

# Models for Probability and Statistical Inference

Theory and Applications

JAMES H. STAPLETON

Michigan State University  
Department of Statistics and Probability  
East Lansing, Michigan



A JOHN WILEY & SONS, INC., PUBLICATION



# Models for Probability and Statistical Inference



---

## THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

---

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

**WILLIAM J. PESCE**  
PRESIDENT AND CHIEF EXECUTIVE OFFICER

**PETER BOOTH WILEY**  
CHAIRMAN OF THE BOARD

---

# Models for Probability and Statistical Inference

Theory and Applications

JAMES H. STAPLETON

Michigan State University  
Department of Statistics and Probability  
East Lansing, Michigan



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2008 by John Wiley & Sons, Inc. All rights reserved.

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

Wiley Bicentennial Logo: Richard J. Pacifico

***Library of Congress Cataloging-in-Publication Data:***

Stapleton, James H., 1931–

Models for probability and statistical inference: theory and applications/James H. Stapleton.  
p. cm.

ISBN 978-0-470-07372-8 (cloth)

1. Probabilities—Mathematical models. 2. Probabilities—Industrial applications. I. Title.

QA273.S7415 2008

519.2—dc22

2007013726

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Alicia, who has made my first home so pleasant  
for almost 44 years.

To Michigan State University and its Department of Statistics  
and Probability, my second home for almost 49 years,  
to which I will always be grateful.



# Contents

<b>Preface</b>	<b>xi</b>
<b>1. Discrete Probability Models</b>	<b>1</b>
1.1. Introduction, 1	
1.2. Sample Spaces, Events, and Probability Measures, 2	
1.3. Conditional Probability and Independence, 15	
1.4. Random Variables, 27	
1.5. Expectation, 37	
1.6. The Variance, 47	
1.7. Covariance and Correlation, 55	
<b>2. Special Discrete Distributions</b>	<b>62</b>
2.1. Introduction, 62	
2.2. The Binomial Distribution, 62	
2.3. The Hypergeometric Distribution, 65	
2.4. The Geometric and Negative Binomial Distributions, 68	
2.5. The Poisson Distribution, 72	
<b>3. Continuous Random Variables</b>	<b>80</b>
3.1. Introduction, 80	
3.2. Continuous Random Variables, 80	
3.3. Expected Values and Variances for Continuous Random Variables, 88	
3.4. Transformations of Random Variables, 93	
3.5. Joint Densities, 97	
3.6. Distributions of Functions of Continuous Random Variables, 104	

<b>4. Special Continuous Distributions</b>	<b>110</b>
4.1. Introduction, 110	
4.2. The Normal Distribution, 111	
4.3. The Gamma Distribution, 117	
<b>5. Conditional Distributions</b>	<b>125</b>
5.1. Introduction, 125	
5.2. Conditional Expectations for Discrete Random Variables, 130	
5.3. Conditional Densities and Expectations for Continuous Random Variables, 136	
<b>6. Moment Generating Functions and Limit Theory</b>	<b>145</b>
6.1. Introduction, 145	
6.2. Moment Generating Functions, 145	
6.3. Convergence in Probability and in Distribution and the Weak Law of Large Numbers, 148	
6.4. The Central Limit Theorem, 155	
<b>7. Estimation</b>	<b>166</b>
7.1. Introduction, 166	
7.2. Point Estimation, 167	
7.3. The Method of Moments, 171	
7.4. Maximum Likelihood, 175	
7.5. Consistency, 182	
7.6. The $\delta$ -Method, 186	
7.7. Confidence Intervals, 191	
7.8. Fisher Information, Cramér–Rao Bound and Asymptotic Normality of MLEs, 201	
7.9. Sufficiency, 207	
<b>8. Testing of Hypotheses</b>	<b>215</b>
8.1. Introduction, 215	
8.2. The Neyman–Pearson Lemma, 222	
8.3. The Likelihood Ratio Test, 228	
8.4. The $p$ -Value and the Relationship between Tests of Hypotheses and Confidence Intervals, 233	
<b>9. The Multivariate Normal, Chi-Square, <math>t</math>, and <math>F</math> Distributions</b>	<b>238</b>
9.1. Introduction, 238	

9.2. The Multivariate Normal Distribution, 238	
9.3. The Central and Noncentral Chi-Square Distributions, 241	
9.4. Student's $t$ -Distribution, 245	
9.5. The $F$ -Distribution, 254	
<b>10. Nonparametric Statistics</b>	<b>260</b>
10.1. Introduction, 260	
10.2. The Wilcoxon Test and Estimator, 262	
10.3. One-Sample Methods, 271	
10.4. The Kolmogorov–Smirnov Tests, 277	
<b>11. Linear Statistical Models</b>	<b>281</b>
11.1. Introduction, 281	
11.2. The Principle of Least Squares, 281	
11.3. Linear Models, 290	
11.4. $F$ -Tests for $H_0: \theta = \beta_1 X_1 + \cdots + \beta_k X_k \in \mathbf{V}_0$ , a Subspace of $V$ , 299	
11.5. Two-Way Analysis of Variance, 308	
<b>12. Frequency Data</b>	<b>319</b>
12.1. Introduction, 319	
12.2. Confidence Intervals on Binomial and Poisson Parameters, 319	
12.3. Logistic Regression, 324	
12.4. Two-Way Frequency Tables, 330	
12.5. Chi-Square Goodness-of-Fit Tests, 340	
<b>13. Miscellaneous Topics</b>	<b>350</b>
13.1. Introduction, 350	
13.2. Survival Analysis, 350	
13.3. Bootstrapping, 355	
13.4. Bayesian Statistics, 362	
13.5. Sampling, 369	
<b>References</b>	<b>378</b>
<b>Appendix</b>	<b>381</b>
<b>Answers to Selected Problems</b>	<b>411</b>
<b>Index</b>	<b>437</b>



# Preface

This book was written over a five to six-year period to serve as a text for the two-semester sequence on probability and statistical inference, STT 861–2, at Michigan State University. These courses are offered for master’s degree students in statistics at the beginning of their study, although only one-half of the students are working for that degree. All students have completed a minimum of two semesters of calculus and one course in linear algebra, although students are encouraged to take a course in analysis so that they have a good understanding of limits. A few exceptional undergraduates have taken the sequence. The goal of the courses, and therefore of the book, is to produce students who have a fundamental understanding of statistical inference. Such students usually follow these courses with specialized courses on sampling, linear models, design of experiments, statistical computing, multivariate analysis, and time series analysis.

