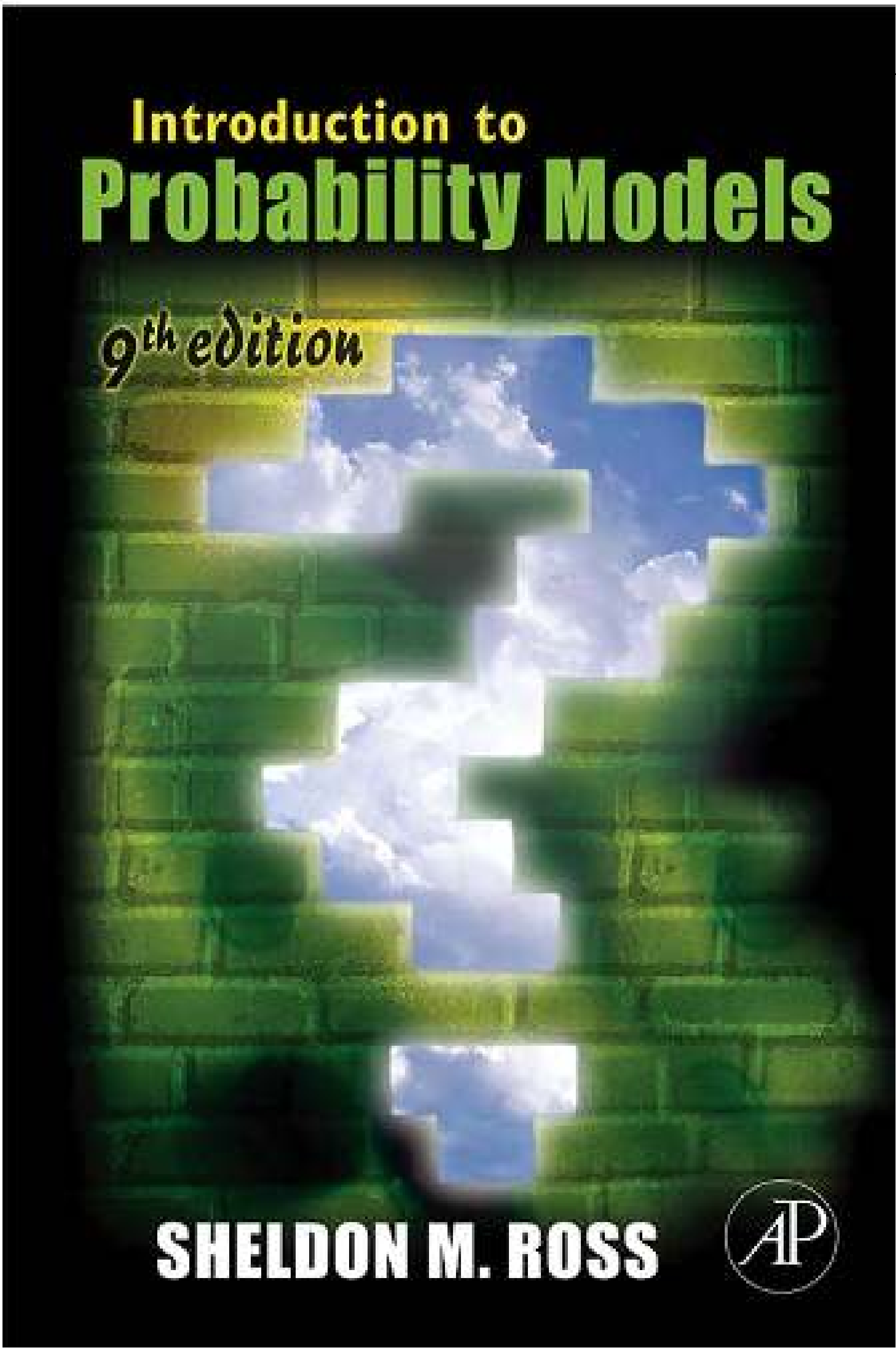

Introduction to
Probability Models
9th edition
SHELDON M. ROSS
AP

Introduction to Probability Models

Ninth Edition

This page intentionally left blank

Introduction to Probability Models

Ninth Edition

Sheldon M. Ross
University of California
Berkeley, California

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Acquisitions Editor	Tom Singer
Project Manager	Sarah M. Hajduk
Marketing Managers	Linda Beattie, Leah Ackerson
Cover Design	Eric DeCicco
Composition	VTEX
Cover Printer	Phoenix Color
Interior Printer	The Maple-Vail Book Manufacturing Group

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ♾

Library of Congress Cataloging-in-Publication Data
Application Submitted

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

ISBN-13: 978-0-12-598062-3
ISBN-10: 0-12-598062-0

For information on all Academic Press publications
visit our Web site at www.books.elsevier.com

