# Probability and its Applications

*Published in association with the Applied Probability Trust*

# Probability and its Applications

*Azencott et al.:* Series of Irregular Observations. Forecasting and Model Building. 1986
*Bass:* Diffusions and Elliptic Operators. 1997
*Bass:* Probabilistic Techniques in Analysis. 1995
*Berglund/Gentz:* Noise-Induced Phenomena in Slow-Fast Dynamical Systems: A Sample-Paths Approach. 2006
*Biagini/Hu/Øksendal/Zhang:* Stochastic Calculus for Fractional Brownian Motion and Applications. 2008
*Chen:* Eigenvalues, Inequalities and Ergodic Theory. 2005
*Costa/Fragoso/Marques:* Discrete-Time Markov Jump Linear Systems. 2005
*Daley/Vere-Jones:* An Introduction to the Theory of Point Processes I: Elementary Theory and Methods. 2nd ed. 2003, corr. 2nd printing 2005
*Daley/Vere-Jones:* An Introduction to the Theory of Point Processes II: General Theory and Structure. 2nd ed. 2008
*de la Peña/Gine:* Decoupling: From Dependence to Independence, Randomly Stopped Processes, U-Statistics and Processes, Martingales and Beyond. 1999
*de la Peña/Lai/Shao:* Self-Normalized Processes. 2009
*Del Moral:* Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications. 2004
*Durrett:* Probability Models for DNA Sequence Evolution. 2002, 2nd ed. 2008
*Ethier:* The Doctrine of Chances. Probabilistic Aspects of Gambling. 2010
*Feng:* The Poisson–Dirichlet Distribution and Related Topics. 2010
*Galambos/Simonelli:* Bonferroni-Type Inequalities with Equations. 1996
*Gani (ed.):* The Craft of Probabilistic Modelling. A Collection of Personal Accounts. 1986
*Gut:* Stopped Random Walks. Limit Theorems and Applications. 1987
*Guyon:* Random Fields on a Network. Modeling, Statistics and Applications. 1995
*Kallenberg:* Foundations of Modern Probability. 1997, 2nd ed. 2002
*Kallenberg:* Probabilistic Symmetries and Invariance Principles. 2005
*Last/Brandt:* Marked Point Processes on the Real Line. 1995
*Molchanov:* Theory of Random Sets. 2005
*Nualart:* The Malliavin Calculus and Related Topics, 1995, 2nd ed. 2006
*Rachev/Rueschendorf:* Mass Transportation Problems. Volume I: Theory and Volume II: Applications. 1998
*Resnick:* Extreme Values, Regular Variation and Point Processes. 1987
*Schmidli:* Stochastic Control in Insurance. 2008
*Schneider/Weil:* Stochastic and Integral Geometry. 2008
*Shedler:* Regeneration and Networks of Queues. 1986
*Silvestrov:* Limit Theorems for Randomly Stopped Stochastic Processes. 2004
*Thorisson:* Coupling, Stationarity and Regeneration. 2000

Stewart N. Ethier

# The Doctrine of Chances

## Probabilistic Aspects of Gambling

Springer

Stewart N. Ethier
Department of Mathematics
University of Utah
155 South 1400 East
Salt Lake City UT 84112-0090
USA
ethier@math.utah.edu

*Series Editors:*



ISSN 1431-7028
ISBN 978-3-540-78782-2 e-ISBN 978-3-540-78783-9
DOI 10.1007/978-3-540-78783-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010927487

Mathematics Subject Classification (2010): 60-02, 91A60, 60G40, 60C05



*Cover design:* WMXDesign

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

> I have found many thousands more readers than I ever looked for. I have no right to say to these, You shall not find fault with my art, or fall asleep over my pages; but I ask you to believe that this person writing strives to tell the truth. If there is not that, there is nothing.
>
> William Makepeace Thackeray, *The History of Pendennis*

This is a monograph/textbook on the probabilistic aspects of gambling, intended for those already familiar with probability at the post-calculus, pre-measure-theory level.

Gambling motivated much of the early development of probability theory (David 1962).[1] Indeed, some of the earliest works on probability include Girolamo Cardano's [1501–1576] *Liber de Ludo Aleae* (*The Book on Games of Chance*, written c. 1565, published 1663), Christiaan Huygens's [1629–1695] "De ratiociniis in ludo aleae" ("On reckoning in games of chance," 1657), Jacob Bernoulli's [1654–1705] *Ars Conjectandi* (*The Art of Conjecturing*, written c. 1690, published 1713), Pierre Rémond de Montmort's [1678–1719] *Essay d'analyse sur les jeux de hasard* (*Analytical Essay on Games of Chance*, 1708, 1713), and Abraham De Moivre's [1667–1754] *The Doctrine of Chances* (1718, 1738, 1756). Gambling also had a major influence on 20th-century probability theory, as it provided the motivation for the concept of a martingale.

Thus, gambling has contributed to probability theory. Conversely, probability theory has contributed much to gambling, from the gambler's ruin formula of Blaise Pascal [1623–1662] to the optimality of bold play due to Lester E. Dubins [1920–2010] and Leonard J. Savage [1917–1971]; from the solution of le her due to Charles Waldegrave to the solution of chemin de fer due to John G. Kemeny [1926–1992] and J. Laurie Snell [1925–]; from the duration-of-play formula of Joseph-Louis Lagrange [1736–1813] to the optimal proportional betting strategy of John L. Kelly, Jr. [1923–1965]; and from

[1] See Maistrov (1974, Chapter 1, Section 2) for a different point of view.

the first evaluation of the banker's advantage at trente et quarante due to Siméon-Denis Poisson [1781–1840] to the first published card-counting system at twenty-one due to Edward O. Thorp [1932–]. Topics such as these are the principal focus of this book.

Is gambling a subject worthy of academic study? Let us quote an authority from the 18th century on this question. In the preface to *The Doctrine of Chances*, De Moivre (1718, p. iii) wrote,

> Another use to be made of this Doctrine of Chances is, that it may serve in Conjunction with the other parts of the Mathematicks, as a fit introduction to the Art of Reasoning; it being known by experience that nothing can contribute more to the attaining of that Art, than the consideration of a long Train of Consequences, rightly deduced from undoubted Principles, of which this Book affords many Examples.

We also quote a 20th-century authority on the same question. In *Le jeu, la chance et le hasard*, Louis Bachelier [1870–1946] (1914, p. 6) wrote,[2]

> It is almost always gambling that enables one to form a fairly clear idea of a manifestation of chance; it is gambling that gave birth to the calculus of probability; it is to gambling that this calculus owes its first faltering utterances and its most recent developments; it is gambling that allows us to conceive of this calculus in the most general way; it is, therefore, gambling that one must strive to understand, but one should understand it in a philosophic sense, free from all vulgar ideas.

Certainly, there are other applications of probability theory on which courses of study could be based, and some of them (e.g., actuarial science, financial engineering) may offer better career prospects than does gambling! But gambling is one of the only applications in which the probabilistic models are often *exactly* correct.[3] This is due to the fundamental simplicity of the nature of the randomness in games of chance. This simplicity translates into an elegance that few other applications enjoy.