For the entire book, simulations and graphs, produced by the statistical package S-Plus, are included to build the intuition of students. For example, Section 1.1 begins with a list of the results of 400 consecutive rolls of a die. Instructors are encouraged to use either S-Plus or R for their courses. Methods for the computer simulation of observations from specified distributions are discussed.

Each section is followed by a selection of problems, from simple to more complex. Answers are provided for many of the problems.

Almost all statements are backed up with proofs, with the exception of the continuity theorem for moment generating functions, and asymptotic theory for logistic and log-linear models. Simulations are provided to show that the asymptotic theory provides good approximations.

The first six chapters are concerned with probability, the last seven with statistical inference. If a few topics covered in the first six chapters were to be omitted, there would be enough time in the first semester to cover at least the first few sections of Chapter Seven, on estimation. There is a bit too much material included on statistical inference for one semester, so that an instructor will need to make judicious choices of sections. For example, this instructor has omitted Section 7.8, on Fisher information, the Cramér–Rao bound, and asymptotic normality of MLEs, perhaps the most difficult material in the book. Section 7.9, on sufficiency, could be omitted.

Chapter One is concerned with discrete models and random variables. In Chapter Two we discuss discrete distributions that are important enough to have names: the binomial, hypergeometric, geometric, negative binomial, and Poisson, and the Poisson process is described. In Chapter Three we introduce continuous distributions, expected values, variances, transformation, and joint densities.

Chapter Four concerns the normal and gamma distributions. The beta distribution is introduced in Problem 4.3.5. Chapter Five, devoted to conditional distributions, could be omitted without much negative effect on statistical inference. Markov chains are discussed briefly in Chapter Five. Chapter Six, on limit theory, is usually the most difficult for students. Modes of convergence of sequences of random variables, with special attention to convergence in distribution, particularly the central limit theorem for independent random variables, are discussed thoroughly.

Statistical inference begins in Chapter Seven with point estimation: first methods of evaluating estimators, then methods of finding estimators: the method of moments and maximum likelihood. The topics of consistency and the  $\delta$ -method are usually a bit more difficult for students because they are often still struggling with limit arguments. Section 7.7, on confidence intervals, is one of the most important topics of the last seven chapters and deserves extra time. The author often asks students to explain the meaning of confidence intervals so that “your mother [or father] would understand.” Students usually fail to produce an adequate explanation the first time. As stated earlier, Section 7.8 is the most difficult and might be omitted. The same could be said for Section 7.9, on sufficiency, although the beauty of the subject should cause instructors to think twice before doing that.

Chapter Eight, on testing hypotheses, is clearly one of the most important chapters. We hope that sufficient time will be devoted to it to “master” the material, since the remaining chapters rely heavily on an understanding of these ideas and those of Section 7.7, on confidence intervals.

Chapter Nine is organized around the distributions defined in terms of the normal: multivariate normal, chi-square,  $t$ , and  $F$  (central and noncentral). The usefulness of each of the latter three distributions is shown immediately by the development of confidence intervals and testing methods for “normal models.” Some of “Student’s” data from the 1908 paper introducing the  $t$ -distribution is used to illustrate the methodology.

Chapter Ten contains descriptions of the two- and one-sample Wilcoxon tests, together with methods of estimation based on these. The Kolmogorov–Smirnov one- and two-sample tests are also discussed.

Chapter Eleven, on linear models, takes the linear space-projection approach. The geometric intuition it provides for multiple regression and the analysis of variance, by which sums of squares are simply squared lengths of vectors, is quite valuable. Examples of S-Plus and SAS printouts are provided.

Chapter Twelve begins with logistic regression. Although the distribution theory is quite different than the linear model theory discussed in Chapter Eleven and is asymptotic, the intuition provided by the vector-space approach carries over to logistic regression. Proofs are omitted in general in the interests of time and the students’ level

of understanding. Two-way frequency tables are discussed for models which suppose that the logs of expected frequencies satisfy a linear model.

Finally, Chapter Thirteen has sections on survival analysis, including the Kaplan–Meier estimator of the cumulative distribution function, bootstrapping, Bayesian statistics, and sampling. Each is quite brief. Instructors will probably wish to select from among these four topics.

I thank the many excellent students in my Statistics 861–2 classes over the last seven years, who provided many corrections to the manuscript as it was being developed. They have been very patient.

JIM STAPLETON

*March 7, 2007*



## CHAPTER ONE

# Discrete Probability Models

### 1.1 INTRODUCTION

The mathematical study of probability can be traced to the seventeenth-century correspondence between Blaise Pascal and Pierre de Fermat, French mathematicians of lasting fame. Chevalier de Mere had posed questions to Pascal concerning gambling, which led to Pascal's correspondence with Fermat. One question was this: Is a gambler equally likely to succeed in the two games: (1) at least one 6 in four throws of one six-sided die, and (2) at least one double-6 (6–6) in 24 throws of two six-sided dice? At that time it seemed to many that the answer was yes. Some believe that de Mere had empirical evidence that the first event was more likely to occur than the second, although we should be skeptical of that, since the probabilities turn out to be 0.5178 and 0.4914, quite close. After students have studied Chapter One they should be able to verify these, then, after Chapter Six, be able to determine how many times de Mere would have to play these games in order to distinguish between the probabilities.

In the eighteenth century, probability theory was applied to astronomy and to the study of errors of measurement in general. In the nineteenth and twentieth centuries, applications were extended to biology, the social sciences, medicine, engineering—to almost every discipline. Applications to genetics, for example, continue to grow rapidly, as probabilistic models are developed to handle the masses of data being collected. Large banks, credit companies, and insurance and marketing firms are all using probability and statistics to help them determine operating rules.