Printed in the United States of America
06 07 08 09 10 9 8 7 6 5 4 3 2 1

Contents

Preface xiii

1. Introduction to Probability Theory 1

1.1. Introduction 1
1.2. Sample Space and Events 1
1.3. Probabilities Defined on Events 4
1.4. Conditional Probabilities 7
1.5. Independent Events 10
1.6. Bayes' Formula 12
Exercises 15
References 21

2. Random Variables 23

2.1. Random Variables 23
2.2. Discrete Random Variables 27
2.2.1. The Bernoulli Random Variable 28
2.2.2. The Binomial Random Variable 29
2.2.3. The Geometric Random Variable 31
2.2.4. The Poisson Random Variable 32
2.3. Continuous Random Variables 34
2.3.1. The Uniform Random Variable 35
2.3.2. Exponential Random Variables 36
2.3.3. Gamma Random Variables 37
2.3.4. Normal Random Variables 37

2.4. Expectation of a Random Variable 38
2.4.1. The Discrete Case 38
2.4.2. The Continuous Case 41
2.4.3. Expectation of a Function of a Random Variable 43
2.5. Jointly Distributed Random Variables 47
2.5.1. Joint Distribution Functions 47
2.5.2. Independent Random Variables 51
2.5.3. Covariance and Variance of Sums of Random Variables 53
2.5.4. Joint Probability Distribution of Functions of Random Variables 61
2.6. Moment Generating Functions 64
2.6.1. The Joint Distribution of the Sample Mean and Sample Variance from a Normal Population 74
2.7. Limit Theorems 77
2.8. Stochastic Processes 83
Exercises 85
References 96

3. Conditional Probability and Conditional Expectation 97

3.1. Introduction 97
3.2. The Discrete Case 97
3.3. The Continuous Case 102
3.4. Computing Expectations by Conditioning 105
3.4.1. Computing Variances by Conditioning 117
3.5. Computing Probabilities by Conditioning 120
3.6. Some Applications 137
3.6.1. A List Model 137
3.6.2. A Random Graph 139
3.6.3. Uniform Priors, Polya's Urn Model, and Bose–Einstein Statistics 147
3.6.4. Mean Time for Patterns 151
3.6.5. The *k*-Record Values of Discrete Random Variables 155
3.7. An Identity for Compound Random Variables 158
3.7.1. Poisson Compounding Distribution 161
3.7.2. Binomial Compounding Distribution 163
3.7.3. A Compounding Distribution Related to the Negative Binomial 164
Exercises 165

4. Markov Chains 185

4.1. Introduction 185
4.2. Chapman–Kolmogorov Equations 189
4.3. Classification of States 193
4.4. Limiting Probabilities 204
4.5. Some Applications 217
    4.5.1. The Gambler's Ruin Problem 217
    4.5.2. A Model for Algorithmic Efficiency 221
    4.5.3. Using a Random Walk to Analyze a Probabilistic Algorithm for the Satisfiability Problem 224
4.6. Mean Time Spent in Transient States 230
4.7. Branching Processes 233
4.8. Time Reversible Markov Chains 236
4.9. Markov Chain Monte Carlo Methods 247
4.10. Markov Decision Processes 252
4.11. Hidden Markov Chains 256
    4.11.1. Predicting the States 261
    Exercises 263
    References 280

5. The Exponential Distribution and the Poisson Process 281

5.1. Introduction 281
5.2. The Exponential Distribution 282
    5.2.1. Definition 282
    5.2.2. Properties of the Exponential Distribution 284
    5.2.3. Further Properties of the Exponential Distribution 291
    5.2.4. Convolutions of Exponential Random Variables 298
5.3. The Poisson Process 302
    5.3.1. Counting Processes 302
    5.3.2. Definition of the Poisson Process 304
    5.3.3. Interarrival and Waiting Time Distributions 307
    5.3.4. Further Properties of Poisson Processes 310
    5.3.5. Conditional Distribution of the Arrival Times 316
    5.3.6. Estimating Software Reliability 328
5.4. Generalizations of the Poisson Process 330
    5.4.1. Nonhomogeneous Poisson Process 330
    5.4.2. Compound Poisson Process 337
    5.4.3. Conditional or Mixed Poisson Processes 343

Exercises 346
References 364

6. Continuous-Time Markov Chains 365

6.1. Introduction 365
6.2. Continuous-Time Markov Chains 366
6.3. Birth and Death Processes 368
6.4. The Transition Probability Function $P_{ij}(t)$ 375
6.5. Limiting Probabilities 384
6.6. Time Reversibility 392
6.7. Uniformization 401
6.8. Computing the Transition Probabilities 404
Exercises 407
References 415

