WILEY SERIES IN PROBABILITY AND STATISTICS

THIRD EDITION

AN INTRODUCTION TO PROBABILITY AND STATISTICS

VIJAY K. ROHATGI
A. K. MD. EHSANES SALEH

WILEY

AN INTRODUCTION TO PROBABILITY AND STATISTICS

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

AN INTRODUCTION TO PROBABILITY AND STATISTICS

Third Edition

VIJAY K. ROHATGI

A. K. Md. EHSANES SALEH

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Rohatgi, V. K., 1939-
An introduction to probability theory and mathematical statistics / Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh. – 3rd edition.
pages cm
Includes index.
ISBN 978-1-118-79964-2 (cloth)
1. Probabilities. 2. Mathematical statistics. I. Saleh, A. K. Md. Ehsanes. II. Title.
QA273.R56 2015
519.5–dc23

2015004848

Set in 10/12pts Times Lt Std by SPi Global, Pondicherry, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

3 2015

To Bina and Shahidara.

CONTENTS

PREFACE TO THE THIRD EDITION **xiii**

PREFACE TO THE SECOND EDITION **xv**

PREFACE TO THE FIRST EDITION **xvii**

ACKNOWLEDGMENTS **xix**

ENUMERATION OF THEOREMS AND REFERENCES **xxi**

1 Probability **1**

1.1 Introduction, 1
1.2 Sample Space, 2
1.3 Probability Axioms, 7
1.4 Combinatorics: Probability on Finite Sample Spaces, 20
1.5 Conditional Probability and Bayes Theorem, 26
1.6 Independence of Events, 31

2 Random Variables and Their Probability Distributions **39**

2.1 Introduction, 39
2.2 Random Variables, 39
2.3 Probability Distribution of a Random Variable, 42
2.4 Discrete and Continuous Random Variables, 47
2.5 Functions of a Random Variable, 55

3 Moments and Generating Functions **67**

3.1 Introduction, 67
3.2 Moments of a Distribution Function, 67
3.3 Generating Functions, 83
3.4 Some Moment Inequalities, 93

4 Multiple Random Variables **99**

4.1 Introduction, 99
4.2 Multiple Random Variables, 99
4.3 Independent Random Variables, 114
4.4 Functions of Several Random Variables, 123
4.5 Covariance, Correlation and Moments, 143
4.6 Conditional Expectation, 157
4.7 Order Statistics and Their Distributions, 164

5 Some Special Distributions **173**

5.1 Introduction, 173
5.2 Some Discrete Distributions, 173
5.2.1 Degenerate Distribution, 173
5.2.2 Two-Point Distribution, 174
5.2.3 Uniform Distribution on *n* Points, 175
5.2.4 Binomial Distribution, 176
5.2.5 Negative Binomial Distribution (Pascal or Waiting Time Distribution), 178
5.2.6 Hypergeometric Distribution, 183
5.2.7 Negative Hypergeometric Distribution, 185
5.2.8 Poisson Distribution, 186
5.2.9 Multinomial Distribution, 189
5.2.10 Multivariate Hypergeometric Distribution, 192
5.2.11 Multivariate Negative Binomial Distribution, 192
5.3 Some Continuous Distributions, 196
5.3.1 Uniform Distribution (Rectangular Distribution), 199
5.3.2 Gamma Distribution, 202
5.3.3 Beta Distribution, 210
5.3.4 Cauchy Distribution, 213
5.3.5 Normal Distribution (the Gaussian Law), 216
5.3.6 Some Other Continuous Distributions, 222
5.4 Bivariate and Multivariate Normal Distributions, 228
5.5 Exponential Family of Distributions, 240

6 Sample Statistics and Their Distributions **245**

6.1 Introduction, 245
6.2 Random Sampling, 246
6.3 Sample Characteristics and Their Distributions, 249

6.4 Chi-Square, t-, and F-Distributions: Exact Sampling Distributions, 262
6.5 Distribution of $(\overline{X}, S^2)$ in Sampling from a Normal Population, 271
6.6 Sampling from a Bivariate Normal Distribution, 276

7 Basic Asymptotics: Large Sample Theory 285

7.1 Introduction, 285
7.2 Modes of Convergence, 285
7.3 Weak Law of Large Numbers, 302
7.4 Strong Law of Large Numbers, 308
7.5 Limiting Moment Generating Functions, 316
7.6 Central Limit Theorem, 321
7.7 Large Sample Theory, 331

8 Parametric Point Estimation 337

8.1 Introduction, 337
8.2 Problem of Point Estimation, 338
8.3 Sufficiency, Completeness and Ancillarity, 342
8.4 Unbiased Estimation, 359
8.5 Unbiased Estimation (*Continued*): A Lower Bound for the Variance of An Estimator, 372
8.6 Substitution Principle (Method of Moments), 386
8.7 Maximum Likelihood Estimators, 388
8.8 Bayes and Minimax Estimation, 401
8.9 Principle of Equivariance, 418

9 Neyman–Pearson Theory of Testing of Hypotheses 429

9.1 Introduction, 429
9.2 Some Fundamental Notions of Hypotheses Testing, 429
9.3 Neyman–Pearson Lemma, 438
9.4 Families with Monotone Likelihood Ratio, 446
9.5 Unbiased and Invariant Tests, 453
9.6 Locally Most Powerful Tests, 459

10 Some Further Results on Hypotheses Testing 463

10.1 Introduction, 463
10.2 Generalized Likelihood Ratio Tests, 463
10.3 Chi-Square Tests, 472
10.4 t-Tests, 484
10.5 F-Tests, 489
10.6 Bayes and Minimax Procedures, 491

11 Confidence Estimation **499**

11.1 Introduction, 499
11.2 Some Fundamental Notions of Confidence Estimation, 499
11.3 Methods of Finding Confidence Intervals, 504
11.4 Shortest-Length Confidence Intervals, 517
11.5 Unbiased and Equivariant Confidence Intervals, 523
11.6 Resampling: Bootstrap Method, 530