We begin with discrete probability theory, for which the events of interest often concern count data. Although many of the examples used to illustrate the theory involve gambling games, students should remember that the theory and methods are applicable to many disciplines.

## 1.2 SAMPLE SPACES, EVENTS, AND PROBABILITY MEASURES

We begin our study of probability by considering the results of 400 consecutive throws of a *fair die*, a six-sided cube for which each of the numbers 1, 2, ..., 6 is equally likely to be the number showing when the die is thrown.

61635	52244	21641	36536	52114	64452	33132	26324	62624	63134
36426	33552	65554	64623	56111	32256	36435	64146	53514	56364
52624	12534	15362	65261	43445	13223	66126	53623	63265	21564
21524	13552	65253	21225	42234	32361	62454	54561	15125	36555
45215	66442	42635	52522	13242	15434	16336	63241	13111	54343
32261	63155	55235	13611	54346	56323	41666	31221	53233	52414
53366	62336	11265	55136	56524	64215	44221	14222	15145	31662
55241	54223	25156	56155	43324	36566	23466	51123	11414	24653

The frequencies are:

1	2	3	4	5	6
60	73	65	58	74	70

We use these data to motivate the definitions and theory to be presented. Consider, for example, the following question: What is the probability that the five numbers appearing in five throws of a die are all different? Among the 80 consecutive sequences of five numbers above, in only four cases were all five numbers different, a relative frequency of  $5/80 = 0.0625$ . In another experiment, with 2000 sequences of five throws each, all were different 183 times, a relative frequency of 0.0915. Is there a way to determine the long-run relative frequency? Put another way, what could we expect the relative frequency to be in 1 million throws of five dice?

It should seem reasonable that all possible sequences of five consecutive integers from 1 to 6 are equally likely. For example, prior to the 400-throw experiment, each of the first two sequences, 61635 and 52244, were equally likely. For this example, such five-digit sequences will be called *outcomes* or *sample points*. The collection of all possible such five-digit sequences will be denoted by  $S$ , the sample space. In more mathematical language,  $S$  is the Cartesian product of the set  $A = \{1, 2, 3, 4, 5, 6\}$  with itself five times. This collection of sequences is often written as  $A^{(5)}$ . Thus,  $S = A^{(5)} = A \times A \times A \times A \times A$ . The number of outcomes (or sample points) in  $S$  is  $6^5 = 7776$ . It should seem reasonable to suppose that all outcomes (five-digit sequences) have probability  $1/6^5$ .

We have already defined a *probability model* for this experiment. As we will see, it is enough in cases in which the sample space is discrete (finite or countably infinite) to assign probabilities, nonnegative numbers summing to 1, to each outcome in the sample space  $S$ . A discrete probability model has been defined for an experiment when (1) a finite or countably infinite sample space has been defined, with each possible result of the experiment corresponding to exactly one outcome; and (2) probabilities, nonnegative numbers, have been assigned to the outcomes in such a way that they

sum to 1. It is not necessary that the probabilities assigned all be the same as they are for this example, although that is often realistic and convenient.

We are interested in the *event*  $A$  that all five digits in an outcome are different. Notice that this event  $A$  is a subset of the sample space  $S$ . We say that an event  $A$  has *occurred* if the outcome is a member of  $A$ . In this case event  $A$  did not occur for any of the eight outcomes in the first row above.

We define the *probability* of the event  $A$ , denoted  $P(A)$ , to be the sum of the probabilities of the outcomes in  $A$ . By defining the probability of an event in this way, we assure that the probability measure  $P$ , defined for all subsets (events, in probability language) of  $S$ , obeys certain axioms for probability measures (to be stated later). Because our probability measure  $P$  has assigned all probabilities of outcomes to be equally likely, to find  $P(A)$  it is enough for us to determine the number of outcomes  $N(A)$  in  $A$ , for then  $P(A) = N(A)[1/N(S)] = N(A)/N(S)$ . Of course, this is the case only because we assigned equal probabilities to all outcomes.

To determine  $N(A)$ , we can apply the multiplication principle.  $A$  is the collection of 5-tuples with all components different. Each outcome in  $A$  corresponds to a way of filling in the boxes of the following cells:



The first cell can hold any of the six numbers. Given the number in the first cell, and given that the outcome must be in  $A$ , the second cell can be any of five numbers, all different from the number in the first cell. Similarly, given the numbers in the first two cells, the third cell can contain any of four different numbers. Continuing in this way, we find that  $N(A) = (6)(5)(4)(3)(2) = 720$  and that  $P(A) = 720/7776 = 0.0926$ , close to the value obtained for 2000 experiments. The number  $N(A) = 720$  is the number of permutations of six things taken five at a time, indicated by  $P(6, 5)$ .

**Example 1.2.1** Consider the following discrete probability model, with sample space  $S = \{a, b, c, d, e, f\}$ .

Outcome $\omega$	$a$	$b$	$c$	$d$	$e$	$f$
$P(\omega)$	0.30	0.20	0.25	0.10	0.10	0.05

Let  $A = \{a, b, d\}$  and  $B = \{b, d, e\}$ . Then  $A \cup B = \{a, b, d, e\}$  and  $P(A \cup B) = 0.3 + 0.2 + 0.1 + 0.1 = 0.7$ . In addition,  $A \cap B = \{b, d\}$ , so that  $P(A \cap B) = 0.2 + 0.1 = 0.3$ . Notice that  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . (Why must this be true?). The *complement* of an event  $D$ , denoted by  $D^c$ , is the collection of outcomes in  $S$  that are not in  $D$ . Thus,  $P(A^c) = P(\{c, e, f\}) = 0.15 + 0.15 + 0.10 = 0.40$ . Notice that  $P(A^c) = 1 - P(A)$ . Why must this be true? □

Let us consider one more example before more formally stating the definitions we have already introduced.

**Example 1.2.2** A penny and a dime are tossed. We are to observe the number  $X$  of heads that occur and determine  $P(X = k)$  for  $k = 0, 1, 2$ . The symbol  $X$ , used here for its convenience in defining the events  $[X = 0]$ ,  $[X = 1]$ , and  $[X = 2]$ , will be called a *random variable* (rv).  $P(X = k)$  is shorthand for  $P([X = k])$ . We delay a more formal discussion of random variables.

