

PROBABILITY

AN INTRODUCTION WITH STATISTICAL APPLICATIONS

SECOND EDITION

JOHN J. KINNEY

WILEY

Probability

Probability

An Introduction with Statistical Applications

Second Edition

John J. Kinney

Colorado Springs, CO

WILEY

Copyright © 2015 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. 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/permissions>.

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.

Library of Congress Cataloging-in-Publication Data:

Kinney, John J.

Probability : an introduction with statistical applications / John Kinney, Colorado Springs, CO. – Second edition.

pages cm

Includes bibliographical references and index.

ISBN 978-1-118-94708-1 (cloth)

1. Probabilities–Textbooks. 2. Mathematical statistics–Textbooks. I. Title.

QA273.K493 2015

519.2–dc23

2014020218

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*This book is for
Cherry and Kaylyn*

Contents

Preface for the First Edition xi

Preface for the Second Edition xv

1. Sample Spaces and Probability 1

1.1.	Discrete Sample Spaces	1
1.2.	Events; Axioms of Probability	7
	Axioms of Probability	8
1.3.	Probability Theorems	10
1.4.	Conditional Probability and Independence	14
	Independence	23
1.5.	Some Examples	28
1.6.	Reliability of Systems	34
	Series Systems	34
	Parallel Systems	35
1.7.	Counting Techniques	39
	Chapter Review	54
	Problems for Review	56
	Supplementary Exercises for Chapter 1	56

2. Discrete Random Variables and Probability Distributions 61

2.1.	Random Variables	61
2.2.	Distribution Functions	68
2.3.	Expected Values of Discrete Random Variables	72
	Expected Value of a Discrete Random Variable	72
	Variance of a Random Variable	75
	Tchebycheff's Inequality	78
2.4.	Binomial Distribution	81
2.5.	A Recursion	82
	The Mean and Variance of the Binomial	84
2.6.	Some Statistical Considerations	88
2.7.	Hypothesis Testing: Binomial Random Variables	92
2.8.	Distribution of A Sample Proportion	98
2.9.	Geometric and Negative Binomial Distributions	102
	A Recursion	108
2.10.	The Hypergeometric Random Variable: Acceptance Sampling	111
	Acceptance Sampling	111
	The Hypergeometric Random Variable	114
	Some Specific Hypergeometric Distributions	116
2.11.	Acceptance Sampling (Continued)	119

Producer's and Consumer's Risks	121
Average Outgoing Quality	122
Double Sampling	124
2.12. The Hypergeometric Random Variable: Further Examples	128
2.13. The Poisson Random Variable	130
Mean and Variance of the Poisson	131
Some Comparisons	132
2.14. The Poisson Process	134
Chapter Review	139
Problems for Review	141
Supplementary Exercises for Chapter 2	142

3. Continuous Random Variables and Probability Distributions

146

3.1. Introduction	146
Mean and Variance	150
A Word on Words	153
3.2. Uniform Distribution	157
3.3. Exponential Distribution	159
Mean and Variance	160
Distribution Function	161
3.4. Reliability	162
Hazard Rate	163
3.5. Normal Distribution	166
3.6. Normal Approximation to the Binomial Distribution	175
3.7. Gamma and Chi-Squared Distributions	178
3.8. Weibull Distribution	184
Chapter Review	186
Problems For Review	189
Supplementary Exercises for Chapter 3	189

4. Functions of Random Variables; Generating Functions; Statistical Applications

194

4.1. Introduction	194
4.2. Some Examples of Functions of Random Variables	195
4.3. Probability Distributions of Functions of Random Variables	196
Expectation of a Function of X	199
4.4. Sums of Random Variables I	203
4.5. Generating Functions	207
4.6. Some Properties of Generating Functions	211
4.7. Probability Generating Functions for Some Specific Probability Distributions	213
Binomial Distribution	213
Poisson's Trials	214
Geometric Distribution	215
Collecting Premiums in Cereal Boxes	216
4.8. Moment Generating Functions	218
4.9. Properties of Moment Generating Functions	223
4.10. Sums of Random Variables-II	224
4.11. The Central Limit Theorem	229
4.12. Weak Law of Large Numbers	233
4.13. Sampling Distribution of the Sample Variance	234

4.14.	Hypothesis Tests and Confidence Intervals for a Single Mean Confidence Intervals, σ Known	240
	Student's t Distribution	242
	p Values	243
4.15.	Hypothesis Tests on Two Samples	248
	Tests on Two Means	248
	Tests on Two Variances	251
4.16.	Least Squares Linear Regression	258
4.17.	Quality Control Chart for \bar{X}	266
	Chapter Review	271
	Problems for Review	275
	Supplementary Exercises for Chapter 4	275

5. Bivariate Probability Distributions **283**

5.1.	Introduction	283
5.2.	Joint and Marginal Distributions	283
5.3.	Conditional Distributions and Densities	293
5.4.	Expected Values and the Correlation Coefficient	298
5.5.	Conditional Expectations	303
5.6.	Bivariate Normal Densities	308
	Contour Plots	310
5.7.	Functions of Random Variables	312
	Chapter Review	316
	Problems for Review	317
	Supplementary Exercises for Chapter 5	317

6. Recursions and Markov Chains **322**

6.1.	Introduction	322
6.2.	Some Recursions and their Solutions	322
	Solution of the Recursion (6.3)	326
	Mean and Variance	329
6.3.	Random Walk and Ruin	334
	Expected Duration of the Game	337
6.4.	Waiting Times for Patterns in Bernoulli Trials	339
	Generating Functions	341
	Average Waiting Times	342
	Means and Variances by Generating Functions	343
6.5.	Markov Chains	344
	Chapter Review	354
	Problems for Review	355
	Supplementary Exercises for Chapter 6	355

7. Some Challenging Problems **357**

7.1.	My Socks and $\sqrt{\pi}$	357
7.2.	Expected Value	359
7.3.	Variance	361
7.4.	Other "Socks" Problems	362
7.5.	Coupon Collection and Related Problems	362
	Three Prizes	363

Permutations	363
An Alternative Approach	363
Altering the Probabilities	364
A General Result	364
Expectations and Variances	366
Geometric Distribution	366
Variances	367
Waiting for Each of the Integers	367
Conditional Expectations	368
Other Expected Values	369
Waiting for All the Sums on Two Dice	370
7.6. Conclusion	372
7.7. Jackknifed Regression and the Bootstrap	372
Jackknifed Regression	372
7.8. Cook's Distance	374
7.9. The Bootstrap	375
7.10. On Waldegrave's Problem	378
Three Players	378
7.11. Probabilities of Winning	378
7.12. More than Three Players	379
$r + 1$ Players	381
Probabilities of Each Player	382
Expected Length of the Series	383
Fibonacci Series	383
7.13. Conclusion	384
7.14. On Huygen's First Problem	384
7.15. Changing the Sums for the Players	384
Decimal Equivalents	386
Another order	387
Bernoulli's Sequence	387
Bibliography	388
Appendix A. Use of Mathematica in Probability and Statistics	390
Appendix B. Answers for Odd-Numbered Exercises	429
Appendix C. Standard Normal Distribution	453
Index	461

Preface for the First Edition

HISTORICAL NOTE

The theory of probability is concerned with events that occur when randomness or chance influences the result. When the data from a sample survey or the occurrence of extreme weather patterns are common enough examples of situations where randomness is involved, we have come to presume that many models of the physical world contain elements of randomness as well. Scientists now commonly suppose that their models contain random components as well as deterministic components. Randomness, of course, does not involve any new physical forces; rather than measuring all the forces involved and thus predicting the exact outcome of an experiment, we choose to combine all these forces and call the result random. The study of random events is the subject of this book.

It is impossible to chronicle the first interest in events involving randomness or chance, but we do know of a correspondence between Blaise Pascal and Pierre de Fermat in the middle of the seventeenth century regarding questions arising in gambling games. Appropriate mathematical tools for the analysis of such situations were not available at that time, but interest continued among some mathematicians. For a long time, the subject was connected only to gambling games and its development was considerably restricted by the situations arising from such considerations. Mathematical techniques suitable for problems involving randomness have produced a theory applicable to not only gambling situations but also more practical situations. It has not been until recent years, however, that scientists and engineers have become increasingly aware of the presence of random factors in their experiments and manufacturing processes and have become interested in measuring or controlling these factors.

It is the realization that the statistical analysis of experimental data, based on the theory of probability, is of great importance to experimenters that has brought the theory to the forefront of applicable mathematics. The history of probability and the statistical analysis it makes possible illustrate a prime example of seemingly useless mathematical research that now has an incredibly wide range of practical application. Mathematical models for experimental situations now commonly involve both deterministic and random terms. It is perhaps a simplification to say that science, while interested in deterministic models to explain the physical world, now is interested as well in separating deterministic factors from random factors and measuring their relative importance.

There are two facts that strike me as most remarkable about the theory of probability. One is the apparent contradiction that random events are in reality well behaved and that there are laws of probability. The outcome on one toss of a coin cannot be predicted, but given 10,000 tosses of the same coin, many events can be predicted with a high degree of accuracy. The second fact, which the reader will soon perceive, is the pervasiveness of a probability distribution known as the normal distribution. This distribution, which will be defined and discussed at some length, arises in situations which at first glance have little in

common: the normal distribution is an essential tool in statistical modeling and is perhaps the single most important concept in statistical inference.

There are reasons for this, and it is my purpose to explain these in this book.

ABOUT THE TEXT

From the author's perspective, the characteristics of this text which most clearly differentiate it from others currently available include the following:

- Applications to a variety of scientific fields, including engineering, appear in every chapter.
- Integration of computer algebra systems such as Mathematica provides insight into both the structure and results of problems in probability.
- A great variety of problems at varying levels of difficulty provides a desirable flexibility in assignments.
- Topics in statistics appear throughout the text so that professors can include or omit these as the nature of their course warrants.
- Some problems are structured and solved using recursions since computers and computer algebra systems facilitate this.
- Significant and practical topics in quality control and quality production are introduced.

It has been my purpose to write a book that is readable by students who have some background in multivariable calculus. Mathematical ideas are often easily understood until one sees formal definitions that frequently obscure such understanding. Examples allow us to explore ideas without the burden of language. Therefore, I often begin with examples and follow with the ideas motivated first by them; this is quite purposeful on my part, since language often obstructs understanding of otherwise simply perceived notions.

I have attempted to give examples that are interesting and often practical in order to show the widespread applicability of the subject. I have sometimes sacrificed exact mathematical precision for the sake of readability; readers who seek a more advanced explication of the subject will have no trouble in finding suitable sources. I have proceeded in the belief that beginning students want most to know what the subject encompasses and for what it may be useful. More theoretical courses may then be chosen as time and opportunity allow. For those interested, the bibliography contains a number of current references.

An author has considerable control over the reader by selecting the material, its order of presentation, and the explication. I am hopeful that I have executed these duties with due regard for the reader. While the author may not be described with any sort of precision as the holder of a tightrope, I have been guided by the admonition: "It's not healthy for the tightrope walker to be misunderstood by the person who's holding the rope."¹

The book makes free use of the now widely available computer algebra systems. I have used Mathematica, Maple, and Derive for various problems and examples in the book, and I hope the reader has access to one of these marvelous mathematical aids. These systems allow us the incredible opportunity to see graphs and surfaces easily, which otherwise would be very difficult and time-consuming to produce. Computer algebra systems make some

¹Smilla's Sense of Snow, by Peter Hoeg (Farrar, Straus and Giroux: New York, 1993).

parts of mathematics visual and thereby add immensely to our understanding. Derivatives, integrals, series expansions, numerical computation, and the solution of recursions are used throughout the book, but the reader will find that only the results are included: in my opinion there is no longer any reason to dwell on calculation of either a numeric or algebraic sort. We can now concentrate on the meaning of the results without being restrained by the often mechanical effort in achieving them; hence our concentration is on the structure of the problem and the insight the solution gives. Graphs are freely drawn and, when appropriate, a geometric view of the problem is given so that the solution and the problem can be visualized. Numerical approximations are given when exact solutions are not feasible. The reader without a computer algebra system can still do the problems; the reader with such a system can reproduce every graph in the book exactly as it appears. I have included a fairly expensive appendix in which computer commands in Mathematica are given for many of the examples in which Mathematica was used; this should also ease the translation to other computer algebra systems. The reader with access to a computer algebra system should refer to Appendix 1 fairly frequently.

Although I hope the book is readable and as completely explanatory as a probability text may be, I know that students often do not read the text, but proceed directly to the problems. There is nothing wrong with this; after all, if the ability to solve practical problems is the goal, then the student who can do this without reading the text is to be admired. Readers are warned, however, that probability problems are rarely repetitive; the solution of one problem does not necessarily give even any sort of hint as to the solution of the next problem. I have included over 840 problems so that a reader who solves the problems can be reasonably assured that the concepts involving them are understood.

The problem sections begin with the easiest problems and gradually work their way up to some reasonably difficult problems while remaining within the scope and level of the book. In discussing a forthcoming examination with my students, I summarize the material and give some suggestions for practice problems, so I have followed each chapter by a Chapter Summary, some suggestions for Review Problems, and finally some Supplementary Problems.

FOR THE INSTRUCTOR

Texts on probability often use generating functions and recursions in the solution of many complex problems; with our use of computer algebra systems, we can determine generating functions, and often their power series expansions, with ease. The structure of generating functions is also used to explain limiting behavior in many situations. Many interesting problems can be best described in terms of recursions; since computer algebra systems allow us to solve such recursions, some discussion of recursive functions is given. Proofs are often given using recursions, a novel feature of the book. Occasionally, the more traditional proofs are given in the exercises.

Although numerous applications of the theory are given in the text and in the problems, the text by no means exhausts the applications of the theory of probability. In addition to solving many practical and varied problems, the theory of probability also provides the basis for the theory of statistical inference and the analysis of data. Statistical analysis is combined with the theory of probability throughout the book. Hypothesis testing, confidence intervals, acceptance sampling, and control charts are considered at various points in

the text. The order in which these topics are to be considered is entirely up to the instructor; the book is quite flexible in allowing sections to be skipped, or delayed, resulting in rearrangement of the material. This book will serve as a first introduction to statistics, but the reader who intends to apply statistics should also elect a course in applied statistics. In my opinion, statistics will be the centerpiece of applied mathematics in the twenty-first century.

Preface for the Second Edition

I am pleased to offer a second edition of this text. The reasons for writing the book remain the same and are indicated in the preface for the first edition. While remaining readable and I hope useful for both the student and the instructor, I want to point out some differences between the two editions.

- The first edition was written when Mathematica was in its fourth release; it is now in its ninth release and while its capabilities have grown, some of the commands, especially those regarding graphs, have changed. Therefore, Appendix 1 is totally new, reflecting the changes in Mathematica.
- Both first and second editions contain about 120 graphs; these have been mostly redrawn.
- The problems are of primary importance to the student. Being able to solve them verifies the student's mastery of the material. The book now contains over 880 problems, 60 or so of which are new.
- Chapter 7, titled “Some Challenging Problems”, is new. Five problems, or sets of problems, some of which have been studied by famous mathematicians, are introduced. Open questions are given, some of which will challenge the reader. Problems are almost always capable of extension; the reader may do this while doing a project regarding one of the major problems.

I have profited from comments from both instructors and students who used the first edition. In a sense I owe a debt to every student of mine at Rose–Hulman Institute of Technology. Heartfelt Thank yous go to Sari Freedman and my editor, Susanne Steitz-Filler of John Wiley & Sons. Sangeetha Parthasarathy of LaserWords has been very helpful and patient during the production process. I have been fortunate to rely on the extensive computer skills of my nephew, Scott Carter to whom I owe a big Thank You. But I owe the greatest debt to my wife, Cherry, who has put up with my long hours in the study. I also owe a pat on the head for Ginger who allowed me to refresh while guiding me on long walks through our Old North End neighborhood.

JOHN J. KINNEY

*March 4, 2014
Colorado Springs*

Chapter 1

Sample Spaces and Probability

1.1 DISCRETE SAMPLE SPACES

Probability theory deals with situations in which there is an element of randomness or chance. Some models of the physical world are *deterministic*, that is, they predict exactly what will happen under certain circumstances. For example, if an object is dropped from a height and given no initial velocity, its distance, s , from the starting point is given by $s = \frac{1}{2} \cdot g \cdot t^2$, where g is the acceleration due to gravity and t is the time. If one tried to apply the formula in a practical situation, one would not find very satisfactory results. The problem is that the formula applies only in a vacuum and ignores the shape of the object and the resistance of the air as well as other factors. Although some of these factors can be determined, we generally combine them and say that the result has a random or chance component. Our model then becomes $s = \frac{1}{2} \cdot g \cdot t^2 + \epsilon$, where ϵ denotes the random component of the model. In contrast with the deterministic model, this model is *stochastic*.

Science often considers stochastic models; in formulating new models, the scientist may try to determine the contributions of both deterministic and random components of the model in predicting accurate results.

The mathematical theory of probability arose in consideration of games of chance, but, as the above-mentioned example shows, it is now widely used in far more practical and applied situations. We encounter other circumstances frequently in everyday life in which we presume that some random factors are at work. Here are some simple examples. What is the chance I will find that all eight traffic lights I pass through on my way to work are green? What are my chances for winning a lottery? I have a ten-volume encyclopedia that I have packed in separate boxes. If the boxes become mixed up and I draw the volumes out at random, what is the chance that my encyclopedia will be in order? My desk lamp has a bulb that is “guaranteed” to last 5000 hours. It has been used for 3000 hours. What is the chance that I must replace it before 2000 more hours are used? Each of these situations involves a random event whose specific outcome is unpredictable in advance.

Probability theory has become important because of the wide variety of practical problems it solves and its role in science. It is also the basis of the statistical analysis of data that is widely used in industry and in experimentation. Consider some examples. A manufacturer of television sets may know that 1% of the television sets manufactured have defects of some kind. What is the chance that a shipment of 200 sets a dealer has received contains 2% defective sets? Solving problems such as these has become important to manufacturers who are anxious to produce high quality products, and indeed such considerations play a central role in what has become known in manufacturing as *statistical process control*.

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

2 Chapter 1 Sample Spaces and Probability

Sample surveys, in which only a portion of a population or reference set is investigated, have become commonplace. A recent survey, for example, showed that two-thirds of welfare recipients in the United States were not old enough to vote. But surely we do not know that exactly two-thirds of all welfare recipients were not old enough to vote; there is some uncertainty, largely dependent on the size of the sample investigated as well as the manner in which the survey was conducted, connected with this result. How is this uncertainty calculated?

As a final example, consider a scientific investigation into say the relationship between temperature, a catalyst, and pressure in creating a chemical compound. A scientist can only carry out a few experiments in which several combinations of temperatures, amount of catalyst, and level of pressure are investigated. Furthermore, there is an element of randomness (largely due to other, unmeasured, factors) that influence the amount of compound produced. How is the scientist to determine which combination of factors maximizes the amount of chemical compound? We will encounter many of these examples in this book.

In some situations, we could measure all the forces involved and predict the outcome precisely but very often choose not to do so. In the traffic light example, we could, by knowledge of the timing of the lights, my speed, and the traffic pattern, predict precisely the color of each light as I approach it. While this is possible, it is probably not worth the effort, so we combine all the forces involved and call the result “chance.” So “chance” as we use it does not imply any new or unknown physical forces; it is simply an umbrella under which we put forces we choose not to measure.

How do we then measure the probability of events such as those described earlier? How do we determine how likely such events are? Such probability problems may be puzzling to us since we lack a framework in which to solve them. We lack a strategy for dealing with the randomness involved in these situations. A sensible way to begin is to consider all the possibilities that could occur. Such a list, or set, is called a *sample space*.

We begin here with some situations that are admittedly much simpler than some of those described earlier; more complex problems will also be encountered in this book.

We will consider situations that we call *experiments*. These are situations that can be repeated under identical circumstances. Those of interest to us will involve some randomness so that the outcomes cannot be precisely predicted in advance. As examples, consider the following:

- Two people are chosen at random from a group of five people.
- Choose one of two brands of breakfast cereal at random.
- Throw two fair dice.
- Take an actuarial examination until it is passed for the first time.
- Any laboratory experiment.

Clearly, the first four of these experiments involve random factors. Laboratory experiments involve random factors as well and we would probably choose not to measure all the factors so as to be able to predict the exact outcome in advance.

Once the conditions for the experiment are set, and we are assured that these conditions can be repeated exactly, we can form the *sample space*, which we define as follows:

Definition A *sample space* is a set of all the possible outcomes from an experiment.

Example 1.1.1

The sample spaces for the first four experiments mentioned above are as follows:

- (a) (Choose two people at random from a group of five people.) Denoting the five people as A, B, C, D , and E , we find, if we disregard the order in which the persons are chosen, that there are ten possible samples of two people:

$$S = \{AB, AC, AD, AE, BC, BD, BE, CD, CE, DE\}.$$

This set, S , then comprises the sample space for the experiment.

If we consider the choice of people as random, we might expect that each of these ten samples occurs about 10% of the time. Further, we see that any particular person, say B , occurs in exactly four of the samples, so we say the *probability* that any particular person is in the sample is $\frac{4}{10} = \frac{2}{5}$. The reader may be interested to show that if three people were selected from a group of five people, then the probability a particular person is in the sample is $\frac{3}{5}$. Here, there is a pattern that we can establish with some results to be developed later in this chapter.

- (b) (Choose one of two brands of breakfast cereal at random.) Denote the brands as K and P . We take the sample space as

$$S = \{K, P\},$$

where the set S contains each of the *elementary* outcomes, K and P .

- (c) (Toss two fair dice.) In contrast with the first two examples, we might consider several different sample spaces. Suppose first that we distinguish the two dice by color, say one is red and the other is green. Then we could write the result of a toss as an ordered pair indicating the outcome on each die, giving say the result on the red die first and the result on the green die second. Let a sample space be

$$S_1 = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (2, 6), \dots, (6, 6)\}.$$

It is useful to see this sample space as a geometric space as in Figure 1.1.

Note that the 36 dots represent the only possible outcomes from the experiment. The sample space is not continuous in any sense in this case and may differ from our notions of a geometric space.

We could also describe all the possible outcomes from the experiment by the set

$$S_2 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

since one of these *sums* must occur when the two dice are thrown.

		Second die						
		6	5	4	3	2	1	
First die	6	•	•	•	•	•	•	
	5	•	•	•	•	•	•	
	4	•	•	•	•	•	•	
	3	•	•	•	•	•	•	
	2	•	•	•	•	•	•	
	1	•	•	•	•	•	•	
1 2 3 4 5 6		Figure 1.1 Sample space for tossing two dice.						

Which sample space should be chosen? Note that each point in S_2 represents at least one point in S_1 . So, while we might consider each of the 36 points in S_1 to occur with equal frequency if we threw the dice a large number of times, we would not consider that to be true if we chose sample space S_2 . A sum of 7, for example, occurs on 6 of the points in S_1 while a sum of 2 occurs at only one point in S_1 . The choice of sample space is largely dependent on what sort of outcomes are of interest when the experiment is performed. It is not uncommon for an experiment to admit more than one sample space. We generally select the sample space most convenient for the analysis of the probabilities involved in the problem.

We continue now with further examples of experiments involving randomness.

- (d) (Take an actuarial examination until it is passed for the first time.) Letting P and F denote passing and failing the examination, respectively, we note that the sample space here is infinite:

$$S = \{P, FP, FFP, FFFF, \dots\}.$$

However, S here is a *countably infinite* sample space since its elements can be counted in the sense that they can be placed in a one-to-one correspondence with the set of natural numbers $\{1, 2, 3, 4, \dots\}$ as follows:

$$P \leftrightarrow 1$$

$$FP \leftrightarrow 2$$

$$FFP \leftrightarrow 3$$

.

.

.

The rule for the one-to-one correspondence is as follows: given an entry in the left column, the corresponding entry in the right column is the number of the attempt on which the examination is passed; given an entry in the right column, say n , consider $n - 1$ F 's followed by P to construct the corresponding entry in the left column. Hence, the correspondence with the set of natural numbers is one-to-one. Such sets are called *countable* or *denumerable*. We will consider countably infinite sets in much the same way that we will consider finite sets. In the next chapter, we will encounter infinite sets that are not countable.

- (e) Sample spaces for laboratory experiments are usually difficult to enumerate and may involve a combination of finite and infinite factors.

Example 1.1.2

As a more difficult example, consider observing single births in a hospital until two girls are born in a row.

The sample space now is a bit more challenging to write down than the sample spaces for the situations considered in Example 1.1.1.

For convenience, we write the points, showing the births in order and grouped by the total number of births.

Number of Births	Sample Points	Number of Sample Points
2	<i>GG</i>	1
3	<i>BGG</i>	1
4	<i>BBGG</i> <i>GBGG</i>	2
5	<i>BBBGG</i> <i>BGBGG</i> <i>GGBGG</i>	4
6	<i>BBBBGG</i> <i>BBGBGG</i> <i>BGBBGG</i> <i>GBBBGG</i> <i>GBGBGG</i>	6

and so on. We note that the number of sample points as we have grouped them follows the sequence 1, 1, 2, 4, 6, ..., which we recognize as the beginning of the Fibonacci sequence. The Fibonacci sequence is found by starting with the sequence 1, 1. Subsequent entries are found by adding the two immediately preceding entries. However, we only have evidence that the Fibonacci sequence applies to a few of the groups of points in the sample space. We will have to establish the general pattern in this example before concluding that the Fibonacci sequence does indeed give the number of sample points in the sample space. The reader may wish to do that before reading the following paragraphs!

Here is the reason the Fibonacci sequence occurs: consider a sequence of *B*'s and *G*'s in which *GG* occurs for the first time at the n th birth. Let a_n denote the number of ways in which this can occur. If *GG* occurs for the first time on the n th birth, there are two possibilities for the beginning of the sequence. These possibilities are mutually exclusive, that is, they cannot occur together.

One possibility is that the sequence begins with a *B* and is followed for the first time by the occurrence of *GG* in $n - 1$ births. Since we are requiring the sequence *GG* to occur for the first time at the $n - 1$ st birth, this can occur in a_{n-1} ways.

The other possibility for the beginning of the sequence is that the sequence begins with *G*, which must then be followed by *B* (else the pattern *GG* will occur in two births) and then the pattern *GG* occurs in $n - 2$ births. This can occur in a_{n-2} ways. Since the sequence begins either with *B* or *G*, it follows that

$$a_n = a_{n-1} + a_{n-2}, \quad n \geq 4,$$

where $a_2 = a_3 = 1$, (1.1)

which describes the Fibonacci sequence.

The sequences for which *GG* occurs for the first time in 7 births can then be found by writing *B* followed by the sequences for 6 births and by writing *GB* followed by *GG* in 5 births:

$$B|BBBBGG$$

$$B|BBGBGG$$

$$B|BGBBGG$$

$$B|GBBBGG$$

$$B|GBGBGG$$

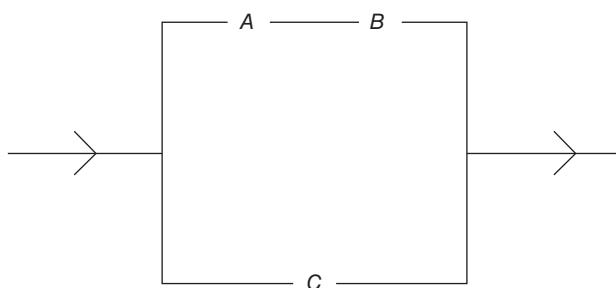
$$GB|BBBGG$$

$$GB|BGBGG$$

$$GB|GBBGG$$

EXERCISES 1.1

1. Show the sample space when 3 people are selected from a group of 5 people. Verify the fact that any particular person in the selected group is $3/5$.
2. In Example 1.1.2, show all the sample points where the births of two girls in a row occur in 8 or 9 births.
3. An experiment consists of drawing two numbered balls from a box of balls numbered from 1 to 9. Describe the sample space if
 - (a) the first ball is not replaced before the second is drawn.
 - (b) the first ball is replaced before the second is drawn.
4. In the diagram below, A, B, and C are switches that may be closed (current flows through the switch) or open (current cannot flow through the switch). Show the sample space indicating all the possible positions of the switches in the circuit.



5. Items being produced on an assembly line can be good (G) or not meeting specifications (N). Show the sample space for the next five items produced by the assembly line.
6. A student decides to take an actuarial examination until it is passed, but will attempt the test at most five times. Show the sample space.
7. In the World Series, games are played until one of the teams has won four games. Show all the points in the sample space in which the American League (A) wins the series over the National League (N) in at most six games.
8. We are interested in the sequence of male and female births in five-child families. Show the sample space.
9. Twelve chips numbered 1 through 12 are mixed in a bowl. Two chips are drawn successively and without replacement. Show the sample space for the experiment.
10. An assembly line is observed until items of both types—good (G) items and items not meeting specification (N)—are observed. Show the sample space.
11. Two numbers are chosen without replacement from the set $\{2, 3, 4, 5, 6, 7\}$, with the additional restriction that the second number chosen must be smaller than the first. Describe an appropriate sample space for the experiment.
12. Computer chips coming off an assembly line are marked defective (D) or nondefective (N). The chips are tested and their condition listed. This is continued until two consecutive defectives are produced or until four chips have been tested, whichever occurs first. Show a sample space for the experiment.
13. A coin is tossed five times and a running count of the heads and tails is kept (so the number of heads and the number of tails tossed so far is recorded at each toss). Show all the sample points where the heads count *always* exceeds the tails count.
14. A sample space consists of all the linear arrangements of the integers 1, 2, 3, 4, and 5. (These linear arrangements are called *permutations*).
 - (a) Use your computer algebra system to list all the sample points.
 - (b) If the sample points are equally likely, what is the probability that the number 3 is in the third position?
 - (c) What is the probability that none of the integers occupies its natural position?

1.2 EVENTS; AXIOMS OF PROBABILITY

After establishing a sample space, we are often interested in particular points, or sets of points, in that sample space. Consider the following examples:

- (a) An item is selected at random from a production line. We are interested in the selection of a good item.
- (b) Two dice are tossed. We are interested in the occurrence of a sum of 5.
- (c) Births are observed until a girl is born. We are interested in this occurring in an even number of births.

Let us begin by defining an *event*.

Definition An *event* is a subset of a sample space.

Events then contain one or more elementary outcomes in the sample space.

In the earlier examples, “a good item is selected,” “the sum is 5,” and “an even number of births was observed” can be described by subsets of the appropriate sample space and are, therefore, *events*.

We say that an event *occurs* if any of the elementary outcomes contained in the event occurs.

We will be interested in the *relative frequency* with which these events occur. In example (a), we would most likely say, if 99% of the items produced in the production line are good, then a good item will be selected about 99% of the time the experiment is performed, but we would expect some variation from this figure. In example (b), such a calculation is more complex since the event “the sum of the spots showing on the dice is 5” comprises several more elementary events. If the sample space distinguishing a red and a green die is

$$S = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\},$$

then the points where the sum is 5 are

$$(1, 4), (2, 3), (3, 2), (4, 1).$$

If the dice are fair, then each of the 36 points in S occurs about $1/36$ of the time, so we conclude that the sum of the spots showing 5 occurs about $4 \cdot \frac{1}{36} = \frac{1}{9}$ of the time.

In example (c), observing births until a girl is born, the event “an even number of births is observed” is much more complex than examples (a) and (b) since there is an infinity of possibilities. How are we to judge the frequency of occurrence of each one? We cannot answer this question at this time, but we will consider it later.

Now we consider a structure so that we can deal with such questions, as well as many others far more complex than those considered so far. We start with some assumptions about any sample space.

Axioms of Probability

We consider the *long-range relative frequency* or *probability* of an event in a sample space. If we perform an experiment 120 times and an event, A , occurs 30 times, then we say that the *relative frequency* of A is $30/120 = 1/4$. In general, if in n trials an event A occurs $n(A)$ times, then we say that the *relative frequency* of A is $\frac{n(A)}{n}$. Of course, if we perform the experiment another n times, we do not expect A to occur exactly the same number of times as before, giving another relative frequency for the event A . We do expect these variable ratios representing relative frequencies to settle down in some manner as n grows large. If A is an event, we denote this limiting relative frequency by the *probability* of A and denote this by $P(A)$.

Definition If A is an event, then the *probability* of A is

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}.$$

We assume at this point that the limit exists. We will discuss this in detail in Chapter 4.

In considering events, it is most convenient to use the language and notation of sets where the following notations are common:

The *union* of sets A and B is denoted by $A \cup B$ where

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\},$$

where the word “or” is used in the inclusive sense, that is, an element in both sets A and B is included in the union of the sets.

The *intersection* of sets A and B is denoted by $A \cap B$ where

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

We will consider the following as axiomatic or self-evident:

- (1) $P(A) \geq 0$, where A is an event,
- (2) $P(S) = 1$, where S is the sample space, and
- (3) If A_1, A_2, \dots are *disjoint* or *mutually exclusive*, that is, they have no sample points in common, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Axioms of probability, of course, should reflect our common intuition about the occurrence of events. Since an event cannot occur with a negative relative frequency, (1) is evident. Since *something* must occur when the experiment is done and since S denotes the entire sample space, S must occur with relative frequency 1, hence assumption (2). Now suppose A and B are events with no sample points in common. We can illustrate events in a graphic manner by drawing a rectangle that represents all the points in S ; events are subsets of this sample space. A diagram showing the event A , that is, the set of all elements of S that are in the event A , is shown in Figure 1.2. Illustrations of sets and their relationships with each other are called *Venn diagrams*.

The event A or B consists of all points in A or in B and so its relative frequency is the sum of the relative frequencies of A or B . This is assumption (3). Figure 1.3 shows a Venn diagram illustrating the disjoint events A and B .

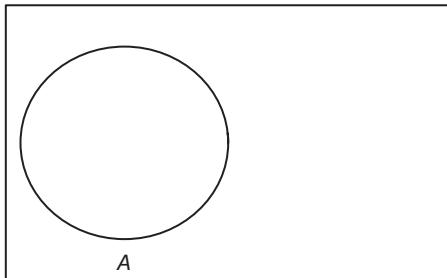


Figure 1.2 Venn diagram showing the event A .

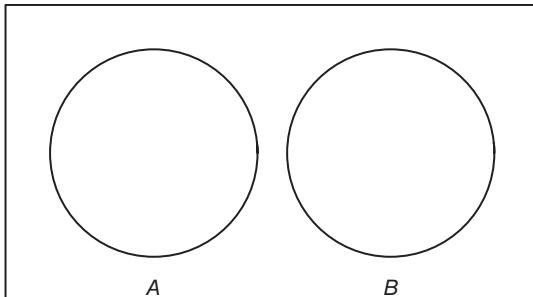


Figure 1.3 Venn diagram showing disjoint events A and B .

No further axioms will be needed in our development of probability theory. We now consider some consequences of these assumptions.

1.3 PROBABILITY THEOREMS

In the above-mentioned example (b), we considered the event that the sum was 5 when two dice were thrown. This event in turn comprises elementary events

$$(1, 4), (2, 3), (3, 2), (4, 1)$$

each of which had probability $\frac{1}{36}$. Since the events $(1, 4), (2, 3), (3, 2)$, and $(4, 1)$ are disjoint, axiom (3) shows that the probability of the event that the sum is 5 is the sum of the probabilities of these four elementary events or $\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{4}{36} = \frac{1}{9}$.

Assumption (3) shows that if A is an event that comprises elementary *disjoint* events $a_1, a_2, a_3, \dots, a_n$, then

Theorem 1:

$$P(A) = \sum_{i=1}^n P(a_i).$$

This fact is often used in the establishment of the theorems we consider in this section. Although we will not do so, all of them can be explained using Theorem 1.

What can we say about $P(A \cup B)$ if A and B have sample points in common? If we find

$$P(A) + P(B)$$

we will have counted the points in the intersection $A \cap B$ twice, as shown in Figure 1.4. So the intersection must be subtracted once giving

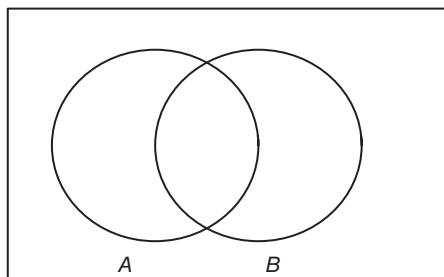


Figure 1.4 Venn diagram showing arbitrary events A and B .

Theorem 2:

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

We call this the *addition theorem* (for two events).

Example 1.3.1

Choose a card from a well-shuffled deck of cards. Let A be the event “the selected card is a heart,” and let B be the event “the selected card is a face card.” Let the sample space S

consist of one point for each of the 52 cards. If the deck is really well shuffled, each point in S can be presumed to have probability $1/52$. The event A contains 13 points and the event B contains 12 points, so $P(A) = 13/52$ and $P(B) = 12/52$. But the events A and B have three sample points in common, those for the King, Queen, and Jack of Hearts. The event $A \cup B$ is then the event “the selected card is a Heart or a face card,” and its probability is

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{22}{52} = \frac{11}{26}. \end{aligned}$$

It is also easy to see by direct counting that the event “the selected card is a Heart or a face card” contains exactly 22 points in the sample space of 52 points.

How can the addition theorem for two events be extended to three or more events? First, consider events A , B , and C in a sample space S . By adding and subtracting probabilities, the reader may be able to see that

Theorem 3:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C), \end{aligned}$$

but we offer another proof as well. This proof will be based on the fact that a correct expression for $P(A \cup B \cup C)$ must count each sample point in the event $A \cup B \cup C$ once and only once. The Venn diagram in Figure 1.5 shows that S comprises 8 disjoint regions labeled as

- 0: points outside $A \cup B \cup C$ (1 region)
- 1: points in A , B , or C alone (3 regions)
- 2: points in exactly two of the events (3 regions)
- 3: points in $A \cap B \cap C$ (1 region).

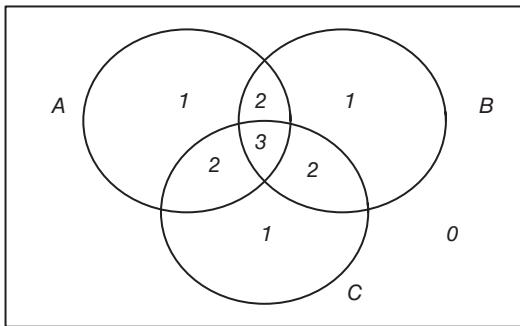


Figure 1.5 Venn diagram showing events, A , B , and C .

Now we show that the right-hand side of Theorem 3 counts each point in the event $A \cup B \cup C$ once and only once. By symmetry, we can consider only four cases:

Case 1. Suppose a point is in event A only. Then its probability is counted only once, in $P(A)$, on the right-hand side of Theorem 3.

Case 2. Suppose a point is in $A \cap B$ only. Then its probability is counted in $P(A)$, $P(B)$, and in $P(A \cap B)$, a net count of one on the right-hand side in Theorem 3.

Case 3. Suppose a point is in $A \cap B \cap C$. Then its probability is counted in each term on the right-hand side of Theorem 3, yielding a net count of 1.

Case 4. If a point is outside $A \cup B \cup C$, then it is not counted on the right-hand side in Theorem 3.

So Theorem 3 must be correct since it counts each point in $A \cup B \cup C$ exactly once and never counts any point outside the event $A \cup B \cup C$. This proof uses a combinatorial principle, that of *inclusion and exclusion*, a principle used in other ways as well in the field of combinatorics. We will make some use of this principle in the remainder of the book. Theorem 2 is of course a special case of Theorem 3.

We would like to extend Theorem 3 to n events, but this requires some combinatorial facts that will be developed later and so we postpone this extension until they are established.

Example 1.3.2

A card is again drawn from a well-shuffled deck. Consider the events

A : the card shows an even number (2, 4, 6, 8, or 10),

B : the card is a Heart, and

C : the card is black.

We use a sample space containing one point for each of the 52 cards in the deck.

Then $P(A) = \frac{20}{52}$, $P(B) = \frac{13}{52}$, $P(C) = \frac{26}{52}$, $P(A \cap B) = \frac{5}{52}$, $P(A \cap C) = \frac{10}{52}$, $P(B \cap C) = 0$,

and $P(A \cap B \cap C) = 0$, so by Theorem 3,

$$P(A \cup B \cup C) = \frac{20}{52} + \frac{13}{52} + \frac{26}{52} - \frac{5}{52} - \frac{10}{52} = \frac{44}{52} = \frac{11}{13}.$$

We will show one more fact in this section. Consider S and an event A in S . Denoting the set of points where the event A does not occur by \bar{A} , it is clear that the events \bar{A} and A are disjoint. So, by Theorem 2, $P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$, which is most often written as

Theorem 4:

$$P(\bar{A}) = 1 - P(A).$$

Example 1.3.3

Throw a pair of fair dice. What is the probability that the dice show different numbers? Here, it is convenient to let A be the event “the dice show different numbers.” Referring to the sample space shown in Figure 1.1, we compute $P(\bar{A})$ since

$$P(\bar{A}) = P(\text{the dice show the same numbers}) = \frac{6}{36} = \frac{1}{6}.$$

$$\text{So } P(A) = 1 - \frac{6}{36} = \frac{5}{6}.$$

This is easier than counting the 30 sample points out of 36 for which the dice show different numbers.

The theorems we have developed so far appear to be fairly simple; the difficulty arises in applying them.

EXERCISES 1.3

1. Verify the probabilities in Example 1.3.2 by specifying the relevant sample points.
2. A fair coin is tossed until a head appears. Find the probability this occurs in four or fewer tosses.
3. A fair coin is tossed five times. Find the probability of obtaining
 - (a) exactly three heads.
 - (b) at most three heads.
4. A manufacturer of pickup trucks is required to recall all the trucks manufactured in a given year for the repair of possible defects in the steering column and defects in the brake linings. Dealers have been notified that 3% of the trucks have defective steering only, and that 6% of the trucks have defective brake linings only. If 87% of the trucks have neither defect, what percentage of the trucks have both defects?
5. A hat contains tags numbered 1, 2, 3, 4, and 5. A tag is drawn from the hat and it is replaced, then a second tag is drawn. Assume that the points in the sample space are equally likely.
 - (a) Show the sample space.
 - (b) Find the probability that the number on the second tag exceeds the number on the first tag.
 - (c) Find the probability that the first tag has a prime number and the second tag has an even number. The number 1 is not considered to be a prime number.
6. A fair coin is tossed four times.
 - (a) Show a sample space for the experiment, showing each possible sequence of tosses.
 - (b) Suppose the sample points are equally likely and that a running count is made of the number of heads and the number of tails tossed. What is the probability the heads count always exceeds the tails count?
 - (c) If the last toss is a tail, what is the probability an even number of heads was tossed?
7. In a sample space of two events is it possible to have $P(A) = 1/2$, $P(A \cap B) = 1/3$ and $P(B) = 1/4$?
8. If A and B are events in a sample space of two events, explain why $P(A \cap \bar{B}) \geq P(A) - P(B)$.
9. In testing the water supply for various cities in a state for two kinds of impurities commonly found in water, it was found that 20% of the water supplies had neither sort of impurity, 40% had an impurity of type A, and 50% had an impurity of type B. If a city is chosen at random, what is the probability its water supply has exactly one type of impurity?
10. A die is loaded so that the probability a face turns up is proportional to the number on that face. If the die is thrown, what is the probability an even number occurs?
11. Show that $P(\bar{A} \cap B) = P(B) - P(A \cap B)$.

12. (a) Explain why $P(A \cup B) \leq P(A) + P(B)$.
 (b) Explain why $P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$.
13. Find a formula for $P(A \text{ or } B)$ using the word “or” in an exclusive sense: that is, A or B means that event A occurs or event B occurs, but not both.
14. The entering class in an engineering college has 34% who intend to major in Mechanical Engineering, 33% who indicate an interest in taking advanced courses in Mathematics as part of their major field of study, and 28% who intend to major in Electrical Engineering, while 23% have other interests. In addition, 59% are known to major in Mechanical Engineering or take advanced Mathematics while 51% intend to major in Electrical Engineering or take advanced Mathematics. Assuming that a student can major in only one field, what percent of the class intends to major in Mechanical Engineering or in Electrical Engineering, but shows no interest in advanced Mathematics?