12 General Linear Hypothesis **535**

12.1 Introduction, 535
12.2 General Linear Hypothesis, 535
12.3 Regression Analysis, 543
12.3.1 Multiple Linear Regression, 543
12.3.2 Logistic and Poisson Regression, 551
12.4 One-Way Analysis of Variance, 554
12.5 Two-Way Analysis of Variance with One Observation Per Cell, 560
12.6 Two-Way Analysis of Variance with Interaction, 566

13 Nonparametric Statistical Inference **575**

13.1 Introduction, 575
13.2 *U*-Statistics, 576
13.3 Some Single-Sample Problems, 584
13.3.1 Goodness-of-Fit Problem, 584
13.3.2 Problem of Location, 590
13.4 Some Two-Sample Problems, 599
13.4.1 Median Test, 601
13.4.2 Kolmogorov–Smirnov Test, 602
13.4.3 The Mann–Whitney–Wilcoxon Test, 604
13.5 Tests of Independence, 608
13.5.1 Chi-square Test of Independence—Contingency Tables, 608
13.5.2 Kendall's Tau, 611
13.5.3 Spearman's Rank Correlation Coefficient, 614
13.6 Some Applications of Order Statistics, 619
13.7 Robustness, 625
13.7.1 Effect of Deviations from Model Assumptions on Some Parametric Procedures, 625
13.7.2 Some Robust Procedures, 631

FREQUENTLY USED SYMBOLS AND ABBREVIATIONS **637**

REFERENCES **641**

STATISTICAL TABLES 647

ANSWERS TO SELECTED PROBLEMS 667

AUTHOR INDEX 677

SUBJECT INDEX 679

PREFACE TO THE THIRD EDITION

The *Third Edition* contains some new material. More specifically, the chapter on large sample theory has been reorganized, repositioned, and re-titled in recognition of the growing role of asymptotic statistics. In Chapter 12 on General Linear Hypothesis, the section on regression analysis has been greatly expanded to include multiple regression and logistic and Poisson regression.

Some more problems and remarks have been added to illustrate the material covered. The basic character of the book, however, remains the same as enunciated in the Preface to the first edition. It remains a solid introduction to first-year graduate students or advanced seniors in mathematics and statistics as well as a reference to students and researchers in other sciences.

We are grateful to the readers for their comments on this book over the past 40 years and would welcome any questions, comments, and suggestions. You can communicate with Vijay K. Rohatgi at vrohatg@bgsu.edu and with A. K. Md. Ehsanes Saleh at esaleh@math.carleton.ca.

Solana Beach, CA VIJAY K. ROHATGI
Ottawa, Canada A. K. Md. EHSANES SALEH

PREFACE TO THE SECOND EDITION

There is a lot that is different about this second edition. First, there is a co-author without whose help this revision would not have been possible. Second, we have benefited from countless letters from readers and colleagues who have pointed out errors and omissions and have made valuable suggestions over the past 25 years. These communications make this revision worth the effort. Third, we have tried to update the content of the book while striving to preserve the character and spirit of the first edition.

Here are some of the numerous changes that have been made.

1. The Introduction section has been removed. We have also removed Chapter 14 on sequential statistical inference.
2. Many parts of the book have gone substantial rewriting. For example, Chapter 4 has many changes, such as inclusion of exchangeability. In Chapter 3, an introduction to characteristic functions has been added. In Chapter 5 some new distributions have been added while in Chapter 6 there have been many changes in proofs.
3. The statistical inference part of the book (Chapters 8 to 13) has been updated. Thus in Chapter 8 we have expanded the coverage of invariance and have included discussions of ancillary statistics and conjugate prior distributions.
4. Similar changes have been made in Chapter 9. A new section on locally most powerful tests has been added.
5. Chapter 11 has been greatly revised and a discussion of invariant confidence intervals has been added.
6. Chapter 13 has been completely rewritten in the light of increased emphasis on nonparametric inference. We have expanded the discussion of U-statistics. Later sections show the connection between commonly used tests and U-statistics.
7. In Chapter 12, the notation has been changed to confirm to the current convention.

8. Many problems and examples have been added.
9. More figures have been added to illustrate examples and proofs.
10. Answers to selected problems have been provided.

We are truly grateful to the readers of the first edition for countless comments and suggestions and hope we will continue to hear from them about this edition.

Special thanks are due Ms. Gillian Murray for her superb word processing of the manuscript, and Dr. Indar Bhatia for figures that appear in the text. Dr. Bhatia spent countless hours preparing the diagrams for publication. We also acknowledge the assistance of Dr. K. Selvavel.

Vijay K. Rohatgi
A. K. Md. Ehsanes Saleh

PREFACE TO THE FIRST EDITION

This book on probability theory and mathematical statistics is designed for a three-quarter course meeting 4 hours per week or a two-semester course meeting 3 hours per week. It is designed primarily for advanced seniors and beginning graduate students in mathematics, but it can also be used by students in physics and engineering with strong mathematical backgrounds. Let me emphasize that this is a mathematics text and not a "cookbook." It should not be used as a text for service courses.

The mathematics prerequisites for this book are modest. It is assumed that the reader has had basic courses in set theory and linear algebra and a solid course in advanced calculus. No prior knowledge of probability and/or statistics is assumed.

My aim is to provide a solid and well-balanced introduction to probability theory and mathematical statistics. It is assumed that students who wish to do graduate work in probability theory and mathematical statistics will be taking, concurrently with this course, a measure-theoretic course in analysis if they have not already had one. These students can go on to take advanced-level courses in probability theory or mathematical statistics after completing this course.