The book consists of two parts. Part I ("Theory") begins with a review of probability, then turns to several probability topics that are often not covered in a first course (conditional expectation, martingales, and Markov chains), then briefly considers game theory, and finally concludes with various gambling topics (house advantage, gambler's ruin, betting systems, bold play, optimal proportional play, and card theory). Part II ("Applications") discusses a variety of casino games, including six games in which successive coups are independent (slot machines, roulette, keno, craps, house-banked poker, and video poker) and four games with dependence among coups (faro, baccarat, trente et quarante, and twenty-one). Within each group, chapters are ordered according to difficulty but are largely independent of one another and can be read in any order. We conclude with a discussion of poker, which is in a class by itself.

---

[2] Translation from Dubins and Savage (1976).

[3] Here, and throughout the book (perhaps with the exception of Section 13.2), we model the ideal, or *benchmark*, game, the game as it is intended to be played by the manufacturer of the dice, cards, wheels, machines, etc.

The only contemporary book with comparable content and prerequisites is Richard A. Epstein's [1927–] *The Theory of Gambling and Statistical Logic* (1967, 1977, 2009). Epstein's book is fun to read but is not entirely suitable as a textbook: It is a compendium of results, often without derivations, and there are few problems or exercises to reinforce the reader's understanding. Our aim was not only to supply the missing material but to provide more-self-contained and more-comprehensive coverage of the principal topics. We have tried to do this without sacrificing the "fun to read" factor.

Although there is enough material here for a two-semester course, the book could be used for a one-semester course, either by covering some subset of the chapters thoroughly (perhaps assigning other chapters as individual projects) or by covering every chapter less than thoroughly. In an NSF-sponsored Research Experience for Undergraduates (REU) summer program at the University of Utah in 2005, we adopted the latter approach using a preliminary draft of the book. Fred M. Hoppe, in a course titled "Probability and Games of Chance" at McMaster University in spring 2009, adopted the former approach, covering Chapters 1, 2, 17, 3, 15, and 6 in that order.

The book is not intended solely for American and Canadian readers. Money is measured in units, not dollars, and European games, such as chemin de fer and trente et quarante, are studied. This is appropriate, inasmuch as France is not only the birthplace of probability theory but also that of roulette, faro, baccarat, trente et quarante, and twenty-one.

With few exceptions, all random variables in the book are discrete.[4] This allows us to provide a mathematically rigorous treatment, while avoiding the need for measure theory except for occasional references to the Appendix. Each chapter contains a collection of problems that range from straightforward to challenging. Some require computing. Answers, but not solutions, will be provided at the author's web page (http://www.math.utah.edu/~ethier/). While we have not hesitated to use computing in the text (in fact, it is a necessity in studying such topics as video poker, twenty-one, and Texas hold'em), we have avoided the use of computer simulation, which seems to us outside the spirit of the subject. Each chapter also contains a set of historical notes, in which credit is assigned wherever possible and to the best of our knowledge. This has necessitated a lengthy bibliography. In many cases we simply do not know who originated a particular idea, so a lack of attribution should not be interpreted as a claim of originality. We frequently refer to the generic gambler, bettor, player, dealer, etc. with the personal pronoun "he," which has the old-fashioned interpretation of, but is much less awkward than, "he or she."

A year or two ago (2008) was the tercentenary of the publication of the first edition of Montmort's *Analytical Essay on Games of Chance*, which can

[4] The only exceptions are nondiscrete limits of sequences of discrete random variables. These may occur, for example, in the martingale convergence theorem.

be regarded as the first published full-length book on probability theory.[5] As Todhunter (1865, Article 136) said of Montmort,

> In 1708 he published his work on Chances, where with the courage of Columbus he revealed a new world to mathematicians.

A decade later De Moivre published his equally groundbreaking work, *The Doctrine of Chances*. Either title would be suitable for the present book; we have chosen the latter because it sounds a little less intimidating.

*Acknowledgments*: I am grateful to Nelson H. F. Beebe for technical advice and assistance and to Davar Khoshnevisan for valuable discussions. Portions of the book were read by Patrik Andersson, R. Michael Canjar, Anthony Curtis, Anirban DasGupta, Persi Diaconis, Marc Estafanous, Robert C. Hannum, Fred M. Hoppe, Robert Muir, Don Schlesinger, and Edward O. Thorp, as well as by several anonymous reviewers for Springer and AMS. I thank them for their input. A fellowship from the Center for Gaming Research at the University of Nevada, Las Vegas, allowed me to spend a month in the Special Collections room of the UNLV Lied Library, and the assistance of the staff is much appreciated.

I would also like to thank several others who helped in various ways during the preparation of this book. These include David R. Bellhouse, István Berkes, Mr. Cacarulo, Renzo Cavalieri, Bob Dancer, Régis Deloche, Edward G. Dunne, Marshall Fey, Carlos Gamez, Susan E. Gevaert, James Grosjean, Norm Hellmers, Thomas M. Kavanagh, David A. Levin, Basil Nestor, Marina Reizakis, Michael W. Shackleford, Larry Shepp, Arnold Snyder, George Stamos, and Zenfighter.

Finally, I am especially grateful to my wife, Kyoko, for her patience throughout this lengthy project.

*Dedication*: The book is dedicated to the memory of gambling historian Russell T. Barnhart [1926–2003] and twenty-one theorist Peter A. Griffin [1937–1998], whom I met in 1984 and 1981, respectively. Their correspondence about gambling matters over the years fills several thick folders (neither used e-mail), and their influence on the book is substantial.

Salt Lake City, December 2009 Stewart N. Ethier

[5] Cardano's *Liber de Ludo Aleae* comprises only 15 (dense) pages of his *Opera omnia* and Huygens's "De ratiociniis in ludo aleae" comprises only 18 pages of van Schooten's *Exercitationum Mathematicarum*.

# Contents

# List of Notation

| symbol | meaning | page of first use |
|---|---|---|
| ♠ | end of proof or end of example[1] | 6 |
| $:=$ | equals by definition ($=:$ is also used) | 3 |
| $\equiv$ | is identically equal to, or is congruent to | 59 |
| $\approx$ | is approximately equal to | 9 |
| $\sim$ | is asymptotic to ($a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$) | 48 |
| $\mathbf{N}$ | the set of positive integers | 58 |
| $\mathbf{Z}_+$ | the set of nonnegative integers | 91 |
| $\mathbf{Z}$ | the set of integers | 58 |
| $\mathbf{Q}$ | the set of rational numbers | 108 |
| $\mathbf{R}$ | the set of real numbers | 21 |
| $\lvert x\rvert$ | absolute value of real $x$; modulus of complex $x$ | 6 |
| $x^+$ | nonnegative part of the real number $x$ ($:= \max(x, 0)$) | 22 |
| $x^-$ | nonpositive part of the real number $x$ ($:= -\min(x, 0)$) | 64 |
| $\boldsymbol{x} \cdot \boldsymbol{y}$ | inner product of $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{R}^d$ | 34 |
| $\lvert\boldsymbol{x}\rvert$ | Euclidean norm of vector $\boldsymbol{x} \in \boldsymbol{R}^d$ ($:= (\boldsymbol{x} \cdot \boldsymbol{x})^{1/2}$) | 177 |
| $\boldsymbol{A}^\top$ | transpose of matrix (or vector) $\boldsymbol{A}$ | 176 |
| $0.4\overline{92}$ | $0.4929292\cdots$ (repeating decimal expansion) | 19 |
| $\mathrm{sgn}(x)$ | sign of $x$ ($:= 1, 0, -1$ if $x > 0, = 0, < 0$) | 170 |
| $\lvert A\rvert$ | cardinality of (number of elements of) the finite set $A$ | 3 |
| $x \in A$ | $x$ is an element of $A$ | 3 |
| $A \subset B$ | $A$ is a (not necessarily proper) subset of $B$ | 3 |
| $B \supset A$ | equivalent to $A \subset B$ | 13 |