Let  $S_1 = \{HH, HT, TH, TT\} = \{H, T\}^{(2)}$ , where the result for the penny and dime are indicated in this order, with H denoting head and T denoting tail. It should seem reasonable to assign equal probabilities  $1/4$  to each of the four outcomes. Denote the resulting probability measure by  $P_1$ . Thus, for  $A =$  [event that the coins give the same result]  $= \{HH, TT\}$ ,  $P_1(A) = 1/4 + 1/4 = 1/2$ .  $\square$

The 400 throws of a die can be used to simulate 400 throws of a coin, and therefore 200 throws of two coins, by considering 1, 2, and 3 as heads and 4, 5, and 6 as tails. For example, using the first 10 throws, proceeding across the first row, we get TH, TH, TT, HH, TT. For all 400 die throws, we get 50 cases of HH, 55 of HT, 47 of TH, and 48 of TT, with corresponding relative proportions 0.250, 0.275, 0.235, and 0.240. For the experiment with 10,000 throws, simulating 5000 pairs of coin tosses, we obtain 1288 HH's, 1215 HT's, 1232 TH's, and 1265 TT's, with relative frequencies 0.2576, 0.2430, 0.2464, and 0.2530. Our model  $(S_1, P_1)$  seems to fit well.

For this model we get  $P_1(X = 0) = 1/4$ ,  $P_1(X = 1) = 1/4 + 1/4 = 1/2$ , and  $P_1(X = 2) = 1/4$ . If we are interested only in  $X$ , we might consider a slightly smaller model, with sample space  $S_2 = \{0, 1, 2\}$ , where these three outcomes represent the numbers of heads occurring. Although it is tempting to make the model simpler by assigning equal probabilities  $1/3, 1/3, 1/3$  to these outcomes, it should be obvious that the empirical results of our experiments with 400 and 10,000 tosses are not consistent with such a model. It should seem reasonable, instead, to assign probabilities  $1/4, 1/2, 1/4$ , thus defining a probability measure  $P_2$  on  $S_2$ . The model  $(S_2, P_2)$  is a *recoding* or *reduction* of the model  $(S_1, P_1)$ , with the outcomes HT and TH of  $S_1$  corresponding to the single outcome  $X = 1$  of  $S_2$ , with corresponding probability determined by adding the probabilities  $1/4$  and  $1/4$  of HT and TH.

The model  $(S_2, P_2)$  is simpler than the model  $(S_1, P_1)$  in the sense that it has fewer outcomes. On the other hand, it is more complex in the sense that the probabilities are unequal. In choosing appropriate probability models, we often have two or more possible models. The choice of a model will depend on its approximation of experimental evidence, consistency with fundamental principles, and mathematical convenience.

Let us stop now to define more formally some of the terms already introduced.

**Definition 1.2.1** A *sample space* is a collection  $S$  of all possible results, called *outcomes*, of an experiment. Each possible result of the experiment must correspond to one and only one outcome in  $S$ . A sample space is *discrete* if it has a finite or countably infinite number of outcomes. (A set is *countably infinite* if it can be put into one-to-one correspondence with the positive integers.)  $\square$

**Definition 1.2.2** An *event* is a subset of a sample space. An event  $A$  is said to *occur* if the outcome of an experiment is a member of  $A$ .  $\square$

**Definition 1.2.3** A *probability measure*  $P$  on a discrete sample space  $S$  is a function defined on the subsets of  $S$  such that:

- (a)  $P(\{\omega\}) \geq 0$  for all points  $\omega \in S$ .
- (b)  $P(A) = \sum_{\omega \in A} P(\omega)$  for all subsets  $A$  of  $S$ .
- (c)  $P(S) = 1$ .

For simplicity, we write  $P(\{\omega\})$  as  $P(\omega)$ .  $\square$

**Definition 1.2.4** A *probability model* is a pair  $(S, P)$ , where  $P$  is a probability measure on  $S$ .  $\square$

In writing  $P(\{\omega\})$  as  $P(\omega)$ , we are abusing notation slightly by using the symbol  $P$  to denote both a function on  $S$  and a function on the subsets of  $S$ . We assume that students are familiar with the notation of set theory: *union*,  $A \cup B$ ; *intersection*  $A \cap B$ ; and *complement*,  $A^c$ . Thus, for events  $A$  and  $B$ , the event  $A \cup B$  is said to occur if the outcome is a member of  $A$  or  $B$  (by “or” we include the case that the outcome is in both  $A$  and  $B$ ). The event  $A \cap B$  is said to occur if both  $A$  and  $B$  occur.  $A^c$ , called a complement, is said to occur if  $A$  does not occur. For convenience we sometimes write  $A \cap B$  as  $AB$ .

We also assume that the student is familiar with the notation for relationships among sets,  $A \subset B$  and  $A \supset B$ . Thus, if  $A \subset B$ , the occurrence of event  $A$  implies that  $B$  must occur. We sometimes use the language “event  $A$  implies event  $B$ .” For the preceding two-coin-toss example, the event  $[X = 1]$  implies the event  $[X \geq 1]$ .

Let  $\emptyset$  denote the *empty event*, the subset of  $S$  consisting of no outcomes. Thus,  $A \cap A^c = \emptyset$ . We say that two events  $A$  and  $B$  are *mutually exclusive* if their intersection is empty. That is,  $A \cap B = \emptyset$ . Thus, if  $A$  and  $B$  are mutually exclusive, the occurrence of one of them implies that the other cannot occur. In set-theoretic language we say that  $A$  and  $B$  are *disjoint*. *DeMorgan’s laws* give relationships among intersection, union, and complement:

$$(1) (A \cap B)^c = A^c \cup B^c \quad \text{and} \quad (2) (A \cup B)^c = A^c \cap B^c.$$

These can be verified from a *Venn diagram* or by showing that any element in the set on the left is a member of the set on the right, and vice versa (see Figure 1.2.1).

### Properties of a Probability Measure $P$ on a Sample Space $S$

1.  $P(\emptyset) = 0$ .
2.  $P(S) = 1$ .
3. For any event  $A$ ,  $P(A^c) = 1 - P(A)$ .