This book consists of essentially three parts, although no such formal divisions are designated in the text. The first part consists of Chapters 1 through 6, which form the core of the probability portion of the course. The second part, Chapters 7 through 11, covers the foundations of statistical inference. The third part consists of the remaining three chapters on special topics. For course sequences that separate probability and mathematical statistics, the first part of the book can be used for a course in probability theory, followed by a course in mathematical statistics based on the second part and, possibly, one or more chapters on special topics.

The reader will find here a wealth of material. Although the topics covered are fairly conventional, the discussions and special topics included are not. Many presentations give far more depth than is usually the case in a book at this level. Some special features of the book are the following:

1. A well-referenced chapter on the preliminaries.
2. About 550 problems, over 350 worked-out examples, about 200 remarks, and about 150 references.
3. An advance warning to reader wherever the details become too involved. They can skip the later portion of the section in question on first reading without destroying the continuity in any way.
4. Many results on characterizations of distributions (Chapter 5).
5. Proof of the central limit theorem by the method of operators and proof of the strong law of large numbers (Chapter 6).
6. A section on minimal sufficient statistics (Chapter 8).
7. A chapter on special tests (Chapter 10).
8. A careful presentation of the theory of confidence intervals, including Bayesian intervals and shortest-length confidence intervals (Chapter 11).
9. A chapter on the general linear hypothesis, which carries linear models through to their use in basic analysis of variance (Chapter 12).
10. Sections on nonparametric estimation and robustness (Chapter 13).
11. Two sections on sequential estimation (Chapter 14).

The contents of this book were used in a 1-year (two-semester) course that I taught three times at the Catholic University of America and once in a three-quarter course at Bowling Green State University. In the fall of 1973 my colleague, Professor Eugene Lukacs, taught the first quarter of this same course on the basis of my notes, which eventually became this book. I have always been able to cover this book (with few omissions) in a 1-year course, lecturing 3 hours a week. An hour-long problem session every week is conducted by a senior graduate student.

In a book of this size there are bound to be some misprints, errors, and ambiguities of presentation. I shall be grateful to any reader who brings these to my attention.

Bowling Green, Ohio V. K. ROHATGI
February 1975

ACKNOWLEDGMENTS

We take this opportunity to thank many correspondents whose comments and criticisms led to improvements in the *Third Edition*. The list below is far from complete since it does not include the names of countless students whose reactions to the book as a text helped the authors in this revised edition. We apologize to those whose names may have been inadvertently omitted from the list because we were not diligent enough to keep a complete record of all the correspondence. For the third edition we wish to thank Professors Yue-Cune Chang, Anirban Das Gupta, A. G. Pathak, Arno Weiershauser, and many other readers who sent their questions and comments. We also wish to acknowledge the assistance of Dr. Pooplasingam Sivakumar in preparation of the manuscript. For the second edition: Barry Arnold, Lennart Bondesson, Harry Cohn, Frank Connonito, Emad El-Neweihi, Ulrich Faigle, Pier Alda Ferrari, Martin Feuerrnan, Xavier Fernando, Z. Govindarajulu, Arjun Gupta, Hassein Hamedani, Thomas Hem, Jin-Sheng Huang, Bill Hudson, Barthel Huff, V. S. Huzurbazar, B. K. Kale, Sam Kotz, Bansi Lal, Sri Gopal Mohanty, M. V. Moorthy, True Nguyen, Tom O'Connor, A. G. Pathak, Edsel Pena, S. Perng, Madan Puri, Prem Puri, J. S. Rao, Bill Raser, Andrew Rukhin, K. Selvavel, Rajinder Singh, R. J. Tomkins; for the first edition, Ralph Baty, Ralph Bradley, Eugene Lukacs, Kae Lea Main, Tom and Carol O'Connor, M. S. Scott Jr., J. Sethuraman, Beatrice Shube, Jeff Spielman, and Robert Tortora.

We thank the publishers of the *American Mathematical Monthly*, the *SIAM Review*, and the *American Statistician* for permission to include many examples and problems that appeared in these journals. Thanks are also due to the following for permission to include tables: Professors E. S. Pearson and L. R. Verdooren (Table ST11), Harvard University Press (Table ST1), Hafner Press (Table ST3), Iowa State University Press (Table ST5), Rand Corporation (Table ST6), the American Statistical Association (Tables ST7 and ST10), the Institute of Mathematical Statistics (Tables ST8 and ST9), Charles Griffin & Co., Ltd. (Tables ST12 and ST13), and John Wiley & Sons (Tables ST1, ST2, ST4, ST10, and ST11).

ENUMERATION OF THEOREMS AND REFERENCES

This book is divided into 13 chapters, numbered 1 through 13. Each chapter is divided into several sections. Lemmas, theorems, equations, definitions, remarks, figures, and so on, are numbered consecutively within each section. Thus Theorem $i.j.k$ refers to the kth theorem in Section j of Chapter i, Section $i.j$ refers to the jth section of Chapter i, and so on. Theorem j refers to the jth theorem of the section in which it appears. A similar convention is used for equations except that equation numbers are enclosed in parentheses. Each section is followed by a set of problems for which the same numbering system is used.

References are given at the end of the book and are denoted in the text by numbers enclosed in square brackets, [ ]. If a citation is to a book, the notation $([i, p. j])$ refers to the jth page of the reference numbered $[i]$.

A word about the proofs of results stated without proof in this book. If a reference appears immediately following or preceding the statement of a result, it generally means that the proof is beyond the scope of this text. If no reference is given, it indicates that the proof is left to the reader. Sometimes the reader is asked to supply the proof as a problem.

1

PROBABILITY

1.1 INTRODUCTION

The theory of probability had its origin in gambling and games of chance. It owes much to the curiosity of gamblers who pestered their friends in the mathematical world with all sorts of questions. Unfortunately this association with gambling contributed to a very slow and sporadic growth of probability theory as a mathematical discipline. The mathematicians of the day took little or no interest in the development of any theory but looked only at the combinatorial reasoning involved in each problem.