[1] In discussions of card games, the symbol ♠ signifies a spade.

| symbol | meaning | page of first use |
|---|---|---|
| $A \cup B$ | union of $A$ and $B$ | 11 |
| $A \cap B$ | intersection of $A$ and $B$ | 11 |
| $A^c$ | complement of $A$ | 11 |
| $A - B$ | set-theoretic difference $(:= A \cap B^c)$ | 11 |
| $A \times B$ | cartesian product of $A$ and $B$ | 15 |
| $A^n$ | $n$-fold cartesian product $A \times \cdots \times A$ | 25 |
| $1_A$ | indicator r.v. of event $A$ or indicator function of set $A$ | 33 |
| $x \vee y$ | $\max(x, y)$ | 56 |
| $x \wedge y$ | $\min(x, y)$ | 6 |
| $\ln x$ | natural (base $e$) logarithm of $x$ | 30 |
| $\log_2 x$ | base-2 logarithm of $x$ | 48 |
| $\lfloor x \rfloor$ | the greatest integer less than or equal to $x$ | 39 |
| $\lceil x \rceil$ | the least integer greater than or equal to $x$ | 30 |
| $\varnothing$ | empty set | 12 |
| $(n)_k$ | $:= n(n-1)\cdots(n-k+1)$ if $k \geq 1$, $(n)_0 := 1$ | 4 |
| $n!$ | $n$ factorial $(:= (n)_n = n(n-1)\cdots 1$ if $n \geq 1$, $0! := 1)$ | 4 |
| $\binom{n}{k}$ | binomial coefficient $n$ choose $k$ $(:= (n)_k/k!)$ | 4 |
| $\binom{n}{n_1,\ldots,n_r}$ | multinomial coefficient $(:= n!/(n_1!\cdots n_r!))$ | 5 |
| $\delta_{i,j}$ | Kronecker delta $(:= 1$ if $i = j$, $:= 0$ otherwise) | 20 |
| $(a, b)$ | open interval $\{x : a < x < b\}$ | 22 |
| $[a, b)$ | half-open interval $\{x : a \leq x < b\}$ | 27 |
| $(a, b]$ | half-open interval $\{x : a < x \leq b\}$ | 140 |
| $[a, b]$ | closed interval $\{x : a \leq x \leq b\}$ | 21 |
| $\stackrel{d}{\to}$ | converges in distribution to | 42 |
| $N(0, 1)$ | standard-normal distribution | 42 |
| $\Phi(x)$ | standard-normal cumulative distribution function | 42 |
| $\phi(x)$ | standard-normal density function $(:= \Phi'(x))$ | 63 |
| a.s. | almost surely | 42 |
| i.o. | infinitely often | 43 |
| i.i.d. | independent and identically distributed | 43 |
| g.c.d. | greatest common divisor | 137 |

Positive, negative, increasing, and decreasing are used only in the strict sense. When the weak sense is intended, we use nonnegative, nonpositive, nondecreasing, and nonincreasing, respectively.

# Part I
# Theory

# Chapter 1
# Review of Probability

> Mr. Arthur Pendennis did not win much money in these transactions with Mr. Bloundell, or indeed gain good of any kind except a knowledge of the odds at hazard, which he might have learned out of books.
>
> William Makepeace Thackeray, *The History of Pendennis*

The reader is assumed to be familiar with basic probability, and here we provide the definitions and theorems, without proofs, for easy reference. We restrict our attention to discrete random variables but not necessarily to discrete sample spaces. A number of examples are worked out in detail, and problems are provided for those who need additional review.

## 1.1 Combinatorics and Probability

The set $\Omega$ (omega) of all possible outcomes of a random experiment is called the *sample space*. Let us first consider the case in which $\Omega$ is finite. Let $n \geq 2$, let $\Omega = \{o_1, o_2, \ldots, o_n\}$, let $p_1, p_2, \ldots, p_n$ be positive numbers that sum to 1, and assign probability $p_i$ to outcome $o_i$ for $i = 1, 2, \ldots, n$. An *event* $E$ is a subset of $\Omega$, and the *probability* of an event $E \subset \Omega$ is defined to be the sum of the probabilities of the outcomes in $E$:

$$\mathrm{P}(E) := \sum_{1 \leq i \leq n:\ o_i \in E} p_i. \tag{1.1}$$

This leads to possibly the oldest result in probability theory:

**Theorem 1.1.1.** *Under the assumption that all outcomes in a finite sample space $\Omega$ are equally likely, the probability of an event $E \subset \Omega$ is given by*

$$\mathrm{P}(E) = \frac{|E|}{|\Omega|}, \tag{1.2}$$

S.N. Ethier, *The Doctrine of Chances*, Probability and its Applications,
DOI 10.1007/978-3-540-78783-9_1, 

*where $|E|$ denotes the cardinality of (or the number of outcomes in) $E$.*

The only difficulty in applying Theorem 1.1.1 is in counting the numbers of outcomes in $E$ and in $\Omega$. This can often be done with the help of *combinatorial analysis*, with which the next seven theorems are concerned.

**Theorem 1.1.2.** *Consider a task that requires completing $r$ subtasks in order, where $r \geq 2$. Suppose there are $n_1$ ways to complete the first subtask; no matter which way is chosen, there are $n_2$ ways to complete the second subtask; no matter which ways are chosen for the first two subtasks, there are $n_3$ ways to complete the third subtask; . . . no matter which ways are chosen for the first $r-1$ subtasks, there are $n_r$ ways to complete the $r$th subtask. Then there are $n_1 n_2 \cdots n_r$ ways to complete the task.*

We define

$$n! := n(n-1)\cdots 2 \cdot 1 \tag{1.3}$$

for each positive integer $n$. The symbol $n!$ is read "$n$ factorial." It will be convenient to define $0! := 1$.

**Theorem 1.1.3.** *The number of permutations of $n$ distinct items taken $k$ at a time (i.e., the number of ways to choose $k$ out of $n$ distinct items, taking the order in which the items are chosen into account) is*

$$(n)_k := n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}, \tag{1.4}$$

*assuming $1 \leq k \leq n$.*

Notice that $n! = (n)_n$. Thus, $n!$ is the number of permutations of $n$ distinct items. It will be convenient to define $(n)_0 := 1$ for each nonnegative integer $n$.