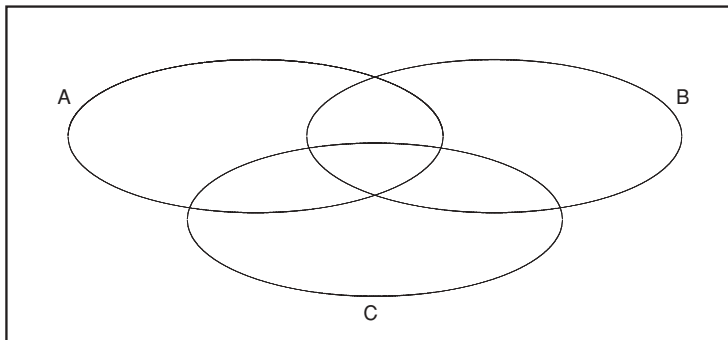


FIGURE 1.2.1 Venn diagram for three events.

4. For any events  $A$  and  $B$ ,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . For three events  $A, B, C$ ,  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$ . This follows from repeated use of the identity for two events. An almost obvious similar result holds for the probability of the union of  $n$  events, with  $2^n - 1$  terms on the right.
5. For events  $A$  and  $B$  with  $A \cap B = \emptyset$ ,  $P(A \cup B) = P(A) + P(B)$ . More generally, if  $A_1, A_2, \dots$ , are disjoint (mutually exclusive) events,  $P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$ . This property of  $P$  is called *countable additivity*. Since  $A_k$  for  $k > n$  could be  $\emptyset$ , the same equality holds when  $\infty$  is replaced by any integer  $n > 0$ .

Let us make use of some of these properties in a few examples.

**Example 1.2.3** Smith and Jones each throw three coins. Let  $X$  denote the number of heads for Smith. Let  $Y$  denote the number of heads for Jones. Find  $P(X = Y)$ .

We can simulate this experiment using the 400-die-tossing example, again letting 1, 2, 3 correspond to heads, 4, 5, 6 correspond to tails. Let the first three tosses be for Smith, the next three for Jones, so that the first six tosses determine one trial of the experiment. Repeating, going across rows, we get 36 trials of the experiment. Among these 36 trials, 10 resulted in  $X = Y$ , suggesting that  $P(X = Y)$  may be approximately  $10/36 = 0.278$ . For the experiment with 9996 tosses, 525 among 1666 trials gave  $X = Y$ , suggesting that  $P(X = Y)$  is close to  $525/1666 = 0.3151$ . Let us now try to find the probability by mathematical methods.

Let  $S_1 = \{H, T\}^{(3)}$ , the collection of 3-tuples of heads and tails.  $S_1$  is the collection of outcomes for Smith. Also, let  $S_2 = \{H, T\}^{(3)} = S_1$ , the collection of outcomes for Jones. Let  $S = S_1 \times S_2$ . This Cartesian product can serve as the sample space for the experiment in which Smith and Jones both toss three coins. One outcome in  $S$ , for example (using shorthand notation), is (HTH, TTH), so that  $X = 2, Y = 1$ . The event  $[X = Y]$  did not occur. Since  $N(S_1) = N(S_2) = 2^3 = 8$ ,  $N(S) = 64$ . Define the probability measure  $P$  on  $S$  by assigning probability  $1/64$  to each outcome. The pair  $(S, P)$  constitutes a probability model for the experiment.

Let  $A_k = [X = Y = k]$  for  $k = 0, 1, 2, 3$ . By this bracket notation we mean the collection of outcomes in  $S$  for which  $X$  and  $Y$  are both  $k$ . We might also have

**TABLE 1.2.1 Box Diagram**

	$R_2$	$R_2^c$	
$R_1$		0.2	0.6
$R_1^c$			
	0.5		1.0

written  $A_k = [X = k, Y = k]$ . The events  $A_0, A_1, A_2, A_3$  are mutually exclusive, and  $[X = Y] = A_0 \cup A_1 \cup A_2 \cup A_3$ . It follows from property 5 above that  $P(X = Y) = P(A_0) + P(A_1) + P(A_2) + P(A_3)$ . Since  $N(A_0) = 1, N(A_1) = 3^2, N(A_2) = 3^2, N(A_3) = 1$ , and  $P(A_k) = N(A_k)/64$ , we find that  $P(X = Y) = 20/64 = 5/16 = 0.3125$ , relatively close to the proportions obtained by experimentation. □

**Example 1.2.4** Suppose that a probability model for the weather for two days has been defined in such a way that  $R_1 = [\text{rain on day 1}], R_2 = [\text{rain on day 2}], P(R_1) = 0.6, P(R_2) = 0.5$ , and  $P(R_1 R_2^c) = 0.2$ . Find  $P(R_1 R_2), P(R_1^c R_2)$ , and  $P(R_1 \cup R_2)$ .

Although a Venn diagram can be used, a *box diagram* (Table 1.2.1) makes things clearer. From the three probabilities given, the other cells may be determined by subtraction. Thus,  $P(R_1^c) = 0.4, P(R_2^c) = 0.5, P(R_1 R_2) = 0.4, P(R_1^c R_2) = 0.1, P(R_1^c R_2^c) = 0.3, P(R_1^c \cup R_2^c) = 0.6$ . Similar tables can be constructed for the three events. □

**Example 1.2.5** A jury of six is to be chosen randomly from a panel of eight men and seven women. Let  $X$  denote the number of women chosen. Let us find  $P(X = k)$  for  $k = 0, 1, \dots, 6$ .

For convenience, name the members of the panel  $1, 2, \dots, 15$ , with the first eight being men. Let  $D = \{1, 2, \dots, 15\}$ . Since the events in which we are interested do not depend on the order in which the people are drawn, the outcomes can be chosen to be subsets of  $D$  of size 6. That is,  $S = \{B \mid B \subset D, N(B) = 6\}$ . We interpret “randomly” to mean that all the outcomes in  $S$  should have equal probability. We need to determine  $N(S)$ . Such subsets are often called *combinations*. The number of combinations of size  $k$  of a set of size  $n$  is denoted by  $\binom{n}{k}$ . Thus,  $N(S) = \binom{15}{6}$ .