The first attempt at some mathematical rigor is credited to Laplace. In his monumental work, *Theorie analytique des probabilités* (1812), Laplace gave the classical definition of the probability of an event that can occur only in a finite number of ways as the proportion of the number of favorable outcomes to the total number of all possible outcomes, provided that all the outcomes are *equally likely*. According to this definition, the computation of the probability of events was reduced to combinatorial counting problems. Even in those days, this definition was found inadequate. In addition to being circular and restrictive, it did not answer the question of what probability is, it only gave a practical method of computing the probabilities of some simple events.

An extension of the classical definition of Laplace was used to evaluate the probabilities of sets of events with infinite outcomes. The notion of *equal likelihood* of certain events played a key role in this development. According to this extension, if Ω is some region with a well-defined measure (length, area, volume, etc.), the probability that a point chosen *at random* lies in a subregion A of Ω is the ratio $\text{measure}(A)/\text{measure}(\Omega)$. Many problems of geometric probability were solved using this extension. The trouble is that one can

An Introduction to Probability and Statistics, Third Edition. Vijay K. Rohatgi and A.K. Md. Ehsanes Saleh.

define "at random" in any way one pleases, and different definitions therefore lead to different answers. Joseph Bertrand, for example, in his book *Calcul des probabilités* (Paris, 1889) cited a number of problems in geometric probability where the result depended on the method of solution. In Example 9 we will discuss the famous Bertrand paradox and show that in reality there is nothing paradoxical about Bertrand's paradoxes; once we define "probability spaces" carefully, the paradox is resolved. Nevertheless difficulties encountered in the field of geometric probability have been largely responsible for the slow growth of probability theory and its tardy acceptance by mathematicians as a mathematical discipline.

The mathematical theory of probability, as we know it today, is of comparatively recent origin. It was A. N. Kolmogorov who axiomatized probability in his fundamental work, *Foundations of the Theory of Probability* (Berlin), in 1933. According to this development, random events are represented by sets and probability is just a *normed measure* defined on these sets. This measure-theoretic development not only provided a logically consistent foundation for probability theory but also, at the same time, joined it to the mainstream of modern mathematics.

In this book we follow Kolmogorov's axiomatic development. In Section 1.2 we introduce the notion of a sample space. In Section 1.3 we state Kolmogorov's axioms of probability and study some simple consequences of these axioms. Section 1.4 is devoted to the computation of probability on finite sample spaces. Section 1.5 deals with conditional probability and Bayes's rule while Section 1.6 examines the independence of events.

1.2 SAMPLE SPACE

In most branches of knowledge, experiments are a way of life. In probability and statistics, too, we concern ourselves with special types of experiments. Consider the following examples.

Example 1. A coin is tossed. Assuming that the coin does not land on the side, there are two possible outcomes of the experiment: heads and tails. On any performance of this experiment one does not know what the outcome will be. The coin can be tossed as many times as desired.

Example 2. A roulette wheel is a circular disk divided into 38 equal sectors numbered from 0 to 36 and 00. A ball is rolled on the edge of the wheel, and the wheel is rolled in the opposite direction. One bets on any of the 38 numbers or some combinations of them. One can also bet on a color, red or black. If the ball lands in the sector numbered 32, say, anybody who bet on 32 or combinations including 32 wins, and so on. In this experiment, all possible outcomes are known in advance, namely 00, 0, 1, 2, ..., 36, but on any performance of the experiment there is uncertainty as to what the outcome will be, provided, of course, that the wheel is not rigged in any manner. Clearly, the wheel can be rolled any number of times.

Example 3. A manufacturer produces footrules. The experiment consists in measuring the length of a footrule produced by the manufacturer as accurately as possible. Because of errors in the production process one does not know what the true length of the footrule selected will be. It is clear, however, that the length will be, say, between 11 and 13 in., or, if one wants to be safe, between 6 and 18 in.

Example 4. The length of life of a light bulb produced by a certain manufacturer is recorded. In this case one does not know what the length of life will be for the light bulb selected, but clearly one is aware in advance that it will be some number between 0 and ∞ hours.

The experiments described above have certain common features. For each experiment, we know in advance all possible outcomes, that is, there are no surprises in store after the performance of any experiment. On any performance of the experiment, however, we do not know what the specific outcome will be, that is, there is uncertainty about the outcome on any performance of the experiment. Moreover, the experiment can be repeated under identical conditions. These features describe a *random* (or a *statistical*) *experiment*.

Definition 1. A random (or a statistical) experiment is an experiment in which

(a) all outcomes of the experiment are known in advance,
(b) any performance of the experiment results in an outcome that is not known in advance, and
(c) the experiment can be repeated under identical conditions.

In probability theory we study this uncertainty of a random experiment. It is convenient to associate with each such experiment a set Ω, the set of all possible outcomes of the experiment. To engage in any meaningful discussion about the experiment, we associate with Ω a σ-field $\mathcal{S}$, of subsets of Ω. We recall that a σ-field is a nonempty class of subsets of Ω that is closed under the formation of countable unions and complements and contains the null set Φ.

Definition 2. The sample space of a statistical experiment is a pair $(\Omega, \mathcal{S})$, where

(a) Ω is the set of all possible outcomes of the experiment and
(b) $\mathcal{S}$ is a σ-field of subsets of Ω.

The elements of Ω are called *sample points*. Any set $A \in \mathcal{S}$ is known as an *event*. Clearly A is a collection of sample points. We say that an event A happens if the outcome of the experiment corresponds to a point in A. Each one-point set is known as a *simple* or an *elementary event*. If the set Ω contains only a finite number of points, we say that $(\Omega, \mathcal{S})$ is a *finite sample space*. If Ω contains at most a countable number of points, we call $(\Omega, \mathcal{S})$ a *discrete sample space*. If, however, Ω contains uncountably many points, we say that $(\Omega, \mathcal{S})$ is an *uncountable sample space*. In particular, if $\Omega = \mathcal{R}_k$ or some rectangle in $\mathcal{R}_k$, we call it a *continuous sample space*.

