a\} = 2(p_0 + r_0/2)(q_0 + r_0/2)$$

Since each member of the next generation will independently be of each of the three gene types with probabilities p, q, r, it follows that the percentages of the members of the next generation that are of type AA, aa, or Aa are respectively p, q, and r.

If we now consider the total gene pool of this next generation, then $p + r/2$, the fraction of its genes that are A, will be unchanged from the previous generation. This follows either by arguing that the total gene pool has not changed from generation to generation or by the following simple algebra:

$$\begin{aligned} p + r/2 &= (p_0 + r_0/2)^2 + (p_0 + r_0/2)(q_0 + r_0/2) \\ &= (p_0 + r_0/2)[p_0 + r_0/2 + q_0 + r_0/2] \\ &= p_0 + r_0/2 \quad \text{since } p_0 + r_0 + q_0 = 1 \\ &= P\{A\} \end{aligned} \tag{4.9}$$

Thus, the fractions of the gene pool that are A and a are the same as in the initial generation. From this it follows that, under random mating, in all successive generations after the initial one the percentages of the population having gene pairs AA, aa, and Aa will remain fixed at the values p, q, and r. This is known as the *Hardy–Weinberg law*.

Suppose now that the gene pair population has stabilized in the percentages p, q, r, and let us follow the genetic history of a single individual and her descendants. (For simplicity, assume that each individual has exactly one offspring.) So, for a given individual, let X_n denote the genetic state of her descendant in the nth generation. The transition probability matrix of this Markov chain, namely,

$$\begin{array}{c} \\ AA \\ aa \\ Aa \end{array} \begin{array}{c} \begin{array}{ccc} AA & aa & Aa \end{array} \\ \left\| \begin{array}{ccc} p + \frac{r}{2} & 0 & q + \frac{r}{2} \\ 0 & q + \frac{r}{2} & p + \frac{r}{2} \\ \frac{p}{2} + \frac{r}{4} & \frac{q}{2} + \frac{r}{4} & \frac{p}{2} + \frac{q}{2} + \frac{r}{2} \end{array} \right\| \end{array}$$

is easily verified by conditioning on the state of the randomly chosen mate. It is quite intuitive (why?) that the limiting probabilities for this Markov chain (which also equal the fractions of the individual's descendants that are in each of the three genetic states) should just be p, q, and r. To verify this we must show that they satisfy Theorem 4.1.

Because one of the equations in Theorem 4.1 is redundant, it suffices to show that

$$p = p\left(p + \frac{r}{2}\right) + r\left(\frac{p}{2} + \frac{r}{4}\right) = \left(p + \frac{r}{2}\right)^2,$$
$$q = q\left(q + \frac{r}{2}\right) + r\left(\frac{q}{2} + \frac{r}{4}\right) = \left(q + \frac{r}{2}\right)^2,$$
$$p + q + r = 1$$

But this follows from Eq. (4.9), and thus the result is established. ■

Example 4.26. Suppose that a production process changes states in accordance with an irreducible, positive recurrent Markov chain having transition probabilities $P_{ij}, i, j = 1, \ldots, n$, and suppose that certain of the states are considered acceptable and the remaining unacceptable. Let A denote the acceptable states and A^c the unacceptable ones. If the production process is said to be "up" when in an acceptable state and "down" when in an unacceptable state, determine

1. the rate at which the production process goes from up to down (that is, the rate of breakdowns);
2. the average length of time the process remains down when it goes down; and
3. the average length of time the process remains up when it goes up.

Solution: Let $\pi_k, k = 1, \ldots, n$, denote the long-run proportions. Now for $i \in A$ and $j \in A^c$ the rate at which the process enters state j from state i is

$$\text{rate enter } j \text{ from } i = \pi_i P_{ij}$$

and so the rate at which the production process enters state j from an acceptable state is

$$\text{rate enter } j \text{ from } A = \sum_{i \in A} \pi_i P_{ij}$$

Hence, the rate at which it enters an unacceptable state from an acceptable one (which is the rate at which breakdowns occur) is

$$\text{rate breakdowns occur} = \sum_{j \in A^c} \sum_{i \in A} \pi_i P_{ij} \tag{4.10}$$

Now let $\bar{U}$ and $\bar{D}$ denote the average time the process remains up when it goes up and down when it goes down. Because there is a single breakdown every $\bar{U} + \bar{D}$ time units on the average, it follows heuristically that

$$\text{rate at which breakdowns occur} = \frac{1}{\bar{U} + \bar{D}}$$

and so from Eq. (4.10),

$$\frac{1}{\bar{U} + \bar{D}} = \sum_{j \in A^c} \sum_{i \in A} \pi_i P_{ij} \tag{4.11}$$

To obtain a second equation relating $\bar{U}$ and $\bar{D}$, consider the percentage of time the process is up, which, of course, is equal to $\sum_{i\in A}\pi_i$. However, since the process is up on the average $\bar{U}$ out of every $\bar{U}+\bar{D}$ time units, it follows (again somewhat heuristically) that the

$$\text{proportion of up time} = \frac{\bar{U}}{\bar{U}+\bar{D}}$$

and so

$$\frac{\bar{U}}{\bar{U}+\bar{D}} = \sum_{i\in A}\pi_i \tag{4.12}$$

Hence, from Eqs. (4.11) and (4.12) we obtain

$$\begin{aligned}\bar{U} &= \frac{\sum_{i\in A}\pi_i}{\sum_{j\in A^c}\sum_{i\in A}\pi_i P_{ij}},\\ \bar{D} &= \frac{1-\sum_{i\in A}\pi_i}{\sum_{j\in A^c}\sum_{i\in A}\pi_i P_{ij}}\\ &= \frac{\sum_{i\in A^c}\pi_i}{\sum_{j\in A^c}\sum_{i\in A}\pi_i P_{ij}}\end{aligned}$$

For example, suppose the transition probability matrix is

$$\mathbf{P} = \begin{Vmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{Vmatrix}$$

where the acceptable (up) states are 1, 2 and the unacceptable (down) ones are 3, 4. The long-run proportions satisfy

$$\begin{aligned}\pi_1 &= \pi_1\tfrac{1}{4} + \pi_3\tfrac{1}{4} + \pi_4\tfrac{1}{4},\\ \pi_2 &= \pi_1\tfrac{1}{4} + \pi_2\tfrac{1}{4} + \pi_3\tfrac{1}{4} + \pi_4\tfrac{1}{4},\\ \pi_3 &= \pi_1\tfrac{1}{2} + \pi_2\tfrac{1}{2} + \pi_3\tfrac{1}{4},\\ \pi_1+\pi_2+\pi_3+\pi_4 &= 1\end{aligned}$$

These solve to yield

$$\pi_1 = \tfrac{3}{16}, \quad \pi_2 = \tfrac{1}{4}, \quad \pi_3 = \tfrac{14}{48}, \quad \pi_4 = \tfrac{13}{48}$$

and thus

$$\begin{aligned}\text{rate of breakdowns} &= \pi_1(P_{13} + P_{14}) + \pi_2(P_{23} + P_{24}) \\ &= \tfrac{9}{32}, \\ \bar{U} &= \tfrac{14}{9} \quad \text{and} \quad \bar{D} = 2\end{aligned}$$

Hence, on the average, breakdowns occur about $\frac{9}{32}$ (or 28 percent) of the time. They last, on the average, 2 time units, and then there follows a stretch of (on the average) $\frac{14}{9}$ time units when the system is up. ■

The long run proportions π_j, $j \geqslant 0$, are often called *stationary* probabilities. The reason being that if the initial state is chosen according to the probabilities π_j, $j \geqslant 0$, then the probability of being in state j at any time n is also equal to π_j. That is, if

$$P\{X_0 = j\} = \pi_j, \quad j \geqslant 0$$

then

$$P\{X_n = j\} = \pi_j \quad \text{for all } n, \ j \geqslant 0$$

The preceding is easily proven by induction, for it is true when $n = 0$ and if we suppose it true for $n - 1$, then writing

$$\begin{aligned}P\{X_n = j\} &= \sum_i P\{X_n = j | X_{n-1} = i\} P\{X_{n-1} = i\} \\ &= \sum_i P_{ij}\pi_i \quad \text{by the induction hypothesis} \\ &= \pi_j \qquad\qquad \text{by Theorem 4.1}\end{aligned}$$

Example 4.27. Suppose the numbers of families that check into a hotel on successive days are independent Poisson random variables with mean λ. Also suppose that the number of days that a family stays in the hotel is a geometric random variable with parameter $p, 0 < p < 1$. (Thus, a family who spent the previous night in the hotel will, independently of how long they have already spent in the hotel, check out the next day with probability p.) Also suppose that all families act independently of each other. Under these conditions it is easy to see that if X_n denotes the number of families that are checked in the hotel at the beginning of day n then $\{X_n, n \geq 0\}$ is a Markov chain. Find

(a) the transition probabilities of this Markov chain;
(b) $E[X_n | X_0 = i]$;
(c) the stationary probabilities of this Markov chain.

Solution: (a) To find $P_{i,j}$, suppose there are i families checked into the hotel at the beginning of a day. Because each of these i families will stay for another day with probability $q = 1 - p$ it follows that R_i, the number of these families that remain another day, is a binomial (i, q) random variable. So, letting N be the number of new families that check in that day, we see that

$$P_{i,j} = P(R_i + N = j)$$

Conditioning on R_i and using that N is Poisson with mean λ, we obtain

$$\begin{aligned}
P_{i,j} &= \sum_{k=0}^{i} P(R_i + N = j | R_i = k) \binom{i}{k} q^k p^{i-k} \\
&= \sum_{k=0}^{i} P(N = j - k | R_i = k) \binom{i}{k} q^k p^{i-k} \\
&= \sum_{k=0}^{\min(i,j)} P(N = j - k) \binom{i}{k} q^k p^{i-k} \\
&= \sum_{k=0}^{\min(i,j)} e^{-\lambda} \frac{\lambda^{j-k}}{(j-k)!} \binom{i}{k} q^k p^{i-k}
\end{aligned}$$

(b) Using the preceding representation $R_i + N$ for the next state from state i, we see that

$$E[X_n | X_{n-1} = i] = E[R_i + N] = iq + \lambda$$

Consequently,

$$E[X_n | X_{n-1}] = X_{n-1} q + \lambda$$

Taking expectations of both sides yields

$$E[X_n] = \lambda + q E[X_{n-1}]$$

Iterating the preceding gives

$$\begin{aligned}
E[X_n] &= \lambda + q E[X_{n-1}] \\
&= \lambda + q(\lambda + q E[X_{n-2}]) \\
&= \lambda + q\lambda + q^2 E[X_{n-2}] \\
&= \lambda + q\lambda + q^2(\lambda + q E[X_{n-3}]) \\
&= \lambda + q\lambda + q^2\lambda + q^3 E[X_{n-3}]
\end{aligned}$$

showing that

$$E[X_n] = \lambda \left(1 + q + q^2 + \ldots + q^{n-1}\right) + q^n E[X_0]$$

and yielding the result

$$E[X_n | X_0 = i] = \frac{\lambda(1 - q^n)}{p} + q^n i$$

(c) To find the stationary probabilities we will not directly use the complicated transition probabilities derived in part (a). Rather we will make use of the fact that

the stationary probability distribution is the only distribution on the initial state that results in the next state having the same distribution. Now, suppose that the initial state X_0 has a Poisson distribution with mean α. That is, assume that the number of families initially in the hotel is Poisson with mean α. Let R denote the number of these families that remain in the hotel at the beginning of the next day. Then, using the result of Example 3.24 that if each of a Poisson distributed (with mean α) number of events occurs with probability q, then the total number of these events that occur is Poisson distributed with mean αq, it follows that R is a Poisson random variable with mean αq. In addition, the number of new families that check in during the day, call it N, is Poisson with mean λ, and is independent of R. Hence, since the sum of independent Poisson random variables is also Poisson distributed, it follows that $R + N$, the number of guests at the beginning of the next day, is Poisson with mean $\lambda + \alpha q$. Consequently, if we choose α so that

$$\alpha = \lambda + \alpha q$$

then the distribution of X_1 would be the same as that of X_0. But this means that when the initial distribution of X_0 is Poisson with mean $\alpha = \frac{\lambda}{p}$, then so is the distribution of X_1, implying that this is the stationary distribution. That is, the stationary probabilities are

$$\pi_i = e^{-\lambda/p}(\lambda/p)^i/i!, \quad i \geq 0$$

The preceding model has an important generalization. Namely, consider an organization whose workers are of r distinct types. For instance, the organization could be a law firm and its lawyers could either be juniors, associates, or partners. Suppose that a worker who is currently type i will in the next period become type j with probability $q_{i,j}$ for $j = 1, \ldots, r$ or will leave the organization with probability $1 - \sum_{j=1}^{r} q_{i,j}$. In addition, suppose that new workers are hired each period, and that the numbers of types $1, \ldots, r$ workers hired are independent Poisson random variables with means $\lambda_1, \ldots, \lambda_r$. If we let $\mathbf{X}_n = (X_n(1), \ldots, X_n(r))$, where $X_n(i)$ is the number of type i workers in the organization at the beginning of period n, then $\mathbf{X}_n, n \geq 0$ is a Markov chain. To compute its stationary probability distribution, suppose that the initial state is chosen so that the number of workers of different types are independent Poisson random variables, with α_i being the mean number of type i workers. That is, suppose that $X_0(1), \ldots, X_0(r)$ are independent Poisson random variables with respective means $\alpha_1, \ldots, \alpha_r$. Also, let $N_j, j = 1, \ldots, r$, be the number of new type j workers hired during the initial period. Now, fix i, and for $j = 1, \ldots, r$, let $M_i(j)$ be the number of the $X_0(i)$ type i workers who become type j in the next period. Then because each of the Poisson number $X_0(i)$ of type i workers will independently become type j with probability $q_{i,j}, j = 1, \ldots, r$, it follows from the remarks following Example 3.24 that $M_i(1), \ldots, M_i(r)$ are independent Poisson random variables with $M_i(j)$ having mean $a_i q_{i,j}$. Because $X_0(1), \ldots, X_0(r)$ are, by assumption, independent, we can also conclude that the random variables $M_i(j), i, j = 1, \ldots, r$ are all independent. Because the sum of independent Poisson random variables is also Poisson

distributed, the preceding yields that the random variables

$$X_1(j) = N_j + \sum_{i=1}^{r} M_i(j), \quad j = 1, \ldots, r$$

are independent Poisson random variables with means

$$E[X_1(j)] = \lambda_j + \sum_{i=1}^{r} \alpha_i q_{i,j}$$

Hence, if $\alpha_1, \ldots, \alpha_r$ satisfied

$$\alpha_j = \lambda_j + \sum_{i=1}^{r} \alpha_i q_{i,j}, \quad j = 1, \ldots, r$$

then $\mathbf{X}_1$ would have the same distribution as $\mathbf{X}_0$. Consequently, if we let $\alpha_1^o, \ldots, \alpha_r^o$ be such that

$$\alpha_j^o = \lambda_j + \sum_{i=1}^{r} \alpha_i^o q_{i,j}, \quad j = 1, \ldots, r$$

then the stationary distribution of the Markov chain is the distribution that takes the number of workers in each type to be independent Poisson random variables with means $\alpha_1^o, \ldots, \alpha_r^o$. That is, the long run proportions are

$$\pi_{k_1,\ldots,k_r} = \prod_{i=1}^{r} e^{-\alpha_i^o} (\alpha_i^o)^{k_i} / k_i!$$

It can be shown that there will be such values α_j^o, $j = 1, \ldots, r$, provided that, with probability 1, each worker eventually leaves the organization. Also, because there is a unique stationary distribution, there can only be one such set of values. ■

The following example exploits the relationship $m_i = 1/\pi_i$, which states that the mean time between visits to a state is the inverse of the long run proportion of time the chain is in that state, to obtain a method for computing the mean time until a specified pattern appears when the data constitutes the successive states of a Markov chain.

Example 4.28 (Mean Pattern Times in Markov Chain Generated Data)**.** Consider an irreducible Markov chain $\{X_n, n \geqslant 0\}$ with transition probabilities $P_{i,j}$ and stationary probabilities π_j, $j \geqslant 0$. Starting in state r, we are interested in determining the expected number of transitions until the pattern $i_1, i_2, \ldots, i_k$ appears. That is, with

$$N(i_1, i_2, \ldots, i_k) = \min\{n \geqslant k\colon X_{n-k+1} = i_1, \ldots, X_n = i_k\}$$

we are interested in

$$E[N(i_1, i_2, \ldots, i_k) | X_0 = r]$$

Note that even if $i_1 = r$, the initial state X_0 is not considered part of the pattern sequence.

Let $\mu(i, i_1)$ be the mean number of transitions for the chain to enter state i_1, given that the initial state is i, $i \geqslant 0$. The quantities $\mu(i, i_1)$ can be determined as the solution of the following set of equations, obtained by conditioning on the first transition out of state i:

$$\mu(i, i_1) = 1 + \sum_{j \neq i_1} P_{i,j}\mu(j, i_1), \quad i \geqslant 0$$

For the Markov chain $\{X_n, n \geqslant 0\}$ associate a corresponding Markov chain, which we will refer to as the k-chain, whose state at any time is the sequence of the most recent k states of the original chain. (For instance, if $k = 3$ and $X_2 = 4, X_3 = 1, X_4 = 1$, then the state of the k-chain at time 4 is (4, 1, 1).) Let $\pi(j_1, \ldots, j_k)$ be the stationary probabilities for the k-chain. Because $\pi(j_1, \ldots, j_k)$ is the proportion of time that the state of the original Markov chain k units ago was j_1 and the following $k - 1$ states, in sequence, were $j_2, \ldots, j_k$, we can conclude that

$$\pi(j_1, \ldots, j_k) = \pi_{j_1} P_{j_1,j_2} \cdots P_{j_{k-1},j_k}$$

Moreover, because the mean number of transitions between successive visits of the k-chain to the state $i_1, i_2, \ldots, i_k$ is equal to the inverse of the stationary probability of that state, we have that

$$E[\text{number of transitions between visits to } i_1, i_2, \ldots, i_k] = \frac{1}{\pi(i_1, \ldots, i_k)} \tag{4.13}$$

Let $A(i_1, \ldots, i_m)$ be the additional number of transitions needed until the pattern appears, given that the first m transitions have taken the chain into states $X_1 = i_1, \ldots, X_m = i_m$.

We will now consider whether the pattern has overlaps, where we say that the pattern $i_1, i_2, \ldots, i_k$ has an overlap of size $j, j < k$, if the sequence of its final j elements is the same as that of its first j elements. That is, it has an overlap of size j if

$$(i_{k-j+1}, \ldots, i_k) = (i_1, \ldots, i_j), \quad j < k$$

Case 1. The pattern $i_1, i_2, \ldots, i_k$ has no overlaps.
Because there is no overlap, Eq. (4.13) yields

$$E[N(i_1, i_2, \ldots, i_k) | X_0 = i_k] = \frac{1}{\pi(i_1, \ldots, i_k)}$$

Because the time until the pattern occurs is equal to the time until the chain enters state i_1 plus the additional time, we may write

$$E[N(i_1, i_2, \ldots, i_k) | X_0 = i_k] = \mu(i_k, i_1) + E[A(i_1)]$$

The preceding two equations imply

$$E[A(i_1)] = \frac{1}{\pi(i_1, \ldots, i_k)} - \mu(i_k, i_1)$$

Using that

$$E[N(i_1, i_2, \ldots, i_k)|X_0 = r] = \mu(r, i_1) + E[A(i_1)]$$

gives the result

$$E[N(i_1, i_2, \ldots, i_k)|X_0 = r] = \mu(r, i_1) + \frac{1}{\pi(i_1, \ldots, i_k)} - \mu(i_k, i_1)$$

where

$$\pi(i_1, \ldots, i_k) = \pi_{i_1} P_{i_1,i_2} \cdots P_{i_{k-1},i_k}$$

Case 2. Now suppose that the pattern has overlaps and let its largest overlap be of size s. In this case the number of transitions between successive visits of the k-chain to the state $i_1, i_2, \ldots, i_k$ is equal to the additional number of transitions of the original chain until the pattern appears given that it has already made s transitions with the results $X_1 = i_1, \ldots, X_s = i_s$. Therefore, from Eq. (4.13)

$$E[A(i_1, \ldots, i_s)] = \frac{1}{\pi(i_1, \ldots, i_k)}$$

But because

$$N(i_1, i_2, \ldots, i_k) = N(i_1, \ldots, i_s) + A(i_1, \ldots, i_s)$$

we have

$$E[N(i_1, i_2, \ldots, i_k)|X_0 = r] = E[N(i_1, i_2, \ldots, i_s)|X_0 = r] + \frac{1}{\pi(i_1, \ldots, i_k)}$$

We can now repeat the same procedure on the pattern $i_1, \ldots, i_s$, continuing to do so until we reach one that has no overlap, and then apply the result from Case 1.

For instance, suppose the desired pattern is 1, 2, 3, 1, 2, 3, 1, 2. Then

$$E[N(1, 2, 3, 1, 2, 3, 1, 2)|X_0 = r] = E[N(1, 2, 3, 1, 2)|X_0 = r] + \frac{1}{\pi(1, 2, 3, 1, 2, 3, 1, 2)}$$

Because the largest overlap of the pattern (1, 2, 3, 1, 2) is of size 2, the same argument as in the preceding gives

$$E[N(1, 2, 3, 1, 2)|X_0 = r] = E[N(1, 2)|X_0 = r] + \frac{1}{\pi(1, 2, 3, 1, 2)}$$

Because the pattern (1, 2) has no overlap, we obtain from Case 1 that

$$E[N(1,2)|X_0 = r] = \mu(r,1) + \frac{1}{\pi(1,2)} - \mu(2,1)$$

Putting it together yields

$$E[N(1,2,3,1,2,3,1,2)|X_0 = r] = \mu(r,1) + \frac{1}{\pi_1 P_{1,2}} - \mu(2,1) + \frac{1}{\pi_1 P_{1,2}^2 P_{2,3} P_{3,1}} + \frac{1}{\pi_1 P_{1,2}^3 P_{2,3}^2 P_{3,1}^2}$$

If the generated data is a sequence of independent and identically distributed random variables, with each value equal to j with probability P_j, then the Markov chain has $P_{i,j} = P_j$. In this case, $\pi_j = P_j$. Also, because the time to go from state i to state j is a geometric random variable with parameter P_j, we have $\mu(i, j) = 1/P_j$. Thus, the expected number of data values that need be generated before the pattern 1, 2, 3, 1, 2, 3, 1, 2 appears would be

$$\frac{1}{P_1} + \frac{1}{P_1 P_2} - \frac{1}{P_1} + \frac{1}{P_1^2 P_2^2 P_3} + \frac{1}{P_1^3 P_2^3 P_3^2} = \frac{1}{P_1 P_2} + \frac{1}{P_1^2 P_2^2 P_3} + \frac{1}{P_1^3 P_2^3 P_3^2}$$ ■

The following result is quite useful.

Proposition 4.6. *Let $\{X_n, n \geqslant 1\}$ be an irreducible Markov chain with stationary probabilities π_j, $j \geqslant 0$, and let r be a bounded function on the state space. Then, with probability 1,*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} r(X_n)}{N} = \sum_{j=0}^{\infty} r(j)\pi_j$$

Proof. If we let $a_j(N)$ be the amount of time the Markov chain spends in state j during time periods $1, \ldots, N$, then

$$\sum_{n=1}^{N} r(X_n) = \sum_{j=0}^{\infty} a_j(N) r(j)$$

Since $a_j(N)/N \to \pi_j$ the result follows from the preceding upon dividing by N and then letting $N \to \infty$. ■

If we suppose that we earn a reward $r(j)$ whenever the chain is in state j, then Proposition 4.6 states that our average reward per unit time is $\sum_j r(j)\pi_j$.

Example 4.29. For the four state Bonus Malus automobile insurance system specified in Example 4.7, find the average annual premium paid by a policyholder whose yearly number of claims is a Poisson random variable with mean 1/2.

Solution: With $a_k = e^{-1/2}\frac{(1/2)^k}{k!}$, we have

$$a_0 = 0.6065, \quad a_1 = 0.3033, \quad a_2 = 0.0758$$

Therefore, the Markov chain of successive states has the following transition probability matrix:

$$\begin{Vmatrix} 0.6065 & 0.3033 & 0.0758 & 0.0144 \\ 0.6065 & 0.0000 & 0.3033 & 0.0902 \\ 0.0000 & 0.6065 & 0.0000 & 0.3935 \\ 0.0000 & 0.0000 & 0.6065 & 0.3935 \end{Vmatrix}$$

The stationary probabilities are given as the solution of

$$\begin{aligned} \pi_1 &= 0.6065\pi_1 + 0.6065\pi_2, \\ \pi_2 &= 0.3033\pi_1 + 0.6065\pi_3, \\ \pi_3 &= 0.0758\pi_1 + 0.3033\pi_2 + 0.6065\pi_4, \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 &= 1 \end{aligned}$$

Rewriting the first three of these equations gives

$$\begin{aligned} \pi_2 &= \frac{1 - 0.6065}{0.6065}\pi_1, \\ \pi_3 &= \frac{\pi_2 - 0.3033\pi_1}{0.6065}, \\ \pi_4 &= \frac{\pi_3 - 0.0758\pi_1 - 0.3033\pi_2}{0.6065} \end{aligned}$$

or

$$\begin{aligned} \pi_2 &= 0.6488\pi_1, \\ \pi_3 &= 0.5697\pi_1, \\ \pi_4 &= 0.4900\pi_1 \end{aligned}$$

Using that $\sum_{i=1}^{4} \pi_i = 1$ gives the solution (rounded to four decimal places)

$$\pi_1 = 0.3692, \quad \pi_2 = 0.2395, \quad \pi_3 = 0.2103, \quad \pi_4 = 0.1809$$

Therefore, the average annual premium paid is

$$200\pi_1 + 250\pi_2 + 400\pi_3 + 600\pi_4 = 326.375$$ ■

4.4.1 Limiting Probabilities

In Example 4.8 we considered a two-state Markov chain with transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

and showed that

$$\mathbf{P}^{(4)} = \begin{pmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix}$$

From this it follows that $\mathbf{P}^{(8)} = \mathbf{P}^{(4)} \cdot \mathbf{P}^{(4)}$ is given (to three significant places) by

$$\mathbf{P}^{(8)} = \begin{pmatrix} 0.571 & 0.429 \\ 0.571 & 0.429 \end{pmatrix}$$

Note that the matrix $\mathbf{P}^{(8)}$is almost identical to the matrix $\mathbf{P}^{(4)}$, and that each of the rows of $\mathbf{P}^{(8)}$ has almost identical values. Indeed, it seems that P_{ij}^n is converging to some value as $n \to \infty$, with this value not depending on i. Moreover, in Example 4.22 we showed that the long-run proportions for this chain are $\pi_0 = 4/7 \approx .571, \pi_1 = 3/7 \approx .429$, thus making it appear that these long-run proportions may also be limiting probabilities. Although this is indeed the case for the preceding chain, it is not always true that the long-run proportions are also limiting probabilities. To see why not, consider a two-state Markov chain having

$$P_{0,1} = P_{1,0} = 1$$

Because this Markov chain continually alternates between states 0 and 1, the long-run proportions of time it spends in these states are

$$\pi_0 = \pi_1 = 1/2$$

However,

$$P_{0,0}^n = \begin{cases} 1, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

and so $P_{0,0}^n$ does not have a limiting value as n goes to infinity. In general, a chain that can only return to a state in a multiple of $d > 1$ steps (where $d = 2$ in the preceding example) is said to be *periodic* and does not have limiting probabilities. However, for an irreducible chain that is not periodic, and such chains are called *aperiodic*, the limiting probabilities will always exist and will not depend on the initial state. Moreover, the limiting probability that the chain will be in state j will equal π_j, the long-run proportion of time the chain is in state j. That the limiting probabilities, when they exist, will equal the long-run proportions can be seen by letting

$$\alpha_j = \lim_{n \to \infty} P(X_n = j)$$

and using that

$$P(X_{n+1} = j) = \sum_{i=0}^{\infty} P(X_{n+1} = j | X_n = i)\, P(X_n = i) = \sum_{i=0}^{\infty} P_{ij} P(X_n = i)$$

and

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

Letting $n \to \infty$ in the preceding two equations yields, upon assuming that we can bring the limit inside the summation, that

$$\alpha_j = \sum_{i=0}^{\infty} \alpha_i P_{ij}$$
$$1 = \sum_{i=0}^{\infty} \alpha_i$$

Hence, $\{\alpha_j, j \geq 0\}$ satisfies the equations for which $\{\pi_j, j \geq 0\}$ is the unique solution, showing that $\alpha_j = \pi_j, j \geq 0$.

An irreducible, positive recurrent, aperiodic Markov chain is said to be *ergodic*.

4.5 Some Applications

4.5.1 The Gambler's Ruin Problem

Consider a gambler who at each play of the game has probability p of winning one unit and probability $q = 1 - p$ of losing one unit. Assuming that successive plays of the game are independent, what is the probability that, starting with i units, the gambler's fortune will reach N before reaching 0?

If we let X_n denote the player's fortune at time n, then the process $\{X_n, n = 0, 1, 2, \ldots\}$ is a Markov chain with transition probabilities

$$P_{00} = P_{NN} = 1,$$
$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i = 1,\ 2, \ldots, N-1$$

This Markov chain has three classes, namely, $\{0\}$, $\{1, 2, \ldots, N-1\}$, and $\{N\}$; the first and third class being recurrent and the second transient. Since each transient state is visited only finitely often, it follows that, after some finite amount of time, the gambler will either attain his goal of N or go broke.

Let $P_i, i = 0, 1, \ldots, N$, denote the probability that, starting with i, the gambler's fortune will eventually reach N. By conditioning on the outcome of the initial play of

the game we obtain

$$P_i = pP_{i+1} + qP_{i-1}, \quad i = 1, 2, \ldots, N-1$$

or equivalently, since $p + q = 1$,

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

or

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}), \quad i = 1, 2, \ldots, N-1$$

Hence, since $P_0 = 0$, we obtain from the preceding line that

$$\begin{aligned}
P_2 - P_1 &= \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1, \\
P_3 - P_2 &= \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1, \\
&\vdots \\
P_i - P_{i-1} &= \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1, \\
&\vdots \\
P_N - P_{N-1} &= \left(\frac{q}{p}\right)(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1
\end{aligned}$$

Adding the first $i - 1$ of these equations yields

$$P_i - P_1 = P_1\left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

or

$$P_i = \begin{cases} \dfrac{1 - (q/p)^i}{1 - (q/p)} P_1, & \text{if } \dfrac{q}{p} \neq 1 \\ iP_1, & \text{if } \dfrac{q}{p} = 1 \end{cases}$$

Now, using the fact that $P_N = 1$, we obtain

$$P_1 = \begin{cases} \dfrac{1 - (q/p)}{1 - (q/p)^N}, & \text{if } p \neq \dfrac{1}{2} \\ \dfrac{1}{N}, & \text{if } p = \dfrac{1}{2} \end{cases}$$

and hence

$$P_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^N}, & \text{if } p \neq \dfrac{1}{2} \\ \dfrac{i}{N}, & \text{if } p = \dfrac{1}{2} \end{cases} \tag{4.14}$$

Note that, as $N \to \infty$,

$$P_i \to \begin{cases} 1-\left(\dfrac{q}{p}\right)^i, & \text{if } p > \dfrac{1}{2} \\ 0, & \text{if } p \leqslant \dfrac{1}{2} \end{cases}$$

Thus, if $p > \frac{1}{2}$, there is a positive probability that the gambler's fortune will increase indefinitely; while if $p \leqslant \frac{1}{2}$, the gambler will, with probability 1, go broke against an infinitely rich adversary.

Example 4.30. Suppose Max and Patty decide to flip pennies; the one coming closest to the wall wins. Patty, being the better player, has a probability 0.6 of winning on each flip. (a) If Patty starts with five pennies and Max with ten, what is the probability that Patty will wipe Max out? (b) What if Patty starts with 10 and Max with 20?

Solution: (a) The desired probability is obtained from Eq. (4.14) by letting $i = 5$, $N = 15$, and $p = 0.6$. Hence, the desired probability is

$$\frac{1-\left(\frac{2}{3}\right)^5}{1-\left(\frac{2}{3}\right)^{15}} \approx 0.87$$

(b) The desired probability is

$$\frac{1-\left(\frac{2}{3}\right)^{10}}{1-\left(\frac{2}{3}\right)^{30}} \approx 0.98$$ ■

For an application of the gambler's ruin problem to drug testing, suppose that two new drugs have been developed for treating a certain disease. Drug i has a cure rate $P_i, i = 1, 2$, in the sense that each patient treated with drug i will be cured with probability P_i. These cure rates, however, are not known, and suppose we are interested in a method for deciding whether $P_1 > P_2$ or $P_2 > P_1$. To decide upon one of these alternatives, consider the following test: Pairs of patients are treated sequentially with one member of the pair receiving drug 1 and the other drug 2. The results for each pair are determined, and the testing stops when the cumulative number of cures using one of the drugs exceeds the cumulative number of cures when using the other by some

fixed predetermined number. More formally, let

$$X_j = \begin{cases} 1, & \text{if the patient in the } j\text{th pair to receive drug number 1 is cured} \\ 0, & \text{otherwise} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{if the patient in the } j\text{th pair to receive drug number 2 is cured} \\ 0, & \text{otherwise} \end{cases}$$

For a predetermined positive integer M the test stops after pair N where N is the first value of n such that either

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = M$$

or

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = -M$$

In the former case we then assert that $P_1 > P_2$, and in the latter that $P_2 > P_1$.

In order to help ascertain whether the preceding is a good test, one thing we would like to know is the probability of it leading to an incorrect decision. That is, for given P_1 and P_2 where $P_1 > P_2$, what is the probability that the test will incorrectly assert that $P_2 > P_1$? To determine this probability, note that after each pair is checked the cumulative difference of cures using drug 1 versus drug 2 will either go up by 1 with probability $P_1(1 - P_2)$—since this is the probability that drug 1 leads to a cure and drug 2 does not—or go down by 1 with probability $(1 - P_1)P_2$, or remain the same with probability $P_1P_2 + (1 - P_1)(1 - P_2)$. Hence, if we only consider those pairs in which the cumulative difference changes, then the difference will go up 1 with probability

$$\begin{aligned} p &= P\{\text{up 1} | \text{up 1 or down 1}\} \\ &= \frac{P_1(1 - P_2)}{P_1(1 - P_2) + (1 - P_1)P_2} \end{aligned}$$

and down 1 with probability

$$q = 1 - p = \frac{P_2(1 - P_1)}{P_1(1 - P_2) + (1 - P_1)P_2}$$

Hence, the probability that the test will assert that $P_2 > P_1$ is equal to the probability that a gambler who wins each (one unit) bet with probability p will go down M before going up M. But Eq. (4.14) with $i = M$, $N = 2M$, shows that this probability is given by

$$\begin{aligned} P\{\text{test asserts that } P_2 > P_1\} &= 1 - \frac{1 - (q/p)^M}{1 - (q/p)^{2M}} \\ &= \frac{1}{1 + (p/q)^M} \end{aligned}$$

Thus, for instance, if $P_1 = 0.6$ and $P_2 = 0.4$ then the probability of an incorrect decision is 0.017 when $M = 5$ and reduces to 0.0003 when $M = 10$.

4.5.2 A Model for Algorithmic Efficiency

The following optimization problem is called a linear program:

$$\begin{aligned} &\text{minimize } \mathbf{cx}, \\ &\text{subject to } \mathbf{Ax} = \mathbf{b}, \\ &\qquad \mathbf{x} \geqslant \mathbf{0} \end{aligned}$$

where $\mathbf{A}$ is an $m \times n$ matrix of fixed constants; $\mathbf{c} = (c_1, \ldots, c_n)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ are vectors of fixed constants; and $\mathbf{x} = (x_1, \ldots, x_n)$ is the n-vector of nonnegative values that is to be chosen to minimize $\mathbf{cx} \equiv \sum_{i=1}^{n} c_i x_i$. Supposing that $n > m$, it can be shown that the optimal $\mathbf{x}$ can always be chosen to have at least $n - m$ components equal to 0—that is, it can always be taken to be one of the so-called extreme points of the feasibility region.

The simplex algorithm solves this linear program by moving from an extreme point of the feasibility region to a better (in terms of the objective function $\mathbf{cx}$) extreme point (via the pivot operation) until the optimal is reached. Because there can be as many as $N \equiv \binom{n}{m}$ such extreme points, it would seem that this method might take many iterations, but, surprisingly to some, this does not appear to be the case in practice.

To obtain a feel for whether or not the preceding statement is surprising, let us consider a simple probabilistic (Markov chain) model as to how the algorithm moves along the extreme points. Specifically, we will suppose that if at any time the algorithm is at the jth best extreme point then after the next pivot the resulting extreme point is equally likely to be any of the $j - 1$ best. Under this assumption, we show that the time to get from the Nth best to the best extreme point has approximately, for large N, a normal distribution with mean and variance equal to the logarithm (base e) of N.

Consider a Markov chain for which $P_{11} = 1$ and

$$P_{ij} = \frac{1}{i-1}, \quad j = 1, \ldots, i-1, \; i > 1$$

and let T_i denote the number of transitions needed to go from state i to state 1. A recursive formula for $E[T_i]$ can be obtained by conditioning on the initial transition:

$$E[T_i] = 1 + \frac{1}{i-1} \sum_{j=1}^{i-1} E[T_j]$$

Starting with $E[T_1] = 0$, we successively see that

$$\begin{aligned} E[T_2] &= 1, \\ E[T_3] &= 1 + \tfrac{1}{2}, \end{aligned}$$

$$E[T_4] = 1 + \tfrac{1}{3}(1 + 1 + \tfrac{1}{2}) = 1 + \tfrac{1}{2} + \tfrac{1}{3}$$

and it is not difficult to guess and then prove inductively that

$$E[T_i] = \sum_{j=1}^{i-1} 1/j$$

However, to obtain a more complete description of T_N, we will use the representation

$$T_N = \sum_{j=1}^{N-1} I_j$$

where

$$I_j = \begin{cases} 1, & \text{if the process ever enters } j \\ 0, & \text{otherwise} \end{cases}$$

The importance of the preceding representation stems from the following:

Proposition 4.7. *$I_1, \ldots, I_{N-1}$ are independent and*

$$P\{I_j = 1\} = 1/j, \quad 1 \leqslant j \leqslant N - 1$$

Proof. Given $I_{j+1}, \ldots, I_N$, let $n = \min\{i: i > j, I_i = 1\}$ denote the lowest numbered state, greater than j, that is entered. Thus we know that the process enters state n and the next state entered is one of the states $1, 2, \ldots, j$. Hence, as the next state from state n is equally likely to be any of the lower number states $1, 2, \ldots, n - 1$ we see that

$$P\{I_j = 1 | I_{j+1}, \ldots, I_N\} = \frac{1/(n-1)}{j/(n-1)} = 1/j$$

Hence, $P\{I_j = 1\} = 1/j$, and independence follows since the preceding conditional probability does not depend on $I_{j+1}, \ldots, I_N$. ■

Corollary 4.8. **(i)** $E[T_N] = \sum_{j=1}^{N-1} 1/j$.

(ii) $\mathrm{Var}(T_N) = \sum_{j=1}^{N-1} (1/j)(1 - 1/j)$.

(iii) *For N large, T_N has approximately a normal distribution with mean* $\log N$ *and variance* $\log N$.

Proof. Parts (i) and (ii) follow from Proposition 4.7 and the representation $T_N = \sum_{j=1}^{N-1} I_j$. Part (iii) follows from the central limit theorem since

$$\int_1^N \frac{dx}{x} < \sum_1^{N-1} 1/j < 1 + \int_1^{N-1} \frac{dx}{x}$$

or

$$\log N < \sum_{1}^{N-1} 1/j < 1 + \log(N-1)$$

and so

$$\log N \approx \sum_{j=1}^{N-1} 1/j$$ ■

Returning to the simplex algorithm, if we assume that n, m, and $n-m$ are all large, we have by Stirling's approximation that

$$N = \binom{n}{m} \sim \frac{n^{n+1/2}}{(n-m)^{n-m+1/2} m^{m+1/2} \sqrt{2\pi}}$$

and so, letting $c = n/m$,

$$\log N \sim \left(mc + \tfrac{1}{2}\right)\log(mc) - \left(m(c-1) + \tfrac{1}{2}\right)\log(m(c-1)) - \left(m + \tfrac{1}{2}\right)\log m - \tfrac{1}{2}\log(2\pi)$$

or

$$\log N \sim m\left[c \log \frac{c}{c-1} + \log(c-1)\right]$$

Now, as $\lim_{x\to\infty} x \log[x/(x-1)] = 1$, it follows that, when c is large,

$$\log N \sim m[1 + \log(c-1)]$$

Thus, for instance, if $n = 8000, m = 1000$, then the number of necessary transitions is approximately normally distributed with mean and variance equal to $1000(1 + \log 7) \approx 3000$. Hence, the number of necessary transitions would be roughly between

$$3000 \pm 2\sqrt{3000} \quad \text{or roughly} \quad 3000 \pm 110$$

95 percent of the time.

4.5.3 Using a Random Walk to Analyze a Probabilistic Algorithm for the Satisfiability Problem

Consider a Markov chain with states $0, 1, \ldots, n$ having

$$P_{0,1} = 1, \quad P_{i,i+1} = p, \quad P_{i,i-1} = q = 1 - p, \quad 1 \leqslant i < n$$

and suppose that we are interested in studying the time that it takes for the chain to go from state 0 to state n. One approach to obtaining the mean time to reach state n

would be to let m_i denote the mean time to go from state i to state n, $i = 0, \ldots, n-1$. If we then condition on the initial transition, we obtain the following set of equations:

$$\begin{aligned} m_0 &= 1 + m_1, \\ m_i &= E[\text{time to reach } n | \text{next state is } i+1]p \\ &\quad + E[\text{time to reach } n | \text{next state is } i-1]q \\ &= (1 + m_{i+1})p + (1 + m_{i-1})q \\ &= 1 + pm_{i+1} + qm_{i-1}, \quad i = 1, \ldots, n-1 \end{aligned}$$

Whereas the preceding equations can be solved for m_i, $i = 0, \ldots, n-1$, we do not pursue their solution; we instead make use of the special structure of the Markov chain to obtain a simpler set of equations. To start, let N_i denote the number of additional transitions that it takes the chain when it first enters state i until it enters state $i+1$. By the Markovian property, it follows that these random variables N_i, $i = 0, \ldots, n-1$ are independent. Also, we can express $N_{0,n}$, the number of transitions that it takes the chain to go from state 0 to state n, as

$$N_{0,n} = \sum_{i=0}^{n-1} N_i \tag{4.15}$$

Letting $\mu_i = E[N_i]$ we obtain, upon conditioning on the next transition after the chain enters state i, that for $i = 1, \ldots, n-1$

$$\mu_i = 1 + E[\text{number of additional transitions to reach } i+1 | \text{chain to } i-1]q$$

Now, if the chain next enters state $i-1$, then in order for it to reach $i+1$ it must first return to state i and must then go from state i to state $i+1$. Hence, we have from the preceding that

$$\mu_i = 1 + E[N_{i-1}^* + N_i^*]q$$

where N_{i-1}^* and N_i^* are, respectively, the additional number of transitions to return to state i from $i-1$ and the number to then go from i to $i+1$. Now, it follows from the Markovian property that these random variables have, respectively, the same distributions as N_{i-1} and N_i. In addition, they are independent (although we will only use this when we compute the variance of $N_{0,n}$). Hence, we see that

$$\mu_i = 1 + q(\mu_{i-1} + \mu_i)$$

or

$$\mu_i = \frac{1}{p} + \frac{q}{p}\mu_{i-1}, \quad i = 1, \ldots, n-1$$

Starting with $\mu_0 = 1$, and letting $\alpha = q/p$, we obtain from the preceding recursion that

$$\mu_1 = 1/p + \alpha,$$

$$\mu_2 = 1/p + \alpha(1/p + \alpha) = 1/p + \alpha/p + \alpha^2,$$
$$\mu_3 = 1/p + \alpha(1/p + \alpha/p + \alpha^2)$$
$$= 1/p + \alpha/p + \alpha^2/p + \alpha^3$$

In general, we see that

$$\mu_i = \frac{1}{p}\sum_{j=0}^{i-1}\alpha^j + \alpha^i, \quad i = 1, \ldots, n-1 \tag{4.16}$$

Using Eq. (4.15), we now get

$$E[N_{0,n}] = 1 + \frac{1}{p}\sum_{i=1}^{n-1}\sum_{j=0}^{i-1}\alpha^j + \sum_{i=1}^{n-1}\alpha^i$$

When $p = \frac{1}{2}$, and so $\alpha = 1$, we see from the preceding that

$$E[N_{0,n}] = 1 + (n-1)n + n - 1 = n^2$$

When $p \neq \frac{1}{2}$, we obtain

$$\begin{aligned} E[N_{0,n}] &= 1 + \frac{1}{p(1-\alpha)}\sum_{i=1}^{n-1}(1-\alpha^i) + \frac{\alpha - \alpha^n}{1-\alpha} \\ &= 1 + \frac{1+\alpha}{1-\alpha}\left[n - 1 - \frac{(\alpha - \alpha^n)}{1-\alpha}\right] + \frac{\alpha - \alpha^n}{1-\alpha} \\ &= 1 + \frac{2\alpha^{n+1} - (n+1)\alpha^2 + n - 1}{(1-\alpha)^2} \end{aligned}$$

where the second equality used the fact that $p = 1/(1+\alpha)$. Therefore, we see that when $\alpha > 1$, or equivalently when $p < \frac{1}{2}$, the expected number of transitions to reach n is an exponentially increasing function of n. On the other hand, when $p = \frac{1}{2}$, $E[N_{0,n}] = n^2$, and when $p > \frac{1}{2}$, $E[N_{0,n}]$ is, for large n, essentially linear in n.

Let us now compute $\mathrm{Var}(N_{0,n})$. To do so, we will again make use of the representation given by Eq. (4.15). Letting $v_i = \mathrm{Var}(N_i)$, we start by determining the v_i recursively by using the conditional variance formula. Let $S_i = 1$ if the first transition out of state i is into state $i+1$, and let $S_i = -1$ if the transition is into state $i-1$, $i = 1, \ldots, n-1$. Then,

given that $S_i = 1$: $\quad N_i = 1$
given that $S_i = -1$: $\quad N_i = 1 + N_{i-1}^* + N_i^*$

Hence,

$$E[N_i | S_i = 1] = 1,$$

$$E[N_i|S_i = -1] = 1 + \mu_{i-1} + \mu_i$$

implying that

$$\begin{aligned}\text{Var}(E[N_i|S_i]) &= \text{Var}(E[N_i|S_i] - 1)\\ &= (\mu_{i-1} + \mu_i)^2 q - (\mu_{i-1} + \mu_i)^2 q^2\\ &= qp(\mu_{i-1} + \mu_i)^2\end{aligned}$$

Also, since N^*_{i-1} and N^*_i, the numbers of transitions to return from state $i-1$ to i and to then go from state i to state $i+1$ are, by the Markovian property, independent random variables having the same distributions as N_{i-1} and N_i, respectively, we see that

$$\begin{aligned}\text{Var}(N_i|S_i = 1) &= 0,\\ \text{Var}(N_i|S_i = -1) &= v_{i-1} + v_i\end{aligned}$$

Hence,

$$E[\text{Var}(N_i|S_i)] = q(v_{i-1} + v_i)$$

From the conditional variance formula, we thus obtain

$$v_i = pq(\mu_{i-1} + \mu_i)^2 + q(v_{i-1} + v_i)$$

or, equivalently,

$$v_i = q(\mu_{i-1} + \mu_i)^2 + \alpha v_{i-1}, \quad i = 1, \ldots, n-1$$

Starting with $v_0 = 0$, we obtain from the preceding recursion that

$$\begin{aligned}v_1 &= q(\mu_0 + \mu_1)^2,\\ v_2 &= q(\mu_1 + \mu_2)^2 + \alpha q(\mu_0 + \mu_1)^2,\\ v_3 &= q(\mu_2 + \mu_3)^2 + \alpha q(\mu_1 + \mu_2)^2 + \alpha^2 q(\mu_0 + \mu_1)^2\end{aligned}$$

In general, we have for $i > 0$,

$$v_i = q\sum_{j=1}^{i} \alpha^{i-j}(\mu_{j-1} + \mu_j)^2 \tag{4.17}$$

Therefore, we see that

$$\text{Var}(N_{0,n}) = \sum_{i=0}^{n-1} v_i = q\sum_{i=1}^{n-1}\sum_{j=1}^{i} \alpha^{i-j}(\mu_{j-1} + \mu_j)^2$$

where μ_j is given by Eq. (4.16).

We see from Eqs. (4.16) and (4.17) that when $p \geqslant \frac{1}{2}$, and so $\alpha \leqslant 1$, that μ_i and v_i, the mean and variance of the number of transitions to go from state i to $i+1$, do not increase too rapidly in i. For instance, when $p = \frac{1}{2}$ it follows from Eqs. (4.16) and (4.17) that

$$\mu_i = 2i + 1$$

and

$$v_i = \frac{1}{2}\sum_{j=1}^{i}(4j)^2 = 8\sum_{j=1}^{i} j^2$$

Hence, since $N_{0,n}$ is the sum of independent random variables, which are of roughly similar magnitudes when $p \geqslant \frac{1}{2}$, it follows in this case from the central limit theorem that $N_{0,n}$ is, for large n, approximately normally distributed. In particular, when $p = \frac{1}{2}$, $N_{0,n}$ is approximately normal with mean n^2 and variance

$$\begin{aligned}
\mathrm{Var}(N_{0,n}) &= 8\sum_{i=1}^{n-1}\sum_{j=1}^{i} j^2 \\
&= 8\sum_{j=1}^{n-1}\sum_{i=j}^{n-1} j^2 \\
&= 8\sum_{j=1}^{n-1}(n-j)j^2 \\
&\approx 8\int_{1}^{n-1}(n-x)x^2\,dx \\
&\approx \tfrac{2}{3}n^4
\end{aligned}$$

Example 4.31 (The Satisfiability Problem)**.** A Boolean variable x is one that takes on either of two values: TRUE or FALSE. If $x_i, i \geqslant 1$ are Boolean variables, then a Boolean clause of the form

$$x_1 + \bar{x}_2 + x_3$$

is TRUE if x_1 is TRUE, or if x_2 is FALSE, or if x_3 is TRUE. That is, the symbol "+" means "or" and $\bar{x}$ is TRUE if x is FALSE and vice versa. A Boolean formula is a combination of clauses such as

$$(x_1 + \bar{x}_2) * (x_1 + x_3) * (x_2 + \bar{x}_3) * (\bar{x}_1 + \bar{x}_2) * (x_1 + x_2)$$

In the preceding, the terms between the parentheses represent clauses, and the formula is TRUE if all the clauses are TRUE, and is FALSE otherwise. For a given Boolean formula, the *satisfiability problem* is either to determine values for the variables that

result in the formula being TRUE, or to determine that the formula is never true. For instance, one set of values that makes the preceding formula TRUE is to set $x_1 =$ TRUE, $x_2 =$ FALSE, and $x_3 =$ FALSE.

Consider a formula of the n Boolean variables $x_1, \ldots, x_n$ and suppose that each clause in this formula refers to exactly two variables. We will now present a *probabilistic algorithm* that will either find values that satisfy the formula or determine *to a high probability* that it is not possible to satisfy it. To begin, start with an arbitrary setting of values. Then, at each stage choose a clause whose value is FALSE, and randomly choose one of the Boolean variables in that clause and change its value. That is, if the variable has value TRUE then change its value to FALSE, and vice versa. If this new setting makes the formula TRUE then stop, otherwise continue in the same fashion. If you have not stopped after $n^2(1 + 4\sqrt{\frac{2}{3}})$ repetitions, then declare that the formula cannot be satisfied. We will now argue that if there is a satisfiable assignment then this algorithm will find such an assignment with a probability very close to 1.

Let us start by assuming that there is a satisfiable assignment of truth values and let $\mathscr{A}$ be such an assignment. At each stage of the algorithm there is a certain assignment of values. Let Y_j denote the number of the n variables whose values at the jth stage of the algorithm agree with their values in $\mathscr{A}$. For instance, suppose that $n = 3$ and $\mathscr{A}$ consists of the settings $x_1 = x_2 = x_3 =$ TRUE. If the assignment of values at the jth step of the algorithm is $x_1 =$ TRUE, $x_2 = x_3 =$ FALSE, then $Y_j = 1$. Now, at each stage, the algorithm considers a clause that is not satisfied, thus implying that at least one of the values of the two variables in this clause does not agree with its value in $\mathscr{A}$. As a result, when we randomly choose one of the variables in this clause then there is a probability of at least $\frac{1}{2}$ that $Y_{j+1} = Y_j + 1$ and at most $\frac{1}{2}$ that $Y_{j+1} = Y_j - 1$. That is, independent of what has previously transpired in the algorithm, at each stage the number of settings in agreement with those in $\mathscr{A}$ will either increase or decrease by 1 and the probability of an increase is at least $\frac{1}{2}$ (it is 1 if both variables have values different from their values in $\mathscr{A}$). Thus, even though the process Y_j, $j \geqslant 0$ is not itself a Markov chain (why not?) it is intuitively clear that both the expectation and the variance of the number of stages of the algorithm needed to obtain the values of $\mathscr{A}$ will be less than or equal to the expectation and variance of the number of transitions to go from state 0 to state n in the Markov chain of Section 4.5.2. Hence, if the algorithm has not yet terminated because it found a set of satisfiable values different from that of $\mathscr{A}$, it will do so within an expected time of at most n^2 and with a standard deviation of at most $n^2\sqrt{\frac{2}{3}}$. In addition, since the time for the Markov chain to go from 0 to n is approximately normal when n is large we can be quite certain that a satisfiable assignment will be reached by $n^2 + 4(n^2\sqrt{\frac{2}{3}})$ stages, and thus if one has not been found by this number of stages of the algorithm we can be quite certain that there is no satisfiable assignment.

Our analysis also makes it clear why we assumed that there are only two variables in each clause. For if there were k, $k > 2$, variables in a clause then as any clause that is not presently satisfied may have only one incorrect setting, a randomly chosen variable whose value is changed might only increase the number of values in agreement with $\mathscr{A}$ with probability $1/k$ and so we could only conclude from our prior Markov chain

results that the mean time to obtain the values in $\mathscr{A}$ is an exponential function of n, which is not an efficient algorithm when n is large. ■

4.6 Mean Time Spent in Transient States

Consider now a finite state Markov chain and suppose that the states are numbered so that $T = \{1, 2, \ldots, t\}$ denotes the set of transient states. Let

$$\mathbf{P}_T = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1t} \\ \vdots & \vdots & \vdots & \vdots \\ P_{t1} & P_{t2} & \cdots & P_{tt} \end{bmatrix}$$

and note that since $\mathbf{P}_T$ specifies only the transition probabilities from transient states into transient states, some of its row sums are less than 1 (otherwise, T would be a closed class of states).

For transient states i and j, let s_{ij} denote the expected number of time periods that the Markov chain is in state j, given that it starts in state i. Let $\delta_{i,j} = 1$ when $i = j$ and let it be 0 otherwise. Condition on the initial transition to obtain

$$\begin{aligned} s_{ij} &= \delta_{i,j} + \sum_k P_{ik} s_{kj} \\ &= \delta_{i,j} + \sum_{k=1}^{t} P_{ik} s_{kj} \end{aligned} \tag{4.18}$$

where the final equality follows since it is impossible to go from a recurrent to a transient state, implying that $s_{kj} = 0$ when k is a recurrent state.

Let $\mathbf{S}$ denote the matrix of values $s_{ij}, i, j = 1, \ldots, t$. That is,

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1t} \\ \vdots & \vdots & \vdots & \vdots \\ s_{t1} & s_{t2} & \cdots & s_{tt} \end{bmatrix}$$

In matrix notation, Eq. (4.18) can be written as

$$\mathbf{S} = \mathbf{I} + \mathbf{P}_T \mathbf{S}$$

where $\mathbf{I}$ is the identity matrix of size t. Because the preceding equation is equivalent to

$$(\mathbf{I} - \mathbf{P}_T)\mathbf{S} = \mathbf{I}$$

we obtain, upon multiplying both sides by $(\mathbf{I} - \mathbf{P}_T)^{-1}$,

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1}$$

That is, the quantities $s_{ij}, i \in T, j \in T$, can be obtained by inverting the matrix $\mathbf{I} - \mathbf{P}_T$. (The existence of the inverse is easily established.)

Example 4.32. Consider the gambler's ruin problem with $p = 0.4$ and $N = 7$. Starting with 3 units, determine

(a) the expected amount of time the gambler has 5 units,
(b) the expected amount of time the gambler has 2 units.

Solution: The matrix $\mathbf{P}_T$, which specifies $P_{ij}, i, j \in \{1, 2, 3, 4, 5, 6\}$, is as follows:

$$\mathbf{P}_T = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 2 & 0.6 & 0 & 0.4 & 0 & 0 & 0 \\ 3 & 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 4 & 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 5 & 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 6 & 0 & 0 & 0 & 0 & 0.6 & 0 \end{array}$$

Inverting $\mathbf{I} - \mathbf{P}_T$ gives

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1} = \begin{bmatrix} 1.6149 & 1.0248 & 0.6314 & 0.3691 & 0.1943 & 0.0777 \\ 1.5372 & 2.5619 & 1.5784 & 0.9228 & 0.4857 & 0.1943 \\ 1.4206 & 2.3677 & 2.9990 & 1.7533 & 0.9228 & 0.3691 \\ 1.2458 & 2.0763 & 2.6299 & 2.9990 & 1.5784 & 0.6314 \\ 0.9835 & 1.6391 & 2.0763 & 2.3677 & 2.5619 & 1.0248 \\ 0.5901 & 0.9835 & 1.2458 & 1.4206 & 1.5372 & 1.6149 \end{bmatrix}$$

Hence,

$$s_{3,5} = 0.9228, \quad s_{3,2} = 2.3677$$ ■

For $i \in T, j \in T$, the quantity f_{ij}, equal to the probability that the Markov chain ever makes a transition into state j given that it starts in state i, is easily determined from $\mathbf{P}_T$. To determine the relationship, let us start by deriving an expression for s_{ij} by conditioning on whether state j is ever entered. This yields

$$\begin{aligned} s_{ij} &= E[\text{time in } j | \text{start in } i, \text{ ever transit to } j] f_{ij} \\ &\quad + E[\text{time in } j | \text{start in } i, \text{ never transit to } j](1 - f_{ij}) \\ &= (\delta_{i,j} + s_{jj}) f_{ij} + \delta_{i,j}(1 - f_{ij}) \\ &= \delta_{i,j} + f_{ij} s_{jj} \end{aligned}$$

since s_{jj} is the expected number of additional time periods spent in state j given that it is eventually entered from state i. Solving the preceding equation yields

$$f_{ij} = \frac{s_{ij} - \delta_{i,j}}{s_{jj}}$$

Example 4.33. In Example 4.32, what is the probability that the gambler ever has a fortune of 1?

Solution: Since $s_{3,1} = 1.4206$ and $s_{1,1} = 1.6149$, then

$$f_{3,1} = \frac{s_{3,1}}{s_{1,1}} = 0.8797$$

As a check, note that $f_{3,1}$ is just the probability that a gambler starting with 3 reaches 1 before 7. That is, it is the probability that the gambler's fortune will go down 2 before going up 4; which is the probability that a gambler starting with 2 will go broke before reaching 6. Therefore,

$$f_{3,1} = 1 - \frac{1 - (0.6/0.4)^2}{1 - (0.6/0.4)^6} = 0.8797$$

which checks with our earlier answer. ■

Suppose we are interested in the expected time until the Markov chain enters some sets of states A, which need not be the set of recurrent states. We can reduce this back to the previous situation by making all states in A absorbing states. That is, reset the transition probabilities of states in A to satisfy

$$P_{i,i} = 1, \quad i \in A$$

This transforms the states of A into recurrent states, and transforms any state outside of A from which an eventual transition into A is possible into a transient state. Thus, our previous approach can be used.

4.7 Branching Processes

In this section we consider a class of Markov chains, known as *branching processes*, which have a wide variety of applications in the biological, sociological, and engineering sciences.

Consider a population consisting of individuals able to produce offspring of the same kind. Suppose that each individual will, by the end of its lifetime, have produced j new offspring with probability P_j, $j \geqslant 0$, independently of the numbers produced by other individuals. We suppose that $P_j < 1$ for all $j \geqslant 0$. The number of individuals initially present, denoted by X_0, is called the size of the zeroth generation. All offspring of the zeroth generation constitute the first generation and their number is denoted by X_1. In general, let X_n denote the size of the nth generation. It follows that $\{X_n, n = 0, 1, \ldots\}$ is a Markov chain having as its state space the set of nonnegative integers.

Note that state 0 is a recurrent state, since clearly $P_{00} = 1$. Also, if $P_0 > 0$, all other states are transient. This follows since $P_{i0} = P_0^i$, which implies that starting with i

individuals there is a positive probability of at least P_0^i that no later generation will ever consist of i individuals. Moreover, since any finite set of transient states $\{1, 2, \ldots, n\}$ will be visited only finitely often, this leads to the important conclusion that, if $P_0 > 0$, *then the population will either die out or its size will converge to infinity*.

Let

$$\mu = \sum_{j=0}^{\infty} j P_j$$

denote the mean number of offspring of a single individual, and let

$$\sigma^2 = \sum_{j=0}^{\infty} (j - \mu)^2 P_j$$

be the variance of the number of offspring produced by a single individual.

Let us suppose that $X_0 = 1$, that is, initially there is a single individual present. We calculate $E[X_n]$ and $\text{Var}(X_n)$ by first noting that we may write

$$X_n = \sum_{i=1}^{X_{n-1}} Z_i$$

where Z_i represents the number of offspring of the ith individual of the $(n-1)$st generation. By conditioning on X_{n-1}, we obtain

$$\begin{aligned} E[X_n] &= E[E[X_n|X_{n-1}]] \\ &= E\left[E\left[\sum_{i=1}^{X_{n-1}} Z_i | X_{n-1}\right]\right] \\ &= E[X_{n-1}\mu] \\ &= \mu E[X_{n-1}] \end{aligned}$$

where we have used the fact that $E[Z_i] = \mu$. Since $E[X_0] = 1$, the preceding yields

$$\begin{aligned} E[X_1] &= \mu, \\ E[X_2] &= \mu E[X_1] = \mu^2, \\ &\vdots \\ E[X_n] &= \mu E[X_{n-1}] = \mu^n \end{aligned}$$

Similarly, $\text{Var}(X_n)$ may be obtained by using the conditional variance formula

$$\text{Var}(X_n) = E[\text{Var}(X_n|X_{n-1})] + \text{Var}(E[X_n|X_{n-1}])$$

Now, given X_{n-1}, X_n is just the sum of X_{n-1} independent random variables each having the distribution $\{P_j, j \geqslant 0\}$. Hence,

$$E[X_n|X_{n-1}] = X_{n-1}\mu, \quad \text{Var}(X_n|X_{n-1}) = X_{n-1}\sigma^2$$

The conditional variance formula now yields

$$\begin{aligned}
\text{Var}(X_n) &= E\big[X_{n-1}\sigma^2\big] + \text{Var}(X_{n-1}\mu) \\
&= \sigma^2\mu^{n-1} + \mu^2\text{Var}(X_{n-1}) \\
&= \sigma^2\mu^{n-1} + \mu^2\big(\sigma^2\mu^{n-2} + \mu^2\text{Var}(X_{n-2})\big) \\
&= \sigma^2\big(\mu^{n-1} + \mu^n\big) + \mu^4\text{Var}(X_{n-2}) \\
&= \sigma^2\big(\mu^{n-1} + \mu^n\big) + \mu^4\big(\sigma^2\mu^{n-3} + \mu^2\text{Var}(X_{n-3})\big) \\
&= \sigma^2\big(\mu^{n-1} + \mu^n + \mu^{n+1}\big) + \mu^6\text{Var}(X_{n-3}) \\
&= \cdots \\
&= \sigma^2\big(\mu^{n-1} + \mu^n + \cdots + \mu^{2n-2}\big) + \mu^{2n}\text{Var}(X_0) \\
&= \sigma^2\big(\mu^{n-1} + \mu^n + \cdots + \mu^{2n-2}\big)
\end{aligned}$$

Therefore,

$$\text{Var}(X_n) = \begin{cases} \sigma^2\mu^{n-1}\left(\frac{1-\mu^n}{1-\mu}\right), & \text{if } \mu \neq 1 \\ n\sigma^2, & \text{if } \mu = 1 \end{cases} \tag{4.19}$$

Let π_0 denote the probability that the population will eventually die out (under the assumption that $X_0 = 1$). More formally,

$$\pi_0 = \lim_{n\to\infty} P\{X_n = 0|X_0 = 1\}$$

The problem of determining the value of π_0 was first raised in connection with the extinction of family surnames by Galton in 1889.

We first note that $\pi_0 = 1$ if $\mu < 1$. This follows since

$$\begin{aligned}
\mu^n = E[X_n] &= \sum_{j=1}^{\infty} j\,P\{X_n = j\} \\
&\geqslant \sum_{j=1}^{\infty} 1 \cdot P\{X_n = j\} \\
&= P\{X_n \geqslant 1\}
\end{aligned}$$

Since $\mu^n \to 0$ when $\mu < 1$, it follows that $P\{X_n \geqslant 1\} \to 0$, and hence $P\{X_n = 0\} \to 1$.

In fact, it can be shown that $\pi_0 = 1$ even when $\mu = 1$. When $\mu > 1$, it turns out that $\pi_0 < 1$, and an equation determining π_0 may be derived by conditioning on the

number of offspring of the initial individual, as follows:

$$\begin{aligned}\pi_0 &= P\{\text{population dies out}\}\\ &= \sum_{j=0}^{\infty} P\{\text{population dies out}|X_1 = j\}P_j\end{aligned}$$

Now, given that $X_1 = j$, the population will eventually die out if and only if each of the j families started by the members of the first generation eventually dies out. Since each family is assumed to act independently, and since the probability that any particular family dies out is just π_0, this yields

$$P\{\text{population dies out}|X_1 = j\} = \pi_0^j$$

and thus π_0 satisfies

$$\pi_0 = \sum_{j=0}^{\infty} \pi_0^j P_j \tag{4.20}$$

In fact when $\mu > 1$, it can be shown that π_0 is the smallest positive number satisfying Eq. (4.20).

Example 4.34. If $P_0 = \frac{1}{2}$, $P_1 = \frac{1}{4}$, $P_2 = \frac{1}{4}$, then determine π_0.

Solution: Since $\mu = \frac{3}{4} \leqslant 1$, it follows that $\pi_0 = 1$. ■

Example 4.35. If $P_0 = \frac{1}{4}$, $P_1 = \frac{1}{4}$, $P_2 = \frac{1}{2}$, then determine π_0.

Solution: π_0 satisfies

$$\pi_0 = \tfrac{1}{4} + \tfrac{1}{4}\pi_0 + \tfrac{1}{2}\pi_0^2$$

or

$$2\pi_0^2 - 3\pi_0 + 1 = 0$$

The smallest positive solution of this quadratic equation is $\pi_0 = \frac{1}{2}$. ■

Example 4.36. In Examples 4.34 and 4.35, what is the probability that the population will die out if it initially consists of n individuals?

Solution: Since the population will die out if and only if the families of each of the members of the initial generation die out, the desired probability is π_0^n. For Example 4.34 this yields $\pi_0^n = 1$, and for Example 4.35, $\pi_0^n = \left(\frac{1}{2}\right)^n$. ■

4.8 Time Reversible Markov Chains

Consider a stationary ergodic Markov chain (that is, an ergodic Markov chain that has been in operation for a long time) having transition probabilities P_{ij} and stationary probabilities π_i, and suppose that starting at some time we trace the sequence of states going backward in time. That is, starting at time n, consider the sequence of states $X_n, X_{n-1}, X_{n-2}, \ldots$. It turns out that this sequence of states is itself a Markov chain with transition probabilities Q_{ij} defined by

$$\begin{aligned} Q_{ij} &= P\{X_m = j | X_{m+1} = i\} \\ &= \frac{P\{X_m = j, X_{m+1} = i\}}{P\{X_{m+1} = i\}} \\ &= \frac{P\{X_m = j\} P\{X_{m+1} = i | X_m = j\}}{P\{X_{m+1} = i\}} \\ &= \frac{\pi_j P_{ji}}{\pi_i} \end{aligned}$$

To prove that the reversed process is indeed a Markov chain, we must verify that

$$P\{X_m = j | X_{m+1} = i, X_{m+2}, X_{m+3}, \ldots\} = P\{X_m = j | X_{m+1} = i\}$$

To see that this is so, suppose that the present time is $m+1$. Now, since $X_0, X_1, X_2, \ldots$ is a Markov chain, it follows that the conditional distribution of the future X_{m+2}, $X_{m+3}, \ldots$ given the present state X_{m+1} is independent of the past state X_m. However, independence is a symmetric relationship (that is, if A is independent of B, then B is independent of A), and so this means that given X_{m+1}, X_m is independent of $X_{m+2}, X_{m+3}, \ldots$. But this is exactly what we had to verify.

Thus, the reversed process is also a Markov chain with transition probabilities given by

$$Q_{ij} = \frac{\pi_j P_{ji}}{\pi_i}$$

If $Q_{ij} = P_{ij}$ for all i, j, then the Markov chain is said to be *time reversible*. The condition for time reversibility, namely, $Q_{ij} = P_{ij}$, can also be expressed as

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \text{for all } i, j \tag{4.21}$$

The condition in Eq. (4.21) can be stated that, for all states i and j, the rate at which the process goes from i to j (namely, $\pi_i P_{ij}$) is equal to the rate at which it goes from j to i (namely, $\pi_j P_{ji}$). It is worth noting that this is an obvious necessary condition for time reversibility since a transition from i to j going backward in time is equivalent to a transition from j to i going forward in time; that is, if $X_m = i$ and $X_{m-1} = j$, then a transition from i to j is observed if we are looking backward, and one from j to i if we are looking forward in time. Thus, the rate at which the forward process makes a transition from j to i is always equal to the rate at which the reverse process makes a

transition from i to j; if time reversible, this must equal the rate at which the forward process makes a transition from i to j.

If we can find nonnegative numbers, summing to one, that satisfy Eq. (4.21), then it follows that the Markov chain is time reversible and the numbers represent the limiting probabilities. This is so since if

$$x_i P_{ij} = x_j P_{ji} \quad \text{for all } i, j, \sum_i x_i = 1 \tag{4.22}$$

then summing over i yields

$$\sum_i x_i P_{ij} = x_j \sum_i P_{ji} = x_j, \quad \sum_i x_i = 1$$

and, because the limiting probabilities π_i are the unique solution of the preceding, it follows that $x_i = \pi_i$ for all i.

Example 4.37. Consider a random walk with states $0, 1, \ldots, M$ and transition probabilities

$$\begin{aligned} P_{i,i+1} &= \alpha_i = 1 - P_{i,i-1}, \quad i = 1, \ldots, M-1, \\ P_{0,1} &= \alpha_0 = 1 - P_{0,0}, \\ P_{M,M} &= \alpha_M = 1 - P_{M,M-1} \end{aligned}$$

Without the need for any computations, it is possible to argue that this Markov chain, which can only make transitions from a state to one of its two nearest neighbors, is time reversible. This follows by noting that the number of transitions from i to $i+1$ must at all times be within 1 of the number from $i+1$ to i. This is so because between any two transitions from i to $i+1$ there must be one from $i+1$ to i (and conversely) since the only way to reenter i from a higher state is via state $i+1$. Hence, it follows that the rate of transitions from i to $i+1$ equals the rate from $i+1$ to i, and so the process is time reversible.

We can easily obtain the limiting probabilities by equating for each state $i = 0, 1, \ldots, M-1$ the rate at which the process goes from i to $i+1$ with the rate at which it goes from $i+1$ to i. This yields

$$\begin{aligned} \pi_0\alpha_0 &= \pi_1(1-\alpha_1), \\ \pi_1\alpha_1 &= \pi_2(1-\alpha_2), \\ &\vdots \\ \pi_i\alpha_i &= \pi_{i+1}(1-\alpha_{i+1}), \quad i = 0, 1, \ldots, M-1 \end{aligned}$$

Solving in terms of π_0 yields

$$\pi_1 = \frac{\alpha_0}{1-\alpha_1}\pi_0,$$

$$\pi_2 = \frac{\alpha_1}{1-\alpha_2}\pi_1 = \frac{\alpha_1\alpha_0}{(1-\alpha_2)(1-\alpha_1)}\pi_0$$

and, in general,

$$\pi_i = \frac{\alpha_{i-1}\cdots\alpha_0}{(1-\alpha_i)\cdots(1-\alpha_1)}\pi_0, \quad i = 1, 2, \ldots, M$$

Since $\sum_0^M \pi_i = 1$, we obtain

$$\pi_0\left[1 + \sum_{j=1}^{M} \frac{\alpha_{j-1}\cdots\alpha_0}{(1-\alpha_j)\cdots(1-\alpha_1)}\right] = 1$$

or

$$\pi_0 = \left[1 + \sum_{j=1}^{M} \frac{\alpha_{j-1}\cdots\alpha_0}{(1-\alpha_j)\cdots(1-\alpha_1)}\right]^{-1} \tag{4.23}$$

and

$$\pi_i = \frac{\alpha_{i-1}\cdots\alpha_0}{(1-\alpha_i)\cdots(1-\alpha_1)}\pi_0, \quad i = 1, \ldots, M \tag{4.24}$$

For instance, if $\alpha_i \equiv \alpha$, then

$$\begin{aligned}\pi_0 &= \left[1 + \sum_{j=1}^{M}\left(\frac{\alpha}{1-\alpha}\right)^j\right]^{-1} \\ &= \frac{1-\beta}{1-\beta^{M+1}}\end{aligned}$$

and, in general,

$$\pi_i = \frac{\beta^i(1-\beta)}{1-\beta^{M+1}}, \quad i = 0, 1, \ldots, M$$

where

$$\beta = \frac{\alpha}{1-\alpha}$$

■

Another special case of Example 4.37 is the following urn model, proposed by the physicists P. and T. Ehrenfest to describe the movements of molecules. Suppose that M molecules are distributed among two urns; and at each time point one of the molecules is chosen at random, removed from its urn, and placed in the other one. The number of molecules in urn I is a special case of the Markov chain of Example 4.37 having

$$\alpha_i = \frac{M-i}{M}, \quad i = 0, 1, \ldots, M$$

Hence, using Eqs. (4.23) and (4.24) the limiting probabilities in this case are

$$\begin{aligned}\pi_0 &= \left[1 + \sum_{j=1}^{M} \frac{(M-j+1)\cdots(M-1)M}{j(j-1)\cdots 1}\right]^{-1} \\ &= \left[\sum_{j=0}^{M} \binom{M}{j}\right]^{-1} \\ &= \left(\frac{1}{2}\right)^{M}\end{aligned}$$

where we have used the identity

$$\begin{aligned}1 &= \left(\frac{1}{2} + \frac{1}{2}\right)^{M} \\ &= \sum_{j=0}^{M} \binom{M}{j} \left(\frac{1}{2}\right)^{M}\end{aligned}$$

Hence, from Eq. (4.24)

$$\pi_i = \binom{M}{i} \left(\frac{1}{2}\right)^{M}, \quad i = 0, 1, \ldots, M$$

Because the preceding are just the binomial probabilities, it follows that in the long run, the positions of each of the M balls are independent and each one is equally likely to be in either urn. This, however, is quite intuitive, for if we focus on any one ball, it becomes quite clear that its position will be independent of the positions of the other balls (since no matter where the other $M-1$ balls are, the ball under consideration at each stage will be moved with probability $1/M$) and by symmetry, it is equally likely to be in either urn.

Example 4.38. Consider an arbitrary connected graph (see Section 3.6 for definitions) having a number w_{ij} associated with arc (i, j) for each arc. One instance of such a graph is given by Fig. 4.1. Now consider a particle moving from node to node in this

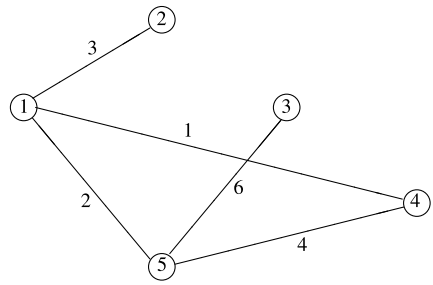

Figure 4.1 A connected graph with arc weights.

manner: If at any time the particle resides at node i, then it will next move to node j with probability P_{ij} where

$$P_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$$

and where w_{ij} is 0 if (i, j) is not an arc. For instance, for the graph of Fig. 4.1, $P_{12} = 3/(3+1+2) = \frac{1}{2}$.

The time reversibility equations

$$\pi_i P_{ij} = \pi_j P_{ji}$$

reduce to

$$\pi_i \frac{w_{ij}}{\sum_j w_{ij}} = \pi_j \frac{w_{ji}}{\sum_i w_{ji}}$$

or, equivalently, since $w_{ij} = w_{ji}$

$$\frac{\pi_i}{\sum_j w_{ij}} = \frac{\pi_j}{\sum_i w_{ji}}$$

which is equivalent to

$$\frac{\pi_i}{\sum_j w_{ij}} = c$$

or

$$\pi_i = c \sum_j w_{ij}$$

or, since $1 = \sum_i \pi_i$

$$\pi_i = \frac{\sum_j w_{ij}}{\sum_i \sum_j w_{ij}}$$

Because the π_is given by this equation satisfy the time reversibility equations, it follows that the process is time reversible with these limiting probabilities.

For the graph of Fig. 4.1 we have that

$$\pi_1 = \tfrac{6}{32}, \quad \pi_2 = \tfrac{3}{32}, \quad \pi_3 = \tfrac{6}{32}, \quad \pi_4 = \tfrac{5}{32}, \quad \pi_5 = \tfrac{12}{32}$$ ■

If we try to solve Eq. (4.22) for an arbitrary Markov chain with states $0, 1, \ldots, M$, it will usually turn out that no solution exists. For example, from Eq. (4.22),

$$x_i P_{ij} = x_j P_{ji},$$
$$x_k P_{kj} = x_j P_{jk}$$

implying (if $P_{ij}P_{jk} > 0$) that

$$\frac{x_i}{x_k} = \frac{P_{ji}P_{kj}}{P_{ij}P_{jk}}$$

which in general need not equal P_{ki}/P_{ik}. Thus, we see that a necessary condition for time reversibility is that

$$P_{ik}P_{kj}P_{ji} = P_{ij}P_{jk}P_{ki} \quad \text{for all } i, j, k \tag{4.25}$$

which is equivalent to the statement that, starting in state i, the path $i \to k \to j \to i$ has the same probability as the reversed path $i \to j \to k \to i$. To understand the necessity of this, note that time reversibility implies that the rate at which a sequence of transitions from i to k to j to i occurs must equal the rate of ones from i to j to k to i (why?), and so we must have

$$\pi_i P_{ik}P_{kj}P_{ji} = \pi_i P_{ij}P_{jk}P_{ki}$$

implying Eq. (4.25) when $\pi_i > 0$.

In fact, we can show the following.

Theorem 4.2. *A stationary Markov chain for which $P_{ij} = 0$ whenever $P_{ji} = 0$ is time reversible if and only if starting in state i, any path back to i has the same probability as the reversed path. That is, if*

$$P_{i,i_1}P_{i_1,i_2}\cdots P_{i_k,i} = P_{i,i_k}P_{i_k,i_{k-1}}\cdots P_{i_1,i} \tag{4.26}$$

for all states $i, i_1, \ldots, i_k$.

Proof. We have already proven necessity. To prove sufficiency, fix states i and j and rewrite (4.26) as

$$P_{i,i_1}P_{i_1,i_2}\cdots P_{i_k,j}P_{ji} = P_{ij}P_{j,i_k}\cdots P_{i_1,i}$$

Summing the preceding over all states $i_1, \ldots, i_k$ yields

$$P_{ij}^{k+1}P_{ji} = P_{ij}P_{ji}^{k+1}$$

Consequently,

$$\frac{P_{ji}\sum_{k=1}^{m}P_{ij}^{k+1}}{m} = \frac{P_{ij}\sum_{k=1}^{m}P_{ji}^{k+1}}{m}$$

Letting $m \to \infty$ yields

$$P_{ji}\pi_j = P_{ij}\pi_i$$

which proves the theorem. ■

Example 4.39. Suppose we are given a set of n elements, numbered 1 through n, which are to be arranged in some ordered list. At each unit of time a request is made to retrieve one of these elements, element i being requested (independently of the past) with probability P_i. After being requested, the element then is put back but not necessarily in the same position. In fact, let us suppose that the element requested is moved one closer to the front of the list; for instance, if the present list ordering is 1, 3, 4, 2, 5 and element 2 is requested, then the new ordering becomes 1, 3, 2, 4, 5. We are interested in the long-run average position of the element requested.

For any given probability vector $P = (P_1, \ldots, P_n)$, the preceding can be modeled as a Markov chain with $n!$ states, with the state at any time being the list order at that time. We shall show that this Markov chain is time reversible and then use this to show that the average position of the element requested when this one-closer rule is in effect is less than when the rule of always moving the requested element to the front of the line is used. The time reversibility of the resulting Markov chain when the one-closer reordering rule is in effect easily follows from Theorem 4.2. For instance, suppose $n = 3$ and consider the following path from state (1, 2, 3) to itself:

$$(1,2,3) \to (2,1,3) \to (2,3,1) \to (3,2,1)$$
$$\to (3,1,2) \to (1,3,2) \to (1,2,3)$$

The product of the transition probabilities in the forward direction is

$$P_2 P_3 P_3 P_1 P_1 P_2 = P_1^2 P_2^2 P_3^2$$

whereas in the reverse direction, it is

$$P_3 P_3 P_2 P_2 P_1 P_1 = P_1^2 P_2^2 P_3^2$$

Because the general result follows in much the same manner, the Markov chain is indeed time reversible. (For a formal argument note that if f_i denotes the number of times element i moves forward in the path, then as the path goes from a fixed state back to itself, it follows that element i will also move backward f_i times. Therefore, since the backward moves of element i are precisely the times that it moves forward in the reverse path, it follows that the product of the transition probabilities for both the path and its reversal will equal

$$\prod_i P_i^{f_i + r_i}$$

where r_i is equal to the number of times that element i is in the first position and the path (or the reverse path) does not change states.)

For any permutation $i_1, i_2, \ldots, i_n$ of $1, 2, \ldots, n$, let $\pi(i_1, i_2, \ldots, i_n)$ denote the limiting probability under the one-closer rule. By time reversibility we have

$$P_{i_{j+1}} \pi(i_1, \ldots, i_j, i_{j+1}, \ldots, i_n) = P_{i_j} \pi(i_1, \ldots, i_{j+1}, i_j, \ldots, i_n) \tag{4.27}$$

for all permutations.

Now the average position of the element requested can be expressed (as in Section 3.6.1) as

$$\begin{aligned}
\text{Average position} &= \sum_i P_i E[\text{Position of element } i] \\
&= \sum_i P_i \left[1 + \sum_{j \neq i} P\{\text{element } j \text{ precedes element } i\} \right] \\
&= 1 + \sum_i \sum_{j \neq i} P_i P\{e_j \text{ precedes } e_i\} \\
&= 1 + \sum_{i<j} \sum [P_i P\{e_j \text{ precedes } e_i\} + P_j P\{e_i \text{ precedes } e_j\}] \\
&= 1 + \sum_{i<j} \sum [P_i P\{e_j \text{ precedes } e_i\} \\
&\qquad + P_j (1 - P\{e_j \text{ precedes } e_i\})] \\
&= 1 + \sum_{i<j} \sum (P_i - P_j) P\{e_j \text{ precedes } e_i\} + \sum_{i<j} \sum P_j
\end{aligned}$$

Hence, to minimize the average position of the element requested, we would want to make $P\{e_j \text{ precedes } e_i\}$ as large as possible when $P_j > P_i$ and as small as possible when $P_i > P_j$. Under the front-of-the-line rule we showed in Section 3.6.1,

$$P\{e_j \text{ precedes } e_i\} = \frac{P_j}{P_j + P_i}$$

(since under the front-of-the-line rule element j will precede element i if and only if the last request for either i or j was for j).

Therefore, to show that the one-closer rule is better than the front-of-the-line rule, it suffices to show that under the one-closer rule

$$P\{e_j \text{ precedes } e_i\} > \frac{P_j}{P_j + P_i} \quad \text{when } P_j > P_i$$

Now consider any state where element i precedes element j, say, $(\ldots, i, i_1, \ldots, i_k, j, \ldots)$. By successive transpositions using Eq. (4.27), we have

$$\pi(\ldots, i, i_1, \ldots, i_k, j, \ldots) = \left(\frac{P_i}{P_j} \right)^{k+1} \pi(\ldots, j, i_1, \ldots, i_k, i, \ldots) \tag{4.28}$$

For instance,

$$\begin{aligned}
\pi(1,2,3) &= \frac{P_2}{P_3} \pi(1,3,2) = \frac{P_2}{P_3} \frac{P_1}{P_3} \pi(3,1,2) \\
&= \frac{P_2}{P_3} \frac{P_1}{P_3} \frac{P_1}{P_2} \pi(3,2,1) = \left(\frac{P_1}{P_3} \right)^2 \pi(3,2,1)
\end{aligned}$$

Now when $P_j > P_i$, Eq. (4.28) implies that

$$\pi(\ldots, i, i_1, \ldots, i_k, j, \ldots) < \frac{P_i}{P_j}\pi(\ldots, j, i_1, \ldots, i_k, i, \ldots)$$

Letting $\alpha(i, j) = P\{e_i \text{ precedes } e_j\}$, we see by summing over all states for which i precedes j and by using the preceding that

$$\alpha(i, j) < \frac{P_i}{P_j}\alpha(j, i)$$

which, since $\alpha(i, j) = 1 - \alpha(j, i)$, yields

$$\alpha(j, i) > \frac{P_j}{P_j + P_i}$$

Hence, the average position of the element requested is indeed smaller under the one-closer rule than under the front-of-the-line rule. ■

The concept of the reversed chain is useful even when the process is not time reversible. To illustrate this, we start with the following proposition whose proof is left as an exercise.

Proposition 4.9. *Consider an irreducible Markov chain with transition probabilities P_{ij}. If we can find positive numbers $\pi_i, i \geqslant 0$, summing to one, and a transition probability matrix $\mathbf{Q} = [Q_{ij}]$ such that*

$$\pi_i P_{ij} = \pi_j Q_{ji} \tag{4.29}$$

then the Q_{ij} are the transition probabilities of the reversed chain and the π_i are the stationary probabilities both for the original and reversed chain.

The importance of the preceding proposition is that, by thinking backward, we can sometimes guess at the nature of the reversed chain and then use the set of Eqs. (4.29) to obtain both the stationary probabilities and the Q_{ij}.

Example 4.40. A single bulb is necessary to light a given room. When the bulb in use fails, it is replaced by a new one at the beginning of the next day. Let X_n equal i if the bulb in use at the beginning of day n is in its ith day of use (that is, if its present age is i). For instance, if a bulb fails on day $n - 1$, then a new bulb will be put in use at the beginning of day n and so $X_n = 1$. If we suppose that each bulb, independently, fails on its ith day of use with probability $p_i, i \geqslant 1$, then it is easy to see that $\{X_n, n \geqslant 1\}$ is a Markov chain whose transition probabilities are as follows:

$$\begin{aligned} P_{i,1} &= P\{\text{bulb, on its } i\text{th day of use, fails}\} \\ &= P\{\text{life of bulb} = i \,|\, \text{life of bulb} \geqslant i\} \\ &= \frac{P\{L = i\}}{P\{L \geqslant i\}} \end{aligned}$$

where L, a random variable representing the lifetime of a bulb, is such that $P\{L = i\} = p_i$. Also,

$$P_{i,i+1} = 1 - P_{i,1}$$

Suppose now that this chain has been in operation for a long (in theory, an infinite) time and consider the sequence of states going backward in time. Since, in the forward direction, the state is always increasing by 1 until it reaches the age at which the item fails, it is easy to see that the reverse chain will always decrease by 1 until it reaches 1 and then it will jump to a random value representing the lifetime of the (in real time) previous bulb. Thus, it seems that the reverse chain should have transition probabilities given by

$$\begin{aligned} Q_{i,i-1} &= 1, \quad i > 1 \\ Q_{1,i} &= p_i, \quad i \geqslant 1 \end{aligned}$$

To check this, and at the same time determine the stationary probabilities, we must see if we can find, with the $Q_{i,j}$ as previously given, positive numbers $\{\pi_i\}$ such that

$$\pi_i P_{i,j} = \pi_j Q_{j,i}$$

To begin, let $j = 1$ and consider the resulting equations:

$$\pi_i P_{i,1} = \pi_1 Q_{1,i}$$

This is equivalent to

$$\pi_i \frac{P\{L = i\}}{P\{L \geqslant i\}} = \pi_1 P\{L = i\}$$

or

$$\pi_i = \pi_1 P\{L \geqslant i\}$$

Summing over all i yields

$$1 = \sum_{i=1}^{\infty} \pi_i = \pi_1 \sum_{i=1}^{\infty} P\{L \geqslant i\} = \pi_1 E[L]$$

and so, for the preceding $Q_{i,j}$ to represent the reverse transition probabilities, it is necessary for the stationary probabilities to be

$$\pi_i = \frac{P\{L \geqslant i\}}{E[L]}, \quad i \geqslant 1$$

To finish the proof that the reverse transition probabilities and stationary probabilities are as given, all that remains is to show that they satisfy

$$\pi_i P_{i,i+1} = \pi_{i+1} Q_{i+1,i}$$

which is equivalent to

$$\frac{P\{L \geqslant i\}}{E[L]}\left(1 - \frac{P\{L = i\}}{P\{L \geqslant i\}}\right) = \frac{P\{L \geqslant i+1\}}{E[L]}$$

and which is true since $P\{L \geqslant i\} - P\{L = i\} = P\{L \geqslant i+1\}$. ■

4.9 Markov Chain Monte Carlo Methods

Let $\mathbf{X}$ be a discrete random vector whose set of possible values is $\mathbf{x}_j$, $j \geqslant 1$. Let the probability mass function of $\mathbf{X}$ be given by $P\{\mathbf{X} = \mathbf{x}_j\}$, $j \geqslant 1$, and suppose that we are interested in calculating

$$\theta = E[h(\mathbf{X})] = \sum_{j=1}^{\infty} h(\mathbf{x}_j) P\{\mathbf{X} = \mathbf{x}_j\}$$

for some specified function h. In situations where it is computationally difficult to evaluate the function $h(\mathbf{x}_j)$, $j \geqslant 1$, we often turn to simulation to approximate θ. The usual approach, called *Monte Carlo simulation*, is to use random numbers to generate a partial sequence of independent and identically distributed random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ having the mass function $P\{\mathbf{X} = \mathbf{x}_j\}$, $j \geqslant 1$ (see Chapter 11 for a discussion as to how this can be accomplished). Since the strong law of large numbers yields

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{h(\mathbf{X}_i)}{n} = \theta \tag{4.30}$$

it follows that we can estimate θ by letting n be large and using the average of the values of $h(\mathbf{X}_i)$, $i = 1, \ldots, n$ as the estimator.

It often, however, turns out that it is difficult to generate a random vector having the specified probability mass function, particularly if $\mathbf{X}$ is a vector of dependent random variables. In addition, its probability mass function is sometimes given in the form $P\{\mathbf{X} = \mathbf{x}_j\} = Cb_j$, $j \geqslant 1$, where the b_j are specified, but C must be computed, and in many applications it is not computationally feasible to sum the b_j so as to determine C. Fortunately, however, there is another way of using simulation to estimate θ in these situations. It works by generating a sequence, not of independent random vectors, but of the successive states of a vector-valued Markov chain $\mathbf{X}_1, \mathbf{X}_2, \ldots$ whose stationary probabilities are $P\{\mathbf{X} = \mathbf{x}_j\}$, $j \geqslant 1$. If this can be accomplished, then it would follow from Proposition 4.6 that Eq. (4.30) remains valid, implying that we can then use $\sum_{i=1}^{n} h(\mathbf{X}_i)/n$ as an estimator of θ.

We now show how to generate a Markov chain with arbitrary stationary probabilities that may only be specified up to a multiplicative constant. Let $b(j)$, $j = 1, 2, \ldots$ be positive numbers whose sum $B = \sum_{j=1}^{\infty} b(j)$ is finite. The following, known as

the *Hastings–Metropolis algorithm*, can be used to generate a time reversible Markov chain whose stationary probabilities are

$$\pi(j) = b(j)/B, \quad j = 1, 2, \ldots$$

To begin, let $\mathbf{Q}$ be any specified irreducible Markov transition probability matrix on the integers, with $q(i, j)$ representing the row i column j element of $\mathbf{Q}$. Now define a Markov chain $\{X_n, n \geqslant 0\}$ as follows. When $X_n = i$, generate a random variable Y such that $P\{Y = j\} = q(i, j)$, $j = 1, 2, \ldots$. If $Y = j$, then set X_{n+1} equal to j with probability $\alpha(i, j)$, and set it equal to i with probability $1 - \alpha(i, j)$. Under these conditions, it is easy to see that the sequence of states constitutes a Markov chain with transition probabilities $P_{i,j}$ given by

$$P_{i,j} = q(i, j)\alpha(i, j), \quad \text{if } j \neq i$$
$$P_{i,i} = q(i, i) + \sum_{k \neq i} q(i, k)(1 - \alpha(i, k))$$

This Markov chain will be time reversible and have stationary probabilities $\pi(j)$ if

$$\pi(i)P_{i,j} = \pi(j)P_{j,i} \quad \text{for } j \neq i$$

which is equivalent to

$$\pi(i)q(i, j)\alpha(i, j) = \pi(j)q(j, i)\alpha(j, i) \tag{4.31}$$

But if we take $\pi(j) = b(j)/B$ and set

$$\alpha(i, j) = \min\left(\frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}, 1\right) \tag{4.32}$$

then Eq. (4.31) is easily seen to be satisfied. For if

$$\alpha(i, j) = \frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}$$

then $\alpha(j, i) = 1$ and Eq. (4.31) follows, and if $\alpha(i, j) = 1$ then

$$\alpha(j, i) = \frac{\pi(i)q(i, j)}{\pi(j)q(j, i)}$$

and again Eq. (4.31) holds, thus showing that the Markov chain is time reversible with stationary probabilities $\pi(j)$. Also, since $\pi(j) = b(j)/B$, we see from (4.32) that

$$\alpha(i, j) = \min\left(\frac{b(j)q(j, i)}{b(i)q(i, j)}, 1\right)$$

which shows that the value of B is not needed to define the Markov chain, because the values $b(j)$ suffice. Also, it is almost always the case that $\pi(j)$, $j \geqslant 1$ will not only be stationary probabilities but will also be limiting probabilities. (Indeed, a sufficient condition is that $P_{i,i} > 0$ for some i.)

Example 4.41. Suppose that we want to generate a uniformly distributed element in $\mathscr{S}$, the set of all permutations $(x_1, \ldots, x_n)$ of the numbers $(1, \ldots, n)$ for which $\sum_{j=1}^{n} j x_j > a$ for a given constant a. To utilize the Hastings–Metropolis algorithm we need to define an irreducible Markov transition probability matrix on the state space $\mathscr{S}$. To accomplish this, we first define a concept of "neighboring" elements of $\mathscr{S}$, and then construct a graph whose vertex set is $\mathscr{S}$. We start by putting an arc between each pair of neighboring elements in $\mathscr{S}$, where any two permutations in $\mathscr{S}$ are said to be neighbors if one results from an interchange of two of the positions of the other. That is, (1, 2, 3, 4) and (1, 2, 4, 3) are neighbors whereas (1, 2, 3, 4) and (1, 3, 4, 2) are not. Now, define the q transition probability function as follows. With $N(s)$ defined as the set of neighbors of $\mathbf{s}$, and $|N(\mathbf{s})|$ equal to the number of elements in the set $N(\mathbf{s})$, let

$$q(\mathbf{s}, \mathbf{t}) = \frac{1}{|N(\mathbf{s})|} \quad \text{if } t \in N(\mathbf{s})$$

That is, the candidate next state from $\mathbf{s}$ is equally likely to be any of its neighbors. Since the desired limiting probabilities of the Markov chain are $\pi(\mathbf{s}) = C$, it follows that $\pi(\mathbf{s}) = \pi(\mathbf{t})$, and so

$$\alpha(\mathbf{s}, \mathbf{t}) = \min(|N(\mathbf{s})|/|N(\mathbf{t})|, 1)$$

That is, if the present state of the Markov chain is $\mathbf{s}$ then one of its neighbors is randomly chosen, say, $\mathbf{t}$. If $\mathbf{t}$ is a state with fewer neighbors than $\mathbf{s}$ (in graph theory language, if the degree of vertex $\mathbf{t}$ is less than that of vertex $\mathbf{s}$), then the next state is $\mathbf{t}$. If not, a uniform $(0, 1)$ random number U is generated and the next state is $\mathbf{t}$ if $U < |(N(\mathbf{s})|/|N(\mathbf{t})|$ and is $\mathbf{s}$ otherwise. The limiting probabilities of this Markov chain are $\pi(\mathbf{s}) = 1/|\mathscr{S}|$, where $|\mathscr{S}|$ is the (unknown) number of permutations in $\mathscr{S}$. ■

The most widely used version of the Hastings–Metropolis algorithm is the *Gibbs sampler*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a discrete random vector with probability mass function $p(\mathbf{x})$ that is only specified up to a multiplicative constant, and suppose that we want to generate a random vector whose distribution is that of $\mathbf{X}$. That is, we want to generate a random vector having mass function

$$p(\mathbf{x}) = Cg(\mathbf{x})$$

where $g(\mathbf{x})$ is known, but C is not. Utilization of the Gibbs sampler assumes that for any i and values x_j, $j \neq i$, we can generate a random variable X having the probability mass function

$$P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$$

It operates by using the Hasting–Metropolis algorithm on a Markov chain with states $\mathbf{x} = (x_1, \ldots, x_n)$, and with transition probabilities defined as follows. Whenever the present state is $\mathbf{x}$, a coordinate that is equally likely to be any of $1, \ldots, n$ is chosen. If coordinate i is chosen, then a random variable X with probability mass function $P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$ is generated. If $X = x$, then the state $\mathbf{y} =$

$(x_1, \ldots x_{i-1}, x, x_{i+1}, \ldots, x_n)$ is considered as the candidate next state. In other words, with $\mathbf{x}$ and $\mathbf{y}$ as given, the Gibbs sampler uses the Hastings–Metropolis algorithm with

$$q(\mathbf{x}, \mathbf{y}) = \frac{1}{n} P\{X_i = x | X_j = x_j, j \neq i\} = \frac{p(\mathbf{y})}{nP\{X_j = x_j, j \neq i\}}$$

Because we want the limiting mass function to be p, we see from Eq. (4.32) that the vector $\mathbf{y}$ is then accepted as the new state with probability

$$\begin{aligned} \alpha(\mathbf{x}, \mathbf{y}) &= \min\left(\frac{p(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{p(\mathbf{x})q(\mathbf{x}, \mathbf{y})}, 1\right) \\ &= \min\left(\frac{p(\mathbf{y})p(\mathbf{x})}{p(\mathbf{x})p(\mathbf{y})}, 1\right) \\ &= 1 \end{aligned}$$

Hence, when utilizing the Gibbs sampler, the candidate state is always accepted as the next state of the chain.

Example 4.42. Suppose that we want to generate n uniformly distributed points in the circle of radius 1 centered at the origin, conditional on the event that no two points are within a distance d of each other, when the probability of this conditioning event is small. This can be accomplished by using the Gibbs sampler as follows. Start with any n points $\mathbf{x_1}, \ldots, \mathbf{x_n}$ in the circle that have the property that no two of them are within d of the other; then generate the value of I, equally likely to be any of the values $1, \ldots, n$. Then continually generate a random point in the circle until you obtain one that is not within d of any of the other $n - 1$ points excluding $\mathbf{x}_I$. At this point, replace $\mathbf{x}_I$ by the generated point and then repeat the operation. After a large number of iterations of this algorithm, the set of n points will approximately have the desired distribution. ■

Example 4.43. Let $X_i, i = 1, \ldots, n$, be independent exponential random variables with respective rates $\lambda_i, i = 1, \ldots, n$. Let $S = \sum_{i=1}^{n} X_i$, and suppose that we want to generate the random vector $\mathbf{X} = (X_1, \ldots, X_n)$, conditional on the event that $S > c$ for some large positive constant c. That is, we want to generate the value of a random vector whose density function is

$$f(x_1, \ldots, x_n) = \frac{1}{P\{S > c\}} \prod_{i=1}^{n} \lambda_i e^{-\lambda_i x_i}, \quad x_i \geqslant 0, \sum_{i=1}^{n} x_i > c$$

This is easily accomplished by starting with an initial vector $\mathbf{x} = (x_1, \ldots, x_n)$ satisfying $x_i > 0, i = 1, \ldots, n, \sum_{i=1}^{n} x_i > c$. Then generate a random variable I that is equally likely to be any of $1, \ldots, n$. Next, generate an exponential random variable X with rate λ_I conditional on the event that $X + \sum_{j \neq I} x_j > c$. This latter step, which calls for generating the value of an exponential random variable given that it exceeds $c - \sum_{j \neq I} x_j$, is easily accomplished by using the fact that an exponential conditioned to be greater than a positive constant is distributed as the constant plus the exponential.

Consequently, to obtain X, first generate an exponential random variable Y with rate λ_I, and then set

$$X = Y + \left(c - \sum_{j \neq I} x_j \right)^+$$

where $a^+ = \max(a, 0)$.
The value of x_I should then be reset as X and a new iteration of the algorithm begun. ■

Remark. As can be seen by Examples 4.42 and 4.43, although the theory for the Gibbs sampler was represented under the assumption that the distribution to be generated was discrete, it also holds when this distribution is continuous.

4.10 Markov Decision Processes

Consider a process that is observed at discrete time points to be in any one of M possible states, which we number by $1, 2, \ldots, M$. After observing the state of the process, an action must be chosen, and we let A, assumed finite, denote the set of all possible actions.

If the process is in state i at time n and action a is chosen, then the next state of the system is determined according to the transition probabilities $P_{ij}(a)$. If we let X_n denote the state of the process at time n and a_n the action chosen at time n, then the preceding is equivalent to stating that

$$P\{X_{n+1} = j | X_0, a_0, X_1, a_1, \ldots, X_n = i, a_n = a\} = P_{ij}(a)$$

Thus, the transition probabilities are functions only of the present state and the subsequent action.

By a policy, we mean a rule for choosing actions. We shall restrict ourselves to policies that are of the form that the action they prescribe at any time depends only on the state of the process at that time (and not on any information concerning prior states and actions). However, we shall allow the policy to be "randomized" in that its instructions may be to choose actions according to a probability distribution. In other words, a policy $\boldsymbol{\beta}$ is a set of numbers $\boldsymbol{\beta} = \{\beta_i(a), a \in A, i = 1, \ldots, M\}$ with the interpretation that if the process is in state i, then action a is to be chosen with probability $\beta_i(a)$. Of course, we need have

$$0 \leqslant \beta_i(a) \leqslant 1, \quad \text{for all } i, a$$
$$\sum_a \beta_i(a) = 1, \quad \text{for all } i$$

Under any given policy $\boldsymbol{\beta}$, the sequence of states $\{X_n,\ n = 0, 1, \ldots\}$ constitutes a Markov chain with transition probabilities $P_{ij}(\boldsymbol{\beta})$ given by[2]

$$\begin{aligned} P_{ij}(\boldsymbol{\beta}) &= P_{\boldsymbol{\beta}}\{X_{n+1} = j | X_n = i\} \\ &= \sum_a P_{ij}(a)\beta_i(a) \end{aligned}$$

where the last equality follows by conditioning on the action chosen when in state i. Let us suppose that for every choice of a policy $\boldsymbol{\beta}$, the resultant Markov chain $\{X_n,\ n = 0, 1, \ldots\}$ is ergodic.

For any policy $\boldsymbol{\beta}$, let π_{ia} denote the limiting (or steady-state) probability that the process will be in state i and action a will be chosen if policy $\boldsymbol{\beta}$ is employed. That is,

$$\pi_{ia} = \lim_{n\to\infty} P_{\boldsymbol{\beta}}\{X_n = i, a_n = a\}$$

The vector $\pi = (\pi_{ia})$ must satisfy

$$\begin{aligned} &\text{(i)} \quad \pi_{ia} \geqslant 0 \quad \text{for all } i, a, \\ &\text{(ii)} \quad \textstyle\sum_i \sum_a \pi_{ia} = 1, \\ &\text{(iii)} \quad \textstyle\sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a) \quad \text{for all } j \end{aligned} \tag{4.33}$$

Eqs. (i) and (ii) are obvious, and Eq. (iii), which is an analogue of Theorem 4.1, follows as the left-hand side equals the steady-state probability of being in state j and the right-hand side is the same probability computed by conditioning on the state and action chosen one stage earlier.

Thus for any policy $\boldsymbol{\beta}$, there is a vector $\boldsymbol{\pi} = (\pi_{ia})$ that satisfies (i)–(iii) and with the interpretation that π_{ia} is equal to the steady-state probability of being in state i and choosing action a when policy $\boldsymbol{\beta}$ is employed. Moreover, it turns out that the reverse is also true. Namely, for any vector $\boldsymbol{\pi} = (\pi_{ia})$ that satisfies (i)–(iii), there exists a policy $\boldsymbol{\beta}$ such that if $\boldsymbol{\beta}$ is used, then the steady-state probability of being in i and choosing action a equals π_{ia}. To verify this last statement, suppose that $\boldsymbol{\pi} = (\pi_{ia})$ is a vector that satisfies (i)–(iii). Then, let the policy $\boldsymbol{\beta} = (\beta_i(a))$ be

$$\begin{aligned} \beta_i(a) &= P\{\boldsymbol{\beta} \text{ chooses } a | \text{state is } i\} \\ &= \frac{\pi_{ia}}{\sum_a \pi_{ia}} \end{aligned}$$

Now let P_{ia} denote the limiting probability of being in i and choosing a when policy $\boldsymbol{\beta}$ is employed. We need to show that $P_{ia} = \pi_{ia}$. To do so, first note that $\{P_{ia}, i = 1, \ldots, M, a \in A\}$ are the limiting probabilities of the two-dimensional Markov chain $\{(X_n, a_n), n \geqslant 0\}$. Hence, by the fundamental Theorem 4.1, they are the unique solu-

[2] We use the notation $P_{\boldsymbol{\beta}}$ to signify that the probability is conditional on the fact that policy $\boldsymbol{\beta}$ is used.

tion of

$$\begin{array}{ll} (\mathrm{i}') & P_{ia} \geqslant 0, \\ (\mathrm{ii}') & \sum_i \sum_a P_{ia} = 1, \\ (\mathrm{iii}') & P_{ja} = \sum_i \sum_{a'} P_{ia'} P_{ij}(a') \beta_j(a) \end{array}$$

where (iii′) follows since

$$P\{X_{n+1} = j, a_{n+1} = a | X_n = i,\ a_n = a'\} = P_{ij}(a')\beta_j(a)$$

Because

$$\beta_j(a) = \frac{\pi_{ja}}{\sum_a \pi_{ja}}$$

we see that (P_{ia}) is the unique solution of

$$\begin{aligned} P_{ia} &\geqslant 0, \\ \sum_i \sum_a P_{ia} &= 1, \\ P_{ja} &= \sum_i \sum_{a'} P_{ia'} P_{ij}(a') \frac{\pi_{ja}}{\sum_a \pi_{ja}} \end{aligned}$$

Hence, to show that $P_{ia} = \pi_{ia}$, we need show that

$$\begin{aligned} \pi_{ia} &\geqslant 0, \\ \sum_i \sum_a \pi_{ia} &= 1, \\ \pi_{ja} &= \sum_i \sum_{a'} \pi_{ia'} P_{ij}(a') \frac{\pi_{ja}}{\sum_a \pi_{ja}} \end{aligned}$$

The top two equations follow from (i) and (ii) of Eq. (4.33), and the third, which is equivalent to

$$\sum_a \pi_{ja} = \sum_i \sum_{a'} \pi_{ia'} P_{ij}(a')$$

follows from condition (iii) of Eq. (4.33).

Thus we have shown that a vector $\boldsymbol{\pi} = (\pi_{ia})$ will satisfy (i), (ii), and (iii) of Eq. (4.33) if and only if there exists a policy $\boldsymbol{\beta}$ such that π_{ia} is equal to the steady-state probability of being in state i and choosing action a when $\boldsymbol{\beta}$ is used. In fact, the policy $\boldsymbol{\beta}$ is defined by $\beta_i(a) = \pi_{ia} / \sum_a \pi_{ia}$.

The preceding is quite important in the determination of "optimal" policies. For instance, suppose that a reward $R(i, a)$ is earned whenever action a is chosen in state i. Since $R(X_i, a_i)$ would then represent the reward earned at time i, the expected average

reward per unit time under policy $\boldsymbol{\beta}$ can be expressed as

$$\text{expected average reward under } \boldsymbol{\beta} = \lim_{n\to\infty} E_{\boldsymbol{\beta}}\left[\frac{\sum_{i=1}^{n} R(X_i, a_i)}{n}\right]$$

Now, if π_{ia} denotes the steady-state probability of being in state i and choosing action a, it follows that the limiting expected reward at time n equals

$$\lim_{n\to\infty} E[R(X_n, a_n)] = \sum_i \sum_a \pi_{ia} R(i, a)$$

which implies that

$$\text{expected average reward under } \boldsymbol{\beta} = \sum_i \sum_a \pi_{ia} R(i, a)$$

Hence, the problem of determining the policy that maximizes the expected average reward is

$$\begin{aligned}
&\underset{\pi=(\pi_{ia})}{\text{maximize}} \sum_i \sum_a \pi_{ia} R(i, a)\\
&\text{subject to } \pi_{ia} \geqslant 0, \quad \text{for all } i, a,\\
&\sum_i \sum_a \pi_{ia} = 1,\\
&\sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a), \quad \text{for all } j
\end{aligned} \tag{4.34}$$

However, the preceding maximization problem is a special case of what is known as a *linear program* and can be solved by a standard linear programming algorithm known as the *simplex algorithm*.[3] If $\boldsymbol{\pi}^* = (\pi_{ia}^*)$ maximizes the preceding, then the optimal policy will be given by $\boldsymbol{\beta}^*$ where

$$\beta_i^*(a) = \frac{\pi_{ia}^*}{\sum_a \pi_{ia}^*}$$

Remarks. **(i)** It can be shown that there is a π^* maximizing Eq. (4.34) that has the property that for each i, π_{ia}^* is zero for all but one value of a, which implies that the optimal policy is nonrandomized. That is, the action it prescribes when in state i is a deterministic function of i.

(ii) The linear programming formulation also often works when there are restrictions placed on the class of allowable policies. For instance, suppose there is a restriction on the fraction of time the process spends in some state, say, state 1. Specifically, suppose that we are allowed to consider only policies having the

[3] It is called a linear program since the objective function $\sum_i \sum_a R(i, a)\pi_{ia}$ and the constraints are all linear functions of the π_{ia}. For a heuristic analysis of the simplex algorithm, see Section 4.5.2.

property that their use results in the process being in state 1 less than 100α percent of time. To determine the optimal policy subject to this requirement, we add to the linear programming problem the additional constraint

$$\sum_a \pi_{1a} \leqslant \alpha$$

since $\sum_a \pi_{1a}$ represents the proportion of time that the process is in state 1.

4.11 Hidden Markov Chains

Let $\{X_n,\ n = 1, 2, \ldots\}$ be a Markov chain with transition probabilities $P_{i,j}$ and initial state probabilities $p_i = P\{X_1 = i\},\ i \geqslant 0$. Suppose that there is a finite set $\mathscr{S}$ of signals, and that a signal from $\mathscr{S}$ is emitted each time the Markov chain enters a state. Further, suppose that when the Markov chain enters state j then, independently of previous Markov chain states and signals, the signal emitted is s with probability $p(s|j)$, $\sum_{s\in\mathscr{S}} p(s|j) = 1$. That is, if S_n represents the nth signal emitted, then

$$P\{S_1 = s|X_1 = j\} = p(s|j),$$
$$P\{S_n = s|X_1, S_1, \ldots, X_{n-1}, S_{n-1}, X_n = j\} = p(s|j)$$

A model of the preceding type in which the sequence of signals $S_1, S_2, \ldots$ is observed, while the sequence of underlying Markov chain states $X_1, X_2, \ldots$ is unobserved, is called a *hidden Markov chain* model.

Example 4.44. Consider a production process that in each period is either in a good state (state 1) or in a poor state (state 2). If the process is in state 1 during a period then, independent of the past, with probability 0.9 it will be in state 1 during the next period and with probability 0.1 it will be in state 2. Once in state 2, it remains in that state forever. Suppose that a single item is produced each period and that each item produced when the process is in state 1 is of acceptable quality with probability 0.99, while each item produced when the process is in state 2 is of acceptable quality with probability 0.96.

If the status, either acceptable or unacceptable, of each successive item is observed, while the process states are unobservable, then the preceding is a hidden Markov chain model. The signal is the status of the item produced, and has value either a or u, depending on whether the item is acceptable or unacceptable. The signal probabilities are

$$p(u|1) = 0.01, \quad p(a|1) = 0.99,$$
$$p(u|2) = 0.04, \quad p(a|2) = 0.96$$

while the transition probabilities of the underlying Markov chain are

$$P_{1,1} = 0.9 = 1 - P_{1,2}, \quad P_{2,2} = 1$$ ■

Although $\{S_n, n \geqslant 1\}$ is not a Markov chain, it should be noted that, conditional on the current state X_n, the sequence $S_n, X_{n+1}, S_{n+1}, \ldots$ of future signals and states is independent of the sequence $X_1, S_1, \ldots, X_{n-1}, S_{n-1}$ of past states and signals.

Let $\mathbf{S}^n = (S_1, \ldots, S_n)$ be the random vector of the first n signals. For a fixed sequence of signals $s_1, \ldots, s_n$, let $\mathbf{s}_k = (s_1, \ldots, s_k), k \leqslant n$. To begin, let us determine the conditional probability of the Markov chain state at time n given that $\mathbf{S}^n = \mathbf{s}_n$. To obtain this probability, let

$$F_n(j) = P\{\mathbf{S}^n = \mathbf{s}_n, X_n = j\}$$

and note that

$$\begin{aligned} P\{X_n = j|\mathbf{S}^n = \mathbf{s}_n\} &= \frac{P\{\mathbf{S}^n = \mathbf{s}_n, X_n = j\}}{P\{\mathbf{S}^n = \mathbf{s}_n\}} \\ &= \frac{F_n(j)}{\sum_i F_n(i)} \end{aligned}$$

Now,

$$\begin{aligned} F_n(j) &= P\{\mathbf{S}^{n-1} = \mathbf{s}_{n-1}, S_n = s_n, X_n = j\} \\ &= \sum_i P\{\mathbf{S}^{n-1} = \mathbf{s}_{n-1}, X_{n-1} = i, X_n = j, S_n = s_n\} \\ &= \sum_i F_{n-1}(i) P\{X_n = j, S_n = s_n|\mathbf{S}^{n-1} = \mathbf{s}_{n-1}, X_{n-1} = i\} \\ &= \sum_i F_{n-1}(i) P\{X_n = j, S_n = s_n|X_{n-1} = i\} \\ &= \sum_i F_{n-1}(i) P_{i,j}\, p(s_n|j) \\ &= p(s_n|j) \sum_i F_{n-1}(i) P_{i,j} \end{aligned} \tag{4.35}$$

where the preceding used that

$$\begin{aligned} &P\{X_n = j, S_n = s_n|X_{n-1} = i\} \\ &\quad = P\{X_n = j|X_{n-1} = i\} \times P\{S_n = s_n|X_n = j, X_{n-1} = i\} \\ &\quad = P_{i,j}\, P\{S_n = s_n|X_n = j\} \\ &\quad = P_{i,j}\, p(s_n|j) \end{aligned}$$

Starting with

$$F_1(i) = P\{X_1 = i, S_1 = s_1\} = p_i\, p(s_1|i)$$

we can use Eq. (4.35) to recursively determine the functions $F_2(i), F_3(i), \ldots$, up to $F_n(i)$.

Example 4.45. Suppose in Example 4.44 that $P\{X_1 = 1\} = 0.8$. It is given that the successive conditions of the first three items produced are a, u, a.

(a) What is the probability that the process was in its good state when the third item was produced?
(b) What is the probability that X_4 is 1?
(c) What is the probability that the next item produced is acceptable?

Solution: With $\mathbf{s}_3 = (a, u, a)$, we have

$$
\begin{aligned}
&F_1(1) = (0.8)(0.99) = 0.792,\\
&F_1(2) = (0.2)(0.96) = 0.192\\
&F_2(1) = 0.01[0.792(0.9) + 0.192(0)] = 0.007128,\\
&F_2(2) = 0.04[0.792(0.1) + 0.192(1)] = 0.010848\\
&F_3(1) = 0.99[(0.007128)(0.9)] \approx 0.006351,\\
&F_3(2) = 0.96[(0.007128)(0.1) + 0.010848] \approx 0.011098
\end{aligned}
$$

Therefore, the answer to part (a) is

$$
P\{X_3 = 1|\mathbf{s}_3\} \approx \frac{0.006351}{0.006351 + 0.011098} \approx 0.364
$$

To compute $P\{X_4 = 1|\mathbf{s}_3\}$, condition on X_3 to obtain

$$
\begin{aligned}
P\{X_4 = 1|\mathbf{s}_3\} &= P\{X_4 = 1|X_3 = 1, \mathbf{s}_3\}P\{X_3 = 1|\mathbf{s}_3\}\\
&\quad + P\{X_4 = 1|X_3 = 2, \mathbf{s}_3\}P\{X_3 = 2|\mathbf{s}_3\}\\
&= P\{X_4 = 1|X_3 = 1, \mathbf{s}_3\}(0.364)\\
&\quad + P\{X_4 = 1|X_3 = 2, \mathbf{s}_3\}(0.636)\\
&= 0.364P_{1,1} + 0.636P_{2,1}\\
&= 0.3276
\end{aligned}
$$

To compute $P\{S_4 = a|\mathbf{s}_3\}$, condition on X_4 to obtain

$$
\begin{aligned}
P\{S_4 = a|\mathbf{s}_3\} &= P\{S_4 = a|X_4 = 1, \mathbf{s}_3\}P\{X_4 = 1|\mathbf{s}_3\}\\
&\quad + P\{S_4 = a|X_4 = 2, \mathbf{s}_3\}P\{X_4 = 2|\mathbf{s}_3\}\\
&= P\{S_4 = a|X_4 = 1\}(0.3276)\\
&\quad + P\{S_4 = a|X_4 = 2\}(1 - 0.3276)\\
&= (0.99)(0.3276) + (0.96)(0.6724) = 0.9698
\end{aligned}
$$

■

To compute $P\{\mathbf{S}^n = \mathbf{s}_n\}$, use the identity $P\{\mathbf{S}^n = \mathbf{s}_n\} = \sum_i F_n(i)$ along with Eq. (4.35). If there are N states of the Markov chain, this requires computing nN quantities $F_n(i)$, with each computation requiring a summation over N terms. This

can be compared with a computation of $P\{\mathbf{S}^n = \mathbf{s}_n\}$ based on conditioning on the first n states of the Markov chain to obtain

$$\begin{aligned} P\{\mathbf{S}^n = \mathbf{s}_n\} &= \sum_{i_1,\ldots,i_n} P\{\mathbf{S}^n = \mathbf{s}_n | X_1 = i_1, \ldots, X_n = i_n\} P\{X_1 = i_1, \ldots, X_n = i_n\} \\ &= \sum_{i_1,\ldots,i_n} p(s_1|i_1) \cdots p(s_n|i_n) p_{i_1} P_{i_1,i_2} P_{i_2,i_3} \cdots P_{i_{n-1},i_n} \end{aligned}$$

The use of the preceding identity to compute $P\{\mathbf{S}^n = \mathbf{s}_n\}$ would thus require a summation over N^n terms, with each term being a product of $2n$ values, indicating that it is not competitive with the previous approach.

The computation of $P\{\mathbf{S}^n = \mathbf{s}_n\}$ by recursively determining the functions $F_k(i)$ is known as the *forward approach*. There also is a *backward approach*, which is based on the quantities $B_k(i)$, defined by

$$B_k(i) = P\{S_{k+1} = s_{k+1}, \ldots, S_n = s_n | X_k = i\}$$

A recursive formula for $B_k(i)$ can be obtained by conditioning on X_{k+1}.

$$\begin{aligned} B_k(i) &= \sum_j P\{S_{k+1} = s_{k+1}, \ldots, S_n = s_n | X_k = i, X_{k+1} = j\} \\ &\quad \times P\{X_{k+1} = j | X_k = i\} \\ &= \sum_j P\{S_{k+1} = s_{k+1}, \ldots, S_n = s_n | X_{k+1} = j\} P_{i,j} \\ &= \sum_j P\{S_{k+1} = s_{k+1} | X_{k+1} = j\} \\ &\quad \times P\{S_{k+2} = s_{k+2}, \ldots, S_n = s_n | S_{k+1} = s_{k+1}, X_{k+1} = j\} P_{i,j} \\ &= \sum_j p(s_{k+1}|j) P\{S_{k+2} = s_{k+2}, \ldots, S_n = s_n | X_{k+1} = j\} P_{i,j} \\ &= \sum_j p(s_{k+1}|j) B_{k+1}(j) P_{i,j} \end{aligned} \tag{4.36}$$

Starting with

$$\begin{aligned} B_{n-1}(i) &= P\{S_n = s_n | X_{n-1} = i\} \\ &= \sum_j P_{i,j} p(s_n|j) \end{aligned}$$

we would then use Eq. (4.36) to determine the function $B_{n-2}(i)$, then $B_{n-3}(i)$, and so on, down to $B_1(i)$. This would then yield $P\{\mathbf{S}^n = \mathbf{s}_n\}$ via

$$P\{\mathbf{S}^n = \mathbf{s}_n\} = \sum_i P\{S_1 = s_1, \ldots, S_n = s_n | X_1 = i\} p_i$$

$$
\begin{aligned}
&= \sum_i P\{S_1 = s_1 | X_1 = i\} \\
&\quad \times P\{S_2 = s_2, \ldots, S_n = s_n | S_1 = s_1, X_1 = i\} p_i \\
&= \sum_i p(s_1|i) P\{S_2 = s_2, \ldots, S_n = s_n | X_1 = i\} p_i \\
&= \sum_i p(s_1|i) B_1(i) p_i
\end{aligned}
$$

Another approach to obtaining $P\{\mathbf{S}^n = \mathbf{s}_n\}$ is to combine both the forward and backward approaches. Suppose that for some k we have computed both functions $F_k(j)$ and $B_k(j)$. Because

$$
\begin{aligned}
P\{\mathbf{S}^n = \mathbf{s}_n, X_k = j\} &= P\{\mathbf{S}^k = \mathbf{s}_k, X_k = j\} \\
&\quad \times P\{S_{k+1} = s_{k+1}, \ldots, S_n = s_n | \mathbf{S}^k = \mathbf{s}_k, X_k = j\} \\
&= P\{\mathbf{S}^k = \mathbf{s}_k, X_k = j\} P\{S_{k+1} = s_{k+1}, \ldots, S_n = s_n | X_k = j\} \\
&= F_k(j) B_k(j)
\end{aligned}
$$

we see that

$$
P\{\mathbf{S}^n = \mathbf{s}_n\} = \sum_j F_k(j) B_k(j)
$$

The beauty of using the preceding identity to determine $P\{\mathbf{S}^n = \mathbf{s}_n\}$ is that we may simultaneously compute the sequence of forward functions, starting with F_1, as well as the sequence of backward functions, starting at B_{n-1}. The parallel computations can then be stopped once we have computed both F_k and B_k for some k.

4.11.1 Predicting the States

Suppose the first n observed signals are $\mathbf{s}_n = (s_1, \ldots, s_n)$, and that given this data we want to predict the first n states of the Markov chain. The best predictor depends on what we are trying to accomplish. If our objective is to maximize the expected number of states that are correctly predicted, then for each $k = 1, \ldots, n$ we need to compute $P\{X_k = j | \mathbf{S}^n = \mathbf{s}_n\}$ and then let the value of j that maximizes this quantity be the predictor of X_k. (That is, we take the mode of the conditional probability mass function of X_k, given the sequence of signals, as the predictor of X_k.) To do so, we must first compute this conditional probability mass function, which is accomplished as follows. For $k \leqslant n$,

$$
\begin{aligned}
P\{X_k = j | \mathbf{S}^n = \mathbf{s}_n\} &= \frac{P\{\mathbf{S}^n = \mathbf{s}_n, X_k = j\}}{P\{\mathbf{S}^n = \mathbf{s}_n\}} \\
&= \frac{F_k(j) B_k(j)}{\sum_j F_k(j) B_k(j)}
\end{aligned}
$$

Thus, given that $\mathbf{S}^n = \mathbf{s}_n$, the optimal predictor of X_k is the value of j that maximizes $F_k(j)B_k(j)$.

A different variant of the prediction problem arises when we regard the sequence of states as a single entity. In this situation, our objective is to choose that sequence of states whose conditional probability, given the sequence of signals, is maximal. For instance, in signal processing, while $X_1, \ldots, X_n$ might be the actual message sent, $S_1, \ldots, S_n$ would be what is received, and so the objective would be to predict the actual message in its entirety.

Letting $\mathbf{X}_k = (X_1, \ldots, X_k)$ be the vector of the first k states, the problem of interest is to find the sequence of states $i_1, \ldots, i_n$ that maximizes $P\{\mathbf{X}_n = (i_1, \ldots, i_n) | \mathbf{S}^n = \mathbf{s}_n\}$. Because

$$P\{\mathbf{X}_n = (i_1, \ldots, i_n) | \mathbf{S}^n = \mathbf{s}_n\} = \frac{P\{\mathbf{X}_n = (i_1, \ldots, i_n), \mathbf{S}^n = \mathbf{s}_n\}}{P\{\mathbf{S}^n = \mathbf{s}_n\}}$$

this is equivalent to finding the sequence of states $i_1, \ldots, i_n$ that maximizes $P\{\mathbf{X}_n = (i_1, \ldots, i_n),\ \mathbf{S}^n = \mathbf{s}_n\}$.

To solve the preceding problem let, for $k \leqslant n$,

$$V_k(j) = \max_{i_1, \ldots, i_{k-1}} P\{\mathbf{X}_{k-1} = (i_1, \ldots, i_{k-1}), X_k = j, \mathbf{S}^k = \mathbf{s}_k\}$$

To recursively solve for $V_k(j)$, use that

$$\begin{aligned}
V_k(j) &= \max_i \max_{i_1, \ldots, i_{k-2}} P\{\mathbf{X}_{k-2} = (i_1, \ldots, i_{k-2}), X_{k-1} = i, X_k = j, \mathbf{S}^k = \mathbf{s}_k\} \\
&= \max_i \max_{i_1, \ldots, i_{k-2}} P\{\mathbf{X}_{k-2} = (i_1, \ldots, i_{k-2}), X_{k-1} = i, \mathbf{S}^{k-1} = \mathbf{s}_{k-1}, \\
&\qquad\qquad X_k = j, S_k = s_k\} \\
&= \max_i \max_{i_1, \ldots, i_{k-2}} P\{\mathbf{X}_{k-2} = (i_1, \ldots, i_{k-2}), X_{k-1} = i, \mathbf{S}^{k-1} = \mathbf{s}_{k-1}\} \\
&\quad \times P\{X_k = j, S_k = s_k | \mathbf{X}_{k-2} = (i_1, \ldots, i_{k-2}), X_{k-1} = i, \mathbf{S}^{k-1} = \mathbf{s}_{k-1}\} \\
&= \max_i \max_{i_1, \ldots, i_{k-2}} P\{\mathbf{X}_{k-2} = (i_1, \ldots, i_{k-2}), X_{k-1} = i, \mathbf{S}^{k-1} = \mathbf{s}_{k-1}\} \\
&\quad \times P\{X_k = j, S_k = s_k | X_{k-1} = i\} \\
&= \max_i P\{X_k = j, S_k = s_k | X_{k-1} = i\} \\
&\quad \times \max_{i_1, \ldots, i_{k-2}} P\{\mathbf{X}_{k-2} = (i_1, \ldots, i_{k-2}), X_{k-1} = i, \mathbf{S}^{k-1} = \mathbf{s}_{k-1}\} \\
&= \max_i P_{i,j}\, p(s_k | j) V_{k-1}(i) \\
&= p(s_k | j) \max_i P_{i,j} V_{k-1}(i) \qquad (4.37)
\end{aligned}$$

Starting with

$$V_1(j) = P\{X_1 = j, S_1 = s_1\} = p_j\, p(s_1 | j)$$

we now use the recursive identity (4.37) to determine $V_2(j)$ for each j; then $V_3(j)$ for each j; and so on, up to $V_n(j)$ for each j.

To obtain the maximizing sequence of states, we work in the reverse direction. Let j_n be the value (or any of the values if there are more than one) of j that maximizes $V_n(j)$. Thus j_n is the final state of a maximizing state sequence. Also, for $k < n$, let $i_k(j)$ be a value of i that maximizes $P_{i,j}V_k(i)$. Then

$$\begin{aligned}
&\max_{i_1,\ldots,i_n} P\{\mathbf{X}_n = (i_1,\ldots,i_n), \mathbf{S}^n = \mathbf{s}_n\} \\
&\quad = \max_j V_n(j) \\
&\quad = V_n(j_n) \\
&\quad = \max_{i_1,\ldots,i_{n-1}} P\{\mathbf{X}_n = (i_1,\ldots,i_{n-1}, j_n), \mathbf{S}^n = \mathbf{s}_n\} \\
&\quad = p(s_n|j_n) \max_i P_{i,j_n} V_{n-1}(i) \\
&\quad = p(s_n|j_n) P_{i_{n-1}(j_n),j_n} V_{n-1}(i_{n-1}(j_n))
\end{aligned}$$

Thus, $i_{n-1}(j_n)$ is the next to last state of the maximizing sequence. Continuing in this manner, the second from the last state of the maximizing sequence is $i_{n-2}(i_{n-1}(j_n))$, and so on.

The preceding approach to finding the most likely sequence of states given a prescribed sequence of signals is known as the *Viterbi Algorithm*.

Exercises

***1.** Three white and three black balls are distributed in two urns in such a way that each contains three balls. We say that the system is in state $i, i = 0, 1, 2, 3$, if the first urn contains i white balls. At each step, we draw one ball from each urn and place the ball drawn from the first urn into the second, and conversely with the ball from the second urn. Let X_n denote the state of the system after the nth step. Explain why $\{X_n, n = 0, 1, 2, \ldots\}$ is a Markov chain and calculate its transition probability matrix.

2. Each individual in a population independently has a random number of offspring that is Poisson distributed with mean λ. Those initially present constitute the zeroth generation. Offspring of zeroth generation people constitute the first generation; their offspring constitute the second generation, and so on. If X_n denotes the size of generation n, is $\{X_n, n \geq 0\}$ a Markov chain. If it is, give its transition probabilities $P_{i,j}$; if it is not, explain why it is not.

3. There are k players, with player i having value $v_i > 0$, $i = 1, \ldots, k$. In every period, two of the players play a game, while the other $k - 2$ wait in an ordered line. The loser of a game joins the end of the line, and the winner then plays a new game against the player who is first in line. Whenever i and j play, i wins with probability $\frac{v_i}{v_i + v_j}$.

(a) Define a Markov chain that is useful in analyzing this model.
(b) How many states does the Markov chain have?
(c) Give the transition probabilities of the chain.

4. Let **P** and **Q** be transition probability matrices on states $1, \ldots, m$, with respective transition probabilities $P_{i,j}$ and $Q_{i,j}$. Consider processes $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ defined as follows:

(a) $X_0 = 1$. A coin that comes up heads with probability p is then flipped. If the coin lands heads, the subsequent states $X_1, X_2, \ldots$, are obtained by using the transition probability matrix **P**; if it lands tails, the subsequent states $X_1, X_2, \ldots$, are obtained by using the transition probability matrix **Q**. (In other words, if the coin lands heads (tails) then the sequence of states is a Markov chain with transition probability matrix **P** (**Q**).) Is $\{X_n, n \geq 0\}$ a Markov chain. If it is, give its transition probabilities. If it is not, tell why not.

(b) $Y_0 = 1$. If the current state is i, then the next state is determined by first flipping a coin that comes up heads with probability p. If the coin lands heads then the next state is j with probability $P_{i,j}$; if it lands tails then the next state is j with probability $Q_{i,j}$. Is $\{Y_n, n \geq 0\}$ a Markov chain. If it is, give its transition probabilities. If it is not, tell why not.

5. A Markov chain $\{X_n, n \geqslant 0\}$ with states 0, 1, 2, has the transition probability matrix

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

If $P\{X_0 = 0\} = P\{X_0 = 1\} = \frac{1}{4}$, find $E[X_3]$.

6. Let the transition probability matrix of a two-state Markov chain be given, as in Example 4.2, by

$$\mathbf{P} = \begin{Vmatrix} p & 1-p \\ 1-p & p \end{Vmatrix}$$

Show by mathematical induction that

$$\mathbf{P}^{(n)} = \begin{Vmatrix} \frac{1}{2} + \frac{1}{2}(2p-1)^n & \frac{1}{2} - \frac{1}{2}(2p-1)^n \\ \frac{1}{2} - \frac{1}{2}(2p-1)^n & \frac{1}{2} + \frac{1}{2}(2p-1)^n \end{Vmatrix}$$

7. In Example 4.4 suppose that it has rained neither yesterday nor the day before yesterday. What is the probability that it will rain tomorrow?

8. An urn initially contains 2 balls, one of which is red and the other blue. At each stage a ball is randomly selected. If the selected ball is red, then it is replaced with a red ball with probability .7 or with a blue ball with probability .3; if the selected ball is blue, then it is equally likely to be replaced by either a red or blue ball.

(a) Let X_n equal 1 if the nth ball selected is red, and let it equal 0 otherwise. Is $\{X_n, n \geq 1\}$ a Markov chain? If so, give its transition probability matrix.

(b) Let Y_n denote the number of red balls in the urn immediately before the nth ball is selected. Is $\{Y_n, n \geq 1\}$ a Markov chain? If so, give its transition probability matrix.

(c) Find the probability that the second ball selected is red.

(d) Find the probability that the fourth ball selected is red.

***9.** In a sequence of independent flips of a coin that comes up heads with probability .6, what is the probability that there is a run of three consecutive heads within the first 10 flips?

10. In Example 4.3, Gary is currently in a cheerful mood. What is the probability that he is not in a glum mood on any of the following three days?

11. In Example 4.13, give the transition probabilities of the Y_n Markov chain in terms of the transition probabilities $P_{i,j}$ of the X_n chain.

12. For a Markov chain $\{X_n, n \geqslant 0\}$ with transition probabilities $P_{i,j}$, consider the conditional probability that $X_n = m$ given that the chain started at time 0 in state i and has not yet entered state r by time n, where r is a specified state not equal to either i or m. We are interested in whether this conditional probability is equal to the n stage transition probability of a Markov chain whose state space does not include state r and whose transition probabilities are

$$Q_{i,j} = \frac{P_{i,j}}{1 - P_{i,r}}, \quad i, j \neq r$$

Either prove the equality

$$P\{X_n = m | X_0 = i, X_k \neq r, k = 1, \ldots, n\} = Q^n_{i,m}$$

or construct a counterexample.

13. Let $\mathbf{P}$ be the transition probability matrix of a Markov chain. Argue that if for some positive integer r, $\mathbf{P}^r$ has all positive entries, then so does $\mathbf{P}^n$, for all integers $n \geqslant r$.

14. Specify the classes of the following Markov chains, and determine whether they are transient or recurrent:

$$\mathbf{P}_1 = \begin{Vmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{Vmatrix}, \qquad \mathbf{P}_2 = \begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{Vmatrix},$$

$$\mathbf{P}_3 = \begin{Vmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{Vmatrix}, \qquad \mathbf{P}_4 = \begin{Vmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 1 & 0 & 0 & 0 & 0 \end{Vmatrix}$$

15. Prove that if the number of states in a Markov chain is M, and if state j can be reached from state i, then it can be reached in M steps or less.

***16.** Show that if state i is recurrent and state i does not communicate with state j, then $P_{ij} = 0$. This implies that once a process enters a recurrent class of states it can never leave that class. For this reason, a recurrent class is often referred to as a *closed* class.

17. For the random walk of Example 4.19 use the strong law of large numbers to give another proof that the Markov chain is transient when $p \neq \frac{1}{2}$.

Hint: Note that the state at time n can be written as $\sum_{i=1}^{n} Y_i$ where the Y_is are independent and $P\{Y_i = 1\} = p = 1 - P\{Y_i = -1\}$. Argue that if $p > \frac{1}{2}$, then, by the strong law of large numbers, $\sum_1^n Y_i \to \infty$ as $n \to \infty$ and hence the initial state 0 can be visited only finitely often, and hence must be transient. A similar argument holds when $p < \frac{1}{2}$.

18. Coin 1 comes up heads with probability 0.6 and coin 2 with probability 0.5. A coin is continually flipped until it comes up tails, at which time that coin is put aside and we start flipping the other one.

(a) What proportion of flips use coin 1?

(b) If we start the process with coin 1 what is the probability that coin 2 is used on the fifth flip?

(c) What proportion of flips land heads?

19. For Example 4.4, calculate the proportion of days that it rains.

20. A transition probability matrix $\mathbf{P}$ is said to be doubly stochastic if the sum over each column equals one; that is,

$$\sum_i P_{ij} = 1, \quad \text{for all } j$$

If such a chain is irreducible and consists of $M + 1$ states $0, 1, \ldots, M$, show that the long-run proportions are given by

$$\pi_j = \frac{1}{M+1}, \quad j = 0, 1, \ldots, M$$

***21.** A DNA nucleotide has any of four values. A standard model for a mutational change of the nucleotide at a specific location is a Markov chain model that supposes that in going from period to period the nucleotide does not change with probability $1 - 3\alpha$, and if it does change then it is equally likely to change to any of the other three values, for some $0 < \alpha < \frac{1}{3}$.

(a) Show that $P_{1,1}^n = \frac{1}{4} + \frac{3}{4}(1-4\alpha)^n$.
(b) What is the long-run proportion of time the chain is in each state?

22. Let Y_n be the sum of n independent rolls of a fair die. Find

$$\lim_{n\to\infty} P\{Y_n \text{ is a multiple of 13}\}$$

Hint: Define an appropriate Markov chain and apply the results of Exercise 20.

23. In a good weather year the number of storms is Poisson distributed with mean 1; in a bad year it is Poisson distributed with mean 3. Suppose that any year's weather conditions depends on past years only through the previous year's condition. Suppose that a good year is equally likely to be followed by either a good or a bad year, and that a bad year is twice as likely to be followed by a bad year as by a good year. Suppose that last year—call it year 0—was a good year.
(a) Find the expected total number of storms in the next two years (that is, in years 1 and 2).
(b) Find the probability there are no storms in year 3.
(c) Find the long-run average number of storms per year.
(d) Find the proportion of years that have no storms.

24. Consider three urns, one colored red, one white, and one blue. The red urn contains 1 red and 4 blue balls; the white urn contains 3 white balls, 2 red balls, and 2 blue balls; the blue urn contains 4 white balls, 3 red balls, and 2 blue balls. At the initial stage, a ball is randomly selected from the red urn and then returned to that urn. At every subsequent stage, a ball is randomly selected from the urn whose color is the same as that of the ball previously selected and is then returned to that urn. In the long run, what proportion of the selected balls are red? What proportion are white? What proportion are blue?

25. Each morning an individual leaves his house and goes for a run. He is equally likely to leave either from his front or back door. Upon leaving the house, he chooses a pair of running shoes (or goes running barefoot if there are no shoes at the door from which he departed). On his return he is equally likely to enter, and leave his running shoes, either by the front or back door. If he owns a total of k pairs of running shoes, what proportion of the time does he run barefooted?

26. Consider the following approach to shuffling a deck of n cards. Starting with any initial ordering of the cards, one of the numbers $1, 2, \ldots, n$ is randomly chosen in such a manner that each one is equally likely to be selected. If number i is chosen, then we take the card that is in position i and put it on top of the deck—that is, we put that card in position 1. We then repeatedly perform the same operation. Show that, in the limit, the deck is perfectly shuffled in the sense that the resultant ordering is equally likely to be any of the $n!$ possible orderings.

***27.** Each individual in a population of size N is, in each period, either active or inactive. If an individual is active in a period then, independent of all else, that individual will be active in the next period with probability α. Similarly, if an

individual is inactive in a period then, independent of all else, that individual will be inactive in the next period with probability β. Let X_n denote the number of individuals that are active in period n.

(a) Argue that $X_n, n \geqslant 0$ is a Markov chain.
(b) Find $E[X_n|X_0 = i]$.
(c) Derive an expression for its transition probabilities.
(d) Find the long-run proportion of time that exactly j people are active.

Hint for (d): Consider first the case where $N = 1$.

28. Every time that the team wins a game, it wins its next game with probability 0.8; every time it loses a game, it wins its next game with probability 0.3. If the team wins a game, then it has dinner together with probability 0.7, whereas if the team loses then it has dinner together with probability 0.2. What proportion of games result in a team dinner?

29. An organization has N employees where N is a large number. Each employee has one of three possible job classifications and changes classifications (independently) according to a Markov chain with transition probabilities

$$\begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

What percentage of employees are in each classification?

30. Three out of every four trucks on the road are followed by a car, while only one out of every five cars is followed by a truck. What fraction of vehicles on the road are trucks?

31. A certain town never has two sunny days in a row. Each day is classified as being either sunny, cloudy (but dry), or rainy. If it is sunny one day, then it is equally likely to be either cloudy or rainy the next day. If it is rainy or cloudy one day, then there is one chance in two that it will be the same the next day, and if it changes then it is equally likely to be either of the other two possibilities. In the long run, what proportion of days are sunny? What proportion are cloudy?

***32.** Each of two switches is either on or off during a day. On day n, each switch will independently be on with probability

$$[1 + \text{number of on switches during day } n - 1]/4$$

For instance, if both switches are on during day $n - 1$, then each will independently be on during day n with probability $3/4$. What fraction of days are both switches on? What fraction are both off?

33. Two players are playing a sequence of points, which begin when one of the players serves. Suppose that player 1 wins each point she serves with probability p, and wins each point her opponent serves with probability q. Suppose the winner of a point becomes the server of the next point.

(a) Find the proportion of points that are won by player 1.
(b) Find the proportion of time that player 1 is the server.

34. A flea moves around the vertices of a triangle in the following manner: Whenever it is at vertex i it moves to its clockwise neighbor vertex with probability p_i and to the counterclockwise neighbor with probability $q_i = 1 - p_i, i = 1, 2, 3$.

(a) Find the proportion of time that the flea is at each of the vertices.

(b) How often does the flea make a counterclockwise move that is then followed by five consecutive clockwise moves?

35. Consider a Markov chain with states 0, 1, 2, 3, 4. Suppose $P_{0,4} = 1$; and suppose that when the chain is in state $i, i > 0$, the next state is equally likely to be any of the states $0, 1, \ldots, i - 1$. Find the limiting probabilities of this Markov chain.

36. The state of a process changes daily according to a two-state Markov chain. If the process is in state i during one day, then it is in state j the following day with probability $P_{i,j}$, where

$$P_{0,0} = 0.4, \quad P_{0,1} = 0.6, \quad P_{1,0} = 0.2, \quad P_{1,1} = 0.8$$

Every day a message is sent. If the state of the Markov chain that day is i then the message sent is "good" with probability p_i and is "bad" with probability $q_i = 1 - p_i, i = 0, 1$

(a) If the process is in state 0 on Monday, what is the probability that a good message is sent on Tuesday?

(b) If the process is in state 0 on Monday, what is the probability that a good message is sent on Friday?

(c) In the long run, what proportion of messages are good?

(d) Let Y_n equal 1 if a good message is sent on day n and let it equal 2 otherwise. Is $\{Y_n, n \geqslant 1\}$ a Markov chain? If so, give its transition probability matrix. If not, briefly explain why not.

37. Show that the stationary probabilities for the Markov chain having transition probabilities $P_{i,j}$ are also the stationary probabilities for the Markov chain whose transition probabilities $Q_{i,j}$ are given by

$$Q_{i,j} = P_{i,j}^k$$

for any specified positive integer k.

38. Capa plays either one or two chess games every day, with the number of games that she plays on successive days being a Markov chain with transition probabilities

$$P_{1,1} = .2, \quad P_{1,2} = .8 \qquad P_{2,1} = .4, \quad P_{2,2} = .6$$

Capa wins each game with probability p. Suppose she plays two games on Monday.

(a) What is the probability that she wins all the games she plays on Tuesday?

(b) What is the expected number of games that she plays on Wednesday?

(c) In the long run, on what proportion of days does Capa win all her games.

39. Consider the one-dimensional symmetric random walk of Example 4.19, which was shown in that example to be recurrent. Let π_i denote the long-run proportion of time that the chain is in state i.
(a) Argue that $\pi_i = \pi_0$ for all i.
(b) Show that $\sum_i \pi_i \neq 1$.
(c) Conclude that this Markov chain is null recurrent, and thus all $\pi_i = 0$.

40. A particle moves on 12 points situated on a circle. At each step it is equally likely to move one step in the clockwise or in the counterclockwise direction. Find the mean number of steps for the particle to return to its starting position.

***41.** Consider a Markov chain with states equal to the nonnegative integers, and suppose its transition probabilities satisfy $P_{i,j} = 0$ if $j \leq i$. Assume $X_0 = 0$, and let e_j be the probability that the Markov chain is ever in state j. (Note that $e_0 = 1$ because $X_0 = 0$.) Argue that for $j > 0$

$$e_j = \sum_{i=0}^{j-1} e_i P_{i,j}$$

If $P_{i,i+k} = 1/3, k = 1, 2, 3$, find e_i for $i = 1, \ldots, 10$.

42. Let A be a set of states, and let A^c be the remaining states.
(a) What is the interpretation of

$$\sum_{i \in A} \sum_{j \in A^c} \pi_i P_{ij}?$$

(b) What is the interpretation of

$$\sum_{i \in A^c} \sum_{j \in A} \pi_i P_{ij}?$$

(c) Explain the identity

$$\sum_{i \in A} \sum_{j \in A^c} \pi_i P_{ij} = \sum_{i \in A^c} \sum_{j \in A} \pi_i P_{ij}$$

43. Each day, one of n possible elements is requested, the ith one with probability $P_i, i \geqslant 1, \sum_1^n P_i = 1$. These elements are at all times arranged in an ordered list that is revised as follows: The element selected is moved to the front of the list with the relative positions of all the other elements remaining unchanged. Define the state at any time to be the list ordering at that time and note that there are $n!$ possible states.
(a) Argue that the preceding is a Markov chain.
(b) For any state $i_1, \ldots, i_n$ (which is a permutation of $1, 2, \ldots, n$), let $\pi(i_1, \ldots, i_n)$ denote the limiting probability. In order for the state to be $i_1, \ldots, i_n$, it is necessary for the last request to be for i_1, the last non-i_1 request for i_2, the last non-i_1 or i_2 request for i_3, and so on. Hence, it

appears intuitive that

$$\pi(i_1, \ldots, i_n) = P_{i_1} \frac{P_{i_2}}{1 - P_{i_1}} \frac{P_{i_3}}{1 - P_{i_1} - P_{i_2}} \cdots \frac{P_{i_{n-1}}}{1 - P_{i_1} - \cdots - P_{i_{n-2}}}$$

Verify when $n = 3$ that the preceding are indeed the limiting probabilities.

44. Suppose that a population consists of a fixed number, say, m, of genes in any generation. Each gene is one of two possible genetic types. If exactly i (of the m) genes of any generation are of type 1, then the next generation will have j type 1 (and $m - j$ type 2) genes with probability

$$\binom{m}{j} \left(\frac{i}{m}\right)^j \left(\frac{m-i}{m}\right)^{m-j}, \quad j = 0, 1, \ldots, m$$

Let X_n denote the number of type 1 genes in the nth generation, and assume that $X_0 = i$.

(a) Find $E[X_n]$.

(b) What is the probability that eventually all the genes will be type 1?

45. Consider an irreducible finite Markov chain with states $0, 1, \ldots, N$.

(a) Starting in state i, what is the probability the process will ever visit state j? Explain!

(b) Let $x_i = P\{\text{visit state } N \text{ before state } 0 | \text{start in } i\}$. Compute a set of linear equations that the x_i satisfy, $i = 0, 1, \ldots, N$.

(c) If $\sum_j j P_{ij} = i$ for $i = 1, \ldots, N - 1$, show that $x_i = i/N$ is a solution to the equations in part (b).

46. An individual possesses r umbrellas that he employs in going from his home to office, and vice versa. If he is at home (the office) at the beginning (end) of a day and it is raining, then he will take an umbrella with him to the office (home), provided there is one to be taken. If it is not raining, then he never takes an umbrella. Assume that, independent of the past, it rains at the beginning (end) of a day with probability p.

(a) Define a Markov chain with $r + 1$ states, which will help us to determine the proportion of time that our man gets wet. (*Note*: He gets wet if it is raining, and all umbrellas are at his other location.)

(b) Show that the limiting probabilities are given by

$$\pi_i = \begin{cases} \dfrac{q}{r+q}, & \text{if } i = 0 \\ \dfrac{1}{r+q}, & \text{if } i = 1, \ldots, r \end{cases} \qquad \text{where } q = 1 - p$$

(c) What fraction of time does our man get wet?

(d) When $r = 3$, what value of p maximizes the fraction of time he gets wet

***47.** Let $\{X_n, n \geqslant 0\}$ denote an ergodic Markov chain with limiting probabilities π_i. Define the process $\{Y_n, n \geqslant 1\}$ by $Y_n = (X_{n-1}, X_n)$. That is, Y_n keeps track of

the last two states of the original chain. Is $\{Y_n, n \geqslant 1\}$ a Markov chain? If so, determine its transition probabilities and find

$$\lim_{n\to\infty} P\{Y_n = (i, j)\}$$

48. Consider a Markov chain in steady state. Say that a k length run of zeroes ends at time m if

$$X_{m-k-1} \neq 0, \quad X_{m-k} = X_{m-k+1} = \ldots = X_{m-1} = 0, X_m \neq 0$$

Show that the probability of this event is $\pi_0(P_{0,0})^{k-1}(1 - P_{0,0})^2$, where π_0 is the limiting probability of state 0.

49. Consider a Markov chain with states 1, 2, 3 having transition probability matrix

$$\begin{pmatrix} .5 & .3 & .2 \\ 0 & .4 & .6 \\ .8 & 0 & .2 \end{pmatrix}$$

(a) If the chain is currently in state 1, find the probability that after two transitions it will be in state 2.

(b) Suppose you receive a reward $r(i) = i^2$ whenever the Markov chain is in state i, $i = 1, 2, 3$. Find your long run average reward per unit time.

Let N_i denote the number of transitions, starting in state i, until the Markov chain enters state 3.

(c) Find $E[N_1]$.

(d) Find $P(N_1 \leq 4)$.

(e) Find $P(N_1 = 4)$.

50. A Markov chain with states $1, \ldots, 6$ has transition probability matrix

$$\begin{pmatrix} .2 & .4 & 0 & .3 & 0 & .1 \\ .1 & .3 & 0 & .4 & 0 & .2 \\ 0 & 0 & .3 & .7 & 0 & 0 \\ 0 & 0 & .6 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & 0 & 0 & .2 & .8 \end{pmatrix}$$

(a) Give the classes and tell which are recurrent and which are transient.

(b) Find $\lim_{n\to\infty} P^n_{1,2}$.

(c) Find $\lim_{n\to\infty} P^n_{5,6}$.

(d) Find $\lim_{n\to\infty} P^n_{1,3}$.

51. In Example 4.3, Gary is in a cheerful mood today. Find the expected number of days until he has been glum for three consecutive days.

52. A taxi driver provides service in two zones of a city. Fares picked up in zone A will have destinations in zone A with probability 0.6 or in zone B with probability 0.4. Fares picked up in zone B will have destinations in zone A with probability 0.3 or in zone B with probability 0.7. The driver's expected

profit for a trip entirely in zone A is 6; for a trip entirely in zone B is 8; and for a trip that involves both zones is 12. Find the taxi driver's average profit per trip.

53. Find the average premium received per policyholder of the insurance company of Example 4.29 if $\lambda = 1/4$ for one-third of its clients, and $\lambda = 1/2$ for two-thirds of its clients.

54. Consider the Ehrenfest urn model in which M molecules are distributed between two urns, and at each time point one of the molecules is chosen at random and is then removed from its urn and placed in the other one. Let X_n denote the number of molecules in urn 1 after the nth switch and let $\mu_n = E[X_n]$. Show that

(a) $\mu_{n+1} = 1 + (1 - 2/M)\mu_n$.

(b) Use (a) to prove that

$$\mu_n = \frac{M}{2} + \left(\frac{M-2}{M}\right)^n \left(E[X_0] - \frac{M}{2}\right)$$

55. Consider a population of individuals each of whom possesses two genes that can be either type A or type a. Suppose that in outward appearance type A is dominant and type a is recessive. (That is, an individual will have only the outward characteristics of the recessive gene if its pair is aa.) Suppose that the population has stabilized, and the percentages of individuals having respective gene pairs AA, aa, and Aa are p, q, and r. Call an individual dominant or recessive depending on the outward characteristics it exhibits. Let S_{11} denote the probability that an offspring of two dominant parents will be recessive; and let S_{10} denote the probability that the offspring of one dominant and one recessive parent will be recessive. Compute S_{11} and S_{10} to show that $S_{11} = S_{10}^2$. (The quantities S_{10} and S_{11} are known in the genetics literature as *Snyder's ratios*.)

56. Suppose that on each play of the game a gambler either wins 1 with probability p or loses 1 with probability $1 - p$. The gambler continues betting until she or he is either up n or down m. What is the probability that the gambler quits a winner?

57. A particle moves among $n + 1$ vertices that are situated on a circle in the following manner. At each step it moves one step either in the clockwise direction with probability p or the counterclockwise direction with probability $q = 1 - p$. Starting at a specified state, call it state 0, let T be the time of the first return to state 0. Find the probability that all states have been visited by time T.

Hint: Condition on the initial transition and then use results from the gambler's ruin problem.

58. In the gambler's ruin problem of Section 4.5.1, suppose the gambler's fortune is presently i, and suppose that we know that the gambler's fortune will eventually reach N (before it goes to 0). Given this information, show that the

probability he wins the next gamble is

$$\begin{cases} \dfrac{p[1-(q/p)^{i+1}]}{1-(q/p)^i}, & \text{if } p \neq \frac{1}{2} \\ \dfrac{i+1}{2i}, & \text{if } p = \frac{1}{2} \end{cases}$$

Hint: The probability we want is

$$P\{X_{n+1} = i+1 | X_n = i, \lim_{m\to\infty} X_m = N\} = \frac{P\{X_{n+1} = i+1, \lim_m X_m = N | X_n = i\}}{P\{\lim_m X_m = N | X_n = i\}}$$

59. For the gambler's ruin model of Section 4.5.1, let M_i denote the mean number of games that must be played until the gambler either goes broke or reaches a fortune of N, given that he starts with i, $i = 0, 1, \ldots, N$. Show that M_i satisfies

$$M_0 = M_N = 0; \quad M_i = 1 + pM_{i+1} + qM_{i-1}, \quad i = 1, \ldots, N-1$$

Solve these equations to obtain

$$\begin{aligned} M_i &= i(N-i), && \text{if } p = \tfrac{1}{2} \\ &= \frac{i}{q-p} - \frac{N}{q-p}\frac{1-(q/p)^i}{1-(q/p)^N}, && \text{if } p \neq \tfrac{1}{2} \end{aligned}$$

60. The following is the transition probability matrix of a Markov chain with states 1, 2, 3, 4

$$\mathbf{P} = \begin{pmatrix} .4 & .3 & .2 & .1 \\ .2 & .2 & .2 & .4 \\ .25 & .25 & .5 & 0 \\ .2 & .1 & .4 & .3 \end{pmatrix}$$

If $X_0 = 1$

(a) find the probability that state 3 is entered before state 4;

(b) find the mean number of transitions until either state 3 or state 4 is entered.

61. Suppose in the gambler's ruin problem that the probability of winning a bet depends on the gambler's present fortune. Specifically, suppose that α_i is the probability that the gambler wins a bet when his or her fortune is i. Given that the gambler's initial fortune is i, let $P(i)$ denote the probability that the gambler's fortune reaches N before 0.

(a) Derive a formula that relates $P(i)$ to $P(i-1)$ and $P(i+1)$.

(b) Using the same approach as in the gambler's ruin problem, solve the equation of part (a) for $P(i)$.

(c) Suppose that i balls are initially in urn 1 and $N - i$ are in urn 2, and suppose that at each stage one of the N balls is randomly chosen, taken from whichever urn it is in, and placed in the other urn. Find the probability that the first urn becomes empty before the second.

***62.** Consider the particle from Exercise 57. What is the expected number of steps the particle takes to return to the starting position? What is the probability that all other positions are visited before the particle returns to its starting state?

63. For the Markov chain with states 1, 2, 3, 4 whose transition probability matrix $\mathbf{P}$ is as specified below find f_{i3} and s_{i3} for $i = 1, 2, 3$.

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.2 & 0.1 & 0.3 \\ 0.1 & 0.5 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

64. Consider a branching process having $\mu < 1$. Show that if $X_0 = 1$, then the expected number of individuals that ever exist in this population is given by $1/(1 - \mu)$. What if $X_0 = n$?

65. In a branching process having $X_0 = 1$ and $\mu > 1$, prove that π_0 is the *smallest* positive number satisfying Eq. (4.20).

Hint: Let π be any solution of $\pi = \sum_{j=0}^{\infty} \pi^j P_j$. Show by mathematical induction that $\pi \geqslant P\{X_n = 0\}$ for all n, and let $n \to \infty$. In using the induction argue that

$$P\{X_n = 0\} = \sum_{j=0}^{\infty} (P\{X_{n-1} = 0\})^j P_j$$

66. For a branching process, calculate π_0 when

(a) $P_0 = \frac{1}{4}, P_2 = \frac{3}{4}$.

(b) $P_0 = \frac{1}{4}, P_1 = \frac{1}{2}, P_2 = \frac{1}{4}$.

(c) $P_0 = \frac{1}{6}, P_1 = \frac{1}{2}, P_3 = \frac{1}{3}$.

67. At all times, an urn contains N balls—some white balls and some black balls. At each stage, a coin having probability $p, 0 < p < 1$, of landing heads is flipped. If heads appears, then a ball is chosen at random from the urn and is replaced by a white ball; if tails appears, then a ball is chosen from the urn and is replaced by a black ball. Let X_n denote the number of white balls in the urn after the nth stage.

(a) Is $\{X_n, n \geqslant 0\}$ a Markov chain? If so, explain why.

(b) What are its classes? What are their periods? Are they transient or recurrent?

(c) Compute the transition probabilities P_{ij}.

(d) Let $N = 2$. Find the proportion of time in each state.

(e) Based on your answer in part (d) and your intuition, guess the answer for the limiting probability in the general case.

(f) Prove your guess in part (e) either by showing that Theorem (4.1) is satisfied or by using the results of Example 4.37.

(g) If $p = 1$, what is the expected time until there are only white balls in the urn if initially there are i white and $N - i$ black?

***68.** **(a)** Show that the limiting probabilities of the reversed Markov chain are the same as for the forward chain by showing that they satisfy the equations

$$\pi_j = \sum_i \pi_i Q_{ij}$$

(b) Give an intuitive explanation for the result of part (a).

69. M balls are initially distributed among m urns. At each stage one of the balls is selected at random, taken from whichever urn it is in, and then placed, at random, in one of the other $m - 1$ urns. Consider the Markov chain whose state at any time is the vector $(n_1, \ldots, n_m)$ where n_i denotes the number of balls in urn i. Guess at the limiting probabilities for this Markov chain and then verify your guess and show at the same time that the Markov chain is time reversible.

70. A total of m white and m black balls are distributed among two urns, with each urn containing m balls. At each stage, a ball is randomly selected from each urn and the two selected balls are interchanged. Let X_n denote the number of black balls in urn 1 after the nth interchange.

(a) Give the transition probabilities of the Markov chain $X_n, n \geqslant 0$.

(b) Without any computations, what do you think are the limiting probabilities of this chain?

(c) Find the limiting probabilities and show that the stationary chain is time reversible.

71. It follows from Theorem 4.2 that for a time reversible Markov chain

$$P_{ij} P_{jk} P_{ki} = P_{ik} P_{kj} P_{ji}, \quad \text{for all } i, j, k$$

It turns out that if the state space is finite and $P_{ij} > 0$ for all i, j, then the preceding is also a sufficient condition for time reversibility. (That is, in this case, we need only check Eq. (4.26) for paths from i to i that have only two intermediate states.) Prove this.

Hint: Fix i and show that the equations

$$\pi_j P_{jk} = \pi_k P_{kj}$$

are satisfied by $\pi_j = c P_{ij} / P_{ji}$, where c is chosen so that $\sum_j \pi_j = 1$.

72. For a time reversible Markov chain, argue that the rate at which transitions from i to j to k occur must equal the rate at which transitions from k to j to i occur.

73. There are k players, with player i having value $v_i > 0$, $i = 1, \ldots, k$. In every period, two of the players play a game. Whoever wins then plays the next game against a randomly chosen one of the other $k - 1$ players (including the one

who has just lost). Suppose that whenever i and j play, i wins with probability $\frac{v_i}{v_i+v_j}$. Let X_n denote the winner of game n.

(a) Give the transition probabilities of the Markov chain $\{X_n, n \geq 1\}$.
(b) Give the stationarity equations that are uniquely satisfied by the π_j.
(c) Give the time reversibility equations.
(d) Find the proportion of all games that are won by j, $j = 1, \ldots, k$.
(e) Find the proportion of games that involve player j as one of the contestants.

74. A group of n processors is arranged in an ordered list. When a job arrives, the first processor in line attempts it; if it is unsuccessful, then the next in line tries it; if it too is unsuccessful, then the next in line tries it, and so on. When the job is successfully processed or after all processors have been unsuccessful, the job leaves the system. At this point we are allowed to reorder the processors, and a new job appears. Suppose that we use the one-closer reordering rule, which moves the processor that was successful one closer to the front of the line by interchanging its position with the one in front of it. If all processors were unsuccessful (or if the processor in the first position was successful), then the ordering remains the same. Suppose that each time processor i attempts a job then, independently of anything else, it is successful with probability p_i.

(a) Define an appropriate Markov chain to analyze this model.
(b) Show that this Markov chain is time reversible.
(c) Find the long-run probabilities.

75. A Markov chain is said to be a tree process if

(i) $P_{ij} > 0$ whenever $P_{ji} > 0$,
(ii) for every pair of states i and j, $i \neq j$, there is a unique sequence of distinct states $i = i_0, i_1, \ldots, i_{n-1}, i_n = j$ such that

$$P_{i_k, i_{k+1}} > 0, \quad k = 0, 1, \ldots, n-1$$

In other words, a Markov chain is a tree process if for every pair of distinct states i and j there is a unique way for the process to go from i to j without reentering a state (and this path is the reverse of the unique path from j to i). Argue that an ergodic tree process is time reversible.

76. On a chessboard compute the expected number of plays it takes a knight, starting in one of the four corners of the chessboard, to return to its initial position if we assume that at each play it is equally likely to choose any of its legal moves. (No other pieces are on the board.)

Hint: Make use of Example 4.38.

77. In a Markov decision problem, another criterion often used, different than the expected average return per unit time, is that of the expected discounted return. In this criterion we choose a number α, $0 < \alpha < 1$, and try to choose a policy so as to maximize $E[\sum_{i=0}^{\infty} \alpha^i R(X_i, a_i)]$ (that is, rewards at time n are discounted at rate α^n). Suppose that the initial state is chosen according to the probabilities b_i. That is,

$$P\{X_0 = i\} = b_i, \quad i = 1, \ldots, n$$

For a given policy $\boldsymbol{\beta}$ let y_{ja} denote the expected discounted time that the process is in state j and action a is chosen. That is,

$$y_{ja} = E_{\boldsymbol{\beta}}\left[\sum_{n=0}^{\infty} \alpha^n I_{\{X_n=j, a_n=a\}}\right]$$

where for any event A the indicator variable I_A is defined by

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that

$$\sum_a y_{ja} = E\left[\sum_{n=0}^{\infty} \alpha^n I_{\{X_n=j\}}\right]$$

or, in other words, $\sum_a y_{ja}$ is the expected discounted time in state j under $\boldsymbol{\beta}$.

(b) Show that

$$\sum_j \sum_a y_{ja} = \frac{1}{1-\alpha},$$

$$\sum_a y_{ja} = b_j + \alpha \sum_i \sum_a y_{ia} P_{ij}(a)$$

Hint: For the second equation, use the identity

$$I_{\{X_{n+1}=j\}} = \sum_i \sum_a I_{\{X_n=i, a_n=a\}} I_{\{X_{n+1}=j\}}$$

Take expectations of the preceding to obtain

$$E\left[I_{X_{n+1}=j\}}\right] = \sum_i \sum_a E\left[I_{\{X_n=i, a_n=a\}}\right] P_{ij}(a)$$

(c) Let $\{y_{ja}\}$ be a set of numbers satisfying

$$\begin{aligned} &\sum_j \sum_a y_{ja} = \frac{1}{1-\alpha}, \\ &\sum_a y_{ja} = b_j + \alpha \sum_i \sum_a y_{ia} P_{ij}(a) \end{aligned} \tag{4.38}$$

Argue that y_{ja} can be interpreted as the expected discounted time that the process is in state j and action a is chosen when the initial state is chosen

according to the probabilities b_j and the policy $\boldsymbol{\beta}$, given by

$$\beta_i(a) = \frac{y_{ia}}{\sum_a y_{ia}}$$

is employed.

Hint: Derive a set of equations for the expected discounted times when policy $\boldsymbol{\beta}$ is used and show that they are equivalent to Eq. (4.38).

(d) Argue that an optimal policy with respect to the expected discounted return criterion can be obtained by first solving the linear program

$$\begin{aligned} \text{maximize} \quad & \sum_j \sum_a y_{ja} R(j,a), \\ \text{such that} \quad & \sum_j \sum_a y_{ja} = \frac{1}{1-\alpha}, \\ & \sum_a y_{ja} = b_j + \alpha \sum_i \sum_a y_{ia} P_{ij}(a), \\ & y_{ja} \geqslant 0, \quad \text{all } j, a; \end{aligned}$$

and then defining the policy β^* by

$$\beta_i^*(a) = \frac{y_{ia}^*}{\sum_a y_{ia}^*}$$

where the y_{ja}^* are the solutions of the linear program.