1.4 CONDITIONAL PROBABILITY AND INDEPENDENCE

Example 1.4.1

Suppose a card is drawn from a well-shuffled deck of 52 cards. What is the probability that the card is a Jack? If the sample space consists of a point for each card in the deck, the answer to the question is $\frac{4}{52}$ since there are four Jacks in the deck.

Now suppose the person choosing the card gives us some additional information. Specifically, suppose we are told that the drawn card is a face card. Now what is the probability that the card is a Jack? An appropriate sample space for the experiment becomes the set of 12 points consisting of all the possible face cards that could be selected:

$$\{JH, QH, KH, JD, QD, KD, JS, QS, KS, JC, QC, KC\}.$$

Considering each of these 12 outcomes to be equally likely, the probability the chosen card is a Jack is now $\frac{4}{12}$. The given additional information that the card is a face card has altered the probability of the event in question. Generally, such additional information, or *conditions*, has the effect of changing the probability of an event as the conditions change. Specifically, the conditions often reduce the sample space and, hence, alter the probabilities on those points that satisfy the conditions.

Let us denote by

A : the event “the chosen card is a Jack”

and

B : the event “the chosen card is a face card”.

Further, we will use the notation $P(A|B)$ to denote the probability of the event A , given that the event B has occurred. We call $P(A|B)$ the *conditional probability* of A given B .

In this example, we see that $P(A|B) = \frac{4}{12}$.

Now we can establish a general result by reasoning as follows. Suppose the event B has occurred; while this reduces the sample space to those points in B , we cannot presume that the probability of the set of points in B is 1. However, if the probability of each point in B is divided by $P(B)$, then the set of points in B has probability 1 and can therefore serve as

a sample space. This division by a constant also preserves the *relative* probabilities of the points in the original sample space; if one point in the original sample space was k times as probable as another, it is still k times as probable as the other point in the new sample space. Clearly, $P(A|B)$ accounts for the points in $A \cap B$ in the new sample space. We have found that

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

where we have presumed of course that $P(B) \neq 0$.

In the earlier example, $P(A \cap B) = \frac{4}{52}$ and $P(B) = \frac{12}{52}$ so $P(A|B) = \frac{4}{12}$ as before.

In this example, $P(A \cap B)$ reduces to $P(A)$, but this will not always be the case.

We can also write this result as

$$P(A \cap B) = P(B) \cdot P(A|B), \text{ or, interchanging } A \text{ and } B,$$

$$P(A \cap B) = P(A) \cdot P(B|A).$$

We call this result the *multiplication theorem*.

Example 1.4.2

A box of transistors has four good transistors mixed up with two bad transistors. A production worker, in order to sample the product, chooses two transistors at random, the first chosen transistor not being replaced before the second transistor is chosen. What is the probability that both transistors are good?

If the events are

A : the first transistor chosen is good

and

B : the second transistor chosen is good,

then we want $P(A \cap B)$.

Now $P(A) = \frac{4}{6}$ while $P(B|A) = \frac{3}{5}$ since the box, after the first good transistor is drawn, contains five transistors, three of which are good transistors. So the probability that both chosen transistors are good is

$$P(A \cap B) = P(A) \cdot P(B|A)$$

$$P(A \cap B) = \frac{4}{6} \cdot \frac{3}{5} = \frac{2}{5}.$$

by the multiplication theorem.

Example 1.4.3

In the context of the earlier example, what is the probability the second transistor chosen is good?

We need $P(B)$. Now B can occur in two mutually exclusive ways: the first transistor is good and the second transistor is also good, or the first transistor is bad and the second transistor is good. So,

$$\begin{aligned} P(B) &= P[(A \cap B) \cup (\bar{A} \cap B)] \\ &= P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A}) \\ P(B) &= \frac{4}{6} \cdot \frac{3}{5} + \frac{2}{6} \cdot \frac{4}{5} = \frac{2}{3}. \end{aligned}$$

We used the fact in this example that

$$P(B) = P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})$$

since B occurs when either A or \bar{A} occurs.

This result can be generalized. Suppose the sample space consists of disjoint events so that

$$S = A_1 \cup A_2 \cup \dots \cup A_n,$$

where A_i and A_j have no sample points in common if $i \neq j$, $i, j = 1, 2, \dots, n$.

Then if B is an event,

$$\begin{aligned} P(B) &= P[(A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B)] \\ &= P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B) \\ &= P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n). \end{aligned}$$

We have then

Theorem: (Law of Total Probability): If $S = A_1 \cup A_2 \cup \dots \cup A_n$ where A_i and A_j have no sample points in common if $i \neq j$, $i, j = 1, 2, \dots, n$, then, if B is an event,

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n) \text{ or}$$

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i).$$

Example 1.4.4

A supplier purchases 10% of its parts from factory A, 20% of its parts from factory B, and the remainder of its parts from factory C. Out of which, 3% of A's parts are defective; 2% of B's parts are defective, and 1/2% of C's parts are defective. What is the probability a randomly selected part is defective?

Let $P(A)$ denote the probability the part is from factory A and define $P(B)$ and $P(C)$ similarly. Let $P(D)$ denote the probability an item is defective. Then, from the law of total probability,

$$P(D) = P(A) \cdot P(D|A) + P(B) \cdot P(D|B) + P(C) \cdot P(D|C) \text{ so}$$

$$P(D) = (0.10) \cdot (0.03) + (0.20) \cdot (0.02) + (0.70) \cdot (0.005) = 0.0105.$$

So 1.05% of the items are defective.

We will encounter other uses of the law of total probability in the following examples.

Example 1.4.5

Suppose, in the context of the previous example, we are given that the second chosen transistor is good. What is the probability the first was also good?

Using the events A and B in the previous example, we want to find $P(A|B)$. That is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

From the previous example, $P(A \cap B) = \frac{4}{6} \cdot \frac{3}{5} = \frac{2}{5}$, and we found in Example 1.4.3 that $P(B) = \frac{2}{3}$, so

$$P(A|B) = \frac{3}{5}.$$

When the earlier results are combined, we see that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})} \quad (1.2)$$

This result is sometimes known as *Bayes' theorem*.

The theorem can easily be extended to three or more mutually disjoint events.

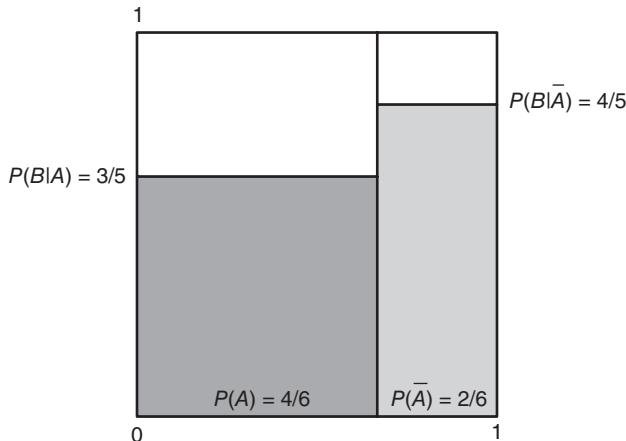
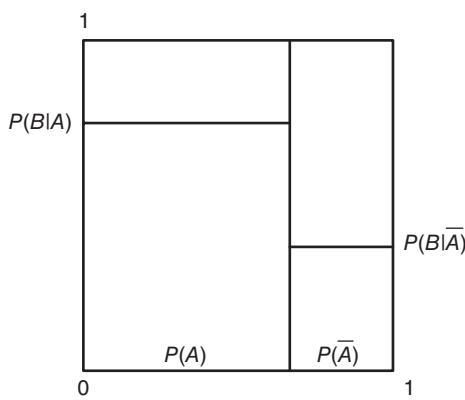
Theorem (Bayes' theorem): If $S = A_1 \cup A_2 \cup \dots \cup A_n$ where A_i and A_j , have no sample points in common if $i \neq j$ then, if B is an event,

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\ P(A_i|B) &= \frac{P(A_i) \cdot P(B|A_i)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n)} \\ \text{and} \\ P(A_i|B) &= \frac{P(A_i) \cdot P(B|A_i)}{\sum_{i=1}^n P(A_i) \cdot P(B|A_i)}. \end{aligned}$$

Rather than remember this result, it is useful to look at Bayes' theorem in a geometric way; it is not nearly as difficult as it may appear. This will first be illustrated using the current example.

Draw a square of side 1; as shown in Figure 1.6, divide the horizontal axis proportional to $P(A)$ and $P(\bar{A})$ —in this case (returning to the context of Example 1.4.5) in the proportions 4/6 to 2/6. Along the vertical axis the conditional probabilities are shown. The vertical axis shows $P(B|A) = 3/5$ and $P(B|\bar{A}) = 4/5$, respectively.

The shaded area above $P(A)$ then shows $P(A) \cdot P(B|A)$. The total shaded area then shows $P(B)P(B) = \frac{4}{6} \cdot \frac{3}{5} + \frac{2}{6} \cdot \frac{4}{5} = \frac{2}{3}$. The doubly shaded region is the proportion of the

**Figure 1.6** Diagram for Example 1.4.5.**Figure 1.7** A geometric view of Bayes' theorem.

shaded area arising from the occurrence of A , which is $P(A|B)$. We see that this is

$$\frac{\frac{4}{6} \cdot \frac{3}{5}}{\frac{4}{6} \cdot \frac{3}{5} + \frac{2}{6} \cdot \frac{4}{5}} = \frac{3}{5}$$

yielding the same result found using Bayes' theorem.

Figure 1.7 shows a geometric view of the general situation.

Bayes' theorem then simply involves the calculation of areas of rectangles.

Example 1.4.6

According to the *New York Times* (September 5, 1987), a test for the presence of the HIV virus exists that gives a positive result (indicating the virus) with certainty if a patient actually has the virus. However, associated with this test, as with most tests, there is a *false positive rate*, that is, the test will sometimes indicate the presence of the virus in patients actually free of the virus. This test has a false

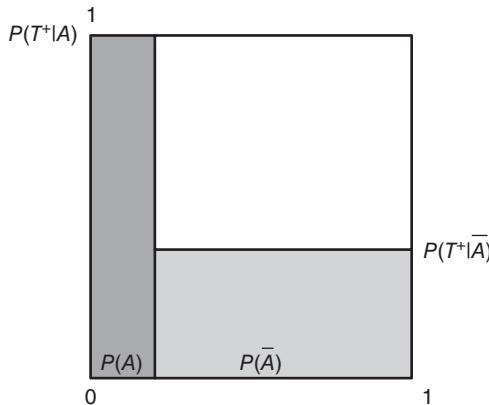


Figure 1.8 AIDS example.

positive rate of 1 in 20,000. So the test would appear to be very sensitive. Assuming now that 1 person in 10,000 is actually HIV positive, what proportion of patients for whom the test indicates HIV actually have the HIV virus? The answer may be surprising.

A picture (greatly exaggerated so that the relevant areas can be seen) is shown in Figure 1.8.

Define the events as

A : patient has AIDS,

T^+ : test indicates patient has AIDS.

Then $P(A) = 0.0001$; $P(T^+|A) = 1$; $P(T^+|\bar{A}) = \frac{1}{20,000}$ from the data given. We are interested in $P(A|T^+)$. So, from Figure 1.8, we see that

$$P(A|T^+) = \frac{(0.0001) \cdot 1}{(0.0001) \cdot 1 + (0.9999) \cdot \frac{1}{20,000}} \text{ or}$$

$$P(A|T^+) = \frac{20,000}{29,999}.$$

We could also of course apply Bayes' theorem to find that

$$\begin{aligned} P(A|T^+) &= \frac{P(A \cap T^+)}{P(T^+)} = \frac{P(A \cap T^+)}{P[(A \cap T^+) \cup (\bar{A} \cap T^+)]} \\ &= \frac{P(A) \cdot P(T^+|A)}{P(A) \cdot P(T^+|A) + P(\bar{A}) \cdot P(T^+|\bar{A})} \\ &= \frac{(0.0001) \cdot 1}{(0.0001) \cdot 1 + (0.9999) \cdot \frac{1}{20,000}} \\ &= \frac{20,000}{29,999}. \end{aligned}$$

giving the same result as that found using simple geometry.

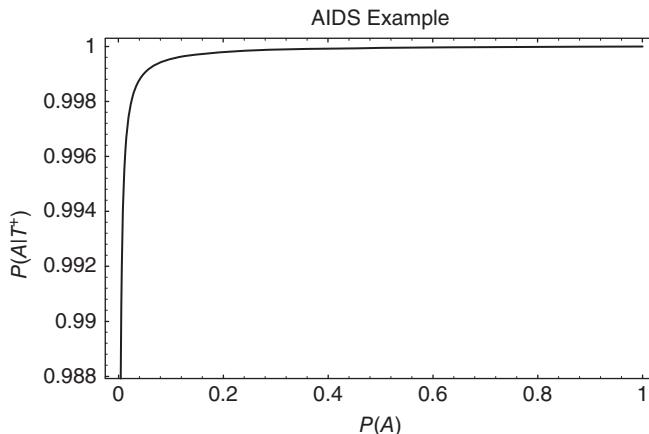


Figure 1.9 $P(A|T^+)$ as a function of $P(A)$.

At first glance, the test would appear to be very sensitive due to its small false positive rate, but only two-thirds of those people testing positive would actually have the virus, showing that widespread use of the test, while detecting many cases of HIV, would also falsely detect the virus in about one-third of the population who test positive. This risk may be unacceptably high.

A graph of $P(A|T^+)$ (shown in Figure 1.9) shows that this probability is highly dependent on $P(A)$.

The graph shows that $P(A|T^+)$ increases as $P(A)$ increases, and that $P(A|T^+)$ is very large even for small values of $P(A)$. For example, if we desire $P(A|T^+)$ to be ≥ 0.9 , then we must have $P(A) \geq 0.0045$.

The sensitivity of the test may incorrectly be associated with $P(T^+|A)$. The patient, however, is concerned with $P(A|T^+)$. This example shows how easy it is to confuse $P(A|T^+)$ with $P(T^+|A)$.

Let us generalize the HIV example to a more general medical test in this way: assume a test has a probability p of indicating a disease among patients actually having the disease; assume also that the test indicates the presence of the disease with probability $1 - p$ among patients not having the disease. Finally, suppose the incidence rate of the disease is r .

If T^+ denotes that the test indicates the disease, and if A denotes the occurrence of the disease, then

$$P(A|T^+) = \frac{r \cdot p}{r \cdot p + (1 - r) \cdot (1 - p)}.$$

For example, if $p = 0.95$ and $r = 0.005$ (indicating that the test is 95% accurate on both those who have the disease and those who do not, and that 5 patients out of 1000 actually have the disease), then $P(A|T^+) = 0.087156$. Since $P(A|T^+) = 0.912844$, a positive result on the test appears to indicate the absence, not the presence of the disease!

This odd result is actually due to the small incidence rate of the disease. Figure 1.10 shows $P(A|T^+)$ as a function of r assuming that $p = 0.95$. We see that $P(A|T^+)$ becomes quite large (≥ 0.8) for $r \geq 0.21$.

It is also interesting to see how r and p , varied together, effect $P(A|T^+)$. The surface is shown in Figure 1.11. The surface shows that $P(A|T^+)$ is large when the test is sensitive, that is, when $P(T^+|A)$ is large, or when the incidence rate $r = P(A)$ is large. But there are also combinations of these values that give large values of $P(A|T^+)$: one of these is $r = 0.2$ and $P(T^+|A) = 0.8$ for then $P(A|T^+) = 1/2$.

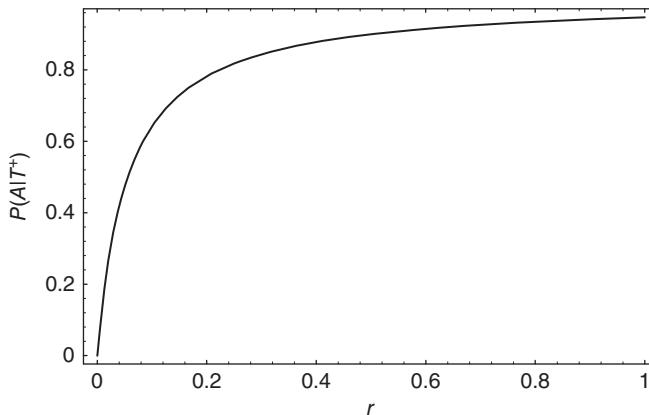


Figure 1.10 $P(A|T^+)$ as a function of the incidence rate, r , if $p = 0.95$.

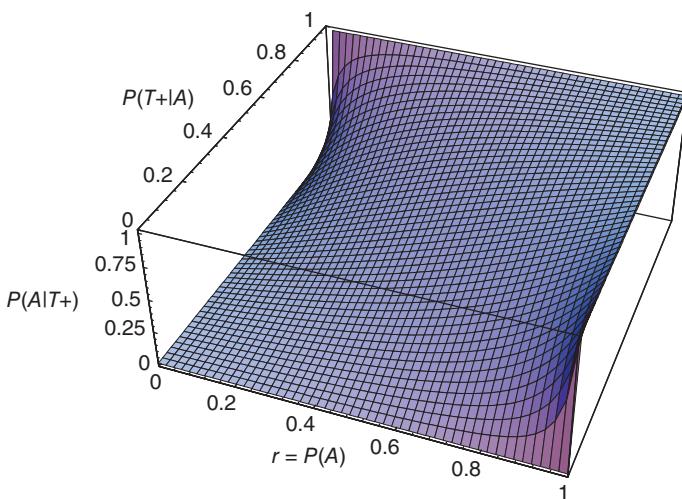


Figure 1.11 $P(A|T^+)$ as a function of r , the incidence rate, and $P(T^+|A)$.

Example 1.4.7

A game show contestant is shown three doors, one of which conceals a valuable prize, while the other two are empty. The contestant is allowed to choose one door. Regardless of the choice made, at least one (i.e., exactly one or perhaps both) of the remaining doors is empty. The show host opens one door to show it empty. The contestant is now given the opportunity to switch doors. Should the contestant switch?

The problem is often called the Monty Hall problem because of its origin on the television show “Let’s Make A Deal.” It has been written about extensively, possibly because of its nonintuitive answer and perhaps because people unwittingly change the problem in the course of thinking about it.

The contestant clearly has a probability of $1/3$ of choosing the prize if a random choice of the door is made. So a probability of $2/3$ rests with the remaining two doors. The fact that one door is opened and revealed empty does not change these probabilities; hence the contestant should switch and will gain the prize with probability $2/3$.

Some think that showing the empty door gives information. Actually, it does not since the contestant already knows that at least one of the doors is empty.

When the empty door is opened, the problem does not suddenly become a choice between two doors (one of which conceals the prize). This change in the problem ignores the fact that the game show host sometimes has a choice of one door to open and sometimes two. Persons changing the problem in this manner may think, incorrectly, that the probability of choosing the prize is now $1/2$, indicating that switching may have no effect in the long run; the strategy in reality has a great effect on the probability of choosing the prize.

To analyze the situation, suppose that the contestant chooses the first door and the host opens the second door. Other possibilities are handled here by symmetry. Let A_i , $i = 1, 2, 3$ denote the event “the prize is behind door i ” and D denote the event “door 2 is opened.” The condition here is then D ; we now calculate the probability the prize is behind door 3, that is, the probability the contestant will win if he switches. We assume that $P(A_1) = P(A_2) = P(A_3) = 1/3$.

Then $P(D|A_1) = 1/2$, $P(D|A_2) = 0$, and $P(D|A_3) = 1$.

The situation is shown in Figure 1.12.

It is clear from the shaded area in Figure 1.12 that the probability the contestant wins if the first choice is switched to door 3 is

$$P(A_3|D) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1} = \frac{2}{3},$$

which verifies our previous analysis.

This example illustrates that some events are highly dependent on others. We now turn our attention to events for which this is not so.

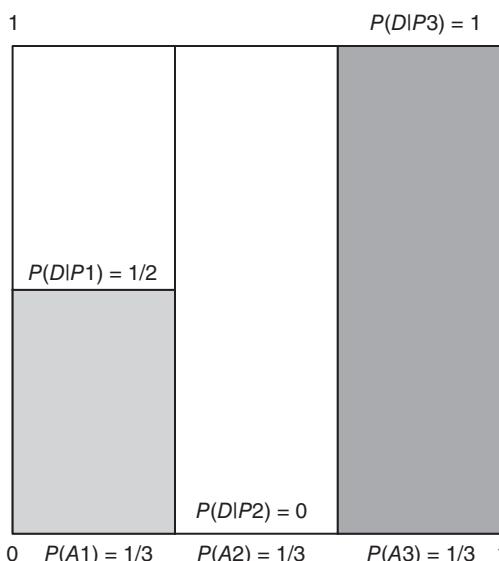


Figure 1.12 Diagram for the Monty Hall problem.

Independence

We have found that $P(A \cap B) = P(A) \cdot P(B|A)$. Occasionally, the occurrence of A has no effect on the occurrence of B so that $P(B|A) = P(B)$. If this is the case, we call A and B as *independent events*. When A and B are independent, we have $P(A \cap B) = P(A) \cdot P(B)$.

Definition Events A and B are called *independent events* if $P(A \cap B) = P(A) \cdot P(B)$.

If we draw cards from a deck, replacing each drawn card before the next card is drawn, then the events denoting the cards drawn are clearly independent since the deck is full before each drawing and each drawing occurs under exactly the same conditions. If the cards are not replaced, however, then the events are not independent.

For three events, say A , B , and C , we define the events as independent if

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B), \\ P(A \cap C) &= P(A) \cdot P(C), \\ P(B \cap C) &= P(B) \cdot P(C), \quad \text{and} \\ P(A \cap B \cap C) &= P(A) \cdot P(B) \cdot P(C). \end{aligned} \tag{1.3}$$

The first three of these conditions establishes that the events are independent in pairs, so we call events satisfying these three conditions as *pairwise independent*. Example 1.4.8 will show that events satisfying these three conditions may not satisfy the fourth condition so pairwise independence does not determine independence.

We also note that there is some confusion between *independent events* and *mutually exclusive events*. Often people speak of these as, “having no effect on each other,” but that is not a precise characterization in either case. Note that while mutually exclusive events cannot occur together, independent events must be able to occur together. To be specific, suppose that neither $P(A)$ nor $P(B)$ is 0, and that A and B are mutually exclusive. Then $P(A \cap B) = 0 \neq P(A) \cdot P(B)$. Hence, A and B cannot be independent. So if A and B are mutually exclusive, then they cannot be independent. This is equivalent to the statement that if A and B are independent, then they cannot be mutually exclusive, but the reader may enjoy establishing this from first principles as well.

Example 1.4.8

This example shows that pairwise independent events are not necessarily independent.

A fair coin is tossed four times. Consider the events A , the first coin shows a head; B , the third coin shows a tail; and C , there are equal numbers of heads and tails. Are these events independent?

Suppose the sample space consists of the 16 points showing the tosses of the coins in order. The sample space, indicating the events that occur at each point, is as follows:

Point	Event
HHHH	A
HHHT	A
HHTH	A, B

Point	Event
<i>THHH</i>	
<i>HTHH</i>	<i>A</i>
<i>HHTT</i>	<i>A, B, C</i>
<i>HTHT</i>	<i>A, C</i>
<i>THHT</i>	<i>C</i>
<i>THTH</i>	<i>B, C</i>
<i>HTTH</i>	<i>A, B, C</i>
<i>TTHH</i>	<i>C</i>
<i>TTTH</i>	<i>B</i>
<i>TTHT</i>	
<i>THTT</i>	<i>B</i>
<i>HTTT</i>	<i>A, B</i>
<i>TTTT</i>	<i>B</i>

Then $P(A) = 1/2$ and $P(B) = 1/2$ while C consists of the 6 points with exactly two heads and two tails, so $P(C) = 6/16 = 3/8$.

Now $P(A \cap B) = \frac{4}{16} = \frac{1}{4} = P(A) \cdot P(B)$; $P(A \cap C) = \frac{3}{16} = P(A) \cdot P(C)$; and $P(B \cap C) = \frac{3}{16} = P(B) \cdot P(C)$, so the events A , B , and C are pairwise independent.

Now $A \cap B \cap C$ consists of the two points *HTTH* and *HHTT* with probability $2/16 = 1/8$.

Hence, $P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$, so A , B , and C are not independent.

Formulas (1.3) also show that establishing only that $P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$ is not sufficient to establish the independence of events A , B , and C .

EXERCISES 1.4

1. In Example 1.4.6, verify $P(A|T^+)$ and $P(\bar{A}|T^+)$.
2. Example 1.4.8 defines the events D: the first head occurs on an even numbered toss and E: at least three heads occur. Are D and E independent?
3. Box I contains 4 green and 5 brown marbles. Box II contains 6 green and 8 brown marbles. A marble is chosen from Box I and placed in Box II, then a marble is drawn from Box II.
 - (a) What is the probability the second marble chosen is green?
 - (b) If the second marble chosen is green, what is the probability a brown marble was transferred?
4. A football team wins its weekly game with probability 0.7. Suppose the outcomes of games on 3 successive weekends are independent. What is the probability the number of wins exceeds the number of losses?
5. Three manufacturers of floppy disks, A, B, and C, produce 15%, 25%, and 60% of the floppy disks made, respectively. Manufacturer A produces 5% defective disks, manufacturer B produces 7% defective disks, and manufacturer C produces 4% defective disks.
 - (a) What proportion of floppy disks are defective?

- (b) If a floppy disk is found to be defective, what is the probability it came from manufacturer B?
6. A chest contains three drawers, each containing a coin. One coin is silver on both sides, one is gold on both sides, and the third is silver on one side and gold on the other side. A drawer is chosen at random and one face of the coin is shown to be silver. What is the probability that the other side is silver also?
7. If A and B are independent events in a sample space, show that

$$P(A \cup B) = P(B) + P(A) \cdot P(\bar{B}) = P(A) + P(\bar{A}) \cdot P(B).$$

8. In a sample space, events A and B are independent: events B and C are mutually exclusive, and A and C are independent. If $P(A \cup B \cup C) = 0.9$ while $P(B) = 0.5$ and $P(C) = 0.3$, find $P(A)$.
9. If $P(A \cup B) = 0.4$ and $P(A) = 0.3$, find $P(B)$ if
- (a) A and B are independent.
 - (b) A and B are mutually exclusive.
10. A coin, loaded so that the probability it shows heads when tossed is $3/4$, is tossed twice. Let the events A , B , and C , be “first toss is heads,” “second toss is heads,” and “tosses show the same face,” respectively.
- (a) Are the events A and B independent?
 - (b) Are the events A and $B \cup C$ independent?
 - (c) Are the events A , B , and C independent?
11. Three missiles, whose probabilities of hitting a target are 0.7, 0.8, and 0.9, respectively, are fired at a target. Assuming independence, what is the probability the target is hit?
12. A student takes a driving test until it is passed. If the probability the test is passed on any attempt is $4/7$ and if the attempts are independent, what is the probability the test is taken an even number of times?
13. (a) Let p be the probability of obtaining a 5 at least once in n independent tosses of a die. What is the least value of n so that $p \geq 1/2$?
- (b) Generalize the result in part (a): suppose an event has probability p of occurring at any one of n independent trials of an experiment. What is the least value of n so that the probability the event occurs at least once is $\geq r$?
- (c) Graph the surface in part (b), showing n as a function of p and r .
14. Box I contains 7 red and 3 black balls; Box II contains 4 red and 5 black balls. After a randomly selected ball is transferred from Box I to Box II, 2 balls are drawn from Box II without replacement. Given that the two balls are red, what is the probability a black ball was transferred?
15. In rolling a fair die, what is the probability of rolling 1 before rolling an even number?
16. (a) There is a fifty-fifty chance that firm A will bid for the construction of a bridge. Firm B submits a bid and the probability that it will get the job is $2/3$, provided firm A does not bid; if firm A submits a bid, the probability firm B gets the job is $1/5$. Firm B is awarded the job; what is the probability firm A did not bid?
- (b) In part (a), suppose now that the probability firm B gets the job if firm A bids on the job is p . Graph the probability that firm A did not bid given that B gets the job as a function of p .

- (c) Generalize parts (a) and (b) further and suppose that the probability that B gets the job given that firm A bids on the job is r . Graph the surface showing the probability firm A did not bid, given that firm B gets the job as a function of p and r .
17. In a sample space, events A and B have probabilities $P(A) = P(B) = 1/2$, and $P(A \cup B) = 2/3$.
- Are A and B mutually exclusive?
 - Are A and B independent?
 - Calculate $P(\bar{A} \cap B)$.
 - Calculate $P(\bar{A} \cap \bar{B})$.
18. Suppose that events A , B , and C are independent with $P(A) = 1/4$, $P(B) = 1/2$, and $P(A \cup B \cup C) = 3/4$. Find $P(C)$.
19. A fair coin is tossed until the same face occurs twice in a row, but it is tossed no more than four times. If the experiment is over no later than the third toss, what is the probability that it is over by the second toss?
20. A collection of 65 coins contains one with two heads; the remainder of the coins are fair. If a coin, selected at random from the collection, turns up heads six times in six tosses, what is the probability that it is the two-headed coin?
21. Three distinct methods, A, B, and C, are available for teaching a certain industrial skill. The failure rates are 30%, 20%, and 10%, respectively. However, due to costs, A is used twice as frequently as B, which is used twice as frequently as C.
- What is the overall failure rate in teaching the skill?
 - A worker is taught the skill, but fails to learn it correctly. What is the probability he was taught by method A?
22. Sixty percent of new drivers have had driver education. During their first year of driving, drivers without driver education have a probability 0.08 of having an accident, but new drivers with driver education have a 0.05 probability of having an accident. What is the probability a new driver with no accidents during the first year had driver education?
23. Events A , B , and C have $P(A) = 0.3$, $P(B) = 0.2$, and $P(C) = 0.4$. Also A and B are mutually exclusive; A and C are independent and B and C are independent. Find the probability that *exactly one* of the events A , B , or C occurs.
24. A set consists of the six possible arrangements of the letters a , b , and c , as well as the points (a, a, a) , (b, b, b) , and (c, c, c) . Let A_k be the event “letter a is in position k ” for $k = 1, 2, 3$. Show that the events A_k are pairwise independent, but that they are not independent.
25. Assume that the probability a first-born child is a boy is p , and that the sex of subsequent children follows a chance mechanism so that the probability the next child is the same sex as the previous child is r .
- Let P_n denote the probability that the n th child is a boy. Find P_i , $i = 1, 2, 3$, in terms of p and r .
 - Are the events A_i : “the i th child is a boy”, $i = 1, 2, 3$ mutually independent?
 - Find a value for r so that A_1 and A_2 are independent.
26. A message is coded into the binary symbols 0 and 1 and the message is sent over a communication channel. The probability a 0 is sent is 0.4 and the probability a 1 is

sent is 0.6. The channel, however, has a random error that changes a 1 to a 0 with probability 0.2 and changes a 0 to a 1 with probability 0.1.

- (a) What is the probability a 0 is received?
- (b) If a 1 is received, what is the probability a 0 was sent?

27. (a) Hospital patients with a certain disease are known to recover with probability $1/2$ if they do not receive a certain drug. The probability of recovery is $3/4$ if the drug is used. Of 100 patients, 10 are selected to receive the drug. If a patient recovers, what is the probability the drug was used?
- (b) In part (a), let the probability the drug is used be p . Graph the probability the drug was used given the patient recovers as a function of p .
- (c) Find p if the probability the drug was used given that the patient recovers is $1/2$.
28. Two people each toss four fair coins. What is the probability they each throw the same number of heads?
29. In sample surveys, people may be asked questions which they regard as sensitive and so they may or may not answer them truthfully. An example might be, “Are you using illegal drugs?” If it is important to discover the real proportion of illegal drug users in the population, the following procedure often called a *randomized response* technique may be used.

The respondent is asked to flip a fair coin and not reveal the result to the questioner. If the result is heads, then the respondent answers the question, “Is your Social Security number even?” If the coin comes up tails, the respondent answers the sensitive question. Clearly the questioner cannot tell whether a response of “yes” is a consequence of illegal drug use or of an even Social Security number. Explain, however, how the results of such a survey to a large number of respondents can be used to find accurately the percentage of the respondents who are users of illegal drugs.

30. (a) The individual events in a series of independent events have probabilities

$$1/2, (1/2)^2, (1/2)^3, \dots, (1/2)^n.$$

Show the probability that at least one of the events occurs approaches 0.711 as $n \rightarrow \infty$.

- (b) Show, if the probabilities of the events are $1/3, (1/3)^2, (1/3)^3, \dots, (1/3)^n$, that the probability at least one of the events occurs approaches 0.440 as $n \rightarrow \infty$.
- (c) Show, if the probabilities of the events are p, p^2, p^3, \dots, p^n , that the probability at least one of the events occurs can be very well approximated by the function $1 - p - p^2 + p^5 + p^7$ for $1/11 \leq p \leq 1/2$.

31. (a) If events A and B are independent, show that

1. \bar{A} and B are independent.
2. A and \bar{B} are independent.
3. \bar{A} and \bar{B} are independent.

- (b) Show that events A and B are independent if and only if $P(A|B) = P(A|\bar{B})$.

32. A lie detector is accurate $3/4$ of the time. That is, if a person is telling the truth, the lie detector indicates he is telling the truth with probability $3/4$ while if the person is lying, the lie detector indicates that he is lying with probability $3/4$. Assume that a person taking the lie detector test is unable to influence its results and also assume that

95% of the people taking the test tell the truth. What is the probability that a person is lying if the lie detector indicates that he is lying?

1.5 SOME EXAMPLES

We now show two examples of probability problems that have interesting results which may counter intuition.

Example 1.5.1 (*The Birthday Problem*)

This problem exists in many variations in the literature on probability and has been written about extensively. The basic problem is this: There are n people in a room; what is the probability that at least two of them have the same birthday?

Let A denote the event “at least two people have the same birthday”; we want to find $P(A)$. It is easier in this case to calculate $P(\bar{A})$ (the probability the birthdays are all distinct) rather than $P(A)$. To find $P(\bar{A})$, note that the first person can have any day as a birthday. The birthday of the next person cannot match that of the first person; this has probability $\frac{364}{365}$; the birthday of the third person cannot match that of either of the first two people; this has probability $\frac{363}{365}$, and so on. So, multiplying these conditional probabilities,

$$P(\bar{A}) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - (n - 1)}{365}.$$

It is easy with a computer algebra system to calculate exact values for $P(A) = 1 - P(\bar{A})$ for various values of n :

n	$P(A)$	n	$P(A)$	n	$P(A)$
2	0.002740	18	0.346911	34	0.795317
3	0.008204	19	0.379119	35	0.814383
4	0.016356	20	0.411438	36	0.832182
5	0.027136	21	0.443688	37	0.848734
6	0.040462	22	0.475695	38	0.864068
7	0.056236	23	0.507297	39	0.878220
8	0.074335	24	0.538344	40	0.891232
9	0.094624	25	0.568700		
10	0.116948	26	0.598241		
11	0.141141	27	0.626859		
12	0.167025	28	0.654461		
13	0.194410	29	0.680969		
14	0.223103	30	0.706316		
15	0.252901	31	0.730455		
16	0.283604	32	0.753348		
17	0.315008	33	0.774972		

We see that $P(A)$ increases rather rapidly; it exceeds $1/2$ for $n = 23$, a fact that surprises many, most people guessing that the value of n to make $P(A) \geq \frac{1}{2}$ is much larger. In thinking about this, note that the problem says that *any* two people in the room can share

any birthday. If some specific date comes to mind, such as August 2, then, since the probability a particular person's birthday is not August 2 is $\frac{364}{365}$, the probability that at least one person in a group of n people has that specific birthday is

$$1 - \left(\frac{364}{365}\right)^n.$$

It is easy to solve this for some specific probability. We find, for example, that for this probability to equal $1/2$, $n = 253$ people are necessary.

We show a graph, in Figure 1.13, of $P(A)$ for $n = 1, 2, 3, \dots, 40$. The graph indicates that $P(A)$ increases quite rapidly as n , the number of people, increases.

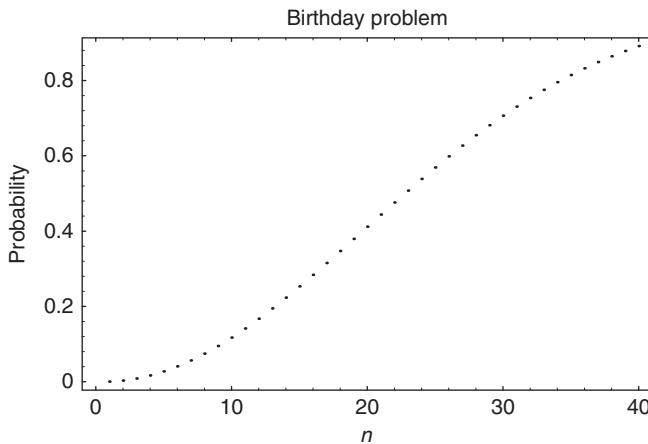


Figure 1.13 The birthday problem as a function of n , the number of people in the group.

It would appear that $P(A)$ might be approximated by a polynomial function of n . To consider how such functions can be constructed would be a diversion now, so we will not discuss it. For now, we state that the least squares approximating function found by applying a principle known as least squares is

$$\begin{aligned} f(n) = & -6.44778 \cdot 10^{-3} - 4.54359 \cdot 10^{-5} \cdot n \\ & + 1.51787 \cdot 10^{-3} \cdot n^2 - 2.40561 \cdot 10^{-5} \cdot n^3. \end{aligned}$$

It can be shown that $f(n)$ fits $P(A)$ quite well in the range $2 \leq n \leq 40$. For example, if $n = 13$, $P(A) = 0.194410$ while $f(13) = 0.196630$; if $n = 27$, $P(A) = 0.626859$ while $f(27) = 0.625357$. A graph of $P(A)$ and the approximating function $f(n)$ is shown in Figure 1.14. The principle of least squares will be considered in Section 4.16.

Example 1.5.2

How many people must be in a group so that the probability at least two of them have birthdays within at most one day of each other is at least $1/2$?

Suppose there are n people in the group, and that A represents the event “at least two people have birthdays within at most one day of each other.” If a person’s birthday is August

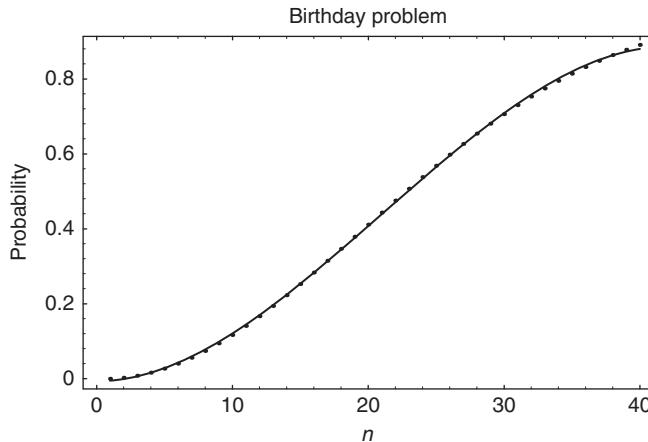


Figure 1.14 Polynomial approximation to the birthday data.

2, for example, then the second person's birthday must not fall on August 1, 2, or 3, giving 362 choices for the second person's birthday. The third person, however, has either 359 or 360 choices, depending on whether the second person's birthday is August 4 or July 31 or some other day that has not previously been excluded from the possibilities. We give then an approximate solution as

$$P(\bar{A}) = \frac{365 \cdot 362 \cdot 359 \cdots (368 - 3n)}{365 \cdot 365 \cdots 365}.$$

We seek $P(A) = 1 - P(\bar{A})$. It is easy to make a table of values of n and $P(A)$ with a computer algebra system.

n	$P(A)$
2	0.008219
3	0.024522
4	0.048575
5	0.079855
6	0.117669
7	0.161181
8	0.209442
9	0.261424
10	0.316058
11	0.372273
12	0.429026
13	0.485341
14	0.540332
15	0.593226
16	0.643376

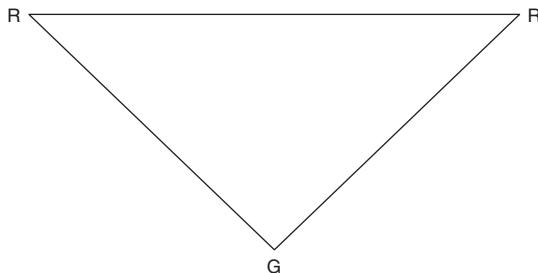
So 14 people are sufficient to make the probability that at least two of the birthdays differ by at most one day exceed 1/2. In the previous example, we found that a group of 23 people was sufficient to make the probability that at least two of them shared the same

birthday to exceed 1/2. The probability is approximately 0.8915 that at least two of these people have birthdays that differ by at most one day.

Example 1.5.3 *(Mowing the Lawn)*

Jack and his daughter, Kaylyn, choose who will mow the lawn by a random process: Jack has one green and two red marbles in his pocket; two are selected at random. If the colors match, Jack mows the lawn, otherwise, Kaylyn mows the lawn. Is the game fair?

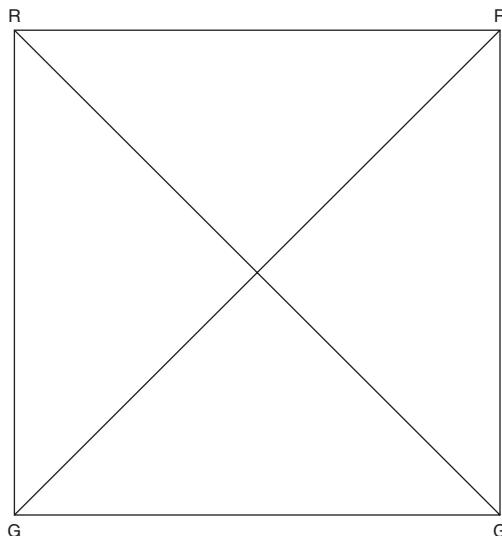
The sample space here is most easily shown by a diagram containing the colors of the marbles as vertices and the edges as the two marbles chosen. Assuming that the three possible samples are equally likely, then two of them lead to Kaylyn mowing the lawn, while Jack only mows it 1/3 of the time. If we mean by the word “fair” that each mows the lawn with probability 1/2, then the game is clearly unfair.



Three marbles in the lawn mowing example.

If we are allowed to add marbles to Jack’s pocket, can the game be made fair? The reader might want to think about this before proceeding.

What if a green marble is added? Then the sample space becomes all the sides and diagonals of a square:



Four marbles in the lawn mowing example.

Although there are now six possible samples, four of them involve different colors while only two of them involve the same colors. So the probability that the colors differ is $\frac{4}{6} = \frac{2}{3}$; the addition of the green marble has not altered the game at all! The reader will easily verify that the addition of a *red* marble, rather than a green marble, will produce a fair game.

The problem of course is that, while the number of red and green marbles is important, the relevant information is the number of sides and diagonals of the figure produced since these represent the samples chosen. If we wish to find other compositions of marbles in Jack's pocket that make the game fair, we need to be able to count these sides and diagonals. We now show how to do this.

Consider a figure with n vertices, as shown in Figure 1.15.

In order to count the number of sides and diagonals, choose one of the n vertices. Now, to choose a side or diagonal, choose any of the other $n - 1$ vertices and join them. We have then $n \cdot (n - 1)$ choices. Since it does not matter which vertex is chosen first, we have counted each side or diagonal twice. We conclude that there are $\frac{n \cdot (n - 1)}{2}$ sides and diagonals.