Remark 1. The choice of $\mathcal{S}$ is an important one, and some remarks are in order. If Ω contains at most a countable number of points, we can always take $\mathcal{S}$ to be the class of all subsets of Ω. This is certainly a σ-field. Each one point set is a member of $\mathcal{S}$ and is the fundamental object of interest. Every subset of Ω is an event. If Ω has uncountably many points, the class of all subsets of Ω is still a σ-field, but it is much too large a class of sets to be of interest. It may not be possible to choose the class of all subsets of Ω as $\mathcal{S}$. One of the most important examples of an uncountable sample space is the case in which $\Omega = \mathcal{R}$ or Ω is an interval in $\mathcal{R}$. In this case we would like all one-point subsets of Ω and all intervals (closed, open, or semiclosed) to be events. We use our knowledge of analysis to specify $\mathcal{S}$. We will not go into details here except to recall that the class of all semiclosed intervals $(a,b]$ generates a class $\mathfrak{B}_1$ which is a σ-field on $\mathcal{R}$. This class contains all one-point sets and all intervals (finite or infinite). We take $\mathcal{S} = \mathfrak{B}_1$. Since we will be dealing mostly with the one-dimensional case, we will write $\mathfrak{B}$ instead of $\mathfrak{B}_1$. There are many subsets of $\mathcal{R}$ that are not in $\mathfrak{B}_1$, but we will not demonstrate this fact here. We refer the reader to Halmos [42], Royden [96], or Kolmogorov and Fomin [54] for further details.

Example 5. Let us toss a coin. The set Ω is the set of symbols H and T, where H denotes head and T represents tail. Also, $\mathcal{S}$ is the class of all subsets of Ω, namely, $\{\{\mathrm{H}\},\{\mathrm{T}\},\{\mathrm{H},\mathrm{T}\},\Phi\}$. If the coin is tossed two times, then

$$\begin{aligned}&\Omega = \{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}, \quad \mathcal{S} = \{\emptyset,\{(\mathrm{H},\mathrm{H})\},\\ &\{(\mathrm{H},\mathrm{T})\},\{(\mathrm{T},\mathrm{H})\},\{(\mathrm{T},\mathrm{T})\},\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T})\},\{(\mathrm{H},\mathrm{H}),(\mathrm{T},\mathrm{H})\},\\ &\{(\mathrm{H},\mathrm{H}),(\mathrm{T},\mathrm{T})\},\{(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H})\},\{(\mathrm{T},\mathrm{T}),(\mathrm{T},\mathrm{H})\},\{(\mathrm{T},\mathrm{T}),\\ &(\mathrm{H},\mathrm{T})\},\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H})\},\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{T})\},\\ &\{(\mathrm{H},\mathrm{H}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\},\{(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\},\Omega\},\end{aligned}$$

where the first element of a pair denotes the outcome of the first toss and the second element, the outcome of the second toss. The event *at least one head* consists of sample points (H,H), (H,T), (T,H). The event *at most one head* is the collection of sample points (H,T), (T,H), (T,T).

Example 6. A die is rolled n times. The sample space is the pair $(\Omega,\mathcal{S})$, where Ω is the set of all n-tuples $(x_1,x_2,\ldots,x_n)$, $x_i \in \{1,2,3,4,5,6\}$, $i = 1,2,\ldots,n$, and $\mathcal{S}$ is the class of all subsets of Ω. Ω contains 6^n elementary events. The event A that 1 shows at least once is the set

$$\begin{aligned}A &= \{(x_1,x_2,\ldots,x_n)\colon \text{at least one of } x_i\text{'s is } 1\}\\ &= \Omega - \{(x_1,x_2,\ldots,x_n)\colon \text{none of the } x_i\text{'s is } 1\}\\ &= \Omega - \{(x_1,x_2,\ldots,x_n)\colon x_i \in \{2,3,4,5,6\},\ i = 1,2,\ldots,n\}.\end{aligned}$$

Example 7. A coin is tossed until the first head appears. Then

$$\Omega = \{\mathrm{H},(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T},\mathrm{T},\mathrm{H}),\ldots\},$$

and $\mathcal{S}$ is the class of all subsets of Ω. An equivalent way of writing Ω would be to look at the number of tosses required for the first head. Clearly, this number can take values

$1, 2, 3, \ldots$, so that Ω is the set of all positive integers. The $\mathcal{S}$ is the class of all subsets of positive integers.

Example 8. Consider a pointer that is free to spin about the center of a circle. If the pointer is spun by an impulse, it will finally come to rest at some point. On the assumption that the mechanism is not rigged in any manner, each point on the circumference is a possible outcome of the experiment. The set Ω consists of all points $0 \le x < 2\pi r$, where r is the radius of the circle. Every one-point set $\{x\}$ is a simple event, namely, that the pointer will come to rest at x. The events of interest are those in which the pointer stops at a point belonging to a specified arc. Here $\mathcal{S}$ is taken to be the Borel σ-field of subsets of $[0, 2\pi r)$.

Example 9. A rod of length l is thrown onto a flat table, which is ruled with parallel lines at distance $2l$. The experiment consists in noting whether the rod intersects one of the ruled lines.