**Theorem 1.1.4.** *The number of combinations of $n$ distinct items taken $k$ at a time (i.e., the number of ways to choose $k$ out of $n$ distinct items, without regard to the order in which the items are chosen) is*

$$\binom{n}{k} := \frac{(n)_k}{k!} = \frac{n!}{k!\,(n-k)!}, \tag{1.5}$$

*assuming $0 \leq k \leq n$.*

The symbol $\binom{n}{k}$ is read "$n$ choose $k$" and is called a *binomial coefficient*. It is useful to be aware that

$$\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{1} = \binom{n}{n-1} = n, \qquad \binom{n}{k} = \binom{n}{n-k}. \tag{1.6}$$

Another useful identity is the one that generates Pascal's triangle, namely

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \tag{1.7}$$

**Table 1.1** The first seven rows of Pascal's triangle. Row $n+1$ contains the $n+1$ binomial coefficients $\binom{n}{0}, \dots, \binom{n}{n}$.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | | | | | | |
| | | | | | 1 | | 1 | | | | | |
| | | | | 1 | | 2 | | 1 | | | | |
| | | | 1 | | 3 | | 3 | | 1 | | | |
| | | 1 | | 4 | | 6 | | 4 | | 1 | | |
| | 1 | | 5 | | 10 | | 10 | | 5 | | 1 | |
| 1 | | 6 | | 15 | | 20 | | 15 | | 6 | | 1 |

for $1 \le k \le n$. See Table 1.1 for Pascal's triangle. The reason that the quantities $\binom{n}{k}$ are called binomial coefficients is that they appear as coefficients in the *binomial theorem*:

**Theorem 1.1.5.** *For all real $a$ and $b$ and positive integers $n$,*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{1.8}$$

*where* $0^0 := 1$.

**Theorem 1.1.6.** *The number of ways in which to partition a set of $n$ distinct items into $r$ specified subsets, the first having $n_1 \ge 0$ elements, the second having $n_2 \ge 0$ elements, …, the $r$th having $n_r \ge 0$ elements, where $n_1 + n_2 + \cdots + n_r = n$, is*

$$\binom{n}{n_1, n_2, \dots, n_r} := \frac{n!}{n_1!\, n_2! \cdots n_r!}. \tag{1.9}$$

*This is also the number of permutations of $n$ items of $r$ distinct types, with $n_1$ of the first type, $n_2$ of the second type, …, $n_r$ of the $r$th type.*

The quantities $\binom{n}{n_1,\dots,n_r}$ are called *multinomial coefficients* because they appear as coefficients in the *multinomial theorem*:

**Theorem 1.1.7.** *For all real $a_1, a_2, \dots, a_r$ and positive integers $n$,*

$$(a_1 + \cdots + a_r)^n = \sum \cdots \sum_{n_1 \ge 0, \dots, n_r \ge 0:\ n_1 + \cdots + n_r = n} \binom{n}{n_1, \dots, n_r} a_1^{n_1} \cdots a_r^{n_r}, \tag{1.10}$$

*where* $0^0 := 1$.

Of course, the special case of the multinomial theorem in which $r = 2$ coincides with the binomial theorem, and multinomial coefficients with $r = 2$

are usually written as binomial coefficients. For example, $\binom{52}{5}$ is preferred to $\binom{52}{5,47}$.

**Theorem 1.1.8.** *The number of terms in the sum in (1.10) is* $\binom{n+r-1}{r-1}$ *or, equivalently,* $\binom{n+r-1}{n}$. *This is also the number of ways to distribute* $n$ *indistinguishable balls into* $r$ *specified urns.*

Of course, the number of ways to distribute $n$ *distinguishable* balls into $r$ specified urns is $r^n$ by Theorem 1.1.2.

**Example 1.1.9.** *Two-dice totals* When rolling a pair of indistinguishable dice (e.g., two red dice), the number of distinguishable outcomes is $\binom{2+6-1}{2} = 21$ by Theorem 1.1.8, but these outcomes are not equally likely. On the other hand, when rolling a pair of distinguishable dice (e.g., one red die and one green die), the number of distinguishable outcomes is $6 \cdot 6 = 36$ by Theorem 1.1.2, and these outcomes *are* equally likely. We list them in Table 1.2, together with the dice totals and their probabilities.

**Table 1.2** The results of tossing two distinguishable dice.

| outcomes | total | probability |
|---|---|---|
| $(1,1)$ | 2 | 1/36 |
| $(1,2),(2,1)$ | 3 | 2/36 |
| $(1,3),(2,2),(3,1)$ | 4 | 3/36 |
| $(1,4),(2,3),(3,2),(4,1)$ | 5 | 4/36 |
| $(1,5),(2,4),(3,3),(4,2),(5,1)$ | 6 | 5/36 |
| $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$ | 7 | 6/36 |
| $(2,6),(3,5),(4,4),(5,3),(6,2)$ | 8 | 5/36 |
| $(3,6),(4,5),(5,4),(6,3)$ | 9 | 4/36 |
| $(4,6),(5,5),(6,4)$ | 10 | 3/36 |
| $(5,6),(6,5)$ | 11 | 2/36 |
| $(6,6)$ | 12 | 1/36 |

We conclude that the probability $\pi_j$ of rolling a total of $j \in \{2,3,4,\ldots,12\}$ is given by

$$\pi_j = \frac{(j-1) \wedge (13-j)}{36} = \frac{6-|j-7|}{36}, \tag{1.11}$$

a formula that will be cited frequently in the sequel. Clearly, the probabilities (1.11) are not affected by the colors of the dice, so (1.11) is equally valid for a pair of indistinguishable dice. ♠

**Example 1.1.10.** *Poker hands*. Poker is played with a standard 52-card deck. By such a deck we mean that each card is described by its *denomination*, namely 2, 3, 4, 5, 6, 7, 8, 9, 10, J (jack), Q (queen), K (king), or A (ace), and its *suit*, namely ♣ (club), ♢ (diamond), ♡ (heart), or ♠ (spade). Sometimes we will denote denomination 10 by T to avoid two-digit numbers. A poker hand consists of five cards. A hand is said to be *in sequence* if its denominations consist (after rearrangement if necessary) of 5-4-3-2-A or 6-5-4-3-2 or ... or A-K-Q-J-T. Notice that the ace plays a special role, in that it can appear as either the low card or the high card in a hand that is in sequence. A *straight flush* contains five cards in sequence and of the same suit, a *flush* contains five cards of the same suit but not in sequence, and a *straight* contains five cards in sequence but not of the same suit. A *royal flush* is an ace-high straight flush.

To describe the other types of hands, we let $d_0, d_1, d_2, d_3, d_4$ denote, respectively, the numbers of denominations in a hand represented $0, 1, 2, 3, 4$ times, and we note that

$$d_0 + d_1 + d_2 + d_3 + d_4 = 13, \qquad d_1 + 2d_2 + 3d_3 + 4d_4 = 5. \tag{1.12}$$

We define $\boldsymbol{d} := (d_0, d_1, d_2, d_3, d_4)$ to be the *denomination multiplicity vector* of the hand. A hand is ranked *four of a kind* if $\boldsymbol{d} = (11, 1, 0, 0, 1)$, *full house* if $\boldsymbol{d} = (11, 0, 1, 1, 0)$, *three of a kind* if $\boldsymbol{d} = (10, 2, 0, 1, 0)$, *two pair* if $\boldsymbol{d} = (10, 1, 2, 0, 0)$, *one pair* if $\boldsymbol{d} = (9, 3, 1, 0, 0)$, and *no pair* if $\boldsymbol{d} = (8, 5, 0, 0, 0)$ and if the five cards are neither in sequence nor of the same suit. For example, the hand consisting of A♠-A♣-8♠-8♣-9♢ has two denominations (A and 8) represented twice and one denomination (9) represented once; the remaining 10 denominations are not represented. Thus, $d_2 = 2$, $d_1 = 1$, and $d_0 = 10$, and we see that this hand is ranked two pair.