78. For the Markov chain of Exercise 5, suppose that $p(s|j)$ is the probability that signal s is emitted when the underlying Markov chain state is j, $j = 0, 1, 2$.

(a) What proportion of emissions are signal s?

(b) What proportion of those times in which signal s is emitted is 0 the underlying state?

79. In Example 4.45, what is the probability that the first 4 items produced are all acceptable?

References

[1] K.L. Chung, Markov Chains with Stationary Transition Probabilities, Springer, Berlin, 1960.

[2] S. Karlin, H. Taylor, A First Course in Stochastic Processes, Second Edition, Academic Press, New York, 1975.

[3] J.G. Kemeny, J.L. Snell, Finite Markov Chains, Van Nostrand Reinhold, Princeton, New Jersey, 1960.

[4] S.M. Ross, Stochastic Processes, Second Edition, John Wiley, New York, 1996.

[5] S. Ross, E. Pekoz, A Second Course in Probability, Probabilitybookstore.com, 2006.

The Exponential Distribution and the Poisson Process

5

5.1 Introduction

In making a mathematical model for a real-world phenomenon it is always necessary to make certain simplifying assumptions so as to render the mathematics tractable. On the other hand, however, we cannot make too many simplifying assumptions, for then our conclusions, obtained from the mathematical model, would not be applicable to the real-world situation. Thus, in short, we must make enough simplifying assumptions to enable us to handle the mathematics but not so many that the mathematical model no longer resembles the real-world phenomenon. One simplifying assumption that is often made is to assume that certain random variables are exponentially distributed. The reason for this is that the exponential distribution is both relatively easy to work with and is often a good approximation to the actual distribution.

The property of the exponential distribution that makes it easy to analyze is that it does not deteriorate with time. By this we mean that if the lifetime of an item is exponentially distributed, then an item that has been in use for ten (or any number of) hours is as good as a new item in regards to the amount of time remaining until the item fails. This will be formally defined in Section 5.2 where it will be shown that the exponential is the only distribution that possesses this property.

In Section 5.3 we shall study counting processes with an emphasis on a kind of counting process known as the Poisson process. Among other things we shall discover about this process is its intimate connection with the exponential distribution.

5.2 The Exponential Distribution

5.2.1 Definition

A continuous random variable X is said to have an *exponential distribution* with parameter λ, $\lambda > 0$, if its probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

or, equivalently, if its cdf is given by

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00010-X

The mean of the exponential distribution, $E[X]$, is given by

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)\,dx \\ &= \int_{0}^{\infty} \lambda x e^{-\lambda x}\,dx \end{aligned}$$

Integrating by parts ($u = x, dv = \lambda e^{-\lambda x} dx$) yields

$$E[X] = -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

The moment generating function $\phi(t)$ of the exponential distribution is given by

$$\begin{aligned} \phi(t) &= E[e^{tX}] \\ &= \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\,dx \\ &= \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda \end{aligned} \tag{5.1}$$

All the moments of X can now be obtained by differentiating Eq. (5.1). For example,

$$\begin{aligned} E[X^2] &= \frac{d^2}{dt^2}\phi(t)\bigg|_{t=0} \\ &= \frac{2\lambda}{(\lambda - t)^3}\bigg|_{t=0} \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Consequently,

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$

Example 5.1 (Exponential Random Variables and Expected Discounted Returns)**.** Suppose that you are receiving rewards at randomly changing rates continuously throughout time. Let $R(x)$ denote the random rate at which you are receiving rewards at time x. For a value $\alpha \geq 0$, called the discount rate, the quantity

$$R = \int_0^{\infty} e^{-\alpha x} R(x)\,dx$$

represents the total discounted reward. (In certain applications, α is a continuously compounded interest rate, and R is the present value of the infinite flow of rewards.)

Whereas

$$E[R] = E\left[\int_0^\infty e^{-\alpha x} R(x)\,dx\right] = \int_0^\infty e^{-\alpha x} E[R(x)]\,dx$$

is the expected total discounted reward, we will show that it is also equal to the expected total reward earned up to an exponentially distributed random time with rate α.

Let T be an exponential random variable with rate α that is independent of all the random variables $R(x)$. We want to argue that

$$\int_0^\infty e^{-\alpha x} E[R(x)]\,dx = E\left[\int_0^T R(x)\,dx\right]$$

To show this define for each $x \geq 0$ a random variable $I(x)$ by

$$I(x) = \begin{cases} 1, & \text{if } x \leq T \\ 0, & \text{if } x > T \end{cases}$$

and note that

$$\int_0^T R(x)\,dx = \int_0^\infty R(x) I(x)\,dx$$

Thus,

$$\begin{aligned} E\left[\int_0^T R(x)\,dx\right] &= E\left[\int_0^\infty R(x) I(x)\,dx\right] \\ &= \int_0^\infty E[R(x) I(x)]\,dx \\ &= \int_0^\infty E[R(x)] E[I(x)]\,dx \qquad \text{by independence} \\ &= \int_0^\infty E[R(x)] P\{T \geq x\}\,dx \\ &= \int_0^\infty e^{-\alpha x} E[R(x)]\,dx \end{aligned}$$

Therefore, the expected total discounted reward is equal to the expected total (undiscounted) reward earned by a random time that is exponentially distributed with a rate equal to the discount factor. ■

5.2.2 Properties of the Exponential Distribution

A random variable X is said to be without memory, or *memoryless*, if

$$P\{X > s + t \mid X > t\} = P\{X > s\} \qquad \text{for all } s, t \geq 0 \tag{5.2}$$

If we think of X as being the lifetime of some instrument, then Eq. (5.2) states that the probability that the instrument lives for at least $s + t$ hours given that it has survived t hours is the same as the initial probability that it lives for at least s hours. In other words, if the instrument is alive at time t, then the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution; that is, the instrument does not remember that it has already been in use for a time t.

The condition in Eq. (5.2) is equivalent to

$$\frac{P\{X > s + t,\ X > t\}}{P\{X > t\}} = P\{X > s\}$$

or

$$P\{X > s + t\} = P\{X > s\}P\{X > t\} \tag{5.3}$$

Since Eq. (5.3) is satisfied when X is exponentially distributed (for $e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t}$), it follows that exponentially distributed random variables are memoryless.

Example 5.2. Suppose that the amount of time one spends in a bank is exponentially distributed with mean ten minutes, that is, $\lambda = \frac{1}{10}$. What is the probability that a customer will spend more than fifteen minutes in the bank? What is the probability that a customer will spend more than fifteen minutes in the bank given that she is still in the bank after ten minutes?

Solution: If X represents the amount of time that the customer spends in the bank, then the first probability is just

$$P\{X > 15\} = e^{-15\lambda} = e^{-3/2} \approx 0.223$$

The second question asks for the probability that a customer who has spent ten minutes in the bank will have to spend at least five more minutes. However, since the exponential distribution does not "remember" that the customer has already spent ten minutes in the bank, this must equal the probability that an entering customer spends at least five minutes in the bank. That is, the desired probability is just

$$P\{X > 5\} = e^{-5\lambda} = e^{-1/2} \approx 0.607$$ ■

Example 5.3. Consider a post office that is run by two clerks. Suppose that when Mr. Smith enters the system he discovers that Mr. Jones is being served by one of the clerks and Mr. Brown by the other. Suppose also that Mr. Smith is told that his service will begin as soon as either Jones or Brown leaves. If the amount of time that a clerk spends with a customer is exponentially distributed with mean $1/\lambda$, what is the probability that, of the three customers, Mr. Smith is the last to leave the post office?

Solution: The answer is obtained by this reasoning: Consider the time at which Mr. Smith first finds a free clerk. At this point either Mr. Jones or Mr. Brown would have just left and the other one would still be in service. However, by the lack of

memory of the exponential, it follows that the amount of time that this other man (either Jones or Brown) would still have to spend in the post office is exponentially distributed with mean $1/\lambda$. That is, it is the same as if he were just starting his service at this point. Hence, by symmetry, the probability that he finishes before Smith must equal $\frac{1}{2}$. ■

Example 5.4. The dollar amount of damage involved in an automobile accident is an exponential random variable with mean 1000. Of this, the insurance company only pays that amount exceeding (the deductible amount of) 400. Find the expected value and the standard deviation of the amount the insurance company pays per accident.

Solution: If X is the dollar amount of damage resulting from an accident, then the amount paid by the insurance company is $(X - 400)^+$, (where a^+ is defined to equal a if $a > 0$ and to equal 0 if $a \leq 0$). Whereas we could certainly determine the expected value and variance of $(X - 400)^+$ from first principles, it is easier to condition on whether X exceeds 400. So, let

$$I = \begin{cases} 1, & \text{if } X > 400 \\ 0, & \text{if } X \leq 400 \end{cases}$$

Let $Y = (X - 400)^+$ be the amount paid. By the lack of memory property of the exponential, it follows that if a damage amount exceeds 400, then the amount by which it exceeds it is exponential with mean 1000. Therefore,

$$\begin{aligned} E[Y|I = 1] &= 1000 \\ E[Y|I = 0] &= 0 \\ \text{Var}(Y|I = 1) &= (1000)^2 \\ \text{Var}(Y|I = 0) &= 0 \end{aligned}$$

which can be conveniently written as

$$E[Y|I] = 10^3 I, \qquad \text{Var}(Y|I) = 10^6 I$$

Because I is a Bernoulli random variable that is equal to 1 with probability $e^{-0.4}$, we obtain

$$E[Y] = E[E[Y|I]] = 10^3 E[I] = 10^3 e^{-0.4} \approx 670.32$$

and, by the conditional variance formula

$$\begin{aligned} \text{Var}(Y) &= E[\text{Var}(Y|I)] + \text{Var}(E[Y|I]) \\ &= 10^6 e^{-0.4} + 10^6 e^{-0.4}(1 - e^{-0.4}) \end{aligned}$$

where the final equality used that the variance of a Bernoulli random variable with parameter p is $p(1 - p)$. Consequently,

$$\sqrt{\text{Var}(Y)} \approx 944.09$$ ■

It turns out that not only is the exponential distribution "memoryless," but it is the unique distribution possessing this property. To see this, suppose that X is memoryless and let $\bar{F}(x) = P\{X > x\}$. Then by Eq. (5.3) it follows that

$$\bar{F}(s+t) = \bar{F}(s)\bar{F}(t)$$

That is, $\bar{F}(x)$ satisfies the functional equation

$$g(s+t) = g(s)g(t)$$

However, it turns out that the only right continuous solution of this functional equation is[1]

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

and since a distribution function is always right continuous we must have

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

or

$$F(x) = P\{X \leq x\} = 1 - e^{-\lambda x}$$

which shows that X is exponentially distributed.

Example 5.5. A store must decide how much of a certain commodity to order so as to meet next month's demand, where that demand is assumed to have an exponential distribution with rate λ. If the commodity costs the store c per pound, and can be sold at a price of $s > c$ per pound, how much should be ordered so as to maximize the store's expected profit? Assume that any inventory left over at the end of the month is worthless and that there is no penalty if the store cannot meet all the demand.

Solution: Let X equal the demand. If the store orders the amount t, then the profit, call it P, is given by

$$P = s\min(X, t) - ct$$

Writing

$$\min(X, t) = X - (X - t)^+$$

[1] This is proven as follows: If $g(s+t) = g(s)g(t)$, then

$$g\left(\frac{2}{n}\right) = g\left(\frac{1}{n} + \frac{1}{n}\right) = g^2\left(\frac{1}{n}\right)$$

and repeating this yields $g(m/n) = g^m(1/n)$. Also,

$$g(1) = g\left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) = g^n\left(\frac{1}{n}\right) \qquad \text{or} \qquad g\left(\frac{1}{n}\right) = (g(1))^{1/n}$$

Hence $g(m/n) = (g(1))^{m/n}$, which implies, since g is right continuous, that $g(x) = (g(1))^x$. Since $g(1) = (g(\frac{1}{2}))^2 \geq 0$ we obtain $g(x) = e^{-\lambda x}$, where $\lambda = -\log(g(1))$.

we obtain, upon conditioning whether $X > t$ and then using the lack of memory property of the exponential, that

$$\begin{aligned} E[(X-t)^+] &= E[(X-t)^+|X>t]P(X>t) \\ &\quad + E[(X-t)^+|X\le t]P(X\le t) \\ &= E[(X-t)^+|X>t]e^{-\lambda t} \\ &= \frac{1}{\lambda}e^{-\lambda t} \end{aligned}$$

where the final equality used the lack of memory property of exponential random variables to conclude that, conditional on X exceeding t, the amount by which it exceeds it is an exponential random variable with rate λ. Hence,

$$E[\min(X,t)] = \frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda t}$$

giving that

$$E[P] = \frac{s}{\lambda} - \frac{s}{\lambda}e^{-\lambda t} - ct$$

Differentiation now yields that the maximal profit is attained when $se^{-\lambda t} - c = 0$; that is, when

$$t = \frac{1}{\lambda}\log(s/c)$$

Now, suppose that all unsold inventory can be returned for the amount $r < \min(s,c)$ per pound; and also that there is a penalty cost p per pound of unmet demand. In this case, using our previously derived expression for $E[P]$, we have

$$E[P] = \frac{s}{\lambda} - \frac{s}{\lambda}e^{-\lambda t} - ct + rE[(t-X)^+] - pE[(X-t)^+]$$

Using that

$$\min(X,t) = t - (t-X)^+$$

we see that

$$E[(t-X)^+] = t - E[\min(X,t)] = t - \frac{1}{\lambda} + \frac{1}{\lambda}e^{-\lambda t}$$

Hence,

$$\begin{aligned} E[P] &= \frac{s}{\lambda} - \frac{s}{\lambda}e^{-\lambda t} - ct + rt - \frac{r}{\lambda} + \frac{r}{\lambda}e^{-\lambda t} - \frac{p}{\lambda}e^{-\lambda t} \\ &= \frac{s-r}{\lambda} + \frac{r-s-p}{\lambda}e^{-\lambda t} - (c-r)t \end{aligned}$$

Calculus now yields that the optimal amount to order is

$$t = \frac{1}{\lambda} \log\left(\frac{s + p - r}{c - r}\right)$$

It is worth noting that the optimal amount to order increases in s, p, and r and decreases in λ and c. (Are these monotonicity properties intuitive?) ■

The memoryless property is further illustrated by the failure rate function (also called the hazard rate function) of the exponential distribution.

Consider a continuous positive random variable X having distribution function F and density f. The *failure* (or *hazard*) *rate* function $r(t)$ is defined by

$$r(t) = \frac{f(t)}{1 - F(t)} \tag{5.4}$$

To interpret $r(t)$, suppose that an item, having lifetime X, has survived for t hours, and we desire the probability that it does not survive for an additional time dt. That is, consider $P\{X \in (t, t + dt) | X > t\}$. Now,

$$\begin{aligned} P\{X \in (t, t + dt) | X > t\} &= \frac{P\{X \in (t, t + dt), X > t\}}{P\{X > t\}} \\ &= \frac{P\{X \in (t, t + dt)\}}{P\{X > t\}} \\ &\approx \frac{f(t)\, dt}{1 - F(t)} = r(t)\, dt \end{aligned}$$

That is, $r(t)$ represents the conditional probability density that a t-year-old item will fail.

Suppose now that the lifetime distribution is exponential. Then, by the memoryless property, it follows that the distribution of remaining life for a t-year-old item is the same as for a new item. Hence, $r(t)$ should be constant. This checks out since

$$\begin{aligned} r(t) &= \frac{f(t)}{1 - F(t)} \\ &= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \end{aligned}$$

Thus, the failure rate function for the exponential distribution is constant. The parameter λ is often referred to as the *rate* of the distribution. (Note that the rate is the reciprocal of the mean, and vice versa.)

It turns out that the failure rate function $r(t)$ uniquely determines the distribution F. To prove this, we note by Eq. (5.4) that

$$r(t) = \frac{\frac{d}{dt} F(t)}{1 - F(t)}$$

Integrating both sides yields

$$\log(1 - F(t)) = -\int_0^t r(t)\,dt + k$$

or

$$1 - F(t) = e^k \exp\left\{-\int_0^t r(t)\,dt\right\}$$

Letting $t = 0$ shows that $k = 0$ and thus

$$F(t) = 1 - \exp\left\{-\int_0^t r(t)\,dt\right\}$$

The preceding identity can also be used to show that exponential random variables are the only ones that are memoryless. Because if X is memoryless, then its failure rate function must be constant. But if $r(t) = c$, then by the preceding equation

$$1 - F(t) = e^{-ct}$$

showing that the random variable is exponential.

Example 5.6. Let $X_1, \ldots, X_n$ be independent exponential random variables with respective rates $\lambda_1, \ldots, \lambda_n$, where $\lambda_i \neq \lambda_j$ when $i \neq j$. Let T be independent of these random variables and suppose that

$$\sum_{j=1}^{n} P_j = 1 \qquad \text{where } P_j = P\{T = j\}$$

The random variable X_T is said to be a *hyperexponential* random variable. To see how such a random variable might originate, imagine that a bin contains n different types of batteries, with a type j battery lasting for an exponential distributed time with rate λ_j, $j = 1, \ldots, n$. Suppose further that P_j is the proportion of batteries in the bin that are type j for each $j = 1, \ldots, n$. If a battery is randomly chosen, in the sense that it is equally likely to be any of the batteries in the bin, then the lifetime of the battery selected will have the hyperexponential distribution specified in the preceding.

To obtain the distribution function F of $X = X_T$, condition on T. This yields

$$\begin{aligned} 1 - F(t) &= P\{X > t\} \\ &= \sum_{i=1}^{n} P\{X > t | T = i\} P\{T = i\} \\ &= \sum_{i=1}^{n} P_i e^{-\lambda_i t} \end{aligned}$$

Differentiation of the preceding yields f, the density function of X.

$$f(t) = \sum_{i=1}^{n} \lambda_i P_i e^{-\lambda_i t}$$

Consequently, the failure rate function of a hyperexponential random variable is

$$r(t) = \frac{\sum_{j=1}^{n} P_j \lambda_j e^{-\lambda_j t}}{\sum_{i=1}^{n} P_i e^{-\lambda_i t}}$$

By noting that

$$\begin{aligned} P\{T = j | X > t\} &= \frac{P\{X > t | T = j\} P\{T = j\}}{P\{X > t\}} \\ &= \frac{P_j e^{-\lambda_j t}}{\sum_{i=1}^{n} P_i e^{-\lambda_i t}} \end{aligned}$$

we see that the failure rate function $r(t)$ can also be written as

$$r(t) = \sum_{j=1}^{n} \lambda_j P\{T = j | X > t\}$$

If $\lambda_1 < \lambda_i$, for all $i > 1$, then

$$\begin{aligned} P\{T = 1 | X > t\} &= \frac{P_1 e^{-\lambda_1 t}}{P_1 e^{-\lambda_1 t} + \sum_{i=2}^{n} P_i e^{-\lambda_i t}} \\ &= \frac{P_1}{P_1 + \sum_{i=2}^{n} P_i e^{-(\lambda_i - \lambda_1) t}} \\ &\to 1 \qquad \text{as } t \to \infty \end{aligned}$$

Similarly, $P\{T = i | X > t\} \to 0$ when $i \neq 1$, thus showing that

$$\lim_{t \to \infty} r(t) = \min_i \lambda_i$$

That is, as a randomly chosen battery ages its failure rate converges to the failure rate of the exponential type having the smallest failure rate, which is intuitive since the longer the battery lasts, the more likely it is a battery type with the smallest failure rate. ■

5.2.3 Further Properties of the Exponential Distribution

Let $X_1, \ldots, X_n$ be independent and identically distributed exponential random variables having mean $1/\lambda$. It follows from the results of Example 2.39 that $X_1 + \cdots + X_n$

has a gamma distribution with parameters n and λ. Let us now give a second verification of this result by using mathematical induction. Because there is nothing to prove when $n = 1$, let us start by assuming that $X_1 + \cdots + X_{n-1}$ has density given by

$$f_{X_1+\cdots+X_{n-1}}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-2}}{(n-2)!}$$

Hence,

$$\begin{aligned} f_{X_1+\cdots+X_{n-1}+X_n}(t) &= \int_0^\infty f_{X_n}(t-s) f_{X_1+\cdots+X_{n-1}}(s)\, ds \\ &= \int_0^t \lambda e^{-\lambda(t-s)} \lambda e^{-\lambda s} \frac{(\lambda s)^{n-2}}{(n-2)!}\, ds \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

Thus, we have proven

Proposition 5.1. *If $X_1, \ldots, X_n$ are independent exponential random variables with common rate λ, then $\sum_{i=1}^n X_i$ is a gamma (n, λ) random variable. That is, its density function is*

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \qquad t > 0$$

Another useful calculation is to determine the probability that one exponential random variable is smaller than another. That is, suppose that X_1 and X_2 are independent exponential random variables with respective means $1/\lambda_1$ and $1/\lambda_2$; what is $P\{X_1 < X_2\}$? This probability is easily calculated by conditioning on X_1:

$$\begin{aligned} P\{X_1 < X_2\} &= \int_0^\infty P\{X_1 < X_2 | X_1 = x\} \lambda_1 e^{-\lambda_1 x}\, dx \\ &= \int_0^\infty P\{x < X_2\} \lambda_1 e^{-\lambda_1 x}\, dx \\ &= \int_0^\infty e^{-\lambda_2 x} \lambda_1 e^{-\lambda_1 x}\, dx \\ &= \int_0^\infty \lambda_1 e^{-(\lambda_1+\lambda_2)x}\, dx \\ &= \frac{\lambda_1}{\lambda_1+\lambda_2} \end{aligned} \tag{5.5}$$

Suppose that $X_1, X_2, \ldots, X_n$ are independent exponential random variables, with X_i having rate μ_i, $i = 1, \ldots, n$. It turns out that the smallest of the X_i is exponential with a rate equal to the sum of the μ_i. This is shown as follows:

$$P\{\text{minimum}(X_1, \ldots, X_n) > x\} = P\{X_i > x \text{ for each } i = 1, \ldots, n\}$$

$$= \prod_{i=1}^{n} P\{X_i > x\} \qquad \text{(by independence)}$$

$$= \prod_{i=1}^{n} e^{-\mu_i x}$$

$$= \exp\left\{-\left(\sum_{i=1}^{n} \mu_i\right) x\right\} \tag{5.6}$$

Example 5.7 (Analyzing Greedy Algorithms for the Assignment Problem)**.** A group of n people is to be assigned to a set of n jobs, with one person assigned to each job. For a given set of n^2 values $C_{ij}, i, j = 1, \ldots, n$, a cost C_{ij} is incurred when person i is assigned to job j. The classical assignment problem is to determine the set of assignments that minimizes the sum of the n costs incurred.

Rather than trying to determine the optimal assignment, let us consider two heuristic algorithms for solving this problem. The first heuristic is as follows. Assign person 1 to the job that results in the least cost. That is, person 1 is assigned to job j_1 where $C(1, j_1) = \text{minimum}_j\, C(1, j)$. Now eliminate that job from consideration and assign person 2 to the job that results in the least cost. That is, person 2 is assigned to job j_2 where $C(2, j_2) = \text{minimum}_{j \neq j_1} C(2, j)$. This procedure is then continued until all n persons are assigned. Since this procedure always selects the best job for the person under consideration, we will call it Greedy Algorithm A.

The second algorithm, which we call Greedy Algorithm B, is a more "global" version of the first greedy algorithm. It considers all n^2 cost values and chooses the pair i_1, j_1 for which $C(i, j)$ is minimal. It then assigns person i_1 to job j_1. It then eliminates all cost values involving either person i_1 or job j_1 (so that $(n-1)^2$ values remain) and continues in the same fashion. That is, at each stage it chooses the person and job that have the smallest cost among all the unassigned people and jobs.

Under the assumption that the C_{ij} constitute a set of n^2 independent exponential random variables each having mean 1, which of the two algorithms results in a smaller expected total cost?

Solution: Suppose first that Greedy Algorithm A is employed. Let C_i denote the cost associated with person $i, i = 1, \ldots, n$. Now C_1 is the minimum of n independent exponentials each having rate 1; so by Eq. (5.6) it will be exponential with rate n. Similarly, C_2 is the minimum of $n-1$ independent exponentials with rate 1, and so is exponential with rate $n-1$. Indeed, by the same reasoning C_i will be exponential with rate $n-i+1,\ i = 1, \ldots, n$. Thus, the expected total cost under Greedy Algorithm A is

$$E_A[\text{total cost}] = E[C_1 + \cdots + C_n]$$

$$= \sum_{i=1}^{n} 1/i$$

Let us now analyze Greedy Algorithm B. Let C_i be the cost of the ith person-job pair assigned by this algorithm. Since C_1 is the minimum of all the n^2 values C_{ij},

it follows from Eq. (5.6) that C_1 is exponential with rate n^2. Now, it follows from the lack of memory property of the exponential that the amounts by which the other C_{ij} exceed C_1 will be independent exponentials with rates 1. As a result, C_2 is equal to C_1 plus the minimum of $(n-1)^2$ independent exponentials with rate 1. Similarly, C_3 is equal to C_2 plus the minimum of $(n-2)^2$ independent exponentials with rate 1, and so on. Therefore, we see that

$$
\begin{aligned}
E[C_1] &= 1/n^2, \\
E[C_2] &= E[C_1] + 1/(n-1)^2, \\
E[C_3] &= E[C_2] + 1/(n-2)^2, \\
&\vdots \\
E[C_j] &= E[C_{j-1}] + 1/(n-j+1)^2, \\
&\vdots \\
E[C_n] &= E[C_{n-1}] + 1
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E[C_1] &= 1/n^2, \\
E[C_2] &= 1/n^2 + 1/(n-1)^2, \\
E[C_3] &= 1/n^2 + 1/(n-1)^2 + 1/(n-2)^2, \\
&\vdots \\
E[C_n] &= 1/n^2 + 1/(n-1)^2 + 1/(n-2)^2 + \cdots + 1
\end{aligned}
$$

Adding up all the $E[C_i]$ yields

$$
\begin{aligned}
E_B[\text{total cost}] &= n/n^2 + (n-1)/(n-1)^2 + (n-2)/(n-2)^2 + \cdots + 1 \\
&= \sum_{i=1}^{n} \frac{1}{i}
\end{aligned}
$$

The expected cost is thus the same for both greedy algorithms. ■

Let $X_1, \ldots, X_n$ be independent exponential random variables, with respective rates $\lambda_1, \ldots, \lambda_n$. A useful result, generalizing Eq. (5.5), is that X_i is the smallest of these with probability $\lambda_i / \sum_j \lambda_j$. This is shown as follows:

$$
\begin{aligned}
P\left\{X_i = \min_j X_j\right\} &= P\left\{X_i < \min_{j \neq i} X_j\right\} \\
&= \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}
\end{aligned}
$$

where the final equality uses Eq. (5.5) along with the fact that $\min_{j\neq i} X_j$ is exponential with rate $\sum_{j\neq i}\lambda_j$.

Another important fact is that $\min_i X_i$ and the rank ordering of the X_i are independent. To see why this is true, consider the conditional probability that $X_{i_1} < X_{i_2} < \cdots < X_{i_n}$ given that the minimal value is greater than t. Because $\min_i X_i > t$ means that all the X_i are greater than t, it follows from the lack of memory property of exponential random variables that their remaining lives beyond t remain independent exponential random variables with their original rates. Consequently,

$$\begin{aligned} P\Big\{X_{i_1} < \cdots < X_{i_n} \Big| \min_i X_i > t\Big\} &= P\Big\{X_{i_1} - t < \cdots < X_{i_n} - t \Big| \min_i X_i > t\Big\} \\ &= P\{X_{i_1} < \cdots < X_{i_n}\} \end{aligned}$$

That is, we have proven the following.

Proposition 5.2. *If $X_1, \ldots, X_n$ are independent exponential random variables with respective rates $\lambda_1, \ldots, \lambda_n$, then $\min_i X_i$ is exponential with rate $\sum_{i=1}^n \lambda_i$. Further, $\min_i X_i$ and the rank order of the variables $X_1, \ldots, X_n$ are independent.*

Example 5.8. Suppose you arrive at a post office having two clerks at a moment when both are busy but there is no one else waiting in line. You will enter service when either clerk becomes free. If service times for clerk i are exponential with rate $\lambda_i, i = 1, 2$, find $E[T]$, where T is the amount of time that you spend in the post office.

Solution: Let R_i denote the remaining service time of the customer with clerk $i, i = 1, 2$, and note, by the lack of memory property of exponentials, that R_1 and R_2 are independent exponential random variables with respective rates λ_1 and λ_2. Conditioning on which of R_1 or R_2 is the smallest yields

$$\begin{aligned} E[T] &= E[T|R_1 < R_2]P\{R_1 < R_2\} + E[T|R_2 \leq R_1]P\{R_2 \leq R_1\} \\ &= E[T|R_1 < R_2]\frac{\lambda_1}{\lambda_1+\lambda_2} + E[T|R_2 \leq R_1]\frac{\lambda_2}{\lambda_1+\lambda_2} \end{aligned}$$

Now, with S denoting your service time

$$\begin{aligned} E[T|R_1 < R_2] &= E[R_1 + S|R_1 < R_2] \\ &= E[R_1|R_1 < R_2] + E[S|R_1 < R_2] \\ &= E[R_1|R_1 < R_2] + \frac{1}{\lambda_1} \\ &= \frac{1}{\lambda_1+\lambda_2} + \frac{1}{\lambda_1} \end{aligned}$$

The final equation used that conditional on $R_1 < R_2$ the random variable R_1 is the minimum of R_1 and R_2 and is thus exponential with rate $\lambda_1 + \lambda_2$; and also that conditional on $R_1 < R_2$ you are served by server 1.

As we can show in a similar fashion that

$$E[T|R_2 \le R_1] = \frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_2}$$

we obtain the result

$$E[T] = \frac{3}{\lambda_1 + \lambda_2}$$

Another way to obtain $E[T]$ is to write T as a sum, take expectations, and then condition where needed. This approach yields

$$\begin{aligned} E[T] &= E[\min(R_1, R_2) + S] \\ &= E[\min(R_1, R_2)] + E[S] \\ &= \frac{1}{\lambda_1 + \lambda_2} + E[S] \end{aligned}$$

To compute $E[S]$, we condition on which of R_1 and R_2 is smallest.

$$\begin{aligned} E[S] &= E[S|R_1 < R_2]\frac{\lambda_1}{\lambda_1 + \lambda_2} + E[S|R_2 \le R_1]\frac{\lambda_2}{\lambda_1 + \lambda_2} \\ &= \frac{2}{\lambda_1 + \lambda_2} \end{aligned}$$

■

Example 5.9. There are n cells in the body, of which cells $1, \ldots, k$ are target cells. Associated with each cell is a weight, with w_i being the weight associated with cell $i, i = 1, \ldots, n$. The cells are destroyed one at a time in a random order, which is such that if S is the current set of surviving cells then, independent of the order in which the cells not in S have been destroyed, the next cell killed is $i, i \in S$, with probability $w_i / \sum_{j \in S} w_j$. In other words, the probability that a given surviving cell is the next one to be killed is the weight of that cell divided by the sum of the weights of all still surviving cells. Let A denote the total number of cells that are still alive at the moment when all the cells $1, 2, \ldots, k$ have been killed, and find $E[A]$.

Solution: Although it would be quite difficult to solve this problem by a direct combinatorial argument, a nice solution can be obtained by relating the order in which cells are killed to a ranking of independent exponential random variables. To do so, let $X_1, \ldots, X_n$ be independent exponential random variables, with X_i having rate $w_i, i = 1, \ldots, n$. Note that X_i will be the smallest of these exponentials with probability $w_i / \sum_j w_j$; further, given that X_i is the smallest, X_r will be the next smallest with probability $w_r / \sum_{j \ne i} w_j$; further, given that X_i and X_r are, respectively, the first and second smallest, X_s, $s \ne i, r$, will be the third smallest with probability $w_s / \sum_{j \ne i,r} w_j$; and so on. Consequently, if we let I_j be the index of the jth smallest of $X_1, \ldots, X_n$—so that $X_{I_1} < X_{I_2} < \cdots < X_{I_n}$—then the order in which the cells are destroyed has the same distribution as $I_1, \ldots, I_n$. So, let us suppose that the order in which the cells are killed is determined by the ordering of

$X_1, \ldots, X_n$. (Equivalently, we can suppose that all cells will eventually be killed, with cell i being killed at time $X_i, i = 1, \ldots, n$.)
If we let A_j equal 1 if cell j is still alive at the moment when all the cells $1, \ldots, k$ have been killed, and let it equal 0 otherwise, then

$$A = \sum_{j=k+1}^{n} A_j$$

Because cell j will be alive at the moment when all the cells $1, \ldots, k$ have been killed if X_j is larger than all the values $X_1, \ldots, X_k$, we see that for $j > k$

$$\begin{aligned}
E[A_j] &= P\{A_j = 1\} \\
&= P\{X_j > \max_{i=1,\ldots,k} X_i\} \\
&= \int_0^\infty P\left\{X_j > \max_{i=1,\ldots,k} X_i | X_j = x\right\} w_j e^{-w_j x}\, dx \\
&= \int_0^\infty P\{X_i < x \text{ for all } i = 1, \ldots, k\} w_j e^{-w_j x}\, dx \\
&= \int_0^\infty \prod_{i=1}^{k} (1 - e^{-w_i x})\, w_j e^{-w_j x}\, dx \\
&= \int_0^1 \prod_{i=1}^{k} (1 - y^{w_i/w_j})\, dy
\end{aligned}$$

where the final equality follows from the substitution $y = e^{-w_j x}$. Thus, we obtain the result

$$E[A] = \sum_{j=k+1}^{n} \int_0^1 \prod_{i=1}^{k} (1 - y^{w_i/w_j})\, dy = \int_0^1 \sum_{j=k+1}^{n} \prod_{i=1}^{k} (1 - y^{w_i/w_j})\, dy \quad \blacksquare$$

Example 5.10. Suppose that customers are in line to receive service that is provided sequentially by a server; whenever a service is completed, the next person in line enters the service facility. However, each waiting customer will only wait an exponentially distributed time with rate θ; if its service has not yet begun by this time then it will immediately depart the system. These exponential times, one for each waiting customer, are independent. In addition, the service times are independent exponential random variables with rate μ. Suppose that someone is presently being served and consider the person who is nth in line.

(a) Find P_n, the probability that this customer is eventually served.
(b) Find W_n, the conditional expected amount of time this person spends waiting in line given that she is eventually served.

Solution: Consider the $n + 1$ random variables consisting of the remaining service time of the person in service along with the n additional exponential departure times with rate θ of the first n in line.

(a) Given that the smallest of these $n+1$ independent exponentials is the departure time of the nth person in line, the conditional probability that this person will be served is 0; on the other hand, given that this person's departure time is not the smallest, the conditional probability that this person will be served is the same as if it were initially in position $n-1$. Since the probability that a given departure time is the smallest of the $n+1$ exponentials is $\theta/(n\theta+\mu)$, we obtain

$$P_n = \frac{(n-1)\theta+\mu}{n\theta+\mu} P_{n-1}$$

Using the preceding with $n-1$ replacing n gives

$$P_n = \frac{(n-1)\theta+\mu}{n\theta+\mu} \frac{(n-2)\theta+\mu}{(n-1)\theta+\mu} P_{n-2} = \frac{(n-2)\theta+\mu}{n\theta+\mu} P_{n-2}$$

Continuing in this fashion yields the result

$$P_n = \frac{\theta+\mu}{n\theta+\mu} P_1 = \frac{\mu}{n\theta+\mu}$$

(b) To determine an expression for W_n, we use the fact that the minimum of independent exponentials is, independent of their rank ordering, exponential with a rate equal to the sum of the rates. Since the time until the nth person in line enters service is the minimum of these $n+1$ random variables plus the additional time thereafter, we see, upon using the lack of memory property of exponential random variables, that

$$W_n = \frac{1}{n\theta+\mu} + W_{n-1}$$

Repeating the preceding argument with successively smaller values of n yields the solution

$$W_n = \sum_{i=1}^{n} \frac{1}{i\theta+\mu}$$

■

5.2.4 Convolutions of Exponential Random Variables

Let $X_i, i = 1, \ldots, n$, be independent exponential random variables with respective rates $\lambda_i, i = 1, \ldots, n$, and suppose that $\lambda_i \neq \lambda_j$ for $i \neq j$. The random variable $\sum_{i=1}^{n} X_i$ is said to be a *hypoexponential* random variable. To compute its probability density function, let us start with the case $n = 2$. Now,

$$\begin{aligned} f_{X_1+X_2}(t) &= \int_0^t f_{X_1}(s) f_{X_2}(t-s)\, ds \\ &= \int_0^t \lambda_1 e^{-\lambda_1 s} \lambda_2 e^{-\lambda_2 (t-s)}\, ds \end{aligned}$$

$$
\begin{aligned}
&= \lambda_1\lambda_2 e^{-\lambda_2 t}\int_0^t e^{-(\lambda_1-\lambda_2)s}\,ds\\
&= \frac{\lambda_1}{\lambda_1-\lambda_2}\lambda_2 e^{-\lambda_2 t}(1-e^{-(\lambda_1-\lambda_2)t})\\
&= \frac{\lambda_1}{\lambda_1-\lambda_2}\lambda_2 e^{-\lambda_2 t}+\frac{\lambda_2}{\lambda_2-\lambda_1}\lambda_1 e^{-\lambda_1 t}
\end{aligned}
$$

Using the preceding, a similar computation yields, when $n = 3$,

$$
f_{X_1+X_2+X_3}(t) = \sum_{i=1}^{3}\lambda_i e^{-\lambda_i t}\Big(\prod_{j\neq i}\frac{\lambda_j}{\lambda_j-\lambda_i}\Big)
$$

which suggests the general result

$$
f_{X_1+\cdots+X_n}(t) = \sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i t}
$$

where

$$
C_{i,n} = \prod_{j\neq i}\frac{\lambda_j}{\lambda_j-\lambda_i}
$$

We will now prove the preceding formula by induction on n. Since we have already established it for $n = 2$, assume it for n and consider $n + 1$ arbitrary independent exponentials X_i with distinct rates λ_i, $i = 1, \ldots, n+1$. If necessary, renumber X_1 and X_{n+1} so that $\lambda_{n+1} < \lambda_1$. Now,

$$
\begin{aligned}
f_{X_1+\cdots+X_{n+1}}(t) &= \int_0^t f_{X_1+\cdots+X_n}(s)\lambda_{n+1}e^{-\lambda_{n+1}(t-s)}\,ds\\
&= \sum_{i=1}^{n} C_{i,n}\int_0^t \lambda_i e^{-\lambda_i s}\lambda_{n+1}e^{-\lambda_{n+1}(t-s)}\,ds\\
&= \sum_{i=1}^{n} C_{i,n}\left(\frac{\lambda_i}{\lambda_i-\lambda_{n+1}}\lambda_{n+1}e^{-\lambda_{n+1}t}+\frac{\lambda_{n+1}}{\lambda_{n+1}-\lambda_i}\lambda_i e^{-\lambda_i t}\right)\\
&= K_{n+1}\lambda_{n+1}e^{-\lambda_{n+1}t}+\sum_{i=1}^{n} C_{i,n+1}\lambda_i e^{-\lambda_i t} \qquad (5.7)
\end{aligned}
$$

where $K_{n+1} = \sum_{i=1}^{n} C_{i,n}\lambda_i/(\lambda_i-\lambda_{n+1})$ is a constant that does not depend on t. But, we also have that

$$
f_{X_1+\cdots+X_{n+1}}(t) = \int_0^t f_{X_2+\cdots+X_{n+1}}(s)\lambda_1 e^{-\lambda_1(t-s)}\,ds
$$

which implies, by the same argument that resulted in Eq. (5.7), that for a constant K_1

$$f_{X_1+\cdots+X_{n+1}}(t) = K_1\lambda_1 e^{-\lambda_1 t} + \sum_{i=2}^{n+1} C_{i,n+1}\lambda_i e^{-\lambda_i t}$$

Equating these two expressions for $f_{X_1+\cdots+X_{n+1}}(t)$ yields

$$K_{n+1}\lambda_{n+1}e^{-\lambda_{n+1}t} + C_{1,n+1}\lambda_1 e^{-\lambda_1 t} = K_1\lambda_1 e^{-\lambda_1 t} + C_{n+1,n+1}\lambda_{n+1}e^{-\lambda_{n+1}t}$$

Multiplying both sides of the preceding equation by $e^{\lambda_{n+1}t}$ and then letting $t \to \infty$ yields [since $e^{-(\lambda_1-\lambda_{n+1})t} \to 0$ as $t \to \infty$]

$$K_{n+1} = C_{n+1,n+1}$$

and this, using Eq. (5.7), completes the induction proof. Thus, we have shown that if $S = \sum_{i=1}^{n} X_i$, then

$$f_S(t) = \sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i t} \tag{5.8}$$

where

$$C_{i,n} = \prod_{j\neq i} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

Integrating both sides of the expression for f_S from t to ∞ yields that the tail distribution function of S is given by

$$P\{S > t\} = \sum_{i=1}^{n} C_{i,n}e^{-\lambda_i t} \tag{5.9}$$

Hence, we obtain from Eqs. (5.8) and (5.9) that $r_S(t)$, the failure rate function of S, is as follows:

$$r_S(t) = \frac{\sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i t}}{\sum_{i=1}^{n} C_{i,n}e^{-\lambda_i t}}$$

If we let $\lambda_j = \min(\lambda_1, \ldots, \lambda_n)$, then it follows, upon multiplying the numerator and denominator of $r_S(t)$ by $e^{\lambda_j t}$, that

$$\lim_{t\to\infty} r_S(t) = \lambda_j$$

From the preceding, we can conclude that the remaining lifetime of a hypoexponentially distributed item that has survived to age t is, for t large, approximately that of an exponentially distributed random variable with a rate equal to the minimum of the rates of the random variables whose sums make up the hypoexponential.

Remark. Although

$$1 = \int_0^\infty f_S(t)\,dt = \sum_{i=1}^{n} C_{i,n} = \sum_{i=1}^{n} \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

it should not be thought that the $C_{i,n}, i = 1, \ldots, n$ are probabilities, because some of them will be negative. Thus, while the form of the hypoexponential density is similar to that of the hyperexponential density (see Example 5.6) these two random variables are very different.

Example 5.11. Let $X_1, \ldots, X_m$ be independent exponential random variables with respective rates $\lambda_1, \ldots, \lambda_m$, where $\lambda_i \neq \lambda_j$ when $i \neq j$. Let N be independent of these random variables and suppose that $\sum_{n=1}^{m} P_n = 1$, where $P_n = P\{N = n\}$. The random variable

$$Y = \sum_{j=1}^{N} X_j$$

is said to be a *Coxian* random variable. Conditioning on N gives its density function:

$$\begin{aligned} f_Y(t) &= \sum_{n=1}^{m} f_Y(t|N=n)\,P_n \\ &= \sum_{n=1}^{m} f_{X_1+\cdots+X_n}(t|N=n)\,P_n \\ &= \sum_{n=1}^{m} f_{X_1+\cdots+X_n}(t)\,P_n \\ &= \sum_{n=1}^{m} P_n \sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i t} \end{aligned}$$

Let

$$r(n) = P\{N = n | N \geq n\}$$

If we interpret N as a lifetime measured in discrete time periods, then $r(n)$ denotes the probability that an item will die in its nth period of use given that it has survived up to that time. Thus, $r(n)$ is the discrete time analog of the failure rate function $r(t)$, and is correspondingly referred to as the discrete time *failure* (or *hazard*) *rate* function.

Coxian random variables often arise in the following manner. Suppose that an item must go through m stages of treatment to be cured. However, suppose that after each stage there is a probability that the item will quit the program. If we suppose that the amounts of time that it takes the item to pass through the successive stages are independent exponential random variables, and that the probability that an item that

has just completed stage n quits the program is (independent of how long it took to go through the n stages) equal to $r(n)$, then the total time that an item spends in the program is a Coxian random variable. ■

5.2.5 The Dirichlet Distribution

Consider an experiment with possible outcomes $1, 2, \ldots, n$, having respective probabilities $P_1, \ldots, P_n$, $\sum_{i=1}^{n} P_i = 1$, and suppose we want to assume a probability distribution on the vector $(P_1, \ldots, P_n)$. Because $\sum_{i=1}^{n} P_i = 1$, we cannot define a density on $P_1, \ldots, P_n$, but what we can do is to define one on $P_1, \ldots, P_{n-1}$ and then take $P_n = 1 - \sum_{i=1}^{n-1} P_i$. The *Dirichlet distribution* assumes that $(P_1, \ldots, P_{n-1})$ is uniformly distributed over the set $S = \{(p_1, \ldots, p_{n-1}) : \sum_{i=1}^{n} p_i < 1,\ 0 < p_i, i = 1, \ldots, n-1\}$. Thus, the Dirichlet joint density function is

$$f_{P_1,\ldots,P_{n-1}}(p_1, \ldots, p_{n-1}) = C, \quad 0 < p_i,\ i = 1, \ldots, n-1,\ \sum_{i=1}^{n-1} p_i < 1$$

Because integrating the preceding density over the set S yields that

$$1 = C\, P(U_1 + \ldots U_{n-1} < 1)$$

where $U_1, \ldots, U_{n-1}$ are independent uniform (0, 1) random variables, it follows from Example 3.28 that $C = (n-1)!$.

There is a relationship between exponential random variables and the Dirichlet distribution.

Proposition 5.3. *Let $X_1, \ldots, X_n$ be independent exponential random variables with rate λ, and let $S = \sum_{i=1}^{n} X_i$. Then, $(\frac{X_1}{S}, \frac{X_2}{S}, \ldots, \frac{X_{n-1}}{S})$ has a Dirichlet distribution.*

Proof. With $f_{X_1,\ldots,X_{n-1}|S}(x_1, \ldots, x_{n-1}|t)$ being the conditional density of $X_1, \ldots, X_{n-1}$ given that $S = t$, we have that

$$f_{X_1,\ldots,X_{n-1}|S}(x_1, \ldots, x_{n-1}|t) = \frac{f_{X_1,\ldots,X_{n-1},S}(x_1, \ldots, x_{n-1}, t)}{f_S(t)} \tag{5.10}$$

Because $X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, S = t$ is equivalent to $X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i$, Eq. (5.10) gives that for $\sum_{i=1}^{n-1} x_i < t, x_i > 0$,

$$\begin{aligned} f_{X_1,\ldots,X_{n-1}|S}(x_1, \ldots, x_{n-1}|t) &= \frac{f_{X_1,\ldots,X_{n-1},X_n}(x_1, \ldots, x_{n-1}, t - \sum_{i=1}^{n-1} x_i)}{f_S(t)} \\ &= \frac{f_{X_1}(x_1) \cdots f_{X_{n-1}}(x_{n-1}) f_{X_n}(t - \sum_{i=1}^{n-1} x_i)}{f_S(t)} \\ &= \frac{\lambda e^{-\lambda x_1} \cdots \lambda e^{-\lambda x_{n-1}} \lambda e^{-\lambda(t - \sum_{i=1}^{n-1} x_i)}}{\lambda e^{-\lambda t} (\lambda t)^{n-1}/(n-1)!} \end{aligned}$$

$$= \frac{(n-1)!}{t^{n-1}}, \qquad \sum_{i=1}^{n-1} x_i < t$$

where the second equality used independence, and the next one used that S, being the sum of n independent exponential random variables with rate λ, has a gamma distribution with parameters n, λ. If we let $Y_i = X_i/t$, $i = 1, \ldots, n-1$ then, as the Jacobian of this transformation is $1/t^{n-1}$, it follows that

$$\begin{aligned} f_{\frac{X_1}{t},\ldots,\frac{X_{n-1}}{t}|S}(y_1, \ldots, y_{n-1}|t) &= f_{X_1,\ldots,X_{n-1}|S}(ty_1, \ldots, ty_{n-1}|t)\, t^{n-1} \\ &= \frac{(n-1)!}{t^{n-1}}\, t^{n-1}, \qquad \sum_{i=1}^{n-1} ty_i < t \\ &= (n-1)!, \qquad \sum_{i=1}^{n-1} y_i < 1 \end{aligned} \tag{5.11}$$

Because, given that $S = t$, the conditional distributions of $\frac{X_1}{S}, \ldots, \frac{X_{n-1}}{S}$ and of $\frac{X_1}{t}, \ldots, \frac{X_{n-1}}{t}$ are identical, it follows from Eq. (5.11) that

$$f_{\frac{X_1}{S},\ldots,\frac{X_{n-1}}{S}|S}(y_1, \ldots, y_{n-1}|t) = (n-1)!, \quad \sum_{i=1}^{n-1} y_i < 1$$

Because the preceding conditional density of $\frac{X_1}{S}, \ldots, \frac{X_{n-1}}{S}$ given that $S = t$ does not depend on t, it follows that it is also the unconditional density of $\frac{X_1}{S}, \ldots, \frac{X_{n-1}}{S}$. That is,

$$f_{\frac{X_1}{S},\ldots,\frac{X_{n-1}}{S}}(y_1, \ldots, y_{n-1}) = (n-1)!, \quad \sum_{i=1}^{n-1} y_i < 1$$

which shows that $(\frac{X_1}{S}, \frac{X_2}{S}, \ldots, \frac{X_{n-1}}{S})$ has a Dirichlet distribution. ■

5.3 The Poisson Process

5.3.1 Counting Processes

A stochastic process $\{N(t), t \geq 0\}$ is said to be a *counting process* if $N(t)$ represents the total number of "events" that occur by time t. Some examples of counting processes are the following:

(a) If we let $N(t)$ equal the number of persons who enter a particular store at or prior to time t, then $\{N(t), t \geq 0\}$ is a counting process in which an event corresponds to a person entering the store. Note that if we had let $N(t)$ equal the number of

persons in the store at time t, then $\{N(t), t \geq 0\}$ would *not* be a counting process (why not?).

(b) If we say that an event occurs whenever a child is born, then $\{N(t), t \geq 0\}$ is a counting process when $N(t)$ equals the total number of people who were born by time t. (Does $N(t)$ include persons who have died by time t? Explain why it must.)

(c) If $N(t)$ equals the number of goals that a given soccer player scores by time t, then $\{N(t), t \geq 0\}$ is a counting process. An event of this process will occur whenever the soccer player scores a goal.

From its definition we see that for a counting process $N(t)$ must satisfy:

(i) $N(t) \geq 0$.

(ii) $N(t)$ is integer valued.

(iii) If $s < t$, then $N(s) \leq N(t)$.

(iv) For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

A counting process is said to possess *independent increments* if the numbers of events that occur in disjoint time intervals are independent. For example, this means that the number of events that occur by time 10 (that is, $N(10)$) must be independent of the number of events that occur between times 10 and 15 (that is, $N(15) - N(10)$).

The assumption of independent increments might be reasonable for example (a), but it probably would be unreasonable for example (b). The reason for this is that if in example (b) $N(t)$ is very large, then it is probable that there are many people alive at time t; this would lead us to believe that the number of new births between time t and time $t + s$ would also tend to be large (that is, it does not seem reasonable that $N(t)$ is independent of $N(t + s) - N(t)$, and so $\{N(t), t \geq 0\}$ would not have independent increments in example (b)). The assumption of independent increments in example (c) would be justified if we believed that the soccer player's chances of scoring a goal today do not depend on "how he's been going." It would not be justified if we believed in "hot streaks" or "slumps."

A counting process is said to possess *stationary increments* if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. In other words, the process has stationary increments if the number of events in the interval $(s, s + t)$ has the same distribution for all s.

The assumption of stationary increments would only be reasonable in example (a) if there were no times of day at which people were more likely to enter the store. Thus, for instance, if there was a rush hour (say, between 12 P.M. and 1 P.M.) each day, then the stationarity assumption would not be justified. If we believed that the earth's population is basically constant (a belief not held at present by most scientists), then the assumption of stationary increments might be reasonable in example (b). Stationary increments do not seem to be a reasonable assumption in example (c) since, for one thing, most people would agree that the soccer player would probably score more goals while in the age bracket 25–30 than he would while in the age bracket 35–40. It may, however, be reasonable over a smaller time horizon, such as one year.

5.3.2 Definition of the Poisson Process

One of the most important types of counting process is the Poisson process. As a prelude to giving its definition, we define the concept of a function $f(\cdot)$ being $o(h)$.

Definition 5.1. The function $f(\cdot)$ is said to be $o(h)$ if

$$\lim_{h\to 0} \frac{f(h)}{h} = 0$$

Example 5.12. **(a)** The function $f(x) = x^2$ is $o(h)$ since

$$\lim_{h\to 0} \frac{f(h)}{h} = \lim_{h\to 0} \frac{h^2}{h} = \lim_{h\to 0} h = 0$$

(b) The function $f(x) = x$ is not $o(h)$ since

$$\lim_{h\to 0} \frac{f(h)}{h} = \lim_{h\to 0} \frac{h}{h} = \lim_{h\to 0} 1 = 1 \neq 0$$

(c) If $f(\cdot)$ is $o(h)$ and $g(\cdot)$ is $o(h)$, then so is $f(\cdot) + g(\cdot)$. This follows since

$$\lim_{h\to 0} \frac{f(h) + g(h)}{h} = \lim_{h\to 0} \frac{f(h)}{h} + \lim_{h\to 0} \frac{g(h)}{h} = 0 + 0 = 0$$

(d) If $f(\cdot)$ is $o(h)$, then so is $g(\cdot) = cf(\cdot)$. This follows since

$$\lim_{h\to 0} \frac{cf(h)}{h} = c \lim \frac{f(h)}{h} = c \cdot 0 = 0$$

(e) From (c) and (d) it follows that any finite linear combination of functions, each of which is $o(h)$, is $o(h)$. ■

In order for the function $f(\cdot)$ to be $o(h)$ it is necessary that $f(h)/h$ go to zero as h goes to zero. But if h goes to zero, the only way for $f(h)/h$ to go to zero is for $f(h)$ to go to zero faster than h does. That is, for h small, $f(h)$ must be small compared with h.

The $o(h)$ notation can be used to make statements more precise. For instance, if X is continuous with density f and failure rate function $\lambda(t)$, then the approximate statements

$$P(t < X < t + h) \approx f(t)\, h$$
$$P(t < X < t + h | X > t) \approx \lambda(t)\, h$$

can be precisely expressed as

$$P(t < X < t + h) = f(t)\, h + o(h)$$
$$P(t < X < t + h | X > t) = \lambda(t)\, h + o(h)$$

We are now in position to define the Poisson process.

Definition 5.2. The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with rate $\lambda > 0$ if the following axioms hold:

(i) $N(0) = 0$
(ii) $\{N(t), t \geq 0\}$ has independent increments
(iii) $P(N(t+h) - N(t) = 1) = \lambda h + o(h)$
(iv) $P(N(t+h) - N(t) \geq 2) = o(h)$

We start our analysis of the Poisson process by first considering the counting process that results when one starts observing the Poisson process at a given time s.

For $s > 0$, let $N_s(t) = N(s+t) - N(s)$. That is, starting at time s, $N_s(t)$ is the number of events of the Poisson process that occur in the next t time units.

Lemma 5.1. *$\{N_s(t), t \geq 0\}$ is a Poisson process with rate λ.*

Proof. To prove this, we check that $\{N_s(t), t \geq 0\}$ satisfies the axioms of a Poisson process with rate λ. Axiom (i) is immediate, and each of the other axioms hold for $\{N_s(t), t \geq 0\}$ because they hold for $\{N(t), t \geq 0\}$. For instance, Axiom (ii) follows because nonoverlapping intervals from time s onward are nonoverlapping; and Axioms (iii) and (iv) follow because $N_s(t+h) - N_s(t) = N(s+t+h) - N(s+t)$. ■

Let T_1 denote the time of the first event of a Poisson process $\{N(t), t \geq 0\}$. That is,

$$T_1 = \min\{t \geq 0 : N(t) = 1\}$$

We now show that T_1 is an exponential random variable with rate λ.

Lemma 5.2. *If T_1 is the time of the first event of the Poisson process $\{N(t), t \geq 0\}$, then*

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

Proof. Let $P_0(t) = P(N(t) = 0)$. Then

$$\begin{aligned} P_0(t+h) &= P(N(t+h) = 0) \\ &= P(N(t) = 0,\ N(t+h) - N(t) = 0) \\ &= P(N(t) = 0)\, P(N(t+h) - N(t) = 0) \qquad \text{by Axiom (ii)} \\ &= P_0(t)\,(1 - \lambda h + o(h)) \qquad \text{by Axioms (iii) and (iv)} \end{aligned}$$

Hence,

$$P_0(t+h) - P_0(t) = -\lambda h P_0(t) + o(h)$$

Dividing by h and then letting $h \to 0$, gives that

$$P_0'(t) = -\lambda P_0(t)$$

or, equivalently

$$\frac{P_0'(t)}{P_0(t)} = -\lambda$$

Integration yields

$$\log(P_0(t)) = -\lambda t + C$$

or

$$P_0(t) = Ke^{-\lambda t}$$

Using that $1 = P_0(0)$ gives that $K = 1$. Because the time of the first event exceeds t if and only if $N(t) = 0$, we see that $P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$. ■

Whereas T_1 is the time of the first event of the Poisson process, for $n > 1$ we define T_n to be the time between the $(n-1)$st and the nth event. For instance, if $T_1 = 5$ and $T_2 = 10$, then the first event of the Poisson process occurred at time 5 and the second at time 15. The sequence $\{T_n, n = 1, 2, \ldots\}$ is called the *sequence of interarrival times*.

Proposition 5.4. *$T_1, T_2, \ldots$ are independent and identically distributed exponential random variables with rate λ.*

Proof. We've already shown that T_1 is exponential with rate λ. Now

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(0 \text{ events in } (s, s+t) | T_1 = s) \\ &= P(0 \text{ events in } (s, s+t)) \qquad \text{by independent increments} \\ &= P(N_s(t) = 0) \\ &= e^{-\lambda t} \end{aligned}$$

where the last equality follows from Lemma 5.2 because $N_s(t)$, $t \geq 0$ is, by Lemma 5.1, a Poisson process with rate λ. Hence, T_2 is exponential with rate λ and, because $P(T_2 > t | T_1 = s)$ does not depend on s, is independent of T_1. Repeating the argument (or using induction) completes the proof. ■

Another quantity of interest is S_n, the time of the nth event. Because the interarrival times are the times between successive events, it is easily seen that

$$S_n = \sum_{i=1}^{n} T_i, \qquad n \geq 1$$

Thus, from Propositions 5.4 and 5.1, it follows that S_n is a gamma (n, λ) random variable with density function

$$f_{S_n}(s) = \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!}, \qquad s > 0$$

We are now ready for the following important theorem.

Theorem 5.1. *If $\{N(t), t \geq 0\}$ is a Poisson process with rate λ, then $N(t)$ is a Poisson random variable with rate λt. That is,*

$$P(N(t) = n) = e^{-\lambda t}(\lambda t)^n / n!, \quad n \geq 0 \tag{5.12}$$

Proof. It was shown in Lemma 5.2 that $P(N(t) = 0) = e^{-\lambda t}$. For $n > 0$, we compute $P(N(t) = n)$ by conditioning on S_n, the time of the nth event. This gives

$$P(N(t) = n) = \int_0^t P(N(t) = n | S_n = s)\, \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!}\, ds \tag{5.13}$$

where the preceding used that $P(N(t) = n | S_n = s) = 0$ when $s > t$. Now, for $0 < s < t$, given that the nth event occurs at time s, there will be a total of n events by time t if the next interarrival time exceeds $t - s$. Hence,

$$\begin{aligned} P(N(t) = n | S_n = s) &= P(T_{n+1} > t - s | T_1 + \ldots T_n = s) \\ &= P(T_{n+1} > t - s) \\ &= e^{-\lambda(t-s)} \end{aligned}$$

where the last two equalities both used Proposition 5.4. Substituting this back into Eq. (5.13) yields that

$$\begin{aligned} P(N(t) = n) &= \int_0^t e^{-\lambda(t-s)}\, \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!}\, ds \\ &= e^{-\lambda t} \lambda^n \int_0^t \frac{s^{n-1}}{(n-1)!}\, ds \\ &= e^{-\lambda t} (\lambda t)^n / n! \end{aligned}$$

■

Remarks. **(i)** Because $\{N_s(t), t \geq 0\}$ is also a Poisson process with rate λ, if follows that $N_s(t) = N(t + s) - N(s)$ is a Poisson random variable with rate λ. Thus the number of events in any fixed interval of length t is Poisson with rate λ.

(ii) A counting process for which the distribution of the number of events in an interval depends only on the length of the interval and not its location is said to have *stationary increments*. Thus, a Poisson process has stationary increments.

(iii) The result that $N(t)$, or more generally $N(t+s) - N(s)$, has a Poisson distribution is a consequence of the Poisson approximation to the binomial distribution (see Section 2.2.4). To see this, subdivide the interval $[0, t]$ into k equal parts where k is very large (Fig. 5.1). Now it can be shown using axiom (iv) of Definition 5.2 that as k increases to ∞ the probability of having two or more events in any of the k subintervals goes to 0. Hence, $N(t)$ will (with a probability going to 1) just equal the number of subintervals in which an event occurs. However, by stationary and independent increments this number will have a binomial distribution with parameters k and $p = \lambda t/k + o(t/k)$. Hence, by the Poisson

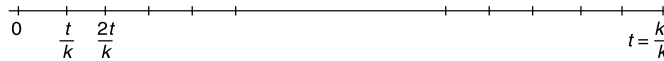

Figure 5.1

approximation to the binomial we see by letting k approach ∞ that $N(t)$ will have a Poisson distribution with mean equal to

$$\lim_{k\to\infty} k\left[\lambda\frac{t}{k} + o\left(\frac{t}{k}\right)\right] = \lambda t + \lim_{k\to\infty} \frac{to(t/k)}{t/k}$$
$$= \lambda t$$

by using the definition of $o(h)$ and the fact that $t/k \to 0$ as $k \to \infty$.

Example 5.13. Suppose that people immigrate into a territory according to a Poisson process with rate $\lambda = 2$ per day.

(a) Find the probability there are 10 arrivals in the following week (of 7 days).
(b) Find the expected number of days until there have been 20 arrivals.

Solution: (a) Because the number of arrivals in 7 days is Poisson with mean $7\lambda = 14$, it follows that the probability there will be 10 arrivals is $e^{-14}(14)^{10}/10!$.
(b) $E[S_{20}] = 20/\lambda = 10$. ■

5.3.3 Further Properties of Poisson Processes

Consider a Poisson process $\{N(t), t \geq 0\}$ having rate λ, and suppose that each time an event occurs it is classified as either a type I or a type II event. Suppose further that each event is classified as a type I event with probability p or a type II event with probability $1 - p$, independently of all other events. For example, suppose that customers arrive at a store in accordance with a Poisson process having rate λ; and suppose that each arrival is male with probability $\frac{1}{2}$ and female with probability $\frac{1}{2}$. Then a type I event would correspond to a male arrival and a type II event to a female arrival.

Let $N_1(t)$ and $N_2(t)$ denote respectively the number of type I and type II events occurring in $[0, t]$. Note that $N(t) = N_1(t) + N_2(t)$.

Proposition 5.5. *$\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ are both Poisson processes having respective rates λp and $\lambda(1 - p)$. Furthermore, the two processes are independent.*

Proof. It is easy to verify that $\{N_1(t),\ t \geq 0\}$ is a Poisson process with rate λp by verifying that it satisfies Definition 5.2.

- $N_1(0) = 0$ follows from the fact that $N(0) = 0$.
- It is easy to see that $\{N_1(t), t \geq 0\}$ inherits the stationary and independent increment properties of the process $\{N(t), t \geq 0\}$. This is true because the distribution of the number of type I events in an interval can be obtained by conditioning on the number of events in that interval, and the distribution of this latter quantity depends only on the length of the interval and is independent of what has occurred in any nonoverlapping interval.
- $$\begin{aligned} P\{N_1(h) = 1\} &= P\{N_1(h) = 1 \mid N(h) = 1\}P\{N(h) = 1\} \\ &\quad + P\{N_1(h) = 1 \mid N(h) \geq 2\}P\{N(h) \geq 2\} \\ &= p(\lambda h + o(h)) + o(h) \\ &= \lambda p h + o(h) \end{aligned}$$

- $P\{N_1(h) \geq 2\} \leq P\{N(h) \geq 2\} = o(h)$

Thus we see that $\{N_1(t), t \geq 0\}$ is a Poisson process with rate λp and, by a similar argument, that $\{N_2(t), t \geq 0\}$ is a Poisson process with rate $\lambda(1-p)$. Because the probability of a type I event in the interval from t to $t+h$ is independent of all that occurs in intervals that do not overlap $(t, t+h)$, it is independent of knowledge of when type II events occur, showing that the two Poisson processes are independent. (For another way of proving independence, see Example 3.24.) ■

Example 5.14. If immigrants to area A arrive at a Poisson rate of ten per week, and if each immigrant is of English descent with probability $\frac{1}{12}$, then what is the probability that no people of English descent will emigrate to area A during the month of February?

Solution: By the previous proposition it follows that the number of Englishmen emigrating to area A during the month of February is Poisson distributed with mean $4 \cdot 10 \cdot \frac{1}{12} = \frac{10}{3}$. Hence, the desired probability is $e^{-10/3}$. ■

Example 5.15. Suppose nonnegative offers to buy an item that you want to sell arrive according to a Poisson process with rate λ. Assume that each offer is the value of a continuous random variable having density function $f(x)$. Once the offer is presented to you, you must either accept it or reject it and wait for the next offer. We suppose that you incur costs at a rate c per unit time until the item is sold, and that your objective is to maximize your expected total return, where the total return is equal to the amount received minus the total cost incurred. Suppose you employ the policy of accepting the first offer that is greater than some specified value y. (Such a type of policy, which we call a y-policy, can be shown to be optimal.) What is the best value of y? What is the maximal expected net return?

Solution: Let us compute the expected total return when you use the y-policy, and then choose the value of y that maximizes this quantity. Let X denote the value of a random offer, and let $\bar{F}(x) = P\{X > x\} = \int_x^\infty f(u)\,du$ be its tail distribution function. Because each offer will be greater than y with probability $\bar{F}(y)$, it follows that such offers occur according to a Poisson process with rate $\lambda \bar{F}(y)$. Hence, the time until an offer is accepted is an exponential random variable with rate $\lambda \bar{F}(y)$. Letting $R(y)$ denote the total return from the policy that accepts the first offer that is greater than y, we have

$$\begin{aligned} E[R(y)] &= E[\text{accepted offer}] - cE[\text{time to accept}] \\ &= E[X|X > y] - \frac{c}{\lambda \bar{F}(y)} \\ &= \int_0^\infty x f_{X|X>y}(x)\,dx - \frac{c}{\lambda \bar{F}(y)} \\ &= \int_y^\infty x \frac{f(x)}{\bar{F}(y)}\,dx - \frac{c}{\lambda \bar{F}(y)} \\ &= \frac{\int_y^\infty x f(x)\,dx - c/\lambda}{\bar{F}(y)} \end{aligned} \tag{5.14}$$

Differentiation yields

$$\frac{d}{dy}E[R(y)] = 0 \Leftrightarrow -\bar{F}(y)yf(y) + \left(\int_y^\infty xf(x)\,dx - \frac{c}{\lambda}\right)f(y) = 0$$

Therefore, the optimal value of y satisfies

$$y\bar{F}(y) = \int_y^\infty xf(x)\,dx - \frac{c}{\lambda}$$

or

$$y\int_y^\infty f(x)\,dx = \int_y^\infty xf(x)\,dx - \frac{c}{\lambda}$$

or

$$\int_y^\infty (x-y)f(x)\,dx = \frac{c}{\lambda}$$

It is not difficult to show that there is a unique value of y that satisfies the preceding. Hence, the optimal policy is the one that accepts the first offer that is greater than y^*, where y^* is such that

$$\int_{y^*}^\infty (x-y^*)\,f(x)\,dx = c/\lambda$$

Putting $y = y^*$ in Eq. (5.14) shows that the maximal expected net return is

$$\begin{aligned} E[R(y^*)] &= \frac{1}{\bar{F}(y^*)}\left(\int_{y^*}^\infty (x - y^* + y^*)\,f(x)\,dx - c/\lambda\right) \\ &= \frac{1}{\bar{F}(y^*)}\left(\int_{y^*}^\infty (x-y^*)\,f(x)\,dx + y^*\int_{y^*}^\infty f(x)\,dx - c/\lambda\right) \\ &= \frac{1}{\bar{F}(y^*)}(c/\lambda + y^*\bar{F}(y^*) - c/\lambda) \\ &= y^* \end{aligned}$$

Thus, the optimal critical value is also the maximal expected net return. To understand why this is so, let m be the maximal expected net return, and note that when an offer is rejected the problem basically starts anew and so the maximal expected additional net return from then on is m. But this implies that it is optimal to accept an offer if and only if it is at least as large as m, showing that m is the optimal critical value. ■

It follows from Proposition 5.5 that if each of a Poisson number of individuals is independently classified into one of two possible groups with respective probabilities p and $1 - p$, then the number of individuals in each of the two groups will be independent Poisson random variables. Because this result easily generalizes to the case where the classification is into any one of r possible groups, we have the following application to a model of employees moving about in an organization.

Example 5.16. Consider a system in which individuals at any time are classified as being in one of r possible states, and assume that an individual changes states in accordance with a Markov chain having transition probabilities $P_{ij}, i, j = 1, \ldots, r$. That is, if an individual is in state i during a time period then, independently of its previous states, it will be in state j during the next time period with probability P_{ij}. The individuals are assumed to move through the system independently of each other. Suppose that the numbers of people initially in states $1, 2, \ldots, r$ are independent Poisson random variables with respective means $\lambda_1, \lambda_2, \ldots, \lambda_r$. We are interested in determining the joint distribution of the numbers of individuals in states $1, 2, \ldots, r$ at some time n.

Solution: For fixed i, let $N_j(i), j = 1, \ldots, r$ denote the number of those individuals, initially in state i, that are in state j at time n. Now each of the (Poisson distributed) number of people initially in state i will, independently of each other, be in state j at time n with probability P_{ij}^n, where P_{ij}^n is the n-stage transition probability for the Markov chain having transition probabilities P_{ij}. Hence, the $N_j(i), j = 1, \ldots, r$ will be independent Poisson random variables with respective means $\lambda_i P_{ij}^n, j = 1, \ldots, r$. Because the sum of independent Poisson random variables is itself a Poisson random variable, it follows that the number of individuals in state j at time n—namely $\sum_{i=1}^r N_j(i)$—will be independent Poisson random variables with respective means $\sum_i \lambda_i P_{ij}^n$, for $j = 1, \ldots, r$. ■

Example 5.17 (The Coupon Collecting Problem)**.** There are m different types of coupons. Each time a person collects a coupon it is, independently of ones previously obtained, a type j coupon with probability $p_j, \sum_{j=1}^m p_j = 1$. Let N denote the number of coupons one needs to collect in order to have a complete collection of at least one of each type. Find $E[N]$.

Solution: If we let N_j denote the number one must collect to obtain a type j coupon, then we can express N as

$$N = \max_{1 \le j \le m} N_j$$

However, even though each N_j is geometric with parameter p_j, the foregoing representation of N is not that useful, because the random variables N_j are not independent.

We can, however, transform the problem into one of determining the expected value of the maximum of *independent* random variables. To do so, suppose that coupons are collected at times chosen according to a Poisson process with rate $\lambda = 1$. Say that an event of this Poisson process is of type $j, 1 \le j \le m$, if the coupon obtained at that time is a type j coupon. If we now let $N_j(t)$ denote the number of type j coupons collected by time t, then it follows from Proposition 5.5 that $\{N_j(t), t \geqslant 0\}, j = 1, \ldots, m$ are independent Poisson processes with respective rates $\lambda p_j = p_j$. Let X_j denote the time of the first event of the jth process, and let

$$X = \max_{1 \le j \le m} X_j$$

denote the time at which a complete collection is amassed. Since the X_j are independent exponential random variables with respective rates p_j, it follows that

$$\begin{aligned} P\{X < t\} &= P\{\max_{1 \le j \le m} X_j < t\} \\ &= P\{X_j < t, \text{ for } j = 1, \ldots, m\} \\ &= \prod_{j=1}^{m} (1 - e^{-p_j t}) \end{aligned}$$

Therefore,

$$\begin{aligned} E[X] &= \int_0^\infty P\{X > t\}\, dt \\ &= \int_0^\infty \left\{ 1 - \prod_{j=1}^{m} (1 - e^{-p_j t}) \right\} dt \end{aligned} \tag{5.15}$$

It remains to relate $E[X]$, the expected time until one has a complete set, to $E[N]$, the expected number of coupons it takes. This can be done by letting T_i denote the ith interarrival time of the Poisson process that counts the number of coupons obtained. Then it is easy to see that

$$X = \sum_{i=1}^{N} T_i$$

Since the T_i are independent exponentials with rate 1, and N is independent of the T_i, we see that

$$E[X|N] = N E[T_i] = N$$

Therefore,

$$E[X] = E[N]$$

and so $E[N]$ is as given in Eq. (5.15).

Let us now compute the expected number of types that appear only once in the complete collection. Letting I_i equal 1 if there is only a single type i coupon in the final set, and letting it equal 0 otherwise, we thus want

$$\begin{aligned} E\left[\sum_{i=1}^{m} I_i\right] &= \sum_{i=1}^{m} E[I_i] \\ &= \sum_{i=1}^{m} P\{I_i = 1\} \end{aligned}$$

Now there will be a single type i coupon in the final set if a coupon of each type has appeared before the second coupon of type i is obtained. Thus, letting S_i denote the time at which the second type i coupon is obtained, we have

$$P\{I_i = 1\} = P\{X_j < S_i, \text{ for all } j \neq i\}$$

Using that S_i has a gamma distribution with parameters $(2, p_i)$, this yields

$$\begin{aligned} P\{I_i = 1\} &= \int_0^\infty P\{X_j < S_i \text{ for all } j \neq i | S_i = x\} p_i e^{-p_i x} p_i x \, dx \\ &= \int_0^\infty P\{X_j < x, \text{ for all } j \neq i\} p_i^2 x e^{-p_i x} \, dx \\ &= \int_0^\infty \prod_{j \neq i} (1 - e^{-p_j x}) p_i^2 x e^{-p_i x} \, dx \end{aligned}$$

Therefore, we have the result

$$\begin{aligned} E\left[\sum_{i=1}^m I_i\right] &= \int_0^\infty \sum_{i=1}^m \prod_{j \neq i} (1 - e^{-p_j x}) p_i^2 x e^{-p_i x} \, dx \\ &= \int_0^\infty x \prod_{j=1}^m (1 - e^{-p_j x}) \sum_{i=1}^m p_i^2 \frac{e^{-p_i x}}{1 - e^{-p_i x}} \, dx \end{aligned}$$

■

The next probability calculation related to Poisson processes that we shall determine is the probability that n events occur in one Poisson process before m events have occurred in a second and independent Poisson process. More formally let $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ be two independent Poisson processes having respective rates λ_1 and λ_2. Also, let S_n^1 denote the time of the nth event of the first process, and S_m^2 the time of the mth event of the second process. We seek

$$P\left\{S_n^1 < S_m^2\right\}$$

Before attempting to calculate this for general n and m, let us consider the special case $n = m = 1$. Since S_1^1, the time of the first event of the $N_1(t)$ process, and S_1^2, the time of the first event of the $N_2(t)$ process, are both exponentially distributed random variables (by Proposition 5.4) with respective means $1/\lambda_1$ and $1/\lambda_2$, it follows from Section 5.2.3 that

$$P\left\{S_1^1 < S_1^2\right\} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{5.16}$$

Let us now consider the probability that two events occur in the $N_1(t)$ process before a single event has occurred in the $N_2(t)$ process. That is, $P\{S_2^1 < S_1^2\}$. To calculate this we reason as follows: In order for the $N_1(t)$ process to have two events before a single event occurs in the $N_2(t)$ process, it is first necessary for the initial event

that occurs to be an event of the $N_1(t)$ process (and this occurs, by Eq. (5.16), with probability $\lambda_1/(\lambda_1+\lambda_2)$). Now, given that the initial event is from the $N_1(t)$ process, the next thing that must occur for S_2^1 to be less than S_1^2 is for the second event also to be an event of the $N_1(t)$ process. However, when the first event occurs both processes start all over again (by the memoryless property of Poisson processes) and hence this conditional probability is also $\lambda_1/(\lambda_1+\lambda_2)$; thus, the desired probability is given by

$$P\left\{S_2^1 < S_1^2\right\} = \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^2$$

In fact, this reasoning shows that *each event that occurs is going to be an event of the $N_1(t)$ process with probability $\lambda_1/(\lambda_1+\lambda_2)$ or an event of the $N_2(t)$ process with probability $\lambda_2/(\lambda_1+\lambda_2)$, independent of all that has previously occurred*. In other words, the probability that the $N_1(t)$ process reaches n before the $N_2(t)$ process reaches m is just the probability that n heads will appear before m tails if one flips a coin having probability $p=\lambda_1/(\lambda_1+\lambda_2)$ of a head appearing. But by noting that this event will occur if and only if the first $n+m-1$ tosses result in n or more heads, we see that our desired probability is given by

$$P\left\{S_n^1 < S_m^2\right\} = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n+m-1-k}$$

5.3.4 Conditional Distribution of the Arrival Times

Suppose we are told that exactly one event of a Poisson process has taken place by time t, and we are asked to determine the distribution of the time at which the event occurred. Now, since a Poisson process possesses stationary and independent increments it seems reasonable that each interval in $[0,t]$ of equal length should have the same probability of containing the event. In other words, the time of the event should be uniformly distributed over $[0,t]$. This is easily checked since, for $s \le t$,

$$\begin{aligned}
P\{T_1 < s | N(t)=1\} &= \frac{P\{T_1 < s, N(t)=1\}}{P\{N(t)=1\}} \\
&= \frac{P\{1 \text{ event in } [0,s), 0 \text{ events in } [s,t]\}}{P\{N(t)=1\}} \\
&= \frac{P\{1 \text{ event in } [0,s)\}P\{0 \text{ events in } [s,t]\}}{P\{N(t)=1\}} \\
&= \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} \\
&= \frac{s}{t}
\end{aligned}$$

This result may be generalized, but before doing so we need to introduce the concept of order statistics.

Let $Y_1, Y_2, \ldots, Y_n$ be n random variables. We say that $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ are the *order statistics* corresponding to $Y_1, Y_2, \ldots, Y_n$ if $Y_{(k)}$ is the kth smallest value among $Y_1, \ldots, Y_n$, $k = 1, 2, \ldots, n$. For instance, if $n = 3$ and $Y_1 = 4$, $Y_2 = 5$, $Y_3 = 1$ then $Y_{(1)} = 1$, $Y_{(2)} = 4$, $Y_{(3)} = 5$. If the $Y_i, i = 1, \ldots, n$, are independent identically distributed continuous random variables with probability density f, then the joint density of the order statistics $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ is given by

$$f(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i), \qquad y_1 < y_2 < \cdots < y_n$$

The preceding follows since

(i) $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ will equal $(y_1, y_2, \ldots, y_n)$ if $(Y_1, Y_2, \ldots, Y_n)$ is equal to any of the $n!$ permutations of $(y_1, y_2, \ldots, y_n)$;

and

(ii) the probability density that $(Y_1, Y_2, \ldots, Y_n)$ is equal to $(y_{i_1}, \ldots, y_{i_n})$ is $\prod_{j=1}^{n} f(y_{i_j}) = \prod_{j=1}^{n} f(y_j)$ when $i_1, \ldots, i_n$ is a permutation of $1, 2, \ldots, n$.

If the $Y_i, i = 1, \ldots, n$, are uniformly distributed over $(0, t)$, then we obtain from the preceding that the joint density function of the order statistics $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ is

$$f(y_1, y_2, \ldots, y_n) = \frac{n!}{t^n}, \qquad 0 < y_1 < y_2 < \cdots < y_n < t$$

We are now ready for the following useful theorem.

Theorem 5.2. *Given that $N(t) = n$, the n arrival times $S_1, \ldots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $(0, t)$.*

Proof. To obtain the conditional density of $S_1, \ldots, S_n$ given that $N(t) = n$ note that for $0 < s_1 < \cdots < s_n < t$ the event that $S_1 = s_1, S_2 = s_2, \ldots, S_n = s_n, N(t) = n$ is equivalent to the event that the first $n + 1$ interarrival times satisfy $T_1 = s_1, T_2 = s_2 - s_1, \ldots, T_n = s_n - s_{n-1}, T_{n+1} > t - s_n$. Hence, using Proposition 5.4, we have that the conditional joint density of $S_1, \ldots, S_n$ given that $N(t) = n$ is as follows:

$$\begin{aligned} f(s_1, \ldots, s_n \mid n) &= \frac{f(s_1, \ldots, s_n, n)}{P\{N(t) = n\}} \\ &= \frac{\lambda e^{-\lambda s_1} \lambda e^{-\lambda(s_2 - s_1)} \cdots \lambda e^{-\lambda(s_n - s_{n-1})} e^{-\lambda(t - s_n)}}{e^{-\lambda t}(\lambda t)^n / n!} \\ &= \frac{n!}{t^n}, \qquad 0 < s_1 < \cdots < s_n < t \end{aligned}$$

which proves the result. ■

Remark. The preceding result is usually paraphrased as stating that, under the condition that n events have occurred in $(0, t)$, the times $S_1, \ldots, S_n$ at which events

occur, considered as unordered random variables, are distributed independently and uniformly in the interval $(0, t)$.

Application of Theorem 5.2 (Sampling a Poisson Process). In Proposition 5.5 we showed that if each event of a Poisson process is independently classified as a type I event with probability p and as a type II event with probability $1 - p$ then the counting processes of type I and type II events are independent Poisson processes with respective rates λp and $\lambda(1 - p)$. Suppose now, however, that there are k possible types of events and that the probability that an event is classified as a type i event, $i = 1, \ldots, k$, depends on the time the event occurs. Specifically, suppose that if an event occurs at time y then it will be classified as a type i event, independently of anything that has previously occurred, with probability $P_i(y), i = 1, \ldots, k$ where $\sum_{i=1}^{k} P_i(y) = 1$. Upon using Theorem 5.2 we can prove the following useful proposition.

Proposition 5.6. *If $N_i(t), i = 1, \ldots, k$, represents the number of type i events occurring by time t then $N_i(t), i = 1, \ldots, k$, are independent Poisson random variables having means*

$$E[N_i(t)] = \lambda \int_0^t P_i(s)\, ds$$

Before proving this proposition, let us first illustrate its use.

Example 5.18 (An Infinite Server Queue)**.** Suppose that customers arrive at a service station in accordance with a Poisson process with rate λ. Upon arrival the customer is immediately served by one of an infinite number of possible servers, and the service times are assumed to be independent with a common distribution G. What is the distribution of $X(t)$, the number of customers that have completed service by time t? What is the distribution of $Y(t)$, the number of customers that are being served at time t?

To answer the preceding questions let us agree to call an entering customer a type I customer if he completes his service by time t and a type II customer if he does not complete his service by time t. Now, if the customer enters at time $s, s \le t$, then he will be a type I customer if his service time is less than $t - s$. Since the service time distribution is G, the probability of this will be $G(t - s)$. Similarly, a customer entering at time $s, s \le t$, will be a type II customer with probability $\bar{G}(t - s) = 1 - G(t - s)$. Hence, from Proposition 5.6 we obtain that the distribution of $X(t)$, the number of customers that have completed service by time t, is Poisson distributed with mean

$$E[X(t)] = \lambda \int_0^t G(t - s)\, ds = \lambda \int_0^t G(y)\, dy \tag{5.17}$$

Similarly, the distribution of $Y(t)$, the number of customers being served at time t is Poisson with mean

$$E[Y(t)] = \lambda \int_0^t \bar{G}(t - s)\, ds = \lambda \int_0^t \bar{G}(y)\, dy \tag{5.18}$$

Furthermore, $X(t)$ and $Y(t)$ are independent.

Suppose now that we are interested in computing the joint distribution of $Y(t)$ and $Y(t+s)$—that is, the joint distribution of the number in the system at time t and at time $t+s$. To accomplish this, say that an arrival is

type 1: if he arrives before time t and completes service between t and $t+s$,
type 2: if he arrives before t and completes service after $t+s$,
type 3: if he arrives between t and $t+s$ and completes service after $t+s$,
type 4: otherwise.

Hence, an arrival at time y will be type i with probability $P_i(y)$ given by

$$P_1(y) = \begin{cases} G(t+s-y) - G(t-y), & \text{if } y < t \\ 0, & \text{otherwise} \end{cases}$$

$$P_2(y) = \begin{cases} \bar{G}(t+s-y), & \text{if } y < t \\ 0, & \text{otherwise} \end{cases}$$

$$P_3(y) = \begin{cases} \bar{G}(t+s-y), & \text{if } t < y < t+s \\ 0, & \text{otherwise} \end{cases}$$

$$P_4(y) = 1 - P_1(y) - P_2(y) - P_3(y)$$

Thus, if $N_i = N_i(s+t), i = 1, 2, 3$, denotes the number of type i events that occur, then from Proposition 5.6, $N_i, i = 1, 2, 3$, are independent Poisson random variables with respective means

$$E[N_i] = \lambda \int_0^{t+s} P_i(y)\, dy, \qquad i = 1, 2, 3$$

Because

$$\begin{aligned} Y(t) &= N_1 + N_2, \\ Y(t+s) &= N_2 + N_3 \end{aligned}$$

it is now an easy matter to compute the joint distribution of $Y(t)$ and $Y(t+s)$. For instance,

$$\begin{aligned} &\text{Cov}[Y(t), Y(t+s)] \\ &\quad = \text{Cov}(N_1 + N_2, N_2 + N_3) \\ &\quad = \text{Cov}(N_2, N_2) \qquad \text{by independence of } N_1, N_2, N_3 \\ &\quad = \text{Var}(N_2) \\ &\quad = \lambda \int_0^t \bar{G}(t+s-y)\, dy = \lambda \int_0^t \bar{G}(u+s)\, du \end{aligned}$$

where the last equality follows since the variance of a Poisson random variable equals its mean, and from the substitution $u = t - y$. Also, the joint distribution of $Y(t)$ and $Y(t+s)$ is as follows:

$$P\{Y(t) = i, Y(t+s) = j\} = P\{N_1 + N_2 = i, N_2 + N_3 = j\}$$

Figure 5.2 Cars enter at point a and depart at b.

$$= \sum_{l=0}^{\min(i,j)} P\{N_2 = l, N_1 = i-l, N_3 = j-l\}$$
$$= \sum_{l=0}^{\min(i,j)} P\{N_2 = l\}P\{N_1 = i-l\}P\{N_3 = j-l\}$$ ■

Example 5.19 (A One Lane Road with No Overtaking)**.** Consider a one lane road with a single entrance and a single exit point which are of distance L from each other (see Fig. 5.2). Suppose that cars enter this road according to a Poisson process with rate λ, and that each entering car has an attached random value V which represents the velocity at which the car will travel, with the proviso that whenever the car encounters a slower moving car it must decrease its speed to that of the slower moving car. Let V_i denote the velocity value of the ith car to enter the road, and suppose that $V_i, i \geq 1$ are independent and identically distributed and, in addition, are independent of the counting process of cars entering the road. Assuming that the road is empty at time 0, we are interested in determining

(a) the probability mass function of $R(t)$, the number of cars on the road at time t; and

(b) the distribution of the road traversal time of a car that enters the road at time y.

Solution: Let $T_i = L/V_i$ denote the time it would take car i to travel the road if it were empty when car i arrived. Call T_i the free travel time of car i, and note that $T_1, T_2, \ldots$ are independent with distribution function

$$G(x) = P(T_i \leq x) = P(L/V_i \leq x) = P(V_i \geq L/x)$$

Let us say that an event occurs each time that a car enters the road. Also, let t be a fixed value, and say that an event that occurs at time s is a type 1 event if both $s \leq t$ and the free travel time of the car entering the road at time s exceeds $t - s$. In other words, a car entering the road is a type 1 event if the car would be on the road at time t even if the road were empty when it entered. Note that, independent of all that occurred prior to time s, an event occurring at time s is a type 1 event with probability

$$P(s) = \begin{cases} \bar{G}(t-s), & \text{if } s \leq t \\ 0, & \text{if } s > t \end{cases}$$

Letting $N_1(y)$ denote the number of type 1 events that occur by time y, it follows from Proposition 5.6 that $N_1(y)$ is, for $y \leq t$, a Poisson random variable with mean

$$E[N_1(y)] = \lambda \int_0^y \bar{G}(t-s)\,ds, \quad y \leq t$$

Because there will be no cars on the road at time t if and only if $N_1(t) = 0$, it follows that

$$P(R(t) = 0) = P(N_1(t) = 0) = e^{-\lambda \int_0^t \bar{G}(t-s)\,ds} = e^{-\lambda \int_0^t \bar{G}(u)\,du}$$

To determine $P(R(t) = n)$ for $n > 0$ we will condition on when the first type 1 event occurs. With X equal to the time of the first type 1 event (or to ∞ if there are no type 1 events), its distribution function is obtained by noting that

$$X \le y \Leftrightarrow N_1(y) > 0$$

thus showing that

$$F_X(y) = P(X \le y) = P(N_1(y) > 0) = 1 - e^{-\lambda \int_0^y \bar{G}(t-s)\,ds}, \quad y \le t$$

Differentiating gives the density function of X:

$$f_X(y) = \lambda \bar{G}(t-y)\, e^{-\lambda \int_0^y \bar{G}(t-s)\,ds}, \quad y \le t$$

To use the identity

$$P(R(t) = n) = \int_0^t P(R(t) = n | X = y)\, f_X(y)\, dy \tag{5.19}$$

note that if $X = y \le t$ then the leading car that is on the road at time t entered at time y. Because all other cars that arrive between y and t will also be on the road at time t, it follows that, conditional on $X = y$, the number of cars on the road at time t will be distributed as 1 plus a Poisson random variable with mean $\lambda(t-y)$. Therefore, for $n > 0$

$$P(R(t) = n | X = y) = \begin{cases} e^{-\lambda(t-y)} \frac{(\lambda(t-y))^{n-1}}{(n-1)!}, & \text{if } y \le t \\ 0, & \text{if } y = \infty \end{cases}$$

Substituting this into Eq. (5.19) yields

$$P(R(t) = n) = \int_0^t e^{-\lambda(t-y)} \frac{(\lambda(t-y))^{n-1}}{(n-1)!} \lambda \bar{G}(t-y)\, e^{-\lambda \int_0^y \bar{G}(t-s)\,ds}\, dy$$

(b) Let T be the free travel time of the car that enters the road at time y, and let $A(y)$ be its actual travel time. To determine $P(A(y) < x)$, let $t = y + x$ and note that $A(y)$ will be less than x if and only if both $T < x$ and there have been no type 1 events (using $t = y + x$) before time y. That is,

$$A(y) < x \;\Leftrightarrow\; T < x,\; N_1(y) = 0$$

Because T is independent of what has occurred prior to time y, the preceding gives

$$P(A(y) < x) = P(T < x) P(N_1(y) = 0)$$

$$= G(x)e^{-\lambda \int_0^y \bar{G}(y+x-s)\,ds}$$
$$= G(x)e^{-\lambda \int_x^{y+x} \bar{G}(u)\,du}$$ ■

Example 5.20 (Tracking the Number of HIV Infections)**.** There is a relatively long incubation period from the time when an individual becomes infected with the HIV virus, which causes AIDS, until the symptoms of the disease appear. As a result, it is difficult for public health officials to be certain of the number of members of the population that are infected at any given time. We will now present a first approximation model for this phenomenon, which can be used to obtain a rough estimate of the number of infected individuals.

Let us suppose that individuals contract the HIV virus in accordance with a Poisson process whose rate λ is unknown. Suppose that the time from when an individual becomes infected until symptoms of the disease appear is a random variable having a known distribution G. Suppose also that the incubation times of different infected individuals are independent.

Let $N_1(t)$ denote the number of individuals who have shown symptoms of the disease by time t. Also, let $N_2(t)$ denote the number who are HIV positive but have not yet shown any symptoms by time t. Now, since an individual who contracts the virus at time s will have symptoms by time t with probability $G(t-s)$ and will not with probability $\bar{G}(t-s)$, it follows from Proposition 5.6 that $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with respective means

$$E[N_1(t)] = \lambda \int_0^t G(t-s)\,ds = \lambda \int_0^t G(y)\,dy$$

and

$$E[N_2(t)] = \lambda \int_0^t \bar{G}(t-s)\,ds = \lambda \int_0^t \bar{G}(y)\,dy$$

Now, if we knew λ, then we could use it to estimate $N_2(t)$, the number of individuals infected but without any outward symptoms at time t, by its mean value $E[N_2(t)]$. However, since λ is unknown, we must first estimate it. Now, we will presumably know the value of $N_1(t)$, and so we can use its known value as an estimate of its mean $E[N_1(t)]$. That is, if the number of individuals who have exhibited symptoms by time t is n_1, then we can estimate that

$$n_1 \approx E[N_1(t)] = \lambda \int_0^t G(y)\,dy$$

Therefore, we can estimate λ by the quantity $\hat{\lambda}$ given by

$$\hat{\lambda} = n_1 \Big/ \int_0^t G(y)\,dy$$

Using this estimate of λ, we can estimate the number of infected but symptomless individuals at time t by

$$\begin{aligned}\text{estimate of } N_2(t) &= \hat{\lambda} \int_0^t \bar{G}(y)\,dy \\ &= \frac{n_1 \int_0^t \bar{G}(y)\,dy}{\int_0^t G(y)\,dy}\end{aligned}$$

For example, suppose that G is exponential with mean μ. Then $\bar{G}(y) = e^{-y/\mu}$, and a simple integration gives that

$$\text{estimate of } N_2(t) = \frac{n_1\mu(1 - e^{-t/\mu})}{t - \mu(1 - e^{-t/\mu})}$$

If we suppose that $t = 16$ years, $\mu = 10$ years, and $n_1 = 220$ thousand, then the estimate of the number of infected but symptomless individuals at time 16 is

$$\text{estimate} = \frac{2{,}200(1 - e^{-1.6})}{16 - 10(1 - e^{-1.6})} = 218.96$$

That is, if we suppose that the foregoing model is approximately correct (and we should be aware that the assumption of a constant infection rate λ that is unchanging over time is almost certainly a weak point of the model), then if the incubation period is exponential with mean 10 years and if the total number of individuals who have exhibited AIDS symptoms during the first 16 years of the epidemic is 220 thousand, then we can expect that approximately 219 thousand individuals are HIV positive though symptomless at time 16. ■

Proof of Proposition 5.6. Let us compute the joint probability $P\{N_i(t) = n_i, i = 1, \ldots, k\}$. To do so note first that in order for there to have been n_i type i events for $i = 1, \ldots, k$ there must have been a total of $\sum_{i=1}^k n_i$ events. Hence, conditioning on $N(t)$ yields

$$\begin{aligned}&P\{N_1(t) = n_1, \ldots, N_k(t) = n_k\} \\ &\quad = P\left\{N_1(t) = n_1, \ldots, N_k(t) = n_k \,\middle|\, N(t) = \sum_{i=1}^k n_i\right\} \\ &\qquad \times P\left\{N(t) = \sum_{i=1}^k n_i\right\}\end{aligned}$$

Now consider an arbitrary event that occurred in the interval $[0, t]$. If it had occurred at time s, then the probability that it would be a type i event would be $P_i(s)$. Hence, since by Theorem 5.2 this event will have occurred at some time uniformly distributed on $[0, t]$, it follows that the probability that this event will be a type i event is

$$P_i = \frac{1}{t} \int_0^t P_i(s)\,ds$$

independently of the other events. Hence,

$$P\left\{N_i(t)=n_i, i=1,\ldots,k \mid N(t)=\sum_{i=1}^{k} n_i\right\}$$

will just equal the multinomial probability of n_i type i outcomes for $i=1,\ldots,k$ when each of $\sum_{i=1}^{k} n_i$ independent trials results in outcome i with probability P_i, $i=1,\ldots,k$. That is,

$$P\left\{N_1(t)=n_1,\ldots,N_k(t)=n_k \mid N(t)=\sum_{i=1}^{k} n_i\right\}=\frac{\left(\sum_{i=1}^{k} n_i\right)!}{n_1!\cdots n_k!}P_1^{n_1}\cdots P_k^{n_k}$$

Consequently,

$$\begin{aligned}
&P\{N_1(t)=n_1,\ldots,N_k(t)=n_k\}\\
&\quad=\frac{(\sum_i n_i)!}{n_1!\cdots n_k!}P_1^{n_1}\cdots P_k^{n_k}e^{-\lambda t}\frac{(\lambda t)^{\sum_i n_i}}{(\sum_i n_i)!}\\
&\quad=\prod_{i=1}^{k} e^{-\lambda t P_i}(\lambda t P_i)^{n_i}/n_i!
\end{aligned}$$

and the proof is complete. ■

We now present some additional examples of the usefulness of Theorem 5.2.

Example 5.21. Insurance claims are made at times distributed according to a Poisson process with rate λ; the successive claim amounts are independent random variables having distribution G with mean μ, and are independent of the claim arrival times. Let S_i and C_i denote, respectively, the time and the amount of the ith claim. The total discounted cost of all claims made up to time t, call it $D(t)$, is defined by

$$D(t)=\sum_{i=1}^{N(t)} e^{-\alpha S_i}C_i$$

where α is the discount rate and $N(t)$ is the number of claims made by time t. To determine the expected value of $D(t)$, we condition on $N(t)$ to obtain

$$E[D(t)]=\sum_{n=0}^{\infty} E[D(t)|N(t)=n]e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

Now, conditional on $N(t)=n$, the claim arrival times $S_1,\ldots,S_n$ are distributed as the ordered values $U_{(1)},\ldots,U_{(n)}$ of n independent uniform $(0,t)$ random variables $U_1,\ldots,U_n$. Therefore,

$$E[D(t)|N(t)=n]=E\left[\sum_{i=1}^{n} C_i e^{-\alpha U_{(i)}}\right]$$

$$= \sum_{i=1}^{n} E[C_i e^{-\alpha U_{(i)}}]$$
$$= \sum_{i=1}^{n} E[C_i] E[e^{-\alpha U_{(i)}}]$$

where the final equality used the independence of the claim amounts and their arrival times. Because $E[C_i] = \mu$, continuing the preceding gives

$$E[D(t)|N(t) = n] = \mu \sum_{i=1}^{n} E[e^{-\alpha U_{(i)}}]$$
$$= \mu E\left[\sum_{i=1}^{n} e^{-\alpha U_{(i)}}\right]$$
$$= \mu E\left[\sum_{i=1}^{n} e^{-\alpha U_i}\right]$$

The last equality follows because $U_{(1)}, \ldots, U_{(n)}$ are the values $U_1, \ldots, U_n$ in increasing order, and so $\sum_{i=1}^{n} e^{-\alpha U_{(i)}} = \sum_{i=1}^{n} e^{-\alpha U_i}$. Continuing the string of equalities yields

$$E[D(t)|N(t) = n] = n\mu E[e^{-\alpha U}]$$
$$= n\frac{\mu}{t}\int_0^t e^{-\alpha x}\, dx$$
$$= n\frac{\mu}{\alpha t}(1 - e^{-\alpha t})$$

Therefore,

$$E[D(t)|N(t)] = N(t)\frac{\mu}{\alpha t}(1 - e^{-\alpha t})$$

Taking expectations yields the result

$$E[D(t)] = \frac{\lambda\mu}{\alpha}(1 - e^{-\alpha t})$$ ■

Example 5.22 (An Optimization Example)**.** Suppose that items arrive at a processing plant in accordance with a Poisson process with rate λ. At a fixed time T, all items are dispatched from the system. The problem is to choose an intermediate time, $t \in (0, T)$, at which all items in the system are dispatched, so as to minimize the total expected wait of all items.

If we dispatch at time t, $0 < t < T$, then the expected total wait of all items will be

$$\frac{\lambda t^2}{2} + \frac{\lambda(T - t)^2}{2}$$

To see why this is true, we reason as follows: The expected number of arrivals in $(0, t)$ is λt, and each arrival is uniformly distributed on $(0, t)$, and hence has expected wait $t/2$. Thus, the expected total wait of items arriving in $(0, t)$ is $\lambda t^2/2$. Similar reasoning holds for arrivals in (t, T), and the preceding follows. To minimize this quantity, we differentiate with respect to t to obtain

$$\frac{d}{dt}\left[\lambda\frac{t^2}{2}+\lambda\frac{(T-t)^2}{2}\right]=\lambda t-\lambda(T-t)$$

and equating to 0 shows that the dispatch time that minimizes the expected total wait is $t = T/2$. ■

We end this section with a result, quite similar in spirit to Theorem 5.2, which states that given S_n, the time of the nth event, then the first $n-1$ event times are distributed as the ordered values of a set of $n-1$ random variables uniformly distributed on $(0, S_n)$.

Proposition 5.7. *Given that $S_n = t$, the set $S_1, \ldots, S_{n-1}$ has the distribution of a set of $n-1$ independent uniform $(0, t)$ random variables.*

Proof. We can prove the preceding in the same manner as we did Theorem 5.2, or we can argue more loosely as follows:

$$\begin{aligned} S_1, \ldots, S_{n-1} \mid S_n = t &\sim S_1, \ldots, S_{n-1} \mid S_n = t,\ N(t^-) = n-1 \\ &\sim S_1, \ldots, S_{n-1} \mid N(t^-) = n-1 \end{aligned}$$

where $\sim$ means "has the same distribution as" and t^- is infinitesimally smaller than t. The result now follows from Theorem 5.2. ■

5.3.5 Estimating Software Reliability

When a new computer software package is developed, a testing procedure is often put into effect to eliminate the faults, or bugs, in the package. One common procedure is to try the package on a set of well-known problems to see if any errors result. This goes on for some fixed time, with all resulting errors being noted. Then the testing stops and the package is carefully checked to determine the specific bugs that were responsible for the observed errors. The package is then altered to remove these bugs. Because we cannot be certain that all the bugs in the package have been eliminated, however, a problem of great importance is the estimation of the error rate of the revised software package.

To model the preceding, let us suppose that initially the package contains an unknown number, m, of bugs, which we will refer to as bug 1, bug 2, ..., bug m. Suppose also that bug i will cause errors to occur in accordance with a Poisson process having an unknown rate λ_i, $i = 1, \ldots, m$. Then, for instance, the number of errors due to bug i that occurs in any s units of operating time is Poisson distributed with mean $\lambda_i s$. Also suppose that these Poisson processes caused by bugs $i, i = 1, \ldots, m$ are independent. In addition, suppose that the package is to be run for t time units with all resulting

errors being noted. At the end of this time a careful check of the package is made to determine the specific bugs that caused the errors (that is, a *debugging*, takes place). These bugs are removed, and the problem is then to determine the error rate for the revised package.

If we let

$$\psi_i(t) = \begin{cases} 1, & \text{if bug } i \text{ has not caused an error by } t \\ 0, & \text{otherwise} \end{cases}$$

then the quantity we wish to estimate is

$$\Lambda(t) = \sum_i \lambda_i \psi_i(t)$$

the error rate of the final package. To start, note that

$$\begin{aligned} E[\Lambda(t)] &= \sum_i \lambda_i E[\psi_i(t)] \\ &= \sum_i \lambda_i e^{-\lambda_i t} \end{aligned} \tag{5.20}$$

Now each of the bugs that is discovered would have been responsible for a certain number of errors. Let us denote by $M_j(t)$ the number of bugs that were responsible for j errors, $j \geq 1$. That is, $M_1(t)$ is the number of bugs that caused exactly one error, $M_2(t)$ is the number that caused two errors, and so on, with $\sum_j j M_j(t)$ equaling the total number of errors that resulted. To compute $E[M_1(t)]$, let us define the indicator variables, $I_i(t), i \geq 1$, by

$$I_i(t) = \begin{cases} 1, & \text{bug } i \text{ causes exactly 1 error} \\ 0, & \text{otherwise} \end{cases}$$

Then,

$$M_1(t) = \sum_i I_i(t)$$

and so

$$E[M_1(t)] = \sum_i E[I_i(t)] = \sum_i \lambda_i t e^{-\lambda_i t} \tag{5.21}$$

Thus, from (5.20) and (5.21) we obtain the intriguing result that

$$E\left[\Lambda(t) - \frac{M_1(t)}{t}\right] = 0 \tag{5.22}$$

Thus suggests the possible use of $M_1(t)/t$ as an estimate of $\Lambda(t)$. To determine whether or not $M_1(t)/t$ constitutes a "good" estimate of $\Lambda(t)$ we shall look at how far

apart these two quantities tend to be. That is, we will compute

$$E\left[\left(\Lambda(t)-\frac{M_1(t)}{t}\right)^2\right]=\mathrm{Var}\left(\Lambda(t)-\frac{M_1(t)}{t}\right) \qquad \text{from (5.22)}$$
$$=\mathrm{Var}(\Lambda(t))-\frac{2}{t}\mathrm{Cov}(\Lambda(t),M_1(t))+\frac{1}{t^2}\mathrm{Var}(M_1(t))$$

Now,

$$\mathrm{Var}(\Lambda(t))=\sum_i \lambda_i^2\mathrm{Var}(\psi_i(t))=\sum_i \lambda_i^2 e^{-\lambda_i t}(1-e^{-\lambda_i t}),$$
$$\mathrm{Var}(M_1(t))=\sum_i \mathrm{Var}(I_i(t))=\sum_i \lambda_i t e^{-\lambda_i t}(1-\lambda_i t e^{-\lambda_i t}),$$
$$\begin{aligned}\mathrm{Cov}(\Lambda(t),M_1(t))&=\mathrm{Cov}\left(\sum_i \lambda_i\psi_i(t),\ \sum_j I_j(t)\right)\\&=\sum_i\sum_j \mathrm{Cov}(\lambda_i\psi_i(t),I_j(t))\\&=\sum_i \lambda_i\mathrm{Cov}(\psi_i(t),I_i(t))\\&=-\sum_i \lambda_i e^{-\lambda_i t}\lambda_i t e^{-\lambda_i t}\end{aligned}$$

where the last two equalities follow since $\psi_i(t)$ and $I_j(t)$ are independent when $i\neq j$ because they refer to different Poisson processes and $\psi_i(t)I_i(t)=0$. Hence, we obtain

$$\begin{aligned}E\left[\left(\Lambda(t)-\frac{M_1(t)}{t}\right)^2\right]&=\sum_i \lambda_i^2 e^{-\lambda_i t}+\frac{1}{t}\sum_i \lambda_i e^{-\lambda_i t}\\&=\frac{E[M_1(t)+2M_2(t)]}{t^2}\end{aligned}$$

where the last equality follows from (5.21) and the identity (which we leave as an exercise)

$$E[M_2(t)]=\frac{1}{2}\sum_i (\lambda_i t)^2 e^{-\lambda_i t}$$

Thus, we can estimate the average square of the difference between $\Lambda(t)$ and $M_1(t)/t$ by the observed value of $M_1(t)+2M_2(t)$ divided by t^2.

Example 5.23. Suppose that in 100 units of operating time 20 bugs are discovered of which two resulted in exactly one, and three resulted in exactly two, errors. Then we would estimate that $\Lambda(100)$ is something akin to the value of a random variable whose mean is equal to $1/50$ and whose variance is equal to $8/10{,}000$. ■

5.4 Generalizations of the Poisson Process

5.4.1 Nonhomogeneous Poisson Process

In this section we consider two generalizations of the Poisson process. The first of these is the nonhomogeneous, also called the nonstationary, Poisson process, which is obtained by allowing the arrival rate at time t to be a function of t.

Definition 5.3. The counting process $\{N(t), t \geq 0\}$ is said to be a *nonhomogeneous Poisson process with intensity function* $\lambda(t),\ t \geq 0$, if

(i) $N(0) = 0$.
(ii) $\{N(t), t \geq 0\}$ has independent increments.
(iii) $P\{N(t+h) - N(t) \geq 2\} = o(h)$.
(iv) $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$.

The function $m(t)$ defined by

$$m(t) = \int_0^t \lambda(y)\, dy$$

is called the *mean value function* of the nonhomogeneous Poisson process.

We start our analysis by proving a lemma analogous to Lemma 5.2.

Lemma 5.3. *If* $\{N(t), t \geq 0\}$ *is a nonhomogeneous Poisson process having intensity function* $\lambda(t)$*, then*

$$P(N(t) = 0) = e^{-m(t)}$$

Proof. Let $P_0(t) = P(N(t) = 0)$. Then

$$\begin{aligned} P_0(t+h) &= P(N(t) = 0,\ N(t+h) - N(t) = 0) \\ &= P_0(t)\, P(N(t+h) - N(t) = 0) \\ &= P_0(t)\,(1 - \lambda(t)h + o(h)) \end{aligned}$$

Hence,

$$P_0(t+h) - P_0(t) = -\lambda(t)h P_0(t) + o(h)$$

Dividing through by h and letting $h \to 0$ now yields

$$P_0'(t) = -\lambda(t) P_0(t)$$

Hence,

$$\int_0^t \frac{P_0'(s)}{P_0(s)} ds = -\int_0^t \lambda(s) ds$$

or

$$\log(P_0(t)) - \log(P_0(0)) = -\int_0^t \lambda(s)ds$$

Using that $P_0(0) = 1$ shows that

$$P_0(t) = e^{-\int_0^t \lambda(s)ds} = e^{-m(t)}$$ ■

Now, if we let T_1 be the time of the first event then

$$P(T_1 > t) = P(N(t) = 0) = e^{-m(t)}$$

Differentiation gives that the density of T_1 is

$$f_{T_1}(t) = \lambda(t)e^{-m(t)}$$

Now, for $s > 0$, let $N_s(t) = N(s + t) - N(s)$. We leave the proof of the following as an exercise.

Lemma 5.4. *If $\{N(t), t \geq 0\}$ is a nonhomogeneous Poisson process having intensity function $\lambda(t)$, then $\{N_s(t), t \geq 0\}$ is a nonhomogeneous Poisson process having intensity function $\lambda_s(t) = \lambda(s + t)$, $t \geq 0$.*

The mean value function of $\{N_s(t), t \geq 0\}$ is

$$\begin{aligned} m_s(t) &= \int_0^t \lambda_s(y)dy \\ &= \int_0^t \lambda(s + y)dy \\ &= \int_s^{s+t} \lambda(u)du \\ &= m(s + t) - m(s) \end{aligned}$$

We are now ready to prove that $N(t)$ is a Poisson random variable with mean $m(t)$.

Theorem 5.3. *If $\{N(t), t \geq 0\}$ is a nonhomogeneous Poisson process having intensity function $\lambda(t)$, then*

$$P(N(t) = n) = e^{-m(t)}(m(t))^n/n!, \quad n \geq 0$$

Proof. We prove this by induction on n. As Lemma 5.3 shows that the result holds when $n = 0$, assume that if $\{N(t), t \geq 0\}$ is a nonhomogeneous Poisson process having intensity function $\lambda(t)$, then for any y

$$P(N(y) = n) = \frac{e^{-m(y)}(m(y))^n}{n!}$$

To complete the induction proof we must now show that the preceding assumption implies that

$$P(N(t) = n+1) = \frac{e^{-m(t)}(m(t))^{n+1}}{(n+1)!}$$

To show this, condition on T_1 to obtain

$$\begin{aligned} P(N(t) = n+1) &= \int_0^\infty P(N(t) = n+1 | T_1 = s) f_{T_1}(s) ds \\ &= \int_0^t P(N(t) = n+1 | T_1 = s) \lambda(s) e^{-m(s)} ds \end{aligned}$$

Now, given that the first event occurs at time s, there will be a total of $n+1$ events by time t if n events occur between times s and t. Thus, from the preceding,

$$\begin{aligned} P(N(t) = n+1) &= \int_0^t P(N(t) - N(s) = n | T_1 = s)\, \lambda(s)\, e^{-m(s)}\, ds \\ &= \int_0^t P(N(t) - N(s) = n)\, \lambda(s)\, e^{-m(s)}\, ds \end{aligned}$$

by independent increments

$$= \int_0^t P(N_s(t-s) = n)\, \lambda(s)\, e^{-m(s)}\, ds$$

Now, from Lemma 5.4 and the induction hypothesis, it follows that

$$\begin{aligned} P(N_s(t-s) = n) &= \frac{e^{-m_s(t-s)}(m_s(t-s))^n}{n!} \\ &= \frac{e^{-(m(t)-m(s))}\,(m(t)-m(s))^n}{n!} \end{aligned}$$

Substituting this back into the previous expression yields

$$\begin{aligned} P(N(t) = n+1) &= \int_0^t \frac{e^{-(m(t)-m(s))}\,(m(t)-m(s))^n}{n!}\, \lambda(s)\, e^{-m(s)}\, ds \\ &= \frac{e^{-m(t)}}{n!} \int_0^t (m(t)-m(s))^n\, \lambda(s)\, ds \end{aligned}$$

Making the change of variables $y = m(t) - m(s)$, $dy = -\lambda(s)\, ds$ gives

$$\begin{aligned} P(N(t) = n+1) &= \frac{e^{-m(t)}}{n!} \int_0^{m(t)} y^n dy \\ &= \frac{e^{-m(t)}(m(t))^{n+1}}{(n+1)!} \end{aligned}$$

and the induction proof is established. ■

Remarks. **(i)** Because $\{N_s(t), t \geq 0\}$ is a nonhomogeneous Poisson process with mean value function $m_s(t) = m(s+t) - m(s)$, it follows from Theorem 5.3 that $N_s(t) = N(s+t) - N(s)$ is Poisson with mean $m(s+t) - m(s)$.

(ii) That $N(s+t) - N(s)$ has a Poisson distribution with mean $\int_s^{s+t} \lambda(y)\,dy$ is a consequence of the Poisson limit of the sum of independent Bernoulli random variables (see Example 2.47). To see why, subdivide the interval $[s, s+t]$ into n subintervals of length $\frac{t}{n}$, where subinterval i goes from $s + (i-1)\frac{t}{n}$ to $s + i\frac{t}{n}$, $i = 1, \ldots, n$. Let $N_i = N(s + i\frac{t}{n}) - N(s + (i-1)\frac{t}{n})$ be the number of events that occur in subinterval i, and note that

$$\begin{aligned} P\{\geq 2 \text{ events in some subinterval}\} &= P\left(\bigcup_{i=1}^{n} \{N_i \geq 2\}\right) \\ &\leq \sum_{i=1}^{n} P\{N_i \geq 2\} \\ &= no(t/n) \qquad \text{by Axiom (iii)} \end{aligned}$$

Because

$$\lim_{n\to\infty} no(t/n) = \lim_{n\to\infty} t\frac{o(t/n)}{t/n} = 0$$

it follows that, as n increases to ∞, the probability of having two or more events in any of the n subintervals goes to 0. Consequently, with a probability going to 1, $N(t)$ will equal the number of subintervals in which an event occurs. Because the probability of an event in subinterval i is $\lambda(s + i\frac{t}{n})\frac{t}{n} + o(\frac{t}{n})$, it follows, because the number of events in different subintervals are independent, that when n is large the number of subintervals that contain an event is approximately a Poisson random variable with mean

$$\sum_{i=1}^{n} \lambda\left(s + i\frac{t}{n}\right)\frac{t}{n} + no(t/n)$$

But,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \lambda\left(s + i\frac{t}{n}\right)\frac{t}{n} + no(t/n) = \int_s^{s+t} \lambda(y)\,dy$$

and the result follows. ■

Time sampling an ordinary Poisson process generates a nonhomogeneous Poisson process. That is, let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ, and suppose that an event occurring at time t is, independently of what has occurred prior to t, counted with probability $p(t)$. With $N_c(t)$ denoting the number of counted events by time t, the counting process $\{N_c(t), t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t) = \lambda p(t)$. This is verified by noting that $\{N_c(t), t \geq 0\}$ satisfies the nonhomogeneous Poisson process axioms.

1. $N_c(0) = 0$.

2. The number of counted events in $(s, s+t)$ depends solely on the number of events of the Poisson process in $(s, s+t)$, which is independent of what has occurred prior to time s. Consequently, the number of counted events in $(s, s+t)$ is independent of the process of counted events prior to s, thus establishing the independent increment property.

3. Let $N_c(t, t+h) = N_c(t+h) - N_c(t)$, with a similar definition of $N(t, t+h)$.

$$P\{N_c(t, t+h) \geq 2\} \leq P\{N(t, t+h) \geq 2\} = o(h)$$

4. To compute $P\{N_c(t, t+h) = 1\}$, condition on $N(t, t+h)$.

$$\begin{aligned} &P\{N_c(t, t+h) = 1\} \\ &\quad = P\{N_c(t, t+h) = 1 | N(t, t+h) = 1\} P\{N(t, t+h) = 1\} \\ &\qquad + P\{N_c(t, t+h) = 1 | N(t, t+h) \geq 2\} P\{N(t, t+h) \geq 2\} \\ &\quad = P\{N_c(t, t+h) = 1 | N(t, t+h) = 1\} \lambda h + o(h) \\ &\quad = p(t) \lambda h + o(h) \end{aligned}$$

The importance of the nonhomogeneous Poisson process resides in the fact that we no longer require the condition of stationary increments. Thus we now allow for the possibility that events may be more likely to occur at certain times than during other times.

Example 5.24. Siegbert runs a hot dog stand that opens at 8 A.M. From 8 until 11 A.M. customers seem to arrive, on the average, at a steadily increasing rate that starts with an initial rate of 5 customers per hour at 8 A.M. and reaches a maximum of 20 customers per hour at 11 A.M. From 11 A.M. until 1 P.M. the (average) rate seems to remain constant at 20 customers per hour. However, the (average) arrival rate then drops steadily from 1 P.M. until closing time at 5 P.M. at which time it has the value of 12 customers per hour. If we assume that the numbers of customers arriving at Siegbert's stand during disjoint time periods are independent, then what is a good probability model for the preceding? What is the probability that no customers arrive between 8:30 A.M. and 9:30 A.M. on Monday morning? What is the expected number of arrivals in this period?

Solution: A good model for the preceding would be to assume that arrivals constitute a nonhomogeneous Poisson process with intensity function $\lambda(t)$ given by

$$\lambda(t) = \begin{cases} 5 + 5t, & 0 \leq t \leq 3 \\ 20, & 3 \leq t \leq 5 \\ 20 - 2(t-5), & 5 \leq t \leq 9 \end{cases}$$

and

$$\lambda(t) = \lambda(t-9) \qquad \text{for } t > 9$$

Note that $N(t)$ represents the number of arrivals during the first t hours that the store is open. That is, we do not count the hours between 5 P.M. and 8 A.M. If for some reason we wanted $N(t)$ to represent the number of arrivals during the first t hours regardless of whether the store was open or not, then, assuming that the process begins at midnight we would let

$$\lambda(t) = \begin{cases} 0, & 0 \le t < 8 \\ 5 + 5(t-8), & 8 \le t \le 11 \\ 20, & 11 \le t \le 13 \\ 20 - 2(t-13), & 13 \le t \le 17 \\ 0, & 17 < t \le 24 \end{cases}$$

and

$$\lambda(t) = \lambda(t-24) \qquad \text{for } t > 24$$

As the number of arrivals between 8:30 A.M. and 9:30 A.M. will be Poisson with mean $m(\frac{3}{2}) - m(\frac{1}{2})$ in the first representation (and $m(\frac{19}{2}) - m(\frac{17}{2})$ in the second representation), we have that the probability that this number is zero is

$$\exp\left\{-\int_{1/2}^{3/2} (5+5t)\,dt\right\} = e^{-10}$$

and the mean number of arrivals is

$$\int_{1/2}^{3/2} (5+5t)\,dt = 10$$ ■

Suppose that events occur according to a Poisson process with rate λ, and suppose that, independent of what has previously occurred, an event at time s is a type 1 event with probability $P_1(s)$ or a type 2 event with probability $P_2(s) = 1 - P_1(s)$. If $N_i(t), t \ge 0$, denotes the number of type i events by time t, then it easily follows from Definition 5.3 that $\{N_1(t), t \ge 0\}$ and $\{N_2(t), t \ge 0\}$ are independent nonhomogeneous Poisson processes with respective intensity functions $\lambda_i(t) = \lambda P_i(t), i = 1, 2$. (The proof mimics that of Proposition 5.5.) This result gives us another way of understanding (or of proving) the time sampling Poisson process result of Proposition 5.6, which states that $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with means $E[N_i(t)] = \lambda \int_0^t P_i(s)\,ds, i = 1, 2$.

Example 5.25 (The Output Process of an Infinite Server Poisson Queue)**.** It turns out that the output process of the $M/G/\infty$ queue—that is, of the infinite server queue having Poisson arrivals and general service distribution G—is a nonhomogeneous Poisson process having intensity function $\lambda(t) = \lambda G(t)$. To verify this claim, let us first argue that the departure process has independent increments. Towards this end, consider nonoverlapping intervals $O_1, \ldots, O_k$; now say that an arrival is type $i, i = 1, \ldots, k$, if that arrival departs in the interval O_i. By Proposition 5.6, it follows that the numbers

of departures in these intervals are independent, thus establishing independent increments. Now, suppose that an arrival is "counted" if that arrival departs between t and $t+h$. Because an arrival at time s, $s < t+h$, will be counted with probability $P(s)$, where

$$P(s) = \begin{cases} G(t+h-s) - G(t-s), & \text{if } s < t \\ G(t+h-s), & \text{if } t < s < t+h \end{cases}$$

it follows from Proposition 5.6 that the number of departures in $(t, t+h)$ is a Poisson random variable with mean

$$\begin{aligned} \lambda \int_0^{t+h} P(s)ds &= \lambda \int_0^{t+h} G(t+h-s)ds - \lambda \int_0^t G(t-s)ds \\ &= \lambda \int_0^{t+h} G(y)dy - \lambda \int_0^t G(y)dy \\ &= \lambda \int_t^{t+h} G(y)dy \\ &= \lambda G(t)h + o(h) \end{aligned}$$

Therefore,

$$P\{1 \text{ departure in } (t, t+h)\} = \lambda G(t)h\, e^{-\lambda G(t)h} + o(h) = \lambda G(t)h + o(h)$$

and

$$P\{\geq 2 \text{ departures in } (t, t+h)\} = o(h)$$

which completes the verification. ■

If we let S_n denote the time of the nth event of the nonhomogeneous Poisson process, then we can obtain its density as follows:

$$\begin{aligned} P\{t < S_n < t+h\} &= P\{N(t) = n-1, \text{ one event in } (t, t+h)\} + o(h) \\ &= P\{N(t) = n-1\}P\{\text{one event in } (t, t+h)\} + o(h) \\ &= e^{-m(t)} \frac{[m(t)]^{n-1}}{(n-1)!} [\lambda(t)h + o(h)] + o(h) \\ &= \lambda(t) e^{-m(t)} \frac{[m(t)]^{n-1}}{(n-1)!} h + o(h) \end{aligned}$$

which implies that

$$f_{S_n}(t) = \lambda(t) e^{-m(t)} \frac{[m(t)]^{n-1}}{(n-1)!}$$

where

$$m(t) = \int_0^t \lambda(s)\, ds$$

5.4.2 Compound Poisson Process

A stochastic process $\{X(t), t \geq 0\}$ is said to be a *compound Poisson process* if it can be represented as

$$X(t) = \sum_{i=1}^{N(t)} Y_i, \qquad t \geq 0 \tag{5.23}$$

where $\{N(t), t \geq 0\}$ is a Poisson process, and $\{Y_i, i \geq 1\}$ is a family of independent and identically distributed random variables that is also independent of $\{N(t), t \geq 0\}$. As noted in Chapter 3, the random variable $X(t)$ is said to be a compound Poisson random variable.

Examples of Compound Poisson Processes

(i) If $Y_i \equiv 1$, then $X(t) = N(t)$, and so we have the usual Poisson process.

(ii) Suppose that buses arrive at a sporting event in accordance with a Poisson process, and suppose that the numbers of fans in each bus are assumed to be independent and identically distributed. Then $\{X(t), t \geq 0\}$ is a compound Poisson process where $X(t)$ denotes the number of fans who have arrived by t. In Eq. (5.23) Y_i represents the number of fans in the ith bus.

(iii) Suppose customers leave a supermarket in accordance with a Poisson process. If the Y_i, the amount spent by the ith customer, $i = 1, 2, \ldots$, are independent and identically distributed, then $\{X(t), t \geq 0\}$ is a compound Poisson process when $X(t)$ denotes the total amount of money spent by time t. ■

Because $X(t)$ is a compound Poisson random variable with Poisson parameter λt, we have from Examples 3.10 and 3.19 that

$$E[X(t)] = \lambda t E[Y_1] \tag{5.24}$$

and

$$\text{Var}(X(t)) = \lambda t E\left[Y_1^2\right] \tag{5.25}$$

Example 5.26. Suppose that families migrate to an area at a Poisson rate $\lambda = 2$ per week. If the number of people in each family is independent and takes on the values 1, 2, 3, 4 with respective probabilities $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}$, then what is the expected value and variance of the number of individuals migrating to this area during a fixed five-week period?

Solution: Letting Y_i denote the number of people in the ith family, we have

$$E[Y_i] = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{3} + 3 \cdot \tfrac{1}{3} + 4 \cdot \tfrac{1}{6} = \tfrac{5}{2},$$
$$E\left[Y_i^2\right] = 1^2 \cdot \tfrac{1}{6} + 2^2 \cdot \tfrac{1}{3} + 3^2 \cdot \tfrac{1}{3} + 4^2 \cdot \tfrac{1}{6} = \tfrac{43}{6}$$

Hence, letting $X(5)$ denote the number of immigrants during a five-week period, we obtain from Eqs. (5.24) and (5.25) that

$$E[X(5)] = 2 \cdot 5 \cdot \tfrac{5}{2} = 25$$

and

$$\text{Var}[X(5)] = 2 \cdot 5 \cdot \tfrac{43}{6} = \tfrac{215}{3}$$ ■

Example 5.27 (Busy Periods in Single-Server Poisson Arrival Queues)**.** Consider a single-server service station in which customers arrive according to a Poisson process having rate λ. An arriving customer is immediately served if the server is free; if not, the customer waits in line (that is, he or she joins the queue). The successive service times are independent with a common distribution.

Such a system will alternate between idle periods when there are no customers in the system, so the server is idle, and busy periods when there are customers in the system, so the server is busy. A busy period will begin when an arrival finds the system empty, and because of the memoryless property of the Poisson arrivals it follows that the distribution of the length of a busy period will be the same for each such period. Let B denote the length of a busy period. We will compute its mean and variance.

To begin, let S denote the service time of the first customer in the busy period and let $N(S)$ denote the number of arrivals during that time. Now, if $N(S) = 0$ then the busy period will end when the initial customer completes his service, and so B will equal S in this case. Now, suppose that one customer arrives during the service time of the initial customer. Then, at time S there will be a single customer in the system who is just about to enter service. Because the arrival stream from time S on will still be a Poisson process with rate λ, it thus follows that the additional time from S until the system becomes empty will have the same distribution as a busy period. That is, if $N(S) = 1$ then

$$B = S + B_1$$

where B_1 is independent of S and has the same distribution as B.

Now, consider the general case where $N(S) = n$, so there will be n customers waiting when the server finishes his initial service. To determine the distribution of remaining time in the busy period note that the order in which customers are served will not affect the remaining time. Hence, let us suppose that the n arrivals, call them $C_1, \ldots, C_n$, during the initial service period are served as follows. Customer C_1 is served first, but C_2 is not served until the only customers in the system are $C_2, \ldots, C_n$. For instance, any customers arriving during C_1's service time will be served before C_2. Similarly, C_3 is not served until the system is free of all customers but $C_3, \ldots, C_n$, and so on. A little thought reveals that the times between the beginnings of service of customers C_i and $C_{i+1}, i = 1, \ldots, n-1$, and the time from the beginning of service of C_n until there are no customers in the system, are independent random variables, each distributed as a busy period.

It follows from the preceding that if we let $B_1, B_2, \ldots$ be a sequence of independent random variables, each distributed as a busy period, then we can express B as

$$B = S + \sum_{i=1}^{N(S)} B_i$$

Hence,

$$E[B|S] = S + E\left[\sum_{i=1}^{N(S)} B_i | S\right]$$

and

$$\text{Var}(B|S) = \text{Var}\left(\sum_{i=1}^{N(S)} B_i | S\right)$$

However, given S, $\sum_{i=1}^{N(S)} B_i$ is a compound Poisson random variable, and thus from Eqs. (5.24) and (5.25) we obtain

$$E[B|S] = S + \lambda S E[B] = (1 + \lambda E[B])S$$
$$\text{Var}(B|S) = \lambda S E[B^2]$$

Hence,

$$E[B] = E[E[B|S]] = (1 + \lambda E[B])E[S]$$

implying, provided that $\lambda E[S] < 1$, that

$$E[B] = \frac{E[S]}{1 - \lambda E[S]}$$

Also, by the conditional variance formula

$$\begin{aligned}\text{Var}(B) &= \text{Var}(E[B|S]) + E[\text{Var}(B|S)] \\ &= (1 + \lambda E[B])^2 \text{Var}(S) + \lambda E[S] E[B^2] \\ &= (1 + \lambda E[B])^2 \text{Var}(S) + \lambda E[S](\text{Var}(B) + E^2[B])\end{aligned}$$

yielding

$$\text{Var}(B) = \frac{\text{Var}(S)(1 + \lambda E[B])^2 + \lambda E[S] E^2[B]}{1 - \lambda E[S]}$$

Using $E[B] = E[S]/(1 - \lambda E[S])$, we obtain

$$\text{Var}(B) = \frac{\text{Var}(S) + \lambda E^3[S]}{(1 - \lambda E[S])^3}$$

■

There is a very nice representation of the compound Poisson process when the set of possible values of the Y_i is finite or countably infinite. So let us suppose that there are numbers α_j, $j \geq 1$, such that

$$P\{Y_i = \alpha_j\} = p_j, \quad \sum_j p_j = 1$$

Now, a compound Poisson process arises when events occur according to a Poisson process and each event results in a random amount Y being added to the cumulative sum. Let us say that the event is a type j event whenever it results in adding the amount α_j, $j \geq 1$. That is, the ith event of the Poisson process is a type j event if $Y_i = \alpha_j$. If we let $N_j(t)$ denote the number of type j events by time t, then it follows from Proposition 5.5 that the random variables $N_j(t)$, $j \geq 1$, are independent Poisson random variables with respective means

$$E[N_j(t)] = \lambda p_j t$$

Since, for each j, the amount α_j is added to the cumulative sum a total of $N_j(t)$ times by time t, it follows that the cumulative sum at time t can be expressed as

$$X(t) = \sum_j \alpha_j N_j(t) \tag{5.26}$$

As a check of Eq. (5.26), let us use it to compute the mean and variance of $X(t)$. This yields

$$\begin{aligned} E[X(t)] &= E\left[\sum_j \alpha_j N_j(t)\right] \\ &= \sum_j \alpha_j E[N_j(t)] \\ &= \sum_j \alpha_j \lambda p_j t \\ &= \lambda t\, E[Y_1] \end{aligned}$$

Also,

$$\begin{aligned} \mathrm{Var}[X(t)] &= \mathrm{Var}\left[\sum_j \alpha_j N_j(t)\right] \\ &= \sum_j \alpha_j^2 \,\mathrm{Var}[N_j(t)] \quad \text{by the independence of the } N_j(t),\ j \geq 1 \\ &= \sum_j \alpha_j^2 \,\lambda p_j t \\ &= \lambda t E[Y_1^2] \end{aligned}$$

where the next to last equality follows since the variance of the Poisson random variable $N_j(t)$ is equal to its mean.

Thus, we see that the representation (5.26) results in the same expressions for the mean and variance of $X(t)$ as were previously derived.

One of the uses of the representation (5.26) is that it enables us to conclude that as t grows large, the distribution of $X(t)$ converges to the normal distribution. To see why, note first that it follows by the central limit theorem that the distribution of a Poisson random variable converges to a normal distribution as its mean increases. (Why is this?) Therefore, each of the random variables $N_j(t)$ converges to a normal random variable as t increases. Because they are independent, and because the sum of independent normal random variables is also normal, it follows that $X(t)$ also approaches a normal distribution as t increases.

Example 5.28. In Example 5.26, find the approximate probability that at least 240 people migrate to the area within the next 50 weeks.

Solution: Since $\lambda = 2$, $E[Y_i] = 5/2$, $E[Y_i^2] = 43/6$, we see that

$$E[X(50)] = 250, \qquad \text{Var}[X(50)] = 4300/6$$

Now, the desired probability is

$$\begin{aligned} P\{X(50) \geq 240\} &= P\{X(50) \geq 239.5\} \\ &= P\left\{\frac{X(50) - 250}{\sqrt{4300/6}} \geq \frac{239.5 - 250}{\sqrt{4300/6}}\right\} \\ &= 1 - \phi(-0.3922) \\ &= \phi(0.3922) \\ &= 0.6525 \end{aligned}$$

where Table 2.3 was used to determine $\phi(0.3922)$, the probability that a standard normal is less than 0.3922. ■

Another useful result is that if $\{X(t), t \geq 0\}$ and $\{Y(t), t \geq 0\}$ are independent compound Poisson processes with respective Poisson parameters and distributions λ_1, F_1 and λ_2, F_2, then $\{X(t) + Y(t), t \geq 0\}$ is also a compound Poisson process. This is true because in this combined process events will occur according to a Poisson process with rate $\lambda_1 + \lambda_2$, and each event independently will be from the first compound Poisson process with probability $\lambda_1/(\lambda_1 + \lambda_2)$. Consequently, the combined process will be a compound Poisson process with Poisson parameter $\lambda_1 + \lambda_2$, and with distribution function F given by

$$F(x) = \frac{\lambda_1}{\lambda_1 + \lambda_2} F_1(x) + \frac{\lambda_2}{\lambda_1 + \lambda_2} F_2(x)$$

5.4.3 Conditional or Mixed Poisson Processes

Let $\{N(t), t \geq 0\}$ be a counting process whose probabilities are defined as follows. There is a positive random variable L such that, conditional on $L = \lambda$, the counting process is a Poisson process with rate λ. Such a counting process is called a *conditional* or a *mixed* Poisson process.

Suppose that L is continuous with density function g. Because

$$\begin{aligned} P\{N(t+s) - N(s) = n\} &= \int_0^\infty P\{N(t+s) - N(s) = n \mid L = \lambda\} g(\lambda)\, d\lambda \\ &= \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} g(\lambda)\, d\lambda \end{aligned} \tag{5.27}$$

we see that a conditional Poisson process has stationary increments. However, because knowing how many events occur in an interval gives information about the possible value of L, which affects the distribution of the number of events in any other interval, it follows that a conditional Poisson process does not generally have independent increments. Consequently, a conditional Poisson process is not generally a Poisson process.

Example 5.29. If g is the gamma density with parameters m and θ,

$$g(\lambda) = \theta e^{-\theta\lambda} \frac{(\theta\lambda)^{m-1}}{(m-1)!}, \qquad \lambda > 0$$

then

$$\begin{aligned} P\{N(t) = n\} &= \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} \theta e^{-\theta\lambda} \frac{(\theta\lambda)^{m-1}}{(m-1)!}\, d\lambda \\ &= \frac{t^n \theta^m}{n!(m-1)!} \int_0^\infty e^{-(t+\theta)\lambda} \lambda^{n+m-1}\, d\lambda \end{aligned}$$

Multiplying and dividing by $\frac{(n+m-1)!}{(t+\theta)^{n+m}}$ gives

$$P\{N(t) = n\} = \frac{t^n \theta^m (n+m-1)!}{n!(m-1)!(t+\theta)^{n+m}} \int_0^\infty (t+\theta) e^{-(t+\theta)\lambda} \frac{((t+\theta)\lambda)^{n+m-1}}{(n+m-1)!}\, d\lambda$$

Because $(t+\theta)e^{-(t+\theta)\lambda}((t+\theta)\lambda)^{n+m-1}/(n+m-1)!$ is the density function of a gamma $(n+m, t+\theta)$ random variable, its integral is 1, giving the result

$$P\{N(t) = n\} = \binom{n+m-1}{n} \left(\frac{\theta}{t+\theta}\right)^m \left(\frac{t}{t+\theta}\right)^n$$

Therefore, the number of events in an interval of length t has the same distribution of the number of failures that occur before a total of m successes are amassed, when each trial is a success with probability $\frac{\theta}{t+\theta}$. ■

To compute the mean and variance of $N(t)$, condition on L. Because, conditional on L, $N(t)$ is Poisson with mean Lt, we obtain

$$\begin{aligned} E[N(t)|L] &= Lt \\ \text{Var}(N(t)|L) &= Lt \end{aligned}$$

where the final equality used that the variance of a Poisson random variable is equal to its mean. Consequently, the conditional variance formula yields

$$\begin{aligned} \text{Var}(N(t)) &= E[Lt] + \text{Var}(Lt) \\ &= tE[L] + t^2\text{Var}(L) \end{aligned}$$

We can compute the conditional distribution function of L, given that $N(t) = n$, as follows.

$$\begin{aligned} P\{L \le x|N(t) = n\} &= \frac{P\{L \le x,\ N(t) = n\}}{P\{N(t) = n\}} \\ &= \frac{\int_0^\infty P\{L \le x, N(t) = n|L = \lambda\}g(\lambda)\,d\lambda}{P\{N(t) = n\}} \\ &= \frac{\int_0^x P\{N(t) = n|L = \lambda\}g(\lambda)\,d\lambda}{P\{N(t) = n\}} \\ &= \frac{\int_0^x e^{-\lambda t}(\lambda t)^n g(\lambda)\,d\lambda}{\int_0^\infty e^{-\lambda t}(\lambda t)^n g(\lambda)\,d\lambda} \end{aligned}$$

where the final equality used Eq. (5.27). In other words, the conditional density function of L given that $N(t) = n$ is

$$f_{L|N(t)}(\lambda \mid n) = \frac{e^{-\lambda t}\lambda^n\, g(\lambda)}{\int_0^\infty e^{-\lambda t}\lambda^n\, g(\lambda)\,d\lambda}, \quad \lambda \ge 0 \tag{5.28}$$

Example 5.30. An insurance company feels that each of its policyholders has a rating value and that a policyholder having rating value λ will make claims at times distributed according to a Poisson process with rate λ, when time is measured in years. The firm also believes that rating values vary from policyholder to policyholder, with the probability distribution of the value of a new policyholder being uniformly distributed over $(0, 1)$. Given that a policyholder has made n claims in his or her first t years, what is the conditional distribution of the time until the policyholder's next claim?

Solution: If T is the time until the next claim, then we want to compute $P\{T > x \mid N(t) = n\}$. Conditioning on the policyholder's rating value gives, upon using Eq. (5.28),

$$P\{T > x \mid N(t) = n\} = \int_0^\infty P\{T > x \mid L = \lambda,\ N(t) = n\}$$

$$\times f_{L|N(t)}(\lambda \mid n)\, d\lambda$$
$$= \frac{\int_0^1 e^{-\lambda x} e^{-\lambda t} \lambda^n \, d\lambda}{\int_0^1 e^{-\lambda t} \lambda^n \, d\lambda} \qquad \blacksquare$$

There is a nice formula for the probability that more than n events occur in an interval of length t. In deriving it we will use the identity

$$\sum_{j=n+1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} = \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!} \, dx \tag{5.29}$$

which follows by noting that it equates the probability that the number of events by time t of a Poisson process with rate λ is greater than n with the probability that the time of the $(n+1)$st event of this process (which has a gamma $(n+1, \lambda)$ distribution) is less than t. Interchanging λ and t in Eq. (5.29) yields the equivalent identity

$$\sum_{j=n+1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} = \int_0^{\lambda} t e^{-tx} \frac{(tx)^n}{n!} \, dx \tag{5.30}$$

Using Eq. (5.27) we now have

$$\begin{aligned}
P\{N(t) > n\} &= \sum_{j=n+1}^{\infty} \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} g(\lambda)\, d\lambda \\
&= \int_0^{\infty} \sum_{j=n+1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} g(\lambda)\, d\lambda \qquad \text{(by interchanging)} \\
&= \int_0^{\infty} \int_0^{\lambda} t e^{-tx} \frac{(tx)^n}{n!} \, dx g(\lambda)\, d\lambda \qquad \text{(using (5.30))} \\
&= \int_0^{\infty} \int_x^{\infty} g(\lambda)\, d\lambda t e^{-tx} \frac{(tx)^n}{n!} \, dx \qquad \text{(by interchanging)} \\
&= \int_0^{\infty} \bar{G}(x) t e^{-tx} \frac{(tx)^n}{n!} \, dx
\end{aligned}$$

5.5 Random Intensity Functions and Hawkes Processes

Whereas the intensity function $\lambda(t)$ of a nonhomogeneous Poisson process is a deterministic function, there are counting processes $\{N(t),\ t \geq 0\}$ whose intensity function value at time t, call it $R(t)$, is a random variable whose value depends on the history of the process up to time t. That is, if we let $\mathcal{H}_t$ denote the "history" of the process up to time t then $R(t)$, the intensity rate at time t, is a random variable whose value is determined by $\mathcal{H}_t$ and which is such that

$$P(N(t+h) - N(t) = 1 | \mathcal{H}_t) = R(t)h + o(h)$$

and

$$P(N(t+h) - N(t) \geq 2 | \mathcal{H}_t) = o(h)$$

The *Hawkes process* is an example of a counting process having a random intensity function. This counting process assumes that there is a base intensity value $\lambda > 0$, and that associated with each event is a nonnegative random variable, called a mark, whose value is independent of all that has previously occurred and has distribution F. Whenever an event occurs, it is supposed that the current value of the random intensity function increases by the amount of the event's mark, with this increase decreasing over time at an exponential rate α. More specifically, if there have been a total of $N(t)$ events by time t, with $S_1 < S_2 < \ldots < S_{N(t)}$ being the event times and M_i being the mark of event $i, i = 1, \ldots, N(t)$, then

$$R(t) = \lambda + \sum_{i=1}^{N(t)} M_i e^{-\alpha(t-S_i)}$$

In other words, a Hawkes process is a counting process in which

1. $R(0) = \lambda$;
2. whenever an event occurs, the random intensity increases by the value of the event's mark;
3. if there are no events between s and $s + t$ then $R(s + t) = \lambda + (R(s) - \lambda)e^{-\alpha t}$.

Because the intensity increases each time an event occurs, the Hawkes process is said to be a *self-exciting* process.

We will derive $E[N(t)]$, the expected number of events of a Hawkes process that occur by time t. To do so, we will need the following lemma, which is valid for all counting processes.

Lemma 5.5. *Let $R(t), t \geq 0$ be the random intensity function of the counting process $\{N(t), t \geq 0\}$ having $N(0) = 0$. Then, with $m(t) = E[N(t)]$*

$$m(t) = \int_0^t E[R(s)]\, ds$$

Proof.

$$E[N(t+h) | N(t), R(t)] = N(t) + R(t)h + o(h)$$

Taking expectations gives

$$E[N(t+h)] = E[N(t)] + E[R(t)]h + o(h)$$

That is,

$$m(t+h) = m(t) + hE[R(t)] + o(h)$$

or

$$\frac{m(t+h)-m(t)}{h} = E[R(t)] + \frac{o(h)}{h}$$

Letting h go to 0 gives

$$m'(t) = E[R(t)]$$

Integrating both sides from 0 to t now gives the result:

$$m(t) = \int_0^t E[R(s)]\,ds$$ ■

Using the preceding, we can now prove the following proposition.

Proposition 5.8. *If μ is the expected value of a mark in a Hawkes process, then for this process*

$$E[N(t)] = \lambda t + \frac{\lambda\mu}{(\mu-\alpha)^2}(e^{(\mu-\alpha)t} - 1 - (\mu-\alpha)t)$$

Proof. To determine the mean value function $m(t)$ it suffices, by the preceding lemma, to determine $E[R(t)]$, which will be accomplished by deriving and then solving a differential equation. To begin note that, with $M_t(h)$ equal to the sum of the marks of all events occurring between t and $t+h$,

$$R(t+h) = \lambda + (R(t)-\lambda)e^{-\alpha h} + M_t(h) + o(h)$$

Letting $g(t) = E[R(t)]$ and taking expectations of the preceding gives

$$g(t+h) = \lambda + (g(t)-\lambda)e^{-\alpha h} + E[M_t(h)] + o(h)$$

Using the identity $e^{-\alpha h} = 1 - \alpha h + o(h)$ shows that

$$\begin{aligned} g(t+h) &= \lambda + (g(t)-\lambda)(1-\alpha h) + E[M_t(h)] + o(h) \\ &= g(t) - \alpha h g(t) + \lambda\alpha h + E[M_t(h)] + o(h) \end{aligned} \tag{5.31}$$

Now, given $R(t)$, there will be 1 event between t and $t+h$ with probability $R(t)h + o(h)$, and there will be 2 or more with probability $o(h)$. Hence, conditioning on the number of events between t and $t+h$ yields, upon using that μ is the expected value of a mark, that

$$E[M_t(h)|R(t)] = \mu R(t)h + o(h)$$

Taking expectations of both sides of the preceding gives that

$$E[M_t(h)] = \mu g(t)h + o(h)$$

Substituting back into Eq. (5.31) gives

$$g(t+h) = g(t) - \alpha h g(t) + \lambda\alpha h + \mu g(t)h + o(h)$$

or, equivalently,

$$\frac{g(t+h) - g(t)}{h} = (\mu - \alpha)g(t) + \lambda\alpha + \frac{o(h)}{h}$$

Letting h go to 0 gives that

$$g'(t) = (\mu - \alpha)g(t) + \lambda\alpha$$

Letting $f(t) = (\mu - \alpha)g(t) + \lambda\alpha$, the preceding can be written as

$$\frac{f'(t)}{\mu - \alpha} = f(t)$$

or

$$\frac{f'(t)}{f(t)} = \mu - \alpha$$

Integration now yields

$$\log(f(t)) = (\mu - \alpha)t + C$$

Now, $g(0) = E[R(0)] = \lambda$ and so $f(0) = \mu\lambda$, showing that $C = \log(\mu\lambda)$ and giving the result

$$f(t) = \mu\lambda e^{(\mu-\alpha)t}$$

Using that $g(t) = \frac{f(t)-\lambda\alpha}{\mu-\alpha} = \frac{f(t)}{\mu-\alpha} + \lambda - \frac{\lambda\mu}{\mu-\alpha}$ gives

$$g(t) = \lambda + \frac{\lambda\mu}{\mu - \alpha}(e^{(\mu-\alpha)t} - 1)$$

Hence, from Lemma 5.5

$$\begin{aligned} E[N(t)] &= \lambda t + \int_0^t \frac{\lambda\mu}{\mu - \alpha}(e^{(\mu-\alpha)s} - 1)\, ds \\ &= \lambda t + \frac{\lambda\mu}{(\mu - \alpha)^2}(e^{(\mu-\alpha)t} - 1 - (\mu - \alpha)t) \end{aligned}$$

and the result is proved. ■

Exercises

1. The time T required to repair a machine is an exponentially distributed random variable with mean $\frac{1}{2}$ (hours).
   (a) What is the probability that a repair time exceeds $\frac{1}{2}$ hour?
   (b) What is the probability that a repair takes at least $12\frac{1}{2}$ hours given that its duration exceeds 12 hours?

2. Suppose that you arrive at a single-teller bank to find five other customers in the bank, one being served and the other four waiting in line. You join the end of the line. If the service times are all exponential with rate μ, what is the expected amount of time you will spend in the bank?

3. Let X be an exponential random variable. Without any computations, tell which one of the following is correct. Explain your answer.
   (a) $E[X^2|X > 1] = E[(X+1)^2]$
   (b) $E[X^2|X > 1] = E[X^2] + 1$
   (c) $E[X^2|X > 1] = (1 + E[X])^2$

4. Consider a post office with two clerks. Three people, A, B, and C, enter simultaneously. A and B go directly to the clerks, and C waits until either A or B leaves before he begins service. What is the probability that A is still in the post office after the other two have left when
   (a) the service time for each clerk is exactly (nonrandom) ten minutes?
   (b) the service times are i with probability $\frac{1}{3}, i = 1, 2, 3$?
   (c) the service times are exponential with mean $1/\mu$?

***5.** If X is exponential with rate λ, show that $Y = [X] + 1$ is geometric with parameter $p = 1 - e^{-\lambda}$, where $[x]$ is the largest integer less than or equal to x.

6. In Example 5.3 if server i serves at an exponential rate $\lambda_i, i = 1, 2$, show that

$$P\{\text{Smith is not last}\} = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^2$$

***7.** If X_1 and X_2 are independent nonnegative continuous random variables, show that

$$P\{X_1 < X_2 | \min(X_1, X_2) = t\} = \frac{r_1(t)}{r_1(t) + r_2(t)}$$

where $r_i(t)$ is the failure rate function of X_i.

8. If X and Y are independent exponential random variables with respective rates λ and μ, what is the conditional distribution of X given that $X < Y$?

9. Machine 1 is currently working. Machine 2 will be put in use at a time t from now. If the lifetime of machine i is exponential with rate $\lambda_i, i = 1, 2$, what is the probability that machine 1 is the first machine to fail?

***10.** Let X and Y be independent exponential random variables with respective rates λ and μ. Let $M = \min(X, Y)$. Find
   (a) $E[MX|M = X]$,

(b) $E[MX|M = Y]$,
(c) $\text{Cov}(X, M)$.

11. Let $X, Y_1, \ldots, Y_n$ be independent exponential random variables; X having rate λ, and Y_i having rate μ. Let A_j be the event that the jth smallest of these $n + 1$ random variables is one of the Y_i. Find $p = P\{X > \max_i Y_i\}$, by using the identity

$$p = P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cdots A_{n-1})$$

Verify your answer when $n = 2$ by conditioning on X to obtain p.

12. If $X_i, i = 1, 2, 3$, are independent exponential random variables with rates λ_i, $i = 1, 2, 3$, find
(a) $P\{X_1 < X_2 < X_3\}$,
(b) $P\{X_1 < X_2 | \max(X_1, X_2, X_3) = X_3\}$,
(c) $E[\max X_i | X_1 < X_2 < X_3]$,
(d) $E[\max X_i]$.

13. Find, in Example 5.10, the expected time until the nth person on line leaves the line (either by entering service or departing without service).

14. I am waiting for two friends to arrive at my house. The time until A arrives is exponentially distributed with rate λ_a, and the time until B arrives is exponentially distributed with rate λ_b. Once they arrive, both will spend exponentially distributed times, with respective rates μ_a and μ_b at my home before departing. The four exponential random variables are independent.
(a) What is the probability that A arrives before and departs after B?
(b) What is the expected time of the last departure?

15. One hundred items are simultaneously put on a life test. Suppose the lifetimes of the individual items are independent exponential random variables with mean 200 hours. The test will end when there have been a total of 5 failures. If T is the time at which the test ends, find $E[T]$ and $\text{Var}(T)$.

16. There are three jobs that need to be processed, with the processing time of job i being exponential with rate μ_i. There are two processors available, so processing on two of the jobs can immediately start, with processing on the final job to start when one of the initial ones is finished.
(a) Let T_i denote the time at which the processing of job i is completed. If the objective is to minimize $E[T_1 + T_2 + T_3]$, which jobs should be initially processed if $\mu_1 < \mu_2 < \mu_3$?
(b) Let M, called the *makespan*, be the time until all three jobs have been processed. With S equal to the time that there is only a single processor working, show that

$$2E[M] = E[S] + \sum_{i=1}^{3} 1/\mu_i$$

For the rest of this problem, suppose that $\mu_1 = \mu_2 = \mu$, $\mu_3 = \lambda$. Also, let $P(\mu)$ be the probability that the last job to finish is either job 1 or job

2, and let $P(\lambda) = 1 - P(\mu)$ be the probability that the last job to finish is job 3.

(c) Express $E[S]$ in terms of $P(\mu)$ and $P(\lambda)$.
Let $P_{i,j}(\mu)$ be the value of $P(\mu)$ when i and j are the jobs that are initially started.

(d) Show that $P_{1,2}(\mu) \leq P_{1,3}(\mu)$.

(e) If $\mu > \lambda$ show that $E[M]$ is minimized when job 3 is one of the jobs that is initially started.

(f) If $\mu < \lambda$ show that $E[M]$ is minimized when processing is initially started on jobs 1 and 2.

17. A set of n cities is to be connected via communication links. The cost to construct a link between cities i and j is $C_{ij}, i \neq j$. Enough links should be constructed so that for each pair of cities there is a path of links that connects them. As a result, only $n - 1$ links need be constructed. A minimal cost algorithm for solving this problem (known as the minimal spanning tree problem) first constructs the cheapest of all the $\binom{n}{2}$ links. Then, at each additional stage it chooses the cheapest link that connects a city without any links to one with links. That is, if the first link is between cities 1 and 2, then the second link will either be between 1 and one of the links $3, \ldots, n$ or between 2 and one of the links $3, \ldots, n$. Suppose that all of the $\binom{n}{2}$ costs C_{ij} are independent exponential random variables with mean 1. Find the expected cost of the preceding algorithm if

(a) $n = 3$,

(b) $n = 4$.

***18.** Let X_1 and X_2 be independent exponential random variables, each having rate μ. Let

$$X_{(1)} = \text{minimum}(X_1, X_2) \quad \text{and} \quad X_{(2)} = \text{maximum}(X_1, X_2)$$

Find

(a) $E[X_{(1)}]$,

(b) $\text{Var}[X_{(1)}]$,

(c) $E[X_{(2)}]$,

(d) $\text{Var}[X_{(2)}]$.

19. In a mile race between A and B, the time it takes A to complete the mile is an exponential random variable with rate λ_a and is independent of the time it takes B to complete the mile, which is an exponential random variable with rate λ_b. The one who finishes earliest is declared the winner and receives $Re^{-\alpha t}$ if the winning time is t, where R and α are constants. If the loser receives 0, find the expected amount that runner A wins.

20. Consider a two-server system in which a customer is served first by server 1, then by server 2, and then departs. The service times at server i are exponential random variables with rates $\mu_i, i = 1, 2$. When you arrive, you find server 1 free and two customers at server 2—customer A in service and customer B waiting in line.

(a) Find P_A, the probability that A is still in service when you move over to server 2.
(b) Find P_B, the probability that B is still in the system when you move over to server 2.
(c) Find $E[T]$, where T is the time that you spend in the system.

Hint: Write

$$T = S_1 + S_2 + W_A + W_B$$

where S_i is your service time at server i, W_A is the amount of time you wait in queue while A is being served, and W_B is the amount of time you wait in queue while B is being served.

21. In a certain system, a customer must first be served by server 1 and then by server 2. The service times at server i are exponential with rate $\mu_i, i = 1, 2$. An arrival finding server 1 busy waits in line for that server. Upon completion of service at server 1, a customer either enters service with server 2 if that server is free or else remains with server 1 (blocking any other customer from entering service) until server 2 is free. Customers depart the system after being served by server 2. Suppose that when you arrive there is one customer in the system and that customer is being served by server 1. What is the expected total time you spend in the system?

22. Suppose in Exercise 21 you arrive to find two others in the system, one being served by server 1 and one by server 2. What is the expected time you spend in the system? Recall that if server 1 finishes before server 2, then server 1's customer will remain with him (thus blocking your entrance) until server 2 becomes free.

***23.** A flashlight needs two batteries to be operational. Consider such a flashlight along with a set of n functional batteries—battery 1, battery 2, ..., battery n. Initially, battery 1 and 2 are installed. Whenever a battery fails, it is immediately replaced by the lowest numbered functional battery that has not yet been put in use. Suppose that the lifetimes of the different batteries are independent exponential random variables each having rate μ. At a random time, call it T, a battery will fail and our stockpile will be empty. At that moment exactly one of the batteries—which we call battery X—will not yet have failed.
(a) What is $P\{X = n\}$?
(b) What is $P\{X = 1\}$?
(c) What is $P\{X = i\}$?
(d) Find $E[T]$.
(e) What is the distribution of T?

24. There are two servers available to process n jobs. Initially, each server begins work on a job. Whenever a server completes work on a job, that job leaves the system and the server begins processing a new job (provided there are still jobs waiting to be processed). Let T denote the time until all jobs have been processed. If the time that it takes server i to process a job is exponentially distributed with rate μ_i, $i = 1, 2$, find $E[T]$ and $\mathrm{Var}(T)$.

25. Customers can be served by any of three servers, where the service times of server i are exponentially distributed with rate μ_i, $i = 1, 2, 3$. Whenever a server becomes free, the customer who has been waiting the longest begins service with that server.
- **(a)** If you arrive to find all three servers busy and no one waiting, find the expected time until you depart the system.
- **(b)** If you arrive to find all three servers busy and one person waiting, find the expected time until you depart the system.

26. Each entering customer must be served first by server 1, then by server 2, and finally by server 3. The amount of time it takes to be served by server i is an exponential random variable with rate μ_i, $i = 1, 2, 3$. Suppose you enter the system when it contains a single customer who is being served by server 3.
- **(a)** Find the probability that server 3 will still be busy when you move over to server 2.
- **(b)** Find the probability that server 3 will still be busy when you move over to server 3.
- **(c)** Find the expected amount of time that you spend in the system. (Whenever you encounter a busy server, you must wait for the service in progress to end before you can enter service.)
- **(d)** Suppose that you enter the system when it contains a single customer who is being served by server 2. Find the expected amount of time that you spend in the system.

27. Show, in Example 5.7, that the distributions of the total cost are the same for the two algorithms.

28. Consider n components with independent lifetimes, which are such that component i functions for an exponential time with rate λ_i. Suppose that all components are initially in use and remain so until they fail.
- **(a)** Find the probability that component 1 is the second component to fail.
- **(b)** Find the expected time of the second failure.

29. Let X and Y be independent exponential random variables with respective rates λ and μ, where $\lambda > \mu$. Let $c > 0$.
- **(a)** Show that the conditional density function of X, given that $X + Y = c$, is

$$f_{X|X+Y}(x|c) = \frac{(\lambda - \mu)e^{-(\lambda-\mu)x}}{1 - e^{-(\lambda-\mu)c}}, \qquad 0 < x < c$$

- **(b)** Use part (a) to find $E[X|X + Y = c]$.
- **(c)** Find $E[Y|X + Y = c]$.

30. The lifetimes of A's dog and cat are independent exponential random variables with respective rates λ_d and λ_c. One of them has just died. Find the expected additional lifetime of the other pet.

31. Suppose $W, X_1, \ldots, X_n$ are independent nonnegative continuous random variables, with W being exponential with rate λ, and with X_i having density function f_i, $i = 1, \ldots, n$.

(a) Show that

$$P(X_i < x_i | W > X_i) = \frac{\int_0^{x_i} e^{-\lambda s} f_i(s) ds}{P(W > X_i)}$$

(b) Show that

$$P(W > \sum_{i=1}^{n} X_i) = \prod_{i=1}^{n} P(W > X_i)$$

(c) Show that

$$P(X_i \le x_i, i = 1, \ldots, n | W > \sum_{i=1}^{n} X_i) = \prod_{i=1}^{n} P(X_i \le x_i | W > X_i)$$

That is, given that $W > \sum_{i=1}^{n} X_i$, the random variables $X_1, \ldots, X_n$ are independent with X_i now being distributed according to its conditional distribution given that it is less than W, $i = 1, \ldots, n$.

32. Let X be a uniform random variable on (0, 1), and consider a counting process where events occur at times $X + i$, for $i = 0, 1, 2, \ldots$.

(a) Does this counting process have independent increments?

(b) Does this counting process have stationary increments?

33. Let X and Y be independent exponential random variables with respective rates λ and μ.

(a) Argue that, conditional on $X > Y$, the random variables $\min(X, Y)$ and $X - Y$ are independent.

(b) Use part (a) to conclude that for any positive constant c

$$\begin{aligned} E[\min(X, Y) | X > Y + c] &= E[\min(X, Y) | X > Y] \\ &= E[\min(X, Y)] = \frac{1}{\lambda + \mu} \end{aligned}$$

(c) Give a verbal explanation of why $\min(X, Y)$ and $X - Y$ are (unconditionally) independent.

34. Two individuals, A and B, both require kidney transplants. If she does not receive a new kidney, then A will die after an exponential time with rate μ_A, and B after an exponential time with rate μ_B. New kidneys arrive in accordance with a Poisson process having rate λ. It has been decided that the first kidney will go to A (or to B if B is alive and A is not at that time) and the next one to B (if still living).

(a) What is the probability that A obtains a new kidney?

(b) What is the probability that B obtains a new kidney?

(c) What is the probability that neither A nor B obtains a new kidney?

(d) What is the probability that both A and B obtain new kidneys?

35. Let T_1 be the time of the first event of $\{N(t), t \geq 0\}$, a Poisson process with rate λ. For another way to show that T_1 is exponential with rate λ, let $\lambda_{T_1}(t)$ be its failure rate function. Using that

$$P(t < T_1 < t + h | T_1 > t) = \lambda_{T_1}(t)h + o(h)$$

show that T_1 is exponential with rate λ.

Hint: Write $P(t < T_1 < t + h | T_1 > t)$ as a conditional probability involving random variables $N(t)$ and $N(t + h)$.

***36.** Let $S(t)$ denote the price of a security at time t. A popular model for the process $\{S(t), t \geq 0\}$ supposes that the price remains unchanged until a "shock" occurs, at which time the price is multiplied by a random factor. If we let $N(t)$ denote the number of shocks by time t, and let X_i denote the ith multiplicative factor, then this model supposes that

$$S(t) = S(0) \prod_{i=1}^{N(t)} X_i$$

where $\prod_{i=1}^{N(t)} X_i$ is equal to 1 when $N(t) = 0$. Suppose that the X_i are independent exponential random variables with rate μ; that $\{N(t), t \geq 0\}$ is a Poisson process with rate λ; that $\{N(t), t \geq 0\}$ is independent of the X_i; and that $S(0) = s$.

(a) Find $E[S(t)]$.

(b) Find $E[S^2(t)]$.

37. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. For $i \leq n$ and $s < t$,

(a) find $P(N(t) = n | N(s) = i)$;

(b) find $P(N(s) = i | N(t) = n)$.

38. Let $\{M_i(t), t \geq 0\}, i = 1, 2, 3$ be independent Poisson processes with respective rates $\lambda_i, i = 1, 2$, and set

$$N_1(t) = M_1(t) + M_2(t), \quad N_2(t) = M_2(t) + M_3(t)$$

The stochastic process $\{(N_1(t), N_2(t)), t \geq 0\}$ is called a bivariate Poisson process.

(a) Find $P\{N_1(t) = n, N_2(t) = m\}$.

(b) Find $\text{Cov}(N_1(t), N_2(t))$.

39. A certain scientific theory supposes that mistakes in cell division occur according to a Poisson process with rate 2.5 per year, and that an individual dies when 196 such mistakes have occurred. Assuming this theory, find

(a) the mean lifetime of an individual,

(b) the variance of the lifetime of an individual.

Also approximate

(c) the probability that an individual dies before age 67.2,

(d) the probability that an individual reaches age 90,

(e) the probability that an individual reaches age 100.

***40.** Show that if $\{N_i(t), t \geq 0\}$ are independent Poisson processes with rate λ_i, $i = 1, 2$, then $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda_1 + \lambda_2$ where $N(t) = N_1(t) + N_2(t)$.

41. In Exercise 40 what is the probability that the first event of the combined process is from the N_1 process?

42. Customers arrive to a single server system according to a Poisson process with rate λ. An arrival that finds the server idle immediately begins service; an arrival that finds the server busy waits. When the server completes a service it then simultaneously serves all those customers who are waiting. The time it takes to serve a group of size i is a random variable with density function g_i, $i \geq 1$. If X_n is the number of customers in the nth service batch, is $\{X_n, n \geq 0\}$ a Markov chain. If it is, give its transition probabilities; if it is not, tell why not.

43. Customers arrive at a two-server service station according to a Poisson process with rate λ. Whenever a new customer arrives, any customer that is in the system immediately departs. A new arrival enters service first with server 1 and then with server 2. If the service times at the servers are independent exponentials with respective rates μ_1 and μ_2, what proportion of entering customers completes their service with server 2?

44. Cars pass a certain street location according to a Poisson process with rate λ. A woman who wants to cross the street at that location waits until she can see that no cars will come by in the next T time units.

(a) Find the probability that her waiting time is 0.

(b) Find her expected waiting time.

Hint: Condition on the time of the first car.

45. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ that is independent of the nonnegative random variable T with mean μ and variance σ^2. Find

(a) $\text{Cov}(T, N(T))$,

(b) $\text{Var}(N(T))$.

46. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ that is independent of the sequence $X_1, X_2, \ldots$ of independent and identically distributed random variables with mean μ and variance σ^2. Find

$$\text{Cov}\left(N(t), \sum_{i=1}^{N(t)} X_i\right)$$

47. Consider a two-server parallel queuing system where customers arrive according to a Poisson process with rate λ, and where the service times are exponential with rate μ. Moreover, suppose that arrivals finding both servers busy immediately depart without receiving any service (such a customer is said to be lost), whereas those finding at least one free server immediately enter service and then depart when their service is completed.

(a) If both servers are presently busy, find the expected time until the next customer enters the system.

(b) Starting empty, find the expected time until both servers are busy.
(c) Find the expected time between two successive lost customers.

48. Consider an n-server parallel queuing system where customers arrive according to a Poisson process with rate λ, where the service times are exponential random variables with rate μ, and where any arrival finding all servers busy immediately departs without receiving any service. If an arrival finds all servers busy, find
(a) the expected number of busy servers found by the next arrival,
(b) the probability that the next arrival finds all servers free,
(c) the probability that the next arrival finds exactly i of the servers free.

49. Events occur according to a Poisson process with rate λ. Each time an event occurs, we must decide whether or not to stop, with our objective being to stop at the last event to occur prior to some specified time T, where $T > 1/\lambda$. That is, if an event occurs at time $t, 0 \le t \le T$, and we decide to stop, then we win if there are no additional events by time T, and we lose otherwise. If we do not stop when an event occurs and no additional events occur by time T, then we lose. Also, if no events occur by time T, then we lose. Consider the strategy that stops at the first event to occur after some fixed time $s, 0 \le s \le T$.
(a) Using this strategy, what is the probability of winning?
(b) What value of s maximizes the probability of winning?
(c) Show that one's probability of winning when using the preceding strategy with the value of s specified in part (b) is $1/e$.

50. The number of hours between successive train arrivals at the station is uniformly distributed on $(0, 1)$. Passengers arrive according to a Poisson process with rate 7 per hour. Suppose a train has just left the station. Let X denote the number of people who get on the next train. Find
(a) $E[X]$,
(b) $\mathrm{Var}(X)$.

51. If an individual has never had a previous automobile accident, then the probability he or she has an accident in the next h time units is $\beta h + o(h)$; on the other hand, if he or she has ever had a previous accident, then the probability is $\alpha h + o(h)$. Find the expected number of accidents an individual has by time t.

52. Teams 1 and 2 are playing a match. The teams score points according to independent Poisson processes with respective rates λ_1 and λ_2. If the match ends when one of the teams has scored k more points than the other, find the probability that team 1 wins.