The number of *permutations* (6-tuples of different people) of 15 people six at a time, is  $P(15, 6) = (15)(14)(13)(12)(11)(10) = 15!/9!$ . The number of ways of ordering six people is  $P(6, 6) = 6!$ . Since (number of subsets of  $D$  of size 6)  $\times$  (number of ways of ordering six people) =  $P(15, 6)$ , we find that  $N(S) = \binom{15}{6} = P(15, 6)/6! = 15!/9!6! = 5005$ . Each outcome is assigned probability  $1/5005$ .

Consider the event  $[X = 2]$ . An outcome in  $[X = 2]$  must include exactly two females and therefore four males. There are  $\binom{7}{2} = (7)(6)/(2)(1) = 21$  such combinations. There are  $\binom{8}{4}$  combinations of four males. There are



Let  $D$  be the set of 52 cards. Let  $S$  be the collection of five-card hands, subsets of five cards. Thus,  $S = \{B \mid B \subset D, N(B) = 5\}$ . “Randomly without replacement” means that all  $N(S) = \binom{52}{5} = 2,598,960$  outcomes are equally likely. Thus, we have defined a probability model.

Let  $F$  be the event [full house]. The rank having three cards can be chosen in 13 ways. For each of these the rank having two cards can be chosen in 12 ways. For each choice of the two ranks there are  $\binom{4}{3}\binom{4}{2} = (4)(6)$  choices for the suits. Thus,  $N(F) = (13)(12)(6)(4) = 3744$ , and  $P(F) = 3744/2,598,960 = 0.0001439 = 1/694$ . Similarly,  $P(\text{straight}) = 10[4^5 - 4]/N(S) = 10,200/N(S) = 0.003925 = 1/255$ , and  $P(2 \text{ pairs}) = \binom{13}{2}\binom{4}{2}\binom{4}{2}(44)/N(S) = 123,552/N(S) = 0.04754 = 1/21.035$ . In general, as the value of a hand increases, the probability of the corresponding category decreases (see Problem 1.2.3).  $\square$

**Example 1.2.7** (The Birthday Problem) A class has  $n$  students. What is the probability that at least one pair of students have the same birthday, not necessarily the same birth year?

So that we can think a bit more clearly about the problem, let the days be numbered 1, . . . , 365, and suppose that  $n = 20$ . Birth dates were randomly chosen using the function “sample” in S-Plus, a statistical computer language.

- (1) 52 283 327 15 110 214 141\* 276 16 43 130 219 337 234 64 262 141\* 336 220 10
- (2) 331 106 364 219 209 70 11 54 192 360 75 228 132 172 30 5 166 15 143 173
- (3) 199 361\* 211 48 86 129 39 202 339 347 22 361\* 208 276 75 115 65 291 57 318
- (4) 300 252 274 135 118 199 254 316 133 192 238 189 94 167 182 5 235 363 160 214
- (5) 110 187 107 47 250 341 49 341 258 273 290 225 31 108 334 118 214 87 315 282
- (6) 195 270^ 24 204# 69 233 38% 204# 12\* 358 38% 138 149 76 71 186 106 270^ 12\* 87
- (7) 105 354 259 10 244 22 70 28 278 127 320 238 60 8 165 339 119 346 295 92
- (8) 359# 289 112 299 201 36 94 75 269 359# 122 288 310 329 133 117 291 61\* 61\* 336
- (9) 300 346 72 296 221 176 109 189 3 114 83 222 292 318 238 215 246 183 220 236
- (10) 337 98 17 357 75 32 138 255 150 12 88 133 135 5 319 198 119 288 183 359

Duplicates are indicated by \*’s, #’s, ^’s, and %’s. Notice that these 10 trials had 1, 0, 0, 3, 0, 2, 0, 0 duplicates. Based on these trials, we estimate the probability of at least one duplicate to be 4/10. This would seem to be a good estimate, since 2000 trials produced 846 cases with at least one duplicate, producing the estimate 0.423. Let us determine the probability mathematically.

Notice the similarity of this example to the die-throw example at the beginning of the chapter. In this case let  $D = \{1, \dots, 365\}$ , the “dates” of the year. Let  $S = D^{(n)}$ , the  $n$ -fold Cartesian product of  $D$  with itself. Assign probability  $1/N(S) = 1/365^n$  to each outcome. We now have a probability model.  $\square$

Let  $A$  be the event of at least one duplicate. As with most “at least one” events, it is easier to determine  $N(A^c)$  than  $N(A)$  directly. In fact,  $N(A^c) = P(365, n) = 365(364) \cdots (365 - n + 1)$ . Let  $G(n) = P(A^c)$ . It follows that  $G(n) = N(A^c)/N(S) = \prod_{k=1}^n [(365 - k + 1)/365] = \prod_{k=1}^n [1 - (k - 1)/365]$ . We can find

TABLE 1.2.3 Probabilities of Coincident Birthdays

	$n$								
	10	20	22	23	30	40	50	60	70
$P(A)$	0.1169	0.4114	0.4757	0.5073	0.7063	0.8912	0.9704	0.9941	0.9991
$h(n)$	0.1160	0.4058	0.4689	0.5000	0.6963	0.8820	0.9651	0.9922	0.9987

a good approximation by taking logarithms and converting the product to a sum.  $\ln G(n) = \sum_{k=1}^n \ln [1 - (k-1)/365]$ . For  $x$  close to zero,  $\ln(1-x)$  is close to  $-x$ , the Taylor series linear approximation. It follows that for  $(n-1)/365$  small,  $\ln(G(n))$  is approximately  $-\sum_{k=1}^n [(k-1)/365] = -[n(n-1)/2]/365 = -n(n-1)/730$ . Hence, a good approximation for  $P(A) = 1 - G(n)$  is  $h(n) = 1 - e^{-n(n-1)/730}$ . Table 1.2.3 compares  $P(A)$  to its approximation  $h(n)$  for various  $n$ . Notice that the relative error in the approximation of  $P(A^c)$  by  $1 - h(n)$  increases as  $n$  increases.