The probability that a randomly dealt five-card poker hand has denomination multiplicity vector $\boldsymbol{d}$ (equal to $(11, 1, 0, 0, 1)$, $(11, 0, 1, 1, 0)$, $(10, 2, 0, 1, 0)$, $(10, 1, 2, 0, 0)$, $(9, 3, 1, 0, 0)$, or $(8, 5, 0, 0, 0)$) is given by

$$\frac{\binom{13}{d_0, d_1, d_2, d_3, d_4} \prod_{i=1}^4 \binom{4}{i}^{d_i}}{\binom{52}{5}}. \tag{1.13}$$

The multinomial coefficient is the number of ways to choose the hand's denominations, while the product of powers of binomial coefficients is the number of ways to choose the suits for the chosen denominations. (We have omitted the $i = 0$ term in the product because it is unnecessary. The $i = 4$ term is also unnecessary, but we have included it for clarity.) See Table 1.3 for the numerators of these probabilities. By a separate argument, the case $\boldsymbol{d} = (8, 5, 0, 0, 0)$ must be broken down into the four ranks straight flush, flush, straight, and no pair. ♠

In Example 1.1.9 we counted outcomes by *enumeration*, that is, by creating a list. In Example 1.1.10 we counted outcomes with the help of combinatorial analysis. Certainly, the latter approach is more elegant. However, the former approach is sometimes the only viable method. The next example illustrates this point.

**Table 1.3** The five-card poker-hand frequencies. For each expression that is the product of two factors, the first is the number of ways to choose the hand's denominations, and the second is the number of ways to choose the suits for the chosen denominations.

| rank | number of ways | |
|---|---|---|
| straight flush* | $\binom{10}{1}\binom{4}{1}$ | 40 |
| four of a kind | $\binom{13}{11,1,0,0,1}\left[\binom{4}{1}\binom{4}{4}\right]$ | 624 |
| full house | $\binom{13}{11,0,1,1,0}\left[\binom{4}{2}\binom{4}{3}\right]$ | 3,744 |
| flush | $\left[\binom{13}{5}-\binom{10}{1}\right]\binom{4}{1}$ | 5,108 |
| straight | $\binom{10}{1}\left[\binom{4}{1}^5-\binom{4}{1}\right]$ | 10,200 |
| three of a kind | $\binom{13}{10,2,0,1,0}\left[\binom{4}{1}^2\binom{4}{3}\right]$ | 54,912 |
| two pair | $\binom{13}{10,1,2,0,0}\left[\binom{4}{1}\binom{4}{2}^2\right]$ | 123,552 |
| one pair | $\binom{13}{9,3,1,0,0}\left[\binom{4}{1}^3\binom{4}{2}\right]$ | 1,098,240 |
| no pair | $\left[\binom{13}{5}-\binom{10}{1}\right]\left[\binom{4}{1}^5-\binom{4}{1}\right]$ | 1,302,540 |
| sum | $\binom{52}{5}$ | 2,598,960 |

*including royal flush

**Example 1.1.11.** *Twenty-one-dealer sequences*. For the purposes of this example, we need to know only a few of the rules of twenty-one, or blackjack. We assume that the game is dealt from a single standard 52-card deck. Aces have value 1 or 11 as specified below, court cards (J, Q, K) have value 10, and every other card has value equal to its nominal value. Suits are irrelevant. The dealer receives two cards initially (one face up) and additional cards one at a time as needed to achieve a total of 17 or greater. The first ace has value 11 unless that would result in a total greater than 21, in which case it has value 1. Every subsequent ace has value 1. A total that includes an ace valued as 11 is called a *soft total*; every other total is called a *hard total*. For example, if the dealer is dealt $(\mathrm{A}, 5)$, he then has a soft total of 16 and requires another card. If his third card is a 6, he then has a hard total of 12 and requires another card. If his fourth card is a 7, he then has a hard total of 19, which is his final total.

Let us define a *twenty-one-dealer sequence* to be a finite sequence $a_1, \ldots, a_k$ of positive integers, none of which exceeds 10, and at most four of which are equal to 1, at most four of which are equal to 2, and so on, such that $k$ is the smallest integer $j \geq 2$ for which

$$a_1 + \cdots + a_j \geq 17 \tag{1.14}$$

or

$$1 \in \{a_1, \ldots, a_j\} \quad \text{and} \quad 7 \leq a_1 + \cdots + a_j \leq 11. \tag{1.15}$$

Observe that (1.14) signifies a hard total of $a_1 + \cdots + a_j$ and that, since 1s play the role of aces, (1.15) signifies a soft total of $a_1 + \cdots + a_j + 10$. Clearly, the order of the terms is crucial: $8, 8, 10$ is a twenty-one-dealer sequence but $10, 8, 8$ and $8, 10, 8$ are not. In general, if $a_1, \ldots, a_k$ is a twenty-one-dealer sequence, then its length $k$ satisfies $2 \leq k \leq 10$.

How many twenty-one-dealer sequences are there? We do not know how to answer this question using combinatorial analysis. Therefore, we resort to the crude but effective method of enumerating all such sequences. By ordering them in reverse-lexicographical order, we ensure that no sequence is overlooked. The list is displayed in Table 1.4, and we see that the answer to our question is 48,532.

Although the twenty-one-dealer sequences are obviously not equally likely, we can nevertheless apply Theorem 1.1.1 to find the probability of each such sequence. Letting

$$k_j := |\{1 \leq i \leq k : a_i = j\}|, \qquad j = 1, 2, \ldots, 10, \tag{1.16}$$

we find that the probability of the twenty-one-dealer sequence $a_1, a_2, \ldots, a_k$ is

$$\frac{(4)_{k_1}(4)_{k_2} \cdots (4)_{k_9}(16)_{k_{10}}}{(52)_k}. \tag{1.17}$$

Here the random experiment consists merely of dealing out $k$ cards in succession.

We now regard these 48,532 sequences as the outcomes of a random experiment and use Table 1.4 and (1.1) to find the probabilities of the various possible dealer final totals. The totals of interest are 17, 18, 19, 20, and 21, with 22–26 collectively describing a dealer *bust*. Further, a two-card 21 (a *natural*) should be distinguished from a 21 comprising three or more cards.

For example,

$$\begin{aligned} \text{P(dealer has two-card 21)} &= \text{P}(10, 1) + \text{P}(1, 10) \\ &= 2\frac{(4)_1(16)_1}{(52)_2} = \frac{32}{663} \approx 0.048265460. \end{aligned} \tag{1.18}$$

The remaining cases require the use of a computer, and results are displayed in Table 1.5. ♠

We have limited our attention so far to finite sample spaces, but this is far too restrictive. We could extend (1.1) to countably infinite sample spaces, but even that is too restrictive. (Consider, for example, the random

**Table 1.4** A partial list of the 48,532 twenty-one-dealer sequences, in reverse lexicographical order. (Rules: single deck, dealer stands on soft 17.)