Hint: Relate this to the gambler's ruin problem.

53. The water level of a certain reservoir is depleted at a constant rate of 1000 units daily. The reservoir is refilled by randomly occurring rainfalls. Rainfalls occur according to a Poisson process with rate 0.2 per day. The amount of water added to the reservoir by a rainfall is 5000 units with probability 0.8 or 8000 units with probability 0.2. The present water level is just slightly below 5000 units.

(a) What is the probability the reservoir will be empty after five days?
(b) What is the probability the reservoir will be empty sometime within the next ten days?

54. A viral linear DNA molecule of length, say, 1 is often known to contain a certain "marked position," with the exact location of this mark being unknown. One approach to locating the marked position is to cut the molecule by agents that break it at points chosen according to a Poisson process with rate λ. It is then possible to determine the fragment that contains the marked position. For instance, letting m denote the location on the line of the marked position, then if L_1 denotes the last Poisson event time before m (or 0 if there are no Poisson events in $[0, m]$), and R_1 denotes the first Poisson event time after m (or 1 if there are no Poisson events in $[m, 1]$), then it would be learned that the marked position lies between L_1 and R_1. Find
(a) $P\{L_1 = 0\}$,
(b) $P\{L_1 < x\},\ 0 < x < m$,
(c) $P\{R_1 = 1\}$,
(d) $P\{R_1 > x\},\ m < x < 1$.
By repeating the preceding process on identical copies of the DNA molecule, we are able to zero in on the location of the marked position. If the cutting procedure is utilized on n identical copies of the molecule, yielding the data $L_i, R_i, i = 1, \ldots, n$, then it follows that the marked position lies between L and R, where

$$L = \max_i L_i, \qquad R = \min_i R_i$$

(e) Find $E[R - L]$, and in doing so, show that $E[R - L] \sim \frac{2}{n\lambda}$.

55. Consider a single server queuing system where customers arrive according to a Poisson process with rate λ, service times are exponential with rate μ, and customers are served in the order of their arrival. Suppose that a customer arrives and finds $n - 1$ others in the system. Let X denote the number in the system at the moment that customer departs. Find the probability mass function of X.

56. An event independently occurs on each day with probability p. Let $N(n)$ denote the total number of events that occur on the first n days, and let T_r denote the day on which the rth event occurs.
(a) What is the distribution of $N(n)$?
(b) What is the distribution of T_1?
(c) What is the distribution of T_r?
(d) Given that $N(n) = r$, show that the set of r days on which events occurred has the same distribution as a random selection (without replacement) of r of the values $1, 2, \ldots, n$.

57. Each round played by a contestant is either a success with probability p or a failure with probability $1 - p$. If the round is a success, then a random amount of money having an exponential distribution with rate λ is won. If the round is a failure, then the contestant loses everything that had been accumulated up to that time and cannot play any additional rounds. After a successful round,

the contestant can either elect to quit playing and keep whatever has already been won or can elect to play another round. Suppose that a newly starting contestant plans on continuing to play until either her total winnings exceeds t or a failure occurs.

(a) What is the distribution of N, equal to the number of successful rounds that it would take until her fortune exceeds t?

(b) What is the probability the contestant will be successful in reaching a fortune of at least t?

(c) Given the contestant is successful, what is her expected winnings?

(d) What is the expected value of the contestant's winnings?

58. There are two types of claims that are made to an insurance company. Let $N_i(t)$ denote the number of type i claims made by time t, and suppose that $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ are independent Poisson processes with rates $\lambda_1 = 10$ and $\lambda_2 = 1$. The amounts of successive type 1 claims are independent exponential random variables with mean \$1000 whereas the amounts from type 2 claims are independent exponential random variables with mean \$5000. A claim for \$4000 has just been received; what is the probability it is a type 1 claim?

59. Cars pass an intersection according to a Poisson process with rate λ. There are 4 types of cars, and each passing car is, independently, type i with probability p_i, $\sum_{i=1}^{4} p_i = 1$.

(a) Find the probability that at least one of each of car types 1, 2, 3 but none of type 4 have passed by time t.

(b) Given that exactly 6 cars of type 1 or 2 passed by time t, find the probability that 4 of them were type 1.

60. People arrive according to a Poisson process with rate λ, with each person independently being equally likely to be either a man or a woman. If a woman (man) arrives when there is at least one man (woman) waiting, then the woman (man) departs with one of the waiting men (women). If there is no member of the opposite sex waiting upon a person's arrival, then that person waits. Let $X(t)$ denote the number waiting at time t. Argue that $E[X(t)] \approx .78\sqrt{2\lambda t}$ when t is large.

Hint: If Z is a standard normal random variable, then $E[|Z|] = \sqrt{2/\pi} \approx .78$.

61. A system has a random number of flaws that we will suppose is Poisson distributed with mean c. Each of these flaws will, independently, cause the system to fail at a random time having distribution G. When a system failure occurs, suppose that the flaw causing the failure is immediately located and fixed.

(a) What is the distribution of the number of failures by time t?

(b) What is the distribution of the number of flaws that remain in the system at time t?

(c) Are the random variables in parts (a) and (b) dependent or independent?

62. Suppose that the number of typographical errors in a new text is Poisson distributed with mean λ. Two proofreaders independently read the text. Suppose that each error is independently found by proofreader i with probability

$p_i, i = 1, 2$. Let X_1 denote the number of errors that are found by proofreader 1 but not by proofreader 2. Let X_2 denote the number of errors that are found by proofreader 2 but not by proofreader 1. Let X_3 denote the number of errors that are found by both proofreaders. Finally, let X_4 denote the number of errors found by neither proofreader.

(a) Describe the joint probability distribution of X_1, X_2, X_3, X_4.

(b) Show that

$$\frac{E[X_1]}{E[X_3]} = \frac{1 - p_2}{p_2} \quad \text{and} \quad \frac{E[X_2]}{E[X_3]} = \frac{1 - p_1}{p_1}$$

Suppose now that λ, p_1, and p_2 are all unknown.

(c) By using X_i as an estimator of $E[X_i], i = 1, 2, 3$, present estimators of p_1, p_2, and λ.

(d) Give an estimator of X_4, the number of errors not found by either proofreader.

63. Consider an infinite server queuing system in which customers arrive in accordance with a Poisson process with rate λ, and where the service distribution is exponential with rate μ. Let $X(t)$ denote the number of customers in the system at time t. Find

(a) $E[X(t + s)|X(s) = n]$;

(b) $\text{Var}(X(t + s)|X(s) = n)$.

Hint: Divide the customers in the system at time $t + s$ into two groups, one consisting of "old" customers and the other of "new" customers.

(c) If there is currently a single customer in the system, find the probability that the system becomes empty when that customer departs.

***64.** Suppose that people arrive at a bus stop in accordance with a Poisson process with rate λ. The bus departs at time t. Let X denote the total amount of waiting time of all those who get on the bus at time t. We want to determine $\text{Var}(X)$. Let $N(t)$ denote the number of arrivals by time t.

(a) What is $E[X|N(t)]$?

(b) Argue that $\text{Var}(X|N(t)) = N(t)t^2/12$.

(c) What is $\text{Var}(X)$?

65. An average of 500 people pass the California bar exam each year. A California lawyer practices law, on average, for 30 years. Assuming these numbers remain steady, roughly how many lawyers would you expect California to have in 2050?

66. Policyholders of a certain insurance company have accidents at times distributed according to a Poisson process with rate λ. The amount of time from when the accident occurs until a claim is made has distribution G.

(a) Find the probability there are exactly n incurred but as yet unreported claims at time t.

(b) Suppose that each claim amount has distribution F, and that the claim amount is independent of the time that it takes to report the claim. Find the expected value of the sum of all incurred but as yet unreported claims at time t.

67. Satellites are launched into space at times distributed according to a Poisson process with rate λ. Each satellite independently spends a random time (having distribution G) in space before falling to the ground. Find the probability that none of the satellites in the air at time t was launched before time s, where $s < t$.

68. Suppose that electrical shocks having random amplitudes occur at times distributed according to a Poisson process $\{N(t), t \geq 0\}$ with rate λ. Suppose that the amplitudes of the successive shocks are independent both of other amplitudes and of the arrival times of shocks, and also that the amplitudes have distribution F with mean μ. Suppose also that the amplitude of a shock decreases with time at an exponential rate α, meaning that an initial amplitude A will have value $Ae^{-\alpha x}$ after an additional time x has elapsed. Let $A(t)$ denote the sum of all amplitudes at time t. That is,

$$A(t) = \sum_{i=1}^{N(t)} A_i e^{-\alpha(t-S_i)}$$

where A_i and S_i are the initial amplitude and the arrival time of shock i.

(a) Find $E[A(t)]$ by conditioning on $N(t)$.

(b) Without any computations, explain why $A(t)$ has the same distribution as does $D(t)$ of Example 5.21.

69. Suppose in Example 5.19 that a car can overtake a slower moving car without any loss of speed. Suppose a car that enters the road at time s has a free travel time equal to t_0. Find the distribution of the total number of other cars that it encounters on the road (either by passing or by being passed).

70. For the infinite server queue with Poisson arrivals and general service distribution G, find the probability that

(a) the first customer to arrive is also the first to depart.

Let $S(t)$ equal the sum of the remaining service times of all customers in the system at time t.

(b) Argue that $S(t)$ is a compound Poisson random variable.

(c) Find $E[S(t)]$.

(d) Find $\text{Var}(S(t))$.

71. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate $\lambda = 2$.

(a) Find $E[N(6)|N(4) = 4]$.

(b) Find $E[N(6)|N(10) = 12]$.

(c) Find $E[N(6)|N(4) = 4, N(10) = 12]$.

72. A cable car starts off with n riders. The times between successive stops of the car are independent exponential random variables with rate λ. At each stop one rider gets off. This takes no time, and no additional riders get on. After a rider gets off the car, he or she walks home. Independently of all else, the walk takes an exponential time with rate μ.

(a) What is the distribution of the time at which the last rider departs the car?

(b) Suppose the last rider departs the car at time t. What is the probability that all the other riders are home at that time?

73. Shocks occur according to a Poisson process with rate λ, and each shock independently causes a certain system to fail with probability p. Let T denote the time at which the system fails and let N denote the number of shocks that it takes.
- **(a)** Find the conditional distribution of T given that $N = n$.
- **(b)** Calculate the conditional distribution of N, given that $T = t$, and notice that it is distributed as 1 plus a Poisson random variable with mean $\lambda(1 - p)t$.
- **(c)** Explain how the result in part (b) could have been obtained without any calculations.

74. The number of missing items in a certain location, call it X, is a Poisson random variable with mean λ. When searching the location, each item will independently be found after an exponentially distributed time with rate μ. A reward of R is received for each item found, and a searching cost of C per unit of search time is incurred. Suppose that you search for a fixed time t and then stop.
- **(a)** Find your total expected return.
- **(b)** Find the value of t that maximizes the total expected return.
- **(c)** The policy of searching for a fixed time is a static policy. Would a dynamic policy, which allows the decision as to whether to stop at each time t, depend on the number already found by t be beneficial?

Hint: How does the distribution of the number of items not yet found by time t depend on the number already found by that time?

75. If $X_1, \ldots, X_n$ are independent exponential random variables with rate λ, find
- **(a)** $P(X_1 < x | X_1 + \ldots + X_n = t)$;
- **(b)** $P(\frac{X_1}{X_1+\ldots+X_n} \le x)$, $0 \le x \le 1$.

Hint: Interpret $X_1, \ldots, X_n$ as the interarrival times of a Poisson process.

76. For the model of Example 5.27, find the mean and variance of the number of customers served in a busy period.

77. Suppose that customers arrive to a system according to a Poisson process with rate λ. There are an infinite number of servers in this system so a customer begins service upon arrival. The service times of the arrivals are independent exponential random variables with rate μ, and are independent of the arrival process. Customers depart the system when their service ends. Let N be the number of arrivals before the first departure.
- **(a)** Find $P(N = 1)$.
- **(b)** Find $P(N = 2)$.
- **(c)** Find $P(N = j)$.
- **(d)** Find the probability that the first to arrive is the first to depart.
- **(e)** Find the expected time of the first departure.

78. A store opens at 8 A.M. From 8 until 10 A.M. customers arrive at a Poisson rate of four an hour. Between 10 A.M. and 12 P.M. they arrive at a Poisson rate of eight an hour. From 12 P.M. to 2 P.M. the arrival rate increases steadily from eight per hour at 12 P.M. to ten per hour at 2 P.M.; and from 2 to 5 P.M. the

arrival rate drops steadily from ten per hour at 2 P.M. to four per hour at 5 P.M. Determine the probability distribution of the number of customers that enter the store on a given day.

***79.** Suppose that events occur according to a nonhomogeneous Poisson process with intensity function $\lambda(t), t > 0$. Further, suppose that an event that occurs at time s is a type 1 event with probability $p(s), s > 0$. If $N_1(t)$ is the number of type 1 events by time t, what type of process is $\{N_1(t), t \geq 0\}$?

80. Let $T_1, T_2, \ldots$ denote the interarrival times of events of a nonhomogeneous Poisson process having intensity function $\lambda(t)$.

(a) Are the T_i independent?

(b) Are the T_i identically distributed?

(c) Find the distribution of T_1.

81. **(a)** Let $\{N(t), t \geq 0\}$ be a nonhomogeneous Poisson process with mean value function $m(t)$. Given $N(t) = n$, show that the unordered set of arrival times has the same distribution as n independent and identically distributed random variables having distribution function

$$F(x) = \begin{cases} \frac{m(x)}{m(t)}, & x \leq t \\ 1, & x \geq t \end{cases}$$

(b) Suppose that workmen incur accidents in accordance with a nonhomogeneous Poisson process with mean value function $m(t)$. Suppose further that each injured man is out of work for a random amount of time having distribution F. Let $X(t)$ be the number of workers who are out of work at time t. By using part (a), find $E[X(t)]$.

82. Let $X_1, X_2, \ldots$ be independent positive continuous random variables with a common density function f, and suppose this sequence is independent of N, a Poisson random variable with mean λ. Define

$$N(t) = \text{number of } i \leq N : X_i \leq t$$

Show that $\{N(t), t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t) = \lambda f(t)$.

83. Prove Lemma 5.4.

***84.** Let $X_1, X_2, \ldots$ be independent and identically distributed nonnegative continuous random variables having density function $f(x)$. We say that a record occurs at time n if X_n is larger than each of the previous values $X_1, \ldots, X_{n-1}$. (A record automatically occurs at time 1.) If a record occurs at time n, then X_n is called a *record value*. In other words, a record occurs whenever a new high is reached, and that new high is called the record value. Let $N(t)$ denote the number of record values that are less than or equal to t. Characterize the process $\{N(t), t \geq 0\}$ when

(a) f is an arbitrary continuous density function.

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

Hint: Finish the following sentence: There will be a record whose value is between t and $t + dt$ if the first X_i that is greater than t lies between ...

85. Let $X(t) = \sum_{i=1}^{N(t)} X_i$ where X_i, $i \geq 1$ are independent and identically distributed with mean $E[X]$, and are independent of $\{N(t), t \geq 0\}$, which is a Poisson process with rate λ. For $s < t$, find
(a) $E[X(t)|X(s)]$;
(b) $E[X(t)|N(s)]$;
(c) $\text{Var}(X(t)|N(s))$;
(d) $E[X(s)|N(t)]$.

86. In good years, storms occur according to a Poisson process with rate 3 per unit time, while in other years they occur according to a Poisson process with rate 5 per unit time. Suppose next year will be a good year with probability 0.3. Let $N(t)$ denote the number of storms during the first t time units of next year.
(a) Find $P\{N(t) = n\}$.
(b) Is $\{N(t)\}$ a Poisson process?
(c) Does $\{N(t)\}$ have stationary increments? Why or why not?
(d) Does it have independent increments? Why or why not?
(e) If next year starts off with three storms by time $t = 1$, what is the conditional probability it is a good year?

87. Determine

$$\text{Cov}(X(t), X(t+s))$$

when $\{X(t), t \geq 0\}$ is a compound Poisson process.

88. Customers arrive at the automatic teller machine in accordance with a Poisson process with rate 12 per hour. The amount of money withdrawn on each transaction is a random variable with mean \$30 and standard deviation \$50. (A negative withdrawal means that money was deposited.) The machine is in use for 15 hours daily. Approximate the probability that the total daily withdrawal is less than \$6000.

89. Some components of a two-component system fail after receiving a shock. Shocks of three types arrive independently and in accordance with Poisson processes. Shocks of the first type arrive at a Poisson rate λ_1 and cause the first component to fail. Those of the second type arrive at a Poisson rate λ_2 and cause the second component to fail. The third type of shock arrives at a Poisson rate λ_3 and causes both components to fail. Let X_1 and X_2 denote the survival times for the two components. Show that the joint distribution of X_1 and X_2 is given by

$$P\{X_1 > s, X_1 > t\} = \exp\{-\lambda_1 s - \lambda_2 t - \lambda_3 \max(s,t)\}$$

This distribution is known as the *bivariate exponential distribution.*

90. In Exercise 89 show that X_1 and X_2 both have exponential distributions.

***91.** Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed exponential random variables. Show that the probability that the largest of them is greater than the sum of the others is $n/2^{n-1}$. That is, if

$$M = \max_j X_j$$

then show

$$P\left\{M > \sum_{i=1}^{n} X_i - M\right\} = \frac{n}{2^{n-1}}$$

Hint: What is $P\{X_1 > \sum_{i=2}^{n} X_i\}$?

92. Prove Eq. (5.22).

93. Prove that

(a) $\max(X_1, X_2) = X_1 + X_2 - \min(X_1, X_2)$ and, in general,

(b)
$$\max(X_1, \ldots, X_n) = \sum_{1}^{n} X_i - \sum\sum_{i<j} \min(X_i, X_j) + \sum\sum\sum_{i<j<k} \min(X_i, X_j, X_k) + \cdots + (-1)^{n-1} \min(X_i, X_j, \ldots, X_n)$$

(c) Show by defining appropriate random variables X_i, $i = 1, \ldots, n$, and by taking expectations in part (b) how to obtain the well-known formula

$$P\left(\bigcup_{1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum\sum_{i<j} P(A_i A_j) + \cdots + (-1)^{n-1} P(A_1 \cdots A_n)$$

(d) Consider n independent Poisson processes—the ith having rate λ_i. Derive an expression for the expected time until an event has occurred in all n processes.

94. A two-dimensional Poisson process is a process of randomly occurring events in the plane such that

(i) for any region of area A the number of events in that region has a Poisson distribution with mean λA, and

(ii) the number of events in nonoverlapping regions are independent.

For such a process, consider an arbitrary point in the plane and let X denote its distance from its nearest event (where distance is measured in the usual Euclidean manner). Show that

(a) $P\{X > t\} = e^{-\lambda \pi t^2}$,

(b) $E[X] = \frac{1}{2\sqrt{\lambda}}$.

95. Let $\{N(t), t \geq 0\}$ be a conditional Poisson process with a random rate L.

(a) Derive an expression for $E[L|N(t) = n]$.

(b) Find, for $s > t$, $E[N(s)|N(t) = n]$.

(c) Find, for $s < t$, $E[N(s)|N(t) = n]$.

96. For the conditional Poisson process, let $m_1 = E[L]$, $m_2 = E[L^2]$. In terms of m_1 and m_2, find $\mathrm{Cov}(N(s), N(t))$ for $s \leq t$.

97. Consider a conditional Poisson process in which the rate L is, as in Example 5.29, gamma distributed with parameters m and p. Find the conditional density function of L given that $N(t) = n$.

98. Let $M(t) = E[D(t)]$ in Example 5.21.

(a) Show that

$$M(t+h) = M(t) + e^{-\alpha t}\lambda h\mu + o(h)$$

(b) Use (a) to show that

$$M'(t) = \lambda\mu e^{-\alpha t}$$

(c) Show that

$$M(t) = \frac{\lambda\mu}{\alpha}(1 - e^{-\alpha t})$$

99. Let X be the time between the first and the second event of a Hawkes process with mark distribution F. Find $P(X > t)$.

References

[1] H. Cramér, M. Leadbetter, Stationary and Related Stochastic Processes, John Wiley, New York, 1966.
[2] S. Ross, Stochastic Processes, Second Edition, John Wiley, New York, 1996.
[3] S. Ross, Probability Models for Computer Science, Academic Press, 2002.

Continuous-Time Markov Chains 6

6.1 Introduction

In this chapter we consider a class of probability models that has a wide variety of applications in the real world. The members of this class are the continuous-time analogs of the Markov chains of Chapter 4 and as such are characterized by the Markovian property that, given the present state, the future is independent of the past.

One example of a continuous-time Markov chain has already been met. This is the Poisson process of Chapter 5. For if we let the total number of arrivals by time t (that is, $N(t)$) be the state of the process at time t, then the Poisson process is a continuous-time Markov chain having states $0, 1, 2, \ldots$ that always proceeds from state n to state $n+1$, where $n \geq 0$. Such a process is known as a *pure birth process* since when a transition occurs the state of the system is always increased by one. More generally, an exponential model that can go (in one transition) only from state n to either state $n-1$ or state $n+1$ is called a *birth and death model*. For such a model, transitions from state n to state $n+1$ are designated as births, and those from n to $n-1$ as deaths. Birth and death models have wide applicability in the study of biological systems and in the study of waiting line systems in which the state represents the number of customers in the system. These models will be studied extensively in this chapter.

In Section 6.2 we define continuous-time Markov chains and then relate them to the discrete-time Markov chains of Chapter 4. In Section 6.3 we consider birth and death processes and in Section 6.4 we derive two sets of differential equations—the forward and backward equations—that describe the probability laws for the system. The material in Section 6.5 is concerned with determining the limiting (or long-run) probabilities connected with a continuous-time Markov chain. In Section 6.6 we consider the topic of time reversibility. We show that all birth and death processes are time reversible, and then illustrate the importance of this observation to queueing systems. In Section 6.7 we introduce the reverse chain, which has important applications even when the chain is not time reversible. The final two sections deal with uniformization and methods for numerically computing transition probabilities.

6.2 Continuous-Time Markov Chains

Suppose we have a continuous-time stochastic process $\{X(t), t \geq 0\}$ taking on values in the set of nonnegative integers. In analogy with the definition of a discrete-time Markov chain, given in Chapter 4, we say that the process $\{X(t), \ t \geq 0\}$ is a *continuous-time Markov chain* if for all $s, t \geq 0$ and nonnegative integers $i, j, x(u), 0 \leq u < s$

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00011-1

$$P\{X(t+s)=j|X(s)=i,\ X(u)=x(u),\ 0\le u<s\}$$
$$=P\{X(t+s)=j|X(s)=i\}$$

In other words, a continuous-time Markov chain is a stochastic process having the Markovian property that the conditional distribution of the future $X(t+s)$ given the present $X(s)$ and the past $X(u), 0 \le u < s$, depends only on the present and is independent of the past. If, in addition,

$$P\{X(t+s)=j|X(s)=i\}$$

is independent of s, then the continuous-time Markov chain is said to have stationary or homogeneous transition probabilities.

All Markov chains considered in this text will be assumed to have stationary transition probabilities.

Suppose that a continuous-time Markov chain enters state i at some time, say, time 0, and suppose that the process does not leave state i (that is, a transition does not occur) during the next ten minutes. What is the probability that the process will not leave state i during the following five minutes? Since the process is in state i at time 10 it follows, by the Markovian property, that the probability that it remains in that state during the interval [10,15] is just the (unconditional) probability that it stays in state i for at least five minutes. That is, if we let T_i denote the amount of time that the process stays in state i before making a transition into a different state, then

$$P\{T_i > 15|T_i > 10\} = P\{T_i > 5\}$$

or, in general, by the same reasoning,

$$P\{T_i > s+t|T_i > s\} = P\{T_i > t\}$$

for all $s, t \ge 0$. Hence, the random variable T_i is *memoryless* and must thus (see Section 5.2.2) be *exponentially* distributed.

In fact, the preceding gives us another way of defining a continuous-time Markov chain. Namely, it is a stochastic process having the properties that each time it enters state i

(i) the amount of time it spends in that state before making a transition into a different state is exponentially distributed with mean, say, $1/v_i$, and

(ii) when the process leaves state i, it next enters state j with some probability, say, P_{ij}. Of course, the P_{ij} must satisfy

$$P_{ii}=0, \quad \text{all } i$$
$$\sum_j P_{ij}=1, \quad \text{all } i$$

In other words, a continuous-time Markov chain is a stochastic process that moves from state to state in accordance with a (discrete-time) Markov chain, but is such that the amount of time it spends in each state, before proceeding to the next state,

is exponentially distributed. In addition, the amount of time the process spends in state i, and the next state visited, must be independent random variables. For if the next state visited were dependent on T_i, then information as to how long the process has already been in state i would be relevant to the prediction of the next state—and this contradicts the Markovian assumption.

Example 6.1 (A Shoe Shine Shop)**.** Consider a shoe shine establishment consisting of two chairs—chair 1 and chair 2. A customer upon arrival goes initially to chair 1 where his shoes are cleaned and polish is applied. After this is done the customer moves on to chair 2 where the polish is buffed. The service times at the two chairs are assumed to be independent random variables that are exponentially distributed with respective rates μ_1 and μ_2. Suppose that potential customers arrive in accordance with a Poisson process having rate λ, and that a potential customer will enter the system only if both chairs are empty.

The preceding model can be analyzed as a continuous-time Markov chain, but first we must decide upon an appropriate state space. Since a potential customer will enter the system only if there are no other customers present, it follows that there will always either be 0 or 1 customers in the system. However, if there is 1 customer in the system, then we would also need to know which chair he was presently in. Hence, an appropriate state space might consist of the three states 0, 1, and 2 where the states have the following interpretation:

State	*Interpretation*
0	system is empty
1	a customer is in chair 1
2	a customer is in chair 2

We leave it as an exercise for you to verify that

$$v_0 = \lambda, \quad v_1 = \mu_1, \quad v_2 = \mu_2,$$
$$P_{01} = P_{12} = P_{20} = 1$$

■

6.3 Birth and Death Processes

Consider a system whose state at any time is represented by the number of people in the system at that time. Suppose that whenever there are n people in the system, then (i) new arrivals enter the system at an exponential rate λ_n, and (ii) people leave the system at an exponential rate μ_n. That is, whenever there are n persons in the system, then the time until the next arrival is exponentially distributed with mean $1/\lambda_n$ and is independent of the time until the next departure, which is itself exponentially distributed with mean $1/\mu_n$. Such a system is called a birth and death process. The parameters $\{\lambda_n\}_{n=0}^{\infty}$ and $\{\mu_n\}_{n=1}^{\infty}$ are called, respectively, the arrival (or birth) and departure (or death) rates.

Thus, a birth and death process is a continuous-time Markov chain with states $\{0, 1, \ldots\}$ for which transitions from state n may go only to either state $n - 1$ or state

$n+1$. The relationships between the birth and death rates and the state transition rates and probabilities are

$$\begin{aligned} v_0 &= \lambda_0, \\ v_i &= \lambda_i + \mu_i, \qquad i > 0 \\ P_{01} &= 1, \\ P_{i,i+1} &= \frac{\lambda_i}{\lambda_i + \mu_i}, \qquad i > 0 \\ P_{i,i-1} &= \frac{\mu_i}{\lambda_i + \mu_i}, \qquad i > 0 \end{aligned}$$

The preceding follows, because if there are i in the system, then the next state will be $i+1$ if a birth occurs before a death, and the probability that an exponential random variable with rate λ_i will occur earlier than an (independent) exponential with rate μ_i is $\lambda_i/(\lambda_i + \mu_i)$. Moreover, the time until either a birth or a death occurs is exponentially distributed with rate $\lambda_i + \mu_i$ (and so, $v_i = \lambda_i + \mu_i$).

Example 6.2 (The Poisson Process)**.** Consider a birth and death process for which

$$\begin{aligned} \mu_n &= 0, \quad \text{for all } n \geq 0 \\ \lambda_n &= \lambda, \quad \text{for all } n \geq 0 \end{aligned}$$

This is a process in which departures never occur, and the time between successive arrivals is exponential with mean $1/\lambda$. Hence, this is just the Poisson process. ■

A birth and death process for which $\mu_n = 0$ for all n is called a pure birth process. Another pure birth process is given by the next example.

Example 6.3 (A Birth Process with Linear Birth Rate)**.** Consider a population whose members can give birth to new members but cannot die. If each member acts independently of the others and takes an exponentially distributed amount of time, with mean $1/\lambda$, to give birth, then if $X(t)$ is the population size at time t, then $\{X(t), t \geq 0\}$ is a pure birth process with $\lambda_n = n\lambda, n \geq 0$. This follows since if the population consists of n persons and each gives birth at an exponential rate λ, then the total rate at which births occur is $n\lambda$. This pure birth process is known as a *Yule process* after G. Yule, who used it in his mathematical theory of evolution. ■

Example 6.4 (A Linear Growth Model with Immigration)**.** A model in which

$$\begin{aligned} \mu_n &= n\mu, \qquad n \geq 1 \\ \lambda_n &= n\lambda + \theta, \quad n \geq 0 \end{aligned}$$

is called a *linear growth process with immigration*. Such processes occur naturally in the study of biological reproduction and population growth. Each individual in the population is assumed to give birth at an exponential rate λ; in addition, there is an exponential rate of increase θ of the population due to an external source such as

immigration. Hence, the total birth rate where there are n persons in the system is $n\lambda + \theta$. Deaths are assumed to occur at an exponential rate μ for each member of the population, so $\mu_n = n\mu$.

Let $X(t)$ denote the population size at time t. Suppose that $X(0) = i$ and let

$$M(t) = E[X(t)]$$

We will determine $M(t)$ by deriving and then solving a differential equation that it satisfies.

We start by deriving an equation for $M(t+h)$ by conditioning on $X(t)$. This yields

$$\begin{aligned} M(t+h) &= E[X(t+h)] \\ &= E[E[X(t+h)|X(t)]] \end{aligned}$$

Now, given the size of the population at time t then, ignoring events whose probability is $o(h)$, the population at time $t+h$ will either increase in size by 1 if a birth or an immigration occurs in $(t, t+h)$, or decrease by 1 if a death occurs in this interval, or remain the same if neither of these two possibilities occurs. That is, given $X(t)$,

$$X(t+h) = \begin{cases} X(t)+1, & \text{with probability } [\theta + X(t)\lambda]h + o(h) \\ X(t)-1, & \text{with probability } X(t)\mu h + o(h) \\ X(t), & \text{with probability } 1 - [\theta + X(t)\lambda + X(t)\mu]h + o(h) \end{cases}$$

Therefore,

$$E[X(t+h)|X(t)] = X(t) + [\theta + X(t)\lambda - X(t)\mu]h + o(h)$$

Taking expectations yields

$$M(t+h) = M(t) + (\lambda - \mu)M(t)h + \theta h + o(h)$$

or, equivalently,

$$\frac{M(t+h) - M(t)}{h} = (\lambda - \mu)M(t) + \theta + \frac{o(h)}{h}$$

Taking the limit as $h \to 0$ yields the differential equation

$$M'(t) = (\lambda - \mu)M(t) + \theta \tag{6.1}$$

If we now define the function $h(t)$ by

$$h(t) = (\lambda - \mu)M(t) + \theta$$

then

$$h'(t) = (\lambda - \mu)M'(t)$$

Therefore, differential equation (6.1) can be rewritten as

$$\frac{h'(t)}{\lambda - \mu} = h(t)$$

or

$$\frac{h'(t)}{h(t)} = \lambda - \mu$$

Integration yields

$$\log[h(t)] = (\lambda - \mu)t + c$$

or

$$h(t) = Ke^{(\lambda-\mu)t}$$

Putting this back in terms of $M(t)$ gives

$$\theta + (\lambda - \mu)M(t) = Ke^{(\lambda-\mu)t}$$

To determine the value of the constant K, we use the fact that $M(0) = i$ and evaluate the preceding at $t = 0$. This gives

$$\theta + (\lambda - \mu)i = K$$

Substituting this back in the preceding equation for $M(t)$ yields the following solution for $M(t)$:

$$M(t) = \frac{\theta}{\lambda - \mu}[e^{(\lambda-\mu)t} - 1] + ie^{(\lambda-\mu)t}$$

Note that we have implicitly assumed that $\lambda \neq \mu$. If $\lambda = \mu$, then differential equation (6.1) reduces to

$$M'(t) = \theta \tag{6.2}$$

Integrating (6.2) and using that $M(0) = i$ gives the solution

$$M(t) = \theta t + i$$ ■

Example 6.5 (The Queueing System $M/M/1$)**.** Suppose that customers arrive at a single-server service station in accordance with a Poisson process having rate λ. That is, the times between successive arrivals are independent exponential random variables having mean $1/\lambda$. Upon arrival, each customer goes directly into service if the server is free; if not, then the customer joins the queue (that is, he waits in line). When the server finishes serving a customer, the customer leaves the system and the next

customer in line, if there are any waiting, enters the service. The successive service times are assumed to be independent exponential random variables having mean $1/\mu$.

The preceding is known as the $M/M/1$ queueing system. The first M refers to the fact that the interarrival process is Markovian (since it is a Poisson process) and the second to the fact that the service distribution is exponential (and, hence, Markovian). The 1 refers to the fact that there is a single server.

If we let $X(t)$ denote the number in the system at time t then $\{X(t), t \geq 0\}$ is a birth and death process with

$$\mu_n = \mu, \quad n \geq 1$$
$$\lambda_n = \lambda, \quad n \geq 0$$

■

Example 6.6 (A Multiserver Exponential Queueing System)**.** Consider an exponential queueing system in which there are s servers available, each serving at rate μ. An entering customer first waits in line and then goes to the first free server. Assuming arrivals are according to a Poisson process having rate λ, this is a birth and death process with parameters

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq s \\ s\mu, & n > s \end{cases}$$
$$\lambda_n = \lambda, \qquad n \geq 0$$

To see why this is true, reason as follows: If there are n customers in the system, where $n \leq s$, then n servers will be busy. Since each of these servers works at rate μ, the total departure rate will be $n\mu$. On the other hand, if there are n customers in the system, where $n > s$, then all s of the servers will be busy, and thus the total departure rate will be $s\mu$. This is known as an $M/M/s$ queueing model. ■

Consider now a general birth and death process with birth rates $\{\lambda_n\}$ and death rates $\{\mu_n\}$, where $\mu_0 = 0$, and let T_i denote the time, starting from state i, it takes for the process to enter state $i + 1, i \geq 0$. We will recursively compute $E[T_i], i \geq 0$, by starting with $i = 0$. Since T_0 is exponential with rate λ_0, we have

$$E[T_0] = \frac{1}{\lambda_0}$$

For $i > 0$, we condition whether the first transition takes the process into state $i - 1$ or $i + 1$. That is, let

$$I_i = \begin{cases} 1, & \text{if the first transition from } i \text{ is to } i + 1 \\ 0, & \text{if the first transition from } i \text{ is to } i - 1 \end{cases}$$

and note that

$$\begin{aligned} E[T_i | I_i = 1] &= \frac{1}{\lambda_i + \mu_i}, \\ E[T_i | I_i = 0] &= \frac{1}{\lambda_i + \mu_i} + E[T_{i-1}] + E[T_i] \end{aligned} \tag{6.3}$$

This follows since, independent of whether the first transition is from a birth or death, the time until it occurs is exponential with rate $\lambda_i + \mu_i$; if this first transition is a birth, then the population size is at $i + 1$, so no additional time is needed; whereas if it is death, then the population size becomes $i - 1$ and the additional time needed to reach $i + 1$ is equal to the time it takes to return to state i (this has mean $E[T_{i-1}]$) plus the additional time it then takes to reach $i + 1$ (this has mean $E[T_i]$). Hence, since the probability that the first transition is a birth is $\lambda_i/(\lambda_i + \mu_i)$, we see that

$$E[T_i] = \frac{1}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i}(E[T_{i-1}] + E[T_i])$$

or, equivalently,

$$E[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E[T_{i-1}], \qquad i \geq 1$$

Starting with $E[T_0] = 1/\lambda_0$, the preceding yields an efficient method to successively compute $E[T_1]$, $E[T_2]$, and so on.

Suppose now that we wanted to determine the expected time to go from state i to state j where $i < j$. This can be accomplished using the preceding by noting that this quantity will equal $E[T_i] + E[T_{i+1}] + \cdots + E[T_{j-1}]$.

Example 6.7. For the birth and death process having parameters $\lambda_i \equiv \lambda$, $\mu_i \equiv \mu$,

$$\begin{aligned} E[T_i] &= \frac{1}{\lambda} + \frac{\mu}{\lambda} E[T_{i-1}] \\ &= \frac{1}{\lambda}(1 + \mu E[T_{i-1}]) \end{aligned}$$

Starting with $E[T_0] = 1/\lambda$, we see that

$$\begin{aligned} E[T_1] &= \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda}\right), \\ E[T_2] &= \frac{1}{\lambda}\left[1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2\right] \end{aligned}$$

and, in general,

$$\begin{aligned} E[T_i] &= \frac{1}{\lambda}\left[1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2 + \cdots + \left(\frac{\mu}{\lambda}\right)^i\right] \\ &= \frac{1 - (\mu/\lambda)^{i+1}}{\lambda - \mu}, \qquad i \geq 0 \end{aligned}$$

The expected time to reach state j, starting at state k, $k < j$, is

$$\begin{aligned} E[\text{time to go from } k \text{ to } j] &= \sum_{i=k}^{j-1} E[T_i] \\ &= \frac{j-k}{\lambda - \mu} - \frac{(\mu/\lambda)^{k+1}}{\lambda - \mu} \frac{\left[1 - (\mu/\lambda)^{j-k}\right]}{1 - \mu/\lambda} \end{aligned}$$

The foregoing assumes that $\lambda \neq \mu$. If $\lambda = \mu$, then

$$E[T_i] = \frac{i+1}{\lambda},$$
$$E[\text{time to go from } k \text{ to } j] = \frac{j(j+1) - k(k+1)}{2\lambda}$$

■

We can also compute the variance of the time to go from 0 to $i+1$ by utilizing the conditional variance formula. First note that Eq. (6.3) can be written as

$$E[T_i|I_i] = \frac{1}{\lambda_i + \mu_i} + (1 - I_i)(E[T_{i-1}] + E[T_i])$$

Thus,

$$\begin{aligned} \text{Var}(E[T_i|I_i]) &= (E[T_{i-1}] + E[T_i])^2 \,\text{Var}(I_i) \\ &= (E[T_{i-1}] + E[T_i])^2 \frac{\mu_i \lambda_i}{(\mu_i + \lambda_i)^2} \end{aligned} \tag{6.4}$$

where $\text{Var}(I_i)$ is as shown since I_i is a Bernoulli random variable with parameter $p = \lambda_i/(\lambda_i + \mu_i)$. Also, note that if we let X_i denote the time until the transition from i occurs, then

$$\begin{aligned} \text{Var}(T_i|I_i = 1) &= \text{Var}(X_i|I_i = 1) \\ &= \text{Var}(X_i) \\ &= \frac{1}{(\lambda_i + \mu_i)^2} \end{aligned} \tag{6.5}$$

where the preceding uses the fact that the time until transition is independent of the next state visited. Also,

$$\begin{aligned} \text{Var}(T_i|I_i = 0) &= \text{Var}(X_i + \text{time to get back to } i + \text{time to then reach } i+1) \\ &= \text{Var}(X_i) + \text{Var}(T_{i-1}) + \text{Var}(T_i) \end{aligned} \tag{6.6}$$

where the foregoing uses the fact that the three random variables are independent. We can rewrite Eqs. (6.5) and (6.6) as

$$\text{Var}(T_i|I_i) = \text{Var}(X_i) + (1 - I_i)[\text{Var}(T_{i-1}) + \text{Var}(T_i)]$$

so

$$E[\text{Var}(T_i|I_i)] = \frac{1}{(\mu_i + \lambda_i)^2} + \frac{\mu_i}{\mu_i + \lambda_i}[\text{Var}(T_{i-1}) + \text{Var}(T_i)] \tag{6.7}$$

Hence, using the conditional variance formula, which states that $\text{Var}(T_i)$ is the sum of Eqs. (6.7) and (6.4), we obtain

$$\text{Var}(T_i) = \frac{1}{(\mu_i + \lambda_i)^2} + \frac{\mu_i}{\mu_i + \lambda_i}[\text{Var}(T_{i-1}) + \text{Var}(T_i)] + \frac{\mu_i \lambda_i}{(\mu_i + \lambda_i)^2}(E[T_{i-1}] + E[T_i])^2$$

or, equivalently,

$$\text{Var}(T_i) = \frac{1}{\lambda_i(\lambda_i + \mu_i)} + \frac{\mu_i}{\lambda_i}\text{Var}(T_{i-1}) + \frac{\mu_i}{\mu_i + \lambda_i}(E[T_{i-1}] + E[T_i])^2$$

Starting with $\text{Var}(T_0) = 1/\lambda_0^2$ and using the former recursion to obtain the expectations, we can recursively compute $\text{Var}(T_i)$. In addition, if we want the variance of the time to reach state j, starting from state k, $k < j$, then this can be expressed as the time to go from k to $k+1$ plus the additional time to go from $k+1$ to $k+2$, and so on. Since, by the Markovian property, these successive random variables are independent, it follows that

$$\text{Var}(\text{time to go from } k \text{ to } j) = \sum_{i=k}^{j-1} \text{Var}(T_i)$$

6.4 The Transition Probability Function $P_{ij}(t)$

Let

$$P_{ij}(t) = P\{X(t+s) = j | X(s) = i\}$$

denote the probability that a process presently in state i will be in state j a time t later. These quantities are often called the *transition probabilities* of the continuous-time Markov chain.

We can explicitly determine $P_{ij}(t)$ in the case of a pure birth process having distinct birth rates. For such a process, let X_k denote the time the process spends in state k before making a transition into state $k+1$, $k \geq 1$. Suppose that the process is presently in state i, and let $j > i$. Then, as X_i is the time it spends in state i before moving to state $i+1$, and X_{i+1} is the time it then spends in state $i+1$ before moving to state $i+2$, and so on, it follows that $\sum_{k=i}^{j-1} X_k$ is the time it takes until the process enters state j. Now, if the process has not yet entered state j by time t, then its state at time t is smaller than j, and vice versa. That is,

$$X(t) < j \Leftrightarrow X_i + \cdots + X_{j-1} > t$$

Therefore, for $i < j$, we have for a pure birth process that

$$P\{X(t) < j | X(0) = i\} = P\left\{\sum_{k=i}^{j-1} X_k > t\right\}$$

However, since $X_i, \ldots, X_{j-1}$ are independent exponential random variables with respective rates $\lambda_i, \ldots, \lambda_{j-1}$, we obtain from the preceding and Eq. (5.9), which gives the tail distribution function of $\sum_{k=i}^{j-1} X_k$, that

$$P\{X(t) < j | X(0) = i\} = \sum_{k=i}^{j-1} e^{-\lambda_k t} \prod_{r \neq k,\ r=i}^{j-1} \frac{\lambda_r}{\lambda_r - \lambda_k}$$

Replacing j by $j+1$ in the preceding gives

$$P\{X(t) < j+1 | X(0) = i\} = \sum_{k=i}^{j} e^{-\lambda_k t} \prod_{r \neq k,\ r=i}^{j} \frac{\lambda_r}{\lambda_r - \lambda_k}$$

Since

$$P\{X(t) = j | X(0) = i\} = P\{X(t) < j+1 | X(0) = i\} - P\{X(t) < j | X(0) = i\}$$

and since $P_{ii}(t) = P\{X_i > t\} = e^{-\lambda_i t}$, we have shown the following.

Proposition 6.1. *For a pure birth process having $\lambda_i \neq \lambda_j$ when $i \neq j$*

$$P_{ij}(t) = \sum_{k=i}^{j} e^{-\lambda_k t} \prod_{r \neq k,\ r=i}^{j} \frac{\lambda_r}{\lambda_r - \lambda_k} - \sum_{k=i}^{j-1} e^{-\lambda_k t} \prod_{r \neq k,\ r=i}^{j-1} \frac{\lambda_r}{\lambda_r - \lambda_k}, \qquad i < j$$
$$P_{ii}(t) = e^{-\lambda_i t}$$

Example 6.8. Consider the Yule process, which is a pure birth process in which each individual in the population independently gives birth at rate λ, and so $\lambda_n = n\lambda$, $n \geq 1$. Letting $i = 1$, we obtain from Proposition 6.1

$$\begin{aligned} P_{1j}(t) &= \sum_{k=1}^{j} e^{-k\lambda t} \prod_{r \neq k,\ r=1}^{j} \frac{r}{r-k} - \sum_{k=1}^{j-1} e^{-k\lambda t} \prod_{r \neq k,\ r=1}^{j-1} \frac{r}{r-k} \\ &= e^{-j\lambda t} \prod_{r=1}^{j-1} \frac{r}{r-j} + \sum_{k=1}^{j-1} e^{-k\lambda t} \left(\prod_{r \neq k,\ r=1}^{j} \frac{r}{r-k} - \prod_{r \neq k,\ r=1}^{j-1} \frac{r}{r-k} \right) \\ &= e^{-j\lambda t} (-1)^{j-1} + \sum_{k=1}^{j-1} e^{-k\lambda t} \left(\frac{j}{j-k} - 1 \right) \prod_{r \neq k,\ r=1}^{j-1} \frac{r}{r-k} \end{aligned}$$

Now,

$$\frac{k}{j-k} \prod_{r \neq k,\ r=1}^{j-1} \frac{r}{r-k} = \frac{(j-1)!}{(1-k)(2-k)\cdots(k-1-k)(j-k)!}$$

$$= (-1)^{k-1}\binom{j-1}{k-1}$$

so

$$\begin{aligned} P_{1j}(t) &= \sum_{k=1}^{j}\binom{j-1}{k-1}e^{-k\lambda t}(-1)^{k-1} \\ &= e^{-\lambda t}\sum_{i=0}^{j-1}\binom{j-1}{i}e^{-i\lambda t}(-1)^{i} \\ &= e^{-\lambda t}(1-e^{-\lambda t})^{j-1} \end{aligned}$$

Thus, starting with a single individual, the population size at time t has a geometric distribution with mean $e^{\lambda t}$. If the population starts with i individuals, then we can regard each of these individuals as starting her own independent Yule process, and so the population at time t will be the sum of i independent and identically distributed geometric random variables with parameter $e^{-\lambda t}$. But this means that the conditional distribution of $X(t)$, given that $X(0) = i$, is the same as the distribution of the number of times that a coin that lands heads on each flip with probability $e^{-\lambda t}$ must be flipped to amass a total of i heads. Hence, the population size at time t has a negative binomial distribution with parameters i and $e^{-\lambda t}$, so

$$P_{ij}(t) = \binom{j-1}{i-1}e^{-i\lambda t}(1-e^{-\lambda t})^{j-i}, \qquad j \geq i \geq 1$$

(We could, of course, have used Proposition 6.1 to immediately obtain an equation for $P_{ij}(t)$, rather than just using it for $P_{1j}(t)$, but the algebra that would have then been needed to show the equivalence of the resulting expression to the preceding result is somewhat involved.) ■

Example 6.9. An urn initially contains one type 1 and one type 2 ball. At each stage, a ball is chosen from the urn, with the chosen ball being equally likely to be any of the balls in the urn. If a type i ball is chosen, then an experiment that is successful with probability p_i is performed; if it is successful then the ball chosen along with a new type i ball are put in the urn, and if it is unsuccessful then only the ball chosen is put in the urn, $i = 1, 2$. We then move to the next stage. We are interested in determining the mean numbers of type 1 and type 2 balls in the urn after n stages.

Solution: To determine the mean numbers, for $i = 1, 2$, let $m_i(j, k : r)$ denote the mean number of type i balls in the urn after the n stages have elapsed, given that there are currently j type 1 and k type 2 balls in the urn, with a total of r additional stages remaining. Also, let $m(j, k : r)$ be the vector

$$m(j, k : r) = (m_1(j, k : r),\ m_2(j, k : r))\,.$$

We need to determine $m(1, 1 : n)$. To start, we derive recursive equations for $m(j, k : r)$ by conditioning on the first ball chosen and whether the resulting ex-

periment is successful. This yields that

$$m(j,k:r) = \frac{j}{j+k}[p_1 m(j+1,k:r-1) + q_1 m(j,k:r-1)] + \frac{k}{j+k}[p_2 m(j,k+1:r-1) + q_2 m(j,k:r-1)]$$

where $q_i = 1 - p_i$, $i = 1, 2$. Now, using that

$$m(j,k:0) = (j,k)$$

we can use the recursion to determine the values $m(j,k:r)$ when $r = 1$, then when $r = 2$, and so on, up to $r = n$.

We can also derive an approximation for the mean numbers of type 1 and type 2 balls in the urn after n stages, by using a "Poissonization" trick. Let us imagine that each ball in the urn, independently of other balls, lights up at times distributed as a Poisson process with rate $\lambda = 1$. Suppose that each time a type i ball lights up, we conduct the experiment that is successful with probability p_i and add a new type i ball to the urn if it is successful, $i = 1, 2$. Each time a ball lights up, say that a new stage has begun. Because, for an urn that currently has j type 1 and k type 2 balls, the next ball to light up will be of type 1 with probability $\frac{j}{j+k}$, the numbers of type 1 and type 2 balls in the urn after successive stages are distributed exactly as in the original model. Now, whenever there are j type 1 balls in the urn, the time until the next type 1 ball lights up is the minimum of j independent exponential random variables with rate 1 and so is exponential with rate j. Because, with probability p_1, this will then result in a new type 1 ball to be added to the urn, it follows that, whenever there are j type 1 balls in the urn, the time until the next type 1 ball is added is distributed as an exponential random variable with rate jp_1. Consequently, the counting process of the number of type 1 balls in the urn is a Yule process with birth parameters $\lambda_1(j) = jp_1$, $j \geq 1$. Similarly, the counting process of the number of type 2 balls in the urn is a Yule process with birth parameters $\lambda_2(j) = jp_2$, $j \geq 1$, with these two Yule processes being independent. Thus, starting with a single type i ball, it follows that $N_i(t)$, defined as the number of type i balls in the urn at time t, is a geometric random variable with parameter $e^{-p_i t}$, $i = 1, 2$. Therefore,

$$E[N_i(t)] = e^{p_i t}, \quad i = 1, 2$$

Also, if $L_i(t)$ denotes the number of times that a type i ball has lit up by time t, then as each light up, independently of all that came earlier, results in a new type i ball being added with probability p_i, it is intuitive that

$$E[N_i(t)] = p_i E[L_i(t)] + 1, \quad i = 1, 2$$

Thus,

$$E[L_i(t)] = \frac{e^{p_i t} - 1}{p_i}, \quad i = 1, 2$$

Hence, the expected number of stages that have passed by time t is

$$E[L_1(t) + L_2(t)] = \frac{e^{p_1 t} - 1}{p_1} + \frac{e^{p_2 t} - 1}{p_2}$$

If we let t_n be the value of t that makes the preceding equal n; that is, t_n is such that

$$\frac{e^{p_1 t_n} - 1}{p_1} + \frac{e^{p_2 t_n} - 1}{p_2} = n$$

then we can approximate the expected number of type i balls in the urn after n stages by $E[N_i(t_n)] = e^{p_i t_n}$, $i = 1, 2$. ■

Remarks. **(i)** That $E[N_i(t)] = p_i E[L_i(t)] + 1$ is not immediate. Because information as to the number of light ups by time t changes the probabilities that the experiments resulting from light ups were successful (for instance, $L_i(t)$ being large makes it more likely that experiments were successful because a successful experiment increases the light up rate)

$$E[N_i(t)|L_i(t)] \neq p_i L_i(t) + 1$$

However, even though the preceding is the case and so can not be used to prove that $E[N_i(t)] = p_i E[L_i(t)] + 1$, the preceding equation is indeed valid and can be proven by using Wald's equation, a technique that will be presented in Section 7.3.

(ii) The preceding example has been applied in drug testing. Imagine there are two drugs with unknown cure probabilities (p_1 and p_2 in the example). At each stage, the choice of the drug to give to a patient is made by randomly choosing a ball from the urn. If a type i ball is chosen, then drug i is used. The result of using this drug is assumed to be immediately learned, and a successful outcome results in another ball of type i being added to the urn, $i = 1, 2$.

(iii) If $p_1 = .7$, $p_2 = .4$, then after $n = 500$ stages, the expected number of type 1 balls in the urn is 288.92 and the expected number of type 2 balls is 36.47. The approximations of these quantities given in the preceding are, respectively, 304.09 and 26.23. After 1000 stages the true means are 600.77 and 58.28, whereas the approximations are 630.37 and 39.79. ■

We shall now derive a set of differential equations that the transition probabilities $P_{ij}(t)$ satisfy in a general continuous-time Markov chain. However, first we need a definition and a pair of lemmas.

For any pair of states i and j, let

$$q_{ij} = v_i P_{ij}$$

Since v_i is the rate at which the process makes a transition when in state i and P_{ij} is the probability that this transition is into state j, it follows that q_{ij} is the rate, when

in state i, at which the process makes a transition into state j. The quantities q_{ij} are called the *instantaneous transition rates*. Since

$$v_i = \sum_j v_i P_{ij} = \sum_j q_{ij}$$

and

$$P_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_j q_{ij}}$$

it follows that specifying the instantaneous transition rates determines the parameters of the continuous-time Markov chain.

Lemma 6.2. **(a)** $\lim_{h\to 0} \dfrac{1 - P_{ii}(h)}{h} = v_i$

(b) $\lim_{h\to 0} \dfrac{P_{ij}(h)}{h} = q_{ij}$ *when* $i \neq j$

Proof. We first note that since the amount of time until a transition occurs is exponentially distributed it follows that the probability of two or more transitions in a time h is $o(h)$. Thus, $1 - P_{ii}(h)$, the probability that a process in state i at time 0 will not be in state i at time h, equals the probability that a transition occurs within time h plus something small compared to h. Therefore,

$$1 - P_{ii}(h) = v_i h + o(h)$$

and part (a) is proven. To prove part (b), we note that $P_{ij}(h)$, the probability that the process goes from state i to state j in a time h, equals the probability that a transition occurs in this time multiplied by the probability that the transition is into state j, plus something small compared to h. That is,

$$P_{ij}(h) = h v_i P_{ij} + o(h)$$

and part (b) is proven. ■

Lemma 6.3. *For all* $s \geq 0, t \geq 0$,

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s) \tag{6.8}$$

Proof. In order for the process to go from state i to state j in time $t + s$, it must be somewhere at time t and thus

$$\begin{aligned} P_{ij}(t+s) &= P\{X(t+s) = j | X(0) = i\} \\ &= \sum_{k=0}^{\infty} P\{X(t+s) = j, X(t) = k | X(0) = i\} \end{aligned}$$

$$= \sum_{k=0}^{\infty} P\{X(t+s) = j | X(t) = k, X(0) = i\} \cdot P\{X(t) = k | X(0) = i\}$$
$$= \sum_{k=0}^{\infty} P\{X(t+s) = j | X(t) = k\} \cdot P\{X(t) = k | X(0) = i\}$$
$$= \sum_{k=0}^{\infty} P_{kj}(s) P_{ik}(t)$$

and the proof is completed. ■

The set of Eqs. (6.8) is known as the *Chapman–Kolmogorov* equations. From Lemma 6.3, we obtain

$$P_{ij}(h+t) - P_{ij}(t) = \sum_{k=0}^{\infty} P_{ik}(h) P_{kj}(t) - P_{ij}(t)$$
$$= \sum_{k \neq i} P_{ik}(h) P_{kj}(t) - [1 - P_{ii}(h)] P_{ij}(t)$$

and thus

$$\lim_{h \to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \lim_{h \to 0} \left\{ \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) - \left[\frac{1 - P_{ii}(h)}{h} \right] P_{ij}(t) \right\}$$

Now, assuming that we can interchange the limit and the summation in the preceding and applying Lemma 6.2, we obtain

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

It turns out that this interchange can indeed be justified and, hence, we have the following theorem.

Theorem 6.1 (Kolmogorov's Backward Equations)**.** *For all states i, j, and times $t \geq 0$,*

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

Example 6.10. The backward equations for the pure birth process become

$$P'_{ij}(t) = \lambda_i P_{i+1,j}(t) - \lambda_i P_{ij}(t)$$

The backward equations for the birth and death process become

$$P'_{0j}(t) = \lambda_0 P_{1j}(t) - \lambda_0 P_{0j}(t),$$

$$P'_{ij}(t) = (\lambda_i + \mu_i)\left[\frac{\lambda_i}{\lambda_i + \mu_i} P_{i+1,j}(t) + \frac{\mu_i}{\lambda_i + \mu_i} P_{i-1,j}(t)\right] - (\lambda_i + \mu_i) P_{ij}(t)$$

or equivalently,

$$\begin{aligned} P'_{0j}(t) &= \lambda_0[P_{1j}(t) - P_{0j}(t)], \\ P'_{ij}(t) &= \lambda_i P_{i+1,j}(t) + \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t), \qquad i > 0 \end{aligned} \tag{6.9}$$

■

Example 6.11 (A Continuous-Time Markov Chain Consisting of Two States)**.** Consider a machine that works for an exponential amount of time having mean $1/\lambda$ before breaking down; and suppose that it takes an exponential amount of time having mean $1/\mu$ to repair the machine. If the machine is in working condition at time 0, then what is the probability that it will be working at time $t = 10$?

To answer this question, we note that the process is a birth and death process (with state 0 meaning that the machine is working and state 1 that it is being repaired) having parameters

$$\begin{aligned} &\lambda_0 = \lambda, && \mu_1 = \mu, \\ &\lambda_i = 0, \quad i \neq 0, && \mu_i = 0, \quad i \neq 1 \end{aligned}$$

We shall derive the desired probability, namely, $P_{00}(10)$ by solving the set of differential equations given in Example 6.10. From Eq. (6.9), we obtain

$$P'_{00}(t) = \lambda[P_{10}(t) - P_{00}(t)], \tag{6.10}$$

$$P'_{10}(t) = \mu P_{00}(t) - \mu P_{10}(t) \tag{6.11}$$

Multiplying Eq. (6.10) by μ and Eq. (6.11) by λ and then adding the two equations yields

$$\mu P'_{00}(t) + \lambda P'_{10}(t) = 0$$

By integrating, we obtain

$$\mu P_{00}(t) + \lambda P_{10}(t) = c$$

However, since $P_{00}(0) = 1$ and $P_{10}(0) = 0$, we obtain $c = \mu$ and hence,

$$\mu P_{00}(t) + \lambda P_{10}(t) = \mu \tag{6.12}$$

or equivalently,

$$\lambda P_{10}(t) = \mu[1 - P_{00}(t)]$$

By substituting this result in Eq. (6.10), we obtain

$$\begin{aligned} P'_{00}(t) &= \mu[1 - P_{00}(t)] - \lambda P_{00}(t) \\ &= \mu - (\mu + \lambda) P_{00}(t) \end{aligned}$$

Letting

$$h(t) = P_{00}(t) - \frac{\mu}{\mu + \lambda}$$

we have

$$\begin{aligned} h'(t) &= \mu - (\mu + \lambda)\left[h(t) + \frac{\mu}{\mu + \lambda}\right] \\ &= -(\mu + \lambda)h(t) \end{aligned}$$

or

$$\frac{h'(t)}{h(t)} = -(\mu + \lambda)$$

By integrating both sides, we obtain

$$\log h(t) = -(\mu + \lambda)t + c$$

or

$$h(t) = Ke^{-(\mu+\lambda)t}$$

and thus

$$P_{00}(t) = Ke^{-(\mu+\lambda)t} + \frac{\mu}{\mu + \lambda}$$

which finally yields, by setting $t = 0$ and using the fact that $P_{00}(0) = 1$,

$$P_{00}(t) = \frac{\lambda}{\mu + \lambda} e^{-(\mu+\lambda)t} + \frac{\mu}{\mu + \lambda}$$

From Eq. (6.12), this also implies that

$$P_{10}(t) = \frac{\mu}{\mu + \lambda} - \frac{\mu}{\mu + \lambda} e^{-(\mu+\lambda)t}$$

Hence, our desired probability is as follows:

$$P_{00}(10) = \frac{\lambda}{\mu + \lambda} e^{-10(\mu+\lambda)} + \frac{\mu}{\mu + \lambda}$$ ■

Another set of differential equations, different from the backward equations, may also be derived. This set of equations, known as *Kolmogorov's forward equations* is derived as follows. From the Chapman–Kolmogorov equations (Lemma 6.3), we have

$$\begin{aligned} P_{ij}(t + h) - P_{ij}(t) &= \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(h) - P_{ij}(t) \\ &= \sum_{k \neq j} P_{ik}(t)P_{kj}(h) - [1 - P_{jj}(h)]P_{ij}(t) \end{aligned}$$

and thus

$$\lim_{h\to 0}\frac{P_{ij}(t+h)-P_{ij}(t)}{h}=\lim_{h\to 0}\left\{\sum_{k\neq j}P_{ik}(t)\frac{P_{kj}(h)}{h}-\left[\frac{1-P_{jj}(h)}{h}\right]P_{ij}(t)\right\}$$

and, assuming that we can interchange limit with summation, we obtain from Lemma 6.2

$$P'_{ij}(t)=\sum_{k\neq j}q_{kj}P_{ik}(t)-v_jP_{ij}(t)$$

Unfortunately, we cannot always justify the interchange of limit and summation and thus the preceding is not always valid. However, they do hold in most models, including all birth and death processes and all finite state models. We thus have the following.

Theorem 6.2 (Kolmogorov's Forward Equations)**.** *Under suitable regularity conditions,*

$$P'_{ij}(t)=\sum_{k\neq j}q_{kj}P_{ik}(t)-v_jP_{ij}(t) \tag{6.13}$$

We shall now solve the forward equations for the pure birth process. For this process, Eq. (6.13) reduces to

$$P'_{ij}(t)=\lambda_{j-1}P_{i,j-1}(t)-\lambda_jP_{ij}(t)$$

However, by noting that $P_{ij}(t)=0$ whenever $j<i$ (since no deaths can occur), we can rewrite the preceding equation to obtain

$$\begin{aligned}&P'_{ii}(t)=-\lambda_iP_{ii}(t),\\&P'_{ij}(t)=\lambda_{j-1}P_{i,j-1}(t)-\lambda_jP_{ij}(t),\qquad j\geq i+1\end{aligned} \tag{6.14}$$

Proposition 6.4. *For a pure birth process,*

$$\begin{aligned}&P_{ii}(t)=e^{-\lambda_it},&&i\geq 0\\&P_{ij}(t)=\lambda_{j-1}e^{-\lambda_jt}\int_0^t e^{\lambda_js}P_{i,j-1}(s)\,ds,&&j\geq i+1\end{aligned}$$

Proof. The fact that $P_{ii}(t)=e^{-\lambda_it}$ follows from Eq. (6.14) by integrating and using the fact that $P_{ii}(0)=1$. To prove the corresponding result for $P_{ij}(t)$, we note by Eq. (6.14) that

$$e^{\lambda_jt}\left[P'_{ij}(t)+\lambda_jP_{ij}(t)\right]=e^{\lambda_jt}\lambda_{j-1}P_{i,j-1}(t)$$

or

$$\frac{d}{dt}\left[e^{\lambda_j t} P_{ij}(t)\right] = \lambda_{j-1} e^{\lambda_j t} P_{i,j-1}(t)$$

Hence, since $P_{ij}(0) = 0$, we obtain the desired results. ■

Example 6.12 (Forward Equations for Birth and Death Process)**.** The forward equations (Eq. (6.13)) for the general birth and death process become

$$\begin{aligned} P'_{i0}(t) &= \sum_{k \neq 0} q_{k0} P_{ik}(t) - \lambda_0 P_{i0}(t) \\ &= \mu_1 P_{i1}(t) - \lambda_0 P_{i0}(t) \end{aligned} \tag{6.15}$$

$$\begin{aligned} P'_{ij}(t) &= \sum_{k \neq j} q_{kj} P_{ik}(t) - (\lambda_j + \mu_j) P_{ij}(t) \\ &= \lambda_{j-1} P_{i,j-1}(t) + \mu_{j+1} P_{i,j+1}(t) - (\lambda_j + \mu_j) P_{ij}(t) \end{aligned} \tag{6.16}$$

■

6.5 Limiting Probabilities

In analogy with a basic result in discrete-time Markov chains, the probability that a continuous-time Markov chain will be in state j at time t often converges to a limiting value that is independent of the initial state. That is, if we call this value P_j, then

$$P_j \equiv \lim_{t \to \infty} P_{ij}(t)$$

where we are assuming that the limit exists and is independent of the initial state i.

To derive a set of equations for the P_j, consider first the set of forward equations

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \tag{6.17}$$

Now, if we let t approach ∞, then assuming that we can interchange limit and summation, we obtain

$$\begin{aligned} \lim_{t \to \infty} P'_{ij}(t) &= \lim_{t \to \infty} \left[\sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \right] \\ &= \sum_{k \neq j} q_{kj} P_k - v_j P_j \end{aligned}$$

However, as $P_{ij}(t)$ is a bounded function (being a probability it is always between 0 and 1), it follows that if $P'_{ij}(t)$ converges, then it must converge to 0 (why is this?). Hence, we must have

$$0 = \sum_{k \neq j} q_{kj} P_k - v_j P_j$$

or

$$v_j P_j = \sum_{k \neq j} q_{kj} P_k, \qquad \text{all states } j \tag{6.18}$$

The preceding set of equations, along with the equation

$$\sum_j P_j = 1 \tag{6.19}$$

can be used to solve for the limiting probabilities.

Remark. **(i)** We have assumed that the limiting probabilities P_j exist. A sufficient condition for this is that

(a) all states of the Markov chain communicate in the sense that starting in state i there is a positive probability of ever being in state j, for all i, j and

(b) the Markov chain is positive recurrent in the sense that, starting in any state, the mean time to return to that state is finite

If conditions (a) and (b) hold, then the limiting probabilities will exist and satisfy Eqs. (6.18) and (6.19). In addition, P_j also will have the interpretation of being the long-run proportion of time that the process is in state j.

(ii) Eqs. (6.18) and (6.19) have a nice interpretation: In any interval $(0, t)$ the number of transitions into state j must equal to within 1 the number of transitions out of state j (why?). Hence, in the long run, the rate at which transitions into state j occur must equal the rate at which transitions out of state j occur. When the process is in state j, it leaves at rate v_j, and, as P_j is the proportion of time it is in state j, it thus follows that

$$v_j P_j = \text{rate at which the process leaves state } j$$

Similarly, when the process is in state k, it enters j at a rate q_{kj}. Hence, as P_k is the proportion of time in state k, we see that the rate at which transitions from k to j occur is just $q_{kj} P_k$; thus

$$\sum_{k \neq j} q_{kj} P_k = \text{rate at which the process enters state } j$$

So, Eq. (6.18) is just a statement of the equality of the rates at which the process enters and leaves state j. Because it balances (that is, equates) these rates, Eq. (6.18) is sometimes referred to as a set of "balance equations."

Let us now determine the limiting probabilities for a birth and death process. From Eq. (6.18) or equivalently, by equating the rate at which the process leaves a state with the rate at which it enters that state, we obtain

State	*Rate at which leave = rate at which enter*
0	$\lambda_0 P_0 = \mu_1 P_1$
1	$(\lambda_1 + \mu_1) P_1 = \mu_2 P_2 + \lambda_0 P_0$
2	$(\lambda_2 + \mu_2) P_2 = \mu_3 P_3 + \lambda_1 P_1$
$n, n \geq 1$	$(\lambda_n + \mu_n) P_n = \mu_{n+1} P_{n+1} + \lambda_{n-1} P_{n-1}$

By adding to each equation the equation preceding it, we obtain

$$\begin{aligned}
&\lambda_0 P_0 = \mu_1 P_1,\\
&\lambda_1 P_1 = \mu_2 P_2,\\
&\lambda_2 P_2 = \mu_3 P_3,\\
&\qquad \vdots\\
&\lambda_n P_n = \mu_{n+1} P_{n+1}, \qquad n \geq 0
\end{aligned}$$

Solving in terms of P_0 yields

$$\begin{aligned}
P_1 &= \frac{\lambda_0}{\mu_1} P_0,\\
P_2 &= \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0,\\
P_3 &= \frac{\lambda_2}{\mu_3} P_2 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} P_0,\\
&\vdots\\
P_n &= \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_1 \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_2 \mu_1} P_0
\end{aligned}$$

And by using the fact that $\sum_{n=0}^{\infty} P_n = 1$, we obtain

$$1 = P_0 + P_0 \sum_{n=1}^{\infty} \frac{\lambda_{n-1} \cdots \lambda_1 \lambda_0}{\mu_n \cdots \mu_2 \mu_1}$$

or

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}$$

and so

$$P_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}\right)}, \qquad n \geq 1 \tag{6.20}$$

The foregoing equations also show us what condition is necessary for these limiting probabilities to exist. Namely, it is necessary that

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty \tag{6.21}$$

This condition also may be shown to be sufficient.

In the multiserver exponential queueing system (Example 6.6), Condition (6.21) reduces to

$$\sum_{n=s+1}^{\infty} \frac{\lambda^n}{(s\mu)^n} < \infty$$

which is equivalent to $\lambda < s\mu$.

For the linear growth model with immigration (Example 6.4), Condition (6.21) reduces to

$$\sum_{n=1}^{\infty} \frac{\theta(\theta+\lambda)\cdots(\theta+(n-1)\lambda)}{n!\mu^n} < \infty$$

Using the ratio test, the preceding will converge when

$$\lim_{n\to\infty} \frac{\theta(\theta+\lambda)\cdots(\theta+n\lambda)}{(n+1)!\mu^{n+1}} \frac{n!\mu^n}{\theta(\theta+\lambda)\cdots(\theta+(n-1)\lambda)} = \lim_{n\to\infty} \frac{\theta+n\lambda}{(n+1)\mu}$$
$$= \frac{\lambda}{\mu} < 1$$

That is, the condition is satisfied when $\lambda < \mu$. When $\lambda \geq \mu$ it is easy to show that Condition (6.21) is not satisfied.

Example 6.13 (A Machine Repair Model)**.** Consider a job shop that consists of M machines and one serviceman. Suppose that the amount of time each machine runs before breaking down is exponentially distributed with mean $1/\lambda$, and suppose that the amount of time that it takes for the serviceman to fix a machine is exponentially distributed with mean $1/\mu$. We shall attempt to answer these questions: (a) What is the average number of machines not in use? (b) What proportion of time is each machine in use?

Solution: If we say that the system is in state n whenever n machines are not in use, then the preceding is a birth and death process having parameters

$$\mu_n = \mu \qquad n \geq 1$$
$$\lambda_n = \begin{cases} (M-n)\lambda, & n \leq M \\ 0, & n > M \end{cases}$$

This is so in the sense that a failing machine is regarded as an arrival and a fixed machine as a departure. If any machines are broken down, then since the serviceman's rate is μ, $\mu_n = \mu$. On the other hand, if n machines are not in use, then since

the $M-n$ machines in use each fail at a rate λ, it follows that $\lambda_n = (M-n)\lambda$. From Eq. (6.20) we have that P_n, the probability that n machines will not be in use, is given by

$$\begin{aligned} P_0 &= \frac{1}{1+\sum_{n=1}^{M}[M\lambda(M-1)\lambda\cdots(M-n+1)\lambda/\mu^n]} \\ &= \frac{1}{1+\sum_{n=1}^{M}(\lambda/\mu)^n M!/(M-n)!}, \\ P_n &= \frac{(\lambda/\mu)^n M!/(M-n)!}{1+\sum_{n=1}^{M}(\lambda/\mu)^n M!/(M-n)!}, \qquad n=0,1,\ldots,M \end{aligned}$$

Hence, the average number of machines not in use is given by

$$\sum_{n=0}^{M} nP_n = \frac{\sum_{n=0}^{M} n(\lambda/\mu)^n M!/(M-n)!}{1+\sum_{n=1}^{M}(\lambda/\mu)^n M!/(M-n)!} \tag{6.22}$$

To obtain the long-run proportion of time that a given machine is working we will compute the equivalent limiting probability of the machine working. To do so, we condition on the number of machines that are not working to obtain

$$\begin{aligned} P\{\text{machine is working}\} &= \sum_{n=0}^{M} P\{\text{machine is working}|n \text{ not working}\}P_n \\ &= \sum_{n=0}^{M} \frac{M-n}{M} P_n \qquad \begin{array}{l}\text{(since if } n \text{ are not working,}\\ \text{then } M-n \text{ are working!)}\end{array} \\ &= 1 - \sum_{n=0}^{M} \frac{nP_n}{M} \end{aligned}$$

where $\sum_{n=0}^{M} nP_n$ is given by Eq. (6.22). ■

Example 6.14 (The *M/M*/1 Queue)**.** In the $M/M/1$ queue $\lambda_n = \lambda$, $\mu_n = \mu$ and thus, from Eq. (6.20),

$$\begin{aligned} P_n &= \frac{(\lambda/\mu)^n}{1+\sum_{n=1}^{\infty}(\lambda/\mu)^n} \\ &= (\lambda/\mu)^n(1-\lambda/\mu), \qquad n \geq 0 \end{aligned}$$

provided that $\lambda/\mu < 1$. It is intuitive that λ must be less than μ for limiting probabilities to exist. Customers arrive at rate λ and are served at rate μ, and thus if $\lambda > \mu$, then they arrive at a faster rate than they can be served and the queue size will go to infinity. The case $\lambda = \mu$ behaves much like the symmetric random walk of Section 4.3, which is null recurrent and thus has no limiting probabilities. ■

Example 6.15. Let us reconsider the shoe shine shop of Example 6.1, and determine the proportion of time the process is in each of the states 0, 1, 2. Because this is not a birth and death process (since the process can go directly from state 2 to state 0), we start with the balance equations for the limiting probabilities.

State	*Rate that the process leaves = rate that the process enters*
0	$\lambda P_0 = \mu_2 P_2$
1	$\mu_1 P_1 = \lambda P_0$
2	$\mu_2 P_2 = \mu_1 P_1$

Solving in terms of P_0 yields

$$P_2 = \frac{\lambda}{\mu_2} P_0, \qquad P_1 = \frac{\lambda}{\mu_1} P_0$$

which implies, since $P_0 + P_1 + P_2 = 1$, that

$$P_0 \left[1 + \frac{\lambda}{\mu_2} + \frac{\lambda}{\mu_1} \right] = 1$$

or

$$P_0 = \frac{\mu_1 \mu_2}{\mu_1 \mu_2 + \lambda(\mu_1 + \mu_2)}$$

and

$$P_1 = \frac{\lambda \mu_2}{\mu_1 \mu_2 + \lambda(\mu_1 + \mu_2)},$$
$$P_2 = \frac{\lambda \mu_1}{\mu_1 \mu_2 + \lambda(\mu_1 + \mu_2)}$$

■

Example 6.16. Consider a set of n components along with a single repairman. Suppose that component i functions for an exponentially distributed time with rate λ_i and then fails. The time it then takes to repair component i is exponential with rate $\mu_i, i = 1, \ldots, n$. Suppose that when there is more than one failed component the repairman always works on the most recent failure. For instance, if there are at present two failed components—say, components 1 and 2 of which 1 has failed most recently—then the repairman will be working on component 1. However, if component 3 should fail before 1's repair is completed, then the repairman would stop working on component 1 and switch to component 3 (that is, a newly failed component preempts service).

To analyze the preceding as a continuous-time Markov chain, the state must represent the set of failed components in the order of failure. That is, the state will be $i_1, i_2, \ldots, i_k$ if $i_1, i_2, \ldots, i_k$ are the k failed components (all the other $n - k$ being functional) with i_1 having been the most recent failure (and is thus presently being repaired), i_2 the second most recent, and so on. Because there are $k!$ possible orderings

for a fixed set of k failed components and $\binom{n}{k}$ choices of that set, it follows that there are

$$\sum_{k=0}^{n} \binom{n}{k} k! = \sum_{k=0}^{n} \frac{n!}{(n-k)!} = n! \sum_{i=0}^{n} \frac{1}{i!}$$

possible states.

The balance equations for the limiting probabilities are as follows:

$$\left(\mu_{i_1} + \sum_{\substack{i \neq i_j \\ j=1,\ldots,k}} \lambda_i \right) P(i_1, \ldots, i_k) = \sum_{\substack{i \neq i_j \\ j=1,\ldots,k}} P(i, i_1, \ldots, i_k)\mu_i + P(i_2, \ldots, i_k)\lambda_{i_1},$$

$$\sum_{i=1}^{n} \lambda_i P(\phi) = \sum_{i=1}^{n} P(i)\mu_i \tag{6.23}$$

where ϕ is the state when all components are working. The preceding equations follow because state $i_1, \ldots, i_k$ can be left either by a failure of any of the additional components or by a repair completion of component i_1. Also, that state can be entered either by a repair completion of component i when the state is $i, i_1, \ldots, i_k$ or by a failure of component i_1 when the state is $i_2, \ldots, i_k$.

However, if we take

$$P(i_1, \ldots, i_k) = \frac{\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}}{\mu_{i_1}\mu_{i_2}\cdots\mu_{i_k}} P(\phi) \tag{6.24}$$

then it is easily seen that Eqs. (6.23) are satisfied. Hence, by uniqueness these must be the limiting probabilities with $P(\phi)$ determined to make their sum equal 1. That is,

$$P(\phi) = \left[1 + \sum_{i_1,\ldots,i_k} \frac{\lambda_{i_1}\cdots\lambda_{i_k}}{\mu_{i_1}\cdots\mu_{i_k}} \right]^{-1}$$

As an illustration, suppose $n = 2$ and so there are five states ϕ, 1, 2, 12, 21. Then from the preceding we would have

$$P(\phi) = \left[1 + \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} + \frac{2\lambda_1\lambda_2}{\mu_1\mu_2} \right]^{-1},$$

$$P(1) = \frac{\lambda_1}{\mu_1} P(\phi),$$

$$P(2) = \frac{\lambda_2}{\mu_2} P(\phi),$$

$$P(1,2) = P(2,1) = \frac{\lambda_1\lambda_2}{\mu_1\mu_2} P(\phi)$$

It is interesting to note, using Eq. (6.24), that given the set of failed components, each of the possible orderings of these components is equally likely. ■

When the limiting probabilities exist we say that the chain is *ergodic*. The limiting probabilities P_j are often called *stationary probabilities* because (as in the case of a discrete time Markov chain) if the initial state of the continuous time Markov chain is chosen according to the probabilities $\{P_j\}$ then the probability of being in state j at time t is P_j for all t and j. To verify this, suppose that the initial state is chosen according to the limiting probabilities P_j. Then,

$$\begin{aligned} P(X(t)=j) &= \sum_k P(X(t)=j|X(0)=k)P(X(0)=k) \\ &= \sum_k P_{k,j}(t)\, P_k \\ &= \sum_k P_{k,j}(t) \lim_{s\to\infty} P_{i,k}(s) \\ &= \lim_{s\to\infty} \sum_k P_{k,j}(t) P_{i,k}(s) \\ &= \lim_{s\to\infty} P_{i,j}(t+s) \\ &= P_j \end{aligned}$$

where we have assumed that the interchange of limit and summation is justified, and where the next to last equality follows from the Chapman–Kolmogorov equations (Lemma 6.3).

6.6 Time Reversibility

Consider a continuous-time Markov chain that is ergodic and let us consider the limiting probabilities P_i from a different point of view than previously. If we consider the sequence of states visited, ignoring the amount of time spent in each state during a visit, then this sequence constitutes a discrete-time Markov chain with transition probabilities P_{ij}. Let us assume that this discrete-time Markov chain, called the embedded chain, is ergodic and denote by π_i its limiting probabilities. That is, the π_i are the unique solution of

$$\pi_i = \sum_j \pi_j P_{ji}, \qquad \text{all } i$$

$$\sum_i \pi_i = 1$$

Now, since π_i represents the proportion of transitions that take the process into state i, and because $1/v_i$ is the mean time spent in state i during a visit, it seems

intuitive that P_i, the proportion of time in state i, should be a weighted average of the π_i where π_i is weighted proportionately to $1/v_i$. That is, it is intuitive that

$$P_i = \frac{\pi_i/v_i}{\sum_j \pi_j/v_j} \tag{6.25}$$

To check the preceding, recall that the limiting probabilities P_i must satisfy

$$v_i P_i = \sum_{j \neq i} P_j q_{ji}, \qquad \text{all } i$$

or equivalently, since $P_{ii} = 0$

$$v_i P_i = \sum_j P_j v_j P_{ji}, \qquad \text{all } i$$

Hence, for the P_is to be given by Eq. (6.25), the following would be necessary:

$$\pi_i = \sum_j \pi_j P_{ji}, \qquad \text{all } i$$

But this, of course, follows since it is in fact the defining equation for the π_is.

Suppose now that the continuous-time Markov chain has been in operation for a long time, and suppose that starting at some (large) time T we trace the process going backward in time. To determine the probability structure of this reversed process, we first note that given we are in state i at some time—say, t—the probability that we have been in this state for an amount of time greater than s is just $e^{-v_i s}$. This is so, since

$$\begin{aligned} &P\{\text{process is in state } i \text{ throughout } [t-s, t] | X(t) = i\} \\ &\quad = \frac{P\{\text{process is in state } i \text{ throughout } [t-s, t]\}}{P\{X(t) = i\}} \\ &\quad = \frac{P\{X(t-s) = i\} e^{-v_i s}}{P\{X(t) = i\}} \\ &\quad = e^{-v_i s} \end{aligned}$$

since for t large $P\{X(t-s) = i\} = P\{X(t) = i\} = P_i$.

In other words, going backward in time, the amount of time the process spends in state i is also exponentially distributed with rate v_i. In addition, as was shown in Section 4.8, the sequence of states visited by the reversed process constitutes a discrete-time Markov chain with transition probabilities Q_{ij} given by

$$Q_{ij} = \frac{\pi_j P_{ji}}{\pi_i}$$

Hence, we see from the preceding that the reversed process is a continuous-time Markov chain with the same transition rates as the forward-time process and with one-stage transition probabilities Q_{ij}. Therefore, the continuous-time Markov chain will

be *time reversible*, in the sense that the process reversed in time has the same probabilistic structure as the original process, if the embedded chain is time reversible. That is, if

$$\pi_i P_{ij} = \pi_j P_{ji}, \qquad \text{for all } i, j$$

Now, using the fact that $P_i = (\pi_i/v_i)/\left(\sum_j \pi_j/v_j\right)$, we see that the preceding condition is equivalent to

$$P_i q_{ij} = P_j q_{ji}, \qquad \text{for all } i, j \tag{6.26}$$

Since P_i is the proportion of time in state i and q_{ij} is the rate when in state i that the process goes to j, the condition of time reversibility is that *the rate at which the process goes directly from state i to state j is equal to the rate at which it goes directly from j to i*. It should be noted that this is exactly the same condition needed for an ergodic discrete-time Markov chain to be time reversible (see Section 4.8).

An application of the preceding condition for time reversibility yields the following proposition concerning birth and death processes.

Proposition 6.5. *An ergodic birth and death process is time reversible.*

Proof. We must show that the rate at which a birth and death process goes from state i to state $i+1$ is equal to the rate at which it goes from $i+1$ to i. In any length of time t the number of transitions from i to $i+1$ must equal to within 1 the number from $i+1$ to i (since between each transition from i to $i+1$ the process must return to i, and this can only occur through $i+1$, and vice versa). Hence, as the number of such transitions goes to infinity as $t \to \infty$, it follows that the rate of transitions from i to $i+1$ equals the rate from $i+1$ to i. ■

Proposition 6.5 can be used to prove the important result that the output process of an $M/M/s$ queue is a Poisson process. We state this as a corollary.

Corollary 6.6. *Consider an $M/M/s$ queue in which customers arrive in accordance with a Poisson process having rate λ and are served by any one of s servers—each having an exponentially distributed service time with rate μ. If $\lambda < s\mu$, then the output process of customers departing is, after the process has been in operation for a long time, a Poisson process with rate λ.*

Proof. Let $X(t)$ denote the number of customers in the system at time t. Since the $M/M/s$ process is a birth and death process, it follows from Proposition 6.5 that $\{X(t), t \geq 0\}$ is time reversible. Going forward in time, the time points at which $X(t)$ increases by 1 constitute a Poisson process since these are just the arrival times of customers. Hence, by time reversibility the time points at which $X(t)$ increases by 1 when we go backward in time also constitute a Poisson process. But these latter points are exactly the points of time when customers depart (see Fig. 6.1). Hence, the departure times constitute a Poisson process with rate λ. ■

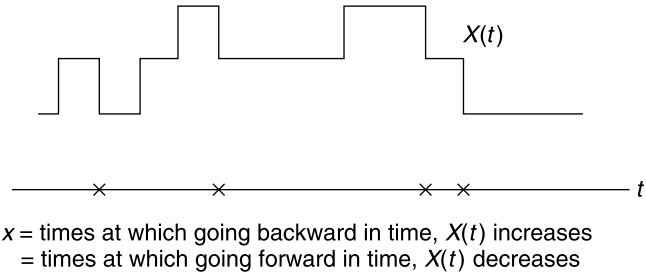

Figure 6.1 The number in the system.

Example 6.17. Consider a first come first serve $M/M/1$ queue, with arrival rate λ and service rate μ, where $\lambda < \mu$, that is in steady state. Given that customer C spends a total of t time units in the system, what is the conditional distribution of the number of others that were present when C arrived?

Solution: Suppose that C arrived at time s and departed at time $s + t$. Because the system is first come first served, the number that were in the system when C arrived is equal to the number of departures of other customers that occur after time s and before time $s + t$, which is equal to the number of arrivals in the reversed process in that interval of time. Now, in the reversed process C would have arrived at time $s + t$ and departed at time s. Because the reversed process is also an $M/M/1$ queueing system, the number of arrivals during that interval of length t is Poisson distributed with mean λt. (For a more direct argument for this result, see Section 8.3.1.) ■

We have shown that a process is time reversible if and only if

$$P_i q_{ij} = P_j q_{ji} \qquad \text{for all } i \neq j$$

Analogous to the result for discrete-time Markov chains, if we can find a probability vector $\mathbf{P}$ that satisfies the preceding then the Markov chain is time reversible and the P_is are the long-run probabilities. That is, we have the following proposition.

Proposition 6.7. *If for some set* $\{P_i\}$

$$\sum_i P_i = 1, \qquad P_i \geq 0$$

and

$$P_i q_{ij} = P_j q_{ji} \qquad \textit{for all } i \neq j \tag{6.27}$$

then the continuous-time Markov chain is time reversible and P_i *represents the limiting probability of being in state* i.

Proof. For fixed i we obtain upon summing Eq. (6.27) over all $j : j \neq i$

$$\sum_{j \neq i} P_i q_{ij} = \sum_{j \neq i} P_j q_{ji}$$

or, since $\sum_{j \neq i} q_{ij} = v_i$,

$$v_i P_i = \sum_{j \neq i} P_j q_{ji}$$

Hence, the P_is satisfy the balance equations and thus represent the limiting probabilities. Because Eq. (6.27) holds, the chain is time reversible. ■

Example 6.18. Consider a set of n machines and a single repair facility to service them. Suppose that when machine $i, i = 1, \ldots, n$, goes down it requires an exponentially distributed amount of work with rate μ_i to get it back up. The repair facility divides its efforts equally among all down components in the sense that whenever there are k down machines $1 \leq k \leq n$ each receives work at a rate of $1/k$ per unit time. Finally, suppose that each time machine i goes back up it remains up for an exponentially distributed time with rate λ_i.

The preceding can be analyzed as a continuous-time Markov chain having 2^n states where the state at any time corresponds to the set of machines that are down at that time. Thus, for instance, the state will be $(i_1, i_2, \ldots, i_k)$ when machines $i_1, \ldots, i_k$ are down and all the others are up. The instantaneous transition rates are as follows:

$$q_{(i_1,\ldots,i_{k-1}),(i_1,\ldots,i_k)} = \lambda_{i_k},$$
$$q_{(i_1,\ldots,i_k),(i_1,\ldots,i_{k-1})} = \mu_{i_k}/k$$

where $i_1, \ldots, i_k$ are all distinct. This follows since the failure rate of machine i_k is always λ_{i_k} and the repair rate of machine i_k when there are k failed machines is μ_{i_k}/k.

Hence, the time reversible equations from (6.27) are

$$P(i_1, \ldots, i_k)\mu_{i_k}/k = P(i_1, \ldots, i_{k-1})\lambda_{i_k}$$

or

$$\begin{aligned} P(i_1, \ldots, i_k) &= \frac{k\lambda_{i_k}}{\mu_{i_k}} P(i_1, \ldots, i_{k-1}) \\ &= \frac{k\lambda_{i_k}}{\mu_{i_k}} \frac{(k-1)\lambda_{i_{k-1}}}{\mu_{i_{k-1}}} P(i_1, \ldots, i_{k-2}) \qquad \text{upon iterating} \\ &= \\ &\vdots \\ &= k! \prod_{j=1}^{k} (\lambda_{i_j}/\mu_{i_j}) P(\phi) \end{aligned}$$

where ϕ is the state in which all components are working. Because

$$P(\phi) + \sum P(i_1, \ldots, i_k) = 1$$

we see that

$$P(\phi) = \left[1 + \sum_{i_1,\ldots,i_k} k! \prod_{j=1}^{k} (\lambda_{i_j}/\mu_{i_j}) \right]^{-1} \tag{6.28}$$

where the preceding sum is over all the $2^n - 1$ nonempty subsets $\{i_1, \ldots, i_k\}$ of $\{1, 2, \ldots, n\}$. Hence, as the time reversible equations are satisfied for this choice of probability vector it follows from Proposition 6.7 that the chain is time reversible and

$$P(i_1, \ldots, i_k) = k! \prod_{j=1}^{k} (\lambda_{i_j}/\mu_{i_j}) P(\phi)$$

with $P(\phi)$ being given by (6.28).

For instance, suppose there are two machines. Then, from the preceding we would have

$$\begin{aligned}
P(\phi) &= \frac{1}{1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2}, \\
P(1) &= \frac{\lambda_1/\mu_1}{1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2}, \\
P(2) &= \frac{\lambda_2/\mu_2}{1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2}, \\
P(1,2) &= \frac{2\lambda_1\lambda_2}{\mu_1\mu_2[1 + \lambda_1/\mu_1 + \lambda_2/\mu_2 + 2\lambda_1\lambda_2/\mu_1\mu_2]}
\end{aligned}$$

■

Consider a continuous-time Markov chain whose state space is S. We say that the Markov chain is truncated to the set $A \subset S$ if q_{ij} is changed to 0 for all $i \in A$, $j \notin A$. That is, transitions out of the class A are no longer allowed, whereas ones in A continue at the same rates as before. A useful result is that if the chain is time reversible, then so is the truncated one.

Proposition 6.8. *A time reversible chain with limiting probabilities P_j, $j \in S$ that is truncated to the set $A \subset S$ and remains irreducible is also time reversible and has limiting probabilities P_j^A given by*

$$P_j^A = \frac{P_j}{\sum_{i \in A} P_i}, \qquad j \in A$$

Proof. By Proposition 6.7 we need to show that, with P_j^A as given,

$$P_i^A q_{ij} = P_j^A q_{ji} \qquad \text{for } i \in A,\ j \in A$$

or, equivalently,

$$P_i q_{ij} = P_j q_{ji} \qquad \text{for } i \in A,\ j \in A$$

But this follows since the original chain is, by assumption, time reversible. ■

Example 6.19. Consider an $M/M/1$ queue in which arrivals finding N in the system do not enter. This finite capacity system can be regarded as a truncation of the $M/M/1$ queue to the set of states $A = \{0, 1, \ldots, N\}$. Since the number in the system in the $M/M/1$ queue is time reversible and has limiting probabilities $P_j = (\lambda/\mu)^j(1 - \lambda/\mu)$ it follows from Proposition 6.8 that the finite capacity model is also time reversible and has limiting probabilities given by

$$P_j = \frac{(\lambda/\mu)^j}{\sum_{i=0}^{N}(\lambda/\mu)^i}, \qquad j = 0, 1, \ldots, N$$ ■

Another useful result is given by the following proposition, whose proof is left as an exercise.

Proposition 6.9. *If $\{X_i(t),\ t \geq 0\}$ are, for $i = 1, \ldots, n$, independent time reversible continuous-time Markov chains, then the vector process $\{(X_1(t), \ldots, X_n(t)),\ t \geq 0\}$ is also a time reversible continuous-time Markov chain.*

Example 6.20. Consider an n-component system where component $i, i = 1, \ldots, n$, functions for an exponential time with rate λ_i and then fails; upon failure, repair begins on component i, with the repair taking an exponentially distributed time with rate μ_i. Once repaired, a component is as good as new. The components act independently except that when there is only one working component the system is temporarily shut down until a repair has been completed; it then starts up again with two working components.

(a) What proportion of time is the system shut down?
(b) What is the (limiting) averaging number of components that are being repaired?

Solution: Consider first the system without the restriction that it is shut down when a single component is working. Letting $X_i(t),\ i = 1, \ldots, n$, equal 1 if component i is working at time t and 0 if it failed, then $\{X_i(t),\ t \geq 0\}$, $i = 1, \ldots, n$, are independent birth and death processes. Because a birth and death process is time reversible, it follows from Proposition 6.9 that the process $\{(X_1(t), \ldots, X_n(t)),\ t \geq 0\}$ is also time reversible. Now, with

$$P_i(j) = \lim_{t \to \infty} P\{X_i(t) = j\}, \qquad j = 0, 1$$

we have

$$P_i(1) = \frac{\mu_i}{\mu_i + \lambda_i}, \qquad P_i(0) = \frac{\lambda_i}{\mu_i + \lambda_i}$$

Also, with

$$P(j_1, \ldots, j_n) = \lim_{t \to \infty} P\{X_i(t) = j_i, i = 1, \ldots, n\}$$

it follows, by independence, that

$$P(j_1, \ldots, j_n) = \prod_{i=1}^{n} P_i(j_i), \qquad j_i = 0, 1, \ i = 1, \ldots, n$$

Now, note that shutting down the system when only one component is working is equivalent to truncating the preceding unconstrained system to the set consisting of all states except the one having all components down. Therefore, with P_T denoting a probability for the truncated system, we have from Proposition 6.8 that

$$P_T(j_1, \ldots, j_n) = \frac{P(j_1, \ldots, j_n)}{1 - C}, \qquad \sum_{i=1}^{n} j_i > 0$$

where

$$C = P(0, \ldots, 0) = \prod_{j=1}^{n} \lambda_j/(\mu_j + \lambda_j)$$

Hence, letting $(\mathbf{0}, 1_i) = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the n vector of zeroes and ones whose single 1 is in the ith place, we have

$$\begin{aligned} P_T(\text{system is shut down}) &= \sum_{i=1}^{n} P_T(\mathbf{0}, 1_i) \\ &= \frac{1}{1 - C} \sum_{i=1}^{n} \left(\frac{\mu_i}{\mu_i + \lambda_i} \right) \prod_{j \neq i} \left(\frac{\lambda_j}{\mu_j + \lambda_j} \right) \\ &= \frac{C \sum_{i=1}^{n} \mu_i/\lambda_i}{1 - C} \end{aligned}$$

Let R denote the number of components being repaired. Then with I_i equal to 1 if component i is being repaired and 0 otherwise, we have for the unconstrained (nontruncated) system that

$$E[R] = E\left[\sum_{i=1}^{n} I_i \right] = \sum_{i=1}^{n} P_i(0) = \sum_{i=1}^{n} \lambda_i/(\mu_i + \lambda_i)$$

But, in addition,

$$\begin{aligned} E[R] = {} & E[R|\text{all components are in repair}]C \\ & + E[R|\text{not all components are in repair}](1 - C) \end{aligned}$$

$$= nC + E_T[R](1 - C)$$

implying that

$$E_T[R] = \frac{\sum_{i=1}^{n} \lambda_i/(\mu_i + \lambda_i) - nC}{1 - C}$$ ■

6.7 The Reversed Chain

Consider an ergodic continuous-time Markov chain whose state space is S and which has instantaneous transition rates q_{ij} and limiting probabilities P_i, $i \in S$, and suppose that this chain that has been in operation for a long (in theory, an infinite) time. Then, it follows from results in the previous section that the process of states going backwards in time is also a continuous time Markov chain, having instantaneous transition rates q_{ij}^* that satisfy

$$P_i q_{ij}^* = P_j q_{ji}, \quad i \neq j$$

The reverse chain is a very useful concept even in cases where it differs from the forward chain (that is, even in cases where the chain is not time reversible).

To begin, note that the amount of time the reverse chain spends in state i during a visit is exponential with rate $v_i^* \equiv \sum_{j \neq i} q_{ij}^*$. Because the amount of time the process spends in a state i during a visit will be the same whether the chain is observed in the usual (forward) or in the reverse direction of time, it follows that the distribution of the time that the reverse chain spends in state i during a visit should be the same as the distribution of the time that the forward chain spends in that state during a visit. That is, we should have that

$$v_i^* = v_i$$

Moreover, because the proportion of time that the chain spends in state i would be the same whether one was observing the chain in the usual (forward) direction of time or in the reverse direction, the two chains should intuitively have the same limiting probabilities.

Proposition 6.10. *Let a continuous-time Markov chain have instantaneous transition rates q_{ij} and limiting probabilities P_i, $i \in S$, and let q_{ij}^* be the instantaneous rates of the reversed chain. Then, with $v_i^* = \sum_{j \neq i} q_{ij}^*$ and $v_i = \sum_{j \neq i} q_{ij}$*

$$v_i^* = v_i$$

Moreover P_i, $i \in S$, are also the limiting probabilities of the reversed chain.

Proof. Using that $P_i q_{ij}^* = P_j q_{ji}$ we see that

$$\sum_{j \neq i} q_{ij}^* = \sum_{j \neq i} P_j q_{ji}/P_i = v_i P_i / P_i = v_i$$

where the preceding used (from (6.18)) that $\sum_{j \neq i} P_j q_{ji} = v_i P_i$.

That the reversed chain has the same limiting probabilities as does the forward chain can be formally proven by showing that the P_j satisfy the balance equations of the reversed chain:

$$v_j^* P_j = \sum_{k \neq j} P_k q_{kj}^*, \quad j \in S$$

Now, because $v_j^* = v_j$ and $P_k q_{kj}^* = P_j q_{jk}$, the preceding equations are equivalent to

$$v_j P_j = \sum_{k \neq j} P_j q_{jk}, \quad j \in S$$

which are just the balance equations for the forward chain, which are known to be satisfied by the P_j. ■

That the long-run proportions for the reverse chain are the same as for the forward chain makes it easy to understand why

$$P_i q_{ij}^* = P_j q_{ji}, \quad i \neq j$$

Because P_i is the proportion of time the reverse chain spends in state i and q_{ij}^* is the rate, when in i, that it makes a transition into state j, it follows that $P_i q_{ij}^*$ is the rate at which the reversed chain makes transitions from i to j. Similarly, $P_j q_{ji}$ is the rate at which the forward chain makes transitions from j to i. Because every transition from j to i in the (forward) Markov chain would be seen as a transition from i to j by someone looking backwards in time, it is evident that $P_i q_{ij}^* = P_j q_{ji}$.

The following proposition shows that if one can find a solution of the "reverse chain equations" then the solution is unique and yields the limiting probabilities.

Proposition 6.11. *Let q_{ij} be the transition rates of an irreducible continuous time Markov chain. If one can find values q_{ij}^* and a collection of positive values P_i that sum to* 1*, such that*

$$P_i q_{ij}^* = P_j q_{ji}, \quad i \neq j \tag{6.29}$$

and

$$\sum_{j \neq i} q_{ij}^* = \sum_{j \neq i} q_{ij}, \quad i \in S \tag{6.30}$$

then q_{ij}^* are the transition rates of the reversed chain and P_i are the limiting probabilities (for both chains).

Proof. We show that the P_i are the limiting probabilities by showing that they satisfy the balance Eqs. (6.18). To show this, sum (6.29) over all j, $j \neq i$, to obtain

$$P_i \sum_{j \neq i} q_{ij}^* = \sum_{j \neq i} P_j q_{ji}, \quad i \in S$$

$$\text{forward transitions :} \quad N \to N-1 \to \cdots \to 2 \to 1 \to 0$$

$$\text{reverse transitions :} \quad 0 \to 1 \to 2 \to \ldots \to N-1 \to N$$

Figure 6.2 Forward and Reverse Transitions.

Using (6.30) now shows that

$$P_i \sum_{j \neq i} q_{ij} = \sum_{j \neq i} P_j q_{ji}$$

Because $\sum_i P_i = 1$ we see that the P_i satisfy the balance equations and are thus the limiting probabilities. Because $P_i q^*_{ij} = P_j q_{ji}$ it also follows that q^*_{ij} are the transition rates of the reversed chain. ■

Suppose now that the structure of the continuous time Markov chain enables us to make a guess as to the transition rates of the reversed chain. Assuming that this guess satisfies Eq. (6.30) of Proposition 6.11, we can then verify the correctness of our guess by seeing whether there are probabilities that satisfy Eqs. (6.29). If there are such probabilities, our guess is correct and we have also found the limiting probabilities; if there are not, our guess is incorrect.

Example 6.21. Consider a continuous-time Markov chain whose states are the non-negative integers. Suppose that a transition out of state 0 goes to state i with probability α_i, $\sum_{i=1}^{\infty} \alpha_i = 1$; whereas a transition out of state $i > 0$ always goes to state $i-1$. That is, the instantaneous transition rates of this chain are, for $i > 0$

$$q_{0i} = v_0 \alpha_i$$
$$q_{i,i-1} = v_i$$

Let N be a random variable having the distribution of the next state from state 0; that is, $P(N = i) = \alpha_i$, $i > 0$. Also, say that a cycle begins each time the chain goes to state 0. Because the forward chain goes from 0 to N and then continually moves one step closer to 0 until reaching that state, it follows that the states in the reverse chain would continually increase by 1 until it reaches N at which point it goes back to state 0 (see Fig. 6.2).

Now, if the chain is currently in state i then the value of N for that cycle must be at least i. Hence, the next state of the reversed chain will be 0 with probability

$$P(N = i | N \geq i) = \frac{P(N = i)}{P(N \geq i)} = \frac{\alpha_i}{P(N \geq i)}$$

and will be $i + 1$ with probability

$$1 - P(N = i | N \geq i) = P(N \geq i + 1 | N \geq i) = \frac{P(N \geq i + 1)}{P(N \geq i)}$$

Because the reversed chain spends the same time in a state during each visit as does the forward chain, it thus appears that the transition rates of the reversed chain are

$$q^*_{i,0} = v_i \frac{\alpha_i}{P(N \geq i)}, \quad i > 0$$
$$q^*_{i,i+1} = v_i \frac{P(N \geq i+1)}{P(N \geq i)}, \quad i \geq 0$$

Based on the preceding guess, the reversed time equations $P_0 q_{0i} = P_i q^*_{i0}$ and $P_i q_{i,i-1} = P_{i-1} q^*_{i-1,i}$ become

$$P_0 v_0 \alpha_i = P_i v_i \frac{\alpha_i}{P(N \geq i)}, \quad i \geq 1 \tag{6.31}$$

and

$$P_i v_i = P_{i-1} v_{i-1} \frac{P(N \geq i)}{P(N \geq i-1)}, \quad i \geq 1 \tag{6.32}$$

The set of Eqs. (6.31) gives

$$P_i = P_0 v_0 P(N \geq i)/v_i, \quad i \geq 1$$

As the preceding equation is also valid when $i = 0$ (since $P(N \geq 0) = 1$), we obtain upon summing over all i that

$$1 = \sum_i P_i = P_0 v_0 \sum_{i=0}^{\infty} P(N \geq i)/v_i$$

Thus,

$$P_i = \frac{P(N \geq i)/v_i}{\sum_{i=0}^{\infty} P(N \geq i)/v_i}, \quad i \geq 0$$

To show that the set of Eqs. (6.32) is also satisfied by the preceding values of P_i, note that, with $C = 1/\sum_{i=0}^{\infty} P(N \geq i)/v_i$,

$$\frac{v_i P_i}{P(N \geq i)} = C = \frac{v_{i-1} P_{i-1}}{P(N \geq i-1)}$$

which immediately shows that Eqs. (6.32) are also satisfied. Because we chose the transition rates of the reversed chain to be such that it spent as much time in state i during a visit as does the forward chain, there is no need to check Condition (6.30) of Proposition 6.11, and so the stationary probabilities are as given. ■

Example 6.22 (A Sequential Queueing System)**.** Consider a two-server queueing system in which customers arrive at server 1 in accordance with a Poisson process with rate λ. An arrival at server 1 either enters service if server 1 is free or joins the queue

if server 1 is busy. After completion of service at server 1 the customer moves over to server 2, where it either enters service if server 2 is free or join its queue otherwise. After completion of service at server 2 a customer departs the system. The service times at servers 1 and 2 are exponential with rates μ_1 and μ_2 respectively. All service times are independent and are also independent of the arrival process.

The preceding model can be analyzed as a continuous-time Markov chain whose state is (n, m) if there are currently n customers with server 1 and m with server 2. The instantaneous transition rates of this chain are

$$\begin{aligned} q_{(n-1,m),(n,m)} &= \lambda, \quad n > 0 \\ q_{(n+1,m-1),(n,m)} &= \mu_1, \quad m > 0 \\ q_{(n,m+1),(n,m)} &= \mu_2 \end{aligned}$$

To find the limiting probabilities, let us first consider the chain going backwards in time. Because in real time the total number in the system decreases at moments when customers depart server 2, looking backwards the total number in the system will at those moments increase by having an added customer at server 2. Similarly while in real time the number will increase when a customer arrives at server 1, in the reverse process at that moment there will be a decrease in the number at server 1. Because the times spent in service at server i will be the same whether looking in forward or in reverse time, it appears that the reverse process is a two-server system in which customers arrive first at server 2, then go to server 1, and then depart the system, with their service times at server i being exponential with rate μ_i, $i = 1, 2$. Now the arrival rate to server 2 in the reverse process is equal to the departure rate from the system in the forward process and this must equal the arrival rate λ of the forward process. (If the departure rate of the forward process was less than the arrival rate, then the queue size would build to infinity and there would not be any limiting probabilities.) Although it is not clear that the arrival process of customers to server 2 in the reverse process is a Poisson process, let us guess that this is the case and then use Proposition 6.11 to determine whether our guess is correct.

So, let us guess that the reverse process is a sequential queue where customers arrive at server 2 according to a Poisson process with rate λ, and after receiving service at server 2 move over to server 1, and after receiving service at server 1 depart the system. In addition, the service times at server i are exponential with rate μ_i, $i = 1, 2$. Now, if this were true then the transition rates of the reverse chain would be

$$\begin{aligned} q^*_{(n,m),(n-1,m)} &= \mu_1, \quad n > 0 \\ q^*_{(n,m),(n+1,m-1)} &= \mu_2, \quad m > 0 \\ q^*_{(n,m),(n,m+1)} &= \lambda \end{aligned}$$

The rate at which a chain with transition rates q^* departs from state (n, m) is

$$q^*_{(n,m),(n-1,m)} + q^*_{(n,m),(n+1,m-1)} + q^*_{(n,m),(n,m+1)} = \mu_1 I\{n > 0\} + \mu_2 I\{m > 0\} + \lambda$$

where $I\{k > 0\}$ is equal to 1 when $k > 0$ and is equal to 0 otherwise. As the preceding is also the rate at which the forward process departs from state (n, m), the Condition (6.30) of Proposition 6.11 is satisfied.

Using the preceding conjectured reverse time transition rates, the reverse time equations would be

$$P_{n-1,m}\,\lambda = P_{n,m}\,\mu_1, \quad n > 0 \tag{6.33}$$

$$P_{n+1,m-1}\,\mu_1 = P_{n,m}\,\mu_2, \quad m > 0 \tag{6.34}$$

$$P_{n,m+1}\,\mu_2 = P_{n,m}\,\lambda \tag{6.35}$$

Writing (6.33) as $P_{n,m} = (\lambda/\mu_1)\,P_{n-1,m}$ and iterating, yields that

$$P_{n,m} = (\lambda/\mu_1)^2\,P_{n-2,m} = \cdots = (\lambda/\mu_1)^n\,P_{0,m}$$

Letting $n = 0, m = m - 1$ in Eq. (6.35) shows that $P_{0,m} = (\lambda/\mu_2)\,P_{0,m-1}$, which yields upon iteration that

$$P_{0,m} = (\lambda/\mu_2)^2\,P_{0,m-2} = \cdots = (\lambda/\mu_2)^m\,P_{0,0}$$

Hence, the conjectured reversed time equations imply that

$$P_{n,m} = (\lambda/\mu_1)^n\,(\lambda/\mu_2)^m\,P_{0,0}$$

Using that $\sum_n \sum_m P_{n,m} = 1$, gives

$$P_{n,m} = (\lambda/\mu_1)^n\,(1 - \lambda/\mu_1)\,(\lambda/\mu_2)^m\,(1 - \lambda/\mu_2)$$

As it is easy to check that all the conjectured reverse time Eqs. (6.33), (6.34), and (6.35) are satisfied for the preceding choice of $P_{n,m}$, it follows that they are the limiting probabilities. Hence, we have shown that in steady state the numbers of customers at the two servers are independent, with the number at server i distributed as the number in the system of an $M/M/1$ queue with Poisson arrival rate λ and exponential service rate μ_i, $i = 1, 2$. (See Example 6.14.)

6.8 Uniformization

Consider a continuous-time Markov chain in which the mean time spent in a state is the same for all states. That is, suppose that $v_i = v$, for all states i. In this case since the amount of time spent in each state during a visit is exponentially distributed with rate v, it follows that if we let $N(t)$ denote the number of state transitions by time t, then $\{N(t), t \geq 0\}$ will be a Poisson process with rate v.

To compute the transition probabilities $P_{ij}(t)$, we can condition on $N(t)$:

$$P_{ij}(t) = P\{X(t) = j | X(0) = i\}$$

$$= \sum_{n=0}^{\infty} P\{X(t) = j | X(0) = i,\ N(t) = n\} P\{N(t) = n | X(0) = i\}$$
$$= \sum_{n=0}^{\infty} P\{X(t) = j | X(0) = i,\ N(t) = n\} e^{-vt} \frac{(vt)^n}{n!}$$

Now, the fact that there have been n transitions by time t tells us something about the amount of time spent in each of the first n states visited, but since the distribution of time spent in each state is the same for all states, it follows that knowing that $N(t) = n$ gives us no information about which states were visited. Hence,

$$P\{X(t) = j | X(0) = i,\ N(t) = n\} = P_{ij}^n$$

where P_{ij}^n is just the n-stage transition probability associated with the discrete-time Markov chain with transition probabilities P_{ij}; and so when $v_i \equiv v$

$$P_{ij}(t) = \sum_{n=0}^{\infty} P_{ij}^n e^{-vt} \frac{(vt)^n}{n!} \tag{6.36}$$

Eq. (6.36) is often useful from a computational point of view since it enables us to approximate $P_{ij}(t)$ by taking a partial sum and then computing (by matrix multiplication of the transition probability matrix) the relevant n stage probabilities P_{ij}^n.

Whereas the applicability of Eq. (6.36) would appear to be quite limited since it supposes that $v_i \equiv v$, it turns out that most Markov chains can be put in that form by the trick of allowing fictitious transitions from a state to itself. To see how this works, consider any Markov chain for which the v_i are bounded, and let v be any number such that

$$v_i \leq v, \qquad \text{for all } i \tag{6.37}$$

When in state i, the process actually leaves at rate v_i; but this is equivalent to supposing that transitions occur at rate v, but only the fraction v_i/v of transitions are real ones (and thus real transitions occur at rate v_i) and the remaining fraction $1 - v_i/v$ are fictitious transitions that leave the process in state i. In other words, any Markov chain satisfying Condition (6.37) can be thought of as being a process that spends an exponential amount of time with rate v in state i and then makes a transition to j with probability P_{ij}^*, where

$$P_{ij}^* = \begin{cases} 1 - \frac{v_i}{v}, & j = i \\ \frac{v_i}{v} P_{ij}, & j \neq i \end{cases} \tag{6.38}$$

Hence, from Eq. (6.36) we have that the transition probabilities can be computed by

$$P_{ij}(t) = \sum_{n=0}^{\infty} P_{ij}^{*n} e^{-vt} \frac{(vt)^n}{n!}$$

where P_{ij}^* are the n-stage transition probabilities corresponding to Eq. (6.38). This technique of uniformizing the rate in which a transition occurs from each state by introducing transitions from a state to itself is known as *uniformization*.

Example 6.23. Let us reconsider Example 6.11, which models the workings of a machine—either on or off—as a two-state continuous-time Markov chain with

$$P_{01} = P_{10} = 1,$$
$$v_0 = \lambda, \qquad v_1 = \mu$$

Letting $v = \lambda + \mu$, the uniformized version of the preceding is to consider it a continuous-time Markov chain with

$$P_{00} = \frac{\mu}{\lambda+\mu} = 1 - P_{01},$$
$$P_{10} = \frac{\mu}{\lambda+\mu} = 1 - P_{11},$$
$$v_i = \lambda + \mu, \qquad i = 1, 2$$

As $P_{00} = P_{10}$, it follows that the probability of a transition into state 0 is equal to $\mu/(\lambda+\mu)$ no matter what the present state. Because a similar result is true for state 1, it follows that the n-stage transition probabilities are given by

$$P_{i0}^n = \frac{\mu}{\lambda+\mu}, \qquad n \geq 1,\ i = 0, 1$$
$$P_{i1}^n = \frac{\lambda}{\lambda+\mu}, \qquad n \geq 1,\ i = 0, 1$$

Hence,

$$\begin{aligned} P_{00}(t) &= \sum_{n=0}^{\infty} P_{00}^n e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!} \\ &= e^{-(\lambda+\mu)t} + \sum_{n=1}^{\infty} \left(\frac{\mu}{\lambda+\mu}\right) e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!} \\ &= e^{-(\lambda+\mu)t} + [1 - e^{-(\lambda+\mu)t}] \frac{\mu}{\lambda+\mu} \\ &= \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \end{aligned}$$

Similarly,

$$\begin{aligned} P_{11}(t) &= \sum_{n=0}^{\infty} P_{11}^n e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!} \\ &= e^{-(\lambda+\mu)t} + [1 - e^{-(\lambda+\mu)t}] \frac{\lambda}{\lambda+\mu} \end{aligned}$$

$$= \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}$$

The remaining probabilities are

$$P_{01}(t) = 1 - P_{00}(t) = \frac{\lambda}{\lambda+\mu}[1 - e^{-(\lambda+\mu)t}],$$
$$P_{10}(t) = 1 - P_{11}(t) = \frac{\mu}{\lambda+\mu}[1 - e^{-(\lambda+\mu)t}]$$ ■

Example 6.24. Consider the two-state chain of Example 6.23 and suppose that the initial state is state 0. Let $O(t)$ denote the total amount of time that the process is in state 0 during the interval $(0, t)$. The random variable $O(t)$ is often called the *occupation time*. We will now compute its mean.

If we let

$$I(s) = \begin{cases} 1, & \text{if } X(s) = 0 \\ 0, & \text{if } X(s) = 1 \end{cases}$$

then we can represent the occupation time by

$$O(t) = \int_0^t I(s)\, ds$$

Taking expectations and using the fact that we can take the expectation inside the integral sign (since an integral is basically a sum), we obtain

$$\begin{aligned} E[O(t)] &= \int_0^t E[I(s)]\, ds \\ &= \int_0^t P\{X(s) = 0\}\, ds \\ &= \int_0^t P_{00}(s)\, ds \\ &= \frac{\mu}{\lambda+\mu} t + \frac{\lambda}{(\lambda+\mu)^2}\{1 - e^{-(\lambda+\mu)t}\} \end{aligned}$$

where the final equality follows by integrating

$$P_{00}(s) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)s}$$

(For another derivation of $E[O(t)]$, see Exercise 46.) ■

6.9 Computing the Transition Probabilities

For any pair of states i and j, let

$$r_{ij} = \begin{cases} q_{ij}, & \text{if } i \neq j \\ -v_i, & \text{if } i = j \end{cases}$$

Using this notation, we can rewrite the Kolmogorov backward equations

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

and the forward equations

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t)$$

as follows:

$$P'_{ij}(t) = \sum_k r_{ik} P_{kj}(t) \qquad \text{(backward)}$$
$$P'_{ij}(t) = \sum_k r_{kj} P_{ik}(t) \qquad \text{(forward)}$$

This representation is especially revealing when we introduce matrix notation. Define the matrices $\mathbf{R}$ and $\mathbf{P}(t)$, $\mathbf{P}'(t)$ by letting the element in row i, column j of these matrices be, respectively, r_{ij}, $P_{ij}(t)$, and $P'_{ij}(t)$. Since the backward equations say that the element in row i, column j of the matrix $\mathbf{P}'(t)$ can be obtained by multiplying the ith row of the matrix $\mathbf{R}$ by the jth column of the matrix $\mathbf{P}(t)$, it is equivalent to the matrix equation

$$\mathbf{P}'(t) = \mathbf{R}\mathbf{P}(t) \tag{6.39}$$

Similarly, the forward equations can be written as

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R} \tag{6.40}$$

Now, just as the solution of the scalar differential equation

$$f'(t) = cf(t)$$

(or, equivalent, $f'(t) = f(t)c$) is

$$f(t) = f(0)e^{ct}$$

it can be shown that the solution of the matrix differential Eqs. (6.39) and (6.40) is given by

$$\mathbf{P}(t) = \mathbf{P}(0)e^{\mathbf{R}t}$$

Since $\mathbf{P}(0) = \mathbf{I}$ (the identity matrix), this yields that

$$\mathbf{P}(t) = e^{\mathbf{R}t} \tag{6.41}$$

where the matrix $e^{\mathbf{R}t}$ is defined by

$$e^{\mathbf{R}t} = \sum_{n=0}^{\infty} \mathbf{R}^n \frac{t^n}{n!} \tag{6.42}$$

with $\mathbf{R}^n$ being the (matrix) multiplication of $\mathbf{R}$ by itself n times.

The direct use of Eq. (6.42) to compute $\mathbf{P}(t)$ turns out to be very inefficient for two reasons. First, since the matrix $\mathbf{R}$ contains both positive and negative elements (remember the off-diagonal elements are the q_{ij} while the ith diagonal element is $-v_i$), there is the problem of computer round-off error when we compute the powers of $\mathbf{R}$. Second, we often have to compute many of the terms in the infinite sum (6.42) to arrive at a good approximation. However, there are certain indirect ways that we can utilize the relation in (6.41) to efficiently approximate the matrix $\mathbf{P}(t)$. We now present two of these methods.

Approximation Method 1 Rather than using Eq. (6.42) to compute $e^{\mathbf{R}t}$, we can use the matrix equivalent of the identity

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$$

which states that

$$e^{\mathbf{R}t} = \lim_{n\to\infty} \left(\mathbf{I} + \mathbf{R}\frac{t}{n}\right)^n$$

Thus, if we let n be a power of 2, say, $n = 2^k$, then we can approximate $\mathbf{P}(t)$ by raising the matrix $\mathbf{M} = \mathbf{I} + \mathbf{R}t/n$ to the nth power, which can be accomplished by k matrix multiplications (by first multiplying $\mathbf{M}$ by itself to obtain $\mathbf{M}^2$ and then multiplying that by itself to obtain $\mathbf{M}^4$ and so on). In addition, since only the diagonal elements of $\mathbf{R}$ are negative (and the diagonal elements of the identity matrix $\mathbf{I}$ are equal to 1), by choosing n large enough we can guarantee that the matrix $\mathbf{I} + \mathbf{R}t/n$ has all nonnegative elements.

Approximation Method 2 A second approach to approximating $e^{\mathbf{R}t}$ uses the identity

$$\begin{aligned} e^{-\mathbf{R}t} &= \lim_{n\to\infty} \left(\mathbf{I} - \mathbf{R}\frac{t}{n}\right)^n \\ &\approx \left(\mathbf{I} - \mathbf{R}\frac{t}{n}\right)^n \qquad \text{for } n \text{ large} \end{aligned}$$

and thus

$$\mathbf{P}(t) = e^{\mathbf{R}t} \approx \left(\mathbf{I} - \mathbf{R}\frac{t}{n}\right)^{-n} = \left[\left(\mathbf{I} - \mathbf{R}\frac{t}{n}\right)^{-1}\right]^{n}$$

Hence, if we again choose n to be a large power of 2, say, $n = 2^k$, we can approximate $\mathbf{P}(t)$ by first computing the inverse of the matrix $\mathbf{I} - \mathbf{R}t/n$ and then raising that matrix to the nth power (by utilizing k matrix multiplications). It can be shown that the matrix $(\mathbf{I} - \mathbf{R}t/n)^{-1}$ will have only nonnegative elements.

Remark. Both of the preceding computational approaches for approximating $\mathbf{P}(t)$ have probabilistic interpretations (see Exercises 49 and 50).

Exercises

1. A population of organisms consists of both male and female members. In a small colony any particular male is likely to mate with any particular female in any time interval of length h, with probability $\lambda h + o(h)$. Each mating immediately produces one offspring, equally likely to be male or female. Let $N_1(t)$ and $N_2(t)$ denote the number of males and females in the population at t. Derive the parameters of the continuous-time Markov chain $\{N_1(t), N_2(t)\}$, i.e., the v_i, P_{ij} of Section 6.2.

***2.** Suppose that a one-celled organism can be in one of two states—either A or B. An individual in state A will change to state B at an exponential rate α; an individual in state B divides into two new individuals of type A at an exponential rate β. Define an appropriate continuous-time Markov chain for a population of such organisms and determine the appropriate parameters for this model.

3. Consider two machines that are maintained by a single repairman. Machine i functions for an exponential time with rate μ_i before breaking down, $i = 1, 2$. The repair times (for either machine) are exponential with rate μ. Can we analyze this as a birth and death process? If so, what are the parameters? If not, how can we analyze it?

***4.** Potential customers arrive at a single-server station in accordance with a Poisson process with rate λ. However, if the arrival finds n customers already in the station, then he will enter the system with probability α_n. Assuming an exponential service rate μ, set this up as a birth and death process and determine the birth and death rates.

5. There are N individuals in a population, some of whom have a certain infection that spreads as follows. Contacts between two members of this population occur in accordance with a Poisson process having rate λ. When a contact occurs, it is equally likely to involve any of the $\binom{N}{2}$ pairs of individuals in the

population. If a contact involves an infected and a noninfected individual, then with probability p the noninfected individual becomes infected. Once infected, an individual remains infected throughout. Let $X(t)$ denote the number of infected members of the population at time t.

(a) Is $\{X(t), t \geq 0\}$ a continuous-time Markov chain?

(b) Specify its type.

(c) Starting with a single infected individual, what is the expected time until all members are infected?

6. Consider a birth and death process with birth rates $\lambda_i = (i+1)\lambda, i \geq 0$, and death rates $\mu_i = i\mu, i \geq 0$.

(a) Determine the expected time to go from state 0 to state 4.

(b) Determine the expected time to go from state 2 to state 5.

(c) Determine the variances in parts (a) and (b).

***7.** Individuals join a club in accordance with a Poisson process with rate λ. Each new member must pass through k consecutive stages to become a full member of the club. The time it takes to pass through each stage is exponentially distributed with rate μ. Let $N_i(t)$ denote the number of club members at time t who have passed through exactly i stages, $i = 1, \ldots, k-1$. Also, let $\mathbf{N}(t) = (N_1(t), N_2(t), \ldots, N_{k-1}(t))$.

(a) Is $\{\mathbf{N}(t), t \geq 0\}$ a continuous-time Markov chain?

(b) If so, give the infinitesimal transition rates. That is, for any state $\mathbf{n} = (n_1, \ldots, n_{k-1})$ give the possible next states along with their infinitesimal rates.

8. Consider two machines, both of which have an exponential lifetime with mean $1/\lambda$. There is a single repairman that can service machines at an exponential rate μ. Set up the Kolmogorov backward equations; you need not solve them.

9. The birth and death process with parameters $\lambda_n = 0$ and $\mu_n = \mu, n > 0$ is called a pure death process. Find $P_{ij}(t)$.

10. Consider two machines. Machine i operates for an exponential time with rate λ_i and then fails; its repair time is exponential with rate $\mu_i, i = 1, 2$. The machines act independently of each other. Define a four-state continuous-time Markov chain that jointly describes the condition of the two machines. Use the assumed independence to compute the transition probabilities for this chain and then verify that these transition probabilities satisfy the forward and backward equations.

***11.** Consider a Yule process starting with a single individual—that is, suppose $X(0) = 1$. Let T_i denote the time it takes the process to go from a population of size i to one of size $i+1$.

(a) Argue that $T_i, i = 1, \ldots, j$, are independent exponentials with respective rates $i\lambda$.

(b) Let $X_1, \ldots, X_j$ denote independent exponential random variables each having rate λ, and interpret X_i as the lifetime of component i. Argue that $\max(X_1, \ldots, X_j)$ can be expressed as

$$\max(X_1, \ldots, X_j) = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_j$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_j$ are independent exponentials with respective rates $j\lambda, (j-1)\lambda, \ldots, \lambda$.

Hint: Interpret ε_i as the time between the $i-1$ and the ith failure.

(c) Using (a) and (b) argue that

$$P\{T_1 + \cdots + T_j \le t\} = (1 - e^{-\lambda t})^j$$

(d) Use (c) to obtain

$$P_{1j}(t) = (1 - e^{-\lambda t})^{j-1} - (1 - e^{-\lambda t})^j = e^{-\lambda t}(1 - e^{-\lambda t})^{j-1}$$

and hence, given $X(0) = 1$, $X(t)$ has a geometric distribution with parameter $p = e^{-\lambda t}$.

(e) Now conclude that

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i}$$

12. Each individual in a biological population is assumed to give birth at an exponential rate λ, and to die at an exponential rate μ. In addition, there is an exponential rate of increase θ due to immigration. However, immigration is not allowed when the population size is N or larger.

(a) Set this up as a birth and death model.

(b) If $N = 3, 1 = \theta = \lambda, \mu = 2$, determine the proportion of time that immigration is restricted.

13. A small barbershop, operated by a single barber, has room for at most two customers. Potential customers arrive at a Poisson rate of three per hour, and the successive service times are independent exponential random variables with mean $\frac{1}{4}$ hour.

(a) What is the average number of customers in the shop?

(b) What is the proportion of potential customers that enter the shop?

(c) If the barber could work twice as fast, how much more business would he do?

14. Consider an irreducible continuous time Markov chain whose state space is the nonnegative integers, having instantaneous transition rates $q_{i,j}$ and stationary probabilities $P_i, i \ge 0$. Let T be a given set of states, and let X_n be the state at the moment of the nth transition into a state in T.

(a) Argue that $\{X_n, n \ge 1\}$ is a Markov chain.

(b) At what rate does the continuous time Markov chain make transitions that go into state j.

(c) For $i \in T$, find the long run proportion of transitions of the Markov chain $\{X_n, n \ge 1\}$ that are into state i.

15. A service center consists of two servers, each working at an exponential rate of two services per hour. If customers arrive at a Poisson rate of three per hour, then, assuming a system capacity of at most three customers,

(a) what fraction of potential customers enter the system?

(b) what would the value of part (a) be if there was only a single server, and his rate was twice as fast (that is, $\mu = 4$)?

***16.** The following problem arises in molecular biology. The surface of a bacterium consists of several sites at which foreign molecules—some acceptable and some not—become attached. We consider a particular site and assume that molecules arrive at the site according to a Poisson process with parameter λ. Among these molecules a proportion α is acceptable. Unacceptable molecules stay at the site for a length of time that is exponentially distributed with parameter μ_1, whereas an acceptable molecule remains at the site for an exponential time with rate μ_2. An arriving molecule will become attached only if the site is free of other molecules. What percentage of time is the site occupied with an acceptable (unacceptable) molecule?

17. Each time a machine is repaired it remains up for an exponentially distributed time with rate λ. It then fails, and its failure is either of two types. If it is a type 1 failure, then the time to repair the machine is exponential with rate μ_1; if it is a type 2 failure, then the repair time is exponential with rate μ_2. Each failure is, independently of the time it took the machine to fail, a type 1 failure with probability p and a type 2 failure with probability $1 - p$. What proportion of time is the machine down due to a type 1 failure? What proportion of time is it down due to a type 2 failure? What proportion of time is it up?

18. After being repaired, a machine functions for an exponential time with rate λ and then fails. Upon failure, a repair process begins. The repair process proceeds sequentially through k distinct phases. First a phase 1 repair must be performed, then a phase 2, and so on. The times to complete these phases are independent, with phase i taking an exponential time with rate $\mu_i, i = 1, \ldots, k$.

(a) What proportion of time is the machine undergoing a phase i repair?

(b) What proportion of time is the machine working?

***19.** A single repairperson looks after both machines 1 and 2. Each time it is repaired, machine i stays up for an exponential time with rate $\lambda_i, i = 1, 2$. When machine i fails, it requires an exponentially distributed amount of work with rate μ_i to complete its repair. The repairperson will always service machine 1 when it is down. For instance, if machine 1 fails while 2 is being repaired, then the repairperson will immediately stop work on machine 2 and start on 1. What proportion of time is machine 2 down?

20. There are two machines, one of which is used as a spare. A working machine will function for an exponential time with rate λ and will then fail. Upon failure, it is immediately replaced by the other machine if that one is in working order, and it goes to the repair facility. The repair facility consists of a single person who takes an exponential time with rate μ to repair a failed machine. At the repair facility, the newly failed machine enters service if the repairperson is free. If the repairperson is busy, it waits until the other machine is fixed; at that time, the newly repaired machine is put in service and repair begins on the other one. Starting with both machines in working condition, find

(a) the expected value and

(b) the variance of the time until both are in the repair facility.

(c) In the long run, what proportion of time is there a working machine?

21. Suppose that when both machines are down in Exercise 20 a second repairperson is called in to work on the newly failed one. Suppose all repair times remain exponential with rate μ. Now find the proportion of time at least one machine is working, and compare your answer with the one obtained in Exercise 20.

22. Customers arrive at a single-server queue in accordance with a Poisson process having rate λ. However, an arrival that finds n customers already in the system will only join the system with probability $1/(n+1)$. That is, with probability $n/(n+1)$ such an arrival will not join the system. Show that the limiting distribution of the number of customers in the system is Poisson with mean λ/μ. Assume that the service distribution is exponential with rate μ.

23. A job shop consists of three machines and two repairmen. The amount of time a machine works before breaking down is exponentially distributed with mean 10. If the amount of time it takes a single repairman to fix a machine is exponentially distributed with mean 8, then

(a) what is the average number of machines not in use?

(b) what proportion of time are both repairmen busy?

***24.** Consider a taxi station where taxis and customers arrive in accordance with Poisson processes with respective rates of one and two per minute. A taxi will wait no matter how many other taxis are present. However, an arriving customer that does not find a taxi waiting leaves. Find

(a) the average number of taxis waiting, and

(b) the proportion of arriving customers that get taxis.

25. Customers arrive at a service station, manned by a single server who serves at an exponential rate μ_1, at a Poisson rate λ. After completion of service the customer then joins a second system where the server serves at an exponential rate μ_2. Such a system is called a *tandem* or *sequential* queueing system. Assuming that $\lambda < \mu_i,\ i = 1, 2$, determine the limiting probabilities.

Hint: Try a solution of the form $P_{n,m} = C\alpha^n\beta^m$, and determine C, α, β.

26. Consider an ergodic $M/M/s$ queue in steady state (that is, after a long time) and argue that the number presently in the system is independent of the sequence of past departure times. That is, for instance, knowing that there have been departures 2, 3, 5, and 10 time units ago does not affect the distribution of the number presently in the system.

27. In the $M/M/s$ queue if you allow the service rate to depend on the number in the system (but in such a way so that it is ergodic), what can you say about the output process? What can you say when the service rate μ remains unchanged but $\lambda > s\mu$?

***28.** If $\{X(t)\}$ and $\{Y(t)\}$ are independent continuous-time Markov chains, both of which are time reversible, show that the process $\{X(t), Y(t)\}$ is also a time reversible Markov chain.

29. Consider a set of n machines and a single repair facility to service these machines. Suppose that when machine $i, i = 1, \ldots, n$, fails it requires an exponentially distributed amount of work with rate μ_i to repair it. The repair

facility divides its efforts equally among all failed machines in the sense that whenever there are k failed machines each one receives work at a rate of $1/k$ per unit time. If there are a total of r working machines, including machine i, then i fails at an instantaneous rate λ_i/r.

(a) Define an appropriate state space so as to be able to analyze the preceding system as a continuous-time Markov chain.

(b) Give the instantaneous transition rates (that is, give the q_{ij}).

(c) Write the time reversibility equations.

(d) Find the limiting probabilities and show that the process is time reversible.

30. Consider a graph with nodes $1, 2, \ldots, n$ and the $\binom{n}{2}$ arcs $(i, j), i \neq j, i, j, = 1, \ldots, n$. (See Section 3.6.2 for appropriate definitions.) Suppose that a particle moves along this graph as follows: Events occur along the arcs (i, j) according to independent Poisson processes with rates λ_{ij}. An event along arc (i, j) causes that arc to become excited. If the particle is at node i at the moment that (i, j) becomes excited, it instantaneously moves to node $j, i, j = 1, \ldots, n$. Let P_j denote the proportion of time that the particle is at node j. Show that

$$P_j = \frac{1}{n}$$

Hint: Use time reversibility.

31. A total of N customers move about among r servers in the following manner. When a customer is served by server i, he then goes over to server j, $j \neq i$, with probability $1/(r-1)$. If the server he goes to is free, then the customer enters service; otherwise he joins the queue. The service times are all independent, with the service times at server i being exponential with rate μ, $i = 1, \ldots, r$. Let the state at any time be the vector $(n_1, \ldots, n_r)$, where n_i is the number of customers presently at server i, $i = 1, \ldots, r, \sum_i n_i = N$.

(a) Argue that if $X(t)$ is the state at time t, then $\{X(t),\ t \geq 0\}$ is a continuous-time Markov chain.

(b) Give the infinitesimal rates of this chain.

(c) Show that this chain is time reversible, and find the limiting probabilities.

32. Customers arrive at a two-server station in accordance with a Poisson process having rate λ. Upon arriving, they join a single queue. Whenever a server completes a service, the person first in line enters service. The service times of server i are exponential with rate μ_i, $i = 1, 2$, where $\mu_1 + \mu_2 > \lambda$. An arrival finding both servers free is equally likely to go to either one. Define an appropriate continuous-time Markov chain for this model, show it is time reversible, and find the limiting probabilities.

***33.** Consider two $M/M/1$ queues with respective parameters $\lambda_i, \mu_i, i = 1, 2$. Suppose they share a common waiting room that can hold at most three customers. That is, whenever an arrival finds her server busy and three customers in the waiting room, she goes away. Find the limiting probability that there will be n queue 1 customers and m queue 2 customers in the system.

Hint: Use the results of Exercise 28 together with the concept of truncation.

34. Four workers share an office that contains four telephones. At any time, each worker is either "working" or "on the phone." Each "working" period of worker i lasts for an exponentially distributed time with rate λ_i, and each "on the phone" period lasts for an exponentially distributed time with rate μ_i, $i = 1, 2, 3, 4$.

(a) What proportion of time are all workers "working"?
Let $X_i(t)$ equal 1 if worker i is working at time t, and let it be 0 otherwise. Let $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t))$.

(b) Argue that $\{\mathbf{X}(t),\ t \geq 0\}$ is a continuous-time Markov chain and give its infinitesimal rates.

(c) Is $\{\mathbf{X}(t)\}$ time reversible? Why or why not?

Suppose now that one of the phones has broken down. Suppose that a worker who is about to use a phone but finds them all being used begins a new "working" period.

(d) What proportion of time are all workers "working"?

35. Consider a time reversible continuous-time Markov chain having infinitesimal transition rates q_{ij} and limiting probabilities $\{P_i\}$. Let A denote a set of states for this chain, and consider a new continuous-time Markov chain with transition rates q_{ij}^* given by

$$q_{ij}^* = \begin{cases} cq_{ij}, & \text{if } i \in A,\ j \notin A \\ q_{ij}, & \text{otherwise} \end{cases}$$

where c is an arbitrary positive number. Show that this chain remains time reversible, and find its limiting probabilities.

36. Consider a system of n components such that the working times of component i, $i = 1, \ldots, n$, are exponentially distributed with rate λ_i. When a component fails, however, the repair rate of component i depends on how many other components are down. Specifically, suppose that the instantaneous repair rate of component i, $i = 1, \ldots, n$, when there are a total of k failed components, is $\alpha^k \mu_i$.

(a) Explain how we can analyze the preceding as a continuous-time Markov chain. Define the states and give the parameters of the chain.

(b) Show that, in steady state, the chain is time reversible and compute the limiting probabilities.

37. A hospital accepts k different types of patients, where type i patients arrive according to a Poisson proccess with rate λ_i, with these k Poisson processes being independent. Type i patients spend an exponentially distributed length of time with rate μ_i in the hospital, $i = 1, \ldots, k$. Suppose that each type i patient in the hospital requires w_i units of resources, and that the hospital will not accept a new patient if it would result in the total of all patient's resource needs exceeding the amount C. Consequently, it is possible to have n_1 type 1 patients, n_2 type 2 patients, $\ldots$, and n_k type k patients in the hospital at the

same time if and only if

$$\sum_{i=1}^{k} n_i w_i \le C$$

(a) Define a continuous-time Markov chain to analyze the preceding.
For parts (b), (c), and (d) suppose that $C = \infty$.
(b) If $N_i(t)$ is the number of type i customers in the system at time t, what type of process is $\{N_i(t),\ t \ge 0\}$? Is it time reversible?
(c) What can be said about the vector process $\{(N_1(t), \ldots, N_k(t)),\ t \ge 0\}$?
(d) What are the limiting probabilities of the process of part (c).
For the remaining parts assume that $C < \infty$.
(e) Find the limiting probabilities for the Markov chain of part (a).
(f) At what rate are type i patients admitted?
(g) What fraction of patients are admitted?

38. Consider an n server system where the service times of server i are exponentially distributed with rate μ_i, $i = 1, \ldots, n$. Suppose customers arrive in accordance with a Poisson process with rate λ, and that an arrival who finds all servers busy does not enter but goes elsewhere. Suppose that an arriving customer who finds at least one idle server is served by a randomly chosen one of that group; that is, an arrival finding k idle servers is equally likely to be served by any of these k.
(a) Define states so as to analyze the preceding as a continuous-time Markov chain.
(b) Show that this chain is time reversible.
(c) Find the limiting probabilities.

39. Suppose in Exercise 38 that an entering customer is served by the server who has been idle the shortest amount of time.
(a) Define states so as to analyze this model as a continuous-time Markov chain.
(b) Show that this chain is time reversible.
(c) Find the limiting probabilities.

***40.** Consider a continuous-time Markov chain with states $1, \ldots, n$, which spends an exponential time with rate v_i in state i during each visit to that state and is then equally likely to go to any of the other $n - 1$ states.
(a) Is this chain time reversible?
(b) Find the long-run proportions of time it spends in each state.

41. Show in Example 6.22 that the limiting probabilities satisfy Eqs. (6.33), (6.34), and (6.35).

42. In Example 6.22 explain why we would have known before analyzing Example 6.22 that the limiting probability there are j customers with server i is $(\lambda/\mu_i)^j(1 - \lambda/\mu_i)$, $i = 1, 2,\ j \ge 0$. (What we would not have known was that the number of customers at the two servers would, in steady state, be independent.)

43. Consider a sequential queueing model with three servers, where customers arrive at server 1 in accordance with a Poisson process with rate λ. After completion at server 1 the customer then moves to server 2; after a service completion at server 2 the customer moves to server 3; after a service completion at server 3 the customer departs the system. Assuming that the service times at server i are exponential with rate μ_i, $i = 1, 2, 3$, find the limiting probabilities of this system by guessing at the reverse chain and then verifying that your guess is correct.

44. A system of N machines operates as follows. Each machine works for an exponentially distributed time with rate λ before failing. Upon failure, a machine must go through two phases of service. Phase 1 service lasts for an exponential time with rate μ, and there are always servers available for phase 1 service. After competing phase 1 service the machine goes to a server that performs phase 2 service. If that server is busy then the machine joins the waiting queue. The time it takes to complete a phase 2 service is exponential with rate v. After completing a phase 2 service the machine goes back to work. Consider the continuous time Markov chain whose state at any time is the triplet of nonnegative numbers $\mathbf{n} = (n_0, n_1, n_2)$ where $n_0 + n_1 + n_2 = N$, with the interpretation that of the N machines, n_0 are working, n_1 are in phase 1 service, and n_2 are in phase 2 service.

(a) Give the instantaneous transition rates of this continuous time Markov chain.

(b) Interpreting the reverse chain as a model of similar type, except that machines go from working to phase 2 and then to phase 1 service, conjecture the transition rates of the reverse chain. In doing so, make sure that your conjecture would result in the rate at which the reverse chain departs state (n, k, j) upon a visit being equal to the rate at which the forward chain departs that state upon a visit.

(c) Prove that your conjecture is correct and find the limiting probabilities.

45. For the continuous-time Markov chain of Exercise 3 present a uniformized version.

46. In Example 6.24, we computed $m(t) = E[O(t)]$, the expected occupation time in state 0 by time t for the two-state continuous-time Markov chain starting in state 0. Another way of obtaining this quantity is by deriving a differential equation for it.

(a) Show that

$$m(t + h) = m(t) + P_{00}(t)h + o(h)$$

(b) Show that

$$m'(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

(c) Solve for $m(t)$.

47. Let $O(t)$ be the occupation time for state 0 in the two-state continuous-time Markov chain. Find $E[O(t)|X(0) = 1]$.

48. Consider the two-state continuous-time Markov chain. Starting in state 0, find $\text{Cov}[X(s), X(t)]$.

49. Let Y denote an exponential random variable with rate λ that is independent of the continuous-time Markov chain $\{X(t)\}$ and let

$$\bar{P}_{ij} = P\{X(Y) = j|X(0) = i\}$$

(a) Show that

$$\bar{P}_{ij} = \frac{1}{v_i + \lambda} \sum_k q_{ik} \bar{P}_{kj} + \frac{\lambda}{v_i + \lambda} \delta_{ij}$$

where δ_{ij} is 1 when $i = j$ and 0 when $i \neq j$.

(b) Show that the solution of the preceding set of equations is given by

$$\bar{\mathbf{P}} = (\mathbf{I} - \mathbf{R}/\lambda)^{-1}$$

where $\bar{\mathbf{P}}$ is the matrix of elements $\bar{P}_{ij}$, $\mathbf{I}$ is the identity matrix, and $\mathbf{R}$ the matrix specified in Section 6.9.

(c) Suppose now that $Y_1, \ldots, Y_n$ are independent exponentials with rate λ that are independent of $\{X(t)\}$. Show that

$$P\{X(Y_1 + \cdots + Y_n) = j|X(0) = i\}$$

is equal to the element in row i, column j of the matrix $\bar{\mathbf{P}}^n$.

(d) Explain the relationship of the preceding to Approximation 2 of Section 6.9.

***50.** **(a)** Show that Approximation 1 of Section 6.9 is equivalent to uniformizing the continuous-time Markov chain with a value v such that $vt = n$ and then approximating $P_{ij}(t)$ by P_{ij}^{*n}.

(b) Explain why the preceding should make a good approximation.

Hint: What is the standard deviation of a Poisson random variable with mean n?

References

[1] D.R. Cox, H.D. Miller, The Theory of Stochastic Processes, Methuen, London, 1965.
[2] A.W. Drake, Fundamentals of Applied Probability Theory, McGraw-Hill, New York, 1967.
[3] S. Karlin, H. Taylor, A First Course in Stochastic Processes, Second Edition, Academic Press, New York, 1975.
[4] E. Parzen, Stochastic Processes, Holden-Day, San Francisco, California, 1962.
[5] S. Ross, Stochastic Processes, Second Edition, John Wiley, New York, 1996.

Renewal Theory and Its Applications

7

7.1 Introduction

We have seen that a Poisson process is a counting process for which the times between successive events are independent and identically distributed exponential random variables. One possible generalization is to consider a counting process for which the times between successive events are independent and identically distributed with an arbitrary distribution. Such a counting process is called a *renewal process*.

Let $\{N(t),\ t \geq 0\}$ be a counting process and let X_n denote the time between the $(n-1)$st and the nth event of this process, $n \geq 1$.

Definition 7.1. If the sequence of nonnegative random variables $\{X_1, X_2, \ldots\}$ is independent and identically distributed, then the counting process $\{N(t),\ t \geq 0\}$ is said to be a *renewal process*.

Thus, a renewal process is a counting process such that the time until the first event occurs has some distribution F, the time between the first and second event has, independently of the time of the first event, the same distribution F, and so on. When an event occurs, we say that a renewal has taken place.

For an example of a renewal process, suppose that we have an infinite supply of lightbulbs whose lifetimes are independent and identically distributed. Suppose also that we use a single lightbulb at a time, and when it fails we immediately replace it with a new one. Under these conditions, $\{N(t),\ t \geq 0\}$ is a renewal process when $N(t)$ represents the number of lightbulbs that have failed by time t.

For a renewal process having interarrival times $X_1, X_2, \ldots$, let

$$S_0 = 0, \quad S_n = \sum_{i=1}^{n} X_i, \quad n \geq 1$$

That is, $S_1 = X_1$ is the time of the first renewal; $S_2 = X_1 + X_2$ is the time until the first renewal plus the time between the first and second renewal, that is, S_2 is the time of the second renewal. In general, S_n denotes the time of the nth renewal (see Fig. 7.1).

We shall let F denote the interarrival distribution and to avoid trivialities, we assume that $F(0) = P\{X_n = 0\} < 1$. Furthermore, we let

$$\mu = E[X_n], \quad n \geq 1$$

be the mean time between successive renewals. It follows from the nonnegativity of X_n and the fact that X_n is not identically 0 that $\mu > 0$.

The first question we shall attempt to answer is whether an infinite number of renewals can occur in a finite amount of time. That is, can $N(t)$ be infinite for some

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00012-3

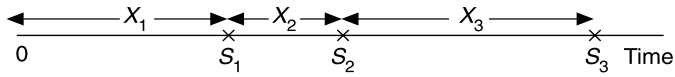

Figure 7.1 Renewal and interarrival times.

(finite) value of t? To show that this cannot occur, we first note that, as S_n is the time of the nth renewal, $N(t)$ may be written as

$$N(t) = \max\{n: \ S_n \le t\} \tag{7.1}$$

To understand why Eq. (7.1) is valid, suppose, for instance, that $S_4 \le t$ but $S_5 > t$. Hence, the fourth renewal had occurred by time t but the fifth renewal occurred after time t; or in other words, $N(t)$, the number of renewals that occurred by time t, must equal 4. Now by the strong law of large numbers it follows that, with probability 1,

$$\frac{S_n}{n} \to \mu \quad \text{as } n \to \infty$$

But since $\mu > 0$ this means that S_n must be going to infinity as n goes to infinity. Thus, S_n can be less than or equal to t for at most a finite number of values of n, and hence by Eq. (7.1), $N(t)$ must be finite.

However, though $N(t) < \infty$ for each t, it is true that, with probability 1,

$$N(\infty) \equiv \lim_{t\to\infty} N(t) = \infty$$

This follows since the only way in which $N(\infty)$, the total number of renewals that occur, can be finite is for one of the interarrival times to be infinite.

Therefore,

$$\begin{aligned} P\{N(\infty) < \infty\} &= P\{X_n = \infty \text{ for some } n\} \\ &= P\left\{\bigcup_{n=1}^{\infty} \{X_n = \infty\}\right\} \\ &\le \sum_{n=1}^{\infty} P\{X_n = \infty\} \\ &= 0 \end{aligned}$$

7.2 Distribution of $N(t)$

The distribution of $N(t)$ can be obtained, at least in theory, by first noting the important relationship that *the number of renewals by time t is greater than or equal to n if and only if the nth renewal occurs before or at time t*. That is,

$$N(t) \ge n \Leftrightarrow S_n \le t \tag{7.2}$$

From Eq. (7.2) we obtain

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geq n\} - P\{N(t) \geq n+1\} \\ &= P\{S_n \leq t\} - P\{S_{n+1} \leq t\} \end{aligned} \tag{7.3}$$

Now, since the random variables $X_i, i \geq 1$, are independent and have a common distribution F, it follows that $S_n = \sum_{i=1}^{n} X_i$ is distributed as F_n, the n-fold convolution of F with itself (Section 2.5). Therefore, from Eq. (7.3) we obtain

$$P\{N(t) = n\} = F_n(t) - F_{n+1}(t)$$

Example 7.1. Suppose that $P\{X_n = i\} = p(1-p)^{i-1}, i \geq 1$. That is, suppose that the interarrival distribution is geometric. Now $S_1 = X_1$ may be interpreted as the number of trials necessary to get a single success when each trial is independent and has a probability p of being a success. Similarly, S_n may be interpreted as the number of trials necessary to attain n successes, and hence has the negative binomial distribution

$$P\{S_n = k\} = \begin{cases} \binom{k-1}{n-1} p^n (1-p)^{k-n}, & k \geq n \\ 0, & k < n \end{cases}$$

Thus, from Eq. (7.3) we have that

$$\begin{aligned} P\{N(t) = n\} = &\sum_{k=n}^{[t]} \binom{k-1}{n-1} p^n (1-p)^{k-n} \\ &- \sum_{k=n+1}^{[t]} \binom{k-1}{n} p^{n+1} (1-p)^{k-n-1} \end{aligned}$$

Equivalently, since an event independently occurs with probability p at each of the times $1, 2, \ldots$

$$P\{N(t) = n\} = \binom{[t]}{n} p^n (1-p)^{[t]-n}$$ ■

Another expression for $P(N(t) = n)$ can be obtained by conditioning on S_n. This yields

$$P\,(N(t) = n) = \int_0^\infty P\,(N(t) = n | S_n = y)\, f_{S_n}(y) dy$$

Now, if the nth event occurred at time $y > t$, then there would have been less than n events by time t. On the other hand, if it occurred at a time $y \leq t$, then there would be exactly n events by time t if the next interarrival exceeds $t - y$. Consequently,

$$P\,(N(t) = n) = \int_0^t P\,(X_{n+1} > t - y | S_n = y)\, f_{S_n}(y) dy$$

$$= \int_0^t \bar{F}(t-y) f_{S_n}(y)\,dy$$

where $\bar{F} = 1 - F$.

Example 7.2. If $F(x) = 1 - e^{\lambda x}$ then S_n, being the sum of n independent exponentials with rate λ, will have a gamma (n, λ) distribution. Consequently, the preceding identity gives

$$\begin{aligned} P\{N(t) = n\} &= \int_0^t e^{-\lambda(t-y)} \frac{\lambda e^{-\lambda y} (\lambda y)^{n-1}}{(n-1)!}\,dy \\ &= \frac{\lambda^n e^{-\lambda t}}{(n-1)!} \int_0^t y^{n-1}\,dy \\ &= e^{-\lambda t} \frac{(\lambda t)^n}{n!} \end{aligned}$$

■

By using Eq. (7.2) we can calculate $m(t)$, the mean value of $N(t)$, as

$$\begin{aligned} m(t) &= E[N(t)] \\ &= \sum_{n=1}^{\infty} P\{N(t) \geq n\} \\ &= \sum_{n=1}^{\infty} P\{S_n \leq t\} \\ &= \sum_{n=1}^{\infty} F_n(t) \end{aligned}$$

where we have used the fact that if X is nonnegative and integer valued, then

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k P\{X = k\} = \sum_{k=1}^{\infty} \sum_{n=1}^{k} P\{X = k\} \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\{X = k\} = \sum_{n=1}^{\infty} P\{X \geq n\} \end{aligned}$$

The function $m(t)$ is known as the *mean-value* or the *renewal function*.

It can be shown that the mean-value function $m(t)$ uniquely determines the renewal process. Specifically, there is a one-to-one correspondence between the interarrival distributions F and the mean-value functions $m(t)$.

Another interesting result that we state without proof is that

$$m(t) < \infty \quad \text{for all } t < \infty$$

Remarks. **(i)** Since $m(t)$ uniquely determines the interarrival distribution, it follows that the Poisson process is the only renewal process having a linear mean-value function.

(ii) Some readers might think that the finiteness of $m(t)$ should follow directly from the fact that, with probability 1, $N(t)$ is finite. However, such reasoning is not valid; consider the following: Let Y be a random variable having the following probability distribution:

$$Y = 2^n \text{ with probability } \left(\tfrac{1}{2}\right)^n, \quad n \geq 1$$

Now,

$$P\{Y < \infty\} = \sum_{n=1}^{\infty} P\{Y = 2^n\} = \sum_{n=1}^{\infty} \left(\tfrac{1}{2}\right)^n = 1$$

But

$$E[Y] = \sum_{n=1}^{\infty} 2^n P\{Y = 2^n\} = \sum_{n=1}^{\infty} 2^n \left(\tfrac{1}{2}\right)^n = \infty$$

Hence, even when Y is finite, it can still be true that $E[Y] = \infty$.

An integral equation satisfied by the renewal function can be obtained by conditioning on the time of the first renewal. Assuming that the interarrival distribution F is continuous with density function f this yields

$$m(t) = E[N(t)] = \int_0^{\infty} E[N(t)|X_1 = x] f(x) dx \tag{7.4}$$

Now suppose that the first renewal occurs at a time x that is less than t. Then, using the fact that a renewal process probabilistically starts over when a renewal occurs, it follows that the number of renewals by time t would have the same distribution as 1 plus the number of renewals in the first $t - x$ time units. Therefore,

$$E[N(t)|X_1 = x] = 1 + E[N(t - x)] \quad \text{if } x < t$$

Since, clearly

$$E[N(t)|X_1 = x] = 0 \quad \text{when } x > t$$

we obtain from Eq. (7.4) that

$$\begin{aligned} m(t) &= \int_0^t [1 + m(t - x)] f(x) dx \\ &= F(t) + \int_0^t m(t - x) f(x) dx \end{aligned} \tag{7.5}$$

Eq. (7.5) is called the *renewal equation* and can sometimes be solved to obtain the renewal function.

Example 7.3. One instance in which the renewal equation can be solved is when the interarrival distribution is uniform—say, uniform on (0, 1). We will now present a solution in this case when $t \leq 1$. For such values of t, the renewal function becomes

$$\begin{aligned} m(t) &= t + \int_0^t m(t-x)dx \\ &= t + \int_0^t m(y)dy \quad \text{by the substitution } y = t - x \end{aligned}$$

Differentiating the preceding equation yields

$$m'(t) = 1 + m(t)$$

Letting $h(t) = 1 + m(t)$, we obtain

$$h'(t) = h(t)$$

or

$$\log h(t) = t + C$$

or

$$h(t) = Ke^t$$

or

$$m(t) = Ke^t - 1$$

Since $m(0) = 0$, we see that $K = 1$, and so we obtain

$$m(t) = e^t - 1, \quad 0 \leq t \leq 1$$ ■

7.3 Limit Theorems and Their Applications

We have shown previously that, with probability 1, $N(t)$ goes to infinity as t goes to infinity. However, it would be nice to know the rate at which $N(t)$ goes to infinity. That is, we would like to be able to say something about $\lim_{t\to\infty} N(t)/t$.

As a prelude to determining the rate at which $N(t)$ grows, let us first consider the random variable $S_{N(t)}$. In words, just what does this random variable represent? Proceeding inductively suppose, for instance, that $N(t) = 3$. Then $S_{N(t)} = S_3$ represents the time of the third event. Since there are only three events that have occurred by time t, S_3 also represents the time of the last event prior to (or at) time t. This is, in fact, what $S_{N(t)}$ represents—namely, the time of the last renewal *prior to or at* time t. Similar reasoning leads to the conclusion that $S_{N(t)+1}$ represents the time of the first renewal *after* time t (see Fig. 7.2). We now are ready to prove the following.

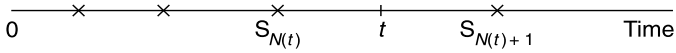

Figure 7.2

Proposition 7.1. *With probability 1,*

$$\frac{N(t)}{t} \to \frac{1}{\mu} \quad \text{as } t \to \infty$$

Proof. Since $S_{N(t)}$ is the time of the last renewal prior to or at time t, and $S_{N(t)+1}$ is the time of the first renewal after time t, we have

$$S_{N(t)} \leq t < S_{N(t)+1}$$

or

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)} \tag{7.6}$$

However, since $S_{N(t)}/N(t) = \sum_{i=1}^{N(t)} X_i/N(t)$ is the average of $N(t)$ independent and identically distributed random variables, it follows by the strong law of large numbers that $S_{N(t)}/N(t) \to \mu$ as $N(t) \to \infty$. But since $N(t) \to \infty$ when $t \to \infty$, we obtain

$$\frac{S_{N(t)}}{N(t)} \to \mu \quad \text{as } t \to \infty$$

Furthermore, writing

$$\frac{S_{N(t)+1}}{N(t)} = \left(\frac{S_{N(t)+1}}{N(t)+1}\right)\left(\frac{N(t)+1}{N(t)}\right)$$

we have that $S_{N(t)+1}/(N(t)+1) \to \mu$ by the same reasoning as before and

$$\frac{N(t)+1}{N(t)} \to 1 \quad \text{as } t \to \infty$$

Hence,

$$\frac{S_{N(t)+1}}{N(t)} \to \mu \quad \text{as } t \to \infty$$

The result now follows by Eq. (7.6) since $t/N(t)$ is between two random variables, each of which converges to μ as $t \to \infty$. ■

Remarks. **(i)** The preceding propositions are true even when μ, the mean time between renewals, is infinite. In this case, we interpret $1/\mu$ to be 0.

(ii) The number $1/\mu$ is called the *rate* of the renewal process.

(iii) Because the average time between renewals is μ, it is quite intuitive that the average rate at which renewals occur is 1 per every μ time units. ■

Example 7.4. Beverly has a radio that works on a single battery. As soon as the battery in use fails, Beverly immediately replaces it with a new battery. If the lifetime of a battery (in hours) is distributed uniformly over the interval $(30, 60)$, then at what rate does Beverly have to change batteries?

Solution: If we let $N(t)$ denote the number of batteries that have failed by time t, we have by Proposition 7.1 that the rate at which Beverly replaces batteries is given by

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu} = \frac{1}{45}$$

That is, in the long run, Beverly will have to replace one battery every 45 hours. ■

Example 7.5. Suppose in Example 7.4 that Beverly does not keep any surplus batteries on hand, and so each time a failure occurs she must go and buy a new battery. If the amount of time it takes for her to get a new battery is uniformly distributed over $(0, 1)$, then what is the average rate that Beverly changes batteries?

Solution: In this case the mean time between renewals is given by

$$\mu = E[U_1] + E[U_2]$$

where U_1 is uniform over $(30, 60)$ and U_2 is uniform over $(0, 1)$. Hence,

$$\mu = 45 + \tfrac{1}{2} = 45\tfrac{1}{2}$$

and so in the long run, Beverly will be putting in a new battery at the rate of $\frac{2}{91}$. That is, she will put in two new batteries every 91 hours. ■

Example 7.6. Suppose that potential customers arrive at a single-server bank in accordance with a Poisson process having rate λ. However, suppose that the potential customer will enter the bank only if the server is free when he arrives. That is, if there is already a customer in the bank, then our arriver, rather than entering the bank, will go home. If we assume that the amount of time spent in the bank by an entering customer is a random variable having distribution G, then

(a) what is the rate at which customers enter the bank?
(b) what proportion of potential customers actually enter the bank?

Solution: In answering these questions, let us suppose that at time 0 a customer has just entered the bank. (That is, we define the process to start when the first customer enters the bank.) If we let μ_G denote the mean service time, then, by the

memoryless property of the Poisson process, it follows that the mean time between entering customers is

$$\mu = \mu_G + \frac{1}{\lambda}$$

Hence, the rate at which customers enter the bank will be given by

$$\frac{1}{\mu} = \frac{\lambda}{1 + \lambda\mu_G}$$

On the other hand, since potential customers will be arriving at a rate λ, it follows that the proportion of them entering the bank will be given by

$$\frac{\lambda/(1 + \lambda\mu_G)}{\lambda} = \frac{1}{1 + \lambda\mu_G}$$

In particular if $\lambda = 2$ and $\mu_G = 2$, then only one customer out of five will actually enter the system. ■

A somewhat unusual application of Proposition 7.1 is provided by our next example.

Example 7.7. A sequence of independent trials, each of which results in outcome number i with probability $P_i, i = 1, \ldots, n$, $\sum_{i=1}^{n} P_i = 1$, is observed until the same outcome occurs k times in a row; this outcome then is declared to be the winner of the game. For instance, if $k = 2$ and the sequence of outcomes is 1, 2, 4, 3, 5, 2, 1, 3, 3, then we stop after nine trials and declare outcome number 3 the winner. What is the probability that i wins, $i = 1, \ldots, n$, and what is the expected number of trials?

Solution: We begin by computing the expected number of coin tosses, call it $E[T]$, until a run of k successive heads occurs when the tosses are independent and each lands on heads with probability p. By conditioning on the time of the first nonhead, we obtain

$$E[T] = \sum_{j=1}^{k} (1 - p)p^{j-1}(j + E[T]) + kp^k$$

Solving this for $E[T]$ yields

$$E[T] = k + \frac{(1 - p)}{p^k} \sum_{j=1}^{k} jp^{j-1}$$

Upon simplifying, we obtain

$$\begin{aligned} E[T] &= \frac{1 + p + \cdots + p^{k-1}}{p^k} \\ &= \frac{1 - p^k}{p^k(1 - p)} \end{aligned} \tag{7.7}$$

Now, let us return to our example, and let us suppose that as soon as the winner of a game has been determined we immediately begin playing another game. For each i let us determine the rate at which outcome i wins. Now, every time i wins, everything starts over again and thus wins by i constitute renewals. Hence, from Proposition 7.1, the

$$\text{rate at which } i \text{ wins} = \frac{1}{E[N_i]}$$

where N_i denotes the number of trials played between successive wins of outcome i. Hence, from Eq. (7.7) we see that

$$\text{rate at which } i \text{ wins} = \frac{P_i^k(1-P_i)}{1-P_i^k} \tag{7.8}$$

Hence, the long-run proportion of games that are won by number i is given by

$$\begin{aligned}\text{proportion of games } i \text{ wins} &= \frac{\text{rate at which } i \text{ wins}}{\sum_{j=1}^n \text{rate at which } j \text{ wins}} \\ &= \frac{P_i^k(1-P_i)/(1-P_i^k)}{\sum_{j=1}^n (P_j^k(1-P_j)/(1-P_j^k))}\end{aligned}$$

However, it follows from the strong law of large numbers that the long-run proportion of games that i wins will, with probability 1, be equal to the probability that i wins any given game. Hence,

$$P\{i \text{ wins}\} = \frac{P_i^k(1-P_i)/(1-P_i^k)}{\sum_{j=1}^n (P_j^k(1-P_j)/(1-P_j^k))}$$

To compute the expected time of a game, we first note that the

$$\begin{aligned}\text{rate at which games end} &= \sum_{i=1}^n \text{rate at which } i \text{ wins} \\ &= \sum_{i=1}^n \frac{P_i^k(1-P_i)}{1-P_i^k} \quad \text{(from Eq. (7.8))}\end{aligned}$$

Now, as everything starts over when a game ends, it follows by Proposition 7.1 that the rate at which games end is equal to the reciprocal of the mean time of a game. Hence,

$$\begin{aligned}E[\text{time of a game}\} &= \frac{1}{\text{rate at which games end}} \\ &= \frac{1}{\sum_{i=1}^n (P_i^k(1-P_i)/(1-P_i^k))}\end{aligned}$$

■

Proposition 7.1 says that the average renewal rate up to time t will, with probability 1, converge to $1/\mu$ as $t \to \infty$. What about the expected average renewal rate? Is it true that $m(t)/t$ also converges to $1/\mu$? This result is known as the *elementary renewal theorem*.

Theorem 7.1 (Elementary Renewal Theorem)**.**

$$\frac{m(t)}{t} \to \frac{1}{\mu} \quad \text{as } t \to \infty$$

As before, $1/\mu$ *is interpreted as 0 when* $\mu = \infty$.

Remark. At first glance it might seem that the elementary renewal theorem should be a simple consequence of Proposition 7.1. That is, since the average renewal rate will, with probability 1, converge to $1/\mu$, should this not imply that the expected average renewal rate also converges to $1/\mu$? We must, however, be careful; consider the next example.

Example 7.8. Let U be a random variable which is uniformly distributed on $(0, 1)$; and define the random variables $Y_n, n \geq 1$, by

$$Y_n = \begin{cases} 0, & \text{if } U > 1/n \\ n, & \text{if } U \leq 1/n \end{cases}$$

Now, since, with probability 1, U will be greater than 0, it follows that Y_n will equal 0 for all sufficiently large n. That is, Y_n will equal 0 for all n large enough so that $1/n < U$. Hence, with probability 1,

$$Y_n \to 0 \quad \text{as } n \to \infty$$

However,

$$E[Y_n] = nP\left\{U \leq \frac{1}{n}\right\} = n\frac{1}{n} = 1$$

Therefore, even though the sequence of random variables Y_n converges to 0, the expected values of the Y_n are all identically 1. ■

To prove the elementary renewal theorem we will make use of an identity known as Wald's equation. Before stating Wald's equation we need to introduce the concept of a stopping time for a sequence of independent random variables.

Definition. The nonnegative integer valued random variable N is said to be a *stopping time* for a sequence of independent random variables $X_1, X_2, \ldots$ if the event that $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$, for all $n = 1, 2, \ldots$.

The idea behind a stopping time is that we imagine that the X_i are observed in sequence, first X_1, then X_2, and so on, and that N denotes the number of them observed before stopping. Because the event that we stop after having observed $X_1, \ldots, X_n$ can only depend on these n values, and not on future unobserved values, it must be independent of these future values.

Example 7.9. Suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with

$$P(X_i = 1) = p = 1 - P(X_i = 0)$$

where $p > 0$. If we define

$$N = \min(n : X_1 + \cdots + X_n = r)$$

then N is a stopping time for the sequence. If we imagine that trials are being performed in sequence and that $X_i = 1$ corresponds to a success on trial i, then N is the number of trials needed until there have been a total of r successes when each trial is independently a success with probability p. ■

Example 7.10. Suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with

$$P(X_i = 1) = 1/2 = 1 - P(X_i = -1)$$

If

$$N = \min(n : X_1 + \cdots + X_n = 1)$$

then N is a stopping time for the sequence. N can be regarded as the stopping time for a gambler who on each play is equally likely to win or lose 1, and who is going to stop the first time he is winning money. (Because the successive winnings of the gambler are a symmetric random walk, which we showed in Chapter 4 to be a recurrent Markov chain, it follows that $P(N < \infty) = 1$.) ■

We are now ready for Wald's equation.