Let r denote the distance from the center of the rod to the nearest ruled line, and let θ be the angle that the axis of the rod makes with this line (Fig. 1). Every outcome of this experiment corresponds to a point (r, θ) in the plane. As Ω we take the set of all points (r, θ) in $\{(r, \theta) \colon 0 \le r \le l, 0 \le \theta < \pi\}$. For $\mathcal{S}$ we take the Borel σ-field, $\mathfrak{B}_2$, of subsets of Ω, that is, the smallest σ-field generated by rectangles of the form

$$\{(x, y) \colon a < x \le b, \quad c < y \le d, \quad 0 \le a < b \le l, \quad 0 \le c < d < \pi\}.$$

Clearly the rod will intersect a ruled line if and only if the center of the rod lies in the area enclosed by the locus of the center of the rod (while one end touches the nearest line) and the nearest line (shaded area in Fig. 2).

Remark 2. From the discussion above it should be clear that in the discrete case there is really no problem. Every one-point set is also an event, and $\mathcal{S}$ is the class of all subsets of Ω.

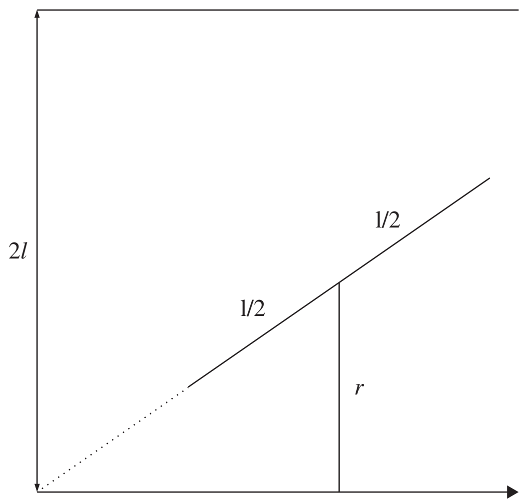

Fig. 1

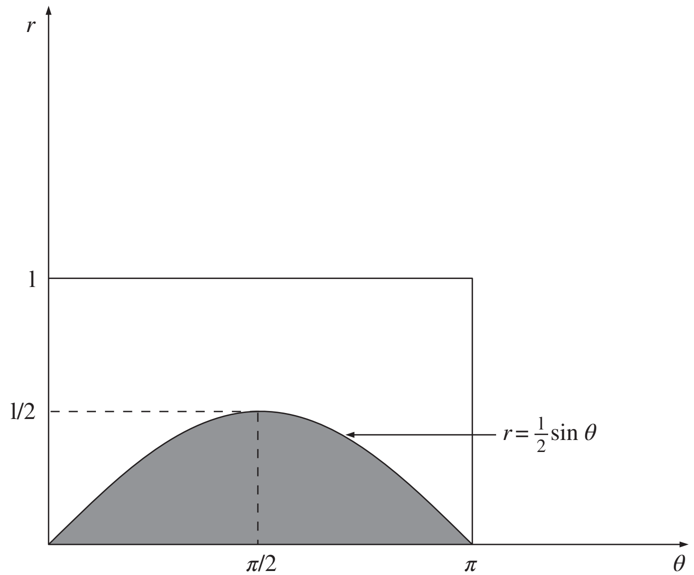

Fig. 2

The problem, if there is any, arises only in regard to uncountable sample spaces. The reader has to remember only that in this case not all subsets of Ω are events. The case of most interest is the one in which $\Omega = \mathcal{R}_k$. In this case, roughly all sets that have a well-defined volume (or area or length) are events. Not every set has the property in question, but sets that lack it are not easy to find and one does not encounter them in practice.

PROBLEMS 1.2

1. A club has five members A, B, C, D, and E. It is required to select a chairman and a secretary. Assuming that one member cannot occupy both positions, write the sample space associated with these selections. What is the event that member A is an office holder?

2. In each of the following experiments, what is the sample space?

(a) In a survey of families with three children, the sexes of the children are recorded in increasing order of age.

(b) The experiment consists of selecting four items from a manufacturer's output and observing whether or not each item is defective.

(c) A given book is opened to any page, and the number of misprints is counted.

(d) Two cards are drawn (i) with replacement and (ii) without replacement from an ordinary deck of cards.

3. Let A, B, C be three arbitrary events on a sample space $(\Omega, \mathcal{S})$. What is the event that only A occurs? What is the event that at least two of A, B, C occur? What is the event that both A and C, but not B, occur? What is the event that at most one of A, B, C occurs?

1.3 PROBABILITY AXIOMS

Let $(\Omega, \mathcal{S})$ be the sample space associated with a statistical experiment. In this section we define a probability set function and study some of its properties.

Definition 1. Let $(\Omega, \mathcal{S})$ be a sample space. A set function P defined on $\mathcal{S}$ is called a probability measure (or simply probability) if it satisfies the following conditions:

(i) $P(A) \geq 0$ for all $A \in \mathcal{S}$.
(ii) $P(\Omega) = 1$.
(iii) Let $\{A_j\}$, $A_j \in \mathcal{S}$, $j = 1, 2, \ldots$, be a disjoint sequence of sets, that is, $A_j \cap A_k = \Phi$ for $j \neq k$ where Φ is the null set. Then

$$P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j), \tag{1}$$

where we have used the notation $\sum_{j=1}^{\infty} A_j$ to denote union of disjoint sets A_j.

We call $P(A)$ the *probability of event A*. If there is no confusion, we will write PA instead of $P(A)$. Property (iii) is called *countable additivity*. That $P\Phi = 0$ and P is also finitely additive follows from it.

Remark 1. If Ω is discrete and contains at most n $(< \infty)$ points, each single-point set $\{\omega_j\}$, $j = 1, 2, \ldots, n$, is an elementary event, and it is sufficient to assign probability to each $\{\omega_j\}$. Then, if $A \in \mathcal{S}$, where $\mathcal{S}$ is the class of all subsets of Ω, $PA = \sum_{\omega \in A} P\{\omega\}$. One such assignment is the *equally likely* assignment or the assignment of *uniform* probabilities. According to this assignment, $P\{\omega_j\} = 1/n$, $j = 1, 2, \ldots, n$. Thus $PA = m/n$ if A contains m elementary events, $1 \leq m \leq n$.