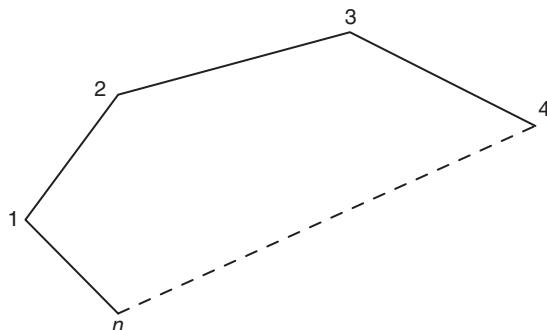


Figure 1.15 n marbles for the lawn mowing problem.

This is also called the number of *combinations of n distinct objects chosen two at a time*, which we denote by the symbol $\binom{n}{2}$. So

$$\binom{n}{2} = \frac{n \cdot (n - 1)}{2}.$$

If the game is to be fair, and if we have r red and g green marbles, then $\binom{r}{2}$ and $\binom{g}{2}$ represent the number of sides and diagonals connecting two red or two green marbles, respectively. We want r and g so that the sum of these is $\frac{1}{2}$ of the total number of sides and diagonals, that is, we want r and g so that

$$\binom{r}{2} + \binom{g}{2} = \left(\frac{1}{2}\right) \cdot \binom{r+g}{2}.$$

The reader can verify that $r = 6$, $g = 3$ will satisfy the above equation as will $r = 10$, $g = 6$. The reader may also enjoy trying to find a general pattern for r and g before reading problem 3 in Exercises 1.5.

EXERCISES 1.5

1. In the birthday problem, verify the probability that, in a group of 23 people, the probability that at least two people have birthdays differing by at most 1 day is 0.8915.
2. In the birthday problem, verify that the values of $f(n)$, the polynomial approximation to $P(A)$, are correct for $f(13)$ and $f(27)$.
3. Show that the “mowing the lawn” game is fair if and only if r and g , the number of red and green marbles, respectively, are consecutive triangular numbers. (The first few triangular numbers are 1, $1 + 2 = 3$, $1 + 2 + 3 = 6$, ...)
4. A fair coin is tossed until a head appears or until six tails have been obtained.
 - (a) What is the probability the experiment ends in an even number of tosses?
 - (b) Answer part (a) if the coin has been loaded so as to show heads with probability p .
5. Let P_r be the probability that among r people, at least two have the same birth month. Make a table of values of P_r for $r = 2, 3, \dots, 12$. Plot a graph of P_r as a function of r .
6. Two defective transistors become mixed up with two good ones. The four transistors are tested one at a time, without replacement, until all the defectives are identified. Find P_r , the probability that the r th transistor tested will be the second defective, for $r = 2, 3, 4$.
7. A coin is tossed four times and the sequence of heads and tails is observed.
 - (a) What is the probability that heads and tails occur equally often if the coin is fair and the tosses are independent?
 - (b) Now suppose the coin is loaded so that $P(H) = 1/3$ and $P(T) = 2/3$ and that the tosses are independent. What is the probability that heads and tails occur equally often, given that the first toss is a head?
8. The following model is sometimes used to model the spread of a contagious disease. Suppose a box contains b black and r red marbles. A marble is drawn and c marbles of that color together with the drawn marble are replaced in the box before the next marble is drawn, so that infected persons infect others while immunity to the disease may also increase.
 - (a) Find the probability that the first three marbles drawn are red.
 - (b) Show that the probability of drawing a black on the second draw is the same as the probability of drawing a black on the first draw.
 - (c) Show by induction that the probability the k th marble is black is the same as the probability of drawing a black on the first draw.
9. A set of 25 items contains five defective items. Items are sampled at random one at a time. What is the probability that the third and fourth defectives occur at the fifth and sixth sample draws if
 - (a) the items are replaced after each is drawn?
 - (b) the items are not replaced after each is drawn?
10. A biased coin has probability $3/8$ of coming up heads. A and B toss this coin with A tossing first.
 - (a) Show that the probability that A gets a head before B gets a tail is very close to $1/2$.
 - (b) How can the coin be loaded so as to make the probability in part (a) $1/2$?

1.6 RELIABILITY OF SYSTEMS

Mechanical and electrical systems are often composed of separate components which may or may not function independently. The space shuttle, for example, comprises hundreds of systems, each of which may have hundreds or thousands of components. The components are, of course, subject to possible failure and these failures in turn may cause individual systems to fail, and ultimately for the entire system to fail. We pause here to consider in some situations how the probability of failure of a component may influence the probability of failure of the system of which it is a part.

In general, we refer to the *reliability*, $R(t)$, of a component as the probability the component will function properly, or survive, for a given period of time. If we denote the event “the component lasts at least t units of time” by $T > t$, then

$$R(t) = P(T > t),$$

where t is fixed.

The reliability of the system depends on two factors: the reliability of its component parts as well as the manner in which they are connected. We will consider some systems in this section, which have few components and elementary patterns of connection.

We will presume that interest centers on the probability an entire system lasts a given period of time; we will calculate this as a function of the probabilities the components last for that amount of time. To do this, we repeatedly use the addition law and multiplication of probabilities.

Series Systems

If a system of two components functions only if both of the components function, then the components are connected in *series*. Such a system is shown in Figure 1.16.

Let p_A and p_B denote the reliabilities of the components A and B , that is,

$$p_A = P(A \text{ survives at least } t \text{ units of time}) \text{ and}$$

$$p_B = P(B \text{ survives at least } t \text{ units of time})$$

for some fixed value t .

If the components function independently, then the reliability of the system, say R , is the product of the individual reliabilities so

$$R = P(A \text{ survives at least } t \text{ units of time and } B \text{ survives at least } t \text{ units of time})$$

$$= P(A \text{ survives at least } t \text{ units of time}) \cdot P(B \text{ survives at least } t \text{ units of time}) \text{ so}$$

$$R = p_A \cdot p_B.$$



Figure 1.16 A series system of two components.

Parallel Systems

If a system of two components functions if either (or both) of the components function, then the components are connected in *parallel*. Such a system is shown in Figure 1.17.

One way to calculate the reliability of the system depends on the fact that at least one of the components must function properly for the given period of time so

$$R = P(A \text{ or } B \text{ survives for a given period of time}) \text{ so,}$$

by the addition law,

$$R = p_A + p_B - p_A \cdot p_B.$$

It is also clear, if the system is to function, that not both of the components can fail so

$$R = 1 - (1 - p_A) \cdot (1 - p_B).$$

These two expressions for R are equivalent.

Figure 1.18 shows the reliability of both series and parallel systems as a function of p_A and p_B . The parallel system is always more reliable than the series system since, for the parallel system to function, at least one of the components must function, while the series system functions only if both components function simultaneously.

Series and parallel systems may be combined in fairly complex ways. We can calculate the reliability of the system from the formulas we have established.

Example 1.6.1

The reliability of the system shown in Figure 1.19 can be calculated by using the addition law and multiplication of probabilities.

The connection of components A and B in the top section can be replaced by a single component with reliability $p_A \cdot p_B$. The parallel connection of switches C and D can be replaced by a single switch with reliability $1 - (1 - p_C) \cdot (1 - p_D)$. The reliability of the resulting parallel system is then

$$1 - (1 - p_A \cdot p_B) \cdot [1 - \{1 - (1 - p_C) \cdot (1 - p_D)\}].$$

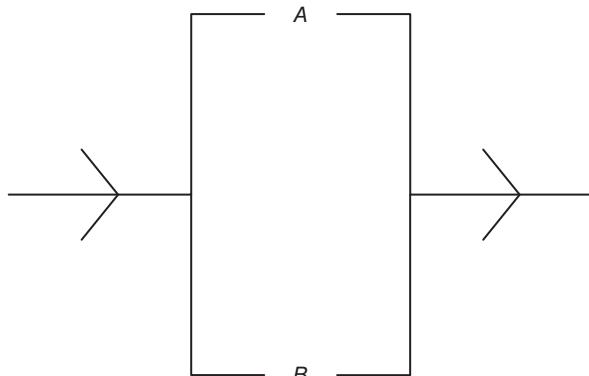


Figure 1.17 A parallel system of two components.

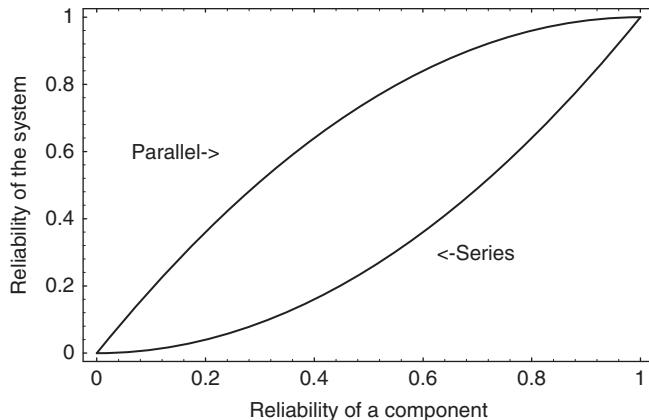


Figure 1.18 Reliability of series and parallel systems.

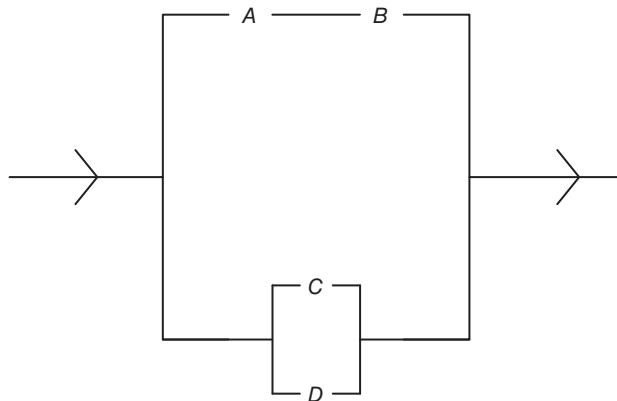


Figure 1.19 System for Example 1.6.1.

A graph of the surface generated, assuming $p_A = p_B$ and $p_C = p_D$, is shown in Figure 1.20.

A contour plot of a surface shows values of p_A and p_C for which the reliability takes on particular values. Figure 1.21 shows a contour plot of the surface for Example 1.6.1, with contours specified at levels 0.80, 0.85, 0.90, 0.95, 0.99, and 0.995 for the reliability. The contour plot shows that if either p_A or p_C is 1, then the reliability is 1. The next contour shows choices of p_A and p_C giving reliability 0.995. The surface indicates that the system is highly reliable if either of the components is highly reliable and that, otherwise, the reliability declines rapidly.

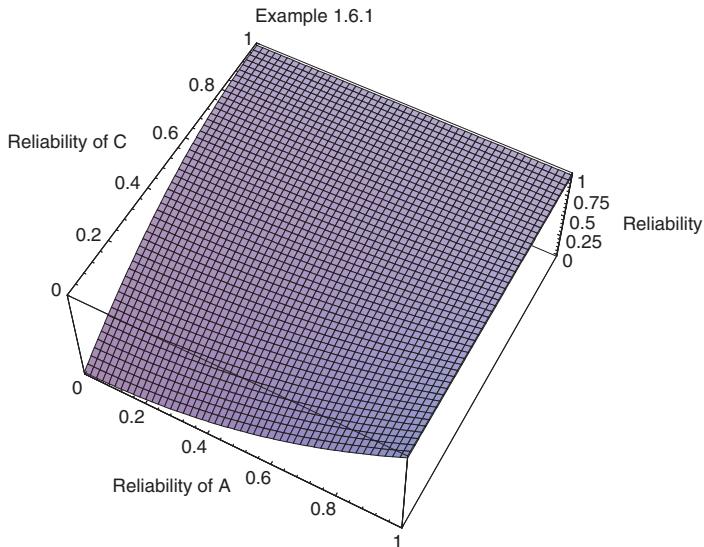


Figure 1.20 Reliability surface for Example 1.6.1.

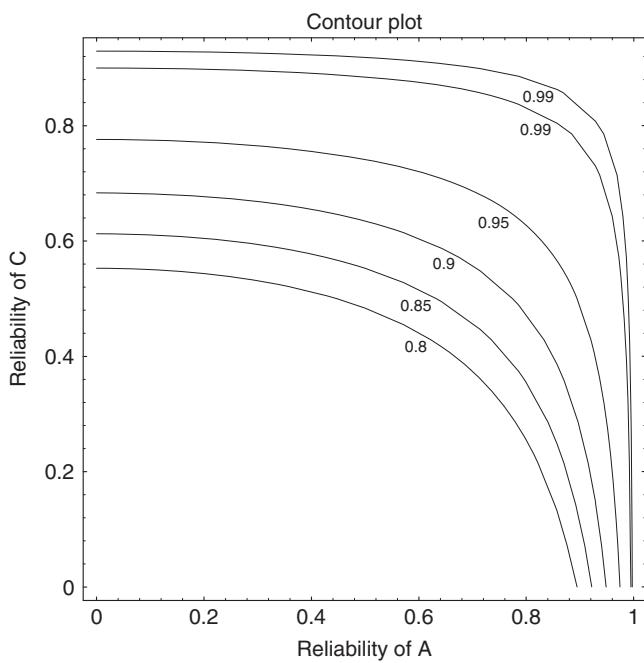
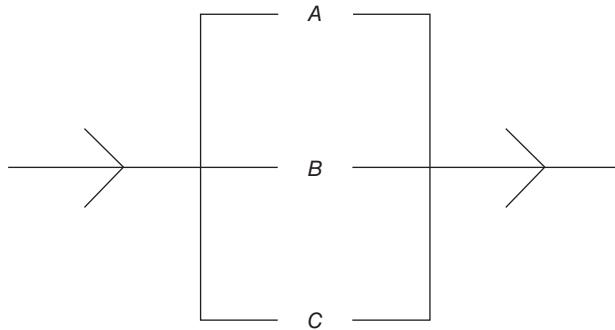


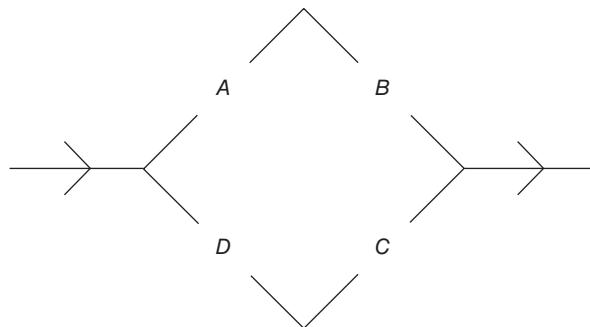
Figure 1.21 Contour plot for the surface in Figure 1.20.

EXERCISES 1.6

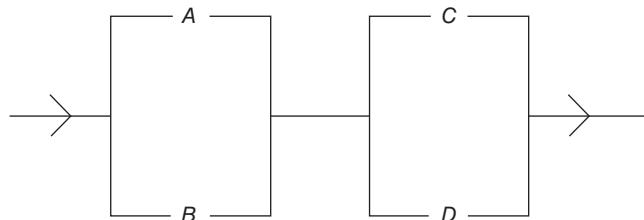
1. In the diagram below, let p_A , p_B , and p_C be the reliabilities of the individual switches. Determine the reliability of the system if
- at least one switch must function.
 - at least two switches must function.



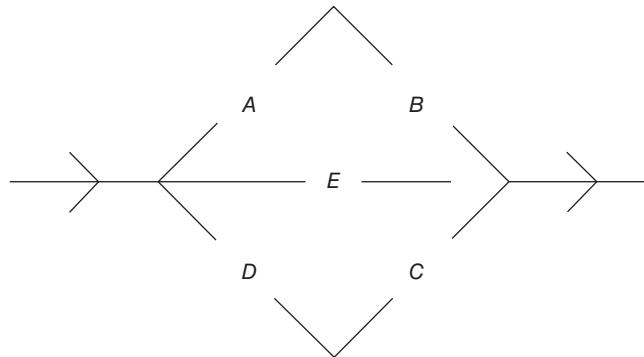
2. Determine the reliability of the system shown below if the reliability of any of the individual components is p .



3. Find the reliability of the system shown below if $p_A = p_B$ and $p_C = p_D$. Then show the surface giving the reliability of the system as a function of p_A and p_C and draw a contour plot of the surface.



4. Find the reliability of the system below if each component has reliability 0.92.



1.7 COUNTING TECHNIQUES

Occasionally, sample spaces are encountered for which the sample points are equally likely. If this is the case, and if the sample space S contains n points, then, since the total probability in the sample space is 1, each point has probability $1/n$. If we denote the mutually exclusive points in A by a_i , $i = 1, 2, 3, \dots, n$, then the probability of an event, A , is the sum of the probabilities of the sample points in A . That is,

$$P(A) = \sum_{a_i \in A} P(a_i) = \sum_{a_i \in A} \frac{1}{n} \text{ so}$$

$$P(A) = \frac{\text{Number of points in } A}{n} = \frac{\text{Number of points in } A}{\text{Number of points in } S}.$$

In order to consider problems leading to sample spaces with equally likely sample points, we pause to consider some techniques for counting sets of points. These techniques provide some challenging problems.

The reader is first cautioned here to beware of concluding that just because a sample space has n points that each point has probability $1/n$. For example, an airplane journey is either safely completed or not. One hopes these do not each have probability $1/2$!

The counting techniques considered here are based on two fundamental counting principles concerning mutually exclusive events A and B :

Principle 1: If events A and B can occur in n and m ways, respectively, then A and B can occur together in $n \cdot m$ ways.

Principle 2: If events A and B can occur in n and m ways, respectively, then A or B (but not both) can occur in $n + m$ ways.

Principle 1 is easily established since A can occur in n ways and then must be followed by each way in which B can occur. A tree diagram, shown in Figure 1.22, illustrates the result. Principle 2 simply uses the word “or” in an exclusive sense.

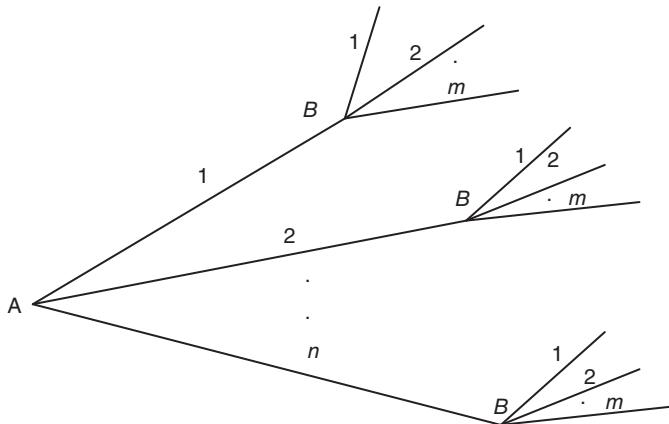


Figure 1.22 Tree diagram showing counting principle 1.

A linear arrangement of n distinct objects is called a *permutation*. For example, three distinct objects, say A , B , and C , can be arranged in six different ways: ABC , ACB , BAC , BCA , CAB , and CBA . So there are six permutations of three distinct objects. To count these permutations for n distinct objects, we use Principle 1. We have n choices for the object in the first position; that object chosen, we have $n - 1$ choices for the object in the second position. Principle 1 tells us that there are $n \cdot (n - 1)$ ways to fill the first two positions. Continuing, we have

$$n \cdot (n - 1) \cdot (n - 2) \cdots 3 \cdot 2 \cdot 1$$

ways to arrange all n of the items. We call this expression $n!$ and note, for example, that $3! = 3 \cdot 2 \cdot 1 = 6$, verifying the number of permutations of A , B , and C above.

The values of $n!$ increase very rapidly: $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$ and $10!$ is over 3 million. If we are interested in the number of permutations of even a small set, we must be prepared to deal with immense quantities. For example, the cards in a deck of 52 cards can be arranged in $52! = 80,658,175,170,943,878,571,660,636,856,403,766,975,289,505,440,883,277,824,000,000,000,000$ different ways. The reader may be surprised to find out how long it would take to enumerate these, even at a rate of 10,000 different permutations per second. This consideration may also persuade us that shuffling a deck so that each of these orders is equally likely is extremely unlikely.

A fact that is useful is that

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

If we wish to permute only r , say, of the n distinct objects, this can be done in

$$n \cdot (n - 1) \cdot (n - 2) \cdots (n - (r - 1)) \text{ ways.}$$

Multiplying and dividing by $(n - r)!$ shows that we can permute r of the n distinct objects in

$$\frac{n!}{(n - r)!} \text{ ways.}$$

So that this formula will work when $r = n$, we define $0! = 1$. If we wish to permute 5 cards chosen from a deck of 52, this can be done in

$$52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = 311,875,200 \text{ ways.}$$

We note, multiplying and dividing by $47!$, that this can also be written as $\frac{52!}{47!}$, a fact that will be useful later.

If a list of permutations is desired, then the reader is advised to do this using a computer algebra system. The $4! = 24$ permutations of the set $\{a, b, c, d\}$ is shown by a computer algebra system to be:

a	b	c	d	c	a	b	d
a	b	d	c	c	a	d	b
a	c	b	d	c	b	a	d
a	c	d	b	c	b	d	a
a	d	b	c	c	d	a	b
a	d	c	b	c	d	b	a
b	a	c	d	d	a	b	c
b	a	d	c	d	a	c	b
b	c	a	d	d	b	a	c
b	c	d	a	d	b	c	a
b	d	a	c	d	c	a	b
b	d	c	a	d	c	b	a

If we regard these permutations as being equally likely and if we want to find the probability that a particular letter, say b , occupies its normal place, we can count the points for which that is true and find that there are six of them. So

$$P(b \text{ is in second place}) = \frac{6}{24}.$$

What if the number of letters is large? An easy way to think about the problem is as follows: b is in its place, and if the set contains n distinct letters, we can arrange the remaining $(n - 1)$ letters in $(n - 1)!$ ways. Since the entire set can be permuted in $n!$ ways, the probability that b , or any of the other particular letters, occupies its own place is $\frac{(n-1)!}{n!} = \frac{1}{n}$.

This raises the question of other letters also occupying their own position. If we arrange the letters entirely at random, what is the probability that at least one of the letters is in its own place? The problem has been posed in the literature in many different ways one of which is this: n men enter a restaurant and each checks his hat; the hats become mixed up during the evening and are passed out at the end of the evening in an entirely random way. What is the probability that at least one man gets his own hat? Equivalently, if we assume some natural order for the cards in a deck, what, after thorough shuffling, is the probability that at least one of the cards is in its own position?

One way to solve the problem is to determine the number of *derangements* (where no object occupies its own place) of a set of objects. For the permutations of the set $\{1, 2, 3, 4\}$, we find the derangements are as follows:

2, 1, 4, 3

2, 3, 4, 1

2, 4, 1, 3

3, 1, 4, 2

3, 4, 1, 2

3, 4, 2, 1

4, 1, 2, 3

4, 3, 1, 2

4, 3, 2, 1

a total of 9 derangements in this case. It follows that the probability that at least one object occupies its own place is $1 - \frac{9}{24} = 15/24 = 0.625$. Surely this is an awkward way to handle larger sets, such as the deck of 52 cards. Surprisingly, we will find that the probability that at least one card occupies its own place after a thorough shuffling of the deck is very close to the probability above for four objects! We will explain this when we return to this problem later in this section.

For now, consider arranging a set of objects when the objects are not all distinct. If we permute the elements in the set $\{a, a, b, c\}$, we find there are 12 permutations:

$$\begin{aligned} &a, a, b, c \\ &a, a, c, b \\ &a, b, a, c \\ &a, b, c, a \\ &a, c, a, b \\ &a, c, b, a \\ &b, a, a, c \\ &b, a, c, a \\ &b, c, a, a \\ &c, a, a, b \\ &c, a, b, a \\ &c, b, a, a \end{aligned}$$

so the set of 24 permutations (where the objects were distinct) has been cut in half. We can arrive at this result by starting with the set $\{a, a, b, c\}$. Let R denote the number of distinct permutations. If we then tag the a 's with subscripts, say as a_1 and a_2 , then each of the permutations in the above list yields $2!$ permutations with the subscripted a 's. Hence, $2! \cdot R = 4!$ so $R = \frac{4!}{2!} = 12$. We could do exactly the same procedure with any set. Consider, for example, the set $\{a, a, b, b, b, c, c, c, c\}$. By subscripting the a 's, b 's, and c 's, respectively, and again letting R denote the number of distinct permutations, we conclude that

$$2! \cdot 3! \cdot 4! \cdot R = 9! \text{ so}$$

$$R = \frac{9!}{2! \cdot 3! \cdot 4!} = 1260.$$

The example is perfectly typical of the general situation: if the set has n_1 objects of one kind, n_2 of another, and so on until we have, say n_k of the k th kind where $\sum_{i=1}^k n_i = n$, then there are

$$\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$$

distinct permutations of the n objects.

Example 1.7.1

In how many distinct ways can 10 A's, 5 B's, and 2 C's be awarded to a class of 17 students?

Put the students in some order. Then each distinct permutation of the letters leads to a different assignment of the grades. So there are

$$\frac{17!}{10! \cdot 5! \cdot 2!} = 408,408$$

different ways to assign the grades.

We turn now to *combinations*, that is, the distinguishable sets or samples of objects that can be chosen from a set of n distinct objects, without regard for order. We denote these combinations of r objects chosen from n distinct objects by $\binom{n}{r}$, which we read " n choose r ." We have already seen that $\binom{n}{2} = \frac{n \cdot (n-1)}{2}$ in Example 1.5.3. Now suppose we have a set of objects and we want to choose a subset or sample of size 3. To be specific, suppose there are four items: a, b, c , and d . It is easy to write down the four combinations of size 3: a, b, c ; a, b, d ; a, c, d ; and b, c, d . However, if we were dealing with larger set, it might be very difficult to write down a complete list without a procedure in mind. As a suggestion, to create the samples of size 3, we could choose each of the samples of size 2 and then attach a third item. The resulting list is as follows:

- a, b, c
- a, b, d
- b, c, a
- b, c, d
- a, b, d
- b, c, a
- a, b, d
- b, c, a .

Since we have 2 choices for the third item, the resulting list contains $2 \cdot \binom{4}{2}$ items. But each of the *combinations* has occurred three times. Therefore,

$$2 \cdot \binom{4}{2} = 3 \cdot \binom{4}{3}, \text{ so}$$

$$\binom{4}{3} = \frac{2 \cdot \binom{4}{2}}{3} = \frac{2 \cdot 6}{3} = 4.$$

This would appear to be a difficult way to arrive at $\binom{4}{3}$. The reasoning here, however, can easily be extended and therein lies its advantage. Suppose we have a set of n distinct items and we wish to choose a sample of size r . If we choose all the possible samples of size $r - 1$ and then attach one of the $n - r + 1$ remaining items to each, the resulting list

has $(n - r + 1) \cdot \binom{n}{r-1}$ items. But this counts each of the $\binom{n}{r}$ combinations r times. So,

$$(n - r + 1) \cdot \binom{n}{r-1} = r \cdot \binom{n}{r}$$

$$\text{or } \binom{n}{r} = \frac{(n - r + 1) \cdot \binom{n}{r-1}}{r}. \quad (1.4)$$

This is a *recurrence formula* since it expresses some values of a function, here $\binom{n}{r}$, in terms of other values of the same function. If we have a starting place, we can calculate any value of the function we want. Here, since $\binom{n}{1} = n$, formula (1.4) shows that

$$\binom{n}{2} = \frac{n - 2 + 1}{2} \cdot \binom{n}{1} = \frac{n \cdot (n - 1)}{2} = \frac{n!}{2! \cdot (n - 2)!}$$

verifying our previous result. We continue to apply formula (1.4) to find

$$\binom{n}{3} = \frac{n - 3 + 1}{3} \cdot \binom{n}{2} = \frac{n \cdot (n - 1) \cdot (n - 2)}{3 \cdot 2} = \frac{n \cdot (n - 1) \cdot (n - 2) \cdot (n - 3)!}{3! \cdot (n - 3)!}$$

$$\binom{n}{3} = \frac{n!}{3! \cdot (n - 3)!}.$$

It can be concluded by an inductive proof that

$$\binom{n}{r} = \frac{n!}{r! \cdot (n - r)!}, \quad r = 0, 1, \dots, n$$

using the recurrence formula above.

If we have a set of n distinct objects and r are chosen, then $n - r$ objects must remain unchosen. Since each time the chosen set is altered, so is the unchosen set, it follows that

$$\binom{n}{r} = \binom{n}{n - r}.$$

The quantities $\binom{n}{r}$ are often called *binomial coefficients* since they occur in the binomial expansion:

Binomial Theorem:

$$(a + b)^n = \sum_{r=0}^n \binom{n}{r} \cdot a^{n-r} \cdot b^r = \sum_{r=0}^n \binom{n}{r} \cdot a^r \cdot b^{n-r}. \quad (1.5)$$

For example,

$$(a + b)^5 = \binom{5}{0} a^5 + \binom{5}{1} a^4 b + \binom{5}{2} a^3 b^2 + \binom{5}{3} a^2 b^3 + \binom{5}{4} a b^4 + \binom{5}{5} b^5$$

$$= a^5 + 5a^4 b + 10a^3 b^2 + 10a^2 b^3 + 5a b^4 + b^5.$$

Many interesting identities are known concerning the binomial coefficients. If $a = 1$ and $b = 1$ are substituted in formula (1.5), the result is

$$(1+1)^n = 2^n = \sum_{r=0}^n \binom{n}{r} = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n}.$$

Each side of this result may be recognized as the number of possible subsets (including the null set) that can be chosen from a set of n distinct items.

If we differentiate (1.5) with respect to a and then let $a = 1$ and $b = 1$, the result is

$$n \cdot 2^n = \sum_{r=0}^n \binom{n}{r} \cdot r = 0 \cdot \binom{n}{0} + 1 \cdot \binom{n}{1} + 2 \cdot \binom{n}{2} + \cdots + n \cdot \binom{n}{n}.$$

We show one more fact concerning the binomial coefficients. Suppose we want to choose a committee of size r chosen from a group of n people, one of whom is Sam. Sam is a member of $\binom{n-1}{r-1}$ committees and he is not a member of $\binom{n-1}{r}$ committees, so, since we have exhausted the possibilities,

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}.$$

This is also often known as *Pascal's identity* since it occurs in Pascal's triangle of binomial coefficients.

It is also necessary for us, although this may seem unnatural to the reader, to ascribe some meaning to a symbol such as $\binom{-7}{3}$. Clearly, we cannot interpret this as the choice of 3 objects from -7 objects! The following definition, while including our previous interpretation of $\binom{n}{r}$, allows us to extend its meaning as well.

Definition: $\binom{n}{r} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}{r!}$ provided that r is a nonnegative integer.

Using the above definition, we have that $\binom{-7}{3} = \frac{(-7) \cdot (-8) \cdot (-9)}{3!} = -84$. We will need facts such as this in subsequent chapters.

Using this definition, the binomial theorem can also be used with negative exponents. For example,

$$\begin{aligned} (a+b)^{-5} &= a^{-5} + \binom{-5}{1} a^{-6}b + \binom{-5}{2} a^{-7}b^2 + \binom{-5}{3} a^{-8}b^3 + \cdots \text{ or} \\ (a+b)^{-5} &= a^{-5} - \binom{5}{1} a^{-6}b + \binom{6}{2} a^{-7}b^2 - \binom{7}{3} a^{-8}b^3 + \cdots \end{aligned}$$

We now use some of the results found here in some examples.

Example 1.7.2

A box of manufactured items contains 8 items that are good and 3 that are not usable. What is the probability that a sample of 5 items contains exactly 1 unusable item?

Suppose that each of the samples has probability $\frac{1}{\binom{11}{5}}$. There are $\binom{8}{4}$ ways to choose the 4 good items and $\binom{3}{1}$ ways to choose the unusable item. The multiplication principle then gives $\binom{3}{1} \cdot \binom{8}{4}$ ways to choose exactly 1 unusable item. So the probability we seek is

$$\frac{\binom{3}{1} \cdot \binom{8}{4}}{\binom{11}{5}} = \frac{5}{11}.$$

Finally, in this chapter, we consider the *general addition law for n events*, having established the addition law for two and for three events. So we seek to prove

Theorem 4: $P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum P(A_i) - \sum P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n)$, where the sums are over all the *distinct* items in the summand, that is, where $i > j > k > \dots$

Proof We again use the principle of inclusion and exclusion. Consider a point in $A_1 \cup A_2 \cup \dots \cup A_n$ which is in exactly k of the events A_i . It will be convenient to renumber the A_i 's if necessary so that the point is in the first k of these events. We will now show that the right-hand side of Theorem 4 counts this point exactly once, showing the theorem to be correct.

The point is counted on the right-hand side of Theorem 4

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \dots \pm \binom{k}{k} \text{ times.}$$

But the binomial expansion of $0 = [1 + (-1)]^k = \sum_{i=0}^k \binom{k}{i} \cdot (-1)^i$ shows that

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \dots \pm \binom{k}{k} = \binom{k}{0} = 1,$$

establishing the result.

Example 1.7.3

We return to the matching problem stated earlier in this section: If n integers are randomly arranged in a row, what is the probability that at least one of them occupies its own place? The general addition law can be used to provide the solution.

Let A_i denote the event, “number i is in the i th place.” We seek $P(A_1 \cup A_2 \cup \dots \cup A_n)$.

Here $P(A_i) = \frac{(n-1)!}{n!}$, since, after i is put in its own place, there are $(n-1)!$ ways to arrange the remaining numbers; $P(A_i \cap A_j) = \frac{(n-2)!}{n!}$, since if i and j occupy their own places we can permute the remaining $n-2$ objects in $(n-2)!$ ways; and, in general, $P(A_1 \cap A_2 \cap \dots \cap A_k) = \frac{(n-k)!}{n!}$. Now we note that there are $\binom{n}{1}$ choices for an individual number i ; there

are $\binom{n}{2}$ choices for pairs of numbers i and j ; and, in general, there are $\binom{n}{k}$ choices for k of the numbers. So, applying Theorem 4,

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \binom{n}{1} \cdot \frac{(n-1)!}{n!} - \binom{n}{2} \cdot \frac{(n-2)!}{n!} \\ &\quad + \binom{n}{3} \cdot \frac{(n-3)!}{n!} - \dots \pm \binom{n}{n} \cdot \frac{(n-n)!}{n!}. \end{aligned}$$

This simplifies to $P(A_1 \cup A_2 \cup \dots \cup A_n) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots \pm \frac{1}{n!}$. A table of values of this expression is shown below.

n	P
1	1.000000
2	0.500000
3	0.666667
4	0.625000
5	0.633333
6	0.631944
7	0.632143
8	0.632118
9	0.632121

To six decimal places, the probability that at least one number is in its natural position remains at 0.632121 for $n \geq 9$. An explanation for this comes from a series expansion for e^x :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

So

$$\begin{aligned} e^{-1} &= 1 - 1 + \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!} + \frac{(-1)^4}{4!} + \dots \text{ or} \\ e^{-1} &= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \dots \end{aligned}$$

So we see that $\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots \pm \frac{1}{n!}$ approaches $1 - \frac{1}{e} = 0.632120559 \dots$ This is our first, but certainly not our last, encounter with e in a probability problem. This also explains why we remarked that the probability at least one card in a shuffled deck of 52 cards was in its natural position differed little from that for a deck consisting of only 9 cards.

We turn now to some examples using the results established in this section.

Example 1.7.4

Five red and four blue marbles are arranged in a row. What is the probability that both the end marbles are blue?

A basic decision in the solution of the problem concerns the type of sample space to be used. Clearly, the problem involves order, but should we consider the marbles to be distinct or not?

Initially, consider the marbles to be alike, except for color of course. There are $\frac{9!}{5!4!} = 126$ possible orderings of the marbles and we consider each of these to be equally likely. Since the blue marbles are indistinct from each other, and since our only choice here is the arrangement of the 7 marbles in the middle, it follows that there are $\frac{7!}{5!2!} = 21$ arrangements with blue marbles at the ends. The probability we seek is then $\frac{21}{126} = \frac{1}{6}$.

Now if we consider each of the marbles to be distinct, there are $9!$ possible arrangements. Of these, we have $\binom{4}{2} \cdot 2! = 12$ ways to arrange the blue marbles at the ends and $7!$ ways to arrange the marbles in the middle. This produces a probability of $\frac{7! \cdot 12}{9!} = \frac{1}{6}$.

The two methods must produce the same result, but the reader may find one method easier to use than another. In any event, it is crucial that the sample space be established as a first step in the solution of the problem and that the events of interest be dealt with consistently for this sample space.

The reader may enjoy showing that, if we have n marbles, r of which are red and b of which are blue, then the probability both ends are blue in a random arrangement of the marbles is given by the product

$$\left(1 - \frac{r}{n}\right) \cdot \left(1 - \frac{r}{n-1}\right).$$

This answer may indicate yet another way to solve the problem, namely this: the probability the first marble is blue is $\left(\frac{n-r}{n}\right)$. Given that the first end is blue, the conditional probability the other end is also blue is $\left(\frac{n-r-1}{n-1}\right)$. Often probability problems involving counting techniques can be solved in a variety of ways.

Example 1.7.5

Ten race cars, numbered from 1 to 10, are running around a track. An observer sees three cars go by. If the cars appear in random order, what is the probability that the largest number seen is 6?

The choice of the sample space here is natural: consider all the $\binom{10}{3}$ samples of three cars that could be observed. If the largest is to be 6, then 6 must be in the sample, together with two cars chosen from the first 5, so the probability of the event “Maximum = 6” is

$$P(\text{Maximum} = 6) = \frac{\binom{1}{1} \cdot \binom{5}{2}}{\binom{10}{3}} = \frac{1}{12}.$$

It is also interesting now to look at the *median* or the number in the middle when the three observed numbers are arranged in order. What is the probability that the median of the group of three is 6?

For the median to be 6, 6 must be chosen and we must choose exactly one number from the set {1, 2, 3, 4, 5} and exactly one number from {7, 8, 9, 10}. Then

$$P(\text{Median} = 6) = \frac{\binom{1}{1} \cdot \binom{5}{1} \cdot \binom{4}{1}}{\binom{10}{3}} = \frac{1}{6}.$$

This can be generalized to

$$\begin{aligned} P(\text{Median} = k) &= \frac{\binom{1}{1} \cdot \binom{k-1}{1} \cdot \binom{10-k}{1}}{\binom{10}{3}} \\ &= \frac{(k-1)(10-k)}{120}, \quad k = 2, 3, \dots, 9. \end{aligned}$$

Figure 1.23 shows a graph of $P(\text{Median} = k)$ for $k = 2, 3, \dots, 9$. It reveals a symmetry in the function around $k = 5.5$.

The problem is easily generalized with a result that may be surprising. Suppose there are 100 cars and we observe a sample of 9 of them. The median of the sample must be at least 5 and can be at most 96. The probability the median is k then is

$$P(\text{Median} = k) = \frac{\binom{1}{1} \binom{k-1}{4} \binom{100-k}{4}}{\binom{100}{9}}, \quad k = 5, 6, \dots, 96.$$

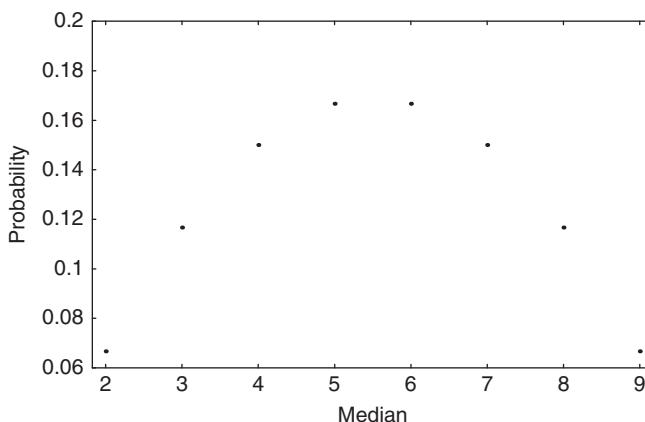


Figure 1.23 $P(\text{Median} = k)$ for a sample of size 3 chosen from 10 cars.

A graph of this function (an eighth degree polynomial in k) is shown in Figure 1.24.

The graph here shows a “bell shape” that, as we will see, is very common in probability problems. The curve is very close to what we will call a *normal curve*. Larger values for the number of cars involved will, surprisingly, not change the approximate normal shape of the curve! An approximation for the actual curve involved here can be found when we study the normal curve thoroughly in Chapter 3.

Example 1.7.6

We can use the result of Example 1.7.3 to count the number of *derangements* of a set of n objects. That is, we want to count the number of permutations in which no object occupies

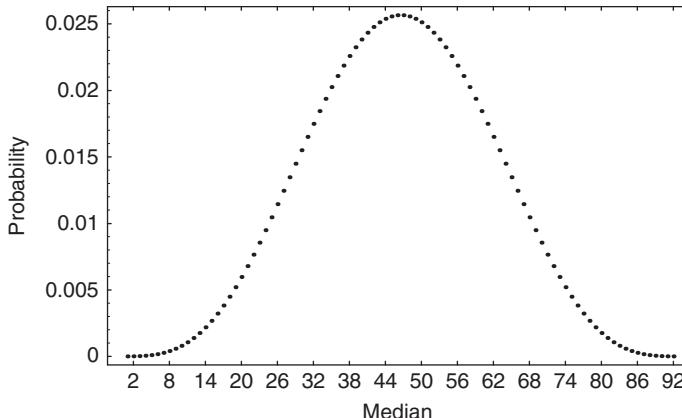


Figure 1.24 $P(\text{Median} = k)$ for a sample of 9 chosen from 100 cars.

its own place. Example 1.7.3 shows that the number of permutations of n distinct objects in which at least one object occupies its own place is

$$n! \left(\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!} \right).$$

It follows that the number of derangements of n distinct objects is

$$n! - n! \left(\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!} \right) = n! \left(\frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{n!} \right). \quad (1.6)$$

Using this formula we find that if $n = 2$, there is 1 derangement; if $n = 3$, there are 2 derangements, and if $n = 4$, there are 9 derangements.

Formula (1.6) also suggests that the number of derangements of n distinct objects is approximately $n! \cdot e^{-1}$ (see the series expansion for e^{-1} in Example 1.7.5). The following table compares the results of formula (1.6) and the approximation:

n	Number of derangements	$n! \cdot e^{-1}$
2	1	0.7358
3	2	2.207
4	9	8.829
5	44	44.146
6	265	264.83
7	1854	1854.11

We see that in every case, the number of derangements is given by $\lfloor n! \cdot e^{-1} + 0.5 \rfloor$ where the symbols indicate the greatest integer function.

Example 1.7.7 (Ken-Ken Puzzles)

The *New York Times* as well as many other newspapers publish a Ken-Ken puzzle daily. The problem consists of a square with 4 rows and 4 columns. The problem is to insert

each of the digits 1, 2, 3, and 4 into each row and each column so that each digit appears exactly once in each row and each column. The reader is given arithmetic clues for some squares. For example, 5+ may indicate the proper entries are 4 1 (but there are many other possibilities. Here is an example of a solved puzzle (without the clues):

$$\begin{array}{cccc} 3 & 4 & 2 & 1 \\ 2 & 1 & 3 & 4 \\ 1 & 3 & 4 & 2 \\ 4 & 2 & 1 & 3 \end{array}$$

Clearly, each row (and hence each column) is a derangement of the integers 1 through 4, but each row (and column) must be a derangement of each of the previous rows (or columns). How many Ken-Ken puzzles are there?