Theorem 7.2 (Wald's Equation)**.** *If $X_1, X_2, \ldots,$ is a sequence of independent and identically distributed random variables with finite expectation $E[X]$, and if N is a stopping time for this sequence such that $E[N] < \infty$, then*

$$E\left[\sum_{n=1}^{N} X_n\right] = E[N]E[X]$$

Proof. For $n = 1, 2, \ldots,$ let

$$I_n = \begin{cases} 1, & \text{if } n \le N \\ 0, & \text{if } n > N \end{cases}$$

and note that

$$\sum_{n=1}^{N} X_n = \sum_{n=1}^{\infty} X_n I_n$$

Taking expectations gives

$$E\left[\sum_{n=1}^{N} X_n\right] = E\left[\sum_{n=1}^{\infty} X_n I_n\right] = \sum_{n=1}^{\infty} E[X_n I_n]$$

Now $I_n = 1$ if $N \geq n$, which means that $I_n = 1$ if we have not yet stopped after having observed $X_1, \ldots, X_{n-1}$. But this implies that the value of I_n is determined before X_n has been observed, and thus X_n is independent of I_n. Consequently,

$$E[X_n I_n] = E[X_n]E[I_n] = E[X]E[I_n]$$

showing that

$$\begin{aligned} E\left[\sum_{n=1}^{N} X_n\right] &= E[X]\sum_{n=1}^{\infty} E[I_n] \\ &= E[X]E\left[\sum_{n=1}^{\infty} I_n\right] \\ &= E[X]E[N] \end{aligned}$$

■

To apply Wald's equation to renewal theory, let $X_1, X_2, \ldots$ be the sequence of interarrival times of a renewal process. If we observe these one at a time and then stop at the first renewal after time t, then we would stop after having observed $X_1, \ldots, X_{N(t)+1}$, showing that $N(t) + 1$ is a stopping time for the sequence of interarrival times. For a more formal argument that $N(t) + 1$ is a stopping time for the sequence of interarrival times, note that $N(t) = n - 1$ if and only if the $(n-1)$st renewal occurs by time t and the nth renewal occurs after time t. That is,

$$N(t) + 1 = n \Leftrightarrow N(t) = n - 1 \Leftrightarrow X_1 + \cdots + X_{n-1} \leq t, X_1 + \cdots + X_n > t$$

showing that the event that $N(t) + 1 = n$ depends only on the values of $X_1, \ldots, X_n$.

We thus have the following corollary of Wald's equation. ■

Proposition 7.2. *If $X_1, X_2, \ldots,$ are the interarrival times of a renewal process then*

$$E[X_1 + \cdots + X_{N(t)+1}] = E[X]E[N(t) + 1]$$

That is,

$$E[S_{N(t)+1}] = \mu[m(t) + 1]$$

We are now ready to prove the elementary renewal theorem.

Proof of Elementary Renewal Theorem. Because $S_{N(t)+1}$ is the time of the first renewal after t, it follows that

$$S_{N(t)+1} = t + Y(t)$$

where $Y(t)$, called the *excess* at time t, is defined as the time from t until the next renewal. Taking expectations of the preceding yields, upon applying Proposition 7.2, that

$$\mu(m(t)+1) = t + E[Y(t)] \tag{7.9}$$

which can be written as

$$\frac{m(t)}{t} = \frac{1}{\mu} + \frac{E[Y(t)]}{t\mu} - \frac{1}{t}$$

Because $Y(t) \geq 0$, the preceding yields that $\frac{m(t)}{t} \geq \frac{1}{\mu} - \frac{1}{t}$, showing that

$$\lim_{t\to\infty} \frac{m(t)}{t} \geq \frac{1}{\mu}$$

To show that $\lim_{t\to\infty} \frac{m(t)}{t} \leq \frac{1}{\mu}$, let us suppose that there is a value $M < \infty$ such that $P(X_i < M) = 1$. Because this implies that $Y(t)$ must also be less than M, we have that $E[Y(t)] < M$, and so

$$\frac{m(t)}{t} \leq \frac{1}{\mu} + \frac{M}{t\mu} - \frac{1}{t}$$

which gives that

$$\lim_{t\to\infty} \frac{m(t)}{t} \leq \frac{1}{\mu}$$

and thus completes the proof of the elementary renewal theorem when the interarrival times are bounded. When the interarrival times $X_1, X_2, \ldots$ are unbounded, fix $M > 0$, and let $N_M(t), t \geq 0$ be the renewal process with interarrival times $\min(X_i, M), i \geq 1$. Because $\min(X_i, M) \leq X_i$ for all i, it follows that $N_M(t) \geq N(t)$ for all t. (That is, because each interarrival time of $N_M(t)$ is smaller than its corresponding interarrival time of $N(t)$, it must have at least as many renewals by time t.) Consequently, $E[N(t)] \leq E[N_M(t)]$, showing that

$$\lim_{t\to\infty} \frac{E[N(t)]}{t} \leq \lim_{t\to\infty} \frac{E[N_M(t)]}{t} = \frac{1}{E[\min(X_i, M)]}$$

where the equality follows because the interarrival times of $N_M(t)$ are bounded. Using that $\lim_{M\to\infty} E[\min(X_i, M)] = E[X_i] = \mu$, we obtain from the preceding upon letting $M \to \infty$ that

$$\lim_{t\to\infty} \frac{m(t)}{t} \leq \frac{1}{\mu}$$

and the proof is completed. ■

Eq. (7.9) shows that if we can determine $E[Y(t)]$, the mean excess at time t, then we can compute $m(t)$ and vice versa.

Example 7.11. Consider the renewal process whose interarrival distribution is the convolution of two exponentials; that is,

$$F = F_1 * F_2, \quad \text{where } F_i(t) = 1 - e^{-\mu_i t}, \quad i = 1, 2$$

We will determine the renewal function by first determining $E[Y(t)]$. To obtain the mean excess at t, imagine that each renewal corresponds to a new machine being put in use, and suppose that each machine has two components—initially component 1 is employed and this lasts an exponential time with rate μ_1, and then component 2, which functions for an exponential time with rate μ_2, is employed. When component 2 fails, a new machine is put in use (that is, a renewal occurs). Now consider the process $\{X(t),\ t \geq 0\}$ where $X(t)$ is i if a type i component is in use at time t. It is easy to see that $\{X(t),\ t \geq 0\}$ is a two-state continuous-time Markov chain, and so, using the results of Example 6.11, its transition probabilities are

$$P_{11}(t) = \frac{\mu_1}{\mu_1 + \mu_2} e^{-(\mu_1+\mu_2)t} + \frac{\mu_2}{\mu_1 + \mu_2}$$

To compute the expected remaining life of the machine in use at time t, we condition on whether it is using its first or second component: for if it is still using its first component, then its remaining life is $1/\mu_1 + 1/\mu_2$, whereas if it is already using its second component, then its remaining life is $1/\mu_2$. Hence, letting $p(t)$ denote the probability that the machine in use at time t is using its first component, we have

$$\begin{aligned} E[Y(t)] &= \left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) p(t) + \frac{1 - p(t)}{\mu_2} \\ &= \frac{1}{\mu_2} + \frac{p(t)}{\mu_1} \end{aligned}$$

But, since at time 0 the first machine is utilizing its first component, it follows that $p(t) = P_{11}(t)$, and so, upon using the preceding expression of $P_{11}(t)$, we obtain

$$E[Y(t)] = \frac{1}{\mu_2} + \frac{1}{\mu_1 + \mu_2} e^{-(\mu_1+\mu_2)t} + \frac{\mu_2}{\mu_1(\mu_1 + \mu_2)} \tag{7.10}$$

Now it follows from Eq. (7.9) that

$$m(t) + 1 = \frac{t}{\mu} + \frac{E[Y(t)]}{\mu} \tag{7.11}$$

where μ, the mean interarrival time, is given in this case by

$$\mu = \frac{1}{\mu_1} + \frac{1}{\mu_2} = \frac{\mu_1 + \mu_2}{\mu_1 \mu_2}$$

Substituting Eq. (7.10) and the preceding equation into (7.11) yields, after simplifying,

$$m(t) = \frac{\mu_1\mu_2}{\mu_1+\mu_2}t - \frac{\mu_1\mu_2}{(\mu_1+\mu_2)^2}[1 - e^{-(\mu_1+\mu_2)t}]$$ ■

Remark. Using the relationship of Eq. (7.11) and results from the two-state continuous-time Markov chain, the renewal function can also be obtained in the same manner as in Example 7.11 for the interarrival distributions

$$F(t) = pF_1(t) + (1-p)F_2(t)$$

and

$$F(t) = pF_1(t) + (1-p)(F_1 * F_2)(t)$$

when $F_i(t) = 1 - e^{-\mu_i t}, t > 0, i = 1, 2.$ ■

Suppose the interarrival times of a renewal process are all positive integer valued. Let

$$I_i = \begin{cases} 1, & \text{if there is a renewal at time } i \\ 0, & \text{otherwise} \end{cases}$$

and note that $N(n)$, the number of renewals by time n, can be expressed as

$$N(n) = \sum_{i=1}^{n} I_i$$

Taking expectations of both sides of the preceding shows that

$$m(n) = E[N(n)] = \sum_{i=1}^{n} P(\text{renewal at time } i)$$

Hence, the elementary renewal theorem yields

$$\frac{\sum_{i=1}^{n} P(\text{renewal at time } i)}{n} \to \frac{1}{E[\text{time between renewals}]}$$

Now, for a sequence of numbers $a_1, a_2, \ldots$ it can be shown that

$$\lim_{n\to\infty} a_n = a \quad \Rightarrow \quad \lim_{n\to\infty} \frac{\sum_{i=1}^{n} a_i}{n} = a$$

Hence, if $\lim_{n\to\infty} P(\text{renewal at time } n)$ exists then that limit must equal $\frac{1}{E[\text{time between renewals}]}$.

Example 7.12. Let $X_i, i \geq 1$ be independent and identically distributed random variables, and set

$$S_0 = 0, \quad S_n = \sum_{i=1}^{n} X_i, \ n > 0$$

The process $\{S_n, n \geq 0\}$ is called a *random walk process*. Suppose that $E[X_i] < 0$. The strong law of large numbers yields

$$\lim_{n\to\infty} \frac{S_n}{n} \to E[X_i]$$

But if S_n divided by n is converging to a negative number, then S_n must be going to minus infinity. Let α be the probability that the random walk is always negative after the initial movement. That is,

$$\alpha = P(S_n < 0 \text{ for all } n \geq 1)$$

To determine α, define a counting process by saying that an event occurs at time n if $S(n) < \min(0, S_1, \ldots, S_{n-1})$. That is, an event occurs each time the random walk process reaches a new low. Now, if an event occurs at time n, then the next event will occur k time units later if

$$X_{n+1} \geq 0, X_{n+1} + X_{n+2} \geq 0, \ldots, X_{n+1} + \cdots + X_{n+k-1} \geq 0,$$
$$X_{n+1} + \cdots + X_{n+k} < 0$$

Because $X_i, i \geq 1$ are independent and identically distributed the preceding event is independent of the values of $X_1, \ldots, X_n$, and its probability of occurrence does not depend on n. Consequently, the times between successive events are independent and identically distributed, showing that the counting process is a renewal process. Now,

$$\begin{aligned} P(\text{renewal at } n) &= P(S_n < 0, S_n < S_1, S_n < S_2, \ldots, S_n < S_{n-1}) \\ &= P(X_1 + \cdots + X_n < 0, X_2 + \cdots + X_n < 0, \\ &\quad X_3 + \cdots + X_n < 0, \ldots, X_n < 0) \end{aligned}$$

Because $X_n, X_{n-1}, \ldots, X_1$ has the same joint distribution as does $X_1, X_2, \ldots, X_n$ it follows that the value of the preceding probability would be unchanged if X_1 were replaced by X_n; X_2 were replaced by X_{n-1}; X_3 were replaced by X_{n-2}; and so on. Consequently,

$$\begin{aligned} P(\text{renewal at } n) &= P(X_n + \cdots + X_1 < 0, X_{n-1} + \cdots + X_1 < 0, \\ &\quad X_{n-2} + \cdots + X_1 < 0, X_1 < 0) \\ &= P(S_n < 0, S_{n-1} < 0, S_{n-2} < 0, \ldots, S_1 < 0) \end{aligned}$$

Hence,

$$\lim_{n\to\infty} P(\text{renewal at } n) = P(S_n < 0 \text{ for all } n \geq 1) = \alpha$$

But, by the elementary renewal theorem, this implies that

$$\alpha = 1/E[T]$$

where T is the time between renewals. That is,

$$T = \min\{n : S_n < 0\}$$

For instance, in the case of left skip free random walks (which are ones for which $\sum_{j=-1}^{\infty} P(X_i = j) = 1$) we showed in Section 3.6.6 that $E[T] = -1/E[X_i]$ when $E[X_i] < 0$, showing that for skip free random walks having a negative mean,

$$P(S_n < 0 \text{ for all } n) = -E[X_i]$$

which verifies a result previously obtained in Section 3.6.6. ■

An important limit theorem is the central limit theorem for renewal processes. This states that, for large t, $N(t)$ is approximately normally distributed with mean t/μ and variance $t\sigma^2/\mu^3$, where μ and σ^2 are, respectively, the mean and variance of the interarrival distribution. That is, we have the following theorem which we state without proof.

Theorem 7.3 (Central Limit Theorem for Renewal Processes)**.**

$$\lim_{t\to\infty} P\left\{\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} < x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-x^2/2}dx$$

We now give a heuristic argument to show, for t large, that the distribution of $N(t)$ is approximately that of a normal random variable with mean t/μ and variance $t\sigma^2/\mu^3$.

Heuristic Argument for Central Limit Theorem for Renewal Processes. To begin, note that by the central limit theorem it follows when n is large that $S_n = \sum_{i=1}^{n} X_i$ is approximately a normal random variable with mean $n\mu$ and variance $n\sigma^2$. Consequently, using that $N(t) < n \Leftrightarrow S_n > t$, we see that when n is large

$$\begin{aligned} P(N(t) < n) &= P(S_n > t) \\ &= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{t - n\mu}{\sigma\sqrt{n}}\right) \\ &\approx P(Z > \frac{t - n\mu}{\sigma\sqrt{n}}) \end{aligned} \tag{7.12}$$

where Z is a standard normal random variable. Now,

$$P(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} < x) = P(N(t) < t/\mu + x\sigma\sqrt{t/\mu^3})$$

Treating $t/\mu + x\sigma\sqrt{t/\mu^3}$ as if it were an integer, we see upon letting $n = t/\mu + x\sigma\sqrt{t/\mu^3}$ in Eq. (7.12) that

$$P\left(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} < x\right) \approx P\left(Z > \frac{t - t - x\sigma\mu\sqrt{t/\mu^3}}{\sigma\sqrt{t/\mu + x\sigma\sqrt{t/\mu^3}}}\right)$$

$$= P\left(Z > \frac{-x\sqrt{t/\mu}}{\sqrt{t/\mu + x\sigma\sqrt{t/\mu^3}}}\right)$$

$$\approx P(Z > -x) \qquad \text{when } t \text{ is large}$$

$$= P(Z < x) \qquad \blacksquare$$

In addition, as might be expected from the central limit theorem for renewal processes, it can be shown that $\text{Var}(N(t))/t$ converges to σ^2/μ^3. That is, it can be shown that

$$\lim_{t\to\infty} \frac{\text{Var}(N(t))}{t} = \sigma^2/\mu^3$$

Example 7.13. Two machines continually process an unending number of jobs. The time that it takes to process a job on machine 1 is a gamma random variable with parameters $n = 4, \lambda = 2$, whereas the time that it takes to process a job on machine 2 is uniformly distributed between 0 and 4. Approximate the probability that together the two machines can process at least 90 jobs by time $t = 100$.

Solution: If we let $N_i(t)$ denote the number of jobs that machine i can process by time t, then $\{N_1(t), t \geq 0\}$ and $\{N_2(t),\ t \geq 0\}$ are independent renewal processes. The interarrival distribution of the first renewal process is gamma with parameters $n = 4, \lambda = 2$, and thus has mean 2 and variance 1. Correspondingly, the interarrival distribution of the second renewal process is uniform between 0 and 4, and thus has mean 2 and variance 16/12.

Therefore, $N_1(100)$ is approximately normal with mean 50 and variance 100/8; and $N_2(100)$ is approximately normal with mean 50 and variance 100/6. Hence, $N_1(100) + N_2(100)$ is approximately normal with mean 100 and variance 175/6. Thus, with Φ denoting the standard normal distribution function, we have

$$P\{N_1(100) + N_2(100) > 89.5\} = P\left\{\frac{N_1(100) + N_2(100) - 100}{\sqrt{175/6}} > \frac{89.5 - 100}{\sqrt{175/6}}\right\}$$

$$\approx 1 - \Phi\left(\frac{-10.5}{\sqrt{175/6}}\right)$$

$$\approx \Phi\left(\frac{10.5}{\sqrt{175/6}}\right)$$

$$\approx \Phi(1.944)$$

$$\approx 0.9741 \qquad \blacksquare$$

A counting process with independent interarrival times in which the time until the first event has distribution function G, whereas all the other interarrivals have distribution F is called a *delayed renewal process*. For instance, consider a waiting line system where customers arrive according to a renewal process, and either enter service if they find a free server or join the queue if all servers are busy. Suppose service times are independent with a distribution H. If we say that an event occurs whenever an arrival finds the system empty, then the process probabilistically starts over after each event (because at that time there will be a single customer in the system and that customer will just be starting service, and the arrival process from then on is a renewal process with interarrival distribution F). However, assuming that the system starts empty of customers, the time until the first event will be the time of the first arrival, which has a different distribution than all the other interarrivals, and thus the counting process of events would be a delayed renewal process. For another example, if one starts observing a renewal process at time t, then the time until the first event will have a different distribution than all the other interarrival times.

All the limiting results we have proven and will prove for renewal processes also hold for delayed renewal processes. (That is, it makes no difference in the limit that the renewal process is "delayed" until the first event occurs.) For instance, let $N_d(t)$ be the number of events that occur by time t in a delayed renewal process. Then, because the counting process from the time of the first event X_1 is a renewal process we have that

$$N_d(t) = 1 + N(t - X_1)$$

where $N(s)$, $s \geq 0$ is a renewal process with interarrival distribution F, and where $N(s) = -1$ if $s < 0$. Hence,

$$\frac{N_d(t)}{t} = \frac{1}{t} + \frac{N(t - X_1)}{t - X_1} \frac{t - X_1}{t}$$

Because X_1 is finite, it follows from Proposition 7.1, the strong law for renewal processes, that

$$\lim_{t \to \infty} \frac{N_d(t)}{t} = \frac{1}{\mu}$$

where $\mu = E[X_i]$, $i > 1$ is the mean of the interarrival distribution F.

7.4 Renewal Reward Processes

A large number of probability models are special cases of the following model. Consider a renewal process $\{N(t),\ t \geq 0\}$ having interarrival times $X_n,\ n \geq 1$, and suppose that each time a renewal occurs we receive a reward. We denote by R_n the reward earned at the time of the nth renewal. We shall assume that the $R_n,\ n \geq 1$, are independent and identically distributed. However, we do allow for the possibility that R_n

may (and usually will) depend on X_n, the length of the nth renewal interval. If we let

$$R(t) = \sum_{n=1}^{N(t)} R_n$$

then $R(t)$ represents the total reward earned by time t. Let

$$E[R] = E[R_n], \quad E[X] = E[X_n]$$

Proposition 7.3. *If $E[R] < \infty$ and $E[X] < \infty$, then*

(a) *with probability 1,* $\lim_{t\to\infty} \frac{R(t)}{t} = \frac{E[R]}{E[X]}$

(b) $\lim_{t\to\infty} \frac{E[R(t)]}{t} = \frac{E[R]}{E[X]}$

Proof. We give the proof for (a) only. To prove this, write

$$\frac{R(t)}{t} = \frac{\sum_{n=1}^{N(t)} R_n}{t} = \left(\frac{\sum_{n=1}^{N(t)} R_n}{N(t)}\right)\left(\frac{N(t)}{t}\right)$$

By the strong law of large numbers we obtain

$$\frac{\sum_{n=1}^{N(t)} R_n}{N(t)} \to E[R] \quad \text{as } t \to \infty$$

and by Proposition 7.1

$$\frac{N(t)}{t} \to \frac{1}{E[X]} \quad \text{as } t \to \infty$$

The result thus follows. ■

Remark. **(i)** If we say that a *cycle* is completed every time a renewal occurs, then Proposition 7.3 states that the long-run average reward per unit time is equal to the expected reward earned during a cycle divided by the expected length of a cycle. For instance, in Example 7.6 if we suppose that the amounts that the successive customers deposit in the bank are independent random variables having a common distribution H, then the rate at which deposits accumulate—that is, $\lim_{t\to\infty}$(total deposits by the time t) $/t$)—is given by

$$\frac{E[\text{deposits during a cycle}]}{E[\text{time of cycle}]} = \frac{\mu_H}{\mu_G + 1/\lambda}$$

where $\mu_G + 1/\lambda$ is the mean time of a cycle, and μ_H is the mean of the distribution H.

(ii) Although we have supposed that the reward is earned at the time of a renewal, the result remains valid when the reward is earned gradually throughout the renewal cycle.

Example 7.14 (A Car Buying Model)**.** The lifetime of a car is a continuous random variable having a distribution H and probability density h. Mr. Brown has a policy that he buys a new car as soon as his old one either breaks down or reaches the age of T years. Suppose that a new car costs C_1 dollars and also that an additional cost of C_2 dollars is incurred whenever Mr. Brown's car breaks down. Under the assumption that a used car has no resale value, what is Mr. Brown's long-run average cost?

If we say that a cycle is complete every time Mr. Brown gets a new car, then it follows from Proposition 7.3 (with costs replacing rewards) that his long-run average cost equals

$$\frac{E[\text{cost incurred during a cycle}]}{E[\text{length of a cycle}]}$$

Now letting X be the lifetime of Mr. Brown's car during an arbitrary cycle, then the cost incurred during that cycle will be given by

$$\begin{aligned} &C_1, && \text{if } X > T \\ &C_1 + C_2, && \text{if } X \le T \end{aligned}$$

so the expected cost incurred over a cycle is

$$C_1 P\{X > T\} + (C_1 + C_2) P\{X \le T\} = C_1 + C_2 H(T)$$

Also, the length of the cycle is

$$\begin{aligned} &X, && \text{if } X \le T \\ &T, && \text{if } X > T \end{aligned}$$

and so the expected length of a cycle is

$$\int_0^T x h(x)dx + \int_T^\infty T h(x)dx = \int_0^T x h(x)dx + T[1 - H(T)]$$

Therefore, Mr. Brown's long-run average cost will be

$$\frac{C_1 + C_2 H(T)}{\int_0^T x h(x)dx + T[1 - H(T)]} \tag{7.13}$$

Now, suppose that the lifetime of a car (in years) is uniformly distributed over $(0, 10)$, and suppose that C_1 is 3 (thousand) dollars and C_2 is $\frac{1}{2}$ (thousand) dollars. What value of T minimizes Mr. Brown's long-run average cost?

If Mr. Brown uses the value T, $T \le 10$, then from Eq. (7.13) his long-run average cost equals

$$\begin{aligned} \frac{3 + \frac{1}{2}(T/10)}{\int_0^T (x/10)dx + T(1 - T/10)} &= \frac{3 + T/20}{T^2/20 + (10T - T^2)/10} \\ &= \frac{60 + T}{20T - T^2} \end{aligned}$$

We can now minimize this by using the calculus. Toward this end, let

$$g(T) = \frac{60 + T}{20T - T^2}$$

then

$$g'(T) = \frac{(20T - T^2) - (60 + T)(20 - 2T)}{(20T - T^2)^2}$$

Equating to 0 yields

$$20T - T^2 = (60 + T)(20 - 2T)$$

or, equivalently,

$$T^2 + 120\,T - 1200 = 0$$

which yields the solutions

$$T \approx 9.25 \quad \text{and} \quad T \approx -129.25$$

Since $T \leq 10$, it follows that the optimal policy for Mr. Brown would be to purchase a new car whenever his old car reaches the age of 9.25 years. ■

Example 7.15 (Dispatching a Train)**.** Suppose that customers arrive at a train depot in accordance with a renewal process having a mean interarrival time μ. Whenever there are N customers waiting in the depot, a train leaves. If the depot incurs a cost at the rate of nc dollars per unit time whenever there are n customers waiting, what is the average cost incurred by the depot?

If we say that a cycle is completed whenever a train leaves, then the preceding is a renewal reward process. The expected length of a cycle is the expected time required for N customers to arrive and, since the mean interarrival time is μ, this equals

$$E[\text{length of cycle}] = N\mu$$

If we let T_n denote the time between the nth and $(n + 1)$st arrival in a cycle, then the expected cost of a cycle may be expressed as

$$E[\text{cost of a cycle}] = E[c\,T_1 + 2c\,T_2 + \cdots + (N - 1)\,c\,T_{N-1}]$$

which, since $E[T_n] = \mu$, equals

$$c\mu\frac{N}{2}(N - 1)$$

Hence, the average cost incurred by the depot is

$$\frac{c\mu\,N(N - 1)}{2N\mu} = \frac{c(N - 1)}{2}$$

Suppose now that each time a train leaves, the depot incurs a cost of six units. What value of N minimizes the depot's long-run average cost when $c = 2$, $\mu = 1$?

In this case, we have that the average cost per unit time is

$$\frac{6 + c\mu N(N-1)/2}{N\mu} = N - 1 + \frac{6}{N}$$

By treating this as a continuous function of N and using the calculus, we obtain that the minimal value of N is

$$N = \sqrt{6} \approx 2.45$$

Hence, the optimal integral value of N is either 2 which yields a value 4, or 3 which also yields the value 4. Hence, either $N = 2$ or $N = 3$ minimizes the depot's average cost. ■

Example 7.16. Suppose that customers arrive at a single-server system in accordance with a Poisson process with rate λ. Upon arriving a customer must pass through a door that leads to the server. However, each time someone passes through, the door becomes locked for the next t units of time. An arrival finding a locked door is lost, and a cost c is incurred by the system. An arrival finding the door unlocked passes through to the server. If the server is free, the customer enters service; if the server is busy, the customer departs without service and a cost K is incurred. If the service time of a customer is exponential with rate μ, find the average cost per unit time incurred by the system.

Solution: The preceding can be considered to be a renewal reward process, with a new cycle beginning each time a customer arrives to find the door unlocked. This is so because whether or not the arrival finds the server free, the door will become locked for the next t time units and the server will be busy for a time X that is exponentially distributed with rate μ. (If the server is free, X is the service time of the entering customer; if the server is busy, X is the remaining service time of the customer in service.) Since the next cycle will begin at the first arrival epoch after a time t has passed, it follows that

$$E[\text{time of a cycle}] = t + 1/\lambda$$

Let C_1 denote the cost incurred during a cycle due to arrivals finding the door locked. Then, since each arrival in the first t time units of a cycle will result in a cost c, we have

$$E[C_1] = \lambda t c$$

Also, let C_2 denote the cost incurred during a cycle due to an arrival finding the door unlocked but the server busy. Then because a cost K is incurred if the server is still busy a time t after the cycle began and, in addition, the next arrival after that

time occurs before the service completion, we see that

$$E[C_2] = Ke^{-\mu t}\frac{\lambda}{\lambda+\mu}$$

Consequently,

$$\text{average cost per unit time} = \frac{\lambda tc + \lambda K e^{-\mu t}/(\lambda+\mu)}{t + 1/\lambda}$$ ■

Example 7.17. Consider a manufacturing process that sequentially produces items, each of which is either defective or acceptable. The following type of sampling scheme is often employed in an attempt to detect and eliminate most of the defective items. Initially, each item is inspected and this continues until there are k consecutive items that are acceptable. At this point 100% inspection ends and each successive item is independently inspected with probability α. This partial inspection continues until a defective item is encountered, at which time 100% inspection is reinstituted, and the process begins anew. If each item is, independently, defective with probability q,

(a) what proportion of items are inspected?
(b) if defective items are removed when detected, what proportion of the remaining items are defective?

Remark. Before starting our analysis, note that the preceding inspection scheme was devised for situations in which the probability of producing a defective item changed over time. It was hoped that 100% inspection would correlate with times at which the defect probability was large and partial inspection when it was small. However, it is still important to see how such a scheme would work in the extreme case where the defect probability remains constant throughout.

Solution: We begin our analysis by noting that we can treat the preceding as a renewal reward process with a new cycle starting each time 100% inspection is instituted. We then have

$$\text{proportion of items inspected} = \frac{E[\text{number inspected in a cycle}]}{E[\text{number produced in a cycle}]}$$

Let N_k denote the number of items inspected until there are k consecutive acceptable items. Once partial inspection begins—that is, after N_k items have been produced—since each inspected item will be defective with probability q, it follows that the expected number that will have to be inspected to find a defective item is $1/q$. Hence,

$$E[\text{number inspected in a cycle}] = E[N_k] + \frac{1}{q}$$

In addition, since at partial inspection each item produced will, independently, be inspected and found to be defective with probability αq, it follows that the number

of items produced until one is inspected and found to be defective is $1/\alpha q$, and so

$$E[\text{number produced in a cycle}] = E[N_k] + \frac{1}{\alpha q}$$

Also, as $E[N_k]$ is the expected number of trials needed to obtain k acceptable items in a row when each item is acceptable with probability $p = 1 - q$, it follows from Example 3.15 that

$$E[N_k] = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k} = \frac{(1/p)^k - 1}{q}$$

Hence, we obtain

$$P_I \equiv \text{proportion of items that are inspected} = \frac{(1/p)^k}{(1/p)^k - 1 + 1/\alpha}$$

To answer (b), note first that since each item produced is defective with probability q it follows that the proportion of items that are both inspected and found to be defective is qP_{I}. Hence, for N large, out of the first N items produced there will be (approximately) NqP_{I} that are discovered to be defective and thus removed. As the first N items will contain (approximately) Nq defective items, it follows that there will be $Nq - NqP_{\text{I}}$ defective items not discovered. Hence,

$$\text{proportion of the nonremoved items that are defective} \approx \frac{Nq(1 - P_{\text{I}})}{N(1 - qP_{\text{I}})}$$

As the approximation becomes exact as $N \to \infty$, we see that

$$\text{proportion of the nonremoved items that are defective} = \frac{q(1 - P_{\text{I}})}{(1 - qP_{\text{I}})} \qquad \blacksquare$$

Example 7.18 (The Average Age of a Renewal Process)**.** Consider a renewal process having interarrival distribution F and define $A(t)$ to be the time at t since the last renewal. If renewals represent old items failing and being replaced by new ones, then $A(t)$ represents the age of the item in use at time t. Since $S_{N(t)}$ represents the time of the last event prior to or at time t, we have

$$A(t) = t - S_{N(t)}$$

We are interested in the average value of the age—that is, in

$$\lim_{s \to \infty} \frac{\int_0^s A(t)dt}{s}$$

To determine this quantity, we use renewal reward theory in the following way: Let us assume that at any time we are being paid money at a rate equal to the age of the renewal process at that time. That is, at time t, we are being paid at rate $A(t)$, and

so $\int_0^s A(t)dt$ represents our total earnings by time s. As everything starts over again when a renewal occurs, it follows that

$$\frac{\int_0^s A(t)dt}{s} \to \frac{E[\text{reward during a renewal cycle}]}{E[\text{time of a renewal cycle}]}$$

Now, since the age of the renewal process a time t into a renewal cycle is just t, we have

$$\begin{aligned}\text{reward during a renewal cycle} &= \int_0^X t\,dt \\ &= \frac{X^2}{2}\end{aligned}$$

where X is the time of the renewal cycle. Hence, we have that

$$\begin{aligned}\text{average value of age} &\equiv \lim_{s\to\infty} \frac{\int_0^s A(t)dt}{s} \\ &= \frac{E[X^2]}{2E[X]} \qquad (7.14)\end{aligned}$$

where X is an interarrival time having distribution function F. ■

Example 7.19 (The Average Excess of a Renewal Process)**.** Another quantity associated with a renewal process is $Y(t)$, the excess or residual time at time t. $Y(t)$ is defined to equal the time from t until the next renewal and, as such, represents the remaining (or residual) life of the item in use at time t. The average value of the excess, namely,

$$\lim_{s\to\infty} \frac{\int_0^s Y(t)dt}{s}$$

also can be easily obtained by renewal reward theory. To do so, suppose that we are paid at time t at a rate equal to $Y(t)$. Then our average reward per unit time will, by renewal reward theory, be given by

$$\begin{aligned}\text{average value of excess} &\equiv \lim_{s\to\infty} \frac{\int_0^s Y(t)dt}{s} \\ &= \frac{E[\text{reward during a cycle}]}{E[\text{length of a cycle}]}\end{aligned}$$

Now, letting X denote the length of a renewal cycle, we have

$$\begin{aligned}\text{reward during a cycle} &= \int_0^X (X-t)dt \\ &= \frac{X^2}{2}\end{aligned}$$

and thus the average value of the excess is

$$\text{average value of excess} = \frac{E[X^2]}{2E[X]}$$

which was the same result obtained for the average value of the age of a renewal process. ■

Example 7.20. Suppose that passengers arrive at a bus stop according to a Poisson process with rate λ. Suppose also that buses arrive according to a renewal process with distribution function F, and that buses pick up all waiting passengers. Assuming that the Poisson process of people arriving and the renewal process of buses arriving are independent, find

(a) the average number of people who are waiting for a bus, averaged over all time; and

(b) the average amount of time that a passenger waits, averaged over all passengers.

Solution: We will solve this by using renewal reward processes. Say that a new cycle begins each time a bus arrives. Let T be the time of a cycle, and note that T has distribution function F. If we suppose that each passenger pays us money at a rate of 1 per unit time while they wait for a bus, then the reward rate at any time is the number waiting at that time, and so the average reward per unit time is the average number of people that are waiting for a bus. Letting R be the reward earned during a cycle, the renewal reward theorem gives

$$\text{Average Number Waiting} = \frac{E[R]}{E[T]}$$

Let N be the number of arrivals during a cycle. To determine $E[R]$, we will condition on the values of both T and N. Now,

$$E[R|T = t, N = n] = nt/2$$

which follows because given there are n arrivals by time t their set of arrival times are distributed as n independent uniform $(0, t)$ random variables, and so the average amount received per passenger is $t/2$. Hence,

$$E[R|T, N] = NT/2$$

Taking expectations yields

$$E[R] = \frac{1}{2}E[NT]$$

To compute $E[NT]$, condition on T to obtain

$$E[NT|T] = TE[N|T] = \lambda T^2$$

where the preceding follows because, given the time T until the bus arrives, the number of people waiting is Poisson distributed with mean λT. Hence, upon taking expectations of the preceding, we obtain

$$E[R] = \frac{1}{2}E[NT] = \lambda E[T^2]/2$$

which gives that

$$\text{Average Number Waiting} = \frac{\lambda E[T^2]}{2E[T]}$$

where T has the interarrival distribution F.

To determine the average amount of time that a passenger waits note that, because each passenger pays 1 per unit time while waiting for a bus, the total amount paid by a passenger is the amount of time the passenger waits. Because R is the total reward earned in a cycle, it thus follows that, with W_i being the waiting time of passenger i,

$$R = W_1 + \cdots + W_N$$

Now, if we consider the rewards earned from successive passengers, namely $W_1, W_2, \ldots$, and imagine that the reward W_i is earned at time i, then this sequence of rewards constitutes a discrete time renewal reward process in which a new cycle begins at time $N + 1$. Consequently, from renewal reward process theory and the preceding identity, we see that

$$\lim_{n\to\infty} \frac{W_1 + \cdots + W_n}{n} = \frac{E[W_1 + \cdots + W_N]}{E[N]} = \frac{E[R]}{E[N]}$$

Using that

$$E[N] = E[E[N|T]] = E[\lambda T] = \lambda E[T]$$

along with the previously derived $E[R] = \lambda E[T^2]/2$ we obtain the result

$$\lim_{n\to\infty} \frac{W_1 + \cdots + W_n}{n} = \frac{E[T^2]}{2E[T]}$$

Because $\frac{E[T^2]}{2E[T]}$ is the average value of the excess for the renewal process of arriving buses, the preceding equation yields the interesting result that the average waiting time of a passenger is equal to the average time until the next bus arrives when we average over all time. Because passengers are arriving according to a Poisson process, this result is a special case of a general result, known as the PASTA principle, to be presented in Chapter 8. The PASTA principle says that a system as seen by Poisson arrivals is the same as the system as averaged over all time. (In our example, the system refers to the time until the next bus.) ■

Consider an irreducible and positive recurrent Markov chain with state space S and with transition probabilities $P_{i,j}$. With π_i denoting the long run proportion of time that the Markov chain is in state i, $i \in S$, we gave a heuristic argument in Chapter 4 that these quantities satisfy the stationarity equations

$$\pi_j = \sum_i \pi_i P_{i,j}$$

$$\sum_j \pi_j = 1$$

For a rigorous argument, fix some state, say state 0, and say that a new cycle starts whenever the Markov chain enters state 0. Now, if for a fixed state i, we suppose that we earn 1 each time the chain enters state i, then the total amount earned by time n is the amount of time the chain is in state i by time n, and thus the average reward per unit time is equal to π_i. But from renewal reward theory, the average reward per unit time is equal to the expected reward earned in a cycle divided by the expected time of a cycle. Hence, if we let N_i denote the number of periods the Markov chain is in state i during a cycle, and let N denote the number of periods in a cycle, then

$$\pi_i = \frac{E[N_i]}{E[N]}.$$

Now, if we let $N_{i,j}$ denote the number of transitions from i to j that occur in a cycle, then as each visit to state i is followed by a transition into j with probability $P_{i,j}$, it seems intuitive that

$$E[N_{i,j}] = E[N_i]P_{i,j} \tag{7.15}$$

Assuming that (7.15) holds, then taking expectations of the identity

$$N_j = \sum_i N_{i,j}$$

would yield that

$$E[N_j] = \sum_i E[N_i]P_{i,j}$$

and dividing by $E[N]$ would give that

$$\pi_j = \sum_i \pi_i P_{i,j}$$

In addition, because $N = \sum_j N_j$, it follows that

$$E[N] = \sum_j E[N_j]$$

Dividing both sides of the preceding by $E[N]$ yields that

$$\sum_j \pi_j = 1$$

Thus, the verification that there are stationary probabilities when the Markov chain is irreducible and positive recurrent will follow once we establish (7.15). To do so, let $I_{i,j}(k)$ equal 1 if the on the kth visit of the chain to state i the next state is j. (Because the chain is recurrent, state i will be visited infinitely often.) It is easy to see that N_i is a stopping time for the sequence $I_{i,j}(k)$, $k \geq 1$. (For instance, suppose that $N_i = 10$, and so the chain spends 10 periods in state i during a cycle. Whereas that gives us some probabilistic information about where those 10 transitions coming from i were into, it gives no information about what will happen the next time the chain enters state i.) Hence, using that

$$N_{i,j} = \sum_{k=1}^{N_i} I_{i,j}(k)$$

we obtain, upon taking expectations and then applying Wald's equation, that

$$E[N_{i,j}] = E[\sum_{k=1}^{N_i} I_{i,j}(k)] = E[N_i]E[I_{i,j}(k)] = E[N_i]\, P_{i,j}$$

which completes the verification. ■

7.5 Regenerative Processes

Consider a stochastic process $\{X(t), t \geq 0\}$ with state space $0, 1, 2, \ldots$, having the property that there exist time points at which the process (probabilistically) restarts itself. That is, suppose that with probability 1, there exists a time T_1, such that the continuation of the process beyond T_1 is a probabilistic replica of the whole process starting at 0. Note that this property implies the existence of further times $T_2, T_3, \ldots$, having the same property as T_1. Such a stochastic process is known as a *regenerative process*.

From the preceding, it follows that $T_1, T_2, \ldots$, constitute the arrival times of a renewal process, and we shall say that a cycle is completed every time a renewal occurs.

Examples. **(1)** A renewal process is regenerative, and T_1 represents the time of the first renewal.

(2) A recurrent Markov chain is regenerative, and T_1 represents the time of the first transition into the initial state.

We are interested in determining the long-run proportion of time that a regenerative process spends in state j. To obtain this quantity, let us imagine that we earn a reward at a rate 1 per unit time when the process is in state j and at rate 0 otherwise. That is, if $I(s)$ represents the rate at which we earn at time s, then

$$I(s) = \begin{cases} 1, & \text{if } X(s) = j \\ 0, & \text{if } X(s) \neq j \end{cases}$$

and

$$\text{total reward earned by } t = \int_0^t I(s)\,ds$$

As the preceding is clearly a renewal reward process that starts over again at the cycle time T_1, we see from Proposition 7.3 that

$$\text{average reward per unit time} = \frac{E[\text{reward by time } T_1]}{E[T_1]}$$

However, the average reward per unit is just equal to the proportion of time that the process is in state j. That is, we have the following.

Proposition 7.4. *For a regenerative process, the long-run*

$$\textit{proportion of time in state } j = \frac{E[\textit{amount of time in } j \textit{ during a cycle}]}{E[\textit{time of a cycle}]}$$

Remark. If the cycle time T_1 is a continuous random variable, then it can be shown by using an advanced theorem called the "key renewal theorem" that the preceding is equal also to the limiting probability that the system is in state j at time t. That is, if T_1 is continuous, then

$$\lim_{t\to\infty} P\{X(t) = j\} = \frac{E[\text{amount of time in } j \text{ during a cycle}]}{E[\text{time of a cycle}]}$$

Example 7.21. Consider a positive recurrent continuous-time Markov chain that is initially in state i. By the Markovian property, each time the process reenters state i it starts over again. Thus returns to state i are renewals and constitute the beginnings of new cycles. By Proposition 7.4, it follows that the long-run

$$\text{proportion of time in state } j = \frac{E[\text{amount of time in } j \text{ during an } i - i \text{ cycle}]}{\mu_{ii}}$$

where μ_{ii} represents the mean time to return to state i. If we take j to equal i, then we obtain

$$\text{proportion of time in state } i = \frac{1/v_i}{\mu_{ii}}$$ ■

Example 7.22 (A Queueing System with Renewal Arrivals)**.** Consider a waiting time system in which customers arrive in accordance with an arbitrary renewal process and are served one at a time by a single server having an arbitrary service distribution. If we suppose that at time 0 the initial customer has just arrived, then $\{X(t), t \geq 0\}$ is a regenerative process, where $X(t)$ denotes the number of customers in the system at time t. The process regenerates each time a customer arrives and finds the server free. ■

Example 7.23. Although a system needs only a single machine to function, it maintains an additional machine as a backup. A machine in use functions for a random time with density function f and then fails. If a machine fails while the other one is in working condition, then the latter is put in use and, simultaneously, repair begins on the one that just failed. If a machine fails while the other machine is in repair, then the newly failed machine waits until the repair is completed; at that time the repaired machine is put in use and, simultaneously, repair begins on the recently failed one. All repair times have density function g. Find P_0, P_1, P_2, where P_i is the long-run proportion of time that exactly i of the machines are in working condition.

Solution: Let us say that the system is in state i whenever i machines are in working condition $i = 0, 1, 2$. It is then easy to see that every time the system enters state 1 it probabilistically starts over. That is, the system restarts every time that a machine is put in use while, simultaneously, repair begins on the other one. Say that a cycle begins each time the system enters state 1. If we let X denote the working time of the machine put in use at the beginning of a cycle, and let R be the repair time of the other machine, then the length of the cycle, call it T_c, can be expressed as

$$T_c = \max(X, R)$$

The preceding follows when $X \leq R$, because, in this case, the machine in use fails before the other one has been repaired, and so a new cycle begins when that repair is completed. Similarly, it follows when $R < X$, because then the repair occurs first, and so a new cycle begins when the machine in use fails. Also, let T_i, $i = 0, 1, 2$, be the amount of time that the system is in state i during a cycle. Then, because the amount of time during a cycle that neither machine is working is $R - X$ provided that this quantity is positive or 0 otherwise, we have

$$T_0 = (R - X)^+$$

Similarly, because the amount of time during the cycle that a single machine is working is $\min(X, R)$, we have

$$T_1 = \min(X, R)$$

Finally, because the amount of time during the cycle that both machines are working is $X - R$ if this quantity is positive or 0 otherwise, we have

$$T_2 = (X - R)^+$$

Hence, we obtain

$$P_0 = \frac{E[(R-X)^+]}{E[\max(X,R)]}$$

$$P_1 = \frac{E[\min(X,R)]}{E[\max(X,R)]}$$

$$P_2 = \frac{E[(X-R)^+]}{E[\max(X,R)]}$$

That $P_0 + P_1 + P_2 = 1$ follows from the easily checked identity

$$\max(x,r) = \min(x,r) + (x-r)^+ + (r-x)^+$$

The preceding expectations can be computed as follows:

$$\begin{aligned}
E[\max(X,R)] &= \int_0^\infty \int_0^\infty \max(x,r) f(x) g(r) dx\, dr \\
&= \int_0^\infty \int_0^r r f(x) g(r) dx\, dr + \int_0^\infty \int_r^\infty x f(x) g(r) dx\, dr \\
E[(R-X)^+] &= \int_0^\infty \int_0^\infty (r-x)^+ f(x) g(r) dx\, dr \\
&= \int_0^\infty \int_0^r (r-x) f(x) g(r) dx\, dr \\
E[\min(X,R)] &= \int_0^\infty \int_0^\infty \min(x,r) f(x) g(r) dx\, dr \\
&= \int_0^\infty \int_0^r x f(x) g(r) dx\, dr + \int_0^\infty \int_r^\infty r f(x) g(r) dx\, dr \\
E[(X-R)^+] &= \int_0^\infty \int_0^x (x-r) f(x) g(r) dr\, dx
\end{aligned}$$

■

7.5.1 Alternating Renewal Processes

Another example of a regenerative process is provided by what is known as an *alternating renewal process*, which considers a system that can be in one of two states: on or off. Initially it is on, and it remains on for a time Z_1; it then goes off and remains off for a time Y_1. It then goes on for a time Z_2; then off for a time Y_2; then on, and so on.

We suppose that the random vectors (Z_n, Y_n), $n \geq 1$ are independent and identically distributed. That is, both the sequence of random variables $\{Z_n\}$ and the sequence $\{Y_n\}$ are independent and identically distributed; but we allow Z_n and Y_n to be dependent. In other words, each time the process goes on, everything starts over again, but when it then goes off, we allow the length of the off time to depend on the previous on time.

Let $E[Z] = E[Z_n]$ and $E[Y] = E[Y_n]$ denote, respectively, the mean lengths of an on and off period.

We are concerned with P_{on}, the long-run proportion of time that the system is on. If we let

$$X_n = Y_n + Z_n, \quad n \geq 1$$

then at time X_1 the process starts over again. That is, the process starts over again after a complete cycle consisting of an on and an off interval. In other words, a renewal occurs whenever a cycle is completed. Therefore, we obtain from Proposition 7.4 that

$$\begin{aligned} P_{\text{on}} &= \frac{E[Z]}{E[Y] + E[Z]} \\ &= \frac{E[\text{on}]}{E[\text{on}] + E[\text{off}]} \end{aligned} \tag{7.16}$$

Also, if we let P_{off} denote the long-run proportion of time that the system is off, then

$$\begin{aligned} P_{\text{off}} &= 1 - P_{\text{on}} \\ &= \frac{E[\text{off}]}{E[\text{on}] + E[\text{off}]} \end{aligned}$$

Example 7.24 (A Production Process)**.** One example of an alternating renewal process is a production process (or a machine) that works for a time Z_1, then breaks down and has to be repaired (which takes a time Y_1), then works for a time Z_2, then is down for a time Y_2, and so on. If we suppose that the process is as good as new after each repair, then this constitutes an alternating renewal process. It is worthwhile to note that in this example it makes sense to suppose that the repair time will depend on the amount of time the process had been working before breaking down. ■

Example 7.25. The rate a certain insurance company charges its policyholders alternates between r_1 and r_0. A new policyholder is initially charged at a rate of r_1 per unit time. When a policyholder paying at rate r_1 has made no claims for the most recent s time units, then the rate charged becomes r_0 per unit time. The rate charged remains at r_0 until a claim is made, at which time it reverts to r_1. Suppose that a given policyholder lives forever and makes claims at times chosen according to a Poisson process with rate λ, and find

(a) P_i, the proportion of time that the policyholder pays at rate r_i, $i = 0, 1$;
(b) the long-run average amount paid per unit time.

Solution: If we say that the system is "on" when the policyholder pays at rate r_1 and "off" when she pays at rate r_0, then this on–off system is an alternating renewal process with a new cycle starting each time a claim is made. If X is the time between successive claims, then the on time in the cycle is the smaller of s and X. (Note that if $X < s$, then the off time in the cycle is 0.) Since X is exponential with rate λ, the preceding yields

$$
\begin{aligned}
E[\text{on time in cycle}] &= E[\min(X, s)] \\
&= \int_0^s x\lambda e^{-\lambda x}\, dx + se^{-\lambda s} \\
&= \frac{1}{\lambda}(1 - e^{-\lambda s})
\end{aligned}
$$

Since $E[X] = 1/\lambda$, we see that

$$
P_1 = \frac{E[\text{on time in cycle}]}{E[X]} = 1 - e^{-\lambda s}
$$

and

$$
P_0 = 1 - P_1 = e^{-\lambda s}
$$

The long-run average amount paid per unit time is

$$
r_0 P_0 + r_1 P_1 = r_1 - (r_1 - r_0)e^{-\lambda s}
$$

■

Example 7.26 (The Age of a Renewal Process)**.** Suppose we are interested in determining the proportion of time that the age of a renewal process is less than some constant c. To do so, let a cycle correspond to a renewal, and say that the system is "on" at time t if the age at t is less than or equal to c, and say it is "off" if the age at t is greater than c. In other words, the system is "on" the first c time units of a renewal interval, and "off" the remaining time. Hence, letting X denote a renewal interval, we have, from Eq. (7.16),

$$
\begin{aligned}
\text{proportion of time age is less than } c &= \frac{E[\min(X, c)]}{E[X]} \\
&= \frac{\int_0^\infty P\{\min(X, c) > x\}dx}{E[X]} \\
&= \frac{\int_0^c P\{X > x\}dx}{E[X]} \\
&= \frac{\int_0^c (1 - F(x))dx}{E[X]} \qquad (7.17)
\end{aligned}
$$

where F is the distribution function of X and where we have used the identity that for a nonnegative random variable Y

$$
E[Y] = \int_0^\infty P\{Y > x\}dx
$$

■

Example 7.27 (The Excess of a Renewal Process)**.** Let us now consider the long-run proportion of time that the excess of a renewal process is less than c. To determine this quantity, let a cycle correspond to a renewal interval and say that the system is on whenever the excess of the renewal process is greater than or equal to c and that it

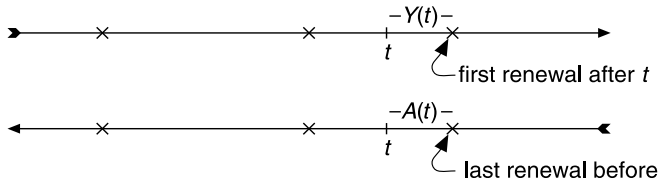

Figure 7.3 Arrowheads indicate direction of time.

is off otherwise. In other words, whenever a renewal occurs the process goes on and stays on until the last c time units of the renewal interval when it goes off. Clearly this is an alternating renewal process, and so we obtain from Eq. (7.16) that

$$\text{long-run proportion of time the excess is less than } c = \frac{E[\text{off time in cycle}]}{E[\text{cycle time}]}$$

If X is the length of a renewal interval, then since the system is off the last c time units of this interval, it follows that the off time in the cycle will equal $\min(X, c)$. Thus,

$$\begin{aligned}\text{long-run proportion of time the excess is less than } c &= \frac{E[\min(X, c)]}{E[X]} \\ &= \frac{\int_0^c (1 - F(x))dx}{E[X]}\end{aligned}$$

where the final equality follows from Eq. (7.17). Thus, we see from the result of Example 7.26 that the long-run proportion of time that the excess is less than c and the long-run proportion of time that the age is less than c are equal. One way to understand this equivalence is to consider a renewal process that has been in operation for a long time and then observe it going backwards in time. In doing so, we observe a counting process where the times between successive events are independent random variables having distribution F. That is, when we observe a renewal process going backwards in time we again observe a renewal process having the same probability structure as the original. Since the excess (age) at any time for the backwards process corresponds to the age (excess) at that time for the original renewal process (see Fig. 7.3), it follows that all long-run properties of the age and the excess must be equal. ■

If μ is the mean interarrival time, then the distribution function F_e, defined by

$$F_e(x) = \int_0^x \frac{1 - F(y)}{\mu} dy$$

is called the *equilibrium distribution* of F. From the preceding, it follows that $F_e(x)$ represents the long-run proportion of time that the age, and the excess, of the renewal process is less than or equal to x.

Example 7.28 (The Busy Period of the $M/G/\infty$ Queue)**.** The infinite server queueing system in which customers arrive according to a Poisson process with rate λ, and have

a general service distribution G, was analyzed in Section 5.3, where it was shown that the number of customers in the system at time t is Poisson distributed with mean $\lambda \int_0^t \bar{G}(y)dy$. If we say that the system is busy when there is at least one customer in the system and is idle when the system is empty, find $E[B]$, the expected length of a busy period.

Solution: If we say that the system is on when there is at least one customer in the system, and off when the system is empty, then we have an alternating renewal process. Because $\int_0^\infty \bar{G}(t)dt = E[S]$, where $E[S]$ is the mean of the service distribution G, it follows from the result of Section 5.3 that

$$\lim_{t\to\infty} P\{\text{system off at } t\} = e^{-\lambda E[S]}$$

Consequently, from alternating renewal process theory we obtain

$$e^{-\lambda E[S]} = \frac{E[\text{off time in cycle}]}{E[\text{cycle time}]}$$

But when the system goes off, it remains off only up to the time of the next arrival, giving that

$$E[\text{off time in cycle}] = 1/\lambda$$

Because

$$E[\text{on time in cycle}] = E[B]$$

we obtain

$$e^{-\lambda E[S]} = \frac{1/\lambda}{1/\lambda + E[B]}$$

or

$$E[B] = \frac{1}{\lambda}\left(e^{\lambda E[S]} - 1\right)$$ ■

Example 7.29 (An Inventory Example)**.** Suppose that customers arrive at a specified store in accordance with a renewal process having interarrival distribution F. Suppose that the store stocks a single type of item and that each arriving customer desires a random amount of this commodity, with the amounts desired by the different customers being independent random variables having the common distribution G. The store uses the following (s, S) ordering policy: If its inventory level falls below s then it orders enough to bring its inventory up to S. That is, if the inventory after serving a customer is x, then the amount ordered is

$$\begin{cases} S - x, & \text{if } x < s \\ 0, & \text{if } x \geq s \end{cases}$$

The order is assumed to be instantaneously filled.

For a fixed value $y, s \leq y \leq S$, suppose that we are interested in determining the long-run proportion of time that the inventory on hand is at least as large as y. To determine this quantity, let us say that the system is "on" whenever the inventory level is at least y and is "off" otherwise. With these definitions, the system will go on each time that a customer's demand causes the store to place an order that results in its inventory level returning to S. Since whenever this occurs a customer must have just arrived it follows that the times until succeeding customers arrive will constitute a renewal process with interarrival distribution F; that is, the process will start over each time the system goes back on. Thus, the on and off periods so defined constitute an alternating renewal process, and from Eq. (7.16) we have that

$$\text{long-run proportion of time inventory} \geq y = \frac{E[\text{on time in a cycle}]}{E[\text{cycle time}]} \tag{7.18}$$

Now, if we let $D_1, D_2, \ldots$ denote the successive customer demands, and let

$$N_x = \min(n : D_1 + \cdots + D_n > S - x) \tag{7.19}$$

then it is the N_y customer in the cycle that causes the inventory level to fall below y, and it is the N_s customer that ends the cycle. As a result, if we let X_i, $i \geq 1$, denote the interarrival times of customers, then

$$\text{on time in a cycle} = \sum_{i=1}^{N_y} X_i \tag{7.20}$$

$$\text{cycle time} = \sum_{i=1}^{N_s} X_i \tag{7.21}$$

Assuming that the interarrival times are independent of the successive demands, we have that

$$\begin{aligned} E\left[\sum_{i=1}^{N_y} X_i\right] &= E\left[E\left[\sum_{i=1}^{Ny} X_i | N_y\right]\right] \\ &= E[N_y E[X]] \\ &= E[X]E[N_y] \end{aligned}$$

Similarly,

$$E\left[\sum_{i=1}^{N_s} X_i\right] = E[X]E[N_s]$$

Therefore, from Eqs. (7.18), (7.20), and (7.21) we see that

$$\text{long-run proportion of time inventory} \geq y = \frac{E[N_y]}{E[N_s]} \tag{7.22}$$

However, as the $D_i, i \geq 1$, are independent and identically distributed nonnegative random variables with distribution G, it follows from Eq. (7.19) that N_x has the same distribution as the index of the first event to occur after time $S - x$ of a renewal process having interarrival distribution G. That is, $N_x - 1$ would be the number of renewals by time $S - x$ of this process. Hence, we see that

$$E[N_y] = m(S - y) + 1,$$
$$E[N_s] = m(S - s) + 1$$

where

$$m(t) = \sum_{n=1}^{\infty} G_n(t)$$

From Eq. (7.22), we arrive at

$$\text{long-run proportion of time inventory} \geq y = \frac{m(S - y) + 1}{m(S - s) + 1}, \quad s \leq y \leq S$$

For instance, if the customer demands are exponentially distributed with mean $1/\mu$, then

$$\text{long-run proportion of time inventory} \geq y = \frac{\mu(S - y) + 1}{\mu(S - s) + 1}, \quad s \leq y \leq S \qquad \blacksquare$$

7.6 Semi-Markov Processes

Consider a process that can be in state 1 or state 2 or state 3. It is initially in state 1 where it remains for a random amount of time having mean μ_1, then it goes to state 2 where it remains for a random amount of time having mean μ_2, then it goes to state 3 where it remains for a mean time μ_3, then back to state 1, and so on. What proportion of time is the process in state i, $i = 1, 2, 3$?

If we say that a cycle is completed each time the process returns to state 1, and if we let the reward be the amount of time we spend in state i during that cycle, then the preceding is a renewal reward process. Hence, from Proposition 7.3 we obtain that P_i, the proportion of time that the process is in state i, is given by

$$P_i = \frac{\mu_i}{\mu_1 + \mu_2 + \mu_3}, \quad i = 1, 2, 3$$

Similarly, if we had a process that could be in any of N states $1, 2, \ldots, N$ and that moved from state $1 \to 2 \to 3 \to \cdots \to N-1 \to N \to 1$, then the long-run proportion of time that the process spends in state i is

$$P_i = \frac{\mu_i}{\mu_1 + \mu_2 + \cdots + \mu_N}, \quad i = 1, 2, \ldots, N$$

where μ_i is the expected amount of time the process spends in state i during each visit.

Let us now generalize the preceding to the following situation. Suppose that a process can be in any one of N states $1, 2, \ldots, N$, and that each time it enters state i it remains there for a random amount of time having mean μ_i and then makes a transition into state j with probability P_{ij}. Such a process is called a *semi-Markov process*. Note that if the amount of time that the process spends in each state before making a transition is identically 1, then the semi-Markov process is just a Markov chain.

Let us calculate P_i for a semi-Markov process. To do so, we first consider π_i, the proportion of transitions that take the process into state i. Now, if we let X_n denote the state of the process after the nth transition, then $\{X_n, n \geq 0\}$ is a Markov chain with transition probabilities $P_{ij}, i, j = 1, 2, \ldots, N$. Hence, π_i will just be the limiting (or stationary) probabilities for this Markov chain (Section 4.4). That is, π_i will be the unique nonnegative solution[1] of

$$\sum_{i=1}^{N} \pi_i = 1,$$

$$\pi_i = \sum_{j=1}^{N} \pi_j P_{ji}, \quad i = 1, 2, \ldots, N \tag{7.23}$$

Now, since the process spends an expected time μ_i in state i whenever it visits that state, it seems intuitive that P_i should be a weighted average of the π_i where π_i is weighted proportionately to μ_i. That is,

$$P_i = \frac{\pi_i \mu_i}{\sum_{j=1}^{N} \pi_j \mu_j}, \quad i = 1, 2, \ldots, N \tag{7.24}$$

where the π_i are given as the solution to Eq. (7.23).

Example 7.30. Consider a machine that can be in one of three states: *good condition, fair condition*, or *broken down*. Suppose that a machine in good condition will remain this way for a mean time μ_1 and then will go to either the fair condition or the broken condition with respective probabilities $\frac{3}{4}$ and $\frac{1}{4}$. A machine in fair condition will remain that way for a mean time μ_2 and then will break down. A broken machine will be repaired, which takes a mean time μ_3, and when repaired will be in good condition

[1] We shall assume that there exists a solution of Eq. (7.23). That is, we assume that all of the states in the Markov chain communicate.

with probability $\frac{2}{3}$ and fair condition with probability $\frac{1}{3}$. What proportion of time is the machine in each state?

Solution: Letting the states be 1, 2, 3, we have by Eq. (7.23) that the π_i satisfy

$$\begin{aligned}\pi_1 + \pi_2 + \pi_3 &= 1,\\ \pi_1 &= \frac{2}{3}\pi_3,\\ \pi_2 &= \frac{3}{4}\pi_1 + \frac{1}{3}\pi_3,\\ \pi_3 &= \frac{1}{4}\pi_1 + \pi_2\end{aligned}$$

The solution is

$$\pi_1 = \frac{4}{15}, \quad \pi_2 = \frac{1}{3}, \quad \pi_3 = \frac{2}{5}$$

Hence, from Eq. (7.24) we obtain that P_i, the proportion of time the machine is in state i, is given by

$$P_1 = \frac{4\mu_1}{4\mu_1 + 5\mu_2 + 6\mu_3},$$

$$P_2 = \frac{5\mu_2}{4\mu_1 + 5\mu_2 + 6\mu_3},$$

$$P_3 = \frac{6\mu_3}{4\mu_1 + 5\mu_2 + 6\mu_3}$$

For instance, if $\mu_1 = 5, \mu_2 = 2, \mu_3 = 1$, then the machine will be in good condition $\frac{5}{9}$ of the time, in fair condition $\frac{5}{18}$ of the time, in broken condition $\frac{1}{6}$ of the time. ■

Remark. When the distributions of the amount of time spent in each state during a visit are continuous, then P_i also represents the limiting (as $t \to \infty$) probability that the process will be in state i at time t.

Example 7.31. Consider a renewal process in which the interarrival distribution is discrete and is such that

$$P\{X = i\} = p_i, \quad i \geq 1$$

where X represents an interarrival random variable. Let $L(t)$ denote the length of the renewal interval that contains the point t (that is, if $N(t)$ is the number of renewals by time t and X_n the nth interarrival time, then $L(t) = X_{N(t)+1}$). If we think of each renewal as corresponding to the failure of a lightbulb (which is then replaced at the beginning of the next period by a new bulb), then $L(t)$ will equal i if the bulb in use at time t dies in its ith period of use.

It is easy to see that $L(t)$ is a semi-Markov process. To determine the proportion of time that $L(t) = j$, note that each time a transition occurs—that is, each time a renewal occurs—the next state will be j with probability p_j. That is, the transition probabilities of the embedded Markov chain are $P_{ij} = p_j$. Hence, the limiting probabilities of this embedded chain are given by

$$\pi_j = p_j$$

and, since the mean time the semi-Markov process spends in state j before a transition occurs is j, it follows that the long-run proportion of time the state is j is

$$P_j = \frac{jp_j}{\sum_i ip_i}$$ ■

7.7 The Inspection Paradox

Suppose that a piece of equipment, say, a battery, is installed and serves until it breaks down. Upon failure it is instantly replaced by a like battery, and this process continues without interruption. Letting $N(t)$ denote the number of batteries that have failed by time t, we have that $\{N(t), t \geq 0\}$ is a renewal process.

Suppose further that the distribution F of the lifetime of a battery is not known and is to be estimated by the following sampling inspection scheme. We fix some time t and observe the total lifetime of the battery that is in use at time t. Since F is the distribution of the lifetime for all batteries, it seems reasonable that it should be the distribution for this battery. However, this is the *inspection paradox* for it turns out that the *battery in use at time t tends to have a larger lifetime than an ordinary battery*.

To understand the preceding so-called paradox, we reason as follows. In renewal theoretic terms what we are interested in is the length of the renewal interval containing the point t. That is, we are interested in $X_{N(t)+1} = S_{N(t)+1} - S_{N(t)}$ (see Fig. 7.2). To calculate the distribution of $X_{N(t)+1}$ we condition on the time of the last renewal prior to (or at) time t. That is,

$$P\{X_{N(t)+1} > x\} = E[P\{X_{N(t)+1} > x | S_{N(t)} = t - s\}]$$

where we recall (Fig. 7.2) that $S_{N(t)}$ is the time of the last renewal prior to (or at) t. Since there are no renewals between $t - s$ and t, it follows that $X_{N(t)+1}$ must be larger than x if $s > x$. That is,

$$P\{X_{N(t)+1} > x | S_{N(t)} = t - s\} = 1 \quad \text{if } s > x \tag{7.25}$$

On the other hand, suppose that $s \leq x$. As before, we know that a renewal occurred at time $t - s$ and no additional renewals occurred between $t - s$ and t, and we ask for the probability that no renewals occur for an additional time $x - s$. That is, we are asking for the probability that an interarrival time will be greater than x given that it is greater

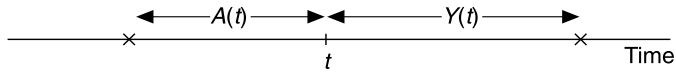

Figure 7.4

than s. Therefore, for $s \leq x$,

$$\begin{aligned} &P\{X_{N(t)+1} > x | S_{N(t)} = t - s\} \\ &\quad = P\{\text{interarrival time} > x | \text{interarrival time} > s\} \\ &\quad = P\{\text{interarrival time} > x\}/P\{\text{interarrival time} > s\} \\ &\quad = \frac{1 - F(x)}{1 - F(s)} \\ &\quad \geq 1 - F(x) \end{aligned} \tag{7.26}$$

Hence, from Eqs. (7.25) and (7.26) we see that, for all s,

$$P\{X_{N(t)+1} > x | S_{N(t)} = t - s\} \geq 1 - F(x)$$

Taking expectations on both sides yields

$$P\{X_{N(t)+1} > x\} \geq 1 - F(x) \tag{7.27}$$

However, $1 - F(x)$ is the probability that an ordinary renewal interval is larger than x, that is, $1 - F(x) = P\{X_n > x\}$, and thus Eq. (7.27) is a statement of the inspection paradox that the length of the renewal interval containing the point t tends to be larger than an ordinary renewal interval.

Remark. To obtain an intuitive feel for the so-called inspection paradox, reason as follows. We think of the whole line being covered by renewal intervals, one of which covers the point t. Is it not more likely that a larger interval, as opposed to a shorter interval, covers the point t?

We can explicitly calculate the distribution of $X_{N(t)+1}$ when the renewal process is a Poisson process. (Note that, in the general case, we did not need to calculate explicitly $P\{X_{N(t)+1} > x\}$ to show that it was at least as large as $1 - F(x)$.) To do so we write

$$X_{N(t)+1} = A(t) + Y(t)$$

where $A(t)$ denotes the time from t since the last renewal, and $Y(t)$ denotes the time from t until the next renewal (see Fig. 7.4). $A(t)$ is the *age* of the process at time t (in our example it would be the age at time t of the battery in use at time t), and $Y(t)$ is the *excess life* of the process at time t (it is the additional time from t until the battery fails). Of course, it is true that $A(t) = t - S_{N(t)}$, and $Y(t) = S_{N(t)+1} - t$.

To calculate the distribution of $X_{N(t)+1}$ we first note the important fact that, for a Poisson process, $A(t)$ and $Y(t)$ are independent. This follows since by the memory-

less property of the Poisson process, the time from t until the next renewal will be exponentially distributed and will be independent of all that has previously occurred (including, in particular, $A(t)$). In fact, this shows that if $\{N(t), t \geq 0\}$ is a Poisson process with rate λ, then

$$P\{Y(t) \leq x\} = 1 - e^{-\lambda x} \tag{7.28}$$

The distribution of $A(t)$ may be obtained as follows

$$\begin{aligned} P\{A(t) > x\} &= \begin{cases} P\{0 \text{ renewals in } [t-x, t]\}, & \text{if } x \leq t \\ 0, & \text{if } x > t \end{cases} \\ &= \begin{cases} e^{-\lambda x}, & \text{if } x \leq t \\ 0, & \text{if } x > t \end{cases} \end{aligned}$$

or, equivalently,

$$P\{A(t) \leq x\} = \begin{cases} 1 - e^{-\lambda x}, & x \leq t \\ 1, & x > t \end{cases} \tag{7.29}$$

Hence, by the independence of $Y(t)$ and $A(t)$ the distribution of $X_{N(t)+1}$ is just the convolution of the exponential distribution seen in Eq. (7.28) and the distribution of Eq. (7.29). It is interesting to note that for t large, $A(t)$ approximately has an exponential distribution. Thus, for t large, $X_{N(t)+1}$ has the distribution of the convolution of two identically distributed exponential random variables, which by Section 5.2.3 is the gamma distribution with parameters $(2, \lambda)$. In particular, for t large, the expected length of the renewal interval containing the point t is approximately *twice* the expected length of an ordinary renewal interval.

Using the results obtained in Examples 7.18 and 7.19 concerning the average values of the age and of the excess, it follows from the identity

$$X_{N(t)+1} = A(t) + Y(t)$$

that the average length of the renewal interval containing a specified point is

$$\lim_{s \to \infty} \frac{\int_0^s X_{N(t)+1}\, dt}{s} = \frac{E[X^2]}{E[X]}$$

where X has the interarrival distribution. Because, except for when X is a constant, $E[X^2] > (E[X])^2$, this average value is, as expected from the inspection paradox, greater than the expected value of an ordinary renewal interval.

We can use an alternating renewal process argument to determine the long-run proportion of time that $X_{N(t)+1}$ is greater than c. To do so, let a cycle correspond to a renewal interval, and say that the system is on at time t if the renewal interval

containing t is of length greater than c (that is, if $X_{N(t)+1} > c$), and say that the system is off at time t otherwise. In other words, the system is always on during a cycle if the cycle time exceeds c or is always off during the cycle if the cycle time is less than c. Thus, if X is the cycle time, we have

$$\text{on time in cycle} = \begin{cases} X, & \text{if } X > c \\ 0, & \text{if } X \le c \end{cases}$$

Therefore, we obtain from alternating renewal process theory that

$$\begin{aligned} \text{long-run proportion of time } X_{N(t)+1} > c &= \frac{E[\text{on time in cycle}]}{E[\text{cycle time}]} \\ &= \frac{\int_c^\infty x f(x)\,dx}{\mu} \end{aligned}$$

where f is the density function of an interarrival.

7.8 Computing the Renewal Function

The difficulty with attempting to use the identity

$$m(t) = \sum_{n=1}^{\infty} F_n(t)$$

to compute the renewal function is that the determination of $F_n(t) = P\{X_1 + \cdots + X_n \le t\}$ requires the computation of an n-dimensional integral. Following, we present an effective algorithm that requires as inputs only one-dimensional integrals.

Let Y be an exponential random variable having rate λ, and suppose that Y is independent of the renewal process $\{N(t), t \ge 0\}$. We start by determining $E[N(Y)]$, the expected number of renewals by the random time Y. To do so, we first condition on X_1, the time of the first renewal. This yields

$$E[N(Y)] = \int_0^\infty E[N(Y)|X_1 = x] f(x)\,dx \tag{7.30}$$

where f is the interarrival density. To determine $E[N(Y)|X_1{=}x]$, we now condition on whether or not Y exceeds x. Now, if $Y < x$, then as the first renewal occurs at time x, it follows that the number of renewals by time Y is equal to 0. On the other hand, if we are given that $x < Y$, then the number of renewals by time Y will equal 1 (the one at x) plus the number of additional renewals between x and Y. But by the memoryless property of exponential random variables, it follows that, given that $Y > x$, the amount by which it exceeds x is also exponential with rate λ, and so given that $Y > x$ the number of renewals between x and Y will have the same distribution as

$N(Y)$. Hence,

$$E[N(Y)|X_1 = x,\ Y < x] = 0,$$
$$E[N(Y)|X_1 = x,\ Y > x] = 1 + E[N(Y)]$$

and so,

$$\begin{aligned} E[N(Y)|X_1 = x] &= E[N(Y)|X_1 = x,\ Y < x]P\{Y < x|X_1 = x\} \\ &\quad + E[N(Y)|X_1 = x,\ Y > x]P\{Y > x|X_1 = x\} \\ &= E[N(Y)|X_1 = x,\ Y > x]P\{Y > x\} \\ &\quad \text{since } Y \text{ and } X_1 \text{ are independent} \\ &= (1 + E[N(Y)])e^{-\lambda x} \end{aligned}$$

Substituting this into Eq. (7.30) gives

$$E[N(Y)] = (1 + E[N(Y)]) \int_0^\infty e^{-\lambda x} f(x)\, dx$$

or

$$E[N(Y)] = \frac{E[e^{-\lambda X}]}{1 - E[e^{-\lambda X}]} \tag{7.31}$$

where X has the renewal interarrival distribution.

If we let $\lambda = 1/t$, then Eq. (7.31) presents an expression for the expected number of renewals (not by time t, but) by a random exponentially distributed time with mean t. However, as such a random variable need not be close to its mean (its variance is t^2), Eq. (7.31) need not be particularly close to $m(t)$. To obtain an accurate approximation suppose that $Y_1, Y_2, \ldots, Y_n$ are independent exponentials with rate λ and suppose they are also independent of the renewal process. Let, for $r = 1, \ldots, n$,

$$m_r = E[N(Y_1 + \cdots + Y_r)]$$

To compute an expression for m_r, we again start by conditioning on X_1, the time of the first renewal:

$$m_r = \int_0^\infty E[N(Y_1 + \cdots + Y_r)|X_1 = x] f(x)\, dx \tag{7.32}$$

To determine the foregoing conditional expectation, we now condition on the number of partial sums $\sum_{i=1}^{j} Y_i$, $j = 1, \ldots, r$, that are less than x. Now, if all r partial sums are less than x—that is, if $\sum_{i=1}^{r} Y_i < x$—then clearly the number of renewals by time $\sum_{i=1}^{r} Y_i$ is 0. On the other hand, given that k, $k < r$, of these partial sums are less than x, it follows from the lack of memory property of the exponential that the number of renewals by time $\sum_{i=1}^{r} Y_i$ will have the same distribution as 1 plus $N(Y_{k+1} + \cdots + Y_r)$.

Hence,

$$E\left[N(Y_1+\cdots+Y_r)\middle|X_1=x,\ k \text{ of the sums } \sum_{i=1}^{j} Y_i \text{ are less than } x\right]$$
$$=\begin{cases}0, & \text{if } k=r\\ 1+m_{r-k}, & \text{if } k<r\end{cases} \tag{7.33}$$

To determine the distribution of the number of the partial sums that are less than x, note that the successive values of these partial sums $\sum_{i=1}^{j} Y_i$, $j=1,\ldots,r$, have the same distribution as the first r event times of a Poisson process with rate λ (since each successive partial sum is the previous sum plus an independent exponential with rate λ). Hence, it follows that, for $k<r$,

$$P\left\{k \text{ of the partial sums } \sum_{i=1}^{j} Y_i \text{ are less than } x\middle|X_1=x\right\}$$
$$=\frac{e^{-\lambda x}(\lambda x)^k}{k!} \tag{7.34}$$

Upon substitution of Eqs. (7.33) and (7.34) into Eq. (7.32), we obtain

$$m_r=\int_0^\infty \sum_{k=0}^{r-1}(1+m_{r-k})\frac{e^{-\lambda x}(\lambda x)^k}{k!}f(x)\,dx$$

or, equivalently,

$$m_r=\frac{\sum_{k=1}^{r-1}(1+m_{r-k})E\left[X^k e^{-\lambda X}\right](\lambda^k/k!)+E\left[e^{-\lambda X}\right]}{1-E\left[e^{-\lambda X}\right]} \tag{7.35}$$

If we set $\lambda=n/t$, then starting with m_1 given by Eq. (7.31), we can use Eq. (7.35) to recursively compute $m_2,\ldots,m_n$. The approximation of $m(t)=E[N(t)]$ is given by $m_n=E[N(Y_1+\cdots+Y_n)]$. Since $Y_1+\cdots+Y_n$ is the sum of n independent exponential random variables each with mean t/n, it follows that it is (gamma) distributed with mean t and variance $nt^2/n^2=t^2/n$. Hence, by choosing n large, $\sum_{i=1}^{n} Y_i$ will be a random variable having most of its probability concentrated about t, and so $E\left[N\left(\sum_{i=1}^{n} Y_i\right)\right]$ should be quite close to $E[N(t)]$. (Indeed, if $m(t)$ is continuous at t, it can be shown that these approximations converge to $m(t)$ as n goes to infinity.)

Example 7.32. Table 7.1 compares the approximation with the exact value for the distributions F_i with densities $f_i, i=1,2,3$, which are given by

$$f_1(x)=xe^{-x},$$
$$1-F_2(x)=0.3e^{-x}+0.7e^{-2x},$$
$$1-F_3(x)=0.5e^{-x}+0.5e^{-5x}$$

■

Table 7.1 Approximating $m(t)$.

F_i		Exact	Approximation				
i	t	$m(t)$	$n=1$	$n=3$	$n=10$	$n=25$	$n=50$
1	1	0.2838	0.3333	0.3040	0.2903	0.2865	0.2852
1	2	0.7546	0.8000	0.7697	0.7586	0.7561	0.7553
1	5	2.250	2.273	2.253	2.250	2.250	2.250
1	10	4.75	4.762	4.751	4.750	4.750	4.750
2	0.1	0.1733	0.1681	0.1687	0.1689	0.1690	–
2	0.3	0.5111	0.4964	0.4997	0.5010	0.5014	–
2	0.5	0.8404	0.8182	0.8245	0.8273	0.8281	0.8283
2	1	1.6400	1.6087	1.6205	1.6261	1.6277	1.6283
2	3	4.7389	4.7143	4.7294	4.7350	4.7363	4.7367
2	10	15.5089	15.5000	15.5081	15.5089	15.5089	15.5089
3	0.1	0.2819	0.2692	0.2772	0.2804	0.2813	–
3	0.3	0.7638	0.7105	0.7421	0.7567	0.7609	–
3	1	2.0890	2.0000	2.0556	2.0789	2.0850	2.0870
3	3	5.4444	5.4000	5.4375	5.4437	5.4442	5.4443

7.9 Applications to Patterns

A counting process with independent interarrival times $X_1, X_2, \ldots$ is said to be a *delayed* or *general* renewal process if X_1 has a different distribution from the identically distributed random variables $X_2, X_3, \ldots$. That is, a delayed renewal process is a renewal process in which the first interarrival time has a different distribution than the others. Delayed renewal processes often arise in practice and it is important to note that all of the limiting theorems about $N(t)$, the number of events by time t, remain valid. For instance, it remains true that

$$\frac{E[N(t)]}{t} \to \frac{1}{\mu} \quad \text{and} \quad \frac{\text{Var}(N(t))}{t} \to \sigma^2/\mu^3 \quad \text{as } t \to \infty$$

where μ and σ^2 are the expected value and variance of the interarrivals X_i, $i > 1$.

7.9.1 Patterns of Discrete Random Variables

Let $X_1, X_2, \ldots$ be independent with $P\{X_i = j\} = p(j)$, $j \geq 0$, and let T denote the first time the pattern $x_1, \ldots, x_r$ occurs. If we say that a renewal occurs at time n, $n \geq r$, if $(X_{n-r+1}, \ldots, X_n) = (x_1, \ldots, x_r)$, then $N(n)$, $n \geq 1$, is a delayed renewal process, where $N(n)$ denotes the number of renewals by time n. It follows that

$$\frac{E[N(n)]}{n} \to \frac{1}{\mu} \quad \text{as } n \to \infty \tag{7.36}$$

$$\frac{\text{Var}(N(n))}{n} \to \frac{\sigma^2}{\mu^3} \quad \text{as } n \to \infty \tag{7.37}$$

where μ and σ are, respectively, the mean and standard deviation of the time between successive renewals. Whereas, in Section 3.6.4, we showed how to compute the expected value of T, we will now show how to use renewal theory results to compute both its mean and its variance.

To begin, let $I(i)$ equal 1 if there is a renewal at time i and let it be 0 otherwise, $i \geq r$. Also, let $p = \prod_{i=1}^{r} p(x_i)$. Since,

$$P\{I(i) = 1\} = P\{X_{i-r+1} = i_1, \ldots, X_i = i_r\} = p$$

it follows that $I(i), i \geq r$, are Bernoulli random variables with parameter p. Now,

$$N(n) = \sum_{i=r}^{n} I(i)$$

so

$$E[N(n)] = \sum_{i=r}^{n} E[I(i)] = (n - r + 1)p$$

Dividing by n and then letting $n \to \infty$ gives, from Eq. (7.36),

$$\mu = 1/p \tag{7.38}$$

That is, the mean time between successive occurrences of the pattern is equal to $1/p$. Also,

$$\begin{aligned}\frac{\text{Var}(N(n))}{n} &= \frac{1}{n}\sum_{i=r}^{n} \text{Var}(I(i)) + \frac{2}{n}\sum_{i=r}^{n-1}\sum_{n \geq j > i} \text{Cov}(I(i), I(j)) \\ &= \frac{n-r+1}{n}p(1-p) + \frac{2}{n}\sum_{i=r}^{n-1}\sum_{i<j\leq \min(i+r-1,n)} \text{Cov}(I(i), I(j))\end{aligned}$$

where the final equality used the fact that $I(i)$ and $I(j)$ are independent, and thus have zero covariance, when $|i - j| \geq r$. Letting $n \to \infty$, and using the fact that $\text{Cov}(I(i), I(j))$ depends on i and j only through $|j - i|$, gives

$$\frac{\text{Var}(N(n))}{n} \to p(1-p) + 2\sum_{j=1}^{r-1} \text{Cov}(I(r), I(r+j))$$

Therefore, using Eqs. (7.37) and (7.38), we see that

$$\sigma^2 = p^{-2}(1-p) + 2p^{-3}\sum_{j=1}^{r-1} \text{Cov}(I(r), I(r+j)) \tag{7.39}$$

Let us now consider the amount of "overlap" in the pattern. The overlap, equal to the number of values at the end of one pattern that can be used as the beginning part of the next pattern, is said to be of size k, $k > 0$, if

$$k = \max\{j < r : (i_{r-j+1}, \ldots, i_r) = (i_1, \ldots, i_j)\}$$

and is of size 0 if for all $k = 1, \ldots, r-1$, $(i_{r-k+1}, \ldots, i_r) \neq (i_1, \ldots, i_k)$. Thus, for instance, the pattern 0, 0, 1, 1 has overlap 0, whereas 0, 0, 1, 0, 0 has overlap 2. We consider two cases.

Case 1 (The Pattern Has Overlap 0)**.** In this case, $N(n)$, $n \geq 1$, is an ordinary renewal process and T is distributed as an interarrival time with mean μ and variance σ^2. Hence, we have the following from Eq. (7.38):

$$E[T] = \mu = \frac{1}{p} \tag{7.40}$$

Also, since two patterns cannot occur within a distance less than r of each other, it follows that $I(r)I(r+j) = 0$ when $1 \leq j \leq r-1$. Hence,

$$\text{Cov}(I(r), I(r+j)) = -E[I(r)]E[I(r+j)] = -p^2, \quad \text{if } 1 \leq j \leq r-1$$

Hence, from Eq. (7.39) we obtain

$$\text{Var}(T) = \sigma^2 = p^{-2}(1-p) - 2p^{-3}(r-1)p^2 = p^{-2} - (2r-1)p^{-1} \tag{7.41}$$

Remark. In cases of "rare" patterns, if the pattern hasn't yet occurred by some time n, then it would seem that we would have no reason to believe that the remaining time would be much less than if we were just beginning from scratch. That is, it would seem that the distribution is approximately memoryless and would thus be approximately exponentially distributed. Thus, since the variance of an exponential is equal to its mean squared, we would expect when μ is large that $\text{Var}(T) \approx E^2[T]$, and this is borne out by the preceding, which states that $\text{Var}(T) = E^2[T] - (2r-1)E[T]$.

Example 7.33. Suppose we are interested in the number of times that a fair coin needs to be flipped before the pattern h, h, t, h, t occurs. For this pattern, $r = 5$, $p = \frac{1}{32}$, and the overlap is 0. Hence, from Eqs. (7.40) and (7.41)

$$E[T] = 32, \quad \text{Var}(T) = 32^2 - 9 \times 32 = 736,$$

and

$$\text{Var}(T)/E^2[T] = 0.71875$$

On the other hand, if $p(i) = i/10$, $i = 1, 2, 3, 4$ and the pattern is 1, 2, 1, 4, 1, 3, 2 then $r = 7$, $p = 3/625{,}000$, and the overlap is 0. Thus, again from Eqs. (7.40) and (7.41), we see that in this case

$$E[T] = 208{,}333.33, \quad \text{Var}(T) = 4.34 \times 10^{10},$$
$$\text{Var}(T)/E^2[T] = 0.99994$$

■

Case 2 (The Overlap Is of Size k)**.** In this case,

$$T = T_{i_1,\ldots,i_k} + T^*$$

where $T_{i_1,\ldots,i_k}$ is the time until the pattern $i_1, \ldots, i_k$ appears and T^*, distributed as an interarrival time of the renewal process, is the additional time that it takes, starting with $i_1, \ldots, i_k$, to obtain the pattern $i_1, \ldots, i_r$. Because these random variables are independent, we have

$$E[T] = E[T_{i_1,\ldots,i_k}] + E[T^*] \tag{7.42}$$

$$\text{Var}(T) = \text{Var}(T_{i_1,\ldots,i_k}) + \text{Var}(T^*) \tag{7.43}$$

Now, from Eq. (7.38)

$$E[T^*] = \mu = p^{-1} \tag{7.44}$$

Also, since no two renewals can occur within a distance $r - k - 1$ of each other, it follows that $I(r)I(r + j) = 0$ if $1 \le j \le r - k - 1$. Therefore, from Eq. (7.39) we see that

$$\begin{aligned}\text{Var}(T^*) = \sigma^2 &= p^{-2}(1-p) + 2p^{-3}\left(\sum_{j=r-k}^{r-1} E[I(r)I(r+j)] - (r-1)p^2\right)\\ &= p^{-2} - (2r-1)p^{-1} + 2p^{-3}\sum_{j=r-k}^{r-1} E[I(r)I(r+j)]\end{aligned} \tag{7.45}$$

The quantities $E[I(r)I(r + j)]$ in Eq. (7.45) can be calculated by considering the particular pattern. To complete the calculation of the first two moments of T, we then compute the mean and variance of $T_{i_1,\ldots,i_k}$ by repeating the same method.

Example 7.34. Suppose that we want to determine the number of flips of a fair coin until the pattern h, h, t, h, h occurs. For this pattern, $r = 5$, $p = \frac{1}{32}$, and the overlap parameter is $k = 2$. Because

$$E[I(5)I(8)] = P\{h,h,t,h,h,t,h,h\} = \tfrac{1}{256},$$
$$E[I(5)I(9)] = P\{h,h,t,h,h,h,t,h,h\} = \tfrac{1}{512}$$

we see from Eqs. (7.44) and (7.45) that

$$E[T^*] = 32,$$
$$\text{Var}(T^*) = (32)^2 - 9(32) + 2(32)^3\left(\tfrac{1}{256} + \tfrac{1}{512}\right) = 1120$$

Hence, from Eqs. (7.42) and (7.43) we obtain

$$E[T] = E[T_{h,h}] + 32, \quad \text{Var}(T) = \text{Var}(T_{h,h}) + 1120$$

Now, consider the pattern h, h. It has $r = 2$, $p = \frac{1}{4}$, and overlap parameter 1. Since, for this pattern, $E[I(2)I(3)] = \frac{1}{8}$, we obtain, as in the preceding, that

$$E[T_{h,h}] = E[T_h] + 4,$$
$$\text{Var}(T_{h,h}) = \text{Var}(T_h) + 16 - 3(4) + 2\left(\frac{64}{8}\right) = \text{Var}(T_h) + 20$$

Finally, for the pattern h, which has $r = 1$, $p = \frac{1}{2}$, we see from Eqs. (7.40) and (7.41) that

$$E[T_h] = 2, \quad \text{Var}(T_h) = 2$$

Putting it all together gives

$$E[T] = 38, \quad \text{Var}(T) = 1142, \quad \text{Var}(T)/E^2[T] = 0.79086 \qquad \blacksquare$$

Example 7.35. Suppose that $P\{X_n = i\} = p_i$, and consider the pattern 0, 1, 2, 0, 1, 3, 0, 1. Then $p = p_0^3 p_1^3 p_2 p_3$, $r = 8$, and the overlap parameter is $k = 2$. Since

$$E[I(8)I(14)] = p_0^5 p_1^5 p_2^2 p_3^2,$$
$$E[I(8)I(15)] = 0$$

we see from Eqs. (7.42) and (7.44) that

$$E[T] = E[T_{0,1}] + p^{-1}$$

and from Eqs. (7.43) and (7.45) that

$$\text{Var}(T) = \text{Var}(T_{0,1}) + p^{-2} - 15p^{-1} + 2p^{-1}(p_0 p_1)^{-1}$$

Now, the r and p values of the pattern 0, 1 are $r(0, 1) = 2$, $p(0, 1) = p_0 p_1$, and this pattern has overlap 0. Hence, from Eqs. (7.40) and (7.41),

$$E[T_{0,1}] = (p_0 p_1)^{-1}, \quad \text{Var}(T_{0,1}) = (p_0 p_1)^{-2} - 3(p_0 p_1)^{-1}$$

For instance, if $p_i = 0.2$, $i = 0, 1, 2, 3$ then

$$E[T] = 25 + 5^8 = 390{,}650$$
$$\text{Var}(T) = 625 - 75 + 5^{16} + 35 \times 5^8 = 1.526 \times 10^{11}$$
$$\text{Var}(T)/E^2[T] = 0.99996 \qquad \blacksquare$$

Remark. It can be shown that T is a type of discrete random variable called *new better than used* (NBU), which loosely means that if the pattern has not yet occurred by some time n then the additional time until it occurs tends to be less than the time it would take the pattern to occur if one started all over at that point. Such a random variable is known to satisfy (see Proposition 9.6.1 of Ref. [4])

$$\text{Var}(T) \le E^2[T] - E[T] \le E^2[T] \qquad \blacksquare$$

Now, suppose that there are s patterns, $A(1), \ldots, A(s)$ and that we are interested in the mean time until one of these patterns occurs, as well as the probability mass function of the one that occurs first. Let us assume, without any loss of generality, that none of the patterns is contained in any of the others. (That is, we rule out such trivial cases as $A(1) = h, h$ and $A(2) = h, h, t$.) To determine the quantities of interest, let $T(i)$ denote the time until pattern $A(i)$ occurs, $i = 1, \ldots, s$, and let $T(i, j)$ denote the additional time, starting with the occurrence of pattern $A(i)$, until pattern $A(j)$ occurs, $i \neq j$. Start by computing the expected values of these random variables. We have already shown how to compute $E[T(i)]$, $i = 1, \ldots, s$. To compute $E[T(i, j)]$, use the same approach, taking into account any "overlap" between the latter part of $A(i)$ and the beginning part of $A(j)$. For instance, suppose $A(1) = 0, 0, 1, 2, 0, 3$, and $A(2) = 2, 0, 3, 2, 0$. Then

$$T(2) = T_{2,0,3} + T(1, 2)$$

where $T_{2,0,3}$ is the time to obtain the pattern 2, 0, 3. Hence,

$$\begin{aligned} E[T(1, 2)] &= E[T(2)] - E[T_{2,0,3}] \\ &= \left(p_2^2 p_0^2 p_3\right)^{-1} + (p_0 p_2)^{-1} - (p_2 p_0 p_3)^{-1} \end{aligned}$$

So, suppose now that all of the quantities $E[T(i)]$ and $E[T(i, j)]$ have been computed. Let

$$M = \min_i T(i)$$

and let

$$P(i) = P\{M = T(i)\}, \quad i = 1, \ldots, s$$

That is, $P(i)$ is the probability that pattern $A(i)$ is the first pattern to occur. Now, for each j we will derive an equation that $E[T(j)]$ satisfies as follows:

$$\begin{aligned} E[T(j)] &= E[M] + E[T(j) - M] \\ &= E[M] + \sum_{i:i \neq j} E[T(i, j)]P(i), \quad j = 1, \ldots, s \end{aligned} \tag{7.46}$$

where the final equality is obtained by conditioning on which pattern occurs first. But Eqs. (7.46) along with the equation

$$\sum_{i=1}^{s} P(i) = 1$$

constitute a set of $s + 1$ equations in the $s + 1$ unknowns $E[M]$, $P(i), i = 1, \ldots, s$. Solving them yields the desired quantities.

Example 7.36. Suppose that we continually flip a fair coin. With $A(1) = h, t, t, h, h$ and $A(2) = h, h, t, h, t$, we have

$$\begin{aligned} E[T(1)] &= 32 + E[T_h] = 34, \\ E[T(2)] &= 32, \\ E[T(1,2)] &= E[T(2)] - E[T_{h,h}] = 32 - (4 + E[T_h]) = 26, \\ E[T(2,1)] &= E[T(1)] - E[T_{h,t}] = 34 - 4 = 30 \end{aligned}$$

Hence, we need, solve the equations

$$\begin{aligned} 34 &= E[M] + 30P(2), \\ 32 &= E[M] + 26P(1), \\ 1 &= P(1) + P(2) \end{aligned}$$

These equations are easily solved, and yield the values

$$P(1) = P(2) = \tfrac{1}{2}, \quad E[M] = 19$$

Note that although the mean time for pattern $A(2)$ is less than that for $A(1)$, each has the same chance of occurring first. ■

Eqs. (7.46) are easy to solve when there are no overlaps in any of the patterns. In this case, for all $i \neq j$

$$E[T(i, j)] = E[T(j)]$$

so Eqs. (7.46) reduce to

$$E[T(j)] = E[M] + (1 - P(j))E[T(j)]$$

or

$$P(j) = E[M]/E[T(j)]$$

Summing the preceding over all j yields

$$E[M] = \frac{1}{\sum_{j=1}^{s} 1/E[T(j)]}, \tag{7.47}$$

$$P(j) = \frac{1/E[T(j)]}{\sum_{j=1}^{s} 1/E[T(j)]} \tag{7.48}$$

In our next example we use the preceding to reanalyze the model of Example 7.7.

Example 7.37. Suppose that each play of a game is, independently of the outcomes of previous plays, won by player i with probability p_i, $i = 1, \ldots, s$. Suppose further that there are specified numbers $n(1), \ldots, n(s)$ such that the first player i to win $n(i)$ consecutive plays is declared the winner of the match. Find the expected number of plays until there is a winner, and also the probability that the winner is i, $i = 1, \ldots, s$.

Solution: Letting $A(i)$, for $i = 1, \ldots, s$, denote the pattern of n_i consecutive values of i, this problem asks for $P(i)$, the probability that pattern $A(i)$ occurs first, and for $E[M]$. Because

$$E[T(i)] = (1/p_i)^{n(i)} + (1/p_i)^{n(i)-1} + \cdots + 1/p_i = \frac{1 - p_i^{n(i)}}{p_i^{n(i)}(1 - p_i)}$$

we obtain, from Eqs. (7.47) and (7.48), that

$$E[M] = \frac{1}{\sum_{j=1}^{s} \left[p_j^{n(j)}(1 - p_j)/\left(1 - p_j^{n(j)}\right)\right]},$$

$$P(i) = \frac{p_i^{n(i)}(1 - p_i)/\left(1 - p_i^{n(i)}\right)}{\sum_{j=1}^{s} \left[p_j^{n(j)}(1 - p_j)/\left(1 - p_j^{n(j)}\right)\right]}$$

■

7.9.2 The Expected Time to a Maximal Run of Distinct Values

Let $X_i, i \geq 1$, be independent and identically distributed random variables that are equally likely to take on any of the values $1, 2, \ldots, m$. Suppose that these random variables are observed sequentially, and let T denote the first time that a run of m consecutive values includes all the values $1, \ldots, m$. That is,

$$T = \min\{n : X_{n-m+1}, \ldots, X_n \text{ are all distinct}\}$$

To compute $E[T]$, define a renewal process by letting the first renewal occur at time T. At this point start over and, without using any of the data values up to T, let the next renewal occur the next time a run of m consecutive values are all distinct, and so on. For instance, if $m = 3$ and the data are

$$1, 3, 3, 2, 1, 2, 3, 2, 1, 3, \ldots, \tag{7.49}$$

then there are two renewals by time 10, with the renewals occurring at times 5 and 9. We call the sequence of m distinct values that constitutes a renewal a *renewal run*.

Let us now transform the renewal process into a delayed renewal reward process by supposing that a reward of 1 is earned at time $n, n \geq m$, if the values $X_{n-m+1}, \ldots, X_n$ are all distinct. That is, a reward is earned each time the previous m data values are all distinct. For instance, if $m = 3$ and the data values are as in (7.49) then unit rewards are earned at times 5, 7, 9, and 10. If we let R_i denote the reward earned at time i, then by Proposition 7.3,

$$\lim_{n\to\infty} \frac{E\left[\sum_{i=1}^{n} R_i\right]}{n} = \frac{E[R]}{E[T]} \tag{7.50}$$

where R is the reward earned between renewal epochs. Now, with A_i equal to the set of the first i data values of a renewal run, and B_i to the set of the first i values

following this renewal run, we have the following:

$$\begin{aligned} E[R] &= 1 + \sum_{i=1}^{m-1} E[\text{reward earned a time } i \text{ after a renewal}] \\ &= 1 + \sum_{i=1}^{m-1} P\{A_i = B_i\} \\ &= 1 + \sum_{i=1}^{m-1} \frac{i!}{m^i} \\ &= \sum_{i=0}^{m-1} \frac{i!}{m^i} \end{aligned} \tag{7.51}$$

Hence, since for $i \geq m$

$$E[R_i] = P\{X_{i-m+1}, \ldots, X_i \text{ are all distinct}\} = \frac{m!}{m^m}$$

it follows from Eq. (7.50) that

$$\frac{m!}{m^m} = \frac{E[R]}{E[T]}$$

Thus, from Eq. (7.51) we obtain

$$E[T] = \frac{m^m}{m!} \sum_{i=0}^{m-1} i!/m^i$$

The preceding delayed renewal reward process approach also gives us another way of computing the expected time until a specified pattern appears. We illustrate by the following example.

Example 7.38. Compute $E[T]$, the expected time until the pattern h, h, h, t, h, h, h appears, when a coin that comes up heads with probability p and tails with probability $q = 1 - p$ is continually flipped.

Solution: Define a renewal process by letting the first renewal occur when the pattern first appears, and then start over. Also, say that a reward of 1 is earned whenever the pattern appears. If R is the reward earned between renewal epochs, we have

$$\begin{aligned} E[R] &= 1 + \sum_{i=1}^{6} E[\text{reward earned } i \text{ units after a renewal}] \\ &= 1 + 0 + 0 + 0 + p^3q + p^3qp + p^3qp^2 \end{aligned}$$

Hence, since the expected reward earned at time i is $E[R_i] = p^6 q$, we obtain the following from the renewal reward theorem:

$$\frac{1 + qp^3 + qp^4 + qp^5}{E[T]} = qp^6$$

or

$$E[T] = q^{-1}p^{-6} + p^{-3} + p^{-2} + p^{-1}$$ ■

7.9.3 Increasing Runs of Continuous Random Variables

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed continuous random variables, and let T denote the first time that there is a string of r consecutive increasing values. That is,

$$T = \min\{n \geq r : X_{n-r+1} < X_{n-r+2} < \cdots < X_n\}$$

To compute $E[T]$, define a renewal process as follows. Let the first renewal occur at T. Then, using only the data values after T, say that the next renewal occurs when there is again a string of r consecutive increasing values, and continue in this fashion. For instance, if $r = 3$ and the first 15 data values are

$$12, 20, 22, 28, 43, 18, 24, 33, 60, 4, 16, 8, 12, 15, 18$$

then 3 renewals would have occurred by time 15, namely, at times 3, 8, and 14. If we let $N(n)$ denote the number of renewals by time n, then by the elementary renewal theorem

$$\frac{E[N(n)]}{n} \to \frac{1}{E[T]}$$

To compute $E[N(n)]$, define a stochastic process whose state at time k, call it S_k, is equal to the number of consecutive increasing values at time k. That is, for $1 \leq j \leq k$

$$S_k = j \quad \text{if } X_{k-j} > X_{k-j+1} < \cdots < X_{k-1} < X_k$$

where $X_0 = \infty$. Note that a renewal will occur at time k if and only if $S_k = ir$ for some $i \geq 1$. For instance, if $r = 3$ and

$$X_5 > X_6 < X_7 < X_8 < X_9 < X_{10} < X_{11}$$

then

$$S_6 = 1, \quad S_7 = 2, \quad S_8 = 3, \quad S_9 = 4, \quad S_{10} = 5, \quad S_{11} = 6$$

and renewals occur at times 8 and 11. Now, for $k > j$

$$P\{S_k = j\} = P\{X_{k-j} > X_{k-j+1} < \cdots < X_{k-1} < X_k\}$$

$$
\begin{aligned}
&= P\{X_{k-j+1} < \cdots < X_{k-1} < X_k\} \\
&\quad - P\{X_{k-j} < X_{k-j+1} < \cdots < X_{k-1} < X_k\} \\
&= \frac{1}{j!} - \frac{1}{(j+1)!} \\
&= \frac{j}{(j+1)!}
\end{aligned}
$$

where the next to last equality follows since all possible orderings of the random variables are equally likely.

From the preceding, we see that

$$
\lim_{k\to\infty} P\{\text{a renewal occurs at time } k\} = \lim_{k\to\infty} \sum_{i=1}^{\infty} P\{S_k = ir\} = \sum_{i=1}^{\infty} \frac{ir}{(ir+1)!}
$$

However,

$$
E[N(n)] = \sum_{k=1}^{n} P\{\text{a renewal occurs at time } k\}
$$

Because we can show that for any numbers $a_k, k \geq 1$, for which $\lim_{k\to\infty} a_k$ exists that

$$
\lim_{n\to\infty} \frac{\sum_{k=1}^{n} a_k}{n} = \lim_{k\to\infty} a_k
$$

we obtain from the preceding, upon using the elementary renewal theorem,

$$
E[T] = \frac{1}{\sum_{i=1}^{\infty} ir/(ir+1)!}
$$

7.10 The Insurance Ruin Problem

Suppose that claims are made to an insurance firm according to a Poisson process with rate λ, and that the successive claim amounts $Y_1, Y_2, \ldots$ are independent random variables having a common distribution function F with density $f(x)$. Suppose also that the claim amounts are independent of the claim arrival times. Thus, if we let $M(t)$ be the number of claims made by time t, then $\sum_{i=1}^{M(t)} Y_i$ is the total amount paid out in claims by time t. Supposing that the firm starts with an initial capital x and receives income at a constant rate c per unit time, we are interested in the probability that the firm's net capital ever becomes negative; that is, we are interested in

$$
R(x) = P\left\{\sum_{i=1}^{M(t)} Y_i > x + ct \text{ for some } t \geq 0\right\}
$$

If the firm's capital ever becomes negative, we say that the firm is ruined; thus $R(x)$ is the probability of ruin given that the firm begins with an initial capital x.

Let $\mu = E[Y_i]$ be the mean claim amount, and let $\rho = \lambda\mu/c$. Because claims occur at rate λ, the long-run rate at which money is paid out is $\lambda\mu$. (A formal argument uses renewal reward processes. A new cycle begins when a claim occurs; the cost for the cycle is the claim amount, and so the long-run average cost is μ, the expected cost incurred in a cycle, divided by $1/\lambda$, the mean cycle time.) Because the rate at which money is received is c, it is clear that $R(x) = 1$ when $\rho > 1$. As $R(x)$ can be shown to also equal 1 when $\rho = 1$ (think of the recurrence of the symmetric random walk), we will suppose that $\rho < 1$.

To determine $R(x)$, we start by deriving a differential equation. To begin, consider what can happen in the first h time units, where h is small. With probability $1 - \lambda h + o(h)$ there will be no claims and the firm's capital at time h will be $x + ch$; with probability $\lambda h + o(h)$ there will be exactly one claim and the firm's capital at time h will be $x + ch - Y_1$; with probability $o(h)$ there will be two or more claims. Therefore, conditioning on what happens during the first h time units yields

$$R(x) = (1 - \lambda h)R(x + ch) + \lambda h E[R(x + ch - Y_1)] + o(h)$$

Equivalently,

$$R(x + ch) - R(x) = \lambda h R(x + ch) - \lambda h E[R(x + ch - Y_1)] + o(h)$$

Dividing through by ch gives

$$\frac{R(x + ch) - R(x)}{ch} = \frac{\lambda}{c} R(x + ch) - \frac{\lambda}{c} E[R(x + ch - Y_1)] + \frac{1}{c}\frac{o(h)}{h}$$

Letting h go to 0 yields the differential equation

$$R'(x) = \frac{\lambda}{c} R(x) - \frac{\lambda}{c} E[R(x - Y_1)]$$

Because $R(u) = 1$ when $u < 0$, the preceding can be written as

$$R'(x) = \frac{\lambda}{c} R(x) - \frac{\lambda}{c} \int_0^x R(x - y) f(y)\, dy - \frac{\lambda}{c} \int_x^\infty f(y)\, dy$$

or, equivalently,

$$R'(x) = \frac{\lambda}{c} R(x) - \frac{\lambda}{c} \int_0^x R(x - y) f(y)\, dy - \frac{\lambda}{c} \bar{F}(x) \tag{7.52}$$

where $\bar{F}(x) = 1 - F(x)$.

We will now use the preceding equation to show that $R(x)$ also satisfies the equation

$$R(x) = R(0) + \frac{\lambda}{c} \int_0^x R(x - y) \bar{F}(y)\, dy - \frac{\lambda}{c} \int_0^x \bar{F}(y)\, dy, \quad x \geq 0 \tag{7.53}$$

To verify Eq. (7.53), we will show that differentiating both sides of it results in Eq. (7.52). (It can be shown that both (7.52) and (7.53) have unique solutions.) To do so, we will need the following lemma, whose proof is given at the end of this section.

Lemma 7.5. *For a function k, and a differentiable function t,*

$$\frac{d}{dx}\int_0^x t(x-y)k(y)\,dy = t(0)k(x) + \int_0^x t'(x-y)k(y)\,dy$$

Differentiating both sides of Eq. (7.53) gives, upon using the preceding lemma,

$$R'(x) = \frac{\lambda}{c}\left[R(0)\bar{F}(x) + \int_0^x R'(x-y)\bar{F}(y)\,dy - \bar{F}(x)\right] \tag{7.54}$$

Differentiation by parts $[u = \bar{F}(y), dv = R'(x-y)\,dy]$ shows that

$$\begin{aligned}\int_0^x R'(x-y)\bar{F}(y)\,dy &= -\bar{F}(y)R(x-y)|_0^x - \int_0^x R(x-y)f(y)\,dy\\ &= -\bar{F}(x)R(0) + R(x) - \int_0^x R(x-y)f(y)\,dy\end{aligned}$$

Substituting this result back in Eq. (7.54) gives Eq. (7.52). Thus, we have established Eq. (7.53).

To obtain a more usable expression for $R(x)$, consider a renewal process whose interarrival times $X_1, X_2, \ldots$ are distributed according to the equilibrium distribution of F. That is, the density function of the X_i is

$$f_e(x) = F_e'(x) = \frac{\bar{F}(x)}{\mu}$$

Let $N(t)$ denote the number of renewals by time t, and let us derive an expression for

$$q(x) = E\left[\rho^{N(x)+1}\right]$$

Conditioning on X_1 gives

$$q(x) = \int_0^\infty E\left[\rho^{N(x)+1}|X_1 = y\right]\frac{\bar{F}(y)}{\mu}dy$$

Because, given that $X_1 = y$, the number of renewals by time x is distributed as $1 + N(x-y)$ when $y \le x$, or is identically 0 when $y > x$, we see that

$$E[\rho^{N(x)+1}|X_1 = y] = \begin{cases}\rho E\left[\rho^{N(x-y)+1}\right], & \text{if } y \le x\\ \rho, & \text{if } y > x\end{cases}$$

Therefore, $q(x)$ satisfies

$$q(x) = \int_0^x \rho q(x-y)\frac{\bar{F}(y)}{\mu}\,dy + \rho\int_x^\infty \frac{\bar{F}(y)}{\mu}\,dy$$

$$= \frac{\lambda}{c}\int_0^x q(x-y)\bar{F}(y)\,dy + \frac{\lambda}{c}\left[\int_0^\infty \bar{F}(y)\,dy - \int_0^x \bar{F}(y)\,dy\right]$$
$$= \frac{\lambda}{c}\int_0^x q(x-y)\bar{F}(y)\,dy + \rho - \frac{\lambda}{c}\int_0^x \bar{F}(y)\,dy$$

Because $q(0) = \rho$, this is exactly the same equation that is satisfied by $R(x)$, namely Eq. (7.53). Therefore, because the solution to (7.53) is unique, we obtain the following.

Proposition 7.6.

$$R(x) = q(x) = E\left[\rho^{N(x)+1}\right]$$

Example 7.39. Suppose that the firm does not start with any initial capital. Then, because $N(0) = 0$, we see that the firm's probability of ruin is $R(0) = \rho$. ■

Example 7.40. If the claim distribution F is exponential with mean μ, then so is F_e. Hence, $N(x)$ is Poisson with mean x/μ, giving the result

$$R(x) = E\left[\rho^{N(x)+1}\right] = \sum_{n=0}^{\infty} \rho^{n+1} e^{-x/\mu}(x/\mu)^n/n!$$
$$= \rho\, e^{-x/\mu}\sum_{n=0}^{\infty}(\rho x/\mu)^n/n!$$
$$= \rho\, e^{-x(1-\rho)/\mu}$$ ■

To obtain some intuition about the ruin probability, let T be independent of the interarrival times X_i of the renewal process having interarrival distribution F_e, and let T have probability mass function

$$P\{T = n\} = \rho^n(1-\rho), \quad n = 0, 1, \ldots$$

Now consider $P\left\{\sum_{i=1}^{T} X_i > x\right\}$, the probability that the sum of the first T of the X_i exceeds x. Because $N(x) + 1$ is the first renewal that occurs after time x, we have

$$N(x) + 1 = \min\left\{n : \sum_{i=1}^{n} X_i > x\right\}$$

Therefore, conditioning on the number of renewals by time x gives

$$P\left\{\sum_{i=1}^{T} X_i > x\right\} = \sum_{j=0}^{\infty} P\left\{\sum_{i=1}^{T} X_i > x \,\middle|\, N(x) = j\right\} P\{N(x) = j\}$$
$$= \sum_{j=0}^{\infty} P\{T \geq j+1 | N(x) = j\} P\{N(x) = j\}$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} P\{T \geq j+1\} P\{N(x) = j\} \\
&= \sum_{j=0}^{\infty} \rho^{j+1} P\{N(x) = j\} \\
&= E\left[\rho^{N(x)+1}\right]
\end{aligned}
$$

Consequently, $P\left\{\sum_{i=1}^{T} X_i > x\right\}$ is equal to the ruin probability. Now, as noted in Example 7.39, the ruin probability of a firm starting with 0 initial capital is ρ. Suppose that the firm starts with an initial capital x, and suppose for the moment that it is allowed to remain in business even if its capital becomes negative. Because the probability that the firm's capital ever falls below its initial starting amount x is the same as the probability that its capital ever becomes negative when it starts with 0, this probability is also ρ. Thus, if we say that a low occurs whenever the firm's capital becomes lower than it has ever previously been, then the probability that a low ever occurs is ρ. Now, if a low does occur, then the probability that there will be another low is the probability that the firm's capital will ever fall below its previous low, and clearly this is also ρ. Therefore, each new low is the final one with probability $1 - \rho$. Consequently, the total number of lows that ever occur has the same distribution as T. In addition, if we let W_i be the amount by which the ith low is less than the low preceding it, it is easy to see that $W_1, W_2, \ldots$ are independent and identically distributed, and are also independent of the number of lows. Because the minimal value over all time of the firm's capital (when it is allowed to remain in business even when its capital becomes negative) is $x - \sum_{i=1}^{T} W_i$, it follows that the ruin probability of a firm that starts with an initial capital x is

$$
R(x) = P\left\{\sum_{i=1}^{T} W_i > x\right\}
$$

Because

$$
R(x) = E\left[\rho^{N(x)+1}\right] = P\left\{\sum_{i=1}^{T} X_i > x\right\}
$$

we can identify W_i with X_i. That is, we can conclude that each new low is lower than its predecessor by a random amount whose distribution is the equilibrium distribution of a claim amount.

Remark. Because the times between successive customer claims are independent exponential random variables with mean $1/\lambda$ while money is being paid to the insurance firm at a constant rate c, it follows that the amounts of money paid in to the insurance company between consecutive claims are independent exponential random variables with mean c/λ. Thus, because ruin can only occur when a claim arises, it follows that

the expression given in Proposition 7.6 for the ruin probability $R(x)$ is valid for any model in which the amounts of money paid to the insurance firm between claims are independent exponential random variables with mean c/λ and the amounts of the successive claims are independent random variables having distribution function F, with these two processes being independent.

Now imagine an insurance model in which customers buy policies at arbitrary times, each customer pays the insurance company a fixed rate c per unit time, the time until a customer makes a claim is exponential with rate λ, and each claim amount has distribution F. Consider the amount of money the insurance firm takes in between claims. Specifically, suppose a claim has just occurred and let X be the amount the insurance company takes in before another claim arises. Note that this amount increases continuously in time until a claim occurs, and suppose that at the present time the amount t has been taken in since the last claim. Let us compute the probability that a claim will be made before the amount taken in increases by an additional amount h, when h is small. To determine this probability, suppose that at the present time the firm has k customers. Because each of these k customers is paying the insurance firm at rate c, it follows that the additional amount taken in by the firm before the next claim occurs will be less than h if and only if a claim is made within the next $\frac{h}{kc}$ time units. Because each of the k customers will register a claim at an exponential rate λ, the time until one of them makes a claim is an exponential random variable with rate $k\lambda$. Calling this random variable $E_{k\lambda}$, it follows that the probability that the additional amount taken in is less than h is

$$\begin{aligned} P(\text{additional amount} < h | k \text{ customers}) &= P\left(E_{k\lambda} < \frac{h}{kc}\right) \\ &= 1 - e^{-\lambda h/c} \\ &= \frac{\lambda}{c} h + o(h) \end{aligned}$$

Thus,

$$P(X < t + h | X > t) = \frac{\lambda}{c} h + o(h)$$

showing that the failure rate function of X is identically $\frac{\lambda}{c}$. But this means that the amounts taken in between claims are exponential random variables with mean $\frac{c}{\lambda}$. Because the amounts of each claim have distribution function F, we can thus conclude that the firm's failure probability in this insurance model is exactly the same as in the previously analyzed classical model. ■

Let us now give the proof of Lemma 7.5.

Proof of Lemma 7.5. Let $G(x) = \int_0^x t(x - y)k(y)\, dy$. Then

$$G(x + h) - G(x) = G(x + h) - \int_0^x t(x + h - y)k(y)\, dy$$

$$
\begin{aligned}
&+ \int_0^x t(x+h-y)k(y)\,dy - G(x) \\
&= \int_x^{x+h} t(x+h-y)k(y)\,dy \\
&\quad + \int_0^x [t(x+h-y) - t(x-y)]k(y)\,dy
\end{aligned}
$$

Dividing through by h gives

$$
\begin{aligned}
\frac{G(x+h)-G(x)}{h} &= \frac{1}{h}\int_x^{x+h} t(x+h-y)k(y)\,dy \\
&\quad + \int_0^x \frac{t(x+h-y)-t(x-y)}{h}k(y)\,dy
\end{aligned}
$$

Letting $h \to 0$ gives the result

$$
G'(x) = t(0)\,k(x) + \int_0^x t'(x-y)\,k(y)\,dy \qquad \blacksquare
$$

Exercises

1. Is it true that
- **(a)** $N(t) < n$ if and only if $S_n > t$?
- **(b)** $N(t) \le n$ if and only if $S_n \ge t$?
- **(c)** $N(t) > n$ if and only if $S_n < t$?

2. Suppose that the interarrival distribution for a renewal process is Poisson distributed with mean μ. That is, suppose

$$
P\{X_n = k\} = e^{-\mu}\frac{\mu^k}{k!}, \quad k = 0, 1, \ldots
$$

- **(a)** Find the distribution of S_n.
- **(b)** Calculate $P\{N(t) = n\}$.

3. Let $\{N_1(t), t \ge 0\}$ and $\{N_2(t), t \ge 0\}$ be independent renewal processes. Let $N(t) = N_1(t) + N_2(t)$.
- **(a)** Are the interarrival times of $\{N(t), t \ge 0\}$ independent?
- **(b)** Are they identically distributed?
- **(c)** Is $\{N(t), t \ge 0\}$ a renewal process?

4. Let $U_1, U_2, \ldots$ be independent uniform $(0, 1)$ random variables, and define N by

$$
N = \min\{n : U_1 + U_2 + \cdots + U_n > 1\}
$$

What is $E[N]$?

***5.** Consider a renewal process $\{N(t), t \geq 0\}$ having a gamma (r, λ) interarrival distribution. That is, the interarrival density is

$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{(r-1)!}, \quad x > 0$$

(a) Show that

$$P\{N(t) \geq n\} = \sum_{i=nr}^{\infty} \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

(b) Show that

$$m(t) = \sum_{i=r}^{\infty} \left[\frac{i}{r}\right] \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

where $[i/r]$ is the largest integer less than or equal to i/r.

Hint: Use the relationship between the gamma (r, λ) distribution and the sum of r independent exponentials with rate λ to define $N(t)$ in terms of a Poisson process with rate λ.

6. Two players are playing a sequence of games, which begin when one of the players serves. Suppose that player 1 wins each game she serves with probability p_1 and wins each game her opponent serves with probability p_2. Further, suppose that the winner of a game becomes the server of the next game. Find the proportion of games that are won by player 1.

7. Mr. Smith works on a temporary basis. The mean length of each job he gets is three months. If the amount of time he spends between jobs is exponentially distributed with mean 2, then at what rate does Mr. Smith get new jobs?

***8.** A machine in use is replaced by a new machine either when it fails or when it reaches the age of T years. If the lifetimes of successive machines are independent with a common distribution F having density f, show that

(a) the long-run rate at which machines are replaced equals

$$\left[\int_0^T x f(x)\, dx + T(1 - F(T))\right]^{-1}$$

(b) the long-run rate at which machines in use fail equals

$$\frac{F(T)}{\int_0^T x f(x)\, dx + T[1 - F(T)]}$$

9. A worker sequentially works on jobs. Each time a job is completed, a new one is begun. Each job, independently, takes a random amount of time having distribution F to complete. However, independently of this, shocks occur according to a Poisson process with rate λ. Whenever a shock occurs, the worker

discontinues working on the present job and starts a new one. In the long run, at what rate are jobs completed?

10. Consider a renewal process with mean interarrival time μ. Suppose that each event of this process is independently "counted" with probability p. Let $N_C(t)$ denote the number of counted events by time $t, t > 0$.
(a) Is $N_C(t), t \geq 0$ a renewal process?
(b) What is $\lim_{t\to\infty} N_C(t)/t$?

11. Events occur according to a Poisson process with rate λ. Any event that occurs within a time d of the event that immediately preceded it is called a d-event. For instance, if $d = 1$ and events occur at times $2, 2.8, 4, 6, 6.6, \ldots$, then the events at times 2.8 and 6.6 would be d-events.
(a) At what rate do d-events occur?
(b) What proportion of all events are d-events?

12. Let $U_1, \ldots, U_n, \ldots$ be independent uniform $(0, 1)$ random variables. Let

$$N = \min\{n : U_n > .8\}$$

and let $S = \sum_{i=1}^{N} U_i$.
(a) Find $E[S]$ by conditioning on the value of U_1.
(b) Find $E[S]$ by conditioning on N.
(c) Find $E[S]$ by using Wald's equation.

13. In each game played one is equally likely to either win or lose 1. Let X be your cumulative winnings if you use the strategy that quits playing if you win the first game, and plays two more games and then quits if you lose the first game.
(a) Use Wald's equation to determine $E[X]$.
(b) Compute the probability mass function of X and use it to find $E[X]$.

14. Consider the gambler's ruin problem where on each bet the gambler either wins 1 with probability p or loses 1 with probability $1 - p$. The gambler will continue to play until his winnings are either $N - i$ or $-i$. (That is, starting with i the gambler will quit when his fortune reaches either N or 0.) Let T denote the number of bets made before the gambler stops. Use Wald's equation, along with the known probability that the gambler's final winnings are $N - i$, to find $E[T]$.

Hint: Let X_j be the gambler's winnings on bet j, $j \geq 1$. What are the possible values of $\sum_{j=1}^{T} X_j$? What is $E\left[\sum_{j=1}^{T} X_j\right]$?

15. Consider a miner trapped in a room that contains three doors. Door 1 leads him to freedom after two days of travel; door 2 returns him to his room after a four-day journey; and door 3 returns him to his room after a six-day journey. Suppose at all times he is equally likely to choose any of the three doors, and let T denote the time it takes the miner to become free.
(a) Define a sequence of independent and identically distributed random variables $X_1, X_2 \ldots$ and a stopping time N such that

$$T = \sum_{i=1}^{N} X_i$$

Note: You may have to imagine that the miner continues to randomly choose doors even after he reaches safety.

(b) Use Wald's equation to find $E[T]$.

(c) Compute $E\left[\sum_{i=1}^{N} X_i | N = n\right]$ and note that it is not equal to $E[\sum_{i=1}^{n} X_i]$.

(d) Use part (c) for a second derivation of $E[T]$.

16. A deck of 52 playing cards is shuffled and the cards are then turned face up one at a time. Let X_i equal 1 if the ith card turned over is an ace, and let it be 0 otherwise, $i = 1, \ldots, 52$. Also, let N denote the number of cards that need be turned over until all four aces appear. That is, the final ace appears on the Nth card to be turned over. Is the equation

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

valid? If not, why is Wald's equation not applicable?

17. In Example 7.6, suppose that potential customers arrive in accordance with a renewal process having interarrival distribution F. Would the number of events by time t constitute a (possibly delayed) renewal process if an event corresponds to a customer

(a) entering the bank?

(b) leaving the bank?

What if F were exponential?

***18.** Compute the renewal function when the interarrival distribution F is such that

$$1 - F(t) = pe^{-\mu_1 t} + (1 - p)e^{-\mu_2 t}$$

19. For the renewal process whose interarrival times are uniformly distributed over $(0, 1)$, determine the expected time from $t = 1$ until the next renewal.

20. For a renewal reward process consider

$$W_n = \frac{R_1 + R_2 + \cdots + R_n}{X_1 + X_2 + \cdots + X_n}$$

where W_n represents the average reward earned during the first n cycles. Show that $W_n \to E[R]/E[X]$ as $n \to \infty$.

21. Consider a single-server bank for which customers arrive in accordance with a Poisson process with rate λ. If a customer will enter the bank only if the server is free when he arrives, and if the service time of a customer has the distribution G, then what proportion of time is the server busy?

***22.** J's car buying policy is to always buy a new car, repair all breakdowns that occur during the first T time units of ownership, and then junk the car and buy a new one at the first breakdown that occurs after the car has reached age T. Suppose that the time until the first breakdown of a new car is exponential with rate λ, and that each time a car is repaired the time until the next breakdown is exponential with rate μ.

(a) At what rate does J buy new cars?
(b) Supposing that a new car costs C and that a cost r is incurred for each repair, what is J's long run average cost per unit time?

23. In a serve and rally competition involving players A and B, each rally that begins with a serve by player A is won by player A with probability p_a and is won by player B with probability $q_a = 1 - p_a$, whereas each rally that begins with a serve by player B is won by player A with probability p_b and is won by player B with probability $q_b = 1 - p_b$. The winner of the rally earns a point and becomes the server of the next rally.
(a) In the long run, what proportion of points are won by A?
(b) What proportion of points are won by A if the protocol is that the players alternate service? That is, if the service protocol is that A serves for the first point, then B for the second, then A for the third point, and so on.
(c) Give the condition under which A wins a higher percentage of points under the winner serves protocol than under the alternating service protocol.

24. Wald's equation can also be proved by using renewal reward processes. Let N be a stopping time for the sequence of independent and identically distributed random variables $X_i, i \geq 1$.
(a) Let $N_1 = N$. Argue that the sequence of random variables X_{N_1+1}, $X_{N_1+2}, \ldots$ is independent of $X_1, \ldots, X_N$ and has the same distribution as the original sequence $X_i, i \geq 1$.
Now treat $X_{N_1+1}, X_{N_1+2}, \ldots$ as a new sequence, and define a stopping time N_2 for this sequence that is defined exactly as N_1 is on the original sequence. (For instance, if $N_1 = \min\{n\colon X_n > 0\}$, then $N_2 = \min\{n\colon X_{N_1+n} > 0\}$.) Similarly, define a stopping time N_3 on the sequence $X_{N_1+N_2+1}, X_{N_1+N_2+2}, \ldots$ that is identically defined on this sequence as N_1 is on the original sequence, and so on.
(b) Is the reward process in which X_i is the reward earned during period i a renewal reward process? If so, what is the length of the successive cycles?
(c) Derive an expression for the average reward per unit time.
(d) Use the strong law of large numbers to derive a second expression for the average reward per unit time.
(e) Conclude Wald's equation.

25. Suppose in Example 7.15 that the arrival process is a Poisson process and suppose that the policy employed is to dispatch the train every t time units.
(a) Determine the average cost per unit time.
(b) Show that the minimal average cost per unit time for such a policy is approximately $c/2$ plus the average cost per unit time for the best policy of the type considered in that example.

26. Consider a train station to which customers arrive in accordance with a Poisson process having rate λ. A train is summoned whenever there are N customers waiting in the station, but it takes K units of time for the train to arrive at the station. When it arrives, it picks up all waiting customers. Assuming that the train station incurs a cost at a rate of nc per unit time whenever there are n customers present, find the long-run average cost.

27. A machine consists of two independent components, the ith of which functions for an exponential time with rate λ_i. The machine functions as long as at least one of these components function. (That is, it fails when both components have failed.) When a machine fails, a new machine having both its components working is put into use. A cost K is incurred whenever a machine failure occurs; operating costs at rate c_i per unit time are incurred whenever the machine in use has i working components, $i = 1, 2$. Find the long-run average cost per unit time.

28. In Example 7.17, what proportion of the defective items produced is discovered?

29. Consider a single-server queueing system in which customers arrive in accordance with a renewal process. Each customer brings in a random amount of work, chosen independently according to the distribution G. The server serves one customer at a time. However, the server processes work at rate i per unit time whenever there are i customers in the system. For instance, if a customer with workload 8 enters service when there are three other customers waiting in line, then if no one else arrives that customer will spend 2 units of time in service. If another customer arrives after 1 unit of time, then our customer will spend a total of 1.8 units of time in service provided no one else arrives.

Let W_i denote the amount of time customer i spends in the system. Also, define $E[W]$ by

$$E[W] = \lim_{n\to\infty} (W_1 + \cdots + W_n)/n$$

and so $E[W]$ is the average amount of time a customer spends in the system. Let N denote the number of customers that arrive in a busy period.

(a) Argue that

$$E[W] = E[W_1 + \cdots + W_N]/E[N]$$

Let L_i denote the amount of work customer i brings into the system; and so the L_i, $i \geq 1$, are independent random variables having distribution G.

(b) Argue that at any time t, the sum of the times spent in the system by all arrivals prior to t is equal to the total amount of work processed by time t.

Hint: Consider the rate at which the server processes work.

(c) Argue that

$$\sum_{i=1}^{N} W_i = \sum_{i=1}^{N} L_i$$

(d) Use Wald's equation (see Exercise 13) to conclude that

$$E[W] = \mu$$

where μ is the mean of the distribution G. That is, the average time that customers spend in the system is equal to the average work they bring to the system.

***30.** For a renewal process, let $A(t)$ be the age at time t. Prove that if $\mu < \infty$, then with probability 1

$$\frac{A(t)}{t} \to 0 \quad \text{as } t \to \infty$$

31. If $A(t)$ and $Y(t)$ are, respectively, the age and the excess at time t of a renewal process having an interarrival distribution F, calculate

$$P\{Y(t) > x | A(t) = s\}$$

32. Determine the long-run proportion of time that $X_{N(t)+1} < c$.

33. In Example 7.16, find the long-run proportion of time that the server is busy.

34. An $M/G/\infty$ queueing system is cleaned at the fixed times T, $2T$, $3T$, ... All customers in service when a cleaning begins are forced to leave early and a cost C_1 is incurred for each customer. Suppose that a cleaning takes time $T/4$, and that all customers who arrive while the system is being cleaned are lost, and a cost C_2 is incurred for each one.

(a) Find the long-run average cost per unit time.

(b) Find the long-run proportion of time the system is being cleaned.

***35.** Satellites are launched according to a Poisson process with rate λ. Each satellite will, independently, orbit the earth for a random time having distribution F. Let $X(t)$ denote the number of satellites orbiting at time t.

(a) Determine $P\{X(t) = k\}$.

Hint: Relate this to the $M/G/\infty$ queue.

(b) If at least one satellite is orbiting, then messages can be transmitted and we say that the system is functional. If the first satellite is orbited at time $t = 0$, determine the expected time that the system remains functional.

Hint: Make use of part (a) when $k = 0$.

36. Each of n skiers continually, and independently, climbs up and then skis down a particular slope. The time it takes skier i to climb up has distribution F_i, and it is independent of her time to ski down, which has distribution H_i, $i = 1, \ldots, n$. Let $N(t)$ denote the total number of times members of this group have skied down the slope by time t. Also, let $U(t)$ denote the number of skiers climbing up the hill at time t.

(a) What is $\lim_{t\to\infty} N(t)/t$?

(b) Find $\lim_{t\to\infty} E[U(t)]$.

(c) If all F_i are exponential with rate λ and all G_i are exponential with rate μ, what is $P\{U(t) = k\}$?

37. There are three machines, all of which are needed for a system to work. Machine i functions for an exponential time with rate λ_i before it fails, $i = 1, 2, 3$. When a machine fails, the system is shut down and repair begins on the failed

machine. The time to fix machine 1 is exponential with rate 5; the time to fix machine 2 is uniform on (0, 4); and the time to fix machine 3 is a gamma random variable with parameters $n = 3$ and $\lambda = 2$. Once a failed machine is repaired, it is as good as new and all machines are restarted.

(a) What proportion of time is the system working?

(b) What proportion of time is machine 1 being repaired?

(c) What proportion of time is machine 2 in a state of suspended animation (that is, neither working nor being repaired)?

38. A truck driver regularly drives round trips from A to B and then back to A. Each time he drives from A to B, he drives at a fixed speed that (in miles per hour) is uniformly distributed between 40 and 60; each time he drives from B to A, he drives at a fixed speed that is equally likely to be either 40 or 60.

(a) In the long run, what proportion of his driving time is spent going to B?

(b) In the long run, for what proportion of his driving time is he driving at a speed of 40 miles per hour?

39. A system consists of two independent machines that each function for an exponential time with rate λ. There is a single repairperson. If the repairperson is idle when a machine fails, then repair immediately begins on that machine; if the repairperson is busy when a machine fails, then that machine must wait until the other machine has been repaired. All repair times are independent with distribution function G and, once repaired, a machine is as good as new. What proportion of time is the repairperson idle?

40. Three marksmen take turns shooting at a target. Marksman 1 shoots until he misses, then marksman 2 begins shooting until he misses, then marksman 3 until he misses, and then back to marksman 1, and so on. Each time marksman i fires he hits the target, independently of the past, with probability $P_i, i = 1, 2, 3$. Determine the proportion of time, in the long run, that each marksman shoots.

41. Consider a waiting line system where customers arrive according to a renewal process, and either enter service if they find a free server or join the queue if all servers are busy. Suppose service times are independent with a distribution H. If we say that an event occurs whenever a departure leaves the system empty, would the counting process of events be a renewal process. If not, would it be a delayed renewal process. If not, when would it be a renewal process.

42. Dry and wet seasons alternate, with each dry season lasting an exponential time with rate λ and each wet season an exponential time with rate μ. The lengths of dry and wet seasons are all independent. In addition, suppose that people arrive to a service facility according to a Poisson process with rate v. Those that arrive during a dry season are allowed to enter; those that arrive during a wet season are lost. Let $N_l(t)$ denote the number of lost customers by time t.

(a) Find the proportion of time that we are in a wet season.

(b) Is $\{N_l(t), t \geq 0\}$ a (possibly delayed) renewal process?

(c) Find $\lim_{t\to\infty} \frac{N_l(t)}{t}$.

43. Individuals arrive two at a time to a 2 server queueing station, with the pairs arriving at times distributed according to a Poisson process with rate λ. A pair

will only enter the system if it finds both servers are free. In that case, one member of the pair enters service with server 1 and the other with server 2. Service times at server i are exponential with rate μ_i, $i = 1, 2$.

(a) Find the rate at which pairs enter the system.

(b) Find the proportion of time exactly one of the servers is busy.

44. Consider a renewal reward process where X_n is the nth interarrival time, and where R_n is the reward earned during the nth renewal interval.

(a) Give an interpretation of the random variable $R_{N(t)+1}$.

(b) Find the average value of $R_{N(t)+1}$. That is, find $\lim_{t\to\infty} \frac{\int_0^t R_{N(s)+1}\, ds}{t}$.

45. Each time a certain machine breaks down it is replaced by a new one of the same type. In the long run, what percentage of time is the machine in use less than one year old if the life distribution of a machine is

(a) uniformly distributed over (0, 2)?

(b) exponentially distributed with mean 1?

***46.** For an interarrival distribution F having mean μ, we defined the equilibrium distribution of F, denoted F_e, by

$$F_e(x) = \frac{1}{\mu}\int_0^x [1 - F(y)]\, dy$$

(a) Show that if F is an exponential distribution, then $F = F_e$.

(b) If for some constant c,

$$F(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$

show that F_e is the uniform distribution on $(0, c)$. That is, if interarrival times are identically equal to c, then the equilibrium distribution is the uniform distribution on the interval $(0, c)$.

(c) The city of Berkeley, California, allows for two hours parking at all nonmetered locations within one mile of the University of California. Parking officials regularly tour around, passing the same point every two hours. When an official encounters a car he or she marks it with chalk. If the same car is there on the official's return two hours later, then a parking ticket is written. If you park your car in Berkeley and return after three hours, what is the probability you will have received a ticket?

47. Consider a renewal process having interarrival distribution F such that

$$\bar{F}(x) = \tfrac{1}{2}e^{-x} + \tfrac{1}{2}e^{-x/2}, \quad x > 0$$

That is, interarrivals are equally likely to be exponential with mean 1 or exponential with mean 2.

(a) Without any calculations, guess the equilibrium distribution F_e.

(b) Verify your guess in part (a).

***48.** In Example 7.20, let π denote the proportion of passengers that wait less than x for a bus to arrive. That is, with W_i equal to the waiting time of passenger i,

if we define

$$X_i = \begin{cases} 1, & \text{if } W_i < x \\ 0, & \text{if } W_i \geq x \end{cases}$$

then $\pi = \lim_{n\to\infty} \sum_{i=1}^{n} X_i/n$.

(a) With N equal to the number of passengers that get on the bus, use renewal reward process theory to argue that

$$\pi = \frac{E[X_1 + \cdots + X_N]}{E[N]}$$

(b) With T equal to the time between successive buses, determine $E[X_1 + \cdots + X_N | T = t]$.

(c) Show that $E[X_1 + \cdots + X_N] = \lambda E[\min(T, x)]$.

(d) Show that

$$\pi = \frac{\int_0^x P(T > t)\, dt}{E[T]} = F_e(x)$$

(e) Using that $F_e(x)$ is the proportion of time that the excess of a renewal process with interarrival times distributed according to T is less than x, relate the result of (d) to the PASTA principle that "Poisson arrivals see the system as it averages over time".

49. Consider a system that can be in either state 1 or 2 or 3. Each time the system enters state i it remains there for a random amount of time having mean μ_i and then makes a transition into state j with probability P_{ij}. Suppose

$$P_{12} = 1, \quad P_{21} = P_{23} = \tfrac{1}{2}, \quad P_{31} = 1$$

(a) What proportion of transitions takes the system into state 1?

(b) If $\mu_1 = 1, \mu_2 = 2, \mu_3 = 3$, then what proportion of time does the system spend in each state?

50. Consider a semi-Markov process in which the amount of time that the process spends in each state before making a transition into a different state is exponentially distributed. What kind of process is this?

51. In a semi-Markov process, let t_{ij} denote the conditional expected time that the process spends in state i given that the next state is j.

(a) Present an equation relating μ_i to the t_{ij}.

(b) Show that the proportion of time the process is in i and will next enter j is equal to $P_i P_{ij} t_{ij}/\mu_i$.

Hint: Say that a cycle begins each time state i is entered. Imagine that you receive a reward at a rate of 1 per unit time whenever the process is in i and heading for j. What is the average reward per unit time?

52. A taxi alternates between three different locations. Whenever it reaches location i, it stops and spends a random time having mean t_i before obtaining

another passenger, $i = 1, 2, 3$. A passenger entering the cab at location i will want to go to location j with probability P_{ij}. The time to travel from i to j is a random variable with mean m_{ij}. Suppose that $t_1 = 1, t_2 = 2, t_3 = 4, P_{12} = 1, P_{23} = 1, P_{31} = \frac{2}{3} = 1 - P_{32}, m_{12} = 10, m_{23} = 20, m_{31} = 15, m_{32} = 25$. Define an appropriate semi-Markov process and determine

(a) the proportion of time the taxi is waiting at location i, and

(b) the proportion of time the taxi is on the road from i to j, $i, j = 1, 2, 3$.

***53.** Consider a renewal process having the gamma (n, λ) interarrival distribution, and let $Y(t)$ denote the time from t until the next renewal. Use the theory of semi-Markov processes to show that

$$\lim_{t\to\infty} P\{Y(t) < x\} = \frac{1}{n}\sum_{i=1}^{n} G_{i,\lambda}(x)$$

where $G_{i,\lambda}(x)$ is the gamma (i, λ) distribution function.

54. To prove Eq. (7.24), define the following notation:

$$X_i^j \equiv \text{time spent in state } i \text{ on the } j\text{th visit to this state;}$$
$$N_i(m) \equiv \text{number of visits to state } i \text{ in the first } m \text{ transitions}$$

In terms of this notation, write expressions for

(a) the amount of time during the first m transitions that the process is in state i;

(b) the proportion of time during the first m transitions that the process is in state i.

Argue that, with probability 1,

(c) $\sum_{j=1}^{N_i(m)} \frac{X_i^j}{N_i(m)} \to \mu_i \quad \text{as } m \to \infty$

(d) $N_i(m)/m \to \pi_i \quad \text{as } m \to \infty.$

(e) Combine parts (a), (b), (c), and (d) to prove Eq. (7.24).

55. In 1984 the country of Morocco in an attempt to determine the average amount of time that tourists spend in that country on a visit tried two different sampling procedures. In one, they questioned randomly chosen tourists as they were leaving the country; in the other, they questioned randomly chosen guests at hotels. (Each tourist stayed at a hotel.) The average visiting time of the 3000 tourists chosen from hotels was 17.8, whereas the average visiting time of the 12,321 tourists questioned at departure was 9.0. Can you explain this discrepancy? Does it necessarily imply a mistake?

56. In Example 7.20, show that if F is exponential with rate μ, then

$$\text{Average Number Waiting} = E[N]$$

That is, when buses arrive according to a Poisson process, the average number of people waiting at the stop, averaged over all time, is equal to the average

number of passengers waiting when a bus arrives. This may seem counterintuitive because the number of people waiting when the bus arrives is at least as large as the number waiting at any time in that cycle.

(b) Can you think of an inspection paradox type explanation for how such a result could be possible?

(c) Explain how this result follows from the PASTA principle.

57. If a coin that comes up heads with probability p is continually flipped until the pattern HTHTHTH appears, find the expected number of flips that land heads.

58. Let $X_i, i \geq 1$, be independent random variables with $p_j = P\{X = j\}$, $j \geq 1$. If $p_j = j/10$, $j = 1, 2, 3, 4$, find the expected time and the variance of the number of variables that need be observed until the pattern 1, 2, 3, 1, 2 occurs.

59. A coin that comes up heads with probability 0.6 is continually flipped. Find the expected number of flips until either the sequence $thht$ or the sequence ttt occurs, and find the probability that ttt occurs first.

60. Random digits, each of which is equally likely to be any of the digits 0 through 9, are observed in sequence.

(a) Find the expected time until a run of 10 distinct values occurs.

(b) Find the expected time until a run of 5 distinct values occurs.

61. Let $h(x) = P\{\sum_{i=1}^{T} X_i > x\}$ where $X_1, X_2, \ldots$ are independent random variables having distribution function F_e and T is independent of the X_i and has probability mass function $P\{T = n\} = \rho^n(1 - \rho), n \geq 0$. Show that $h(x)$ satisfies Eq. (7.53).

Hint: Start by conditioning on whether $T = 0$ or $T > 0$.

References

The results in Section 7.9.1 concerning the computation of the variance of the time until a specified pattern appears are new, as are the results of Section 7.9.2. The results of Section 7.9.3 are from Ref. [3].

[1] D.R. Cox, Renewal Theory, Methuen, London, 1962.

[2] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley, New York, 1966.

[3] F. Hwang, D. Trietsch, A Simple Relation Between the Pattern Probability and the Rate of False Signals in Control Charts, Probability in the Engineering and Informational Sciences 10 (1996) 315–323.

[4] S. Ross, Stochastic Processes, Second Edition, John Wiley, New York, 1996.

[5] H.C. Tijms, Stochastic Models, An Algorithmic Approach, John Wiley, New York, 1994.

Queueing Theory

8

8.1 Introduction

In this chapter we will study a class of models in which customers arrive in some random manner at a service facility. Upon arrival they are made to wait in queue until it is their turn to be served. Once served they are generally assumed to leave the system. For such models we will be interested in determining, among other things, such quantities as the average number of customers in the system (or in the queue) and the average time a customer spends in the system (or spends waiting in the queue).

In Section 8.2 we derive a series of basic queueing identities that are of great use in analyzing queueing models. We also introduce three different sets of limiting probabilities that correspond to what an arrival sees, what a departure sees, and what an outside observer would see.

In Section 8.3 we deal with queueing systems in which all of the defining probability distributions are assumed to be exponential. For instance, the simplest such model is to assume that customers arrive in accordance with a Poisson process (and thus the interarrival times are exponentially distributed) and are served one at a time by a single server who takes an exponentially distributed length of time for each service. These exponential queueing models are special examples of continuous-time Markov chains and so can be analyzed as in Chapter 6. However, at the cost of a (very) slight amount of repetition we shall not assume that you are familiar with the material of Chapter 6, but rather we shall redevelop any needed material. Specifically we shall derive anew (by a heuristic argument) the formula for the limiting probabilities.

In Section 8.4 we consider models in which customers move randomly among a network of servers. The model of Section 8.4.1 is an open system in which customers are allowed to enter and depart the system, whereas the one studied in Section 8.4.2 is closed in the sense that the set of customers in the system is constant over time.

In Section 8.5 we study the model $M/G/1$, which while assuming Poisson arrivals, allows the service distribution to be arbitrary. To analyze this model we first introduce in Section 8.5.1 the concept of work, and then use this concept in Section 8.5.2 to help analyze this system. In Section 8.5.3 we derive the average amount of time that a server remains busy between idle periods.

In Section 8.6 we consider some variations of the model $M/G/1$. In particular in Section 8.6.1 we suppose that bus loads of customers arrive according to a Poisson process and that each bus contains a random number of customers. In Section 8.6.2 we suppose that there are two different classes of customers—with type 1 customers receiving service priority over type 2.

In Section 8.6.3 we present an $M/G/1$ optimization example. We suppose that the server goes on break whenever she becomes idle, and then determine, under certain cost assumptions, the optimal time for her to return to service.

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00013-5

In Section 8.7 we consider a model with exponential service times but where the interarrival times between customers is allowed to have an arbitrary distribution. We analyze this model by use of an appropriately defined Markov chain. We also derive the mean length of a busy period and of an idle period for this model.

In Section 8.8 we consider a single-server system whose arrival process results from return visits of a finite number of possible sources. Assuming a general service distribution, we show how a Markov chain can be used to analyze this system.

In the final section of the chapter we talk about multiserver systems. We start with loss systems, in which arrivals finding all servers busy are assumed to depart and as such are lost to the system. This leads to the famous result known as Erlang's loss formula, which presents a simple formula for the number of busy servers in such a model when the arrival process in Poisson and the service distribution is general. We then discuss multiserver systems in which queues are allowed. However, except in the case where exponential service times are assumed, there are very few explicit formulas for these models. We end by presenting an approximation for the average time a customer waits in queue in a k-server model that assumes Poisson arrivals but allows for a general service distribution.

8.2 Preliminaries

In this section we will derive certain identities that are valid in the great majority of queueing models.

8.2.1 Cost Equations

Some fundamental quantities of interest for queueing models are

L the average number of customers in the system;
L_Q the average number of customers waiting in queue;
W the average amount of time a customer spends in the system;
W_Q the average amount of time a customer spends waiting in queue.

A large number of interesting and useful relationships between the preceding and other quantities of interest can be obtained by making use of the following idea: Imagine that entering customers are forced to pay money (according to some rule) to the system. We would then have the following basic cost identity:

$$\begin{aligned}&\text{average rate at which the system earns}\\&\quad = \lambda_a \times \text{average amount an entering customer pays}\end{aligned} \tag{8.1}$$

where λ_a is defined to be average arrival rate of entering customers. That is, if $N(t)$ denotes the number of customer arrivals by time t, then

$$\lambda_a = \lim_{t\to\infty} \frac{N(t)}{t}$$

We now present a heuristic proof of Eq. (8.1).

Heuristic Proof of Eq. (8.1). Let T be a fixed large number. In two different ways, we will compute the average amount of money the system has earned by time T. On one hand, this quantity approximately can be obtained by multiplying the average rate at which the system earns by the length of time T. On the other hand, we can approximately compute it by multiplying the average amount paid by an entering customer by the average number of customers entering by time T (this latter factor is approximately $\lambda_a T$). Hence, both sides of Eq. (8.1) when multiplied by T are approximately equal to the average amount earned by T. The result then follows by letting $T \to \infty$.[1]

By choosing appropriate cost rules, many useful formulas can be obtained as special cases of Eq. (8.1). For instance, by supposing that each customer pays \$1 per unit time while in the system, Eq. (8.1) yields the so-called Little's formula,

$$L = \lambda_a W \tag{8.2}$$

This follows since, under this cost rule, the rate at which the system earns is just the number in the system, and the amount a customer pays is just equal to its time in the system.

Similarly if we suppose that each customer pays \$1 per unit time while in queue, then Eq. (8.1) yields

$$L_Q = \lambda_a W_Q \tag{8.3}$$

By supposing the cost rule that each customer pays \$1 per unit time while in service we obtain from Eq. (8.1) that the

$$\text{average number of customers in service} = \lambda_a E[S] \tag{8.4}$$

where $E[S]$ is defined as the average amount of time a customer spends in service.

It should be emphasized that Eqs. (8.1) through (8.4) are valid for almost all queueing models regardless of the arrival process, the number of servers, or queue discipline. ■

8.2.2 Steady-State Probabilities

Let $X(t)$ denote the number of customers in the system at time t and define $P_n, n \geq 0$, by

$$P_n = \lim_{t \to \infty} P\{X(t) = n\}$$

where we assume the preceding limit exists. In other words, P_n is the limiting or long-run probability that there will be exactly n customers in the system. It is sometimes

[1] This can be made into a rigorous proof provided we assume that the queueing process is regenerative in the sense of Section 7.5. Most models, including all the ones in this chapter, satisfy this condition.

referred to as the *steady-state probability* of exactly n customers in the system. It also usually turns out that P_n equals the (long-run) proportion of time that the system contains exactly n customers. For example, if $P_0 = 0.3$, then in the long run, the system will be empty of customers for 30 percent of the time. Similarly, $P_1 = 0.2$ would imply that for 20 percent of the time the system would contain exactly one customer.[2]

Two other sets of limiting probabilities are $\{a_n, n \geq 0\}$ and $\{d_n, n \geq 0\}$, where

$$a_n = \text{proportion of customers that find } n \text{ in the system when they arrive, and}$$

$$d_n = \text{proportion of customers leaving behind } n \text{ in the system when they depart}$$

That is, P_n is the proportion of time during which there are n in the system; a_n is the proportion of arrivals that find n; and d_n is the proportion of departures that leave behind n. That these quantities need not always be equal is illustrated by the following example.

Example 8.1. Consider a queueing model in which all customers have service times equal to 1, and where the times between successive customers are always greater than 1 (for instance, the interarrival times could be uniformly distributed over (1, 2)). Hence, as every arrival finds the system empty and every departure leaves it empty, we have

$$a_0 = d_0 = 1$$

However,

$$P_0 \neq 1$$

as the system is not always empty of customers. ■

It was, however, no accident that a_n equaled d_n in the previous example. That arrivals and departures always see the same number of customers is always true as is shown in the next proposition.

Proposition 8.1. *In any system in which customers arrive and depart one at a time*

the rate at which arrivals find n = *the rate at which departures leave* n

and

$$a_n = d_n$$

Proof. An arrival will see n in the system whenever the number in the system goes from n to $n + 1$; similarly, a departure will leave behind n whenever the number in the

[2] A sufficient condition for the validity of the dual interpretation of P_n is that the queueing process be regenerative.

system goes from $n+1$ to n. Now in any interval of time T the number of transitions from n to $n+1$ must equal to within 1 the number from $n+1$ to n. (Between any two transitions from n to $n+1$, there must be one from $n+1$ to n, and conversely.) Hence, the rate of transitions from n to $n+1$ equals the rate from $n+1$ to n; or, equivalently, the rate at which arrivals find n equals the rate at which departures leave n. Now a_n, the proportion of arrivals finding n, can be expressed as

$$a_n = \frac{\text{the rate at which arrivals find } n}{\text{overall arrival rate}}$$

Similarly,

$$d_n = \frac{\text{the rate at which departures leave } n}{\text{overall departure rate}}$$

Thus, if the overall arrival rate is equal to the overall departure rate, then the preceding shows that $a_n = d_n$. On the other hand, if the overall arrival rate exceeds the overall departure rate, then the queue size will go to infinity, implying that $a_n = d_n = 0$. ■

Hence, on the average, arrivals and departures always see the same number of customers. However, as Example 8.1 illustrates, they do not, in general, see time averages. One important exception where they do is in the case of Poisson arrivals.

Proposition 8.2. *Poisson arrivals always see time averages. In particular, for Poisson arrivals,*

$$P_n = a_n$$

To understand why Poisson arrivals always see time averages, consider an arbitrary Poisson arrival. If we knew that it arrived at time t, then the conditional distribution of what it sees upon arrival is the same as the unconditional distribution of the system state at time t. For knowing that an arrival occurs at time t gives us no information about what occurred prior to t. (Since the Poisson process has independent increments, knowing that an event occurred at time t does not affect the distribution of what occurred prior to t.) Hence, an arrival would just see the system according to the limiting probabilities.

Contrast the foregoing with the situation of Example 8.1 where knowing that an arrival occurred at time t tells us a great deal about the past; in particular it tells us that there have been no arrivals in $(t-1, t)$. Thus, in this case, we cannot conclude that the distribution of what an arrival at time t observes is the same as the distribution of the system state at time t.

For a second argument as to why Poisson arrivals see time averages, note that the total time the system is in state n by time T is (roughly) $P_n T$. Hence, as Poisson arrivals always arrive at rate λ no matter what the system state, it follows that the number of arrivals in $[0, T]$ that find the system in state n is (roughly) $\lambda P_n T$. In the long run, therefore, the rate at which arrivals find the system in state n is λP_n and, as λ is the overall arrival rate, it follows that $\lambda P_n/\lambda = P_n$ is the proportion of arrivals that find the system in state n.

The result that Poisson arrivals see time averages is called the *PASTA* principle.

Example 8.2. People arrive at a bus stop according to a Poisson process with rate λ. Buses arrive at the stop according to a Poisson process with rate μ, with each arriving bus picking up all the currently waiting people. Let W_Q be the average amount of time that a person waits at the stop for a bus. Because the waiting time of each person is equal to the time from when they arrive until the next bus, which is exponentially distributed with rate μ, we see that

$$W_Q = 1/\mu$$

Using $L_Q = \lambda_a W_Q$, now shows that L_Q, the average number of people waiting at the bus stop, averaged over all time, is

$$L_Q = \lambda/\mu$$

If we let X_i be the number of people picked up by the ith bus, then with T_i equal to the time between the $(i-1)$st and the ith bus arrival,

$$E[X_i|T_i] = \lambda T_i$$

which follows because the number of people that arrive at the stop in any time interval is Poisson with a mean equal to λ times the length of the interval. Because T_i is exponential with rate μ, it follows upon taking expectations of both sides of the preceding that

$$E[X_i] = \lambda E[T_i] = \lambda/\mu$$

Thus, the average number of people picked up by a bus is equal to the time average number of people waiting for a bus, an illustration of the PASTA principle. That is, because buses arrive according to a Poisson process, it follows from PASTA that the average number of waiting people seen by arriving buses is the same as the average number of people waiting when we average over all time. ■

8.3 Exponential Models

8.3.1 A Single-Server Exponential Queueing System

Suppose that customers arrive at a single-server service station in accordance with a Poisson process having rate λ. That is, the times between successive arrivals are independent exponential random variables having mean $1/\lambda$. Each customer, upon arrival, goes directly into service if the server is free and, if not, the customer joins the queue. When the server finishes serving a customer, the customer leaves the system, and the next customer in line, if there is any, enters service. The successive service times are assumed to be independent exponential random variables having mean $1/\mu$.

The preceding is called the $M/M/1$ queue. The two Ms refer to the fact that both the interarrival and the service distributions are exponential (and thus memoryless, or Markovian), and the 1 to the fact that there is a single server. To analyze it, we shall begin by determining the limiting probabilities P_n, for $n = 0, 1, \ldots$. To do so, think along the following lines. Suppose that we have an infinite number of rooms numbered $0, 1, 2, \ldots$, and suppose that we instruct an individual to enter room n whenever there are n customers in the system. That is, he would be in room 2 whenever there are two customers in the system; and if another were to arrive, then he would leave room 2 and enter room 3. Similarly, if a service would take place he would leave room 2 and enter room 1 (as there would now be only one customer in the system).

Now suppose that in the long run our individual is seen to have entered room 1 at the rate of ten times an hour. Then at what rate must he have left room 1? Clearly, at this same rate of ten times an hour. For the total number of times that he enters room 1 must be equal to (or one greater than) the total number of times he leaves room 1. This sort of argument thus yields the general principle that will enable us to determine the state probabilities. Namely, for each $n \geq 0$, *the rate at which the process enters state n equals the rate at which it leaves state n*. Let us now determine these rates. Consider first state 0. When in state 0 the process can leave only by an arrival as clearly there cannot be a departure when the system is empty. Since the arrival rate is λ and the proportion of time that the process is in state 0 is P_0, it follows that the rate at which the process leaves state 0 is λP_0. On the other hand, state 0 can only be reached from state 1 via a departure. That is, if there is a single customer in the system and he completes service, then the system becomes empty. Since the service rate is μ and the proportion of time that the system has exactly one customer is P_1, it follows that the rate at which the process enters state 0 is μP_1.

Hence, from our rate-equality principle we get our first equation,

$$\lambda P_0 = \mu P_1$$

Now consider state 1. The process can leave this state either by an arrival (which occurs at rate λ) or a departure (which occurs at rate μ). Hence, when in state 1, the process will leave this state at a rate of $\lambda + \mu$.[3] Since the proportion of time the process is in state 1 is P_1, the rate at which the process leaves state 1 is $(\lambda + \mu)P_1$. On the other hand, state 1 can be entered either from state 0 via an arrival or from state 2 via a departure. Hence, the rate at which the process enters state 1 is $\lambda P_0 + \mu P_2$. Because the reasoning for other states is similar, we obtain the following set of equations:

$$\begin{array}{cc} \textit{State} & \textit{Rate at which the process leaves} = \textit{rate at which it enters} \\ 0 & \lambda P_0 = \mu P_1 \\ n, n \geq 1 & (\lambda + \mu) P_n = \lambda P_{n-1} + \mu P_{n+1} \end{array} \tag{8.5}$$

Eqs. (8.5), which balance the rate at which the process enters each state with the rate at which it leaves that state are known as *balance equations*.

[3] If one event occurs at a rate λ and another occurs at rate μ, then the total rate at which either event occurs is $\lambda + \mu$. Suppose one man earns \$2 per hour and another earns \$3 per hour; then together they clearly earn \$5 per hour.

In order to solve Eqs. (8.5), we rewrite them to obtain

$$P_1 = \frac{\lambda}{\mu} P_0,$$
$$P_{n+1} = \frac{\lambda}{\mu} P_n + \left(P_n - \frac{\lambda}{\mu} P_{n-1} \right), \qquad n \geq 1$$

Solving in terms of P_0 yields

$$P_0 = P_0,$$
$$P_1 = \frac{\lambda}{\mu} P_0,$$
$$P_2 = \frac{\lambda}{\mu} P_1 + \left(P_1 - \frac{\lambda}{\mu} P_0 \right) = \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu} \right)^2 P_0,$$
$$P_3 = \frac{\lambda}{\mu} P_2 + \left(P_2 - \frac{\lambda}{\mu} P_1 \right) = \frac{\lambda}{\mu} P_2 = \left(\frac{\lambda}{\mu} \right)^3 P_0,$$
$$P_4 = \frac{\lambda}{\mu} P_3 + \left(P_3 - \frac{\lambda}{\mu} P_2 \right) = \frac{\lambda}{\mu} P_3 = \left(\frac{\lambda}{\mu} \right)^4 P_0,$$
$$P_{n+1} = \frac{\lambda}{\mu} P_n + \left(P_n - \frac{\lambda}{\mu} P_{n-1} \right) = \frac{\lambda}{\mu} P_n = \left(\frac{\lambda}{\mu} \right)^{n+1} P_0$$

To determine P_0 we use the fact that the P_n must sum to 1, and thus

$$1 = \sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n P_0 = \frac{P_0}{1 - \lambda/\mu}$$

or

$$P_0 = 1 - \frac{\lambda}{\mu},$$
$$P_n = \left(\frac{\lambda}{\mu} \right)^n \left(1 - \frac{\lambda}{\mu} \right), \qquad n \geq 1 \tag{8.6}$$

Notice that for the preceding equations to make sense, it is necessary for λ/μ to be less than 1. For otherwise $\sum_{n=0}^{\infty} (\lambda/\mu)^n$ would be infinite and all the P_n would be 0. Hence, we shall assume that $\lambda/\mu < 1$. Note that it is quite intuitive that there would be no limiting probabilities if $\lambda > \mu$. For suppose that $\lambda > \mu$. Since customers arrive at a Poisson rate λ, it follows that the expected total number of arrivals by time t is λt. On the other hand, what is the expected number of customers served by time t? If there were always customers present, then the number of customers served would be a Poisson process having rate μ since the time between successive services would be independent exponentials having mean $1/\mu$. Hence, the expected number of customers served by time t is no greater than μt; and, therefore, the expected number in the

system at time t is at least

$$\lambda t - \mu t = (\lambda - \mu)t$$

Now, if $\lambda > \mu$, then the preceding number goes to infinity as t becomes large. That is, $\lambda/\mu > 1$, the queue size increases without limit and there will be no limiting probabilities. Note also that the condition $\lambda/\mu < 1$ is equivalent to the condition that the mean service time be less than the mean time between successive arrivals. This is the general condition that must be satisfied for limited probabilities to exist in most single-server queueing systems.

Remarks. **(i)** In solving the balance equations for the $M/M/1$ queue, we obtained as an intermediate step the set of equations

$$\lambda P_n = \mu P_{n+1}, \qquad n \geq 0$$

These equations could have been directly argued from the general queueing result (shown in Proposition 8.1) that the rate at which arrivals find n in the system—namely λP_n—is equal to the rate at which departures leave behind n—namely, μP_{n+1}.

(ii) We can also prove that $P_n = (\lambda/\mu)^n(1-\lambda/\mu)$ by using a queueing cost identity. Suppose that, for a fixed $n > 0$, whenever there are at least n customers in the system the nth oldest customer (with age measured from when the customer arrived) pays 1 per unit time. Letting X be the steady state number of customers in the system, because the system earns 1 per unit time whenever X is at least n, it follows that

$$\text{average rate at which the system earns} = P\{X \geq n\}$$

Also, because a customer who finds fewer than $n-1$ in the system when it arrives will pay 0, while an arrival who finds at least $n-1$ in the system will pay 1 per unit time for an exponentially distributed time with rate μ,

$$\text{average amount a customer pays} = \frac{1}{\mu} P\{X \geq n-1\}$$

Therefore, the queueing cost identity yields

$$P\{X \geq n\} = (\lambda/\mu) P\{X \geq n-1\}, \qquad n > 0$$

Iterating this gives

$$\begin{aligned} P\{X \geq n\} &= (\lambda/\mu) P\{X \geq n-1\} \\ &= (\lambda/\mu)^2 P\{X \geq n-2\} \\ &= \cdots \\ &= (\lambda/\mu)^n P\{X \geq 0\} \end{aligned}$$

$$= (\lambda/\mu)^n$$

Therefore,

$$P\{X = n\} = P\{X \geq n\} - P\{X \geq n+1\} = (\lambda/\mu)^n(1 - \lambda/\mu) \quad \blacksquare$$

Now let us attempt to express the quantities L, L_Q, W, and W_Q in terms of the limiting probabilities P_n. Since P_n is the long-run probability that the system contains exactly n customers, the average number of customers in the system clearly is given by

$$L = \sum_{n=0}^{\infty} n P_n = \sum_{n=1}^{\infty} n(\lambda/\mu)^n(1 - \lambda/\mu).$$

To compute $\sum_{n=1}^{\infty} n(\lambda/\mu)^n$, we relate it to the mean of a geometric random variable. Now, if X is geometric with parameter $1 - p$, then

$$\begin{aligned} \frac{1}{1-p} = E[X] &= \sum_{n=1}^{\infty} np^{n-1}(1-p) \\ &= \frac{1-p}{p} \sum_{n=1}^{\infty} np^n \end{aligned}$$

showing that

$$\sum_{n=1}^{\infty} np^n = \frac{p}{(1-p)^2}. \tag{8.7}$$

Consequently,

$$L = \frac{\lambda/\mu}{(1-\lambda/\mu)^2}(1 - \lambda/\mu) = \frac{\lambda}{\mu - \lambda} \tag{8.8}$$

The quantities W, W_Q, and L_Q now can be obtained with the help of Eqs. (8.2) and (8.3). That is, since $\lambda_a = \lambda$, we have from Eq. (8.8) that

$$\begin{aligned} W &= \frac{L}{\lambda} \\ &= \frac{1}{\mu - \lambda}, \\ W_Q &= W - E[S] \\ &= W - \frac{1}{\mu} \\ &= \frac{\lambda}{\mu(\mu - \lambda)}, \end{aligned}$$

$$L_Q = \lambda W_Q$$
$$= \frac{\lambda^2}{\mu(\mu - \lambda)}$$

Example 8.3. Suppose that customers arrive at a Poisson rate of one per every 12 minutes, and that the service time is exponential at a rate of one service per 8 minutes. What are L and W?