7. Renewal Theory and Its Applications 417

7.1. Introduction 417
7.2. Distribution of $N(t)$ 419
7.3. Limit Theorems and Their Applications 423
7.4. Renewal Reward Processes 433
7.5. Regenerative Processes 442
7.5.1. Alternating Renewal Processes 445
7.6. Semi-Markov Processes 452
7.7. The Inspection Paradox 455
7.8. Computing the Renewal Function 458
7.9. Applications to Patterns 461
7.9.1. Patterns of Discrete Random Variables 462
7.9.2. The Expected Time to a Maximal Run of Distinct Values 469
7.9.3. Increasing Runs of Continuous Random Variables 471
7.10. The Insurance Ruin Problem 473
Exercises 479
References 492

8. Queueing Theory 493

8.1. Introduction 493
8.2. Preliminaries 494
8.2.1. Cost Equations 495
8.2.2. Steady-State Probabilities 496

8.3. Exponential Models 499
8.3.1. A Single-Server Exponential Queueing System 499
8.3.2. A Single-Server Exponential Queueing System Having Finite Capacity 508
8.3.3. A Shoeshine Shop 511
8.3.4. A Queueing System with Bulk Service 514
8.4. Network of Queues 517
8.4.1. Open Systems 517
8.4.2. Closed Systems 522
8.5. The System $M/G/1$ 528
8.5.1. Preliminaries: Work and Another Cost Identity 528
8.5.2. Application of Work to $M/G/1$ 529
8.5.3. Busy Periods 530
8.6. Variations on the $M/G/1$ 531
8.6.1. The $M/G/1$ with Random-Sized Batch Arrivals 531
8.6.2. Priority Queues 533
8.6.3. An $M/G/1$ Optimization Example 536
8.6.4. The $M/G/1$ Queue with Server Breakdown 540
8.7. The Model $G/M/1$ 543
8.7.1. The $G/M/1$ Busy and Idle Periods 548
8.8. A Finite Source Model 549
8.9. Multiserver Queues 552
8.9.1. Erlang's Loss System 553
8.9.2. The $M/M/k$ Queue 554
8.9.3. The $G/M/k$ Queue 554
8.9.4. The $M/G/k$ Queue 556
Exercises 558
References 570

9. Reliability Theory 571

9.1. Introduction 571
9.2. Structure Functions 571
9.2.1. Minimal Path and Minimal Cut Sets 574
9.3. Reliability of Systems of Independent Components 578
9.4. Bounds on the Reliability Function 583
9.4.1. Method of Inclusion and Exclusion 584
9.4.2. Second Method for Obtaining Bounds on $r(\mathbf{p})$ 593
9.5. System Life as a Function of Component Lives 595
9.6. Expected System Lifetime 604
9.6.1. An Upper Bound on the Expected Life of a Parallel System 608

9.7. Systems with Repair 610
9.7.1. A Series Model with Suspended Animation 615
Exercises 617
References 624

10. Brownian Motion and Stationary Processes 625

10.1. Brownian Motion 625
10.2. Hitting Times, Maximum Variable, and the Gambler's Ruin Problem 629
10.3. Variations on Brownian Motion 631
10.3.1. Brownian Motion with Drift 631
10.3.2. Geometric Brownian Motion 631
10.4. Pricing Stock Options 632
10.4.1. An Example in Options Pricing 632
10.4.2. The Arbitrage Theorem 635
10.4.3. The Black-Scholes Option Pricing Formula 638
10.5. White Noise 644
10.6. Gaussian Processes 646
10.7. Stationary and Weakly Stationary Processes 649
10.8. Harmonic Analysis of Weakly Stationary Processes 654
Exercises 657
References 662

11. Simulation 663

11.1. Introduction 663
11.2. General Techniques for Simulating Continuous Random Variables 668
11.2.1. The Inverse Transformation Method 668
11.2.2. The Rejection Method 669
11.2.3. The Hazard Rate Method 673
11.3. Special Techniques for Simulating Continuous Random Variables 677
11.3.1. The Normal Distribution 677
11.3.2. The Gamma Distribution 680
11.3.3. The Chi-Squared Distribution 681
11.3.4. The Beta (n, m) Distribution 681
11.3.5. The Exponential Distribution—The Von Neumann Algorithm 682
11.4. Simulating from Discrete Distributions 685
11.4.1. The Alias Method 688

11.5. Stochastic Processes 692
11.5.1. Simulating a Nonhomogeneous Poisson Process 693
11.5.2. Simulating a Two-Dimensional Poisson Process 700
11.6. Variance Reduction Techniques 703
11.6.1. Use of Antithetic Variables 704
11.6.2. Variance Reduction by Conditioning 708
11.6.3. Control Variates 712
11.6.4. Importance Sampling 714
11.7. Determining the Number of Runs 720
11.8. Coupling from the Past 720
Exercises 723
References 731

Appendix: Solutions to Starred Exercises 733

Index 775

This page intentionally left blank

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 which 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 ninth edition contains the following new sections.