Since we will be permuting the rows and columns later, we might just as well start with the row 1 2 3 4. For the second row, we must select one of the 9 derangements of the integers 1, 2, 3, 4, as shown in Example 1.7.1. We will choose 2 4 1 3, so now we have $\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{array}$. By examining the 9 derangements again, we find only two choices for the third row: 3 1 4 2 or 4 3 2 1. When one of these is chosen, there is only one choice for the fourth row—the derangement that was not selected for the third row. Selecting the first choice for the third row, we have

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 3 & 1 & 4 & 2 \\ 4 & 3 & 2 & 1 \end{array}$$

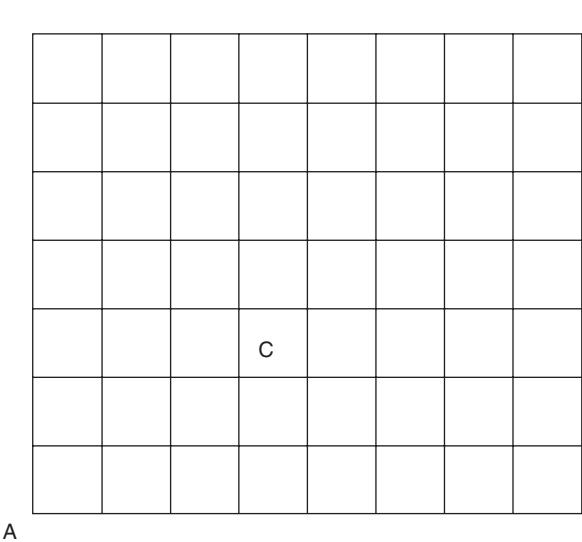
Now the rows and the columns may be permuted in $4! * 4!$ ways, so the total number of Ken-Ken puzzles with 4 rows and 4 columns is $9 * 2 * 4! * 4! = 10368$.

EXERCISES 1.7

1. The integers 1, 2, 3, ..., 9 are arranged in a row, resulting in a nine-digit integer. What is the probability that
 - (a) the integer resulting is even?
 - (b) the integer resulting is divisible by 5?
 - (c) the digits 6 and 4 are next to each other?
2. License plates in Indiana consist of a number from 1 to 99 (indicating the county of registration), a letter of the alphabet, and finally an integer from 1 to 9999. How many cars may be licensed in Indiana?
3. Prove that at least two people in Colorado Springs, Colorado, have the same three initials.
4. In a small school, 5 centers, 8 guards, and 6 forwards try out for the basketball team.
 - (a) How many five-member teams can be formed from these players? (Assume a team has two guards, two forwards, and one center.)
 - (b) Intercollegiate regulations require that no more than 8 players can be listed for the team roster. How many rosters can be formed consisting of exactly 8 players?

5. A restaurant offers 5 appetizers, 7 main courses, and 8 desserts. How many meals can be ordered?
 - (a) assuming all three courses are ordered?
 - (b) not assuming all three courses are necessarily ordered?
6. A club of 56 people has 40 men and 16 women. What is the probability the board of directors, consisting of 8 members, contains no women?
7. In a controlled experiment, 12 patients are to be randomly assigned to each of three different drug regimens. In how many ways can this be done if each drug is to be tested on 4 patients?
8. In the game Keno, the casino draws 20 balls from a set of balls numbered from 1 to 80. A player must choose 10 numbers in advance of this drawing. What is the probability the player has exactly five of the 20 numbers drawn?
9. A lot of 10 refrigerators contains 3 which are defective. The refrigerators are randomly chosen and shipped to customers. What is the probability that by the seventh shipment, none of the defective refrigerators remain?
10. In how many different ways can the letters in the word “repetition” be arranged?
11. In a famous correspondence in the very early history of probability, the Chevalier de Mere wrote to the mathematician Blaise Pascal and asked the following question, “Which is more likely—at least one six in four rolls of a fair die or at least one sum of 12 in 24 rolls of a pair of dice?”
 - (a) Show that the two questions above have nearly equal answers. Which is more likely?
 - (b) A generalization of the Pascal–de Mere problem is: what is the probability that the sum $6n$ occurs at least once in $4 \cdot 6^{n-1}$ rolls of n fair dice? Show that the answer is very nearly $1/2$ for $n \leq 5$.
 - (c) Show that in part (b) the probability approaches $1 - e^{-2/3}$ as $n \rightarrow \infty$.
12. A box contains 8 red and 5 yellow marbles from which a sample of 3 is drawn.
 - (a) Find the probability that the sample contains no yellow marbles if
 - (1) the sampling is done without replacement; and,
 - (2) if the sampling is done with replacement.
 - (b) Now suppose the box contains 24 red and 15 yellow marbles (so that the ratio of reds to yellows is the same as in part (a)). Calculate the answers to part (a). What do you expect to happen as the number of marbles in the box increases but the ratio of reds to yellows remains the same?
13. (a) From a group of 20 people, two samples of size 3 are chosen, the first sample being replaced before the second sample is chosen. What is the probability the samples have at least one person in common?
 - (b) Show that two bridge hands, the first being replaced before the second is drawn, are virtually certain to contain at least one card in common.
14. A shipment of 20 components will be accepted by a buyer if a random sample of 3 (chosen without replacement) contains no defectives. What is the probability the shipment will be rejected if actually 2 of the components are defective?
15. A deck of cards is shuffled and the cards turned up one at a time. What is the probability that all the aces will appear before any of the 10’s?

16. In how many distinguishable ways can 6 A's, 4 B's, and 8 C's be assigned as grades to 18 students?
17. What is the probability a poker hand (5 cards drawn from a deck of 52 cards) has exactly 2 aces?
18. In how many ways can 6 students be seated in 10 chairs?
19. Ten children are to be grouped into two clubs, say the Lions and the Tigers, with five children in each club. Each club is then to elect a president and a secretary. In how many ways can this be done?
20. A small pond contains 50 fish, 10 of which have been tagged. If a catch of 7 fish is made, in how many ways can the catch contain exactly 2 tagged fish?
21. From a fleet of 12 limousines, 6 are to go to hotel I, 4 to hotel II, and the remainder to hotel III. In how many different ways can this be done?
22. The grid shows a region of city blocks defined by 7 streets running North–South and 8 streets running East–West. Joe will walk from corner A to corner B. At each corner between A and B, Joe will choose to walk either North or East.



- (a) How many possible routes are there?
- (b) Assuming that each route is equally likely, find the probability that Joe will pass through intersection C.
23. Suppose that N people are arranged in a line. What is the probability that two particular people, say A and B, are not next to each other?
24. The Hawaiian language has only 12 letters: the vowels a, e, i, o, and u and the consonants h, k, l, m, n, p, and w.
 - (a) How many possible three-letter Hawaiian “words” are there? (Some of these may be nonsense words.)
 - (b) How many three-letter “words” have no repeated letter?
 - (c) What is the probability a randomly selected three-letter “word” begins with a consonant and ends with 2 different vowels?

- (d) What is the probability that a randomly selected three-letter “word” contains all vowels?
25. How many partial derivatives of order 4 are there for a function of 4 variables?
26. A set of 15 marbles contains 4 red and 11 green marbles. They are selected, one at a time, without replacement. In how many ways can the last red marble be drawn on the seventh selection?
27. A true–false test has four questions. A student is not prepared for the test and so must guess the answer to each question.
- (a) What is the probability the student answers at least half of the questions correctly?
- (b) Now suppose, in a sudden flash of insight, he knows the answer to question 2 is “false.” What is the probability he answers at least half of the questions correctly?
28. What is the probability of being dealt a bridge hand (13 cards selected from a deck of 52 cards) that does not contain a heart?
29. Explain why the number of derangements of n distinct objects is given by $\lfloor n! \cdot e^{-1} + 0.5 \rfloor$. Explain why $n! \cdot e^{-1}$ sometimes underestimates the number of derangements and sometimes overestimates the number of derangements. $\lfloor x \rfloor$ denotes the greatest integer in x .
30. Find the number of Ken–Ken puzzles if the grid is 5×5 for the integers 1, 2, 3, 4, 5.

CHAPTER REVIEW

In dealing with an experiment or situation involving random or chance elements, it is reasonable to begin an analysis of the situation by asking the question, “What *can* happen?” An enumeration of all the possibilities is called a *sample space*. Generally, situations admit of more than one sample space; the appropriate one chosen is usually governed by the probabilities that one wants to compute. Several examples of sample spaces are given in this chapter, each of them *discrete*, that is, either the sample space has a finite number of points or a countably infinite number of points.

Tossing two dice yields a sample space with a finite number of points; observing births until a girl is born gives a sample space with an infinite (but countable) number of points. In the next chapter, we will encounter *continuous* sample spaces that are characterized by a noncountably infinite number of points.

Assessing the *long-range relative frequency*, or *probability*, of any of the points or sets of points (which we refer to as *events*) is the primary goal of this chapter. We use the set symbols \cup for the union of two events and \cap for the intersection of two events. We begin with three assumptions or axioms concerning sample spaces:

- (1) $P(A) \geq 0$, where A is an event;
- (2) $P(S) = 1$, where S is the entire sample space; and,
- (3) $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$ if A and B are disjoint, or mutually exclusive, they have no sample points in common.

From these assumptions, we derived several theorems concerning probability, among them:

$$(1) P(A) = \sum_{a_i \in A} P(a_i), \text{ where the } a_i \text{ are distinct point in } S.$$

(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (the addition law for two events).

(3) $P(\bar{A}) = 1 - P(A)$

We showed the Law of Total Probability.

Theorem (Law of Total Probability): If the sample space $S = A_1 \cup A_2 \cup \dots \cup A_n$ where A_i and A_j have no sample points in common if $i \neq j$, then, if B is an event,

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n)$$

We then turned our attention to problems of *conditional probability* where we sought the probability of some event, say A , on the condition that some other event, say B , has occurred. We showed that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})}.$$

This can be generalized using the Law of Total Probability as follows:

Theorem (Bayes' Theorem): If $S = A_1 \cup A_2 \cup \dots \cup A_n$ where A_i and A_j have no sample points in common for $i \neq j$, then, if B is an event,

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\ P(A_i|B) &= \frac{P(A_i) \cdot P(A_i|B)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n)} \\ P(A_i|B) &= \frac{P(A_i) \cdot P(A_i|B)}{\sum_{i=1}^n P(A_i) \cdot P(A_i|B)} \end{aligned}$$

Bayes' theorem has a simple geometric interpretation. The chapter contains many examples of this.

We defined the *independence* of two events, A and B as follows:

$$A \text{ and } B \text{ are independent if } P(A \cap B) = P(A) \cdot P(B)$$

We then applied the results of this chapter to some specific probability problems, such as the well-known birthday problem and a geometric problem involving the sides and diagonals of a polygonal figure.

Finally, we considered some very special counting techniques which are useful, it is to be emphasized, only if the points in the sample space are equally likely. If that is so, then the probability of an event, say A , is

$$P(A) = \frac{\text{Number of points in } A}{\text{Number of points in } S}.$$

If order is important, then all the *permutations* of objects may well comprise the sample space. We showed that there are $n! = n \cdot (n - 1) \cdot (n - 2) \cdots 3 \cdot 2 \cdot 1$ permutations of n distinct objects.

If order is not important, then the sample space may well comprise various *combinations* of items. We showed that there are

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

samples of size r that can be selected from n distinct objects and applied this formula to several examples. A large number of identities are known concerning these combinations, or *binomial coefficients*, among them:

$$(1) \sum_{r=0}^n \binom{n}{r} = 2^n.$$

$$(2) \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}.$$

One very important result from this section is the *general addition law*:

Theorem:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum P(A_i) - \sum P(A_i \cap A_j) \\ + \sum P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n),$$

where the summations are over $i > j > k > \dots$

PROBLEMS FOR REVIEW

Exercises 1.1 # 1, 2, 5, 7, 9, 11

Exercises 1.3 # 1, 2, 6, 7, 9, 13

Exercises 1.4 # 1, 2, 3, 6, 10, 15, 16, 18, 19, 21, 24

Exercises 1.5 # 2, 3, 6, 7

Exercises 1.6 # 1, 3

Exercises 1.7 # 1, 6, 8, 10, 12, 13, 16, 17, 20, 23, 28

SUPPLEMENTARY EXERCISES FOR CHAPTER 1

1. A hat contains slips of paper on which each of the integers 1, 2, ..., 20 is written. A sample of size 6 is drawn (without replacement) and the sample values, x_i , put in order so that $x_1 < x_2 < \dots < x_6$. Find the probability that $x_3 = 12$.
2. Show that $(n-k) \binom{n}{n-k} = (k+1) \binom{n}{k+1}$.
3. Suppose that events A , B , and C are independent with $P(A) = 1/3$, $P(B) = 1/4$, and $P(A \cup B \cup C) = 3/4$. Find $P(C)$.
4. Events A and B are such that $P(A \cup B) = 0.8$ and $P(A) = 0.2$. For what value of $P(B)$ are
 - (a) A and B independent?
 - (b) A and B mutually exclusive?
5. Events A , B , and C in a sample space have $P(A) = 0.2$, $P(B) = 0.4$, and $P(A \cup B \cup C) = 0.9$. Find $P(C)$ if A and B are mutually exclusive, A and C are independent, and B and C are independent.

6. How many distinguishable arrangements of the letters in **PROBABILITY** are there?
7. How many people must be in a group so that the probability that at least two were born on the same day of the week is at least $1/2$?
8. A and B are special dice. The faces on die A are 2, 2, 5, 5, 5 and the faces on die B are 3, 3, 3, 6, 6, 6. The two dice are rolled. What is the probability that the number showing on die B is greater than the number showing on die A?
9. A committee of 5 is chosen from a group of 8 men and 4 women. What is the probability the group contains a majority of women?
10. A college senior finds he needs one more course for graduation and finds only courses in Mathematics, Chemistry, and Computer Science available. On the basis of interest, he assigns probabilities of 0.1, 0.6 and 0.3, respectively, to the events of choosing each of these. After considering his past performance, his advisor estimates his probabilities of passing these courses as 0.8, 0.7, and 0.6, respectively, regarding the passing of courses as independent events.
 - (a) What is the probability he passes the course if he chooses a course at random?
 - (b) Later we find that the student graduated. What is the probability he took Chemistry?
11. A number, X , is chosen at random from the set $\{10, 11, 12, \dots, 99\}$.
 - (a) Find the probability that the 10's digit in X is less than the units digit.
 - (b) Find the probability that X is at least 50.
 - (c) Find the probability that the 10's digit in X is the square of the units digit.
12. If the integers 1, 2, 3, and 4 are randomly permuted, what is the probability that 4 is to the left of 2?
13. In a sample space, events A and B are such that $P(A) = P(B)$, $P(\overline{A} \cap \overline{B}) = P(A \cap B) = 1/6$. Find
 - (a) $P(A)$.
 - (b) $P(\overline{A} \cup \overline{B})$.
 - (c) $P(\text{Exactly one of the events } A \text{ or } B)$.
14. A fair coin is tossed four times. Let A be the event “2nd toss is heads,” B be the event “Exactly 3 heads,” and C be the event “4th toss is tails if the 2nd toss is heads.” Are A , B , and C independent?
15. An instructor has decided to grade each of his students A , B , or C . He wants the probability a student receives a grade of B or better to be 0.7 and the probability a student receives at most a grade of B to be 0.8. Is this possible? If so, what proportions of each letter grade must be assigned?
16. How many bridge hands are there containing 3 hearts, 4 clubs, and 6 spades?
17. A day's production of 100 fuses is inspected by a quality control inspector who tests 10 fuses at random, sampling without replacement. If he finds 2 or fewer defective fuses, he accepts the entire lot of 100 fuses. What is the probability the lot is accepted if it actually contains 20 defective fuses?
18. Suppose that A and B are events for which $P(A) = a$, $P(B) = b$ and $P(A \cap B) = c$. Express each of the following in terms of a , b , and c .
 - (a) $P(\overline{A} \cup \overline{B})$
 - (b) $P(\overline{A} \cap B)$

- (c) $P(\overline{A} \cup B)$
 (d) $P(\overline{A} \cap \overline{B})$
 (e) $P(\text{exactly one of } A \text{ or } B \text{ occurs}).$
19. An elevator starts with 10 people on the first floor of an eight-story building and stops at each floor.
- (a) In how many ways can all the people get off the elevator?
 (b) How many ways are there for everyone to get off if no one gets off on some two specific floors?
 (c) In how many ways are there for everyone to get off if at least one person gets off at each floor?
20. A manufacturer of calculators buys integrated circuits from suppliers A , B , and C . Fifty percent of the circuits come from A , 30% from B , and 20% from C . One percent of the circuits supplied by A have been defective in the past, 3% of B 's have been defective, and 4% of C 's have been defective. A circuit is selected at random and found to be defective. What is the probability it was manufactured by B ?
21. Suppose that E and T are independent events with $P(E) = P(T)$ and $P(E \cup T) = 1/2$. What is $P(E)$?
22. A quality control inspector draws parts one at a time and without replacement from a set containing 5 defective and 10 good parts. What is the probability the third defective is found on the eighth drawing?
23. If A , B , and C are independent events, show that the events A and $B \cup C$ are independent.
24. Bean seeds from supplier A have an 85% germination rate and those from supplier B have a 75% germination rate. A seed company purchases 40% of their bean seeds from supplier A and the remaining 60% from supplier B and mixes these together. If a seed germinates, what is the probability it came from supplier A ?
25. An experiment consists of choosing two numbers without replacement from the set {1, 2, 3, 4, 5, 6} with the restriction that the second number chosen must be greater than the first.
- (a) Describe the sample space.
 (b) What is the probability the second number is even?
 (c) What is the probability the sum of the two numbers is at least 5?
26. What is the probability a poker hand contains exactly one pair?
27. A box contains 6 good and 8 defective light bulbs. The bulbs are drawn out one at a time, without replacement, and tested. What is the probability that the fifth good item is found on the ninth test?
28. An individual tried by a three-judge panel is declared guilty if at least two judges cast votes of guilty. Suppose that when the defendant is, in fact, guilty, each judge will independently vote guilty with probability 0.7 but, if the defendant is, in fact, innocent, each judge will independently vote guilty with probability 0.2. Assume that 70% of the defendants are actually guilty. If a defendant is judged guilty by the panel of judges, what is the probability he is actually innocent?
29. What is the probability a bridge hand is missing cards in at least one suit?
30. Suppose 0.1% of the population is infected with a certain disease. On a medical test for the disease, 98% of those infected give a positive result while 1% of those not infected

give a positive result. If a randomly chosen person is tested and gives a positive result, what is the probability the person has the disease?

31. A committee of 50 politicians is to be chosen from the 100 US Senators (2 are from each state). If the selection is done at random, what is the probability that each state will be represented?
32. In a roll of a pair of dice (one red and one green), let A be the event “red die shows 3, 4, or 5,” B the event “green die shows a 1 or a 2,” and C the event “dice total 7.” Show that A , B , and C are independent.
33. An oil wildcatter thinks there is a 50-50 chance that oil is on the property he purchased. He has a test for oil that is 80% reliable: that is, if there is oil, it indicates this with probability 0.80 and if there is no oil, it indicates that with probability 0.80. The test indicates oil on the property. What is the probability there really is oil on the property?
34. Given: A and B are events with $P(A) = 0.3$, $P(B) = 0.7$ and $P(A \cup B) = 0.9$. Find
 - (a) $P(A \cap B)$
 - (b) $P(B|\bar{A})$
35. Two good transistors become mixed up with three defective transistors. A person is assigned to sampling the mixture by drawing out three items without replacement. However, the instructions are not followed and the first item is replaced, but the second and third items are not replaced.
 - (a) What is the probability the sample contains exactly two items that test as good?
 - (b) What is the probability the two items finally drawn are both good transistors?
36. How many lines are determined by 8 points, no three of which are collinear?
37. Show that if A and B are independent, then \bar{A} and \bar{B} are independent.
38. How many tosses of a fair coin are needed so that the probability of at least one head is at least 0.99?
39. A lot of 24 tubes contains 13 defective ones. The lot is randomly divided into two equal groups, and each group is placed in a box.
 - (a) What is the probability that one box contains only defective tubes?
 - (b) Suppose the tubes were divided so that one box contains only defective tubes. A box is chosen at random and one tube is chosen from the chosen box and is found to be defective. What is the probability a second tube chosen from the same box is also defective?
40. A machine is composed of two components, A and B , which function (or fail) independently. The machine works only if both components work. It is known that component A is 98% reliable and the machine is 95% reliable. How reliable is component B ?
41. Suppose A and B are events. Explain why $P(\text{exactly one of events } A, B \text{ occurs}) = P(A) + P(B) - 2P(A \cap B)$.
42. A box contains 8 red, 3 white, and 9 blue balls. Three balls are to be drawn, without replacement. What is the probability that more blues than whites are drawn?
43. A marksman, whose probability of hitting a moving target is 0.6, fires three shots. Suppose the shots are independent.
 - (a) What is the probability the target is hit?
 - (b) How many shots must be fired to make the probability at least 0.99 that the target will be hit?

44. A box contains 6 green and 11 yellow balls. Three are chosen at random. The first and third balls are yellow. Which method of sampling—with replacement or without replacement—gives the higher probability of this event?
45. A box contains slips of paper numbered from 1 to m . One slip is drawn from the box; if it is 1, it is kept; otherwise, it is returned to the box. A second slip is drawn from the box. What is the probability the second slip is numbered 2?
46. Three integers are selected at random from the set {1, 2, ..., 10}. What is the probability the largest of these is 5?
47. A pair of dice is rolled until a 5 or a 7 appears. What is the probability a 5 occurs first?
48. The probability is 1 that a fisherman will say he had a good day when, in fact, he did, but the probability is only 0.6 that he will say he had a good day when, in fact, he did not. Only 1/4 of his fishing days are actually good days. What is the probability he had a good day if he says he had a good day?
49. An inexperienced employee mistakenly samples n items from a lot of N items, with replacement. What is the probability the sample contains at least one duplicate?
50. A roulette wheel has 38 slots—18 red, 18 black, and 2 green (the house wins on green). Suppose the spins of the wheel are independent and that the wheel is fair. The wheel is spun twice and we know that at least one spin is green. What is the probability that both spins are green?
51. A “rook” deck of cards consists of four suits of cards: red, green, black, and yellow, each suit having 14 cards. In addition, the deck has an uncolored “rook” card. A hand contains 14 cards.
- How many different hands are possible?
 - How many hands have the rook card?
 - How many hands contain only two colors with equal numbers of cards of each color?
 - How many hands have at most three colors and no rook card?
52. Find the probability a poker hand contains 3 of a kind (exactly 3 cards of one face value and 2 cards of different face values).
53. A box contains tags numbered 1, 2, ..., n . Two tags are chosen without replacement. What is the probability they are consecutive integers?
54. In how many different ways can n people be seated around a circular table?
55. A production lot has 100 units of which 25 are known to be defective. A random sample of 4 units is chosen without replacement. What is the probability that the sample will contain no more than 2 defective units?
56. A recent issue of a newspaper said that given a 5% probability of an unusual event in a 1-year study, one should expect a 35% probability in a 7-year study. This is obviously faulty. What is the correct probability?
57. Independent events A and B have probabilities p_A and p_B , respectively. Show that the probability of either two successes or two failures in two trials has probability 1/2 if and only if at least one of p_A and p_B is 1/2.

Chapter 2

Discrete Random Variables and Probability Distributions

At this point, we have considered discrete sample spaces and we have derived theorems concerning probabilities for any discrete sample space and some of the events within it. Often, however, events are most easily described by performing some *operation* on the sample points. For example, if two dice are tossed, we might consider the *sum* showing on the two dice; but when we find the sum, we have operated on the sample point seen. Other operations, as we will see, are commonly encountered.

We want to consider some properties of the sum; we start with the sample space. In this example, a natural sample space shows the result on each die and, if the dice are fair, leads to equally likely sample points. Then, the sample space consists of the 36 points in S_1 :

$$S_1 = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}.$$

These points are shown in Figure 2.1.

If we consider the *sum* on the two dice, then a sample space

$$S_2 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

might be considered, but now the sample points are not equally likely.

We call the sum in this example a *random variable*.

Definition: A *random variable* is a real-valued function defined on the points of a sample space.

Various functions occur commonly and we will be interested in a variety of them; sums are among the most interesting of these functions as we will see. We will soon determine the probabilities of various sums, but the determination of these is probably evident now to the reader. We first need, for this problem as well as for others, some ideas and some notation.

2.1 RANDOM VARIABLES

Since we have considered only discrete sample spaces to this point, we discuss discrete random variables in this chapter.

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

First, consider another example. It is convenient to let X denote the number of times an examination is attempted until it is passed. X in this case denotes a random variable; we will use capital letters to denote random variables and small letters to denote values of random variables. Following are some of the infinite sample space, indicating the value of X, x at each point.

Event	x
P	1
FP	2
FFP	3
$FFFP$	4
.	.
.	.
.	.

Clearly, we see that the event “ $X = 3$ ” is equivalent to the event “ FFP ” and so their probabilities must be equal. Therefore,

$$P(X = 3) = P(FFP) = \frac{1}{8}.$$

The terminology “random variable” is curious since we could, in the earlier example, define a variable, say Y , to be 6 regardless of the outcome of the experiment. Y would carry no information whatsoever, and it would be neither random nor variable! There are other curiosities with terminology in probability theory as well, but they have become, alas, standard in the field and so we accept them. What we call here a “random variable” is in reality a *function* whose domain is the sample space and whose range is the real line. The random variable here, as in all cases, provides a mapping from the sample space to the real line. While being technically incorrect, the phrase “random variable” seems to convey the correct idea. This perhaps becomes a bit more clear when we use functional notation to define a function $f(x)$ to be

$$f(x) = P(X = x),$$

where x denotes a *value* of the random variable X .

In the earlier example, we could then write $f(3) = 1/8$.

The function $f(x)$ is called a *probability distribution function* (abbreviated as pdf) for the random variable X .

Since probabilities must be nonnegative and since the probabilities must sum to 1, we see that

$$[1] f(x) \geq 0 \text{ and}$$

$$[2] \sum_S f(x) = 1 \text{ where } S \text{ is the sample space.}$$

We turn now to some examples of random variables.

Example 2.1.1

Throw a fair die once and let X denote the result. The random variable X can assume the values 1, 2, 3, 4, 5, 6, and so

$$P(X = x) = \begin{cases} \frac{1}{6} & \text{for } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

A graph of this function is of course flat; it is shown in Figure 2.1. This is an example of a *discrete uniform probability distribution*.

The use of a computer algebra system for sampling from this distribution is explained in Appendix A.

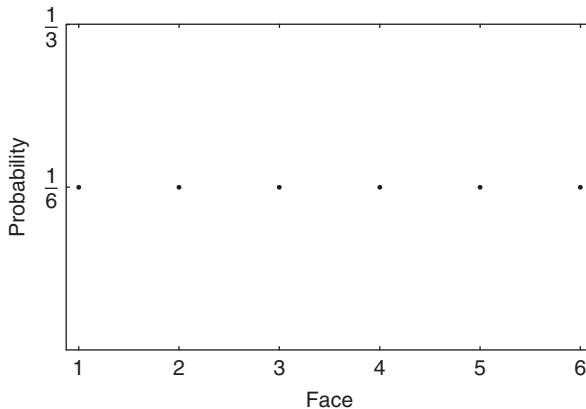


Figure 2.1 Discrete uniform probability distribution.

Example 2.1.2

In the previous example, the die is fair, so now we consider an unfair die. In particular, could the die be weighted so that the probability a face appears is proportional to the face?

Suppose that X denotes the face that appears and let $P(X = x) = k \cdot x$ where k denotes the constant of proportionality. The probability distribution function is then

$$P(X = x) = \begin{cases} k & \text{if } x = 1 \\ 2k & \text{if } x = 2 \\ 3k & \text{if } x = 3 \\ 4k & \text{if } x = 4 \\ 5k & \text{if } x = 5 \\ 6k & \text{if } x = 6 \end{cases}$$

The sum of these probabilities must be 1, so

$$k + 2k + 3k + 4k + 5k + 6k = 1,$$

hence $k = 1/21$ and the weighting is possible.

The probability distribution function is then

$$P(X = x) = \begin{cases} \frac{x}{21} & x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

A procedure for selecting a random sample from this distribution is explained in Appendix A.

Example 2.1.3

Now we return to the experiment consisting of throwing two fair dice. We want to investigate the probabilities of the various sums that can occur. Let the random variable X denote the sum that appears. Then, for example,

$$\begin{aligned} P(X = 5) &= P[(1, 4) \text{ or } (2, 3) \text{ or } (3, 2) \text{ or } (4, 1)] \\ &= \frac{4}{36} = \frac{1}{9}. \end{aligned}$$

So we have determined the probability of one sum. Others can be determined on a similar way.

The experiment could then be described by giving all the values for the probability distribution function (or pdf), $P(X = x)$, where, as earlier, x denotes a *value* for the random variable X , as we saw in Example 2.1.2.

In this example, it is easy to find that

$$P(X = x) = \begin{cases} \frac{1}{36} & \text{if } x = 2 \text{ or } 12 \\ \frac{2}{36} & \text{if } x = 3 \text{ or } 11 \\ \frac{3}{36} & \text{if } x = 4 \text{ or } 10 \\ \frac{4}{36} & \text{if } x = 5 \text{ or } 9 \\ \frac{5}{36} & \text{if } x = 6 \text{ or } 8 \\ \frac{6}{36} & \text{if } x = 7 \\ 0 & \text{otherwise} \end{cases}.$$

We see that

$$P(X = x) = P(X = 14 - x) = \frac{x - 1}{36} \text{ for } x = 2, 3, 4, 5, 6, 7, \text{ and}$$

$$P(X = x) = 0 \text{ otherwise.}$$

A graph of this function shows a tent-like shape shown in Figure 2.2.

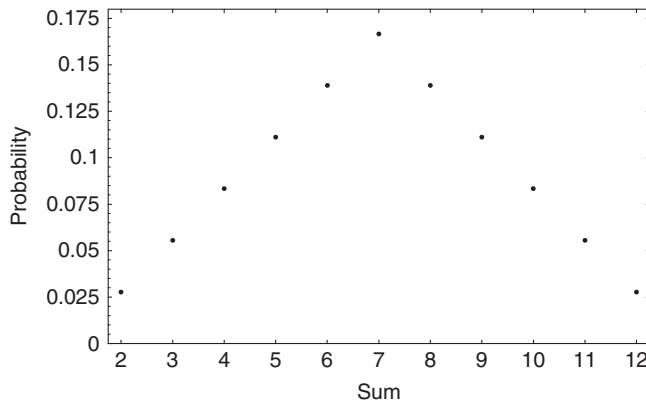


Figure 2.2 Sums on two fair dice.

The sums when two dice are thrown then behave quite differently from the behavior of the individual dice. In fact, we note that if the random variable X_1 denotes the result showing on the first die and the random variable X_2 denotes the result showing on the second die, then $X = X_1 + X_2$. The random variable X can be expressed as a *sum* of random variables. While X_1 and X_2 are uniform, X most decidedly is *not* uniform. There is a theoretical reason for this behavior, which will be discussed in a later chapter. It is sufficient to note here that this is, in fact, not unusual, but very typical behavior for a sum of random variables.

A natural inquiry at this point is, “What is the probability distribution of the sum on three fair dice?”

It is more difficult to work out the distribution here than it was for two dice. Although we will show another solution later, we give one approach to the problem at this time. Consider, for example, a sum of 10 on three dice. The sum could have arisen from these *combinations* of results showing on the individual dice (which do not indicate which die showed which face):

$$(2, 2, 6), (3, 3, 4), (2, 4, 4),$$

$$(3, 1, 6), (3, 2, 5), (5, 1, 4).$$

Each of the first three of these combinations could occur in three different orders (corresponding to the three different dice), while each of the last three could occur in six different orders. This gives a total of 27 possibilities, each of which has probability $\frac{1}{216}$. Therefore, $P(X = 10) = \frac{27}{216}$. A similar process could be followed for other values of the sum; the complete probability distribution can be found to be

$$P(X = x) = \begin{cases} \frac{1}{216} & \text{if } x = 3 \text{ or } 18 \\ \frac{3}{216} & \text{if } x = 4 \text{ or } 17 \\ \frac{6}{216} & \text{if } x = 5 \text{ or } 16 \\ \frac{10}{216} & \text{if } x = 6 \text{ or } 15 \\ \frac{15}{216} & \text{if } x = 7 \text{ or } 14 \\ \frac{21}{216} & \text{if } x = 8 \text{ or } 13 \\ \frac{25}{216} & \text{if } x = 9 \text{ or } 12 \\ \frac{27}{216} & \text{if } x = 10 \text{ or } 11 \\ 0 & \text{otherwise} \end{cases}.$$

A computer algebra system may also be used to find the probability distribution for X . Many systems will give all the permutations, each of which may be summed and the relative frequencies recorded. This is shown in Appendix A. There are other methods that can be used to solve the problem; one of these will be discussed in Chapter 4.

A graph of this function is shown in Figure 2.3. It begins to show what we will call a *normal* probability distribution shape. As the number of dice increases, the “curve” the

eye sees smooths out to resemble a *normal* probability distribution; the distribution for 6 or more dice is remarkably close to the normal distribution. We will discuss the normal distribution in Chapter 3.

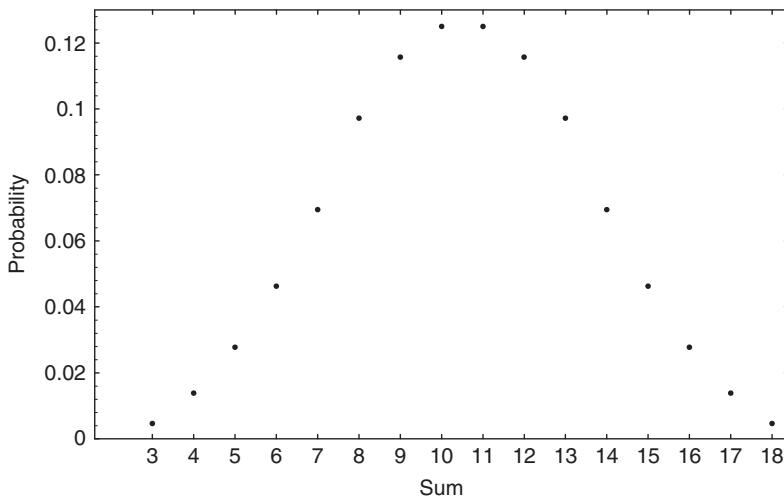


Figure 2.3 Sums on three fair dice.

Example 2.1.4

We saw in Example 2.1.2 that a single die could be loaded so that the probability of the occurrence of a face is proportional to the face. Can we load a die so that when the die is thrown twice the probability of a sum is proportional to the sum?

If $P(X = i)$ is denoted by P_i , for $i = 1, 2, 3, 4, 5, 6$, and if k is the constant of proportionality, then $P_1^2 = 2k$, $2P_1P_2 = 3k$, $2P_1P_3 + P_2^2 = 4k$, and so on, together with the restriction that $\sum_{i=1}^6 P_i = 1$, giving a system of 12 equations in 7 unknowns. Unfortunately, this set of equations has no solution, however, so we cannot load the die in the manner suggested.

Example 2.1.5

Let us look now at the sum when two loaded dice are thrown. First, let each die be loaded so that the probability a face occurs is proportional to that face, as is Example 2.1.2. The sample space of 36 points can be used to determine the probabilities of the various sums. Figure 2.4 shows these probabilities. We see that the symmetry we noticed in Figures 2.1 and 2.3 is now gone.

Now suppose one die is loaded so that the probability a face appears is proportional to that face while a second die is loaded so that the probability face i appears is proportional to $7 - i$, $i = 1, 2, \dots, 6$. The probabilities of various sums are then shown in Figure 2.5. Now symmetry around $x = 7$ has returned.

The appearance, once more, of the normal-like shape is striking. The reader with access to a computer algebra system may want to find the probability distribution of the sums on four dice, two loaded in each manner as in this example. The result is remarkably normal.

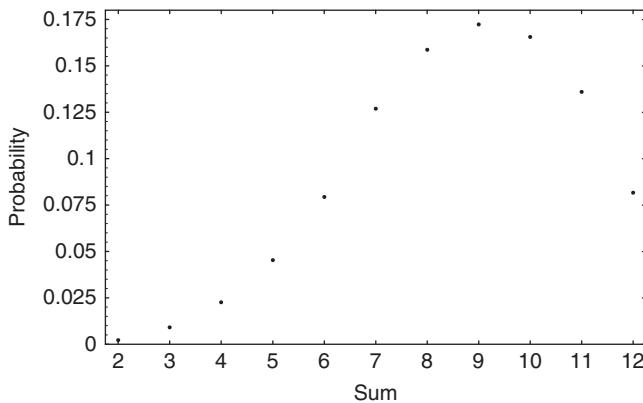


Figure 2.4 Sums on two similarly loaded dice.

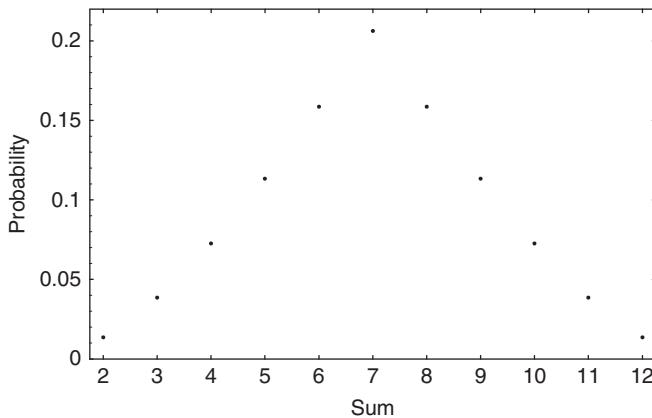


Figure 2.5 Sums on two differently loaded dice.

Example 2.1.6

Sample spaces in the examples in this chapter so far have been finite. Our final example involves a countably infinite sample space. Consider observing single births until a girl is born. Let the random variable X denote the number of births necessary. Assuming the births to be independent,

$$P(X = x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, 3, \dots$$

To check that $P(X = x)$ is a probability distribution, note that $P(X = x) \geq 0$ for all x . The sum of all the probabilities is

$$S = \sum_{x=1}^{\infty} P(X = x) = \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \dots$$

To calculate this sum, note that

$$\left(\frac{1}{2}\right)S = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 \dots$$

Subtracting the second series from the first series gives

$$\left(\frac{1}{2}\right)S = \left(\frac{1}{2}\right) \text{ so} \\ S = 1.$$

Another way to sum the series is to recognize that it is an infinite geometric series of the form

$S = a + ar + ar^2 + \dots$ and the sum of this series is known to be

$$S = \frac{a}{1 - r}, \text{ if } |r| < 1.$$

In this case, a is $\frac{1}{2}$ and r is also $\frac{1}{2}$, so the sum is 1.

Here, X is called a *geometric* random variable. A graph of $P(X = x)$ appears in Figure 2.6.

Since $P(X = x + 1) = \left(\frac{1}{2}\right)P(X = x)$, the probabilities decline rapidly in size.

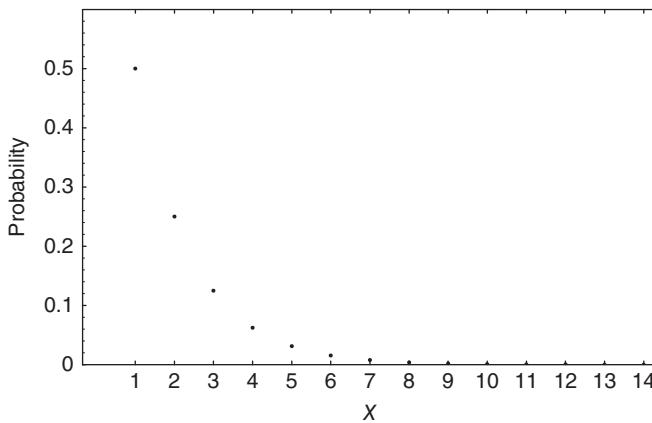


Figure 2.6 Geometric distribution.

2.2 DISTRIBUTION FUNCTIONS

Another function often useful in probability problems is called the *distribution function*. For a discrete random variable, we denote this function by $F(x)$ where

$$F(x) = P(X \leq x), \text{ so}$$

$$F(x) = \sum_{t \leq x} f(t).$$

$F(x)$ is also known as a *cumulative distribution function* (abbreviated *cdf*) since it accumulates probabilities. Note the distinction now between $f(x)$, the *probability distribution function (pdf)*, and $F(x)$, the *cumulative distribution function (cdf)*.

In Chapter 1, we used the *reliability* of a component where $R(t) = P(T > t)$ so

$$R(t) = 1 - F(t),$$

establishing a relationship between $R(t)$ and the distribution function.

Example 2.2.1

For the fair die whose probability distribution function is given in Example 2.1.1, we find

$$F(1) = 1/6, F(2) = 2/6, F(3) = 3/6, F(4) = 4/6, F(5) = 5/6, F(6) = 1.$$

It is also customary to show this function for *any* value of the random variable X . Here, for example, $F(3.4) = P(X \leq 3.4) = 3/6$. Since $F(x)$ is defined for any value of X , we draw a continuous graph, unlike the graph of the probability distribution function. We see that in this case

$$F(x) = \begin{cases} 0, & \text{if } x < 1 \\ \frac{1}{6}, & \text{if } 1 \leq x < 2 \\ \frac{2}{6}, & \text{if } 2 \leq x < 3 \\ \frac{3}{6}, & \text{if } 3 \leq x < 4 \\ \frac{4}{6}, & \text{if } 4 \leq x < 5 \\ \frac{5}{6}, & \text{if } 5 \leq x < 6 \\ 1, & \text{if } 6 \leq x \end{cases}.$$

A graph of this function is shown in Figure 2.7. It is a series of step functions, since, when $f(x)$ is scanned from the right, $F(x)$ can increase only at those points where $f(x)$ is not zero.

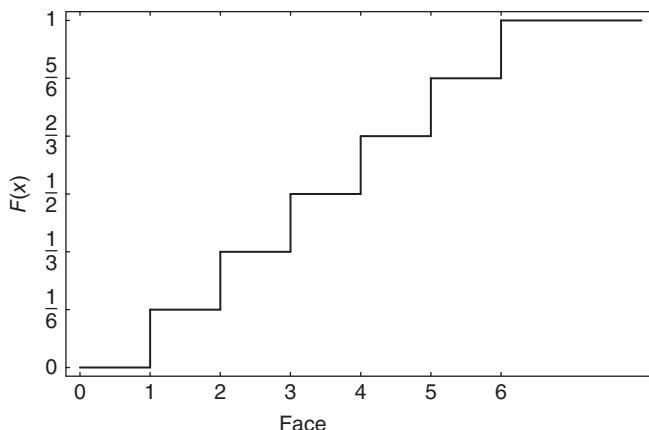


Figure 2.7 Distribution function for one toss of a fair die.

It is clear from the definition of $F(x)$ and from the fact that probabilities are in the interval $[0,1]$ that

$$0 \leq F(x) \leq 1 \text{ and that}$$

$$F(a) \geq F(b) \text{ if } a \geq b.$$

It is also true, for *discrete* random variables taking integer values, that

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a - 1).$$

Individual probabilities, say $P(X = a)$, can be found by

$$P(X = a) = P(X \leq a) - P(X \leq a - 1) = F(a) - F(a - 1).$$

These probabilities are then the size of the “steps” in the distribution function.

EXERCISES 2.2

1. Suppose the probability distribution function for a random variable, X , is $P(X = x) = 1/5$ for $x = 1, 2, 3, 4, 5$.
 - Find $P(X > 3)$.
 - Find $P(X \text{ is even})$.
2. Draw a graph of the cumulative distribution function in problem 1.
3. A fair coin is tossed four times.
 - Show a sample space for the experiment and assign probabilities to the sample points.
 - Suppose a count of the total number of heads (X) and the total number of tails (Y) is made after each toss. What is the probability that X always exceeds Y ?
 - What is the probability, after four tosses, that X is even if we know that $Y \geq 1$?
4. A single expensive electronic part is to be manufactured, but the manufacture of a successful part is not guaranteed. The first attempt costs \$100 and has a 0.7 probability of success. Each attempt thereafter costs \$60 and has a 0.9 probability of success. The outcomes of various attempts are independent, but at most three attempts can be made at successful manufacture. The finished part sells for \$500. Find the probability distribution for N , the net profit.
5. An automobile dealer has found that X , the number of cars customers buy each week, follows the probability distribution

$$f(x) = \begin{cases} \frac{kx^2}{x!}, & x = 1, 2, 3, 4, \\ 0, & \text{otherwise} \end{cases}$$

- Find k .
- Find the probability the dealer sells at least two cars in a week.
- Find $F(x)$, the cumulative distribution function.