Solution: Since $\lambda = \frac{1}{12}$, $\mu = \frac{1}{8}$, we have

$$L = 2, \qquad W = 24$$

Hence, the average number of customers in the system is 2, and the average time a customer spends in the system is 24 minutes.
Now suppose that the arrival rate increases 20 percent to $\lambda = \frac{1}{10}$. What is the corresponding change in L and W? Using Eq. (8.7) and $L = \lambda W$, gives

$$L = 4, \qquad W = 40$$

Hence, an increase of 20 percent in the arrival rate *doubled* the average number of customers in the system.
To understand this better, note that

$$L = \frac{\lambda/\mu}{1 - \lambda/\mu},$$
$$W = \frac{1/\mu}{1 - \lambda/\mu}$$

From these equations we can see that when λ/μ is near 1, a slight increase in λ/μ will lead to a large increase in L and W. ■

Example 8.4. Suppose customers arrive to a two server system according to a Poisson process with rate λ, and suppose that each arrival is, independently, sent either to server 1 with probability α or to server 2 with probability $1 - \alpha$. Further, suppose that no matter which server is used, a service time is exponential with rate μ. Letting $\lambda_1 = \lambda\alpha$ and $\lambda_2 = \lambda(1 - \alpha)$, then because arrivals to server i follow a Poisson process with rate λ_i, it follows that the system as it relates to server i, $i = 1, 2$, is an $M/M/1$ system with arrival rate λ_i and service rate μ. Hence, provided that $\lambda_i < \mu$, the average time a customer sent to server i spends in the system is $W_i = \frac{1}{\mu - \lambda_i}$, $i = 1, 2$. Because the fraction of all arrivals that go to server 1 is α and the fraction that go to server 2 is $1 - \alpha$, this shows that the average time that a customer spends in the system, call it $W(\alpha)$, is

$$W(\alpha) = \alpha W_1 + (1 - \alpha) W_2$$
$$= \frac{\alpha}{\mu - \lambda\alpha} + \frac{1 - \alpha}{\mu - \lambda(1 - \alpha)}$$

Suppose now that we want to find the value of α that minimizes $W(\alpha)$. To do so, let

$$f(\alpha) = \frac{\alpha}{\mu - \lambda\alpha}$$

and note that

$$W(\alpha) = f(\alpha) + f(1-\alpha)$$

Differentiation yields that

$$f'(\alpha) = \frac{\mu - \lambda\alpha + \lambda\alpha}{(\mu - \lambda\alpha)^2} = \mu(\mu - \lambda\alpha)^{-2}$$

and

$$f''(\alpha) = 2\lambda\mu(\mu - \lambda\alpha)^{-3}$$

Because $\mu > \lambda\alpha$, we see that $f''(\alpha) > 0$. Similarly, because $\mu > \lambda(1-\alpha)$, we have that $f''(1-\alpha) > 0$. Hence,

$$W''(\alpha) = f''(\alpha) + f''(1-\alpha) > 0$$

Equating

$$W'(\alpha) = f'(\alpha) - f'(1-\alpha)$$

to 0 yields the solution $\alpha = 1 - \alpha$, or $\alpha = 1/2$. Hence, $W(\alpha)$ is minimized when $\alpha = 1/2$, with minimal value

$$\min_{0 \le \alpha \le 1} W(\alpha) = W(1/2) = \frac{1}{\mu - \lambda/2}$$ ■

A Technical Remark. We have used the fact that if one event occurs at an exponential rate λ, and another independent event at an exponential rate μ, then together they occur at an exponential rate $\lambda + \mu$. To check this formally, let T_1 be the time at which the first event occurs, and T_2 the time at which the second event occurs. Then

$$P\{T_1 \le t\} = 1 - e^{-\lambda t},$$
$$P\{T_2 \le t\} = 1 - e^{-\mu t}$$

Now if we are interested in the time until either T_1 or T_2 occurs, then we are interested in $T = \min(T_1, T_2)$. Now,

$$\begin{aligned} P\{T \le t\} &= 1 - P\{T > t\} \\ &= 1 - P\{\min(T_1, T_2) > t\} \end{aligned}$$

However, $\min(T_1, T_2) > t$ if and only if both T_1 and T_2 are greater than t; hence,

$$P\{T \le t\} = 1 - P\{T_1 > t, T_2 > t\}$$

$$
\begin{aligned}
&= 1 - P\{T_1 > t\}P\{T_2 > t\} \\
&= 1 - e^{-\lambda t}e^{-\mu t} \\
&= 1 - e^{-(\lambda+\mu)t}
\end{aligned}
$$

Thus, T has an exponential distribution with rate $\lambda + \mu$, and we are justified in adding the rates. ■

Given that an $M/M/1$ steady-state customer—that is, a customer who arrives after the system has been in operation a long time—spends a total of t time units in the system, let us determine the conditional distribution of N, the number of others that were present when that customer arrived. That is, letting W^* be the amount of time a customer spends in the system, we will find $P\{N = n|W^* = t\}$. Now,

$$
\begin{aligned}
P\{N = n|W^* = t\} &= \frac{f_{N,W^*}(n, t)}{f_{W^*}(t)} \\
&= \frac{P\{N = n\} f_{W^*|N}(t|n)}{f_{W^*}(t)}
\end{aligned}
$$

where $f_{W^*|N}(t|n)$ is the conditional density of W^* given that $N = n$, and $f_{W^*}(t)$ is the unconditional density of W^*. Now, given that $N = n$, the time that the customer spends in the system is distributed as the sum of $n+1$ independent exponential random variables with a common rate μ, implying that the conditional distribution of W^* given that $N = n$ is the gamma distribution with parameters $n + 1$ and μ. Therefore, with $C = 1/f_{W^*}(t)$,

$$
\begin{aligned}
P\{N = n|W^* = t\} &= C P\{N = n\}\mu e^{-\mu t}\frac{(\mu t)^n}{n!} \\
&= C(\lambda/\mu)^n(1 - \lambda/\mu)\mu e^{-\mu t}\frac{(\mu t)^n}{n!} \qquad \text{(by PASTA)} \\
&= K\frac{(\lambda t)^n}{n!}
\end{aligned}
$$

where $K = C(1 - \lambda/\mu)\mu e^{-\mu t}$ does not depend on n. Summing over n yields

$$
1 = \sum_{n=0}^{\infty} P\{N = n|T = t\} = K\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!} = Ke^{\lambda t}
$$

Thus, $K = e^{-\lambda t}$, showing that

$$
P\{N = n|W^* = t\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}
$$

Therefore, the conditional distribution of the number seen by an arrival who spends a total of t time units in the system is the Poisson distribution with mean λt.

In addition, as a by-product of our analysis, we have

$$
f_{W^*}(t) = 1/C
$$

$$= \frac{1}{K}(1-\lambda/\mu)\mu e^{-\mu t}$$
$$= (\mu-\lambda)e^{-(\mu-\lambda)t}$$

In other words, W^*, the amount of time a customer spends in the system, is an exponential random variable with rate $\mu-\lambda$. (As a check, we note that $E[W^*] = 1/(\mu-\lambda)$, which checks with Eq. (8.8) since $W = E[W^*]$.)

Remark. Another argument as to why W^* is exponential with rate $\mu-\lambda$ is as follows. If we let N denote the number of customers in the system as seen by an arrival, then this arrival will spend $N+1$ service times in the system before departing. Now,

$$P\{N+1=j\} = P\{N=j-1\} = (\lambda/\mu)^{j-1}(1-\lambda/\mu), \qquad j \geq 1$$

In words, the number of services that have to be completed before the arrival departs is a geometric random variable with parameter $1-\lambda/\mu$. Therefore, after each service completion our customer will be the one departing with probability $1-\lambda/\mu$. Thus, no matter how long the customer has already spent in the system, the probability he will depart in the next h time units is $\mu h + o(h)$, the probability that a service ends in that time, multiplied by $1-\lambda/\mu$. That is, the customer will depart in the next h time units with probability $(\mu-\lambda)h + o(h)$, which says that the hazard rate function of W^* is the constant $\mu-\lambda$. But only the exponential has a constant hazard rate, and so we can conclude that W^* is exponential with rate $\mu-\lambda$.

Our next example illustrates the inspection paradox.

Example 8.5. For an $M/M/1$ queue in steady state, what is the probability that the next arrival finds n in the system?

Solution: Although it might initially seem, by the PASTA principle, that this probability should just be $(\lambda/\mu)^n(1-\lambda/\mu)$, we must be careful. Because if t is the current time, then the time from t until the next arrival is exponentially distributed with rate λ, and is independent of the time from t since the last arrival, which (in the limit, as t goes to infinity) is also exponential with rate λ. Thus, although the times between successive arrivals of a Poisson process are exponential with rate λ, the time between the previous arrival before t and the first arrival after t is distributed as the sum of two independent exponentials. (This is an illustration of the inspection paradox, which results because the length of an interarrival interval that contains a specified time tends to be longer than an ordinary interarrival interval—see Section 7.7.)

Let N_a denote the number found by the next arrival, and let X be the number currently in the system. Conditioning on X yields

$$P\{N_a=n\} = \sum_{k=0}^{\infty} P\{N_a=n|X=k\}P\{X=k\}$$
$$= \sum_{k=0}^{\infty} P\{N_a=n|X=k\}(\lambda/\mu)^k(1-\lambda/\mu)$$

$$= \sum_{k=n}^{\infty} P\{N_a = n | X = k\}(\lambda/\mu)^k(1 - \lambda/\mu)$$
$$= \sum_{i=0}^{\infty} P\{N_a = n | X = n + i\}(\lambda/\mu)^{n+i}(1 - \lambda/\mu)$$

Now, for $n > 0$, given there are currently $n + i$ in the system, the next arrival will find n if we have i services before an arrival and then an arrival before the next service completion. By the lack of memory property of exponential interarrival random variables, this gives

$$P\{N_a = n | X = n + i\} = \left(\frac{\mu}{\lambda + \mu}\right)^i \frac{\lambda}{\lambda + \mu}, \qquad n > 0$$

Consequently, for $n > 0$,

$$P\{N_a = n\} = \sum_{i=0}^{\infty} \left(\frac{\mu}{\lambda + \mu}\right)^i \frac{\lambda}{\lambda + \mu} \left(\frac{\lambda}{\mu}\right)^{n+i} (1 - \lambda/\mu)$$
$$= (\lambda/\mu)^n(1 - \lambda/\mu)\frac{\lambda}{\lambda + \mu} \sum_{i=0}^{\infty} \left(\frac{\lambda}{\lambda + \mu}\right)^i$$
$$= (\lambda/\mu)^{n+1}(1 - \lambda/\mu)$$

On the other hand, the probability that the next arrival will find the system empty, when there are currently i in the system, is the probability that there are i services before the next arrival. Therefore, $P\{N_a = 0 | X = i\} = (\frac{\mu}{\lambda+\mu})^i$, giving

$$P\{N_a = 0\} = \sum_{i=0}^{\infty} \left(\frac{\mu}{\lambda + \mu}\right)^i \left(\frac{\lambda}{\mu}\right)^i (1 - \lambda/\mu)$$
$$= (1 - \lambda/\mu) \sum_{i=0}^{\infty} \left(\frac{\lambda}{\lambda + \mu}\right)^i$$
$$= (1 + \lambda/\mu)(1 - \lambda/\mu)$$

As a check, note that

$$\sum_{n=0}^{\infty} P\{N_a = n\} = (1 - \lambda/\mu)\left[1 + \lambda/\mu + \sum_{n=1}^{\infty} (\lambda/\mu)^{n+1}\right]$$
$$= (1 - \lambda/\mu) \sum_{i=0}^{\infty} (\lambda/\mu)^i$$
$$= 1$$

Note that $P\{N_a = 0\}$ is larger than $P_0 = 1 - \lambda/\mu$, showing that the next arrival is more likely to find an empty system than is an average arrival, and thus illustrating the inspection paradox that when the next customer arrives the elapsed time since the previous arrival is distributed as the sum of two independent exponentials with rate λ. Also, we might expect because of the inspection paradox that $E[N_a]$ is less than L, the average number of customers seen by an arrival. That this is indeed the case is seen from

$$E[N_a] = \sum_{n=1}^{\infty} n(\lambda/\mu)^{n+1}(1 - \lambda/\mu) = \frac{\lambda}{\mu} L < L$$ ■

8.3.2 A Single-Server Exponential Queueing System Having Finite Capacity

In the previous model, we assumed that there was no limit on the number of customers that could be in the system at the same time. However, in reality there is always a finite system capacity N, in the sense that there can be no more than N customers in the system at any time. By this, we mean that if an arriving customer finds that there are already N customers present, then he does not enter the system.

As before, we let $P_n, 0 \le n \le N$, denote the limiting probability that there are n customers in the system. The rate-equality principle yields the following set of balance equations:

State	*Rate at which the process leaves = rate at which it enters*
0	$\lambda P_0 = \mu P_1$
$1 \le n \le N - 1$	$(\lambda + \mu) P_n = \lambda P_{n-1} + \mu P_{n+1}$
N	$\mu P_N = \lambda P_{N-1}$

The argument for state 0 is exactly as before. Namely, when in state 0, the process will leave only via an arrival (which occurs at rate λ) and hence the rate at which the process leaves state 0 is λP_0. On the other hand, the process can enter state 0 only from state 1 via a departure; hence, the rate at which the process enters state 0 is μP_1. The equation for state n, where $1 \le n < N$, is the same as before. The equation for state N is different because now state N can only be left via a departure since an arriving customer will not enter the system when it is in state N; also, state N can now only be entered from state $N - 1$ (as there is no longer a state $N + 1$) via an arrival.

We could now either solve the balance equations exactly as we did for the infinite capacity model, or we could save a few lines by directly using the result that the rate at which departures leave behind $n - 1$ is equal to the rate at which arrivals find $n - 1$. Invoking this result yields

$$\mu P_n = \lambda P_{n-1}, \qquad n = 1, \ldots, N$$

giving

$$P_n = \frac{\lambda}{\mu} P_{n-1} = \left(\frac{\lambda}{\mu}\right)^2 P_{n-2} = \cdots = \left(\frac{\lambda}{\mu}\right)^n P_0, \qquad n = 1, \ldots, N$$

By using the fact that $\sum_{n=0}^{N} P_n = 1$ we obtain

$$\begin{aligned} 1 &= P_0 \sum_{n=0}^{N} \left(\frac{\lambda}{\mu}\right)^n \\ &= P_0 \left[\frac{1 - (\lambda/\mu)^{N+1}}{1 - \lambda/\mu}\right] \end{aligned}$$

or

$$P_0 = \frac{(1 - \lambda/\mu)}{1 - (\lambda/\mu)^{N+1}}$$

and hence from the preceding we obtain

$$P_n = \frac{(\lambda/\mu)^n (1 - \lambda/\mu)}{1 - (\lambda/\mu)^{N+1}}, \qquad n = 0, 1, \ldots, N$$

Note that in this case, there is no need to impose the condition that $\lambda/\mu < 1$. The queue size is, by definition, bounded so there is no possibility of its increasing indefinitely.

As before, L may be expressed in terms of P_n to yield

$$\begin{aligned} L &= \sum_{n=0}^{N} n P_n \\ &= \frac{(1 - \lambda/\mu)}{1 - (\lambda/\mu)^{N+1}} \sum_{n=0}^{N} n \left(\frac{\lambda}{\mu}\right)^n \end{aligned}$$

which after some algebra yields

$$L = \frac{\lambda[1 + N(\lambda/\mu)^{N+1} - (N+1)(\lambda/\mu)^N]}{(\mu - \lambda)(1 - (\lambda/\mu)^{N+1})}$$

In deriving W, the expected amount of time a customer spends in the system, we must be a little careful about what we mean by a customer. Specifically, are we including those "customers" who arrive to find the system full and thus do not spend any time in the system? Or, do we just want the expected time spent in the system by a customer who actually entered the system? The two questions lead, of course, to different answers. In the first case, we have $\lambda_a = \lambda$; whereas in the second case, since the fraction of arrivals that actually enter the system is $1 - P_N$, it follows that

$\lambda_a = \lambda(1 - P_N)$. Once it is clear what we mean by a customer, W can be obtained from

$$W = \frac{L}{\lambda_a}$$

Example 8.6. Suppose that it costs $c\mu$ dollars per hour to provide service at a rate μ. Suppose also that we incur a gross profit of A dollars for each customer served. If the system has a capacity N, what service rate μ maximizes our total profit?

Solution: To solve this, suppose that we use rate μ. Let us determine the amount of money coming in per hour and subtract from this the amount going out each hour. This will give us our profit per hour, and we can choose μ so as to maximize this.

Now, potential customers arrive at a rate λ. However, a certain proportion of them do not join the system—namely, those who arrive when there are N customers already in the system. Hence, since P_N is the proportion of time that the system is full, it follows that entering customers arrive at a rate of $\lambda(1 - P_N)$. Since each customer pays $\$A$, it follows that money comes in at an hourly rate of $\lambda(1 - P_N)A$ and since it goes out at an hourly rate of $c\mu$, it follows that our total profit per hour is given by

$$\begin{aligned}\text{profit per hour} &= \lambda(1 - P_N)A - c\mu \\ &= \lambda A\left[1 - \frac{(\lambda/\mu)^N(1 - \lambda/\mu)}{1 - (\lambda/\mu)^{N+1}}\right] - c\mu \\ &= \frac{\lambda A[1 - (\lambda/\mu)^N]}{1 - (\lambda/\mu)^{N+1}} - c\mu\end{aligned}$$

For instance if $N = 2, \lambda = 1, A = 10, c = 1$, then

$$\begin{aligned}\text{profit per hour} &= \frac{10[1 - (1/\mu)^2]}{1 - (1/\mu)^3} - \mu \\ &= \frac{10(\mu^3 - \mu)}{\mu^3 - 1} - \mu\end{aligned}$$

In order to maximize profit we differentiate to obtain

$$\frac{d}{d\mu}[\text{profit per hour}] = 10\frac{(2\mu^3 - 3\mu^2 + 1)}{(\mu^3 - 1)^2} - 1$$

The value of μ that maximizes our profit now can be obtained by equating to zero and solving numerically. ■

We say that a queueing system alternates between idle periods when there are no customers in the system and busy periods in which there is at least one customer in the system. We will end this section by determining the expected value and variance

of the number of lost customers in a busy period, where a customer is said to be lost if it arrives when the system is at capacity.

To determine the preceding quantities, let L_n denote the number of lost customers in a busy period of a finite capacity $M/M/1$ queue in which an arrival finding n others does not join the system. To derive an expression for $E[L_n]$ and $\text{Var}(L_n)$, suppose a busy period has just begun and condition on whether the next event is an arrival or a departure. Now, with

$$I = \begin{cases} 0, & \text{if service completion occurs before next arrival} \\ 1, & \text{if arrival before service completion} \end{cases}$$

note that if $I = 0$ then the busy period will end before the next arrival and so there will be no lost customers in that busy period. As a result

$$E[L_n|I = 0] = \text{Var}(L_n|I = 0) = 0$$

Now suppose that the next arrival appears before the end of the first service time, and so $I = 1$. Then if $n = 1$ that arrival will be lost and it will be as if the busy period were just beginning anew at that point, yielding that the conditional number of lost customers has the same distribution as does $1 + L_1$. On the other hand, if $n > 1$ then at the moment of the arrival there will be two customers in the system, the one in service and the "second customer" who has just arrived. Because the distribution of the number of lost customers in a busy period does not depend on the order in which customers are served, let us suppose that the "second customer" is put aside and does not receive any service until it is the only remaining customer. Then it is easy to see that the number of lost customers until that "second customer" begins service has the same distribution as the number of lost customers in a busy period when the system capacity is $n - 1$. Moreover, the additional number of lost customers in the busy period starting when service begins on the "second customer" has the distribution of the number of lost customers in a busy period when the system capacity is n. Consequently, given $I = 1$, L_n has the distribution of the sum of two independent random variables: one of which is distributed as L_{n-1} and represents the number of lost customers before there is again only a single customer in the system, and the other which is distributed as L_n and represents the additional number of lost customers from the moment when there is again a single customer until the busy period ends. Hence,

$$E[L_n|I = 1] = \begin{cases} 1 + E[L_1], & \text{if } n = 1 \\ E[L_{n-1}] + E[L_n], & \text{if } n > 1 \end{cases}$$

and

$$\text{Var}(L_n|I = 1) = \begin{cases} \text{Var}(L_1), & \text{if } n = 1 \\ \text{Var}(L_{n-1}) + \text{Var}(L_n), & \text{if } n > 1 \end{cases}$$

Letting

$$m_n = E[L_n] \qquad \text{and} \qquad v_n = \text{Var}(L_n)$$

then, with $m_0 = 1$, $v_0 = 0$, the preceding equations can be rewritten as

$$E[L_n|I] = I(m_{n-1} + m_n), \tag{8.9}$$

$$\text{Var}(L_n|I) = I(v_{n-1} + v_n) \tag{8.10}$$

Using that $P(I = 1) = P(\text{arrival before service}) = \frac{\lambda}{\lambda+\mu} = 1 - P(I = 0)$, we obtain upon taking expectations of both sides of Eq. (8.9) that

$$m_n = \frac{\lambda}{\lambda + \mu}[m_n + m_{n-1}]$$

or

$$m_n = \frac{\lambda}{\mu} m_{n-1}$$

Starting with $m_1 = \lambda/\mu$, this yields the result

$$m_n = (\lambda/\mu)^n$$

To determine v_n, we use the conditional variance formula. Using Eqs. (8.9) and (8.10) it gives

$$\begin{aligned} v_n &= (v_n + v_{n-1})E[I] + (m_n + m_{n-1})^2 \text{Var}(I) \\ &= \frac{\lambda}{\lambda + \mu}(v_n + v_{n-1}) + [(\lambda/\mu)^n + (\lambda/\mu)^{n-1}]^2 \frac{\lambda}{\lambda + \mu} \frac{\mu}{\lambda + \mu} \\ &= \frac{\lambda}{\lambda + \mu}(v_n + v_{n-1}) + (\lambda/\mu)^{2n-2} \left(\frac{\lambda}{\mu} + 1\right)^2 \frac{\lambda\mu}{(\lambda + \mu)^2} \\ &= \frac{\lambda}{\lambda + \mu}(v_n + v_{n-1}) + (\lambda/\mu)^{2n-1} \end{aligned}$$

Hence,

$$\mu v_n = \lambda v_{n-1} + (\lambda + \mu)(\lambda/\mu)^{2n-1}$$

or, with $\rho = \lambda/\mu$

$$v_n = \rho v_{n-1} + \rho^{2n-1} + \rho^{2n}$$

Therefore,

$$\begin{aligned} v_1 &= \rho + \rho^2, \\ v_2 &= \rho^2 + 2\rho^3 + \rho^4, \\ v_3 &= \rho^3 + 2\rho^4 + 2\rho^5 + \rho^6, \\ v_4 &= \rho^4 + 2\rho^5 + 2\rho^6 + 2\rho^7 + \rho^8 \end{aligned}$$

and, in general,

$$v_n = \rho^n + 2\sum_{j=n+1}^{2n-1} \rho^j + \rho^{2n}$$

8.3.3 Birth and Death Queueing Models

An exponential queueing system in which the arrival rates and the departure rates depend on the number of customers in the system is known as a *birth and death* queueing model. Let λ_n denote the arrival rate and let μ_n denote the departure rate when there are n customers in the system. Loosely speaking, when there are n customers in the system then the time until the next arrival is exponential with rate λ_n and is independent of the time of the next departure, which is exponential with rate μ_n. Equivalently, and more formally, whenever there are n customers in the system, the time until either the next arrival or the next departure occurs is an exponential random variable with rate $\lambda_n + \mu_n$ and, independent of how long it takes for this occurrence, it will be an arrival with probability $\frac{\lambda_n}{\lambda_n+\mu_n}$. We now give some examples of birth and death queues.

(a) The $M/M/1$ Queueing System

Because the arrival rate is always λ, and the departure rate is μ when the system is nonempty, the $M/M/1$ is a birth and death model with

$$\lambda_n = \lambda, \quad n \geq 0$$
$$\mu_n = \mu, \quad n \geq 1$$

(b) The $M/M/1$ Queueing System with Balking

Consider the $M/M/1$ system but now suppose that a customer that finds n others in the system upon its arrival will only join the system with probability α_n. (That is, with probability $1 - \alpha_n$ it balks at joining the system.) Then this system is a birth and death model with

$$\lambda_n = \lambda\alpha_n, \quad n \geq 0$$
$$\mu_n = \mu, \quad n \geq 1$$

The $M/M/1$ with finite capacity N is the special case where

$$\alpha_n = \begin{cases} 1, & \text{if } n < N \\ 0, & \text{if } n \geq N \end{cases}$$

(c) The $M/M/k$ Queueing System

Consider a k server system in which customers arrive according to a Poisson process with rate λ. An arriving customer immediately enters service if any of the k servers are free. If all k servers are busy, then the arrival joins the queue. When a server completes a service the customer served departs the system and if

there are any customers in queue then the one who has been waiting longest enters service with that server. All service times are exponential random variables with rate μ. Because customers are always arriving at rate λ,

$$\lambda_n = \lambda, \quad n \geq 0$$

Now, when there are $n \leq k$ customers in the system then each customer will be receiving service and so the time until a departure will be the minimum of n independent exponentials each having rate μ, and so will be exponential with rate $n\mu$. On the other hand if there are $n > k$ in the system then only k of the n will be in service, and so the departure rate in this case is $k\mu$. Hence, the $M/M/k$ is a birth and death queueing model with arrival rates

$$\lambda_n = \lambda, \quad n \geq 0$$

and departure rates

$$\mu_n = \begin{cases} n\mu, & \text{if } n \leq k \\ k\mu, & \text{if } n \geq k \end{cases}$$

■

To analyze the general birth and death queueing model, let P_n denote the long-run proportion of time there are n in the system. Then, either as a consequence of the balance equations given by

State	*Rate at which process leaves = rate at which process enters*
$n = 0$	$\lambda_0 P_0 = \mu_1 P_1$
$n \geq 1$	$(\lambda_n + \mu_n) P_n = \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1}$

or by directly using the result that the rate at which arrivals find n in the system is equal to the rate at which departures leave behind n, we obtain

$$\lambda_n P_n = \mu_{n+1} P_{n+1}, \qquad n \geq 0$$

or, equivalently, that

$$P_{n+1} = \frac{\lambda_n}{\mu_{n+1}} P_n, \qquad n \geq 0$$

Thus,

$$\begin{aligned} P_0 &= P_0, \\ P_1 &= \frac{\lambda_0}{\mu_1} P_0, \\ P_2 &= \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0, \\ P_3 &= \frac{\lambda_2}{\mu_3} P_2 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} P_0 \end{aligned}$$

and, in general

$$P_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}P_0, \quad n \geq 1$$

Using that $\sum_{n=0}^{\infty} P_n = 1$ shows that

$$1 = P_0\left[1 + \sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\right]$$

Hence,

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}}$$

and

$$P_n = \frac{\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}}{1 + \sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}}, \quad n \geq 1$$

The necessary and sufficient conditions for the long-run probabilities to exist is that the denominator in the preceding is finite. That is, we need have that

$$\sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} < \infty$$

Example 8.7. For the $M/M/k$ system

$$\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} = \begin{cases} \frac{(\lambda/\mu)^n}{n!}, & \text{if } n \leq k \\ \frac{\lambda^n}{\mu^n k!k^{n-k}}, & \text{if } n > k \end{cases}$$

Hence, using that $\frac{\lambda^n}{\mu^n k!k^{n-k}} = (\lambda/k\mu)^n k^k/k!$ we see that

$$P_0 = \frac{1}{1 + \sum_{n=1}^{k}(\lambda/\mu)^n/n! + \sum_{n=k+1}^{\infty}(\lambda/k\mu)^n k^k/k!},$$

$$P_n = P_0(\lambda/\mu)^n/n!, \qquad \text{if} \quad n \leq k$$

$$P_n = P_0(\lambda/k\mu)^n k^k/k!, \quad \text{if} \quad n > k$$

It follows from the preceding that the condition needed for the limiting probabilities to exist is $\lambda < k\mu$. Because $k\mu$ is the service rate when all servers are busy, the preceding is just the intuitive condition that for limiting probabilities to exist the service rate needs to be larger than the arrival rate when there are many customers in the system. ■

Example 8.8. Find the average amount of time a customer spends in the system for an $M/M/2$ system.

Solution: Letting $\mu_2 = 2\mu$, the long run proportions for the $M/M/2$ system can be expressed as

$$P_n = 2\,(\lambda/\mu_2)^n P_0, \quad n \geq 1$$

This yields that

$$\begin{aligned} 1 &= \sum_{n=0}^{\infty} P_n \\ &= P_0 \left(1 + 2\sum_{n=1}^{\infty} (\lambda/\mu_2)^n\right) \\ &= P_0 \left(1 + \frac{\lambda/\mu}{1 - \lambda/\mu_2}\right) \\ &= P_0 \left(\frac{1 + \lambda/\mu_2}{1 - \lambda/\mu_2}\right) \end{aligned}$$

Thus,

$$P_0 = \frac{1 - \lambda/\mu_2}{1 + \lambda/\mu_2}$$

To determine W, we first compute L. This gives

$$L = \sum_{n=1}^{\infty} n P_n = 2P_0 \sum_{n=1}^{\infty} n(\lambda/\mu_2)^n$$

Using the identity (8.7) yields that

$$L = 2P_0 \frac{\lambda/\mu_2}{(1 - \lambda/\mu_2)^2} = \frac{\lambda/\mu}{(1 - \lambda/\mu_2)(1 + \lambda/\mu_2)}$$

Because $L = \lambda W$, the preceding gives

$$W = \frac{1}{(\mu - \lambda/2)(1 + \lambda/\mu_2)}$$

It is interesting to contrast the average time in the system when there is a single queue as in the $M/M/2$, with when arrivals are randomly sent to be served by either server. As shown in Example 8.4, the average time in the system in the latter case is minimized when each customer is equally likely to be sent to either server, with the average time being equal to $\frac{1}{\mu - \lambda/2}$ in this case. Hence, the average time that a customer spends in the system when using a single queue as in the $M/M/2$ system is $\frac{1}{1+\lambda/\mu_2}$ multiplied by what it would be if each customer were

equally likely to be sent to either server's queue. For instance, if $\lambda = \mu = 1$, then $\lambda/\mu_2 = 1/2$, and the use of a single queue results in the customer average time in the system being equal to $2/3$ times what it would be if two separate queues were used. When $\lambda = 1.5\mu$, the reduction factor becomes $4/7$; and when $\lambda = 1.9\mu$, it is $20/39$. ■

Example 8.9 (M/M/1 Queue with Impatient Customers)**.** Consider a single-server queue where customers arrive according to a Poisson process with rate λ and where the service distribution is exponential with rate μ, but now suppose that each customer will only spend an exponential time with rate α in queue before quitting the system. Assume that the impatient times are independent of all else, and that a customer who enters service always remains until its service is completed. This system can be modeled as a birth and death process with birth and death rates

$$\lambda_n = \lambda, \quad n \geq 0$$
$$\mu_n = \mu + (n-1)\alpha, \quad n \geq 1$$

Using the previously obtained limiting probabilities enables us to answer a variety of questions about this system. For instance, suppose we wanted to determine the proportion of arrivals that receive service. Calling this quantity π_s, it can be obtained by letting λ_s be the average rate at which customers are served and noting that

$$\pi_s = \frac{\lambda_s}{\lambda}$$

To verify the preceding equation, let $N_a(t)$ and $N_s(t)$ denote, respectively, the number of arrivals and the number of services by time t. Then,

$$\pi_s = \lim_{t\to\infty} \frac{N_s(t)}{N_a(t)} = \lim_{t\to\infty} \frac{N_s(t)/t}{N_a(t)/t} = \frac{\lambda_s}{\lambda}$$

Because the service departure rate is 0 when the system is empty and is μ when the system is nonempty, it follows that $\lambda_s = \mu(1 - P_0)$, yielding that

$$\pi_s = \frac{\mu(1-P_0)}{\lambda}$$

■

Remark. As illustrated in the previous example, often the easiest way of determining the proportion of all events that are of a certain type A is to determine the rates at which events of type A occur and the rate at which all events occur, and then use that

$$\text{proportion of events that are type } A = \frac{\text{rate at which type } A \text{ events occur}}{\text{rate at which all events occur}}$$

For instance, if people arrive at rate λ and women arrive at rate λ_w, then the proportion of arrivals that are women is λ_w/λ. ■

To determine W, the average time that a customer spends in the system, for the birth and death queueing system, we employ the fundamental queueing identity $L = \lambda_a W$. Because L is the average number of customers in the system,

$$L = \sum_{n=0}^{\infty} n P_n$$

Also, because the arrival rate when there are n in the system is λ_n and the proportion of time in which there are n in the system is P_n, we see that the average arrival rate of customers is

$$\lambda_a = \sum_{n=0}^{\infty} \lambda_n P_n$$

Consequently,

$$W = \frac{\sum_{n=0}^{\infty} n P_n}{\sum_{n=0}^{\infty} \lambda_n P_n}$$

Now consider a_n equal to the proportion of arrivals that find n in the system. Since arrivals are at rate λ_n whenever there are n in system it follows that the rate at which arrivals find n is $\lambda_n P_n$. Hence, in a large time T approximately $\lambda_n P_n T$ of the approximately $\lambda_a T$ arrivals will encounter n. Letting T go to infinity shows that the long-run proportion of arrivals finding n in the system is

$$a_n = \frac{\lambda_n P_n}{\lambda_a}$$

Let us now consider the average length of a busy period, where we say that the system alternates between idle periods when there are no customers in the system and busy periods in which there is at least one customer in the system. Now, an idle period begins when the system is empty and ends when the next customer arrives. Because the arrival rate when the system is empty is λ_0, it thus follows that, independent of all that previously occurred, the length of an idle period is exponential with rate λ_0. Because a busy period always begins when there is one in the system and ends when the system is empty, it is easy to see that the lengths of successive busy periods are independent and identically distributed. Let I_j and B_j denote, respectively, the lengths of the jth idle and the jth busy period, $j \geq 1$. Now, in the first $\sum_{j=1}^{n}(I_j + B_j)$ time units the system will be empty for a time $\sum_{j=1}^{n} I_j$. Consequently, P_0, the long-run proportion of time in which the system is empty, can be expressed as

$$\begin{aligned} P_0 &= \text{long-run proportion of time empty} \\ &= \lim_{n\to\infty} \frac{I_1 + \ldots + I_n}{I_1 + \ldots + I_n + B_1 + \ldots + B_n} \\ &= \lim_{n\to\infty} \frac{(I_1 + \ldots + I_n)/n}{(I_1 + \ldots + I_n)/n + (B_1 + \ldots + B_n)/n} \end{aligned}$$

$$= \frac{E[I]}{E[I] + E[B]} \tag{8.11}$$

where I and B represent, respectively, the lengths of an idle and of a busy period, and where the final equality follows from the strong law of large numbers. Hence, using that $E[I] = 1/\lambda_0$, we see that

$$P_0 = \frac{1}{1 + \lambda_0 E[B]}$$

or,

$$E[B] = \frac{1 - P_0}{\lambda_0 P_0} \tag{8.12}$$

For instance, in the $M/M/1$ queue, this yields $E[B] = \frac{\lambda/\mu}{\lambda(1-\lambda/\mu)} = \frac{1}{\mu-\lambda}$.

Another quantity of interest is T_n, the amount of time during a busy period that there are n in the system. To determine its mean, note that $E[T_n]$ is the average amount of time there are n in the system in intervals between successive busy periods. Because the average time between successive busy periods is $E[B] + E[I]$, it follows that

$$\begin{aligned} P_n &= \text{long-run proportion of time there are } n \text{ in system} \\ &= \frac{E[T_n]}{E[I] + E[B]} \\ &= \frac{E[T_n]P_0}{E[I]} \qquad \text{from (8.11)} \end{aligned}$$

Hence,

$$E[T_n] = \frac{P_n}{\lambda_0 P_0} = \frac{\lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}$$

As a check, note that

$$B = \sum_{n=1}^{\infty} T_n$$

and thus,

$$E[B] = \sum_{n=1}^{\infty} E[T_n] = \frac{1}{\lambda_0 P_0} \sum_{n=1}^{\infty} P_n = \frac{1 - P_0}{\lambda_0 P_0}$$

which is in agreement with (8.12).

For the $M/M/1$ system, the preceding gives $E[T_n] = \lambda^{n-1}/\mu^n$.

Whereas in exponential birth and death queueing models the state of the system is just the number of customers in the system, there are other exponential models in which a more detailed state space is needed. To illustrate, we consider some examples.

8.3.4 A Shoe Shine Shop

Consider a shoe shine shop consisting of two chairs, with each chair having its own server. Suppose that an entering customer first will go to chair 1. When his work is completed in chair 1, he will go either to chair 2 if that chair is empty or else wait in chair 1 until chair 2 becomes empty. Suppose that a potential customer will enter this shop as long as chair 1 is empty. (Thus, for instance, a potential customer might enter even if there is a customer in chair 2.)

If we suppose that potential customers arrive in accordance with a Poisson process at rate λ, and that the service times for the two chairs are independent and have respective exponential rates of μ_1 and μ_2, then

(a) what proportion of potential customers enters the system?
(b) what is the mean number of customers in the system?
(c) what is the average amount of time that an entering customer spends in the system?
(d) Find π_b, equal to the fraction of entering customers that are blockers? That is, find the fraction of entering customers that will have to wait after completing service with server 1 before they can enter chair 2.

To begin we must first decide upon an appropriate state space. It is clear that the state of the system must include more information than merely the number of customers in the system. For instance, it would not be enough to specify that there is one customer in the system as we would also have to know which chair was in. Further, if we only know that there are two customers in the system, then we would not know if the person in chair 1 is still being served or that customer is just waiting for the person in chair 2 to finish. To account for these points, the following state space, consisting of the five states $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$, and $(b, 1)$, will be used. The states have the following interpretation:

State	*Interpretation*
$(0, 0)$	There are no customers in the system.
$(1, 0)$	There is one customer in the system, and that customer is in chair 1.
$(0, 1)$	There is one customer in the system, and that customer is in chair 2.
$(1, 1)$	There are two customers in the system, and both are presently being served.
$(b, 1)$	There are two customers in the system, but the customer in the first chair has completed his work in that chair and is waiting for the second chair to become free.

It should be noted that when the system is in state $(b, 1)$, the person in chair 1, though not being served, is nevertheless "blocking" potential arrivals from entering the system.

As a prelude to writing down the balance equations, it is usually worthwhile to make a transition diagram. This is done by first drawing a circle for each state and then drawing an arrow labeled by the rate at which the process goes from one state to another. The transition diagram for this model is shown in Fig. 8.1. The explanation

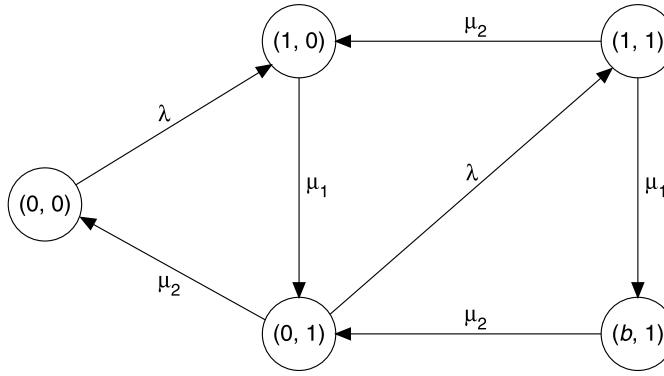

Figure 8.1 A transition diagram.

for the diagram is as follows: The arrow from state $(0, 0)$ to state $(1, 0)$ that is labeled λ means that when the process is in state $(0, 0)$, that is, when the system is empty, then it goes to state $(1, 0)$ at a rate λ, that is, via an arrival. The arrow from $(0, 1)$ to $(1, 1)$ is similarly explained.

When the process is in state $(1, 0)$, it will go to state $(0, 1)$ when the customer in chair 1 is finished and this occurs at a rate μ_1; hence the arrow from $(1, 0)$ to $(0, 1)$ labeled μ_1. The arrow from $(1, 1)$ to $(b, 1)$ is similarly explained.

When in state $(b, 1)$ the process will go to state $(0, 1)$ when the customer in chair 2 completes his service (which occurs at rate μ_2); hence the arrow from $(b, 1)$ to $(0, 1)$ labeled μ_2. Also, when in state $(1, 1)$ the process will go to state $(1, 0)$ when the man in chair 2 finishes; hence the arrow from $(1, 1)$ to $(1, 0)$ labeled μ_2. Finally, if the process is in state $(0, 1)$, then it will go to state $(0, 0)$ when the man in chair 2 completes his service; hence the arrow from $(0, 1)$ to $(0, 0)$ labeled μ_2.

Because there are no other possible transitions, this completes the transition diagram.

To write the balance equations we equate the sum of the arrows (multiplied by the probability of the states where they originate) coming into a state with the sum of the arrows (multiplied by the probability of the state) going out of that state. This gives

State	*Rate that the process leaves = rate that it enters*
$(0, 0)$	$\lambda P_{00} = \mu_2 P_{01}$
$(1, 0)$	$\mu_1 P_{10} = \lambda P_{00} + \mu_2 P_{11}$
$(0, 1)$	$(\lambda + \mu_2) P_{01} = \mu_1 P_{10} + \mu_2 P_{b1}$
$(1, 1)$	$(\mu_1 + \mu_2) P_{11} = \lambda P_{01}$
$(b, 1)$	$\mu_2 P_{b1} = \mu_1 P_{11}$

These along with the equation

$$P_{00} + P_{10} + P_{01} + P_{11} + P_{b1} = 1$$

may be solved to determine the limiting probabilities. Though it is easy to solve the preceding equations, the resulting solutions are quite involved and hence will not be explicitly presented. However, it is easy to answer our questions in terms of these limiting probabilities. To answer (a), note that a potential arrival will only enter if it finds the system in either state (0, 0) or (0, 1). Because all arrivals, including those that are lost, arrive according to a Poisson process, it follows by PASTA that the proportion of arrivals that find the system in either of those states is the proportion of time the system is in either of those states; namely, $P_{00} + P_{01}$.

To answer (b), note that since there is one customer in the system whenever the state is (0, 1) or (1, 0) and two customers in the system whenever the state is (1, 1) or $(b, 1)$, it follows that L, the average number in the system, is given by

$$L = P_{01} + P_{10} + 2(P_{11} + P_{b1})$$

To derive the average amount of time that an entering customer spends in the system, we use the relationship $W = L/\lambda_a$. Since a potential customer will enter the system when the state is either (0, 0) or (0, 1), it follows that $\lambda_a = \lambda(P_{00} + P_{01})$ and hence

$$W = \frac{P_{01} + P_{10} + 2(P_{11} + P_{b1})}{\lambda(P_{00} + P_{01})}$$

One way to determine the proportion of entering customers that are blockers is to condition on the state seen by the customer. Because the state seen by an entering customer is either (0, 0) or (0, 1), the probability that an entering customers finds the system in state (0, 1) is $P(01 \mid 00 \text{ or } 01) = \frac{P_{01}}{P_{0,0}+P_{0,1}}$. As an entering customer will be a blocker if he or she enters the system when the state is (0, 1) and then completes service at 1 before server 2 has finished its service, we see that

$$\pi_b = \frac{P_{01}}{P_{00} + P_{01}} \frac{\mu_1}{\mu_1 + \mu_2}$$

Another way to obtain the proportion of entering customers that are blockers is to let λ_b be the rate at which customers become blockers, and then use that the proportion of entering customers that are blockers is λ_b/λ_a. Because blockers originate when the state is (1, 1) and a service at 1 occurs, it follows that $\lambda_b = \mu_1 P_{11}$, and so

$$\pi_b = \frac{\mu_1 P_{11}}{\lambda(P_{00} + P_{01})}$$

That the two solutions agree follows from the balance equation for state (1, 1). ■

8.3.5 Queueing Systems with Bulk Service

Our next example refers to a system in which a server is able to simultaneously serve all waiting customers.

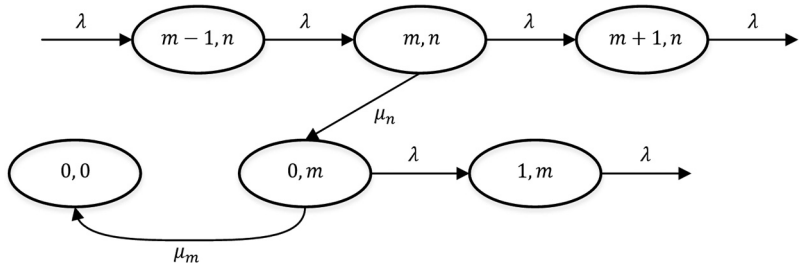

Figure 8.2

Example 8.10. Suppose that customers arrive to a single server system according to a Poisson process with rate λ, and that arrivals finding the server free immediately begin service, whereas those finding the server busy join the queue. Upon completing a service, the server then simultaneously serves all customers waiting in queue. The service time to serve a group of n customers is exponentially distributed with rate μ_n, $n \geq 1$.

To analyze this system, let the state be (m, n) if there are m customers is queue and n in service. If the state is $(0, 0)$ and a customer arrives, then that arrival will instantly begin service and the state will be $(0, 1)$. If an arrival comes when the state is (m, n), $n > 0$, then that arrival will join the queue and so the state will become $(m + 1, n)$. If a service is completed when the state is $(0, n)$ then the state becomes $(0, 0)$. If a service is completed when the state is (m, n), $m > 0$, then the m customers in queue will all enter service and so the new state will be $(0, m)$. The transition diagram for this system is given in Fig. 8.2.

Thus, we have the following balance equations equating the rates at which the system leaves and enters each state:

$$\begin{array}{cc} \textit{State} & \textit{Rate leave} = \textit{Rate enter} \\ (0,0) & \lambda P_{0,0} = \sum_{n=1}^{\infty} \mu_n P_{0,n} \\ (0,n), n > 0 & (\lambda + \mu_n) P_{0,n} = \sum_{m=1}^{\infty} \mu_m P_{n,m} \\ (m,n), mn > 0 & (\lambda + \mu_n) P_{m,n} = \lambda P_{m-1,n} \end{array}$$

$$\sum_{m,n} P_{m,n} = 1$$

In terms of the solution of these equations, determine

(a) the average amount of time that a customer spends in service;
(b) the average amount of time that a customer spends in the system;
(c) the proportion of services that are done on n customers;
(d) the proportion of customers that are served in a batch of size n.

Solution: To determine the average amount of time that a customer spends in service, we can use the fundamental identity that the

$$\begin{aligned} &\text{average number of customers in service} \\ &= \lambda_a \times \text{average time a customer spends in service} \end{aligned}$$

Because there are n customers in service when the state is (m, n), it follows that

$$P(n \text{ customers in service}) = \sum_{m=0}^{\infty} P_{m,n}$$

and thus

$$\text{average time a customer spends in service} = \frac{\sum_{n=1}^{\infty} n \sum_{m=0}^{\infty} P_{m,n}}{\lambda}$$

(b) To determine W, we use the identity $L = \lambda W$. Because there are $m + n$ customers in the system when the state is (m, n), it follows that

$$W = \frac{L}{\lambda} = \frac{\sum_{m=0}^{\infty} \sum_{n=1}^{\infty} (m+n) P_{m,n}}{\lambda} \tag{8.13}$$

(c) Calling a service that is performed on n customers a type n service, note that such services are completed whenever the state is (m, n) and a service occurs. Because the service rate when the state is (m, n) is μ_n, it follows that the rate at which type n services are completed is $\mu_n \sum_{m=0}^{\infty} P_{m,n}$. Because the rate at which services are completed is the sum of the rates at which type n services are completed, we see that

$$\begin{aligned}
&\text{proportion of all services that are type } n \\
&\quad = \frac{\text{rate at which type } n \text{ services occur}}{\text{rate at which services occur}} \\
&\quad = \frac{\mu_n \sum_{m=0}^{\infty} P_{m,n}}{\sum_{n=1}^{\infty} \mu_n \sum_{m=0}^{\infty} P_{m,n}}
\end{aligned}$$

(d) To determine the proportion of customers that are served in a batch of size n, call such a customer a type n customer. Because n customers depart each time there is a type n service, it follows that

$$\begin{aligned}
&\text{proportion of customers that are type } n \\
&\quad = \frac{\text{rate at which type } n \text{ customers depart}}{\text{rate at which customers depart}} \\
&\quad = \frac{n\mu_n \sum_{m=0}^{\infty} P_{m,n}}{\lambda}
\end{aligned}$$

where the final equality used that the rate at which customers depart is equal to the rate at which they arrive. ■

The needed computations of the preceding example simplify significantly when the service distribution is the same no matter how many customers are being served.

Example 8.11. If in Example 8.10, the service times are all exponential with rate μ no matter how many customers are being simultaneously served, then we can simplify

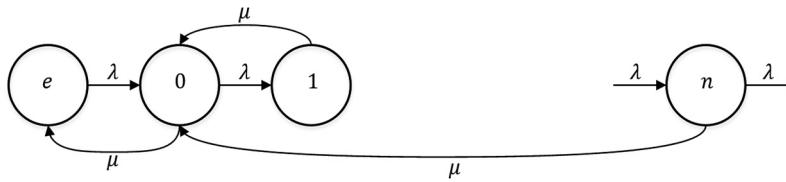

Figure 8.3

the state space by only keeping track of the number in queue. Since when there is no one in queue we would need to know whether or not the server was busy (so as to know whether a new service would begin if an arrival came) we define the following states:

State	*Interpretation*
e	system empty
$n, n \geq 0$	n in queue, server busy

The transition diagram for this system is given in Fig. 8.3.

The balance equations are

State	*Rate leave = Rate enter*
e	$\lambda P_e = \mu P_0$
0	$(\lambda + \mu) P_0 = \lambda P_e + \sum_{n=1}^{\infty} \mu P_n$
$n, n > 0$	$(\lambda + \mu) P_n = \lambda P_{n-1}$

These equations are easily solved. Using that the sum of all the probabilities is 1, the second equation can be rewritten as

$$(\lambda + \mu) P_0 = \lambda P_e + \mu(1 - P_e - P_0)$$

In conjunction with the balance equation for state e, this yields that

$$P_0 = \frac{\lambda\mu}{\lambda^2 + \lambda\mu + \mu^2}, \qquad P_e = \frac{\mu}{\lambda} P_0 = \frac{\mu^2}{\lambda^2 + \lambda\mu + \mu^2}$$

Also, from the balance equation for state n, $n > 0$,

$$P_n = \frac{\lambda}{\lambda + \mu} P_{n-1} = (\frac{\lambda}{\lambda + \mu})^2 P_{n-2} = \ldots = (\frac{\lambda}{\lambda + \mu})^n P_0$$

Now, the amount of time a customer will spend in queue is 0 if that customer finds the system empty, and is exponential with rate μ otherwise. By PASTA, the proportion of arrivals finding the system empty is P_e, yielding that the average time that a customer spends in queue is

$$W_Q = \frac{1 - P_e}{\mu} = \frac{\lambda^2 + \lambda\mu}{\mu(\lambda^2 + \lambda\mu + \mu^2)} \tag{8.14}$$

We can now determine L_Q, W, L from (8.14) by using that

$$L_Q = \lambda W_Q, \qquad W = W_Q + 1/\mu, \qquad L = \lambda W$$

Suppose now that we want to determine Π_n, the proportion of services that are on a batch of size n. To determine this, first note that because the proportion of time the server is busy is $1 - P_e$, and the service rate is μ, it follows that services are completed at rate $\mu(1 - P_e)$. Also, a service involving n people originates, when $n > 1$, whenever a service is completed while there are n in queue; and when $n = 1$, whenever either a service is completed when there is 1 in queue or when an arrival finds the system empty. Hence, the rate at which n customer services occur is μP_n when $n > 1$, and is $\mu P_1 + \lambda P_e$ when $n = 1$. Thus,

$$\Pi_n = \begin{cases} \frac{\mu P_1 + \lambda P_e}{\mu(1-P_e)}, & \text{if } n = 1 \\ \frac{\mu P_n}{\mu(1-P_e)}, & \text{if } n > 1 \end{cases}$$ ■

8.4 Network of Queues

8.4.1 Open Systems

Consider a two-server system in which customers arrive at a Poisson rate λ at server 1. After being served by server 1 they then join the queue in front of server 2. We suppose there is infinite waiting space at both servers. Each server serves one customer at a time with server i taking an exponential time with rate μ_i for a service, $i = 1, 2$. Such a system is called a *tandem* or *sequential* system (see Fig. 8.3).

To analyze this system we need to keep track of the number of customers at server 1 and the number at server 2. So let us define the state by the pair (n, m)—meaning that there are n customers at server 1 and m at server 2. The balance equations are

State	*Rate that the process leaves = rate that it enters*
$0, 0$	$\lambda P_{0,0} = \mu_2 P_{0,1}$
$n, 0; n > 0$	$(\lambda + \mu_1) P_{n,0} = \mu_2 P_{n,1} + \lambda P_{n-1,0}$
$0, m; m > 0$	$(\lambda + \mu_2) P_{0,m} = \mu_2 P_{0,m+1} + \mu_1 P_{1,m-1}$
$n, m; nm > 0$	$(\lambda + \mu_1 + \mu_2) P_{n,m} = \mu_2 P_{n,m+1} + \mu_1 P_{n+1,m-1} + \lambda P_{n-1,m}$

(8.15)

Rather than directly attempting to solve these (along with the equation $\sum_{n,m} P_{n,m} = 1$) we shall guess at a solution and then verify that it indeed satisfies the preceding. We first note that the situation at server 1 is just as in an $M/M/1$ model. Similarly, as it was shown in Section 6.6 that the departure process of an $M/M/1$ queue is a Poisson process with rate λ, it follows that what server 2 faces is also an $M/M/1$ queue. Hence,

the probability that there are n customers at server 1 is

$$P\{n \text{ at server } 1\} = \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right)$$

and, similarly,

$$P\{m \text{ at server } 2\} = \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right)$$

Now, if the numbers of customers at servers 1 and 2 were independent random variables, then it would follow that

$$P_{n,m} = \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right) \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right) \tag{8.16}$$

To verify that $P_{n,m}$ is indeed equal to the preceding (and thus that the number of customers at server 1 is independent of the number at server 2), all we need do is verify that the preceding satisfies Eqs. (8.15)—this suffices since we know that the $P_{n,m}$ are the unique solution of Eqs. (8.15). Now, for instance, if we consider the first equation of (8.15), we need to show that

$$\lambda \left(1 - \frac{\lambda}{\mu_1}\right)\left(1 - \frac{\lambda}{\mu_2}\right) = \mu_2 \left(1 - \frac{\lambda}{\mu_1}\right)\left(\frac{\lambda}{\mu_2}\right)\left(1 - \frac{\lambda}{\mu_2}\right)$$

which is easily verified. We leave it as an exercise to show that the $P_{n,m}$, as given by Eq. (8.16), satisfy all of the equations of (8.15), and are thus the limiting probabilities.

From the preceding we see that L, the average number of customers in the system, is given by

$$\begin{aligned} L &= \sum_{n,m} (n+m) P_{n,m} \\ &= \sum_n n \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right) + \sum_m m \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right) \\ &= \frac{\lambda}{\mu_1 - \lambda} + \frac{\lambda}{\mu_2 - \lambda} \end{aligned}$$

and from this we see that the average time a customer spends in the system is

$$W = \frac{L}{\lambda} = \frac{1}{\mu_1 - \lambda} + \frac{1}{\mu_2 - \lambda}$$

Remarks. **(i)** The result (Eqs. (8.15)) could have been obtained as a direct consequence of the time reversibility of an $M/M/1$ (see Section 6.6). For not only does time reversibility imply that the output from server 1 is a Poisson process, but it also implies (Exercise 26 of Chapter 6) that the number of customers at server 1 is independent of the past departure times from server 1. As these past

departure times constitute the arrival process to server 2, the independence of the numbers of customers in the two systems follows.

(ii) Since a Poisson arrival sees time averages, it follows that in a tandem queue the numbers of customers an arrival (to server 1) sees at the two servers are independent random variables. However, it should be noted that this does not imply that the waiting times of a given customer at the two servers are independent. For a counter example suppose that λ is very small with respect to $\mu_1 = \mu_2$, and thus almost all customers have zero wait in queue at both servers. However, given that the wait in queue of a customer at server 1 is positive, his wait in queue at server 2 also will be positive with probability at least as large as $\frac{1}{2}$ (why?). Hence, the waiting times in queue are not independent. Remarkably enough, however, it turns out that the total times (that is, service time plus wait in queue) that an arrival spends at the two servers are indeed independent random variables.

The preceding result can be substantially generalized. To do so, consider a system of k servers. Customers arrive from outside the system to server $i, i = 1, \ldots, k$, in accordance with independent Poisson processes at rate r_i; they then join the queue at i until their turn at service comes. Once a customer is served by server i, he then joins the queue in front of server j, $j = 1, \ldots, k$, with probability P_{ij}. Hence, $\sum_{j=1}^{k} P_{ij} \leq 1$, and $1 - \sum_{j=1}^{k} P_{ij}$ represents the probability that a customer departs the system after being served by server i.

If we let λ_j denote the total arrival rate of customers to server j, then the λ_j can be obtained as the solution of

$$\lambda_j = r_j + \sum_{i=1}^{k} \lambda_i P_{ij}, \qquad i = 1, \ldots, k \tag{8.17}$$

Eq. (8.17) follows since r_j is the arrival rate of customers to j coming from outside the system and, as λ_i is the rate at which customers depart server i (rate in must equal rate out), $\lambda_i P_{ij}$ is the arrival rate to j of those coming from server i.

It turns out that the number of customers at each of the servers is independent and of the form

$$P\{n \text{ customers at server } j\} = \left(\frac{\lambda_j}{\mu_j}\right)^n \left(1 - \frac{\lambda_j}{\mu_j}\right), \qquad n \geq 1$$

where μ_j is the exponential service rate at server j and the λ_j are the solution to Eq. (8.17). Of course, it is necessary that $\lambda_j/\mu_j < 1$ for all j. To prove this, we first note that it is equivalent to asserting that the limiting probabilities $P(n_1, n_2, \ldots, n_k) = P\{n_j \text{ at server } j, j = 1, \ldots, k\}$ are given by

$$P(n_1, n_2, \ldots, n_k) = \prod_{j=1}^{k} \left(\frac{\lambda_j}{\mu_j}\right)^{n_j} \left(1 - \frac{\lambda_j}{\mu_j}\right) \tag{8.18}$$

which can be verified by showing that it satisfies the balance equations for this model.

The average number of customers in the system is

$$L = \sum_{j=1}^{k} \text{average number at server } j$$
$$= \sum_{j=1}^{k} \frac{\lambda_j}{\mu_j - \lambda_j}$$

The average time a customer spends in the system can be obtained from $L = \lambda W$ with $\lambda = \sum_{j=1}^{k} r_j$. (Why not $\lambda = \sum_{j=1}^{k} \lambda_j$?) This yields

$$W = \frac{\sum_{j=1}^{k} \lambda_j/(\mu_j - \lambda_j)}{\sum_{j=1}^{k} r_j}$$

Remark. The result embodied in Eq. (8.18) is rather remarkable in that it says that the distribution of the number of customers at server i is the same as in an $M/M/1$ system with rates λ_i and μ_i. What is remarkable is that in the network model the arrival process at node i need *not* be a Poisson process. For if there is a possibility that a customer may visit a server more than once (a situation called *feedback*), then the arrival process will not be Poisson. An easy example illustrating this is to suppose that there is a single server whose service rate is very large with respect to the arrival rate from outside. Suppose also that with probability $p = 0.9$ a customer upon completion of service is fed back into the system. Hence, at an arrival time epoch there is a large probability of another arrival in a short time (namely, the feedback arrival); whereas at an arbitrary time point there will be only a very slight chance of an arrival occurring shortly (since λ is so very small). Hence, the arrival process does not possess independent increments and so cannot be Poisson.

Thus, we see that when feedback is allowed the steady-state probabilities of the number of customers at any given station have the same distribution as in an $M/M/1$ model even though the model is not $M/M/1$. (Presumably such quantities as the joint distribution of the number at the station at two different time points will not be the same as for an $M/M/1$.)

Example 8.12. Consider a system of two servers where customers from outside the system arrive at server 1 at a Poisson rate 4 and at server 2 at a Poisson rate 5. The service rates of 1 and 2 are respectively 8 and 10. A customer upon completion of service at server 1 is equally likely to go to server 2 or to leave the system (i.e., $P_{11} = 0$, $P_{12} = \frac{1}{2}$); whereas a departure from server 2 will go 25 percent of the time to server 1 and will depart the system otherwise (i.e., $P_{21} = \frac{1}{4}$, $P_{22} = 0$). Determine the limiting probabilities, L, and W.

Solution: The total arrival rates to servers 1 and 2—call them λ_1 and λ_2—can be obtained from Eq. (8.17). That is, we have

$$\lambda_1 = 4 + \tfrac{1}{4}\lambda_2,$$

$$\lambda_2 = 5 + \tfrac{1}{2}\lambda_1$$

implying that

$$\lambda_1 = 6, \qquad \lambda_2 = 8$$

Hence,

$$\begin{aligned} P\{n \text{ at server } 1, m \text{ at server } 2\} &= \left(\tfrac{3}{4}\right)^n \tfrac{1}{4} \left(\tfrac{4}{5}\right)^m \tfrac{1}{5} \\ &= \tfrac{1}{20} \left(\tfrac{3}{4}\right)^n \left(\tfrac{4}{5}\right)^m \end{aligned}$$

and

$$L = \frac{6}{8-6} + \frac{8}{10-8} = 7,$$
$$W = \frac{L}{9} = \frac{7}{9}$$

■

8.4.2 Closed Systems

The queueing systems described in Section 8.4.1 are called *open systems* since customers are able to enter and depart the system. A system in which new customers never enter and existing ones never depart is called a *closed system*.

Let us suppose that we have m customers moving among a system of k servers, where the service times at server i are exponential with rate $\mu_i, i = 1, \ldots, k$. When a customer completes service at server i, she then joins the queue in front of server $j, j = 1, \ldots, k$, with probability P_{ij}, where we now suppose that $\sum_{j=1}^{k} P_{ij} = 1$ for all $i = 1, \ldots, k$. That is, $\mathbf{P} = [P_{ij}]$ is a Markov transition probability matrix, which we shall assume is irreducible. Let $\pi = (\pi_1, \ldots, \pi_k)$ denote the stationary probabilities for this Markov chain; that is, π is the unique positive solution of

$$\begin{aligned} \pi_j &= \sum_{i=1}^{k} \pi_i P_{ij}, \\ \sum_{j=1}^{k} \pi_j &= 1 \end{aligned} \tag{8.19}$$

If we denote the average arrival rate (or equivalently the average service completion rate) at server j by $\lambda_m(j)$, $j = 1, \ldots, k$ then, analogous to Eq. (8.17), the $\lambda_m(j)$ satisfy

$$\lambda_m(j) = \sum_{i=1}^{k} \lambda_m(i) P_{ij}$$

Hence, from (8.19) we can conclude that

$$\lambda_m(j) = \lambda_m \pi_j, \qquad j = 1, 2, \ldots, k \tag{8.20}$$

where

$$\lambda_m = \sum_{j=1}^{k} \lambda_m(j) \tag{8.21}$$

From Eq. (8.21), we see that λ_m is the average service completion rate of the entire system, that is, it is the system *throughput* rate.[4]

If we let $P_m(n_1, n_2, \ldots, n_k)$ denote the limiting probabilities

$$P_m(n_1, n_2, \ldots, n_k) = P\{n_j \text{ customers at server } j, j = 1, \ldots, k\}$$

then, by verifying that they satisfy the balance equation, it can be shown that

$$P_m(n_1, n_2, \ldots, n_k) = \begin{cases} K_m \prod_{j=1}^{k} (\lambda_m(j)/\mu_j)^{n_j}, & \text{if } \sum_{j=1}^{k} n_j = m \\ 0, & \text{otherwise} \end{cases}$$

But from Eq. (8.20) we thus obtain

$$P_m(n_1, n_2, \ldots, n_k) = \begin{cases} C_m \prod_{j=1}^{k} (\pi_j/\mu_j)^{n_j}, & \text{if } \sum_{j=1}^{k} n_j = m \\ 0, & \text{otherwise} \end{cases} \tag{8.22}$$

where

$$C_m = \left[\sum_{\substack{n_1, \ldots, n_k: \\ \sum n_j = m}} \prod_{j=1}^{k} (\pi_j/\mu_j)^{n_j} \right]^{-1} \tag{8.23}$$

Eq. (8.22) is not as useful as we might suppose, for in order to utilize it we must determine the normalizing constant C_m given by Eq. (8.23), which requires summing the products $\Pi_{j=1}^{k} (\pi_j/\mu_j)^{n_j}$ over all the feasible vectors $(n_1, \ldots, n_k)$: $\sum_{j=1}^{k} n_j = m$. Hence, since there are $\binom{m+k-1}{m}$ vectors this is only computationally feasible for relatively small values of m and k.

We will now present an approach that will enable us to determine recursively many of the quantities of interest in this model without first computing the normalizing constants. To begin, consider a customer who has just left server i and is headed to server j, and let us determine the probability of the system as seen by this customer. In particular, let us determine the probability that this customer observes, at that moment, n_l customers at server $l, l = 1, \ldots, k$, $\sum_{l=1}^{k} n_l = m - 1$. This is done as follows:

$$\begin{aligned} &P\{\text{customer observes } n_l \text{ at server } l, l = 1, \ldots, k \mid \text{customer goes from } i \text{ to } j\} \\ &\quad = \frac{P\{\text{state is } (n_1, \ldots, n_i + 1, \ldots, n_j, \ldots, n_k),\ \text{customer goes from } i \text{ to } j\}}{P\{\text{customer goes from } i \text{ to } j\}} \end{aligned}$$

[4] We are just using the notation $\lambda_m(j)$ and λ_m to indicate the dependence on the number of customers in the closed system. This will be used in recursive relations we will develop.

$$
\begin{aligned}
&= \frac{P_m(n_1, \ldots, n_i + 1, \ldots, n_j, \ldots, n_k)\mu_i P_{ij}}{\sum_{n:\sum n_j = m-1} P_m(n_1, \ldots, n_i + 1, \ldots, n_k)\mu_i P_{ij}} \\
&= \frac{(\pi_i/\mu_i)\prod_{j=1}^{k}(\pi_j/\mu_j)^{n_j}}{K} \qquad \text{from (8.22)} \\
&= C\prod_{j=1}^{k}(\pi_j/\mu_j)^{n_j}
\end{aligned}
$$

where C does not depend on $n_1, \ldots, n_k$. But because the preceding is a probability density on the set of vectors $(n_1, \ldots, n_k)$, $\sum_{j=1}^{k} n_j = m - 1$, it follows from (8.22) that it must equal $P_{m-1}(n_1, \ldots, n_k)$. Hence,

$$
\begin{aligned}
&P\{\text{customer observes } n_l \text{ at server } l, l = 1, \ldots, k \mid \text{customer goes from } i \text{ to } j\} \\
&= P_{m-1}(n_1, \ldots, n_k), \qquad \sum_{i=1}^{k} n_i = m - 1
\end{aligned} \tag{8.24}
$$

As (8.24) is true for all i, we thus have proven the following proposition, known as the arrival theorem.

Proposition 8.3 (The Arrival Theorem)**.** *In the closed network system with m customers, the system as seen by arrivals to server j is distributed as the stationary distribution in the same network system when there are only $m - 1$ customers.*

Denote by $L_m(j)$ and $W_m(j)$ the average number of customers and the average time a customer spends at server j when there are m customers in the network. Upon conditioning on the number of customers found at server j by an arrival to that server, it follows that

$$
\begin{aligned}
W_m(j) &= \frac{1 + E_m[\text{number at server } j \text{ as seen by an arrival}]}{\mu_j} \\
&= \frac{1 + L_{m-1}(j)}{\mu_j}
\end{aligned} \tag{8.25}
$$

where the last equality follows from the arrival theorem. Now when there are $m - 1$ customers in the system, then, from Eq. (8.20), $\lambda_{m-1}(j)$, the average arrival rate to server j, satisfies

$$
\lambda_{m-1}(j) = \lambda_{m-1}\pi_j
$$

Now, applying the basic cost identity Eq. (8.1) with the cost rule being that each customer in the network system of $m - 1$ customers pays one per unit time while at server j, we obtain

$$
L_{m-1}(j) = \lambda_{m-1}\pi_j W_{m-1}(j) \tag{8.26}
$$

Using Eq. (8.25), this yields

$$W_m(j) = \frac{1 + \lambda_{m-1}\pi_j W_{m-1}(j)}{\mu_j} \tag{8.27}$$

Also using the fact that $\sum_{j=1}^{k} L_{m-1}(j) = m - 1$ (why?) we obtain, from Eq. (8.26), the following:

$$m - 1 = \lambda_{m-1} \sum_{j=1}^{k} \pi_j W_{m-1}(j)$$

or

$$\lambda_{m-1} = \frac{m-1}{\sum_{i=1}^{k} \pi_i W_{m-1}(i)} \tag{8.28}$$

Hence, from Eq. (8.27), we obtain the recursion

$$W_m(j) = \frac{1}{\mu_j} + \frac{(m-1)\pi_j W_{m-1}(j)}{\mu_j \sum_{i=1}^{k} \pi_i W_{m-1}(i)} \tag{8.29}$$

Starting with the stationary probabilities π_j, $j = 1, \ldots, k$, and $W_1(j) = 1/\mu_j$ we can now use Eq. (8.29) to determine recursively $W_2(j), W_3(j), \ldots, W_m(j)$. We can then determine the throughput rate λ_m by using Eq. (8.28), and this will determine $L_m(j)$ by Eq. (8.26). This recursive approach is called *mean value analysis*.

Example 8.13. Consider a k-server network in which the customers move in a cyclic permutation. That is,

$$P_{i,i+1} = 1, \qquad i = 1, 2 \ldots, k-1, \qquad P_{k,1} = 1$$

Let us determine the average number of customers at server j when there are two customers in the system. Now, for this network,

$$\pi_i = 1/k, \qquad i = 1, \ldots, k$$

and as

$$W_1(j) = \frac{1}{\mu_j}$$

we obtain from Eq. (8.29) that

$$\begin{aligned} W_2(j) &= \frac{1}{\mu_j} + \frac{(1/k)(1/\mu_j)}{\mu_j \sum_{i=1}^{k} (1/k)(1/\mu_i)} \\ &= \frac{1}{\mu_j} + \frac{1}{\mu_j^2 \sum_{i=1}^{k} 1/\mu_i} \end{aligned}$$

Hence, from Eq. (8.28),

$$\lambda_2 = \frac{2}{\sum_{l=1}^{k} \frac{1}{k} W_2(l)} = \frac{2k}{\sum_{l=1}^{k} \left(\frac{1}{\mu_l} + \frac{1}{\mu_l^2 \sum_{i=1}^{k} 1/\mu_i} \right)}$$

and finally, using Eq. (8.26),

$$L_2(j) = \lambda_2 \frac{1}{k} W_2(j) = \frac{2\left(\frac{1}{\mu_j} + \frac{1}{\mu_j^2 \sum_{i=1}^{k} 1/\mu_i} \right)}{\sum_{l=1}^{k} \left(\frac{1}{\mu_l} + \frac{1}{\mu_l^2 \sum_{i=1}^{k} 1/\mu_i} \right)}$$

■

Another approach to learning about the stationary probabilities specified by Eq. (8.22), which finesses the computational difficulties of computing the constant C_m, is to use the Gibbs sampler of Section 4.9 to generate a Markov chain having these stationary probabilities. To begin, note that since there are always a total of m customers in the system, Eq. (8.22) may equivalently be written as a joint mass function of the numbers of customers at each of the servers $1, \ldots, k-1$, as follows:

$$\begin{aligned} P_m(n_1, \ldots, n_{k-1}) &= C_m (\pi_k/\mu_k)^{m - \sum n_j} \prod_{j=1}^{k-1} (\pi_j/\mu_j)^{n_j} \\ &= K \prod_{j=1}^{k-1} (a_j)^{n_j}, \qquad \sum_{j=1}^{k-1} n_j \le m \end{aligned}$$

where $a_j = (\pi_j \mu_k)/(\pi_k \mu_j)$, $j = 1, \ldots, k-1$. Now, if $\mathbf{N} = (N_1, \ldots, N_{k-1})$ has the preceding joint mass function then

$$\begin{aligned} &P\{N_i = n | N_1 = n_1, \ldots, N_{i-1} = n_{i-1}, N_{i+1} = n_{i+1}, \ldots, N_{k-1} = n_{k-1}\} \\ &= \frac{P_m(n_1, \ldots, n_{i-1}, n, n_{i+1}, \ldots, n_{k-1})}{\sum_r P_m(n_1, \ldots, n_{i-1}, r, n_{i+1}, \ldots, n_{k-1})} \\ &= C a_i^n, \qquad n \le m - \sum_{j \ne i} n_j \end{aligned}$$

It follows from the preceding that we may use the Gibbs sampler to generate the values of a Markov chain whose limiting probability mass function is $P_m(n_1, \ldots, n_{k-1})$ as follows:

1. Let $(n_1, \ldots, n_{k-1})$ be arbitrary nonnegative integers satisfying $\sum_{j=1}^{k-1} n_j \le m$.
2. Generate a random variable I that is equally likely to be any of $1, \ldots, k-1$.
3. If $I = i$, set $s = m - \sum_{j \ne i} n_j$, and generate the value of a random variable X having probability mass function

$$P\{X = n\} = C a_i^n, \qquad n = 0, \ldots, s$$

4. Let $n_I = X$ and go to step 2.

The successive values of the state vector $(n_1, \ldots, n_{k-1}, m - \sum_{j=1}^{k-1} n_j)$ constitute the sequence of states of a Markov chain with the limiting distribution P_m. All quantities of interest can be estimated from this sequence. For instance, the average of the values of the jth coordinate of these vectors will converge to the mean number of individuals at station j, the proportion of vectors whose jth coordinate is less than r will converge to the limiting probability that the number of individuals at station j is less than r, and so on.

Other quantities of interest can also be obtained from the simulation. For instance, suppose we want to estimate W_j, the average amount of time a customer spends at server j on each visit. Then, as noted in the preceding, L_j, the average number of customers at server j, can be estimated. To estimate W_j, we use the identity

$$L_j = \lambda_j W_j$$

where λ_j is the rate at which customers arrive at server j. Setting λ_j equal to the service completion rate at server j shows that

$$\lambda_j = P\{j \text{ is busy}\}\mu_j$$

Using the Gibbs sampler simulation to estimate $P\{j \text{ is busy}\}$ then leads to an estimator of W_j.

8.5 The System $M/G/1$

8.5.1 Preliminaries: Work and Another Cost Identity

For an arbitrary queueing system, let us define the work in the system at any time t to be the sum of the remaining service times of all customers in the system at time t. For instance, suppose there are three customers in the system—the one in service having been there for three of his required five units of service time, and both people in queue having service times of six units. Then the work at that time is $2 + 6 + 6 = 14$. Let V denote the (time) average work in the system.

Now recall the fundamental cost equation (8.1), which states that the

$$\begin{aligned}&\text{average rate at which the system earns}\\ &\quad = \lambda_a \times \text{average amount a customer pays}\end{aligned}$$

and consider the following cost rule: *Each customer pays at a rate of y/unit time when his remaining service time is y, whether he is in queue or in service*. Thus, the rate at which the system earns is just the work in the system; so the basic identity yields

$$V = \lambda_a E[\text{amount paid by a customer}]$$

Now, let S and W_Q^* denote respectively the service time and the time a given customer spends waiting in queue. Then, since the customer pays at a constant rate of S per unit time while he waits in queue and at a rate of $S - x$ after spending an amount of time x in service, we have

$$E[\text{amount paid by a customer}] = E\left[SW_Q^* + \int_0^S (S-x)dx\right]$$

and thus

$$V = \lambda_a E[SW_Q^*] + \frac{\lambda_a E[S^2]}{2} \tag{8.30}$$

It should be noted that the preceding is a basic queueing identity (like Eqs. (8.2)–(8.4)) and as such is valid in almost all models. In addition, if a customer's service time is independent of his wait in queue (as is usually, but not always the case),[5] then we have from Eq. (8.30) that

$$V = \lambda_a E[S]W_Q + \frac{\lambda_a E[S^2]}{2} \tag{8.31}$$

8.5.2 Application of Work to $M/G/1$

The $M/G/1$ model assumes (i) Poisson arrivals at rate λ; (ii) a general service distribution; and (iii) a single server. In addition, we will suppose that customers are served in the order of their arrival.

Now, for an arbitrary customer in an $M/G/1$ system,

$$\text{customer's wait in queue} = \text{work in the system when he arrives} \tag{8.32}$$

This follows since there is only a single server (think about it!). Taking expectations of both sides of Eq. (8.32) yields

$$W_Q = \text{average work as seen by an arrival}$$

But, due to Poisson arrivals, the average work as seen by an arrival will equal V, the time average work in the system. Hence, for the model $M/G/1$,

$$W_Q = V$$

The preceding in conjunction with the identity

$$V = \lambda E[S]W_Q + \frac{\lambda E[S^2]}{2}$$

[5] For an example where it is not true, see Section 8.6.2.

yields the so-called *Pollaczek–Khintchine formula*,

$$W_Q = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} \tag{8.33}$$

where $E[S]$ and $E[S^2]$ are the first two moments of the service distribution.

The quantities L, L_Q, and W can be obtained from Eq. (8.33) as

$$\begin{aligned} L_Q &= \lambda W_Q = \frac{\lambda^2 E[S^2]}{2(1-\lambda E[S])}, \\ W &= W_Q + E[S] = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} + E[S], \\ L &= \lambda W = \frac{\lambda^2 E[S^2]}{2(1-\lambda E[S])} + \lambda E[S] \end{aligned} \tag{8.34}$$

Remarks. **(i)** For the preceding quantities to be finite, we need $\lambda E[S] < 1$. This condition is intuitive since we know from renewal theory that if the server was always busy, then the departure rate would be $1/E[S]$ (see Section 7.3), which must be larger than the arrival rate λ to keep things finite.

(ii) Since $E[S^2] = \text{Var}(S) + (E[S])^2$, we see from Eqs. (8.33) and (8.34) that, for fixed mean service time, L, L_Q, W, and W_Q all increase as the variance of the service distribution increases.

(iii) Another approach to obtain W_Q is presented in Exercise 42.

Example 8.14. Suppose that customers arrive to a single server system in accordance with a Poisson process with rate λ, and that each customer is one of r types. Further, suppose that, independently of all that has previously transpired, each new arrival is type i with probability α_i, $\sum_{i=1}^{r} \alpha_i = 1$. Also, suppose that the amount of time it takes to serve a type i customer has distribution function F_i, with mean μ_i and variance σ_i^2.

(a) Find the average amount of time a type j customer spends in the system, $j = 1, \ldots, r$.

(b) Find the average number of type j customers in the system, $j = 1, \ldots, r$.

Solution: First note that this model is a special case of the $M/G/1$ model, where if S is the service time of a customer, than the service distribution G is obtained by conditioning on the type of the customer:

$$\begin{aligned} G(x) &= P(S \le x) \\ &= \sum_{i=1}^{n} P(S \le x | \text{customer is type } i)\, \alpha_i \\ &= \sum_{i=1}^{n} F_i(x) \alpha_i \end{aligned}$$

To compute $E[S]$ and $E[S^2]$, we condition on the customer's type. This yields

$$\begin{aligned} E[S] &= \sum_{i=1}^{n} E[S|\text{type } i]\,\alpha_i \\ &= \sum_{i=1}^{n} \mu_i \alpha_i \end{aligned}$$

and

$$\begin{aligned} E[S^2] &= \sum_{i=1}^{n} E[S^2|\text{type } i]\alpha_i \\ &= \sum_{i=1}^{n} (\mu_i^2 + \sigma_i^2)\alpha_i \end{aligned}$$

where the final equality used that $E[X^2] = E^2[X] + \text{Var}(X)$. Now, because the time that a customer spends in queue is equal to the work in the system when that customer arrives, it follows that the average time that a type j customer spends in queue, call it $W_Q(j)$, is equal to the average work seen by a time j arrival. However, because type j customers arrive according to a Poisson process with rate $\lambda\alpha_j$ it follows, from the PASTA principle, that the work seen by a type j arrival has the same distribution as the work as it averages over time, and thus the average work seen by a type j arrival is equal to V. Consequently,

$$W_Q(j) = V = \frac{\lambda E[S^2]}{2(1 - \lambda E[S])} = \frac{\lambda \sum_{i=1}^{n} (\mu_i^2 + \sigma_i^2)\alpha_i}{2(1 - \lambda \sum_{i=1}^{n} \mu_i \alpha_i)}$$

With $W(j)$ being the average time that a type j customer spends in the system, we have

$$W(j) = W_Q(j) + \mu_j$$

Finally, using that the average number of type j customers in the system is the average arrival rate of type j customers times the average time they spend in the system ($L = \lambda_a W$ applied to type j customers), we see that $L(j)$, the average number of type j customers in the system, is

$$L(j) = \lambda\alpha_j W(j)$$ ■

8.5.3 Busy Periods

The system alternates between idle periods (when there are no customers in the system, and so the server is idle) and busy periods (when there is at least one customer in the system, and so the server is busy).

Let I and B represent, respectively, the length of an idle and of a busy period. Because I represents the time from when a customer departs and leaves the system empty until the next arrival, it follows, since arrivals are according to a Poisson process with rate λ, that I is exponential with rate λ and thus

$$E[I] = \frac{1}{\lambda} \tag{8.35}$$

To determine $E[B]$ we argue, as in Section 8.3.3, that the long-run proportion of time the system is empty is equal to the ratio of $E[I]$ to $E[I] + E[B]$. That is,

$$P_0 = \frac{E[I]}{E[I] + E[B]} \tag{8.36}$$

To compute P_0, we note from Eq. (8.4) (obtained from the fundamental cost equation by supposing that a customer pays at a rate of one per unit time while in service) that

$$\text{average number of busy servers} = \lambda E[S]$$

However, as the left-hand side of the preceding equals $1 - P_0$ (why?), we have

$$P_0 = 1 - \lambda E[S] \tag{8.37}$$

and, from Eqs. (8.35)–(8.37),

$$1 - \lambda E[S] = \frac{1/\lambda}{1/\lambda + E[B]}$$

or

$$E[B] = \frac{E[S]}{1 - \lambda E[S]}$$

Another quantity of interest is C, the number of customers served in a busy period. The mean of C can be computed by noting that, on the average, for every $E[C]$ arrivals exactly one will find the system empty (namely, the first customer in the busy period). Hence,

$$a_0 = \frac{1}{E[C]}$$

and, as $a_0 = P_0 = 1 - \lambda E[S]$ because of Poisson arrivals, we see that

$$E[C] = \frac{1}{1 - \lambda E[S]}$$

8.6 Variations on the $M/G/1$

8.6.1 The $M/G/1$ with Random-Sized Batch Arrivals

Suppose that, as in the $M/G/1$, arrivals occur in accordance with a Poisson process having rate λ. But now suppose that each arrival consists not of a single customer but of a random number of customers. As before there is a single server whose service times have distribution G.

Let us denote by α_j, $j \geq 1$, the probability that an arbitrary batch consists of j customers; and let N denote a random variable representing the size of a batch and so $P\{N = j\} = \alpha_j$. Since $\lambda_a = \lambda E[N]$, the basic formula for work (Eq. (8.31)) becomes

$$V = \lambda E[N]\left[E(S)W_Q + \frac{E[S^2]}{2}\right] \tag{8.38}$$

To obtain a second equation relating V to W_Q, consider an average customer. We have that

$$\begin{aligned}\text{his wait in queue} &= \text{work in system when he arrives} \\ &\quad + \text{his waiting time due to those in his batch}\end{aligned}$$

Taking expectations and using the fact that Poisson arrivals see time averages yields

$$\begin{aligned}W_Q &= V + E[\text{waiting time due to those in his batch}] \\ &= V + E[W_B]\end{aligned} \tag{8.39}$$

Now, $E(W_B)$ can be computed by conditioning on the number in the batch, but we must be careful because the probability that our average customer comes from a batch of size j is *not* α_j. For α_j is the proportion of batches that are of size j, and if we pick a customer at random, it is more likely that he comes from a larger rather than a smaller batch. (For instance, suppose $\alpha_1 = \alpha_{100} = \frac{1}{2}$, then half the batches are of size 1 but $100/101$ of the customers will come from a batch of size 100!)

To determine the probability that our average customer came from a batch of size j we reason as follows: Let M be a large number. Then of the first M batches approximately $M\alpha_j$ will be of size j, $j \geq 1$, and thus there would have been approximately $jM\alpha_j$ customers that arrived in a batch of size j. Hence, the proportion of arrivals in the first M batches that were from batches of size j is approximately $jM\alpha_j / \sum_j jM\alpha_j$. This proportion becomes exact as $M \to \infty$, and so we see that

$$\begin{aligned}\text{proportion of customers from batches of size } j &= \frac{j\alpha_j}{\sum_j j\alpha_j} \\ &= \frac{j\alpha_j}{E[N]}\end{aligned}$$

We are now ready to compute $E(W_B)$, the expected wait in queue due to others in the batch:

$$E[W_B] = \sum_j E[W_B \mid \text{batch of size } j] \frac{j\alpha_j}{E[N]} \tag{8.40}$$

Now if there are j customers in his batch, then our customer would have to wait for $i-1$ of them to be served if he was ith in line among his batch members. As he is equally likely to be either 1st, 2nd, ..., or jth in line we see that

$$\begin{aligned} E[W_B \mid \text{batch is of size } j] &= \sum_{i=1}^{j} (i-1)E(S)\frac{1}{j} \\ &= \frac{j-1}{2}E[S] \end{aligned}$$

Substituting this in Eq. (8.40) yields

$$\begin{aligned} E[W_B] &= \frac{E[S]}{2E[N]} \sum_j (j-1)j\alpha_j \\ &= \frac{E[S](E[N^2] - E[N])}{2E[N]} \end{aligned}$$

and from Eqs. (8.38) and (8.39) we obtain

$$W_Q = \frac{E[S](E[N^2] - E[N])/(2E[N]) + \lambda E[N]E[S^2]/2}{1 - \lambda E[N]E[S]}$$

Remarks. **(i)** Note that the condition for W_Q to be finite is that

$$\lambda E[N] < \frac{1}{E[S]}$$

which again says that the arrival rate must be less than the service rate (when the server is busy).

(ii) For fixed value of $E[N]$, W_Q is increasing in $\text{Var}(N)$, again indicating that "single-server queues do not like variation."

(iii) The other quantities L, L_Q, and W can be obtained by using

$$\begin{aligned} W &= W_Q + E[S], \\ L &= \lambda_a W = \lambda E[N]W, \\ L_Q &= \lambda E[N]W_Q \end{aligned}$$

8.6.2 Priority Queues

Priority queueing systems are ones in which customers are classified into types and then given service priority according to their type. Consider the situation where there

are two types of customers, which arrive according to independent Poisson processes with respective rates λ_1 and λ_2, and have service distributions G_1 and G_2. We suppose that type 1 customers are given service priority, in that service will never begin on a type 2 customer if a type 1 is waiting. However, if a type 2 is being served and a type 1 arrives, we assume that the service of the type 2 is continued until completion. That is, there is no preemption once service has begun.

Let W_Q^i denote the average wait in queue of a type i customer, $i = 1, 2$. Our objective is to compute the W_Q^i.

First, note that the total work in the system at any time would be exactly the same no matter what priority rule was employed (as long as the server is always busy whenever there are customers in the system). This is so since the work will always decrease at a rate of one per unit time when the server is busy (no matter who is in service) and will always jump by the service time of an arrival. Hence, the work in the system is exactly as it would be if there was no priority rule but rather a first-come, first-served (called FIFO) ordering. However, under FIFO the preceding model is just $M/G/1$ with

$$\begin{aligned} \lambda &= \lambda_1 + \lambda_2, \\ G(x) &= \frac{\lambda_1}{\lambda} G_1(x) + \frac{\lambda_2}{\lambda} G_2(x) \end{aligned} \tag{8.41}$$

which follows since the combination of two independent Poisson processes is itself a Poisson process whose rate is the sum of the rates of the component processes. The service distribution G can be obtained by conditioning on which priority class the arrival is from—as is done in Eq. (8.41).

Hence, from the results of Section 8.5, it follows that V, the average work in the priority queueing system, is given by

$$\begin{aligned} V &= \frac{\lambda E[S^2]}{2(1 - \lambda E[S])} \\ &= \frac{\lambda((\lambda_1/\lambda)E[S_1^2] + (\lambda_2/\lambda)E[S_2^2])}{2[1 - \lambda((\lambda_1/\lambda)E[S_1] + (\lambda_2/\lambda)E[S_2])]} \\ &= \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2(1 - \lambda_1 E[S_1] - \lambda_2 E[S_2])} \end{aligned} \tag{8.42}$$

where S_i has distribution G_i, $i = 1, 2$.

Continuing in our quest for W_Q^i let us note that S and W_Q^*, the service and wait in queue of an arbitrary customer, are not independent in the priority model since knowledge about S gives us information as to the type of customer, which in turn gives us information about W_Q^*. To get around this we will compute separately the average amount of type 1 and type 2 work in the system. Denoting V^i as the average amount of type i work we have, exactly as in Section 8.5.1,

$$V^i = \lambda_i E[S_i] W_Q^i + \frac{\lambda_i E[S_i^2]}{2}, \qquad i = 1, 2 \tag{8.43}$$

If we define

$$V_Q^i \equiv \lambda_i E[S_i] W_Q^i,$$
$$V_S^i \equiv \frac{\lambda_i E[S_i^2]}{2}$$

then we may interpret V_Q^i as the average amount of type i work in queue, and V_S^i as the average amount of type i work in service (why?).

Now we are ready to compute W_Q^1. To do so, consider an arbitrary type 1 arrival. Then

his delay = amount of type 1 work in the system when he arrives
+ amounts of type 2 work in service when he arrives

Taking expectations and using the fact that Poisson arrivals see time average yields

$$\begin{aligned} W_Q^1 &= V^1 + V_S^2 \\ &= \lambda_1 E[S_1] W_Q^1 + \frac{\lambda_1 E[S_1^2]}{2} + \frac{\lambda_2 E[S_2^2]}{2} \end{aligned} \tag{8.44}$$

or

$$W_Q^1 = \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2(1 - \lambda_1 E[S_1])} \tag{8.45}$$

To obtain W_Q^2 we first note that since $V = V^1 + V^2$, we have from Eqs. (8.42) and (8.43) that

$$\begin{aligned} \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2(1 - \lambda_1 E[S_1] - \lambda_2 E[S_2])} &= \lambda_1 E[S_1] W_Q^1 + \lambda_2 E[S_2] W_Q^2 \\ &\quad + \frac{\lambda_1 E[S_1^2]}{2} + \frac{\lambda_2 E[S_2^2]}{2} \\ &= W_Q^1 + \lambda_2 E[S_2] W_Q^2 \qquad \text{from Eq. (8.44)} \end{aligned}$$

Now, using Eq. (8.45), we obtain

$$\lambda_2 E[S_2] W_Q^2 = \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2} \left[\frac{1}{1 - \lambda_1 E[S_1] - \lambda_2 E[S_2]} - \frac{1}{1 - \lambda_1 E[S_1]} \right]$$

or

$$W_Q^2 = \frac{\lambda_1 E[S_1^2] + \lambda_2 E[S_2^2]}{2(1 - \lambda_1 E[S_1] - \lambda_2 E[S_2])(1 - \lambda_1 E[S_1])} \tag{8.46}$$

Remarks. **(i)** Note that from Eq. (8.45), the condition for W_Q^1 to be finite is that $\lambda_1 E[S_1] < 1$, which is independent of the type 2 parameters. (Is this intuitive?)

For W_Q^2 to be finite, we need, from Eq. (8.46), that

$$\lambda_1 E[S_1] + \lambda_2 E[S_2] < 1$$

Since the arrival rate of all customers is $\lambda = \lambda_1 + \lambda_2$, and the average service time of a customer is $(\lambda_1/\lambda)E[S_1] + (\lambda_2/\lambda)E[S_2]$, the preceding condition is just that the average arrival rate be less than the average service rate.

(ii) If there are n types of customers, we can solve for V^j, $j = 1, \ldots, n$, in a similar fashion. First, note that the total amount of work in the system of customers of types $1, \ldots, j$ is independent of the internal priority rule concerning types $1, \ldots, j$ and only depends on the fact that each of them is given priority over any customers of types $j+1, \ldots, n$. (Why is this? Reason it out!) Hence, $V^1 + \cdots + V^j$ is the same as it would be if types $1, \ldots, j$ were considered as a single type I priority class and types $j+1, \ldots, n$ as a single type II priority class. Now, from Eqs. (8.43) and (8.45),

$$V^{\mathrm{I}} = \frac{\lambda_{\mathrm{I}} E[S_{\mathrm{I}}^2] + \lambda_{\mathrm{I}}\lambda_{\mathrm{II}} E[S_{\mathrm{I}}]E[S_{\mathrm{II}}^2]}{2(1 - \lambda_{\mathrm{I}} E[S_{\mathrm{I}}])}$$

where

$$\begin{aligned}
\lambda_{\mathrm{I}} &= \lambda_1 + \cdots + \lambda_j, \\
\lambda_{\mathrm{II}} &= \lambda_{j+1} + \cdots + \lambda_n, \\
E[S_{\mathrm{I}}] &= \sum_{i=1}^{j} \frac{\lambda_i}{\lambda_{\mathrm{I}}} E[S_i], \\
E[S_{\mathrm{I}}^2] &= \sum_{i=1}^{j} \frac{\lambda_i}{\lambda_{\mathrm{I}}} E[S_i^2], \\
E[S_{\mathrm{II}}^2] &= \sum_{i=j+1}^{n} \frac{\lambda_i}{\lambda_{\mathrm{II}}} E[S_i^2]
\end{aligned}$$

Hence, as $V^{\mathrm{I}} = V^1 + \cdots + V^j$, we have an expression for $V^1 + \cdots + V^j$, for each $j = 1, \ldots, n$, which then can be solved for the individual $V^1, V^2, \ldots, V^n$. We now can obtain W_Q^i from Eq. (8.43). The result of all this (which we leave for an exercise) is that

$$W_Q^i = \frac{\lambda_1 E[S_1^2] + \cdots + \lambda_n E[S_n^2]}{2\prod_{j=i-1}^{i}(1 - \lambda_1 E[S_1] - \cdots - \lambda_j E[S_j])}, \qquad i = 1, \ldots, n \quad (8.47)$$

8.6.3 An M/G/1 Optimization Example

Consider a single-server system where customers arrive according to a Poisson process with rate λ, and where the service times are independent and have distribution

function G. Let $\rho = \lambda E[S]$, where S represents a service time random variable, and suppose that $\rho < 1$. Suppose that the server departs whenever a busy period ends and does not return until there are n customers waiting. At that time the server returns and continues serving until the system is once again empty. If the system facility incurs costs at a rate of c per unit time per customer in the system, as well as a cost K each time the server returns, what value of $n, n \geq 1$, minimizes the long-run average cost per unit time incurred by the facility, and what is this minimal cost?

To answer the preceding, let us first determine $A(n)$, the average cost per unit time for the policy that returns the server whenever there are n customers waiting. To do so, say that a new cycle begins each time the server returns. As it is easy to see that everything probabilistically starts over when a cycle begins, it follows from the theory of renewal reward processes that if $C(n)$ is the cost incurred in a cycle and $T(n)$ is the time of a cycle, then

$$A(n) = \frac{E[C(n)]}{E[T(n)]}$$

To determine $E[C(n)]$ and $E[T(n)]$, consider the time interval of length, say, T_i, starting from the first time during a cycle that there are a total of i customers in the system until the first time afterward that there are only $i-1$. Therefore, $\sum_{i=1}^{n} T_i$ is the amount of time that the server is busy during a cycle. Adding the additional mean idle time until n customers are in the system gives

$$E[T(n)] = \sum_{i=1}^{n} E[T_i] + n/\lambda$$

Now, consider the system at the moment when a service is about to begin and there are $i-1$ customers waiting in queue. Since service times do not depend on the order in which customers are served, suppose that the order of service is last come first served, implying that service does not begin on the $i-1$ presently in queue until these $i-1$ are the only ones in the system. Thus, we see that the time that it takes to go from i customers in the system to $i-1$ has the same distribution as the time it takes the $M/G/1$ system to go from a single customer (just beginning service) to empty; that is, its distribution is that of B, the length of an $M/G/1$ busy period. (Essentially the same argument was made in Example 5.25.) Hence,

$$E[T_i] = E[B] = \frac{E[S]}{1-\rho}$$

implying that

$$E[T(n)] = \frac{nE[S]}{1-\lambda E[S]} + \frac{n}{\lambda} = \frac{n}{\lambda(1-\rho)} \tag{8.48}$$

To determine $E[C(n)]$, let C_i denote the cost incurred during the interval of length T_i that starts with $i-1$ in queue and a service just beginning and ends when the $i-1$

in queue are the only customers in the system. Thus, $K + \sum_{i=1}^{n} C_i$ represents the total cost incurred during the busy part of the cycle. In addition, during the idle part of the cycle there will be i customers in the system for an exponential time with rate $\lambda, i = 1, \ldots, n-1$, resulting in an expected cost of $c(1+\cdots+n-1)/\lambda$. Consequently,

$$E[C(n)] = K + \sum_{i=1}^{n} E[C_i] + \frac{n(n-1)c}{2\lambda} \tag{8.49}$$

To find $E[C_i]$, consider the moment when the interval of length T_i begins, and let W_i be the sum of the initial service time plus the sum of the times spent in the system by all the customers that arrive (and are served) until the moment when the interval ends and there are only $i-1$ customers in the system. Then,

$$C_i = (i-1)cT_i + cW_i$$

where the first term refers to the cost incurred due to the $i-1$ customers in queue during the interval of length T_i. As it is easy to see that W_i has the same distribution as W_b, the sum of the times spent in the system by all arrivals in an $M/G/1$ busy period, we obtain

$$E[C_i] = (i-1)c\frac{E[S]}{1-\rho} + cE[W_b] \tag{8.50}$$

Using Eq. (8.49), this yields

$$\begin{aligned} E[C(n)] &= K + \frac{n(n-1)cE[S]}{2(1-\rho)} + ncE[W_b] + \frac{n(n-1)c}{2\lambda} \\ &= K + ncE[W_b] + \frac{n(n-1)c}{2\lambda}\left(\frac{\rho}{1-\rho} + 1\right) \\ &= K + ncE[W_b] + \frac{n(n-1)c}{2\lambda(1-\rho)} \end{aligned}$$

Utilizing the preceding in conjunction with Eq. (8.48) shows that

$$A(n) = \frac{K\lambda(1-\rho)}{n} + \lambda c(1-\rho)E[W_b] + \frac{c(n-1)}{2} \tag{8.51}$$

To determine $E[W_b]$, we use the result that the average amount of time spent in the system by a customer in the $M/G/1$ system is

$$W = W_Q + E[S] = \frac{\lambda E[S^2]}{2(1-\rho)} + E[S]$$

However, if we imagine that on day $j, j \geq 1$, we earn an amount equal to the total time spent in the system by the jth arrival at the $M/G/1$ system, then it follows from

renewal reward processes (since everything probabilistically restarts at the end of a busy period) that

$$W = \frac{E[W_b]}{E[N]}$$

where N is the number of customers served in an $M/G/1$ busy period. Since $E[N] = 1/(1-\rho)$ we see that

$$(1-\rho)E[W_b] = W = \frac{\lambda E[S^2]}{2(1-\rho)} + E[S]$$

Therefore, using Eq. (8.51), we obtain

$$A(n) = \frac{K\lambda(1-\rho)}{n} + \frac{c\lambda^2 E[S^2]}{2(1-\rho)} + c\rho + \frac{c(n-1)}{2}$$

To determine the optimal value of n, treat n as a continuous variable and differentiate the preceding to obtain

$$A'(n) = \frac{-K\lambda(1-\rho)}{n^2} + \frac{c}{2}$$

Setting this equal to 0 and solving yields that the optimal value of n is

$$n^* = \sqrt{\frac{2K\lambda(1-\rho)}{c}}$$

and the minimal average cost per unit time is

$$A(n^*) = \sqrt{2\lambda K(1-\rho)c} + \frac{c\lambda^2 E[S^2]}{2(1-\rho)} + c\rho - \frac{c}{2}$$

It is interesting to see how close we can come to the minimal average cost when we use a simpler policy of the following form: Whenever the server finds the system empty of customers she departs and then returns after a fixed time t has elapsed. Let us say that a new cycle begins each time the server departs. Both the expected costs incurred during the idle and the busy parts of a cycle are obtained by conditioning on $N(t)$, the number of arrivals in the time t that the server is gone. With $\bar{C}(t)$ being the cost incurred during a cycle, we obtain

$$\begin{aligned} E[\bar{C}(t) \mid N(t)] &= K + \sum_{i=1}^{N(t)} E[C_i] + cN(t)\frac{t}{2} \\ &= K + \frac{N(t)(N(t)-1)cE[S]}{2(1-\rho)} + N(t)cE[W_b] + cN(t)\frac{t}{2} \end{aligned}$$

The final term of the first equality is the conditional expected cost during the idle time in the cycle and is obtained by using that, given the number of arrivals in the time t, the

arrival times are independent and uniformly distributed on $(0, t)$; the second equality used Eq. (8.50). Since $N(t)$ is Poisson with mean λt, it follows that $E[N(t)(N(t) - 1)] = E[N^2(t)] - E[N(t)] = \lambda^2 t^2$. Thus, taking the expected value of the preceding gives

$$\begin{aligned} E[\bar{C}(t)] &= K + \frac{\lambda^2 t^2 c E[S]}{2(1-\rho)} + \lambda t c E[W_b] + \frac{c\lambda t^2}{2} \\ &= K + \frac{c\lambda t^2}{2(1-\rho)} + \lambda t c E[W_b] \end{aligned}$$

Similarly, if $\bar{T}(t)$ is the time of a cycle, then

$$\begin{aligned} E[\bar{T}(t)] &= E[E[\bar{T}(t)|N(t)]] \\ &= E[t + N(t)E[B]] \\ &= t + \frac{\rho t}{1-\rho} \\ &= \frac{t}{1-\rho} \end{aligned}$$

Hence, the average cost per unit time, call it $\bar{A}(t)$, is

$$\begin{aligned} \bar{A}(t) &= \frac{E[\bar{C}(t)]}{E[\bar{T}(t)]} \\ &= \frac{K(1-\rho)}{t} + \frac{c\lambda t}{2} + c\lambda(1-\rho)E[W_b] \end{aligned}$$

Thus, from Eq. (8.51), we see that

$$\bar{A}(n/\lambda) - A(n) = c/2$$

which shows that allowing the return decision to depend on the number presently in the system can reduce the average cost only by the amount $c/2$. ■

8.6.4 The M/G/1 Queue with Server Breakdown

Consider a single server queue in which customers arrive according to a Poisson process with rate λ, and where the amount of service time required by each customer has distribution G. Suppose, however, that when working the server breaks down at an exponential rate α. That is, the probability a working server will be able to work for an additional time t without breaking down is $e^{-\alpha t}$. When the server breaks down, it immediately goes to the repair facility. The repair time is a random variable with distribution H. Suppose that the customer in service when a breakdown occurs has its service continue, when the server returns, from the point it was at when the breakdown occurred. (Therefore, the total amount of time a customer is actually receiving service from a working server has distribution G.)

By letting a customer's "service time" include the time that the customer is waiting for the server to come back from being repaired, the preceding is an $M/G/1$ queue. If we let T denote the amount of time from when a customer first enters service until it departs the system, then T is a service time random variable of this $M/G/1$ queue. The average amount of time a customer spends waiting in queue before its service first commences is, thus,

$$W_Q = \frac{\lambda E[T^2]}{2(1-\lambda E[T])}$$

To compute $E[T]$ and $E[T^2]$, let S, having distribution G, be the service requirement of the customer; let N denote the number of times that the server breaks down while the customer is in service; let $R_1, R_2, \ldots$ be the amounts of time the server spends in the repair facility on its successive visits. Then,

$$T = \sum_{i=1}^{N} R_i + S$$

Conditioning on S yields

$$E[T|S=s] = E\left[\sum_{i=1}^{N} R_i|S=s\right] + s,$$
$$Var(T|S=s) = \text{Var}\left(\sum_{i=1}^{N} R_i|S=s\right)$$

Now, a working server always breaks down at an exponential rate α. Therefore, given that a customer requires s units of service time, it follows that the number of server breakdowns while that customer is being served is a Poisson random variable with mean αs. Consequently, conditional on $S=s$, the random variable $\sum_{i=1}^{N} R_i$ is a compound Poisson random variable with Poisson mean αs. Using the results from Examples 3.10 and 3.19, we thus obtain

$$E\left[\sum_{i=1}^{N} R_i|S=s\right] = \alpha s E[R], \qquad \text{Var}\left(\sum_{i=1}^{N} R_i|S=s\right) = \alpha s E[R^2]$$

where R has the repair distribution H. Therefore,

$$E[T|S] = \alpha S E[R] + S = S(1+\alpha E[R]),$$
$$\text{Var}(T|S) = \alpha S E[R^2]$$

Thus,

$$E[T] = E[E[T|S]] = E[S](1+\alpha E[R])$$

and, by the conditional variance formula,

$$\begin{aligned}\text{Var}(T) &= E[\text{Var}(T|S)] + \text{Var}(E[T|S]) \\ &= \alpha E[S]E[R^2] + (1+\alpha E[R])^2 \text{Var}(S)\end{aligned}$$

Therefore,

$$\begin{aligned}E[T^2] &= \text{Var}(T) + (E[T])^2 \\ &= \alpha E[S]E[R^2] + (1+\alpha E[R])^2 E[S^2]\end{aligned}$$

Consequently, assuming that $\lambda E[T] = \lambda E[S](1+\alpha E[R]) < 1$, we obtain

$$W_Q = \frac{\lambda\alpha E[S]E[R^2] + \lambda(1+\alpha E[R])^2 E[S^2]}{2(1-\lambda E[S](1+\alpha E[R]))}$$

From the preceding, we can now obtain

$$\begin{aligned}L_Q &= \lambda W_Q, \\ W &= W_Q + E[T], \\ L &= \lambda W\end{aligned}$$

Some other quantities we might be interested in are

(i) P_w, the proportion of time the server is working;
(ii) P_r, the proportion of time the server is being repaired;
(iii) P_I, the proportion of time the server is idle.

These quantities can all be obtained by using the queueing cost identity. For instance, if we suppose that customers pay 1 per unit time while actually being served, then

$$\begin{aligned}&\text{average rate at which system earns} = P_w, \\ &\quad \text{average amount a customer pays} = E[S]\end{aligned}$$

Therefore, the identity yields

$$P_w = \lambda E[S]$$

To determine P_r, suppose a customer whose service is interrupted pays 1 per unit time while the server is being repaired. Then,

$$\begin{aligned}&\text{average rate at which system earns} = P_r, \\ &\quad \text{average amount a customer pays} = E\left[\sum_{i=1}^{N} R_i\right] = \alpha E[S]E[R]\end{aligned}$$

yielding

$$P_r = \lambda\alpha E[S]E[R]$$

Of course, P_I can be obtained from

$$P_I = 1 - P_w - P_r$$

Remark. The quantities P_w and P_r could also have been obtained by first noting that $1 - P_I = \lambda E[T]$ is the proportion of time the server is either working or in repair. Thus,

$$P_w = \lambda E[T]\frac{E[S]}{E[T]} = \lambda E[S],$$
$$P_r = \lambda E[T]\frac{E[T] - E[S]}{E[T]} = \lambda E[S]\alpha E[R]$$ ■

8.7 The Model $G/M/1$

The model $G/M/1$ assumes that the times between successive arrivals have an arbitrary distribution G. The service times are exponentially distributed with rate μ and there is a single server.

The immediate difficulty in analyzing this model stems from the fact that the number of customers in the system is not informative enough to serve as a state space. For in summarizing what has occurred up to the present we would need to know not only the number in the system, but also the amount of time that has elapsed since the last arrival (since G is not memoryless). (Why need we not be concerned with the amount of time the person being served has already spent in service?) To get around this problem we shall only look at the system when a customer arrives; and so let us define $X_n, n \geq 1$, by

$X_n \equiv$ the number in the system as seen by the nth arrival

It is easy to see that the process $\{X_n, n \geq 1\}$ is a Markov chain. To compute the transition probabilities P_{ij} for this Markov chain let us first note that, as long as there are customers to be served, the number of services in any length of time t is a Poisson random variable with mean μt. This is true since the time between successive services is exponential and, as we know, this implies that the number of services thus constitutes a Poisson process. Hence,

$$P_{i,i+1-j} = \int_0^\infty e^{-\mu t}\frac{(\mu t)^j}{j!}\,dG(t), \qquad j = 0, 1, \ldots, i$$

which follows since if an arrival finds i in the system, then the next arrival will find $i + 1$ minus the number served, and the probability that j will be served is easily seen to equal the right side of the preceding (by conditioning on the time between the successive arrivals).

The formula for P_{i0} is a little different (it is the probability that *at least* $i + 1$ Poisson events occur in a random length of time having distribution G) and can be

obtained from

$$P_{i0} = 1 - \sum_{j=0}^{i} P_{i,i+1-j}$$

The limiting probabilities $\pi_k, k = 0, 1, \ldots,$ can be obtained as the unique solution of

$$\pi_k = \sum_{i=0}^{\infty} \pi_i P_{ik}, \qquad k \geq 0,$$

$$\sum_{k=0}^{\infty} \pi_k = 1$$

which, in this case, reduce to

$$\pi_k = \sum_{i=k-1}^{\infty} \pi_i \int_0^{\infty} e^{-\mu t} \frac{(\mu t)^{i+1-k}}{(i+1-k)!}\, dG(t), \qquad k \geq 1,$$

$$\sum_{k=0}^{\infty} \pi_k = 1 \tag{8.52}$$

(We have not included the equation $\pi_0 = \sum \pi_i P_{i0}$ since one of the equations is always redundant.)

To solve the preceding, let us try a solution of the form $\pi_k = c\beta^k$. Substitution into Eq. (8.52) leads to

$$\begin{aligned} c\beta^k &= c \sum_{i=k-1}^{\infty} \beta^i \int_0^{\infty} e^{-\mu t} \frac{(\mu t)^{i+1-k}}{(i+1-k)!}\, dG(t) \\ &= c \int_0^{\infty} e^{-\mu t} \beta^{k-1} \sum_{i=k-1}^{\infty} \frac{(\beta\mu t)^{i+1-k}}{(i+1-k)!}\, dG(t) \end{aligned} \tag{8.53}$$

However,

$$\begin{aligned} \sum_{i=k-1}^{\infty} \frac{(\beta\mu t)^{i+1-k}}{(i+1-k)!} &= \sum_{j=0}^{\infty} \frac{(\beta\mu t)^j}{j!} \\ &= e^{\beta\mu t} \end{aligned}$$

and thus Eq. (8.53) reduces to

$$\beta^k = \beta^{k-1} \int_0^{\infty} e^{-\mu t(1-\beta)} dG(t)$$

or

$$\beta = \int_0^{\infty} e^{-\mu t(1-\beta)} dG(t) \tag{8.54}$$

The constant c can be obtained from $\sum_k \pi_k = 1$, which implies that

$$c\sum_{k=0}^{\infty}\beta^k = 1$$

or

$$c = 1-\beta$$

As (π_k) is the *unique* solution to Eq. (8.52), and $\pi_k = (1-\beta)\beta^k$ satisfies, it follows that

$$\pi_k = (1-\beta)\beta^k, \qquad k = 0, 1, \ldots$$

where β is the solution of Eq. (8.54). (It can be shown that if the mean of G is greater than the mean service time $1/\mu$, then there is a unique value of β satisfying Eq. (8.54) which is between 0 and 1.) The exact value of β usually can only be obtained by numerical methods.

As π_k is the limiting probability that an arrival sees k customers, it is just the a_k as defined in Section 8.2. Hence,

$$\alpha_k = (1-\beta)\beta^k, \qquad k \geq 0 \tag{8.55}$$

We can obtain W by conditioning on the number in the system when a customer arrives. This yields

$$\begin{aligned} W &= \sum_k E[\text{time in system} \mid \text{arrival sees } k](1-\beta)\beta^k \\ &= \sum_k \frac{k+1}{\mu}(1-\beta)\beta^k \qquad \text{(Since if an arrival sees } k \text{ then he spends } k+1 \text{ service periods in the system)} \\ &= \frac{1}{\mu(1-\beta)} \qquad \left(\text{by using } \sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2}\right) \end{aligned}$$

and

$$\begin{aligned} W_Q &= W - \frac{1}{\mu} = \frac{\beta}{\mu(1-\beta)}, \\ L &= \lambda W = \frac{\lambda}{\mu(1-\beta)}, \\ L_Q &= \lambda W_Q = \frac{\lambda\beta}{\mu(1-\beta)} \end{aligned} \tag{8.56}$$

where λ is the reciprocal of the mean interarrival time. That is,

$$\frac{1}{\lambda} = \int_0^{\infty} x\, dG(x)$$

In fact, in exactly the same manner as shown for the $M/M/1$ in Section 8.3.1 and Exercise 6 we can show that

$$W^* \text{ is exponential with rate } \mu(1-\beta),$$

$$W_Q^* = \begin{cases} 0 & \text{with probability } 1-\beta \\ \text{exponential with rate } \mu(1-\beta) & \text{with probability } \beta \end{cases}$$

where W^* and W_Q^* are the amounts of time that a customer spends in system and queue, respectively (their means are W and W_Q).

Whereas $a_k = (1-\beta)\beta^k$ is the probability that an arrival sees k in the system, it is not equal to the proportion of time during which there are k in the system (since the arrival process is not Poisson). To obtain the P_k we first note that the rate at which the number in the system changes from $k-1$ to k must equal the rate at which it changes from k to $k-1$ (why?). Now the rate at which it changes from $k-1$ to k is equal to the arrival rate λ multiplied by the proportion of arrivals finding $k-1$ in the system. That is,

$$\text{rate number in system goes from } k-1 \text{ to } k = \lambda a_{k-1}$$

Similarly, the rate at which the number in the system changes from k to $k-1$ is equal to the proportion of time during which there are k in the system multiplied by the (constant) service rate. That is,

$$\text{rate number in system goes from } k \text{ to } k-1 = P_k \mu$$

Equating these rates yields

$$P_k = \frac{\lambda}{\mu} a_{k-1}, \qquad k \geq 1$$

and so, from Eq. (8.55),

$$P_k = \frac{\lambda}{\mu}(1-\beta)\beta^{k-1}, \qquad k \geq 1$$

and, as $P_0 = 1 - \sum_{k=1}^{\infty} P_k$, we obtain

$$P_0 = 1 - \frac{\lambda}{\mu}$$

Remarks. In the foregoing analysis we guessed at a solution of the stationary probabilities of the Markov chain of the form $\pi_k = c\beta^k$, then verified such a solution by substituting in the stationary Eq. (8.52). However, it could have been argued directly that the stationary probabilities of the Markov chain are of this form. To do so, define β_i to be the expected number of times that state $i+1$ is visited in the Markov chain

between two successive visits to state $i, i \geq 0$. Now it is not difficult to see (and we will let you argue it out for yourself) that

$$\beta_0 = \beta_1 = \beta_2 = \cdots = \beta$$

Now it can be shown by using renewal reward processes that

$$\begin{aligned}\pi_{i+1} &= \frac{E[\text{number of visits to state } i+1 \text{ in an } i\text{–}i \text{ cycle}]}{E[\text{number of transitions in an } i\text{–}i \text{ cycle}]} \\ &= \frac{\beta_i}{1/\pi_i}\end{aligned}$$

and so,

$$\pi_{i+1} = \beta_i \pi_i = \beta \pi_i, \qquad i \geq 0$$

implying, since $\sum_{i=0}^{\infty} \pi_i = 1$, that

$$\pi_i = \beta^i (1 - \beta), \qquad i \geq 0$$

8.7.1 The G/M/1 Busy and Idle Periods

Suppose that an arrival has just found the system empty—and so initiates a busy period—and let N denote the number of customers served in that busy period. Since the Nth arrival (after the initiator of the busy period) will also find the system empty, it follows that N is the number of transitions for the Markov chain (of Section 8.7) to go from state 0 to state 0. Hence, $1/E[N]$ is the proportion of transitions that take the Markov chain into state 0; or equivalently, it is the proportion of arrivals that find the system empty. Therefore,

$$E[N] = \frac{1}{a_0} = \frac{1}{1 - \beta}$$

Also, as the next busy period begins after the Nth interarrival, it follows that the cycle time (that is, the sum of a busy and idle period) is equal to the time until the Nth interarrival. In other words, the sum of a busy and idle period can be expressed as the sum of N interarrival times. Thus, if T_i is the ith interarrival time after the busy period begins, then

$$\begin{aligned}E[\text{Busy}] + E[\text{Idle}] &= E\left[\sum_{i=1}^{N} T_i\right] \\ &= E[N]E[T] \qquad \text{(by Wald's equation)} \\ &= \frac{1}{\lambda(1 - \beta)}\end{aligned} \tag{8.57}$$

For a second relation between $E[\text{Busy}]$ and $E[\text{Idle}]$, we can use the same argument as in Section 8.5.3 to conclude that

$$1 - P_0 = \frac{E[\text{Busy}]}{E[\text{Idle}] + E[\text{Busy}]}$$

and since $P_0 = 1 - \lambda/\mu$, we obtain, upon combining this with (8.57), that

$$E[\text{Busy}] = \frac{1}{\mu(1-\beta)},$$
$$E[\text{Idle}] = \frac{\mu - \lambda}{\lambda\mu(1-\beta)}$$

8.8 A Finite Source Model

Consider a system of m machines, whose working times are independent exponential random variables with rate λ. Upon failure, a machine instantly goes to a repair facility that consists of a single repairperson. If the repairperson is free, repair begins on the machine; otherwise, the machine joins the queue of failed machines. When a machine is repaired it becomes a working machine, and repair begins on a new machine from the queue of failed machines (provided the queue is nonempty). The successive repair times are independent random variables having density function g, with mean

$$\mu_R = \int_0^\infty x g(x)\, dx$$

To analyze this system, so as to determine such quantities as the average number of machines that are down and the average time that a machine is down, we will exploit the exponentially distributed working times to obtain a Markov chain. Specifically, let X_n denote the number of failed machines immediately after the nth repair occurs, $n \geq 1$. Now, if $X_n = i > 0$, then the situation when the nth repair has just occurred is that repair is about to begin on a machine, there are $i - 1$ other machines waiting for repair, and there are $m - i$ working machines, each of which will (independently) continue to work for an exponential time with rate λ. Similarly, if $X_n = 0$, then all m machines are working and will (independently) continue to do so for exponentially distributed times with rate λ. Consequently, any information about earlier states of the system will not affect the probability distribution of the number of down machines at the moment of the next repair completion; hence, $\{X_n, n \geq 1\}$ is a Markov chain. To determine its transition probabilities $P_{i,j}$, suppose first that $i > 0$. Conditioning on R, the length of the next repair time, and making use of the independence of the $m - i$ remaining working times, yields that for $j \leq m - i$

$$\begin{aligned} P_{i,i-1+j} &= P\{j \text{ failures during } R\} \\ &= \int_0^\infty P\{j \text{ failures during } R \mid R = r\} g(r)\, dr \end{aligned}$$

$$= \int_0^\infty \binom{m-i}{j} (1-e^{-\lambda r})^j (e^{-\lambda r})^{m-i-j} g(r)\, dr$$

If $i = 0$, then, because the next repair will not begin until one of the machines fails,

$$P_{0,j} = P_{1,j}, \qquad j \le m-1$$

Let π_j, $j = 0, \ldots, m-1$, denote the stationary probabilities of this Markov chain. That is, they are the unique solution of

$$\pi_j = \sum_i \pi_i P_{i,j},$$

$$\sum_{j=0}^{m-1} \pi_j = 1$$

Therefore, after explicitly determining the transition probabilities and solving the preceding equations, we would know the value of π_0, the proportion of repair completions that leaves all machines working. Let us say that the system is "on" when all machines are working and "off" otherwise. (Thus, the system is on when the repairperson is idle and off when he is busy.) As all machines are working when the system goes back on, it follows from the lack of memory property of the exponential that the system probabilistically starts over when it goes on. Hence, this on–off system is an alternating renewal process. Suppose that the system has just become on, thus starting a new cycle, and let $R_i, i \ge 1$, be the time of the ith repair from that moment. Also, let N denote the number of repairs in the off (busy) time of the cycle. Then, it follows that B, the length of the off period, can be expressed as

$$B = \sum_{i=1}^{N} R_i$$

Although N is not independent of the sequence $R_1, R_2, \ldots$, it is easy to check that it is a stopping time for this sequence, and thus by Wald's equation (see Exercise 13 of Chapter 7) we have

$$E[B] = E[N]E[R] = E[N]\mu_R$$

Also, since an on time will last until one of the machines fails, and since the minimum of independent exponential random variables is exponential with a rate equal to the sum of their rates, it follows that $E[I]$, the mean on (idle) time in a cycle, is given by

$$E[I] = 1/(m\lambda)$$

Hence, P_B, the proportion of time that the repairperson is busy, satisfies

$$P_B = \frac{E[N]\mu_R}{E[N]\mu_R + 1/(m\lambda)}$$

However, since, on average, one out of every $E[N]$ repair completions will leave all machines working, it follows that

$$\pi_0 = \frac{1}{E[N]}$$

Consequently,

$$P_B = \frac{\mu_R}{\mu_R + \pi_0/(m\lambda)} \tag{8.58}$$

Now focus attention on one of the machines, call it machine number 1, and let $P_{1,R}$ denote the proportion of time that machine 1 is being repaired. Since the proportion of time that the repairperson is busy is P_B, and since all machines fail at the same rate and have the same repair distribution, it follows that

$$P_{1,R} = \frac{P_B}{m} = \frac{\mu_R}{m\mu_R + \pi_0/\lambda} \tag{8.59}$$

However, machine 1 alternates between time periods when it is working, when it is waiting in queue, and when it is in repair. Let W_i, Q_i, S_i denote, respectively, the ith working time, the ith queueing time, and the ith repair time of machine 1, $i \geq 1$. Then, the proportion of time that machine 1 is being repaired during its first n working–queue–repair cycles is as follows:

$$\begin{aligned}
&\text{proportion of time in the first } n \text{ cycles that machine 1 is being repaired} \\
&= \frac{\sum_{i=1}^n S_i}{\sum_{i=1}^n W_i + \sum_{i=1}^n Q_i + \sum_{i=1}^n S_i} \\
&= \frac{\sum_{i=1}^n S_i/n}{\sum_{i=1}^n W_i/n + \sum_{i=1}^n Q_i/n + \sum_{i=1}^n S_i/n}
\end{aligned}$$

Letting $n \to \infty$ and using the strong law of large numbers to conclude that the averages of the W_i and of the S_i converge, respectively, to $1/\lambda$ and μ_R, yields

$$P_{1,R} = \frac{\mu_R}{1/\lambda + \bar{Q} + \mu_R}$$

where $\bar{Q}$ is the average amount of time that machine 1 spends in queue when it fails. Using Eq. (8.59), the preceding gives

$$\frac{\mu_R}{m\mu_R + \pi_0/\lambda} = \frac{\mu_R}{1/\lambda + \bar{Q} + \mu_R}$$

or, equivalently, that

$$\bar{Q} = (m-1)\mu_R - (1-\pi_0)/\lambda$$

Moreover, since all machines are probabilistically equivalent it follows that $\bar{Q}$ is equal to W_Q, the average amount of time that a failed machine spends in queue. To determine

the average number of machines in queue, we will make use of the basic queueing identity

$$L_Q = \lambda_a W_Q = \lambda_a \bar{Q}$$

where λ_a is the average rate at which machines fail. To determine λ_a, again focus attention on machine 1 and suppose that we earn one per unit time whenever machine 1 is being repaired. It then follows from the basic cost identity of Eq. (8.1) that

$$P_{1,R} = r_1 \mu_R$$

where r_1 is the average rate at which machine 1 fails. Thus, from Eq. (8.59), we obtain

$$r_1 = \frac{1}{m\mu_R + \pi_0/\lambda}$$

Because all m machines fail at the same rate, the preceding implies that

$$\lambda_a = mr_1 = \frac{m}{m\mu_R + \pi_0/\lambda}$$

which gives that the average number of machines in queue is

$$L_Q = \frac{m(m-1)\mu_R - m(1-\pi_0)/\lambda}{m\mu_R + \pi_0/\lambda}$$

Since the average number of machines being repaired is P_B, the preceding, along with Eq. (8.58), shows that the average number of down machines is

$$L = L_Q + P_B = \frac{m^2\mu_R - m(1-\pi_0)/\lambda}{m\mu_R + \pi_0/\lambda}$$

8.9 Multiserver Queues

By and large, systems that have more than one server are much more difficult to analyze than those with a single server. In Section 8.9.1 we start first with a Poisson arrival system in which no queue is allowed, and then consider in Section 8.9.2 the infinite capacity $M/M/k$ system. For both of these models we are able to present the limiting probabilities. In Section 8.9.3 we consider the model $G/M/k$. The analysis here is similar to that of the $G/M/1$ (Section 8.7) except that in place of a single quantity β given as the solution of an integral equation, we have k such quantities. We end in Section 8.9.4 with the model $M/G/k$ for which unfortunately our previous technique (used in $M/G/1$) no longer enables us to derive W_Q, and we content ourselves with an approximation.

8.9.1 Erlang's Loss System

A loss system is a queueing system in which arrivals that find all servers busy do not enter but rather are lost to the system. The simplest such system is the $M/M/k$ loss system in which customers arrive according to a Poisson process having rate λ, enter the system if at least one of the k servers is free, and then spend an exponential amount of time with rate μ being served. The balance equations for this system are

$$\begin{array}{cc} \textit{State} & \textit{Rate leave} = \textit{Rate enter} \\ 0 & \lambda P_0 = \mu P_1 \\ 1 & (\lambda+\mu)P_1 = 2\mu P_2 + \lambda P_0 \\ 2 & (\lambda+2\mu)P_2 = 3\mu P_3 + \lambda P_1 \\ i, 0 < i < k & (\lambda + i\mu)P_i = (i+1)\mu P_{i+1} + \lambda P_{i-1} \\ k & k\mu P_k = \lambda P_{k-1} \end{array}$$

Rewriting gives

$$\begin{aligned} \lambda P_0 &= \mu P_1, \\ \lambda P_1 &= 2\mu P_2, \\ \lambda P_2 &= 3\mu P_3, \\ &\vdots \\ \lambda P_{k-1} &= k\mu P_k \end{aligned}$$

or

$$\begin{aligned} P_1 &= \frac{\lambda}{\mu} P_0, \\ P_2 &= \frac{\lambda}{2\mu} P_1 = \frac{(\lambda/\mu)^2}{2} P_0, \\ P_3 &= \frac{\lambda}{3\mu} P_2 = \frac{(\lambda/\mu)^3}{3!} P_0, \\ &\vdots \\ P_k &= \frac{\lambda}{k\mu} P_{k-1} = \frac{(\lambda/\mu)^k}{k!} P_0 \end{aligned}$$

and using $\sum_{i=0}^{k} P_i = 1$, we obtain

$$P_i = \frac{(\lambda/\mu)^i / i!}{\sum_{j=0}^{k} (\lambda/\mu)^j / j!}, \qquad i = 0, 1, \ldots, k$$

Since $E[S] = 1/\mu$, where $E[S]$ is the mean service time, the preceding can be written as

$$P_i = \frac{(\lambda E[S])^i / i!}{\sum_{j=0}^{k} (\lambda E[S])^j / j!}, \qquad i = 0, 1, \ldots, k \tag{8.60}$$

Consider now the same system except that the service distribution is general—that is, consider the $M/G/k$ with no queue allowed. This model is sometimes called the *Erlang loss system*. It can be shown (though the proof is advanced) that Eq. (8.60) (which is called *Erlang's loss formula*) remains valid for this more general system.

Remark. It is easy to see that Eq. (8.60) is valid when $k = 1$. For in this case, $L = P_1$, $W = E[S]$, and $\lambda_a = \lambda P_0$. Using that $L = \lambda_a W$ gives

$$P_1 = \lambda P_0 E[S]$$

which implies, since $P_0 + P_1 = 1$, that

$$P_0 = \frac{1}{1 + \lambda E[S]}, \qquad P_1 = \frac{\lambda E[S]}{1 + \lambda E[S]}$$

■

8.9.2 The M/M/k Queue

The $M/M/k$ infinite capacity queue can be analyzed by the balance equation technique. We leave it for you to verify that

$$P_i = \begin{cases} \dfrac{\dfrac{(\lambda/\mu)^i}{i!}}{\displaystyle\sum_{i=0}^{k-1} \frac{(\lambda/\mu)^i}{i!} + \frac{(\lambda/\mu)^k}{k!} \frac{k\mu}{k\mu - \lambda}}, & i \le k \\ \dfrac{(\lambda/k\mu)^i k^k}{k!} P_0, & i > k \end{cases}$$

We see from the preceding that we need to impose the condition $\lambda < k\mu$.

8.9.3 The G/M/k Queue

In this model we again suppose that there are k servers, each of whom serves at an exponential rate μ. However, we now allow the time between successive arrivals to have an arbitrary distribution G. To ensure that a steady-state (or limiting) distribution exists, we assume the condition $1/\mu_G < k\mu$ where μ_G is the mean of G.[6]

[6] It follows from the renewal theory (Proposition 7.1) that customers arrive at rate $1/\mu_G$, and as the maximum service rate is $k\mu$, we clearly need that $1/\mu_G < k\mu$ for limiting probabilities to exist.

The analysis for this model is similar to that presented in Section 8.7 for the case $k = 1$. Namely, to avoid having to keep track of the time since the last arrival, we look at the system only at arrival epochs. Once again, if we define X_n as the number in the system at the moment of the nth arrival, then $\{X_n, n \geq 0\}$ is a Markov chain.

To derive the transition probabilities of the Markov chain, it helps to first note the relationship

$$X_{n+1} = X_n + 1 - Y_n, \qquad n \geq 0$$

where Y_n denotes the number of departures during the interarrival time between the nth and $(n + 1)$ st arrival. The transition probabilities P_{ij} can now be calculated as follows:

Case 1. $j > i + 1$.

In this case it easily follows that $P_{ij} = 0$.

Case 2. $j \leq i + 1 \leq k$.

In this case if an arrival finds i in the system, then as $i < k$ the new arrival will also immediately enter service. Hence, the next arrival will find j if of the $i + 1$ services exactly $i + 1 - j$ are completed during the interarrival time. Conditioning on the length of this interarrival time yields

$$\begin{aligned} P_{ij} &= P\{i + 1 - j \text{ of } i + 1 \text{ services are completed in an interarrival time}\} \\ &= \int_0^\infty P\{i + 1 - j \text{ of } i + 1 \text{ are completed } | \text{interarrival time is } t\}\, dG(t) \\ &= \int_0^\infty \binom{i+1}{j} (i - e^{-\mu t})^{i+1-j} (e^{-\mu t})^j\, dG(t) \end{aligned}$$

where the last equality follows since the number of service completions in a time t will have a binomial distribution.

Case 3. $i + 1 \geq j \geq k$.

To evaluate P_{ij} in this case we first note that when all servers are busy, the departure process is a Poisson process with rate $k\mu$ (why?). Hence, again conditioning on the interarrival time we have

$$\begin{aligned} P_{ij} &= P\{i + 1 - j \text{ departures}\} \\ &= \int_0^\infty P\{i + 1 - j \text{ departures in time } t\}\, dG(t) \\ &= \int_0^\infty e^{-k\mu t} \frac{(k\mu t)^{i+1-j}}{(i + 1 - j)!}\, dG(t) \end{aligned}$$

Case 4. $i + 1 \geq k > j$.

In this case since when all servers are busy the departure process is a Poisson process, it follows that the length of time until there will only be k in the system will

have a gamma distribution with parameters $i+1-k, k\mu$ (the time until $i+1-k$ events of a Poisson process with rate $k\mu$ occur is gamma distributed with parameters $i+1-k, k\mu$). Conditioning first on the interarrival time and then on the time until there are only k in the system (call this latter random variable T_k) yields

$$\begin{aligned}
P_{ij} &= \int_0^\infty P\{i+1-j \text{ departures in time } t\}\, dG(t) \\
&= \int_0^\infty \int_0^t P\{i+1-j \text{ departures in } t \mid T_k = s\} k\mu e^{-k\mu s} \frac{(k\mu s)^{i-k}}{(i-k)!}\, ds\, dG(t) \\
&= \int_0^\infty \int_0^t \binom{k}{j} \left(1-e^{-\mu(t-s)}\right)^{k-j} \left(e^{-\mu(t-s)}\right)^{j} k\mu e^{-k\mu s} \frac{(k\mu s)^{i-k}}{(i-k)!}\, ds\, dG(t)
\end{aligned}$$

where the last equality follows since of the k people in service at time s the number whose service will end by time t is binomial with parameters k and $1-e^{-\mu(t-s)}$.

We now can verify either by a direct substitution into the equations $\pi_j = \sum_i \pi_i P_{ij}$, or by the same argument as presented in the remark at the end of Section 8.7, that the limiting probabilities of this Markov chain are of the form

$$\pi_{k-1+j} = c\beta^j, \qquad j = 0, 1, \ldots.$$

Substitution into any of the equations $\pi_j = \sum_i \pi_i P_{ij}$ when $j > k$ yields that β is given as the solution of

$$\beta = \int_0^\infty e^{-k\mu t(1-\beta)}\, dG(t)$$

The values $\pi_0, \pi_1, \ldots, \pi_{k-2}$ can be obtained by recursively solving the first $k-1$ of the steady-state equations, and c can then be computed by using $\sum_{i=0}^\infty \pi_i = 1$.

If we let W_Q^* denote the amount of time that a customer spends in queue, then in exactly the same manner as in $G/M/1$ we can show that

$$W_Q^* = \begin{cases} 0, & \text{with probability } \sum_{i=0}^{k-1} \pi_i = 1 - \frac{c\beta}{1-\beta} \\ \text{Exp}(k\mu(1-\beta)), & \text{with probability } \sum_{i=k}^{\infty} \pi_i = \frac{c\beta}{1-\beta} \end{cases}$$

where $\text{Exp}(k\mu(1-\beta))$ is an exponential random variable with rate $k\mu(1-\beta)$.

8.9.4 The M/G/k Queue

In this section we consider the $M/G/k$ system in which customers arrive at a Poisson rate λ and are served by any of k servers, each of whom has the service distribution G. If we attempt to mimic the analysis presented in Section 8.5 for the $M/G/1$ system, then we would start with the basic identity

$$V = \lambda E[S] W_Q + \lambda E[S^2]/2 \tag{8.61}$$

and then attempt to derive a second equation relating V and W_Q.

Now if we consider an arbitrary arrival, then we have the following identity:

$$\text{work in system when customer arrives} = k \times \text{time customer spends in queue} + R \tag{8.62}$$

where R is the sum of the remaining service times of all other customers in service at the moment when our arrival enters service.

The foregoing follows because while the arrival is waiting in queue, work is being processed at a rate k per unit time (since all servers are busy). Thus, an amount of work $k \times$ time in queue is processed while he waits in queue. Now, all of this work was present when he arrived and in addition the remaining work on those still being served when he enters service was also present when he arrived—so we obtain Eq. (8.62). For an illustration, suppose that there are three servers all of whom are busy when the customer arrives. Suppose, in addition, that there are no other customers in the system and also that the remaining service times of the three people in service are 3, 6, and 7. Hence, the work seen by the arrival is $3 + 6 + 7 = 16$. Now the arrival will spend 3 time units in queue, and at the moment he enters service, the remaining times of the other two customers are $6 - 3 = 3$ and $7 - 3 = 4$. Hence, $R = 3 + 4 = 7$ and as a check of Eq. (8.62) we see that $16 = 3 \times 3 + 7$.

Taking expectations of Eq. (8.62) and using the fact that Poisson arrivals see time averages, we obtain

$$V = kW_Q + E[R]$$

which, along with Eq. (8.61), would enable us to solve for W_Q if we could compute $E[R]$. However there is no known method for computing $E[R]$ and in fact, there is no known exact formula for W_Q. The following approximation for W_Q was obtained in Reference 6 by using the foregoing approach and then approximating $E[R]$:

$$W_Q \approx \frac{\lambda^k E[S^2](E[S])^{k-1}}{2(k-1)!(k-\lambda E[S])^2\left[\sum_{n=0}^{k-1}\frac{(\lambda E[S])^n}{n!} + \frac{(\lambda E[S])^k}{(k-1)!(k-\lambda E[S])}\right]} \tag{8.63}$$

The preceding approximation has been shown to be quite close to W_Q when the service distribution is gamma. It is also exact when G is exponential.

Exercises

1. For the $M/M/1$ queue, compute
   (a) the expected number of arrivals during a service period and
   (b) the probability that no customers arrive during a service period.

   Hint: "Condition."

***2.** Machines in a factory break down at an exponential rate of six per hour. There is a single repairman who fixes machines at an exponential rate of eight per hour. The cost incurred in lost production when machines are out of service is \$10 per hour per machine. What is the average cost rate incurred due to failed machines?

3. The manager of a market can hire either Mary or Alice. Mary, who gives service at an exponential rate of 20 customers per hour, can be hired at a rate of \$3 per hour. Alice, who gives service at an exponential rate of 30 customers per hour, can be hired at a rate of \$$C$ per hour. The manager estimates that, on the average, each customer's time is worth \$1 per hour and should be accounted for in the model. Assume customers arrive at a Poisson rate of 10 per hour

(a) What is the average cost per hour if Mary is hired? If Alice is hired?

(b) Find C if the average cost per hour is the same for Mary and Alice.

4. In the $M/M/1$ system, derive P_0 by equating the rate at which customers arrive with the rate at which they depart.

5. Suppose customers arrive to a two server system according to a Poisson process with rate λ, and suppose that each arrival is, independently, sent either to server 1 with probability α or to server 2 with probability $1-\alpha$. Suppose the service time at server i is exponential with rate μ_i, $i = 1, 2$.

(a) Find $W(\alpha)$, the average amount of time a customer spends in the system.

(b) If $\lambda = 1$ and $\mu_i = i$, $i = 1, 2$, find the value of α that minimizes $W(\alpha)$.

6. Suppose that a customer of the $M/M/1$ system spends the amount of time $x > 0$ waiting in queue before entering service.

(a) Show that, conditional on the preceding, the number of other customers that were in the system when the customer arrived is distributed as $1 + P$, where P is a Poisson random variable with mean λ.

(b) Let W_Q^* denote the amount of time that an $M/M/1$ customer spends in queue. As a by-product of your analysis in part (a), show that

$$P\{W_Q^* \le x\} = \begin{cases} 1 - \frac{\lambda}{\mu} & \text{if } x = 0 \\ 1 - \frac{\lambda}{\mu} + \frac{\lambda}{\mu}(1 - e^{-(\mu-\lambda)x}) & \text{if } x > 0 \end{cases}$$

7. It follows from Exercise 6 that if, in the $M/M/1$ model, W_Q^* is the amount of time that a customer spends waiting in queue, then

$$W_Q^* = \begin{cases} 0, & \text{with probability } 1 - \lambda/\mu \\ \text{Exp}(\mu - \lambda), & \text{with probability } \lambda/\mu \end{cases}$$

where $\text{Exp}(\mu - \lambda)$ is an exponential random variable with rate $\mu - \lambda$. Using this, find $\text{Var}(W_Q^*)$.

***8.** Show that W is smaller in an $M/M/1$ model having arrivals at rate λ and service at rate 2μ than it is in a two-server $M/M/2$ model with arrivals at rate λ and with each server at rate μ. Can you give an intuitive explanation for this result? Would it also be true for W_Q?

9. Consider the $M/M/1$ queue with impatient customers model as presented in Example 8.9. Give your answers in terms of the limiting probabilities $P_n, n \geq 0$.
- **(a)** What is the average amount of time that a customer spends in queue.
- **(b)** If e_n denotes the probability that a customer who finds n others in the system upon arrival will be served, find $e_n, n \geq 0$.
- **(c)** Find the conditional probability that a served customer found n in the system upon arrival. That is, find $P(\text{arrival finds } n | \text{ arrival is served})$.
- **(d)** Find the average amount of time spent in queue by a customer that is served.
- **(e)** Find the average amount of time spent in queue by a customer that departs before entering service.

10. A facility produces items according to a Poisson process with rate λ. However, it has shelf space for only k items and so it shuts down production whenever k items are present. Customers arrive at the facility according to a Poisson process with rate μ. Each customer wants one item and will immediately depart either with the item or empty handed if there is no item available.
- **(a)** Find the proportion of customers that go away empty handed.
- **(b)** Find the average time that an item is on the shelf.
- **(c)** Find the average number of items on the shelf.

11. A group of n customers moves around among two servers. Upon completion of service, the served customer then joins the queue (or enters service if the server is free) at the other server. All service times are exponential with rate μ. Find the proportion of time that there are j customers at server 1, $j = 0, \ldots, n$.

12. A group of m customers frequents a single-server station in the following manner. When a customer arrives, he or she either enters service if the server is free or joins the queue otherwise. Upon completing service the customer departs the system, but then returns after an exponential time with rate θ. All service times are exponentially distributed with rate μ.
- **(a)** Find the average rate at which customers enter the station.
- **(b)** Find the average time that a customer spends in the station per visit.

***13.** Families arrive at a taxi stand according to a Poisson process with rate λ. An arriving family finding N other families waiting for a taxi does not wait. Taxis arrive at the taxi stand according to a Poisson process with rate μ. A taxi finding M other taxis waiting does not wait. Derive expressions for the following quantities.
- **(a)** The proportion of time there are no families waiting.
- **(b)** The proportion of time there are no taxis waiting.
- **(c)** The average amount of time that a family waits.
- **(d)** The average amount of time that a taxi waits.
- **(e)** The fraction of families that take taxis.

Now redo the problem if we assume that $N = M = \infty$ and that each family will only wait for an exponential time with rate α before seeking other transportation, and each taxi will only wait for an exponential time with rate β before departing without a fare.

14. Customers arrive to a single server system in accordance with a Poisson process with rate λ. Arrivals only enter if the server is free. Each customer is either a type 1 customer with probability p or a type 2 customer with probability $1 - p$. The time it takes to serve a type i customer is exponential with rate μ_i, $i = 1, 2$. Find the average amount of time an entering customer spends in the system.

15. Customers arrive to a two server system in accordance with a Poisson process with rate λ. Server 1 is the preferred server, and an arrival finding server 1 free enters service with 1; an arrival finding 1 busy but 2 free, enters service with 2. Arrivals finding both servers busy do not enter. A customer who is with server 2 at a moment when server 1 becomes free, immediately leaves server 2 and moves over to server 1. After completing a service (with either server) the customer departs. The service times at server i are exponential with rate μ_i, $i = 1, 2$.

(a) Define states and give the transition diagram.
(b) Find the long run proportion of time the system is in each state.
(c) Find the proportion of all arrivals that enter the system.
(d) Find the average time that an entering customer spends in the system.
(e) Find the proportion of entering customers that complete service with server 2.

16. Consider a 2 server system where customers arrive according to a Poisson process with rate λ, and where each arrival is sent to the server currently having the shortest queue. (If they have the same length queue then the choice is made at random.) The service time at either server is exponential with rate μ, where $\lambda < 2\mu$. For $n \geq 0$, say that the state is (n, n) if both servers currently have n customers, and say that the state is (n, m), $n < m$, if one of the servers has n customers and the other has m.

(a) Write down the balance equation equating the rate at which the process enters and leaves a state for state $(0, 0)$.
(b) Write down the balance equations equating the rate at which the process enters and leaves states of the form $(0, m)$, $m > 0$.
(c) Write down the balance equations for the states (n, n), $n > 0$.
(d) Write down the balance equations for the states (n, m), $0 < n < m$.
(e) In terms of the solution of the balance equations, find the average time a customer spends in the system.

17. Two customers move about among three servers. Upon completion of service at server i, the customer leaves that server and enters service at whichever of the other two servers is free. (Therefore, there are always two busy servers.) If the service times at server i are exponential with rate $\mu_i, i = 1, 2, 3$, what proportion of time is server i idle?

18. Consider a queueing system having two servers and no queue. There are two types of customers. Type 1 customers arrive according to a Poisson process having rate λ_1, and will enter the system if either server is free. The service time of a type 1 customer is exponential with rate μ_1. Type 2 customers arrive according to a Poisson process having rate λ_2. A type 2 customer requires the

simultaneous use of both servers; hence, a type 2 arrival will only enter the system if both servers are free. The time that it takes (the two servers) to serve a type 2 customer is exponential with rate μ_2. Once a service is completed on a customer, that customer departs the system.

(a) Define states to analyze the preceding model.

(b) Give the balance equations.

In terms of the solution of the balance equations, find

(c) the average amount of time an entering customer spends in the system;

(d) the fraction of served customers that are type 1.

19. Consider a sequential-service system consisting of two servers, A and B. Arriving customers will enter this system only if server A is free. If a customer does enter, then he is immediately served by server A. When his service by A is completed, he then goes to B if B is free, or if B is busy, he leaves the system. Upon completion of service at server B, the customer departs. Assume that the (Poisson) arrival rate is two customers an hour, and that A and B serve at respective (exponential) rates of four and two customers an hour.

(a) What proportion of customers enter the system?

(b) What proportion of entering customers receive service from B?

(c) What is the average number of customers in the system?

(d) What is the average amount of time that an entering customer spends in the system?

20. Customers arrive at a two-server system according to a Poisson process having rate $\lambda = 5$. An arrival finding server 1 free will begin service with that server. An arrival finding server 1 busy and server 2 free will enter service with server 2. An arrival finding both servers busy goes away. Once a customer is served by either server, he departs the system. The service times at server i are exponential with rates μ_i, where $\mu_1 = 4, \mu_2 = 2$.

(a) What is the average time an entering customer spends in the system?

(b) What proportion of time is server 2 busy?

21. Customers arrive at a two-server station in accordance with a Poisson process with a rate of two per hour. Arrivals finding server 1 free begin service with that server. Arrivals finding server 1 busy and server 2 free begin service with server 2. Arrivals finding both servers busy are lost. When a customer is served by server 1, she then either enters service with server 2 if 2 is free or departs the system if 2 is busy. A customer completing service at server 2 departs the system. The service times at server 1 and server 2 are exponential random variables with respective rates of four and six per hour.

(a) What fraction of customers do not enter the system?

(b) What is the average amount of time that an entering customer spends in the system?

(c) What fraction of entering customers receives service from server 1?

22. Arrivals to a three-server system are according to a Poisson process with rate λ. Arrivals finding server 1 free enter service with 1. Arrivals finding 1 busy but 2 free enter service with 2. Arrivals finding both 1 and 2 busy do not join the system. After completion of service at either 1 or 2 the customer will then

either go to server 3 if 3 is free or depart the system if 3 is busy. After service at 3 customers depart the system. The service times at i are exponential with rate μ_i, $i = 1, 2, 3$.

(a) Define states to analyze the above system.
(b) Give the balance equations.
(c) In terms of the solution of the balance equations, what is the average time that an entering customer spends in the system?
(d) Find the probability that a customer who arrives when the system is empty is served by server 3.

23. The economy alternates between good and bad periods. During good times customers arrive at a certain single-server queueing system in accordance with a Poisson process with rate λ_1, and during bad times they arrive in accordance with a Poisson process with rate λ_2. A good time period lasts for an exponentially distributed time with rate α_1, and a bad time period lasts for an exponential time with rate α_2. An arriving customer will only enter the queueing system if the server is free; an arrival finding the server busy goes away. All service times are exponential with rate μ.

(a) Define states so as to be able to analyze this system.
(b) Give a set of linear equations whose solution will yield the long-run proportion of time the system is in each state.
In terms of the solutions of the equations in part (b),
(c) what proportion of time is the system empty?
(d) what is the average rate at which customers enter the system?

24. There are two types of customers. Type 1 and 2 customers arrive in accordance with independent Poisson processes with respective rate λ_1 and λ_2. There are two servers. A type 1 arrival will enter service with server 1 if that server is free; if server 1 is busy and server 2 is free, then the type 1 arrival will enter service with server 2. If both servers are busy, then the type 1 arrival will go away. A type 2 customer can only be served by server 2; if server 2 is free when a type 2 customer arrives, then the customer enters service with that server. If server 2 is busy when a type 2 arrives, then that customer goes away. Once a customer is served by either server, he departs the system. Service times at server i are exponential with rate μ_i, $i = 1, 2$.
Suppose we want to find the average number of customers in the system.

(a) Define states.
(b) Give the balance equations. Do not attempt to solve them.
In terms of the long-run probabilities, what is
(c) the average number of customers in the system?
(d) the average time a customer spends in the system?

***25.** Suppose in Exercise 24 we want to find out the proportion of time there is a type 1 customer with server 2. In terms of the long-run probabilities given in Exercise 24, what is

(a) the rate at which a type 1 customer enters service with server 2?
(b) the rate at which a type 2 customer enters service with server 2?
(c) the fraction of server 2's customers that are type 1?

(d) the proportion of time that a type 1 customer is with server 2?

26. Customers arrive at a single-server station in accordance with a Poisson process with rate λ. All arrivals that find the server free immediately enter service. All service times are exponentially distributed with rate μ. An arrival that finds the server busy will leave the system and roam around "in orbit" for an exponential time with rate θ at which time it will then return. If the server is busy when an orbiting customer returns, then that customer returns to orbit for another exponential time with rate θ before returning again. An arrival that finds the server busy and N other customers in orbit will depart and not return. That is, N is the maximum number of customers in orbit.

(a) Define states.

(b) Give the balance equations.

In terms of the solution of the balance equations, find

(c) the proportion of all customers that are eventually served;

(d) the average time that a served customer spends waiting in orbit.

27. Consider the $M/M/1$ system in which customers arrive at rate λ and the server serves at rate μ. However, suppose that in any interval of length h in which the server is busy there is a probability $\alpha h + o(h)$ that the server will experience a breakdown, which causes the system to shut down. All customers that are in the system depart, and no additional arrivals are allowed to enter until the breakdown is fixed. The time to fix a breakdown is exponentially distributed with rate β.

(a) Define appropriate states.

(b) Give the balance equations.

In terms of the long-run probabilities,

(c) what is the average amount of time that an entering customer spends in the system?

(d) what proportion of entering customers complete their service?

(e) what proportion of customers arrive during a breakdown?

***28.** Reconsider Exercise 27, but this time suppose that a customer that is in the system when a breakdown occurs remains there while the server is being fixed. In addition, suppose that new arrivals during a breakdown period are allowed to enter the system. What is the average time a customer spends in the system?

29. Poisson (λ) arrivals join a queue in front of two parallel servers A and B, having exponential service rates μ_A and μ_B (see Fig. 8.4). When the system is empty, arrivals go into server A with probability α and into B with probability $1 - \alpha$. Otherwise, the head of the queue takes the first free server.

(a) Define states and set up the balance equations. Do not solve.

(b) In terms of the probabilities in part (a), what is the average number in the system? Average number of servers idle?

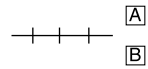

Figure 8.4

(c) In terms of the probabilities in part (a), what is the probability that an arbitrary arrival will get serviced in A?

30. In a queue with unlimited waiting space, arrivals are Poisson (parameter λ) and service times are exponentially distributed (parameter μ). However, the server waits until K people are present before beginning service on the first customer; thereafter, he services one at a time until all K units, and all subsequent arrivals, are serviced. The server is then "idle" until K new arrivals have occurred.

(a) Define an appropriate state space, draw the transition diagram, and set up the balance equations.

(b) In terms of the limiting probabilities, what is the average time a customer spends in queue?

(c) What conditions on λ and μ are necessary?

31. Consider a single-server exponential system in which ordinary customers arrive at a rate λ and have service rate μ. In addition, there is a special customer who has a service rate μ_1. Whenever this special customer arrives, she goes directly into service (if anyone else is in service, then this person is bumped back into queue). When the special customer is not being serviced, she spends an exponential amount of time (with mean $1/\theta$) out of the system.

(a) What is the average arrival rate of the special customer?

(b) Define an appropriate state space and set up balance equations.

(c) Find the probability that an ordinary customer is bumped n times.

***32.** Let D denote the time between successive departures in a stationary $M/M/1$ queue with $\lambda < \mu$. Show, by conditioning on whether or not a departure has left the system empty, that D is exponential with rate λ.

Hint: By conditioning on whether or not the departure has left the system empty we see that

$$D = \begin{cases} \text{Exponential}(\mu), & \text{with probability } \lambda/\mu \\ \text{Exponential}(\lambda) * \text{Exponential}(\mu), & \text{with probability } 1 - \lambda/\mu \end{cases}$$

where Exponential(λ) $*$ Exponential(μ) represents the sum of two independent exponential random variables having rates μ and λ. Now use moment-generating functions to show that D has the required distribution.

Note that the preceding does not prove that the departure process is Poisson. To prove this we need show not only that the interdeparture times are all exponential with rate λ, but also that they are independent.

33. Potential customers arrive to a single-server hair salon according to a Poisson process with rate λ. A potential customer who finds the server free enters the system; a potential customer who finds the server busy goes away. Each potential customer is type i with probability p_i, where $p_1 + p_2 + p_3 = 1$. Type 1 customers have their hair washed by the server; type 2 customers have their hair cut by the server; and type 3 customers have their hair first washed and then cut by the server. The time that it takes the server to wash hair is exponentially distributed with rate μ_1, and the time that it takes the server to cut hair is exponentially distributed with rate μ_2.

(a) Explain how this system can be analyzed with four states.
(b) Give the equations whose solution yields the proportion of time the system is in each state.
In terms of the solution of the equations of (b), find
(c) the proportion of time the server is cutting hair;
(d) the average arrival rate of entering customers.

34. For the tandem queue model verify that

$$P_{n,m} = (\lambda/\mu_1)^n(1 - \lambda/\mu_1)(\lambda/\mu_2)^m(1 - \lambda/\mu_2)$$

satisfies the balance Eqs. (8.15).

35. Consider a network of three stations with a single server at each station. Customers arrive at stations 1, 2, 3 in accordance with Poisson processes having respective rates 5, 10, and 15. The service times at the three stations are exponential with respective rates 10, 50, and 100. A customer completing service at station 1 is equally likely to (i) go to station 2, (ii) go to station 3, or (iii) leave the system. A customer departing service at station 2 always goes to station 3. A departure from service at station 3 is equally likely to either go to station 2 or leave the system.
(a) What is the average number of customers in the system (consisting of all three stations)?
(b) What is the average time a customer spends in the system?

36. Consider a closed queueing network consisting of two customers moving among two servers, and suppose that after each service completion the customer is equally likely to go to either server—that is, $P_{1,2} = P_{2,1} = \frac{1}{2}$. Let μ_i denote the exponential service rate at server $i, i = 1, 2$.
(a) Determine the average number of customers at each server.
(b) Determine the service completion rate for each server.

37. Explain how a Markov chain Monte Carlo simulation using the Gibbs sampler can be utilized to estimate
(a) the distribution of the amount of time spent at server j on a visit.

Hint: Use the arrival theorem.

(b) the proportion of time a customer is with server j (i.e., either in server j's queue or in service with j).

38. For open queueing networks
(a) state and prove the equivalent of the arrival theorem;
(b) derive an expression for the average amount of time a customer spends waiting in queues.

39. Customers arrive at a single-server station in accordance with a Poisson process having rate λ. Each customer has a value. The successive values of customers are independent and come from a uniform distribution on $(0, 1)$. The service time of a customer having value x is a random variable with mean $3 + 4x$ and variance 5.
(a) What is the average time a customer spends in the system?
(b) What is the average time a customer having value x spends in the system?

***40.** Compare the $M/G/1$ system for first-come, first-served queue discipline with one of last-come, first-served (for instance, in which units for service are taken from the top of a stack). Would you think that the queue size, waiting time, and busy-period distribution differ? What about their means? What if the queue discipline was always to choose at random among those waiting? Intuitively, which discipline would result in the smallest variance in the waiting time distribution?

41. In an $M/G/1$ queue,

(a) what proportion of departures leave behind 0 work?

(b) what is the average work in the system as seen by a departure?

42. For the $M/G/1$ queue, let X_n denote the number in the system left behind by the nth departure.

(a) If

$$X_{n+1} = \begin{cases} X_n - 1 + Y_n, & \text{if } X_n \geq 1 \\ Y_n, & \text{if } X_n = 0 \end{cases}$$

what does Y_n represent?

(b) Rewrite the preceding as

$$X_{n+1} = X_n - 1 + Y_n + \delta_n \tag{8.64}$$

where

$$\delta_n = \begin{cases} 1, & \text{if } X_n = 0 \\ 0, & \text{if } X_n \geq 1 \end{cases}$$

Take expectations and let $n \to \infty$ in Eq. (8.64) to obtain

$$E[\delta_\infty] = 1 - \lambda E[S]$$

(c) Square both sides of Eq. (8.64), take expectations, and then let $n \to \infty$ to obtain

$$E[X_\infty] = \frac{\lambda^2 E[S^2]}{2(1 - \lambda E[S])} + \lambda E[S]$$

(d) Argue that $E[X_\infty]$, the average number as seen by a departure, is equal to L.

***43.** Consider an $M/G/1$ system in which the first customer in a busy period has the service distribution G_1 and all others have distribution G_2. Let C denote the number of customers in a busy period, and let S denote the service time of a customer chosen at random.

Argue that

(a) $a_0 = P_0 = 1 - \lambda E[S]$.

(b) $E[S] = a_0 E[S_1] + (1 - a_0) E[S_2]$ where S_i has distribution G_i.

(c) Use (a) and (b) to show that $E[B]$, the expected length of a busy period, is given by

$$E[B] = \frac{E[S_1]}{1 - \lambda E[S_2]}$$

(d) Find $E[C]$.

44. Consider a $M/G/1$ system with $\lambda E[S] < 1$.

(a) Suppose that service is about to begin at a moment when there are n customers in the system.

(i) Argue that the additional time until there are only $n - 1$ customers in the system has the same distribution as a busy period.

(ii) What is the expected additional time until the system is empty?

(b) Suppose that the work in the system at some moment is A. We are interested in the expected additional time until the system is empty—call it $E[T]$. Let N denote the number of arrivals during the first A units of time.

(i) Compute $E[T|N]$.

(ii) Compute $E[T]$.

45. Carloads of customers arrive at a single-server station in accordance with a Poisson process with rate 4 per hour. The service times are exponentially distributed with rate 20 per hour. If each carload contains either 1, 2, or 3 customers with respective probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, compute the average customer delay in queue.

46. In the two-class priority queueing model of Section 8.6.2, what is W_Q? Show that W_Q is less than it would be under FIFO if $E[S_1] < E[S_2]$ and greater than under FIFO if $E[S_1] > E[S_2]$.

47. In a two-class priority queueing model suppose that a cost of C_i per unit time is incurred for each type i customer that waits in queue, $i = 1, 2$. Show that type 1 customers should be given priority over type 2 (as opposed to the reverse) if

$$\frac{E[S_1]}{C_1} < \frac{E[S_2]}{C_2}$$

48. Consider the priority queueing model of Section 8.6.2 but now suppose that if a type 2 customer is being served when a type 1 arrives then the type 2 customer is bumped out of service. This is called the preemptive case. Suppose that when a bumped type 2 customer goes back in service his service begins at the point where it left off when he was bumped.

(a) Argue that the work in the system at any time is the same as in the non-preemptive case.

(b) Derive W_Q^1.

Hint: How do type 2 customers affect type 1s?

(c) Why is it not true that

$$V_Q^2 = \lambda_2 E[S_2] W_Q^2$$

(d) Argue that the work seen by a type 2 arrival is the same as in the nonpreemptive case, and so

$$W_Q^2 = W_Q^2(\text{nonpreemptive}) + E[\text{extra time}]$$

where the extra time is due to the fact that he may be bumped.

(e) Let N denote the number of times a type 2 customer is bumped. Why is

$$E[\text{extra time}|N] = \frac{NE[S_1]}{1 - \lambda_1 E[S_1]}$$

Hint: When a type 2 is bumped, relate the time until he gets back in service to a "busy period."

(f) Let S_2 denote the service time of a type 2. What is $E[N|S_2]$?

(g) Combine the preceding to obtain

$$W_Q^2 = W_Q^2(\text{nonpreemptive}) + \frac{\lambda_1 E[S_1]E[S_2]}{1 - \lambda_1 E[S_1]}$$

***49.** Calculate explicitly (not in terms of limiting probabilities) the average time a customer spends in the system in Exercise 28.

50. In the $G/M/1$ model if G is exponential with rate λ show that $\beta = \lambda/\mu$.

51. In the k server Erlang loss model, suppose that $\lambda = 1$ and $E[S] = 4$. Find L if $P_k = .2$.

52. Verify the formula given for the P_i of the $M/M/k$.

53. In the Erlang loss system suppose the Poisson arrival rate is $\lambda = 2$, and suppose there are three servers, each of whom has a service distribution that is uniformly distributed over $(0, 2)$. What proportion of potential customers is lost?

54. In the $M/M/k$ system,

(a) what is the probability that a customer will have to wait in queue?

(b) determine L and W.

55. Verify the formula for the distribution of W_Q^* given for the $G/M/k$ model.

***56.** Consider a system where the interarrival times have an arbitrary distribution F, and there is a single server whose service distribution is G. Let D_n denote the amount of time the nth customer spends waiting in queue. Interpret S_n, T_n so that

$$D_{n+1} = \begin{cases} D_n + S_n - T_n, & \text{if } D_n + S_n - T_n \geq 0 \\ 0, & \text{if } D_n + S_n - T_n < 0 \end{cases}$$

57. Consider a model in which the interarrival times have an arbitrary distribution F, and there are k servers each having service distribution G. What condition on F and G do you think would be necessary for there to exist limiting probabilities?

Reliability Theory

9

9.1 Introduction

Reliability theory is concerned with determining the probability that a system, possibly consisting of many components, will function. We shall suppose that whether or not the system functions is determined solely from a knowledge of which components are functioning. For instance, a *series* system will function if and only if all of its components are functioning, while a *parallel* system will function if and only if at least one of its components is functioning. In Section 9.2, we explore the possible ways in which the functioning of the system may depend upon the functioning of its components. In Section 9.3, we suppose that each component will function with some known probability (independently of each other) and show how to obtain the probability that the system will function. As this probability often is difficult to explicitly compute, we also present useful upper and lower bounds in Section 9.4. In Section 9.5 we look at a system dynamically over time by supposing that each component initially functions and does so for a random length of time at which it fails. We then discuss the relationship between the distribution of the amount of time that a system functions and the distributions of the component lifetimes. In particular, it turns out that if the amount of time that a component functions has an *increasing failure rate on the average* (IFRA) distribution, then so does the distribution of system lifetime. In Section 9.6 we consider the problem of obtaining the mean lifetime of a system. In the final section we analyze the system when failed components are subjected to repair.

9.2 Structure Functions

Consider a system consisting of n components, and suppose that each component is either functioning or has failed. To indicate whether or not the ith component is functioning, we define the indicator variable x_i by

$$x_i = \begin{cases} 1, & \text{if the } i\text{th component is functioning} \\ 0, & \text{if the } i\text{th component has failed} \end{cases}$$

The vector $\mathbf{x} = (x_1, \ldots, x_n)$ is called the *state vector*. It indicates which of the components are functioning and which have failed.

We further suppose that whether or not the system as a whole is functioning is completely determined by the state vector $\mathbf{x}$. Specifically, it is supposed that there exists a function $\phi(\mathbf{x})$ such that

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if the system is functioning when the state vector is } \mathbf{x} \\ 0, & \text{if the system has failed when the state vector is } \mathbf{x} \end{cases}$$

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00014-7

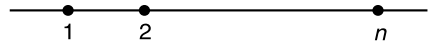

Figure 9.1 A series system.

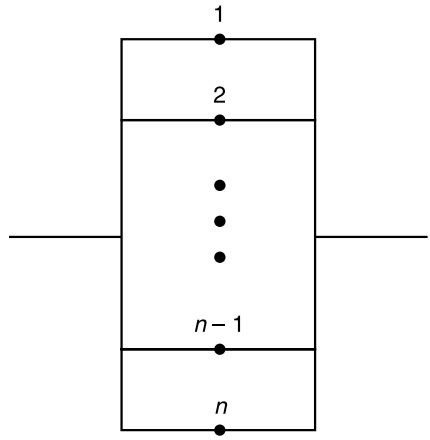

Figure 9.2 A parallel system.

The function $\phi(\mathbf{x})$ is called the *structure function* of the system.

Example 9.1 (The Series Structure)**.** A series system functions if and only if all of its components are functioning. Hence, its structure function is given by

$$\phi(\mathbf{x}) = \min(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i$$

We shall find it useful to represent the structure of a system in terms of a diagram. The relevant diagram for the series structure is shown in Fig. 9.1. The idea is that if a signal is initiated at the left end of the diagram then in order for it to successfully reach the right end, it must pass through all of the components; hence, they must all be functioning. ■

Example 9.2 (The Parallel Structure)**.** A parallel system functions if and only if at least one of its components is functioning. Hence, its structure function is given by

$$\phi(\mathbf{x}) = \max(x_1, \ldots, x_n)$$

A parallel structure may be pictorially illustrated by Fig. 9.2. This follows since a signal at the left end can successfully reach the right end as long as at least one component is functioning. ■

Example 9.3 (The k-out-of-n Structure)**.** The series and parallel systems are both special cases of a k-out-of-n system. Such a system functions if and only if at least

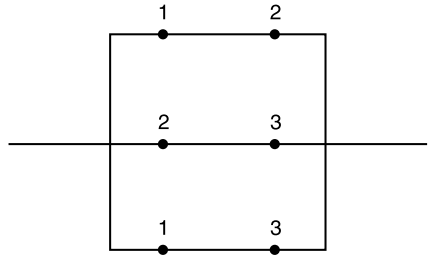

Figure 9.3 A two-out-of-three system.

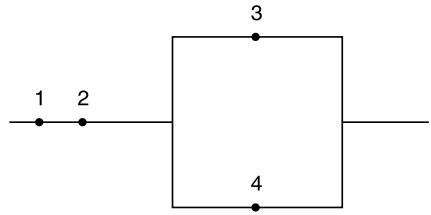

Figure 9.4

k of the n components are functioning. As $\sum_{i=1}^{n} x_i$ equals the number of functioning components, the structure function of a k-out-of-n system is given by

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^{n} x_i \geqslant k \\ 0, & \text{if } \sum_{i=1}^{n} x_i < k \end{cases}$$

Series and parallel systems are respectively n-out-of-n and 1-out-of-n systems.

The two-out-of-three system may be diagrammed as shown in Fig. 9.3. ■

Example 9.4 (A Four-Component Structure)**.** Consider a system consisting of four components, and suppose that the system functions if and only if components 1 and 2 both function and at least one of components 3 and 4 function. Its structure function is given by

$$\phi(\mathbf{x}) = x_1 x_2 \max(x_3, x_4)$$

Pictorially, the system is as shown in Fig. 9.4. A useful identity, easily checked, is that for binary variables,[1] $x_i, i = 1, \ldots, n$,

$$\max(x_1, \ldots, x_n) = 1 - \prod_{i=1}^{n} (1 - x_i)$$

[1] A binary variable is one that assumes either the value 0 or 1.

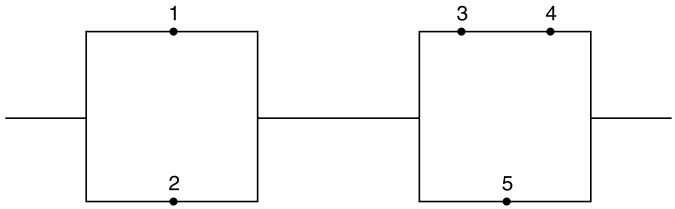

Figure 9.5

When $n = 2$, this yields

$$\max(x_1, x_2) = 1 - (1 - x_1)(1 - x_2) = x_1 + x_2 - x_1x_2$$

Hence, the structure function in the example may be written as

$$\phi(\mathbf{x}) = x_1x_2(x_3 + x_4 - x_3x_4)$$ ■

It is natural to assume that replacing a failed component by a functioning one will never lead to a deterioration of the system. In other words, it is natural to assume that the structure function $\phi(\mathbf{x})$ is an increasing function of $\mathbf{x}$, that is, if $x_i \leqslant y_i, i = 1, \ldots, n$, then $\phi(\mathbf{x}) \leqslant \phi(\mathbf{y})$. Such an assumption shall be made in this chapter and the system will be called *monotone*.

9.2.1 Minimal Path and Minimal Cut Sets

In this section we show how any system can be represented both as a series arrangement of parallel structures and as a parallel arrangement of series structures. As a preliminary, we need the following concepts.

A state vector $\mathbf{x}$ is called a *path vector* if $\phi(\mathbf{x}) = 1$. If, in addition, $\phi(\mathbf{y}) = 0$ for all $\mathbf{y} < \mathbf{x}$, then $\mathbf{x}$ is said to be a *minimal path vector*.[2] If $\mathbf{x}$ is a minimal path vector, then the set $A = \{i : x_i = 1\}$ is called a *minimal path set*. In other words, a minimal path set is a minimal set of components whose functioning ensures the functioning of the system.

Example 9.5. Consider a five-component system whose structure is illustrated by Fig. 9.5. Its structure function equals

$$\begin{aligned}\phi(\mathbf{x}) &= \max(x_1, x_2)\ \max(x_3x_4, x_5)\\ &= (x_1 + x_2 - x_1x_2)(x_3x_4 + x_5 - x_3x_4x_5)\end{aligned}$$

There are four minimal path sets, namely, $\{1, 3, 4\}$, $\{2, 3, 4\}$, $\{1, 5\}$, $\{2, 5\}$. ■

Example 9.6. In a k-out-of-n system, there are $\binom{n}{k}$ minimal path sets, namely, all of the sets consisting of exactly k components. ■

[2] We say that $\mathbf{y} < \mathbf{x}$ if $y_i \leqslant x_i, i = 1, \ldots, n$, with $y_i < x_i$ for some i.