### Pascal's Triangle: An Interesting Identity for Combinatorics

Consider a set of five elements  $A = \{a_1, a_2, \dots, a_5\}$ .  $A$  has  $\binom{5}{3} = 10$  subsets of size 3. These are of two types: those that contain element  $a_1$  and those that do not. The number that contains  $a_1$  is the number of subsets  $\binom{4}{2} = 6$  of  $\{a_2, \dots, a_5\}$  of size 2. The number of subsets of  $A$  that do not contain  $a_1$  is  $\binom{4}{3} = 4$ . Thus,  $\binom{5}{3} = \binom{4}{2} + \binom{4}{3}$ .

More generally, if  $A$  has  $n$  elements  $\{a_1, a_2, \dots, a_n\}$ ,  $A$  has  $\binom{n}{k}$  subsets of size  $k$  for  $0 < k \leq n$ . These subsets are of two types, those that contain  $a_1$  and those that do not. It follows by the same reasoning that  $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ . The same equality can be proved by manipulation of factorials. Pascal, in the mid-seventeenth century, represented this in the famous *Pascal triangle* (Figure 1.2.3). Each row begins and ends with 1, and each interior value is the sum of the two immediately above. The  $n$ th row for  $n = 0, 1, \dots$  has  $\binom{n}{k}$  in the  $k$ th place for  $k = 0, 1, \dots, n$ . Row  $n = 4$  has elements 1, 4, 6, 4, 1. Notice that these sum to  $16 = 2^4$ .

### The Equality $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n$

The collection  $B$  of subsets of a set with  $n$  elements is in one-to-one correspondence to the set  $C = \{0, 1\}^{(n)}$ . For example, for the set  $A = \{a_1, a_2, a_3, a_4\}$ , the point  $(0, 1, 1, 0)$  in  $C$  corresponds to the subset  $\{a_2, a_3\}$ , and  $\{1, 0, 1, 1\}$  corresponds to the subset  $\{a_1, a_3, a_4\}$ . Thus,  $N(B) = N(C) = 2^k$ . But we can count the elements in  $B$  another way. There are those with no elements, those with one, those with 2, and so on. The



$S_2$  such that  $P_1(g^{-1}(A)) = P_2(A)$  for any subset  $A$  of  $S_2$ .  $(S_1, P_1)$  is said to be an *expansion* of  $(S_2, P_2)$ .  $\square$

If  $S_1$  is finite or countably infinite,  $P_1(g^{-1}(A)) = P_2(A)$  is assured if it holds whenever  $A$  is a one-point set. This follows from the fact that  $g^{-1}(A) = \{g^{-1}(\omega) \mid \omega \in A\}$  is a union of mutually exclusive events.

If  $X$  is a discrete random variable defined on  $(S_1, P_1)$ , let  $S_2 = \{k \mid P(X = k) > 0\}$ . Let  $P_2(k) = P_1(X = k)$ . Then  $X$  plays the role of  $g$  in the definition so that  $(S_2, P_2)$  is a reduction of  $(S_1, P_1)$ . This is the most common way to reduce a probability model. More generally, if  $X_1, \dots, X_n$  are random variables defined on  $(S_1, P_1)$ ,  $\mathbf{X} = (X_1, \dots, X_n)$ , then we can take  $S_2 = \{\mathbf{x} \in R_n \mid P(\mathbf{X} = \mathbf{x}) > 0\}$  and assign  $P_2(\mathbf{x}) = P_1(\mathbf{X} = \mathbf{x})$ .

**Example 1.2.9** A husband and wife and two other couples are seated at random around a round table with six seats. What is the probability that the husband and wife in a particular couple, say  $C_1$ , are seated in adjacent seats?

Let the people be  $a, b, \dots, g$ , let the seats be numbered  $1, \dots, 6$ , reading clockwise around the table, and let  $(x_1, \dots, x_6)$ , where each  $x_i$  is one of these letters, all different, correspond to the outcome in which person  $x_i$  is seated in seat  $i$ ,  $i = 1, 2, \dots, 6$ . Let  $S_1$  be the collection of such arrangements. Let  $P_1$  assign probability  $1/6!$  to each outcome. Let  $A$  be the collection of outcomes for which  $f$  and  $g$  are adjacent. If  $f$  and  $g$  are the husband and wife in  $C_1$ , then  $fg$  in this order may be in seats  $12, 23, \dots, 56, 61$ . They may also be in the order of  $21, 32, \dots, 16$ . For each of these the other four people may be seated in  $4!$  ways. Thus,  $N(A) = (2)(6)(4!) = 288$  and  $P(A) = 2/5$ .

We may instead let an outcome designate only the seats given to the husband and wife in  $C_1$ , and let  $S_2$  be the set of pairs  $(x, y)$ ,  $x \neq y$ . We have combined all  $4!$  seating arrangements in  $S_1$  which lead to the same seats for the husband and wife in  $C_1$ . Thus,  $N(S_2) = (6)(5)$ . Let  $P_2$  assign equal probability  $1/[(5)(6)]$  to each outcome,  $4! = 24$  times as large as for the outcomes in  $S_1$ . Let  $B = [\text{husband and wife in } C_1 \text{ are seated together}] = \{12, 23, \dots, 61, 21, \dots, 16\}$ , a subset of  $S_2$ . Then  $P(B) = (2)(6)/(6)(5) = 2/5$ , as before. Of course, if we were asked the probability of the event  $D$  that all three couples are seated together, each wife next to her husband, we could not answer the question using  $(S_2, P_2)$ , although we could using the model  $(S_1, P_1)$ .  $P_1(D) = (2)(3!)(2^3)/6! = 96/720 = 2/15$ . (Why?)  $D$  is an event with respect to  $S_1$  (a subset of  $S_1$ ), but there is no corresponding subset of  $S_2$ .  $\square$