6. Job interviews last 1/2 hour. The interviewer knows that the probability an applicant is qualified for the job is 0.8. The first person interviewed who is qualified is selected for the job. If the qualifications of any one applicant is independent of the qualifications of any other applicant, what is the probability that 2 hours is sufficient time to select a person for the job?
7. Verify the probability distribution for the sum on three fair dice as given in Example 2.1.3.
8. (a) Since $\left(\frac{1}{2} + \frac{1}{2}\right)^5 = 1$ and since each term in the binomial expansion of $\left(\frac{1}{2} + \frac{1}{2}\right)^5$ is greater than 0, it follows that the individual terms in the binomial expansion are probabilities. Suggest an experiment and a sample space for which these terms represent probabilities of the sample points.
(b) Answer part (a) for $(p + q)^n$, $q = 1 - p$, $0 \leq p \leq 1$.
9. Two loaded dice are tossed. Each die is loaded so that the probability a face, i , appears is proportional to $7 - i$. Find the probability distribution for the sum that appears. Draw a graph of the probability distribution function.
10. Suppose that X is a random variable giving the number of tosses necessary for a fair coin to turn up heads. Find the probability that X is even.
11. The random variable Y has the probability distribution $g(y) = \frac{1}{4}$ if $y = 2, 3, 4$, or 5. Find $G(y)$, the distribution function for Y .
12. Find the distribution function for the geometric distribution $f(x) = \left(\frac{1}{2}\right)^x$, $x = 1, 2, 3 \dots$
13. A random variable, X , has the distribution function

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{3}, & -1 \leq x < 0 \\ \frac{5}{6}, & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}.$$

Find the probability distribution function, $f(x)$.

14. A random variable X is defined on the integers 0, 1, 2, 3, ... and has distribution function $F(x)$. Find expressions, in terms of $F(x)$, for the following:
- (a) $P(a < X < b)$
(b) $P(a \leq X < b)$
(c) $P(a < X \leq b)$
(d) $P(a \leq X \leq b)$.
15. If $f(x) = 1/n$, $x = 1, 2, 3, \dots, n$ (so that each value of X has the same probability), then X is called a *discrete uniform* random variable. Find the distribution function for this random variable.

2.3 EXPECTED VALUES OF DISCRETE RANDOM VARIABLES

Expected Value of a Discrete Random Variable

Random variables are easily distinguished by their probability distribution functions. They are also often characterized or described by measures that summarize these distributions. Usually, “average” values, or measures of centrality, and some measure of their dispersion, or variability, are found as values characteristic of the distribution.

We begin with the definition of an average value for a discrete random variable, X , denoted by $E(X)$, or μ_x , which we will call the *expectation*, or *expected value*, or *mean*, or *mean value* (all of these terms are in common usage) of X :

$$\text{Definition: } E(X) = \mu_x = \sum_x x \cdot P(X = x),$$

provided the sum converges, where the summation occurs over all the discrete values of the random variable, X . Note that each value of the random variable X is weighted by its probability in the sum.

The provision that the sum be convergent cautions us that the sum may, indeed, be infinite. There are random variables, otherwise seemingly well behaved, which have no mean value.

This definition is, in reality, a simple extension of what the reader would recognize as an average value. Consider an example:

Example 2.3.1

A student has examination grades of 82, 91, 79, and 96 in a course in probability. We would no doubt calculate the average grade as

$$\frac{82 + 91 + 79 + 96}{4} = 87.$$

This could also be calculated as

$$82 \cdot \frac{1}{4} + 91 \cdot \frac{1}{4} + 79 \cdot \frac{1}{4} + 96 \cdot \frac{1}{4} = 87,$$

where the examination scores have now been equally weighted. Should the instructor decide to weight the fourth examination three times as much as any one of the other examinations, this simply changes the weights and the average examination grade is then

$$82 \cdot \frac{1}{6} + 91 \cdot \frac{1}{6} + 79 \cdot \frac{1}{6} + 96 \cdot \frac{3}{6} = 90.$$

So the idea of adding scores multiplied by their probabilities is not a new one. This is exactly what we do when we calculate $E(X)$.

Example 2.3.2

If a fair die is thrown once, as in Example 2.1.1, the average result is

$$\mu_x = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

So we recognize $7/2$, or 3.5, as the average result, although 3.5 is not a possible value for the face showing on the die. What is the meaning of this? The interpretation is as follows: if we threw a fair die a large number of times, we would expect each of the faces from 1 to 6 to occur about 1/6th of the time, so the average result would be given by μ_x . We could, of course, expect some deviation from this result in actual practice, the size of the deviation decreases as the number of tosses of the die increases. Later we will see that a deviation of more than about 0.11 in the average is highly unlikely in 1000 tosses of the die, that is, the average is almost certain to fall in the interval from 3.39 to 3.61. If the deviation is more than 0.11, we would no doubt conclude that the die is an unfair one.

Example 2.3.3

What is the average result on the loaded die where $P(X = i) = i/21$, for $i = 1, 2, 3, 4, 5, 6$?

$$\text{Here, } E(X) = 1 \cdot \frac{1}{21} + 2 \cdot \frac{2}{21} + 3 \cdot \frac{3}{21} + 4 \cdot \frac{4}{21} + 5 \cdot \frac{5}{21} + 6 \cdot \frac{6}{21} = 13/3.$$

Example 2.3.4

In Example 2.1.3, we determined the probability distribution for X , the sum showing on two fair dice. Then, we find

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \dots + 12 \cdot \frac{1}{36} = 7.$$

Now let X_1 denote the face showing on the first die and let X_2 denote the face showing on the second die. We found in Example 2.3.2 that $E(X_i) = \frac{7}{2}$, for $i = 1, 2$. We note here that

$$E(X) = E(X_1) + E(X_2),$$

so that the expectation of the sum is the sum of the expectations of the sum's components; this is in fact generally true and so is no coincidence. We will discuss this further in Chapter 5.

Example 2.3.5

Sometimes, the calculation of an expected value will involve an infinite series. Suppose we toss a coin, loaded to come up heads with probability p , until heads occur. Since the

tosses are independent, and since the event, “First head on toss x ,” is equivalent to $x - 1$ tails followed by heads, it follows that

$$P(X = x) = q^{x-1}p, \quad x = 1, 2, 3, \dots, \text{ where } q = 1 - p.$$

We check first that $\sum_x P(X = x) = 1$. Here,

$$\begin{aligned}\sum_x P(X = x) &= p + q \cdot p + q^2 \cdot p + q^3 \cdot p + \dots \\ &= p \cdot (1 + q + q^2 + q^3 + \dots) \\ &= p \cdot \frac{1}{1 - q} = 1.\end{aligned}$$

Then,

$$E(X) = \sum_{x=1}^{\infty} x \cdot q^{x-1}p = p + 2 \cdot q \cdot p + 3 \cdot q^2 \cdot p + 4 \cdot q^3 \cdot p + \dots$$

To simplify this, notice that

$$q \cdot E(X) = q \cdot p + 2 \cdot q^2 \cdot p + 3 \cdot q^3 \cdot p + 4 \cdot q^4 \cdot p + \dots$$

By subtracting $q \cdot E(X)$ from $E(X)$, we find that

$$E(X) - q \cdot E(X) = p + q \cdot p + q^2 \cdot p + q^3 \cdot p + q^4 \cdot p + \dots$$

where the right-hand side is $\sum_x P(X = x) = 1$. So

$$(1 - q) \cdot E(X) = 1, \text{ hence}$$

$$E(X) = \frac{1}{p}.$$

(The reader is cautioned that the “trick” above for summing the series is valid only because the series is absolutely convergent. $E(X)$ could also be found by integrating, with respect to q , the series for $E(X)$ term by term.)

With a fair coin, then, since $p = 1/2$, an average of two tosses is necessary to find the first occurrence of heads. Since $P(X = x)$ involves a geometric series, X here, as in Example 2.1.6, is often called a *geometric* random variable.

Mean values generally show a central value for the random variable. Now we turn to a discussion of the dispersion, or variability, of the random variable.

Variance of a Random Variable

Figure 2.8 shows two random variables with the same mean value, $\mu = 3$. These graphs are continuous; the reader may regard them as idealized discrete random variables. Continuous random variables will be discussed in Chapter 3. If we did not know μ and wanted to estimate μ by selecting an observation from one of these probability distributions, we would no doubt choose Y since the values of Y are less disperse and generally closer to μ than those for X .

There are many ways to measure the fact that Y is less disperse than X . We could look at the *range* (the largest possible value minus the smallest possible value); another possibility is to calculate the deviation of each value of X from μ and then calculate the average value of these deviations from the mean, $E(X - \mu)$. This, however, is 0 for any random variable and hence carries absolutely no information whatsoever regarding X . Here is a demonstration that this is so:

$$\begin{aligned} E(X - \mu) &= \sum_x (x - \mu) \cdot P(X = x) \\ &= \sum_x x \cdot P(X = x) - \mu \cdot \sum_x P(X = x) \\ &= \mu - \mu = 0. \end{aligned}$$

So the positive deviations from the mean exactly compensate for the negative deviations.

One way to avoid this is to consider the *mean deviation*, $E|X - \mu|$, but this is not commonly done. Yet another way to prevent the positive deviations from compensating for the negative deviations is to square each value of $X - \mu$ and then sum the result. This is the usual solution; we call the result the *variance*, denoted by σ^2 , which we define as

Definition:

$$\begin{aligned} \sigma^2 &= \text{Var}(X) = E(X - \mu)^2 \text{ so} \\ \sigma^2 &= \sum_x (x - \mu)^2 \cdot P(X = x), \end{aligned} \tag{2.1}$$

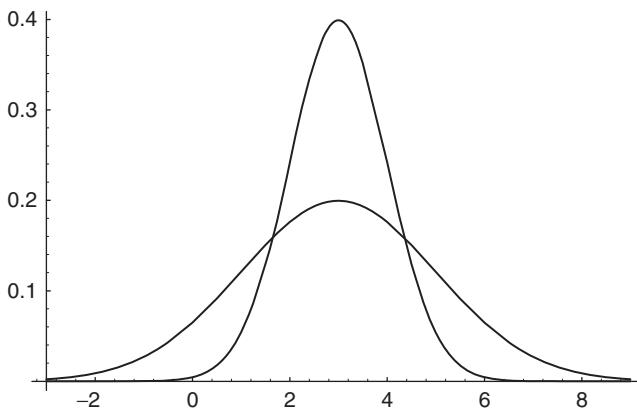


Figure 2.8 Two random variables with the same mean value.

provided the sum converges, and where the summation is over all the possible values of X .

The quantity σ^2 is then a weighted average of the squared deviations of the values of X from its mean value. The variance may appear to be much more complex than the range or mean deviation. This is true, but the variance also has remarkable properties that we cannot describe now and which do not hold for the range or for the mean deviation.

Example 2.3.6

Consider the random variable X with probability distribution function:

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{3} & \text{if } x = 2 \\ \frac{1}{6} & \text{if } x = 3. \end{cases}$$

Here, $E(X) = \mu = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{6} = \frac{5}{3}$, so

$$E(X - \mu)^2 = \sigma^2 = \left(1 - \frac{5}{3}\right)^2 \cdot \frac{1}{2} + \left(2 - \frac{5}{3}\right)^2 \cdot \frac{1}{3} + \left(3 - \frac{5}{3}\right)^2 \cdot \frac{1}{6} = \frac{5}{9}.$$

Before turning to some more examples, we show another formula for σ^2 . This formula is often very useful.

Expand (2.1) as follows:

$$\begin{aligned} \sigma^2 &= \sum_x (x - \mu)^2 \cdot P(X = x) \\ &= \sum_x (x - \mu)^2 \cdot f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) \cdot f(x) \\ &= \sum_x x^2 \cdot f(x) - 2\mu \sum_x x \cdot f(x) + \mu^2 \end{aligned}$$

since $\sum_x f(x) = 1$.

Now $\sum_x x \cdot f(x) = \mu$, so

$$\sigma^2 = \sum_x x^2 \cdot f(x) - \mu^2 \tag{2.2}$$

$$\text{So } \sigma^2 = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2.$$

Formula (2.2) is often easier to use for computational purposes than formula (2.1). σ is called the *standard deviation of X* .

Example 2.3.7

Refer again to throwing a single die, as in Examples 2.1.1 and 2.2.2. We calculate

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6} \text{ so that}$$

$$\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Example 2.3.8

What is the variance of the geometric random variable whose probability distribution function is $P(X = x) = q^{x-1} \cdot p$, $x = 1, 2, 3, \dots$?

Starting with $\sigma^2 = E(X^2) - \mu^2$, since we know that $\mu = \frac{1}{p}$, we only need to compute $E(X^2)$:

$$E(X)^2 = \sum_{x=1}^{\infty} x^2 q^{x-1} p = p(1^2 + 2^2 q + 3^2 q^2 + \dots)$$

from which no easily seen pattern emerges.

Another thought is to consider $E[X(X - 1)]$. If we write

$$\begin{aligned} E[X(X - 1)] &= \sum_{x=1}^{\infty} (x^2 - x) \cdot P(X = x), \text{ we see that} \\ E[X(X - 1)] &= \sum_{x=1}^{\infty} x^2 \cdot P(X = x) - \sum_{x=1}^{\infty} x \cdot P(X = x) \quad \text{or} \\ E(X^2 - X) &= E(X^2) - E(X). \end{aligned}$$

So if we know $E[X(X - 1)]$, we can find $E(X^2)$ and hence calculate σ^2 . In this example, a trick will help as it did in determining $E(X)$:

$$E[X(X - 1)] = 1 \cdot 0 \cdot p + 2 \cdot 1 \cdot q \cdot p + 3 \cdot 2 \cdot q^2 \cdot p + 4 \cdot 3 \cdot q^3 \cdot p + \dots$$

so multiplying through by q , we have

$$q \cdot E[X(X - 1)] = 2 \cdot 1 \cdot q^2 \cdot p + 3 \cdot 2 \cdot q^3 \cdot p + 4 \cdot 3 \cdot q^4 \cdot p + \dots$$

Subtract the second series from the first series and, since $p = 1 - q$, it follows that

$$\begin{aligned} p \cdot E[X(X - 1)] &= 2 \cdot q \cdot p + 4 \cdot q^2 \cdot p + 6 \cdot q^3 \cdot p + \dots \\ &= 2q(1p + 2qp + 3q^2p + \dots) \end{aligned}$$

Thus,

$$p \cdot E[X(X - 1)] = 2q \cdot E(X) = \frac{2q}{p}.$$

$$\text{So } E[X(X - 1)] = \frac{2q}{p^2},$$

$$\text{and } E(X^2) = \frac{2q}{p^2} + \frac{1}{p},$$

$$\text{giving } \sigma^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

The value of the variance is quite difficult to interpret at this point but, as we proceed, we will find more and more uses for the variance. Patience is requested of the reader now, with the promise that these calculations are in fact useful and meaningful. We pause to consider the question, “Does σ measure variability?” We can show a general result, albeit a very crude one, in the following inequality.

Tchebycheff's Inequality

Theorem 1: Suppose the random variable X has mean μ and standard deviation σ . Choose a positive quantity, k . Then,

$$P(|X - \mu| \leq k \cdot \sigma) \geq 1 - \frac{1}{k^2}.$$

Tchebycheff's inequality gives a lower bound on the probability a value of X is within $k \cdot \sigma$ units of the mean, μ .

Before offering a proof, we consider some special cases. If $k = 2$, the inequality is

$$P(|X - \mu| \leq 2 \cdot \sigma) \geq 1 - \frac{1}{2^2} = \frac{3}{4},$$

so 3/4 of *any* probability distribution lies within two standard deviations, that is, 2σ units of the mean while, if $k = 3$, the inequality states that

$$P(|X - \mu| \leq 3 \cdot \sigma) \geq 1 - \frac{1}{3^2} = \frac{8}{9},$$

showing that 8/9 of *any* probability distribution lies within 3σ units of the mean. We will see later that if the specific distribution is known, these inequalities can be sharpened considerably. Now we show a proof.

Proof: Let $P(X = x) = f(x)$. Consider two sets of points:

$$A = \{x | |x - \mu| \geq k \cdot \sigma\} \text{ and}$$

$$B = \{x | |x - \mu| < k \cdot \sigma\}.$$

We could then write the variance as

$$\sigma^2 = \sum_{x \in A} (x - \mu)^2 \cdot f(x) + \sum_{x \in B} (x - \mu)^2 \cdot f(x).$$

Now for every point x in A , replace $|x - \mu|$ by $k \cdot \sigma$ and in B replace $|x - \mu|$ by 0. The crudity of the result is now evident! So

$$\sigma^2 \geq \sum_{x \in A} (k \cdot \sigma)^2 f(x) + \sum_{x \in B} 0^2 \cdot f(x).$$

Since $\sum_{x \in A} f(x) = P(A) = P(|X - \mu| \geq k \cdot \sigma)$,

$\sigma^2 \geq k^2 \cdot \sigma^2 \cdot P(|X - \mu| \geq k \cdot \sigma)$ from which we conclude that

$$P(|X - \mu| \geq k \cdot \sigma) \leq \frac{1}{k^2} \text{ or}$$

$$P(|X - \mu| \leq k \cdot \sigma) \geq 1 - \frac{1}{k^2}.$$

While the theorem is far from precise, it does verify that as we move farther away from the mean, in terms of standard deviations, the more of the probability distribution we cover; hence σ is indeed a measure of variability.

EXERCISES 2.3

1. If X is the outcome when a loaded die with $P(X = x) = x/21$ for $x = 1, 2, 3, 4, 5, 6$, find μ and σ^2 .
2. Verify Tchebycheff's inequality in problem 1.
3. A small manufacturing firm sells 1 machine per month with probability 0.3; it sells 2 machines per month with probability 0.1; it never sells more than 2 machines per month. If X represents the number of machines sold per month,
 - (a) find the mean and variance of X .
 - (b) If the monthly profit is $2X^2 + 3X + 1$ (in thousands of dollars), find the expected monthly profit.
4. Bolts are packaged in boxes so that the mean number of bolts per box is 100 with standard deviation 3. Use Tchebycheff's inequality to find a bound on the probability that the box has between 95 and 105 bolts.
5. Graduates of a distinguished undergraduate mathematics program received graduate school fellowships as follows: 20% received \$10,000; 10% received \$12,000; 30% received \$14,000; 30% received \$13,000; 5% received \$15,000; and 5% received \$17,000. Find the mean and the variance of the value of a graduate fellowship.
6. A fair coin is tossed four times; let X denote the number of heads that occur. Find the mean and variance of X .
7. A batch of 15 electric motors actually contains three defective motors. An inspector chooses 3 (without replacement). Find the mean and variance of X , the number of defective motors in the sample.
8. A coin, loaded to show heads with probability $2/3$, is tossed until heads appear or until 5 tosses have been made. Let X denote the number of tosses made. Find the mean and variance of X .
9. Suppose X is a discrete uniform random variable so that $f(x) = 1/n, x = 1, 2, 3, \dots, n$. Find the mean and variance of X .

10. In problem 5, suppose the batch of motors is accepted if no more than 1 defective motor is in the sample. If each motor costs \$100 to manufacture, how much should the manufacturer charge for each motor in order to make the expected profit for the batch be \$200?
11. A physicist makes several independent measurements of the specific gravity of a substance. The limitations of his equipment are such that the standard deviation of each measurement is σ units. Suppose μ is the true specific gravity of the substance. Approximate the probability that one of the measurements is within $5\sigma/4$ units of μ .
12. A manufacturer ships parts in lots of 1000 and makes a profit of \$50 per lot sold. The purchaser, however, subjects the product to a sampling inspection plan as follows: 10 parts are selected at random. If none of these parts is defective, the lot is purchased; if 1 part is defective, the manufacturer returns \$10 to the buyer; if 2 or more parts are found to be defective, the entire lot is returned at a net loss of \$25 to the manufacturer. What is the manufacturer's expected profit if 10% of the parts are defective? (Assume that the sampling is done with replacement.)
13. In a lot of six batteries, one is worn out. A technician tests the batteries one at a time until the worn out battery is found. Tested batteries are put aside, but after every third test the tester takes a break and another worker, unaware of the test, returns one of the tested batteries to the set of batteries not yet tested.
 - (a) Find the probability distribution for X , the number of tests required to identify the worn out battery.
 - (b) Assume the first test of each set of three tests costs \$5, and that each of the next two tests in each set of three tests costs \$2. Find the increase in the expected cost of locating the worn out battery due to the unaware worker.
14. A carnival game consists of hitting a lever with a sledge hammer to propel a weight upward toward a bell. Because the hammer is quite heavy, the chance of ringing the bell declines with the number of attempts; in particular, the probability of ringing the bell on the i th attempt is $(3/4)^i$. For a fee, the carnival sells you the privilege of swinging the hammer until the bell rings or until you have made three attempts, whichever occurs first.
 - (a) Find the probability distribution of X , the number of hits taken.
 - (b) The prize for ringing the bell on the i th try is $\$(4 - i)$, $i = 1, 2, 3$. How much should the carnival charge for playing the game if it wants an expected profit of \$1 per customer?
15. Suppose X is a random variable defined on the points $x = 0, 1, 2, 3, \dots$. Calculate

$$\sum_{x=0}^{\infty} P(X > x).$$

There are many very important specific discrete probability distribution functions that arise in practical applications. Having established some general properties, we now turn to discussions of several of the most important of these distributions.

Occasionally, random variables in apparently different situations actually arise from common assumptions and hence lead to the same probability distribution function. We now investigate some of these special circumstances and the probability distribution functions which result.

2.4 BINOMIAL DISTRIBUTION

Among all discrete probability distribution functions, the most commonly occurring one, arising in a great variety of applications, is called the *binomial* probability distribution function. It is attributed to James Bernoulli.

Consider an experiment where, on each trial of the experiment, one of only two outcomes occurs, which we describe as *success* (*S*) or *failure* (*F*). For example, a manufactured part is either good or does not meet specifications; a student's examination score is passing or it is not; a team wins a basketball game or it does not—these are some examples of binomial variables and the reader can no doubt think of many more. One of these outcomes can be associated with success and the other with failure; it does not matter which is which.

In addition to the restriction that there be two and only two outcomes on each trial of the experiment, suppose further that the trials are independent and that the probabilities of success or failure at each trial remain constant from trial to trial and do not change with subsequent performances of the experiment.

The individual trials of such an experiment are often called *Bernoulli trials*.

Consider, as a specific example, 5 independent trials with probability $\frac{2}{3}$ of success at any trial. Then, if interest centers on the occurrence of exactly 3 successes, we note that exactly 3 successes can occur in 10 different ways:

$$SSSFF, SSFSF, SFSSF, FSSSF, SFSFS,$$

$$SSFFS, FSSFS, SFFSS, FSFSS, FFSSS.$$

There are $\binom{5}{3} = 10$ of these mutually exclusive orders. Each has probability $\left(\frac{2}{3}\right)^3 \cdot \left(\frac{1}{3}\right)^2$, so

$$P(\text{Exactly 3 } S\text{'s in 5 trials}) = \binom{5}{3} \cdot \left(\frac{2}{3}\right)^3 \cdot \left(\frac{1}{3}\right)^2 = \frac{80}{243}.$$

Now return to the general situation. Let the probabilities be $P(S) = p$ and $P(F) = q = 1 - p$, and let the random variable X denote the number of successes in n trials of the experiment. Any specific sequence of exactly x successes and $n - x$ failures has probability $p^x \cdot q^{n-x}$. The successes in such a sequence can occur at $\binom{n}{x}$ positions so, since the sequences are mutually exclusive,

$$P(X = x) = \binom{n}{x} p^x \cdot q^{n-x}, \quad x = 0, 1, 2, \dots, n, \quad (2.3)$$

giving the probability distribution function for a binomial random variable.

Although the binomial random variable occurs in many different situations, a perfect model for any binomial situation is that of observing the number of heads when a coin loaded so that the probability of heads is p and that of tails is $q = 1 - p$ is tossed n times.

Now does (2.3) define a probability distribution? Since $P(X = x) \geq 0$ and $\sum_{x=0}^n P(X = x) = \sum_{x=0}^n \binom{n}{x} p^x \cdot q^{n-x} = (q + p)^n = 1$ by the binomial theorem, we conclude that (2.3) defines a probability distribution.

It is interesting to note then that individual terms in the binomial expansion of $(q + p)^n$, if $p + q = 1$, represent binomial probabilities.

Example 2.4.1

A student has no knowledge whatsoever of the material to be tested on a true—false examination, so he flips a fair coin in order to determine his response to each question. What is the probability he scores at least 60% on a ten-item examination?

Here, the binomial variable X the number of correct responses, has $n = 10$ and $p = q = 1/2$. We need

$$P(X \geq 6) = \sum_{x=6}^{10} \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x}.$$

$$\text{Now we find that } P(X \geq 6) = \frac{193}{512} = 0.376953.$$

The above-mentioned calculations can easily be done with a pocket computer. If we want to investigate the probability that at least 60% of the questions were answered correctly as the number of items on the examination increases, then use of a computer algebra system is recommended for aiding in the calculation. Many computer algebra systems contain the binomial probability distribution as a defined probability distribution; for other systems the probability distribution function may be entered directly. The following results can be found where n is the number of trials and P is the probability of at least 60% correct:

n	10	40	80	100
P	0.376953	0.134094	0.0464559	0.028444

Clearly, guessing is not a sensible strategy on a test with a large number of items.

Example 2.4.2

Graphs of $P(X = x) = \binom{n}{x} p^x \cdot q^{n-x}$ for $x = 0, 1, 2, \dots, n$ are interesting. The graphs of $P(X = x)$ for $n = 10$ and also for $n = 100$ with $p = 1/2$ in each case in shown in Figure 2.9. We see that each curve is bell-shaped or normal-like, and the distributions are symmetric about $x = 5$ and $x = 50$, respectively.

Again we find the bell-shaped or normal appearance here, but the reader may wonder if the appearance is still normal for $p \neq 1/2$. Figure 2.10 shows a graph of $P(X = x)$ for $n = 50$ and $p = 3/4$. This curve indicates that the bell shape survives even though $p \neq 1/2$. The maximum point on the curve has shifted to the right, however.

We will discuss the reason for the normal appearance of the binomial distribution in the next chapter. Appendix A contains a procedure for selecting a sample from a binomial distribution and for simulating an experiment consisting of flipping a loaded coin.

2.5 A RECURSION

If a computer algebra system is not available, calculating values of $P(X = x) = \binom{n}{x} p^x \cdot q^{n-x}$ can certainly become difficult, especially for large values of n and small values of p . In any event, $\binom{n}{x}$ becomes large while $p^x \cdot q^{n-x}$ becomes small. By calculating the ratio of successive terms we find an interesting result, which will aid in making these

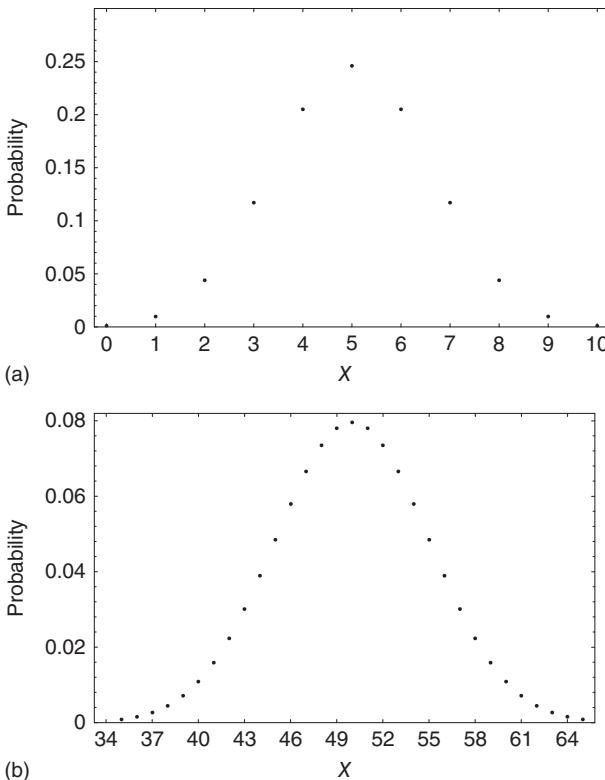


Figure 2.9 (a) Binomial distribution, $n = 10$, $p = 1/2$. (b) Binomial distribution, $n = 100$, $p = 1/2$.

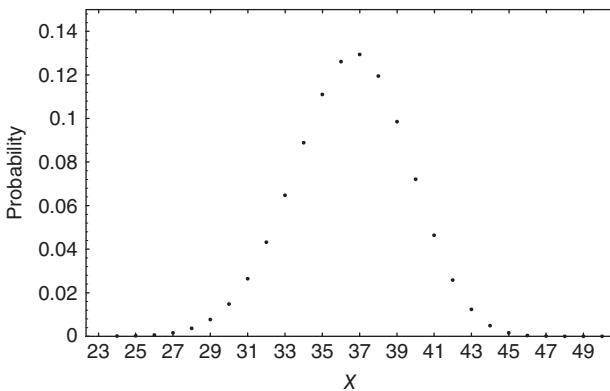


Figure 2.10 Binomial distribution, $n = 50$, $p = 3/4$

calculations (and which has other interesting consequences as well). We divide $P(X = x)$ by $P(X = x - 1)$:

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{\binom{n}{x} p^x \cdot q^{n-x}}{\binom{n}{x-1} p^{x-1} \cdot q^{n-x+1}}, x = 1, 2, \dots, n.$$

This can be simplified to $\frac{P(X = x)}{P(X = x - 1)} = \frac{n - x + 1}{x} \cdot \frac{p}{q}$,

$$\text{so } P(X = x) = \frac{n - x + 1}{x} \cdot \frac{p}{q} \cdot P(X = x - 1), x = 1, 2, \dots, n. \quad (2.4)$$

Formula (2.4) is another example of a *recursion* since it expresses one value of a function, here $P(X = x)$, in terms of another value of the function, here $P(X = x - 1)$. Given a starting point and the recursion, any value of the function can be computed. In this case, since n failures have probability q^n , $P(X = 0) = q^n$ is a natural starting value. We then find that

$$P(X = 0) = q^n,$$

so

$$P(X = 1) = n \cdot \frac{p}{q} \cdot P(X = 0) = n \cdot \frac{p}{q} \cdot q^n = \binom{n}{1} \cdot p \cdot q^{n-1}$$

and

$$\begin{aligned} P(X = 2) &= \frac{(n-1)}{2} \cdot \frac{p}{q} \cdot P(X = 1) = \frac{(n-1)}{2} \cdot \frac{p}{q} \cdot n \cdot p \cdot q^{n-1} \\ &= \binom{n}{2} \cdot p^2 \cdot q^{n-2}, \end{aligned}$$

and so on, giving the expected result that $P(X = x) = \binom{n}{x} p^x \cdot q^{n-x}$, $x = 0, 1, \dots, n$. So we can recover the probability distribution function from the recursion.

Recursions can be easily programmed and recursions such as (2.4) are also of some interest for theoretical purposes. For example, consider locating the maximum, or most frequently occurring value, of $P(X = x)$.

If we require that $P(X = x) \geq P(X = x - 1)$, then, from (2.4), $\frac{n-x+1}{x} \cdot \frac{p}{q} \geq 1$.

This reduces to $x \leq p \cdot (n + 1)$, so we can conclude that the value of X with the maximum probability is $X = \lfloor p \cdot (n + 1) \rfloor$ where $\lfloor x \rfloor$ denotes the largest integer in x .

The Mean and Variance of the Binomial

The recursion (2.4) can be used to determine the mean and variance of a binomial random variable. Consider first $\mu = \sum_{x=0}^n x \cdot P(X = x)$. Recursion (2.4) is

$$P(X = x) = \frac{n - x + 1}{x} \cdot \frac{p}{q} \cdot P(X = x - 1), x = 1, 2, \dots, n.$$

Multiplying through by x and summing from 1 to n gives

$$\sum_{x=1}^n x \cdot P(X = x) = \sum_{x=1}^n [n - (x - 1)] \cdot \frac{p}{q} \cdot P(X = x - 1), \text{ so}$$

$$\mu = \frac{p}{q} \cdot n \cdot [1 - P(X = n)] - \frac{p}{q} \cdot \sum_{x=1}^n (x - 1) \cdot P(X = x - 1) \text{ or}$$

$$\mu = \frac{p}{q} \cdot n \cdot (1 - p^n) - \frac{p}{q} \cdot [\mu - n \cdot P(X = n)] \text{ which reduces to}$$

$$\mu = n \cdot p.$$

This result makes a good deal of intuitive sense: if we toss a coin, loaded to come up heads with probability $3/4$, 1000 times, we expect $1000 \cdot \frac{3}{4} = 750$ heads. So in n trials of a binomial experiment with p as the probability of success, we expect $n \cdot p$ successes.

The variance can also be found using (2.4). We first calculate $E(X^2)$:

$$\begin{aligned} E(X^2) &= \sum_{x=1}^n x^2 \cdot P(X = x) = \sum_{x=1}^n x \cdot [n - (x - 1)] \cdot \frac{p}{q} \cdot P(X = x - 1), \\ &= n \cdot \frac{p}{q} \cdot \sum_{x=1}^n [(x - 1) + 1] \cdot P(X = x - 1) - \frac{p}{q} \cdot \sum_{x=1}^n x \cdot (x - 1) \cdot P(X = x - 1). \end{aligned}$$

Then since $\sum_{x=1}^n (x - 1) \cdot P(X = x - 1) = \mu - n \cdot P(X = n)$ and since

$$\sum_{x=1}^n x \cdot (x - 1) \cdot P(X = x - 1) = \sum_{x=1}^n [(x - 1)^2 + (x - 1)] \cdot P(X = x - 1),$$

it follows that

$$E(X^2) = p \cdot (n - 1) \cdot (np - np^n) + n \cdot p \cdot (1 - p^n) + n^2 \cdot p^{n+1},$$

and this reduces to $E(X^2) = np^2(n - 1) + np$. Therefore,

$$\sigma^2 = E(X^2) - [E(X)]^2 = np^2(n - 1) + np - (np)^2 = npq.$$

Example 2.5.1

We apply the above-mentioned results to a binomial experiment in which $p = q = 1/2$ and $n = 100$ trials. Here, $E(X) = \mu = n \cdot p = 50$ and $\sigma^2 = n \cdot p \cdot q = 25$. Tchebycheff's inequality with $k = 3$ then gives

$$P[n \cdot p - k \cdot \sqrt{n \cdot p \cdot q} \leq X \leq n \cdot p + k \cdot \sqrt{n \cdot p \cdot q}] \geq 1 - \frac{1}{k^2}$$

$$\text{so } P[50 - 3 \cdot 5 \leq X \leq 50 + 3 \cdot 5] \geq \frac{8}{9} \text{ or}$$

$$P[35 \leq X \leq 65] \geq \frac{8}{9}.$$

But we find exactly that

$$\sum_{x=35}^{65} \binom{100}{x} \cdot \left(\frac{1}{2}\right)^{100} = 0.99821,$$

verifying Tchebycheff's inequality in this case.

EXERCISES 2.5

1. Suppose a binomial random variable X assumes only the values 0 and 1 and $P(X = 1) = p$. Verify the mean and variance of X directly.
2. For a binomial random variable with probability p and $n = 5$, find all the probabilities for the probability distribution function and draw a graph of them.
3. A test is conducted to determine the concentration of a chemical in a lawn weed killer, which will effectively kill dandelions. It is found that a given concentration of the chemical will kill on average 80% of the dandelions in 24 hours. A test is performed on 20 dandelions. Find the probability that
 - (a) exactly 14 are killed in 24 hours.
 - (b) at least 10 are killed in 24 hours.
4. A fair die is rolled 240 times. Find the probability that the number of 2's or 3's is between 75 and 83, inclusive.
5. A manufacturer of dry cells actually makes two batteries that appear to be identical. Batteries of Type A last more than 600 hours with probability 0.30 and batteries of Type B last more than 600 hours with probability 0.40.
 - (a) What is the probability that 5 out of 10 of Type A batteries last more than 600 hours?
 - (b) Of 50 Type B batteries, how many are expected to last at least 600 hours?
 - (c) What is the probability that three Type A batteries have more batteries lasting 600 hours than two Type B batteries?
6. X and Y play the following game. X tosses 2 fair coins and Y tosses 3. The player throwing the greater number of heads wins. In case of a tie, the throws are repeated until a winner is determined.
 - (a) What is the probability that X wins on the first play?
 - (b) What is the probability that X wins the game?
7. In a political race, it is known that 40% of the voters favor candidate C. In a random sample of 100 voters, what is the probability that
 - (a) between 30 and 45 voters favor C?
 - (b) exactly 36 voters favor C?
8. A gambling game is played as follows. A player, who pays \$4 to play the game, tosses a fair coin five times. The player wins as many dollars as heads are tossed.
 - (a) Find the probability distribution for N , the player's net winnings.
 - (b) Find the mean and variance of the player's net winnings.
9. A red die is fair and a green die is loaded so that the probability it comes up 6 is 1/10.
 - (a) What is the probability of rolling exactly 3 sixes in 3 rolls with the red die?
 - (b) What is the probability of at least 30 sixes in 100 rolls of the red die?
 - (c) The green die is thrown five times and the red die is thrown four times. Find the probability that a total of 3 sixes occurs.
10. What is the probability of one head twice in three tosses of four fair coins?
11. A commuter's drive to work includes seven stoplights. Assume the probability a light is red when the commuter reaches it is 0.20 and that the lights are far enough apart to operate independently.

- (a) If X is the number of red lights the commuter stops for, find the probability distribution function for X .
 (b) Find $P(X \geq 5)$.
 (c) Find $P(X \geq 5|X \geq 3)$.
12. The probability of being able to log on a computer system from a remote terminal during a busy period is 0.7. Suppose that 10 independent attempts are made and that X denotes the number of successful attempts.
 (a) Write an expression for the probability distribution function, $f(x)$.
 (b) Find $P(X \geq 5)$.
 (c) Now suppose that Y represents the number of attempts up to and including the first successful attempt. Write an expression for the probability distribution function, $g(y)$.
13. An experimental rocket is launched five times. The probability of a successful launch is 0.9. Let X denote the number of successful launches. A study has shown that the net cost of the experiment, in thousands of dollars, is $2 - 3X^2$. Find the expected net cost of the experiment.
14. Twenty percent of the integrated circuit (IC) chips made in a plant are defective. Assume that a binomial model is appropriate.
 (a) Find the probability that at most 13 defective chips occur in a sample of 100.
 (b) Find the probability that two samples, each of size 100, will have a total of exactly 26 defective chips.
15. A coin, loaded to come up heads with probability $2/3$, is tossed five times. If the number of heads is odd, the player is paid \$5. If the number of heads is 2 or 4 the player wins nothing; if no heads occur, the player tosses the coin five more times and wins, in dollars, the number of heads thrown. If the game costs the player \$3 to play, find the probability distribution of N , his net winnings.
16. (a) Show that the probability of being dealt a full house (3 cards of one value and 2 cards of another value) in poker is about 0.0014.
 (b) Find the probability that in 1000 hands of poker, you will be dealt at least 2 full houses.
17. An airline knows that 10% of the people holding reservations on a given flight will not appear. The plane holds 90 people.
 (a) If 95 reservations have been sold, find the probability that the airline will be able to accommodate everyone appearing for the flight.
 (b) How many reservations should be sold so that the airline can accommodate everyone who appears for the flight 99% of the time?
18. The probability an individual seed of a certain type will germinate is 0.9. A nurseryman sells flats of this type of plant and wants to “guarantee” (with probability 0.99) that at least 100 plants in the flat will germinate. How many plants should he put in each flat?
19. A coin with $P(H) = 1/2$ is flipped four times, and then a coin with $P(H) = 2/3$ is tossed twice. What is the probability that a total of five heads occurs?
20. (a) Each of two persons tosses three fair coins. What is the probability that each gets the same number of heads?

- (b) In part (a), what is the probability that $X_1 + X_2$ is odd, where X_1 is the number of heads the first person tosses and X_2 is the number of tosses the second person tosses?
- (c) Repeat part (a) if each person tosses n fair coins. Simplify the result as much as possible.
21. Find the probability that more than 520 heads occur in 1000 tosses of a fair coin.
22. How many times must a fair coin be tossed if the probability of obtaining at least 40 heads is at least 0.95?
23. Samples of 100 are selected each hour from a production process that produces items, 20% of which are defective.
- (a) What is the probability that at most 15 defectives are found in an hour?
- (b) What is the probability that a total of 47 defectives is found in the first 2 hours?
24. A small engineering college would like to have an entering class of 360 students. Past data indicate that 85% of those accepted actually enroll in the class. How many students should be accepted if the probability the class will be at least 360 is to be approximately 0.95?
25. A fair coin is tossed repeatedly. What is the probability the number of heads tossed reaches 6 before the number of tails tossed reaches 4?
26. Evaluate the sums

$$\sum_{x=0}^n x \cdot \binom{n}{x} p^x \cdot q^{n-x} \text{ and } \sum_{x=0}^n x \cdot (x-1) \cdot \binom{n}{x} p^x \cdot q^{n-x}$$

directly and use these to verify the formulas for μ and σ^2 for the binomial distribution.

[Note that $\sum_{x=0}^n \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = [p + (1-p)]^n = 1$.]

27. In problem 6, show that the game is fair if X wins if he tosses at least as many heads as Y .

2.6 SOME STATISTICAL CONSIDERATIONS

We pause here and in the next section to show some statistical applications of the probability theory we have developed so far. From time to time in this book, we will show some applications of probability theory to statistics and the statistical analysis of data as well as to other applied situations; this is our first consideration of statistical problems.

From the previous section, we know what can happen when n observations are taken from a binomial distribution with known parameter p . Generally, however, p is unknown. We might, for example, be interested in the proportion of unacceptable items arising from a production line. Usually, this proportion would not be known. So we suppose now that p is unknown. How can we *estimate* the unknown p ? We certainly would observe the binomial process the production line represents; the result of this would be a number of good items from the process, say X , and we would surely use X in some way to estimate p . How precisely can we use X to estimate p ?

It would appear natural to estimate p by the proportion of good items in the sample, $\frac{X}{n}$. Since X is a random variable, so is $\frac{X}{n}$. We can calculate the expected value of this random variable as follows:

$$E\left[\frac{X}{n}\right] = \sum_{x=0}^n \frac{x}{n} \cdot P(X=x) = \frac{1}{n} \cdot \sum_{x=0}^n x \cdot P(X=x)$$

so $E\left[\frac{X}{n}\right] = \frac{1}{n} \cdot n \cdot p = p.$

This indicates that, on average, our estimate for p gives the true value, p . We say that our estimator, $\frac{X}{n}$, is an *unbiased estimator* for p .

This gives us a way of estimating p by a single value. This single value is dependent on the sample and if we choose another sample, we are likely to find another value of X , and hence arrive at another estimate of p . Could we also find a “likely” range for the value of p ?

To answer this, consider a related question. If we have a binomial situation with probability p and sample size n , what is a “likely” range for the observed values of the random variable, X ? The answer of course depends on the meaning of the word “likely.” Suppose that a likely range for the values of a random variable is a range in which the values of the variable occur with probability 0.95.