Let $A_1, \ldots, A_s$ denote the minimal path sets of a given system. We define $\alpha_j(\mathbf{x})$, the indicator function of the jth minimal path set, by

$$\begin{aligned}\alpha_j(\mathbf{x}) &= \begin{cases} 1, & \text{if all the components of } A_j \text{ are functioning} \\ 0, & \text{otherwise} \end{cases} \\ &= \prod_{i \in A_j} x_i\end{aligned}$$

By definition, it follows that the system will function if all the components of at least one minimal path set are functioning; that is, if $\alpha_j(\mathbf{x}) = 1$ for some j. On the other hand, if the system functions, then the set of functioning components must include a minimal path set. Therefore, *a system will function if and only if all the components of at least one minimal path set are functioning*. Hence,

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \alpha_j(\mathbf{x}) = 1 \text{ for some } j \\ 0, & \text{if } \alpha_j(\mathbf{x}) = 0 \text{ for all } j \end{cases}$$

or equivalently,

$$\begin{aligned}\phi(\mathbf{x}) &= \max_j \alpha_j(\mathbf{x}) \\ &= \max_j \prod_{i \in A_j} x_i \end{aligned} \tag{9.1}$$

Since $\alpha_j(\mathbf{x})$ is a series structure function of the components of the jth minimal path set, Eq. (9.1) expresses an arbitrary system as a parallel arrangement of series systems.

Example 9.7. Consider the system of Example 9.5. Because its minimal path sets are $A_1 = \{1, 3, 4\}$, $A_2 = \{2, 3, 4\}$, $A_3 = \{1, 5\}$, and $A_4 = \{2, 5\}$, we have by Eq. (9.1) that

$$\begin{aligned}\phi(\mathbf{x}) &= \max\{x_1x_3x_4, x_2x_3x_4, x_1x_5, x_2x_5\} \\ &= 1 - (1 - x_1x_3x_4)(1 - x_2x_3x_4)(1 - x_1x_5)(1 - x_2x_5)\end{aligned}$$

You should verify that this equals the value of $\phi(\mathbf{x})$ given in Example 9.5. (Make use of the fact that, since x_i equals 0 or 1, $x_i^2 = x_i$.) This representation may be pictured as shown in Fig. 9.6. ■

Example 9.8. The system whose structure is as pictured in Fig. 9.7 is called the *bridge system*. Its minimal path sets are $\{1, 4\}$, $\{1, 3, 5\}$, $\{2, 5\}$, and $\{2, 3, 4\}$. Hence, by Eq. (9.1), its structure function may be expressed as

$$\begin{aligned}\phi(\mathbf{x}) &= \max\{x_1x_4, x_1x_3x_5, x_2x_5, x_2x_3x_4\} \\ &= 1 - (1 - x_1x_4)(1 - x_1x_3x_5)(1 - x_2x_5)(1 - x_2x_3x_4)\end{aligned}$$

This representation $\phi(\mathbf{x})$ is diagrammed as shown in Fig. 9.8. ■

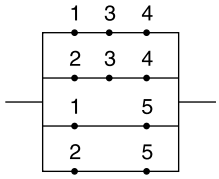

Figure 9.6

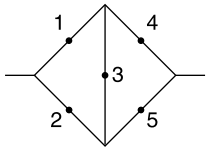

Figure 9.7 The bridge system.

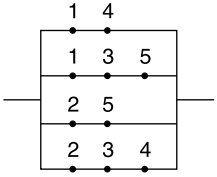

Figure 9.8

A state vector $\mathbf{x}$ is called a *cut vector* if $\phi(\mathbf{x}) = 0$. If, in addition, $\phi(\mathbf{y}) = 1$ for all $\mathbf{y} > \mathbf{x}$, then $\mathbf{x}$ is said to be a *minimal cut vector*. If $\mathbf{x}$ is a minimal cut vector, then the set $C = \{i : x_i = 0\}$ is called a *minimal cut set*. In other words, a minimal cut set is a minimal set of components whose failure ensures the failure of the system.

Let $C_1, \ldots, C_k$ denote the minimal cut sets of a given system. We define $\beta_j(\mathbf{x})$, the indicator function of the jth minimal cut set, by

$$\beta_j(\mathbf{x}) = \begin{cases} 1, & \text{if at least one component of the } j\text{th minimal} \\ & \text{cut set is functioning} \\ 0, & \text{if all of the components of the } j\text{th minimal} \\ & \text{cut set are not functioning} \end{cases}$$

$$= \max_{i \in C_j} x_i$$

Since a system is not functioning if and only if all the components of at least one minimal cut set are not functioning, it follows that

$$\phi(\mathbf{x}) = \prod_{j=1}^{k} \beta_j(\mathbf{x}) = \prod_{j=1}^{k} \max_{i \in C_j} x_i \tag{9.2}$$

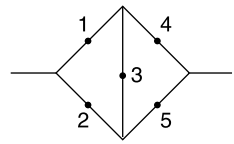

Figure 9.9

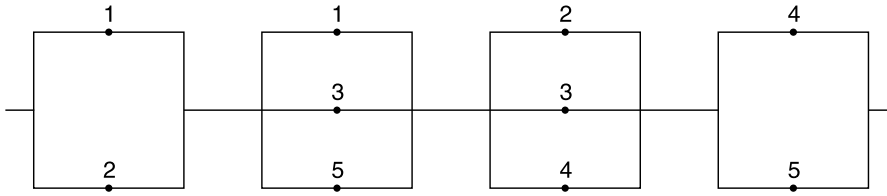

Figure 9.10 Minimal cut representation of the bridge system.

Since $\beta_j(\mathbf{x})$ is a parallel structure function of the components of the jth minimal cut set, Eq. (9.2) represents an arbitrary system as a series arrangement of parallel systems.

Example 9.9. The minimal cut sets of the bridge structure shown in Fig. 9.9 are $\{1, 2\}$, $\{1, 3, 5\}$, $\{2, 3, 4\}$, and $\{4, 5\}$. Hence, from Eq. (9.2), we may express $\phi(\mathbf{x})$ by

$$\begin{aligned}\phi(\mathbf{x}) &= \max(x_1, x_2)\ \max(x_1, x_3, x_5)\ \max(x_2, x_3, x_4)\ \max(x_4, x_5)\\ &= [1-(1-x_1)(1-x_2)][1-(1-x_1)(1-x_3)(1-x_5)]\\ &\quad \times\ [1-(1-x_2)(1-x_3)(1-x_4)][1-(1-x_4)(1-x_5)]\end{aligned}$$

This representation of $\phi(\mathbf{x})$ is pictorially expressed as Fig. 9.10. ■

9.3 Reliability of Systems of Independent Components

In this section, we suppose that X_i, the state of the ith component, is a random variable such that

$$P\{X_i = 1\} = p_i = 1 - P\{X_i = 0\}$$

The value p_i, which equals the probability that the ith component is functioning, is called the *reliability* of the ith component. If we define r by

$$r = P\{\phi(\mathbf{X}) = 1\}, \quad \text{where } \mathbf{X} = (X_1, \ldots, X_n)$$

then r is called the *reliability* of the system. When the components, that is, the random variables $X_i, i = 1, \ldots, n$, are independent, we may express r as a function of the

component reliabilities. That is,

$$r = r(\mathbf{p}), \quad \text{where } \mathbf{p} = (p_1, \ldots, p_n)$$

The function $r(\mathbf{p})$ is called the *reliability function*. We shall assume throughout the remainder of this chapter that the components are independent.

Example 9.10 (The Series System)**.** The reliability function of the series system of n independent components is given by

$$\begin{aligned} r(\mathbf{p}) &= P\{\phi(\mathbf{X}) = 1\} \\ &= P\{X_i = 1 \text{ for all } i = 1, \ldots, n\} \\ &= \prod_{i=1}^{n} p_i \end{aligned}$$

■

Example 9.11 (The Parallel System)**.** The reliability function of the parallel system of n independent components is given by

$$\begin{aligned} r(\mathbf{p}) &= P\{\phi(\mathbf{X}) = 1\} \\ &= P\{X_i = 1 \text{ for some } i = 1, \ldots, n\} \\ &= 1 - P\{X_i = 0 \text{ for all } i = 1, \ldots, n\} \\ &= 1 - \prod_{i=1}^{n} (1 - p_i) \end{aligned}$$

■

Example 9.12 (The k-out-of-n System with Equal Probabilities)**.** Consider a k-out-of-n system. If $p_i = p$ for all $i = 1, \ldots, n$, then the reliability function is given by

$$\begin{aligned} r(p, \ldots, p) &= P\{\phi(\mathbf{X}) = 1\} \\ &= P\left\{\sum_{i=1}^{n} X_i \geqslant k\right\} \\ &= \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

■

Example 9.13 (The Two-out-of-Three System)**.** The reliability function of a two-out-of-three system is given by

$$\begin{aligned} r(\mathbf{p}) &= P\{\phi(\mathbf{X}) = 1\} \\ &= P\{\mathbf{X} = (1,1,1)\} + P\{\mathbf{X} = (1,1,0)\} \\ &\quad + P\{\mathbf{X} = (1,0,1)\} + P\{\mathbf{X} = (0,1,1)\} \\ &= p_1 p_2 p_3 + p_1 p_2 (1 - p_3) + p_1 (1 - p_2) p_3 + (1 - p_1) p_2 p_3 \\ &= p_1 p_2 + p_1 p_3 + p_2 p_3 - 2 p_1 p_2 p_3 \end{aligned}$$

■

Example 9.14 (The Three-out-of-Four System)**.** The reliability function of a three-out-of-four system is given by

$$\begin{aligned} r(\mathbf{p}) &= P\{\mathbf{X} = (1,1,1,1)\} + P\{\mathbf{X} = (1,1,1,0)\} + P\{\mathbf{X} = (1,1,0,1)\} \\ &\quad + P\{\mathbf{X} = (1,0,1,1)\} + P\{\mathbf{X} = (0,1,1,1)\} \\ &= p_1p_2p_3p_4 + p_1p_2p_3(1-p_4) + p_1p_2(1-p_3)p_4 \\ &\quad + p_1(1-p_2)p_3p_4 + (1-p_1)p_2p_3p_4 \\ &= p_1p_2p_3 + p_1p_2p_4 + p_1p_3p_4 + p_2p_3p_4 - 3p_1p_2p_3p_4 \end{aligned}$$

■

Example 9.15 (A Five-Component System)**.** Consider a five-component system that functions if and only if component 1, component 2, and at least one of the remaining components function. Its reliability function is given by

$$\begin{aligned} r(\mathbf{p}) &= P\{X_1 = 1, X_2 = 1, \max(X_3, X_4, X_5) = 1\} \\ &= P\{X_1 = 1\}P\{X_2 = 1\}P\{\max(X_3, X_4, X_5) = 1\} \\ &= p_1p_2[1 - (1-p_3)(1-p_4)(1-p_5)] \end{aligned}$$

■

Since $\phi(\mathbf{X})$ is a 0–1 (that is, a Bernoulli) random variable, we may also compute $r(\mathbf{p})$ by taking its expectation. That is,

$$\begin{aligned} r(\mathbf{p}) &= P\{\phi(\mathbf{X}) = 1\} \\ &= E[\phi(\mathbf{X})] \end{aligned}$$

Example 9.16 (A Four-Component System)**.** A four-component system that functions when both components 1 and 4, and at least one of the other components function has its structure function given by

$$\phi(\mathbf{x}) = x_1x_4 \max(x_2, x_3)$$

Hence,

$$\begin{aligned} r(\mathbf{p}) &= E[\phi(\mathbf{X})] \\ &= E[X_1X_4(1 - (1-X_2)(1-X_3))] \\ &= p_1p_4[1 - (1-p_2)(1-p_3)] \end{aligned}$$

■

An important and intuitive property of the reliability function $r(\mathbf{p})$ is given by the following proposition.

Proposition 9.1. *If $r(\mathbf{p})$ is the reliability function of a system of independent components, then $r(\mathbf{p})$ is an increasing function of* $\mathbf{p}$.

Proof. By conditioning on X_i and using the independence of the components, we obtain

$$r(\mathbf{p}) = E[\phi(\mathbf{X})]$$

$$
\begin{aligned}
&= p_i E[\phi(\mathbf{X}) \mid X_i = 1] + (1 - p_i) E[\phi(\mathbf{X}) \mid X_i = 0] \\
&= p_i E[\phi(1_i, \mathbf{X})] + (1 - p_i) E[\phi(0_i, \mathbf{X})]
\end{aligned}
$$

where

$$
\begin{aligned}
(1_i, \mathbf{X}) &= (X_1, \ldots, X_{i-1}, 1, X_{i+1}, \ldots, X_n), \\
(0_i, \mathbf{X}) &= (X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)
\end{aligned}
$$

Thus,

$$
r(\mathbf{p}) = p_i E[\phi(1_i, \mathbf{X}) - \phi(0_i, \mathbf{X})] + E[\phi(0_i, \mathbf{X})]
$$

However, since ϕ is an increasing function, it follows that

$$
E[\phi(1_i, \mathbf{X}) - \phi(0_i, \mathbf{X})] \geqslant 0
$$

and so the preceding is increasing in p_i for all i. Hence, the result is proven. ■

Let us now consider the following situation: A system consisting of n different components is to be built from a stockpile containing exactly two of each type of component. How should we use the stockpile so as to maximize our probability of attaining a functioning system? In particular, should we build two separate systems, in which case the probability of attaining a functioning one would be

$$
\begin{aligned}
&P\{\text{at least one of the two systems function}\} \\
&\quad = 1 - P\{\text{neither of the systems function}\} \\
&\quad = 1 - [(1 - r(\mathbf{p}))(1 - r(\mathbf{p}'))]
\end{aligned}
$$

where $p_i\,(p_i')$ is the probability that the first (second) number i component functions; or should we build a single system whose ith component functions if at least one of the number i components functions? In this latter case, the probability that the system will function equals

$$
r[\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')]
$$

since $1 - (1 - p_i)(1 - p_i')$ equals the probability that the ith component in the single system will function.[3] We now show that replication at the component level is more effective than replication at the system level.

Theorem 9.1. *For any reliability function r and vectors $\mathbf{p}, \mathbf{p}'$,*

$$
r[\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')] \geqslant 1 - [1 - r(\mathbf{p})][1 - r(\mathbf{p}')]
$$

[3] Notation: If $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, then $\mathbf{xy} = (x_1 y_1, \ldots, x_n y_n)$. Also, $\max(\mathbf{x}, \mathbf{y}) = (\max(x_1, y_1), \ldots, \max(x_n, y_n))$ and $\min(\mathbf{x}, \mathbf{y}) = (\min(x_1, y_1), \ldots, \min(x_n, y_n))$.

Proof. Let $X_1, \ldots, X_n, X'_1, \ldots, X'_n$ be mutually independent 0–1 random variables with

$$p_i = P\{X_i = 1\}, \quad p'_i = P\{X'_i = 1\}$$

Since $P\{\max(X_i, X'_i) = 1\} = 1 - (1 - p_i)(1 - p'_i)$, it follows that

$$r[\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')] = E[\phi[\max(\mathbf{X}, \mathbf{X}')]]$$

However, by the monotonicity of ϕ, we have that $\phi[\max(\mathbf{X}, \mathbf{X}')]$ is greater than or equal to both $\phi(\mathbf{X})$ and $\phi(\mathbf{X}')$ and hence is at least as large as $\max[\phi(\mathbf{X}), \phi(\mathbf{X}')]$. Hence, from the preceding we have

$$\begin{aligned} r[\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')] &\geqslant E[\max(\phi(\mathbf{X}), \phi(\mathbf{X}'))] \\ &= P\{\max[\phi(\mathbf{X}), \phi(\mathbf{X}')] = 1\} \\ &= 1 - P\{\phi(\mathbf{X}) = 0, \phi(\mathbf{X}') = 0\} \\ &= 1 - [1 - r(\mathbf{p})][1 - r(\mathbf{p}')] \end{aligned}$$

where the first equality follows from the fact that $\max[\phi(\mathbf{X}), \phi(\mathbf{X}')]$ is a 0–1 random variable and hence its expectation equals the probability that it equals 1. ■

As an illustration of the preceding theorem, suppose that we want to build a series system of two different types of components from a stockpile consisting of two of each of the kinds of components. Suppose that the reliability of each component is $\frac{1}{2}$. If we use the stockpile to build two separate systems, then the probability of attaining a working system is

$$1 - \left(\tfrac{3}{4}\right)^2 = \tfrac{7}{16}$$

while if we build a single system, replicating components, then the probability of attaining a working system is

$$\left(\tfrac{3}{4}\right)^2 = \tfrac{9}{16}$$

Hence, replicating components leads to a higher reliability than replicating systems (as, of course, it must by Theorem 9.1).

9.4 Bounds on the Reliability Function

Consider the bridge system of Example 9.8, which is represented by Fig. 9.11. Using the minimal path representation, we have

$$\phi(\mathbf{x}) = 1 - (1 - x_1x_4)(1 - x_1x_3x_5)(1 - x_2x_5)(1 - x_2x_3x_4)$$

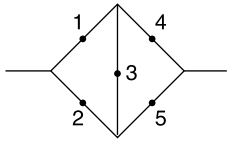

Figure 9.11

Hence,

$$r(\mathbf{p}) = 1 - E[(1 - X_1X_4)(1 - X_1X_3X_5)(1 - X_2X_5)(1 - X_2X_3X_4)]$$

However, since the minimal path sets overlap (that is, they have components in common), the random variables $(1-X_1X_4)$, $(1-X_1X_3X_5)$, $(1-X_2X_5)$, and $(1-X_2X_3X_4)$ are not independent, and thus the expected value of their product is not equal to the product of their expected values. Therefore, in order to compute $r(\mathbf{p})$, we must first multiply the four random variables and then take the expected value. Doing so, using that $X_i^2 = X_i$, we obtain

$$\begin{aligned} r(\mathbf{p}) &= E[X_1X_4 + X_2X_5 + X_1X_3X_5 + X_2X_3X_4 - X_1X_2X_3X_4 \\ &\quad - X_1X_2X_3X_5 - X_1X_2X_4X_5 - X_1X_3X_4X_5 - X_2X_3X_4X_5 \\ &\quad + 2X_1X_2X_3X_4X_5] \\ &= p_1p_4 + p_2p_5 + p_1p_3p_5 + p_2p_3p_4 - p_1p_2p_3p_4 - p_1p_2p_3p_5 \\ &\quad - p_1p_2p_4p_5 - p_1p_3p_4p_5 - p_2p_3p_4p_5 + 2p_1p_2p_3p_4p_5 \end{aligned}$$

As can be seen by the preceding example, it is often quite tedious to evaluate $r(\mathbf{p})$, and thus it would be useful if we had a simple way of obtaining bounds. We now consider two methods for this.

9.4.1 Method of Inclusion and Exclusion

The following is a well-known formula for the probability of the union of the events $E_1, E_2, \ldots, E_n$:

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{i=1}^{n} P(E_i) - \sum\sum_{i<j} P(E_iE_j) + \sum\sum\sum_{i<j<k} P(E_iE_jE_k) \\ &\quad - \cdots + (-1)^{n+1} P(E_1E_2\cdots E_n) \end{aligned} \tag{9.3}$$

A result, not as well known, is the following set of inequalities:

$$P\left(\bigcup_{1}^{n} E_i\right) \leqslant \sum_{i=1}^{n} P(E_i),$$

$$P\left(\bigcup_{1}^{n} E_i\right) \geqslant \sum_{i} P(E_i) - \sum_{i<j} P(E_iE_j),$$

$$P\left(\bigcup_1^n E_i\right) \leqslant \sum_i P(E_i) - \sum_{i<j}\sum P(E_iE_j) + \sum_{i<j<k}\sum\sum P(E_iE_jE_k),$$
$$\geqslant \cdots$$
$$\leqslant \cdots \tag{9.4}$$

where the inequality always changes direction as we add an additional term of the expansion of $P(\bigcup_{i=1}^n E_i)$.

Eq. (9.3) is usually proven by induction on the number of events. However, let us now present another approach that will not only prove Eq. (9.3) but also establish Inequalities (9.4).

To begin, define the indicator variables I_j, $j = 1, \ldots, n$, by

$$I_j = \begin{cases} 1, & \text{if } E_j \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

Letting

$$N = \sum_{j=1}^n I_j$$

then N denotes the number of the E_j, $1 \leqslant j \leqslant n$, that occur. Also, let

$$I = \begin{cases} 1, & \text{if } N > 0 \\ 0, & \text{if } N = 0 \end{cases}$$

Then, as

$$1 - I = (1-1)^N$$

we obtain, upon application of the binomial theorem, that

$$1 - I = \sum_{i=0}^N \binom{N}{i} (-1)^i$$

or

$$I = N - \binom{N}{2} + \binom{N}{3} - \cdots \pm \binom{N}{N} \tag{9.5}$$

We now make use of the following combinatorial identity (which is easily established by induction on i):

$$\binom{n}{i} - \binom{n}{i+1} + \cdots \pm \binom{n}{n} = \binom{n-1}{i-1} \geqslant 0, \quad i \leqslant n$$

The preceding thus implies that

$$\binom{N}{i} - \binom{N}{i+1} + \cdots \pm \binom{N}{N} \geqslant 0 \tag{9.6}$$

From Eqs. (9.5) and (9.6) we obtain

$$\begin{aligned} &I \leqslant N, && \text{by letting } i = 2 \text{ in (9.6)} \\ &I \geqslant N - \binom{N}{2}, && \text{by letting } i = 3 \text{ in (9.6)} \\ &I \leqslant N - \binom{N}{2} + \binom{N}{3}, \\ &\quad \vdots \end{aligned} \tag{9.7}$$

and so on. Now, since $N \leqslant n$ and $\binom{m}{i} = 0$ whenever $i > m$, we can rewrite Eq. (9.5) as

$$I = \sum_{i=1}^{n} \binom{N}{i} (-1)^{i+1} \tag{9.8}$$

Eq. (9.3) and Inequalities (9.4) now follow upon taking expectations of (9.7) and (9.8). This is the case since

$$E[I] = P\{N > 0\} = P\{\text{at least one of the } E_j \text{ occurs}\} = P\left(\bigcup_{1}^{n} E_j\right),$$

$$E[N] = E\left[\sum_{j=1}^{n} I_j\right] = \sum_{j=1}^{n} P(E_j)$$

Also,

$$\begin{aligned} E\left[\binom{N}{2}\right] &= E[\text{number of pairs of the } E_j \text{ that occur}] \\ &= E\left[\sum\sum_{i<j} I_i I_j\right] \\ &= \sum\sum_{i<j} P(E_i E_j) \end{aligned}$$

and, in general

$$E\left[\binom{N}{i}\right] = E[\text{number of sets of size } i \text{ that occur}]$$

$$= E\left[\sum_{j_1<j_2<\cdots<j_i}\sum\sum I_{j_1} I_{j_2}\cdots I_{j_i}\right]$$
$$= \sum_{j_1<j_2<\cdots<j_i}\sum\sum P(E_{j_1} E_{j_2}\cdots E_{j_i})$$

The bounds expressed in (9.4) are commonly called the *inclusion–exclusion bounds*. To apply them in order to obtain bounds on the reliability function, let $A_1 A_2, \ldots, A_s$ denote the minimal path sets of a given structure ϕ, and define the events $E_1, E_2, \ldots, E_s$ by

$$E_i = \{\text{all components in } A_i \text{ function}\}$$

Now, since the system functions if and only if at least one of the events E_i occur, we have

$$r(\mathbf{p}) = P\left(\bigcup_{i=1}^{s} E_i\right)$$

Applying (9.4) yields the desired bounds on $r(\mathbf{p})$. The terms in the summation are computed thusly:

$$P(E_i) = \prod_{l\in A_i} p_l,$$
$$P(E_i E_j) = \prod_{l\in A_i\cup A_j} p_l,$$
$$P(E_i E_j E_k) = \prod_{l\in A_i\cup A_j\cup A_k} p_l$$

and so forth for intersections of more than three of the events. (The preceding follows since, for instance, in order for the event $E_i E_j$ to occur, all of the components in A_i and all of them in A_j must function; or, in other words, all components in $A_i \cup A_j$ must function.)

When the p_is are small the probabilities of the intersection of many of the events E_i should be quite small and the convergence should be relatively rapid.

Example 9.17. Consider the bridge structure with identical component probabilities. That is, take p_i to equal p for all i. Letting $A_1 = \{1, 4\}$, $A_2 = \{1, 3, 5\}$, $A_3 = \{2, 5\}$, and $A_4 = \{2, 3, 4\}$ denote the minimal path sets, we have

$$P(E_1) = P(E_3) = p^2,$$
$$P(E_2) = P(E_4) = p^3$$

Also, because exactly five of the six $= \binom{4}{2}$ unions of A_i and A_j contain four components (the exception being $A_2 \cup A_4$, which contains all five components), we have

$$P(E_1E_2) = P(E_1E_3) = P(E_1E_4) = P(E_2E_3) = P(E_3E_4) = p^4,$$
$$P(E_2E_4) = p^5$$

Hence, the first two inclusion–exclusion bounds yield

$$2(p^2 + p^3) - 5p^4 - p^5 \leqslant r(p) \leqslant 2(p^2 + p^3)$$

where $r(p) = r(p, p, p, p, p)$. For instance, when $p = 0.2$, we have

$$0.08768 \leqslant r(0.2) \leqslant 0.09600$$

and, when $p = 0.1$,

$$0.02149 \leqslant r(0.1) \leqslant 0.02200$$ ■

Just as we can define events in terms of the minimal path sets whose union is the event that the system functions, so can we define events in terms of the minimal cut sets whose union is the event that the system fails. Let $C_1, C_2, \ldots, C_r$ denote the minimal cut sets and define the events $F_1, \ldots, F_r$ by

$$F_i = \{\text{all components in } C_i \text{ are failed}\}$$

Now, because the system is failed if and only if all of the components of at least one minimal cut set are failed, we have

$$1 - r(\mathbf{p}) = P\left(\bigcup_1^r F_i\right),$$
$$1 - r(\mathbf{p}) \leqslant \sum_i P(F_i),$$
$$1 - r(\mathbf{p}) \geqslant \sum_i P(F_i) - \sum_{i<j}\sum P(F_iF_j),$$
$$1 - r(\mathbf{p}) \leqslant \sum_i P(F_i) - \sum_{i<j}\sum P(F_iF_j) + \sum_{i<j<k}\sum\sum P(F_iF_jF_k),$$

and so on. As

$$P(F_i) = \prod_{l \in C_i}(1 - p_l),$$
$$P(F_iF_j) = \prod_{l \in C_i \cup C_j}(1 - p_l),$$

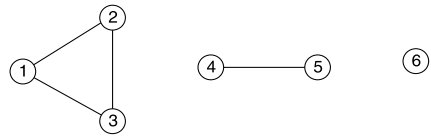

Figure 9.12

$$P(F_i F_j F_k) = \prod_{l \in C_i \cup C_j \cup C_k} (1 - p_l)$$

the convergence should be relatively rapid when the p_is are large.

Example 9.18 (A Random Graph)**.** Let us recall from Section 3.6.2 that a graph consists of a set N of nodes and a set A of pairs of nodes, called arcs. For any two nodes i and j we say that the sequence of arcs $(i, i_1)(i_1, i_2), \ldots, (i_k, j)$ constitutes an i–j path. If there is an i–j path between all the $\binom{n}{2}$ pairs of nodes i and $j, i \neq j$, then the graph is said to be connected. If we think of the nodes of a graph as representing geographical locations and the arcs as representing direct communication links between the nodes, then the graph will be connected if any two nodes can communicate with each other—if not directly, then at least through the use of intermediary nodes.

A graph can always be subdivided into nonoverlapping connected subgraphs called components. For instance, the graph in Fig. 9.12 with nodes $N = \{1, 2, 3, 4, 5, 6\}$ and arcs $A = \{(1, 2), (1, 3), (2, 3), (4, 5)\}$ consists of three components (a graph consisting of a single node is considered to be connected).

Consider now the random graph having nodes $1, 2, \ldots, n$, which is such that there is an arc from node i to node j with probability P_{ij}. Assume in addition that the occurrences of these arcs constitute independent events. That is, assume that the $\binom{n}{2}$ random variables $X_{ij}, i \neq j$, are independent where

$$X_{ij} = \begin{cases} 1, & \text{if } (i, j) \text{ is an arc} \\ 0, & \text{otherwise} \end{cases}$$

We are interested in the probability that this graph will be connected.

We can think of the preceding as being a reliability system of $\binom{n}{2}$ components—each component corresponding to a potential arc. The component is said to work if the corresponding arc is indeed an arc of the network, and the system is said to work if the corresponding graph is connected. As the addition of an arc to a connected graph cannot disconnect the graph, it follows that the structure so defined is monotone.

Let us start by determining the minimal path and minimal cut sets. It is easy to see that a graph will not be connected if and only if the set of nodes can be partitioned into two nonempty subsets X and X^c in such a way that there is no arc connecting a node from X with one from X^c. For instance, if there are six nodes and if there are no arcs connecting any of the nodes 1, 2, 3, 4 with either 5 or 6, then clearly the graph will not be connected. Thus, we see that any partition of the nodes into two nonempty

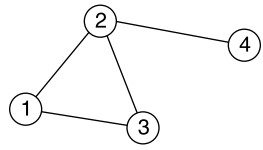

Figure 9.13

Figure 9.14

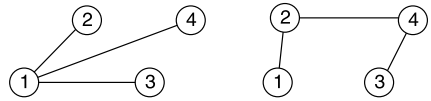

Figure 9.15 Two spanning trees (minimal path sets) when $n = 4$.

subsets X and X^c corresponds to the minimal cut set defined by

$$\{(i, j) : i \in X, j \in X^c\}$$

As there are $2^{n-1} - 1$ such partitions (there are $2^n - 2$ ways of choosing a nonempty proper subset X and, as the partition X, X^c is the same as X^c, X, we must divide by 2) there are therefore this number of minimal cut sets.

To determine the minimal path sets, we must characterize a minimal set of arcs that results in a connected graph. The graph in Fig. 9.13 is connected but it would remain connected if any one of the arcs from the cycle shown in Fig. 9.14 were removed. In fact it is not difficult to see that the minimal path sets are exactly those sets of arcs that result in a graph being connected but not having any cycles (a cycle being a path from a node to itself). Such sets of arcs are called *spanning trees* (Fig. 9.15). It is easily verified that any spanning tree contains exactly $n - 1$ arcs, and it is a famous result in graph theory (due to Cayley) that there are exactly n^{n-2} of these minimal path sets.

Because of the large number of minimal path and minimal cut sets (n^{n-2} and $2^{n-1} - 1$, respectively), it is difficult to obtain any useful bounds without making further restrictions. So, let us assume that all the P_{ij} equal the common value p. That is, we suppose that each of the possible arcs exists, independently, with the same probability p. We shall start by deriving a recursive formula for the probability that the graph is connected, which is computationally useful when n is not too large, and then we shall present an asymptotic formula for this probability when n is large.

Let us denote by P_n the probability that the random graph having n nodes is connected. To derive a recursive formula for P_n we first concentrate attention on a single node—say, node 1—and try to determine the probability that node 1 will be part of a component of size k in the resultant graph. Now, for a given set of $k - 1$ other nodes these nodes along with node 1 will form a component if

(i) there are no arcs connecting any of these k nodes with any of the remaining $n-k$ nodes;

(ii) the random graph restricted to these k nodes (and $\binom{k}{2}$ potential arcs—each independently appearing with probability p) is connected.

The probability that (i) and (ii) both occur is

$$q^{k(n-k)} P_k$$

where $q = 1 - p$. As there are $\binom{n-1}{k-1}$ ways of choosing $k-1$ other nodes (to form along with node 1 a component of size k) we see that

$$\begin{aligned} &P\{\text{node 1 is part of a component of size } k\} \\ &\quad = \binom{n-1}{k-1} q^{k(n-k)} P_k, \qquad k = 1, 2, \ldots, n \end{aligned}$$

Now, since the sum of the foregoing probabilities as k ranges from 1 through n clearly must equal 1, and as the graph is connected if and only if node 1 is part of a component of size n, we see that

$$P_n = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{k(n-k)} P_k, \qquad n = 2, 3, \ldots \tag{9.9}$$

Starting with $P_1 = 1$, $P_2 = p$, Eq. (9.9) can be used to determine P_n recursively when n is not too large. It is particularly suited for numerical computation.

To determine an asymptotic formula for P_n when n is large, first note from Eq. (9.9) that since $P_k \leqslant 1$, we have

$$1 - P_n \leqslant \sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{k(n-k)}$$

As it can be shown that for $q < 1$ and n sufficiently large,

$$\sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{k(n-k)} \leqslant (n+1) q^{n-1}$$

we have that for n large

$$1 - P_n \leqslant (n+1) q^{n-1} \tag{9.10}$$

To obtain a bound in the other direction, we concentrate our attention on a particular type of minimal cut set—namely, those that separate one node from all others in the graph. Specifically, define the minimal cut set C_i as

$$C_i = \{(i, j) : j \neq i\}$$

and define F_i to be the event that all arcs in C_i are not working (and thus, node i is isolated from the other nodes). Now,

$$1 - P_n = P(\text{graph is not connected}) \geqslant P\left(\bigcup_i F_i\right)$$

since, if any of the events F_i occur, then the graph will be disconnected. By the inclusion–exclusion bounds, we have

$$P\left(\bigcup_i F_i\right) \geqslant \sum_i P(F_i) - \sum_{i<j}\sum P(F_i F_j)$$

As $P(F_i)$ and $P(F_i F_j)$ are just the respective probabilities that a given set of $n-1$ arcs and a given set of $2n-3$ arcs are not in the graph (why?), it follows that

$$P(F_i) = q^{n-1},$$
$$P(F_i F_j) = q^{2n-3}, \quad i \neq j$$

and so

$$1 - P_n \geqslant nq^{n-1} - \binom{n}{2} q^{2n-3}$$

Combining this with Eq. (9.10) yields that for n sufficiently large,

$$nq^{n-1} - \binom{n}{2} q^{2n-3} \leqslant 1 - P_n \leqslant (n+1)q^{n-1}$$

and as

$$\binom{n}{2} \frac{q^{2n-3}}{nq^{n-1}} \to 0$$

as $n \to \infty$, we see that, for large n,

$$1 - P_n \approx nq^{n-1}$$

Thus, for instance, when $n = 20$ and $p = \frac{1}{2}$, the probability that the random graph will be connected is approximately given by

$$P_{20} \approx 1 - 20\left(\tfrac{1}{2}\right)^{19} = 0.99996$$ ■

9.4.2 Second Method for Obtaining Bounds on r(p)

Our second approach to obtaining bounds on $r(\mathbf{p})$ is based on expressing the desired probability as the probability of the intersection of events. To do so, let $A_1, A_2, \ldots, A_s$ denote the minimal path sets as before, and define the events, $D_i, i = 1, \ldots, s$ by

$$D_i = \{\text{at least one component in } A_i \text{ has failed}\}$$

Now since the system will have failed if and only if at least one component in each of the minimal path sets has failed we have

$$\begin{aligned} 1 - r(\mathbf{p}) &= P(D_1 D_2 \cdots D_s) \\ &= P(D_1)P(D_2 \mid D_1) \cdots P(D_s \mid D_1 D_2 \cdots D_{s-1}) \end{aligned} \tag{9.11}$$

Now it is quite intuitive that the information that at least one component of A_1 is down can only increase the probability that at least one component of A_2 is down (or else leave the probability unchanged if A_1 and A_2 do not overlap). Hence, intuitively

$$P(D_2 \mid D_1) \geqslant P(D_2)$$

To prove this inequality, we write

$$P(D_2) = P(D_2 \mid D_1)P(D_1) + P(D_2 \mid D_1^c)(1 - P(D_1)) \tag{9.12}$$

and note that

$$\begin{aligned} P(D_2 \mid D_1^c) &= P\{\text{at least one failed in } A_2 \mid \text{all functioning in } A_1\} \\ &= 1 - \prod_{\substack{j \in A_2 \\ j \notin A_1}} p_j \\ &\leqslant 1 - \prod_{j \in A_2} p_j \\ &= P(D_2) \end{aligned}$$

Hence, from Eq. (9.12) we see that

$$P(D_2) \leqslant P(D_2 \mid D_1)P(D_1) + P(D_2)(1 - P(D_1))$$

or

$$P(D_2 \mid D_1) \geqslant P(D_2)$$

By the same reasoning, it also follows that

$$P(D_i \mid D_1 \cdots D_{i-1}) \geqslant P(D_i)$$

and so from Eq. (9.11) we have

$$1 - r(\mathbf{p}) \geqslant \prod_i P(D_i)$$

or, equivalently,

$$r(\mathbf{p}) \leqslant 1 - \prod_i \left(1 - \prod_{j \in A_i} p_j\right)$$

To obtain a bound in the other direction, let $C_1, \ldots, C_r$ denote the minimal cut sets and define the events $U_1, \ldots, U_r$ by

$$U_i = \{\text{at least one component in } C_i \text{ is functioning}\}$$

Then, since the system will function if and only if all of the events U_i occur, we have

$$\begin{aligned} r(\mathbf{p}) &= P(U_1 U_2 \cdots U_r) \\ &= P(U_1) P(U_2 \mid U_1) \cdots P(U_r \mid U_1 \cdots U_{r-1}) \\ &\geqslant \prod_i P(U_i) \end{aligned}$$

where the last inequality is established in exactly the same manner as for the D_i. Hence,

$$r(\mathbf{p}) \geqslant \prod_i \left[1 - \prod_{j \in C_i} (1 - p_j)\right]$$

and we thus have the following bounds for the reliability function:

$$\prod_i \left[1 - \prod_{j \in C_i} (1 - p_j)\right] \leqslant r(\mathbf{p}) \leqslant 1 - \prod_i \left(1 - \prod_{j \in A_i} p_j\right) \tag{9.13}$$

It is to be expected that the upper bound should be close to the actual $r(\mathbf{p})$ if there is not too much overlap in the minimal path sets, and the lower bound to be close if there is not too much overlap in the minimal cut sets.

Example 9.19. For the three-out-of-four system the minimal path sets are $A_1 = \{1, 2, 3\}$, $A_2 = \{1, 2, 4\}$, $A_3 = \{1, 3, 4\}$, and $A_4 = \{2, 3, 4\}$; and the minimal cut sets are $C_1 = \{1, 2\}$, $C_2 = \{1, 3\}$, $C_3 = \{1, 4\}$, $C_4 = \{2, 3\}$, $C_5 = \{2, 4\}$, and $C_6 = \{3, 4\}$. Hence, by Eq. (9.13) we have

$$\begin{aligned} &(1 - q_1 q_2)(1 - q_1 q_3)(1 - q_1 q_4)(1 - q_2 q_3)(1 - q_2 q_4)(1 - q_3 q_4) \\ &\quad \leqslant r(\mathbf{p}) \leqslant 1 - (1 - p_1 p_2 p_3)(1 - p_1 p_2 p_4)(1 - p_1 p_3 p_4)(1 - p_2 p_3 p_4) \end{aligned}$$

where $q_i \equiv 1 - p_i$. For instance, if $p_i = \frac{1}{2}$ for all i, then the preceding yields

$$0.18 \leqslant r\left(\tfrac{1}{2}, \ldots, \tfrac{1}{2}\right) \leqslant 0.59$$

The exact value for this structure is easily computed to be

$$r\left(\tfrac{1}{2}, \ldots, \tfrac{1}{2}\right) = \tfrac{5}{16} = 0.31$$

■

9.5 System Life as a Function of Component Lives

For a random variable having distribution function G, we define $\bar{G}(a) \equiv 1 - G(a)$ to be the probability that the random variable is greater than a.

Consider a system in which the ith component functions for a random length of time having distribution F_i and then fails. Once failed it remains in that state forever. Assuming that the individual component lifetimes are independent, how can we express the distribution of system lifetime as a function of the system reliability function $r(\mathbf{p})$ and the individual component lifetime distributions $F_i, i = 1, \ldots, n$?

To answer this we first note that the system will function for a length of time t or greater if and only if it is still functioning at time t. That is, letting F denote the distribution of system lifetime, we have

$$\begin{aligned}\bar{F}(t) &= P\{\text{system life} > t\} \\ &= P\{\text{system is functioning at time } t\}\end{aligned}$$

But, by the definition of $r(\mathbf{p})$ we have

$$P\{\text{system is functioning at time } t\} = r(P_1(t), \ldots, P_n(t))$$

where

$$\begin{aligned}P_i(t) &= P\{\text{component } i \text{ is functioning at } t\} \\ &= P\{\text{lifetime of } i > t\} \\ &= \bar{F}_i(t)\end{aligned}$$

Hence, we see that

$$\bar{F}(t) = r(\bar{F}_1(t), \ldots, \bar{F}_n(t)) \tag{9.14}$$

Example 9.20. In a series system, $r(\mathbf{p}) = \prod_1^n p_i$ and so from Eq. (9.14)

$$\bar{F}(t) = \prod_1^n \bar{F}_i(t)$$

which is, of course, quite obvious since for a series system the system life is equal to the minimum of the component lives and so will be greater than t if and only if all component lives are greater than t. ■

Example 9.21. In a parallel system $r(\mathbf{p}) = 1 - \prod_1^n (1 - p_i)$ and so

$$\bar{F}(t) = 1 - \prod_1^n F_i(t)$$

The preceding is also easily derived by noting that, in the case of a parallel system, the system life is equal to the maximum of the component lives. ■

For a continuous distribution G, we define $\lambda(t)$, the *failure rate function* of G, by

$$\lambda(t) = \frac{g(t)}{\bar{G}(t)}$$

where $g(t) = d/dtG(t)$. In Section 5.2.2, it is shown that if G is the distribution of the lifetime of an item, then $\lambda(t)$ represents the probability intensity that a t-year-old item will fail. We say that G is an *increasing failure rate* (IFR) distribution if $\lambda(t)$ is an increasing function of t. Similarly, we say that G is a *decreasing failure rate* (DFR) distribution if $\lambda(t)$ is a decreasing function of t.

Example 9.22 (The Weibull Distribution)**.** A random variable is said to have the *Weibull* distribution if its distribution is given, for some $\lambda > 0, \alpha > 0$, by

$$G(t) = 1 - e^{-(\lambda t)^\alpha}, \qquad t \geqslant 0$$

The failure rate function for a Weibull distribution equals

$$\lambda(t) = \frac{e^{-(\lambda t)^\alpha}\alpha(\lambda t)^{\alpha-1}\lambda}{e^{-(\lambda t)^\alpha}} = \alpha\lambda(\lambda t)^{\alpha-1}$$

Thus, the Weibull distribution is IFR when $\alpha \geqslant 1$, and DFR when $0 < \alpha \leqslant 1$; when $\alpha = 1$, $G(t) = 1 - e^{-\lambda t}$, the exponential distribution, which is both IFR and DFR. ■

Example 9.23 (The Gamma Distribution)**.** A random variable is said to have a *gamma* distribution if its density is given, for some $\lambda > 0, \alpha > 0$, by

$$g(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\Gamma(\alpha)} \qquad \text{for } t \geqslant 0$$

where

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

For the gamma distribution,

$$\begin{aligned}\frac{1}{\lambda(t)} = \frac{\bar{G}(t)}{g(t)} &= \frac{\int_t^\infty \lambda e^{-\lambda x}(\lambda x)^{\alpha-1}dx}{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}} \\ &= \int_t^\infty e^{-\lambda(x-t)}\left(\frac{x}{t}\right)^{\alpha-1}dx\end{aligned}$$

With the change of variables $u = x - t$, we obtain

$$\frac{1}{\lambda(t)} = \int_0^\infty e^{-\lambda u}\left(1 + \frac{u}{t}\right)^{\alpha-1}du$$

Hence, G is IFR when $\alpha \geqslant 1$ and is DFR when $0 < \alpha \leqslant 1$. ■

Suppose that the lifetime distribution of each component in a monotone system is IFR. Does this imply that the system lifetime is also IFR? To answer this, let us at first suppose that each component has the same lifetime distribution, which we denote by G. That is, $F_i(t) = G(t), i = 1, \ldots, n$. To determine whether the system lifetime is IFR, we must compute $\lambda_F(t)$, the failure rate function of F. Now, by definition,

$$\begin{aligned}\lambda_F(t) &= \frac{(d/dt)F(t)}{\bar{F}(t)} \\ &= \frac{(d/dt)[1 - r(\bar{G}(t))]}{r(\bar{G}(t))}\end{aligned}$$

where

$$r(\bar{G}(t)) \equiv r(\bar{G}(t), \ldots, \bar{G}(t))$$

Hence,

$$\begin{aligned}\lambda_F(t) &= \frac{r'(\bar{G}(t))}{r(\bar{G}(t))} G'(t) \\ &= \frac{\bar{G}(t) r'(\bar{G}(t))}{r(\bar{G}(t))} \frac{G'(t)}{\bar{G}(t)} \\ &= \lambda_G(t) \frac{p r'(p)}{r(p)}\bigg|_{p=\bar{G}(t)}\end{aligned} \tag{9.15}$$

Since $\bar{G}(t)$ is a decreasing function of t, it follows from Eq. (9.15) that *if each component of a coherent system has the same IFR lifetime distribution, then the distribution of system lifetime will be IFR if $pr'(p)/r(p)$ is a decreasing function of p.*

Example 9.24 (The k-out-of-n System with Identical Components)**.** Consider the k-out-of-n system, which will function if and only if k or more components function. When each component has the same probability p of functioning, the number of functioning components will have a binomial distribution with parameters n and p. Hence,

$$r(p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

which, by continual integration by parts, can be shown to be equal to

$$r(p) = \frac{n!}{(k-1)!(n-k)!} \int_0^p x^{k-1}(1-x)^{n-k} dx$$

Upon differentiation, we obtain

$$r'(p) = \frac{n!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k}$$

Therefore,

$$\frac{pr'(p)}{r(p)} = \left[\frac{r(p)}{pr'(p)}\right]^{-1}$$
$$= \left[\frac{1}{p}\int_0^p \left(\frac{x}{p}\right)^{k-1}\left(\frac{1-x}{1-p}\right)^{n-k} dx\right]^{-1}$$

Letting $y = x/p$ yields

$$\frac{pr'(p)}{r(p)} = \left[\int_0^1 y^{k-1}\left(\frac{1-yp}{1-p}\right)^{n-k} dy\right]^{-1}$$

Since $(1-yp)/(1-p)$ is increasing in p, it follows that $pr'(p)/r(p)$ is decreasing in p. Thus, if a k-out-of-n system is composed of independent, like components having an increasing failure rate, the system itself has an increasing failure rate. ■

It turns out, however, that for a k-out-of-n system, in which the independent components have different IFR lifetime distributions, the system lifetime need not be IFR. Consider the following example of a two-out-of-two (that is, a parallel) system.

Example 9.25 (A Parallel System That Is Not IFR)**.** The life distribution of a parallel system of two independent components, the ith component having an exponential distribution with mean $1/i, i = 1, 2$, is given by

$$\bar{F}(t) = 1 - (1 - e^{-t})(1 - e^{-2t})$$
$$= e^{-2t} + e^{-t} - e^{-3t}$$

Therefore,

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}$$
$$= \frac{2e^{-2t} + e^{-t} - 3e^{-3t}}{e^{-2t} + e^{-t} - e^{-3t}}$$

It easily follows upon differentiation that the sign of $\lambda'(t)$ is determined by $e^{-5t} - e^{-3t} + 3e^{-4t}$, which is positive for small values and negative for large values of t. Therefore, $\lambda(t)$ is initially strictly increasing, and then strictly decreasing. Hence, F is not IFR. ■

Remark. The result of the preceding example is quite surprising at first glance. To obtain a better feel for it we need the concept of a mixture of distribution functions. The distribution function G is said to be a *mixture* of the distributions G_1 and G_2 if for some $p, 0 < p < 1$,

$$G(x) = pG_1(x) + (1 - p)G_2(x) \tag{9.16}$$

Mixtures occur when we sample from a population made up of two distinct groups. For example, suppose we have a stockpile of items of which the fraction p are type 1 and the fraction $1 - p$ are type 2. Suppose that the lifetime distribution of type 1 items is G_1 and of type 2 items is G_2. If we choose an item at random from the stockpile, then its life distribution is as given by Eq. (9.16).

Consider now a mixture of two exponential distributions having rates λ_1 and λ_2 where $\lambda_1 < \lambda_2$. We are interested in determining whether or not this mixture distribution is IFR. To do so, we note that if the item selected has survived up to time t, then its distribution of remaining life is still a mixture of the two exponential distributions. This is so since its remaining life will still be exponential with rate λ_1 if it is type 1 or with rate λ_2 if it is a type 2 item. However, the probability that it is a type 1 item is no longer the (prior) probability p but is now a conditional probability given that it has survived to time t. In fact, its probability of being a type 1 is

$$\begin{aligned} P\{\text{type 1} \mid \text{life} > t\} &= \frac{P\{\text{type 1, life} > t\}}{P\{\text{life} > t\}} \\ &= \frac{pe^{-\lambda_1 t}}{pe^{-\lambda_1 t} + (1-p)e^{-\lambda_2 t}} \end{aligned}$$

As the preceding is increasing in t, it follows that the larger t is, the more likely it is that the item in use is a type 1 (the better one, since $\lambda_1 < \lambda_2$). Hence, the older the item is, the less likely it is to fail, and thus the mixture of exponentials far from being IFR is, in fact, DFR.

Now, let us return to the parallel system of two exponential components having respective rates λ_1 and λ_2. The lifetime of such a system can be expressed as the sum of two independent random variables, namely,

$$\text{system life} = \text{Exp}(\lambda_1 + \lambda_2) + \begin{cases} \text{Exp}(\lambda_1) \text{ with probability } \dfrac{\lambda_2}{\lambda_1 + \lambda_2} \\ \text{Exp}(\lambda_2) \text{ with probability } \dfrac{\lambda_1}{\lambda_1 + \lambda_2} \end{cases}$$

The first random variable whose distribution is exponential with rate $\lambda_1 + \lambda_2$ represents the time until one of the components fails, and the second, which is a mixture of exponentials, is the additional time until the other component fails. (Why are these two random variables independent?)

Now, given that the system has survived a time t, it is very unlikely when t is large that both components are still functioning, but instead it is far more likely that one of the components has failed. Hence, for large t, the distribution of remaining life is basically a mixture of two exponentials—and so as t becomes even larger its failure rate should decrease (as indeed occurs). ■

Recall that the failure rate function of a distribution $F(t)$ having density $f(t) = F'(t)$ is defined by

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

By integrating both sides, we obtain

$$\begin{aligned}\int_0^t \lambda(s)\,ds &= \int_0^t \frac{f(s)}{1-F(s)}ds \\ &= -\log\,\bar{F}(t)\end{aligned}$$

Hence,

$$\bar{F}(t) = e^{-\Lambda(t)} \tag{9.17}$$

where

$$\Lambda(t) = \int_0^t \lambda(s)\,ds$$

The function $\Lambda(t)$ is called the *hazard function* of the distribution F.

Definition 9.1. A distribution F is said to have *increasing failure on the average* (IFRA) if

$$\frac{\Lambda(t)}{t} = \frac{\int_0^t \lambda(s)\,ds}{t} \tag{9.18}$$

increases in t for $t \geqslant 0$.

In other words, Eq. (9.18) states that the average failure rate up to time t increases as t increases. It is not difficult to show that if F is IFR, then F is IFRA; but the reverse need not be true.

Note that F is IFRA if $\Lambda(s)/s \leqslant \Lambda(t)/t$ whenever $0 \leqslant s \leqslant t$, which is equivalent to

$$\frac{\Lambda(\alpha t)}{\alpha t} \leqslant \frac{\Lambda(t)}{t} \quad \text{for } 0 \leqslant \alpha \leqslant 1,\ \text{all } t \geqslant 0$$

But by Eq. (9.17) we see that $\Lambda(t) = -\log \bar{F}(t)$, and so the preceding is equivalent to

$$-\log \bar{F}(\alpha t) \leqslant -\alpha \log \bar{F}(t)$$

or equivalently,

$$\log \bar{F}(\alpha t) \geqslant \log \bar{F}^{\alpha}(t)$$

which, since $\log x$ is a monotone function of x, shows that F is *IFRA if and only if*

$$\bar{F}(\alpha t) \geqslant \bar{F}^{\alpha}(t) \qquad \text{for } 0 \leqslant \alpha \leqslant 1,\ \text{all } t \geqslant 0 \tag{9.19}$$

For a vector $\mathbf{p} = (p_1, \ldots, p_n)$ we define $\mathbf{p}^{\alpha} = (p_1^{\alpha}, \ldots, p_n^{\alpha})$. We shall need the following proposition.

Proposition 9.2. *Any reliability function* $r(\mathbf{p})$ *satisfies*

$$r(\mathbf{p}^\alpha) \geqslant [r(\mathbf{p})]^\alpha, \qquad 0 \leqslant \alpha \leqslant 1$$

Proof. We prove this by induction on n, the number of components in the system. If $n = 1$, then either $r(p) \equiv 0, r(p) \equiv 1,$ or $r(p) \equiv p$. Hence, the proposition follows in this case.

Assume that Proposition 9.2 is valid for all monotone systems of $n-1$ components and consider a system of n components having structure function ϕ. By conditioning upon whether or not the nth component is functioning, we obtain

$$r(\mathbf{p}^\alpha) = p_n^\alpha r(1_n, \mathbf{p}^\alpha) + (1 - p_n^\alpha) r(0_n, \mathbf{p}^\alpha) \tag{9.20}$$

Now consider a system of components 1 through $n-1$ having a structure function $\phi_1(\mathbf{x}) = \phi(1_n, \mathbf{x})$. The reliability function for this system is given by $r_1(\mathbf{p}) = r(1_n, \mathbf{p})$; hence, from the induction assumption (valid for all monotone systems of $n-1$ components), we have

$$r(1_n, \mathbf{p}^\alpha) \geqslant [r(1_n, \mathbf{p})]^\alpha$$

Similarly, by considering the system of components 1 through $n-1$ and structure function $\phi_0(\mathbf{x}) = \phi(0_n, \mathbf{x})$, we obtain

$$r(0_n, \mathbf{p}^\alpha) \geqslant [r(0_n, \mathbf{p})]^\alpha$$

Thus, from Eq. (9.20), we obtain

$$r(\mathbf{p}^\alpha) \geqslant p_n^\alpha [r(1_n, \mathbf{p})]^\alpha + (1 - p_n^\alpha)[r(0_n, \mathbf{p})]^\alpha$$

which, by using the lemma to follow (with $\lambda = p_n, x = r(1_n, \mathbf{p}), y = r(0_n, \mathbf{p})$), implies that

$$\begin{aligned} r(\mathbf{p}^\alpha) &\geqslant [p_n r(1_n, \mathbf{p}) + (1 - p_n) r(0_n, \mathbf{p})]^\alpha \\ &= [r(\mathbf{p})]^\alpha \end{aligned}$$

which proves the result. ■

Lemma 9.3. *If* $0 \leqslant \alpha \leqslant 1, 0 \leqslant \lambda \leqslant 1$, *then*

$$h(y) = \lambda^\alpha x^\alpha + (1 - \lambda^\alpha) y^\alpha - (\lambda x + (1 - \lambda) y)^\alpha \geqslant 0$$

for all $0 \leqslant y \leqslant x$.

Proof. The proof is left as an exercise. ■

We are now ready to prove the following important theorem.

Theorem 9.2. *For a monotone system of independent components, if each component has an IFRA lifetime distribution, then the distribution of system lifetime is itself IFRA.*

Proof. The distribution of system lifetime F is given by

$$\bar{F}(\alpha t) = r(\bar{F}_1(\alpha t), \ldots, \bar{F}_n(\alpha t))$$

Hence, since r is a monotone function, and since each of the component distributions $\bar{F}_i$ is IFRA, we obtain from Eq. (9.19)

$$\begin{aligned}\bar{F}(\alpha t) &\geqslant r(\bar{F}_1^{\alpha}(t), \ldots, \bar{F}_n^{\alpha}(t)) \\ &\geqslant [r(\bar{F}_1(t), \ldots, \bar{F}_n(t))]^{\alpha} \\ &= \bar{F}^{\alpha}(t)\end{aligned}$$

which by Eq. (9.19) proves the theorem. The last inequality followed, of course, from Proposition 9.2. ■

9.6 Expected System Lifetime

In this section, we show how the mean lifetime of a system can be determined, at least in theory, from a knowledge of the reliability function $r(\mathbf{p})$ and the component lifetime distributions $F_i, i = 1, \ldots, n$.

Since the system's lifetime will be t or larger if and only if the system is still functioning at time t, we have

$$P\{\text{system life} > t\} = r\left(\bar{\mathbf{F}}(t)\right)$$

where $\bar{\mathbf{F}}(t) = (\bar{F}_1(t), \ldots, \bar{F}_n(t))$. Hence, by a well-known formula that states that for any nonnegative random variable X,

$$E[X] = \int_0^{\infty} P\{X > x\}\, dx,$$

we see that[4]

$$E[\text{system life}] = \int_0^{\infty} r\left(\bar{\mathbf{F}}(t)\right)\, dt \tag{9.21}$$

Example 9.26 (A Series System of Uniformly Distributed Components)**.** Consider a series system of three independent components each of which functions for an amount of time (in hours) uniformly distributed over (0, 10). Hence, $r(\mathbf{p}) = p_1 p_2 p_3$ and

$$F_i(t) = \begin{cases} t/10, & 0 \leqslant t \leqslant 10 \\ 1, & t > 10 \end{cases} \qquad i = 1, 2, 3$$

[4] That $E[X] = \int_0^{\infty} P\{X > x\}\, dx$ can be shown as follows when X has density f:

$$\int_0^{\infty} P\{X > x\}\, dx = \int_0^{\infty} \int_x^{\infty} f(y)\, dy\, dx = \int_0^{\infty} \int_0^{y} f(y)\, dx\, dy = \int_0^{\infty} y f(y)\, dy = E[X]$$

Therefore,

$$r\big(\bar{\mathbf{F}}(t)\big) = \begin{cases} \left(\dfrac{10-t}{10}\right)^3, & 0 \leqslant t \leqslant 10 \\ 0, & t > 10 \end{cases}$$

and so from Eq. (9.21) we obtain

$$\begin{aligned} E[\text{system life}] &= \int_0^{10} \left(\frac{10-t}{10}\right)^3 dt \\ &= 10 \int_0^1 y^3\, dy \\ &= \tfrac{5}{2} \end{aligned}$$

■

Example 9.27 (A Two-out-of-Three System)**.** Consider a two-out-of-three system of independent components, in which each component's lifetime is (in months) uniformly distributed over (0, 1). As was shown in Example 9.13, the reliability of such a system is given by

$$r(\mathbf{p}) = p_1p_2 + p_1p_3 + p_2p_3 - 2p_1p_2p_3$$

Since

$$F_i(t) = \begin{cases} t, & 0 \leqslant t \leqslant 1 \\ 1, & t > 1 \end{cases}$$

we see from Eq. (9.21) that

$$\begin{aligned} E[\text{system life}] &= \int_0^1 [3(1-t)^2 - 2(1-t)^3]\, dt \\ &= \int_0^1 (3y^2 - 2y^3)\, dy \\ &= 1 - \tfrac{1}{2} \\ &= \tfrac{1}{2} \end{aligned}$$

■

Example 9.28 (A Four-Component System)**.** Consider the four-component system that functions when components 1 and 2 and at least one of components 3 and 4 functions. Its structure function is given by

$$\phi(\mathbf{x}) = x_1x_2(x_3 + x_4 - x_3x_4)$$

and thus its reliability function equals

$$r(\mathbf{p}) = p_1p_2(p_3 + p_4 - p_3p_4)$$

Let us compute the mean system lifetime when the ith component is uniformly distributed over $(0, i)$, $i = 1, 2, 3, 4$. Now,

$$\bar{F}_1(t) = \begin{cases} 1 - t, & 0 \leqslant t \leqslant 1 \\ 0, & t > 1 \end{cases}$$

$$\bar{F}_2(t) = \begin{cases} 1 - t/2, & 0 \leqslant t \leqslant 2 \\ 0, & t > 2 \end{cases}$$

$$\bar{F}_3(t) = \begin{cases} 1 - t/3, & 0 \leqslant t \leqslant 3 \\ 0, & t > 3 \end{cases}$$

$$\bar{F}_4(t) = \begin{cases} 1 - t/4, & 0 \leqslant t \leqslant 4 \\ 0, & t > 4 \end{cases}$$

Hence,

$$r\big(\bar{\mathbf{F}}(t)\big) = \begin{cases} (1-t)\left(\dfrac{2-t}{2}\right)\left[\dfrac{3-t}{3} + \dfrac{4-t}{4} - \dfrac{(3-t)(4-t)}{12}\right], & 0 \leqslant t \leqslant 1 \\ 0, & t > 1 \end{cases}$$

Therefore,

$$\begin{aligned} E[\text{system life}] &= \frac{1}{24}\int_0^1 (1-t)(2-t)(12-t^2)\,dt \\ &= \frac{593}{(24)(60)} \\ &\approx 0.41 \end{aligned}$$

■

We end this section by obtaining the mean lifetime of a k-out-of-n system of independent identically distributed exponential components. If θ is the mean lifetime of each component, then

$$\bar{F}_i(t) = e^{-t/\theta}$$

Hence, since for a k-out-of-n system,

$$r(p, p, \ldots, p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

we obtain from Eq. (9.21)

$$E[\text{system life}] = \int_0^\infty \sum_{i=k}^{n} \binom{n}{i} (e^{-t/\theta})^i (1 - e^{-t/\theta})^{n-i} dt$$

Making the substitution

$$y = e^{-t/\theta}, \qquad dy = -\frac{1}{\theta} e^{-t/\theta} dt = -\frac{y}{\theta} dt$$

yields

$$E[\text{system life}] = \theta \sum_{i=k}^{n} \binom{n}{i} \int_0^1 y^{i-1}(1-y)^{n-i}\,dy$$

Now, it is not difficult to show that[5]

$$\int_0^1 y^n(1-y)^m\,dy = \frac{m!n!}{(m+n+1)!} \tag{9.22}$$

Thus, the foregoing equals

$$\begin{aligned} E[\text{system life}] &= \theta \sum_{i=k}^{n} \frac{n!}{(n-i)!i!}\frac{(i-1)!(n-i)!}{n!} \\ &= \theta \sum_{i=k}^{n} \frac{1}{i} \end{aligned} \tag{9.23}$$

Remark. Eq. (9.23) could have been proven directly by making use of special properties of the exponential distribution. First note that the lifetime of a k-out-of-n system can be written as $T_1 + \cdots + T_{n-k+1}$, where T_i represents the time between the $(i-1)$st and ith failure. This is true since $T_1 + \cdots + T_{n-k+1}$ equals the time at which the $(n-k+1)$st component fails, which is also the first time that the number of functioning components is less than k. Now, when all n components are functioning, the rate at which failures occur is n/θ. That is, T_1 is exponentially distributed with mean θ/n. Similarly, since T_i represents the time until the next failure when there are $n-(i-1)$ functioning components, it follows that T_i is exponentially distributed with mean $\theta/(n-i+1)$. Hence, the mean system lifetime equals

$$E[T_1 + \cdots + T_{n-k+1}] = \theta\left[\frac{1}{n} + \cdots + \frac{1}{k}\right]$$

Note also that it follows, from the lack of memory of the exponential, that the T_i, $i = 1, \ldots, n-k+1$, are independent random variables.

9.6.1 An Upper Bound on the Expected Life of a Parallel System

Consider a parallel system of n components, whose lifetimes are not necessarily independent. The system lifetime can be expressed as

$$\text{system life} = \max_i X_i$$

[5] Let

$$C(n,m) = \int_0^1 y^n(1-y)^m\,dy$$

Integration by parts yields $C(n,m) = [m/(n+1)]C(n+1, m-1)$. Starting with $C(n,0) = 1/(n+1)$, Eq. (9.22) follows by mathematical induction.

where X_i is the lifetime of component $i, i = 1, \ldots, n$. We can bound the expected system lifetime by making use of the following inequality. Namely, for any constant c

$$\max_i X_i \leqslant c + \sum_{i=1}^{n} (X_i - c)^+ \tag{9.24}$$

where x^+, the positive part of x, is equal to x if $x > 0$ and is equal to 0 if $x \leqslant 0$. The validity of Inequality (9.24) is immediate since if $\max X_i < c$ then the left side is equal to $\max X_i$ and the right side is equal to c. On the other hand, if $X_{(n)} = \max X_i > c$ then the right side is at least as large as $c + (X_{(n)} - c) = X_{(n)}$. It follows from Inequality (9.24), upon taking expectations, that

$$E\big[\max_i X_i\big] \leqslant c + \sum_{i=1}^{n} E[(X_i - c)^+] \tag{9.25}$$

Now, $(X_i - c)^+$ is a nonnegative random variable and so

$$\begin{aligned} E[(X_i - c)^+] &= \int_0^\infty P\{(X_i - c)^+ > x\}\, dx \\ &= \int_0^\infty P\{X_i - c > x\}\, dx \\ &= \int_c^\infty P\{X_i > y\}\, dy \end{aligned}$$

Thus, we obtain

$$E\big[\max_i X_i\big] \leqslant c + \sum_{i=1}^{n} \int_c^\infty P\{X_i > y\}\, dy \tag{9.26}$$

Because the preceding is true for all c, it follows that we obtain the best bound by letting c equal the value that minimizes the right side of the preceding. To determine that value, differentiate the right side of the preceding and set the result equal to 0, to obtain

$$1 - \sum_{i=1}^{n} P\{X_i > c\} = 0$$

That is, the minimizing value of c is that value c^* for which

$$\sum_{i=1}^{n} P\{X_i > c^*\} = 1$$

Since $\sum_{i=1}^{n} P\{X_i > c\}$ is a decreasing function of c, the value of c^* can be easily approximated and then utilized in Inequality (9.26). Also, it is interesting to note that

c^* is such that the expected number of the X_i that exceed c^* is equal to 1 (see Exercise 32). That the optimal value of c has this property is interesting and somewhat intuitive in as much as Inequality (9.24) is an equality when exactly one of the X_i exceeds c.

Example 9.29. Suppose the lifetime of component i is exponentially distributed with rate λ_i, $i = 1, \ldots, n$. Then the minimizing value of c is such that

$$1 = \sum_{i=1}^{n} P\{X_i > c^*\} = \sum_{i=1}^{n} e^{-\lambda_i c^*}$$

and the resulting bound of the mean system life is

$$\begin{aligned} E\big[\max_i X_i\big] &\leqslant c^* + \sum_{i=1}^{n} E[(X_i - c^*)^+] \\ &= c^* + \sum_{i=1}^{n} \big(E[(X_i - c^*)^+ \mid X_i > c^*] P\{X_i > c^*\} \\ &\quad + E[(X_i - c^*)^+ \mid X_i \leqslant c^*]\, P\{X_i \leqslant c^*\}\big) \\ &= c^* + \sum_{i=1}^{n} \frac{1}{\lambda_i} e^{-\lambda_i c^*} \end{aligned}$$

In the special case where all the rates are equal, say, $\lambda_i = \lambda$, $i = 1, \ldots, n$, then

$$1 = ne^{-\lambda c^*} \quad \text{or} \quad c^* = \frac{1}{\lambda} \log(n)$$

and the bound is

$$E\big[\max_i X_i\big] \leqslant \frac{1}{\lambda}(\log(n) + 1)$$

That is, if $X_1, \ldots, X_n$ are identically distributed exponential random variables with rate λ, then the preceding gives a bound on the expected value of their maximum. In the special case where these random variables are also independent, the following exact expression, given by Eq. (9.25), is not much less than the preceding upper bound:

$$E\big[\max_i X_i\big] = \frac{1}{\lambda} \sum_{i=1}^{n} 1/i \approx \frac{1}{\lambda} \int_1^n \frac{1}{x}\, dx \approx \frac{1}{\lambda} \log(n) \qquad \blacksquare$$

9.7 Systems with Repair

Consider an n-component system having reliability function $r(\mathbf{p})$. Suppose that component i functions for an exponentially distributed time with rate λ_i and then fails;

once failed it takes an exponential time with rate μ_i to be repaired, $i = 1, \ldots, n$. All components act independently.

Let us suppose that all components are initially working, and let

$$A(t) = P\{\text{system is working at } t\}$$

$A(t)$ is called the *availability* at time t. Since the components act independently, $A(t)$ can be expressed in terms of the reliability function as follows:

$$A(t) = r(A_1(t), \ldots, A_n(t)) \tag{9.27}$$

where

$$A_i(t) = P\{\text{component } i \text{ is functioning at } t\}$$

Now the state of component i—either on or off—changes in accordance with a two-state continuous time Markov chain. Hence, from the results of Example 6.11 we have

$$A_i(t) = P_{00}(t) = \frac{\mu_i}{\mu_i + \lambda_i} + \frac{\lambda_i}{\mu_i + \lambda_i} e^{-(\lambda_i + \mu_i)t}$$

Thus, we obtain

$$A(t) = r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\mu} + \boldsymbol{\lambda}} + \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu} + \boldsymbol{\lambda}} e^{-(\boldsymbol{\lambda} + \boldsymbol{\mu})t}\right)$$

If we let t approach ∞, then we obtain the limiting availability—call it A—which is given by

$$A = \lim_{t \to \infty} A(t) = r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\lambda} + \boldsymbol{\mu}}\right)$$

Remarks. **(i)** If the on and off distribution for component i are arbitrary continuous distributions with respective means $1/\lambda_i$ and $1/\mu_i, i = 1, \ldots, n$, then it follows from the theory of alternating renewal processes (see Section 7.5.1) that

$$A_i(t) \to \frac{1/\lambda_i}{1/\lambda_i + 1/\mu_i} = \frac{\mu_i}{\mu_i + \lambda_i}$$

and thus using the continuity of the reliability function, it follows from (9.27) that the limiting availability is

$$A = \lim_{t \to \infty} A(t) = r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\mu} + \boldsymbol{\lambda}}\right)$$

Hence, A depends only on the on and off distributions through their means.

(ii) It can be shown (using the theory of regenerative processes as presented in Section 7.5) that A will also equal the long-run proportion of time that the system will be functioning.

Example 9.30. For a series system, $r(\mathbf{p}) = \prod_{i=1}^{n} p_i$ and so

$$A(t) = \prod_{i=1}^{n} \left[\frac{\mu_i}{\mu_i + \lambda_i} + \frac{\lambda_i}{\mu_i + \lambda_i} e^{-(\lambda_i + \mu_i)t} \right]$$

and

$$A = \prod_{i=1}^{n} \frac{\mu_i}{\mu_i + \lambda_i}$$ ■

Example 9.31. For a parallel system, $r(\mathbf{p}) = 1 - \prod_{i=1}^{n} (1 - p_i)$ and thus

$$A(t) = 1 - \prod_{i=1}^{n} \left[\frac{\lambda_i}{\mu_i + \lambda_i} (1 - e^{-(\lambda_i + \mu_i)t}) \right]$$

and

$$A(t) = 1 - \prod_{i=1}^{n} \frac{\lambda_i}{\mu_i + \lambda_i}$$ ■

The preceding system will alternate between periods when it is up and periods when it is down. Let us denote by U_i and $D_i, i \geqslant 1$, the lengths of the ith up and down period respectively. For instance in a two-out-of-three system, U_1 will be the time until two components are down; D_1, the additional time until two are up; U_2 the additional time until two are down, and so on. Let

$$\bar{U} = \lim_{n \to \infty} \frac{U_1 + \cdots + U_n}{n},$$
$$\bar{D} = \lim_{n \to \infty} \frac{D_1 + \cdots + D_n}{n}$$

denote the average length of an up and down period respectively.[6]

To determine $\bar{U}$ and $\bar{D}$, note first that in the first n up–down cycles—that is, in time $\sum_{i=1}^{n} (U_i + D_i)$—the system will be up for a time $\sum_{i=1}^{n} U_i$. Hence, the proportion of time the system will be up in the first n up–down cycles is

$$\frac{U_1 + \cdots + U_n}{U_1 + \cdots + U_n + D_1 + \cdots + D_n} = \frac{\sum_{i=1}^{n} U_i / n}{\sum_{i=1}^{n} U_i / n + \sum_{i=1}^{n} D_i / n}$$

As $n \to \infty$, this must converge to A, the long-run proportion of time the system is up. Hence,

$$\frac{\bar{U}}{\bar{U} + \bar{D}} = A = r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\lambda} + \boldsymbol{\mu}} \right) \tag{9.28}$$

[6] It can be shown using the theory of regenerative processes that, with probability 1, the preceding limits will exist and will be constants.

However, to solve for $\bar{U}$ and $\bar{D}$ we need a second equation. To obtain one consider the rate at which the system fails. As there will be n failures in time $\sum_{i=1}^{n}(U_i + D_i)$, it follows that the rate at which the system fails is

$$\begin{aligned}\text{rate at which system fails} &= \lim_{n\to\infty} \frac{n}{\sum_1^n U_i + \sum_1^n D_i} \\ &= \lim_{n\to\infty} \frac{1}{\sum_1^n U_i/n + \sum_1^n D_i/n} = \frac{1}{\bar{U}+\bar{D}} \end{aligned} \tag{9.29}$$

That is, the foregoing yields the intuitive result that, on average, there is one failure every $\bar{U} + \bar{D}$ time units. To utilize this, let us determine the rate at which a failure of component i causes the system to go from up to down. Now, the system will go from up to down when component i fails if the states of the other components $x_1, \ldots, x_{i-1}, x_{i-1}, \ldots, x_n$ are such that $\phi(1_i, \mathbf{x}) = 1, \phi(0_i, \mathbf{x}) = 0$. That is, the states of the other components must be such that

$$\phi(1_i, \mathbf{x}) - \phi(0_i, \mathbf{x}) = 1 \tag{9.30}$$

Since component i will, on average, have one failure every $1/\lambda_i + 1/\mu_i$ time units, it follows that the rate at which component i fails is equal to $(1/\lambda_i + 1/\mu_i)^{-1} = \lambda_i\mu_i/(\lambda_i + \mu_i)$. In addition, the states of the other components will be such that (9.30) holds with probability

$$\begin{aligned} &P\{\phi(1_i, X(\infty)) - \phi(0_i, X(\infty)) = 1\} \\ &\quad = E[\phi(1_i, X(\infty)) - \phi(0_i, X(\infty))] \qquad \text{since } \phi(1_i, X(\infty)) - \phi(0_i, X(\infty)) \text{ is a Bernoulli random variable} \\ &\quad = r\left(1_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right) - r\left(0_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right) \end{aligned}$$

Hence, putting the preceding together we see that

$$\text{rate at which component } i \text{ causes the system to fail} = \frac{\lambda_i\mu_i}{\lambda_i+\mu_i}\left[r\left(1_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right) - r\left(0_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)\right]$$

Summing this over all components i thus gives

$$\text{rate at which system fails} = \sum_i \frac{\lambda_i\mu_i}{\lambda_i+\mu_i}\left[r\left(1_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right) - r\left(0_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)\right]$$

Finally, equating the preceding with (9.29) yields

$$\frac{1}{\bar{U}+\bar{D}} = \sum_i \frac{\lambda_i\mu_i}{\lambda_i+\mu_i}\left[r\left(1_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right) - r\left(0_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)\right] \tag{9.31}$$

Solving (9.28) and (9.31), we obtain

$$\bar{U} = \frac{r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)}{\sum_{i=1}^{n} \frac{\lambda_i \mu_i}{\lambda_i+\mu_i}\left[r\left(1_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right) - r\left(0_i, \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)\right]}, \tag{9.32}$$

$$\bar{D} = \frac{\left[1 - r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)\right]\bar{U}}{r\left(\frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}+\boldsymbol{\mu}}\right)} \tag{9.33}$$

Also, (9.31) yields the rate at which the system fails.

Remark. In establishing the formulas for $\bar{U}$ and $\bar{D}$, we did not make use of the assumption of exponential on and off times and in fact, our derivation is valid and Eqs. (9.32) and (9.33) hold whenever $\bar{U}$ and $\bar{D}$ are well defined (a sufficient condition is that all on and off distributions are continuous). The quantities $\lambda_i, \mu_i, i = 1, \ldots, n$, will represent, respectively, the reciprocals of the mean lifetimes and mean repair times.

Example 9.32. For a series system,

$$\bar{U} = \frac{\prod_i \frac{\mu_i}{\mu_i+\lambda_i}}{\sum_i \frac{\lambda_i \mu_i}{\lambda_i+\mu_i}\prod_{j\neq i}\frac{\mu_j}{\mu_j+\lambda_j}} = \frac{1}{\sum_i \lambda_i},$$

$$\bar{D} = \frac{1 - \prod_i \frac{\mu_i}{\mu_i+\lambda_i}}{\prod_i \frac{\mu_i}{\mu_i+\lambda_i}} \times \frac{1}{\sum_i \lambda_i}$$

whereas for a parallel system,

$$\bar{U} = \frac{1 - \prod_i \frac{\lambda_i}{\mu_i+\lambda_i}}{\sum_i \frac{\lambda_i \mu_i}{\lambda_i+\mu_i}\prod_{j\neq i}\frac{\lambda_j}{\mu_j+\lambda_j}} = \frac{1 - \prod_i \frac{\lambda_i}{\mu_i+\lambda_i}}{\prod_j \frac{\lambda_j}{\mu_j+\lambda_j}} \times \frac{1}{\sum_i \mu_i},$$

$$\bar{D} = \frac{\prod_i \frac{\lambda_i}{\mu_i+\lambda_i}}{1 - \prod_i \frac{\lambda_i}{\mu_i+\lambda_i}}\bar{U} = \frac{1}{\sum_i \mu_i}$$

The preceding formulas hold for arbitrary continuous up and down distributions with $1/\lambda_i$ and $1/\mu_i$ denoting respectively the mean up and down times of component $i, i = 1, \ldots, n$. ■

9.7.1 A Series Model with Suspended Animation

Consider a series consisting of n components, and suppose that whenever a component (and thus the system) goes down, repair begins on that component and each of the other components enters a state of suspended animation. That is, after the down component is repaired, the other components resume operation in exactly the same condition as when the failure occurred. If two or more components go down simultaneously, one of them is arbitrarily chosen as being the failed component and repair on that component begins; the others that went down at the same time are considered to be in a state of suspended animation, and they will instantaneously go down when the repair is completed. We suppose that (not counting any time in suspended animation) the distribution of time that component i functions is F_i with mean u_i, whereas its repair distribution is G_i with mean $d_i, i = 1, \ldots, n$.

To determine the long-run proportion of time this system is working, we reason as follows. To begin, consider the time, call it T, at which the system has been up for a time t. Now, when the system is up, the failure times of component i constitute a renewal process with mean interarrival time u_i. Therefore, it follows that

$$\text{number of failures of } i \text{ in time } T \approx \frac{t}{u_i}$$

As the average repair time of i is d_i, the preceding implies that

$$\text{total repair time of } i \text{ in time } T \approx \frac{td_i}{u_i}$$

Therefore, in the period of time in which the system has been up for a time t, the total system downtime has approximately been

$$t \sum_{i=1}^{n} d_i/u_i$$

Hence, the proportion of time that the system has been up is approximately

$$\frac{t}{t + t\sum_{i=1}^{n} d_i/u_i}$$

Because this approximation should become exact as we let t become larger, it follows that

$$\text{proportion of time the system is up} = \frac{1}{1 + \sum_i d_i/u_i} \tag{9.34}$$

which also shows that

$$\begin{aligned}\text{proportion of time the system is down} &= 1 - \text{proportion of time the system is up} \\ &= \frac{\sum_i d_i/u_i}{1 + \sum_i d_i/u_i}\end{aligned}$$

Moreover, in the time interval from 0 to T, the proportion of the repair time that has been devoted to component i is approximately

$$\frac{td_i/u_i}{\sum_i td_i/u_i}$$

Thus, in the long run,

$$\text{proportion of down time that is due to component } i = \frac{d_i/u_i}{\sum_i d_i/u_i}$$

Multiplying the preceding by the proportion of time the system is down gives

$$\text{proportion of time component } i \text{ is being repaired} = \frac{d_i/u_i}{1+\sum_i d_i/u_i}$$

Also, since component j will be in suspended animation whenever any of the other components is in repair, we see that

$$\text{proportion of time component } j \text{ is in suspended animation} = \frac{\sum_{i\neq j} d_i/u_i}{1+\sum_i d_i/u_i}$$

Another quantity of interest is the long-run rate at which the system fails. Since component i fails at rate $1/u_i$ when the system is up, and does not fail when the system is down, it follows that

$$\begin{aligned}\text{rate at which } i \text{ fails} &= \frac{\text{proportion of time system is up}}{u_i}\\ &= \frac{1/u_i}{1+\sum_i d_i/u_i}\end{aligned}$$

Since the system fails when any of its components fail, the preceding yields that

$$\text{rate at which the system fails} = \frac{\sum_i 1/u_i}{1+\sum_i d_i/u_i} \tag{9.35}$$

If we partition the time axis into periods when the system is up and those when it is down, we can determine the average length of an up period by noting that if $U(t)$ is the total amount of time that the system is up in the interval $[0, t]$, and if $N(t)$ is the number of failures by time t, then

$$\begin{aligned}\text{average length of an up period} &= \lim_{t\to\infty} \frac{U(t)}{N(t)}\\ &= \lim_{t\to\infty} \frac{U(t)/t}{N(t)/t}\\ &= \frac{1}{\sum_i 1/u_i}\end{aligned}$$

where the final equality used Eqs. (9.34) and (9.35). Also, in a similar manner it can be shown that

$$\text{average length of a down period} = \frac{\sum_i d_i/u_i}{\sum_i 1/u_i} \tag{9.36}$$

Exercises

1. Prove that, for any structure function ϕ,

$$\phi(\mathbf{x}) = x_i \phi(1_i, \mathbf{x}) + (1 - x_i)\phi(0_i, \mathbf{x})$$

where

$$(1_i, \mathbf{x}) = (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n),$$
$$(0_i, \mathbf{x}) = (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$$

2. Show that

(a) if $\phi(0, 0, \ldots, 0) = 0$ and $\phi(1, 1, \ldots, 1) = 1$, then

$$\min x_i \leqslant \phi(\mathbf{x}) \leqslant \max x_i$$

(b) $\phi(\max(\mathbf{x}, \mathbf{y})) \geqslant \max(\phi(\mathbf{x}), \phi(\mathbf{y}))$

(c) $\phi(\min(\mathbf{x}, \mathbf{y})) \leqslant \min(\phi(\mathbf{x}), \phi(\mathbf{y}))$

3. For any structure function ϕ, we define the dual structure ϕ^{D} by

$$\phi^{\mathrm{D}}(\mathbf{x}) = 1 - \phi(\mathbf{1} - \mathbf{x})$$

(a) Show that the dual of a parallel (series) system is a series (parallel) system.

(b) Show that the dual of a dual structure is the original structure.

(c) What is the dual of a k-out-of-n structure?

(d) Show that a minimal path (cut) set of the dual system is a minimal cut (path) set of the original structure.

***4.** Write the structure function corresponding to the following:

(a) See Fig. 9.16:

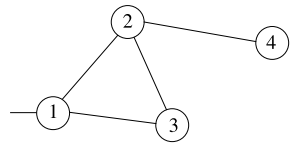

Figure 9.16

(b) See Fig. 9.17:

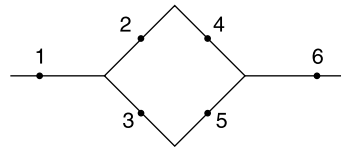

Figure 9.17

(c) See Fig. 9.18:

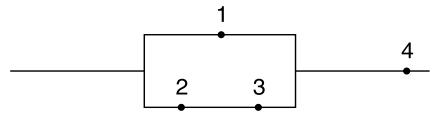

Figure 9.18

5. Find the minimal path and minimal cut sets for:
(a) See Fig. 9.19:

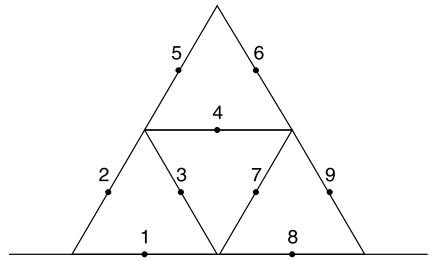

Figure 9.19

(b) See Fig. 9.20:

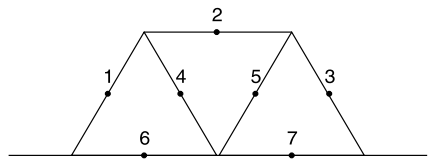

Figure 9.20

***6.** The minimal path sets are $\{1, 2, 4\}$, $\{1, 3, 5\}$, and $\{5, 6\}$. Give the minimal cut sets.

7. The minimal cut sets are $\{1, 2, 3\}$, $\{2, 3, 4\}$, and $\{3, 5\}$. What are the minimal path sets?

8. Give the minimal path sets and the minimal cut sets for the structure given by Fig. 9.21.

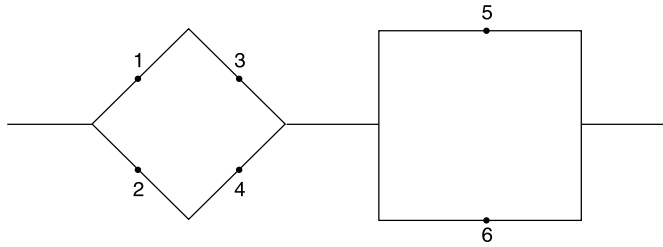

Figure 9.21

9. Component i is said to be *relevant* to the system if for some state vector $\mathbf{x}$,

$$\phi(1_i, \mathbf{x}) = 1, \qquad \phi(0_i, \mathbf{x}) = 0$$

Otherwise, it is said to be *irrelevant*.

(a) Explain in words what it means for a component to be irrelevant.

(b) Let $A_1, \ldots, A_s$ be the minimal path sets of a system, and let S denote the set of components. Show that $S = \bigcup_{i=1}^{s} A_i$ if and only if all components are relevant.

(c) Let $C_1, \ldots, C_k$ denote the minimal cut sets. Show that $S = \bigcup_{i=1}^{k} C_i$ if and only if all components are relevant.

10. Let t_i denote the time of failure of the ith component; let $\tau_\phi(t)$ denote the time to failure of the system ϕ as a function of the vector $\mathbf{t} = (t_1, \ldots, t_n)$. Show that

$$\max_{1 \leqslant j \leqslant s} \min_{i \in A_j} t_i = \tau_\phi(\mathbf{t}) = \min_{1 \leqslant j \leqslant k} \max_{i \in C_j} t_i$$

where $C_1, \ldots, C_k$ are the minimal cut sets, and $A_1, \ldots, A_s$ the minimal path sets.

11. Give the reliability function of the structure of Exercise 8.

***12.** Give the minimal path sets and the reliability function for the structure in Fig. 9.22.

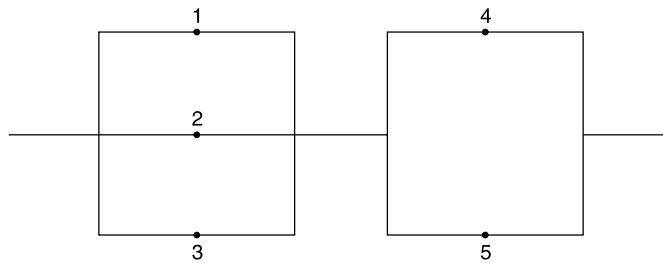

Figure 9.22

13. Let $r(\mathbf{p})$ be the reliability function. Show that

$$r(\mathbf{p}) = p_i r(1_i, \mathbf{p}) + (1 - p_i) r(0_i, \mathbf{p})$$

14. Compute the reliability function of the bridge system (see Fig. 9.11) by conditioning upon whether or not component 3 is working.

15. Compute upper and lower bounds of the reliability function (using Method 2) for the systems given in Exercise 4, and compare them with the exact values when $p_i \equiv \frac{1}{2}$.

16. Compute the upper and lower bounds of $r(\mathbf{p})$ using both methods for the
- **(a)** two-out-of-three system and
- **(b)** two-out-of-four system.
- **(c)** Compare these bounds with the exact reliability when
  - **(i)** $p_i \equiv 0.5$
  - **(ii)** $p_i \equiv 0.8$
  - **(iii)** $p_i \equiv 0.2$

***17.** Let N be a nonnegative, integer-valued random variable. Show that

$$P\{N > 0\} \geqslant \frac{(E[N])^2}{E[N^2]}$$

and explain how this inequality can be used to derive additional bounds on a reliability function.

Hint:

$$\begin{aligned} E[N^2] &= E[N^2 \mid N > 0]P\{N > 0\} \qquad \text{(Why?)} \\ &\geqslant (E[N \mid N > 0])^2 P\{N > 0\} \qquad \text{(Why?)} \end{aligned}$$

Now multiply both sides by $P\{N > 0\}$.

18. Consider a structure in which the minimal path sets are $\{1, 2, 3\}$ and $\{3, 4, 5\}$.
- **(a)** What are the minimal cut sets?
- **(b)** If the component lifetimes are independent uniform $(0, 1)$ random variables, determine the probability that the system life will be less than $\frac{1}{2}$.

19. Let $X_1, X_2, \ldots, X_n$ denote independent and identically distributed random variables and define the order statistics $X_{(1)}, \ldots, X_{(n)}$ by

$$X_{(i)} \equiv i\text{th smallest of } X_1, \ldots, X_n$$

Show that if the distribution of X_j is IFR, then so is the distribution of $X_{(i)}$.
Hint: Relate this to one of the examples of this chapter.

20. Let F be a continuous distribution function. For some positive α, define the distribution function G by

$$\bar{G}(t) = (\bar{F}(t))^{\alpha}$$

Find the relationship between $\lambda_G(t)$ and $\lambda_F(t)$, the respective failure rate functions of G and F.

21. Consider the following four structures:

(i) See Fig. 9.23:

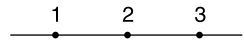

Figure 9.23

(ii) See Fig. 9.24:

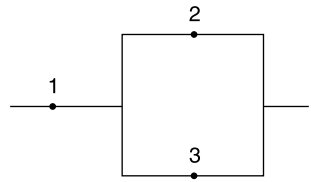

Figure 9.24

(iii) See Fig. 9.25:

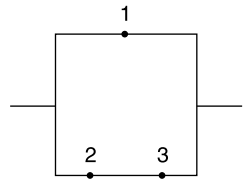

Figure 9.25

(iv) See Fig. 9.26:

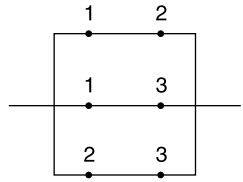

Figure 9.26

Let F_1, F_2, and F_3 be the corresponding component failure distributions; each of which is assumed to be IFR (increasing failure rate). Let F be the system failure distribution. All components are independent.

(a) For which structures is F necessarily IFR if $F_1 = F_2 = F_3$? Give reasons.

(b) For which structures is F necessarily IFR if $F_2 = F_3$? Give reasons.

(c) For which structures is F necessarily IFR if $F_1 \neq F_2 \neq F_3$? Give reasons.

***22.** Let X denote the lifetime of an item. Suppose the item has reached the age of t. Let X_t denote its remaining life and define

$$\bar{F}_t(a) = P\{X_t > a\}$$

In words, $\bar{F}_t(a)$ is the probability that a t-year-old item survives an additional time a. Show that

(a) $\bar{F}_t(a) = \bar{F}(t+a)/\bar{F}(t)$ where F is the distribution function of X.

(b) Another definition of IFR is to say that F is IFR if $\bar{F}_t(a)$ decreases in t, for all a. Show that this definition is equivalent to the one given in the text when F has a density.

23. Show that if each (independent) component of a series system has an IFR distribution, then the system lifetime is itself IFR by

(a) showing that

$$\lambda_F(t) = \sum_i \lambda_i(t)$$

where $\lambda_F(t)$ is the failure rate function of the system; and $\lambda_i(t)$ the failure rate function of the lifetime of component i.

(b) using the definition of IFR given in Exercise 22.

24. Show that if F is IFR, then it is also IFRA, and show by counterexample that the reverse is not true.

***25.** We say that ζ is a p-percentile of the distribution F if $F(\zeta) = p$. Show that if ζ is a p-percentile of the IFRA distribution F, then

$$\bar{F}(x) \leqslant e^{-\theta x}, \quad x \geqslant \zeta$$
$$\bar{F}(x) \geqslant e^{-\theta x}, \quad x \leqslant \zeta$$

where

$$\theta = \frac{-\log(1-p)}{\zeta}$$

26. Prove Lemma 9.3.

Hint: Let $x = y + \delta$. Note that $f(t) = t^\alpha$ is a concave function when $0 \leqslant \alpha \leqslant 1$, and use the fact that for a concave function $f(t+h) - f(t)$ is decreasing in t.

27. Let $r(p) = r(p, p, \ldots, p)$. Show that if $r(p_0) = p_0$, then

$$r(p) \geqslant p \quad \text{for } p \geqslant p_0$$
$$r(p) \leqslant p \quad \text{for } p \leqslant p_0$$

Hint: Use Proposition 9.2.

28. Find the mean lifetime of a series system of two components when the component lifetimes are respectively uniform on (0, 1) and uniform on (0, 2). Repeat for a parallel system.

29. Show that the mean lifetime of a parallel system of two components is

$$\frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{(\mu_1 + \mu_2)\mu_2} + \frac{\mu_2}{(\mu_1 + \mu_2)\mu_1}$$

when the first component is exponentially distributed with mean $1/\mu_1$ and the second is exponential with mean $1/\mu_2$.

***30.** Compute the expected system lifetime of a three-out-of-four system when the first two component lifetimes are uniform on $(0, 1)$ and the second two are uniform on $(0, 2)$.

31. Show that the variance of the lifetime of a k-out-of-n system of components, each of whose lifetimes is exponential with mean θ, is given by

$$\theta^2 \sum_{i=k}^{n} \frac{1}{i^2}$$

32. In Section 9.6.1 show that the expected number of X_i that exceed c^* is equal to 1.

33. Let X_i be an exponential random variable with mean $8 + 2i$, for $i = 1, 2, 3$. Use the results of Section 9.6.1 to obtain an upper bound on $E[\max X_i]$, and then compare this with the exact result when the X_i are independent.

34. For the model of Section 9.7, compute for a k-out-of-n structure (i) the average up time, (ii) the average down time, and (iii) the system failure rate.

35. Prove the combinatorial identity

$$\binom{n-1}{i-1} = \binom{n}{i} - \binom{n}{i+1} + \cdots \pm \binom{n}{n}, \quad i \leqslant n$$

(a) by induction on i;

(b) by a backwards induction argument on i—that is, prove it first for $i = n$, then assume it for $i = k$ and show that this implies that it is true for $i = k - 1$.

36. Verify Eq. (9.36).

References

[1] R.E. Barlow, F. Proschan, Statistical Theory of Reliability and Life Testing, Holt, New York, 1975.

[2] H. Frank, I. Frisch, Communication, Transmission, and Transportation Network, Addison-Wesley, Reading, Massachusetts, 1971.

[3] I.B. Gertsbakh, Statistical Reliability Theory, Marcel Dekker, New York and Basel, 1989.

Brownian Motion and Stationary Processes

10

10.1 Brownian Motion

Let us start by considering the symmetric random walk, which in each time unit is equally likely to take a unit step either to the left or to the right. That is, it is a Markov chain with $P_{i,i+1} = \frac{1}{2} = P_{i,i-1}$, $i = 0, \pm 1, \ldots$. Now suppose that we speed up this process by taking smaller and smaller steps in smaller and smaller time intervals. If we now go to the limit in the right manner what we obtain is Brownian motion.

More precisely, suppose that each Δt time unit we take a step of size Δx either to the left or the right with equal probabilities. If we let $X(t)$ denote the position at time t then

$$X(t) = \Delta x(X_1 + \cdots + X_{[t/\Delta t]}) \tag{10.1}$$

where

$$X_i = \begin{cases} +1, & \text{if the } i\text{th step of length } \Delta x \text{ is to the right} \\ -1, & \text{if it is to the left} \end{cases}$$

$[t/\Delta t]$ is the largest integer less than or equal to $t/\Delta t$, and the X_i are assumed independent with

$$P\{X_i = 1\} = P\{X_i = -1\} = \tfrac{1}{2}$$

As $E[X_i] = 0$, $\text{Var}(X_i) = E[X_i^2] = 1$, we see from Eq. (10.1) that

$$\begin{aligned} E[X(t)] &= 0, \\ \text{Var}(X(t)) &= (\Delta x)^2 \left[\frac{t}{\Delta t}\right] \end{aligned} \tag{10.2}$$

We shall now let Δx and Δt go to 0. However, we must do it in a way such that the resulting limiting process is nontrivial (for instance, if we let $\Delta x = \Delta t$ and let $\Delta t \to 0$, then from the preceding we see that $E[X(t)]$ and $\text{Var}(X(t))$ would both converge to 0 and thus $X(t)$ would equal 0 with probability 1). If we let $\Delta x = \sigma\sqrt{\Delta t}$ for some positive constant σ then from Eq. (10.2) we see that as $\Delta t \to 0$

$$\begin{aligned} E[X(t)] &= 0, \\ \text{Var}(X(t)) &\to \sigma^2 t \end{aligned}$$

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00015-9

We now list some intuitive properties of this limiting process obtained by taking $\Delta x = \sigma\sqrt{\Delta t}$ and then letting $\Delta t \to 0$. From Eq. (10.1) and the central limit theorem the following seems reasonable:

(i) $X(t)$ is normal with mean 0 and variance $\sigma^2 t$. In addition, because the changes of value of the random walk in nonoverlapping time intervals are independent,

(ii) $\{X(t), t \geq 0\}$ has independent increments, in that for all $t_1 < t_2 < \cdots < t_n$

$$X(t_n) - X(t_{n-1}), X(t_{n-1}) - X(t_{n-2}), \ldots, X(t_2) - X(t_1), X(t_1)$$

are independent. Finally, because the distribution of the change in position of the random walk over any time interval depends only on the length of that interval, it would appear that

(iii) $\{X(t), t \geq 0\}$ has stationary increments, in that the distribution of $X(t+s) - X(t)$ does not depend on t. We are now ready for the following formal definition.

Definition 10.1. A stochastic process $\{X(t),\ t \geq 0\}$ is said to be a *Brownian motion* process if

(i) $X(0) = 0$;

(ii) $\{X(t), t \geq 0\}$ has stationary and independent increments;

(iii) for every $t > 0$, $X(t)$ is normally distributed with mean 0 and variance $\sigma^2 t$.

The Brownian motion process, sometimes called the Wiener process, is one of the most useful stochastic processes in applied probability theory. It originated in physics as a description of Brownian motion. This phenomenon, named after the English botanist Robert Brown who discovered it, is the motion exhibited by a small particle that is totally immersed in a liquid or gas. Since then, the process has been used beneficially in such areas as statistical testing of goodness of fit, analyzing the price levels on the stock market, and quantum mechanics.

The first explanation of the phenomenon of Brownian motion was given by Einstein in 1905. He showed that Brownian motion could be explained by assuming that the immersed particle was continually being subjected to bombardment by the molecules of the surrounding medium. However, the preceding concise definition of this stochastic process underlying Brownian motion was given by Wiener in a series of papers originating in 1918.

When $\sigma = 1$, the process is called *standard Brownian motion*. Because any Brownian motion can be converted to the standard process by letting $B(t) = X(t)/\sigma$ we shall, unless otherwise stated, suppose throughout this chapter that $\sigma = 1$.

The interpretation of Brownian motion as the limit of the random walks (Eq. (10.1)) suggests that $X(t)$ should be a continuous function of t, which turns out to be true. To prove this, we must show that with probability 1

$$\lim_{h \to 0} (X(t+h) - X(t)) = 0$$

Although a rigorous proof of the preceding is beyond the scope of this text, a plausibility argument is obtained by noting that the random variable $X(t+h) - X(t)$ has

mean 0 and variance h, and so would seem to converge to a random variable with mean 0 and variance 0 as $h \to 0$. That is, it seems reasonable that $X(t+h) - X(t)$ converges to 0, thus yielding continuity.

Although $X(t)$ will, with probability 1, be a continuous function of t, it possesses the interesting property of being nowhere differentiable. To see why this might be the case, note that $\frac{X(t+h)-X(t)}{h}$ has mean 0 and variance $1/h$. Because the variance of $\frac{X(t+h)-X(t)}{h}$ converges to ∞ as $h \to 0$, it is not surprising that the ratio does not converge.

As $X(t)$ is normal with mean 0 and variance t, its density function is given by

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

To obtain the joint density function of $X(t_1), X(t_2), \ldots, X(t_n)$ for $t_1 < \cdots < t_n$, note first that the set of equalities

$$\begin{aligned} X(t_1) &= x_1, \\ X(t_2) &= x_2, \\ &\vdots \\ X(t_n) &= x_n \end{aligned}$$

is equivalent to

$$\begin{aligned} X(t_1) &= x_1, \\ X(t_2) - X(t_1) &= x_2 - x_1, \\ &\vdots \\ X(t_n) - X(t_{n-1}) &= x_n - x_{n-1} \end{aligned}$$

However, by the independent increment assumption it follows that $X(t_1), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$, are independent and, by the stationary increment assumption, that $X(t_k) - X(t_{k-1})$ is normal with mean 0 and variance $t_k - t_{k-1}$. Hence, the joint density of $X(t_1), \ldots, X(t_n)$ is given by

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \cdots f_{t_n-t_{n-1}}(x_n - x_{n-1}) \\ &= \frac{\exp\left\{-\frac{1}{2}\left[\frac{x_1^2}{t_1} + \frac{(x_2-x_1)^2}{t_2-t_1} + \cdots + \frac{(x_n-x_{n-1})^2}{t_n-t_{n-1}}\right]\right\}}{(2\pi)^{n/2}[t_1(t_2-t_1)\cdots(t_n-t_{n-1})]^{1/2}} \end{aligned} \tag{10.3}$$

From this equation, we can compute in principle any desired probabilities. For instance, suppose we require the conditional distribution of $X(s)$ given that $X(t) = B$ where $s < t$. The conditional density is

$$f_{s|t}(x|B) = \frac{f_s(x) f_{t-s}(B-x)}{f_t(B)}$$

$$
\begin{aligned}
&= K_1 \exp\{-x^2/2s - (B-x)^2/2(t-s)\} \\
&= K_2 \exp\left\{-x^2\left(\frac{1}{2s} + \frac{1}{2(t-s)}\right) + \frac{Bx}{t-s}\right\} \\
&= K_2 \exp\left\{-\frac{t}{2s(t-s)}\left(x^2 - 2\frac{sB}{t}x\right)\right\} \\
&= K_3 \exp\left\{-\frac{(x - Bs/t)^2}{2s(t-s)/t}\right\}
\end{aligned}
$$

where K_1, K_2, and K_3 do not depend on x. Hence, we see from the preceding that the conditional distribution of $X(s)$ given that $X(t) = B$ is, for $s < t$, normal with mean and variance given by

$$
\begin{aligned}
E[X(s)|X(t) = B] &= \frac{s}{t}B, \\
\text{Var}[X(s)|X(t) = B] &= \frac{s}{t}(t-s)
\end{aligned}
\tag{10.4}
$$

Example 10.1. In a bicycle race between two competitors, let $Y(t)$ denote the amount of time (in seconds) by which the racer that started in the inside position is ahead when $100t$ percent of the race has been completed, and suppose that $\{Y(t),\ 0 \le t \le 1\}$ can be effectively modeled as a Brownian motion process with variance parameter σ^2.

(a) If the inside racer is leading by σ seconds at the midpoint of the race, what is the probability that she is the winner?

(b) If the inside racer wins the race by a margin of σ seconds, what is the probability that she was ahead at the midpoint?

Solution:

(a)
$$
\begin{aligned}
&P\{Y(1) > 0|Y(1/2) = \sigma\} \\
&\quad = P\{Y(1) - Y(1/2) > -\sigma|Y(1/2) = \sigma\} \\
&\quad = P\{Y(1) - Y(1/2) > -\sigma\} \qquad \text{by independent increments} \\
&\quad = P\{Y(1/2) > -\sigma\} \qquad \text{by stationary increments} \\
&\quad = P\left\{\frac{Y(1/2)}{\sigma/\sqrt{2}} > -\sqrt{2}\right\} \\
&\quad = \Phi(\sqrt{2}) \\
&\quad \approx 0.9213
\end{aligned}
$$

where $\Phi(x) = P\{N(0,1) \le x\}$ is the standard normal distribution function.

(b) Because we must compute $P\{Y(1/2) > 0|Y(1) = \sigma\}$, let us first determine the conditional distribution of $Y(s)$ given that $Y(t) = C$, when $s < t$. Now, since $\{X(t),\ t \ge 0\}$ is standard Brownian motion when $X(t) = Y(t)/\sigma$, we obtain from Eq. (10.4) that the conditional distribution of $X(s)$, given that $X(t) = C/\sigma$, is normal with mean $sC/t\sigma$ and variance $s(t-s)/t$. Hence, the conditional distribution of $Y(s) = \sigma X(s)$ given that $Y(t) = C$ is normal

with mean sC/t and variance $\sigma^2 s(t-s)/t$. Hence,

$$\begin{aligned} P\{Y(1/2) > 0 | Y(1) = \sigma\} &= P\{N(\sigma/2, \sigma^2/4) > 0\} \\ &= \Phi(1) \\ &\approx 0.8413 \end{aligned}$$

■

10.2 Hitting Times, Maximum Variable, and the Gambler's Ruin Problem

Let T_a denote the first time the Brownian motion process hits a. When $a > 0$ we will compute $P\{T_a \le t\}$ by considering $P\{X(t) \ge a\}$ and conditioning on whether or not $T_a \le t$. This gives

$$\begin{aligned} P\{X(t) \ge a\} = &P\{X(t) \ge a | T_a \le t\} P\{T_a \le t\} \\ &+ P\{X(t) \ge a | T_a > t\} P\{T_a > t\} \end{aligned} \tag{10.5}$$

Now if $T_a \le t$, then the process hits a at some point in $[0, t]$ and, by symmetry, it is just as likely to be above a or below a at time t. That is,

$$P\{X(t) \ge a | T_a \le t\} = \tfrac{1}{2}$$

As the second right-hand term of Eq. (10.5) is clearly equal to 0 (since, by continuity, the process value cannot be greater than a without having yet hit a), we see that

$$\begin{aligned} P\{T_a \le t\} &= 2P\{X(t) \ge a\} \\ &= \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t}\, dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty e^{-y^2/2}\, dy, \qquad a > 0 \end{aligned} \tag{10.6}$$

For $a < 0$, the distribution of T_a is, by symmetry, the same as that of T_{-a}. Hence, from Eq. (10.6) we obtain

$$P\{T_a \le t\} = \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^\infty e^{-y^2/2}\, dy \tag{10.7}$$

Another random variable of interest is the maximum value the process attains in $[0, t]$. Its distribution is obtained as follows: For $a > 0$

$$\begin{aligned} P\left\{\max_{0 \le s \le t} X(s) \ge a\right\} &= P\{T_a \le t\} \qquad \text{by continuity} \\ &= 2P\{X(t) \ge a\} \qquad \text{from (10.6)} \end{aligned}$$

$$= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{\infty} e^{-y^2/2}\, dy$$

Let us now consider the probability that Brownian motion hits A before $-B$ where $A > 0,\ B > 0$. To compute this we shall make use of the interpretation of Brownian motion as being a limit of the symmetric random walk. To start let us recall from the results of the gambler's ruin problem (see Section 4.5.1) that the probability that the symmetric random walk goes up A before going down B when each step is equally likely to be either up or down a distance Δx is (by Eq. (4.14) with $N = (A + B)/\Delta x, i = B/\Delta x$) equal to $B\Delta x/(A + B)\Delta x = B/(A + B)$.

Hence, upon letting $\Delta x \to 0$, we see that

$$P\{\text{up } A \text{ before down } B\} = \frac{B}{A + B}$$

10.3 Variations on Brownian Motion

10.3.1 Brownian Motion with Drift

We say that $\{X(t),\ t \geq 0\}$ is a Brownian motion process with drift coefficient μ and variance parameter σ^2 if

(i) $X(0) = 0$;
(ii) $\{X(t),\ t \geq 0\}$ has stationary and independent increments;
(iii) $X(t)$ is normally distributed with mean μt and variance $t\sigma^2$.

An equivalent definition is to let $\{B(t),\ t \geq 0\}$ be standard Brownian motion and then define

$$X(t) = \sigma B(t) + \mu t$$

It follows from this representation that $X(t)$ will also be a continuous function of t.

10.3.2 Geometric Brownian Motion

If $\{Y(t),\ t \geq 0\}$ is a Brownian motion process with drift coefficient μ and variance parameter σ^2, then the process $\{X(t),\ t \geq 0\}$ defined by

$$X(t) = e^{Y(t)}$$

is called *geometric Brownian motion*.

For a geometric Brownian motion process $\{X(t)\}$, let us compute the expected value of the process at time t given the history of the process up to time s. That is, for $s < t$, consider $E[X(t)|X(u),\ 0 \leq u \leq s]$. Now,

$$E[X(t)|X(u),\ 0 \leq u \leq s] = E[e^{Y(t)}|Y(u),\ 0 \leq u \leq s]$$

$$
\begin{aligned}
&= E[e^{Y(s)+Y(t)-Y(s)} | Y(u),\ 0 \le u \le s] \\
&= e^{Y(s)} E[e^{Y(t)-Y(s)} | Y(u),\ 0 \le u \le s] \\
&= X(s) E[e^{Y(t)-Y(s)}]
\end{aligned}
$$

where the next to last equality follows from the fact that $Y(s)$ is given, and the last equality from the independent increment property of Brownian motion. Now, the moment generating function of a normal random variable W is given by

$$E[e^{aW}] = e^{aE[W]+a^2 \mathrm{Var}(W)/2}$$

Hence, since $Y(t) - Y(s)$ is normal with mean $\mu(t-s)$ and variance $(t-s)\sigma^2$, it follows by setting $a = 1$ that

$$E[e^{Y(t)-Y(s)}] = e^{\mu(t-s)+(t-s)\sigma^2/2}$$

Thus, we obtain

$$E[X(t) | X(u),\ 0 \le u \le s] = X(s) e^{(t-s)(\mu+\sigma^2/2)} \tag{10.8}$$

Geometric Brownian motion is useful in the modeling of stock prices over time when you feel that the percentage changes are independent and identically distributed. For instance, suppose that X_n is the price of some stock at time n. Then, it might be reasonable to suppose that $X_n/X_{n-1}, n \ge 1$, are independent and identically distributed. Let

$$Y_n = X_n / X_{n-1}$$

and so

$$X_n = Y_n X_{n-1}$$

Iterating this equality gives

$$
\begin{aligned}
X_n &= Y_n Y_{n-1} X_{n-2} \\
&= Y_n Y_{n-1} Y_{n-2} X_{n-3} \\
&\ \ \vdots \\
&= Y_n Y_{n-1} \cdots Y_1 X_0
\end{aligned}
$$

Thus,

$$\log(X_n) = \sum_{i=1}^{n} \log(Y_i) + \log(X_0)$$

Since $\log(Y_i), i \ge 1$ are independent and identically distributed, $\{\log(X_n)\}$ will, when suitably normalized, approximately be Brownian motion with a drift, and so $\{X_n\}$ will be approximately geometric Brownian motion.

10.4 Pricing Stock Options

10.4.1 An Example in Options Pricing

In situations in which money is to be received or paid out in differing time periods, we must take into account the time value of money. That is, to be given the amount v a time t in the future is not worth as much as being given v immediately. The reason for this is that if we were immediately given v, then it could be loaned out with interest and so be worth more than v at time t. To take this into account, we will suppose that the time 0 value, also called the *present value*, of the amount v to be earned at time t is $ve^{-\alpha t}$. The quantity α is often called the discount factor. In economic terms, the assumption of the discount function $e^{-\alpha t}$ is equivalent to the assumption that we can earn interest at a continuously compounded rate of 100α percent per unit time.

We will now consider a simple model for pricing an option to purchase a stock at a future time at a fixed price.

Suppose the present price of a stock is \$100 per unit share, and suppose we know that after one time period it will be, in present value dollars, either \$200 or \$50 (see Fig. 10.1). It should be noted that the prices at time 1 are the present value (or time 0) prices. That is, if the discount factor is α, then the actual possible prices at time 1 are either $200e^{\alpha}$ or $50e^{\alpha}$. To keep the notation simple, we will suppose that all prices given are time 0 prices.

Suppose that for any y, at a cost of cy, you can purchase at time 0 the option to buy y shares of the stock at time 1 at a (time 0) cost of \$150 per share. Thus, for instance, if you do purchase this option and the stock rises to \$200, then you would exercise the option at time 1 and realize a gain of $\$200 - 150 = \$50$ for each of the y option units purchased. On the other hand, if the price at time 1 was \$50, then the option would be worthless at time 1. In addition, at a cost of $100x$ you can purchase x units of the stock at time 0, and this will be worth either $200x$ or $50x$ at time 1.

We will suppose that both x or y can be either positive or negative (or zero). That is, you can either buy or sell both the stock and the option. For instance, if x were negative then you would be selling $-x$ shares of the stock, yielding you a return of $-100x$, and you would then be responsible for buying $-x$ shares of the stock at time 1 at a cost of either \$200 or \$50 per share.

We are interested in determining the appropriate value of c, the unit cost of an option. Specifically, we will show that unless $c = 50/3$ there will be a combination of purchases that will always result in a positive gain.

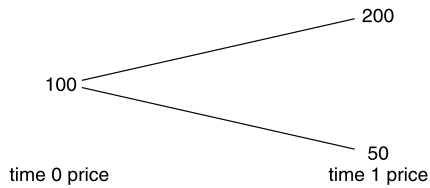

Figure 10.1

To show this, suppose that at time 0 we

buy x units of stock, and
buy y units of options

where x and y (which can be either positive or negative) are to be determined. The value of our holding at time 1 depends on the price of the stock at that time; and it is given by the following

$$\text{value} = \begin{cases} 200x + 50y, & \text{if price is 200} \\ 50x, & \text{if price is 50} \end{cases}$$

The preceding formula follows by noting that if the price is 200 then the x units of the stock are worth $200x$, and the y units of the option to buy the stock at a unit price of 150 are worth $(200 - 150)y$. On the other hand, if the stock price is 50, then the x units are worth $50x$ and the y units of the option are worthless. Now, suppose we choose y to be such that the preceding value is the same no matter what the price at time 1. That is, we choose y so that

$$200x + 50y = 50x$$

or

$$y = -3x$$

(Note that y has the opposite sign of x, and so if x is positive and as a result x units of the stock are purchased at time 0, then $3x$ units of stock options are also *sold* at that time. Similarly, if x is negative, then $-x$ units of stock are sold and $-3x$ units of stock options are purchased at time 0.)

Thus, with $y = -3x$, the value of our holding at time 1 is

$$\text{value} = 50x$$

Since the original cost of purchasing x units of the stock and $-3x$ units of options is

$$\text{original cost} = 100x - 3xc,$$

we see that our gain on the transaction is

$$\text{gain} = 50x - (100x - 3xc) = x(3c - 50)$$

Thus, if $3c = 50$, then the gain is 0; on the other hand if $3c \neq 50$, we can guarantee a positive gain (no matter what the price of the stock at time 1) by letting x be positive when $3c > 50$ and letting it be negative when $3c < 50$.

For instance, if the unit cost per option is $c = 20$, then purchasing 1 unit of the stock ($x = 1$) and simultaneously selling 3 units of the option ($y = -3$) initially costs us $100 - 60 = 40$. However, the value of this holding at time 1 is 50 whether the stock goes up to 200 or down to 50. Thus, a guaranteed profit of 10 is attained. Similarly,

if the unit cost per option is $c = 15$, then selling 1 unit of the stock ($x = -1$) and buying 3 units of the option ($y = 3$) leads to an initial gain of $100 - 45 = 55$. On the other hand, the value of this holding at time 1 is -50. Thus, a guaranteed profit of 5 is attained.

A sure win betting scheme is called an *arbitrage*. Thus, as we have just seen, the only option cost c that does not result in an arbitrage is $c = 50/3$.

10.4.2 The Arbitrage Theorem

Consider an experiment whose set of possible outcomes is $S = \{1, 2, \ldots, m\}$. Suppose that n wagers are available. If the amount x is bet on wager i, then the return $xr_i(j)$ is earned if the outcome of the experiment is j. In other words, $r_i(\cdot)$ is the return function for a unit bet on wager i. The amount bet on a wager is allowed to be either positive or negative or zero.

A betting scheme is a vector $\mathbf{x} = (x_1, \ldots, x_n)$ with the interpretation that x_1 is bet on wager 1, x_2 on wager 2, $\ldots$, and x_n on wager n. If the outcome of the experiment is j, then the return from the betting scheme $\mathbf{x}$ is

$$\text{return from } \mathbf{x} = \sum_{i=1}^{n} x_i r_i(j)$$

The following theorem states that either there exists a probability vector $\mathbf{p} = (p_1, \ldots, p_m)$ on the set of possible outcomes of the experiment under which each of the wagers has expected return 0, or else there is a betting scheme that guarantees a positive win.

Theorem 10.1 (The Arbitrage Theorem)**.** *Exactly one of the following is true: Either*

(i) *there exists a probability vector* $\mathbf{p} = (p_1, \ldots, p_m)$ *for which*

$$\sum_{j=1}^{m} p_j r_i(j) = 0, \qquad \text{for all } i = 1, \ldots, n$$

or

(ii) *there exists a betting scheme* $\mathbf{x} = (x_1, \ldots, x_n)$ *for which*

$$\sum_{i=1}^{n} x_i r_i(j) > 0, \qquad \text{for all } j = 1, \ldots, m$$

In other words, if X is the outcome of the experiment, then the arbitrage theorem states that either there is a probability vector $\mathbf{p}$ for X such that

$$E_{\mathbf{p}}[r_i(X)] = 0, \qquad \text{for all } i = 1, \ldots, n$$

or else there is a betting scheme that leads to a sure win.

Remark. This theorem is a consequence of the (linear algebra) theorem of the separating hyperplane, which is often used as a mechanism to prove the duality theorem of linear programming.

The theory of linear programming can be used to determine a betting strategy that guarantees the greatest return. Suppose that the absolute value of the amount bet on each wager must be less than or equal to 1. To determine the vector $\mathbf{x}$ that yields the greatest guaranteed win—call this win v—we need to choose $\mathbf{x}$ and v so as to maximize v, subject to the constraints

$$\begin{aligned} \sum_{i=1}^{n} x_i r_i(j) &\geq v, \qquad \text{for } j = 1, \ldots, m \\ -1 \leq x_i &\leq 1, \qquad i = 1, \ldots, n \end{aligned}$$

This optimization problem is a linear program and can be solved by standard techniques (such as by using the simplex algorithm). The arbitrage theorem yields that the optimal v will be positive unless there is a probability vector $\mathbf{p}$ for which $\sum_{j=1}^{m} p_j r_i(j) = 0$ for all $i = 1, \ldots, n$.

Example 10.2. In some situations, the only types of wagers allowed are to choose one of the outcomes $i, i = 1, \ldots, m$, and bet that i is the outcome of the experiment. The return from such a bet is often quoted in terms of "odds." If the odds for outcome i are o_i (often written as "o_i to 1") then a 1-unit bet will return o_i if the outcome of the experiment is i and will return -1 otherwise. That is,

$$r_i(j) = \begin{cases} o_i, & \text{if } j = i \\ -1 & \text{otherwise} \end{cases}$$

Suppose the odds $o_1, \ldots, o_m$ are posted. In order for there not to be a sure win there must be a probability vector $\mathbf{p} = (p_1, \ldots, p_m)$ such that

$$0 \equiv E_{\mathbf{p}}[r_i(X)] = o_i p_i - (1 - p_i)$$

That is, we must have

$$p_i = \frac{1}{1 + o_i}$$

Since the p_i must sum to 1, this means that the condition for there not to be an arbitrage is that

$$\sum_{i=1}^{m} (1 + o_i)^{-1} = 1$$

Thus, if the posted odds are such that $\sum_i (1+o_i)^{-1} \neq 1$, then a sure win is possible. For instance, suppose there are three possible outcomes and the odds are as follows:

Outcome	Odds
1	1
2	2
3	3

That is, the odds for outcome 1 are $1-1$, the odds for outcome 2 are $2-1$, and that for outcome 3 are $3-1$. Since

$$\tfrac{1}{2}+\tfrac{1}{3}+\tfrac{1}{4} > 1$$

a sure win is possible. One possibility is to bet -1 on outcome 1 (and so you either win 1 if the outcome is not 1 and lose 1 if the outcome is 1) and bet -0.7 on outcome 2, and -0.5 on outcome 3. If the experiment results in outcome 1, then we win $-1+0.7+0.5=0.2$; if it results in outcome 2, then we win $1-1.4+0.5=0.1$; if it results in outcome 3, then we win $1+0.7-1.5=0.2$. Hence, in all cases we win a positive amount. ■

Remark. If $\sum_i (1+o_i)^{-1} \neq 1$, then the betting scheme

$$x_i = \frac{(1+o_i)^{-1}}{1-\sum_i (1+o_i)^{-1}}, \qquad i=1,\ldots,n$$

will always yield a gain of exactly 1.

Example 10.3. Let us reconsider the option pricing example of the previous section, where the initial price of a stock is 100 and the present value of the price at time 1 is either 200 or 50. At a cost of c per share we can purchase at time 0 the option to buy the stock at time 1 at a present value price of 150 per share. The problem is to set the value of c so that no sure win is possible.

In the context of this section, the outcome of the experiment is the value of the stock at time 1. Thus, there are two possible outcomes. There are also two different wagers: to buy (or sell) the stock, and to buy (or sell) the option. By the arbitrage theorem, there will be no sure win if there is a probability vector $(p, 1-p)$ that makes the expected return under both wagers equal to 0.

Now, the return from purchasing 1 unit of the stock is

$$\text{return} = \begin{cases} 200-100=100, & \text{if the price is 200 at time 1} \\ 50-100=-50, & \text{if the price is 50 at time 1} \end{cases}$$

Hence, if p is the probability that the price is 200 at time 1, then

$$E[\text{return}] = 100p - 50(1-p)$$

Setting this equal to 0 yields

$$p = \tfrac{1}{3}$$

That is, the only probability vector $(p, 1-p)$ for which wager 1 yields an expected return 0 is the vector $(\frac{1}{3}, \frac{2}{3})$.

Now, the return from purchasing one share of the option is

$$\text{return} = \begin{cases} 50 - c, & \text{if price is } 200 \\ -c, & \text{if price is } 50 \end{cases}$$

Hence, the expected return when $p = \frac{1}{3}$ is

$$\begin{aligned} E[\text{return}] &= (50 - c)\tfrac{1}{3} - c\tfrac{2}{3} \\ &= \tfrac{50}{3} - c \end{aligned}$$

Thus, it follows from the arbitrage theorem that the only value of c for which there will not be a sure win is $c = \frac{50}{3}$, which verifies the result of Section 10.4.1. ■

10.4.3 The Black–Scholes Option Pricing Formula

Suppose the present price of a stock is $X(0) = x_0$, and let $X(t)$ denote its price at time t. Suppose we are interested in the stock over the time interval 0 to T. Assume that the discount factor is α (equivalently, the interest rate is 100α percent compounded continuously), and so the present value of the stock price at time t is $e^{-\alpha t}X(t)$.

We can regard the evolution of the price of the stock over time as our experiment, and thus the outcome of the experiment is the value of the function $X(t), 0 \le t \le T$. The types of wagers available are that for any $s < t$ we can observe the process for a time s and then buy (or sell) shares of the stock at price $X(s)$ and then sell (or buy) these shares at time t for the price $X(t)$. In addition, we will suppose that we may purchase any of N different options at time 0. Option i, costing c_i per share, gives us the option of purchasing shares of the stock at time t_i for the fixed price of K_i per share, $i = 1, \ldots, N$.

Suppose we want to determine values of the c_i for which there is no betting strategy that leads to a sure win. Assuming that the arbitrage theorem can be generalized (to handle the preceding situation, where the outcome of the experiment is a function), it follows that there will be no sure win if and only if there exists a probability measure over the set of outcomes under which all of the wagers have expected return 0. Let $\mathbf{P}$ be a probability measure on the set of outcomes. Consider first the wager of observing the stock for a time s and then purchasing (or selling) one share with the intention of selling (or purchasing) it at time $t, 0 \le s < t \le T$. The present value of the amount paid for the stock is $e^{-\alpha s}X(s)$, whereas the present value of the amount received is $e^{-\alpha t}X(t)$. Hence, in order for the expected return of this wager to be 0 when $\mathbf{P}$ is the probability measure on $X(t), 0 \le t \le T$, we must have

$$E_{\mathbf{P}}[e^{-\alpha t}X(t) | X(u), 0 \le u \le s] = e^{-\alpha s}X(s) \tag{10.9}$$

Consider now the wager of purchasing an option. Suppose the option gives us the right to buy one share of the stock at time t for a price K. At time t, the worth of this option

will be as follows:

$$\text{worth of option at time } t = \begin{cases} X(t) - K, & \text{if } X(t) \geq K \\ 0, & \text{if } X(t) < K \end{cases}$$

That is, the time t worth of the option is $(X(t) - K)^+$. Hence, the present value of the worth of the option is $e^{-\alpha t}(X(t) - K)^+$. If c is the (time 0) cost of the option, we see that, in order for purchasing the option to have expected (present value) return 0, we must have

$$E_{\mathbf{P}}[e^{-\alpha t}(X(t) - K)^+] = c \tag{10.10}$$

By the arbitrage theorem, if we can find a probability measure $\mathbf{P}$ on the set of outcomes that satisfies Eq. (10.9), then if c, the cost of an option to purchase one share at time t at the fixed price K, is as given in Eq. (10.10), then no arbitrage is possible. On the other hand, if for given prices $c_i, i = 1, \ldots, N$, there is no probability measure $\mathbf{P}$ that satisfies both (10.9) and the equality

$$c_i = E_{\mathbf{P}}[e^{-\alpha t_i}(X(t_i) - K_i)^+], \qquad i = 1, \ldots, N$$

then a sure win is possible.

We will now present a probability measure $\mathbf{P}$ on the outcome $X(t), 0 \leq t \leq T$, that satisfies Eq. (10.9).

Suppose that

$$X(t) = x_0 e^{Y(t)}$$

where $\{Y(t), t \geq 0\}$ is a Brownian motion process with drift coefficient μ and variance parameter σ^2. That is, $\{X(t), t \geq 0\}$ is a geometric Brownian motion process (see Section 10.3.2). From Eq. (10.8) we have that, for $s < t$,

$$E[X(t)|X(u), 0 \leq u \leq s] = X(s)e^{(t-s)(\mu+\sigma^2/2)}$$

Hence, if we choose μ and σ^2 so that

$$\mu + \sigma^2/2 = \alpha$$

then Eq. (10.9) will be satisfied. That is, by letting $\mathbf{P}$ be the probability measure governing the stochastic process $\{x_0 e^{Y(t)}, 0 \leq t \leq T\}$, where $\{Y(t)\}$ is Brownian motion with drift parameter μ and variance parameter σ^2, and where $\mu + \sigma^2/2 = \alpha$, Eq. (10.9) is satisfied.

It follows from the preceding that if we price an option to purchase a share of the stock at time t for a fixed price K by

$$c = E_{\mathbf{P}}[e^{-\alpha t}(X(t) - K)^+]$$

then no arbitrage is possible. Since $X(t) = x_0 e^{Y(t)}$, where $Y(t)$ is normal with mean μt and variance $t\sigma^2$, we see that

$$\begin{aligned} ce^{\alpha t} &= \int_{-\infty}^{\infty} (x_0 e^y - K)^+ \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(y-\mu t)^2/2t\sigma^2}\, dy \\ &= \int_{\log(K/x_0)}^{\infty} (x_0 e^y - K) \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(y-\mu t)^2/2t\sigma^2}\, dy \end{aligned}$$

Making the change of variable $w = (y - \mu t)/(\sigma t^{1/2})$ yields

$$ce^{\alpha t} = x_0 e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{\sigma w\sqrt{t}} e^{-w^2/2}\, dw - K \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-w^2/2}\, dw \tag{10.11}$$

where

$$a = \frac{\log(K/x_0) - \mu t}{\sigma\sqrt{t}}$$

Now,

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{\sigma w\sqrt{t}} e^{-w^2/2}\, dw &= e^{t\sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-(w-\sigma\sqrt{t})^2/2}\, dw \\ &= e^{t\sigma^2/2} P\{N(\sigma\sqrt{t}, 1) \geq a\} \\ &= e^{t\sigma^2/2} P\{N(0,1) \geq a - \sigma\sqrt{t}\} \\ &= e^{t\sigma^2/2} P\{N(0,1) \leq -(a - \sigma\sqrt{t})\} \\ &= e^{t\sigma^2/2} \phi(\sigma\sqrt{t} - a) \end{aligned}$$

where $N(m, v)$ is a normal random variable with mean m and variance v, and ϕ is the standard normal distribution function.

Thus, we see from Eq. (10.11) that

$$ce^{\alpha t} = x_0 e^{\mu t + \sigma^2 t/2} \phi(\sigma\sqrt{t} - a) - K\phi(-a)$$

Using that

$$\mu + \sigma^2/2 = \alpha$$

and letting $b = -a$, we can write this as follows:

$$c = x_0 \phi(\sigma\sqrt{t} + b) - Ke^{-\alpha t} \phi(b) \tag{10.12}$$

where

$$b = \frac{\alpha t - \sigma^2 t/2 - \log(K/x_0)}{\sigma\sqrt{t}}$$

The option price formula given by Eq. (10.12) depends on the initial price of the stock x_0, the option exercise time t, the option exercise price K, the discount (or interest rate) factor α, and the value σ^2. Note that for any value of σ^2, if the options are priced according to the formula of Eq. (10.12) then no arbitrage is possible. However, as many people believe that the price of a stock actually follows a geometric Brownian motion—that is, $X(t) = x_0 e^{Y(t)}$ where $Y(t)$ is Brownian motion with parameters μ and σ^2—it has been suggested that it is natural to price the option according to the formula of Eq. (10.12) with the parameter σ^2 taken equal to the estimated value (see the remark that follows) of the variance parameter under the assumption of a geometric Brownian motion model. When this is done, the formula of Eq. (10.12) is known as the *Black–Scholes option cost valuation*. It is interesting that this valuation does not depend on the value of the drift parameter μ but only on the variance parameter σ^2.

If the option itself can be traded, then the formula of Eq. (10.12) can be used to set its price in such a way so that no arbitrage is possible. If at time s the price of the stock is $X(s) = x_s$, then the price of a (t, K) option—that is, an option to purchase one unit of the stock at time t for a price K—should be set by replacing t by $t - s$ and x_0 by x_s in Eq. (10.12).

Remark. If we observe a Brownian motion process with variance parameter σ^2 over any time interval, then we could theoretically obtain an arbitrarily precise estimate of σ^2. For suppose we observe such a process $\{Y(s)\}$ for a time t. Then, for fixed h, let $N = [t/h]$ and set

$$\begin{aligned} W_1 &= Y(h) - Y(0), \\ W_2 &= Y(2h) - Y(h), \\ &\vdots \\ W_N &= Y(Nh) - Y(Nh - h) \end{aligned}$$

Then random variables $W_1, \ldots, W_N$ are independent and identically distributed normal random variables having variance $h\sigma^2$. We now use the fact (see Section 3.6.4) that $(N-1)S^2/(\sigma^2 h)$ has a chi-squared distribution with $N-1$ degrees of freedom, where S^2 is the sample variance defined by

$$S^2 = \sum_{i=1}^{N} (W_i - \bar{W})^2/(N-1)$$

Since the expected value and variance of a chi-squared with k degrees of freedom are equal to k and $2k$, respectively, we see that

$$E[(N-1)S^2/(\sigma^2 h)] = N - 1$$

and

$$\text{Var}[(N-1)S^2/(\sigma^2 h)] = 2(N-1)$$

From this, we see that

$$E[S^2/h] = \sigma^2$$

and

$$\text{Var}[S^2/h] = 2\sigma^4/(N-1)$$

Hence, as we let h become smaller (and so $N = [t/h]$ becomes larger) the variance of the unbiased estimator of σ^2 becomes arbitrarily small. ■

Eq. (10.12) is not the only way in which options can be priced so that no arbitrage is possible. Let $\{X(t), 0 \le t \le T\}$ be any stochastic process satisfying, for $s < t$,

$$E[e^{-\alpha t}X(t)|X(u),\ 0 \le u \le s] = e^{-\alpha s}X(s) \tag{10.13}$$

(that is, Eq. (10.9) is satisfied). By setting c, the cost of an option to purchase one share of the stock at time t for price K, equal to

$$c = E[e^{-\alpha t}(X(t) - K)^+] \tag{10.14}$$

it follows that no arbitrage is possible.

Another type of stochastic process, aside from geometric Brownian motion, that satisfies Eq. (10.13) is obtained as follows. Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables having a common mean μ, and suppose that this process is independent of $\{N(t),\ t \ge 0\}$, which is a Poisson process with rate λ. Let

$$X(t) = x_0 \prod_{i=1}^{N(t)} Y_i$$

Using the identity

$$X(t) = x_0 \prod_{i=1}^{N(s)} Y_i \prod_{j=N(s)+1}^{N(t)} Y_j$$

and the independent increment assumption of the Poisson process, we see that, for $s < t$,

$$E[X(t)|X(u),\ 0 \le u \le s] = X(s)\ E\left[\prod_{j=N(s)+1}^{N(t)} Y_j\right]$$

Conditioning on the number of events between s and t yields

$$E\left[\prod_{j=N(s)+1}^{N(t)} Y_j\right] = \sum_{n=0}^{\infty} \mu^n e^{-\lambda(t-s)}[\lambda(t-s)]^n/n!$$

$$= e^{-\lambda(t-s)(1-\mu)}$$

Hence,

$$E[X(t)|X(u),\ 0 \le u \le s] = X(s)e^{-\lambda(t-s)(1-\mu)}$$

Thus, if we choose λ and μ to satisfy

$$\lambda(1-\mu) = -\alpha$$

then Eq. (10.13) is satisfied. Therefore, if for any value of λ we let the Y_i have any distributions with a common mean equal to $\mu = 1 + \alpha/\lambda$ and then price the options according to Eq. (10.14), then no arbitrage is possible.

Remark. If $\{X(t), t \ge 0\}$ satisfies Eq. (10.13), then the process $\{e^{-\alpha t}X(t),\ t \ge 0\}$ is called a *Martingale*. Thus, any pricing of options for which the expected gain on the option is equal to 0 when $\{e^{-\alpha t}X(t)\}$ follows the probability law of some Martingale will result in no arbitrage possibilities.

That is, if we choose any Martingale process $\{Z(t)\}$ and let the cost of a (t, K) option be

$$\begin{aligned} c &= E[e^{-\alpha t}(e^{\alpha t}Z(t) - K)^+] \\ &= E[(Z(t) - Ke^{-\alpha t})^+] \end{aligned}$$

then there is no sure win.

In addition, while we did not consider the type of wager where a stock that is purchased at time s is sold not at a fixed time t but rather at some random time that depends on the movement of the stock, it can be shown using results about Martingales that the expected return of such wagers is also equal to 0.

Remark. A variation of the arbitrage theorem was first noted by de Finetti in 1937. A more general version of de Finetti's result, of which the arbitrage theorem is a special case, is given in Reference [3].

10.5 The Maximum of Brownian Motion with Drift

For $\{X(y), y \ge 0\}$ being a Brownian motion process with drift coefficient μ and variance parameter σ^2, define

$$M(t) = \max_{0 \le y \le t} X(y)$$

to be the maximal value of the process up to time t.

We will determine the distribution of $M(t)$ by deriving the conditional distribution of $M(t)$ given the value of $X(t)$. To do so, we first show that the conditional distribution of $X(y), 0 \le y \le t$, given the value of $X(t)$, does not depend on μ. That is, given

the value of the process at time t, the distribution of its history up to time t does not depend on μ.

We start with a lemma.

Lemma 10.1. *If $Y_1, \ldots, Y_n$ are independent and identically distributed normal random variables with mean θ and variance v^2, then the conditional distribution of $Y_1, \ldots, Y_n$ given that $\sum_{i=1}^n Y_i = x$ does not depend on θ.*

Proof. Because, given $\sum_{i=1}^n Y_i = x$, the value of Y_n is determined by knowledge of those of $Y_1, \ldots, Y_{n-1}$, it suffices to consider the conditional density of $Y_1, \ldots, Y_{n-1}$ given that $\sum_{i=1}^n Y_i = x$. Letting $X = \sum_{i=1}^n Y_i$, this is obtained as follows.

$$f_{Y_1,\ldots,Y_{n-1}|X}(y_1, \ldots, y_{n-1}|x) = \frac{f_{Y_1,\ldots,Y_{n-1},X}(y_1, \ldots, y_{n-1}, x)}{f_X(x)}$$

Now, because

$$Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}, X = x \Leftrightarrow Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}, Y_n = x - \sum_{i=1}^{n-1} y_i$$

it follows that

$$\begin{aligned} f_{Y_1,\ldots,Y_{n-1},X}(y_1, \ldots, y_{n-1}, x) &= f_{Y_1,\ldots,Y_{n-1},Y_n}(y_1, \ldots, y_{n-1}, x - \sum_{i=1}^{n-1} y_i) \\ &= f_{Y_1}(y_1) \cdots f_{Y_{n-1}}(y_{n-1}) f_{Y_n}(x - \sum_{i=1}^{n-1} y_i) \end{aligned}$$

where the last equality used that $Y_1, \ldots, Y_n$ are independent. Hence, using that $X = \sum_{i=1}^n Y_i$ is normal with mean $n\theta$ and variance nv^2, we obtain

$$\begin{aligned} f_{Y_1,\ldots,Y_{n-1}|X}(y_1, \ldots, y_{n-1}|x) &= \frac{f_{Y_n}(x - \sum_{i=1}^{n-1} y_i) f_{Y_1}(y_1) \cdots f_{Y_{n-1}}(y_{n-1})}{f_X(x)} \\ &= K \frac{e^{-(x-\sum_{i=1}^{n-1} y_i - \theta)^2/2v^2} \prod_{i=1}^{n-1} e^{-(y_i-\theta)^2/2v^2}}{e^{-(x-n\theta)^2/2nv^2}} \\ &= K \exp\{-\frac{1}{2v^2}[(x - \sum_{i=1}^{n-1} y_i - \theta)^2 \\ &\quad + \sum_{i=1}^{n-1} (y_i - \theta)^2 - (x - n\theta)^2/n]\} \end{aligned}$$

where K does not depend on θ. Expanding the squares in the preceding, and treating everything that does not depend on θ as a constant, shows that

$$
\begin{aligned}
&f_{Y_1,\ldots,Y_{n-1}|X}(y_1,\ldots,y_{n-1}|x)\\
&= K' \exp\{-\frac{1}{2v^2}[-2\theta(x-\sum_{i=1}^{n-1} y_i) + \theta^2 - 2\theta\sum_{i=1}^{n-1} y_i + (n-1)\theta^2 + 2\theta x - n\theta^2]\}\\
&= K'
\end{aligned}
$$

where $K' = K'(v, y_1, \ldots, y_{n-1}, x)$ is a function that does not depend on θ. Thus the result is proven. ■

Remark. Suppose that the distribution of random variables $Y_1, \ldots, Y_n$ depends on some parameter θ. Further, suppose that there is some function $D(Y_1, \ldots, Y_n)$ of $Y_1, \ldots, Y_n$ such that the conditional distribution of $Y_1, \ldots, Y_n$ given the value of $D(Y_1, \ldots, Y_n)$ does not depend on θ. Then it is said in statistical theory that $D(Y_1, \ldots, Y_n)$ is a *sufficient statistic* for θ. For suppose we wanted to use the data $Y_1, \ldots, Y_n$ to estimate the value of θ. Because, given the value of $D(Y_1, \ldots, Y_n)$, the conditional distribution of the data $Y_1, \ldots, Y_n$ does not depend on θ, it follows that if the value of $D(Y_1, \ldots, Y_n)$ is known then no additional information about θ can be obtained from knowing all the data values $Y_1, \ldots, Y_n$. Thus our preceding lemma proves that the sum of the data values of independent and identically distributed normal random variables is a sufficient statistic for their mean. (Because knowing the value of the sum is equivalent to knowing the value of $\sum_{i=1}^{n} Y_i/n$, called the *sample mean*, the common terminology in statistics is that the sample mean is a sufficient statistic for the mean of a normal population.) ■

Theorem 10.2. *Let $X(t), t \geq 0$ be a Brownian motion process with drift coefficient μ and variance parameter σ^2. Given that $X(t) = x$, the conditional distribution of $X(y), 0 \leq y \leq t$ is the same for all values of μ.*

Proof. Fix n and set $t_i = it/n$, $i = 1, \ldots, n$. To prove the theorem we will show for any n that the conditional distribution of $X(t_1), \ldots, X(t_n)$ given the value of $X(t)$ does not depend on μ. To do so, let $Y_1 = X(t_1)$, $Y_i = X(t_i) - X(t_{i-1}), i = 2, \ldots, n$ and note that $Y_1, \ldots, Y_n$ are independent and identically distributed normal random variables with mean $\theta = \mu t/n$. Because $\sum_{i=1}^{n} Y_i = X(t)$ it follows from Lemma 10.1 that the conditional distribution of $Y_1, \ldots, Y_n$ given $X(t)$ does not depend on μ. Because knowing $Y_1, \ldots, Y_n$ is equivalent to knowing $X(t_1), \ldots, X(t_n)$ the result follows. ■

We now derive the conditional distribution of $M(t)$ given the value of $X(t)$.

Theorem 10.3. *For $y > x$*

$$P(M(t) \geq y | X(t) = x) = e^{-2y(y-x)/t\sigma^2}, \quad y \geq 0$$

Proof. Because $X(0) = 0$ it follows that $M(t) \geq 0$, and so the result is true when $y = 0$ (since both sides are equal to 1 in this case). So suppose that $y > 0$. Because it follows from Theorem 10.2 that $P(M(t) \geq y | X(t) = x)$ does not depend on the value of μ, let us suppose that $\mu = 0$. Now, let T_y denote the first time that the Brownian motion reaches the value y, and note that it follows from the continuity property of Brownian

motion that the event that $M(t) \geq y$ is equivalent to the event that $T_y \leq t$. This is true because before the process can exceed the positive value y it must, by continuity, first pass through that value. Now, let h be a small positive number for which $y > x + h$. Then

$$\begin{aligned} P(M(t) \geq y,\ x \leq X(t) \leq x + h) &= P(T_y \leq t,\ x \leq X(t) \leq x + h) \\ &= P(x \leq X(t) \leq x + h | T_y \leq t) P(T_y \leq t) \end{aligned}$$

Now, given $T_y \leq t$, the event $x \leq X(t) \leq x + h$ will occur if after hitting y the process will decrease by an amount between $y - x - h$ and $y - x$ in the time between T_y and t. But because $\mu = 0$, in any period of time the process is just as likely to increase as it is to decrease by an amount between $y - x - h$ and $y - x$. Consequently,

$$P(x \leq X(t) \leq x + h | T_y \leq t) = P(2y - x - h \leq X(t) \leq 2y - x | T_y \leq t)$$

which gives that

$$\begin{aligned} P(M(t) \geq y,\ x \leq X(t) \leq x + h) &= P(2y - x - h \leq X(t) \leq 2y - x | T_y \leq t) \\ &\quad \times P(T_y \leq t) \\ &= P(2y - x - h \leq X(t) \leq 2y - x,\ T_y \leq t) \\ &= P(2y - x - h \leq X(t) \leq 2y - x) \end{aligned}$$

where the final equation follows because the assumption $y > x + h$ implies that $2y - x - h > y$ and so, by the continuity of Brownian motion, if $2y - x - h \leq X(t)$ then $T_y \leq t$. Hence,

$$\begin{aligned} P(M(t) \geq y | x \leq X(t) \leq x + h) &= \frac{P(2y - x - h \leq X(t) \leq 2y - x)}{P(x \leq X(t) \leq x + h)} \\ &= \frac{f_{X(t)}(2y - x)\, h + o(h)}{f_{X(t)}(x)\, h + o(h)} \\ &= \frac{f_{X(t)}(2y - x) + o(h)/h}{f_{X(t)}(x) + o(h)/h} \end{aligned}$$

where $f_{X(t)}$, the density function of $X(t)$, is the density function of a normal random variable with mean 0 and variance $t\sigma^2$. Letting $h \to 0$ in the preceding gives

$$\begin{aligned} P(M(t) \geq y | X(t) = x) &= \frac{f_{X(t)}(2y - x)}{f_{X(t)}(x)} \\ &= \frac{e^{-(2y-x)^2/2t\sigma^2}}{e^{-x^2/2t\sigma^2}} \\ &= e^{-2y(y-x)/t\sigma^2} \end{aligned}$$

■

With Z being a standard normal random variable, and Φ its distribution function, let

$$\bar{\Phi}(x) = 1 - \Phi(x) = P(Z > x)$$

We now have

Corollary 10.1.

$$P(M(t) \geq y) = e^{2y\mu/\sigma^2}\bar{\Phi}\left(\frac{\mu t + y}{\sigma\sqrt{t}}\right) + \bar{\Phi}\left(\frac{y - \mu t}{\sigma\sqrt{t}}\right)$$

Proof. Conditioning on $X(t)$ and using Theorem 10.3 yields

$$\begin{aligned}
P(M(t) \geq y) &= \int_{-\infty}^{\infty} P(M(t) \geq y | X(t) = x) f_{X(t)}(x) dx \\
&= \int_{-\infty}^{y} P(M(t) \geq y | X(t) = x) f_{X(t)}(x) dx + \int_{y}^{\infty} f_{X(t)}(x) dx \\
&= \int_{-\infty}^{y} e^{-2y(y-x)/t\sigma^2} \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-(x-\mu t)^2/2t\sigma^2} dx + P(X(t) > y) \\
&= \frac{1}{\sqrt{2\pi t}\,\sigma} e^{-2y^2/t\sigma^2} e^{-\mu^2 t^2/2t\sigma^2} \int_{-\infty}^{y} \exp\left\{-\frac{1}{2t\sigma^2}(x^2 - 2\mu t x \right. \\
&\quad \left. -4yx)\right\} dx + P(X(t) > y) \\
&= \frac{1}{\sqrt{2\pi t}\,\sigma} e^{-(4y^2+\mu^2t^2)/2t\sigma^2} \\
&\quad \times \int_{-\infty}^{y} \exp\left\{-\frac{1}{2t\sigma^2}(x^2 - 2x(\mu t + 2y))\right\} dx + P(X(t) > y)
\end{aligned}$$

Now,

$$x^2 - 2x(\mu t + 2y) = (x - (\mu t + 2y))^2 - (\mu t + 2y)^2$$

giving that

$$\begin{aligned}
P(M(t) \geq y) &= e^{-(4y^2+\mu^2t^2-(\mu t+2y)^2)/2t\sigma^2} \frac{1}{\sqrt{2\pi t}\,\sigma} \int_{-\infty}^{y} e^{-(x-\mu t-2y)^2/2t\sigma^2} dx \\
&\quad + P(X(t) > y)
\end{aligned}$$

Making the change of variable

$$w = \frac{x - \mu t - 2y}{\sigma\sqrt{t}}, \qquad dx = \sigma\sqrt{t}\, dw$$

gives

$$\begin{aligned}
P(M(t) \geq y) &= e^{2y\mu/\sigma^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\mu t - y}{\sigma\sqrt{t}}} e^{-w^2/2} dw + P(X(t) > y) \\
&= e^{2y\mu/\sigma^2}\Phi\left(\frac{-\mu t - y}{\sigma\sqrt{t}}\right) + P(X(t) > y)
\end{aligned}$$

$$= e^{2y\mu/\sigma^2}\bar{\Phi}\left(\frac{\mu t + y}{\sigma\sqrt{t}}\right) + \bar{\Phi}\left(\frac{y - \mu t}{\sigma\sqrt{t}}\right)$$

and the proof is complete. ■

In the proof of Theorem 10.3 we let T_y denote the first time the Brownian motion is equal to y. In addition, as previously noted, the continuity of Brownian motion implies that, for $y > 0$, the process would have hit y by time t if and only if the maximum of the process by time t was at least y. Consequently, for $y > 0$,

$$T_y \le t \Leftrightarrow M(t) \ge y$$

which, using Corollary 10.1, gives

$$P(T_y \le t) = e^{2y\mu/\sigma^2}\bar{\Phi}\left(\frac{y + \mu t}{\sigma\sqrt{t}}\right) + \bar{\Phi}\left(\frac{y - \mu t}{\sigma\sqrt{t}}\right), \quad y > 0$$

10.6 White Noise

Let $\{X(t), t \ge 0\}$ denote a standard Brownian motion process and let f be a function having a continuous derivative in the region $[a, b]$. The stochastic integral $\int_a^b f(t)\,dX(t)$ is defined as follows:

$$\int_a^b f(t)\,dX(t) \equiv \lim_{\substack{n\to\infty \\ \max(t_i - t_{i-1})\to 0}} \sum_{i=1}^{n} f(t_{i-1})[X(t_i) - X(t_{i-1})] \tag{10.15}$$

where $a = t_0 < t_1 < \cdots < t_n = b$ is a partition of the region $[a, b]$. Using the identity (the integration by parts formula applied to sums)

$$\sum_{i=1}^{n} f(t_{i-1})[X(t_i) - X(t_{i-1})]$$
$$= f(b)X(b) - f(a)X(a) - \sum_{i=1}^{n} X(t_i)[f(t_i) - f(t_{i-1})]$$

we see that

$$\int_a^b f(t)\,dX(t) = f(b)X(b) - f(a)X(a) - \int_a^b X(t)\,df(t) \tag{10.16}$$

Eq. (10.16) is usually taken as the definition of $\int_a^b f(t)\,dX(t)$.

By using the right side of Eq. (10.16) we obtain, upon assuming the interchangeability of expectation and limit, that

$$E\left[\int_a^b f(t)\,dX(t)\right] = 0$$

Also,

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^{n} f(t_{i-1})[X(t_i) - X(t_{i-1})]\right) &= \sum_{i=1}^{n} f^2(t_{i-1})\text{Var}[X(t_i) - X(t_{i-1})] \\ &= \sum_{i=1}^{n} f^2(t_{i-1})(t_i - t_{i-1})\end{aligned}$$

where the top equality follows from the independent increments of Brownian motion. Hence, we obtain from Eq. (10.15) upon taking limits of the preceding that

$$\text{Var}\left[\int_a^b f(t)\,dX(t)\right] = \int_a^b f^2(t)\,dt$$

Remark. The preceding gives operational meaning to the family of quantities $\{dX(t), 0 \le t < \infty\}$ by viewing it as an operator that carries functions f into the values $\int_a^b f(t)\,dX(t)$. This is called a *white noise transformation*, or more loosely $\{dX(t), 0 \le t < \infty\}$ is called white noise since it can be imagined that a time varying function f travels through a white noise medium to yield the output (at time b) $\int_a^b f(t)\,dX(t)$.

Example 10.4. Consider a particle of unit mass that is suspended in a liquid and suppose that, due to the liquid, there is a viscous force that retards the velocity of the particle at a rate proportional to its present velocity. In addition, let us suppose that the velocity instantaneously changes according to a constant multiple of white noise. That is, if $V(t)$ denotes the particle's velocity at t, suppose that

$$V'(t) = -\beta V(t) + \alpha X'(t)$$

where $\{X(t),\ t \ge 0\}$ is standard Brownian motion. This can be written as follows:

$$e^{\beta t}[V'(t) + \beta V(t)] = \alpha e^{\beta t} X'(t)$$

or

$$\frac{d}{dt}[e^{\beta t} V(t)] = \alpha e^{\beta t} X'(t)$$

Hence, upon integration, we obtain

$$e^{\beta t} V(t) = V(0) + \alpha \int_0^t e^{\beta s} X'(s)\,ds$$

or

$$V(t) = V(0)e^{-\beta t} + \alpha \int_0^t e^{-\beta(t-s)}\, dX(s)$$

Hence, from Eq. (10.16),

$$V(t) = V(0)e^{-\beta t} + \alpha\left[X(t) - \int_0^t X(s)\beta e^{-\beta(t-s)}\, ds\right] \qquad \blacksquare$$

10.7 Gaussian Processes

We start with the following definition.

Definition 10.2. A stochastic process $X(t), t \geq 0$ is called a *Gaussian*, or a *normal*, process if $X(t_1), \ldots, X(t_n)$ has a multivariate normal distribution for all $t_1, \ldots, t_n$.

If $\{X(t), t \geq 0\}$ is a Brownian motion process, then because each of $X(t_1)$, $X(t_2)$, $\ldots, X(t_n)$ can be expressed as a linear combination of the independent normal random variables $X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ it follows that Brownian motion is a Gaussian process.

Because a multivariate normal distribution is completely determined by the marginal mean values and the covariance values (see Section 2.6) it follows that standard Brownian motion could also be defined as a Gaussian process having $E[X(t)] = 0$ and, for $s \leq t$,

$$\begin{aligned}
\text{Cov}(X(s), X(t)) &= \text{Cov}(X(s), X(s) + X(t) - X(s)) \\
&= \text{Cov}(X(s), X(s)) + \text{Cov}(X(s), X(t) - X(s)) \\
&= \text{Cov}(X(s), X(s)) \qquad \text{by independent increments} \\
&= s \qquad \text{since } \text{Var}(X(s)) = s
\end{aligned} \tag{10.17}$$

Let $\{X(t),\ t \geq 0\}$ be a standard Brownian motion process and consider the process values between 0 and 1 conditional on $X(1) = 0$. That is, consider the conditional stochastic process $\{X(t),\ 0 \leq t \leq 1 | X(1) = 0\}$. Since the conditional distribution of $X(t_1), \ldots, X(t_n)$ is multivariate normal it follows that this conditional process, known as the *Brownian bridge* (as it is tied down both at 0 and at 1), is a Gaussian process. Let us compute its covariance function. As, from Eq. (10.4),

$$E[X(s)|X(1) = 0] = 0, \qquad \text{for } s < 1$$

we have that, for $s < t < 1$,

$$\begin{aligned}
&\text{Cov}[(X(s), X(t))|X(1) = 0] \\
&\quad = E[X(s)X(t)|X(1) = 0] \\
&\quad = E[E[X(s)X(t)|X(t), X(1) = 0]|X(1) = 0]
\end{aligned}$$

$$
\begin{aligned}
&= E[X(t)E[X(s)|X(t)]|X(1) = 0] \\
&= E\left[X(t)\frac{s}{t}X(t)|X(1) = 0\right] \qquad \text{by (10.4)} \\
&= \frac{s}{t}E[X^2(t)|X(1) = 0] \\
&= \frac{s}{t}t(1-t) \qquad \text{by (10.4)} \\
&= s(1-t)
\end{aligned}
$$

Thus, the Brownian bridge can be defined as a Gaussian process with mean value 0 and covariance function $s(1-t), s \le t$. This leads to an alternative approach to obtaining such a process.

Proposition 10.1. *If* $\{X(t),\ t \ge 0\}$ *is standard Brownian motion, then* $\{Z(t),\ 0 \le t \le 1\}$ *is a Brownian bridge process when* $Z(t) = X(t) - tX(1)$.

Proof. As it is immediate that $\{Z(t),\ t \ge 0\}$ is a Gaussian process, all we need verify is that $E[Z(t)] = 0$ and $\text{Cov}(Z(s), Z(t)) = s(1-t)$, when $s \le t$. The former is immediate and the latter follows from

$$
\begin{aligned}
\text{Cov}(Z(s), Z(t)) &= \text{Cov}(X(s) - sX(1), X(t) - tX(1)) \\
&= \text{Cov}(X(s), X(t)) - t\text{Cov}(X(s), X(1)) \\
&\quad - s\,\text{Cov}(X(1), X(t)) + st\,\text{Cov}(X(1), X(1)) \\
&= s - st - st + st \\
&= s(1-t)
\end{aligned}
$$

and the proof is complete. ■

If $\{X(t),\ t \ge 0\}$ is Brownian motion, then the process $\{Z(t),\ t \ge 0\}$ defined by

$$
Z(t) = \int_0^t X(s)\,ds \tag{10.18}
$$

is called *integrated Brownian motion*. As an illustration of how such a process may arise in practice, suppose we are interested in modeling the price of a commodity throughout time. Letting $Z(t)$ denote the price at t then, rather than assuming that $\{Z(t)\}$ is Brownian motion (or that log $Z(t)$ is Brownian motion), we might want to assume that the rate of change of $Z(t)$ follows a Brownian motion. For instance, we might suppose that the rate of change of the commodity's price is the current inflation rate, which is imagined to vary as Brownian motion. Hence,

$$
\begin{aligned}
&\frac{d}{dt}Z(t) = X(t), \\
&Z(t) = Z(0) + \int_0^t X(s)\,ds
\end{aligned}
$$

It follows from the fact that Brownian motion is a Gaussian process that $\{Z(t), t \geq 0\}$ is also Gaussian. To prove this, first recall that $W_1, \ldots, W_n$ is said to have a multivariate normal distribution if they can be represented as

$$W_i = \sum_{j=1}^{m} a_{ij} U_j, \qquad i = 1, \ldots, n$$

where U_j, $j = 1, \ldots, m$ are independent normal random variables. From this it follows that any set of partial sums of $W_1, \ldots, W_n$ are also jointly normal. The fact that $Z(t_1), \ldots, Z(t_n)$ is multivariate normal can now be shown by writing the integral in Eq. (10.18) as a limit of approximating sums.

As $\{Z(t),\ t \geq 0\}$ is Gaussian it follows that its distribution is characterized by its mean value and covariance function. We now compute these when $\{X(t),\ t \geq 0\}$ is standard Brownian motion.

$$\begin{aligned} E[Z(t)] &= E\left[\int_0^t X(s)\, ds\right] \\ &= \int_0^t E[X(s)]\, ds \\ &= 0 \end{aligned}$$

For $s \leq t$,

$$\begin{aligned} \mathrm{Cov}[Z(s), Z(t)] &= E[Z(s)Z(t)] \\ &= E\left[\int_0^t X(y)\, dy \int_0^s X(u)\, du\right] \\ &= E\left[\int_0^s \int_0^t X(y)X(u)\, dy\, du\right] \\ &= \int_0^s \int_0^t E[X(y)X(u)]\, dy\, du \\ &= \int_0^s \int_0^t \min(y, u)\, dy\, du \qquad \text{by (10.17)} \\ &= \int_0^s \left(\int_0^u y\, dy + \int_u^t u\, dy\right) du = s^2\left(\frac{t}{2} - \frac{s}{6}\right) \end{aligned}$$

■

10.8 Stationary and Weakly Stationary Processes

A stochastic process $\{X(t), t \geq 0\}$ is said to be a *stationary process* if for all n, s, $t_1, \ldots, t_n$ the random vectors $X(t_1), \ldots, X(t_n)$ and $X(t_1 + s), \ldots, X(t_n + s)$ have the same joint distribution. In other words, a process is stationary if, in choosing any fixed point s as the origin, the ensuing process has the same probability law. Two examples of stationary processes are:

(i) An ergodic continuous-time Markov chain $\{X(t),\ t \geq 0\}$ when

$$P\{X(0) = j\} = P_j, \qquad j \geq 0$$

where $\{P_j,\ j \geq 0\}$ are the limiting probabilities.

(ii) $\{X(t),\ t \geq 0\}$ when $X(t) = N(t + L) - N(t),\ t \geq 0$, where $L > 0$ is a fixed constant and $\{N(t),\ t \geq 0\}$ is a Poisson process having rate λ.

The first one of these processes is stationary for it is a Markov chain whose initial state is chosen according to the limiting probabilities, and it can thus be regarded as an ergodic Markov chain that we start observing at time ∞. Hence, the continuation of this process at time s after observation begins is just the continuation of the chain starting at time $\infty + s$, which clearly has the same probability for all s. That the second example—where $X(t)$ represents the number of events of a Poisson process that occur between t and $t + L$—is stationary follows from the stationary and independent increment assumption of the Poisson process, which implies that the continuation of a Poisson process at any time s remains a Poisson process.

Example 10.5 (The Random Telegraph Signal Process)**.** Let $\{N(t), t \geq 0\}$ denote a Poisson process, and let X_0 be independent of this process and be such that $P\{X_0 = 1\} = P\{X_0 = -1\} = \frac{1}{2}$. Defining $X(t) = X_0(-1)^{N(t)}$ then $\{X(t), t \geq 0\}$ is called a *random telegraph signal* process. To see that it is stationary, note first that starting at any time t, no matter what the value of $N(t)$, as X_0 is equally likely to be either plus or minus 1, it follows that $X(t)$ is equally likely to be either plus or minus 1. Hence, because the continuation of a Poisson process beyond any time remains a Poisson process, it follows that $\{X(t), t \geq 0\}$ is a stationary process.

Let us compute the mean and covariance function of the random telegraph signal.

$$\begin{aligned}
E[X(t)] &= E[X_0(-1)^{N(t)}] \\
&= E[X_0]E[(-1)^{N(t)}] \qquad \text{by independence} \\
&= 0 \qquad \text{since } E[X_0] = 0, \\
\text{Cov}[X(t), X(t+s)] &= E[X(t)X(t+s)] \\
&= E[X_0^2(-1)^{N(t)+N(t+s)}] \\
&= E[(-1)^{2N(t)}(-1)^{N(t+s)-N(t)}] \\
&= E[(-1)^{N(t+s)-N(t)}] \\
&= E[(-1)^{N(s)}] \\
&= \sum_{i=0}^{\infty} (-1)^i e^{-\lambda s} \frac{(\lambda s)^i}{i!} \\
&= e^{-2\lambda s}
\end{aligned} \tag{10.19}$$

For an application of the random telegraph signal consider a particle moving at a constant unit velocity along a straight line and suppose that collisions involving this particle occur at a Poisson rate λ. Also suppose that each time the particle suffers a

collision it reverses direction. Therefore, if X_0 represents the initial velocity of the particle, then its velocity at time t—call it $X(t)$—is given by $X(t) = X_0(-1)^{N(t)}$, where $N(t)$ denotes the number of collisions involving the particle by time t. Hence, if X_0 is equally likely to be plus or minus 1, and is independent of $\{N(t),\ t \geq 0\}$, then $\{X(t),\ t \geq 0\}$ is a random telegraph signal process. If we now let

$$D(t) = \int_0^t X(s)\,ds$$

then $D(t)$ represents the displacement of the particle at time t from its position at time 0. The mean and variance of $D(t)$ are obtained as follows:

$$\begin{aligned}
E[D(t)] &= \int_0^t E[X(s)]\,ds = 0,\\
\mathrm{Var}[D(t)] &= E[D^2(t)]\\
&= E\left[\int_0^t X(y)\,dy \int_0^t X(u)\,du\right]\\
&= \int_0^t \int_0^t E[X(y)X(u)]\,dy\,du\\
&= 2 \iint\limits_{0<y<u<t} E[X(y)X(u)]\,dy\,du\\
&= 2\int_0^t \int_0^u e^{-2\lambda(u-y)}\,dy\,du \qquad \text{by (10.19)}\\
&= \frac{1}{\lambda}\left(t - \frac{1}{2\lambda} + \frac{1}{2\lambda}e^{-2\lambda t}\right)
\end{aligned}$$

■

The condition for a process to be stationary is rather stringent and so we define the process $\{X(t),\ t \geq 0\}$ to be a *second-order stationary* or a *weakly stationary* process if $E[X(t)] = c$ and $\mathrm{Cov}[X(t), X(t+s)]$ does not depend on t. That is, a process is second-order stationary if the first two moments of $X(t)$ are the same for all t and the covariance between $X(s)$ and $X(t)$ depends only on $|t-s|$. For a second-order stationary process, let

$$R(s) = \mathrm{Cov}[X(t), X(t+s)]$$

As the finite dimensional distributions of a Gaussian process (being multivariate normal) are determined by their means and covariance, it follows that a second-order stationary Gaussian process is stationary.

Example 10.6 (The Ornstein–Uhlenbeck Process)**.** Let $\{X(t),\ t \geq 0\}$ be a standard Brownian motion process, and define, for $\alpha > 0$,

$$V(t) = e^{-\alpha t/2} X(e^{\alpha t})$$

The process $\{V(t),\ t \geq 0\}$ is called the *Ornstein–Uhlenbeck process*. It has been proposed as a model for describing the velocity of a particle immersed in a liquid or gas, and as such is useful in statistical mechanics. Let us compute its mean and covariance function.

$$\begin{aligned} E[V(t)] &= 0, \\ \mathrm{Cov}[V(t), V(t+s)] &= e^{-\alpha t/2} e^{-\alpha(t+s)/2} \\ \mathrm{Cov}[X(e^{\alpha t}), X(e^{\alpha(t+s)})] &= e^{-\alpha t} e^{-\alpha s/2} e^{\alpha t} \qquad \text{by Eq. (10.17)} \\ &= e^{-\alpha s/2} \end{aligned}$$

Hence, $\{V(t),\ t \geq 0\}$ is weakly stationary and as it is clearly a Gaussian process (since Brownian motion is Gaussian) we can conclude that it is stationary. It is interesting to note that (with $\alpha = 4\lambda$) it has the same mean and covariance function as the random telegraph signal process, thus illustrating that two quite different processes can have the same second-order properties. (Of course, if two Gaussian processes have the same mean and covariance functions then they are identically distributed.) ■

As the following examples show, there are many types of second-order stationary processes that are not stationary.

Example 10.7 (An Autoregressive Process)**.** Let $Z_0, Z_1, Z_2, \ldots$ be uncorrelated random variables with $E[Z_n] = 0, n \geq 0$ and

$$\mathrm{Var}(Z_n) = \begin{cases} \sigma^2/(1-\lambda^2), & n = 0 \\ \sigma^2, & n \geq 1 \end{cases}$$

where $\lambda^2 < 1$. Define

$$\begin{aligned} X_0 &= Z_0, \\ X_n &= \lambda X_{n-1} + Z_n, \qquad n \geq 1 \end{aligned} \tag{10.20}$$

The process $\{X_n,\ n \geq 0\}$ is called a *first-order autoregressive process*. It says that the state at time n (that is, X_n) is a constant multiple of the state at time $n-1$ plus a random error term Z_n.