Remark 2. If Ω is discrete and contains a countable number of points, one cannot make an equally likely assignment of probabilities. It suffices to make the assignment for each elementary event. If $A \in \mathcal{S}$, where $\mathcal{S}$ is the class of all subsets of Ω, define $PA = \sum_{\omega \in A} P\{\omega\}$.

Remark 3. If Ω contains uncountably many points, each one-point set is an elementary event, and again one cannot make an equally likely assignment of probabilities. Indeed, one cannot assign positive probability to each elementary event without violating the axiom $P\Omega = 1$. In this case one assigns probabilities to compound events consisting of intervals. For example, if $\Omega = [0, 1]$ and $\mathcal{S}$ is the Borel σ-field of all subsets of Ω, the assignment $P[I] =$ length of I, where I is a subinterval of Ω, defines a probability.

Definition 2. The triple $(\Omega, \mathcal{S}, P)$ is called a probability space.

Definition 3. Let $A \in \mathcal{S}$. We say that the odds for A are a to b if $PA = a/(a+b)$, and then the odds against A are b to a.

In many games of chance, probability is often stated in terms of odds against an event. Thus in horse racing a two dollar bet on a horse to win with odds of 2 to 1 (against) pays approximately six dollars if the horse wins the race. In this case the probability of winning is 1/3.

Example 1. Let us toss a coin. The sample space is $(\Omega, \mathcal{S})$, where $\Omega = \{\mathrm{H}, \mathrm{T}\}$, and $\mathcal{S}$ is the σ-field of all subsets of Ω. Let us define P on $\mathcal{S}$ as follows.

$$P\{\mathrm{H}\} = 1/2, \quad P\{\mathrm{T}\} = 1/2.$$

Then P clearly defines a probability. Similarly, $P\{\mathrm{H}\} = 2/3$, $P\{\mathrm{T}\} = 1/3$, and $P\{\mathrm{H}\} = 1$, $P\{\mathrm{T}\} = 0$ are probabilities defined on $\mathcal{S}$. Indeed,

$$P\{\mathrm{H}\} = p \quad \text{and} \quad P\{\mathrm{T}\} = 1 - p \quad (0 \leq p \leq 1)$$

defines a probability on $(\Omega, \mathcal{S})$.

Example 2. Let $\Omega = \{1, 2, 3, \ldots\}$ be the set of positive integers, and let $\mathcal{S}$ be the class of all subsets of Ω. Define P on $\mathcal{S}$ as follows:

$$P\{i\} = \frac{1}{2^i}, \quad i = 1, 2, \ldots.$$

Then $\sum_{i=1}^{\infty} P\{i\} = 1$, and P defines a probability.

Example 3. Let $\Omega = (0, \infty)$ and $\mathcal{S} = \mathfrak{B}$, the Borel σ-Field on Ω. Define P as follows: for each interval $I \subseteq \Omega$,

$$PI = \int_I e^{-x}\, dx.$$

Clearly $PI \geq 0$, $P\Omega = 1$, and P is countably additive by properties of integrals.

Theorem 1. P is monotone and subtractive; that is, if $A, B \in \mathcal{S}$ and $A \subseteq B$, then $PA \leq PB$ and $P(B - A) = PB - PA$, where $B - A = B \cap A^c$, A^c being the complement of the event A.

Proof. If $A \subseteq B$, then

$$B = (A \cap B) + (B - A) = A + (B - A).$$

and it follows that $PB = PA + P(B - A)$.

Corollary. For all $A \in \mathcal{S}$, $0 \leq PA \leq 1$.

Remark 4. We wish to emphasize that, if $PA = 0$ for some $A \in \mathcal{S}$, we call A an event with *zero probability* or a *null* event. However, it does not follow that $A = \Phi$. Similarly, if $PB = 1$ for some $B \in \mathcal{S}$, we call B a *certain event* but it does not follow that $B = \Omega$.

Theorem 2 (The Addition Rule). If $A, B \in \mathcal{S}$, then

$$P(A \cup B) = PA + PB - P(A \cap B). \tag{2}$$

Proof. Clearly

$$A \cup B = (A - B) + (B - A) + (A \cap B)$$

and

$$A = (A \cap B) + (A - B), B = (A \cap B) + (B - A).$$

The result follows by countable additivity of P.

Corollary 1. P is subadditive, that is, if $A, B \in \mathcal{S}$, then

$$P(A \cup B) \leq PA + PB. \tag{3}$$

Corollary 1 can be extended to an arbitrary number of events A_j,

$$P\left(\bigcup_j A_j\right) \leq \sum_j PA_j. \tag{4}$$

Corollary 2. If $B = A^c$, then A and B are disjoint and

$$PA = 1 - PA^c. \tag{5}$$

The following generalization of (2) is left as an exercise.

Theorem 3 (The Principle of Inclusion–Exclusion). Let $A_1, A_2, \ldots, A_n \in \mathcal{S}$. Then

$$\begin{aligned} P\left(\bigcup_{k=1}^{n} A_k\right) &= \sum_{k=1}^{n} PA_k - \sum_{k_1<k_2}^{n} P(A_{k_1} \cap A_{k_2}) \\ &\quad + \sum_{k_1<k_2<k_3}^{n} P(A_{k_1} \cap A_{k_2} \cap A_{k_3}) \\ &\quad + \cdots + (-1)^{n+1} P\left(\bigcap_{k=1}^{n} A_k\right). \end{aligned} \tag{6}$$