With the considerable aid of our computer algebra system, we can consider a number of different binomial distributions. We vary n , the number of observations, and p , the probability of success. In each case, we find the proportion of the values of X that lie within two standard deviations of the mean, that is, the proportion of the values of X that lie in the interval $\mu \pm 2\sigma = n \cdot p \pm 2\sqrt{n \cdot p \cdot (1-p)}$. We selected the constant 2 because we need to find a range that includes a large portion—95%—of the values of X and 2 would appear to be a reasonable multiplier for the standard deviation. Table 2.1 shows the results of these

Table 2.1 Exact probability of binomial intervals around the mean for various values of n and p

n	p	σ	$\mu \pm 2\sigma$	P
36	1/2	3	12, 24	0.971183
64	1/2	4	24, 40	0.967234
100	1/2	5	40, 60	0.964800
144	1/2	6	60, 84	0.963148
196	1/2	7	84, 112	0.961530
18	1/3	2	2, 10	0.978800
72	1/3	4	16, 32	0.967288
162	1/3	6	42, 66	0.963177
288	1/3	8	80, 112	0.961066
48	1/4	3	6, 18	0.971345
192	1/4	6	36, 60	0.963214
432	1/4	9	90, 126	0.960373
10000	1/2	50	4900, 5100	0.954494
11250	1/3	50	3650, 3850	0.954497
13872	1/4	51	3366, 3570	0.954499

calculations. Here, P represents the probability an observed value of the random variable X lies in the interval $\mu \pm 2\sigma = n \cdot p \pm 2\sqrt{n \cdot p \cdot (1-p)}$. The values of n and p have been chosen so that the end points of the intervals are integers.

We are led to believe from the table, regardless of the value of p , that at least 95% of the values of the variable X lie in the interval $\mu \pm 2\sigma$. (Later we will show, for large values of n , regardless of the value of p , that the probability is approximately 0.9545, a result supported by our calculations.) So we have

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \geq 0.95. \quad (2.5)$$

Solving the inequalities for μ , we have

$$P(X - 2\sigma \leq \mu \leq X + 2\sigma) \geq 0.95. \quad (2.6)$$

Replacing μ and σ by $n \cdot p$ and $\sqrt{n \cdot p \cdot q}$, respectively, (2.6) becomes

$$P(X - 2\sqrt{n \cdot p \cdot q} \leq n \cdot p \leq X + 2\sqrt{n \cdot p \cdot q}) \geq 0.95. \quad (2.7)$$

The inequalities in (2.7) can now be solved for p . The result is

$$\begin{aligned} P\left(\frac{nX + 2n - 2\sqrt{n^2X + n^2 - nX^2}}{n^2 + 4n} \leq p \leq \frac{nX + 2n + 2\sqrt{n^2X + n^2 - nX^2}}{n^2 + 4n}\right) \\ \geq 0.95 \end{aligned} \quad (2.8)$$

Our thinking here is as follows: if we find an interval that contains at least 95% of the values of X and if p is unknown, then those same values of X will produce an interval in which p , in some sense, is likely to lie. The end points produced by formula (2.8) comprise what we call a *95% confidence interval for p* .

While (2.5) gives a likely range of values of X if p is known, (2.8) gives a likely range of values of p if X is known. So we have a response to a variant of our first question: If X successes are observed in n binomial trials, what is a likely value for p ?

We note that (2.5) is a legitimate probability statement since X is a random variable and 95% of its values lie in the stated interval. However, (2.8) is *not* a probability statement! Why not? The reason is that p is an unknown constant. Either it lies in the stated interval or it does not. Then, what does the 95% mean?

Here is a way of looking at this. Consider samples of fixed size, say $n = 100$. If we find 25 successes in these 100 trials (where $X = 25$), then (2.8) gives the interval $0.174152 \leq p \leq 0.345079$. However, the next time we perform the experiment, we are most likely to find another value of X , and hence another confidence interval. For example, if $X = 30$, the confidence interval is $0.217492 \leq p \leq 0.397893$. From (2.8), we see that these confidence intervals are centered about the value $\frac{X+2}{n+4}$ and have width $\frac{4\sqrt{n^2X+n^2-nX^2}}{n^2+4n}$, so both the center and width of the intervals change as X changes for a fixed sample size n . This gives us a proper interpretation of (2.8): 95% of these intervals will contain the unknown, and fixed, value p .

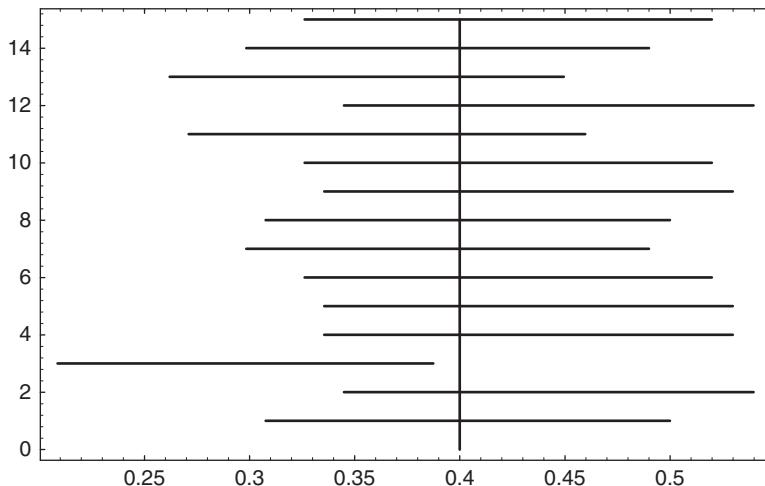


Figure 2.11 Some confidence intervals.

As another example, 15 observations were taken from a binomial distribution with $n = 100$ and gave the following values for X : 40, 44, 29, 43, 43, 42, 39, 40, 43, 42, 36, 44, 35, 39, and 42. Formula (2.8) was then used to compute a confidence interval for p for each of these values of X .

Figure 2.11 shows these confidence intervals. As expected, they vary in both position and width. The actual value of p used to generate the X values was 0.40. As it happens here, $p = 0.40$ is contained in 14 of the 15 confidence intervals, but in larger samples we would expect that 0.40 would be contained in about 95% of the confidence intervals produced.

EXERCISES 2.6

1. In the above-mentioned text, we drew 15 observations from a binomial distribution with $n = 100$. Calculate the end points of a 95% confidence interval for $X = 40$ as shown in Figure 2.11.
2. If ten 95% confidence intervals for an unknown binomial p are calculated for samples of size 50, what is the probability that p is contained in exactly 6 of them?
3. If a sample of size 30 is chosen from a binomial distribution with $p = 1/2$, and if X denotes the number of successes obtained, find an interval in which 95% of the values of X will lie.
4. Use your computer algebra system to verify the results in Table 2.1 for
 - (a) $p = 1/2$, $n = 36$
 - (b) $p = 1/3$, $n = 18$
 - (c) $p = 1/4$, $n = 48$

5. Use your computer algebra system to verify the result in Table 2.1 for
 - (a) $p = 1/2, n = 10000$
 - (b) $p = 1/3, n = 11250$
 - (c) $p = 1/4, n = 13872$
6. A survey of 300 college students found that 50 are thinking about changing their majors. Find a 95% confidence interval for the true proportion of college students thinking about changing their majors.
7. A random sample of 1250 voters was asked whether or not they voted in favor of a school bond issue. Out of which, 325 replied that they favored the issue. Find a 95% confidence interval for the true proportion of voters who favor the school bond issue.
8. Find 90% confidence intervals by constructing a table similar to Table 2.1. One should find that $P(\mu - 1.645\sigma \leq X \leq \mu + 1.645\sigma) = 0.90$.
9. A newspaper survey of 125 of its subscribers found that 40% of the respondents knew someone who was killed or injured by a drunk driver. Find a 90% confidence interval for the true proportion of people in the population who know someone who was killed or injured by a drunk driver.
10. As a project in a probability course, a student discovered that among a random sample of 80 families, 25% did not have checking accounts. Use this information to construct a 90% confidence interval for the true proportion of families in the population who do not have checking accounts.
11. A study showed that 1/8th of American workers worked in management or in administration, while 1/27th of Japanese workers worked in management or administration. The study was based on 496 American workers and 810 Japanese workers.

Is it possible that the same proportion of American and Japanese workers are in management or administration and that the apparent differences found by the study are simply due to the variation inherent in sampling? [Hint: Compare 90% confidence intervals.]
12. n values of X , the number of successes in a binomial process, are used to compute n 95% confidence intervals for the unknown parameter p . Find the probability that p lies in exactly k of the n confidence intervals.

2.7 HYPOTHESIS TESTING: BINOMIAL RANDOM VARIABLES

In the previous section, we considered confidence intervals for binomial random variables. The problem of estimating a parameter, in this case the value of p by means of an interval, is part of *statistics* or *statistical inference*. Statistical inference, in simplest terms, is concerned with drawing inferences from data that have been gathered by a sampling process. Statistical inference comprises the theory of *estimation* and that of *hypothesis testing*. In the preceding section, we considered the construction of a confidence interval that is part of the theory of estimation. The remaining portion of the theory of drawing inferences from samples is called *hypothesis testing*. We begin with a somewhat artificial example in order to fix ideas and define some vocabulary before proceeding to other applications.

Example 2.7.1

The manufacturing process of a sensitive component has been producing items of which 20% must be reworked before they can be used. A recent sample of 20 items shows 6 items that must be reworked. Has the manufacturing process changed so that 30% of the items must be reworked?

Assume that the production process is binomial, with p , which is of course unknown to us, denoting the probability an item must be reworked. We begin with a *hypothesis* or conjecture about the binomial process, which has not in fact changed, and that the proportion of items that must be reworked is 20%. We denote this by H_0 and we call this the *null hypothesis*. As a result of a *test* – in this case the result of a sample of the items – this hypothesis will be *accepted* (i.e., we will believe that H_0 is true) or it will be *rejected* (i.e., we will believe that H_0 is not true). In the latter case, when the null hypothesis is rejected, we agree to accept an *alternative hypothesis*, H_a . Here, the hypotheses are chosen as follows:

$$H_0: p = 0.20$$

$$H_a: p = 0.30.$$

How are sample results (in this case 6 items that must be reworked) to be interpreted? Does this information lead to the acceptance or the rejection of H_0 ? We must then decide what sample results lead to the acceptance of H_0 and what sample results lead to its rejection (and hence the acceptance of H_a).

The sampling is, of course, subject to variability and our conclusions cannot be reached without running the risk of error. There are two risks: that we will reject H_0 even though it is, in reality, true, or, we accept H_0 even though it is, in reality, false. The following table may help in seeing the four possibilities that exist whenever a hypothesis is tested:

		Reality	
		H_0 True	H_0 False
H_0 Rejected	Type I error (α)	Correct decision	
	Correct decision	Type II error (β)	
H_0 Accepted			

We never will know reality, but the table does indicate the consequences of the decision process. It is customary to denote the two types of errors by

$$\alpha = \text{Probability of a Type I error}$$

$$= P[H_0 \text{ is rejected when it is true}]$$

$$\beta = \text{Probability of a Type II error}$$

$$= P[H_0 \text{ is accepted when it is false}].$$

Both α and β are conditional probabilities, and each is highly dependent on the set of sample values that lead to the rejection of the hypothesis. This set of values is called the *critical region*.

What should the critical region be? We are free to choose any critical region we want; it would appear sensible in this case to conclude that the percentage of product to be reworked had increased when the number of items to be reworked in the sample is large. Therefore, we

arbitrarily take as a critical region $\{x|x \geq 9\}$, where X is the random variable denoting the number of items in the sample that must be reworked.

What are the consequences of this choice for the critical region? We can calculate α , the size of the Type I error:

$$\begin{aligned}\alpha &= P[X \geq 9 \text{ if } H_0 \text{ is true}] \\ &= P[X \geq 9 \text{ if } p = 0.2] \\ &= \sum_{x=9}^{20} \binom{20}{x} (0.2)^x (0.8)^{20-x} \\ &= 0.00998179 \approx 0.01.\end{aligned}$$

So about 1% of the time, this critical region will reject a true hypothesis. This means that the manufacturing process is such that if $p = 0.20$, about 1% of the time it will behave as if $p = 0.30$ with this critical region. α is called the *size* or the *significance level* of the test.

What is β ?

$$\begin{aligned}\beta &= P[\text{accept } H_0 \text{ if it is false}] \\ &= P[X < 9 \text{ if } H_0 \text{ is false}] \\ &= P[X < 9 \text{ if } p = 0.30] \\ &= \sum_{x=0}^8 \binom{20}{x} (0.30)^x (0.70)^{20-x} \\ &= 0.886669.\end{aligned}$$

These calculations are shown in Appendix A.

So, with this critical region, about 89% of the time a process producing 30% items to be reworked behaves as if it were producing only 20% of such items. This might appear to be a very high risk. Can it be reduced? One way to reduce β would be to change the critical region to say $\{x|x \geq 8\}$. We now find that

$$\begin{aligned}\beta &= \sum_{x=0}^7 \binom{20}{x} (0.30)^x (0.70)^{20-x} = 0.772272 \\ \text{but then } \alpha &= \sum_{x=8}^{20} \binom{20}{x} (0.20)^x (0.80)^{20-x} = 0.032147.\end{aligned}$$

So the cost in decreasing β comes at the cost of an increase in α . We will see later than one way to decrease both errors is to increase the sample size.

What are the consequences of other choices for the critical region? We could choose $x = 0$ for the critical region so that the hypothesis is rejected only if $x = 0$. Then

$$\begin{aligned}\alpha &= P[X = 0 \text{ if } p = 0.20] \\ &= (0.8)^{20} \\ &= 0.0115292,\end{aligned}$$

producing a Type I error of about the same size as it was before. But then

$$\begin{aligned}\beta &= \sum_{x=1}^{20} \binom{20}{x} (0.30)^x (0.70)^{20-x} \\ &= 0.999202.\end{aligned}$$

These two critical regions then have roughly equal Type I errors, but β is larger for the second choice of critical region.

We will choose one more critical region whose Type I error is about 0.01: the critical region $X = 9, 10, \text{ or } 11$. Then

$$\begin{aligned}\alpha &= P[X = 9, 10, \text{ or } 11 \text{ if } p = 0.20] \\ \alpha &= \sum_{x=9}^{11} \binom{20}{x} (0.20)^x (0.80)^{20-x} \\ &= 0.00998,\end{aligned}$$

again roughly 0.01.

β , however, is now

$$\begin{aligned}\beta &= 1 - \sum_{x=9}^{11} \binom{20}{x} (0.30)^x (0.70)^{20-x} \\ &= 0.891807.\end{aligned}$$

The earlier four cases illustrate that there are several choices for critical regions that give the same size for Type I error; we will call the critical region *best*, if for a given Type I error, it *minimizes* Type II error. In this case, the best critical region for a test with $\alpha \approx 0.01$ is $\{x|x \geq 9\}$. Best critical regions can often, but not always, be constructed.

So, to return to the original problem where the sample yielded six items for reworking, we conclude that the process has not changed since it is not in the critical region $\{x|x \geq 9\}$ for $\alpha \approx 0.01$.

Finally, we note that the size of Type II error, β , is a function of the alternative, $p = 0.30$, in this example. If the alternative hypothesis were $H_a: p > .20$, then β could be calculated for any *particular* alternative in H_a . That is, if $p > 0.20$, then

$$\beta = \sum_{x=0}^8 \binom{20}{x} p^x (1-p)^{20-x}, \text{ a function of } p.$$

As p increases, β decreases quite rapidly, reflecting the fact that it is increasingly unlikely that the hypothesis will be accepted if it is false. A graph of β as a function of p is shown in Figure 2.12.

It is customary to graph $1 - \beta = P[\text{a false } H_0 \text{ is rejected}]$. This is called the *power function* for the test.

The hypothesis $H_0 : p = 0.20$ is called a *simple hypothesis* since it completely specifies the probability distribution of the variable under consideration. The hypothesis $H_a: p > 0.20$ is composed of an infinity of simple hypotheses. It is called a *composite hypothesis*.

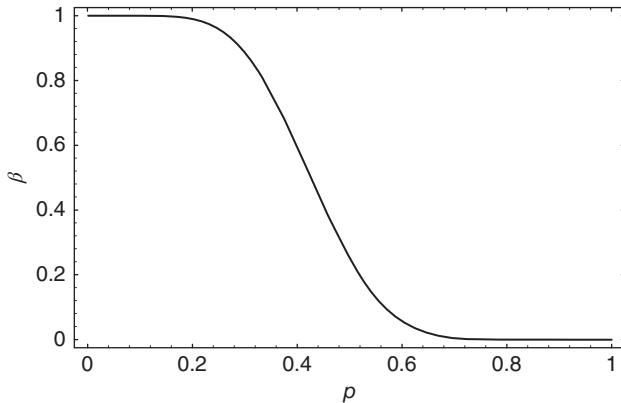


Figure 2.12 β as a function of p for Example 2.7.1.

Example 2.7.2

In the previous example, the critical region was specified and then values for α and β were found. It is common, however, for experimenters to specify α and β before the experiment is done; often the sample size necessary to achieve these probabilities can be found, at least approximately. One of the consequences of the binomial model in the preceding example is that a change in the critical region by a single unit produces large changes in α and β . Suppose, in the preceding example, that it is desired to have, approximately, $\alpha = 0.05$ and $\beta = 0.10$. If we assume that the best critical region is of the form $\{x|x \geq k\}$, then

$$\alpha = \sum_{x=k}^n \binom{n}{x} (0.20)^x (0.80)^{n-x} = 0.05$$

and

$$\beta = \sum_{x=0}^{k-1} \binom{n}{x} (0.30)^x (0.70)^{n-x} = 0.10,$$

These equations are difficult to solve without the aid of extensive binomial tables or a computer algebra system. We find that

$$\alpha = \sum_{x=40}^{156} \binom{156}{x} (0.20)^x (0.80)^{156-x} = 0.05145$$

$$\text{and } \beta = \sum_{x=0}^{39} \binom{156}{x} (0.30)^x (0.70)^{156-x} = 0.09962,$$

so $n \approx 156$ and $k \approx 40$. These values are probably close enough for all practical purposes. Other solutions are of course possible, depending on the closeness with which we want to solve the equations for α and β . It may well be that we cannot carry out an experiment with this large sample size; such a restriction would obviously then have implications for the sizes of α and β that can be entertained.

EXERCISES 2.7

1. A manufacturer of a new electronic tablet device wants to determine the proportion of current tablet users who would purchase a new version of the tablet. The manufacturer thinks that 15% of current users would purchase the new tablet. For a test, the hypotheses are as follows:

$$H_0: p = 0.15$$

$$H_a: p > 0.15.$$

- (a) Find α if the critical region is $X > 30$ for a sample of 150 tablet users.
 (b) Find β for $H_a: p = 0.25$.
2. A new car dealer tests customers who will pay \$1000 down for free financing for 2 years. A sample of 20 buyers is taken; X is the number of customers who will take the financing deal. The hypotheses are as follows:

$$H_0: p = 0.40$$

$$H_a: p > 0.40.$$

- (a) Find α if the critical region is $X < 8$.
 (b) Find β for the alternative $p = 0.50$.
3. It is thought that 80% of VCR owners do not know how to program their VCR for taping a TV program. To test this hypothesis, a sample of 20 VCR owners is chosen and the proportion, p , who can program a VCR is recorded. The hypotheses are as follows:

$$H_0: p = 0.80$$

$$H_a: p < 0.80.$$

- (a) Find α if the critical region is $X < 14$ where X is the number in the sample who cannot program a VCR.
 (b) Find β for the alternative $H_a: p = 0.70$.
 (c) Graph β as a function of p , $0 \leq p \leq 0.80$.
4. A researcher speculates that 20% of the people in a very large group under study is left-handed, a proportion much larger than the 10% of people who are left-handed in the population. A sample is chosen to test

$$H_0: p = 0.10$$

$$H_a: p = 0.20.$$

The critical region is $X \geq k$, where X is the number of left-handed people in the sample. It is desired to have $\alpha = 0.07$ and $\beta = 0.13$, approximately. How large a sample should be chosen?

5. In exercise 4, show that β is larger for the critical region $X \leq c$ where c is chosen so that the test has size α .

6. A drug is thought to cure 2/3 of the patients with a disease; without the drug, 1/3 of the patients recover. The hypothesis

$$H_0: p = 1/3 \text{ is tested against}$$

$$H_a: p = 2/3$$

on the basis of a sample of 12 patients. H_0 is rejected if X , the number of patients in the sample who recover, is greater than 5. Find α and β for this test.

7. In exercise 6, find the sample size for which $\alpha = 0.05$ and $\beta = 0.13$, approximately.
8. A recent survey showed that 46% of Americans feel that they are “being left behind by technology.” To test this hypothesis, a sample of 36 Americans showed that 18 of them agreed that they were being left behind by technology. Does the data support the hypothesis $H_0: p = 0.46$ against the alternative $H_a: p > 0.46$? (use $\alpha = 0.05$.)
9. A publisher thinks that 57% of the magazines on newsstands are unsold. To test this hypothesis, a sample of 1000 magazines put on the newsstand resulted in 495 unsold magazines. Does this data support $H_0: p = 0.57$ or the alternative $H_a: p < 0.57$ if $\alpha = 0.05$?
10. A survey indicates that 41% of the people interviewed think that holders of Ph.D. degrees have attended medical school. In a sample of 88 people, 50 agreed that Ph.D.’s attended medical school. Is this evidence, using $\alpha = 0.05$, that the percentage of people thinking that Ph.D.’s are M.D.’s is greater than 41%?
11. In a survey of questions concerning health issues, 59% of the respondents thought that at some time in their life they would develop cancer. If a sample of 200 people showed that 89 agreed that they would develop cancer at some time, is this evidence to support the hypothesis that the percentage thinking they will develop cancer is less than 59% (use $\alpha = 0.05$).
12. Among Americans earning more than \$50,000 per year, 2/3 people agree that Americans are “materialistic.” If 70 people out of 100 people interviewed agree that Americans are materialistic, is this evidence that the true proportion thinking Americans are materialistic is greater than 2/3 (use $\alpha = 0.05$).

2.8 DISTRIBUTION OF A SAMPLE PROPORTION

Before considering some important probability distributions in addition to the binomial distribution, we consider here a common problem: a sample survey of n individuals indicates that the proportion p_s of the respondents favors a certain candidate in an election. p_s is clearly a random variable since our sampling will not always produce exactly the same proportion of voters favoring the candidate if the sampling is repeated, p_s is called a *sample proportion*. What is its probability distribution? How can we expect p_s to vary from sample to sample? If we observe a value of p_s – say 51% of the voters favor a candidate – what does this tell us about the true proportion of all voters who favor the candidate, say p ? We consider these questions now.

Let us suppose that in reality the proportion p of the voters favor a candidate. Let us also assume that the sample is taken so that the responses can be assumed to be independent among the people interviewed. The *number* of voters favoring the candidate, say X , is then

a binomial random variable since a voter favors either the candidate or the opponent. The sample proportion favoring the candidate, P_s , is also a random variable. If we take a random sample of size n , then

$$P_s = \frac{X}{n}.$$

So our random variable P_s is related to a binomial random variable. We considered confidence intervals for binomial random variables in Section 2.6. We now extend that theory somewhat.

We now calculate the mean and variance of the variable P_s . We let the sample proportion be $p_s = \frac{x}{n}$. Clearly,

$$P(P_s = p_s) = P\left(\frac{X}{n} = p_s\right) = P(X = n \cdot p_s) = P(X = x)$$

so

$$\begin{aligned} E(P_s) &= \sum_{p_s=0}^1 p_s \cdot P(P_s = p_s) \\ &= \sum_{x=0}^n \frac{x}{n} \cdot P\left(\frac{X}{n} = p_s\right) \\ &= \frac{1}{n} \sum_{x=0}^n x \cdot P(X = x) \\ \text{Therefore, } E(P_s) &= \frac{1}{n} \cdot E(X) = \frac{n \cdot p}{n} = p. \end{aligned}$$

So, as might be expected, the average value of the variable P_s is the true proportion, p . This is precisely the same result we saw in Section 2.6.

The variance of P_s can be calculated using the variance of a binomial random variable as follows:

$$\begin{aligned} \text{Var}(P_s) &= \text{Var}\left(\frac{X}{n}\right) = \sum_{p_s=0}^1 (P_s - p)^2 \cdot P(P_s = p_s) \\ &= \sum_{x=0}^n \left(\frac{x}{n} - p\right)^2 \cdot P\left(\frac{X}{n} = p_s\right) \\ &= \frac{1}{n^2} \sum_{x=0}^n (x - n \cdot p)^2 \cdot P(X = n \cdot p_s) \\ &= \frac{1}{n^2} \sum_{x=0}^n (x - n \cdot p)^2 \cdot P(X = x) \end{aligned}$$

showing that

$$\text{Var}(P_s) = \frac{1}{n^2} \cdot \text{Var}(X)$$

or that

$$\text{Var}(P_s) = \frac{1}{n^2} \cdot n \cdot p \cdot q = \frac{p \cdot q}{n}.$$

The earlier considerations also show a more general result: if random variables X and Y are related by $Y = k \cdot X$, where k is a constant, then $E(Y) = k \cdot E(X)$ and $\text{Var}(Y) = k^2 \cdot \text{Var}(X)$.

Using the facts we derived in Section 2.6 regarding binomial confidence intervals, we can say that

$$P\left(p_s - 2 \cdot \sqrt{\frac{p \cdot q}{n}} \leq p \leq p_s + 2 \cdot \sqrt{\frac{p \cdot q}{n}}\right) \geq 0.95$$

giving a 95% confidence interval for the true population proportion, p . But, as occurred in the binomial situation, the standard deviation is a function of the unknown p , so we must solve for p . There are two ways to do this. One method is to solve the quadratic equations that arise exactly. However, if $0.3 \leq p \leq 0.7$, then a good approximation to $p \cdot q$ is $\frac{1}{4}$. This approximation is far from exact, but often yields acceptable results when p is in the indicated range.

Example 2.8.1

A sample survey of 400 voters showed that 51% of the voters favored a certain candidate. Find a 95% confidence interval for p , the true proportion of voters in the population favoring the candidate.

We have that

$$P\left(0.51 - 2 \cdot \sqrt{\frac{p \cdot q}{400}} \leq p \leq 0.51 + 2 \cdot \sqrt{\frac{p \cdot q}{400}}\right) \geq 0.95.$$

If we solve the inequalities for p , noting that $q = 1 - p$, we find that

$$\frac{n \cdot p_s + 2 - 2\sqrt{1 + n \cdot p_s - n \cdot p_s^2}}{n + 4} \leq p \leq \frac{n \cdot p_s + 2 + 2\sqrt{1 + n \cdot p_s - n \cdot p_s^2}}{n + 4}.$$

This result is equivalent to formula (2.8) in Section 2.6.

Substituting $n = 400$ and $p_s = 0.51$ gives $P(0.46016 \leq p \leq 0.55964) \geq 0.95$ while using the approximation $p \cdot q = 1/4$ gives $P(0.46 \leq p \leq 0.56) \geq 0.95$.

The difference in the confidence intervals is very small, but this is because the observed proportion, 0.51, is close to 1/2. The two confidence intervals will deviate more markedly as the difference between p_s and 1/2 increases. The candidate certainly cannot feel confident of winning the election on the basis of the sample, but we can only make this observation since we have created a confidence interval for p . In the popular press, half the width of the confidence interval is referred to as the *sampling error*. So a survey may be reported with a sampling error of 3% meaning that a 95% confidence interval for p is $p_s \pm 0.03$.

If the sampling error is given, then the sample size can be inferred. If the sampling error is stated as 3%, then

$$2 \cdot \sqrt{\frac{p \cdot q}{n}} = 0.03$$

where, of course, the difficulty is that p is unknown. Note that $p \cdot q \approx 1/4$ if $0.3 \leq p \leq 0.7$. Using this approximation here, we conclude that $\sqrt{\frac{1}{n}} \approx 0.03$ so that $n \approx 1111$.

The approximation $p \cdot q = 1/4$ is usually used only if p is in the interval $0.3 \leq p \leq 0.7$; otherwise p is replaced by the sample proportion, p_s , in determining sample size.

We presumed earlier here that the sample of voters is a simple random one, and we further presumed that the people sampled will actually vote and that they have been candid with the interviewer concerning their voting preference. Samplers commonly call these presumptions into question and have a variety of ways of dealing with them. In addition, such samples are rarely simple random samples; all we can say here is that these variations in the sampling design have some effect on the sampling error.

EXERCISES 2.8

1. A random sample of 200 automobile registrations shows that 22 are Subarus. Find a 95% confidence interval for the true proportion of Subaru registrations.
2. Compare the result in exercise 1 by estimating $p = 1/4$.
3. A survey of 300 paperback novels showed that 47% could be classified as romance novels. Find an approximate 95% confidence interval for p , the true proportion of romance paperback novels.
4. Records indicate that 1/8 of American children receive welfare payments. If this survey was based on 250 records, find an approximate 95% confidence interval for the true proportion of children who receive welfare payments.
5. A random sample of 300 voters showed that 48% favored a candidate. Does an approximate 95% confidence interval indicate that it is possible for the candidate to win the election?
6. A survey of 423 workers found that 1/9 were union members. Find an approximate 95% confidence interval for the true proportion of union workers.
7. The sampling error of a survey in a magazine was stated to be 5%. What was the sample size for the survey?
8. A student conducted a project for a statistics course and found that 2/3 of the respondents in interviews of 120 people did not know that the Bill of Rights is the first ten amendments to the Constitution. Find an approximate 90% confidence interval for the true proportion of people who do not know that the Bill of Rights is the first ten amendments to the Constitution.
9. A magazine devoted to health issues discovered that 3/5 of the time a visit to a physician resulted in a prescription. The survey was based on 130 telephone interviews. Use this data to construct an approximate 90% confidence interval for the true proportion of patients, given a prescription as a result of a visit to their physician.
10. According to a recent study, 81% of college students say that they favor drug testing in the workplace. The study was conducted among 400 college students. Find an approximate 90% confidence interval for the true proportion of college students who favor drug testing in the workplace.

11. Interviews of 150 patients recently tested for the HIV virus indicate that among those whose tests indicate the presence of the virus, 1/2 did not know they had the virus prior to testing. Find an approximate 95% confidence interval for the proportion of people in the population whose tests indicate they have the HIV virus and who did not know this.
12. A California automobile dealer knows that 1/10 of California residents own convertibles. Is the dealer likely (with probability 0.95) to sell at least 200 convertibles in the next 1000 sales?

2.9 GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

We considered *geometric* random variables in Examples 2.3.5 and 2.3.8 where the random variable of interest was the waiting time for the occurrence of a binomial event. A perfect model for the geometric random variable is tossing a coin, loaded so that the probability of coming up heads is p , until heads appear. If X denotes the number of tosses necessary and if $q = 1 - p$, we have seen that

$$P(X = x) = q^{x-1} \cdot p, \quad x = 1, 2, 3, \dots$$

and that

$$E(X) = \frac{1}{p} \text{ and } \text{Var}(X) = \frac{q}{p^2}.$$

Now suppose we wait until the *second* head appears when the loaded coin is tossed. Let X denote the number of trials necessary for this event to occur. We want $P(X = x)$, the probability distribution for X . Since the last trial must be heads, the first $x - 1$ trials must contain exactly one head and $x - 2$ tails; since the trials are independent, and since the single head can occur in any of $x - 1$ places, it follows that

$$P(\text{first } x - 1 \text{ trials have exactly 1 heads and } x - 2 \text{ tails}) = \binom{x-1}{1} \cdot q^{x-2} \cdot p$$

So, since the last trial must be heads,

$$P(X = x) = \binom{x-1}{1} \cdot q^{x-2} \cdot p \cdot p, \quad x = 2, 3, 4, \dots \quad (2.9)$$

Since formula (2.9) exhausts the possibilities, it must be that $\sum_{x=2}^{\infty} P(X = x) = 1$. One way to verify this is to notice that

$$\sum_{x=2}^{\infty} \binom{x-1}{1} \cdot q^{x-2} \cdot p^2 = p^2 \sum_{x=2}^{\infty} \binom{x-1}{1} \cdot q^{x-2} = p^2 \cdot (1-q)^{-2} = p^2 \cdot p^{-2} = 1$$

by the binomial theorem with a negative exponent. This series will arise again in our work. We have established the probability distribution for the waiting time for the second head.

What is the average waiting time for the second head? We might reason as follows: we flip the coin until the first head appears; the average number of flips is $1/p$. But then the situation is exactly the same as it was for the first flip of the coin; the fact that we flipped the coin and waited for the first head has absolutely no influence on subsequent tosses of

the coin. We must wait an average of $1/p$ flips again until the second head appears. So the average waiting time for the second head to appear is $\frac{1}{p} + \frac{1}{p} = \frac{2}{p}$. It follows that if we were to wait for the r th head to appear, the average total waiting time would be r/p . We will give a more formal derivation of this result later.

What is the probability distribution function for the r th head to appear? Let X denote the number of tosses until the r th head appears. Since, again, the last toss must be heads and the first $x - 1$ tosses must contain exactly $r - 1$ heads:

$$P(X = x) = \binom{x-1}{r-1} \cdot p^{r-1} \cdot q^{x-r} \cdot p, \quad x = r, r+1, r+2, \dots \quad (2.10)$$

Since $P(X = x) \geq 0$, we must check the sum of the probabilities to see that we have a probability distribution function.

$$\text{But } \sum_{x=r}^{\infty} \binom{x-1}{r-1} \cdot p^{r-1} \cdot q^{x-r} \cdot p = p^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} q^{x-r} = p^r (1-q)^{-r} = 1,$$

so $P(X = x)$ is a probability distribution.

If $r = 1$ in (2.10), we find that $P(X = x)$ reduces to the geometric probability distribution function. The result in (2.10) is called the *negative binomial distribution* because of the occurrence of the binomial expansion with a negative exponent.

We now calculate the mean and the variance of this negative binomial random variable. We reasoned that the mean is $\frac{r}{p}$ and we now give another derivation of this.

By the definition of expected value,

$$\begin{aligned} E(X) &= \sum_{x=r}^{\infty} x \cdot \binom{x-1}{r-1} \cdot p^r \cdot q^{x-r} \\ &= \sum_{x=r}^{\infty} r \cdot \frac{x!}{r! \cdot (x-r)!} \cdot p^r \cdot q^{x-r} \\ &= r \cdot p^r \cdot \sum_{x=r}^{\infty} \binom{x}{r} \cdot q^{x-r} \\ &= r \cdot p^r \cdot [1 + \binom{r+1}{1} \cdot q + \binom{r+2}{2} \cdot q^2 + \dots] \\ &= r \cdot p^r \cdot (1-q)^{-(r+1)} \\ &= \frac{r \cdot p^r}{p^{r+1}} = \frac{r}{p}. \end{aligned}$$

Now we seek the variance of this negative binomial random variable. Since $E(X^2)$ is difficult to find directly, we resort to the fact that

$$\text{Var}(X) = E[X(X + 1)] - E(X) - [E(X)]^2.$$

Now

$$\begin{aligned}
 E[X(X+1)] &= \sum_{x=r}^{\infty} x(x+1) \cdot \binom{x-1}{r-1} \cdot p^r \cdot q^{x-r} \\
 &= r(r+1) \cdot p^r \sum_{x=r}^{\infty} \binom{x+1}{r+1} \cdot q^{x-r} \\
 &= r(r+1) \cdot p^r \cdot (1-q)^{-(r+2)} \\
 &= \frac{r(r+1)}{p^2}.
 \end{aligned}$$

Since $E(X) = r/p$, it follows that

$$\text{Var}(X) = \frac{r(r+1)}{p^2} - \frac{r}{p} - \left(\frac{r}{p}\right)^2 = \frac{r \cdot q}{p^2}.$$

It is also useful to view the above-mentioned random variable X as a sum of other random variables. Let X_1 denote the number of trials up to and including the first success, X_2 denote the number of trials after the first success until the second success, and so on. It follows that

$$X = X_1 + X_2 + \dots + X_r.$$

Each of the X_i 's has mean $1/p$ and variance q/p^2 . We see that

$$\begin{aligned}
 E(X) &= \frac{r}{p} = E\left(\sum_{i=1}^r X_i\right) = \sum_{i=1}^r E(X_i) \text{ and in this case} \\
 \text{Var}(X) &= \frac{r \cdot q}{p^2} = \text{Var}\left(\sum_{i=1}^r X_i\right) = \sum_{i=1}^r \text{Var}(X_i),
 \end{aligned}$$

verifying results that were previously obtained.

The fact that the expectation of a sum is the sum of the expectations is generally true; the fact that the variance of a sum is the sum of the variances requires independence of the summands. We will discuss these facts in a more thorough manner in Chapter 5.

In Figure 2.13, we show a graph of the negative binomial distribution with $r = 5$ and $p = 1/2$. It shows that the probabilities increase to a maximum and then decline to become asymptotic to the x -axis as (2.10) would lead us to suspect.

It is also interesting to consider the *total* number of failures that precede the last success. If Y denotes the number of failures preceding the r th success, then

$$P(Y = y) = \binom{y+r-1}{y} \cdot p^r \cdot q^y, \quad y = 0, 1, 2, \dots,$$

which is also a negative binomial distribution. Here, $E(Y) = E(X - r) = E(X) - r = \frac{r}{p} - r = \frac{r \cdot q}{p}$ and $\text{Var}(Y) = \frac{q}{p^2}$.

We now consider three fairly complex examples involving the negative binomial distribution. Each involves special techniques.

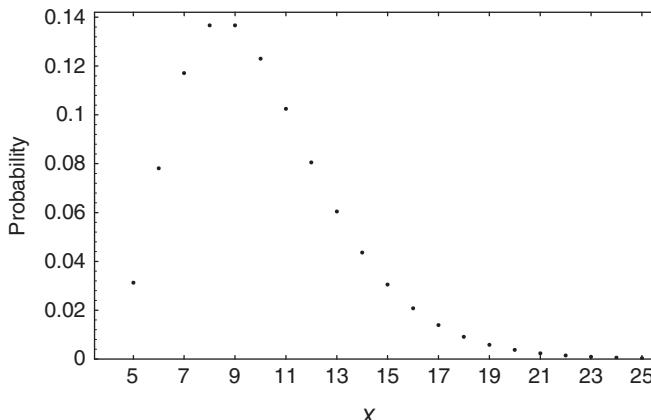


Figure 2.13 A negative binomial distribution.

Example 2.9.1 *All Heads*

I have some fair coins. I toss them once, together, and set aside any that come up heads. I continue to toss the coins remaining, on each toss removing those that come up heads, until all of the coins have come up heads. On average, how many (group) tosses will I have to make?

The problem is probably a bit hard at this point, so let us analyze the situation with only two fair coins.

Since the waiting time for heads with either coin is a geometric variable, we are interested in the maximum value of two geometric variables. Let Y be the random variable denoting the number of group tosses that must be made. We seek $P(Y = y)$. The last head can occur at the y th toss in two mutually exclusive ways:

1. Both coins come up tails for $y - 1$ tosses and then both come up heads on the y th toss
or
2. Exactly one of the coins comes up heads on one of the first $y - 1$ tosses, followed by a head on the remaining coin on the y th toss.

The first of these possibilities has probability $\left(\frac{1}{4}\right)^{y-1} \cdot \left(\frac{1}{4}\right)$. To calculate the second, suppose first that there are $j - 1$ tosses where both coins show tails. Then one of the coins comes up heads on the j th toss. Finally, the single remaining coin is tossed giving $y - j - 1$ tails followed by heads on the y th toss.

This sequence of events has probability

$$\left(\frac{1}{4}\right)^{j-1} \cdot \left\{ \binom{2}{1} \cdot \frac{1}{2} \cdot \frac{1}{2} \right\} \cdot \left(\frac{1}{2}\right)^{y-j-1} \cdot \frac{1}{2}.$$

To find the probability for the second possibility, we must sum the earlier expression over all possible values of j . Thus, the second possibility has probability

$$\sum_{j=1}^{y-1} \binom{2}{1} \cdot \left(\frac{1}{4}\right)^j \cdot \left(\frac{1}{2}\right)^{y-j}.$$

So, putting these results together,

$$\begin{aligned}
 P(Y = y) &= \left(\frac{1}{4}\right)^y + \sum_{j=1}^{y-1} \binom{2}{1} \cdot \left(\frac{1}{4}\right)^j \cdot \left(\frac{1}{2}\right)^{y-j}, \quad y = 1, 2, 3, \dots \\
 &= \left(\frac{1}{4}\right)^y + 2 \cdot \left(\frac{1}{2}\right)^{y-1} \sum_{j=1}^{y-1} \left(\frac{1}{2}\right)^{2j} \cdot \left(\frac{1}{2}\right)^{-j} \\
 &= \left(\frac{1}{4}\right)^y + 2 \cdot \left(\frac{1}{2}\right)^{y-1} \sum_{j=1}^{y-1} \left(\frac{1}{2}\right)^j \\
 &= \left(\frac{1}{4}\right)^y + 2 \cdot \left(\frac{1}{2}\right)^y \left[1 - \left(\frac{1}{2}\right)^{y-1}\right].
 \end{aligned}$$

This reduces to

$$P(Y = y) = \frac{2^{y+1} - 3}{4^y}, \quad y = 1, 2, 3, \dots$$

A computer algebra system shows that the mean, and also the variance, of this distribution is $8/3$.

Example 2.9.2

A fair coin is tossed repeatedly, and a running count of the number of heads and tails obtained is made. What is the probability the heads count reaches 5 before the tails count reaches 3?

Clearly the last toss must result in the fifth head that can be preceded by exactly 0, or 1 or 2 tails. Each of these probabilities is a negative binomial probability.

Let X denote the total number of tosses necessary and let j denote the number of tails. Then, by the negative binomial distribution,

$$\begin{aligned}
 P(5 \text{ heads before } 3 \text{ tails}) &= \sum_{j=0}^2 \binom{4+j}{4} \cdot \left(\frac{1}{2}\right)^{5+j} \\
 &= \left(\frac{1}{2}\right)^5 + \binom{5}{4} \cdot \left(\frac{1}{2}\right)^6 + \binom{6}{4} \cdot \left(\frac{1}{2}\right)^7 \\
 &= \frac{29}{128}.
 \end{aligned}$$

It may be easier to see the structure of the answer if the coin was loaded. Let p denote the probability of heads. Then, reasoning as above,

$$P(5 \text{ heads before } 3 \text{ tails}) = \sum_{j=0}^2 \binom{4+j}{4} p^5 q^j$$

and

$$P(\text{heads count reaches } h \text{ before the tails count reaches } t)$$

$$= \sum_{j=0}^{t-1} \binom{h-1+j}{h-1} q^j p^h = p^h \sum_{j=0}^{t-1} \binom{h-1+j}{h-1} q^j.$$

Example 2.9.3
Candy Jars

A professor has two jars of candy on his desk. When a student enters his office, he or she is invited to choose a jar at random and then select a piece of candy. After sometime, one of the jars will be found empty. At that time, on average, how many pieces of candy are in the remaining jar?

The problem appears in the literature as *Banach's Match Book Problem* after the famous Polish mathematician. It is an instance of Example 2.9.2.

We specialize the problem to two jars, each jar initially containing n pieces of candy and we further suppose that each jar is selected with probability 1/2.

Consider either of the jars; call it, for convenience, the first jar; suppose we empty it and then at some subsequent selection, choose it again and find that it is empty. Suppose further that the remaining jar at that point has X pieces of candy in it. Thus, the first $n + (n - x)$ selections involve choosing the first jar exactly n times and the last choice must be the first jar. Since the jars are symmetric and it makes no difference, which we designate as the first jar,

$$P(X = x) = 2 \cdot \binom{2n-x}{n} \cdot \left(\frac{1}{2}\right)^{2n-x+1}, x = 0, 1, 2, \dots, n. \quad (2.11)$$

A graph of this probability distribution function, for $n = 15$, is shown in Figure 2.14. It shows that the most probable value for X is $x = 0$ or $x = 1$, and that the probabilities decrease steadily as x increases.

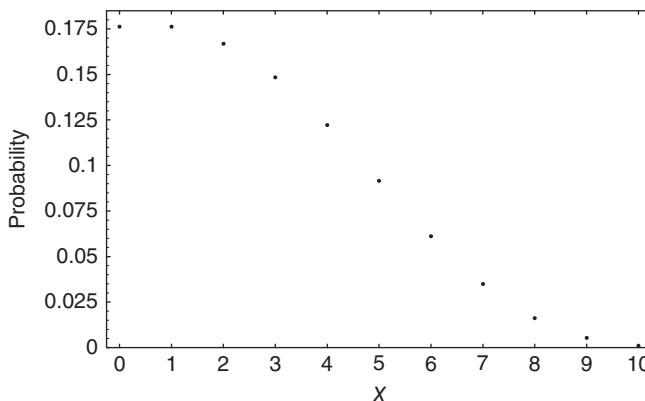


Figure 2.14 The candy jars problem for $n = 15$.