| seq. no. | sequence | total | probability |
|---|---|---|---|
| 1 | $10, 10$ | 20 | $(16)_2/(52)_2$ |
| 2 | $10, 9$ | 19 | $(4)_1(16)_1/(52)_2$ |
| 3 | $10, 8$ | 18 | $(4)_1(16)_1/(52)_2$ |
| 4 | $10, 7$ | 17 | $(4)_1(16)_1/(52)_2$ |
| 5 | $10, 6, 10$ | 26 | $(4)_1(16)_2/(52)_3$ |
| 6 | $10, 6, 9$ | 25 | $(4)_1(4)_1(16)_1/(52)_3$ |
| ⋮ | | | |
| 286 | $10, 2, 1, 1, 1, 1, 3$ | 19 | $(4)_4(4)_1(4)_1(16)_1/(52)_7$ |
| 287 | $10, 2, 1, 1, 1, 1, 2$ | 18 | $(4)_4(4)_2(16)_1/(52)_7$ |
| 288 | $10, 1$ | 21 | $(4)_1(16)_1/(52)_2$ |
| 289 | $9, 10$ | 19 | $(4)_1(16)_1/(52)_2$ |
| 290 | $9, 9$ | 18 | $(4)_2/(52)_2$ |
| ⋮ | | | |
| 15,110 | $4, 2, 2, 2, 2, 1, 1, 1, 1, 4$ | 20 | $(4)_4(4)_4(4)_2/(52)_{10}$ |
| 15,111 | $4, 2, 2, 2, 2, 1, 1, 1, 1, 3$ | 19 | $(4)_4(4)_4(4)_1(4)_1/(52)_{10}$ |
| 15,112 | $4, 2, 2, 2, 1$ | 21 | $(4)_1(4)_3(4)_1/(52)_5$ |
| 15,113 | $4, 2, 2, 1$ | 19 | $(4)_1(4)_2(4)_1/(52)_4$ |
| 15,114 | $4, 2, 1$ | 17 | $(4)_1(4)_1(4)_1/(52)_3$ |
| 15,115 | $4, 1, 10, 10$ | 25 | $(4)_1(4)_1(16)_2/(52)_4$ |
| 15,116 | $4, 1, 10, 9$ | 24 | $(4)_1(4)_1(4)_1(16)_1/(52)_4$ |
| ⋮ | | | |
| 42,532 | $2, 1, 1, 1, 1, 6, 2, 2, 3$ | 19 | $(4)_4(4)_3(4)_1(4)_1/(52)_9$ |
| 42,533 | $2, 1, 1, 1, 1, 6, 2, 2, 2$ | 18 | $(4)_4(4)_4(4)_1/(52)_9$ |
| 42,534 | $2, 1, 1, 1, 1, 5$ | 21 | $(4)_4(4)_1(4)_1/(52)_6$ |
| 42,535 | $2, 1, 1, 1, 1, 4$ | 20 | $(4)_4(4)_1(4)_1/(52)_6$ |
| 42,536 | $2, 1, 1, 1, 1, 3$ | 19 | $(4)_4(4)_1(4)_1/(52)_6$ |
| 42,537 | $2, 1, 1, 1, 1, 2$ | 18 | $(4)_4(4)_2/(52)_6$ |
| 42,538 | $1, 10$ | 21 | $(4)_1(16)_1/(52)_2$ |
| 42,539 | $1, 9$ | 20 | $(4)_1(4)_1/(52)_2$ |
| ⋮ | | | |
| 48,527 | $1, 1, 1, 1, 2, 6, 2, 2, 3$ | 19 | $(4)_4(4)_3(4)_1(4)_1/(52)_9$ |
| 48,528 | $1, 1, 1, 1, 2, 6, 2, 2, 2$ | 18 | $(4)_4(4)_4(4)_1/(52)_9$ |
| 48,529 | $1, 1, 1, 1, 2, 5$ | 21 | $(4)_4(4)_1(4)_1/(52)_6$ |
| 48,530 | $1, 1, 1, 1, 2, 4$ | 20 | $(4)_4(4)_1(4)_1/(52)_6$ |
| 48,531 | $1, 1, 1, 1, 2, 3$ | 19 | $(4)_4(4)_1(4)_1/(52)_6$ |
| 48,532 | $1, 1, 1, 1, 2, 2$ | 18 | $(4)_4(4)_2/(52)_6$ |

experiment consisting of an infinite sequence of coin tosses.) Therefore, we take an axiomatic approach in what follows. We begin by introducing the required definitions.

**Table 1.5** The probabilities of the twenty-one dealer's various final totals, rounded to nine decimal places. (Rules: single deck, dealer stands on soft 17.)

| dealer total | no. of sequences | probability |
|---|---|---|
| 17 | 5,134 | .145 829 659 |
| 18 | 5,243 | .138 063 176 |
| 19 | 5,433 | .134 820 214 |
| 20 | 5,455 | .175 806 476 |
| 21[3] | 5,433 | .073 629 613 |
| 21[2] | 2 | .048 265 460 |
| bust | 21,832 | .283 585 403 |
| sum | 48,532 | 1.000 000 000 |

[3]three or more cards [2]two cards (natural)

We define the *union* of two events $E$ and $F$ by

$$E \cup F := \{o \in \Omega : o \in E \text{ or } o \in F \text{ (or both)}\} \tag{1.19}$$

and the *intersection* by

$$E \cap F := \{o \in \Omega : o \in E \text{ and } o \in F\}. \tag{1.20}$$

The *complement* of $E$ is

$$E^c := \{o \in \Omega : o \notin E\}. \tag{1.21}$$

We will also occasionally use

$$F - E := F \cap E^c. \tag{1.22}$$

We can extend the binary operations, union and intersection, to finite or countably infinite collections of events. Given events $E_1, E_2, \ldots$, their *union* is given by

$$E_1 \cup E_2 \cup \cdots = \bigcup_{i=1}^{\infty} E_i := \{o \in \Omega : o \in E_i \text{ for some } i\}, \tag{1.23}$$

and their *intersection* is given by

$$E_1 \cap E_2 \cap \cdots = \bigcap_{i=1}^{\infty} E_i := \{o \in \Omega : o \in E_i \text{ for every } i\}. \tag{1.24}$$

Incidentally, unions, intersections, and complements apply also to arbitrary sets (not just events) and will occasionally be used in that way. The operations (1.21), (1.23), and (1.24) are related by *De Morgan's laws*:

$$\left(\bigcup_{i=1}^{\infty} E_i\right)^c = \bigcap_{i=1}^{\infty} E_i^c, \qquad \left(\bigcap_{i=1}^{\infty} E_i\right)^c = \bigcup_{i=1}^{\infty} E_i^c. \tag{1.25}$$

Given events $E_1, E_2, \ldots,$ we say that they are *mutually exclusive* (or *pairwise disjoint*) if no two of them can occur simultaneously, that is, if $E_i \cap E_j = \varnothing$ for all $i \neq j$. We can now state the four *axioms of probability*:

**Axiom 1.1.12.** *The collection of events contains the sample space $\Omega$ and is closed under complementation and under countable unions.*

**Axiom 1.1.13.** $\mathrm{P}(E) \geq 0$ *for every event $E$.*

**Axiom 1.1.14.** *If $E_1, E_2, \ldots$ are mutually exclusive events, then*

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(E_i). \tag{1.26}$$

**Axiom 1.1.15.** $\mathrm{P}(\Omega) = 1$.

When $\Omega$ is finite or countably infinite, it is possible to define the collection of events to be the collection of all subsets of $\Omega$. But if $\Omega$ is uncountable, such a definition leads to complications, so instead we simply adopt Axiom 1.1.12. By that axiom and De Morgan's laws, the collection of events is also closed under countable intersections. Axiom 1.1.14 is called *countable additivity*. If we take $E_1 = E_2 = \cdots = \varnothing$ in Axiom 1.1.14 and use Axiom 1.1.15, we find that $\mathrm{P}(\varnothing) = 0$. It follows that countable additivity implies *finite additivity*: If $n \geq 2$ and $E_1, E_2, \ldots, E_n$ are mutually exclusive events, then

$$\mathrm{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathrm{P}(E_i). \tag{1.27}$$

Notice that our definition (1.1), with every subset $E \subset \Omega$ being an event, satisfies the axioms.