Iterating Eq. (10.20) yields

$$\begin{aligned} X_n &= \lambda(\lambda X_{n-2} + Z_{n-1}) + Z_n \\ &= \lambda^2 X_{n-2} + \lambda Z_{n-1} + Z_n \\ &\ \ \vdots \\ &= \sum_{i=0}^{n} \lambda^{n-i} Z_i \end{aligned}$$

and so

$$\begin{aligned}\text{Cov}(X_n, X_{n+m}) &= \text{Cov}\left(\sum_{i=0}^{n}\lambda^{n-i}Z_i, \sum_{i=0}^{n+m}\lambda^{n+m-i}Z_i\right)\\ &= \sum_{i=0}^{n}\lambda^{n-i}\lambda^{n+m-i}\text{Cov}(Z_i, Z_i)\\ &= \sigma^2\lambda^{2n+m}\left(\frac{1}{1-\lambda^2} + \sum_{i=1}^{n}\lambda^{-2i}\right)\\ &= \frac{\sigma^2\lambda^m}{1-\lambda^2}\end{aligned}$$

where the preceding uses the fact that Z_i and Z_j are uncorrelated when $i \neq j$. As $E[X_n] = 0$, we see that $\{X_n,\ n \geq 0\}$ is weakly stationary (the definition for a discrete time process is the obvious analog of that given for continuous time processes). ■

Example 10.8. If, in the random telegraph signal process, we drop the requirement that $P\{X_0 = 1\} = P\{X_0 = -1\} = \frac{1}{2}$ and only require that $E[X_0] = 0$, then the process $\{X(t),\ t \geq 0\}$ need no longer be stationary. (It will remain stationary if X_0 has a symmetric distribution in the sense that $-X_0$ has the same distribution as X_0.) However, the process will be weakly stationary since

$$\begin{aligned}E[X(t)] &= E[X_0]E[(-1)^{N(t)}] = 0,\\ \text{Cov}[X(t), X(t+s)] &= E[X(t)X(t+s)]\\ &= E[X_0^2]E[(-1)^{N(t)+N(t+s)}]\\ &= E[X_0^2]e^{-2\lambda s} \qquad \text{from (10.19)}\end{aligned}$$ ■

Example 10.9. Let $W_0, W_1, W_2, \ldots$ be uncorrelated with $E[W_n] = \mu$ and $\text{Var}(W_n) = \sigma^2, n \geq 0$, and for some positive integer k define

$$X_n = \frac{W_n + W_{n-1} + \cdots + W_{n-k}}{k+1}, \qquad n \geq k$$

The process $\{X_n, n \geq k\}$, which at each time keeps track of the arithmetic average of the most recent $k+1$ values of the Ws, is called a *moving average process*. Using the fact that the $W_n, n \geq 0$ are uncorrelated, we see that

$$\text{Cov}(X_n, X_{n+m}) = \begin{cases}\frac{(k+1-m)\sigma^2}{(k+1)^2}, & \text{if } 0 \leq m \leq k\\ 0, & \text{if } m > k\end{cases}$$

Hence, $\{X_n,\ n \geq k\}$ is a second-order stationary process. ■

Let $\{X_n,\ n \geq 1\}$ be a second-order stationary process with $E[X_n] = \mu$. An important question is when, if ever, does $\bar{X}_n \equiv \sum_{i=1}^{n} X_i/n$ converge to μ? The following

proposition, which we state without proof, shows that $E[(\bar{X}_n - \mu)^2] \to 0$ if and only if $\sum_{i=1}^{n} R(i)/n \to 0$. That is, the expected square of the difference between $\bar{X}_n$ and μ will converge to 0 if and only if the limiting average value of $R(i)$ converges to 0.

Proposition 10.2. *Let* $\{X_n, n \geq 1\}$ *be a second-order stationary process having mean* μ *and covariance function* $R(i) = \text{Cov}(X_n, X_{n+i})$, *and let* $\bar{X}_n \equiv \sum_{i=1}^{n} X_i/n$. *Then* $\lim_{n\to\infty} E[(\bar{X}_n - \mu)^2] = 0$ *if and only if* $\lim_{n\to\infty} \sum_{i=1}^{n} R(i)/n = 0$.

10.9 Harmonic Analysis of Weakly Stationary Processes

Suppose that the stochastic processes $\{X(t), \ -\infty < t < \infty\}$ and $\{Y(t), -\infty < t < \infty\}$ are related as follows:

$$Y(t) = \int_{-\infty}^{\infty} X(t-s)h(s)\,ds \tag{10.21}$$

We can imagine that a signal, whose value at time t is $X(t)$, is passed through a physical system that distorts its value so that $Y(t)$, the received value at t, is given by Eq. (10.21). The processes $\{X(t)\}$ and $\{Y(t)\}$ are called, respectively, the input and output processes. The function h is called the *impulse response* function. If $h(s) = 0$ whenever $s < 0$, then h is also called a weighting function since Eq. (10.21) expresses the output at t as a weighted integral of all the inputs prior to t with $h(s)$ representing the weight given the input s time units ago.

The relationship expressed by Eq. (10.21) is a special case of a time invariant linear filter. It is called a filter because we can imagine that the input process $\{X(t)\}$ is passed through a medium and then filtered to yield the output process $\{Y(t)\}$. It is a linear filter because if the input processes $\{X_i(t)\}, i = 1, 2$, result in the output processes $\{Y_i(t)\}$—that is, if $Y_i(t) = \int_0^\infty X_i(t-s)h(s)\,ds$—then the output process corresponding to the input process $\{aX_1(t) + bX_2(t)\}$ is just $\{aY_1(t) + bY_2(t)\}$. It is called time invariant since lagging the input process by a time τ—that is, considering the new input process $\bar{X}(t) = X(t+\tau)$—results in a lag of τ in the output process since

$$\int_0^\infty \bar{X}(t-s)h(s)\,ds = \int_0^\infty X(t+\tau-s)h(s)\,ds = Y(t+\tau)$$

Let us now suppose that the input process $\{X(t), \ -\infty < t < \infty\}$ is weakly stationary with $E[X(t)] = 0$ and covariance function

$$R_X(s) = \text{Cov}[X(t), X(t+s)]$$

Let us compute the mean value and covariance function of the output process $\{Y(t)\}$.

Assuming that we can interchange the expectation and integration operations (a sufficient condition being that $\int |h(s)| < \infty$ [1] and, for some $M < \infty$, $E[|X(t)|] < M$ for

[1] The range of all integrals in this section is from $-\infty$ to $+\infty$.

all t) we obtain

$$E[Y(t)] = \int E[X(t-s)]h(s)\,ds = 0$$

Similarly,

$$\begin{aligned}\mathrm{Cov}[Y(t_1), Y(t_2)] &= \mathrm{Cov}\left[\int X(t_1 - s_1)h(s_1)\,ds_1,\ \int X(t_2 - s_2)h(s_2)\,ds_2\right] \\ &= \iint \mathrm{Cov}[X(t_1 - s_1), X(t_2 - s_2)]h(s_1)h(s_2)\,ds_1\,ds_2 \\ &= \iint R_X(t_2 - s_2 - t_1 + s_1)h(s_1)h(s_2)\,ds_1\,ds_2 \qquad (10.22)\end{aligned}$$

Hence, $\mathrm{Cov}[Y(t_1),\ Y(t_2)]$ depends on t_1, t_2 only through $t_2 - t_1$; thus showing that $\{Y(t)\}$ is also weakly stationary.

The preceding expression for $R_Y(t_2 - t_1) = \mathrm{Cov}[Y(t_1),\ Y(t_2)]$ is, however, more compactly and usefully expressed in terms of Fourier transforms of R_X and R_Y. Let, for $i = \sqrt{-1}$,

$$\widetilde{R}_X(w) = \int e^{-iws} R_X(s)\,ds$$

and

$$\widetilde{R}_Y(w) = \int e^{-iws} R_Y(s)\,ds$$

denote the Fourier transforms, respectively, of R_X and R_Y. The function $\widetilde{R}_X(w)$ is also called the *power spectral density* of the process $\{X(t)\}$. Also, let

$$\widetilde{h}(w) = \int e^{-iws} h(s)\,ds$$

denote the Fourier transform of the function h. Then, from Eq. (10.22),

$$\begin{aligned}\widetilde{R}_Y(w) &= \iiint e^{iws} R_X(s - s_2 + s_1)h(s_1)h(s_2)\,ds_1\,ds_2\,ds \\ &= \iiint e^{iw(s-s_2+s_1)} R_X(s - s_2 + s_1)\,ds e^{-iws_2} h(s_2)\,ds_2 e^{iws_1} h(s_1)\,ds_1 \\ &= \widetilde{R}_X(w)\widetilde{h}(w)\widetilde{h}(-w) \qquad (10.23)\end{aligned}$$

Now, using the representation

$$\begin{aligned}e^{ix} &= \cos x + i \sin x, \\ e^{-ix} &= \cos(-x) + i \sin(-x) = \cos x - i \sin x\end{aligned}$$

we obtain

$$\begin{aligned}\widetilde{h}(w)\widetilde{h}(-w) &= \left[\int h(s)\cos(ws)\,ds - i\int h(s)\sin(ws)\,ds\right] \\ &\quad \times \left[\int h(s)\cos(ws)\,ds + i\int h(s)\sin(ws)\,ds\right] \\ &= \left[\int h(s)\cos(ws)\,ds\right]^2 + \left[\int h(s)\sin(ws)\,ds\right]^2 \\ &= \left|\int h(s)e^{-iws}\,ds\right|^2 = |\widetilde{h}(w)|^2\end{aligned}$$

Hence, from Eq. (10.23) we obtain

$$\widetilde{R}_Y(w) = \widetilde{R}_X(w)|\widetilde{h}(w)|^2$$

In words, the Fourier transform of the covariance function of the output process is equal to the square of the amplitude of the Fourier transform of the impulse function multiplied by the Fourier transform of the covariance function of the input process.

Exercises

In the following exercises $\{B(t),\ t \geq 0\}$ is a standard Brownian motion process and T_a denotes the time it takes this process to hit a.

***1.** What is the distribution of $B(s) + B(t)$, $s \leq t$?

2. Compute the conditional distribution of $B(s)$ given that $B(t_1) = A$ and $B(t_2) = B$, where $0 < t_1 < s < t_2$.

***3.** Compute $E[B(t_1)B(t_2)B(t_3)]$ for $t_1 < t_2 < t_3$.

4. Show that

$$\begin{aligned}&P\{T_a < \infty\} = 1, \\ &E[T_a] = \infty, \qquad a \neq 0\end{aligned}$$

***5.** What is $P\{T_1 < T_{-1} < T_2\}$?

6. Suppose you own one share of a stock whose price changes according to a standard Brownian motion process. Suppose that you purchased the stock at a price $b + c$, $c > 0$, and the present price is b. You have decided to sell the stock either when it reaches the price $b + c$ or when an additional time t goes by (whichever occurs first). What is the probability that you do not recover your purchase price?

7. Compute an expression for

$$P\left\{\max_{t_1 \leq s \leq t_2} B(s) > x\right\}$$

8. Consider the random walk that in each Δt time unit either goes up or down the amount $\sqrt{\Delta t}$ with respective probabilities p and $1-p$, where $p = \frac{1}{2}(1 + \mu\sqrt{\Delta t})$.
   (a) Argue that as $\Delta t \to 0$ the resulting limiting process is a Brownian motion process with drift rate μ.
   (b) Using part (a) and the results of the gambler's ruin problem (Section 4.5.1), compute the probability that a Brownian motion process with drift rate μ goes up A before going down B, $A > 0$, $B > 0$.

9. Let $\{X(t),\ t \geq 0\}$ be a Brownian motion process with drift coefficient μ and variance parameter σ^2. What is the joint density function of $X(s)$ and $X(t)$, $s < t$?

***10.** Let $\{X(t),\ t \geq 0\}$ be a Brownian motion process with drift coefficient μ and variance parameter σ^2. What is the conditional distribution of $X(t)$ given that $X(s) = c$ when
   (a) $s < t$?
   (b) $t < s$?

11. Consider a process whose value changes every h time units; its new value being its old value multiplied either by the factor $e^{\sigma\sqrt{h}}$ with probability $p = \frac{1}{2}(1 + \frac{\mu}{\sigma}\sqrt{h})$, or by the factor $e^{-\sigma\sqrt{h}}$ with probability $1-p$. As h goes to zero, show that this process converges to geometric Brownian motion with drift coefficient μ and variance parameter σ^2.

12. A stock is presently selling at a price of \$50 per share. After one time period, its selling price will (in present value dollars) be either \$150 or \$25. An option to purchase y units of the stock at time 1 can be purchased at cost cy.
   (a) What should c be in order for there to be no sure win?
   (b) If $c = 4$, explain how you could guarantee a sure win.
   (c) If $c = 10$, explain how you could guarantee a sure win.
   (d) Use the arbitrage theorem to verify your answer to part (a).

13. Verify the statement made in the remark following Example 10.2.

14. The present price of a stock is 100. The price at time 1 will be either 50, 100, or 200. An option to purchase y shares of the stock at time 1 for the (present value) price ky costs cy.
   (a) If $k = 120$, show that an arbitrage opportunity occurs if and only if $c > 80/3$.
   (b) If $k = 80$, show that there is not an arbitrage opportunity if and only if $20 \leq c \leq 40$.

15. The current price of a stock is 100. Suppose that the logarithm of the price of the stock changes according to a Brownian motion process with drift coefficient $\mu = 2$ and variance parameter $\sigma^2 = 1$. Give the Black–Scholes cost of an option to buy the stock at time 10 for a cost of
   (a) 100 per unit.
   (b) 120 per unit.
   (c) 80 per unit.

Assume that the continuously compounded interest rate is 5 percent.

A stochastic process $\{Y(t),\ t \geq 0\}$ is said to be a *Martingale* process if, for $s < t$,

$$E[Y(t)|Y(u),\ 0 \leq u \leq s] = Y(s)$$

16. If $\{Y(t),\ t \geq 0\}$ is a Martingale, show that

$$E[Y(t)] = E[Y(0)]$$

17. Show that standard Brownian motion is a Martingale.

18. Show that $\{Y(t),\ t \geq 0\}$ is a Martingale when

$$Y(t) = B^2(t) - t$$

What is $E[Y(t)]$?

Hint: First compute $E[Y(t)|B(u),\ 0 \leq u \leq s]$.

***19.** Show that $\{Y(t), t \geq 0\}$ is a Martingale when

$$Y(t) = \exp\{cB(t) - c^2t/2\}$$

where c is an arbitrary constant. What is $E[Y(t)]$?

An important property of a Martingale is that if you continually observe the process and then stop at some time T, then, subject to some technical conditions (which will hold in the problems to be considered),

$$E[Y(T)] = E[Y(0)]$$

The time T usually depends on the values of the process and is known as a *stopping time* for the Martingale. This result, that the expected value of the stopped Martingale is equal to its fixed time expectation, is known as the *Martingale stopping theorem*.

***20.** Let

$$T = \text{Min}\{t:\ B(t) = 2 - 4t\}$$

That is, T is the first time that standard Brownian motion hits the line $2 - 4t$. Use the Martingale stopping theorem to find $E[T]$.

21. Let $\{X(t),\ t \geq 0\}$ be Brownian motion with drift coefficient μ and variance parameter σ^2. That is,

$$X(t) = \sigma B(t) + \mu t$$

Let $\mu > 0$, and for a positive constant x let

$$T = \text{Min}\{t:\ X(t) = x\}$$

$$= \text{Min}\left\{t:\ B(t) = \frac{x - \mu t}{\sigma}\right\}$$

That is, T is the first time the process $\{X(t),\ t \geq 0\}$ hits x. Use the Martingale stopping theorem to show that

$$E[T] = x/\mu$$

22. Let $X(t) = \sigma B(t) + \mu t$, and for given positive constants A and B, let p denote the probability that $\{X(t),\ t \geq 0\}$ hits A before it hits $-B$.

(a) Define the stopping time T to be the first time the process hits either A or $-B$. Use this stopping time and the Martingale defined in Exercise 19 to show that

$$E[\exp\{c(X(T) - \mu T)/\sigma - c^2 T/2\}] = 1$$

(b) Let $c = -2\mu/\sigma$, and show that

$$E[\exp\{-2\mu X(T)/\sigma\}] = 1$$

(c) Use part (b) and the definition of T to find p.

Hint: What are the possible values of $\exp\{-2\mu X(T)/\sigma^2\}$?

23. Let $X(t) = \sigma B(t) + \mu t$, and define T to be the first time the process $\{X(t),\ t \geq 0\}$ hits either A or $-B$, where A and B are given positive numbers. Use the Martingale stopping theorem and part (c) of Exercise 22 to find $E[T]$.

***24.** Let $\{X(t),\ t \geq 0\}$ be Brownian motion with drift coefficient μ and variance parameter σ^2. Suppose that $\mu > 0$. Let $x > 0$ and define the stopping time T (as in Exercise 21) by

$$T = \text{Min}\{t:\ X(t) = x\}$$

Use the Martingale defined in Exercise 18, along with the result of Exercise 21, to show that

$$\text{Var}(T) = x\sigma^2/\mu^3$$

In Exercises 25 to 27, $\{X(t), t \geq 0\}$ is a Brownian motion process with drift parameter μ and variance parameter σ^2.

25. Suppose every Δ time units a process either increases by the amount $\sigma\sqrt{\Delta}$ with probability p or decreases by the amount $\sigma\sqrt{\Delta}$ with probability $1 - p$ where

$$p = \frac{1}{2}(1 + \frac{\mu}{\sigma}\sqrt{\Delta}).$$

Show that as Δ goes to 0, this process converges to a Brownian motion process with drift parameter μ and variance parameter σ^2.

26. Let T_y be the first time that the process is equal to y. For $y > 0$, show that

$$P(T_y < \infty) = \begin{cases} 1, & \text{if } \mu \geq 0 \\ e^{2y\mu/\sigma^2}, & \text{if } \mu < 0 \end{cases}$$

Let $M = \max_{0 \leq t < \infty} X(t)$ be the maximal value ever attained. Explain why the preceding implies that, when $\mu < 0$, M is an exponential random variable with rate $-2\mu/\sigma^2$.

27. Determine the distribution function of $\min_{0 \leq y \leq t} X(y)$.

28. Compute the mean and variance of

(a) $\int_0^1 t\, dB(t)$

(b) $\int_0^1 t^2\, dB(t)$

29. Let $Y(t) = tB(1/t), t > 0$ and $Y(0) = 0$.

(a) What is the distribution of $Y(t)$?

(b) Compare $\text{Cov}(Y(s), Y(t))$.

(c) Argue that $\{Y(t),\ t \geq 0\}$ is a standard Brownian motion process.

30. Let $Y(t) = B(a^2t)/a$ for $a > 0$. Argue that $\{Y(t)\}$ is a standard Brownian motion process.

31. For $s < t$, argue that $B(s) - \frac{s}{t}B(t)$ and $B(t)$ are independent.

32. Let $\{Z(t),\ t \geq 0\}$ denote a Brownian bridge process. Show that if

$$Y(t) = (t+1)Z(t/(t+1))$$

then $\{Y(t),\ t \geq 0\}$ is a standard Brownian motion process.

33. Let $X(t) = N(t+1) - N(t)$ where $\{N(t),\ t \geq 0\}$ is a Poisson process with rate λ. Compute

$$\text{Cov}[X(t), X(t+s)]$$

34. Let $\{N(t), t \geq 0\}$ denote a Poisson process with rate λ and define $Y(t)$ to be the time from t until the next Poisson event.

(a) Argue that $\{Y(t),\ t \geq 0\}$ is a stationary process.

(b) Compute $\text{Cov}[Y(t), Y(t+s)]$.

35. Let $\{X(t), -\infty < t < \infty\}$ be a weakly stationary process having covariance function $R_X(s) = \text{Cov}[X(t), X(t+s)]$.

(a) Show that

$$\text{Var}(X(t+s) - X(t)) = 2R_X(0) - 2R_X(t)$$

(b) If $Y(t) = X(t+1) - X(t)$ show that $\{Y(t),\ -\infty < t < \infty\}$ is also weakly stationary having a covariance function $R_Y(s) = \text{Cov}[Y(t), Y(t+s)]$ that satisfies

$$R_Y(s) = 2R_X(s) - R_X(s-1) - R_X(s+1)$$

36. Let Y_1 and Y_2 be independent unit normal random variables and for some constant w set

$$X(t) = Y_1 \cos wt + Y_2 \sin wt, \qquad -\infty < t < \infty$$

(a) Show that $\{X(t)\}$ is a weakly stationary process.
(b) Argue that $\{X(t)\}$ is a stationary process.

37. Let $\{X(t),\ -\infty < t < \infty\}$ be weakly stationary with covariance function $R(s) = \text{Cov}(X(t), X(t+s))$ and let $\widetilde{R}(w)$ denote the power spectral density of the process.
(i) Show that $\widetilde{R}(w) = \widetilde{R}(-w)$. It can be shown that

$$R(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widetilde{R}(w) e^{iws}\, dw$$

(ii) Use the preceding to show that

$$\int_{-\infty}^{\infty} \widetilde{R}(w)\, dw = 2\pi E[X^2(t)]$$

References

[1] M.S. Bartlett, An Introduction to Stochastic Processes, Cambridge University Press, London, 1954.
[2] U. Grenander, M. Rosenblatt, Statistical Analysis of Stationary Time Series, John Wiley, New York, 1957.
[3] D. Heath, W. Sudderth, On a Theorem of De Finetti, Oddsmaking, and Game Theory, Ann. Math. Stat. 43 (1972) 2072–2077.
[4] S. Karlin, H. Taylor, A Second Course in Stochastic Processes, Academic Press, Orlando, FL, 1981.
[5] L.H. Koopmans, The Spectral Analysis of Time Series, Academic Press, Orlando, FL, 1974.
[6] S. Ross, Stochastic Processes, Second Edition, John Wiley, New York, 1996.

Simulation

11

11.1 Introduction

Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote a random vector having a given density function $f(x_1, \ldots, x_n)$ and suppose we are interested in computing

$$E[g(\mathbf{X})] = \iint \cdots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n$$

for some n-dimensional function g. For instance, g could represent the total delay in queue of the first $[n/2]$ customers when the X values represent the first $[n/2]$ interarrival and service times.[1] In many situations, it is not analytically possible either to compute the preceding multiple integral exactly or even to numerically approximate it within a given accuracy. One possibility that remains is to approximate $E[g(\mathbf{X})]$ by means of simulation.

To approximate $E[g(\mathbf{X})]$, start by generating a random vector $\mathbf{X}^{(1)} = \left(X_1^{(1)}, \ldots, X_n^{(1)}\right)$ having the joint density $f(x_1, \ldots, x_n)$ and then compute $Y^{(1)} = g\left(\mathbf{X}^{(1)}\right)$. Now generate a second random vector (independent of the first) $\mathbf{X}^{(2)}$ and compute $Y^{(2)} = g\left(\mathbf{X}^{(2)}\right)$. Keep on doing this until r, a fixed number of independent and identically distributed random variables $Y^{(i)} = g\left(\mathbf{X}^{(i)}\right), i = 1, \ldots, r$ have been generated. Now by the strong law of large numbers, we know that

$$\lim_{r \to \infty} \frac{Y^{(1)} + \cdots + Y^{(r)}}{r} = E[Y^{(i)}] = E[g(\mathbf{X})]$$

and so we can use the average of the generated Ys as an estimate of $E[g(\mathbf{X})]$. This approach to estimating $E[g(\mathbf{X})]$ is called the *Monte Carlo simulation* approach.

Clearly there remains the problem of how to generate, or *simulate*, random vectors having a specified joint distribution. The first step in doing this is to be able to generate random variables from a uniform distribution on $(0, 1)$. One way to do this would be to take 10 identical slips of paper, numbered $0, 1, \ldots, 9$, place them in a hat and then successively select n slips, with replacement, from the hat. The sequence of digits obtained (with a decimal point in front) can be regarded as the value of a uniform $(0, 1)$ random variable rounded off to the nearest $\left(\frac{1}{10}\right)^n$. For instance, if the sequence of digits selected is 3, 8, 7, 2, 1, then the value of the uniform $(0, 1)$ random variable is 0.38721 (to the nearest 0.00001). Tables of the values of uniform $(0, 1)$ random variables, known as random number tables, have been extensively published (for instance, see The RAND Corporation, *A Million Random Digits with 100,000 Normal Deviates* (New York: The Free Press, 1955)). Table 11.1 is such a table.

[1] We are using the notation $[a]$ to represent the largest integer less than or equal to a.

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00016-0

Table 11.1 A Random Number Table.

04839	96423	24878	82651	66566	14778	76797	14780	13300	87074
68086	26432	46901	20848	89768	81536	86645	12659	92259	57102
39064	66432	84673	40027	32832	61362	98947	96067	64760	64584
25669	26422	44407	44048	37937	63904	45766	66134	75470	66520
64117	94305	26766	25940	39972	22209	71500	64568	91402	42416
87917	77341	42206	35126	74087	99547	81817	42607	43808	76655
62797	56170	86324	88072	76222	36086	84637	93161	76038	65855
95876	55293	18988	27354	26575	08625	40801	59920	29841	80150
29888	88604	67917	48708	18912	82271	65424	69774	33611	54262
73577	12908	30883	18317	28290	35797	05998	41688	34952	37888
27958	30134	04024	86385	29880	99730	55536	84855	29080	09250
90999	49127	20044	59931	06115	20542	18059	02008	73708	83517
18845	49618	02304	51038	20655	58727	28168	15475	56942	53389
94824	78171	84610	82834	09922	25417	44137	48413	25555	21246
35605	81263	39667	47358	56873	56307	61607	49518	89356	20103
33362	64270	01638	92477	66969	98420	04880	45585	46565	04102
88720	82765	34476	17032	87589	40836	32427	70002	70663	88863
39475	46473	23219	53416	94970	25832	69975	94884	19661	72828
06990	67245	68350	82948	11398	42878	80287	88267	47363	46634
40980	07391	58745	25774	22987	80059	39911	96189	41151	14222
83974	29992	65381	38857	50490	83765	55657	14361	31720	57375
33339	31926	14883	24413	59744	92351	97473	89286	35931	04110
31662	25388	61642	34072	81249	35648	56891	69352	48373	45578
93526	70765	10592	04542	76463	54328	02349	17247	28865	14777
20492	38391	91132	21999	59516	81652	27195	48223	46751	22923
04153	53381	79401	21438	83035	92350	36693	31238	59649	91754
05520	91962	04739	13092	97662	24822	94730	06496	35090	04822
47498	87637	99016	71060	88824	71013	18735	20286	23153	72924
23167	49323	45021	33132	12544	41035	80780	45393	44812	12515
23792	14422	15059	45799	22716	19792	09983	74353	68668	30429
85900	98275	32388	52390	16815	69298	82732	38480	73817	32523
42559	78985	05300	22164	24369	54224	35083	19687	11062	91491
14349	82674	66523	44133	00697	35552	35970	19124	63318	29686
17403	53363	44167	64486	64758	75366	76554	31601	12614	33072
23632	27889	47914	02584	37680	20801	72152	39339	34806	08930

However, this is not the way in which digital computers simulate uniform $(0, 1)$ random variables. In practice, they use pseudo random numbers instead of truly random ones. Most random number generators start with an initial value X_0, called the seed, and then recursively compute values by specifying positive integers a, c, and m, and then letting

$$X_{n+1} = (aX_n + c) \quad \text{modulo } m, \quad n \geq 0$$

where the preceding means that $aX_n + c$ is divided by m and the remainder is taken as the value of X_{n+1}. Thus each X_n is either $0, 1, \ldots,$ or $m-1$ and the quantity X_n/m is taken as an approximation to a uniform $(0, 1)$ random variable. It can be shown that subject to suitable choices for a, c, m, the preceding gives rise to a sequence of numbers that looks as if it were generated from independent uniform $(0, 1)$ random variables.

As our starting point in the simulation of random variables from an arbitrary distribution, we shall suppose that we can simulate from the uniform $(0, 1)$ distribution, and we shall use the term "random numbers" to mean independent random variables from this distribution. In Sections 11.2 and 11.3 we present both general and special techniques for simulating continuous random variables; and in Section 11.4 we do the same for discrete random variables. In Section 11.5 we discuss the simulation both of jointly distributed random variables and stochastic processes. Particular attention is given to the simulation of nonhomogeneous Poisson processes, and in fact three different approaches for this are discussed. Simulation of two-dimensional Poisson processes is discussed in Section 11.5.2. In Section 11.6 we discuss various methods for increasing the precision of the simulation estimates by reducing their variance; and in Section 11.7 we consider the problem of choosing the number of simulation runs needed to attain a desired level of precision. Before beginning this program, however, let us consider two applications of simulation to combinatorial problems.

Example 11.1 (Generating a Random Permutation)**.** Suppose we are interested in generating a permutation of the numbers $1, 2, \ldots, n$ that is such that all $n!$ possible orderings are equally likely. The following algorithm will accomplish this by first choosing one of the numbers $1, \ldots, n$ at random and then putting that number in position n; it then chooses at random one of the remaining $n-1$ numbers and puts that number in position $n-1$; it then chooses at random one of the remaining $n-2$ numbers and puts it in position $n-2$, and so on (where choosing a number at random means that each of the remaining numbers is equally likely to be chosen). However, so that we do not have to consider exactly which of the numbers remain to be positioned, it is convenient and efficient to keep the numbers in an ordered list and then randomly choose the position of the number rather than the number itself. That is, starting with any initial ordering $p_1, p_2, \ldots, p_n$, we pick one of the positions $1, \ldots, n$ at random and then interchange the number in that position with the one in position n. Now we randomly choose one of the positions $1, \ldots, n-1$ and interchange the number in this position with the one in position $n-1$, and so on.

To implement the preceding, we need to be able to generate a random variable that is equally likely to take on any of the values $1, 2, \ldots, k$. To accomplish this, let U denote a random number—that is, U is uniformly distributed over $(0, 1)$—and note that kU is uniform on $(0, k)$ and so

$$P\{i-1 < kU < i\} = \frac{1}{k}, \quad i = 1, \ldots, k$$

Hence, the random variable $I = [kU] + 1$ will be such that

$$P\{I = i\} = P\{[kU] = i-1\} = P\{i-1 < kU < i\} = \frac{1}{k}$$

The preceding algorithm for generating a random permutation can now be written as follows:

Step 1: Let $p_1, p_2, \ldots, p_n$ be any permutation of $1, 2, \ldots, n$ (for instance, we can choose $p_j = j$, $j = 1, \ldots, n$).
Step 2: Set $k = n$.
Step 3: Generate a random number U and let $I = [kU] + 1$.
Step 4: Interchange the values of p_I and p_k.
Step 5: Let $k = k - 1$ and if $k > 1$ go to step 3.
Step 6: $p_1, \ldots, p_n$ is the desired random permutation.

For instance, suppose $n = 4$ and the initial permutation is 1, 2, 3, 4. If the first value of I (which is equally likely to be either 1, 2, 3, 4) is $I = 3$, then the new permutation is 1, 2, 4, 3. If the next value of I is $I = 2$ then the new permutation is 1, 4, 2, 3. If the final value of I is $I = 2$, then the final permutation is 1, 4, 2, 3, and this is the value of the random permutation.

One very important property of the preceding algorithm is that it can also be used to generate a random subset, say of size r, of the integers $1, \ldots, n$. Namely, just follow the algorithm until the positions $n, n-1, \ldots, n-r+1$ are filled. The elements in these positions constitute the random subset. ■

Example 11.2 (Estimating the Number of Distinct Entries in a Large List)**.** Consider a list of n entries where n is very large, and suppose we are interested in estimating d, the number of distinct elements in the list. If we let m_i denote the number of times that the element in position i appears on the list, then we can express d by

$$d = \sum_{i=1}^{n} \frac{1}{m_i}$$

To estimate d, suppose that we generate a random value X equally likely to be either $1, 2, \ldots, n$ (that is, we take $X = [nU] + 1$) and then let $m(X)$ denote the number of times the element in position X appears on the list. Then

$$E\left[\frac{1}{m(X)}\right] = \sum_{i=1}^{n} \frac{1}{m_i}\frac{1}{n} = \frac{d}{n}$$

Hence, if we generate k such random variables $X_1, \ldots, X_k$ we can estimate d by

$$d \approx \frac{n\sum_{i=1}^{k} 1/m(X_i)}{k}$$

Suppose now that each item in the list has a value attached to it—$v(i)$ being the value of the ith element. The sum of the values of the distinct items—call it v—can be expressed as

$$v = \sum_{i=1}^{n} \frac{v(i)}{m(i)}$$

Now if $X = [nU] + 1$, where U is a random number, then

$$E\left[\frac{v(X)}{m(X)}\right] = \sum_{i=1}^{n} \frac{v(i)}{m(i)} \frac{1}{n} = \frac{v}{n}$$

Hence, we can estimate v by generating $X_1, \ldots, X_k$ and then estimating v by

$$v \approx \frac{n}{k} \sum_{i=1}^{k} \frac{v(X_i)}{m(X_i)}$$

For an important application of the preceding, let $A_i = \{a_{i,1}, \ldots, a_{i,n_i}\}, i = 1, \ldots, s$ denote events, and suppose we are interested in estimating $P\left(\bigcup_{i=1}^{s} A_i\right)$. Since

$$P\left(\bigcup_{i=1}^{s} A_i\right) = \sum_{a \in \cup A_i} P(a) = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \frac{P(a_{i,j})}{m(a_{i,j})}$$

where $m(a_{i,j})$ is the number of events to which the point $a_{i,j}$ belongs, the preceding method can be used to estimate $P\left(\bigcup_{1}^{s} A_i\right)$.

Note that the preceding procedure for estimating v can be effected without prior knowledge of the set of values $\{v_1, \ldots, v_n\}$. That is, it suffices that we can determine the value of an element in a specific place and the number of times that element appears on the list. When the set of values is *a priori* known, there is another approach available as will be shown in Example 11.11. ■

11.2 General Techniques for Simulating Continuous Random Variables

In this section we present three methods for simulating continuous random variables.

11.2.1 The Inverse Transformation Method

A general method for simulating a random variable having a continuous distribution—called the *inverse transformation method*—is based on the following proposition.

Proposition 11.1. *Let U be a uniform $(0, 1)$ random variable. For any continuous distribution function F if we define the random variable X by*

$$X = F^{-1}(U)$$

then the random variable X has distribution function F. ($F^{-1}(u)$ is defined to equal that value x for which $F(x) = u$.)

Proof.

$$\begin{aligned} F_X(a) &= P\{X \le a\} \\ &= P\{F^{-1}(U) \le a\} \end{aligned} \tag{11.1}$$

Now, since $F(x)$ is a monotone function, it follows that $F^{-1}(U) \le a$ if and only if $U \le F(a)$. Hence, from Eq. (11.1), we see that

$$\begin{aligned} F_X(a) &= P\{U \le F(a)\} \\ &= F(a) \end{aligned}$$

■

Hence, we can simulate a random variable X from the continuous distribution F, when F^{-1} is computable, by simulating a random number U and then setting $X = F^{-1}(U)$.

Example 11.3 (Simulating an Exponential Random Variable)**.** If $F(x) = 1 - e^{-x}$, then $F^{-1}(u)$ is that value of x such that

$$1 - e^{-x} = u$$

or

$$x = -\log(1 - u)$$

Hence, if U is a uniform $(0, 1)$ variable, then

$$F^{-1}(U) = -\log(1 - U)$$

is exponentially distributed with mean 1. Since $1 - U$ is also uniformly distributed on $(0, 1)$ it follows that $-\log U$ is exponential with mean 1. Since cX is exponential with mean c when X is exponential with mean 1, it follows that $-c \log U$ is exponential with mean c. ■

11.2.2 The Rejection Method

Suppose that we have a method for simulating a random variable having density function $g(x)$. We can use this as the basis for simulating from the continuous distribution having density $f(x)$ by simulating Y from g and then accepting this simulated value with a probability proportional to $f(Y)/g(Y)$.

Specifically, let c be a constant such that

$$\frac{f(y)}{g(y)} \le c \quad \text{for all } y$$

We then have the following technique for simulating a random variable having density f.

Rejection Method

Step 1: Simulate Y having density g and simulate a random number U.
Step 2: If $U \leq f(Y)/cg(Y)$ set $X = Y$. Otherwise return to step 1.

Proposition 11.2. *The random variable X generated by the rejection method has density function f.*

Proof. Let X be the value obtained, and let N denote the number of necessary iterations. Then

$$\begin{aligned} P\{X \leq x\} &= P\{Y_N \leq x\} \\ &= P\{Y \leq x | U \leq f(Y)/cg(Y)\} \\ &= \frac{P\{Y \leq x, U \leq f(Y)/cg(Y)\}}{K} \\ &= \frac{\int P\{Y \leq x, U \leq f(Y)/cg(Y) | Y = y\} g(y)\, dy}{K} \\ &= \frac{\int_{-\infty}^{x} (f(y)/cg(y)) g(y)\, dy}{K} \\ &= \frac{\int_{-\infty}^{x} f(y)\, dy}{Kc} \end{aligned}$$

where $K = P\{U \leq f(Y)/cg(Y)\}$. Letting $x \to \infty$ shows that $K = 1/c$ and the proof is complete. ■

Remarks. **(i)** The preceding method was originally presented by Von Neumann in the special case where g was positive only in some finite interval (a, b), and Y was chosen to be uniform over (a, b) (that is, $Y = a + (b - a)U$).

(ii) Note that the way in which we "accept the value Y with probability $f(Y)/cg(Y)$" is by generating a uniform $(0, 1)$ random variable U and then accepting Y if $U \leq f(Y)/cg(Y)$.

(iii) Since each iteration of the method will, independently, result in an accepted value with probability $P\{U \leq f(Y)/cg(Y)\} = 1/c$ it follows that the number of iterations is geometric with mean c.

(iv) Actually, it is not necessary to generate a new uniform random number when deciding whether or not to accept, since at a cost of some additional computation, a single random number, suitably modified at each iteration, can be used throughout. To see how, note that the actual value of U is not used—only whether or not $U \leq f(Y)/cg(Y)$. Hence, if Y is rejected—that is, if $U > f(Y)/cg(Y)$—we can use the fact that, given Y,

$$\frac{U - f(Y)/cg(Y)}{1 - f(Y)/cg(Y)} = \frac{cUg(Y) - f(Y)}{cg(Y) - f(Y)}$$

is uniform on $(0, 1)$. Hence, this may be used as a uniform random number in the next iteration. As this saves the generation of a random number at the cost

of the preceding computation, whether it is a net savings depends greatly upon the method being used to generate random numbers. ■

Example 11.4. Let us use the rejection method to generate a random variable having density function

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1$$

Since this random variable (which is beta with parameters 2, 4) is concentrated in the interval (0, 1), let us consider the rejection method with

$$g(x) = 1, \quad 0 < x < 1$$

To determine the constant c such that $f(x)/g(x) \leq c$, we use calculus to determine the maximum value of

$$\frac{f(x)}{g(x)} = 20x(1-x)^3$$

Differentiation of this quantity yields

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = 20\Big[(1-x)^3 - 3x(1-x)^2\Big]$$

Setting this equal to 0 shows that the maximal value is attained when $x = \frac{1}{4}$, and thus

$$\frac{f(x)}{g(x)} \leq 20\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^3 = \frac{135}{64} \equiv c$$

Hence,

$$\frac{f(x)}{cg(x)} = \frac{256}{27}\, x(1-x)^3$$

and thus the rejection procedure is as follows:

Step 1: Generate random numbers U_1 and U_2.
Step 2: If $U_2 \leq \frac{256}{27}U_1(1-U_1)^3$, stop and set $X = U_1$. Otherwise return to step 1.

The average number of times that step 1 will be performed is $c = \frac{135}{64}$. ■

Example 11.5 (Simulating a Normal Random Variable)**.** To simulate a standard normal random variable Z (that is, one with mean 0 and variance 1) note first that the absolute value of Z has density function

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad 0 < x < \infty \tag{11.2}$$

We will start by simulating from the preceding density by using the rejection method with

$$g(x) = e^{-x}, \quad 0 < x < \infty$$

Now, note that

$$\frac{f(x)}{g(x)} = \sqrt{2e/\pi}\exp\{-(x-1)^2/2\} \le \sqrt{2e/\pi}$$

Hence, using the rejection method we can simulate from Eq. (11.2) as follows:

(a) Generate independent random variables Y and U, Y being exponential with rate 1 and U being uniform on $(0, 1)$.

(b) If $U \le \exp\{-(Y-1)^2/2\}$, or equivalently, if

$$-\log U \ge (Y-1)^2/2$$

set $X = Y$. Otherwise return to step (a).

Once we have simulated a random variable X having Density Function (11.2) we can then generate a standard normal random variable Z by letting Z be equally likely to be either X or $-X$.

To improve upon the foregoing, note first that from Example 11.3 it follows that $-\log U$ will also be exponential with rate 1. Hence, steps (a) and (b) are equivalent to the following:

(**a**′) Generate independent exponentials with rate 1, Y_1, and Y_2.

(**b**′) Set $X = Y_1$ if $Y_2 \ge (Y_1 - 1)^2/2$. Otherwise return to step (a′).

Now suppose that we accept step (b′). It then follows by the lack of memory property of the exponential that the amount by which Y_2 exceeds $(Y_1 - 1)^2/2$ will also be exponential with rate 1.

Hence, summing up, we have the following algorithm which generates an exponential with rate 1 and an independent standard normal random variable:

Step 1: Generate Y_1, an exponential random variable with rate 1.

Step 2: Generate Y_2, an exponential with rate 1.

Step 3: If $Y_2 - (Y_1 - 1)^2/2 > 0$, set $Y = Y_2 - (Y_1 - 1)^2/2$ and go to step 4. Otherwise go to step 1.

Step 4: Generate a random number U and set

$$Z = \begin{cases} Y_1, & \text{if } U \le \frac{1}{2} \\ -Y_1, & \text{if } U > \frac{1}{2} \end{cases}$$

The random variables Z and Y generated by the preceding are independent with Z being normal with mean 0 and variance 1 and Y being exponential with rate 1. (If we want the normal random variable to have mean μ and variance σ^2, just take $\mu + \sigma Z$.) ■

Remarks. **(i)** Since $c = \sqrt{2e/\pi} \approx 1.32$, the preceding requires a geometric distributed number of iterations of step 2 with mean 1.32.

(ii) The final random number of step 4 need not be separately simulated but rather can be obtained from the first digit of any random number used earlier. That is, suppose we generate a random number to simulate an exponential; then we can

strip off the initial digit of this random number and just use the remaining digits (with the decimal point moved one step to the right) as the random number. If this initial digit is 0, 1, 2, 3, or 4 (or 0 if the computer is generating binary digits), then we take the sign of Z to be positive and take it to be negative otherwise.

(iii) If we are generating a sequence of standard normal random variables, then we can use the exponential obtained in step 3 as the initial exponential needed in step 1 for the next normal to be generated. Hence, on the average, we can simulate a unit normal by generating 1.64 exponentials and computing 1.32 squares.

11.2.3 The Hazard Rate Method

Let F be a continuous distribution function with $\bar{F}(0) = 1$. Recall that $\lambda(t)$, the hazard rate function of F, is defined by

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}, \quad t \geq 0$$

(where $f(t) = F'(t)$ is the density function). Recall also that $\lambda(t)$ represents the instantaneous probability intensity that an item having life distribution F will fail at time t given it has survived to that time.

Suppose now that we are given a bounded function $\lambda(t)$, such that $\int_0^\infty \lambda(t)\, dt = \infty$, and we desire to simulate a random variable S having $\lambda(t)$ as its hazard rate function.

To do so let λ be such that

$$\lambda(t) \leq \lambda \quad \text{for all } t \geq 0$$

To simulate from $\lambda(t), t \geq 0$, we will

(a) simulate a Poisson process having rate λ. We will then only "accept" or "count" certain of these Poisson events. Specifically we will

(b) count an event that occurs at time t, independently of all else, with probability $\lambda(t)/\lambda$.

We now have the following proposition.

Proposition 11.3. *The time of the first counted event—call it S— is a random variable whose distribution has hazard rate function $\lambda(t), t \geq 0$.*

Proof.

$$\begin{aligned} &P\{t < S < t + dt | S > t\} \\ &\quad = P\{\text{first counted event in } (t, t + dt) | \text{no counted events prior to } t\} \\ &\quad = P\{\text{Poisson event in } (t, t + dt), \text{ it is counted} | \text{no counted events prior to } t\} \\ &\quad = P\{\text{Poisson event in } (t, t + dt), \text{ it is counted}\} \end{aligned}$$

$$= [\lambda\, dt + o(dt)]\frac{\lambda(t)}{\lambda} = \lambda(t)\, dt + o(dt)$$

which completes the proof. Note that the next to last equality follows from the independent increment property of Poisson processes. ■

Because the interarrival times of a Poisson process having rate λ are exponential with rate λ, it thus follows from Example 11.3 and the previous proposition that the following algorithm will generate a random variable having hazard rate function $\lambda(t), t \geq 0$.

Hazard Rate Method for Generating S: $\lambda_s(t) = \lambda(t)$

Let λ be such that $\lambda(t) \leq \lambda$ for all $t \geq 0$. Generate pairs of random variables U_i, X_i, $i \geq 1$, with X_i being exponential with rate λ and U_i being uniform $(0, 1)$, stopping at

$$N = \min\left\{n\colon\ U_n \leq \lambda\left(\sum_{i=1}^{n} X_i\right)\Big/\lambda\right\}$$

Set

$$S = \sum_{i=1}^{N} X_i$$ ■

To compute $E[N]$ we need the result, known as Wald's equation, which states that if $X_1, X_2, \ldots$ are independent and identically distributed random variables that are observed in sequence up to some random time N then

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

More precisely let $X_1, X_2, \ldots$ denote a sequence of independent random variables and consider the following definition.

Definition 11.1. An integer-valued random variable N is said to be a *stopping time* for the sequence $X_1, X_2, \ldots$ if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$ for all $n = 1, 2, \ldots$.

Intuitively, we observe the X_ns in sequential order and N denotes the number observed before stopping. If $N = n$, then we have stopped after observing $X_1, \ldots, X_n$ and before observing $X_{n+1}, X_{n+2}, \ldots$ for all $n = 1, 2, \ldots$.

Example 11.6. Let $X_n, n = 1, 2, \ldots$, be independent and such that

$$P\{X_n = 0\} = P\{X_n = 1\} = \tfrac{1}{2}, \quad n = 1, 2, \ldots$$

If we let

$$N = \min\{n\colon X_1 + \cdots + X_n = 10\}$$

then N is a stopping time. We may regard N as being the stopping time of an experiment that successively flips a fair coin and then stops when the number of heads reaches 10. ■

Proposition 11.4 (Wald's Equation). *If $X_1, X_2, \ldots$ are independent and identically distributed random variables having finite expectations, and if N is a stopping time for $X_1, X_2, \ldots$ such that $E[N] < \infty$, then*

$$E\left[\sum_1^N X_n\right] = E[N]E[X]$$

Proof. Letting

$$I_n = \begin{cases} 1, & \text{if } N \geq n \\ 0, & \text{if } N < n \end{cases}$$

we have

$$\sum_{n=1}^{N} X_n = \sum_{n=1}^{\infty} X_n I_n$$

Hence,

$$E\left[\sum_{n=1}^{N} X_n\right] = E\left[\sum_{n=1}^{\infty} X_n I_n\right] = \sum_{n=1}^{\infty} E[X_n I_n] \tag{11.3}$$

However, $I_n = 1$ if and only if we have not stopped after successively observing $X_1, \ldots, X_{n-1}$. Therefore, I_n is determined by $X_1, \ldots, X_{n-1}$ and is thus independent of X_n. From Eq. (11.3) we thus obtain

$$\begin{aligned} E\left[\sum_{n=1}^{N} X_n\right] &= \sum_{n=1}^{\infty} E[X_n]E[I_n] \\ &= E[X]\sum_{n=1}^{\infty} E[I_n] \\ &= E[X]E\left[\sum_{n=1}^{\infty} I_n\right] \\ &= E[X]E[N] \end{aligned}$$

■

Returning to the hazard rate method, we have

$$S = \sum_{i=1}^{N} X_i$$

As $N = \min\{n: U_n \leq \lambda \left(\sum_1^n X_i\right)/\lambda\}$ it follows that the event that $N = n$ is independent of $X_{n+1}, X_{n+2}, \ldots$. Hence, by Wald's equation,

$$E[S] = E[N]E[X_i]$$

$$= \frac{E[N]}{\lambda}$$

or

$$E[N] = \lambda E[S]$$

where $E[S]$ is the mean of the desired random variable.

11.3 Special Techniques for Simulating Continuous Random Variables

Special techniques have been devised to simulate from most of the common continuous distributions. We now present certain of these.

11.3.1 The Normal Distribution

Let X and Y denote independent standard normal random variables and thus have the joint density function

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \quad -\infty < x < \infty, -\infty < y < \infty$$

Consider now the polar coordinates of the point (X, Y). As shown in Fig. 11.1,

$$R^2 = X^2 + Y^2,$$

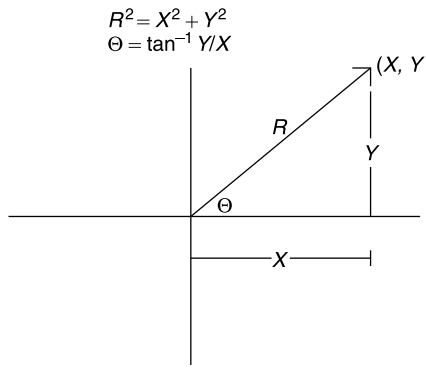

Figure 11.1

$$\Theta = \tan^{-1} Y/X$$

To obtain the joint density of R^2 and Θ, consider the transformation

$$d = x^2 + y^2, \quad \theta = \tan^{-1} y/x$$

The Jacobian of this transformation is

$$J = \begin{vmatrix} \frac{\partial d}{\partial x} & \frac{\partial d}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} = \begin{vmatrix} 2x & 2y \\ \frac{1}{1+y^2/x^2}\left(\frac{-y}{x^2}\right) & \frac{1}{1+y^2/x^2}\left(\frac{1}{x}\right) \end{vmatrix}$$
$$= 2\begin{vmatrix} x & y \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix} = 2$$

Hence, from Section 2.5.3 the joint density of R^2 and Θ is given by

$$f_{R^2,\Theta}(d,\theta) = \frac{1}{2\pi} e^{-d/2} \frac{1}{2}$$
$$= \frac{1}{2} e^{-d/2} \frac{1}{2\pi}, \quad 0 < d < \infty, 0 < \theta < 2\pi$$

Thus, we can conclude that R^2 and Θ are independent with R^2 having an exponential distribution with rate $\frac{1}{2}$ and Θ being uniform on $(0, 2\pi)$.

Let us now go in reverse from the polar to the rectangular coordinates. From the preceding if we start with W, an exponential random variable with rate $\frac{1}{2}$ (W plays the role of R^2) and with V, independent of W and uniformly distributed over $(0, 2\pi)$ (V plays the role of Θ) then $X = \sqrt{W}\cos V$, $Y = \sqrt{W}\sin V$ will be independent standard normals. Hence, using the results of Example 11.3 we see that if U_1 and U_2 are independent uniform $(0, 1)$ random numbers, then

$$\begin{aligned} X &= (-2\log\ U_1)^{1/2}\cos(2\pi U_2), \\ Y &= (-2\log\ U_1)^{1/2}\sin(2\pi U_2) \end{aligned} \tag{11.4}$$

are independent standard normal random variables.

Remark. The fact that $X^2 + Y^2$ has an exponential distribution with rate $\frac{1}{2}$ is quite interesting for, by the definition of the chi-square distribution, $X^2 + Y^2$ has a chi-squared distribution with two degrees of freedom. Hence, these two distributions are identical.

The preceding approach to generating standard normal random variables is called the *Box–Muller approach*. Its efficiency suffers somewhat from its need to compute

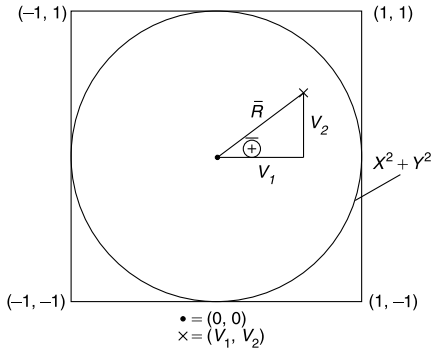

Figure 11.2

the preceding sine and cosine values. There is, however, a way to get around this potentially time-consuming difficulty. To begin, note that if U is uniform on $(0, 1)$, then $2U$ is uniform on $(0, 2)$, and so $2U - 1$ is uniform on $(-1, 1)$. Thus, if we generate random numbers U_1 and U_2 and set

$$V_1 = 2U_1 - 1,$$
$$V_2 = 2U_2 - 1$$

then (V_1, V_2) is uniformly distributed in the square of area 4 centered at $(0, 0)$ (see Fig. 11.2).

Suppose now that we continually generate such pairs (V_1, V_2) until we obtain one that is contained in the circle of radius 1 centered at $(0, 0)$—that is, until (V_1, V_2) is such that $V_1^2 + V_2^2 \leq 1$. It now follows that such a pair (V_1, V_2) is uniformly distributed in the circle. If we let $\bar{R}, \bar{\Theta}$ denote the polar coordinates of this pair, then it is easy to verify that $\bar{R}$ and $\bar{\Theta}$ are independent, with $\bar{R}^2$ being uniformly distributed on $(0, 1)$, and $\bar{\Theta}$ uniformly distributed on $(0, 2\pi)$.

Since

$$\sin\bar{\Theta} = V_2/\bar{R} = \frac{V_2}{\sqrt{V_1^2 + V_2^2}},$$
$$\cos\bar{\Theta} = V_1/\bar{R} = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$

it follows from Eq. (11.4) that we can generate independent standard normals X and Y by generating another random number U and setting

$$X = (-2\log\ U)^{1/2} V_1/\bar{R},$$
$$Y = (-2\log\ U)^{1/2} V_2/\bar{R}$$

In fact, since (conditional on $V_1^2 + V_2^2 \leq 1$) $\bar{R}^2$ is uniform on (0, 1) and is independent of $\bar{\Theta}$, we can use it instead of generating a new random number U; thus showing that

$$X = (-2\log \bar{R}^2)^{1/2} V_1/\bar{R} = \sqrt{\frac{-2\log S}{S}} V_1,$$
$$Y = (-2\log \bar{R}^2)^{1/2} V_2/\bar{R} = \sqrt{\frac{-2\log S}{S}} V_2$$

are independent standard normals, where

$$S = \bar{R}^2 = V_1^2 + V_2^2$$

Summing up, we thus have the following approach to generating a pair of independent standard normals:

Step 1: Generate random numbers U_1 and U_2.
Step 2: Set $V_1 = 2U_1 - 1,\ V_2 = 2U_2 - 1,\ S = V_1^2 + V_2^2$.
Step 3: If $S > 1$, return to step 1.
Step 4: Return the independent unit normals

$$X = \sqrt{\frac{-2\log S}{S}} V_1, \quad Y = \sqrt{\frac{-2\log S}{S}} V_2$$

The preceding is called the *polar method*. Since the probability that a random point in the square will fall within the circle is equal to $\pi/4$ (the area of the circle divided by the area of the square), it follows that, on average, the polar method will require $4/\pi = 1.273$ iterations of step 1. Hence, it will, on average, require 2.546 random numbers, 1 logarithm, 1 square root, 1 division, and 4.546 multiplications to generate 2 independent standard normals.

11.3.2 The Gamma Distribution

To simulate from a gamma distribution with parameters (n, λ), where n is an integer, we use the fact that the sum of n independent exponential random variables each having rate λ has this distribution. Hence, if $U_1, \ldots, U_n$ are independent uniform (0, 1) random variables,

$$X = \frac{1}{\lambda} \sum_{i=1}^{n} \log\ U_i = -\frac{1}{\lambda} \log\left(\prod_{i=1}^{n} U_i\right)$$

has the desired distribution.

When n is large, there are other techniques available that do not require so many random numbers. One possibility is to use the rejection procedure with $g(x)$ being taken as the density of an exponential random variable with mean n/λ (as this is the mean of the gamma). It can be shown that for large n the average number of iterations needed by the rejection algorithm is $e[(n-1)/2\pi]^{1/2}$. In addition, if we wanted to

generate a series of gammas, then, just as in Example 11.4, we can arrange things so that upon acceptance we obtain not only a gamma random variable but also, for free, an exponential random variable that can then be used in obtaining the next gamma (see Exercise 8).

11.3.3 The Chi-Squared Distribution

The chi-squared distribution with n degrees of freedom is the distribution of $\chi_n^2 = Z_1^2 + \cdots + Z_n^2$ where $Z_i, i = 1, \ldots, n$ are independent standard normals. Using the fact noted in the remark at the end of Section 3.1 we see that $Z_1^2 + Z_2^2$ has an exponential distribution with rate $\frac{1}{2}$. Hence, when n is even—say, $n = 2k$—χ_{2k}^2 has a gamma distribution with parameters $\left(k, \frac{1}{2}\right)$. Hence, $-2\log\left(\prod_{i=1}^{k} U_i\right)$ has a chi-squared distribution with $2k$ degrees of freedom. We can simulate a chi-squared random variable with $2k + 1$ degrees of freedom by first simulating a standard normal random variable Z and then adding Z^2 to the preceding. That is,

$$\chi_{2k+1}^2 = Z^2 - 2\log\left(\prod_{i=1}^{k} U_i\right)$$

where $Z, U_1, \ldots, U_n$ are independent with Z being a standard normal and the others being uniform (0, 1) random variables.

11.3.4 The Beta (n, m) Distribution

The random variable X is said to have a beta distribution with parameters n, m if its density is given by

$$f(x) = \frac{(n+m-1)!}{(n-1)!(m-1)!} x^{n-1}(1-x)^{m-1}, \quad 0 < x < 1$$

One approach to simulating from the preceding distribution is to let $U_1, \ldots, U_{n+m-1}$ be independent uniform (0, 1) random variables and consider the nth smallest value of this set—call it $U_{(n)}$. Now $U_{(n)}$ will equal x if, of the $n + m - 1$ variables,

(i) $n - 1$ are smaller than x,
(ii) one equals x,
(iii) $m - 1$ are greater than x.

Hence, if the $n + m - 1$ uniform random variables are partitioned into three subsets of sizes $n - 1$, 1, and $m - 1$ the probability (density) that each of the variables in the first set is less than x, the variable in the second set equals x, and all the variables in the third set are greater than x is given by

$$(P\{U < x\})^{n-1} f_U(x) (P\{U > x\})^{m-1} = x^{n-1}(1-x)^{m-1}$$

Hence, as there are $(n+m-1)!/(n-1)!(m-1)!$ possible partitions, it follows that $U_{(n)}$ is beta with parameters (n, m).

Thus, one way to simulate from the beta distribution is to find the nth smallest of a set of $n+m-1$ random numbers. However, when n and m are large, this procedure is not particularly efficient.

For another approach consider a Poisson process with rate 1, and recall that given S_{n+m}, the time of the $(n+m)$th event, the set of the first $n+m-1$ event times is distributed independently and uniformly on $(0, S_{n+m})$. Hence, given S_{n+m}, the nth smallest of the first $n+m-1$ event times—that is, S_n—is distributed as the nth smallest of a set of $n+m-1$ uniform $(0, S_{n+m})$ random variables. But from the preceding we can thus conclude that S_n/S_{n+m} has a beta distribution with parameters (n, m). Therefore, if $U_1, \ldots, U_{n+m}$ are random numbers,

$$\frac{-\log \prod_{i=1}^{n} U_i}{-\log \prod_{i=1}^{m+n} U_i} \text{ is beta with parameters } (n, m)$$

By writing the preceding as

$$\frac{-\log \prod_{i=1}^{n} U_i}{-\log \prod_{1}^{n} U_i - \log \prod_{n+1}^{n+m} U_i}$$

we see that it has the same distribution as $X/(X+Y)$ where X and Y are independent gamma random variables with respective parameters $(n, 1)$ and $(m, 1)$. Hence, when n and m are large, we can efficiently simulate a beta by first simulating two gamma random variables.

11.3.5 The Exponential Distribution—The Von Neumann Algorithm

As we have seen, an exponential random variable with rate 1 can be simulated by computing the negative of the logarithm of a random number. Most computer programs for computing a logarithm, however, involve a power series expansion, and so it might be useful to have at hand a second method that is computationally easier. We now present such a method due to Von Neumann.

To begin let $U_1, U_2, \ldots$ be independent uniform (0, 1) random variables and define N, $N \geq 2$, by

$$N = \min\{n: U_1 \geq U_2 \geq \cdots \geq U_{n-1} < U_n\}$$

That is, N is the index of the first random number that is greater than its predecessor. Let us now compute the joint distribution of N and U_1.

$$\begin{aligned} P\{N > n, U_1 \leq y\} &= \int_0^1 P\{N > n, U_1 \leq y | U_1 = x\}\, dx \\ &= \int_0^y P\{N > n | U_1 = x\}\, dx \end{aligned}$$

Now, given that $U_1 = x$, N will be greater than n if $x \geq U_2 \geq \cdots \geq U_n$ or, equivalently, if

(a) $U_i \leq x, \quad i = 2, \ldots, n$

and

(b) $U_2 \geq \cdots \geq U_n$

Now, (a) has probability x^{n-1} of occurring and given (a), since all of the $(n-1)!$ possible rankings of $U_2, \ldots, U_n$ are equally likely, (b) has probability $1/(n-1)!$ of occurring. Hence,

$$P\{N > n | U_1 = x\} = \frac{x^{n-1}}{(n-1)!}$$

and so

$$P\{N > n, U_1 \leq y\} = \int_0^y \frac{x^{n-1}}{(n-1)!}\, dx = \frac{y^n}{n!}$$

which yields

$$\begin{aligned} P\{N = n, U_1 \leq y\} &= P\{N > n-1, U_1 \leq y\} - P\{N > n, U_1 \leq y\} \\ &= \frac{y^{n-1}}{(n-1)!} - \frac{y^n}{n!} \end{aligned}$$

Upon summing over all the even integers, we see that

$$\begin{aligned} P\{N \text{ is even}, \ U_1 \leq y\} &= y - \frac{y^2}{2!} + \frac{y^3}{3!} - \frac{y^4}{4!} - \cdots \\ &= 1 - e^{-y} \end{aligned} \tag{11.5}$$

We are now ready for the following algorithm for generating an exponential random variable with rate 1.

Step 1: Generate uniform random numbers $U_1, U_2, \ldots$ stopping at $N = \min\{n: U_1 \geq \cdots \geq U_{n-1} < U_n\}$.

Step 2: If N is even accept that run, and go to step 3. If N is odd reject the run, and return to step 1.

Step 3: Set X equal to the number of failed runs plus the first random number in the successful run.

To show that X is exponential with rate 1, first note that the probability of a successful run is, from Eq. (11.5) with $y = 1$,

$$P\{N \text{ is even}\} = 1 - e^{-1}$$

Now, in order for X to exceed x, the first $[x]$ runs must all be unsuccessful and the next run must either be unsuccessful or be successful but have $U_1 > x - [x]$ (where

$[x]$ is the largest integer not exceeding x). As

$$\begin{aligned} P\{N \text{ even},\ U_1 > y\} &= P\{N \text{ even}\} - P\{N \text{ even},\ U_1 \le y\} \\ &= 1 - e^{-1} - (1 - e^{-y}) = e^{-y} - e^{-1} \end{aligned}$$

we see that

$$P\{X > x\} = e^{-[x]}[e^{-1} + e^{-(x-[x])} - e^{-1}] = e^{-x}$$

which yields the result.

Let T denote the number of trials needed to generate a successful run. As each trial is a success with probability $1 - e^{-1}$ it follows that T is geometric with mean $1/(1 - e^{-1})$. If we let N_i denote the number of uniform random variables used on the ith run, $i \ge 1$, then T (being the first run i for which N_i is even) is a stopping time for this sequence. Hence, by Wald's equation, the mean number of uniform random variables needed by this algorithm is given by

$$E\left[\sum_{i=1}^{T} N_i\right] = E[N]E[T]$$

Now,

$$\begin{aligned} E[N] &= \sum_{n=0}^{\infty} P\{N > n\} \\ &= 1 + \sum_{n=1}^{\infty} P\{U_1 \ge \cdots \ge U_n\} \\ &= 1 + \sum_{n=1}^{\infty} 1/n! = e \end{aligned}$$

and so

$$E\left[\sum_{i=1}^{T} N_i\right] = \frac{e}{1 - e^{-1}} \approx 4.3$$

Hence, this algorithm, which computationally speaking is quite easy to perform, requires on the average about 4.3 random numbers to execute.

11.4 Simulating from Discrete Distributions

All of the general methods for simulating from continuous distributions have analogs in the discrete case. For instance, if we want to simulate a random variable X having

probability mass function

$$P\{X = x_j\} = P_j, \quad j = 1, 2, \ldots, \quad \sum_j P_j = 1$$

we can use the following discrete time analog of the inverse transform technique:

To simulate X for which $P\{X = x_j\} = P_j$

let U be uniformly distributed over $(0, 1)$, and set

$$X = \begin{cases} x_1, & \text{if } U < P_1 \\ x_2, & \text{if } P_1 < U < P_1 + P_2 \\ \vdots & \\ x_j, & \text{if } \sum_1^{j-1} P_i < U < \sum_i^j P_i \\ \vdots & \end{cases}$$

As,

$$P\{X = x_j\} = P\left\{\sum_1^{j-1} P_i < U < \sum_1^j P_i\right\} = P_j$$

we see that X has the desired distribution.

Example 11.7 (The Geometric Distribution)**.** Suppose we want to simulate X such that

$$P\{X = i\} = p(1-p)^{i-1}, \quad i \geq 1$$

As

$$\sum_{i=1}^{j-1} P\{X = i\} = 1 - P\{X > j-1\} = 1 - (1-p)^{j-1}$$

we can simulate such a random variable by generating a random number U and then setting X equal to that value j for which

$$1 - (1-p)^{j-1} < U < 1 - (1-p)^j$$

or, equivalently, for which

$$(1-p)^j < 1 - U < (1-p)^{j-1}$$

As $1 - U$ has the same distribution as U, we can thus define X by

$$X = \min\{j\colon (1-p)^j < U\} = \min\left\{j\colon j > \frac{\log U}{\log(1-p)}\right\}$$

$$= 1 + \left[\frac{\log U}{\log(1-p)} \right] \qquad \blacksquare$$

As in the continuous case, special simulation techniques have been developed for the more common discrete distributions. We now present certain of these.

Example 11.8 (Simulating a Binomial Random Variable)**.** A binomial (n, p) random variable can be most easily simulated by recalling that it can be expressed as the sum of n independent Bernoulli random variables. That is, if $U_1, \ldots, U_n$ are independent uniform $(0, 1)$ variables, then letting

$$X_i = \begin{cases} 1, & \text{if } U_i < p \\ 0, & \text{otherwise} \end{cases}$$

it follows that $X \equiv \sum_{i=1}^{n} X_i$ is a binomial random variable with parameters n and p.

One difficulty with this procedure is that it requires the generation of n random numbers. To show how to reduce the number of random numbers needed, note first that this procedure does not use the actual value of a random number U but only whether or not it exceeds p. Using this and the result that the conditional distribution of U given that $U < p$ is uniform on $(0, p)$ and the conditional distribution of U given that $U > p$ is uniform on $(p, 1)$, we now show how we can simulate a binomial (n, p) random variable using only a single random number:

Step 1: Let $\alpha = 1/p, \beta = 1/(1-p)$.
Step 2: Set $k = 0$.
Step 3: Generate a uniform random number U.
Step 4: If $k = n$ stop. Otherwise reset k to equal $k + 1$.
Step 5: If $U \leq p$ set $X_k = 1$ and reset U to equal αU. If $U > p$ set $X_k = 0$ and reset U to equal $\beta(U - p)$. Return to step 4.

This procedure generates $X_1, \ldots, X_n$ and $X = \sum_{i=1}^{n} X_i$ is the desired random variable. It works by noting whether $U_k \leq p$ or $U_k > p$; in the former case it takes U_{k+1} to equal U_k/p, and in the latter case it takes U_{k+1} to equal $(U_k - p)/(1-p)$.[2] $\blacksquare$

Example 11.9 (Simulating a Poisson Random Variable)**.** To simulate a Poisson random variable with mean λ, generate independent uniform $(0, 1)$ random variables $U_1, U_2, \ldots$ stopping at

$$N + 1 = \min \left\{ n: \prod_{i=1}^{n} U_i < e^{-\lambda} \right\}$$

The random variable N has the desired distribution, which can be seen by noting that

$$N = \max \left\{ n: \sum_{i=1}^{n} -\log U_i < \lambda \right\}$$

[2] Because of computer round-off errors, a single random number should not be continuously used when n is large.

But $-\log U_i$ is exponential with rate 1, and so if we interpret $-\log U_i, i \geq 1$, as the interarrival times of a Poisson process having rate 1, we see that $N = N(\lambda)$ would equal the number of events by time λ. Hence N is Poisson with mean λ.

When λ is large we can reduce the amount of computation in the preceding simulation of $N(\lambda)$, the number of events by time λ of a Poisson process having rate 1, by first choosing an integer m and simulating S_m, the time of the mth event of the Poisson process, and then simulating $N(\lambda)$ according to the conditional distribution of $N(\lambda)$ given S_m. Now the conditional distribution of $N(\lambda)$ given S_m is as follows:

$$N(\lambda)|S_m = s \sim m + \text{Poisson}(\lambda - s), \qquad \text{if } s < \lambda$$
$$N(\lambda)|S_m = s \sim \text{Binomial}\left(m - 1, \frac{\lambda}{s}\right), \qquad \text{if } s > \lambda$$

where $\sim$ means "has the distribution of." This follows since if the mth event occurs at time s, where $s < \lambda$, then the number of events by time λ is m plus the number of events in (s, λ). On the other hand given that $S_m = s$ the set of times at which the first $m - 1$ events occur has the same distribution as a set of $m - 1$ uniform $(0, s)$ random variables (see Section 5.3.4). Hence, when $\lambda < s$, the number of these that occur by time λ is binomial with parameters $m - 1$ and λ/s. Hence, we can simulate $N(\lambda)$ by first simulating S_m and then simulating, either $P(\lambda - S_m)$, a Poisson random variable with mean $\lambda - S_m$, when $S_m < \lambda$, or simulating $\text{Bin}(m - 1, \lambda/S_m)$, a binomial random variable with parameters $m - 1$ and λ/S_m, when $S_m > \lambda$; and then setting

$$N(\lambda) = \begin{cases} m + P(\lambda - S_m), & \text{if } S_m < \lambda \\ \text{Bin}(m - 1, \lambda/S_m), & \text{if } S_m > \lambda \end{cases}$$

In the preceding it has been found computationally effective to let m be approximately $\frac{7}{8}\lambda$. Of course, S_m is simulated by simulating from a gamma (m, λ) distribution via an approach that is computationally fast when m is large (see Section 11.3.3). ■

There are also rejection and hazard rate methods for discrete distributions but we leave their development as exercises. However, there is a technique available for simulating finite discrete random variables—called the *alias method*—which, though requiring some setup time, is very fast to implement.

11.4.1 The Alias Method

In what follows, the quantities $\mathbf{P}, \mathbf{P}^{(k)}, \mathbf{Q}^{(k)}, k \leq n - 1$ will represent probability mass functions on the integers $1, 2, \ldots, n$—that is, they will be n-vectors of nonnegative numbers summing to 1. In addition, the vector $\mathbf{P}^{(k)}$ will have at most k nonzero components, and each of the $\mathbf{Q}^{(k)}$ will have at most two nonzero components. We show that any probability mass function $\mathbf{P}$ can be represented as an equally weighted mixture of $n - 1$ probability mass functions $\mathbf{Q}$ (each having at most two nonzero components).

That is, we show that for suitably defined $\mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(n-1)}$, $\mathbf{P}$ can be expressed as

$$\mathbf{P} = \frac{1}{n-1} \sum_{k=1}^{n-1} \mathbf{Q}^{(k)} \tag{11.6}$$

As a prelude to presenting the method for obtaining this representation, we will need the following simple lemma whose proof is left as an exercise.

Lemma 11.5. *Let* $\mathbf{P} = \{P_i, i = 1, \ldots, n\}$ *denote a probability mass function, then*

(a) *there exists an* $i, 1 \le i \le n$, *such that* $P_i < 1/(n-1)$, *and*
(b) *for this* i, *there exists a* $j, j \ne i$, *such that* $P_i + P_j \ge 1/(n-1)$.

Before presenting the general technique for obtaining the representation of Eq. (11.6), let us illustrate it by an example.

Example 11.10. Consider the three-point distribution $\mathbf{P}$ with $P_1 = \frac{7}{16}$, $P_2 = \frac{1}{2}$, $P_3 = \frac{1}{16}$. We start by choosing i and j such that they satisfy the conditions of Lemma 11.5. As $P_3 < \frac{1}{2}$ and $P_3 + P_2 > \frac{1}{2}$, we can work with $i = 3$ and $j = 2$. We will now define a two-point mass function $\mathbf{Q}^{(1)}$ putting all of its weight on 3 and 2 and such that $\mathbf{P}$ will be expressible as an equally weighted mixture between $\mathbf{Q}^{(1)}$ and a second two-point mass function $\mathbf{Q}^{(2)}$. Secondly, all of the mass of point 3 will be contained in $\mathbf{Q}^{(1)}$. As we will have

$$P_j = \frac{1}{2}(Q_j^{(1)} + Q_j^{(2)}), \quad j = 1, 2, 3 \tag{11.7}$$

and, by the preceding, $Q_3^{(2)}$ is supposed to equal 0, we must therefore take

$$Q_3^{(1)} = 2P_3 = \frac{1}{8}, \quad Q_2^{(1)} = 1 - Q_3^{(1)} = \frac{7}{8}, \quad Q_1^{(1)} = 0$$

To satisfy Eq. (11.7), we must then set

$$Q_3^{(2)} = 0, \quad Q_2^{(2)} = 2P_2 - \frac{7}{8} = \frac{1}{8}, \quad Q_1^{(2)} = 2P_1 = \frac{7}{8}$$

Hence, we have the desired representation in this case. Suppose now that the original distribution was the following four-point mass function:

$$P_1 = \frac{7}{16}, \quad P_2 = \frac{1}{4}, \quad P_3 = \frac{1}{8}, \quad P_4 = \frac{3}{16}$$

Now, $P_3 < \frac{1}{3}$ and $P_3 + P_1 > \frac{1}{3}$. Hence our initial two-point mass function—$\mathbf{Q}^{(1)}$—will concentrate on points 3 and 1 (giving no weights to 2 and 4). As the final representation will give weight $\frac{1}{3}$ to $\mathbf{Q}^{(1)}$ and in addition the other $\mathbf{Q}^{(j)}$, $j = 2, 3$, will not give any mass to the value 3, we must have

$$\frac{1}{3} Q_3^{(1)} = P_3 = \frac{1}{8}$$

Hence,

$$Q_3^{(1)} = \frac{3}{8}, \quad Q_1^{(1)} = 1 - \frac{3}{8} = \frac{5}{8}$$

Also, we can write

$$\mathbf{P} = \frac{1}{3}\mathbf{Q}^{(1)} + \frac{2}{3}\mathbf{P}^{(3)}$$

where $\mathbf{P}^{(3)}$, to satisfy the preceding, must be the vector

$$\begin{aligned}
\mathbf{P}_1^{(3)} &= \frac{3}{2}\left(P_1 - \frac{1}{3}Q_1^{(1)}\right) = \frac{1}{3}\frac{1}{2}, \\
\mathbf{P}_2^{(3)} &= \frac{3}{2}P_2 = \frac{3}{8}, \\
\mathbf{P}_3^{(3)} &= 0, \\
\mathbf{P}_4^{(3)} &= \frac{3}{2}P_4 = \frac{9}{32}
\end{aligned}$$

Note that $\mathbf{P}^{(3)}$ gives no mass to the value 3. We can now express the mass function $\mathbf{P}^{(3)}$ as an equally weighted mixture of two-point mass functions $\mathbf{Q}^{(2)}$ and $\mathbf{Q}^{(3)}$, and we will end up with

$$\begin{aligned}
\mathbf{P} &= \frac{1}{3}\mathbf{Q}^{(1)} + \frac{2}{3}\left(\frac{1}{2}\mathbf{Q}^{(2)} + \frac{1}{2}\mathbf{Q}^{(3)}\right) \\
&= \frac{1}{3}(\mathbf{Q}^{(1)} + \mathbf{Q}^{(2)} + \mathbf{Q}^{(3)})
\end{aligned}$$

(We leave it as an exercise for you to fill in the details.) ■

The preceding example outlines the following general procedure for writing the n-point mass function $\mathbf{P}$ in the form of Eq. (11.6) where each of the $\mathbf{Q}^{(i)}$ are mass functions giving all their mass to at most two points. To start, we choose i and j satisfying the conditions of Lemma 11.5. We now define the mass function $\mathbf{Q}^{(1)}$ concentrating on the points i and j and which will contain all of the mass for point i by noting that, in the representation of Eq. (11.6), $Q_i^{(k)} = 0$ for $k = 2, \ldots, n-1$, implying that

$$Q_i^{(1)} = (n-1)P_i, \quad \text{and so} \quad Q_j^{(1)} = 1 - (n-1)P_i$$

Writing

$$\mathbf{P} = \frac{1}{n-1}\mathbf{Q}^{(1)} + \frac{n-2}{n-1}\mathbf{P}^{(n-1)} \tag{11.8}$$

where $\mathbf{P}^{(n-1)}$ represents the remaining mass, we see that

$$P_i^{(n-1)} = 0,$$

$$P_j^{(n-1)} = \frac{n-1}{n-2}\left(P_j - \frac{1}{n-1}Q_j^{(1)}\right) = \frac{n-1}{n-2}\left(P_i + P_j - \frac{1}{n-1}\right),$$
$$P_k^{(n-1)} = \frac{n-1}{n-2}P_k, \quad k \neq i \text{ or } j$$

That the foregoing is indeed a probability mass function is easily checked—for instance, the nonnegativity of $P_j^{(n-1)}$ follows from the fact that j was chosen so that $P_i + P_j \geq 1/(n-1)$.

We may now repeat the foregoing procedure on the $(n-1)$-point probability mass function $\mathbf{P}^{(n-1)}$ to obtain

$$\mathbf{P}^{(n-1)} = \frac{1}{n-2}\mathbf{Q}^{(2)} + \frac{n-3}{n-2}\mathbf{P}^{(n-2)}$$

and thus from Eq. (11.8) we have

$$\mathbf{P} = \frac{1}{n-1}\mathbf{Q}^{(1)} + \frac{1}{n-1}\mathbf{Q}^{(2)} + \frac{n-3}{n-1}\mathbf{P}^{(n-2)}$$

We now repeat the procedure on $\mathbf{P}^{(n-2)}$ and so on until we finally obtain

$$\mathbf{P} = \frac{1}{n-1}(\mathbf{Q}^{(1)} + \cdots + \mathbf{Q}^{(n-1)})$$

In this way we are able to represent $\mathbf{P}$ as an equally weighted mixture of $n-1$ two-point mass functions. We can now easily simulate from $\mathbf{P}$ by first generating a random integer N equally likely to be either $1, 2, \ldots,$ or $n-1$. If the resulting value N is such that $\mathbf{Q}^{(N)}$ puts positive weight only on the points i_N and j_N, then we can set X equal to i_N if a second random number is less than $Q_{i_N}^{(N)}$ and equal to j_N otherwise. The random variable X will have probability mass function $\mathbf{P}$. That is, we have the following procedure for simulating from $\mathbf{P}$:

Step 1: Generate U_1 and set $N = 1 + [(n-1)U_1]$.
Step 2: Generate U_2 and set

$$X = \begin{cases} i_N, & \text{if } U_2 < Q_{i_N}^{(N)} \\ j_N, & \text{otherwise} \end{cases}$$

Remarks. **(i)** The preceding is called the alias method because by a renumbering of the $\mathbf{Q}$s we can always arrange things so that for each k, $Q_k^{(k)} > 0$. (That is, we can arrange things so that the kth two-point mass function gives positive weight to the value k.) Hence, the procedure calls for simulating N, equally likely to be $1, 2, \ldots,$ or $n-1$, and then if $N = k$ it either accepts k as the value of X, or it accepts for the value of X the "alias" of k (namely, the other value that $\mathbf{Q}^{(k)}$ gives positive weight).

(ii) Actually, it is not necessary to generate a new random number in step 2. Because $N-1$ is the integer part of $(n-1)U_1$, it follows that the remainder $(n-1)U_1 -$

$(N-1)$ is independent of U_1 and is uniformly distributed in $(0, 1)$. Hence, rather than generating a new random number U_2 in step 2, we can use $(n-1)U_1 - (N-1) = (n-1)U_1 - [(n-1)U_1]$.

Example 11.11. Let us return to the problem of Example 11.2, which considers a list of n, not necessarily distinct, items. Each item has a value—$v(i)$ being the value of the item in position i—and we are interested in estimating

$$v = \sum_{i=1}^{n} v(i)/m(i)$$

where $m(i)$ is the number of times the item in position i appears on the list. In words, v is the sum of the values of the (distinct) items on the list.

To estimate v, note that if X is a random variable such that

$$P\{X = i\} = v(i) \Big/ \sum_{1}^{n} v(j), \quad i = 1, \ldots, n$$

then

$$E[1/m(X)] = \frac{\sum_i v(i)/m(i)}{\sum_j v(j)} = v \Big/ \sum_{j=1}^{n} v(j)$$

Hence, we can estimate v by using the alias (or any other) method to generate independent random variables $X_1, \ldots, X_k$ having the same distribution as X and then estimating v by

$$v \approx \frac{1}{k} \sum_{j=1}^{n} v(j) \sum_{i=1}^{k} 1/m(X_i)$$

■

11.5 Stochastic Processes

We can easily simulate a stochastic process by simulating a sequence of random variables. For instance, to simulate the first t time units of a renewal process having interarrival distribution F we can simulate independent random variables $X_1, X_2, \ldots$ having distribution F, stopping at

$$N = \min\{n: X_1 + \cdots + X_n > t\}$$

The $X_i, i \geq 1$, represent the interarrival times of the renewal process and so the preceding simulation yields $N - 1$ events by time t—the events occurring at times $X_1, X_1 + X_2, \ldots, X_1 + \cdots + X_{N-1}$.

Actually there is another approach for simulating a Poisson process that is quite efficient. Suppose we want to simulate the first t time units of a Poisson process having rate λ. To do so, we can first simulate $N(t)$, the number of events by t, and then use the result that given the value of $N(t)$, the set of $N(t)$ event times is distributed as a set of n independent uniform $(0, t)$ random variables. Hence, we start by simulating $N(t)$, a Poisson random variable with mean λt (by one of the methods given in Example 11.9). Then, if $N(t) = n$, generate a new set of n random numbers—call them $U_1, \ldots, U_n$—and $\{tU_1, \ldots, tU_n\}$ will represent the set of $N(t)$ event times. If we could stop here this would be much more efficient than simulating the exponentially distributed interarrival times. However, we usually desire the event times in increasing order—for instance, for $s < t$,

$$N(s) = \text{number of } U_i : tU_i \leq s$$

and so to compute the function $N(s), s \leq t$, it is best to first order the values $U_i, i = 1, \ldots, n$ before multiplying by t. However, in doing so you should not use an all-purpose sorting algorithm, such as quick sort (see Example 3.14), but rather one that takes into account that the elements to be sorted come from a uniform (0, 1) population. Such a sorting algorithm of n uniform (0, 1) variables is as follows: Rather than a single list to be sorted of length n we will consider n ordered, or linked, lists of random size. The value U will be put in list i if its value is between $(i - 1)/n$ and i/n—that is, U is put in list $[nU] + 1$. The individual lists are then ordered, and the total linkage of all the lists is the desired ordering. As almost all of the n lists will be of relatively small size (for instance, if $n = 1000$ the mean number of lists of size greater than 4 is (using the Poisson approximation to the binomial) approximately equal to $1000(1 - \frac{65}{24}e^{-1}) \simeq 4$) the sorting of individual lists will be quite quick, and so the running time of such an algorithm will be proportional to n (rather than to $n \log n$ as in the best all-purpose sorting algorithms).

An extremely important counting process for modeling purposes is the nonhomogeneous Poisson process, which relaxes the Poisson process assumption of stationary increments. Thus it allows for the possibility that the arrival rate need not be constant but can vary with time. However, there are few analytical studies that assume a nonhomogeneous Poisson arrival process for the simple reason that such models are not usually mathematically tractable. (For example, there is no known expression for the average customer delay in the single-server exponential service distribution queueing model that assumes a nonhomogeneous arrival process.)[3] Clearly such models are strong candidates for simulation studies.

11.5.1 Simulating a Nonhomogeneous Poisson Process

We now present three methods for simulating a nonhomogeneous Poisson process having intensity function $\lambda(t), 0 \leq t < \infty$.

[3] One queueing model that assumes a nonhomogeneous Poisson arrival process and is mathematically tractable is the infinite server model.

Method 1. Sampling a Poisson Process

To simulate the first T time units of a nonhomogeneous Poisson process with intensity function $\lambda(t)$, let λ be such that

$$\lambda(t) \leq \lambda \quad \text{for all } t \leq T$$

Now, as shown in Chapter 5, such a nonhomogeneous Poisson process can be generated by a random selection of the event times of a Poisson process having rate λ. That is, if an event of a Poisson process with rate λ that occurs at time t is counted (independently of what has transpired previously) with probability $\lambda(t)/\lambda$ then the process of counted events is a nonhomogeneous Poisson process with intensity function $\lambda(t), 0 \leq t \leq T$. Hence, by simulating a Poisson process and then randomly counting its events, we can generate the desired nonhomogeneous Poisson process. We thus have the following procedure:

Generate independent random variables $X_1, U_1, X_2, U_2, \ldots$ where the X_i are exponential with rate λ and the U_i are random numbers, stopping at

$$N = \min\left\{n: \sum_{i=1}^{n} X_i > T\right\}$$

Now let, for $j = 1, \ldots, N-1$,

$$I_j = \begin{cases} 1, & \text{if } U_j \leq \lambda\left(\sum_{i=1}^{j} X_i\right)/\lambda \\ 0, & \text{otherwise} \end{cases}$$

and set

$$J = \{j: I_j = 1\}$$

Thus, the counting process having events at the set of times $\{\sum_{i=1}^{j} X_i: j \in J\}$ constitutes the desired process.

The foregoing procedure, referred to as the thinning algorithm (because it "thins" the homogeneous Poisson points) will clearly be most efficient, in the sense of having the fewest number of rejected event times, when $\lambda(t)$ is near λ throughout the interval. Thus, an obvious improvement is to break up the interval into subintervals and then use the procedure over each subinterval. That is, determine appropriate values $k, 0 < t_1 < t_2 < \cdots < t_k < T, \lambda_1, \ldots, \lambda_{k+1}$, such that

$$\lambda(s) \leq \lambda_i \quad \text{when } t_{i-1} \leq s < t_i, i = 1, \ldots, k+1 \quad (\text{where } t_0 = 0,\ t_{k+1} = T) \tag{11.9}$$

Now simulate the nonhomogeneous Poisson process over the interval (t_{i-1}, t_i) by generating exponential random variables with rate λ_i and accepting the generated event occurring at time $s, s \in (t_{i-1}, t_i)$, with probability $\lambda(s)/\lambda_i$. Because of the memoryless property of the exponential and the fact that the rate of an exponential can be changed upon multiplication by a constant, it follows that there is no loss of efficiency

in going from one subinterval to the next. In other words, if we are at $t \in [t_{i-1}, t_i)$ and generate X, an exponential with rate λ_i, which is such that $t + X > t_i$ then we can use $\lambda_i[X - (t_i - t)]/\lambda_{i+1}$ as the next exponential with rate λ_{i+1}. Thus, we have the following algorithm for generating the first t time units of a nonhomogeneous Poisson process with intensity function $\lambda(s)$ when the relations (11.9) are satisfied. In the algorithm, t will represent the present time and I the present interval (that is, $I = i$ when $t_{i-1} \le t < t_i$).

Step 1: $t = 0, I = 1$.

Step 2: Generate an exponential random variable X having rate λ_I.

Step 3: If $t + X < t_I$, reset $t = t + X$, generate a random number U, and accept the event time t if $U \le \lambda(t)/\lambda_I$. Return to step 2.

Step 4: (Step reached if $t + X \ge t_I$.) Stop if $I = k + 1$. Otherwise, reset $X = (X - t_I + t)\lambda_I/\lambda_{I+1}$. Also reset $t = t_I$ and $I = I + 1$, and go to step 3.

Suppose now that over some subinterval (t_{i-1}, t_i) it follows that $\underline{\lambda}_i > 0$ where

$$\underline{\lambda}_i \equiv \text{infimum } \{\lambda(s): t_{i-1} \le s < t_i\}$$

In such a situation, we should not use the thinning algorithm directly but rather should first simulate a Poisson process with rate $\underline{\lambda}_i$ over the desired interval and then simulate a nonhomogeneous Poisson process with the intensity function $\lambda(s) - \underline{\lambda}_i$ when $s \in (t_{i-1}, t_i)$. (The final exponential generated for the Poisson process, which carries one beyond the desired boundary, need not be wasted but can be suitably transformed so as to be reusable.) The superposition (or, merging) of the two processes yields the desired process over the interval. The reason for doing it this way is that it saves the need to generate uniform random variables for a Poisson distributed number, with mean $\underline{\lambda}_i(t_i - t_{i-1})$ of the event times. For instance, consider the case where

$$\lambda(s) = 10 + s, \quad 0 < s < 1$$

Using the thinning method with $\lambda = 11$ would generate an expected number of 11 events each of which would require a random number to determine whether or not to accept it. On the other hand, to generate a Poisson process with rate 10 and then merge it with a generated nonhomogeneous Poisson process with rate $\lambda(s) = s, 0 < s < 1$, would yield an equally distributed number of event times but with the expected number needing to be checked to determine acceptance being equal to 1.

Another way to make the simulation of nonhomogeneous Poisson processes more efficient is to make use of superpositions. For instance, consider the process where

$$\lambda(t) = \begin{cases} \exp\{t^2\}, & 0 < t < 1.5 \\ \exp\{2.25\}, & 1.5 < t < 2.5 \\ \exp\{(4-t)^2\}, & 2.5 < t < 4 \end{cases}$$

A plot of this intensity function is given in Fig. 11.3. One way of simulating this process up to time 4 is to first generate a Poisson process with rate 1 over this interval; then generate a Poisson process with rate $e - 1$ over this interval, accept all events in

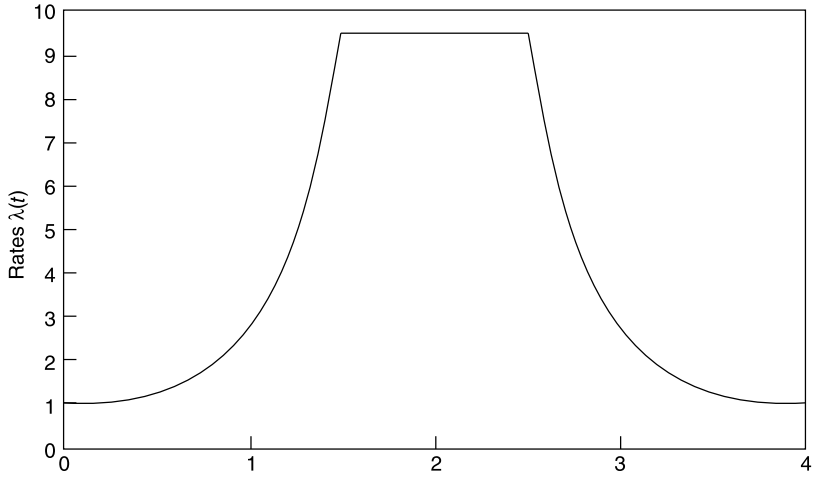

Figure 11.3

(1, 3), and only accept an event at time t that is not contained in (1, 3) with probability $[\lambda(t) - 1]/(e - 1)$; then generate a Poisson process with rate $e^{2.25} - e$ over the interval (1, 3), accepting all event times between 1.5 and 2.5 and any event time t outside this interval with probability $[\lambda(t) - e]/(e^{2.25} - e)$. The superposition of these processes is the desired nonhomogeneous Poisson process. In other words, what we have done is to break up $\lambda(t)$ into the following nonnegative parts:

$$\lambda(t) = \lambda_1(t) + \lambda_2(t) + \lambda_3(t), \quad 0 < t < 4$$

where

$$\lambda_1(t) \equiv 1,$$

$$\lambda_2(t) = \begin{cases} \lambda(t) - 1, & 0 < t < 1 \\ e - 1, & 1 < t < 3 \\ \lambda(t) - 1, & 3 < t < 4 \end{cases}$$

$$\lambda_3(t) = \begin{cases} \lambda(t) - e, & 1 < t < 1.5 \\ e^{2.25} - e, & 1.5 < t < 2.5 \\ \lambda(t) - e, & 2.5 < t < 3 \\ 0, & 3 < t < 4 \end{cases}$$

and where the thinning algorithm (with a single interval in each case) was used to simulate the constituent nonhomogeneous processes.

Method 2. Conditional Distribution of the Arrival Times

Recall the result for a Poisson process having rate λ that given the number of events by time T the set of event times are independent and identically distributed uniform $(0, T)$

random variables. Now suppose that each of these events is independently counted with a probability that is equal to $\lambda(t)/\lambda$ when the event occurred at time t. Hence, given the number of counted events, it follows that the set of times of these counted events are independent with a common distribution given by $F(s)$, where

$$\begin{aligned} F(s) &= P\{\text{time} \le s | \text{counted}\} \\ &= \frac{P\{\text{time} \le s, \text{counted}\}}{P\{\text{counted}\}} \\ &= \frac{\int_0^T P\{\text{time} \le s, \text{counted} | \text{time} = x\}\, dx/T}{P\{\text{counted}\}} \\ &= \frac{\int_0^s \lambda(x)\, dx}{\int_0^T \lambda(x)\, dx} \end{aligned}$$

The preceding (somewhat heuristic) argument thus shows that given n events of a nonhomogeneous Poisson process by time T the n event times are independent with a common density function

$$f(s) = \frac{\lambda(s)}{m(T)}, \quad 0 < s < T, \quad m(T) = \int_0^T \lambda(s)\, ds \tag{11.10}$$

Since $N(T)$, the number of events by time T, is Poisson distributed with mean $m(T)$, we can simulate the nonhomogeneous Poisson process by first simulating $N(T)$ and then simulating $N(T)$ random variables from the density function of (11.10).

Example 11.12. If $\lambda(s) = cs$, then we can simulate the first T time units of the nonhomogeneous Poisson process by first simulating $N(T)$, a Poisson random variable having mean $m(T) = \int_0^T cs\, ds = cT^2/2$, and then simulating $N(T)$ random variables having distribution

$$F(s) = \frac{s^2}{T^2}, \quad 0 < s < T$$

Random variables having the preceding distribution either can be simulated by use of the inverse transform method (since $F^{-1}(U) = T\sqrt{U}$) or by noting that F is the distribution function of $\max(TU_1, TU_2)$ when U_1 and U_2 are independent random numbers. ■

If the distribution function specified by Eq. (11.10) is not easily invertible, we can always simulate from (11.10) by using the rejection method where we either accept or reject simulated values of uniform $(0, T)$ random variables. That is, let $h(s) = 1/T, 0 < s < T$. Then

$$\frac{f(s)}{h(s)} = \frac{T\lambda(s)}{m(T)} \le \frac{\lambda T}{m(T)} \equiv C$$

where λ is a bound on $\lambda(s), 0 \le s \le T$. Hence, the rejection method is to generate random numbers U_1 and U_2 then accept TU_1 if

$$U_2 \le \frac{f(TU_1)}{Ch(TU_1)}$$

or, equivalently, if

$$U_2 \le \frac{\lambda(TU_1)}{\lambda}$$

Method 3. Simulating the Event Times

The third method we shall present for simulating a nonhomogeneous Poisson process having intensity function $\lambda(t), t \ge 0$ is probably the most basic approach—namely, to simulate the successive event times. So let $X_1, X_2, \ldots$ denote the event times of such a process. As these random variables are dependent we will use the conditional distribution approach to simulation. Hence, we need the conditional distribution of X_i given $X_1, \ldots, X_{i-1}$.

To start, note that if an event occurs at time x then, independent of what has occurred prior to x, the time until the next event has the distribution F_x given by

$$\begin{aligned}\bar{F}_x(t) &= P\{0 \text{ events in } (x, x+t) | \text{event at } x\} \\ &= P\{0 \text{ events in } (x, x+t)\} \quad \text{by independent increments} \\ &= \exp\left\{-\int_0^t \lambda(x+y)\,dy\right\}\end{aligned}$$

Differentiation yields that the density corresponding to F_x is

$$f_x(t) = \lambda(x+t)\exp\left\{-\int_0^t \lambda(x+y)\,dy\right\}$$

implying that the hazard rate function of F_x is

$$r_x(t) = \frac{f_x(t)}{\bar{F}_x(t)} = \lambda(x+t)$$

We can now simulate the event times $X_1, X_2, \ldots$ by simulating X_1 from F_0; then if the simulated value of X_1 is x_1, simulate X_2 by adding x_1 to a value generated from F_{x_1}, and if this sum is x_2 simulate X_3 by adding x_2 to a value generated from F_{x_2}, and so on. The method used to simulate from these distributions should depend, of course, on the form of these distributions. However, it is interesting to note that if we let λ be such that $\lambda(t) \le \lambda$ and use the hazard rate method to simulate, then we end up with the approach of Method 1 (we leave the verification of this fact as an exercise). Sometimes, however, the distributions F_x can be easily inverted and so the inverse transform method can be applied.

Example 11.13. Suppose that $\lambda(x) = 1/(x+a)$, $x \geq 0$. Then

$$\int_0^t \lambda(x+y)\,dy = \log\left(\frac{x+a+t}{x+a}\right)$$

Hence,

$$F_x(t) = 1 - \frac{x+a}{x+a+t} = \frac{t}{x+a+t}$$

and so

$$F_x^{-1}(u) = (x+a)\frac{u}{1-u}$$

We can, therefore, simulate the successive event times $X_1, X_2, \ldots$ by generating $U_1, U_2, \ldots$ and then setting

$$X_1 = \frac{aU_1}{1-U_1},$$
$$X_2 = (X_1 + a)\frac{U_2}{1-U_2} + X_1$$

and, in general,

$$X_j = (X_{j-1} + a)\frac{U_j}{1-U_j} + X_{j-1}, \quad j \geq 2$$ ■

11.5.2 Simulating a Two-Dimensional Poisson Process

A point process consisting of randomly occurring points in the plane is said to be a two-dimensional Poisson process having rate λ if

(a) the number of points in any given region of area A is Poisson distributed with mean λA; and
(b) the numbers of points in disjoint regions are independent.

For a given fixed point **O** in the plane, we now show how to simulate events occurring according to a two-dimensional Poisson process with rate λ in a circular region of radius r centered about **O**. Let $R_i, i \geq 1$, denote the distance between **O** and its ith nearest Poisson point, and let $C(a)$ denote the circle of radius a centered at **O**. Then

$$P\left\{\pi R_1^2 > b\right\} = P\left\{R_1 > \sqrt{\frac{b}{\pi}}\right\} = P\left\{\text{no points in } C\left(\sqrt{b/\pi}\right)\right\} = e^{-\lambda b}$$

Also, with $C(a_2) - C(a_1)$ denoting the region between $C(a_2)$ and $C(a_1)$:

$$P\left\{\pi R_2^2 - \pi R_1^2 > b | R_1 = r\right\}$$

$$
\begin{aligned}
&= P\left\{R_2 > \sqrt{(b+\pi r^2)/\pi} \,|\, R_1 = r\right\} \\
&= P\left\{\text{no points in } C\left(\sqrt{(b+\pi r^2)/\pi}\right) - C(r) \,|\, R_1 = r\right\} \\
&= P\left\{\text{no points in } C\left(\sqrt{(b+\pi r^2)/\pi}\right) - C(r)\right\} \quad \text{by (b)} \\
&= e^{-\lambda b}
\end{aligned}
$$

In fact, the same argument can be repeated to obtain the following.

Proposition 11.6. *With* $R_0 = 0$,

$$\pi R_i^2 - \pi R_{i-1}^2, \quad i \geq 1,$$

are independent exponentials with rate λ.

In other words, the amount of area that needs to be traversed to encompass a Poisson point is exponential with rate λ. Since, by symmetry, the respective angles of the Poisson points are independent and uniformly distributed over $(0, 2\pi)$, we thus have the following algorithm for simulating the Poisson process over a circular region of radius r about **O**:

Step 1: Generate independent exponentials with rate 1, $X_1, X_2, \ldots$, stopping at

$$N = \min\left\{n: \frac{X_1 + \cdots + X_n}{\lambda\pi} > r^2\right\}$$

Step 2: If $N = 1$, stop. There are no points in $C(r)$. Otherwise, for $i = 1, \ldots, N-1$, set

$$R_i = \sqrt{(X_1 + \cdots + X_i)/\lambda\pi}$$

Step 3: Generate independent uniform (0, 1) random variables $U_1, \ldots, U_{N-1}$.

Step 4: Return the $N-1$ Poisson points in $C(r)$ whose polar coordinates are

$$(R_i, 2\pi U_i), \quad i = 1, \ldots, N-1$$

The preceding algorithm requires, on average, $1+\lambda\pi r^2$ exponentials and an equal number of uniform random numbers. Another approach to simulating points in $C(r)$ is to first simulate N, the number of such points, and then use the fact that, given N, the points are uniformly distributed in $C(r)$. This latter procedure requires the simulation of N, a Poisson random variable with mean $\lambda\pi r^2$; we must then simulate N uniform points on $C(r)$, by simulating R from the distribution $F_R(a) = a^2/r^2$ (see Exercise 25) and θ from uniform $(0, 2\pi)$ and must then sort these N uniform values in increasing order of R. The main advantage of the first procedure is that it eliminates the need to sort.

The preceding algorithm can be thought of as the fanning out of a circle centered at **O** with a radius that expands continuously from 0 to r. The successive radii at which

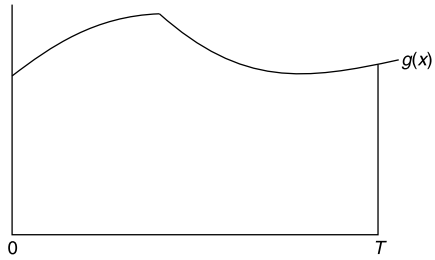

Figure 11.4

Poisson points are encountered is simulated by noting that the additional area necessary to encompass a Poisson point is always, independent of the past, exponential with rate λ. This technique can be used to simulate the process over noncircular regions. For instance, consider a nonnegative function $g(x)$, and suppose we are interested in simulating the Poisson process in the region between the x-axis and g with x going from 0 to T (see Fig. 11.4). To do so we can start at the left-hand end and fan vertically to the right by considering the successive areas $\int_0^a g(x)\,dx$. Now if $X_1 < X_2 < \cdots$ denote the successive projections of the Poisson points on the x-axis, then analogous to Proposition 11.6, it will follow that (with $X_0 = 0$) $\lambda \int_{X_{i-1}}^{X_i} g(x)\,dx, i \geq 1$, will be independent exponentials with rate 1. Hence, we should simulate $\epsilon_1, \epsilon_2, \ldots$, independent exponentials with rate 1, stopping at

$$N = \min\left\{n: \epsilon_1 + \cdots + \epsilon_n > \lambda \int_0^T g(x)\,dx\right\}$$

and determine $X_1, \ldots, X_{N-1}$ by

$$\begin{aligned}
\lambda \int_0^{X_1} g(x)\,dx &= \epsilon_1,\\
\lambda \int_{X_1}^{X_2} g(x)\,dx &= \epsilon_2,\\
&\vdots\\
\lambda \int_{X_{N-2}}^{X_{N-1}} g(x)\,dx &= \epsilon_{N-1}
\end{aligned}$$

If we now simulate $U_1, \ldots, U_{N-1}$—independent uniform $(0, 1)$ random numbers—then as the projection on the y-axis of the Poisson point whose x-coordinate is X_i is uniform on $(0, g(X_i))$, it follows that the simulated Poisson points in the interval are $(X_i, U_i g(X_i)), i = 1, \ldots, N-1$.

Of course, the preceding technique is most useful when g is regular enough so that the foregoing equations can be solved for the X_i. For instance, if $g(x) = y$ (and so the

region of interest is a rectangle), then

$$X_i = \frac{\epsilon_1 + \cdots + \epsilon_i}{\lambda y}, \quad i = 1, \ldots, N-1$$

and the Poisson points are

$$(X_i, yU_i), \quad i = 1, \ldots, N-1$$

11.6 Variance Reduction Techniques

Let $X_1, \ldots, X_n$ have a given joint distribution, and suppose we are interested in computing

$$\theta \equiv E[g(X_1, \ldots, X_n)]$$

where g is some specified function. It is often the case that it is not possible to analytically compute the preceding, and when such is the case we can attempt to use simulation to estimate θ. This is done as follows: Generate $X_1^{(1)}, \ldots, X_n^{(1)}$ having the same joint distribution as $X_1, \ldots, X_n$ and set

$$Y_1 = g\left(X_1^{(1)}, \ldots, X_n^{(1)}\right)$$

Now, simulate a second set of random variables (independent of the first set) $X_1^{(2)}, \ldots, X_n^{(2)}$ having the distribution of $X_1, \ldots, X_n$ and set

$$Y_2 = g\left(X_1^{(2)}, \ldots, X_n^{(2)}\right)$$

Continue this until you have generated k (some predetermined number) sets, and so have also computed $Y_1, Y_2, \ldots, Y_k$. Now, $Y_1, \ldots, Y_k$ are independent and identically distributed random variables each having the same distribution of $g(X_1, \ldots, X_n)$. Thus, if we let $\bar{Y}$ denote the average of these k random variables—that is,

$$\bar{Y} = \sum_{i=1}^{k} Y_i / k$$

then

$$E[\bar{Y}] = \theta,$$
$$E\left[(\bar{Y} - \theta)^2\right] = \mathrm{Var}(\bar{Y})$$

Hence, we can use $\bar{Y}$ as an estimate of θ. As the expected square of the difference between $\bar{Y}$ and θ is equal to the variance of $\bar{Y}$, we would like this quantity to be as

small as possible. In the preceding situation, $\text{Var}(\bar{Y}) = \text{Var}(Y_i)/k$, which is usually not known in advance but must be estimated from the generated values $Y_1, \ldots, Y_n$. We now present three general techniques for reducing the variance of our estimator.

11.6.1 Use of Antithetic Variables

In the preceding situation, suppose that we have generated Y_1 and Y_2, identically distributed random variables having mean θ. Now,

$$\begin{aligned}\text{Var}\left(\frac{Y_1 + Y_2}{2}\right) &= \frac{1}{4}[\text{Var}(Y_1) + \text{Var}(Y_2) + 2\text{Cov}(Y_1, Y_2)] \\ &= \frac{\text{Var}(Y_1)}{2} + \frac{\text{Cov}(Y_1, Y_2)}{2}\end{aligned}$$

Hence, it would be advantageous (in the sense that the variance would be reduced) if Y_1 and Y_2 rather than being independent were negatively correlated. To see how we could arrange this, let us suppose that the random variables $X_1, \ldots, X_n$ are independent and, in addition, that each is simulated via the inverse transform technique. That is, X_i is simulated from $F_i^{-1}(U_i)$ where U_i is a random number and F_i is the distribution of X_i. Hence, Y_1 can be expressed as

$$Y_1 = g\left(F_1^{-1}(U_1), \ldots, F_n^{-1}(U_n)\right)$$

Now, since $1 - U$ is also uniform over $(0, 1)$ whenever U is a random number (and is negatively correlated with U) it follows that Y_2 defined by

$$Y_2 = g\left(F_1^{-1}(1 - U_1), \ldots, F_n^{-1}(1 - U_n)\right)$$

will have the same distribution as Y_1. Hence, if Y_1 and Y_2 were negatively correlated, then generating Y_2 by this means would lead to a smaller variance than if it were generated by a new set of random numbers. (In addition, there is a computational savings since rather than having to generate n additional random numbers, we need only subtract each of the previous n from 1.) The following theorem will be the key to showing that this technique—known as the use of *antithetic variables*—will lead to a reduction in variance whenever g is a monotone function.

Theorem 11.1. *If $X_1, \ldots, X_n$ are independent, then, for any increasing functions f and g of n variables,*

$$E[f(\mathbf{X})g(\mathbf{X})] \geq E[f(\mathbf{X})]E[g(\mathbf{X})] \tag{11.11}$$

where $\mathbf{X} = (X_1, \ldots, X_n)$.

Proof. The proof is by induction on n. To prove it when $n = 1$, let f and g be increasing functions of a single variable. Then, for any x and y,

$$(f(x) - f(y))(g(x) - g(y)) \geq 0$$

since if $x \geq y$ ($x \leq y$) then both factors are nonnegative (nonpositive). Hence, for any random variables X and Y,

$$(f(X) - f(Y))(g(X) - g(Y)) \geq 0$$

implying that

$$E[(f(X) - f(Y))(g(X) - g(Y))] \geq 0$$

or, equivalently,

$$E[f(X)g(X)] + E[f(Y)g(Y)] \geq E[f(X)g(Y)] + E[f(Y)g(X)]$$

If we suppose that X and Y are independent and identically distributed, as in this case, then

$$E[f(X)g(X)] = E[f(Y)g(Y)],$$
$$E[f(X)g(Y)] = E[f(Y)g(X)] = E[f(X)]E[g(X)]$$

and so we obtain the result when $n = 1$.

So assume that (11.11) holds for $n - 1$ variables, and now suppose that $X_1, \ldots, X_n$ are independent and f and g are increasing functions. Then

$$\begin{aligned}
&E[f(\mathbf{X})g(\mathbf{X})|X_n = x_n] \\
&\quad = E[f(X_1, \ldots, X_{n-1}, x_n)g(X_1, \ldots, X_{n-1}, x_n)|X_n = x] \\
&\quad = E[f(X_1, \ldots, X_{n-1}, x_n)g(X_1, \ldots, X_{n-1}, x_n)] \quad \text{by independence} \\
&\quad \geq E[f(X_1, \ldots, X_{n-1}, x_n)]E[g(X_1, \ldots, X_{n-1}, x_n)] \\
&\qquad \text{by the induction hypothesis} \\
&\quad = E[f(\mathbf{X})|X_n = x_n]E[g(\mathbf{X})|X_n = x_n]
\end{aligned}$$

Hence,

$$E[f(\mathbf{X})g(\mathbf{X})|X_n] \geq E[f(\mathbf{X})|X_n]E[g(\mathbf{X})|X_n]$$

and, upon taking expectations of both sides,

$$\begin{aligned}
E[f(\mathbf{X})g(\mathbf{X})] &\geq E\big[E[f(\mathbf{X})|X_n]E[g(\mathbf{X})|X_n]\big] \\
&\geq E[f(\mathbf{X})]E[g(\mathbf{X})]
\end{aligned}$$

The last inequality follows because $E[f(\mathbf{X})|X_n]$ and $E[g(\mathbf{X})|X_n]$ are both increasing functions of X_n, and so, by the result for $n = 1$,

$$\begin{aligned}
E\big[E[f(\mathbf{X})|X_n]E[g(\mathbf{X})|X_n]\big] &\geq E\big[E[f(\mathbf{X})|X_n]\big]\,E\big[E[g(\mathbf{X})|X_n]\big] \\
&= E[f(\mathbf{X})]E[g(\mathbf{X})]
\end{aligned}$$

■

Corollary 11.7. *If $U_1, \ldots, U_n$ are independent, and k is either an increasing or decreasing function, then*

$$\text{Cov}(k(U_1, \ldots, U_n), k(1 - U_1, \ldots, 1 - U_n)) \leq 0$$

Proof. Suppose k is increasing. As $-k(1 - U_1, \ldots, 1 - U_n)$ is increasing in $U_1, \ldots, U_n$, then, from Theorem 11.1,

$$\text{Cov}(k(U_1, \ldots, U_n), -k(1 - U_1, \ldots, 1 - U_n)) \geq 0$$

When k is decreasing just replace k by its negative. ■

Since $F_i^{-1}(U_i)$ is increasing in U_i (as F_i, being a distribution function, is increasing) it follows that $g(F_1^{-1}(U_1), \ldots, F_n^{-1}(U_n))$ is a monotone function of $U_1, \ldots, U_n$ whenever g is monotone. Hence, if g is monotone the antithetic variable approach of twice using each set of random numbers $U_1, \ldots, U_n$ by first computing $g(F_1^{-1}(U_1), \ldots, F_n^{-1}(U_n))$ and then $g(F_1^{-1}(1 - U_1), \ldots, F_n^{-1}(1 - U_n))$ will reduce the variance of the estimate of $E[g(X_1, \ldots, X_n)]$. That is, rather than generating k sets of n random numbers, we should generate $k/2$ sets and use each set twice.

Example 11.14 (Simulating the Reliability Function)**.** Consider a system of n components in which component i, independently of other components, works with probability $p_i, i = 1, \ldots, n$. Letting

$$X_i = \begin{cases} 1, & \text{if component } i \text{ works} \\ 0, & \text{otherwise} \end{cases}$$

suppose there is a monotone structure function ϕ such that

$$\phi(X_1, \ldots, X_n) = \begin{cases} 1, & \text{if the system works under } X_1, \ldots, X_n \\ 0, & \text{otherwise} \end{cases}$$

We are interested in using simulation to estimate

$$r(p_1, \ldots, p_n) \equiv E[\phi(X_1, \ldots, X_n)] = P\{\phi(X_1, \ldots, X_n) = 1\}$$

Now, we can simulate the X_i by generating uniform random numbers $U_1, \ldots, U_n$ and then setting

$$X_i = \begin{cases} 1, & \text{if } U_i < p_i \\ 0, & \text{otherwise} \end{cases}$$

Hence, we see that

$$\phi(X_1, \ldots, X_n) = k(U_1, \ldots, U_n)$$

where k is a decreasing function of $U_1, \ldots, U_n$. Hence,

$$\text{Cov}(k(\mathbf{U}), k(\mathbf{1} - \mathbf{U})) \leq 0$$

and so the antithetic variable approach of using $U_1, \ldots, U_n$ to generate both $k(U_1, \ldots, U_n)$ and $k(1 - U_1, \ldots, 1 - U_n)$ results in a smaller variance than if an independent set of random numbers was used to generate the second k. ■

Example 11.15 (Simulating a Queueing System)**.** Consider a given queueing system, let D_i denote the delay in queue of the ith arriving customer, and suppose we are interested in simulating the system so as to estimate

$$\theta = E[D_1 + \cdots + D_n]$$

Let $X_1, \ldots, X_n$ denote the first n interarrival times and $S_1, \ldots, S_n$ the first n service times of this system, and suppose these random variables are all independent. Now in most systems $D_1 + \cdots + D_n$ will be a function of $X_1, \ldots, X_n, S_1, \ldots, S_n$—say,

$$D_1 + \cdots + D_n = g(X_1, \ldots, X_n, S_1, \ldots, S_n)$$

Also, g will usually be increasing in S_i and decreasing in $X_i, i = 1, \ldots, n$. If we use the inverse transform method to simulate $X_i, S_i, i = 1, \ldots, n$—say, $X_i = F_i^{-1}(1 - U_i)$, $S_i = G_i^{-1}(\bar{U}_i)$ where $U_1, \ldots, U_n, \bar{U}_1, \ldots, \bar{U}_n$ are independent uniform random numbers—then we may write

$$D_1 + \cdots + D_n = k(U_1, \ldots, U_n, \bar{U}_1, \ldots, \bar{U}_n)$$

where k is increasing in its variates. Hence, the antithetic variable approach will reduce the variance of the estimator of θ. (Thus, we would generate $U_i, \bar{U}_i, i = 1, \ldots, n$ and set $X_i = F_i^{-1}(1 - U_i)$ and $Y_i = G_i^{-1}(\bar{U}_i)$ for the first run, and $X_i = F_i^{-1}(U_i)$ and $Y_i = G_i^{-1}(1 - \bar{U}_i)$ for the second.) As all the U_i and $\bar{U}_i$ are independent, however, this is equivalent to setting $X_i = F_i^{-1}(U_i)$, $Y_i = G_i^{-1}(\bar{U}_i)$ in the first run and using $1 - U_i$ for U_i and $1 - \bar{U}_i$ for $\bar{U}_i$ in the second. ■

11.6.2 Variance Reduction by Conditioning

Let us start by recalling (see Proposition 3.1) the conditional variance formula

$$\text{Var}(Y) = E[\text{Var}(Y|Z)] + \text{Var}(E[Y|Z]) \tag{11.12}$$

Now suppose we are interested in estimating $E[g(X_1, \ldots, X_n)]$ by simulating $\mathbf{X} = (X_1, \ldots, X_n)$ and then computing $Y = g(X_1, \ldots, X_n)$. Now, if for some random variable Z we can compute $E[Y|Z]$ then, as $\text{Var}(Y|Z) \geq 0$, it follows from the conditional variance formula that

$$\text{Var}(E[Y|Z]) \leq \text{Var}(Y)$$

implying, since $E[E[Y|Z]] = E[Y]$, that $E[Y|Z]$ is a better estimator of $E[Y]$ than is Y.

In many situations, there are a variety of Z_i that can be conditioned on to obtain an improved estimator. Each of these estimators $E[Y|Z_i]$ will have mean $E[Y]$ and

smaller variance than does the raw estimator Y. We now show that for any choice of weights $\lambda_i, \lambda_i \geq 0, \sum_i \lambda_i = 1$, $\sum_i \lambda_i E[Y|Z_i]$ is also an improvement over Y.

Proposition 11.8. *For any* $\lambda_i \geq 0, \sum_{i=1}^{\infty} \lambda_i = 1$,

(a) $E\left[\sum_i \lambda_i E[Y|Z_i]\right] = E[Y]$,

(b) $\text{Var}\left(\sum_i \lambda_i E[Y|Z_i]\right) \leq \text{Var}(Y)$.

Proof. The proof of (a) is immediate. To prove (b), let N denote an integer valued random variable independent of all the other random variables under consideration and such that

$$P\{N = i\} = \lambda_i, \quad i \geq 1$$

Applying the conditional variance formula twice yields

$$\begin{aligned}\text{Var}(Y) &\geq \text{Var}(E[Y|N, Z_N]) \\ &\geq \text{Var}(E[E[Y|N, Z_N]|Z_1, \ldots]) \\ &= \text{Var}\left(\sum_i \lambda_i E[Y|Z_i]\right)\end{aligned}$$

■

Example 11.16. Consider a queueing system having Poisson arrivals and suppose that any customer arriving when there are already N others in the system is lost. Suppose that we are interested in using simulation to estimate the expected number of lost customers by time t. The raw simulation approach would be to simulate the system up to time t and determine L, the number of lost customers for that run. A better estimate, however, can be obtained by conditioning on the total time in $[0, t]$ that the system is at capacity. Indeed, if we let T denote the time in $[0, t]$ that there are N in the system, then

$$E[L|T] = \lambda T$$

where λ is the Poisson arrival rate. Hence, a better estimate for $E[L]$ than the average value of L over all simulation runs can be obtained by multiplying the average value of T per simulation run by λ. If the arrival process were a nonhomogeneous Poisson process, then we could improve over the raw estimator L by keeping track of those time periods for which the system is at capacity. If we let $I_1, \ldots, I_C$ denote the time intervals in $[0, t]$ in which there are N in the system, then

$$E[L|I_1, \ldots, I_C] = \sum_{i=1}^{C} \int_{I_i} \lambda(s)\, ds$$

where $\lambda(s)$ is the intensity function of the nonhomogeneous Poisson arrival process. The use of the right side of the preceding would thus lead to a better estimate of $E[L]$ than the raw estimator L. ■

Example 11.17. Suppose that we wanted to estimate the expected sum of the times in the system of the first n customers in a queueing system. That is, if W_i is the time that the ith customer spends in the system, then we are interested in estimating

$$\theta = E\left[\sum_{i=1}^{n} W_i\right]$$

Let Y_i denote the "state of the system" at the moment at which the ith customer arrives. It can be shown[4] that for a wide class of models the estimator $\sum_{i=1}^{n} E[W_i|Y_i]$ has (the same mean and) a smaller variance than the estimator $\sum_{i=1}^{n} W_i$. (It should be noted that whereas it is immediate that $E[W_i|Y_i]$ has smaller variance than W_i, because of the covariance terms involved it is not immediately apparent that $\sum_{i=1}^{n} E[W_i|Y_i]$ has smaller variance than $\sum_{i=1}^{n} W_i$.) For instance, in the model $G/M/1$

$$E[W_i|Y_i] = (N_i + 1)/\mu$$

where N_i is the number in the system encountered by the ith arrival and $1/\mu$ is the mean service time; the result implies that $\sum_{i=1}^{n}(N_i + 1)/\mu$ is a better estimate of the expected total time in the system of the first n customers than is the raw estimator $\sum_{i=1}^{n} W_i$. ■

Example 11.18 (Estimating the Renewal Function by Simulation)**.** Consider a queueing model in which customers arrive daily in accordance with a renewal process having interarrival distribution F. However, suppose that at some fixed time T, for instance 5 P.M., no additional arrivals are permitted and those customers that are still in the system are serviced. At the start of the next and each succeeding day customers again begin to arrive in accordance with the renewal process. Suppose we are interested in determining the average time that a customer spends in the system. Upon using the theory of renewal reward processes (with a cycle starting every T time units), it can be shown that

$$\text{average time that a customer spends in the system} = \frac{E[\text{sum of the times in the system of arrivals in } (0, T)]}{m(T)}$$

where $m(T)$ is the expected number of renewals in $(0, T)$.

If we were to use simulation to estimate the preceding quantity, a run would consist of simulating a single day, and as part of a simulation run, we would observe the quantity $N(T)$, the number of arrivals by time T. Since $E[N(T)] = m(T)$, the natural simulation estimator of $m(T)$ would be the average (over all simulated days) value of $N(T)$ obtained. However, $\text{Var}(N(T))$ is, for large T, proportional to T (its asymptotic form being $T\sigma^2/\mu^3$, where σ^2 is the variance and μ the mean of the interarrival distribution F), and so, for large T, the variance of our estimator would be

[4] S. M. Ross, "Simulating Average Delay—Variance Reduction by Conditioning," *Probability in the Engineering and Informational Sciences* 2(3), (1988), pp. 309–312.

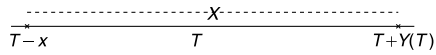

Figure 11.5 $A(T) = x$.

large. A considerable improvement can be obtained by using the analytic formula (see Section 7.3)

$$m(T) = \frac{T}{\mu} - 1 + \frac{E[Y(T)]}{\mu} \tag{11.13}$$

where $Y(T)$ denotes the time from T until the next renewal—that is, it is the excess life at T. Since the variance of $Y(T)$ does not grow with T (indeed, it converges to a finite value provided the moments of F are finite), it follows that for T large, we would do much better by using the simulation to estimate $E[Y(T)]$ and then using Eq. (11.13) to estimate $m(T)$.

However, by employing conditioning, we can improve further on our estimate of $m(T)$. To do so, let $A(T)$ denote the age of the renewal process at time T—that is, it is the time at T since the last renewal. Then, rather than using the value of $Y(T)$, we can reduce the variance by considering $E[Y(T)|A(T)]$. Now, knowing that the age at T is equal to x is equivalent to knowing that there was a renewal at time $T - x$ and the next interarrival time X is greater than x. Since the excess at T will equal $X - x$ (see Fig. 11.5), it follows that

$$\begin{aligned} E[Y(T)|A(T) = x] &= E[X - x|X > x] \\ &= \int_0^\infty \frac{P\{X - x > t\}}{P\{X > x\}}\, dt \\ &= \int_0^\infty \frac{1 - F(t + x)}{1 - F(x)}\, dt \end{aligned}$$

which can be numerically evaluated if necessary.

As an illustration of the preceding note that if the renewal process is a Poisson process with rate λ, then the raw simulation estimator $N(T)$ will have variance λT; since $Y(T)$ will be exponential with rate λ, the estimator based on (11.13) will have variance λ^2 Var $\{Y(T)\} = 1$. On the other hand, since $Y(T)$ will be independent of $A(T)$ (and $E[Y(T)|A(T)] = 1/\lambda$), it follows that the variance of the improved estimator $E[Y(T)|A(T)]$ is 0. That is, conditioning on the age at time T yields, in this case, the exact answer. ■

Example 11.19. Consider the $M/G/1$ queueing system where customers arrive in accordance with a Poisson process with rate λ to a single server having service distribution G with mean $E[S]$. Suppose that, for a specified time t_0, the server will take a break at the first time $t \geq t_0$ at which the system is empty. That is, if $X(t)$ is the number of customers in the system at time t, then the server will take a break at time

$$T = \min\{t \geq t_0 \colon X(t) = 0\}$$

To efficiently use simulation to estimate $E[T]$, generate the system to time t_0; let R denote the remaining service time of the customer in service at time t_0, and let X_Q equal the number of customers waiting in queue at time t_0. (Note that R is equal to 0 if $X(t_0) = 0$, and $X_Q = (X(t_0) - 1)^+$.) Now, with N equal to the number of customers that arrive in the remaining service time R, it follows that if $N = n$ and $X_Q = n_Q$, then the additional amount of time from $t_0 + R$ until the server can take a break is equal to the amount of time that it takes until the system, starting with $n + n_Q$ customers, becomes empty. Because this is equal to the sum of $n + n_Q$ busy periods, it follows from Section 8.5.3 that

$$E[T|R, N, X_Q] = t_0 + R + (N + X_Q)\frac{E[S]}{1 - \lambda E[S]}$$

Consequently,

$$\begin{aligned} E[T|R, X_Q] &= E\big[E[T|R, N, X_Q]|R, X_Q\big] \\ &= t_0 + R + (E[N|R, X_Q] + X_Q)\frac{E[S]}{1 - \lambda E[S]} \\ &= t_0 + R + (\lambda R + X_Q)\frac{E[S]}{1 - \lambda E[S]} \end{aligned}$$

Thus, rather than using the generated value of T as the estimator from a simulation run, it is better to stop the simulation at time t_0 and use the estimator $t_0 + (\lambda R + X_Q)\frac{E[S]}{1-\lambda E[S]}$. ■

11.6.3 Control Variates

Again suppose we want to use simulation to estimate $E[g(\mathbf{X})]$ where $\mathbf{X} = (X_1, \ldots, X_n)$. But now suppose that for some function f the expected value of $f(\mathbf{X})$ is known—say, $E[f(\mathbf{X})] = \mu$. Then for any constant a we can also use

$$W = g(\mathbf{X}) + a(f(\mathbf{X}) - \mu)$$

as an estimator of $E[g(\mathbf{X})]$. Now,

$$\mathrm{Var}(W) = \mathrm{Var}(g(\mathbf{X})) + a^2\,\mathrm{Var}(f(\mathbf{X})) + 2a\,\mathrm{Cov}(g(\mathbf{X}), f(\mathbf{X}))$$

Simple calculus shows that the preceding is minimized when

$$a = \frac{-\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))}$$

and, for this value of a,

$$\mathrm{Var}(W) = \mathrm{Var}(g(\mathbf{X})) - \frac{[\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))]^2}{\mathrm{Var}(f(\mathbf{X}))}$$

Because $\text{Var}(f(\mathbf{X}))$ and $\text{Cov}(f(\mathbf{X}), g(\mathbf{X}))$ are usually unknown, the simulated data should be used to estimate these quantities.

Dividing the preceding equation by $\text{Var}(g(\mathbf{X}))$ shows that

$$\frac{\text{Var}(W)}{\text{Var}(g(\mathbf{X}))} = 1 - \text{Corr}^2(f(\mathbf{X}), g(\mathbf{X}))$$

where $\text{Corr}(X, Y)$ is the correlation between X and Y. Consequently, the use of a control variate will greatly reduce the variance of the simulation estimator whenever $f(\mathbf{X})$ and $g(\mathbf{X})$ are strongly correlated.

Example 11.20. Consider a continuous-time Markov chain that, upon entering state i, spends an exponential time with rate v_i in that state before making a transition into some other state, with the transition being into state j with probability $P_{i,j}$, $i \geq 0$, $j \neq i$. Suppose that costs are incurred at rate $C(i) \geq 0$ per unit time whenever the chain is in state i, $i \geq 0$. With $X(t)$ equal to the state at time t, and α being a constant such that $0 < \alpha < 1$, the quantity

$$W = \int_0^\infty e^{-\alpha t} C(X(t))\, dt$$

represents the total discounted cost. For a given initial state, suppose we want to use simulation to estimate $E[W]$. Whereas at first it might seem that we cannot obtain an unbiased estimator without simulating the continuous-time Markov chain for an infinite amount of time (which is clearly impossible), we can make use of the results of Example 5.1, which gives the equivalent expression for $E[W]$:

$$E[W] = E\left[\int_0^T C(X(t))\, dt\right]$$

where T is an exponential random variable with rate α that is independent of the continuous-time Markov chain. Therefore, we can first generate the value of T, then generate the states of the continuous-time Markov chain up to time T, to obtain the unbiased estimator $\int_0^T C(X(t))\, dt$. Because all the cost rates are nonnegative this estimator is strongly positively correlated with T, which will thus make an effective control variate. ■

Example 11.21 (A Queueing System)**.** Let D_{n+1} denote the delay in queue of the $n + 1$ customer in a queueing system in which the interarrival times are independent and identically distributed (i.i.d.) with distribution F having mean μ_F and are independent of the service times, which are i.i.d. with distribution G having mean μ_G. If X_i is the interarrival time between arrival i and $i + 1$, and if S_i is the service time of customer i, $i \geq 1$, we may write

$$D_{n+1} = g(X_1, \ldots, X_n, S_1, \ldots, S_n)$$

To take into account the possibility that the simulated variables X_i, S_i may by chance be quite different from what might be expected we can let

$$f(X_1, \ldots, X_n, S_1, \ldots, S_n) = \sum_{i=1}^{n} (S_i - X_i)$$

As $E[f(\mathbf{X}, \mathbf{S})] = n(\mu_G - \mu_F)$ we could use

$$g(\mathbf{X}, \mathbf{S}) + a[f(\mathbf{X}, \mathbf{S}) - n(\mu_G - \mu_F)]$$

as an estimator of $E[D_{n+1}]$. Since D_{n+1} and f are both increasing functions of $S_i, -X_i, i = 1, \ldots, n$ it follows from Theorem 11.1 that $f(\mathbf{X}, \mathbf{S})$ and D_{n+1} are positively correlated, and so the simulated estimate of a should turn out to be negative.

If we wanted to estimate the expected sum of the delays in queue of the first $N(T)$ arrivals, then we could use $\sum_{i=1}^{N(T)} S_i$ as our control variable. Indeed as the arrival process is usually assumed independent of the service times, it follows that

$$E\left[\sum_{i=1}^{N(T)} S_i\right] = E[S]E[N(T)]$$

where $E[N(T)]$ can either be computed by the method suggested in Section 7.8 or estimated from the simulation as in Example 11.18. This control variable could also be used if the arrival process were a nonhomogeneous Poisson with rate $\lambda(t)$; in this case,

$$E[N(T)] = \int_0^T \lambda(t)\, dt$$ ■

11.6.4 Importance Sampling

Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote a vector of random variables having a joint density function (or joint mass function in the discrete case) $f(\mathbf{x}) = f(x_1, \ldots, x_n)$, and suppose that we are interested in estimating

$$\theta = E[h(\mathbf{X})] = \int h(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$$

where the preceding is an n-dimensional integral. (If the X_i are discrete, then interpret the integral as an n-fold summation.)

Suppose that a direct simulation of the random vector $\mathbf{X}$, so as to compute values of $h(\mathbf{X})$, is inefficient, possibly because (a) it is difficult to simulate a random vector having density function $f(\mathbf{x})$, or (b) the variance of $h(\mathbf{X})$ is large, or (c) a combination of (a) and (b).

Another way in which we can use simulation to estimate θ is to note that if $g(\mathbf{x})$ is another probability density such that $f(\mathbf{x}) = 0$ whenever $g(\mathbf{x}) = 0$, then we can

express θ as

$$\begin{aligned}\theta &= \int \frac{h(\mathbf{x})f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x})\, d\mathbf{x} \\ &= E_g\left[\frac{h(\mathbf{X})f(\mathbf{X})}{g(\mathbf{X})}\right] \end{aligned} \tag{11.14}$$

where we have written E_g to emphasize that the random vector $\mathbf{X}$ has joint density $g(\mathbf{x})$.

It follows from Eq. (11.14) that θ can be estimated by successively generating values of a random vector $\mathbf{X}$ having density function $g(\mathbf{x})$ and then using as the estimator the average of the values of $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$. If a density function $g(\mathbf{x})$ can be chosen so that the random variable $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$ has a small variance then this approach—referred to as *importance sampling*—can result in an efficient estimator of θ.

Let us now try to obtain a feel for why importance sampling can be useful. To begin, note that $f(\mathbf{X})$ and $g(\mathbf{X})$ represent the respective likelihoods of obtaining the vector $\mathbf{X}$ when $\mathbf{X}$ is a random vector with respective densities f and g. Hence, if $\mathbf{X}$ is distributed according to g, then it will usually be the case that $f(\mathbf{X})$ will be small in relation to $g(\mathbf{X})$ and thus when $\mathbf{X}$ is simulated according to g the likelihood ratio $f(\mathbf{X})/g(\mathbf{X})$ will usually be small in comparison to 1. However, it is easy to check that its mean is 1:

$$E_g\left[\frac{f(\mathbf{X})}{g(\mathbf{X})}\right] = \int \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x})\, dx = \int f(\mathbf{x})\, d\mathbf{x} = 1$$

Thus we see that even though $f(\mathbf{X})/g(\mathbf{X})$ is usually smaller than 1, its mean is equal to 1; thus implying that it is occasionally large and so will tend to have a large variance. So how can $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$ have a small variance? The answer is that we can sometimes arrange to choose a density g such that those values of $\mathbf{x}$ for which $f(\mathbf{x})/g(\mathbf{x})$ is large are precisely the values for which $h(\mathbf{x})$ is exceedingly small, and thus the ratio $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$ is always small. Since this will require that $h(\mathbf{x})$ sometimes be small, importance sampling seems to work best when estimating a small probability; for in this case the function $h(\mathbf{x})$ is equal to 1 when $\mathbf{x}$ lies in some set and is equal to 0 otherwise.

We will now consider how to select an appropriate density g. We will find that the so-called tilted densities are useful. Let $M(t) = E_f[e^{tX}] = \int e^{tx} f(x)\, dx$ be the moment generating function corresponding to a one-dimensional density f.

Definition 11.2. A density function

$$f_t(x) = \frac{e^{tx} f(x)}{M(t)}$$

is called a *tilted* density of f, $-\infty < t < \infty$.

A random variable with density f_t tends to be larger than one with density f when $t > 0$ and tends to be smaller when $t < 0$.

In certain cases the tilted distributions f_t have the same parametric form as does f.

Example 11.22. If f is the exponential density with rate λ then

$$f_t(x) = Ce^{tx}\lambda e^{-\lambda x} = \lambda C e^{-(\lambda - t)x}$$

where $C = 1/M(t)$ does not depend on x. Therefore, for $t \leq \lambda$, f_t is an exponential density with rate $\lambda - t$.

If f is a Bernoulli probability mass function with parameter p, then

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Hence, $M(t) = E_f[e^{tX}] = pe^t + 1 - p$ and so

$$\begin{aligned} f_t(x) &= \frac{1}{M(t)}(pe^t)^x(1-p)^{1-x} \\ &= \left(\frac{pe^t}{pe^t + 1 - p}\right)^x \left(\frac{1-p}{pe^t + 1 - p}\right)^{1-x} \end{aligned} \tag{11.15}$$

That is, f_t is the probability mass function of a Bernoulli random variable with parameter

$$p_t = \frac{pe^t}{pe^t + 1 - p}$$

We leave it as an exercise to show that if f is a normal density with parameters μ and σ^2 then f_t is a normal density with mean $\mu + \sigma^2 t$ and variance σ^2. ■

In certain situations the quantity of interest is the sum of the independent random variables $X_1, \ldots, X_n$. In this case the joint density f is the product of one-dimensional densities. That is,

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

where f_i is the density function of X_i. In this situation it is often useful to generate the X_i according to their tilted densities, with a common choice of t employed.

Example 11.23. Let $X_1, \ldots, X_n$ be independent random variables having respective probability density (or mass) functions f_i, for $i = 1, \ldots, n$. Suppose we are interested in approximating the probability that their sum is at least as large as a, where a is much larger than the mean of the sum. That is, we are interested in

$$\theta = P\{S \geq a\}$$

where $S = \sum_{i=1}^n X_i$, and where $a > \sum_{i=1}^n E[X_i]$. Letting $I\{S \geq a\}$ equal 1 if $S \geq a$ and letting it be 0 otherwise, we have that

$$\theta = E_{\mathbf{f}}[I\{S \geq a\}]$$

where $\mathbf{f} = (f_1, \ldots, f_n)$. Suppose now that we simulate X_i according to the tilted mass function $f_{i,t}, i = 1, \ldots, n$, with the value of $t, t > 0$ left to be determined. The importance sampling estimator of θ would then be

$$\hat{\theta} = I\{S \geq a\} \prod \frac{f_i(X_i)}{f_{i,t}(X_i)}$$

Now,

$$\frac{f_i(X_i)}{f_{i,t}(X_i)} = M_i(t)e^{-tX_i}$$

and so

$$\hat{\theta} = I\{S \geq a\} M(t)e^{-tS}$$

where $M(t) = \prod M_i(t)$ is the moment generating function of S. Since $t > 0$ and $I\{S \geq a\}$ is equal to 0 when $S < a$, it follows that

$$I\{S \geq a\}e^{-tS} \leq e^{-ta}$$

and so

$$\hat{\theta} \leq M(t)e^{-ta}$$

To make the bound on the estimator as small as possible we thus choose $t, t > 0$, to minimize $M(t)e^{-ta}$. In doing so, we will obtain an estimator whose value on each iteration is between 0 and $\min_t M(t)e^{-ta}$. It can be shown that the minimizing t, call it t^*, is such that

$$E_{t^*}[S] = E_{t^*}\left[\sum_{i=1}^{n} X_i\right] = a$$

where, in the preceding, we mean that the expected value is to be taken under the assumption that the distribution of X_i is f_{i,t^*} for $i = 1, \ldots, n$.

For instance, suppose that $X_1, \ldots, X_n$ are independent Bernoulli random variables having respective parameters p_i, for $i = 1, \ldots, n$. Then, if we generate the X_i according to their tilted mass functions $p_{i,t}, i = 1, \ldots, n$, the importance sampling estimator of $\theta = P\{S \geq a\}$ is

$$\hat{\theta} = I\{S \geq a\}e^{-tS} \prod_{i=1}^{n} \left(p_i e^t + 1 - p_i\right)$$

Since $p_{i,t}$ is the mass function of a Bernoulli random variable with parameter $p_i e^t/(p_i e^t + 1 - p_i)$ it follows that

$$E_t\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \frac{p_i e^t}{p_i e^t + 1 - p_i}$$

The value of t that makes the preceding equal to a can be numerically approximated and then utilized in the simulation.

As an illustration, suppose that $n = 20$, $p_i = 0.4$, and $a = 16$. Then

$$E_t[S] = 20\frac{0.4e^t}{0.4e^t + 0.6}$$

Setting this equal to 16 yields, after a little algebra,

$$e^{t^*} = 6$$

Thus, if we generate the Bernoullis using the parameter

$$\frac{0.4e^{t^*}}{0.4e^{t^*} + 0.6} = 0.8$$

then because

$$M(t^*) = (0.4e^{t^*} + 0.6)^{20} \quad \text{and} \quad e^{-t^*S} = (1/6)^S$$

we see that the importance sampling estimator is

$$\hat{\theta} = I\{S \geq 16\}(1/6)^S 3^{20}$$

It follows from the preceding that

$$\hat{\theta} \leq (1/6)^{16}3^{20} = 81/2^{16} = 0.001236$$

That is, on each iteration the value of the estimator is between 0 and 0.001236. Since, in this case, θ is the probability that a binomial random variable with parameters 20, 0.4 is at least 16, it can be explicitly computed with the result $\theta = 0.000317$. Hence, the raw simulation estimator I, which on each iteration takes the value 0 if the sum of the Bernoullis with parameter 0.4 is less than 16 and takes the value 1 otherwise, will have variance

$$\text{Var}(I) = \theta(1 - \theta) = 3.169 \times 10^{-4}$$

On the other hand, it follows from the fact that $0 \leq \hat{\theta} \leq 0.001236$ that (see Exercise 33)

$$\text{Var}(\hat{\theta}) \leq 2.9131 \times 10^{-7}$$ ■

Example 11.24. Consider a single-server queue in which the times between successive customer arrivals have density function f and the service times have density g. Let D_n denote the amount of time that the nth arrival spends waiting in queue and

suppose we are interested in estimating $\alpha = P\{D_n \geq a\}$ when a is much larger than $E[D_n]$. Rather than generating the successive interarrival and service times according to f and g, respectively, they should be generated according to the densities f_{-t} and g_t, where t is a positive number to be determined. Note that using these distributions as opposed to f and g will result in smaller interarrival times (since $-t < 0$) and larger service times. Hence, there will be a greater chance that $D_n > a$ than if we had simulated using the densities f and g. The importance sampling estimator of α would then be

$$\hat{\alpha} = I\{D_n > a\}e^{t(S_n - Y_n)}[M_f(-t)M_g(t)]^n$$

where S_n is the sum of the first n interarrival times, Y_n is the sum of the first n service times, and M_f and M_g are the moment generating functions of the densities f and g, respectively. The value of t used should be determined by experimenting with a variety of different choices. ■

11.7 Determining the Number of Runs

Suppose that we are going to use simulation to generate r independent and identically distributed random variables $Y^{(1)}, \ldots, Y^{(r)}$ having mean μ and variance σ^2. We are then going to use

$$\bar{Y}_r = \frac{Y^{(1)} + \cdots + Y^{(r)}}{r}$$

as an estimate of μ. The precision of this estimate can be measured by its variance

$$\begin{aligned} \text{Var}(\bar{Y}_r) &= E[(\bar{Y}_r - \mu)^2] \\ &= \sigma^2/r \end{aligned}$$

Hence, we would want to choose r, the number of necessary runs, large enough so that σ^2/r is acceptably small. However, the difficulty is that σ^2 is not known in advance. To get around this, you should initially simulate k runs (where $k \geq 30$) and then use the simulated values $Y^{(1)}, \ldots, Y^{(k)}$ to estimate σ^2 by the sample variance

$$\sum_{i=1}^{k} (Y^{(i)} - \bar{Y}_k)^2/(k-1)$$

Based on this estimate of σ^2 the value of r that attains the desired level of precision can now be determined and an additional $r - k$ runs can be generated.

11.8 Generating from the Stationary Distribution of a Markov Chain

11.8.1 Coupling from the Past

Consider an irreducible Markov chain with states $1, \ldots, m$ and transition probabilities $P_{i,j}$ and suppose we want to generate the value of a random variable whose distribution is that of the stationary distribution of this Markov chain. Whereas we could *approximately* generate such a random variable by arbitrarily choosing an initial state, simulating the resulting Markov chain for a large fixed number of time periods, and then choosing the final state as the value of the random variable, we will now present a procedure that generates a random variable whose distribution is *exactly* that of the stationary distribution.

If, in theory, we generated the Markov chain starting at time $-\infty$ in any arbitrary state, then the state at time 0 would have the stationary distribution. So imagine that we do this, and suppose that a different person is to generate the next state at each of these times. Thus, if $X(-n)$, the state at time $-n$, is i, then person $-n$ would generate a random variable that is equal to j with probability $P_{i,j}$, $j = 1, \ldots, m$, and the value generated would be the state at time $-(n-1)$. Now suppose that person -1 wants to do his random variable generation early. Because he does not know what the state at time -1 will be, he generates a sequence of random variables $N_{-1}(i), i = 1, \ldots, m$, where $N_{-1}(i)$, the next state if $X(-1) = i$, is equal to j with probability $P_{i,j}$, $j = 1, \ldots, m$. If it results that $X(-1) = i$, then person -1 would report that the state at time 0 is

$$S_{-1}(i) = N_{-1}(i), \quad i = 1, \ldots, m$$

(That is, $S_{-1}(i)$ is the simulated state at time 0 when the simulated state at time -1 is i.)

Now suppose that person -2, hearing that person -1 is doing his simulation early, decides to do the same thing. She generates a sequence of random variables $N_{-2}(i), i = 1, \ldots, m$, where $N_{-2}(i)$ is equal to j with probability $P_{i,j}$, $j = 1, \ldots, m$. Consequently, if it is reported to her that $X(-2) = i$, then she will report that $X(-1) = N_{-2}(i)$. Combining this with the early generation of person -1 shows that if $X(-2) = i$, then the simulated state at time 0 is

$$S_{-2}(i) = S_{-1}(N_{-2}(i)), \quad i = 1, \ldots, m$$

Continuing in the preceding manner, suppose that person -3 generates a sequence of random variables $N_{-3}(i), i = 1, \ldots, m$, where $N_{-3}(i)$ is to be the generated value of the next state when $X(-3) = i$. Consequently, if $X(-3) = i$ then the simulated state at time 0 would be

$$S_{-3}(i) = S_{-2}(N_{-3}(i)), \quad i = 1, \ldots, m$$

Now suppose we continue the preceding, and so obtain the simulated functions

$$S_{-1}(i), S_{-2}(i), S_{-3}(i), \ldots, \quad i = 1, \ldots, m$$

Going backward in time in this manner, we will at some time, say $-r$, have a simulated function $S_{-r}(i)$ that is a constant function. That is, for some state j, $S_{-r}(i)$ will equal j for all states $i = 1, \ldots, m$. But this means that no matter what the simulated values from time $-\infty$ to $-r$, we can be certain that the simulated value at time 0 is j. Consequently, j can be taken as the value of a generated random variable whose distribution is exactly that of the stationary distribution of the Markov chain.

Example 11.25. Consider a Markov chain with states 1, 2, 3 and suppose that simulation yielded the values

$$N_{-1}(i) = \begin{cases} 3, & \text{if } i = 1 \\ 2, & \text{if } i = 2 \\ 2, & \text{if } i = 3 \end{cases}$$

and

$$N_{-2}(i) = \begin{cases} 1, & \text{if } i = 1 \\ 3, & \text{if } i = 2 \\ 1, & \text{if } i = 3 \end{cases}$$

Then

$$S_{-2}(i) = \begin{cases} 3, & \text{if } i = 1 \\ 2, & \text{if } i = 2 \\ 3, & \text{if } i = 3 \end{cases}$$

If

$$N_{-3}(i) = \begin{cases} 3, & \text{if } i = 1 \\ 1, & \text{if } i = 2 \\ 1, & \text{if } i = 3 \end{cases}$$

then

$$S_{-3}(i) = \begin{cases} 3, & \text{if } i = 1 \\ 3, & \text{if } i = 2 \\ 3, & \text{if } i = 3 \end{cases}$$

Therefore, no matter what the state is at time -3, the state at time 0 will be 3. ■

Remark. The procedure developed in this section for generating a random variable whose distribution is the stationary distribution of the Markov chain is called *coupling from the past*.

11.8.2 Another Approach

Consider a Markov chain whose state space is the nonnegative integers. Suppose the chain has stationary probabilities, and denote them by $\pi_i, i \geq 0$. We now present another way of simulating a random variable whose distribution is given by the $\pi_i, i \geq 0$, which can be utilized if the chain satisfies the following property. Namely, that for some state, which we will call state 0, and some positive number α

$$P_{i,0} \geq \alpha > 0$$

for all states i. That is, whatever the current state, the probability that the next state will be 0 is at least some positive value α.

To simulate a random variable distributed according to the stationary probabilities, start by simulating the Markov chain in the obvious manner. Namely, whenever the chain is in state i, generate a random variable that is equal to j with probability $P_{i,j}, j \geq 0$, and then set the next state equal to the generated value of this random variable. In addition, however, whenever a transition into state 0 occurs a coin, whose probability of coming up heads depends on the state from which the transition occurred, is flipped. Specifically, if the transition into state 0 was from state i, then the coin flipped has probability $\alpha/P_{i,0}$ of coming up heads. Call such a coin an i-coin, $i \geq 0$. If the coin comes up heads then we say that an event has occurred. Consequently, each transition of the Markov chain results in an event with probability α, implying that events occur at rate α. Now say that an event is an i-event if it resulted from a transition out of state i; that is, an event is an i-event if it resulted from the flip of an i-coin. Because π_i is the proportion of transitions that are out of state i, and each such transition will result in an i-event with probability α, it follows that the rate at which i-events occur is $\alpha\pi_i$. Therefore, the proportion of all events that are i-events is $\alpha\pi_i/\alpha = \pi_i, i \geq 0$.

Now, suppose that $X_0 = 0$. Fix i, and let I_j equal 1 if the jth event that occurs is an i-event, and let I_j equal 0 otherwise. Because an event always leaves the chain in state 0 it follows that $I_j, j \geq 1$, are independent and identically distributed random variables. Because the proportion of the I_j that are equal to 1 is π_i, we see that

$$\begin{aligned}\pi_i &= \lim_{n\to\infty} \frac{I_1 + \ldots + I_n}{n} \\ &= E[I_1] \\ &= P(I_1 = 1)\end{aligned}$$

where the second equality follows from the strong law of large numbers. Hence, if we let

$$T = \min\{n > 0 : \text{an event occurs at time } n\}$$

denote the time of the first event, then it follows from the preceding that

$$\pi_i = P(I_1 = 1) = P(X_{T-1} = i)$$

As the preceding is true for all states i, it follows that X_{T-1}, the state of the Markov chain at time $T-1$, has the stationary distribution.

Exercises

***1.** Suppose it is relatively easy to simulate from the distributions $F_i, i = 1, 2, \ldots, n$. If n is small, how can we simulate from

$$F(x) = \sum_{i=1}^{n} P_i F_i(x), \quad P_i \geq 0, \qquad \sum_i P_i = 1?$$

Give a method for simulating from

$$F(x) = \begin{cases} \dfrac{1 - e^{-2x} + 2x}{3}, & 0 < x < 1 \\ \dfrac{3 - e^{-2x}}{3}, & 1 < x < \infty \end{cases}$$

2. Give a method for simulating a negative binomial random variable.

***3.** Give a method for simulating a hypergeometric random variable.

4. Suppose we want to simulate a point located at random in a circle of radius r centered at the origin. That is, we want to simulate X, Y having joint density

$$f(x, y) = \frac{1}{\pi r^2}, \quad x^2 + y^2 \leq r^2$$

(a) Let $R = \sqrt{X^2 + Y^2}, \theta = \tan^{-1} Y/X$ denote the polar coordinates. Compute the joint density of R, θ and use this to give a simulation method. Another method for simulating X, Y is as follows:

Step 1: Generate independent random numbers U_1, U_2 and set $Z_1 = 2rU_1 - r$, $Z_2 = 2rU_2 - r$. Then Z_1, Z_2 is uniform in the square whose sides are of length $2r$ and which encloses, the circle of radius r (see Fig. 11.6).

Step 2: If (Z_1, Z_2) lies in the circle of radius r—that is, if $Z_1^2 + Z_2^2 \leq r^2$—set $(X, Y) = (Z_1, Z_2)$. Otherwise return to step 1.

(b) Prove that this method works, and compute the distribution of the number of random numbers it requires.

5. Suppose it is relatively easy to simulate from F_i for each $i = 1, \ldots, n$. How can we simulate from

(a) $F(x) = \prod_{i=1}^{n} F_i(x)$?

(b) $F(x) = 1 - \prod_{i=1}^{n} (1 - F_i(x))$?

(c) Give two methods for simulating from the distribution $F(x) = x^n, 0 < x < 1$.

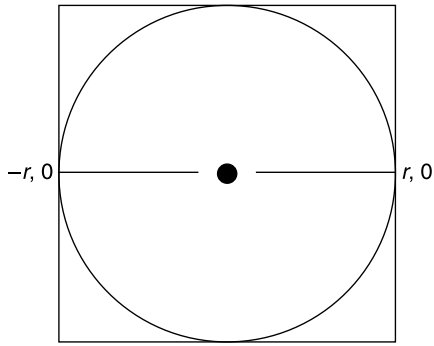

Figure 11.6

***6.** In Example 11.5 we simulated the absolute value of a standard normal by using the Von Neumann rejection procedure on exponential random variables with rate 1. This raises the question of whether we could obtain a more efficient algorithm by using a different exponential density—that is, we could use the density $g(x) = \lambda e^{-\lambda x}$. Show that the mean number of iterations needed in the rejection scheme is minimized when $\lambda = 1$.

7. Give an algorithm for simulating a random variable having density function

$$f(x) = 30(x^2 - 2x^3 + x^4), \quad 0 < x < 1$$

8. Consider the technique of simulating a gamma (n, λ) random variable by using the rejection method with g being an exponential density with rate λ/n.

(a) Show that the average number of iterations of the algorithm needed to generate a gamma is $n^n e^{1-n}/(n-1)!$.

(b) Use Stirling's approximation to show that for large n the answer to part (a) is approximately equal to $e[(n-1)/(2\pi)]^{1/2}$.

(c) Show that the procedure is equivalent to the following:

Step 1: Generate Y_1 and Y_2, independent exponentials with rate 1.

Step 2: If $Y_1 < (n-1)[Y_2 - \log(Y_2) - 1]$, return to step 1.

Step 3: Set $X = nY_2/\lambda$.

(d) Explain how to obtain an independent exponential along with a gamma from the preceding algorithm.

9. Set up the alias method for simulating from a binomial random variable with parameters $n = 6$, $p = 0.4$.

10. Explain how we can number the $\mathbf{Q}^{(k)}$ in the alias method so that k is one of the two points that $\mathbf{Q}^{(k)}$ gives weight.

Hint: Rather than giving the initial $\mathbf{Q}$ the name $\mathbf{Q}^{(1)}$, what else could we call it?

11. Complete the details of Example 11.10.

12. Let $X_1, \ldots, X_k$ be independent with

$$P\{X_i = j\} = \frac{1}{n}, \quad j = 1, \ldots, n, \ i = 1, \ldots, k$$

If D is the number of distinct values among $X_1, \ldots, X_k$ show that

$$E[D] = n\left[1 - \left(\frac{n-1}{n}\right)^k\right]$$
$$\approx k - \frac{k^2}{2n} \quad \text{when } \frac{k^2}{n} \text{ is small}$$

13. *The Discrete Rejection Method:* Suppose we want to simulate X having probability mass function $P\{X = i\} = P_i, i = 1, \ldots, n$ and suppose we can easily simulate from the probability mass function $Q_i, \sum_i Q_i = 1, Q_i \geq 0$. Let C be such that $P_i \leq C Q_i, i = 1, \ldots, n$. Show that the following algorithm generates the desired random variable:

Step 1: Generate Y having mass function Q and U an independent random number.

Step 2: If $U \leq P_Y / C Q_Y$, set $X = Y$. Otherwise return to step 1.

14. *The Discrete Hazard Rate Method:* Let X denote a nonnegative integer valued random variable. The function $\lambda(n) = P\{X = n \mid X \geq n\}, n \geq 0$, is called the *discrete hazard rate function*.

(a) Show that $P\{X = n\} = \lambda(n) \prod_{i=0}^{n-1} (1 - \lambda(i))$.

(b) Show that we can simulate X by generating random numbers $U_1, U_2, \ldots$ stopping at

$$X = \min\{n: U_n \leq \lambda(n)\}$$

(c) Apply this method to simulating a geometric random variable. Explain, intuitively, why it works.

(d) Suppose that $\lambda(n) \leq p < 1$ for all n. Consider the following algorithm for simulating X and explain why it works: Simulate $X_i, U_i, i \geq 1$ where X_i is geometric with mean $1/p$ and U_i is a random number. Set $S_k = X_1 + \cdots + X_k$ and let

$$X = \min\{S_k: U_k \leq \lambda(S_k)/p\}$$

15. Suppose you have just simulated a normal random variable X with mean μ and variance σ^2. Give an easy way to generate a second normal variable with the same mean and variance that is negatively correlated with X.

***16.** Suppose n balls having weights $w_1, w_2, \ldots, w_n$ are in an urn. These balls are sequentially removed in the following manner: At each selection, a given ball in the urn is chosen with a probability equal to its weight divided by the sum of the weights of the other balls that are still in the urn. Let $I_1, I_2, \ldots, I_n$ denote the order in which the balls are removed—thus $I_1, \ldots, I_n$ is a random permutation with weights.

(a) Give a method for simulating $I_1, \ldots, I_n$.

(b) Let X_i be independent exponentials with rates $w_i, i = 1, \ldots, n$. Explain how X_i can be utilized to simulate $I_1, \ldots, I_n$.

17. *Order Statistics:* Let $X_1, \ldots, X_n$ be i.i.d. from a continuous distribution F, and let $X_{(i)}$ denote the ith smallest of $X_1, \ldots, X_n, i = 1, \ldots, n$. Suppose we want to simulate $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. One approach is to simulate n values from F, and then order these values. However, this ordering, or *sorting*, can be time consuming when n is large.

(a) Suppose that $\lambda(t)$, the hazard rate function of F, is bounded. Show how the hazard rate method can be applied to generate the n variables in such a manner that no sorting is necessary.

Suppose now that F^{-1} is easily computed.

(b) Argue that $X_{(1)}, \ldots, X_{(n)}$ can be generated by simulating $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$—the ordered values of n independent random numbers—and then setting $X_{(i)} = F^{-1}(U_{(i)})$. Explain why this means that $X_{(i)}$ can be generated from $F^{-1}(\beta_i)$ where β_i is beta with parameters $i, n + i + 1$.

(c) Argue that $U_{(1)}, \ldots, U_{(n)}$ can be generated, without any need for sorting, by simulating i.i.d. exponentials $Y_1, \ldots, Y_{n+1}$ and then setting

$$U_{(i)} = \frac{Y_1 + \cdots + Y_i}{Y_1 + \cdots + Y_{n+1}}, \quad i = 1, \ldots, n$$

Hint: Given the time of the $(n + 1)$st event of a Poisson process, what can be said about the set of times of the first n events?

(d) Show that if $U_{(n)} = y$ then $U_{(1)}, \ldots, U_{(n-1)}$ has the same joint distribution as the order statistics of a set of $n - 1$ uniform $(0, y)$ random variables.

(e) Use part (d) to show that $U_{(1)}, \ldots, U_{(n)}$ can be generated as follows:

Step 1: Generate random numbers $U_1, \ldots, U_n$.

Step 2: Set

$$U_{(n)} = U_1^{1/n},$$
$$U_{(j-1)} = U_{(j)}(U_{n-j+2})^{1/(j-1)}, \quad j = 2, \ldots, n-1$$

18. Let $X_1, \ldots, X_n$ be independent exponential random variables each having rate 1. Set

$$W_1 = X_1/n,$$
$$W_i = W_{i-1} + \frac{X_i}{n-i+1}, \quad i = 2, \ldots, n$$

Explain why $W_1, \ldots, W_n$ has the same joint distribution as the order statistics of a sample of n exponentials each having rate 1.

19. Suppose we want to simulate a large number n of independent exponentials with rate 1—call them $X_1, X_2, \ldots, X_n$. If we were to employ the inverse

transform technique we would require one logarithmic computation for each exponential generated. One way to avoid this is to first simulate S_n, a gamma random variable with parameters $(n, 1)$ (say, by the method of Section 11.3.3). Now interpret S_n as the time of the nth event of a Poisson process with rate 1 and use the result that given S_n the set of the first $n-1$ event times is distributed as the set of $n-1$ independent uniform $(0, S_n)$ random variables. Based on this, explain why the following algorithm simulates n independent exponentials:

Step 1: Generate S_n, a gamma random variable with parameters $(n, 1)$.
Step 2: Generate $n-1$ random numbers $U_1, U_2, \ldots, U_{n-1}$.
Step 3: Order the $U_i, i = 1, \ldots, n-1$ to obtain $U_{(1)} < U_{(2)} < \cdots < U_{(n-1)}$.
Step 4: Let $U_{(0)} = 0, U_{(n)} = 1$, and set $X_i = S_n(U_{(i)} - U_{(i-1)}), i = 1, \ldots, n$.

When the ordering (step 3) is performed according to the algorithm described in Section 11.5, the preceding is an efficient method for simulating n exponentials when all n are simultaneously required. If memory space is limited, however, and the exponentials can be employed sequentially, discarding each exponential from memory once it has been used, then the preceding may not be appropriate.

20. Consider the following procedure for randomly choosing a subset of size k from the numbers $1, 2, \ldots, n$: Fix p and generate the first n time units of a renewal process whose interarrival distribution is geometric with mean $1/p$—that is, $P\{\text{interarrival time} = k\} = p(1-p)^{k-1}, k = 1, 2, \ldots$. Suppose events occur at times $i_1 < i_2 < \cdots < i_m \leq n$. If $m = k$, stop; $i_1, \ldots, i_m$ is the desired set. If $m > k$, then randomly choose (by some method) a subset of size k from $i_1, \ldots, i_m$ and then stop. If $m < k$, take $i_1, \ldots, i_m$ as part of the subset of size k and then select (by some method) a random subset of size $k - m$ from the set $\{1, 2, \ldots, n\} - \{i_1, \ldots, i_m\}$. Explain why this algorithm works. As $E[N(n)] = np$ a reasonable choice of p is to take $p \approx k/n$. (This approach is due to Dieter.)

21. Consider the following algorithm for generating a random permutation of the elements $1, 2, \ldots, n$. In this algorithm, $P(i)$ can be interpreted as the element in position i.

Step 1: Set $k = 1$.
Step 2: Set $P(1) = 1$.
Step 3: If $k = n$, stop. Otherwise, let $k = k + 1$.
Step 4: Generate a random number U, and let

$$P(k) = P([kU] + 1),$$
$$P([kU] + 1) = k.$$

Go to step 3.

(a) Explain in words what the algorithm is doing.
(b) Show that at iteration k—that is, when the value of $P(k)$ is initially set—that $P(1), P(2), \ldots, P(k)$ is a random permutation of $1, 2, \ldots, k$.

Hint: Use induction and argue that

$$\begin{aligned} P_k\{i_1, i_2, \ldots, i_{j-1}, k, i_j, \ldots, i_{k-2}, i\} &= P_{k-1}\{i_1, i_2, \ldots, i_{j-1}, i, i_j, \ldots, i_{k-2}\}\frac{1}{k} \\ &= \frac{1}{k!} \quad \text{by the induction hypothesis} \end{aligned}$$

The preceding algorithm can be used even if n is not initially known.

22. Verify that if we use the hazard rate approach to simulate the event times of a nonhomogeneous Poisson process whose intensity function $\lambda(t)$ is such that $\lambda(t) \le \lambda$, then we end up with the approach given in method 1 of Section 11.5.

***23.** For a nonhomogeneous Poisson process with intensity function $\lambda(t), t \ge 0$, where $\int_0^\infty \lambda(t)\,dt = \infty$, let $X_1, X_2, \ldots$ denote the sequence of times at which events occur.

(a) Show that $\int_0^{X_1} \lambda(t)\,dt$ is exponential with rate 1.

(b) Show that $\int_{X_{i-1}}^{X_i} \lambda(t)\,dt, i \ge 1$, are independent exponentials with rate 1, where $X_0 = 0$.

In words, independent of the past, the additional amount of hazard that must be experienced until an event occurs is exponential with rate 1.

24. Give an efficient method for simulating a nonhomogeneous Poisson process with intensity function

$$\lambda(t) = b + \frac{1}{t+a}, \quad t \ge 0$$

25. Let (X, Y) be uniformly distributed in a circle of radius r about the origin. That is, their joint density is given by

$$f(x, y) = \frac{1}{\pi r^2}, \quad 0 \le x^2 + y^2 \le r^2$$

Let $R = \sqrt{X^2 + Y^2}$ and $\theta = \arctan Y/X$ denote their polar coordinates. Show that R and θ are independent with θ being uniform on $(0, 2\pi)$ and $P\{R < a\} = a^2/r^2, 0 < a < r$.

26. Let R denote a region in the two-dimensional plane. Show that for a two-dimensional Poisson process, given that there are n points located in R, the points are independently and uniformly distributed in R—that is, their density is $f(x, y) = c, (x, y) \in R$ where c is the inverse of the area of R.

27. Let $X_1, \ldots, X_n$ be independent random variables with $E[X_i] = \theta, \text{Var}(X_i) = \sigma_i^2\ i = 1, \ldots, n$, and consider estimates of θ of the form $\sum_{i=1}^n \lambda_i X_i$ where $\sum_{i=1}^n \lambda_i = 1$. Show that $\text{Var}\left(\sum_{i=1}^n \lambda_i X_i\right)$ is minimized when

$$\lambda_i = (1/\sigma_i^2) \Big/ \left(\sum_{j=1}^n 1/\sigma_j^2\right), \quad i = 1, \ldots, n.$$

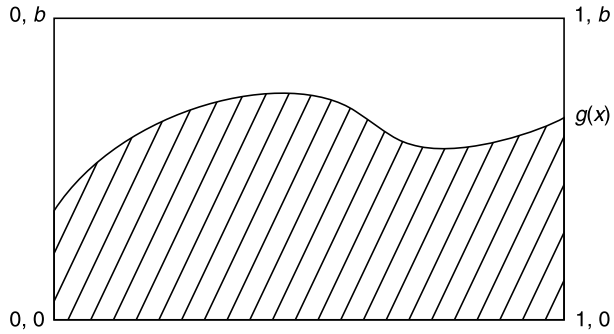

Figure 11.7

Possible Hint: If you cannot do this for general n, try it first when $n = 2$.

The following two problems are concerned with the estimation of $\int_0^1 g(x)\,dx = E[g(U)]$ where U is uniform $(0, 1)$.

28. *The Hit–Miss Method:* Suppose g is bounded in $[0, 1]$—for instance, suppose $0 \le g(x) \le b$ for $x \in [0, 1]$. Let U_1, U_2 be independent random numbers and set $X = U_1, Y = bU_2$—so the point (X, Y) is uniformly distributed in a rectangle of length 1 and height b. Now set

$$I = \begin{cases} 1, & \text{if } Y < g(X) \\ 0, & \text{otherwise} \end{cases}$$

That is, accept (X, Y) if it falls in the shaded area of Fig. 11.7.

(a) Show that $E[bI] = \int_0^1 g(x)\,dx$.

(b) Show that $\text{Var}(bI) \ge \text{Var}(g(U))$, and so hit–miss has larger variance than simply computing g of a random number.

29. *Stratified Sampling:* Let $U_1, \ldots, U_n$ be independent random numbers and set $\bar{U}_i = (U_i + i - 1)/n, i = 1, \ldots, n$. Hence, $\bar{U}_i, i \ge 1$, is uniform on $((i-1)/n, i/n)$. $\sum_{i=1}^n g(\bar{U}_i)/n$ is called the stratified sampling estimator of $\int_0^1 g(x)\,dx$.

(a) Show that $E[\sum_{i=1}^n g(\bar{U}_i)/n] = \int_0^1 g(x)\,dx$.

(b) Show that $\text{Var}[\sum_{i=1}^n g(\bar{U}_i)/n] \le \text{Var}[\sum_{i=1}^n g(U_i)/n]$.