From the arguments used to establish (2.11), it follows that

$$\sum_{x=0}^n 2 \cdot \binom{2n-x}{n} \cdot \left(\frac{1}{2}\right)^{2n-x+1} = 1.$$

A direct analytic proof of this is challenging. Finding the mean and variance is similarly difficult, so we show a way to find these using a recursion. (This method was also used to establish the mean and variance of the binomial distribution and is generally applicable to other discrete distributions.)

A Recursion

It is easy to use (2.11) to show that

$$\frac{P(X=x)}{P(X=x-1)} = 2 \cdot \frac{n-x+1}{2n-x+1}, \quad x = 1, 2, \dots, n. \quad (2.12)$$

This can also be written as

$$\frac{P(X=x)}{P(X=x-1)} = 1 - \frac{x-1}{2n-(x-1)}, \quad x = 1, 2, \dots, n,$$

showing that the probabilities decrease as x increases and that the most probable value is $x = 0$ or $x = 1$.

Now we seek the mean and the variance. Rearranging and summing (2.12) from 1 to n (the region of validity for the recursion), we have

$$\sum_{x=1}^n (2n-x+1) \cdot P(X=x) = 2 \cdot \sum_{x=1}^n (n-x+1) \cdot P(X=x-1).$$

This in turn can be written as

$$(2n+1) \cdot [1 - P(X=0)] - E(X) = 2n \cdot [1 - P(X=n)] \\ - 2 \cdot [E(X) - n \cdot P(X=n)].$$

Simplifying and rearranging give

$$E(X) = (2n+1) \cdot \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} - 1.$$

$E(X)$ is approximately a linear function of n as Figure 2.15 shows.

To find the variance of X , we first find $E(X^2)$. It follows from recursion (2.12) that

$$\sum_{x=1}^n x \cdot (2n-x+1) \cdot P(X=x) = 2 \sum_{x=1}^n x \cdot (n-x+1) \cdot P(X=x-1).$$

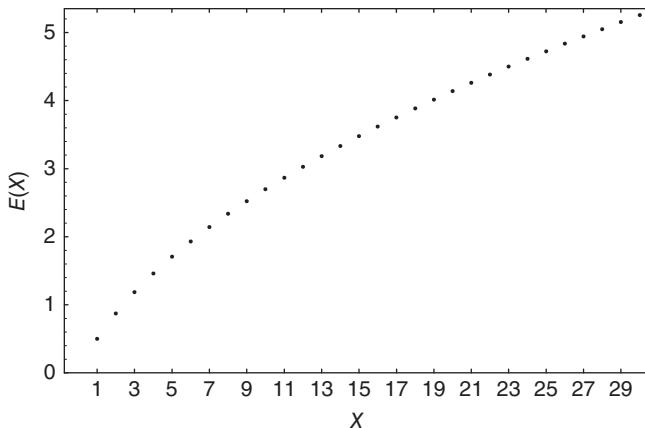


Figure 2.15 $E(X)$ for the candy jars problem.

The left-hand side reduces to $(2n + 1) \cdot E(X) - E(X^2)$ while the right-hand side can be written as

$$\begin{aligned} & 2n \cdot \sum_{x=1}^n (x - 1) \cdot P(X = x - 1) + 2n \cdot \sum_{x=1}^n P(X = x - 1) \\ & - 2 \sum_{x=1}^n x \cdot (x - 1) \cdot P(X = x - 1), \end{aligned}$$

which becomes

$$\begin{aligned} & 2n \cdot [E(X) - n \cdot P(X = n)] + 2n \cdot [1 - P(X = n)] - 2E[X(X + 1)] \\ & + 2n(n + 1)P(X = n). \end{aligned}$$

It then follows that

$$\text{Var}(X) = 2(n + 1) - (2n + 1) \cdot \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} - \left[(2n + 1) \cdot \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n}\right]^2$$

This is an increasing function of n . A graph is shown in Figure 2.16.

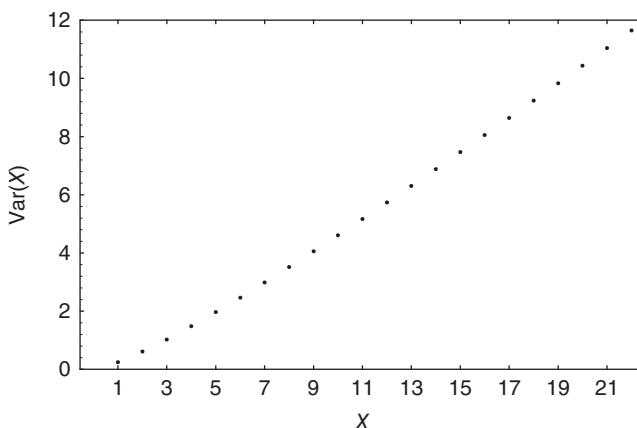


Figure 2.16 Variance in the candy jars problem.

EXERCISES 2.9

1. A fair die is thrown until a 6 appears. Find the probability this occurs in 5 tosses.
2. I have 6 pairs of socks randomly distributed in a drawer. They are drawn out one at a time until a pair occurs. Find the probability this happens in 3 draws. (The reader may also wish to consult Chapter 7.)
3. A coin, loaded to come up heads $2/3$ of the time, is thrown until heads appear. What is the probability an odd number of tosses is necessary?
4. The coin in problem 3 is now tossed until the fifth head appears. What is the probability this will occur in at most 9 tosses?
5. The probability of a successful rocket launching is 0.8, the process following the binomial assumptions.
 - (a) Find the probability the first successful launch occurs at the fourth attempt.
 - (b) Suppose now that attempts are made until 3 successful launchings have occurred. What is the probability that exactly 6 attempts will be necessary?
6. A box of manufactured parts contains four good and three defective parts. They are drawn out one at a time, without replacement. Let X denote the number of the drawing on which the first defective part occurs.
 - (a) Find the probability distribution for X .
 - (b) Find $E(X)$.
7. The probability a player wins a game at a single trial is $1/3$. Assume the trials follow the binomial assumptions. If the player plays until he wins, find the probability the number of trials is divisible by 4.
8. The probability a new driver will pass a driving test is 0.8.
 - (a) One student takes the test until she passes it. What is the probability it will take at least two attempts to pass the test?
 - (b) Now suppose three students take the driving test until each has passed it. What is the probability that exactly one of the three will take at least two attempts before passing the test? (Assume independence.)
9. To become an actuary, one must pass a series of 9 examinations. Suppose that 60% of those taking each examination pass it and that passing the examinations are independent of each other. What is the probability a person passes the 9th examination, and so has passed all the examinations, on the 15th attempt?
10. A quality control inspector on a production line samples items until a defective item is found.
 - (a) If the probability an item is defective is 0.08, what is the probability that at least 10 items must be inspected?
 - (b) Suppose now that the 16th item inspected is the first defective item found. If p is the probability an item is defective, what is the value of p that makes the probability that the 16th item inspected is the first defective item found most likely?
11. A fair coin is tossed. What is the probability the fourth head is preceded by at most two tails?
12. A TV interviewer must conduct five interviews. Suppose the probability a person agrees to be interviewed is $2/3$.

- (a) What is the probability the interviewer will ask 9 people in all to be interviewed?
 (b) How many people can the interviewer expect to ask to be interviewed?
13. In August, the probability a thunderstorm will occur on any particular day is 0.1. What is the probability the first thunderstorm in August will occur on August 12?
14. In a manufacturing process, the probability a produced item is good is 0.97. Assuming the items produced are independent, what is the probability that exactly five defective items precede the 100th good item?
15. A box contains six good and four defective items. Items are drawn out one at a time, without replacement.
 (a) Find the probability the third defective item occurs on the fifth draw.
 (b) On what drawing is it most likely for the third defective to occur?
16. A coin, loaded to come up heads with probability $3/4$, is tossed until heads appear or until it has been tossed five times. Find the probability the experiment will end in an odd numbered toss, given that the experiment takes more than one toss.
17. Suppose you are allowed to flip a fair coin until the first head appears. Let X denote the total number of flips required.
 (a) Suppose you win $\$2^X$ if $X \leq 19$ and $\$2^{20}$ if $X \geq 20$ for playing the game. A game is *fair* if the amount paid to play the game equals the expected winnings. How much should you pay to play this game if it is fair?
 (b) Suppose now that you win $\$2^X$ regardless of the number of flips. Can the game be made fair?
18. Use the recursion (2.12) to find the most likely number of pieces of candy remaining when one of the candy jars is found empty.
19. X is a negative binomial random variable with p as the probability of success at any trial. Suppose the r th success occurs at trial t . Find the value of p that makes this event most likely.

2.10 THE HYPERGEOMETRIC RANDOM VARIABLE: ACCEPTANCE SAMPLING

Acceptance Sampling

Products produced from industrial processes are often subjected to sampling inspection before they are delivered to the customer. This sampling is done to insure a level of quality in delivered manufactured products and to insure some uniformity in the product. Usually, unacceptable product (product which does not meet the manufacturer's specifications) becomes mixed up with acceptable product due to changes in the manufacturing process and random events in that process. Modern techniques of statistical process control have greatly improved the quality of manufactured products and while it is best to produce only flawless products, often the quality of the product can only be determined through sampling. However, determining whether a product is acceptable or unacceptable may destroy the product. Because of the time and money involved in inspecting the product in its entirety even if destruction of the product is not involved, *sampling plans*, which inspect only a sample of the product, are often employed. It has also been found that sampling is often more

more accurate than 100% inspection since the inspection of each and every item demands constant attention. Boredom or lack of care often sets in which is not the case when smaller samples are randomly chosen at random times. As we will see, probability theory renders 100% inspection unnecessary even when it is possible, so total inspection of a manufactured product has become rare.

Due to the emphasis on quality in manufacturing and statistical process control, probability theory has become extremely important in industry.

The chance a sample has a given composition can be determined from probability theory. As an example, suppose we have a lot (a number of produced items) containing eight acceptable, or good, items as well as four unacceptable items. A sample of three items is drawn. What is the probability the sample contains exactly one unacceptable item?

The sampling is done without replacement (since one would not want to inspect the same item repeatedly!), and since the order in which the items are drawn is of no importance, there are $\binom{12}{3} = 220$ samples comprising the sample space. If the sampling plan, that is, the manner in which the sampled items are drawn, is appropriate, we consider each of these samples to be equally likely. Now, we must count the number of samples containing exactly one unacceptable item (and so exactly two acceptable items). There are $\binom{4}{1} \cdot \binom{8}{2} = 112$ such samples. So the probability the sample contains exactly one unacceptable item is

$$\frac{\binom{4}{1} \cdot \binom{8}{2}}{\binom{12}{3}} = \frac{112}{220} = 0.509.$$

The probability that the sample contains no defective items is

$$\frac{\binom{4}{0} \cdot \binom{8}{3}}{\binom{12}{3}} = \frac{14}{55} = 0.255$$

so the probability that the sampling plan will detect *at least* one unacceptable item is

$$1 - \frac{\binom{8}{3}}{\binom{12}{3}} = 0.745.$$

Our sampling plan is then likely to detect at least one of the unacceptable items in the lot, but it is not certain to do so.

Let us suppose that we carry out the earlier inspection plan and decide to sell the entire lot only if no unacceptable items are found in the sample. The probability this lot survives this sampling plan and is sold is 0.255. So about 26% of the time, lots with $4/12 = 33\frac{1}{3}\%$ unacceptable items will be sold.

Usually, then the sampling plan will determine some unacceptable items, which are not sent to the customer. One of two courses of action is generally pursued at this point.

Either the unacceptable items in the sample are replaced with good items or the entire lot is inspected and any unacceptable items in the lot are replaced by good items. Either of these plans will improve the quality of the product sold, the second being the better if it can be carried out. In case the testing is destructive, only the first plan can be executed. Let us compare the plans in this case, assuming that either can be carried out.

We start by replacing only the unacceptable items in the sample.

The sample contains no unacceptable items with probability $\frac{\binom{8}{3}}{\binom{12}{3}} = \frac{14}{55}$, so the outgoing lot will contain $4/12$ or $1/3$ unacceptable items with this probability.

The sample contains exactly one unacceptable item with probability $\frac{\binom{8}{2} \cdot \binom{4}{1}}{\binom{12}{3}} = \frac{28}{55}$, producing $3/12$ or $1/4$ unacceptable items in the outgoing lot.

The sample contains exactly two unacceptable items with probability $\frac{\binom{8}{1} \cdot \binom{4}{2}}{\binom{12}{3}} = \frac{12}{55}$, producing $2/12$ or $1/6$ unacceptable items in the outgoing lot.

Finally, the sample contains exactly three unacceptable items with probability $\frac{\binom{4}{3}}{\binom{12}{3}} = \frac{1}{55}$, resulting in $1/12$ unacceptable items in the outgoing lot.

The result of this plan is that, on average, the percentage of unacceptable items the lot will contain is

$$\frac{14}{55} \cdot \frac{1}{3} + \frac{28}{55} \cdot \frac{1}{4} + \frac{12}{55} \cdot \frac{1}{6} + \frac{1}{55} \cdot \frac{1}{12} = 25\%.$$

This is considerably less than the $33\frac{1}{3}\%$ unacceptable items in the lot. Sampling cannot improve the quality of the product manufactured, but it can, and does, improve the quality of the product sold. In fact, dramatic gains can be made by this process, which we will call *acceptance sampling*.

Even greater gains can be attained if, when the sample contains at least one unacceptable item, the entire lot is inspected and any unacceptable items in the lot are replaced by good items. In that circumstance, either the lot sold is 100% good (with probability 0.745) or the lot contains $4/12 = 33\frac{1}{3}\%$ unacceptable items. Then the average percentage of unacceptable items sold is

$$0\% \cdot 0.745 + 33\frac{1}{3}\% \cdot 0.255 = 8.5\%.$$

This is a dramatic gain and, as we shall see, is often possible if acceptance sampling is employed.

The average percentage of unacceptable product sold is called the *average outgoing quality* (AOQ). The AOQ, if only unacceptable items in the sample are replaced before the lot is sold, is 25%.

Lots are rarely so small as in our example, so we must investigate the behavior of the above-mentioned sampling plan when the lots are large. Before doing that, we define the relevant random variable and determine some of its properties.

The Hypergeometric Random Variable

We generalize the earlier situation to a lot of N items, D of which are unacceptable. Let X denote the number of unacceptable items in the randomly chosen sample of n items. Then

$$P(X = x) = \frac{\binom{D}{x} \cdot \binom{N-D}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \dots, \min\{n, D\} \quad (2.13)$$

We assume that $\min\{n, D\} = n$ in what follows. The argument is similar if $\min\{n, D\} = D$.

If X has the probability distribution given by (2.13), then X is called a *hypergeometric* random variable.

Since $\sum_{x=0}^n \binom{D}{x} \cdot \binom{N-D}{n-x}$ represents all the mutually exclusive ways in which x unacceptable items and $n - x$ acceptable items can be chosen from a group of N items, this sum must be $\binom{N}{n}$, showing that the sum of the probabilities in (2.13) must be 1.

We will use a recursion to find the mean and variance. Let $G = N - D$. Then, from (2.13),

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{(D - x + 1)(n - x + 1)}{x(G - n + x)}, \quad x = 1, 2, \dots, n. \quad (2.14)$$

So

$$\begin{aligned} (G - n) \sum_{x=1}^n x P(X = x) + \sum_{x=1}^n x^2 P(X = x) \\ = \sum_{x=1}^n (D - x + 1)(n - x + 1) P(X = x - 1). \end{aligned}$$

After expanding and simplifying the sums involved, we find that

$$E(X) = n \cdot \frac{D}{N}.$$

This result is analogous to the mean of the binomial: np , but here D/N is the probability that the first item drawn only is unacceptable. It is surprising that the nonreplacement does not affect the mean value. The drawings for the hypergeometric are clearly dependent, a fact that will affect the variance.

To find $E(X^2)$, multiply (2.14) through by x giving

$$\begin{aligned} (G - n) \sum_{x=1}^n x^2 P(X = x) + \sum_{x=1}^n x^3 P(X = x) \\ = \sum_{x=1}^n x \cdot (D - x + 1)(n - x + 1) P(X = x - 1). \end{aligned}$$

These quantities can be expanded and simplified using the result for $E(X)$. We find that

$$E(X^2) = \frac{nD}{N(N-1)} \cdot (nD - n - D + N),$$

from which it follows that

$$\text{Var}(X) = n \cdot \frac{D}{N} \cdot \frac{N-D}{N} \cdot \frac{N-n}{N-1}.$$

This result is analogous to the variance, $n \cdot p \cdot q$, of the binomial but involves a factor, $\frac{N-n}{N-1}$, often called a *finite population correction factor*, due to the fact that the drawings are not independent.

The correction factor, however, approaches 1 as $N \rightarrow \infty$ and so the variance of the hypergeometric approaches that of the binomial. This result, together with the mean value, suggests that the hypergeometric distribution can be approximated by the binomial distribution as the population size, N , increases. This is due to the fact that as N increases, the nonreplacement of the items drawn has less and less effect on the probabilities involved. We pause here to show that is indeed the case.

We begin with

$$P(X = x) = \frac{\binom{D}{x} \cdot \binom{N-D}{n-x}}{\binom{N}{n}},$$

which can be written as

$$P(X = x) = \frac{D(D-1)(D-2) \cdots (D-x+1)}{x!}.$$

$$\frac{(N-D)(N-D-1) \cdots (N-D-n+x+1)}{(n-x)!}.$$

$$\frac{n!}{N(N-1)(N-2) \cdots (N-n+1)}.$$

This in turn can be rearranged as

$$P(X = x) = \binom{n}{x} \cdot \frac{D}{N} \cdot \frac{D-1}{N-1} \cdots \frac{D-x+1}{N-x+1}.$$

$$\frac{N-D}{N-x} \cdot \frac{N-D-1}{N-x-1} \cdots \frac{N-D-n+x+1}{N-n+1}.$$

Approximating each of the factors $\frac{D}{N}, \frac{D-1}{N-1}, \dots, \frac{D-x+1}{N-x+1}$ by $\frac{D}{N}$ and each of the factors $\frac{N-D}{N-x}, \frac{N-D-1}{N-x-1}, \dots, \frac{N-D-n+x+1}{N-n+1}$ by $\frac{N-D}{N}$, we see that

$$P(X = x) \approx \binom{n}{x} \cdot \left(\frac{D}{N}\right)^x \cdot \left(\frac{N-D}{N}\right)^{n-x},$$

which is the binomial distribution.

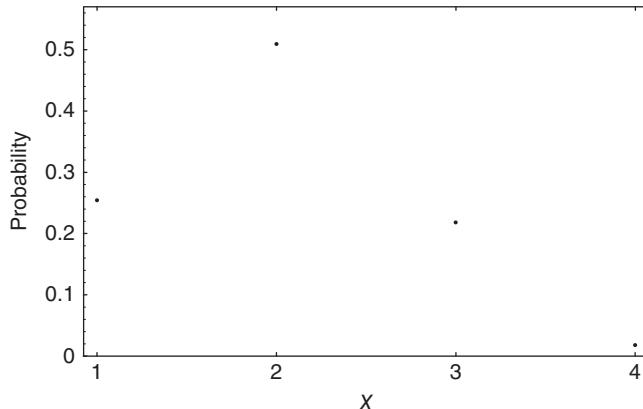


Figure 2.17 Hypergeometric distribution with $N = 12$, $n = 3$, and $D = 4$.

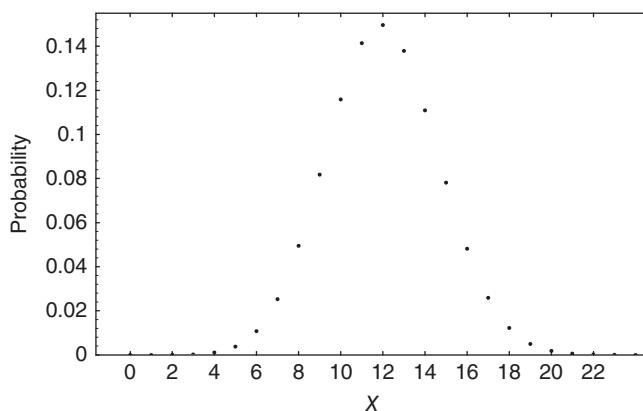


Figure 2.18 Hypergeometric distribution with $N = 1000$, $n = 30$, and $D = 400$.

Some Specific Hypergeometric Distributions

It is useful at this point to look at some specific hypergeometric distributions. Our initial example, in Section 2.10 had $N = 12$, $n = 3$, and $D = 4$. A graph of the probability distribution is shown in Figure 2.17.

As the population size increases, we expect the hypergeometric distribution to appear more binomial or normal-like. Figure 2.18 shows that this is the case. Here, $N = 1000$, $D = 400$, and $n = 30$.

While Figure 2.17 shows no particular features, Figure 2.18 shows again the now familiar normal appearance.

EXERCISES 2.10

1. A carton of 12 light bulbs contains 1 defective bulb. A sample of 3 bulbs is chosen. What is the probability the sample contains the defective bulb?
2. Let X denote the number of defective bulbs in the sample in problem 1. Find $E[X]$.

3. A lot of 50 fuses is known to contain 7 defectives. A random sample of size 10 is drawn without replacement. What is the probability the sample contains at least 1 defective fuse?
4. A collection of 30 gems, all of which are identical in appearance and are supposed to be genuine diamonds, actually contains 8 worthless stones. The genuine diamonds are valued at \$1200 each. Two gems are selected.
 - (a) Let X denote the total actual value of the gems selected. Find the probability distribution function for X .
 - (b) Find $E(X)$.
5. (a) A box contains three red and five blue marbles. The marbles are drawn out one at a time and without replacement, until all of the red marbles have been selected. Let X denote the number of drawings necessary. Find the probability distribution function for X .
 - (b) Find the mean and variance for X .
6. (a) A box contains three red and five blue marbles. The marbles are drawn out one at a time and without replacement, until all the marbles left in the box are of the same color. Let X denote the number of drawings necessary. Find the probability distribution function for X .
 - (b) Find the mean and variance for X .
7. A lot of 400 automobile tires contains 10 with blemishes that cannot be sold at full price. A sampling inspection plan chooses 5 tires at random and accepts the lot only if the sample contains no tires with blemishes.
 - (a) Find the probability the lot is accepted.
 - (b) Suppose any tires with blemishes in the sample are replaced by good tires if the lot is rejected. Find the AOQ of the lot.
8. A sample of size 4 is chosen from a lot of 25 items of which D are defective. Draw the curve showing the probability the lot is accepted as a function of D if the lot is accepted only when the sample contains no defective items.
9. A lot of 250 items which contains 15 defective items is subject to an acceptance sampling plan that calls for a sample of size 6 to be drawn. The lot is accepted if the sample contains at most 1 defective item.
 - (a) Find the probability the lot is accepted.
 - (b) Suppose any defective items in the sample are replaced by good items. Find the AOQ.
10. In problem 5, suppose now that the entire lot is inspected and any blemished tires replaced by good tires if the lot is rejected by the sample. Find the AOQ.
11. In problem 7 if any defective items in the lot are replaced by good items when the sample rejects the entire lot, find the AOQ.
12. Exercises 5 and 6 can be generalized. Suppose a box has a red and b blue marbles and that X is the number of drawings necessary to draw out all of the red marbles.
 - (a) Show that

$$P(X = x) = \frac{\binom{x-1}{a-1}}{\binom{a+b}{a}}, \quad x = a, a+1, \dots, a+b.$$

- (b) Using the result in part (a), show that a recursion can be simplified to

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{x - 1}{x - a}, \quad x = a + 1, a + 2, \dots, a + b.$$

- (c) Show that the recursion in part (b) leads to

$$\sum_{x=a+1}^{a+b} x \cdot (x - a) \cdot P(X = x) = \sum_{x=a+1}^{a+b} x \cdot (x - 1) \cdot P(X = x - 1).$$

From this, conclude that

$$E(X) = a \cdot \frac{a + b + 1}{a + 1}.$$

- (d) Show that

$$V(X) = \frac{a \cdot b \cdot (a + b + 1)}{(a + 1)^2 \cdot (a + 2)}.$$

- 13.** (Exercise 12 continued) Now suppose X represents the number of drawings until all the marbles remaining in the box are of the same color. Show that

$$P(X = x) = \frac{\binom{x-1}{a-1} + \binom{x-1}{b-1}}{\binom{a+b}{a}}, \quad x = \min[a, b], \dots, a + b - 1,$$

and that

$$E(X) = \frac{a \cdot b}{a + 1} + \frac{a \cdot b}{b + 1}.$$

- 14.** A box contains three red and five blue marbles. The marbles are drawn out one at a time without replacement until a red marble is drawn. Let X denote the total number of drawings necessary.

- (a) Find the probability distribution function for X .
(b) Find the mean and the variance of X .

- 15.** Exercise 14 is generalized here. Suppose a box contains a red and b blue marbles, and that X denotes the total number of drawings made without replacement until a red marble is drawn.

- (a) Show that

$$P(X = x) = \frac{\binom{a+b-x}{a-1}}{\binom{a+b}{a}}, \quad x = 1, 2, \dots, b + 1.$$

- (b) Using the result in part (a), show that a recursion can be simplified to

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{b - x + 2}{a + b - x + 1}, \quad x = 2, 3, \dots, b + 1.$$

- (c) Use the recursion in part (b) to show that

$$E(X) = \frac{a+b+1}{a+1}$$

and $V(X) = \frac{a \cdot b \cdot (a+b+1)}{(a+1)^2 \cdot (a+2)}$.

- (d) Show that the mean and variance in part (c) approach the mean and variance of the geometric random variable as both a and b become large.

2.11 ACCEPTANCE SAMPLING (CONTINUED)

We considered an acceptance sampling plan in section “Acceptance Sampling”, and we saw that some gains can be made with respect to the average quality delivered when the unacceptable items in either the sample or in the entire lot are replaced with good items. We can now discuss some specific results, dealing with lots that are usually large. We first consider the effect of the size of the sample on the process.

Example 2.11.1

A lot of 200 items is inspected by drawing a sample of size n without replacement; the lot is accepted only if all the items in the sample are good. Suppose the lot contains 2%, or 4, unacceptable items. Then the probability the lot is accepted by this sampling plan is

$$\frac{\binom{196}{n}}{\binom{200}{n}}.$$

This is a steadily decreasing function of n , as we would expect. We find that if $n = 5$, the probability the lot is accepted is 0.903, while if $n = 30$, this probability is 0.519. A graph of this function is shown in Figure 2.19.

Not surprisingly, large samples yield more accurate results than small samples.

Example 2.11.2

Now we consider the effect of the quality of the lot on the probability of acceptance. Suppose $p\%$ of a lot of 1000 items is unacceptable. The sampling plan is this: select a sample of 100 and accept the lot if the sample contains at most 4 unacceptable items. The probability the lot is accepted is then

$$\sum_{x=0}^4 \frac{\binom{1000p}{x} \cdot \binom{1000 - 1000p}{100 - x}}{\binom{1000}{100}}.$$

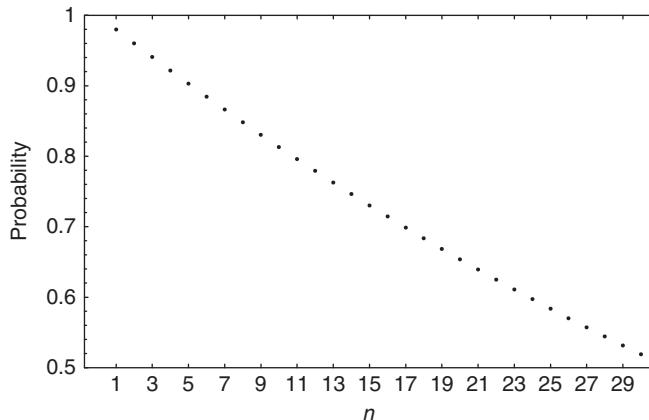


Figure 2.19 Effect of sample size, n , on a sampling plan.

This is a decreasing function of the percentage of unacceptable items in the lot. These values are easily calculated. If, for example, the lot contains 10 unacceptable items, then the probability the lot is accepted is 0.9985.

A graph of this probability as a function of p is shown in Figure 2.20.

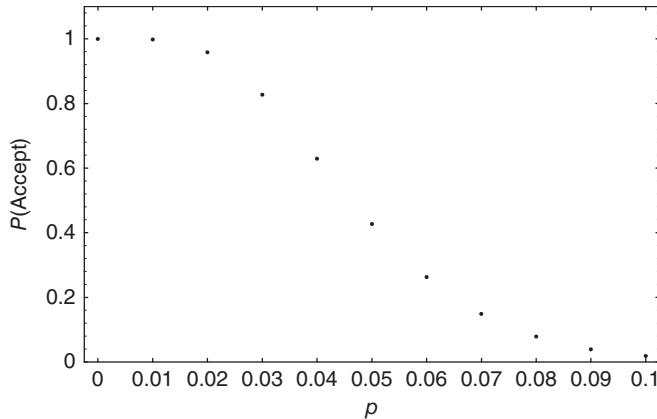


Figure 2.20 Effect of quality in the lot on the probability of acceptance.

The curve in Figure 2.20 is called the *operating characteristic* (or *OC*) curve for the sampling plan. Sampling plans are often compared by comparing the rapidity with which the OC curves for different plans decrease.

In this case, the sample size is small relative to the population size, so we would expect that the nonreplacement of the sample items will have little effect on the probability the lot is accepted. A binomial model approximates the probability the lot is accepted if in fact it contains 10 unacceptable items as 0.9966 (we found the exact probability above to be 0.9985).

Producer's and Consumer's Risks

Acceptance sampling involves two types of risk: the producer would like to guard against a “good” lot being rejected, although this cannot be guaranteed; the consumer, on the other hand, wants to guard against a “poor” lot being accepted by the sampling plan, although, again, this cannot be guaranteed.

The words “good” and “poor” of course must be decided in the context of the practical situation. Often when these are defined and the probability of the risks set, a sampling plan can be devised (specifically, a sample size can be determined) that, at least approximately, meets the risks set.

Consider Example 2.11.1 again. Here the lot size is 200, but suppose that D of these items are unacceptable. Again we draw a sample of size n and accept the lot when the sample contains no unacceptable items. So

$$P(\text{lot is accepted}) = \frac{\binom{200-D}{n}}{\binom{200}{n}}.$$

Figure 2.21 shows this probability as a function of the sample size, n , where D has been varied from 0 to 25.

Since the curves are monotonically decreasing, it is often possible to select a curve (thus, determining a sample size) that passes through two given points. If the producer would like lots with exactly 1 unacceptable item rejected with probability 0.10 (so such lots are accepted with probability 0.90) and if the consumer would like lots with 24 unacceptable items rejected with probability 0.95 (so such lots are accepted with probability 0.05), we find a sample size of 22 will approximate these restrictions. To check this, note that

$$\frac{\binom{199}{22}}{\binom{200}{22}} = 0.89$$

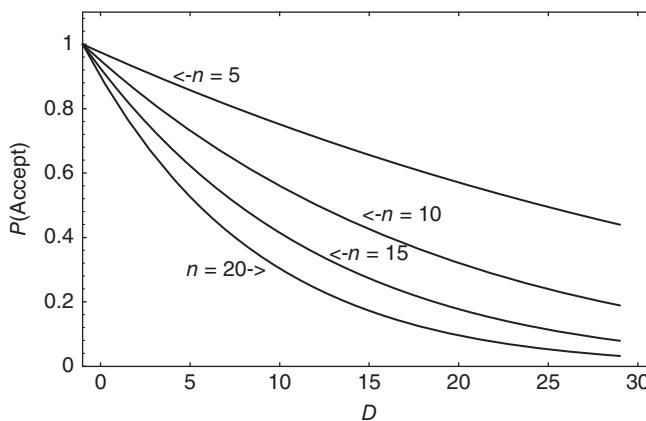


Figure 2.21 Some operating characteristic curves.

and

$$\frac{\binom{176}{22}}{\binom{200}{22}} = 0.05.$$

A computer algebra system here is of great help in finding an approximate solution to the problem.

Average Outgoing Quality

We saw that considerable improvement in the quality of the product sold can be made if any items in the sample are replaced by good items. This is the most sensible strategy we can follow if the sampling is destructive; in that case we have little choice but to take a sample since destroyed products must be replaced by others. Recall that the AOQ is the percentage of unacceptable items sold to the buyer. We want to consider the behavior of the AOQ in this section.

Example 2.11.3

Suppose a lot of 100 items actually contains 4 unacceptable items. A sample of 5 items is drawn and any unacceptable item in the sample is replaced by a good item. On average, what proportion of unacceptable items is sold using this sampling plan?

Let X denote the number of unacceptable items in the sample. Then X is a hypergeometric random variable and

$$P(X = x) = \frac{\binom{4}{x} \cdot \binom{96}{5-x}}{\binom{100}{5}}, \quad x = 0, 1, 2, 3, 4.$$

So the AOQ is

$$\text{AOQ} = \sum_{x=0}^4 \frac{(4-x)}{100} \cdot \frac{\binom{4}{x} \cdot \binom{96}{5-x}}{\binom{100}{5}}$$

since $4 - X$ unacceptable items will be sold. But

$$\text{AOQ} = \frac{4}{100} - \frac{1}{100} \sum_{x=0}^4 x \cdot \frac{\binom{4}{x} \cdot \binom{96}{5-x}}{\binom{100}{5}},$$

where the summation is the mean value of a hypergeometric random variable with $N = 100$, $n = 5$, and $D = 4$. It follows that

$$\text{AOQ} = \frac{4}{100} - \frac{1}{100} \cdot 5 \cdot \frac{4}{100} = 0.038.$$

This is less than the population percentage of unacceptable items, 0.04, but not greatly less. The sampling has improved the quality of the product sold, but not by much. The effect the sampling plan has will increase as the percentage of unacceptable product in the lot increases.

Another possible plan is to replace each unacceptable item in the lot with a good item if the sample contains any unacceptable items. Now we deliver either all the unacceptable items in the lot or none of them. It follows that

$$\begin{aligned} \text{AOQ} &= \frac{4}{100} \cdot P(\text{sample contains no unacceptable items}) \\ &= \frac{4}{100} \cdot \frac{\binom{96}{5}}{\binom{100}{5}} = 0.032475. \end{aligned}$$

So the gain is greater if we happen to inspect the entire lot.

These conclusions are probably not surprising. But more lurks behind the scenes here! Let us consider a general example.

Example 2.11.4

From a lot of N items that contain D unacceptable items, we draw a sample of size n . If the sample contains any unacceptable items, we inspect the entire lot, replacing each unacceptable item with an acceptable, or good, item. The resulting lot then contains either D or 0 unacceptable items. The AOQ is then

$$\begin{aligned} \text{AOQ} &= \frac{D}{N} \cdot P(\text{sample contains no unacceptable items}) \\ &= \frac{D}{N} \cdot \frac{\binom{N-D}{n}}{\binom{N}{n}}. \end{aligned} \tag{2.15}$$

What happens to this product as D increases? Since $\frac{D}{N}$ increases as D increases and since $\frac{\binom{N-D}{n}}{\binom{N}{n}}$ is a decreasing function of D , it follows that the product in (2.15) attains a maximum value. This is true, regardless of the size of D ! So, no matter what D is, there is a limit for the percentage of unacceptable product sold! This is called the *AOQ limit*. We illustrate this phenomenon in the next example.

Example 2.11.5

Consider again the situation when $N = 1000$ and $n = 100$. Then

$$\text{AOQ} = \frac{D}{1000} \cdot \frac{\binom{1000-D}{100}}{\binom{1000}{100}}.$$

A graph of the AOQ is shown in Figure 2.22, showing that the maximum value of the AOQ is about 0.35%.

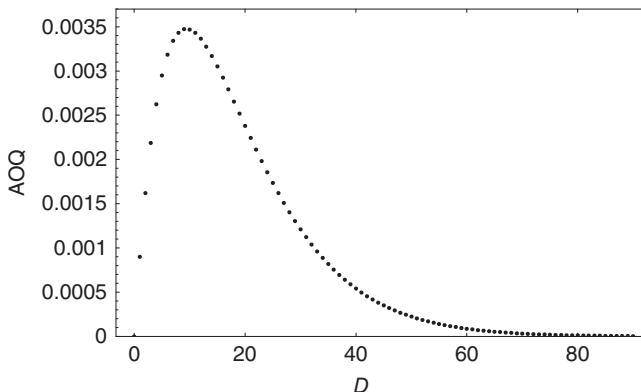


Figure 2.22 AOQ as a function of the number of unacceptable items in the lot.

Double Sampling

Occasionally, lots are accepted if the sample contains, say, at most c unacceptable items and are subject to total inspection if a sample has d or more unacceptable items, where $d > c$. Often if the number of unacceptable items falls between c and d , another sample is taken. We illustrate this procedure with a concrete example.

A lot of 500 items contains 40 unacceptable items. A sample of 50 is taken and the lot accepted if the sample contains no more than 3 unacceptable items. If the sample contains 4 or 5 unacceptable items, an additional sample of 30 is taken; the lot is accepted only if this additional sample contains no unacceptable items. Otherwise, the lot is rejected. We want the probability the lot is accepted.

Let X denote the number of unacceptable items in the first sample. Then the probability the lot is accepted on the basis of the first sample is

$$P(X \leq 3) = \sum_{x=0}^3 \frac{\binom{40}{x} \cdot \binom{460}{50-x}}{\binom{500}{50}}.$$

Now if the first sample has 4 or 5 unacceptable items, then the second sample is taken. The lot now contains 450 items of which $40 - X$ are unacceptable while $450 - (40 - X) = 410 + X$ are good items. The second sample of size 30 must contain only good items. So the probability the lot is accepted on the basis of the second sample is

$$\sum_{x=4}^5 \frac{\binom{40}{x} \cdot \binom{460}{50-x}}{\binom{500}{50}} \cdot \frac{\binom{410+x}{30}}{\binom{450}{30}}.$$

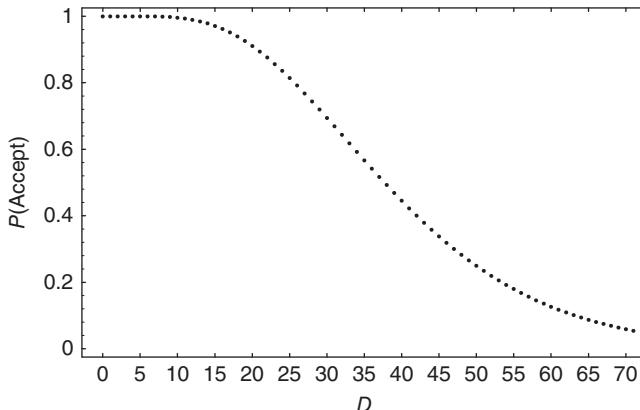


Figure 2.23 Probability lot is accepted in a double sampling plan.

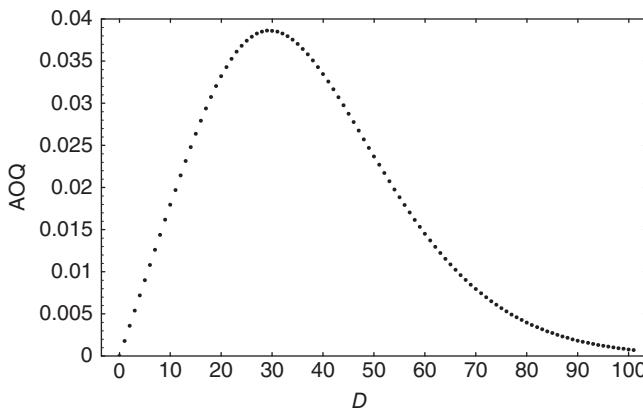


Figure 2.24 Average outgoing quality for a double sampling plan.

The probability the lot is accepted is then

$$\sum_{x=0}^3 \frac{\binom{40}{x} \cdot \binom{460}{50-x}}{\binom{500}{50}} + \sum_{x=4}^5 \frac{\binom{40}{x} \cdot \binom{460}{50-x}}{\binom{500}{50}} \cdot \frac{\binom{410+x}{30}}{\binom{450}{30}} = 0.445334,$$

and, assuming that none of the unacceptable items found in the samples are sold, the AOQ is

$$\begin{aligned} & \sum_{x=0}^3 \frac{(40-x) \cdot \binom{40}{x} \cdot \binom{460}{50-x}}{500 \cdot \binom{500}{50}} \\ & + \sum_{x=4}^5 \frac{(40-x) \cdot \binom{40}{x} \cdot \binom{460}{50-x}}{500 \cdot \binom{500}{50}} \cdot \frac{\binom{410+x}{30}}{\binom{450}{30}} \\ & = 0.0334579. \end{aligned}$$

Here, the sampling plan has reduced the percentage unacceptable item sold from $40/500 = 0.08$ to 0.033 on average, so the plan is quite effective.

With the aid of a computer we can easily vary the number of unacceptable items in the lot and observe the effect this has on both the probability the lot is accepted and the AOQ. These graphs are shown in Figures 2.23 and 2.24.

The maximum value for the AOQ is 0.0386071 and occurs when the lot contains 29 unacceptable items.

EXERCISES 2.11

1. A lot of 100 produced items contains 10 defective items. A sample of 5 is chosen and all the defective items in the sample are replaced by good items.
 - (a) Find the probability the sample contains at least 1 defective item.
 - (b) Find the proportion of defective items sold under this sampling plan.
2. In problem 1, find the AOQ if all the defective items in the lot are replaced by good items when the sample contains any defective items.
3. A day's production of 25 television sets from a small company has 4 sets that have defects and cannot be sold. The company inspects its product by selecting 2 sets; if at most 1 of these has defects, the lot is shipped. What is the probability the lot is shipped?
4. A shipment of 1500 washers contains 400 defective and 1100 nondefective items. Two hundred washers are chosen at random, without replacement.
 - (a) Find the probability that exactly 90 defective items are found.
 - (b) Approximate the probability in part (a) by using the binomial distribution.
5. A lot of 100 fuses is inspected by a quality control engineer who tests 10 fuses selected at random. If 2 or fewer defective fuses are discovered, the entire lot is accepted. Find the probability the lot is accepted if it actually contains 20 defective fuses.
6. A lot of 25 items contains 4 defective items. A sample of size 2 is chosen; the lot is accepted if the sample shows no defective items.
 - (a) Find the probability the lot is shipped.
 - (b) If any defective items in the sample are replaced by good items before the lot is shipped, find the AOQ.
 - (c) Now suppose the lot contains D defective items and that the entire lot is rectified if the sample shows any defective items. Plot the OC curve.
7. A bakery has a batch of 100 cookies, 5 of which are burned. A sample of 3 cookies is chosen and the batch put out for sale if none of the cookies in the sample is burned.
 - (a) What is the probability the batch of cookies is put out for sale?
 - (b) Find the AOQ if any burned cookies in the sample are replaced by good cookies.
 - (c) Assuming that the batch contains B burned cookies and that the entire batch is rectified if any of the cookies in the sample is burned, show the OC curve.
8. In Exercise 6, suppose the number of defective items is unknown and also suppose a rejected lot is subject to 100% inspection and that any defective item in the population is replaced by a good item. Estimate the AOQ limit from a graph of the AOQ.
9. In Exercise 7, suppose that the entire batch of cookies is inspected if the sample should reject the batch. Estimate the AOQ limit from a graph.