These axioms allow us to establish several useful theorems, the first of which is concerned with the monotonicity of probability.

**Theorem 1.1.16.** *If $E$ and $F$ are events with $E \subset F$, then* $\mathrm{P}(E) \leq \mathrm{P}(F)$, *and in fact* $\mathrm{P}(F - E) = \mathrm{P}(F) - \mathrm{P}(E)$.

**Corollary 1.1.17.** $\mathrm{P}(E^c) = 1 - \mathrm{P}(E)$ *for every event $E$.*

This simple corollary is called the *complementation law*. The next result is known as the *inclusion-exclusion law*. It generalizes the familiar formula

$$\mathrm{P}(E_1 \cup E_2) = \mathrm{P}(E_1) + \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2). \tag{1.28}$$

**Theorem 1.1.18.** *Given events* $E_1, E_2, \ldots, E_n$, *define*

$$S_1 := \sum_{i=1}^{n} \mathrm{P}(E_i), \qquad S_2 := \sum\sum_{1 \le i < j \le n} \mathrm{P}(E_i \cap E_j), \tag{1.29}$$

*and so on. More generally, for* $1 \le m \le n$, *define*

$$S_m := \sum \cdots \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} \mathrm{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}). \tag{1.30}$$

*Then*

$$\mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{m=1}^{n} (-1)^{m-1} S_m. \tag{1.31}$$

The next result contains several inequalities related to the inclusion-exclusion law. The first is often called *Boole's inequality.*

**Theorem 1.1.19.** *Under the assumptions and notation of Theorem 1.1.18,*

$$\mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \le S_1, \tag{1.32}$$
$$\mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \ge S_1 - S_2, \tag{1.33}$$
$$\mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \le S_1 - S_2 + S_3, \tag{1.34}$$
$$\mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \ge S_1 - S_2 + S_3 - S_4, \tag{1.35}$$

*and so on.*

The last theorem in this section is the first that requires countable additivity.

**Theorem 1.1.20.** (a) *Given a sequence of events satisfying* $E_1 \subset E_2 \subset \cdots$,

$$\mathrm{P}\bigg(\bigcup_{i=1}^{\infty} E_i\bigg) = \lim_{i \to \infty} \mathrm{P}(E_i). \tag{1.36}$$

(b) *Given a sequence of events satisfying* $E_1 \supset E_2 \supset \cdots$,

$$\mathrm{P}\bigg(\bigcap_{i=1}^{\infty} E_i\bigg) = \lim_{i \to \infty} \mathrm{P}(E_i). \tag{1.37}$$

Finally, we generalize the first inequality in Theorem 1.1.19. The result is called *countable subadditivity*.

**Corollary 1.1.21.** *Given an arbitrary sequence of events* $E_1, E_2, \ldots$,

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathrm{P}(E_i). \tag{1.38}$$

**Example 1.1.22.** *Méré's problem.* In 1654, the Chevalier de Méré raised the question of whether the probability of at least one six in four tosses of a single die is equal to the probability of at least one double six in 24 tosses of a pair of dice. We can easily evaluate both probabilities using the complementation law and Theorems 1.1.1 and 1.1.2. The probability of at least one six in four tosses of a single die is

$$\begin{aligned} &1 - \mathrm{P}(\text{no sixes in four tosses of a single die}) \\ &\quad = 1 - \frac{5^4}{6^4} = \frac{671}{1{,}296} \approx 0.517747, \end{aligned} \tag{1.39}$$

while the probability of at least one double six in 24 tosses of a pair of dice is

$$\begin{aligned} &1 - \mathrm{P}(\text{no double sixes in 24 tosses of a pair of dice}) \\ &\quad = 1 - \frac{(35)^{24}}{(36)^{24}} \approx 0.491404. \end{aligned} \tag{1.40}$$

(The second probability is the ratio of two 38-digit integers, but it does not seem useful to display them.) Méré had predicted the nonequality of the two probabilities based on empirical evidence. ♠

**Example 1.1.23.** *Rencontre.* The game of rencontre ("encounter" or "coincidence" in French) has been studied by Montmort, De Moivre, Laplace, Euler, and others. There are several versions of this game, but the one described by Montmort, or actually his simplification of it, is as follows. Consider a deck of $n$ distinct cards, which for convenience we will assume are labeled $1, 2, \ldots, n$. For specificity, we also label the positions of the cards in the deck as follows: With the cards face down, the top card in the deck is in position 1, the second card is in position 2, and so on. The cards are well shuffled and cut, and then dealt out one by one. The dealer is said to win if, for some $j \in \{1, 2, \ldots, n\}$, the card labeled $j$ is in position $j$. What is the probability $P_n$ that the dealer wins?

For $j = 1, 2, \ldots, n$, let $E_j$ be the event that the card labeled $j$ is in position $j$. The problem is to evaluate $P_n := \mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n)$. We use the inclusion-exclusion law. If $1 \leq m \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, then

$$\mathrm{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = \frac{(n-m)!}{n!}, \tag{1.41}$$

and hence, for $m = 1, 2, \ldots, n$, $S_m$ of (1.30) is given by

$$S_m = \binom{n}{m} \frac{(n-m)!}{n!} = \frac{1}{m!}. \tag{1.42}$$

We conclude from Theorem 1.1.18 that

$$P_n = \mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{m=1}^{n} \frac{(-1)^{m-1}}{m!} = 1 - \sum_{m=0}^{n} \frac{(-1)^m}{m!}. \quad (1.43)$$

Notice that $P_n$ converges rapidly to $1 - e^{-1} \approx 0.632120559$ as $n \to \infty$. ♠

Probabilities are frequently stated in terms of odds, and it is occasionally necessary to convert from one to the other. Given an event $E$, to say that the *odds against* $E$ are $\beta$ to $\alpha$ (or the *odds in favor of* $E$ are $\alpha$ to $\beta$) simply means that $\mathrm{P}(E) = \alpha/(\alpha+\beta)$. Here $\alpha$ and $\beta$ are positive numbers (typically, but not necessarily, integers). Notice that the odds factors $\alpha$ and $\beta$ can be scaled arbitrarily, that is, both can be multiplied by the same positive number without effect. For example, to say that the odds against $E$ are $\beta$ to $\alpha$ is equivalent to saying that they are $\beta/\alpha$ to 1.

The odds just defined are often referred to as the *true odds*, to distinguish them from the payoff odds. Suppose that an event $E$ offers *payoff odds* of $\beta$ to $\alpha$ (briefly, $E$ pays $\beta$ to $\alpha$), and that the bettor stakes 1 unit on $E$. If $E$ occurs, he wins $\beta/\alpha$ units, otherwise he loses 1 unit. In the case of a win, the casino returns his stake of 1 unit together with his profit of $\beta/\alpha$ units, for a total of $(\alpha+\beta)/\alpha$ units. In particular, the payoff odds of $\beta$ *to* $\alpha$ are sometimes stated as $\alpha + \beta$ *for* $\alpha$. Here $\alpha$ and $\beta$ are positive numbers (typically, but not necessarily, integers). Again, notice that the odds factors $\alpha$ and $\beta$ can be scaled arbitrarily. For example, to say that an event pays $\beta$ to $\alpha$ is equivalent to saying that it pays $\beta/\alpha$ to 1. If an event pays 1 to 1, it is said to pay *even money*.