## Problems for Section 1.2

- 1.2.1** Consider the sample space  $S = \{a, b, c, d, e, f\}$ . Let  $A = \{a, b, c\}$ ,  $B = \{b, c, d\}$ , and  $C = \{a, f\}$ . For each outcome  $x$  in  $S$ , let  $P(\{x\}) = p(x)$ , where  $p(a) = 0.20$ ,  $p(b) = 0.15$ ,  $p(c) = 0.20$ ,  $p(d) = 0.10$ ,  $p(e) = 0.30$ . Find  $p(f)$ ,  $P(A)$ ,  $P(B)$ ,  $P(A \cup B)$ ,  $P(A \cup B^c)$ ,  $P(A \cup B^c \cup C)$ . Verify that  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

- 1.2.2** Box 1 has four red and three white balls. Box 2 has three red and two white balls. A ball is drawn randomly from each box. Let  $(x, y)$  denote an outcome for which the ball drawn from box 1 has color  $x$  and the ball drawn from box 2 has color  $y$ . Let  $C = \{\text{red, white}\}$  and let  $S = C \times C$ .
- (a) Assign probabilities to the outcomes in  $S$  in a reasonable way. [In 1000 trials of this experiment the outcome (red, white) occurred 241 times.]
  - (b) Let  $X =$  (no. red balls drawn). Find  $P(X = k)$  for  $k = 0, 1, 2$ . (In 1000 trials the events  $[X = 0]$ ,  $[X = 1]$ , and  $[X = 2]$  occurred 190, 473, and 337 times.)
- 1.2.3** For the game of poker, find the probabilities of the events [straight flush], [4-of-a-kind], [flush], [3-of-a-kind], [one pair].
- 1.2.4** Find the elements in the rows labeled  $n = 6$  and  $n = 7$  in Pascal's triangle. Verify that their sums are  $2^6$  and  $2^7$ .
- 1.2.5** A coin is tossed five times.
- (a) Give a probability model so that  $S$  is a Cartesian product.
  - (b) Let  $X =$  (no. heads). Determine  $P(X = 2)$ .
  - (c) Use the die-toss data at the beginning of the chapter to simulate this experiment and verify that the relative frequency of cases for which the event  $[X = 2]$  occurs is close to  $P(X = 2)$  for your model.
- 1.2.6** For a Venn diagram with three events  $A, B, C$ , indicate the following events by darkening the corresponding region:
- (a)  $A \cup B^c \cup C$ .
  - (b)  $A^c \cup (B^c \cap C)$ .
  - (c)  $(A \cup B^c) \cap (B^c \cup C)$ .
  - (d)  $(A \cup B^c \cap C)^c$ .
- 1.2.7** Two six-sided fair dice are thrown.
- (a) Let  $X =$  (total for the two dice). State a reasonable model and determine  $P(X = k)$  for  $k = 2, 3, \dots, 12$ . (Different reasonable people may have sample spaces with different numbers of outcomes, but their answers should be the same.)
  - (b) Let  $Y =$  (maximum for the two dice). Find  $P(Y = j)$  for  $j = 1, 2, \dots, 6$ .
- 1.2.8**
- (a) What is the (approximate) probability that at least two among five nonrelated people celebrate their birthdays in the same month? State a model first. In 100,000 simulations the event occurred 61,547 times.
  - (b) What is the probability that at least two of five cards chosen randomly without replacement from the deck of 48 cards formed by omitting the aces are of the same rank? Intuitively, should the probability be larger

or smaller than the answer to part (a)? Why? In 100,000 simulations the event occurred 52,572 times.

- 1.2.9** A small town has six houses on three blocks,  $B_1 = \{a, b, c\}$ ,  $B_2 = \{d, e\}$ ,  $B_3 = \{f\}$ . A random sample of two houses is to be chosen according to two different methods. Under method 1, all possible pairs of houses are written on slips of paper, the slips are thoroughly mixed, and one slip is chosen. Under method 2, two of the three blocks are chosen randomly without replacement, then one house is chosen randomly from each of the blocks chosen. For each of these two methods state a probability model, then use it to determine the probabilities of the events [house  $a$  is chosen], [house  $d$  is chosen], [house  $f$  is chosen], and [at least one of houses  $a, d$  is chosen].
- 1.2.10** Four married couples attend a dance. For the first dance the partners for the women are randomly assigned among the men. What is the probability that at least one woman must dance with her husband?
- 1.2.11** From among nine men and seven women a jury of six is chosen randomly. What is the probability that two or fewer of those chosen are men?
- 1.2.12** A six-sided die is thrown three times.
- (a) What is the probability that the numbers appearing are in increasing order? *Hint:* There is a one-to-one correspondence between subsets of size 3 and increasing sequences from  $\{1, 2, \dots, 6\}$ . In 10,000 simulations the event occurred 934 times.
- (b) What is the probability that the three numbers are in nondecreasing order?  $(2, 4, 4)$  is not in increasing order, but *is* in nondecreasing order. Use the first 60 throws given at the beginning of the chapter to simulate the experiment 20 times. For the 10,000 simulations, the event occurred 2608 times.
- 1.2.13** Let  $(S, P)$  be a probability model and let  $A, B, C$  be three events such that  $P(A) = 0.55$ ,  $P(B) = 0.60$ ,  $P(C) = 0.45$ ,  $P(A \cap B) = 0.25$ ,  $P(A \cap C) = 0.20$ ,  $P(B^c \cap C) = 0.15$ , and  $P(A \cap B \cap C) = 0.10$ .
- (a) Present a box diagram with  $2^3 = 8$  cells giving the probabilities of all events of the form  $A^* \cap B^* \cap C^*$ , where  $A^*$  is either  $A$  or  $A^c$ , and  $B^*$  and  $C^*$  are defined similarly.
- (b) Draw a Venn diagram indicating the same probabilities.
- (c) Find  $P(A^c \cap B \cap C^c)$  and  $P(A \cup B^c \cup C)$ . *Hint:* Use one of DeMorgan's laws for the case of three events.
- 1.2.14** (*The Matching Problem*)
- (a) Let  $A_1, \dots, A_n$  be  $n$  events, subsets of the sample space  $S$ . Let  $S_k$  be the sum of the probabilities of the intersections of all  $\binom{n}{k}$  choices of these  $n$  events, taken  $k$  at a time. For example, for  $n = 4$ ,