Hint: Let U be uniform $(0, 1)$ and define N by $N = i$ if $(i-1)/n < U < i/n, i = 1, \ldots, n$. Now use the conditional variance formula to obtain

$$\begin{aligned} \text{Var}(g(U)) &= E[\text{Var}(g(U)|N)] + \text{Var}(E[g(U)|N]) \\ &\ge E[\text{Var}(g(U)|N)] \\ &= \sum_{i=1}^n \frac{\text{Var}(g(U)|N = i)}{n} = \sum_{i=1}^n \frac{\text{Var}[g(\bar{U}_i)]}{n} \end{aligned}$$

30. If f is the density function of a normal random variable with mean μ and variance σ^2, show that the tilted density f_t is the density of a normal random variable with mean $\mu + \sigma^2 t$ and variance σ^2.

31. Consider a queueing system in which each service time, independent of the past, has mean μ. Let W_n and D_n denote, respectively, the amounts of time customer n spends in the system and in queue. Hence, $D_n = W_n - S_n$ where S_n is the service time of customer n. Therefore,

$$E[D_n] = E[W_n] - \mu$$

If we use simulation to estimate $E[D_n]$, should we

(a) use the simulated data to determine D_n, which is then used as an estimate of $E[D_n]$; or

(b) use the simulated data to determine W_n and then use this quantity minus μ as an estimate of $E[D_n]$?

Repeat for when we want to estimate $E[W_n]$.

***32.** Show that if X and Y have the same distribution then

$$\text{Var}((X + Y)/2) \leq \text{Var}(X)$$

Hence, conclude that the use of antithetic variables can never increase variance (though it need not be as efficient as generating an independent set of random numbers).

33. If $0 \leq X \leq a$, show that

(a) $E[X^2] \leq aE[X]$,

(b) $\text{Var}(X) \leq E[X](a - E[X])$,

(c) $\text{Var}(X) \leq a^2/4$.

34. Suppose in Example 11.19 that no new customers are allowed in the system after time t_0. Give an efficient simulation estimator of the expected additional time after t_0 until the system becomes empty.

35. Suppose we are able to simulate independent random variables X and Y. If we simulate $2k$ independent random variables $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$, where the X_i have the same distribution as does X, and the Y_j have the same distribution as does Y, how would you use them to estimate $P(X < Y)$?

36. If U_1, U_2, U_3 are independent uniform $(0, 1)$ random variables, find $P\left(\prod_{i=1}^{3} U_i > 0.1\right)$.

Hint: Relate the desired probability to one about a Poisson process.

References

[1] J. Banks, J. Carson, Discrete Event System Simulation, Prentice Hall, Englewood Cliffs, New Jersey, 1984.

[2] G. Fishman, Principles of Discrete Event Simulation, John Wiley, New York, 1978.

[3] D. Knuth, Semi Numerical Algorithms, The Art of Computer Programming, vol. 2, Second Edition, Addison-Wesley, Reading, Massachusetts, 1981.
[4] A. Law, W. Kelton, Simulation Modelling and Analysis, Second Edition, McGraw-Hill, New York, 1992.
[5] J. Propp, D. Wilson, Coupling From The Past: A User's Guide, in: Workshop on Microsurveys in Discrete Probability, Princeton, NJ, 1997.
[6] S. Ross, Simulation, Fifth Edition, Academic Press, San Diego, 2013.
[7] R. Rubenstein, Simulation and the Monte Carlo Method, John Wiley, New York, 1981.

Coupling

12

12.1 A Brief Introduction

In this chapter we will introduce the concept of *coupling* and show how it can be effectively employed in showing stochastic order relations between random variables and between processes, in bounding the distance between distributions, in bounding the error obtained when utilizing the Poisson paradigm, in determining stochastic optimization results, and in other areas of applied probability. Occasionally, for convenience we will repeat arguments given earlier in the text.

For an event A, we will use the notation $I\{A\}$ to stand for the indicator random variable for A, defined to equal 1 when A occurs and to equal 0 otherwise.

12.2 Coupling and Stochastic Order Relations

We say that (X', Y') is a *coupling* of the pair of random variables (X, Y) if X' has the same distribution as X and Y' has the same distribution as Y. That is, if X has distribution F and Y has distribution G, then any pair of random variables X', Y' having respective distributions F and G is a coupling of X, Y.

Couplings are useful in many areas of probability. We start by indicating their use in establishing stochastic order relations. To begin, we define the concept of one random variable being stochastically larger than another.

Definition. Say that the random variable X is stochastically larger than the random variable Y, written as $X \geq_{st} Y$, if

$$P(X > x) \geq P(Y > x) \qquad \text{for all } x$$

Because the preceding inequality is equivalent to $P(X \leq x) \leq P(Y \leq x)$, it follows that if X and Y have respective distribution functions F and G, then $X \geq_{st} Y$ if $F(x) \leq G(x)$ for all x.

One way to establish that $X \geq_{st} Y$ is to find a coupling (X', Y') of (X, Y) such that $X' \geq Y'$ with probability 1. For suppose such a coupling existed. Then, because $Y' > x \Rightarrow X' > x$ we have that

$$P(Y > x) = P(Y' > x) \leq P(X' > x) = P(X > x)$$

showing that $X \geq_{st} Y$.

Example 12.1. Suppose that X is Poisson with mean λ and that Y is Poisson with mean μ, where $\lambda > \mu$. We could attempt to show that $X \geq_{st} Y$ by directly showing

Introduction to Probability Models. https://doi.org/10.1016/B978-0-12-814346-9.00017-2

that, for all k,

$$P(X \le k) = \sum_{i=0}^{k} e^{-\lambda}\lambda^i/i! \le \sum_{i=0}^{k} e^{-\mu}\mu^i/i! = P(Y \le k)$$

That is, we could try to show that $\sum_{i=0}^{k} e^{-\lambda}\lambda^i/i!$ is, for all k, a decreasing function of λ. However, another way is to let Z be a Poisson random variable with mean $\lambda - \mu$ that is independent of Y. Because the sum of independent Poisson random variables is also Poisson, we have that $Z + Y$ is Poisson with mean λ. Hence, $(Z + Y, Y)$ is a coupling of (X, Y). Because $Z + Y \ge Y$, we can conclude that a Poisson random variable increases stochastically in its mean. ■

Example 12.2. We now show that a binomial random variable $X(n, p)$ with parameters (n, p) stochastically increases in both n and p. That is, for all $k \ge 0$

$$P(X(n, p) \ge k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1 - p)^{n-i}$$

is an increasing function of both n and p. To use coupling to show that it is stochastically increasing in n, let $X_1, X_2, \ldots$ be a sequence of independent Bernoulli random variables with $P(X_i = 1) = p = 1 - P(X_i = 0)$. Then

$$X_1 + \ldots X_n + X_{n+1} \ge X_1 + \ldots X_n$$

which proves the result since $X_1 + \ldots X_r$ is binomial with parameters (r, p). To show that $X(n, p)$ is stochastically increasing in p, we will argue that $X(n, p) \ge_{st} X(n, \alpha p)$ for $0 < \alpha < 1$. To do so, let $U_1, \ldots, U_n$ be independent uniform $(0, 1)$ random variables, and for $i = 1, \ldots, n$, let $X_i = I\{U_i \le p\}$, and let $Y_i = I\{U_i \le \alpha p\}$. That is X_i is 1 if $U_i \le p$ and is 0 otherwise, whereas Y_i is 1 if $U_i \le \alpha p$ and is 0 otherwise. Because $\alpha p < p$, we see that $X_i \ge Y_i$, and thus

$$X_1 + \ldots + X_n \ge Y_1 + \ldots + Y_n$$

As $X_1 + \ldots + X_n$ is binomial with parameters (n, p) whereas $Y_1 + \ldots + Y_n$ is binomial with parameters $(n, \alpha p)$ the result follows. ■

It turns out that if $X \ge_{st} Y$ then there is always a coupling (X', Y') such that $X' \ge Y'$.

Proposition 12.1. *$X \ge_{st} Y$ if and only if there is a coupling (X', Y') of (X, Y) such that $P(X' \ge Y') = 1$.*

Proof. We have already argued that if there is a coupling (X', Y') such that $P(X' \ge Y') = 1$, then $X \ge_{st} Y$. So, now suppose that $X \ge_{st} Y$. We show that this implies that there is a coupling (X', Y') such that $X' \ge Y'$ in the case where X and Y are continuous random variables, with respective distributions F and G. (The proof in the

general case is similar.) Recall that if $h^{-1}(x)$ is the inverse of the function $h(x)$, then $h^{-1}(x) = y$ if $h(y) = x$.

To obtain the desired coupling, let U be uniform over $(0, 1)$. Because F is an increasing function it follows that

$$F^{-1}(U) \le x \Leftrightarrow F(F^{-1}(U)) \le F(x).$$

Hence,

$$\begin{aligned} P(F^{-1}(U) \le x) &= P\left(F(F^{-1}(U)) \le F(x)\right) \\ &= P\left(U \le F(x)\right) \\ &= F(x) \end{aligned}$$

Thus, $F^{-1}(U)$ has distribution F. Similarly, $G^{-1}(U)$ has distribution G. Because $X \ge_{st} Y$ is equivalent to $F(x) \le G(x)$, from which it follows that $F^{-1}(x) \ge G^{-1}(x)$, we obtain an (X, Y) coupling $\left(X' = F^{-1}(U),\ Y' = G^{-1}(U)\right)$ for which $X' \ge Y'$. ■

An equivalent definition of stochastically larger is given by the following proposition.

Proposition 12.2. *$X \ge_{st} Y$ if and only if $E[h(X)] \ge E[h(Y)]$ for all increasing functions h.*

Proof. If $X \ge_{st} Y$ then, by Proposition 12.1, we can couple them so that $X \ge Y$ with probability 1. But as h is increasing this yields that $h(X) \ge h(Y)$ and taking expectations shows that $E[h(X)] \ge E[h(Y)]$. To go the other way, suppose that $E[h(X)] \ge E[h(Y)]$ for any increasing function h. To show that this implies that $P(X > x) \ge P(Y > x)$ for all x, fix x and define the function h by

$$h(y) = I\{y > x\} = \begin{cases} 0, & \text{if } y \le x \\ 1, & \text{if } y > x \end{cases}$$

Because h is an increasing function, it follows that $E[h(X)] \ge E[h(Y)]$, which proves the result because

$$E[h(X)] = P(X > x), \quad E[h(Y)] = P(Y > x).$$ ■

We now define the concept of one random n-vector being stochastically larger than another.

Definition. Say that $\mathbf{X} = (X_1, \ldots, X_n)$ is stochastically larger than $\mathbf{Y} = (Y_1, \ldots, Y_n)$ if

$$h(\mathbf{X}) \ge_{st} h(\mathbf{Y})$$

for all increasing functions h.

Proposition 12.3. *Suppose that $X_1, \ldots, X_n$ are independent, that $Y_1, \ldots, Y_n$ are independent, and that $X_i \geq_{st} Y_i$ for all i. Then $(X_1, \ldots, X_n) \geq_{st} (Y_1, \ldots, Y_n)$.*

Proof. To prove this, for each i couple X_i and Y_i so that $X_i \geq Y_i$. It is easy to see that this can be done in such a way that the vectors $(X_i, Y_i), i = 1, \ldots, n$ are independent. (For instance, do the coupling via the approach used to prove Proposition 12.1, using n independent uniforms for the n couplings.) Now let h be an increasing function. Then, $h(X_1, \ldots, X_n) \geq h(Y_1, \ldots, Y_n)$, which implies that $h(X_1, \ldots, X_n) \geq_{st} h(Y_1, \ldots, Y_n)$. ■

Definition. We say that X is stochastically equal to Y, written as $X =_{st} Y$, if X and Y have the same distribution function.

12.3 Stochastic Ordering of Stochastic Processes

We start with a definition.

Definition. Say that the stochastic process $\{X(t), t \in T\}$ is stochastically larger than the stochastic process $\{Y(t), t \in T\}$ if, for any n and values $t_1, \ldots, t_n \in T$, $(X(t_1), \ldots, X(t_n)) \geq_{st} (Y(t_1), \ldots, Y(t_n))$.

To use coupling to show that $\{X(t), t \in T\}$ is stochastically larger than $\{Y(t), t \in T\}$, we try to find stochastic processes $\{X'(t), t \in T\}$ having the same probability law as $\{X(t), t \in T\}$ and $\{Y'(t), t \in T\}$ having the same probability law as $\{Y(t), t \in T\}$ such that $X'(t) \geq Y'(t)$ for all t.

Our first result gives a sufficient condition for one discrete time Markov chain to be stochastically larger than another. Letting $\mathbf{X} = \{X_n, n \geq 0\}$ and $\mathbf{Y} = \{Y_n, n \geq 0\}$ be discrete time Markov chains with respective transition probabilities $P_{i,j}$ and $Q_{i,j}$, our objective is to determine a sufficient condition on these transition probabilities so that $\{X_n, n \geq 0\} \geq_{st} \{Y_n, n \geq 0\}$ whenever $X_0 \geq_{st} Y_0$. To state our result, let $N_x(i)$ be a random variable having the distribution of the next state from state i of the Markov chain $\mathbf{X}$, and $N_y(i)$ be one having the distribution of the next state from state i of the Markov chain $\mathbf{Y}$. That is,

$$P(N_x(i) = k) = P_{i,k}, \quad P(N_y(i) = k) = Q_{i,k}$$

Proposition 12.4. *If $X_0 \geq_{st} Y_0$ and $N_x(i) \geq_{st} N_y(j)$ for all $i \geq j$, then $\mathbf{X} \geq_{st} \mathbf{Y}$.*

Proof. Assume the conditions of the proposition. We will prove the result by showing how the two Markov chains can be coupled so that $X_n \geq Y_n$ for every n. Because $X_0 \geq_{st} Y_0$ we can couple them so that $X_0 \geq Y_0$. But then, by the conditions of the proposition, we have that

$$X_1 =_{st} N_x(X_0) \geq_{st} N_y(Y_0) =_{st} Y_1$$

Hence, $X_1 \geq_{st} Y_1$ and so they can be coupled so that $X_1 \geq Y_1$. Continuing in this manner shows that there are coupled versions of the two Markov chains such that $X_n \geq Y_n$ for every n, thus proving the result. ■

Corollary 12.5. *If $X_0 \geq_{st} Y_0$, $N_x(i) \geq_{st} N_y(i)$ for all i, and either $N_x(i)$ or $N_y(i)$ is stochastically increasing in i, then $\mathbf{X} \geq_{st} \mathbf{Y}$.*

Proof. By Proposition 12.4 it suffices to prove that $N_x(i) \geq_{st} N_y(j)$ for all $i \geq j$. Now, suppose that $N_x(i) \geq_{st} N_y(i)$ for all i. Then, for $i \geq j$,

$$N_x(i) \uparrow_{st} i \Rightarrow N_x(i) \geq_{st} N_x(j) \geq_{st} N_y(j)$$

whereas

$$N_y(i) \uparrow_{st} i \Rightarrow N_x(i) \geq_{st} N_y(i) \geq_{st} N_y(j)$$

Hence, the conditions of Proposition 12.4 are met, which proves the result. ■

Remark. By letting the transition probabilities of the two chains be identical, the preceding shows that if $N_x(i)$ stochastically increases in i, then the Markov chain $\{X_n, n \geq 0\}$ is stochastically increasing in its initial state.

Recall that a birth and death process is a continuous time Markov chain with integer states in which a transition always either increases or decreases the state by 1. Let $\mathbf{X} = \{X(t), t \geq 0\}$ be such a process. We now show that $\mathbf{X}$ stochastically increases in its initial state.

Proposition 12.6. $\{X(t), t \geq 0\}$ *increases stochastically in its initial state.*

Proof. Let $\mathbf{X} = \{X(t), t \geq 0\}$ and $\mathbf{Y} = \{Y(t), t \geq 0\}$ be independent birth and death processes having the same transition probabilities, but with $X(0) > Y(0)$. Because the birth and death processes are independent and their change points are continuous, it follows that either the processes are equal at some point or that $X(t) > Y(t)$ for all t. Consequently, if we let T be the first time their value is the same, then

$$T = \begin{cases} \infty, & \text{if } X(t) > Y(t) \text{ for all } t \\ \min\{t : X(t) = Y(t)\} & \text{otherwise} \end{cases}$$

Now, define $Z(t)$ by

$$Z(t) = \begin{cases} X(t), & \text{if } t \leq T \\ Y(t), & \text{if } t > T \end{cases}$$

Because the continuations of the $\mathbf{X}$ and the $\mathbf{Y}$ processes beyond time T are identically distributed, it follows that the $\mathbf{Z}$ process, which follows the $\mathbf{X}$ process up to time T and then the $\mathbf{Y}$ process from then on, has the same distribution as does the $\mathbf{X}$ process and, moreover, is never below the $\mathbf{Y}$ process (see Fig. 12.1). Consequently, $(\mathbf{Z}, \mathbf{Y})$ is a coupling of $(\mathbf{X}, \mathbf{Y})$ having $Z(t) \geq Y(t)$ for all t, thus proving that $\{X(t), t \geq 0\} \geq_{st} \{Y(t), t \geq 0\}$. ■

$$Z(t) = \begin{cases} X(t), & \text{if } t \le T \\ Y(t), & \text{if } t > T \end{cases}$$

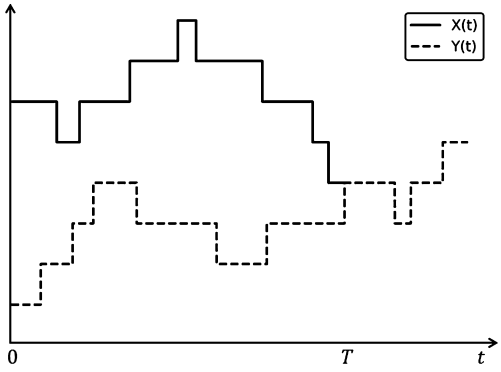

Figure 12.1

We next show that if the birth and death process has only nonnegative states, then $X(0) = 0$ implies that $X(t)$ stochastically increases in t.

Proposition 12.7. *Let $\{X(t), t \ge 0\}$ be a birth and death process with nonnegative states. If $X(0) = 0$, then $X(t)$ stochastically increases in t.*

Proof. To show, for $s > 0$, that $P(X(t+s) > j | X(0) = 0) \ge P(X(t) > j | X(0) = 0)$, condition on $X(s)$ to obtain

$$\begin{aligned}
&P(X(t+s) > j | X(0) = 0) \\
&\quad = \sum_{i=0}^{\infty} P(X(t+s) > j | X(0) = 0,\ X(s) = i)\, P(X(s) = i | X(0) = 0) \\
&\quad = \sum_{i=0}^{\infty} P(X(t+s) > j | X(s) = i)\, P(X(s) = i | X(0) = 0) \\
&\quad = \sum_{i=0}^{\infty} P(X(t) > j | X(0) = i)\, P(X(s) = i | X(0) = 0) \\
&\quad \ge \sum_{i=0}^{\infty} P(X(t) > j | X(0) = 0)\, P(X(s) = i | X(0) = 0) \quad \text{by Proposition 12.6} \\
&\quad = P(X(t) > j | X(0) = 0) \sum_{i=0}^{\infty} P(X(s) = i | X(0) = 0) \\
&\quad = P(X(t) > j | X(0) = 0)
\end{aligned}$$

■

12.4 Maximum Couplings, Total Variation Distance, and the Coupling Identity

When X has distribution function F and Y has distribution function G, we call the pair (X, Y) an F, G couple. With C denoting the set of all F, G couples, say that $(\hat{X}, \hat{Y}) \in C$ is a *maximum F, G couple* if

$$P(\hat{X} = \hat{Y}) = \max_{(X,Y)\in C} P(X = Y)$$

That is, $(\hat{X}, \hat{Y})$ is a maximum F, G couple if among all such couples its components are most likely to be equal.

Proposition 12.8. *A maximum F, G couple always exists. If $(\hat{X}, \hat{Y})$ is a maximum F, G couple, then*

(a) *if F and G are continuous with respective densities $F' = f$ and $G' = g$*

$$P(\hat{X} = \hat{Y}) = \int_x m(x)dx$$

where $m(x) = \min(f(x), g(x))$;

(b) *if F and G are discrete with respective mass functions $\{p_i\}$ and $\{q_i\}$, then*

$$P(\hat{X} = \hat{Y}) = \sum_i m(i)$$

where $m(i) = \min(p_i, q_i)$.

Proof. To prove (a), let $p = \int_{-\infty}^{\infty} m(x)dx$. Also, let

$$A = \{x : f(x) > g(x)\}$$

and note that

$$m(x) = \begin{cases} g(x), & \text{if } x \in A \\ f(x), & \text{if } x \in A^c \end{cases}$$

Now, for any random variables X and Y, with respective distributions F and G

$$\begin{aligned} P(X = Y) &= P(X = Y \in A) + P(X = Y \notin A) \\ &\le P(Y \in A) + P(X \notin A) \\ &= \int_A g(x)dx + \int_{A^c} f(x)dx \\ &= \int_A m(x)dx + \int_{A^c} m(x)dx \\ &= p \end{aligned}$$

Hence, for any F, G coupling, $P(X = Y) \leq p$. To prove that we obtain an equality for the maximum coupling, we exhibit an F, G coupling (X, Y) for which $P(X = Y) = p$. To do so, let V_1, V_2, V_3, U be independent random variables with respective density functions

$$f_{V_1}(x) = \frac{m(x)}{p}$$
$$f_{V_2}(x) = \frac{f(x) - m(x)}{1 - p}$$
$$f_{V_3}(x) = \frac{g(x) - m(x)}{1 - p}$$
$$f_U(x) = 1\,,\ 0 < x < 1$$

Now, define X and Y as follows:

$$U \leq p \quad \Rightarrow \quad X = Y = V_1$$
$$U > p \quad \Rightarrow \quad X = V_2,\ Y = V_3$$

Because $P(V_2 = V_3) = 0$, we see that

$$P(X = Y) = P(U \leq p) = p$$

and so the result will follow if we show that (X, Y) is an F, G couple. To prove this, condition on whether $U \leq p$ to obtain

$$f_X(x) = pf_{V_1}(x) + (1 - p) f_{V_2}(x) = m(x) + f(x) - m(x) = f(x)$$

and, similarly, that

$$f_Y(x) = pf_{V_1}(x) + (1 - p) f_{V_3}(x) = m(x) + g(x) - m(x) = g(x)$$

and the proof in the continuous case is complete. The proof in the discrete case is identical with mass functions replacing densities and sums replacing integrals. ■

Example 12.3. Suppose that X_1 and X_2 are such that

$$X_i = \begin{cases} 1, & \text{with probability } p_i \\ 0, & \text{with probability } 1 - p_i \end{cases}$$

where $p_1 > p_2$. The usual way to couple X_1 and X_2 is to let U be uniform on $(0, 1)$ and set

$$X_i = 1 \Leftrightarrow U < p_i$$

Because $U < p_2 \Rightarrow U < p_1$ it follows that $X_2 = 1 \Rightarrow X_1 = 1$ and thus that $X_1 \geq X_2$. One might wonder if this is the maximal couple. To determine whether this is so,

note that $X = Y$ if either $U < p_2$ or $U > p_1$. Because these two events are mutually exclusive (since $p_1 > p_2$), we see that

$$P(X = Y) = P(U < p_2) + P(U > p_1) = p_2 + 1 - p_1$$

Because,

$$\sum_j \min\left(P(X_1 = j), P(X_2 = j)\right) = \min(1 - p_1, 1 - p_2) + \min(p_1, p_2)$$
$$= 1 - p_1 + p_2$$

it follows by Proposition 12.8 that the preceding is indeed a maximum couple. ■

There is a relationship between how closely two random variables can be coupled and how close they are in distribution. One common measure of distance between the distributions of two random variables X and Y is the *total variation distance*, defined as

$$\rho(X, Y) = \max_B |P(X \in B) - P(Y \in B)|.$$

We next show the link between total variation distance and couplings.

Proposition 12.9 (The Coupling Identity)**.** *If* $(\hat{X}, \hat{Y})$ *is a maximal coupling for* (X, Y)*, then* $\rho(X, Y) = P(\hat{X} \neq \hat{Y})$.

Proof. The result will be proven under the assumption that X, Y are continuous with respective density functions f, g. Let $m(x) = \min(f(x), g(x))$. Now, note that

$$\rho(X, Y) = \max\left(\max_B \left(P(X \in B) - P(Y \in B)\right), \max_B \left(P(Y \in B) - P(X \in B)\right)\right)$$

Let $A = \{x : f(x) > g(x)\}$. Because $P(X \in B) - P(Y \in B)$ is increased when adding to B points in A and is decreased by adding to B points not in A, it follows that

$$\max_B \left(P(X \in B) - P(Y \in B)\right) = P(X \in A) - P(Y \in A)$$

and, similarly, that

$$\begin{aligned}\max_B \left(P(Y \in B) - P(X \in B)\right) &= P(Y \in A^c) - P(X \in A^c) \\ &= 1 - P(Y \in A) - 1 + P(X \in A) \\ &= P(X \in A) - P(Y \in A)\end{aligned}$$

Hence,

$$\begin{aligned}\rho(X, Y) &= P(X \in A) - P(Y \in A) \\ &= 1 - P(X \notin A) - P(Y \in A)\end{aligned}$$

$$
\begin{aligned}
&= 1 - \int_{A^c} f(x)\,dx - \int_A g(x)\,dx \\
&= 1 - \int_{A^c} m(x)\,dx - \int_A m(x)\,dx \\
&= 1 - \int m(x)\,dx \\
&= 1 - P(\hat{X} = \hat{Y})
\end{aligned}
$$

where the final equality used Proposition 12.8. ■

12.5 Applications of the Coupling Identity

The coupling identity can often be used to effectively bound total variation distance. For instance, let $(\hat{X}, \hat{Y})$ be a maximum couple of (X, Y), and let (X', Y') be any other (X, Y) couple. Because the maximum couple has the largest probability of being equal, it thus has the smallest probability of being unequal. Consequently, Proposition 12.9 implies that

$$\rho(X, Y) = P(\hat{X} \neq \hat{Y}) \leq P(X' \neq Y')$$

12.5.1 Applications to Markov Chains

Consider a Markov chain $\{X_n, n \geq 0\}$, with state space S and transition probabilities $P_{i,j}$. Recall that the set of nonnegative values π_j, $j \in S$ is said to be a *stationary probability vector* for the Markov chain if

$$
\begin{aligned}
\pi_j &= \sum_i \pi_i P_{i,j}, \quad j \in S \\
\sum_j \pi_j &= 1
\end{aligned}
$$

Proposition 12.10. *Let π_j, $j \in S$ be a stationary probability vector for the Markov chain. If $P(X_0 = j) = \pi_j$, $j \in S$ then $P(X_n = j) = \pi_j$ for all n and j.*

Proof. The proof is by induction on n. As it is true by assumption when $n = 0$, assume that $P(X_{n-1} = i) = \pi_i$ for all i. Then

$$
\begin{aligned}
P(X_n = j) &= \sum_i P(X_n = j | X_{n-1} = i) P(X_{n-1} = i) \\
&= \sum_i P_{i,j} \pi_i \\
&= \pi_j
\end{aligned}
$$

and the proof is complete. ■

Corollary 12.11. *If π_j, $j \in S$ is a stationary probability vector for the Markov chain $\{X_n, n \geq 0\}$, then for any n*

$$\pi_j = \sum_i \pi_i P_{i,j}^n$$

Proof. Suppose that $P(X_0 = i) = \pi_i$, $i \in S$. Then

$$\begin{aligned}\pi_j &= P(X_n = j) \\ &= \sum_i P(X_n = j | X_0 = i) P(X_0 = i) \\ &= \sum_i P_{i,j}^n \pi_i\end{aligned}$$

■

Definition. A Markov chain is said to be *stationary* when the probability of its initial state is a stationary probability vector.

Proposition 12.12. *If π_j, $j \in S$ is a stationary probability vector for an irreducible Markov chain, then $\pi_j > 0$ for all j.*

Proof. Because $\pi_i \geq 0$ and $\sum_i \pi_i = 1$, it follows that $\pi_i > 0$ for at least one i. So, suppose that $\pi_k > 0$. To show that $\pi_j > 0$ for all j, fix j. Because j is accessible from k, there exists a value n for which $P_{k,j}^n > 0$. The result now follows upon using that

$$\pi_j = \sum_i \pi_i P_{ij}^n \geq \pi_k P_{kj}^n > 0$$

■

Proposition 12.13. *Let $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ be Markov chains with the same transition probabilities $P_{i,j}$. Suppose X_0 has an arbitrary distribution, whereas $P(Y_0 = j) = \pi_j$ where π_j, $j \in S$ is a stationary probability vector for the chain. Suppose also that the chain is both irreducible and aperiodic. Then*

$$\rho(X_n, Y_n) \to 0 \quad as \quad n \to \infty$$

Proof. Let the two chains be independent, and define

$$N = \min\{n : X_n = Y_n\}$$

where $N = \infty$ if $X_n \neq Y_n$ for all n. Now, define

$$Z_n = \begin{cases} X_n, & \text{if } n < N \\ Y_n, & \text{if } n \geq N \end{cases}$$

Because $Z_{N+k} =_{st} X_{N+k}$ for any $k \geq 0$, it follows that $Z_n =_{st} X_n$. Hence, (Z_n, Y_n) is an (X_n, Y_n) couple, and so by the coupling identity

$$\begin{aligned}\rho(X_n, Y_n) &\le P(Z_n \neq Y_n)\\ &= P(N > n)\end{aligned}$$

Thus we must show that $P(N > n) \to 0$, or (by the continuity property of probability as a set function) that $P(N < \infty) = 1$. We now argue that it suffices to prove that $P(N < \infty) = 1$ when $X_n, n \ge 0$ is also stationary; that is, when $P(X_0 = j) = \pi_j$, $j \in S$. To see why, suppose that $P(X_0 = j) = \pi_j$, $j \in S$, then

$$P(N < \infty) = \sum_j P(N < \infty | X_0 = j)\pi_j$$

Hence, if $P(N < \infty | X_0 = i) < 1$ for some i, then as $\pi_i > 0$ the preceding would imply that

$$\begin{aligned}P(N < \infty) &< \pi_i + \sum_{j \neq i} P(N < \infty | X_0 = j)\pi_j\\ &\le \pi_i + \sum_{j \neq i} \pi_j\\ &= 1\end{aligned}$$

Hence, $P(N < \infty | X_0 = i) < 1$ for some i implies that $P(N < \infty) < 1$, showing that $P(N < \infty) = 1$ implies that $P(N < \infty | X_0 = i) = 1$ for all i.

So let us assume that $P(X_0 = j) = \pi_j$, $j \in S$. Because $\{X_n, n \ge 0\}$ and $\{Y_n, n \ge 0\}$ are independent, it follows that $\{(X_n, Y_n), n \ge 0\}$ is also a Markov chain. Because the individual chains are irreducible and aperiodic it can be shown that the bivariate chain $\{(X_n, Y_n), n \ge 0\}$. is also irreducible. By independence,

$$P\left((X_n, Y_n) = (i, j)\right) = P(X_n = i)P(Y_n = j) = \pi_i \pi_j > 0$$

which shows that the chain $\{(X_n, Y_n)\}$ is not transient. (Recall that if a state is transient then the limiting probability of being in that state is 0.) Thus the bivariate chain is recurrent, and so, with probability 1, it will eventually enter state (i, i), which shows that $P(N < \infty) = 1$. ■

Remarks. **(a)** Because

$$\begin{aligned}\rho(X_n, Y_n) &= \max_B |P(X_n \in B) - P(Y_n \in B)|\\ &\ge |P(X_n = j) - P(Y_n = j)|\\ &= \left|P(X_n = j) - \pi_j\right|\end{aligned}$$

the preceding shows that for any distribution on X_0

$$\pi_j = \lim_{n \to \infty} P(X_n = j)$$

(b) It follows from the preceding that if there can be at most one stationary probability vector.

(c) It can be shown that if a Markov chain with transition probabilities $P_{i,j}$ is irreducible and aperiodic, then for any states i and j, $P_{i,j}^n > 0$ for all sufficiently large n. Using this, it is easy to show that the bivariate chain introduced in the proof of Proposition 12.13 is irreducible. ■

The next proposition uses the coupling identity to prove that if $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ are Markov chains with the same transition probabilities then the distance between X_n and Y_n decreases in n.

Proposition 12.14. *If $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ are Markov chains with the same transition probabilities then, for any distributions on their initial states X_0 and Y_0,*

$$\rho(X_n, Y_n) \text{ decreases in } n$$

Proof. Fix n and let $(\hat{X}_n, \hat{Y}_n)$ be the maximum coupling of (X_n, Y_n), and note by the coupling identity that

$$\rho(X_n, Y_n) = P(\hat{X}_n \neq \hat{Y}_n)$$

Let $\hat{X}_n$ and $\hat{Y}_n$ be the states of the two Markov chains at time n, and let the chains evolve independently from that point on. Let

$$N = \min\{k \geq n : X_k = Y_k\}$$

be the first time from n on until the chains are in the same state (and let it be ∞ if they are never equal from n on). Also, for $m \geq n$, define

$$Z_m = \begin{cases} X_m, & m < N \\ Y_m, & m \geq N \end{cases}$$

Noting that Z_m has the same distribution as X_m, it follows that for $m > n$

$$\begin{aligned} \rho(X_m, Y_m) &\leq P(Z_m \neq Y_m) \\ &= P(N > m) \\ &\leq P(N > n) \\ &= P(\hat{X}_n \neq \hat{Y}_n) \\ &= \rho(X_n, Y_n) \end{aligned}$$

■

Our next result uses the coupling identity to bound the Poisson approximation of a sum of independent but not necessarily identically distributed Bernoulli random variables.

Proposition 12.15. *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with*

$$P(X_i = 1) = p_i = 1 - P(X_i = 0), \quad i = 1, \ldots, n$$

and let Y be a Poisson random variable with mean $\sum_{i=1}^{n} p_i$. Then

$$\rho\left(\sum_{i=1}^{n} X_i, Y\right) \le \sum_{i=1}^{n} p_i^2$$

Proof. Let Y_i be Poisson with mean p_i and let $(\hat{X}_i, \hat{Y}_i)$ be a maximum coupling of (X_i, Y_i), $i = 1, \ldots, n$. Let these n maximum couplings $(\hat{X}_i, \hat{Y}_i)$, $i = 1, \ldots, n$ be independent. By Proposition 12.8,

$$\begin{aligned}
P(\hat{X}_i \neq \hat{Y}_i) &= 1 - P(\hat{X}_i = \hat{Y}_i) \\
&= 1 - \sum_j \min\left(P(X_i = j), P(Y_i = j)\right) \\
&= 1 - \sum_{j=0}^{1} \min\left(P(X_i = j), P(Y_i = j)\right) \\
&= 1 - \min(1 - p_i, e^{-p_i}) - \min(p_i, p_i e^{-p_i}) \\
&= p_i - p_i e^{-p_i}
\end{aligned}$$

where the preceding used the inequality $e^{-p_i} \ge 1 - p_i$. Because the sum of independent Poisson random variables is also Poisson, it follows that $\sum_{i=1}^{n} \hat{Y}_i$ is Poisson with mean $\sum_{i=1}^{n} p_i$. Hence, by the coupling identity

$$\rho\left(\sum_{i=1}^{n} X_i, Y\right) \le P\left(\sum_{i=1}^{n} \hat{X}_i \neq \sum_{i=1}^{n} \hat{Y}_i\right)$$

Because $\sum_{i=1}^{n} \hat{X}_i \neq \sum_{i=1}^{n} \hat{Y}_i$ implies that $\hat{X}_i \neq \hat{Y}_i$ for some i, the preceding gives that

$$\begin{aligned}
\rho\left(\sum_{i=1}^{n} X_i, Y\right) &\le P(\hat{X}_i \neq \hat{Y}_i \text{ for some } i) \\
&= P(\cup_{i=1}^{n} \{\hat{X}_i \neq \hat{Y}_i\}) \\
&\le \sum_{i=1}^{n} P(\hat{X}_i \neq \hat{Y}_i) \\
&= \sum_{i=1}^{n} p_i (1 - e^{-p_i}) \\
&\le \sum_{i=1}^{n} p_i^2
\end{aligned}$$

where final inequality again used the inequality $e^{-p_i} \ge 1 - p_i$. ■

Our next two examples deal with shuffling and the one dimensional symmetric random walk.

Example 12.4. In random to top shuffling of a deck of k cards numbered 1 thru k, at each stage one of the k positions is randomly chosen, with the chosen one being equally likely to be any of the k positions, and the card in that position is moved to the top of the deck. (An equivalent description is that at each stage one of the k cards is randomly chosen and that card is moved to the top of the deck.) If we let X_n denote the ordering of the deck after stage n, then $\{X_n, n \geq 0\}$ is a Markov chain with $k!$ states. It is clear by symmetry (or by noting that this Markov chain is doubly stochastic) that in the limit all $k!$ possible orderings are equally likely. We would like to bound the total variation distance between the limiting distribution and the distribution of the state after shuffle n.

So let X_n denote the state at time n when X_0 has an arbitrary distribution and let Y_n denote the state when Y_0 is equally likely to be any of the $k!$ orderings. To determine $\rho(X_n, Y_n)$, couple $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ by choosing at each shuffle the same card (not the same position) to be moved to the top of the deck. Using this coupling, the coupling identity gives that

$$\rho(X_n, Y_n) \leq P(X_n \neq Y_n)$$

Now, once a card is chosen it will be moved to the top of the deck in both chains and, furthermore, it will from then on be in the same position in both chains. Consequently, if we let N denote the number of shuffles until all cards have been chosen at least once, then the two deck orderings will be identical from time N on, showing that $P(X_n \neq Y_n) \leq P(N > n)$. Therefore,

$$\rho(X_n, Y_n) \leq P(N > n)$$

To bound $P(N > n)$, which is the probability that it takes more than n coupons to obtain a complete set in the coupon collectors problem with k equally likely coupon types, let A_i be the event that card i is not selected in the first n shuffles. Then,

$$\begin{aligned} P(N > n) &= P(\cup_{i=1}^{k} A_i) \\ &\leq \sum_{i=1}^{k} P(A_i) \\ &= k(1 - \frac{1}{k})^n \end{aligned}$$

Thus the distribution of X_n converges to the equally likely limiting distribution exponentially fast. ■

Our next example shows that in the one-dimensional symmetric random walk the effect of the initial value goes to 0 as the number of transitions increase.

Example 12.5. The Markov chain whose state space is the set of all integers and whose transition probabilities are $P_{i,i+1} = P_{i,i-1} = 1/2$ is called a symmetric random walk. As shown in Example 4.19 and Exercise 39 of Chapter 4, it is a null recurrent

Markov chain and so does not have stationary probabilities. Let $\{X_n\}$ be a symmetric random walk with $X_0 = 0$ and let $\{Y_n\}$ be one with $Y_0 = 2k$. We show that

$$\rho(X_n, Y_n) \to 0$$

To show the preceding, let $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ be independent. Define

$$N = \min\{n : X_n = Y_n\}$$

and set

$$Z_n = \begin{cases} X_n, & \text{if } n < N \\ Y_n, & \text{if } n \geq N \end{cases}$$

Because $Z_n =_{st} X_n$ we have, by the coupling identity, that

$$\begin{aligned} \rho(X_n, Y_n) &\leq P(Z_n \neq Y_n) \\ &= P(N > n) \end{aligned}$$

Hence, we must show that $\lim_{n\to\infty} P(N > n) = 0$ or, equivalently, that $P(N < \infty) = 1$. To do so, let

$$W_n = X_n - Y_n, \ n \geq 0$$

and note that we need to show that $P(W_n = 0 \text{ for some } n) = 1$. Because

$$W_{n+1} - W_n = \begin{cases} -2, & \text{with probability } 1/4 \\ 0, & \text{with probability } 1/2 \\ 2, & \text{with probability } 1/4 \end{cases}$$

it follows that $\{W_n, n \geq 0\}$ is, if we ignore those transitions that leave it in the same state, itself a symmetric random walk. Hence, $\{W_n, n \geq 0\}$ is recurrent, showing that with probability 1 it will eventually equal 0. ■

12.6 Coupling and Stochastic Optimization

Stochastic optimization problems are typically of two types: either static or dynamic in nature. A static problem results when all decisions are made at a single time, whereas a dynamic problem results when decisions are made sequentially in time. Coupling has important applications in both types of stochastic optimization problems. In this section we briefly indicate the potential uses of coupling, first in some dynamic stochastic optimization problems and then in a static one.

Example 12.6. Suppose one has an asset to sell, and that in each period an offer for that asset is presented, with the values of successive offers being independent and hav-

ing a known distribution function F. Suppose that after being presented with an offer, the decision maker must decide whether to accept it and thus end the problem or reject it and wait for the next offer. Suppose that a cost c per offer is incurred, and that the objective is to maximize the expected net return, where the net return is the accepted price minus c multiplied by the number of offers presented before the item is sold.

Let V_F denote the maximal expected net return when the offer distribution is specified to be F, and suppose that we want to prove that $V_F \geq V_G$ when $F \leq G$. That is, in two problems, in which the offers in the first problem are stochastically larger than in the second, we want to show that the optimal expected return in the first problem is at least as large as in the second. To prove this via a coupling argument, suppose the two problems run concurrently, and let X_n and Y_n denote the offers in period $n, n \geq 1$. Because $X_n \geq_{st} Y_n$ for each n, we can couple the two sequences $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ so that the first sequence is that of independent and identical random variables having distribution F and the second is that of independent and identical random variables having distribution G and, in addition, $X_n \geq Y_n$ for every n. Suppose that the optimal policy is employed in Problem 2 (where the offers are the Y values). Let the decision maker in the Problem 1 only accept an offer at a time when an offer is accepted in Problem 2. Thus, if offer Y_N is accepted in Problem 2 then offer X_N will be accepted in Problem 1. Call this policy π. Because $X_N \geq Y_N$, it follows that the net return obtained in Problem 1 when using policy π is at least as large as the net return obtained in Problem 2 when using the optimal policy for Problem 2. Taking expectations shows that the expected net return in Problem 1 when π is employed is at least as large as the optimal expected net return in Problem 2. Because the optimal expected net return in Problem 1 is at least as large as the expected return when the policy π is used, we can conclude that $V_F \geq V_G$. ■

Example 12.7. Consider an asset selling problem where one initially has n items to sell. Suppose that in each period an offer vector $(Y_1, \ldots, Y_n)$ is presented, with the interpretation being that the maker of the offer is bidding for all items and is willing to pay Y_i for item i. Upon receiving such an offer vector, the decision maker can either reject the offer or can elect to sell any subset S of as yet unsold items for the amount $\sum_{i \in S} Y_i$. Suppose that all offer vectors are independent with a known joint distribution function F, that a cost c is incurred each period until all items are sold, and that the objective is to maximize the expected net return, defined as the sum of the selling prices for each of the n items minus c multiplied by the number of periods until all items are sold.

An intuitive result for this model is that if the optimal policy would call for item i to be sold when the set of unsold items is $S, i \in S$, and the offer vector is $(y_1, \ldots, y_n)$ then it would also call for i to be sold if when the set of unsold items is $S, i \in S$, and the offer vector is $(w_1, \ldots, w_n)$ when $w_j \geq y_j, j = 1, \ldots, n$. That is, if it would be optimal to sell item i then it would be optimal to sell it if everything else remained the same and the offer vector was even larger. However, although this is quite intuitive it is not so immediate to prove. One method of proving this result first establishes the inequality

$$V(S \cup T) + V(S \cap T) \geq V(S) + V(T) \tag{12.1}$$

where $V(U)$ is the maximal expected net additional return when the set of unsold items is U. To give a coupling proof of the inequality (12.1), let us change the problem interpretation a bit by imagining that there are n types of items and that an offer vector $(y_1, \ldots, y_n)$ means that the bidder is willing to buy any amount of each of the item types at a price of y_i per type i item, $i = 1, \ldots, n$. Consider two sellers having the same $|S| + |T|$ items for sale. Namely, two of each of item types in $S \cap T$, one of each of the types in $S \cap T^c$, and one of each of the types in $S^c \cap T$. Suppose that both of these sellers are required to divide their items into two groups, with each of their groups being sold in a separate room, and with each room costing the seller c per offer until all items in that room are sold. Suppose that seller one divides his items so that one group consists of one item of each of the types in S, and the second consists of one item of each of the types in T. On the other hand, suppose that seller two divides her items so that one group consists of one item of each of the types in $S \cup T$, and the other of one item of each of the types in $S \cap T$. Couple the offer vectors for the two sellers to be identical. Suppose that seller one uses the optimal policy based on his division. If so, then that seller's expected net return from the items in the group consisting of the item types in S is $V(S)$, and his expected net return from the items in the other group is $V(T)$. Hence, using the optimal policy yields seller one an expected net return $V(S) + V(T)$. Suppose that seller two always sells exactly the same items that seller one does, but when there is a choice of groups from which to sell (that is, for instance, when both of seller one's groups contain a type i item and seller one makes the decision to sell only one of them) then seller two sells the item of that type that resides in the group that initially consisted of one of each of the types in $S \cap T$. Using this policy, it is easy to see that the amounts received by the two sellers for their $|S| + |T|$ items are identical, and that seller two will empty one of her rooms at a time that is either the same as or earlier than the time at which seller one empties one of his rooms. Thus, the net return for seller two is at least as large as that for seller one. Taking expectations shows that there is a policy for seller two whose expected net return is at least $V(S) + V(T)$. Because the expected return for seller two when using this policy cannot be higher than the maximal expected return for seller two, which is $V(S \cup T) + V(S \cap T)$, the inequality (12.1) is established. ■

Example 12.8. There are n initially empty boxes, labeled $1, \ldots, n$. At each stage a ball appears. Attached to each ball is a vector of binary values, say $(x_1, \ldots, x_n)$ with the interpretation that the ball is eligible to be put in box i if $x_i = 1$ and is ineligible if $x_i = 0$. If there are no empty boxes for which the ball is eligible, then that ball is discarded; if there are empty boxes for which the ball is eligible, then a decision must be made as to which box the ball is put. The problem continues until all boxes are nonempty. Let N denote the number of stages needed until there are no nonempty boxes. Assuming that each new ball is independently eligible for box i with probability $p_i, i = 1, \ldots, n$, we are looking for a policy that minimizes $E[N]$.

Without loss of generality, let us suppose that the boxes are numbered so that $p_1 \leq p_2 \leq \ldots \leq p_n$. Because it is hardest to fill a box i having a small eligibility probability p_i, it is intuitive that the optimal policy is the one that always puts a ball into the smallest indexed empty box for which it is eligible. To prove this result, consider any

policy that does not always make the preceding choice, and consider a situation where it does something different. That is, suppose that the policy, call it π, would put a ball into box j when it could have been put in box i and $i < j$. We compare what happens in this case, called scenario one, with what transpires in a second scenario in which the ball is put in box i. In comparing these scenarios, we couple all following eligibility vectors and use a policy in scenario two so that there are never more nonempty boxes in scenario two than in scenario one. Because we must be careful that there is not a resulting eligibility vector that allows one to put a ball into urn i in scenario one but does not allow for it to be put in urn j in scenario two, we cannot couple the eligibility vectors in the two scenarios to be identical. Instead we do the following: We let $U_1, \ldots, U_n$ be independent uniform $(0, 1)$ random variables. The eligibility vector $(X_1^{(1)}, \ldots, X_n^{(1)})$ in scenario one is then defined by

$$X_k^{(1)} = 1 \Leftrightarrow U_k \leq p_k,\ k = 1, \ldots, n$$

whereas the eligibility vector $(X_1^{(2)}, \ldots, X_n^{(2)})$ in scenario two is

$$X_k^{(2)} = 1 \Leftrightarrow U_k \leq p_k,\ k \neq i, j$$
$$X_i^{(2)} = 1 \Leftrightarrow U_j \leq p_i$$
$$X_j^{(2)} = 1 \Leftrightarrow U_i \leq p_j$$

Because $p_i \leq p_j$ it follows from the preceding that $X_i^{(1)} = 1 \Rightarrow X_j^{(2)} = 1$. That is, if the ball is eligible for box i in scenario one then it is also eligible for box j in scenario two. Now we couple the decisions made in scenario two with those in scenario one in the following manner: if the ball is put into box i in scenario one then we put it in box j in scenario two; otherwise whatever box π puts the ball into in scenario one is the box we put the ball into in scenario two. It is easy to see that all boxes are filled in scenario two at least as quickly as in scenario one, showing that we need not consider any policy that does not always put balls in the lowest indexed empty and eligible box, which proves that $E[N]$ is minimized by one that always put a ball in the empty and eligible box whose eligibility probability is lowest. In fact, the preceding argument can be used to show that not only does this policy minimize $E[N]$ but it also stochastically minimizes N, in that it maximizes $P(N < k)$ for all k. ■

Our next example deals with a static optimization problem.

Example 12.9. Consider the coupon collector's problem where each new coupon is independently any of n types, with p_i being the probability that it is type i, $\sum_{i=1}^{n} p_i = 1$. We continue to collect coupons until we have at least one of each type. Letting $N(p_1, \ldots, p_n)$ equal the number of coupons needed when the probability vector is $(p_1, \ldots, p_n)$, we claim that $N(p_1, \ldots, p_n)$ is stochastically minimized when $p_i = 1/n, i = 1, \ldots, n$.

To show the preceding, first suppose that $n = 2$, and let $p_1 = p$, $p_2 = 1 - p$. Because $N > m$ if the first m coupons are all of the same type, we have that

$$P(N > m) = p^m + (1 - p)^m$$

Differentiating and setting equal to 0 gives that

$$mp^{m-1} = m(1-p)^{m-1}$$

or

$$\left(\frac{p}{1-p}\right)^{m-1} = 1$$

which shows that the minimizing value occurs when $p = 1 - p$. Thus, the result holds when $n = 2$. Now consider $N(p_1, \ldots, p_n)$ where at least two of the p_i are not equal, say that $p_1 \neq p_2$. We will show that for any $m \geq n$

$$P(N(p_1, p_2, p_3, \ldots, p_n) > m) \geq P(N(p_a, p_a, p_3, \ldots, p_n) > m)$$

where $p_a = \frac{p_1+p_2}{2}$. To show the preceding, let N_{p_1,p_2} be the number of coupons of type either 1 or 2 that need be collected until there have been at least one of each of these two types, when the probability of these types is p_1, p_2; and let N_{p_a,p_a} be the number of coupons of type 1 or 2 that need be collected until there have been at least one of each of these two types, when the probabilities of these types are p_a, p_a. Also, let N' be the number of types $3, \ldots, n$ that need to be collected to obtain at least one of each of the types $3, \ldots, n$ when the probabilities of types $3, \ldots, n$ are $p_3, \ldots, p_n$. Because N_{p_1,p_2} is the number of coupons needed for a complete set when there are two types of coupons, with coupon type probabilities $\frac{p_i}{p_1+p_2}, i = 1, 2$; and N_{p_a,p_a} is the number needed when the two types have probabilities $1/2$, $1/2$, it follows from the result when $n = 2$ that $N_{p_1,p_2} \geq_{st} N_{p_a,p_a}$. So let us couple N_{p_1,p_2} and N_{p_a,p_a} so that $N_{p_1,p_2} \geq N_{p_a,p_a}$. Letting N' be independent of N_{p_1,p_2} and N_{p_a,p_a}, it follows that $N(p_1, p_2, \ldots, p_n)$ has the distribution of the number of coupons needed until there have been at least N_{p_1,p_2} that are either of types 1 or 2, and at least N' that are one of the types $3, \ldots, n$; whereas $N(p_a, p_a, \ldots, p_n)$ is the number of coupons needed until there have been at least N_{p_a,p_a} that are either of types 1 or 2 and at least N' that are one of the types $3, \ldots, n$. Because $N_{p_1,p_2} \geq N_{p_a,p_a}$, this yields a coupling for which $N(p_1, p_2, \ldots, p_n) \geq N(p_a, p_a, \ldots, p_n)$, which shows that $N(p_1, p_2, p_3, \ldots, p_n) \geq_{st} N(p_a, p_a, p_3, \ldots, p_n)$. Taking the limit of continual repetitions of this argument proves that $N(p_1, p_2, \ldots, p_n) \geq_{st} N(1/n, \ldots, 1/n)$. ■

12.7 Chen–Stein Poisson Approximation Bounds

Let $X_1, X_2, \ldots, X_n$ be Bernoulli random variables with respective means $\lambda_1, \lambda_2, \ldots, \lambda_n$. That is,

$$P(X_i = 1) = \lambda_i = 1 - P(X_i = 0), \qquad i = 1, \ldots, n$$

Set $W = \sum_{i=1}^{n} X_i$, and let $\lambda = E[W] = \sum_{i=1}^{n} \lambda_i$. For Z being a Poisson random variable with mean λ, the Chen–Stein method often enables us to bound

$$\rho(W, Z) = \max_A \{|P(W \in A) - \sum_{i \in A} e^{-\lambda} \lambda^i / i!|\}$$

the total variation distance between W and Z. The germ of the method is that for any function f

$$\begin{aligned} E[Zf(Z)] &= \sum_{i=0}^{\infty} i f(i) e^{-\lambda} \lambda^i / i! \\ &= \sum_{i=1}^{\infty} f(i) e^{-\lambda} \lambda^i / (i-1)! \\ &= \lambda \sum_{j=0}^{\infty} f(j+1) e^{-\lambda} \lambda^j / j! \qquad \text{(by letting } j = i-1) \\ &= \lambda E[f(Z+1)] \end{aligned}$$

Motivated by the preceding, for any given set A, we will define a function $f_A(j)$, $j \geq 0$, such that

$$E[\lambda f_A(W+1) - W f_A(W)] = P(W \in A) - \sum_{i \in A} e^{-\lambda} \lambda^i / i! \tag{12.2}$$

To do so, let $f_A(0) = 0$, and then recursively define f_A by letting

$$\lambda f_A(j+1) = j f_A(j) + I\{j \in A\} - \sum_{i \in A} e^{-\lambda} \lambda^i / i!, \quad j \geq 0$$

where $I\{j \in A\}$ is the indicator function of the event that $j \in A$. As the above holds for all $j \geq 0$, it follows that

$$\lambda f_A(W+1) - W f_A(W) = I\{W \in A\} - \sum_{i \in A} e^{-\lambda} \lambda^i / i!$$

Taking expectations of both sides of the preceding yields (12.2).

We state the following technical Lemma without giving a proof.

Lemma 12.16. *For any set A*

$$|f_A(j) - f_A(i)| \leq \frac{1 - e^{-\lambda}}{\lambda} |j - i| \tag{12.3}$$

To utilize (12.2), note that

$$E[\lambda f_A(W+1)] = E[\sum_{i=1}^{n} \lambda_i f_A(W+1)]$$

$$= \sum_{i=1}^{n} \lambda_i E[f_A(W+1)] \tag{12.4}$$

In addition,

$$\begin{aligned}
E[Wf_A(W)] &= E[\sum_{i=1}^{n} X_i f_A(W)] \\
&= \sum_{i=1}^{n} E[X_i f_A(W)] \\
&= \sum_{i=1}^{n} (E[X_i f_A(W)|X_i = 1]\lambda_i + E[X_i f_A(W)|X_i = 0](1-\lambda_i)) \\
&= \sum_{i=1}^{n} E[f_A(W)|X_i = 1]\,\lambda_i \\
&= \sum_{i=1}^{n} E[f_A(1 + \sum_{j\neq i}^{n} X_j)|X_i = 1]\,\lambda_i \\
&= \sum_{i=1}^{n} E[f_A(1+V_i)]\,\lambda_i
\end{aligned} \tag{12.5}$$

where V_i is any random variable whose distribution is that of the conditional distribution of $\sum_{j\neq i} X_j$ given that $X_i = 1$. That is, V_i is any random variable such that

$$V_i =_{st} \sum_{j\neq i} X_j | X_i = 1$$

Hence, from (12.4) and (12.5) we see that

$$\begin{aligned}
E[\lambda f_A(W+1) - Wf_A(W)] &= \sum_{i=1}^{n} \lambda_i \,(E[f_A(W+1)] - E[f_A(1+V_i)]) \\
&= \sum_{i=1}^{n} \lambda_i \, E[f_A(W+1) - f_A(1+V_i)]
\end{aligned}$$

Taking absolute values of both sides yields, upon using the triangle inequality, that

$$\begin{aligned}
|E[\lambda f_A(W+1) - Wf_A(W)]| &\leq \sum_{i=1}^{n} \lambda_i \,|E[f_A(W+1) - f_A(1+V_i)]| \\
&\leq \sum_{i=1}^{n} \lambda_i \, E[|f_A(W+1) - f_A(1+V_i)|]
\end{aligned}$$

$$\leq \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \lambda_i E[|W - V_i|]$$

where the second inequality used that $|E[Y]| \leq E[|Y|]$ for any random variable Y, and the final inequality used Lemma 12.16.

Hence, we have proven the following:

Theorem 12.17 (Chen–Stein Poisson Approximation Bound Theorem)**.** *Let* $W = \sum_{j=1}^{n} X_j$ *where* X_j *is Bernoulli with mean* λ_j, $j = 1, \ldots, n$, *and let* Z *be Poisson with mean* $\lambda = \sum_{i=1}^{n} \lambda_i$. *Then for any random variables* $V_i, i = 1, \ldots, n$ *for which* $V_i =_{st} \sum_{j \neq i} X_j | X_i = 1$

$$\rho(W, Z) \leq \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \lambda_i E[|W - V_i|]$$

Example 12.10. If $X_1, \ldots, X_n$ are independent, then

$$\begin{aligned} V_i &=_{st} \sum_{j \neq i} X_j | X_i = 1 \\ &=_{st} \sum_{j \neq i} X_j \end{aligned}$$

where the final equation is true because $\sum_{j \neq i} X_j$ and X_i are independent. Hence, we can let $V_i = \sum_{j \neq i} X_j$, and apply the Chen–Stein bound to obtain

$$\begin{aligned} \rho(W, Z) &\leq \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \lambda_i E[|W - V_i|] \\ &= \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \lambda_i E[|X_i|] \\ &= \frac{1-e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \lambda_i^2 \end{aligned}$$

The preceding inequality is stronger than the inequality given in Proposition 12.15, which stated that $\rho(W, Z) \leq \sum_{i=1}^{n} \lambda_i^2$. For instance, if $n = 100$ and $\lambda_i \equiv .1$ then Proposition 12.15 gives that $\rho(W, Z) \leq 1$, whereas the Chen–Stein bound yields that $\rho(W, Z) \leq .1$. ■

Suppose now that $W \geq_{st} V_i$ for all i. In this case there is, by the inverse transform argument, a coupling for which $W \geq V_i$ for each $i = 1, \ldots, n$. For this coupling

$$E[|W - V_i|] = E[W - V_i] = E[W] - E[V_i] = \lambda - E[V_i]$$

Thus, we have the following corollary to the Chen–Stein bound.

Corollary 12.18. *If $W \geq_{st} V_i$ for all $i = 1, \ldots, n$, then*

$$\rho(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \lambda_i (\lambda - E[V_i]) = \frac{1 - e^{-\lambda}}{\lambda} (\lambda^2 - \sum_{i=1}^{n} \lambda_i E[V_i])$$

Remark. In cases where there is a negative dependence between $X_1, \ldots, X_n$, in the sense that knowing that $X_i = 1$ makes it less likely that $X_j = 1$ for $j \neq i$, we might expect that $W \geq_{st} V_i$.

Example 12.11. Suppose that k balls are independently distributed among n urns, with each ball going into urn j with probability p_j, $\sum_{j=1}^{n} p_j = 1$. Let X_j be the indicator for the event that none of the balls go into urn j, and let $W = \sum_{j=1}^{n} X_j$ denote the number of empty urns. Because having no balls in urn i makes it less likely that there will be no balls in box j, $j \neq i$, it seems intuitive that $W \geq_{st} V_i$. To prove this, we show how to couple W and V_i so that $W \geq V_i$. First distribute the balls into the urns as previously described, and let W denote the number of empty urns. If we now take any ball that was put in urn i, and move it to one of the other urns, choosing urn j with probability $\frac{p_j}{1-p_i}$, $j \neq i$, then the number of urns j, $j \neq i$, that are empty has the distribution of V_i. Because the redistribution of the balls that were initially in urn i cannot increase the number of the urns j, $j \neq i$, that are empty, it follows that $W \geq V_i$ for this coupling, which shows that $W \geq_{st} V_i$. Because

$$\lambda_i = E[X_i] = (1 - p_i)^k$$

$$\lambda = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} (1 - p_i)^k$$

$$E[V_i] = \sum_{j \neq i} E[X_j | X_i = 1] = \sum_{j \neq i} (1 - \frac{p_j}{1 - p_i})^k$$

we obtain from Corollary 12.18 that

$$\rho(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} (\lambda^2 - \sum_{i=1}^{n} \lambda_i \sum_{j \neq i} (1 - \frac{p_j}{1 - p_i})^k)$$

where Z is Poisson with mean λ. ■

Example 12.12. Suppose, in the scenario of Example 12.11, that we are interested in the probability that there is an urn that contains at least m balls. To analyze this, let N_i be the number of balls in box i, and let $X_i = I\{N_i \geq m\}$ be the indicator of the event that box i contains at least m balls. Letting $B(r, p)$ represent a binomial random variable with parameters (r, p), it follows, because N_i is binomial with parameters (k, p_i), that

$$\lambda_i = P(X_i = 1) = P(N_i \geq m) = P(B(k, p_i) \geq m), \quad i = 1, \ldots, n.$$

It is easily shown (see Exercise 10) that $N_i|\{N_i \geq m\} \geq_{st} N_i$. (That is, N_i becomes stochastically larger when we are told that $N_i \geq m$.) Consequently, the information that $X_i = 1$ stochastically increases the number of balls in box i, and thus stochastically reduces the number of balls available to go in the other boxes, making it likely that a coupling such that $W \geq V_i$ is possible. Indeed, the coupling can be accomplished by coupling random variables N_i^*, having the conditional distribution of a binomial (k, p_i) random variable given that it is at least m, and N_i, a binomial (k, p_i) random variable, so that $N_i^* \geq N_i$. Now consider two scenarios: in scenario 1 put N_i balls in urn i and in scenario 2 put N_i^* balls in urn i. Then distribute $n - N_i^*$ balls to the same urns in both scenarios, putting each one in urn j, $j \neq i$, with probability $\frac{p_j}{1-p_i}$. Then in scenario 1, independently put each of an additional $N_i^* - N_i$ balls into urn j with probability $\frac{p_j}{1-p_i}$, $j \neq i$. Because every urn j, $j \neq i$, contains at least as many balls in scenario 1 as it does in scenario 2, it follows that the number of urns containing at least m balls in scenario 1 is at least as large as the number of urns j, $j \neq i$, containing at least m balls in scenario 2. Because the number of urns with at least m balls in scenario 1 is distributed as W, whereas the number of urns j, $j \neq i$, having at least m balls in scenario 2 is distributed as V_i, this coupling shows that $W \geq_{st} V_i$. Now,

$$E[V_i] = \sum_{j \neq i} E[X_j|X_i = 1]$$

To determine $E[X_j|X_i = 1]$ we condition on N_i. This yields

$$\begin{aligned} E[X_j|X_i = 1] &= E[X_j|N_i \geq m] \\ &= \sum_{r=m}^{k} E[X_j|N_i = r]P(N_i = r|N_i \geq m) \\ &= \sum_{r=m}^{k} P(N_j \geq m|N_i = r)\, P(B(k, p_i) = r|B(k, p_i) \geq m) \\ &= \sum_{r=m}^{k} P(B(k-r, \frac{p_j}{1-p_i}) \geq m)\, P(B(k, p_i) = r|B(k, p_i) \geq m) \end{aligned}$$

In cases where λ_i is small, the approximation

$$E[X_j|X_i = 1] = E[X_j|N_i \geq m] \approx E[X_j|N_i = m] = P(B(k-m, \frac{p_j}{1-p_i}) \geq m)$$

should be quite precise. Hence, in this case where all λ_i are small

$$E[V_i] \approx \sum_{j \neq i} P(B(k-m, \frac{p_j}{1-p_i}) \geq m)$$

yielding, by Corollary 12.18, the error bound

$$\rho(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} (\lambda^2 - \sum_{i=1}^{n} \lambda_i \sum_{j \neq i} P(B(k-m, \frac{p_j}{1-p_i}) \geq m))$$

One application of the preceding is the generalized birthday problem, which supposes that each of k people independently has birthday j with probability p_j, $j = 1, \ldots, n$, and we are interested in the probability that among the k people there is a set of m all having the same birthday. If we let X_i, $i = 1, \ldots, n$ be the indicator of the event that at least m of the k people were born on day i, then the desired probability is $P(W > 0)$, where $W = \sum_{i=1}^{n} X_i$. For instance, suppose that $n = 365$ and $p_i = 1/365$ for all i. Then, when $m = 3$, we are asking for the probability that among k people, whose birthdays are assumed independent and equally likely to be any of the 365 days, there will be a group of size three that all share the same birthday. Because $\lambda_i = P(X_i = 1) = P(B(k, 1/365) \geq 3)$ we have when $k = 88$ that $\lambda_i \approx .0018966$. Hence, with W being the number of days of the year that are the birthdays of at least 3 people, the Poisson approximation yields that W is approximately Poisson with mean $\lambda = 365(.0018966) = .69226$. Thus, the Poisson approximation of the probability that there will be a set of size three having the same birthday is

$$P(W > 0) \approx 1 - e^{-.69226} \approx .49956.$$

Using that $E[V_i] \approx 364\, P(B(85, 1/364) \geq 3) \approx .63006$, we see that the error of the Poisson approximation is bounded by

$$\rho(W, Z) \leq \frac{1 - e^{-.69226}}{.69226} (.69226^2 - .69226 \times .63006) \approx .031.$$

(It was shown in Exercise 20 of Chapter 2 that, to 3 decimal places, $P(W > 0) = .504$.) ■

In cases where there is a positive dependence of $X_1, \ldots, X_n$ which results in $V_i \geq_{st} \sum_{j \neq i} X_j$ for all i, we can implement the Chen–Stein approach by coupling V_i and $X_1, \ldots, X_n$ so that $V_i \geq \sum_{j \neq i} X_j$. With this coupling, we then have

$$\begin{aligned} |V_i - W| &= |V_i - \sum_{j \neq i} X_j - X_i| \\ &\leq |V_i - \sum_{j \neq i} X_j| + |X_i| \qquad \text{(by the triangle inequality)} \\ &= V_i - \sum_{j \neq i} X_j + X_i \\ &= V_i - W + 2X_i \end{aligned}$$

which yields

$$E[|V_i - W|] \leq E[V_i] - \lambda + 2\lambda_i$$

Using this, along with the Chen–Stein Theorem yields the following.

Proposition 12.19. *If $V_i \geq_{st} \sum_{j \neq i} X_j$ for all $i = 1, \ldots, n$, then*

$$\rho(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} \left(\sum_i \lambda_i E[V_i] - \lambda^2 + 2 \sum_i \lambda_i^2 \right) \tag{12.6}$$

Example 12.13. Consider an m component system where component j is failed with probability q_j, $j = 1, \ldots, m$ and where the components are independent. Let $C_1, \ldots, C_n$ be subsets, none of which is contained in another, such that the system is failed if and only if all of the components of at least one of these subsets is failed. ($C_1, \ldots, C_n$ are called the *minimal cut sets* of the system.) With X_k being the indicator of the event that all components in C_k are failed, $\lambda_k = E[X_k] = \prod_{j \in C_k} q_j$. The system will be failed if $W \equiv \sum_{k=1}^n X_k > 0$. To bound the error involved when we approximate the distribution of W by a Poisson distribution with mean $\lambda = \sum_{k=1}^n \lambda_k$, we use that we can couple $X_1, \ldots, X_n$ and V_i so that $V_i \geq \sum_{k \neq i} X_k$. The coupling is obtained by first letting $Y_1, \ldots, Y_m$ be independent Bernoulli random variables with means $q_1, \ldots, q_m$. Now, set $X_k = \prod_{j \in C_k} Y_j$, $k = 1, \ldots, n$. Also, let

$$Y_j^* = \begin{cases} Y_j, & \text{if } j \notin C_i \\ 1, & \text{if } j \in C_i \end{cases}$$

Set $X_k^* = \prod_{j \in C_k} Y_j^*$, $k = 1, \ldots, n$, and let $V_i = \sum_{k \neq i} X_k^*$. Because $Y_j^* \geq Y_j$ for all j, it follows that $X_k^* \geq X_k$ for all k, thus yielding a coupling where $V_i \geq \sum_{k \neq i} X_k$. ■

Exercises

1. Show that a normal random variable is stochastically increasing in its mean. That is, with $N(\mu, \sigma)$ being a normal random variable with mean μ and variance σ^2, show that $N(\mu_1, \sigma) \geq_{st} N(\mu_2, \sigma)$ when $\mu_1 > \mu_2$.
2. If $\sigma_1 \neq \sigma_2$, is it possible to have $N(\mu_1, \sigma_1) \geq_{st} N(\mu_2, \sigma_2)$.
3. Show that a gamma (n, λ) random variable, whose density is

$$f(x) = \lambda e^{-\lambda x} (\lambda x)^{n-1} / (n-1)!, \qquad x > 0$$

   is stochastically increasing in n and stochastically decreasing in λ.
4. Let $\mathbf{N_i} = \{N_i(t), t \geq 0\}$ be a renewal process with interarrival distribution F_i, $i = 1, 2$. If $F_1 \leq F_2$, show that $\mathbf{N_1} \leq_{st} \mathbf{N_2}$.
5. Let $\mathbf{N_i} = \{N_i(t), t \geq 0\}$, $i = 1, 2$, be nonhomogeneous Poisson processes with respective intensity functions $\lambda_i(t)$, $i = 1, 2$. Suppose $\lambda_1(t) \geq \lambda_2(t)$ for all t. Let A_j, $j = 1, \ldots, n$ be arbitrary subsets of the real line, and for $i = 1, 2$, let $N_i(A_j)$ be the number of points of the process $\mathbf{N_i}$ that are in A_j, $j = 1, \ldots, n$. Show that $(N_1(A_1), \ldots, N_1(A_n)) \geq_{st} (N_2(A_1), \ldots, N_2(A_n))$.

6. A new item will fail on its ith day of use with probability p_i, $\sum_{i=1}^{\infty} p_i = 1$. An item that fails during a period is replaced by a new one at the beginning of the next period. Let A_n denote the age of the item in use at the beginning of period n. That is, $A_n = i$ if the item in use is beginning its ith day. The random variables A_n can be interpreted as the age at time n of a renewal process whose interarrival times have mass function $\{p_i, i \geq 1\}$, with $A_n = 1$ signifying that a renewal occurs at time n.

(a) Argue that $\{A_n, n \geq 1\}$ is a Markov chain and give its transition probabilities.

(b) Suppose $A_0 = 1$. If $\frac{p_i}{\sum_{j=i}^{\infty} p_j}$ decreases in i, show that A_n stochastically increases in n.

7. If X is a positive integer valued random variable, with mass function $p_i = P(X = i)$, $i \geq 1$, then the function

$$\lambda(i) = P(X = i | X \geq i)$$

is called the (discrete) hazard rate function of X.

(a) Express $P(X > n)$ in terms of the values $\lambda(i), i \geq 1$.

(b) If $\lambda(i)$ is increasing (decreasing) in i then the random variable X is said to have increasing (decreasing) failure rate. Let X_n^* be a random variable whose distribution is that of the conditional distribution of $X - n$ given that $X \geq n$. That is,

$$P(X_n^* = j) = P(X = n + j | X \geq n).$$

Show that if X has increasing (decreasing) failure rate it and only if X_n^* stochastically decreases (increases) in n.

8. Consider two renewal processes: $N_x = \{N_x(t), t \geq 0\}$ and $N_y = \{N_y(t), t \geq 0\}$ whose interarrival distributions are discrete with, respective, hazard rate functions $\lambda_x(i)$ and $\lambda_y(i)$. For any set of points A, let $N_x(A)$ and $N_y(A)$ denote, respectively, the numbers of renewals that occur at time points in A for the two processes. If $\lambda_x(i) \leq \lambda_y(i)$ for all i and either $\lambda_x(i)$ or $\lambda_y(i)$ is decreasing, show that $N_x(A) \leq_{st} N_y(A)$ for any A.

9. A discrete time birth and death process is a Markov chain $\{X_n, n \geq 0\}$ with transition probabilities of the form $P_{i,i+1} = p_i = 1 - P_{i,i-1}$. Prove or give a counterexample to the claim that $\{X_n, n \geq 0 | X_0 = i\}$ is stochastically increasing in i.

10. If X_a is a random variable whose distribution is that of the conditional distribution of X given that $X > a$, show that $X_a \geq_{st} X$ for every a.

11. Let $\{N(t), t \geq 0\}$ be a renewal process whose interarrival times $X_i, i \geq 1$, have distribution F.

(a) The random variable $X_{N(t)+1}$ is the length of the renewal interval that does what.

(b) Show that $X_{N(t)+1} \geq_{st} X_i$.

12. Let $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$ be independent irreducible Markov chains with states $0, 1, \ldots, m$, and with respective transition probabilities $P_{i,j}$ and $Q_{i,j}$.

(a) Give the transition probabilities of the Markov chain $\{(X_n, Y_n),\ n \geq 0\}$.
(b) Show by giving a counterexample that $\{(X_n, Y_n),\ n \geq 0\}$ is not necessarily irreducible.

13. If X and Y are discrete integer valued random variables with respective mass functions p_i and q_i, show that

$$\rho(X, Y) = \frac{1}{2} \sum_i |p_i - q_i|$$

14. With W and V_i as defined in Section 12.7, show that
(a) $\sum_{i=1}^{n} \lambda_i E[1 + V_i] = E[W^2]$
(b) If for each $i = 1, \ldots, n$, W and V_i can be coupled so that $W \geq V_i$, show that

$$\rho(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} (\lambda - \text{Var}(W))$$

15. A coin with probability p of coming up heads is flipped $n + k$ times. Let R_k denote the event that a run of k consecutive heads occurs at least once. Let X_1 be the indicator variable of the event that flips $1, \ldots, k$ all land heads, and for $i = 2, \ldots, n + 1$, let X_i be the indicator variable of the event that flip $i - 1$ lands tails and flips $i, \ldots, i + k - 1$ all land heads. With $W = \sum_{i=1}^{n+1} X_i$ show that
(a) $P(R_k) = P(W > 0)$.
(b) Approximate $P(W > 0)$.
(c) Bound the error of the approximation.

16. Show that $|E[X]| \leq E[|X|]$.

17. In Example 12.12 show that $E[X_j | N_i = m]$, the approximation of $E[X_j | X_i = 1]$ when λ_i is small, is an upper bound. That is, show that $E[X_j | X_i = 1] \leq E[X_j | N_i = m]$.

18. In a group of size 101 each pair of individuals are, independently, friends with probability .01. With N_4 equal to the number of individuals that have at least 4 friends, approximate the probability that $P(N_4 \geq 3)$, and give a bound on the error of your approximation.

Solutions to Starred Exercises

Chapter 1

2. $S = \{(r, g), (r, b), (g, r), (g, b), (b, r), (b, g)\}$ where, for instance, (r, g) means that the first marble drawn was red and the second one green. The probability of each one of these outcomes is $\frac{1}{6}$.

5. $\frac{3}{4}$. If he wins, he only wins \$1; if he loses, he loses \$3.

9. $F = E \cup FE^c$, implying since E and FE^c are disjoint that $P(F) = P(E) + P(FE^c)$.

17.
$$\begin{aligned} P\{\text{end}\} &= 1 - P\{\text{continue}\} \\ &= 1 - [\text{Prob}(H, H, H) + \text{Prob}(T, T, T)] \end{aligned}$$

$$\text{Fair coin: } P\{\text{end}\} = 1 - \left[\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right] = \frac{3}{4}$$

$$\text{Biased coin: } P\{\text{end}\} = 1 - \left[\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4} + \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}\right] = \frac{9}{16}$$

19.
$$E = \text{event at least 1 six}$$

$$P(E) = \frac{\text{number of ways to get } E}{\text{number of sample points}} = \frac{11}{36}$$

$$D = \text{event two faces are different}$$

$$P(D) = 1 - P(\text{two faces the same}) = 1 - \frac{6}{36} = \frac{5}{6}$$

$$P(E|D) = \frac{P(ED)}{P(D)} = \frac{10/36}{5/6} = \frac{1}{3}$$

25. **(a)**
$$\begin{aligned} P\{\text{pair}\} &= P\{\text{second card is same denomination as first}\} \\ &= \frac{3}{51} \end{aligned}$$

(b)
$$\begin{aligned} P\{\text{pair} \mid \text{different suits}\} &= \frac{P\{\text{pair, different suits}\}}{P\{\text{different suits}\}} \\ &= \frac{P\{\text{pair}\}}{P\{\text{different suits}\}} \end{aligned}$$

$$= \frac{3/51}{39/51} = \frac{1}{13}$$

27. $P(E_1) = 1$

$$P(E_2|E_1) = \frac{39}{51}$$

since 12 cards are in the ace of spades pile and 39 are not.

$$P(E_3|E_1E_2) = \frac{26}{50}$$

since 24 cards are in the piles of the two aces and 26 are in the other two piles.

$$P(E_4|E_1E_2E_3) = \frac{13}{49}$$

So

$$P\{\text{each pile has an ace}\} = \left(\frac{39}{51}\right)\left(\frac{26}{50}\right)\left(\frac{13}{49}\right)$$

30. **(a)**

$$P\{\text{George} \mid \text{exactly 1 hit}\} = \frac{P\{\text{George, not Bill}\}}{P\{\text{exactly 1}\}}$$

$$= \frac{P\{\text{G, not B}\}}{P\{\text{G, not B}\} + P\{\text{B, not G}\}}$$

$$= \frac{(0.4)(0.3)}{(0.4)(0.3) + (0.7)(0.6)}$$

$$= \frac{2}{9}$$

(b)

$$P\{\text{G} \mid \text{hit}\} = \frac{P\{\text{G, hit}\}}{P\{\text{hit}\}}$$

$$= \frac{P\{\text{G}\}}{P\{\text{hit}\}} = \frac{0.4}{1 - (0.3)(0.6)} = \frac{20}{41}$$

32. Let E_i = event person i selects own hat.

$$\begin{aligned}
&P \text{ (no one selects hat)} \\
&= 1 - P(E_1 \cup E_2 \cup \cdots \cup E_n) \\
&= 1 - \left[\sum_{i_1} P(E_{i_1}) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots + (-1)^{n+1} P(E_1E_2\cdots E_n)\right] \\
&= 1 - \sum_{i_1} P(E_{i_1}) + \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) - \sum_{i_1<i_2<i_3} P(E_{i_1}E_{i_2}E_{i_3}) + \cdots \\
&\quad + (-1)^n P(E_1E_2\cdots E_n)
\end{aligned}$$

Let $k \in \{1, 2, \ldots, n\}$. $P(E_{i_1} E_{i_2} E_{i_k})$ = number of ways k specific men can select own hats $\div$ total number of ways hats can be arranged $= (n-k)!/n!$. Number of terms in summation $\sum_{i_1<i_2<\cdots<i_k}$ = number of ways to choose k variables out of n variables $= \binom{n}{k} = n!/k!(n-k)!$. Thus,

$$\begin{aligned}\sum_{i_1<\cdots<i_k} P(E_{i_1} E_{i_2} \cdots E_{i_k}) &= \sum_{i_1<\cdots<i_k} \frac{(n-k)!}{n!} \\ &= \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!} \\ \therefore P(\text{no one selects own hat}) &= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \\ &= \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\end{aligned}$$

40. **(a)** F = event fair coin flipped; U = event two-headed coin flipped.

$$\begin{aligned}P(F|H) &= \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|U)P(U)} \\ &= \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}\end{aligned}$$

(b)
$$\begin{aligned}P(F|HH) &= \frac{P(HH|F)P(F)}{P(HH|F)P(F) + P(HH|U)P(U)} \\ &= \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{1}{4} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}} = \frac{\frac{1}{8}}{\frac{5}{8}} = \frac{1}{5}\end{aligned}$$

(c)
$$\begin{aligned}P(F|HHT) &= \frac{P(HHT|F)P(F)}{P(HHT|F)P(F) + P(HHT|U)P(U)} \\ &= \frac{P(HHT|F)P(F)}{P(HHT|F)P(F) + 0} = 1\end{aligned}$$

since the fair coin is the only one that can show tails.

43. Let B be the event that Flo has a blue eyed gene. Using that Jo and Joe both have one blue-eyed gene yields, upon letting X be the number of blue-eyed genes a daughter of possessed by a daughter of theirs, that

$$P(B) = P(X = 1 | X < 2) = \frac{1/2}{3/4} = 2/3$$

Hence, with C being the event that Flo's daughter is blue eyed, we obtain

$$P(C) = P(CB) = P(B)P(C|B) = 1/3$$

45. Let B_i = event ith ball is black; R_i = event ith ball is red.

$$\begin{aligned} P(B_1|R_2) &= \frac{P(R_2|B_1)P(B_1)}{P(R_2|B_1)P(B_1) + P(R_2|R_1)P(R_1)} \\ &= \frac{\dfrac{r}{b+r+c} \cdot \dfrac{b}{b+r}}{\dfrac{r}{b+r+c} \cdot \dfrac{b}{b+r} + \dfrac{r+c}{b+r+c} \cdot \dfrac{r}{b+r}} \\ &= \frac{rb}{rb + (r+c)r} \\ &= \frac{b}{b+r+c} \end{aligned}$$

48. Let C be the event that the randomly chosen family owns a car, and let H be the event that the randomly chosen family owns a house.

$$P(CH^c) = P(C) - P(CH) = 0.6 - 0.2 = 0.4$$

and

$$P(C^c H) = P(H) - P(CH) = 0.3 - 0.2 = 0.1$$

giving the result

$$P(CH^c) + P(C^c H) = 0.5$$

Chapter 2

4. **(a)** $1, 2, 3, 4, 5, 6$.
(b) $1, 2, 3, 4, 5, 6$.
(c) $2, 3, \ldots, 11, 12$.
(d) $-5, 4, \ldots, 4, 5$.

11. $\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8}$.

16. $1 - (0.95)^{52} - 52(0.95)^{51}(0.05)$.

18. **(a)**

$$\begin{aligned} &P(X_i = x_i, i = 1, \ldots, r-1 | X_r = j) \\ &\quad = \frac{P(X_i = x_i, i = 1, \ldots, r-1, X_r = j)}{P(X_r = j)} \\ &\quad = \frac{\frac{n!}{x_1! \cdots x_{r-1}! j!} p_1^{x_1} \cdots p_{r-1}^{x_{r-1}} p_r^j}{\frac{n!}{j!(n-j)!} p_r^j (1-p_r)^{n-j}} \\ &\quad = \frac{(n-j)!}{x_1! \cdots x_{r-1}!} \prod_{i=1}^{r-1} \left(\frac{p_i}{1-p_r}\right)^{x_i} \end{aligned}$$

(b) The conditional distribution of $X_1, \ldots, X_{r-1}$ given that $X_r = j$ is multinomial with parameters $n - j$, $\frac{p_i}{1-p_r}$, $i = 1, \ldots, r-1$.

(c) The preceding is true because given that $X_r = j$, each of the $n - j$ trials that did not result in outcome r resulted in outcome i with probability $\frac{p_i}{1-p_r}$, $i = 1, \ldots, r-1$.

23. In order for X to equal n, the first $n-1$ flips must have $r-1$ heads, and then the nth flip must land heads. By independence the desired probability is thus

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

27.

$$\begin{aligned} P\{\text{same number of heads}\} &= \sum_i P\{A = i, B = i\} \\ &= \sum_i \binom{k}{i}\left(\frac{1}{2}\right)^k \binom{n-k}{i}\left(\frac{1}{2}\right)^{n-k} \\ &= \sum_i \binom{k}{i}\binom{n-k}{i}\left(\frac{1}{2}\right)^n \\ &= \sum_i \binom{k}{k-i}\binom{n-k}{i}\left(\frac{1}{2}\right)^n \\ &= \binom{n}{k}\left(\frac{1}{2}\right)^n \end{aligned}$$

Another argument is as follows:

$$\begin{aligned} &P\{\#\text{ heads of } A = \#\text{ heads of B}\} \\ &\quad = P\{\#\text{ tails of } A = \#\text{ heads of B}\} \qquad \text{since coin is fair} \\ &\quad = P\{k - \#\text{ heads of } A = \#\text{ heads of B}\} \\ &\quad = P\{k = \text{total }\#\text{ heads}\} \end{aligned}$$

38. $c = 2$, $\qquad P\{X > 2\} = \int_2^\infty 2e^{-2x}dx = e^{-4}$

47. Let X_i be 1 if trial i is a success and 0 otherwise.

(a) The largest value is 0.6. If $X_1 = X_2 = X_3$, then

$$1.8 = E[X] = 3E[X_1] = 3P\{X_1 = 1\}$$

and so $P\{X = 3\} = P\{X_1 = 1\} = 0.6$. That this is the largest value is seen by Markov's inequality, which yields

$$P\{X \geq 3\} \leq E[X]/3 = 0.6$$

(b) The smallest value is 0. To construct a probability scenario for which $P\{X = 3\} = 0$, let U be a uniform random variable on $(0, 1)$, and define

$$X_1 = \begin{cases} 1, & \text{if } U \le 0.6 \\ 0, & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if } U \ge 0.4 \\ 0, & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1, & \text{if either } U \le 0.3 \text{ or } U \ge 0.7 \\ 0, & \text{otherwise} \end{cases}$$

It is easy to see that

$$P\{X_1 = X_2 = X_3 = 1\} = 0$$

49. $E[X^2] - (E[X])^2 = \text{Var}(X) = E[(X - E[X])^2] \ge 0$. There is equality when $\text{Var}(X) = 0$, that is, when X is constant.

64. For the matching problem, letting $X = X_1 + \cdots + X_N$, where

$$X_i = \begin{cases} 1, & \text{if } i\text{th man selects his own hat} \\ 0, & \text{otherwise} \end{cases}$$

we obtain

$$\text{Var}(X) = \sum_{i=1}^{N} \text{Var}(X_i) + 2 \sum \sum_{i<j} \text{Cov}(X_i, X_j)$$

Since $P\{X_i = 1\} = 1/N$, we see

$$\text{Var}(X_i) = \frac{1}{N}\left(1 - \frac{1}{N}\right) = \frac{N-1}{N^2}$$

Also,

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

Now,

$$X_i X_j = \begin{cases} 1, & \text{if the } i\text{th and } j\text{th men both select their own hats} \\ 0, & \text{otherwise} \end{cases}$$

and thus

$$\begin{aligned} E[X_i X_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\} P\{X_j = 1 | X_i = 1\} \\ &= \frac{1}{N} \frac{1}{N-1} \end{aligned}$$

Hence,

$$\text{Cov}(X_i, X_j) = \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)^2 = \frac{1}{N^2(N-1)}$$

and

$$\begin{aligned}\text{Var}(X) &= \frac{N-1}{N} + 2\binom{N}{2}\frac{1}{N^2(N-1)} \\ &= \frac{N-1}{N} + \frac{1}{N} \\ &= 1\end{aligned}$$

66. Letting B_i be the event that $X_i \in A_i$, $i = 1, \ldots, n$, we have

$$P(B_1 \cdots B_n) = P(B_1)\prod_{i=2}^{n} P(B_i|B_1 \cdots B_{i-1}) = P(B_1)\prod_{i=2}^{n} P(B_i)$$

71. See Section 5.2.3 of Chapter 5. Another way is to use moment generating functions. The moment generating function of the sum of n independent exponentials with rate λ is equal to the product of their moment generating functions. That is, it is $[\lambda/(\lambda - t)]^n$. But this is precisely the moment generating function of a gamma with parameters n and λ.

74. $$E[e^{-uX}] = \sum_n e^{-un}e^{-\lambda}\lambda^n/n! = e^{-\lambda}\sum_n (\lambda e^{-u})^n/n! = e^{\lambda(e^{-u}-1)}$$

80. Let X_i be Poisson with mean 1. Then

$$P\left\{\sum_1^n X_i \le n\right\} = e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$$

But for n large $\sum_1^n X_i - n$ has approximately a normal distribution with mean 0, and so the result follows.

85. **(a)** Using that $\text{Var}\left(\frac{W}{\sigma_W}\right) = 1$ along with the formula for the variance of a sum gives

$$2 + 2\frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y} \ge 0$$

(b) Start with $\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \ge 0$, and proceed as in part (a).

(c) Squaring both sides yields that the inequality is equivalent to

$$\text{Var}(X + Y) \le \text{Var}(X) + \text{Var}(Y) + 2\sigma_X\sigma_Y$$

or, using the formula for the variance of a sum

$$\text{Cov}(X, Y) \le \sigma_X\sigma_Y$$

which is part (b).

86. Let X_i be the time it takes to process book i. With Z being a standard normal

(a) $P\left(\sum_{i=1}^{40} X_i > 420\right) \approx P\left(Z > \frac{420-400}{\sqrt{9 \cdot 40}}\right)$

(b) $P\left(\sum_{i=1}^{25} X_i < 240\right) \approx P\left(Z < \frac{240-250}{\sqrt{9 \cdot 25}}\right) = P(Z > 2/3)$

Chapter 3

2. Intuitively it would seem that the first head would be equally likely to occur on any of trials $1, \ldots, n-1$. That is, it is intuitive that

$$P\{X_1 = i | X_1 + X_2 = n\} = \frac{1}{n-1}, \qquad i = 1, \ldots, n-1$$

Formally,

$$\begin{aligned} P\{X_1 = i | X_1 + X_2 = n\} &= \frac{P\{X_1 = i, X_1 + X_2 = n\}}{P\{X_1 + X_2 = n\}} \\ &= \frac{P\{X_1 = i, X_2 = n - i\}}{P\{X_1 + X_2 = n\}} \\ &= \frac{p(1-p)^{i-1} p(1-p)^{n-i-1}}{\binom{n-1}{1} p(1-p)^{n-2} p} \\ &= \frac{1}{n-1} \end{aligned}$$

In the preceding, the next to last equality uses the independence of X_1 and X_2 to evaluate the numerator and the fact that $X_1 + X_2$ has a negative binomial distribution to evaluate the denominator.

6.

$$\begin{aligned} p_{X|Y}(1 \,|\, 3) &= \frac{P\{X = 1, Y = 3\}}{P\{Y = 3\}} \\ &= \frac{P\{1 \text{ white, } 3 \text{ black, } 2 \text{ red}\}}{P\{3 \text{ black}\}} \\ &= \frac{\dfrac{6!}{1!3!2!}\left(\dfrac{3}{14}\right)^1\left(\dfrac{5}{14}\right)^3\left(\dfrac{6}{14}\right)^2}{\dfrac{6!}{3!3!}\left(\dfrac{5}{14}\right)^3\left(\dfrac{9}{14}\right)^3} \\ &= \frac{4}{9} \\ p_{X|Y}(0 \,|\, 3) &= \frac{8}{27} \\ p_{X|Y}(2 \,|\, 3) &= \frac{2}{9} \end{aligned}$$

$$p_{X|Y}(3\,|\,3) = \frac{1}{27}$$
$$E[X\,|\,Y=1] = \frac{5}{3}$$

13. The conditional density of X given that $X > 1$ is

$$f_{X|X>1}(X) = \frac{f(x)}{P\{X>1\}} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda}} \qquad \text{when } x > 1$$
$$E[X\,|\,X>1] = e^{\lambda}\int_1^{\infty} x\lambda e^{-\lambda x}dx = 1 + 1/\lambda$$

by integration by parts. This latter result also follows immediately by the lack of memory property of the exponential.

19.
$$\begin{aligned}\int E[X\,|\,Y=y]f_Y(y)\,dy &= \iint x f_{X|Y}(x|y)\,dx f_Y(y)\,dy\\ &= \iint x\frac{f(x,y)}{f_Y(y)}\,dx f_Y(y)\,dy\\ &= \int x\int f(x,y)\,dy\,dx\\ &= \int x f_X(x)\,dx\\ &= E[X]\end{aligned}$$

23. Let X denote the first time a head appears. Let us obtain an equation for $E[N\,|\,X]$ by conditioning on the next two flips after X. This gives

$$\begin{aligned}E[N\,|\,X] = {} & E[N\,|\,X,h,h]p^2 + E[N\,|\,X,h,t]pq + E[N\,|\,X,t,h]pq\\ & + E[N\,|\,X,t,t]q^2\end{aligned}$$

where $q = 1-p$. Now

$$E[N\,|\,X,h,h] = X+1, \qquad E[N\,|\,X,h,t] = X+1$$
$$E[N\,|\,X,t,h] = X+2, \qquad E[N\,|\,X,t,t] = X+2+E[N]$$

Substituting back gives

$$E[N|X] = (X+1)(p^2+pq) + (X+2)pq + (X+2+E[N])\,q^2$$

Taking expectations, and using the fact that X is geometric with mean $1/p$, we obtain

$$E[N] = 1 + p + q + 2pq + q^2/p + 2q^2 + q^2E[N]$$

Solving for $E[N]$ yields

$$E[N] = \frac{2+2q+q^2/p}{1-q^2}$$

38. $E[X] = E[E[X|Y]] = E[Y/2] = 1/4$

$$\begin{aligned}\text{Var}(X) &= E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) \\ &= E[Y^2/12] + \text{Var}(Y/2) \\ &= 1/36 + 1/48 = 1/12\end{aligned}$$

Suppose Y is uniformly distributed on $(0, 1)$, and that the conditional distribution of X given that $Y = y$ is uniform on $(0, y)$. Find $E[X]$ and $\text{Var}(X)$.

41. Condition on whether any of the workers is eligible and then use symmetry. This gives

$$P(1) = P(1|\text{someone is eligible})\,P(\text{someone is eligible}) = \frac{1}{n}[1 - (1-p)^n]$$

42. **(a)**

$$\begin{aligned}E[e^{tX^2}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-(x-\mu)^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-(x^2 - 2\mu x + \mu^2 - 2tx^2)/2\} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2} \int_{-\infty}^{\infty} \exp\{-(x^2(1-2t) - 2\mu x)/2\} dx\end{aligned}$$

Thus, with $\sigma^2 = \frac{1}{1-2t}$

$$E[e^{tX^2}] = \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2} \int_{-\infty}^{\infty} \exp\{-(x^2 - 2\sigma^2\mu x)/2\sigma^2\} dx$$

Using that

$$x^2 - 2\sigma^2\mu x = (x - \sigma^2\mu)^2 - \mu^2\sigma^4$$

we have

$$\begin{aligned}E[e^{tX^2}] &= e^{-\mu^2/2 + \mu^2\sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-(x - \sigma^2\mu)^2/2\sigma^2\} dx \\ &= e^{-(1-\sigma^2)\mu^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-y^2/2\sigma^2\} dy \\ &= \sigma e^{-(1-\sigma^2)\mu^2/2} \\ &= (1-2t)^{-1/2} \exp\left\{-\left(1 - \frac{1}{1-2t}\right)\mu^2/2\right\} \\ &= (1-2t)^{-1/2} e^{\frac{t\mu^2}{1-2t}}\end{aligned}$$

(b)
$$E\left[\exp\left\{t\sum_{i=1}^{n}X_i^2\right\}\right]=\prod_{i=1}^{n}E\left[e^{tX_i^2}\right]$$
$$=(1-2t)^{-n/2}\exp\left\{\frac{t}{1-2t}\sum_{i=1}^{n}\mu_i^2\right\}$$

(c)
$$\frac{d}{dt}(1-2t)^{-n/2}=n(1-2t)^{-n/2-1}$$
$$\frac{d^2}{dt^2}(1-2t)^{-n/2}=2n(n/2+1)(1-2t)^{-n/2-2}$$

Hence, if χ_n^2 is chi-squared with n degrees of freedom then evaluating the preceding at $t=0$ gives

$$E\left[\chi_n^2\right]=n,\quad \mathrm{Var}\left(\chi_n^2\right)=n^2+2n-n^2=2n$$

(d) Conditioning on K yields

$$\begin{aligned}E\left[e^{tW}\right]&=\sum_{k=0}^{\infty}E\left[e^{tW}|K=k\right]e^{-\theta/2}(\theta/2)^k/k!\\&=\sum_{k=0}^{\infty}(1-2t)^{-(n+2k)/2}e^{-\theta/2}(\theta/2)^k/k!\\&=(1-2t)^{-n/2}e^{-\theta/2}\sum_{k=0}^{\infty}(1-2t)^{-k}(\theta/2)^k/k!\\&=(1-2t)^{-n/2}e^{-\theta/2}\sum_{k=0}^{\infty}\left(\frac{\theta}{2(1-2t)}\right)^k\Big/k!\\&=(1-2t)^{-n/2}\exp\left\{-\frac{\theta}{2}+\frac{\theta}{2(1-2t)}\right\}\\&=(1-2t)^{-n/2}\exp\left\{\frac{t\theta}{1-2t}\right\}\end{aligned}$$

Because the preceding is the moment generating function of a noncentral chi-squared random variable with parameters n and θ, and the moment generating function uniquely determines the distribution, the result is proven.

(e) From the preceding, we have

$$E[W|K=k]=E[\chi_{n+2k}^2]=n+2k$$
$$\mathrm{Var}(W|K=k)=\mathrm{Var}(\chi_{n+2k}^2)=2n+4k$$

Hence,

$$E[W] = E[E[W|K]] = E[n + 2K] = n + 2E[K] = n + \theta$$

and the conditional variance formula yields

$$\text{Var}(W) = E[2n + 4K] + \text{Var}(n + 2K) = 2n + 2\theta + 2\theta = 2n + 4\theta$$

43. With $I = I\{Y \in A\}$

$$E[XI] = E[XI|I = 1]P(I = 1) + E[XI|I = 0]P(I = 0) = E[X|I = 1]P(I = 1)$$

47.
$$\begin{aligned} E[X^2Y^2|X] &= X^2E[Y^2|X] \\ &\geq X^2(E[Y|X])^2 = X^2 \end{aligned}$$

The inequality follows since for any random variable U, $E[U^2] \geq (E[U])^2$ and this remains true when conditioning on some other random variable X. Taking expectations of the preceding shows that

$$E[(XY)^2] \geq E[X^2]$$

As

$$E[XY] = E[E[XY|X]] = E[XE[Y|X]] = E[X]$$

the results follow.

53.
$$\begin{aligned} P\{X = n\} &= \int_0^\infty P\{X = n|\lambda\}e^{-\lambda}d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!}e^{-\lambda}d\lambda \\ &= \int_0^\infty e^{-2\lambda}\lambda^n\frac{d\lambda}{n!} \\ &= \int_0^\infty e^{-t}t^n\frac{dt}{n!}\left(\frac{1}{2}\right)^{n+1} \end{aligned}$$

The results follow since $\int_0^\infty e^{-t}t^n dt = \Gamma(n + 1) = n!$

58. **(a)** r/λ;

(b) $E[\text{Var}(N|Y)] + \text{Var}(E[N|Y]) = E[Y] + \text{Var}(Y) = \frac{r}{\lambda} + \frac{r}{\lambda^2}$

(c) With $p = \frac{\lambda}{\lambda+1}$

$$\begin{aligned} P(N = n) &= \int P(N = n|Y = y)\, f_Y(y)\, dy \\ &= \int e^{-y}\frac{y^n}{n!}\frac{\lambda e^{-\lambda y}(\lambda y)^{r-1}}{(r-1)!}\, dy \end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda^r}{n!(r-1)!} \int e^{-(\lambda+1)y} y^{n+r-1}\, dy \\
&= \frac{\lambda^r}{n!(r-1)!(\lambda+1)^{n+r}} \int e^{-x} x^{n+r-1}\, dx \\
&= \frac{\lambda^r (n+r-1)!}{n!(r-1)!(\lambda+1)^{n+r}} \\
&= \binom{n+r-1}{r-1} p^r (1-p)^n
\end{aligned}
$$

(d) The total number of failures before the rth success when each trial is independently a success with probability p is distributed as $X - r$ where X, equal to the number of trials until the rth success, is negative binomial. Hence,

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

60. **(a)** Intuitive that $f(p)$ is increasing in p, since the larger p is the greater is the advantage of going first.

(b) 1.

(c) $\frac{1}{2}$ since the advantage of going first becomes nil.

(d) Condition on the outcome of the first flip:

$$
\begin{aligned}
f(p) &= P\{I \text{ wins} \mid h\} p + P\{I \text{ wins} \mid t\}(1-p) \\
&= p + [1 - f(p)](1-p)
\end{aligned}
$$

Therefore,

$$
f(p) = \frac{1}{2-p}
$$

67. Part (a) is proven by noting that a run of j successive heads can occur within the first n flips in two mutually exclusive ways. Either there is a run of j successive heads within the first $n-1$ flips; or there is no run of j successive heads within the first $n-j-1$ flips, flip $n-j$ is not a head, and flips $n-j+1$ through n are all heads.

Let A be the event that a run of j successive heads occurs within the first n, $(n \geq j)$, flips. Conditioning on X, the trial number of the first non-head, gives the following

$$
\begin{aligned}
P_j(n) &= \sum_k P(A|X=k) p^{k-1}(1-p) \\
&= \sum_{k=1}^{j} P(A|X=k) p^{k-1}(1-p) + \sum_{k=j+1}^{\infty} P(A|X=k) p^{k-1}(1-p)
\end{aligned}
$$

$$= \sum_{i=1}^{j} P_j(n-k)p^{k-1}(1-p) + \sum_{k=j+1}^{\infty} p^{k-1}(1-p)$$

$$= \sum_{i=1}^{j} P_j(n-k)p^{k-1}(1-p) + p^j$$

73. Condition on the value of the sum prior to going over 100. In all cases the most likely value is 101. (For instance, if this sum is 98 then the final sum is equally likely to be either 101, 102, 103, or 104. If the sum prior to going over is 95, then the final sum is 101 with certainty.)

84. Suppose in Example 3.33 that a point is only won if the winner of the rally was the server of that rally.

(a) If A is currently serving, what is the probability that A wins the next point?

(b) Explain how to obtain the final score probabilities.

93. **(a)** By symmetry, for any value of $(T_1, \ldots, T_m)$, the random vector $(I_1, \ldots, I_m)$ is equally likely to be any of the $m!$ permutations.

(b)
$$E[N] = \sum_{i=1}^{m} E[N|X=i]P\{X=i\}$$

$$= \frac{1}{m}\sum_{i=1}^{m} E[N|X=i]$$

$$= \frac{1}{m}\left(\sum_{i=1}^{m-1}(E[T_i]+E[N]) + E[T_{m-1}]\right)$$

where the final equality used the independence of X and T_i. Therefore,

$$E[N] = E[T_{m-1}] + \sum_{i=1}^{m-1} E[T_i]$$

(c) $E[T_i] = \sum_{j=1}^{i} \frac{m}{m+1-j}$

(d)
$$E[N] = \sum_{j=1}^{m-1} \frac{m}{m+1-j} + \sum_{i=1}^{m-1}\sum_{j=1}^{i} \frac{m}{m+1-j}$$

$$= \sum_{j=1}^{m-1} \frac{m}{m+1-j} + \sum_{j=1}^{m-1}\sum_{i=j}^{m-1} \frac{m}{m+1-j}$$

$$= \sum_{j=1}^{m-1} \frac{m}{m+1-j} + \sum_{j=1}^{m-1} \frac{m(m-j)}{m+1-j}$$

$$= \sum_{j=1}^{m-1} \left(\frac{m}{m+1-j} + \frac{m(m-j)}{m+1-j} \right)$$
$$= m(m-1)$$

97. Let X be geometric with parameter p. To compute Var (X), we will use the conditional variance formula, conditioning on the outcome of the first trial. Let I equal 1 if the first trial is a success, and let it equal 0 otherwise. If $I = 1$, then $X = 1$; since the variance of a constant is 0, this gives

$$\text{Var}(X|I = 1) = 0$$

On the other hand, if $I = 0$ then the conditional distribution of X given that $I = 0$ is the same as the unconditional distribution of 1 (the first trial) plus a geometric with parameter p (the number of additional trials needed for a success). Therefore,

$$\text{Var}(X|I = 0) = \text{Var}(X)$$

yielding

$$E[\text{Var}(X|I)] = \text{Var}(X|I = 1)P(I = 1) + \text{Var}(X|I = 0)P(I = 0) = (1 - p)\text{Var}(X)$$

Similarly,

$$E[X|I = 1] = 1, \quad E[X|I = 0] = 1 + E[X] = 1 + \frac{1}{p}$$

which can be written as

$$E[X|I] = 1 + \frac{1}{p}(1 - I)$$

yielding

$$\text{Var}(E[X|I]) = \frac{1}{p^2}\text{Var}(I) = \frac{1}{p^2}p(1 - p) = \frac{1 - p}{p}$$

The conditional variance formula now gives

$$\text{Var}(X) = E[\text{Var}(X|I)] + \text{Var}(E[X|I])$$
$$= (1 - p)\text{Var}(X) + \frac{1 - p}{p}$$

or

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

Chapter 4

1. $P_{01} = 1, P_{10} = \frac{1}{9}, \quad P_{21} = \frac{4}{9}, \quad P_{32} = 1$
$P_{11} = \frac{4}{9}, \quad P_{22} = \frac{4}{9}$
$P_{12} = \frac{4}{9}, \quad P_{23} = \frac{1}{9}$

9. $P_{0,3}^{10} = .5078.$

16. If P_{ij} were (strictly) positive, then P_{ji}^n would be 0 for all n (otherwise, i and j would communicate). But then the process, starting in i, has a positive probability of at least P_{ij} of never returning to i. This contradicts the recurrence of i. Hence $P_{ij} = 0$.

21. The transition probabilities are

$$P_{i,j} = \begin{cases} 1 - 3\alpha, & \text{if } j = i \\ \alpha, & \text{if } j \neq i \end{cases}$$

By symmetry,

$$P_{ij}^n = \frac{1}{3}(1 - P_{ii}^n), \qquad j \neq i$$

So, let us prove by induction that

$$P_{i,j}^n = \begin{cases} \frac{1}{4} + \frac{3}{4}(1 - 4\alpha)^n & \text{if } j = i \\ \frac{1}{4} - \frac{1}{4}(1 - 4\alpha)^n & \text{if } j \neq i \end{cases}$$

As the preceding is true for $n = 1$, assume it for n. To complete the induction proof, we need to show that

$$P_{i,j}^{n+1} = \begin{cases} \frac{1}{4} + \frac{3}{4}(1 - 4\alpha)^{n+1} & \text{if } j = i \\ \frac{1}{4} - \frac{1}{4}(1 - 4\alpha)^{n+1} & \text{if } j \neq i \end{cases}$$

Now,

$$\begin{aligned} P_{i,i}^{n+1} &= P_{i,i}^n P_{i,i} + \sum_{j \neq i} P_{i,j}^n P_{j,i} \\ &= \left(\frac{1}{4} + \frac{3}{4}(1 - 4\alpha)^n\right)(1 - 3\alpha) + 3\left(\frac{1}{4} - \frac{1}{4}(1 - 4\alpha)^n\right)\alpha \\ &= \frac{1}{4} + \frac{3}{4}(1 - 4\alpha)^n(1 - 3\alpha - \alpha) \\ &= \frac{1}{4} + \frac{3}{4}(1 - 4\alpha)^{n+1} \end{aligned}$$

By symmetry, for $j \neq i$

$$P_{ij}^{n+1} = \frac{1}{3}(1 - P_{ii}^{n+1}) = \frac{1}{4} - \frac{1}{4}(1 - 4\alpha)^{n+1}$$

and the induction is complete.

By letting $n \to \infty$ in the preceding, or by using that the transition probability matrix is doubly stochastic, or by just using a symmetry argument, we obtain that $\pi_i = 1/4, i = 1, 2, 3, 4$.

27. **(a)** It is a Markov chain because each individual's state the next period depends only on its current state and not on any information about earlier times.

(b) If i of the N individuals are currently active, then the number of actives in the next period is the sum of two independent random variables; R_i, the number of the i currently active who remain active in the next period; and B_i, the number of the $N - i$ inactives who become active in the next period. Because R_i is binomial (i, α), and B_i is binomial $(N - i, 1 - \beta)$, we see that

$$E[X_n|X_{n-1} = i] = i\alpha + (N - i)(1 - \beta) = N(1 - \beta) + (\alpha + \beta - 1)i$$

Hence,

$$E[X_n|X_{n-1}] = N(1 - \beta) + (\alpha + \beta - 1)X_{n-1}$$

giving that

$$E[X_n] = N(1 - \beta) + (\alpha + \beta - 1)E[X_{n-1}]$$

Letting $a = N(1 - \beta), b = \alpha + \beta - 1$, the preceding gives

$$\begin{aligned} E[X_n] &= a + bE[X_{n-1}] \\ &= a + b(a + bE[X_{n-2}]) = a + ba + b^2E[X_{n-2}] \\ &= a + ba + b^2a + b^3E[X_{n-3}] \end{aligned}$$

Continuing this, we arrive at

$$E[X_n] = a\left(1 + b + \cdots + b^{n-1}\right) + b^n E\left[X_0\right]$$

Thus,

$$E[X_n|X_0 = i] = a\left(1 + b + \cdots + b^{n-1}\right) + b^n i$$

Note that

$$\lim_{n\to\infty} E[X_n] = \frac{a}{1 - b} = N\frac{1 - \beta}{2 - \alpha - \beta}$$

(c) With R_i, B_i as previously defined

$$P_{i,j} = P(R_i + B_i = j)$$

$$= \sum_k P(R_i + B_i = j | R_i = k) \binom{i}{k} \alpha^i (1-\alpha)^{i-k}$$
$$= \sum_k \binom{N-i}{j-k} (1-\beta)^{j-k} \beta^{N-i-j+k} \binom{i}{k} \alpha^i (1-\alpha)^{i-k}$$

where $\binom{m}{r} = 0$ if $r < 0$ or $r > m$.

(d) Suppose $N = 1$. Then, with 1 standing for active and 0 for inactive, the limiting probabilities are such that

$$\pi_0 = \pi_0\beta + \pi_1(1-\alpha)$$
$$\pi_1 = \pi_0(1-\beta) + \pi_1\alpha$$
$$\pi_0 + \pi_1 = 1$$

Solving yields

$$\pi_1 = \frac{1-\beta}{2-\alpha-\beta}, \quad \pi_0 = \frac{1-\alpha}{2-\alpha-\beta}$$

Now consider the case of population size N. Because each member will, in steady state, be active with probability π_1 and because each of the members changes states independently of each other it follows that the steady state number of actives has a binomial (N, π_1) distribution. Hence, the long-run proportion of time that exactly j people are active is

$$\pi_j(N) = \binom{N}{j} \left(\frac{1-\beta}{2-\alpha-\beta}\right)^j \left(\frac{1-\alpha}{2-\alpha-\beta}\right)^{N-j}$$

Note that the steady state expected number of actives is $N\frac{1-\alpha}{2-\alpha-\beta}$, in accord with what we saw in part (b).

32. With the state being the number of on switches this is a three-state Markov chain. The equations for the long-run proportions are

$$\pi_0 = \frac{9}{16}\pi_0 + \frac{1}{4}\pi_1 + \frac{1}{16}\pi_2,$$
$$\pi_1 = \frac{3}{8}\pi_0 + \frac{1}{2}\pi_1 + \frac{3}{8}\pi_2,$$
$$\pi_0 + \pi_1 + \pi_2 = 1$$

This gives the solution

$$\pi_0 = \frac{2}{7}, \qquad \pi_1 = \frac{3}{7}, \qquad \pi_2 = \frac{2}{7}$$

41. $$e_j = \sum_{i=0}^{j-1} P(\text{enters } j \text{ directly from } i) = \sum_{i=0}^{j-1} e_i P_{i,j}$$

$$e_1 = 1/3$$
$$e_2 = 1/3 + 1/3(1/3) = 4/9$$
$$e_3 = 1/3 + 1/3(1/3) + 4/9(1/3) = 16/27$$
$$e_4 = 1/3(1/3) + 4/9(1/3) + 16/27(1/3) = 37/81$$
$$e_5 = 4/9(1/3) + 16/27(1/3) + 37/81(1/3) = 158/243$$

47. $\{Y_n, n \geq 1\}$ is a Markov chain with states (i, j).

$$P_{(i,j),(k,l)} = \begin{cases} 0, & \text{if } j \neq k \\ P_{jl}, & \text{if } j = k \end{cases}$$

where P_{jl} is the transition probability for $\{X_n\}$.

$$\begin{aligned} \lim_{n\to\infty} P\{Y_n = (i, j)\} &= \lim_n P\{X_n = i, X_{n+1} = j\} \\ &= \lim_n [P\{X_n = i\} P_{ij}] \\ &= \pi_i P_{ij} \end{aligned}$$

62. It is easy to verify that the stationary probabilities are $\pi_i = \frac{1}{n+1}$. Hence, the mean time to return to the initial position is $n + 1$.

68. **(a)** $$\sum_i \pi_i Q_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j$$

(b) Whether perusing the sequence of states in the forward direction of time or in the reverse direction, the proportion of time the state is i will be the same.

Chapter 5

5. $P(Y = n) = P(n - 1 < X < n) = e^{-\lambda(n-1)} - e^{-\lambda n} = (e^{-\lambda})^{n-1}(1 - e^{-\lambda})$

7.
$$\begin{aligned} &P\{X_1 < X_2 | \min(X_1, X_2) = t\} \\ &= \frac{P\{X_1 < X_2, \min(X_1, X_2) = t\}}{P\{\min(X_1, X_2) = t\}} \\ &= \frac{P\{X_1 = t, X_2 > t\}}{P\{X_1 = t, X_2 > t\} + P\{X_2 = t, X_1 > t\}} \\ &= \frac{f_1(t)[1 - F_2(t)]}{f_1(t)[1 - F_2(t)] + f_2(t)[1 - F_1(t)]} \end{aligned}$$

Dividing through by $[1 - F_1(t)][1 - F_2(t)]$ yields the result. (Of course, f_i and F_i are the density and distribution function of $X_i, i = 1, 2$.) To make the preceding derivation rigorous, we should replace "$= t$" by $\in (t, t + \varepsilon)$ throughout and then let $\varepsilon \to 0$.

10. **(a)**
$$\begin{aligned} E[MX|M = X] &= E[M^2|M = X] \\ &= E[M^2] \\ &= \frac{2}{(\lambda + \mu)^2} \end{aligned}$$

(b) By the memoryless property of exponentials, given that $M = Y, X$ is distributed as $M + X'$ where X' is an exponential with rate λ that is independent of M. Therefore,

$$\begin{aligned} E[MX|M = Y] &= E[M(M + X')] \\ &= E[M^2] + E[M]E[X'] \\ &= \frac{2}{(\lambda + \mu)^2} + \frac{1}{\lambda(\lambda + \mu)} \end{aligned}$$

(c)
$$\begin{aligned} E[MX] &= E[MX|M = X]\frac{\lambda}{\lambda + \mu} + E[MX|M = Y]\frac{\mu}{\lambda + \mu} \\ &= \frac{2\lambda + \mu}{\lambda(\lambda + \mu)^2} \end{aligned}$$

Therefore,

$$\mathrm{Cov}(X, M) = \frac{\lambda}{\lambda(\lambda + \mu)^2}$$

18. **(a)** $1/(2\mu)$.

(b) $1/(4\mu^2)$, since the variance of an exponential is its mean squared.

(c), (d) By the lack of memory property of the exponential it follows that A, the amount by which $X_{(2)}$ exceeds $X_{(1)}$, is exponentially distributed with rate μ and is independent of $X_{(1)}$. Therefore,

$$\begin{aligned} E[X_{(2)}] &= E[X_{(1)} + A] = \frac{1}{2\mu} + \frac{1}{\mu} \\ \mathrm{Var}(X_{(2)}) &= \mathrm{Var}(X_{(1)} + A) = \frac{1}{4\mu^2} + \frac{1}{\mu^2} = \frac{5}{4\mu^2} \end{aligned}$$

23. **(a)** $\frac{1}{2}$.

(b) $(\frac{1}{2})^{n-1}$. Whenever battery 1 is in use and a failure occurs the probability is $\frac{1}{2}$ that it is not battery 1 that has failed.

(c) $(\frac{1}{2})^{n-i+1}, i > 1$.

(d) T is the sum of $n-1$ independent exponentials with rate 2μ (since each time a failure occurs the time until the next failure is exponential with rate 2μ).

(e) Gamma with parameters $n-1$ and 2μ.

36.
$$\begin{aligned} E[S(t)|N(t)=n] &= sE\left[\prod_{i=1}^{N(t)} X_i | N(t)=n\right] \\ &= sE\left[\prod_{i=1}^{n} X_i | N(t)=n\right] \\ &= sE\left[\prod_{i=1}^{n} X_i\right] \\ &= s(E[X])^n \\ &= s(1/\mu)^n \end{aligned}$$

Thus,

$$\begin{aligned} E[S(t)] &= s\sum_n (1/\mu)^n e^{-\lambda t}(\lambda t)^n/n! \\ &= se^{-\lambda t}\sum_n (\lambda t/\mu)^n/n! \\ &= se^{-\lambda t+\lambda t/\mu} \end{aligned}$$

By the same reasoning

$$E[S^2(t)|N(t)=n] = s^2(E[X^2])^n = s^2(2/\mu^2)^n$$

and

$$E[S^2(t)] = s^2 e^{-\lambda t+2\lambda t/\mu^2}$$

40. The easiest way is to use Definition 5.3. It is easy to see that $\{N(t), t \geq 0\}$ will also possess stationary and independent increments. Since the sum of two independent Poisson random variables is also Poisson, it follows that $N(t)$ is a Poisson random variable with mean $(\lambda_1+\lambda_2)t$.

57. **(a)** e^{-2}.

(b) 2 P.M.

(c) $1-5e^{-4}$.

60. **(a)** $\frac{1}{9}$.

(b) $\frac{5}{9}$.

64. **(a)** Since, given $N(t)$, each arrival is uniformly distributed on $(0,t)$ it follows that

$$E[X|N(t)] = N(t)\int_0^t (t-s)\frac{ds}{t} = N(t)\frac{t}{2}$$

(b) Let $U_1, U_2, \ldots$ be independent uniform $(0, t)$ random variables. Then

$$\begin{aligned}\text{Var}(X|N(t) = n) &= \text{Var}\left[\sum_{i=1}^{n}(t - U_i)\right] \\ &= n\ \text{Var}(U_i) = n\frac{t^2}{12}\end{aligned}$$

(c) By parts (a) and (b) and the conditional variance formula,

$$\begin{aligned}\text{Var}(X) &= \text{Var}\left(\frac{N(t)t}{2}\right) + E\left[\frac{N(t)t^2}{12}\right] \\ &= \frac{\lambda t t^2}{4} + \frac{\lambda t t^2}{12} = \frac{\lambda t^3}{3}\end{aligned}$$

79. It is a nonhomogeneous Poisson process with intensity function $p(t)\lambda(t)$, $t > 0$.

84. There is a record whose value is between t and $t + dt$ if the first X larger than t lies between t and $t + dt$. From this we see that, independent of all record values less than t, there will be one between t and $t + dt$ with probability $\lambda(t)\, dt$ where $\lambda(t)$ is the failure rate function given by

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

Since the counting process of record values has, by the preceding, independent increments we can conclude (since there cannot be multiple record values because the X_i are continuous) that it is a nonhomogeneous Poisson process with intensity function $\lambda(t)$. When f is the exponential density, $\lambda(t) = \lambda$ and so the counting process of record values becomes an ordinary Poisson process with rate λ.

91. To begin, note that

$$\begin{aligned}P\left\{X_1 > \sum_{2}^{n} X_i\right\} &= P\{X_1 > X_2\}P\{X_1 - X_2 > X_3 | X_1 > X_2\} \\ &\quad \times P\{X_1 - X_2 - X_3 > X_4 | X_1 > X_2 + X_3\} \cdots \\ &\quad \times P\{X_1 - X_2 \cdots - X_{n-1} > X_n | X_1 > X_2 + \cdots + X_{n-1}\} \\ &= \left(\frac{1}{2}\right)^{n-1} \qquad \text{by lack of memory}\end{aligned}$$

Hence,

$$P\left\{M > \sum_{i=1}^{n} X_i - M\right\} = \sum_{i=1}^{n} P\left\{X_i > \sum_{j \neq i} X_j\right\} = \frac{n}{2^{n-1}}$$

Chapter 6

2. Let $N_A(t)$ be the number of organisms in state A and let $N_B(t)$ be the number of organisms in state B. Then $\{N_A(t), N_B(t)\}$ is a continuous-Markov chain with

$$\begin{aligned} v_{\{n,m\}} &= \alpha n + \beta m \\ P_{\{n,m\},\{n-1,m+1\}} &= \frac{\alpha n}{\alpha n + \beta m} \\ P_{\{n,m\},\{n+2,m-1\}} &= \frac{\beta m}{\alpha n + \beta m} \end{aligned}$$

4. Let $N(t)$ denote the number of customers in the station at time t. Then $\{N(t)\}$ is a birth and death process with

$$\lambda_n = \lambda\alpha_n, \qquad \mu_n = \mu$$

7. **(a)** Yes!

(b) For $\mathbf{n} = (n_1, \ldots, n_i, n_{i+1}, \ldots, n_{k-1})$ let

$$\begin{aligned} S_i(\mathbf{n}) &= (n_1, \ldots, n_i - 1, n_{i+1} + 1, \ldots, n_{k-1}), \qquad i = 1, \ldots, k-2 \\ S_{k-1}(\mathbf{n}) &= (n_1, \ldots, n_i, n_{i+1}, \ldots, n_{k-1} - 1), \\ S_0(\mathbf{n}) &= (n_1 + 1, \ldots, n_i, n_{i+1}, \ldots, n_{k-1}). \end{aligned}$$

Then

$$\begin{aligned} q_{\mathbf{n},S_i(\mathbf{n})} &= n_i\mu, \qquad i = 1, \ldots, k-1 \\ q_{\mathbf{n},S_0(n)} &= \lambda \end{aligned}$$

11. **(b)** Follows from the hint about using the lack of memory property and the fact that ε_i, the minimum of $j - (i-1)$ independent exponentials with rate λ, is exponential with rate $(j - i + 1)\lambda$.

(c) From parts (a) and (b)

$$P\{T_1 + \cdots + T_j \le t\} = P\left\{\max_{1\le i\le j} X_i \le t\right\} = (1 - e^{-\lambda t})^j$$

(d) With all probabilities conditional on $X(0) = 1$,

$$\begin{aligned} P_{1j}(t) &= P\{X(t) = j\} \\ &= P\{X(t) \ge j\} - P\{X(t) \ge j + 1\} \\ &= P\{T_1 + \cdots + T_j \le t\} - P\{T_1 + \cdots + T_{j+1} \le t\} \end{aligned}$$

(e) The sum of i independent geometrics, each having parameter $p = e^{-\lambda t}$, is a negative binomial with parameters i, p. The result follows since starting with an initial population of i is equivalent to having i independent Yule processes, each starting with a single individual.

16. Let the state be

2: an acceptable molecule is attached
0: no molecule attached
1: an unacceptable molecule is attached.

Then, this is a birth and death process with balance equations

$$\mu_1 P_1 = \lambda(1-\alpha)P_0$$
$$\mu_2 P_2 = \lambda\alpha P_0$$

Since $\sum_0^2 P_i = 1$, we get

$$P_2 = \left[1 + \frac{\mu_2}{\lambda\alpha} + \frac{1-\alpha}{\alpha}\frac{\mu_2}{\mu_1}\right]^{-1} = \frac{\lambda\alpha\mu_1}{\lambda\alpha\mu_1 + \mu_1\mu_2 + \lambda(1-\alpha)\mu_2}$$

where P_2 is the percentage of time the site is occupied by an acceptable molecule. The percentage of time the site is occupied by an unacceptable molecule is

$$P_1 = \frac{1-\alpha}{\alpha}\frac{\mu_2}{\mu_1}P_1 = \frac{\lambda(1-\alpha)\mu_2}{\lambda\alpha\mu_1 + \mu_1\mu_2 + \lambda(1-\alpha)\mu_2}$$

19. There are four states. Let state 0 mean that no machines are down, state 1 that machine 1 is down and 2 is up, state 2 that machine 1 is up and 2 is down, and state 3 that both machines are down. The balance equations are as follows:

$$\begin{aligned}
(\lambda_1+\lambda_2)P_0 &= \mu_1 P_1 + \mu_2 P_2\\
(\mu_1+\lambda_2)P_1 &= \lambda_1 P_0\\
(\lambda_1+\mu_2)P_2 &= \lambda_2 P_0 + \mu_1 P_3\\
\mu_1 P_3 &= \lambda_2 P_1 + \lambda_1 P_2\\
P_0 + P_1 + P_2 + P_3 &= 1
\end{aligned}$$

The equations are easily solved and the proportion of time machine 2 is down is $P_2 + P_3$.

24. We will let the state be the number of taxis waiting. Then, we get a birth and death process with $\lambda_n = 1, \mu_n = 2$. This is an $M/M/1$. Therefore:

(a) Average number of taxis waiting $= \dfrac{1}{\mu-\lambda} = \dfrac{1}{2-1} = 1.$

(b) The proportion of arriving customers that gets taxis is the proportion of arriving customers that find at least one taxi waiting. The rate of arrival of such customers is $2(1-P_0)$. The proportion of such arrivals is therefore

$$\frac{2(1-P_0)}{2} = 1 - P_0 = 1 - \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu} = \frac{1}{2}$$

28. Let P_{ij}^x, v_i^x denote the parameters of the $X(t)$ and P_{ij}^y, v_i^y of the $Y(t)$ process; and let the limiting probabilities be P_i^x, P_i^y, respectively. By independence we have that for the Markov chain $\{X(t), Y(t)\}$ its parameters are

$$\begin{aligned} v_{(i,l)} &= v_i^x + v_l^y, \\ P_{(i,l)(j,l)} &= \frac{v_i^x}{v_i^x + v_l^y} P_{ij}^x, \\ P_{(i,l)(i,k)} &= \frac{v_l^y}{v_i^x + v_l^y} P_{lk}^y, \end{aligned}$$

and

$$\lim_{t\to\infty} P\{(X(t), Y(t)) = (i, j)\} = P_i^x P_j^y$$

Hence, we need to show that

$$P_i^x P_l^y v_i^x P_{ij}^x = P_j^x P_l^y v_j^x P_{ji}^x$$

(That is, the rate from (i, l) to (j, l) equals the rate from (j, l) to (i, l).) But this follows from the fact that the rate from i to j in $X(t)$ equals the rate from j to i; that is,

$$P_i^x v_i^x P_{ij}^x = P_j^x v_j^x P_{ji}^x$$

The analysis is similar in looking at pairs (i, l) and (i, k).

33. Suppose first that the waiting room is of infinite size. Let $X_i(t)$ denote the number of customers at server $i, i = 1, 2$. Then since each of the $M/M/1$ processes $\{X_1(t)\}$ is time reversible, it follows from Exercise 28 that the vector process $\{(X_1(t), (X(t)), t \geq 0\}$ is a time reversible Markov chain. Now the process of interest is just the truncation of this vector process to the set of states A where

$$A = \{(0, m): m \leq 4\} \cup \{(n, 0): n \leq 4\} \cup \{(n, m): nm > 0, n + m \leq 5\}$$

Hence, the probability that there are n with server 1 and m with server 2 is

$$\begin{aligned} P_{n,m} &= k \left(\frac{\lambda_1}{\mu_1}\right)^n \left(1 - \frac{\lambda_1}{\mu_1}\right) \left(\frac{\lambda_2}{\mu_2}\right)^m \left(1 - \frac{\lambda_2}{\mu_2}\right) \\ &= C \left(\frac{\lambda_1}{\mu_1}\right)^n \left(\frac{\lambda_2}{\mu_2}\right)^m, \qquad (n, m) \in A \end{aligned}$$

The constant C is determined from

$$\sum P_{n,m} = 1$$

where the sum is over all (n, m) in A.

40. The time reversible equations are

$$P(i)\frac{v_i}{n-1} = P(j)\frac{v_j}{n-1}$$

yielding the solution

$$P(j) = \frac{1/v_j}{\sum_{i=1}^{n} 1/v_i}$$

Hence, the chain is time reversible with long run proportions given by the preceding.

41. Show in Example 6.22 that the limiting probabilities satisfy Eqs. (6.33), (6.34), and (6.35).

49. **(a)** The matrix $\mathbf{P}^*$ can be written as

$$\mathbf{P}^* = \mathbf{I} + \mathbf{R}/v$$

and so P_{ij}^{*n} can be obtained by taking the i, j element of $(\mathbf{I}+\mathbf{R}/v)^n$, which gives the result when $v = n/t$.

(b) Uniformization shows that $P_{ij}(t) = E[P_{ij}^{*N}]$, where N is independent of the Markov chain with transition probabilities P_{ij}^* and is Poisson distributed with mean vt. Since a Poisson random variable with mean vt has standard deviation $(vt)^{1/2}$, it follows that for large values of vt it should be near vt. (For instance, a Poisson random variable with mean 10^6 has standard deviation 10^3 and thus will, with high probability, be within 3000 of 10^6.) Hence, since for fixed i and j, P_{ij}^{*m} should not vary much for values of m about vt where vt is large, it follows that, for large vt,

$$E[P_{ij}^{*N}] \approx P_{ij}^{*n} \qquad \text{where } n = vt$$

Chapter 7

6. **(a)** Consider a Poisson process having rate λ and say that an event of the renewal process occurs whenever one of the events numbered $r, 2r, 3r, \ldots$ of the Poisson process occurs. Then

$$\begin{aligned} P\{N(t) \geq n\} &= P\{nr \text{ or more Poisson events by } t\} \\ &= \sum_{i=nr}^{\infty} e^{-\lambda t}(\lambda t)^i / i! \end{aligned}$$

(b)

$$\begin{aligned} E[N(t)] &= \sum_{n=1}^{\infty} P\{N(t) \geq n\} = \sum_{n=1}^{\infty} \sum_{i=nr}^{\infty} e^{-\lambda t}(\lambda t)^i / i! \\ &= \sum_{i=r}^{\infty} \sum_{n=1}^{[i/r]} e^{-\lambda t}(\lambda t)^i / i! = \sum_{i=r}^{\infty} [i/r] e^{-\lambda t}(\lambda t)^i / i! \end{aligned}$$

8. **(a)** The number of replaced machines by time t constitutes a renewal process. The time between replacements equals T, if the lifetime of new machine is $\geq T$; x, if the lifetime of new machine is $x, x < T$. Hence,

$$E[\text{time between replacements}] = \int_0^T xf(x)\,dx + T[1 - F(T)]$$

and the result follows by Proposition 7.1.

(b) The number of machines that have failed in use by time t constitutes a renewal process. The mean time between in-use failures, $E[F]$, can be calculated by conditioning on the lifetime of the initial machine as $E[F] = E[E[F|\text{ lifetime of initial machine}]]$. Now

$$E[F|\text{lifetime of machine is } x] = \begin{cases} x, & \text{if } x \leq T \\ T + E[F], & \text{if } x > T \end{cases}$$

Hence,

$$E[F] = \int_0^T xf(x)\,dx + (T + E[F])[1 - F(T)]$$

or

$$E[F] = \frac{\int_0^T xf(x)\,dx + T[1 - F(T)]}{F(T)}$$

and the result follows from Proposition 7.1.

18. We can imagine that a renewal corresponds to a machine failure, and each time a new machine is put in use its life distribution will be exponential with rate μ_1 with probability p, and exponential with rate μ_2 otherwise. Hence, if our state is the index of the exponential life distribution of the machine presently in use, then this is a two-state continuous-time Markov chain with intensity rates

$$q_{1,2} = \mu_1(1 - p), \qquad q_{2,1} = \mu_2 p$$

Hence,

$$P_{11}(t) = \frac{\mu_1(1-p)}{\mu_1(1-p) + \mu_2 p} \exp\{-[\mu_1(1-p) + \mu_2 p]t\} + \frac{\mu_2 p}{\mu_1(1-p) + \mu_2 p}$$

with similar expressions for the other transition probabilities ($P_{12}(t) = 1 - P_{11}(t)$, and $P_{22}(t)$ is the same with $\mu_2 p$ and $\mu_1(1-p)$ switching places). Conditioning on the initial machine now gives

$$
\begin{aligned}
E[Y(t)] &= pE[Y(t)|X(0)=1] + (1-p)E[Y(t)|X(0)=2] \\
&= p\left[\frac{P_{11}(t)}{\mu_1} + \frac{P_{12}(t)}{\mu_2}\right] + (1-p)\left[\frac{P_{21}(t)}{\mu_1} + \frac{P_{22}(t)}{\mu_2}\right]
\end{aligned}
$$

Finally, we can obtain $m(t)$ from

$$\mu[m(t)+1] = t + E[Y(t)]$$

where

$$\mu = p/\mu_1 + (1-p)/\mu_2$$

is the mean interarrival time.

22. **(a)** Let X denote the length of time that J keeps a car. Let I equal 1 if there is a breakdown by time T and equal 0 otherwise. Then

$$
\begin{aligned}
E[X] &= E[X|I=1](1-e^{-\lambda T}) + E[X|I=0]e^{-\lambda T} \\
&= \left(T + \frac{1}{\mu}\right)(1-e^{-\lambda T}) + \left(T + \frac{1}{\lambda}\right)e^{-\lambda T} \\
&= T + \frac{1-e^{-\lambda T}}{\mu} + \frac{e^{-\lambda T}}{\lambda}
\end{aligned}
$$

$1/E[X]$ is the rate that J buys a new car.

(b) Let W equal to the total cost involved with purchasing a car. Then, with Y equal to the time of the first breakdown

$$
\begin{aligned}
E[W] &= \int_0^\infty E[W|Y=y]\lambda e^{-\lambda y}dy \\
&= C + \int_0^T r(1+\mu(T-y)+1)\lambda e^{-\lambda y}dy + \int_T^\infty r\lambda e^{-\lambda y}\,dy \\
&= C + r(2-e^{-\lambda T}) + r\int_0^T \mu(T-y)\lambda e^{-\lambda y}\,dy
\end{aligned}
$$

J's long run average cost is $E[W]/E[X]$.

30.

$$
\begin{aligned}
\frac{A(t)}{t} &= \frac{t - S_{N(t)}}{t} \\
&= 1 - \frac{S_{N(t)}}{t} \\
&= 1 - \frac{S_{N(t)}}{N(t)}\frac{N(t)}{t}
\end{aligned}
$$

The result follows since $S_{N(t)}/N(t) \to \mu$ (by the strong law of large numbers) and $N(t)/t \to 1/\mu$.

35. **(a)** We can view this as an $M/G/\infty$ system where a satellite launching corresponds to an arrival and F is the service distribution. Hence,

$$P\{X(t) = k\} = e^{-\lambda(t)}[\lambda(t)]^k/k!$$

where $\lambda(t) = \lambda \int_0^t (1 - F(s))\, ds$.

(b) By viewing the system as an alternating renewal process that is on when there is at least one satellite orbiting, we obtain

$$\lim P\{X(t) = 0\} = \frac{1/\lambda}{1/\lambda + E[T]}$$

where T, the on time in a cycle, is the quantity of interest. From part (a)

$$\lim P\{X(t) = 0\} = e^{-\lambda\mu}$$

where $\mu = \int_0^\infty (1 - F(s))\, ds$ is the mean time that a satellite orbits. Hence,

$$e^{-\lambda\mu} = \frac{1/\lambda}{1/\lambda + E[T]}$$

so

$$E[T] = \frac{1 - e^{-\lambda\mu}}{\lambda e^{-\lambda\mu}}$$

46. **(a)** $F_e(x) = \dfrac{1}{\mu} \displaystyle\int_0^x e^{-y/\mu}\, dy = 1 - e^{-x/\mu}.$

(b) $F_e(x) = \dfrac{1}{c} \displaystyle\int_0^x dy = \frac{x}{c}, \qquad 0 \le x \le c.$

(c) You will receive a ticket if, starting when you park, an official appears within one hour. From Example 7.27 the time until the official appears has the distribution F_e which, by part (a), is the uniform distribution on $(0, 2)$. Thus, the probability is equal to $\frac{1}{2}$.

48. **(a)** Let N_i denote the number of passengers that get on bus i. If we interpret X_i as the reward incurred at time i then we have a renewal reward process whose ith cycle is of length N_i, and has reward $X_{N_1+\cdots+N_{i-1}+1} + \cdots + X_{N_1+\cdots+N_i}$. Hence, part (a) follows because N is the time and $X_1 + \cdots + X_N$ is the cost of the first cycle.

(b) Condition on $N(t)$ and use that conditional on $N(t) = n$ the n arrival times are independently and uniformly distributed on $(0, t)$. As $S \equiv X_1 + \cdots + X_N$ is the number of these n passengers whose waiting time is less than x, this gives

$$E[S|T=t, N(t)=n] = \begin{cases} nx/t, & \text{if } x < t \\ n, & \text{if } x > t \end{cases}$$

That is, $E[S|T=t, N(t)] = N(t)\min(x,t)/t$. Taking expectations yields

$$E[S|T=t] = \lambda \min(x,t)$$

(c) From (b), $E[S|T] = \lambda \min(x, T)$ and (c) follows upon taking expectations.

(d) This follows from parts (a) and (c) using that

$$E[\min(x,T)] = \int_0^\infty P(\min(x,T) > t)dt = \int_0^x P(T > t)dt$$

along with the identity $E[N] = \lambda E[T]$.

(e) Because the waiting time for an arrival is the time until the next bus, the preceding result yields the PASTA result that the proportion of arrivals who see the excess life of the renewal process of bus arrivals to be less than x is equal to the proportion of time it is less than x.

53. Think of each interarrival time as consisting of n independent phases—each of which is exponentially distributed with rate λ—and consider the semi-Markov process whose state at any time is the phase of the present interarrival time. Hence, this semi-Markov process goes from state 1 to 2 to 3 ... to n to 1, and so on. Also the time spent in each state has the same distribution. Thus, clearly the limiting probability of this semi-Markov chain is $P_i = 1/n, i = 1, \ldots, n$. To compute $\lim P\{Y(t) < x\}$, we condition on the phase at time t and note that if it is $n - i + 1$, which will be the case with probability $1/n$, then the time until a renewal occurs will be sum of i exponential phases, which will thus have a gamma distribution with parameters i and λ.

Chapter 8

2. This problem can be modeled by an $M/M/1$ queue in which $\lambda = 6, \mu = 8$. The average cost rate will be

$$\$10 \text{ per hour per machine} \times \text{average number of broken machines}$$

The average number of broken machines is just L, which can be computed from Eq. (3.2):

$$L = \frac{\lambda}{\mu - \lambda}$$
$$= \frac{6}{2} = 3$$

Hence, the average cost rate = \$30/hour.

8. To compute W for the $M/M/2$, set up balance equations as follows:

$$\begin{aligned}
\lambda P_0 &= \mu P_1 \qquad \text{(each server has rate } \mu\text{)} \\
(\lambda+\mu)P_1 &= \lambda P_0 + 2\mu P_2 \\
(\lambda+2\mu)P_n &= \lambda P_{n-1} + 2\mu P_{n+1}, \qquad n \geq 2
\end{aligned}$$

These have solutions $P_n = (\rho^n/2^{n-1})P_0$ where $\rho = \lambda/\mu$. The boundary condition $\sum_{n=0}^{\infty} P_n = 1$ implies

$$P_0 = \frac{1-\rho/2}{1+\rho/2} = \frac{(2-\rho)}{(2+\rho)}$$

Now we have P_n, so we can compute L, and hence W from $L = \lambda W$:

$$\begin{aligned}
L &= \sum_{n=0}^{\infty} n P_n = \rho P_0 \sum_{n=0}^{\infty} n \left(\frac{\rho}{2}\right)^{n-1} \\
&= 2P_0 \sum_{n=0}^{\infty} n \left(\frac{\rho}{2}\right)^{n} \\
&= 2\frac{(2-\rho)}{(2+\rho)} \frac{(\rho/2)}{(1-\rho/2)^2} \qquad \text{(See derivation of Eq. (8.7).)} \\
&= \frac{4\rho}{(2+\rho)(2-\rho)} \\
&= \frac{4\mu\lambda}{(2\mu+\lambda)(2\mu-\lambda)}
\end{aligned}$$

From $L = \lambda W$ we have

$$W = W(M/M/2) = \frac{4\mu}{(2\mu+\lambda)(2\mu-\lambda)}$$

The $M/M/1$ queue with service rate 2μ has

$$W(M/M/1) = \frac{1}{2\mu-\lambda}$$

from Eq. (8.8). We assume that in the $M/M/1$ queue, $2\mu > \lambda$ so that the queue is stable. But then $4\mu > 2\mu + \lambda$, or $4\mu/(2\mu+\lambda) > 1$, which implies $W(M/M/2) > W(M/M/1)$. The intuitive explanation is that if one finds the queue empty in the $M/M/2$ case, it would do no good to have two servers. One would be better off with one faster server. Now let $W_Q^1 = W_Q(M/M/1)$ and $W_Q^2 = W_Q(M/M/2)$. Then,

$$W_Q^1 = W(M/M/1) - 1/2\mu$$
$$W_Q^2 = W(M/M/2) - 1/\mu$$

So,

$$W_Q^1 = \frac{\lambda}{2\mu(2\mu - \lambda)} \qquad \text{from Eq. (8.8)}$$

and

$$W_Q^2 = \frac{\lambda^2}{\mu(2\mu - \lambda)(2\mu + \lambda)}$$

Then,

$$W_Q^1 > W_Q^2 \Leftrightarrow \frac{1}{2} > \frac{\lambda}{(2\mu + \lambda)}$$
$$\lambda < 2\mu$$

Since we assume $\lambda < 2\mu$ for stability in the $M/M/1$ case, $W_Q^2 < W_Q^1$ whenever this comparison is possible, that is, whenever $\lambda < 2\mu$.

13. Let the state be (n, m) if there are n families and m taxis waiting, $nm = 0$. The time reversibility equations are

$$P_{n-1,0}\lambda = P_{n,0}\mu, \quad n = 1, \ldots, N$$
$$P_{0,m-1}\mu = P_{0,m}\lambda, \quad m = 1, \ldots, M$$

Solving yields

$$P_{n,0} = (\lambda/\mu)^n P_{0,0}, \quad n = 0, 1, \ldots, N$$
$$P_{0,m} = (\mu/\lambda)^m P_{0,0}, \quad m = 0, 1, \ldots, M$$

where

$$\frac{1}{P_{0,0}} = \sum_{n=0}^{N} (\lambda/\mu)^n + \sum_{m=1}^{M} (\mu/\lambda)^m$$

(a) $\sum_{m=0}^{M} P_{0,m}$

(b) $\sum_{n=0}^{N} P_{n,0}$

(c) $\frac{\sum_{n=0}^{N} n P_{n,0}}{\lambda(1-P_{N,0})}$

(d) $\frac{\sum_{m=0}^{M} m P_{0,m}}{\mu(1-P_{0,M})}$

(e) $1 - P_{N,0}$

When $N = M = \infty$ the time reversibility equations become

$$P_{n-1,0}\lambda = P_{n,0}(\mu + n\alpha), \quad n \geq 1$$

$$P_{0,m-1}\mu = P_{0,m}(\lambda + m\beta), \quad m \geq 1$$

which yields

$$P_{n,0} = P_{0,0}\prod_{i=1}^{n}\frac{\lambda}{\mu + i\alpha}, \quad n \geq 1$$

$$P_{0,m} = P_{0,0}\prod_{i=1}^{m}\frac{\mu}{\lambda + i\beta}, \quad m \geq 1$$

The rest is similar to the preceding.

25. **(a)** $\lambda_1 P_{10}$.

(b) $\lambda_2(P_0 + P_{10})$.

(c) $\lambda_1 P_{10}/[\lambda_1 P_{10} + \lambda_2(P_0 + P_{10})]$.

(d) This is equal to the fraction of server 2's customers that are type 1 multiplied by the proportion of time server 2 is busy. (This is true since the amount of time server 2 spends with a customer does not depend on which type of customer it is.) By (c) the answer is thus

$$\frac{(P_{01} + P_{11})\lambda_1 P_{10}}{\lambda_1 P_{10} + \lambda_2(P_0 + P_{10})}$$

28. The states are now $n, n \geq 0$, and $n', n \geq 1$ where the state is n when there are n in the system and no breakdown, and n' when there are n in the system and a breakdown is in progress. The balance equations are

$$\begin{aligned}
\lambda P_0 &= \mu P_1 \\
(\lambda + \mu + \alpha)P_n &= \lambda P_{n-1} + \mu P_{n+1} + \beta P_{n'}, \qquad n \geq 1 \\
(\beta + \lambda)P_{1'} &= \alpha P_1 \\
(\beta + \lambda)P_{n'} &= \alpha P_n + \lambda P_{(n-1)'}, \qquad n \geq 2 \\
\sum_{n=0}^{\infty} P_n + \sum_{n=1}^{\infty} P_{n'} &= 1
\end{aligned}$$

In terms of the solution to the preceding,

$$L = \sum_{n=1}^{\infty} n(P_n + P_{n'})$$

and so

$$W = \frac{L}{\lambda_a} = \frac{L}{\lambda}$$

32. If a customer leaves the system busy, the time until the next departure is the time of a service. If a customer leaves the system empty, the time until the next departure is the time until an arrival *plus* the time of a service.

Using moment generating functions we get

$$\begin{aligned} E\left\{e^{sD}\right\} &= \frac{\lambda}{\mu} E\{e^{sD} | \text{system left busy}\} \\ &\quad + \left(1 - \frac{\lambda}{\mu}\right) E\{e^{sD} | \text{system left empty}\} \\ &= \left(\frac{\lambda}{\mu}\right)\left(\frac{\mu}{\mu - s}\right) + \left(1 - \frac{\lambda}{\mu}\right) E\{e^{s(X+Y)}\} \end{aligned}$$

where X has the distribution of interarrival times, Y has the distribution of service times, and X and Y are independent. Then

$$\begin{aligned} E\left[e^{s(X+Y)}\right] &= E\left[e^{sX} e^{sY}\right] \\ &= E\left[e^{sX}\right] E\left[e^{sY}\right] \qquad \text{by independence} \\ &= \left(\frac{\lambda}{\lambda - s}\right)\left(\frac{\mu}{\mu - s}\right) \end{aligned}$$

So,

$$\begin{aligned} E\left\{e^{sD}\right\} &= \left(\frac{\lambda}{\mu}\right)\left(\frac{\mu}{\mu - s}\right) + \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\lambda - s}\right)\left(\frac{\mu}{\mu - s}\right) \\ &= \frac{\lambda}{(\lambda - s)} \end{aligned}$$

By the uniqueness of generating functions, it follows that D has an exponential distribution with parameter λ.

40. The distributions of the queue size and busy period are the same for all three disciplines; that of the waiting time is different. However, the means are identical. This can be seen by using $W = L/\lambda$, since L is the same for all. The smallest variance in the waiting time occurs under first-come, first-served and the largest under last-come, first-served.

43. **(a)** $a_0 = P_0$ due to Poisson arrivals. Assuming that each customer pays 1 per unit time while in service the cost identity of Eq. (8.1) states that

$$\text{average number in service} = \lambda E[S]$$

or

$$1 - P_0 = \lambda E[S]$$

(b) Since a_0 is the proportion of arrivals that have service distribution G_1 and $1 - a_0$ the proportion having service distribution G_2, the result follows.

(c) We have

$$P_0 = \frac{E[I]}{E[I] + E[B]}$$

and $E[I] = 1/\lambda$ and thus,

$$E[B] = \frac{1 - P_0}{\lambda P_0}$$
$$= \frac{E[S]}{1 - \lambda E[S]}$$

Now from parts (a) and (b) we have

$$E[S] = (1 - \lambda E[S])E[S_1] + \lambda E[S]E[S_2]$$

or

$$E[S] = \frac{E[S_1]}{1 + \lambda E[S_1] + \lambda E[S_2]}$$

Substituting into $E[B] = E[S]/(1 - \lambda E[S])$ now yields the result.

(d) $a_0 = 1/E[C]$, implying that

$$E[C] = \frac{E[S_1] + 1/\lambda - E[S_2]}{1/\lambda - E[S_2]}$$

49. By regarding any breakdowns that occur during a service as being part of that service, we see that this is an $M/G/1$ model. We need to calculate the first two moments of a service time. Now the time of a service is the time T until something happens (either a service completion or a breakdown) plus any additional time A. Thus,

$$E[S] = E[T + A]$$
$$= E[T] + E[A]$$

To compute $E[A]$, we condition upon whether the happening is a service or a breakdown. This gives

$$E[A] = E[A|\text{service}]\frac{\mu}{\mu + \alpha} + E[A|\text{breakdown}]\frac{\alpha}{\mu + \alpha}$$
$$= E[A|\text{breakdown}]\frac{\alpha}{\mu + \alpha}$$
$$= \left(\frac{1}{\beta} + E[S]\right)\frac{\alpha}{\mu + \alpha}$$

Since $E[T] = 1/(\alpha + \mu)$ we obtain

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

or

$$E[S] = \frac{1}{\mu} + \frac{\alpha}{\mu\beta}$$

We also need $E[S^2]$, which is obtained as follows:

$$\begin{aligned} E[S^2] &= E[(T+A)^2] \\ &= E[T^2] + 2E[AT] + E[A^2] \\ &= E[T^2] + 2E[A]E[T] + E[A^2] \end{aligned}$$

The independence of A and T follows because the time of the first happening is independent of whether the happening was a service or a breakdown. Now,

$$\begin{aligned} E[A^2] &= E[A^2|\text{breakdown}]\frac{\alpha}{\mu+\alpha} \\ &= \frac{\alpha}{\mu+\alpha} E[(\text{downtime} + S^*)^2] \\ &= \frac{\alpha}{\mu+\alpha}\left\{E[\text{down}^2] + 2E[\text{down}]E[S] + E[S^2]\right\} \\ &= \frac{\alpha}{\mu+\alpha}\left\{\frac{2}{\beta^2} + \frac{2}{\beta}\left[\frac{1}{\mu} + \frac{\alpha}{\mu\beta}\right] + E[S^2]\right\} \end{aligned}$$

Hence,

$$\begin{aligned} E[S^2] = {} & \frac{2}{(\mu+\beta)^2} + 2\left[\frac{\alpha}{\beta(\mu+\alpha)} + \frac{\alpha}{\mu+\alpha}\left(\frac{1}{\mu} + \frac{\alpha}{\mu\beta}\right)\right] \\ & + \frac{\alpha}{\mu+\alpha}\left\{\frac{2}{\beta^2} + \frac{2}{\beta}\left[\frac{1}{\mu} + \frac{\alpha}{\mu\beta}\right] + E[S^2]\right\} \end{aligned}$$

Now solve for $E[S^2]$. The desired answer is

$$W_Q = \frac{\lambda E[S^2]}{2(1-\lambda E[S])}$$

In the preceding, S^* is the additional service needed after the breakdown is over and S^* has the same distribution as S. The preceding also uses the fact that the expected square of an exponential is twice the square of its mean.
Another way of calculating the moments of S is to use the representation

$$S = \sum_{i=1}^{N}(T_i + B_i) + T_{N+1}$$

where N is the number of breakdowns while a customer is in service, T_i is the time starting when service commences for the ith time until a happening occurs, and B_i is the length of the ith breakdown. We now use the fact that, given N, all of the random variables in the representation are independent exponentials with the T_i having rate $\mu+\alpha$ and the B_i having rate β. This yields

$$E[S|N] = \frac{N+1}{\mu+\alpha} + \frac{N}{\beta},$$

$$\text{Var}(S|N) = \frac{N+1}{(\mu+\alpha)^2} + \frac{N}{\beta^2}$$

Therefore, since $1+N$ is geometric with mean $(\mu+\alpha)/\mu$ (and variance $\alpha(\alpha+\mu)/\mu^2$) we obtain

$$E[S] = \frac{1}{\mu} + \frac{\alpha}{\mu\beta}$$

and, using the conditional variance formula,

$$\text{Var}(S) = \left[\frac{1}{\mu+\alpha} + \frac{1}{\beta}\right]^2 \frac{\alpha(\alpha+\mu)}{\mu^2} + \frac{1}{\mu(\mu+\alpha)} + \frac{\alpha}{\mu\beta^2}$$

56. S_n is the service time of the nth customer; T_n is the time between the arrival of the nth and $(n+1)$st customer.

Chapter 9

4. **(a)** $\phi(x) = x_1 \max(x_2, x_3, x_4)x_5$.
(b) $\phi(x) = x_1 \max(x_2x_4, x_3x_5)x_6$.
(c) $\phi(x) = \max(x_1, x_2x_3)x_4$.

6. A minimal cut set has to contain at least one component of each minimal path set. There are six minimal cut sets: $\{1,5\}, \{1,6\}, \{2,5\}, \{2,3,6\}, \{3,4,6\}, \{4,5\}$.

12. The minimal path sets are $\{1,4\}$, $\{1,5\}$, $\{2,4\}$, $\{2,5\}$, $\{3,4\}$, $\{3,5\}$. With $q_i = 1 - p_i$, the reliability function is

$$\begin{aligned} r(\mathbf{p}) &= P\{\text{either of 1, 2, or 3 works}\}P\{\text{either of 4 or 5 works}\} \\ &= (1 - q_1q_2q_3)(1 - q_4q_5) \end{aligned}$$

17.

$$\begin{aligned} E[N^2] &= E[N^2|N>0]P\{N>0\} \\ &\geq (E[N|N>0])^2 P\{N>0\}, \qquad \text{since } E[X^2] \geq (E[X])^2 \end{aligned}$$

Thus,

$$\begin{aligned} E[N^2]P\{N>0\} &\geq (E[N|N>0]P[N>0])^2 \\ &= (E[N])^2 \end{aligned}$$

Let N denote the number of minimal path sets having all of its components functioning. Then $r(\boldsymbol{p}) = P\{N>0\}$. Similarly, if we define N as the number of minimal cut sets having all of its components failed, then $1 - r(\boldsymbol{p}) = P\{N>0\}$. In both cases we can compute expressions for $E[N]$ and $E[N^2]$ by writing N as the sum of indicator (i.e., Bernoulli) random variables. Then we can use the inequality to derive bounds on $r(\boldsymbol{p})$.

22. **(a)**
$$\begin{aligned}\bar{F}_t(a) &= P\{X > t + a | X > t\} \\ &= \frac{P[X > t + a\}}{P\{X > t\}} = \frac{\bar{F}(t + a)}{\bar{F}(t)}\end{aligned}$$

(b) Suppose $\lambda(t)$ is increasing. Recall that

$$\bar{F}(t) = e^{-\int_0^t \lambda(s)\, ds}$$

Hence,

$$\frac{\bar{F}(t + a)}{\bar{F}(t)} = \exp\left\{-\int_t^{t+a} \lambda(s) ds\right\}$$

which decreases in t since $\lambda(t)$ is increasing. To go the other way, suppose $\bar{F}(t + a)/\bar{F}(t)$ decreases in t. Now when a is small

$$\frac{\bar{F}(t + a)}{\bar{F}(t)} \approx e^{-a\lambda(t)}$$

Hence, $e^{-a\lambda(t)}$ must decrease in t and thus $\lambda(t)$ increases.

25. For $x \geq \xi$,

$$1 - p = \bar{F}(\xi) = \bar{F}(x(\xi/x)) \geq [\bar{F}(x)]^{\xi/x}$$

since IFRA. Hence, $\bar{F}(x) \leq (1 - p)^{x/\xi} = e^{-\theta x}$.
For $x \leq \xi$,

$$\bar{F}(x) = \bar{F}(\xi(x/\xi)) \geq [\bar{F}(\xi)]^{x/\xi}$$

since IFRA. Hence, $\bar{F}(x) \geq (1 - p)^{x/\xi} = e^{-\theta x}$.

30. $r(\boldsymbol{p}) = p_1p_2p_3 + p_1p_2p_4 + p_1p_3p_4 + p_2p_3p_4 - 3p_1p_2p_3p_4$

$$r(\mathbf{1} - \boldsymbol{F}(t)) = \begin{cases} 2(1-t)^2(1-t/2) + 2(1-t)(1-t/2)^2 \\ \quad -3(1-t)^2(1-t/2)^2, & 0 \leq t \leq 1 \\ 0, & 1 \leq t \leq 2 \end{cases}$$

$$E[\text{lifetime}] = \int_0^1 \Big[2(1-t)^2(1-t/2) + 2(1-t)(1-t/2)^2 - 3(1-t)^2(1-t/2)^2\Big]dt$$
$$= \frac{31}{60}$$

Chapter 10

1. $B(s) + B(t) = 2B(s) + B(t) - B(s)$. Now $2B(s)$ is normal with mean 0 and variance $4s$ and $B(t) - B(s)$ is normal with mean 0 and variance $t - s$. Because $B(s)$ and $B(t) - B(s)$ are independent, it follows that $B(s) + B(t)$ is normal with mean 0 and variance $4s + t - s = 3s + t$.

3.
$$\begin{aligned}
E[B(t_1)B(t_2)B(t_3)] &= E[E[B(t_1)B(t_2)B(t_3)|B(t_1), B(t_2)]] \\
&= E[B(t_1)B(t_2)E[B(t_3)|B(t_1), B(t_2)]] \\
&= E[B(t_1)B(t_2)B(t_2)] \\
&= E[E[B(t_1)B^2(t_2)|B(t_1)]] \\
&= E[B(t_1)E[B^2(t_2)|B(t_1)]] \\
&= E[B(t_1)\{(t_2 - t_1) + B^2(t_1)\}] \qquad (*) \\
&= E[B^3(t_1)] + (t_2 - t_1)E[B(t_1)] \\
&= 0
\end{aligned}$$

where the equality $(*)$ follows since given $B(t_1)$, $B(t_2)$ is normal with mean $B(t_1)$ and variance $t_2 - t_1$. Also, $E[B^3(t)] = 0$ since $B(t)$ is normal with mean 0.

5.
$$\begin{aligned}
P\{T_1 < T_{-1} < T_2\} &= P\{\text{hit 1 before } -1 \text{ before 2}\} \\
&= P\{\text{hit 1 before } -1\} \\
&\quad \times P\{\text{hit } -1 \text{ before 2}|\text{hit 1 before } -1\} \\
&= \tfrac{1}{2}P\{\text{down 2 before up 1}\} \\
&= \frac{1}{2}\frac{1}{3} = \frac{1}{6}
\end{aligned}$$

The next to last equality follows by looking at the Brownian motion when it first hits 1.

10. **(a)** Writing $X(t) = X(s) + X(t) - X(s)$ and using independent increments, we see that given $X(s) = c$, $X(t)$ is distributed as $c + X(t) - X(s)$. By stationary increments this has the same distribution as $c + X(t - s)$, and is thus normal with mean $c + \mu(t - s)$ and variance $(t - s)\sigma^2$.

(b) Use the representation $X(t) = \sigma B(t) + \mu t$, where $\{B(t)\}$ is standard Brownian motion. Using Eq. (10.4), but reversing s and t, we see that the conditional distribution of $B(t)$ given that $B(s) = (c - \mu s)/\sigma$ is normal

with mean $t(c-\mu s)/(\sigma s)$ and variance $t(s-t)/s$. Thus, the conditional distribution of $X(t)$ given that $X(s)=c$, $s>t$, is normal with mean

$$\sigma\left[\frac{t(c-\mu s)}{\sigma s}\right]+\mu t=\frac{(c-\mu s)t}{s}+\mu t$$

and variance

$$\frac{\sigma^2 t(s-t)}{s}$$

19. Since knowing the value of $Y(t)$ is equivalent to knowing $B(t)$, we have

$$\begin{aligned}E[Y(t)|Y(u),0\le u\le s]&=e^{-c^2t/2}E[e^{cB(t)}|B(u),0\le u\le s]\\&=e^{-c^2t/2}E[e^{cB(t)}|B(s)]\end{aligned}$$

Now, given $B(s)$, the conditional distribution of $B(t)$ is normal with mean $B(s)$ and variance $t-s$. Using the formula for the moment generating function of a normal random variable we see that

$$\begin{aligned}e^{-c^2t/2}E[e^{cB(t)}|B(s)]&=e^{-c^2t/2}e^{cB(s)+(t-s)c^2/2}\\&=e^{-c^2s/2}e^{cB(s)}\\&=Y(s)\end{aligned}$$

Thus $\{Y(t)\}$ is a Martingale.

$$E[Y(t)]=E[Y(0)]=1$$

20. By the Martingale stopping theorem

$$E[B(T)]=E[B(0)]=0$$

However, $B(T)=2-4T$ and so $2-4E[T]=0$, or $E[T]=\frac{1}{2}$.

24. It follows from the Martingale stopping theorem and the result of Exercise 18 that

$$E[B^2(T)-T]=0$$

where T is the stopping time given in this problem and

$$B(t)=\frac{X(t)-\mu t}{\sigma}$$

Therefore,

$$E\left[\frac{(X(T)-\mu T)^2}{\sigma^2}-T\right]=0$$

However, $X(T) = x$ and so the preceding gives that

$$E[(x - \mu T)^2] = \sigma^2 E[T]$$

But, from Exercise 21, $E[T] = x/\mu$ and so the preceding is equivalent to

$$\text{Var}(\mu T) = \sigma^2 \frac{x}{\mu} \quad \text{or} \quad \text{Var}(T) = \sigma^2 \frac{x}{\mu^3}$$

Chapter 11

1. **(a)** Let U be a random number. If $\sum_{j=1}^{i-1} P_j < U \leq \sum_{j=1}^{i} P_j$ then simulate from F_i. (In the preceding $\sum_{j=1}^{i-1} P_j \equiv 0$ when $i = 1$.)

(b) Note that

$$F(x) = \frac{1}{3} F_1(x) + \frac{2}{3} F_2(x)$$

where

$$F_1(x) = 1 - e^{2x}, \quad 0 < x < \infty$$

$$F_2(x) = \begin{cases} x, & 0 < x < 1 \\ 1, & 1 < x \end{cases}$$

Hence, using part (a), let U_1, U_2, U_3 be random numbers and set

$$X = \begin{cases} \dfrac{-\log U_2}{2}, & \text{if } U_1 < \frac{1}{3} \\ U_3, & \text{if } U_1 > \frac{1}{3} \end{cases}$$

The preceding uses the fact that $-\log U_2/2$ is exponential with rate 2.

3. If a random sample of size n is chosen from a set of $N + M$ items of which N are acceptable, then X, the number of acceptable items in the sample, is such that

$$P\{X = k\} = \binom{N}{k} \binom{M}{n-k} \Big/ \binom{N+M}{k}$$

To simulate X, note that if

$$I_j = \begin{cases} 1, & \text{if the } j\text{th selection is acceptable} \\ 0, & \text{otherwise} \end{cases}$$

then

$$P\{I_j = 1 | I_1, \ldots, I_{j-1}\} = \frac{N - \sum_1^{j-1} I_i}{N + M - (j-1)}$$

Hence, we can simulate $I_1, \ldots, I_n$ by generating random numbers $U_1, \ldots, U_n$ and then setting

$$I_j = \begin{cases} 1, & \text{if } U_j < \dfrac{N - \sum_1^{j-1} I_i}{N + M - (j-1)} \\ 0, & \text{otherwise} \end{cases}$$

and $X = \sum_{j=1}^{n} I_j$ has the desired distribution.
Another way is to let

$$X_j = \begin{cases} 1, & \text{the } j\text{th acceptable item is in the sample} \\ 0, & \text{otherwise} \end{cases}$$

and then simulate $X_1, \ldots, X_N$ by generating random numbers $U_1, \ldots, U_N$ and then setting

$$X_j = \begin{cases} 1, & \text{if } U_j < \dfrac{n - \sum_{i=1}^{j-1} X_i}{N + M - (j-1)} \\ 0, & \text{otherwise} \end{cases}$$

and $X = \sum_{j=1}^{N} X_j$ then has the desired distribution.
The former method is preferable when $n \le N$ and the latter when $N \le n$.

6. Let

$$\begin{aligned} c(\lambda) &= \max_x \left\{ \frac{f(x)}{\lambda e^{-\lambda x}} \right\} = \frac{2}{\lambda\sqrt{2\pi}} \max_x \left[\exp\left\{ \frac{-x^2}{2} + \lambda x \right\} \right] \\ &= \frac{2}{\lambda\sqrt{2\pi}} \exp\left\{ \frac{\lambda^2}{2} \right\} \end{aligned}$$

Hence,

$$\frac{d}{d\lambda} c(\lambda) = \sqrt{2/\pi} \exp\left\{ \frac{\lambda^2}{2} \right\} \left[1 - \frac{1}{\lambda^2} \right]$$

Hence $(d/d\lambda)c(\lambda) = 0$ when $\lambda = 1$ and it is easy to check that this yields the minimal value of $c(\lambda)$.

16. **(a)** They can be simulated in the same sequential fashion in which they are defined. That is, first generate the value of a random variable I_1 such that

$$P\{I_1 = i\} = \frac{w_i}{\sum_{j=1}^{n} w_j}, \qquad i = 1, \ldots, n$$

Then, if $I_1 = k$, generate the value of I_2 where

$$P\{I_2 = i\} = \frac{w_i}{\sum_{j \ne k} w_j}, \qquad i \ne k$$

and so on. However, the approach given in part (b) is more efficient.

(b) Let I_j denote the index of the jth smallest X_i.

23. Let $m(t) = \int_0^t \lambda(s)\,ds$, and let $m^{-1}(t)$ be the inverse function. That is, $m(m^{-1}(t)) = t$.

(a)
$$\begin{aligned} P\{m(X_1) > x\} &= P\{X_1 > m^{-1}(x)\} \\ &= P\{N(m^{-1}(x)) = 0\} \\ &= e^{-m(m^{-1}(x))} \\ &= e^{-x} \end{aligned}$$

(b)
$$\begin{aligned} P\{m(X_i) - m(X_{i-1}) > x | m(X_1), \ldots, m(X_{i-1}) - m(X_{i-2})\} &= P\{m(X_i) - m(X_{i-1}) > x | X_1, \ldots, X_{i-1}\} \\ &= P\{m(X_i) - m(X_{i-1}) > x | X_{i-1}\} \\ &= P\{m(X_i) - m(X_{i-1}) > x | m(X_{i-1})\} \end{aligned}$$

Now,

$$\begin{aligned} &P\{m(X_i) - m(X_{i-1}) > x | X_{i-1} = y\} \\ &= P\left\{\int_y^{X_i} \lambda(t)dt > x | X_{i-1} = y\right\} \\ &= P\{X_i > c | X_{i-1} = y\} \qquad \text{where } \int_y^c \lambda(t)\,dt = x \\ &= P\{N(c) - N(y) = 0 | X_{i-1} = y\} \\ &= P\{N(c) - N(y) = 0\} \\ &= \exp\left\{-\int_y^c \lambda(t)\,dt\right\} \\ &= e^{-x} \end{aligned}$$

32.
$$\begin{aligned} \text{Var}[(X+Y)/2] &= \tfrac{1}{4}[\text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)] \\ &= \frac{\text{Var}(X) + \text{Cov}(X,Y)}{2} \end{aligned}$$

Now it is always true that

$$\frac{\text{Cov}(V,W)}{\sqrt{\text{Var}(V)\text{Var}(W)}} \le 1$$

and so when X and Y have the same distribution $\text{Cov}(X,Y) \le \text{Var}(X)$.