Consider, for example, a single number (zero, say) on an unbiased 37-number roulette wheel. The probability that zero will occur at the next coup is $1/37$, so the odds against zero occurring are 36 to 1. However, the payoff odds for the occurrence of zero are only 35 to 1, which can also be stated as 36 *for* 1.

## 1.2 Independence and Conditional Probability

Consider two random experiments that are unrelated to each other, and assume that Theorem 1.1.1 on p. 3 applies to both. Let $\Omega_1$ and $\Omega_2$ be the two sample spaces, and let $E_1 \subset \Omega_1$ and $F_2 \subset \Omega_2$ be events. If we define

$$\Omega := \Omega_1 \times \Omega_2, \quad E := E_1 \times \Omega_2, \quad F := \Omega_1 \times F_2, \quad (1.44)$$

then $\Omega$ is the sample space for the joint random experiment, to which Theorem 1.1.1 on p. 3 still applies, $E$ is the event that $E_1$ occurs in the first random experiment, and $F$ is the event that $F_2$ occurs in the second random experiment. Furthermore,

$$P(E \cap F) = P(E_1 \times F_2) = \frac{|E_1 \times F_2|}{|\Omega_1 \times \Omega_2|}$$
$$= \frac{|E_1 \times \Omega_2|}{|\Omega_1 \times \Omega_2|} \frac{|\Omega_1 \times F_2|}{|\Omega_1 \times \Omega_2|} = \frac{|E|}{|\Omega|} \frac{|F|}{|\Omega|} = P(E)P(F). \quad (1.45)$$

Although this is not the most general situation under which events $E$ and $F$ are unrelated, we use (1.45) to motivate the next definition.

In general, events $E$ and $F$ are said to be *independent* if $P(E \cap F) = P(E)P(F)$. More generally, events $E_1, E_2, \ldots, E_n$ $(n \geq 2)$ are said to be *independent* if

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = P(E_{i_1})P(E_{i_2}) \cdots P(E_{i_m}) \quad (1.46)$$

whenever $2 \leq m \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_m \leq n$. Finally, a countably infinite collection of events $E_1, E_2, \ldots$ is said to be *independent* if $E_1, E_2, \ldots, E_n$ are independent for every $n \geq 2$.

**Example 1.2.1.** *Outcome 1 before outcome 2 in repeated independent trials.* Given a random experiment that has exactly three possible outcomes, referred to as outcomes 1, 2, and 3, with probabilities $p_1 > 0$, $p_2 > 0$, and $p_3 > 0$ $(p_1 + p_2 + p_3 = 1)$, consider a sequence of independent trials, at each of which the given random experiment is performed. What is the probability that outcome 1 occurs at least once before the first occurrence of outcome 2?

For $n = 1, 2, \ldots$, let $E_n$ be the event that outcome 1 occurs for the first time at trial $n$ and prior to the first occurrence of outcome 2. Then $E_1, E_2, \ldots$ are mutually exclusive, so by Axiom 1.1.14 on p. 12,

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} p_3^{n-1} p_1 = \frac{p_1}{1 - p_3} = \frac{p_1}{p_1 + p_2}. \quad (1.47)$$

To justify the second equality, we can write $E_n = F_1 \cap \cdots \cap F_{n-1} \cap G_n$, where $F_j$ $(1 \leq j \leq n-1)$ is the event that outcome 3 occurs at trial $j$ and $G_n$ is the event that outcome 1 occurs at trial $n$. Then $F_1, \ldots, F_{n-1}, G_n$ are independent events by the assumed independence of the trials, and therefore $P(E_n) = p_3^{n-1} p_1$.

This result is more useful than it may at first appear. In particular, the sample space for the original experiment may have more than three possible outcomes, and the roles of outcomes 1, 2, and 3 above may be played by three mutually exclusive events whose union is $\Omega$ and whose probabilities are $p_1$, $p_2$, and $p_3$.

For example, in repeated rolls of a pair of dice, the probability of rolling a total of 6 at least once before the first occurrence of a total of 7 is $\pi_6/(\pi_6 + \pi_7) = 5/11$, where we are using (1.11) on p. 6. ♠

The *conditional probability* of an event $E$, given the occurrence of an event $D$, is defined by

$$P(E \mid D) := \frac{P(D \cap E)}{P(D)}, \tag{1.48}$$

provided that $P(D) > 0$. Notice that, if $D$ and $E$ are independent, then $P(E \mid D) = P(E)$, that is, the conditional probability of an event $E$, given an independent event $D$, is equal to the (unconditional) probability of $E$.

We rarely use definition (1.48) to evaluate conditional probabilities. Instead, we can usually evaluate them as unconditional probabilities. Two examples should suffice to explain the idea.

**Example 1.2.2.** *Pass-line bet at craps.* The game of craps is played by rolling a pair of dice repeatedly. Except for some less important wagers, only the total of the two dice matters, and the probabilities of the various totals are given by (1.11) on p. 6. The principal bet at craps is called the *pass-line bet* and is initiated prior to the initial roll of the dice, which is called the *come-out roll*. The bet is won if the shooter rolls 7 or 11 (a *natural*) on the come-out roll. It is lost if the shooter rolls 2, 3, or 12 (a *craps number*) on the come-out roll. The only other possibility is that the shooter rolls a number belonging to

$$\mathscr{P} := \{4, 5, 6, 8, 9, 10\} \tag{1.49}$$

on the come-out roll, which establishes that number as the shooter's *point*. He continues to roll the dice until he either wins by repeating his point or loses by rolling 7. A win pays even money.

Let us introduce some events. For $j = 2, 3, 4, \ldots, 12$, we let $D_j$ be the event that $j$ is rolled on the come-out roll. We let $E$ be the event that the pass-line bet is won, and, for each $j \in \mathscr{P}$, we let $E_j$ be the event that, beginning with the second roll of the dice, $j$ appears before 7. Then

$$P(E \mid D_j) = P(E_j \mid D_j) = P(E_j) = \frac{\pi_j}{\pi_j + \pi_7}, \qquad j \in \mathscr{P}. \tag{1.50}$$

Here the first equality uses the fact that, given that point $j$ is established on the come-out roll, events $E$ and $E_j$ are equivalent (i.e., $D_j \cap E = D_j \cap E_j$). The second equality is a consequence of the independence of $D_j$ and $E_j$; this independence is due to the fact that $D_j$ depends on only the result of the come-out roll, while $E_j$ depends on only the results of subsequent rolls. Finally, we use Example 1.2.1 (together with the notation (1.11) on p. 6) for the third equality.

We continue with this example in Example 1.2.8 below. ♠

**Example 1.2.3.** *Drawing to a four-card flush.* Consider the game of five-card draw poker (or video poker). Given that a player is dealt four cards of one suit and a fifth card of another, what is the conditional probability of completing the flush (or straight flush) with a one-card draw? (Here the card of the odd suit is replaced by a card drawn from the residual 47-card deck.) We let $D$ be the event that the player is dealt four cards of one suit and a fifth card of another, and we let $E$ be the event that he completes the flush