Example 4. A die is rolled twice. Let all the elementary events in $\Omega = \{(i,j)\colon\, i,j = 1,2,\ldots,6\}$ be assigned the same probability. Let A be the event that the first throw shows a number ≤ 2, and B, the event that the second throw shows at least 5. Then

$$\begin{aligned} A &= \{(i,j)\colon\, 1 \leq i \leq 2,\ j = 1,2,\ldots,6\}, \\ B &= \{(i,j)\colon\, 5 \leq j \leq 6,\ i = 1,2,\ldots,6\}, \\ A \cap B &= \{(1,5),(1,6),(2,5),(2,6)\}; \end{aligned}$$

$$\begin{aligned} P(A \cup B) &= PA + PB - P(A \cap B) \\ &= \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{4}{36} = \tfrac{5}{9}. \end{aligned}$$

Example 5. A coin is tossed three times. Let us assign equal probability to each of the 2^3 elementary events in Ω. Let A be the event that at least one head shows up in three throws. Then

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - P(\text{no heads}) \\ &= 1 - P(\text{TTT}) = \tfrac{7}{8}. \end{aligned}$$

We next derive two useful inequalities.

Theorem 4 (Bonferroni's Inequality). Given n (> 1) events $A_1, A_2, \ldots, A_n$,

$$\sum_{i=1}^{n} PA_i - \sum_{i<j} P(A_i \cap A_j) \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} PA_i. \tag{7}$$

Proof. In view of (4) it suffices to prove the left side of (7). The proof is by induction. The inequality on the left is true for $n = 2$ since

$$PA_1 + PA_2 - P(A_1 \cap A_2) = P(A_1 \cup A_2).$$

For $n = 3$,

$$P\left(\bigcup_{i=1}^{3} A_i\right) = \sum_{i=1}^{3} PA_i - \sum_{i<j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3),$$

and the result holds. Assuming that (7) holds for $3 < m \leq n-1$, we show that it holds also for $m+1$:

$$\begin{aligned} P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \cup A_{m+1}\right) \\ &= P\left(\bigcup_{i=1}^{m} A_i\right) + PA_{m+1} - P\left(A_{m+1} \cap \left(\bigcup_{1}^{m} A_i\right)\right) \end{aligned}$$

$$\geq \sum_{i=1}^{m+1} PA_i - \sum_{i<j}^{m} P(A_i \cap A_j) - P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right)$$
$$\geq \sum_{i=1}^{m+1} PA_i - \sum_{i<j}^{m} P(A_i \cap A_j) - \sum_{i=1}^{m} P(A_i \cap A_{m+1})$$
$$= \sum_{i=1}^{m+1} PA_i - \sum_{i<j}^{m+1} P(A_i \cap A_j).$$

Theorem 5 (Boole's Inequality). For any two events, A and B,

$$P(A \cap B) \geq 1 - PA^c - PB^c. \tag{8}$$

Corollary 1. Let $\{A_j\}, j = 1, 2, \ldots,$ be a countable sequence of events; then

$$P(\cap A_j) \geq 1 - \sum P(A_j^c). \tag{9}$$

Proof. Take

$$B = \bigcap_{j=2}^{\infty} A_j \quad \text{and} \quad A = A_1$$

in (8).

Corollary 2 (The Implication Rule). If $A, B, C \in \mathcal{S}$ and A and B imply C, then

$$PC^c \leq PA^c + PB^c. \tag{10}$$

Let $\{A_n\}$ be a sequence of sets. The set of all points $\omega \in \Omega$ that belong to A_n for infinitely many values of n is known as the *limit superior* of the sequence and is denoted by

$$\limsup_{n\to\infty} A_n \quad \text{or} \quad \overline{\lim_{n\to\infty}} A_n.$$

The set of all points that belong to A_n for all but a finite number of values of n is known as the *limit inferior* of the sequence $\{A_n\}$ and is denoted by

$$\liminf_{n\to\infty} A_n \quad \text{or} \quad \underline{\lim}_{n\to\infty} A_n.$$

If

$$\underline{\lim}_{n\to\infty} A_n = \overline{\lim_{n\to\infty}} A_n,$$

we say that the limit exists and write $\lim_{n\to\infty} A_n$ for the common set and call it the *limit set*.

We have

$$\varliminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k \subseteq \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \varlimsup_{n\to\infty} A_n.$$

If the sequence $\{A_n\}$ is such that $A_n \subseteq A_{n+1}$, for $n = 1, 2, \ldots$, it is called *nondecreasing*; if $A_n \supseteq A_{n+1}$, $n = 1, 2, \ldots$, it is called *nonincreasing*. If the sequence A_n is nondecreasing, we write $A_n \not\downarrow$; if A_n is *nonincreasing*, we write $A_n \not\uparrow$. Clearly, if $A_n \not\uparrow$ or $A_n \not\downarrow$, the limit exists and we have

$$\lim_n A_n = \bigcup_{n=1}^{\infty} A_n \quad \text{if } A_n \not\downarrow$$

and

$$\lim_n A_n = \bigcap_{n=1}^{\infty} A_n \quad \text{if } A_n \not\uparrow .$$

Theorem 6. Let $\{A_n\}$ be a nondecreasing sequence of events in $\mathcal{S}$, that is, $A_n \in \mathcal{S}$, $n = 1, 2, \ldots$, and

$$A_n \supseteq A_{n-1}, \quad n = 2, 3, \ldots.$$

Then

$$\lim_{n\to\infty} PA_n = P\left(\lim_{n\to\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right). \tag{11}$$

Proof. Let

$$A = \bigcup_{j=1}^{\infty} A_j.$$

Then

$$A = A_n + \sum_{j=n}^{\infty} (A_{j+1} - A_j).$$

By countable additivity we have

$$PA = PA_n + \sum_{j=n}^{\infty} P(A_{j+1} - A_j).$$

and letting $n \to \infty$, we see that

$$PA = \lim_{n\to\infty} PA_n + \lim_{n\to\infty} \sum_{j=n}^{\infty} P(A_{j+1} - A_j).$$