- Section 3.7 is concerned with compound random variables of the form $S_N = \sum_{i=1}^{N} X_i$, where N is independent of the sequence of independent and identically distributed random variables X_i, $i \geqslant 1$. It starts by deriving a general identity concerning compound random variables, as well as a corollary of that identity in the case where the X_i are positive and integer valued. The corollary is then used in subsequent subsections to obtain recursive formulas for the probability mass function of S_N, when N is a Poisson distribution (Subsection 3.7.1), a binomial distribution (Subsection 3.7.2), or a negative binomial distribution (Subsection 3.7.3).

- Section 4.11 deals with hidden Markov chains. These models suppose that a random signal is emitted each time a Markov chain enters a state, with the distribution of the signal depending on the state entered. The Markov chain is hidden in the sense that it is supposed that only the signals and not the underlying states of the chain are observable. As part of our analysis of these models we present, in Subsection 4.11.1, the Viterbi algorithm for determining the most probable sequence of first n states, given the first n signals.
- Section 8.6.4 analyzes the Poisson arrival single server queue under the assumption that the working server will randomly break down and need repair.

There is also new material in almost all chapters. Some of the more significant additions being the following.

- Example 5.9, which is concerned with the expected number of normal cells that survive until all cancer cells have been killed. The example supposes that each cell has a weight, and the probability that a given surviving cell is the next cell killed is proportional to its weight.
- A new approach—based on time sampling of a Poisson process—is presented in Subsection 5.4.1 for deriving the probability mass function of the number of events of a nonhomogeneous Poisson process that occur in any specified time interval.
- There is additional material in Section 8.3 concerning the $M/M/1$ queue. Among other things, we derive the conditional distribution of the number of customers originally found in the system by a customer who spends a time t in the system before departing. (The conditional distribution is Poisson.) In Example 8.3, we illustrate the inspection paradox, by obtaining the probability distribution of the number in the system as seen by the first arrival after some specified time.

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. Subsection 2.6.1 gives a simple derivation of the joint distribution of the sample mean and sample variance of a normal data sample.

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. Subsection 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. Subsection 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. Subsection 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 6.7 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 Subsection 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 Subsection 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 Subsection 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 Subsection 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. Subsection 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 (9.7.1) analyzing 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 program 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. Subsection 11.6.4 introduces the important simulation technique of importance sampling, and indicates the usefulness of tilted distributions when applying this method.

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

Tapas Das, University of South Florida

Israel David, Ben-Gurion University

Jay Devore, California Polytechnic Institute

Eugene Feinberg, State University of New York, Stonybrook

Ramesh Gupta, University of Maine

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

George Michailidis, University of Michigan

Krzysztof Osfaszewski, University of Illinois

Erol Pekoz, Boston University

Evgeny Poletsky, Syracuse University

Anthony Quas, University of Victoria

David Scollnik, University of Calgary

Mary Shepherd, Northwest Missouri State University

Galen Shorack, University of Washington, Seattle

Osnat Stramer, University of Iowa

Gabor Szekeley, Bowling Green State University

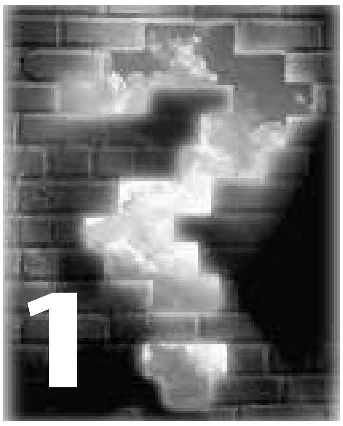
Marlin Thomas, Purdue University

Zhenyuan Wang, University of Binghamton

Julie Zhou, University of Victoria

Special thanks to Donald Minassian of Butler University, for his extremely careful reading of the text.

Introduction to Probability Theory



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 experiments 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 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:

$$S = \left\{\begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{matrix}\right\}$$

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,

$$S = [0, \infty)^* \quad ■$$

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.

*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 $a \leqslant x < b$.

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. ■

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

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

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 that event which 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 $S^c = \emptyset$.

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 $E_1, E_2, \dots$ that are mutually exclusive, that is, events for which $E_n E_m = \emptyset$ 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

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

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

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

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, Equation (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 Equation (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.