10. In inspecting a lot of 500 items, it is desired to accept the lot if the lot contains 1 defective item with probability 0.95 and it is desired to accept the lot if the lot contains 20 defective items with probability 0.05. Suppose the lot is accepted only if the sample contains no defective items. What sample size is necessary?
11. A producer inspects a lot of 400 items and wants the probability the lot is accepted if the lot contains 1% defectives to be 0.90; the consumer wants the probability a lot containing 5% defective items to be accepted with probability 0.60. Suppose the lot is accepted only if the sample contains no defective items. Find the sample size so that the sampling plan meets these risks.
12. A double sampling plan is carried out from a lot of 500 items. A sample of 10 is selected and the lot is accepted if this sample contains no unacceptable items; if this sample contains 3 or more unacceptable items, the lot is rejected. If the sample contains 1 or 2 unacceptable items, a second sample of 20 is drawn; the lot is then accepted if the total number of unacceptable items in the two samples combined is at most 3. Suppose that at any stage an unacceptable item is replaced by a good item.
 - (a) Find the probability a lot containing 15 unacceptable items is accepted.
 - (b) Graph the probability in part (a) as a function of D , the number of unacceptable items in the lot.
 - (c) Find the AOQL for this double sampling plan if unacceptable lots are rectified.
 - (d) Approximate the probability that the lot is accepted using the binomial distribution.
13. A lot of 400 items that actually contains 3 defective items is subject to the following double sampling plan: the lot is accepted if a first sample of 5 contains no defectives; the lot is rejected if this sample contains 2 or more defectives; if the first sample contains 1 defective, a second sample of 7 is drawn; the lot is accepted if the second sample contains no more than 1 defective, otherwise the lot is rejected. Suppose that at any stage an unacceptable item is replaced by a good item.
 - (a) What is the probability the lot is accepted?
 - (b) What is the AOQ if all the defective items in unacceptable lots are replaced by good items?
 - (c) Show the OC curve for the sampling plan.
14. A day's production of 200 compact disks is inspected as follows. If an initial sample of 15 shows at most 2 defective disks, the lot is accepted and is subject to no more sampling. However, if the first sample shows 3 or more defective disks, then a second sample of 20 disks is chosen and the lot is accepted if the total number of defectives in the two samples is no more than 4.
 - (a) Find the probability the lot is accepted if, in fact, it contains 10 defective disks.
 - (b) Find the AOQ.
 - (c) Plot the OC curve.
15. A random sample of 100 items is chosen from a lot of 4500 items, which is 2% defective. If the sample contains no more than 4 defective items, the lot is accepted; otherwise, the remainder of the lot is inspected and defective items are replaced by good items.
 - (a) What is the average number of items inspected?
 - (b) Graph the average number of items inspected as a function of the percentage defective in the lot.

2.12 THE HYPERGEOMETRIC RANDOM VARIABLE: FURTHER EXAMPLES

Example 2.12.1 A Lottery

Lottery games have become popular in many states. The game is played as follows: a player chooses five different white balls numbered from 1, 2, ..., 59. Another red ball is then chosen from balls numbered from 1, 2, ..., 35. This choice, called a *powerball*, may match one of the first five white balls chosen. Lottery officials then choose the five integers and the powerball.

The number of integers the player correctly chooses from among the first five is a hypergeometric random variable, which we will call X . Then

$$P(X = x) = \frac{\binom{5}{x} \cdot \binom{54}{5-x}}{\binom{59}{5}}, \quad x = 0, 1, \dots, 5.$$

Let Y denote the number of correct powerball choices made. Then

$$P(Y = y) = \frac{\binom{1}{y} \cdot \binom{34}{1-y}}{\binom{35}{1}}, \quad y = 0, 1.$$

Since the choices are independent, it follows that

$$P(X = x \text{ and } Y = y) = \frac{\binom{5}{x} \cdot \binom{54}{5-x}}{\binom{59}{5}} \cdot \frac{\binom{1}{y} \cdot \binom{34}{1-y}}{\binom{35}{1}}, \quad x = 0, 1, \dots, 5; \quad y = 0, 1.$$

Here is a table of values of X and Y giving the probabilities with which the possible values occur. The jackpot, J here, may vary from week to week.

X	Y	Probability	Payoff
5	1	$\frac{1}{175,223,510} = 5.7070 \times 10^{-9}$	J
5	0	$\frac{17}{87611755} = \frac{1}{5,153,633} = 1.9404 \times 10^{-7}$	\$1,000,000
4	1	$\frac{27}{17522351} = \frac{1}{648,976} = 1.5409 \times 10^{-6}$	\$10,000
4	0	$\frac{918}{17522351} = \frac{1}{19,079} = 5.2414 \times 10^{-5}$	\$100

3 1	$\frac{1431}{17522351} = \frac{1}{12,244} = 8.1673 \times 10^{-5}$	\$100
3 0	$\frac{48654}{17522351} = \frac{1}{361} = 2.7701 \times 10^{-3}$	\$7
2 1	$\frac{24804}{17522351} = \frac{1}{707} = 1.4144 \times 10^{-3}$	\$7
1 1	$\frac{316251}{35044702} = \frac{1}{110} = 9.0909 \times 10^{-3}$	\$4
0 1	$\frac{316251}{17522351} = \frac{1}{55} = 1.8182 \times 10^{-2}$	\$4

The probability the player selects at least one of the five integers correctly is

$$1 - \frac{\binom{5}{0} \cdot \binom{54}{5}}{\binom{59}{5}} = 0.3683.$$

The probability the player wins something is $\frac{1}{175223510} + \frac{17}{87611755} + \frac{27}{17522351} + \frac{918}{17522351} + \frac{1431}{17522351} + \frac{48654}{17522351} + \frac{24804}{17522351} + \frac{316251}{35044702} + \frac{316251}{17522351} = 0.0314007$. What is the value of a ticket? The expected value of a ticket is $\frac{1}{175223510} * J + \frac{17}{87611755} * 1000000 + \frac{27}{17522351} * 10000 + \frac{918}{17522351} * 100 + \frac{1431}{17522351} * 100 + \frac{48654}{17522351} * 7 + \frac{24804}{17522351} * 7 + \frac{316251}{35044702} * 4 + \frac{316251}{17522351} * 4 = \frac{1}{175223510} * J + 0.36049$.

So for the expected value of a ticket to be \$3, the cost of a ticket including the guess for the powerball, the jackpot, J , must be \$462,504,000.

Example 2.12.2 A Card Game

A bridge hand consists of 13 cards chosen without replacement from a deck of 52 cards. What is the most likely distribution of suits in the bridge hand?

In the hypergeometric random variable, the sampling is done from a population containing two kinds of items; here, we generalize the distribution somewhat to sample from a population containing four kinds of items, namely, the suits in the deck.

It would appear, since the suits all occur with equal frequency in the deck, that the most likely distribution of suits might be four of one suit and three each of the remaining three suits. This has probability

$$P(4, 3, 3, 3) = \frac{\binom{4}{3} \cdot \binom{13}{3}^3 \cdot \binom{13}{4}}{\binom{52}{13}}$$

since we first choose three suits; then we choose three cards from each of those suits; finally, we choose four cards from the remaining suit.

However, the distribution of four cards from each of two suits, three cards from one suit, and two cards from the remaining suit has probability

$$P(4, 4, 3, 2) = \frac{\binom{4}{2} \cdot \binom{13}{4}^2 \cdot \binom{2}{1} \cdot \binom{13}{3} \cdot \binom{13}{2}}{\binom{52}{13}}.$$

We find that $\frac{P(4, 4, 3, 2)}{P(4, 3, 3, 3)} = \frac{45}{22}$ so the distribution of four cards from each of two suits, three cards from another suit, and the remaining two cards from the fourth suit is over twice as likely as the more uniform distribution of suits. Any other combination of suits is less likely than the combination found here.

2.13 THE POISSON RANDOM VARIABLE

In Section 2.4, we considered the binomial random variable whose probability distribution is

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}, \quad x = 0, 1, 2, \dots, n$$

where $q = 1 - p$.

Events that have small probability – rare events – are of particular interest and we turn our attention to them.

We calculated the recursion from the binomial probability distribution function:

$$P(X = x) = \frac{p}{q} \cdot \frac{n - x + 1}{x} \cdot P(X = x - 1), \quad x = 1, 2, \dots, n. \quad (2.16)$$

Now notice that the recursion could also be written as

$$P(X = x) = \frac{np - p(x - 1)}{(1 - p) \cdot x} \cdot P(X = x - 1).$$

In this form of the recursion, we fix x and let $n \rightarrow \infty$ and $p \rightarrow 0$ while keeping np , which we denote by λ , fixed. These presumptions allow us to concentrate on events that are rare. We see that under these conditions, in the limit we have

$$P(X = x) = \frac{\lambda}{x} \cdot P(X = x - 1), \quad x = 1, 2, 3, \dots \quad (2.17)$$

Our task now is to determine the function $P(X = x)$, which satisfies the recursion (2.17). Applying the recursion repeatedly, we find that

$$\begin{aligned} P(X = 1) &= \frac{\lambda}{1} \cdot P(X = 0), \\ P(X = 2) &= \frac{\lambda}{2} \cdot P(X = 1) = \frac{\lambda^2}{1 \cdot 2} \cdot P(X = 0), \\ P(X = 3) &= \frac{\lambda}{3} \cdot P(X = 2) = \frac{\lambda^3}{1 \cdot 2 \cdot 3} \cdot P(X = 0), \end{aligned}$$

and so on. Since $\sum_{n=0}^{\infty} P(X = n) = 1$, it follows that

$$P(X = 0) \cdot \left\{ 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right\} = 1$$

or

$$\begin{aligned} P(X = 0) \cdot e^{\lambda} &= 1, \\ \text{so } P(X = 0) &= e^{-\lambda}, \\ P(X = 1) &= \lambda \cdot e^{-\lambda}, \end{aligned}$$

and

$$P(X = 2) = \frac{\lambda^2 \cdot e^{-\lambda}}{2!}.$$

We conjecture then that

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (2.18)$$

Formula (2.18) defines the *Poisson* probability distribution with parameter λ . The reader should check that (2.18) satisfies (2.17). It is also easy to check that (2.18) is a probability distribution. Obviously, $P(X = x) \geq 0$ and

$$\begin{aligned} \sum_{x=0}^{\infty} P(X = x) &= \sum_{x=0}^{\infty} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) = e^{-\lambda} \cdot e^{\lambda} = 1. \end{aligned}$$

So (2.18) is a probability distribution function.

Mean and Variance of the Poisson

It should be no surprise that the mean of the Poisson distribution is np since we found the Poisson by taking the limit of the binomial with mean np and keeping np fixed. The calculation is as follows:

$$\begin{aligned} \mu &= \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} \\ &= e^{-\lambda} \cdot \lambda \cdot \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda. \end{aligned}$$

For the variance, it is easiest to calculate $E[X \cdot (X - 1)]$ and then make use of the fact that

$$\text{Var}(X) = E[X \cdot (X - 1)] + E(X) - [E(X)]^2.$$

Here,

$$\begin{aligned} E[X \cdot (X - 1)] &= \sum_{x=0}^{\infty} x \cdot (x - 1) \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} \\ &= e^{-\lambda} \cdot \lambda^2 \cdot e^{\lambda} = \lambda^2, \end{aligned}$$

from which it follows that

$$\text{Var}(X) = \lambda.$$

That should not really be much of a surprise. After all, the variance of the binomial is $n \cdot p \cdot q = n \cdot p \cdot (1 - p) = \lambda \cdot (1 - p)$. To find the Poisson, we let λ stay fixed and $p \rightarrow 0$. So $\lambda \cdot (1 - p) \rightarrow \lambda$.

Some Comparisons

The Poisson distribution was derived here as an approximation to the binomial distribution. It is now interesting to compare some binomial distributions and their Poisson approximations to measure, to some extent, how close the approximation is. We use a computer algebra system to make the calculations and the graphs.

First, consider a binomial variable with $n = 20$ and $p = 0.03$. Here, the value of n is not particularly large nor is p particularly small.

We show here $P(X = x)$ using both the binomial distribution and the Poisson approximation.

X	Binomial	Poisson
0	0.548812	0.543794
1	0.329287	0.336358
2	0.0987861	0.0988297
3	0.0197572	0.0183395
4	0.00296358	0.00241061
5	0.00035563	0.000238576

So the values are very close. The graphs in Figure 2.25 reveal the same observation.

As n increases, the approximation is generally very good. Figure 2.26 shows a comparison between a binomial with $n = 100$ and $p = 0.030$ and a Poisson distribution with $\lambda = 3$.

The curves, which exhibit again the normal-like shape, are remarkably close and differ the most at the maximum point; this difference is 0.00343232.

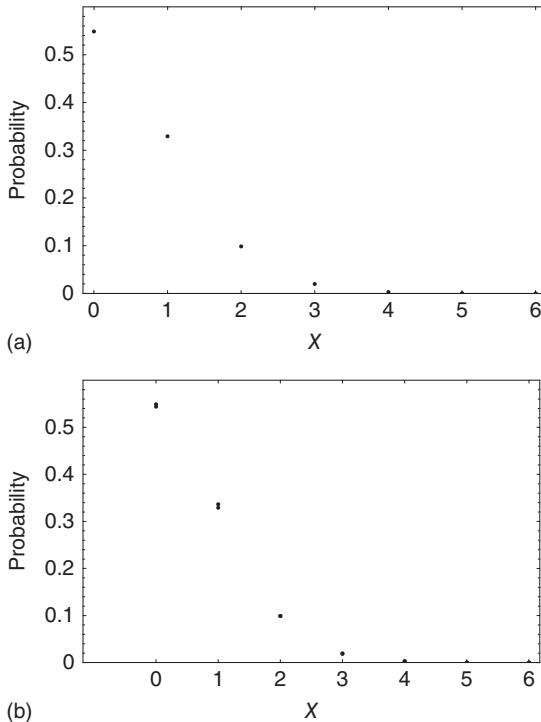


Figure 2.25 (a) Poisson distribution with parameter 0.60. (b) Binomial distribution with $n = 100$, $p = 0.03$.

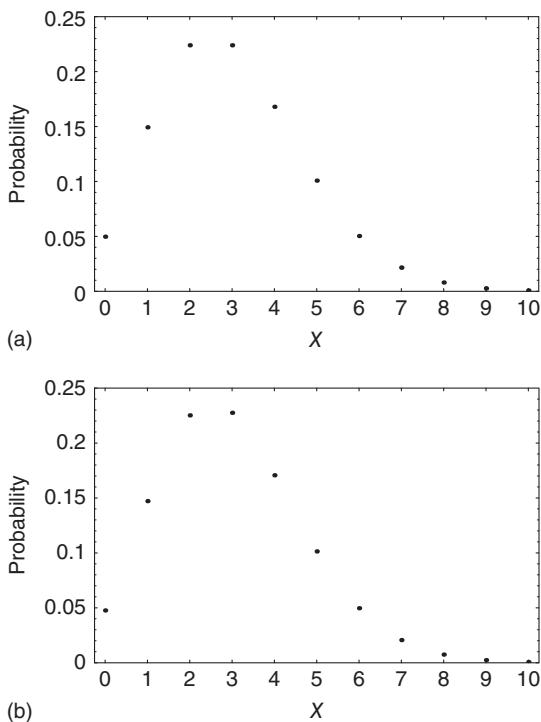


Figure 2.26 (a) Poisson distribution with parameter 3. (b) Binomial distribution with $n = 100$, $p = 0.03$.

Example 2.13.1

An acceptance sampling plan selects 5 items from a population of 500 items, 16 of which are unacceptable. The lot is accepted if at most 2 of the sampled items are unacceptable. Here, we compare the exact (hypergeometric) probability with both binomial and Poisson approximations:

The hypergeometric probability is

$$P(X \leq 2) = \sum_{x=0}^2 \frac{\binom{16}{x} \cdot \binom{484}{5-x}}{\binom{500}{5}} = 0.99974.$$

The binomial approximation uses $n = 5$ and $p = 16/500$ and is equivalent to no replacement, so

$$P(X \leq 2) = \sum_{x=0}^2 \binom{5}{x} \cdot \left(\frac{16}{500}\right)^x \cdot \left(\frac{484}{500}\right)^{5-x} = 0.999688.$$

Finally, the Poisson distribution with $\lambda = 5 \cdot \frac{16}{500} = 0.16$ gives

$$P(X \leq 2) = \sum_{x=0}^2 \frac{e^{-0.16} \cdot (0.16)^x}{x!} = 0.999394.$$

The approximations continue to be very good. Actually, the error in the Poisson distribution when approximating the binomial is not easily characterized, but many advise that its use is best when $np \leq 5$, a rule that generally works fairly well.

Figure 2.26 compares the binomial with $n = 100$ and $p = 0.03$ to the Poisson distribution with $\lambda = np = 3$. The Poisson approximation is seen to be remarkably good.

2.14 THE POISSON PROCESS

The Poisson distribution serves as an approximation to the binomial distribution. The binomial distribution models situations where we can observe both the number of successes and the number of failures for a certain number of trials of an experiment.

We now turn our attention to other events that occur in time or space. We may be interested in the following examples: the number of faults in a fixed length of optic cable; the number of customers arriving at a checkout counter in a store; the number of telephone calls received at a telephone switchboard; or the number of messages received at a computer terminal. In each of these examples, we can count the number of occurrences of an event (such as the number of faults in the optic cable), but we cannot count the number of failures. How can such phenomena be modeled?

Consider a continuous interval of length or time or space and suppose that the following are true for the events we wish to observe the following:

1. The number of events in intervals having no points in common is independent.

2. Consider a short interval of length h . Suppose the probability of exactly one event in this interval is $\lambda \cdot h$, where λ is a constant of proportionality.
3. The probability of more than one event in the interval of length h is 0.

Now divide an interval of unit length into n mutually exclusive parts. By assumption (2), the probability of exactly one event in this interval is $\lambda \cdot (1/n)$, and so the probability of no events in this interval is $1 - \lambda/n$. Letting X denote the number of events in the unit interval, assumptions (1) and (3) allow us to calculate

$$P(X = x) = \binom{n}{x} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

But we have shown that this can be approximated by a Poisson variable with mean value $n \cdot \frac{\lambda}{n} = \lambda$. So the Poisson distribution can be used in situations for which assumptions (1)–(3) hold. We see that λ is the expected number of events in a period of time or in an interval of space.

Example 2.14.1

Calls come into a telephone switchboard at a rate of 4 per minute.

- (a) Find the probability of exactly 6 calls in an interval of 2 minutes.
- (b) Find the probability of at least 3 calls in 3 minutes.

Here, the interval of interest changes. We present two different solutions.

Solution 1

In part (a) the interval is of length 2 minutes so we might suspect that λ , the expected number of events in that interval, is 8. Proceeding on that assumption, and letting X denote the number of calls received in that interval, we have

$$P(X = 6) = \frac{e^{-8} \cdot 8^6}{6!} = 0.122138.$$

In part (b), the interval is 3 minutes and so $\lambda = 12$ here. We find

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{x=0}^2 \frac{e^{-12} \cdot 12^x}{x!} = 0.999478.$$

It is not clear, however, that we can change the interval and retain a Poisson variable.

Solution 2

- (a) Let the random variables X_1 and X_2 be defined as follows:

Let X_1 denote the number of calls received during the first minute and let X_2 denote the number of calls received during the second minute.

Since the number of calls received during the first and second minutes are independent,

$$\begin{aligned} P(X_1 + X_2 = 6) &= \sum_{x_1=0}^6 P(X_1 = x_1) \cdot P(X_2 = 6 - x_1) \\ &= \sum_{x_1=0}^6 \frac{e^{-4} \cdot 4^{x_1}}{x_1!} \cdot \frac{e^{-4} \cdot 4^{6-x_1}}{(6-x_1)!} \end{aligned}$$

Multiply and divide by 6! to obtain

$$P(X_1 + X_2 = 6) = \frac{e^{-8}}{6!} \sum_{x_1=0}^6 \binom{6}{x_1} 4^{x_1} \cdot 4^{6-x_1}$$

and now by the binomial theorem,

$$P(X_1 + X_2 = 6) = \frac{e^{-8}}{6!} (4+4)^6 = \frac{e^{-8}}{6!} \cdot 8^6,$$

giving the same result as in Solution 1.

Solution 2 indicates that the sum of independent Poisson random variables is again a Poisson random variable. We will return to a discussion of this fact and related facts in Chapter 5 where Solution 1 will be completely justified.

EXERCISES 2.14

1. Show the Poisson approximation to the binomial distribution with $n = 5$ and probability 0.2 and draw a graph of these probabilities.
2. Show a recursion for the Poisson distribution in problem 1 and use it to calculate the probabilities in problem 1.
3. A Poisson random variable has

$$P(X = 2) = \frac{2}{3}P(X = 1).$$

Find $P(X = 3)$.

4. Deaths in a small city occur at a rate of 5 per week and are known to follow a Poisson distribution.
 - What is the expected number of deaths in a 3-day period?
 - What is the probability no one dies in a 3-day period?
 - What is the probability that at least 250 people die in 52 weeks?
5. Traffic accidents at an intersection are assumed to follow a Poisson distribution with 4 accidents expected in a period of 1 year.
 - What is the probability of at most 1 accident in a given year?
 - What is the probability of exactly 3 accidents in 6 months?
 - It is expected that 2 accidents occur during a year at another intersection. What is the probability that there is a total of at least 3 accidents in a given year at the two intersections?

6. The number of typographical errors per page in a book follows a Poisson distribution with parameter $3/4$. What is the probability that there is a total of 10 errors on 10 randomly selected pages in the book?
7. Twenty percent of the IC chips made in a plant are nonfunctional. Assume that a binomial model is appropriate.
 - (a) Find the probability that at most 13 nonfunctional chips occur in a sample of 100 chips.
 - (b) Use the Poisson distribution to approximate the result in part (a).
8. Let X , the number of hits in a baseball game, be a Poisson variable with parameter α . If the probability of a no-hit game is $1/3$, what is α ?
9. An insurance company has discovered that about 0.1% of the population is involved in a certain type of accident each year. If the 10,000 policy holders of the company are randomly selected from the population, what is the probability that not more than 5 of its clients are involved in such an accident next year?
10. A study of customers entering a grocery store shows that all the arrivals are Poisson with males entering on an average rate of 3 per minute and females at an average rate of 5 per minute. Find the probability that at least 20 customers enter the store in the next 5 minutes.
11. Computer programs run on a certain computer are executed during an interval of 1 minute according to a Poisson process with mean 12. Twenty-five percent of these programs utilize a plotter.
 - (a) What is the probability there will be a demand for at least 15 programs run in a given minute?
 - (b) The plotter takes 10 seconds to execute a plot. What is the expected number of seconds the plotter is in use during a given minute?
12. A multiple choice examination contains 4 choices for each of 100 questions.
 - (a) Find the exact probability a student who guesses misses at most 4 questions.
 - (b) Approximate the probability in part (a) using the Poisson distribution.
13. The number of earthquakes of destructive magnitude in California follows a Poisson distribution with one such earthquake expected each year. What is the probability of at least three such earthquakes in a 6-month period?
14. A quality control inspector follows the following plan in inspecting soccer balls that are produced according to a Poisson process with four soccer balls expected each minute. The produced balls fall into a bin that automatically empties at the end of each minute. If the bin collects exactly three balls, the inspector takes them out for possible inspection of 10 seconds each. He flips a fair coin for each and inspects them only if heads appear. If the bin should contain five balls, he spends 5 seconds inspecting each ball. Otherwise, the inspector does not inspect the output. What is the average amount of time per minute spent in inspecting the soccer balls?
15. Major crimes are reported at an average rate of 5 per night in a given police precinct. The number of these crimes is assumed to follow a Poisson distribution.
 - (a) What is the probability that on a given night no more than three major crimes will be reported?
 - (b) What is the chance that a full week will pass with no more than three major crimes reported on any of the seven nights?

- 16.** An airline knows that 10% of the people holding reservations for a certain flight will not appear. The plane holds 90 people. Use the Poisson approximation in answering the following questions:
- If 95 reservations have been sold, what is the probability that everyone who appears for the flight can be accommodated?
 - How many reservations should be sold so that the probability the airline can accommodate everyone who appears is at least 0.99?
- 17.** Molecules of a rare gas occur at an average rate of 3 per cubic foot of air and follow a Poisson distribution.
- What is the probability that a cubic foot of air contains none of the molecules?
 - What is the probability that 3 cubic feet of air contain exactly four of the molecules?
 - How much air must be taken as a sample to make the probability at least 0.99 so that at least one molecule will be found?
- 18.** A librarian shelves 1000 books per day. If the probability any particular book is misshelved is 0.001 and if the books are shelved independently of each other,
- What is the probability that at most 2 books are misshelved?
 - Approximate the probability in part (a) using the Poisson distribution.
- 19.** A popular chocolate chip cookie “guarantees” at least 16 chocolate chips per cookie. The actual number of chocolate chips per cookie, however, is a Poisson random variable. What must be the average number of chips per cookie if approximately 95% or more of the cookies are to meet the guarantee?
- 20.** A bakery makes a batch of 1000 chocolate chip cookies and adds n chocolate chips to the batter for each batch and mixes the batter well. Under these assumptions, the number of chocolate chips per cookie should follow a Poisson distribution.
- If $n = 4900$, what is the probability that at least 2 chips are in a randomly selected cookie?
 - If $n = 4900$, what is the number of cookies in each batch that are expected to contain exactly 3 chocolate chips?
 - FDA regulations declare that at most 1% of cookies labeled “chocolate chip” can fail to contain a single chocolate chip. What is the minimum value for n for the bakery to be within the law?
- 21.** A truck repair shop has facilities for the repair of 3 large trucks per day. The trucks arrive according to a Poisson process with 2 trucks expected per day. If more than 3 trucks arrive, the excess is turned away.
- Find the probability exactly 3 trucks arrive in 1 day.
 - Find the probability that trucks are turned away.
 - Find the probability distribution for X , the number of trucks serviced per day.
 - Find the expected number of trucks turned away each day.
 - The shop decides to add facilities so that it can service the trucks arriving during a day about 95% of the time. How many trucks must it be able to service in a day?
- 22.** Calls come into a very busy switchboard at a rate of 6 per minute according to a Poisson process. Unfortunately some new electronic switching devices work imperfectly and the probability a received call is switched to the proper extension is only 0.8. It has been observed that the calls are switched independently, however.

- (a) If X represents the number of calls correctly switched, find $P(X = k)$ for some 1-minute period.
- (b) Simplify the result in part (a) and show that X is Poisson with parameter 4.8.
23. Telephone calls coming into a busy switchboard follow a Poisson distribution with 4 calls expected in a 1-minute period. The switchboard, however, can answer at most 6 calls in a 1-minute interval; any calls exceeding 6 during that period receive a busy signal.
- (a) Let Y denote the number of calls answered in a 1-minute period. Find the probability distribution for Y .
- (b) Find $E(Y)$.

CHAPTER REVIEW

This chapter considers several discrete probability distributions whose importance arises from the fact that they have various applications. Each of the distributions in this chapter arises in one way or another from the binomial distribution.

We began by defining a *random variable* as a real-valued function defined on the points of a sample space. A typical example is throwing two dice and then recording the sum that appears. The sum is a random variable since it is a function, in this case the sum, of the outcomes of the particular sample point that occurs.

If X is a random variable, we defined the *probability distribution function*, or *pdf*, as

$$f(x) = P(X = x).$$

A related function is the *distribution function*, defined as

$$F(x) = P(X \leq x).$$

The distribution function is not often used in this chapter, but has very important applications in the work to come.

Probability distributions are often distinguished and described by the values of their *mean*, μ_x , and their variance, σ_x^2 . These are defined as

$$\begin{aligned}\mu_x &= E(X) = \sum_x x \cdot f(x) \quad \text{and} \\ \sigma_x^2 &= \text{Var}(X) = E(X - \mu_x)^2 = \sum_x (x - \mu_x)^2 f(x)\end{aligned}$$

provided, of course, that the sums exist. The variance, σ_x^2 , can also be calculated as

$$\sigma_x^2 = E(X^2) - [E(X)]^2.$$

As a (crude) indication that σ actually measures the variation, or dispersion in a random variable, we proved Tchebycheff's inequality:

$$P(|X - \mu| \leq k \cdot \sigma) \geq 1 - \frac{1}{k^2},$$

where k is some positive quantity.

We then turned to some specific discrete probability distributions. Of these, the single most important probability distribution is the *binomial distribution* whose pdf is given by

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n \text{ where } q = 1 - p.$$

This random variable arises from an experiment of n independent trials on each of which the result is one of two outcomes (usually denoted by “success” or “failure”), where p denotes the probability of success and X denotes the total number of successes.

We used a recursion to find that, for the binomial distribution,

$$\begin{aligned}\mu &= n \cdot p \\ \text{and } \sigma^2 &= n \cdot p \cdot q.\end{aligned}$$

We then considered some statistical problems. We first considered the construction of a *confidence interval* when sampling from a binomial distribution with known values of n and p . Frequently, however, p is unknown. We found an approximate 95% confidence interval for p to be

$$\begin{aligned}P\left(\frac{nX + 2n - 2\sqrt{n^2X + n^2 - nX^2}}{n^2 + 4n} \leq p \leq \frac{nX + 2n + 2\sqrt{n^2X + n^2 - nX^2}}{n^2 + 4n}\right) \\ = 0.95\end{aligned}$$

where X is the observed number of successes in the binomial process with n trials.

Tests of hypotheses were then considered. We examined tests of

$$H_0: p = p_0 \text{ against the alternative}$$

$$H_a: p = p_a.$$

The two types of error in testing a hypothesis are

$$\alpha = \text{Probability of a Type I error}$$

$$= P[H_0 \text{ is rejected when it is true}] \text{ and}$$

$$\beta = \text{Probability of a Type II error}$$

$$= P[H_0 \text{ is accepted when it is false}].$$

We considered the effect of the *critical region* – the set of observed values leading to the rejection of the null hypothesis – on the size of β and discussed $1 - \beta$, the *power* of the test. This is the probability that a false H_0 is correctly rejected.

We derived the mean and the variance of a *sample proportion* arising from a sample survey. Using these results, we found that

$$P\left(p_s - 2 \cdot \sqrt{\frac{p \cdot q}{n}} \leq p \leq p_s + 2 \cdot \sqrt{\frac{p \cdot q}{n}}\right) = 0.95$$

is a 95% confidence interval for the unknown true proportion p based on a sample proportion p_s .

The negative binomial distribution arises when we, in a binomial experiment, wait for, say, the r th success. The probability distribution function is

$$P(X = x) = \binom{x-1}{r-1} \cdot p^r \cdot q^{x-r}, \quad x = r, r+1, r+2, \dots$$

with mean $\mu = \frac{r}{p}$ and variance $\sigma^2 = \frac{r \cdot q}{p^2}$. In the special case where $r = 1$, so that we wait for the first binomial success, X is called a *geometric* random variable.

A common situation in which the *hypergeometric* random variable arises is that of *acceptance sampling*. Here a lot, or a collection of a product manufactured over a given period of time, is sampled, but, unlike the binomial distribution, the sampling is done without replacement. If the lot actually contains D unacceptable items and $N - D$ acceptable items and if X denotes the number of unacceptable items in a sample of size n , then

$$P(X = x) = \frac{\binom{D}{x} \cdot \binom{N-D}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \dots, \min\{n, D\}.$$

We found that $\mu = n \cdot \frac{D}{N}$ and that $\sigma^2 = n \cdot \frac{N-n}{N-1} \cdot \frac{D}{N} \cdot \frac{N-D}{N}$. We showed that the hypergeometric random variable is approximated by the binomial random variable when the sample size, n , is small in comparison to the lot size, N . Examples of acceptance sampling were given, and we considered two plans for improving the quality of the lot of items sent to the buyer. In one plan we replaced any unacceptable items in the sample by good items; in the second plan, if the sample so indicated, we replaced every unacceptable item in the lot by a good item. Each plan leads to gains with respect to the quality of the outgoing product; under the second plan there is a limit of the percentage of unacceptable product that can be sold. This is known as the *AOQ limit*.

Finally, we considered a *Poisson* random variable that can be regarded in two ways: we first found the distribution as a limit of the binomial distribution when n is large and p is small. We also considered the *Poisson process* in which events occur over a period of time or space in an independent manner so that the probability of more than one independent event in a given interval is negligible, and that the probability of an event in some interval is proportional to the length of the interval. These assumptions yield the same distribution as the limiting binomial distribution. The Poisson distribution has a variety of applications, many of which were given in the exercises.

In the next chapter, we will consider some important continuous probability distributions.

PROBLEMS FOR REVIEW

Exercises 2.2 # 1, 3, 4, 8, 11

Exercises 2.3 # 1, 2, 5, 6, 7, 11

Exercises 2.5 # 1, 2, 4, 7, 8, 9, 12, 15, 17, 19, 22

- Exercises 2.6 # 1, 4, 5, 8, 9
- Exercises 2.7 # 1, 4, 6, 7, 10
- Exercises 2.8 # 1, 3, 4, 8, 9
- Exercises 2.9 # 1, 2, 5, 7, 9, 13
- Exercises 2.10 # 1, 3, 5, 6, 8, 10
- Exercises 2.11 # 2, 4, 5, 6, 9, 11
- Exercises 2.14 # 2, 3, 5, 6, 8, 13, 18, 20

SUPPLEMENTARY EXERCISES FOR CHAPTER 2

1. Calls come into a telephone exchange at a rate of 1.5 per minute. Assuming that the number of calls received follows a Poisson distribution, find the probability that at least 3 calls are received in the next 4 minutes.
2. Twenty percent of the IC chips made in a plant are defective. Assume that the chips are produced according to a binomial process.
 - (a) Find the probability that at most 13 defectives occur in a sample of 100 IC chips.
 - (b) Approximate the probability in part (a) by a Poisson random variable.
3. A manufacturer of soft drink bottles turns out defectives with probability 0.10. Assume that the bottles are produced according to a binomial process.
 - (a) Find the probability that there are 4 defective bottles among the next 10 bottles produced.
 - (b) Find the probability that there are at least 4 defective bottles among the next 10 bottles produced.
 - (c) How many bottles must be produced to make the probability that at least one bottle among them is defective to be at least 0.95?
4. Earthquakes in a certain part of California occur according to a Poisson process with three earthquakes expected each century.
 - (a) What is the probability of exactly four earthquakes in a century?
 - (b) What is the probability of at least two earthquakes in a 50-year period?
 - (c) Let X be the number of earthquakes in a century. Compare the exact value of $P(\mu - \sigma \leq X \leq \mu + \sigma)$ with the approximation given by Tchebycheff's inequality.
5. Suppose an event has probability p of occurring and that several independent trials are observed. What value of p maximizes the probability that the first failure occurs on the fifth trial?
6. Suppose that X and Y are independent observations of a Poisson random variable with parameter $\mu = 1$. Find the probability that the smallest of the two observations is 1.
7. A series of trials in which success or failure occurs on each trial has probability of success at the i th trial as $\frac{1}{i+1}$. In three trials, find the probability of exactly 2 successes.
8. A manufacturer makes a lot of 10 items a day. Two items are drawn (without replacement) and inspected. The lot is accepted if the sample contains at most 1 defective item. Find the probability a lot containing 3 defective items is accepted.
9. (a) What is the probability that a poker hand contains exactly 2 aces?

- (b) How many poker hands must be selected to make the probability of having at least one hand containing at least 2 aces be at least 0.99?
10. A store sells chocolate donuts at a rate of 16 per hour, the number sold following a Poisson distribution. Find the probability that the store sells at least 3 chocolate donuts in 15 minutes.
11. Five defective transistors are mixed up with 10 good transistors. They are inspected one after another until all the good transistors have been found. What is the probability the last good transistor will be found on the 12th test?
12. Errors are known to occur in a digitized message in a communications channel; the probability an individual bit is incorrectly transmitted is 0.001 and the errors are assumed to be independent.
- (a) Find the probability that at most 2 errors occur in a sequence of 10 bits.
 - (b) Find the mean and variance of the number of errors.
 - (c) Find the probability of at most 2 errors in a message of 10,000 bits.
13. In a small voting precinct, 100 voters favor candidate A and 80 voters oppose candidate A. What is the probability that a majority of a random sample of 4 voters will oppose candidate A?
14. Customers arrive at a checkout counter in a supermarket at a rate of 20 per half hour, the number following a Poisson distribution. What is the probability that at most 5 customers arrive in a period of 15 minutes?
15. A manufacturer produces items that are good or defective, according to a binomial process where p is the probability an item is defective. Let X denote the number of items produced up to and including the second defective item.
- (a) Find an expression for the probability that X is even.
 - (b) Now suppose that the sixth item is the second defective item produced. What is the most likely value for p ?
16. Thirty percent of the applicants for a position have advanced training in computer programming. Three jobs requiring advanced training are open. Find the probability that the third qualified applicant is found on the fifth interview, supposing that the applicants are interviewed sequentially and at random.
17. A fair die is tossed until a 5 or a 6 appears. Compute the probability that the number of tosses is a multiple of 4.
18. From a lot of 25 items, 5 of which are defective and 4 are chosen at random. Let X be the number of defectives found. Find the probability distribution of X if
- (a) 1. the items are chosen with replacement.
2. the items are chosen without replacement.
 - (b) In part (a) assume that the items are chosen with replacement until a defective item is found. What is the probability an odd number of drawings is necessary?
19. Customers arrive at a computer store according to a Poisson process with 5 customers expected per hour. The sales force can accommodate at most 10 customers per hour; if more than 10 customers appear in an hour, the excess must be turned away.
- (a) What is the probability that customers are turned away in a 1-hour period?
 - (b) Consider two independent 1-hour intervals. Let X denote the number of arrivals during the first hour and Y the number of arrivals during the second hour. Find $P(X + Y \leq 8)$.

- 20.** Fifty chocolate chip cookies are to be made using 150 chocolate chips. The number of chocolate chips per cookie is a Poisson random variable.
- What is the probability a cookie has at least 4 chocolate chips?
 - How many chocolate chips must be used in order to make the probability a cookie has at least one chocolate chip be at least 0.90?
- 21.** A pair of fair dice is rolled 180 times each hour in a dice game at a casino. What is the probability that at least 25 rolls give a sum of 7 during 1 hour?
- 22.** Telephone calls come into an answering service at an average rate of 3 per hour, the number of calls following a Poisson distribution. During the noon hour, only the first 3 calls are answered. What is the expected number of calls answered during the noon hour?
- 23.** A box contains 4 bad and 6 good tubes. The tubes are checked by drawing a tube at random and not replacing it in the box. In how many ways can the fourth bad tube be found on the seventh drawing?
- 24.** A box contains three blue and four yellow marbles. Marbles are drawn out one at a time, the drawn marbles not being replaced. Drawings are made until all the marbles remaining in the box are of the same color.
- Assign probabilities to the sample points and verify that their sum is 1.
 - What is the probability that only yellow marbles remain in the box when the sampling is finished?
- 25.** A tosses three coins that have probability p_A of coming up heads while B tosses two coins that have probability p_B of coming up heads.
- Find an expression for the probability that A tosses more heads than B.
 - Show that the game is fair if the coins are fair.
- 26.** A player pays \$A to play the following game: a coin, loaded to come up heads with probability $2/3$, is tossed five times. Let X denote the number of heads. The player wins $\$(X + 1)$ if X is even and wins $\$(X - 1)$ if X is odd. Find A so that the game is fair.
- 27.** A machine, producing defective parts with probability $1/10$, has produced five parts. Unknown to the operator of the machine, an adjustment to the machine increases this probability to $1/5$. Ten parts are produced after the adjustment. What is the probability the output contains at least 2 defectives? Assume the parts are produced according to a binomial process.
- 28.** Past studies have shown that $2/3$ of professional football players will sustain a permanent injury before retiring. To see if this proportion is true for current players, a sample of 100 retired professional football players showed that 80 of them had sustained permanent injuries. Using $\alpha = 0.05$, test $H_0: p = 2/3$ against $H_a: p > 2/3$.
- 29.** In problem 28, find the size of β for the alternative $p = 0.72$.
- 30.** A study of 1200 college students showed that 44% of them said that their political views were similar to those of their parents. Find a 95% confidence interval for the true proportion of college students whose political views are similar to those of their parents.
- 31.** A drug is thought to be effective in 10% of patients with a certain condition. To test this hypothesis, the drug is given to 100 randomly chosen patients with the condition. If 8 or more show some improvement, $H_0: p = 0.10$ is accepted; otherwise, $H_a: p < 0.10$ is accepted. Find the size of the test.

- 32.** Jack thinks that he can guess the correct answer to a multiple choice question with probability $1/2$. Kaylyn thinks his probability is $1/3$. To decide who is correct, Jack takes a multiple choice test, guessing the answer to each question. If he answers at least 40 out of 100 questions correctly, it will be decided that Jack is correct. Find α and β for this test.
- 33.** A survey of 300 workers showed that 100 are self-employed. Find a 90% confidence interval for the proportion of workers who are self-employed.
- 34.** A management study showed that $1/3$ of American office workers has his or her own office while $1/33$ of Japanese office workers has his or her own office. The study was based on 300 American workers and 300 Japanese workers. Could the difference in these proportions only be apparent and due to sampling variability? [Use 90% confidence intervals.]
- 35.** The Internal Revenue Service says that the chance a United States Corporation will have its income tax return audited is 1 in 15. A sample of 75 corporate income tax returns showed that 6 were audited. Does the data support the Internal Revenue Service's claim? Use $\alpha = 0.05$.
- 36.** A survey of 400 children showed that $1/8$ of them were on welfare. Find a 95% confidence interval for the true proportion of children on welfare.
- 37.** How large a sample is necessary to estimate the proportion of people who do not know whose picture is on the \$1 bill to within 0.02 with probability 0.90?
- 38.** Three marbles are drawn without replacement from a bag containing three white, three red, and five green marbles. \$1 is won for each red selected and \$1 is lost for each white selected. No payoff is associated with the green marbles. Let X denote the net winnings from the game. Find the probability distribution function for X .
- 39.** Three fair dice are rolled. You as the bettor are allowed to bet \$1 on the occurrence of one of the integers 1, 2, 3, 4, 5, or 6. If you bet on X and X occurs k times ($k = 1, 2, 3$), then you win $$k$; otherwise, you lose the \$1 you bet. Let W represent the net winnings per play.
- Find the probability distribution for W .
 - Find $E(W)$.
 - If you could roll m dice, instead of 3 dice, what would your choice of m be?
- 40.** (a) Suppose that X is a Poisson random variable with parameter λ . Find λ if $P(X = 2) = P(X = 3)$.
- (b) Show if X is a Poisson random variable with parameter λ , where λ is an integer, then some two consecutive values of X have equal probabilities.
- 41.** Calls come into an office according to a Poisson process with 3 calls expected per hour. Suppose that the calls are answered independently, with the probability that a call is answered as $3/4$. Find the probability that exactly 4 calls are answered in a 1-hour period.
- 42.** Let X be Poisson with parameter λ .
- Find a recursion for $P(X = x + 1)$ in terms of $P(X = x)$.
 - Use the recursion in part (a) to find μ and σ^2 .
- 43.** Ten people are wearing badges numbered 1, 2, ... 10. Three people are asked to leave the room. What is the probability that the smallest badge number among the three is 5?

Chapter 3

Continuous Random Variables and Probability Distributions

3.1 INTRODUCTION

Discrete random variables were discussed in Chapter 2. However, it is not always possible to describe all the possible outcomes of an experiment with a finite, or countably infinite, sample space. As an example, consider the wheel shown in Figure 3.1 where the numbers from 0 to 1 have been marked on the outside edge.

The experiment consists of spinning the spinner and recording where the arrow stops. It would be natural here to consider the sample space, S , to be

$$S = \{x | 0 \leq x \leq 1\}.$$

S is infinite, but not countably infinite.

Now the question arises, “What probability should be put on each of the points in S ? ” Surely, if the wheel is fair, each point should receive the same probability and the total probability should be 1. What value should that probability be?

Suppose, for the sake of argument, that a probability of $0.00000000000000000000000000000001 = 10^{-22}$ is put on each point. It is easy to show that the circumference of the wheel contains more than 10^{22} points, so we have used up more than the allotted probability of 1. So we conclude that the only possible assignment of probabilities is

$$P(X = x) = 0 \text{ for any } x \text{ in } S.$$

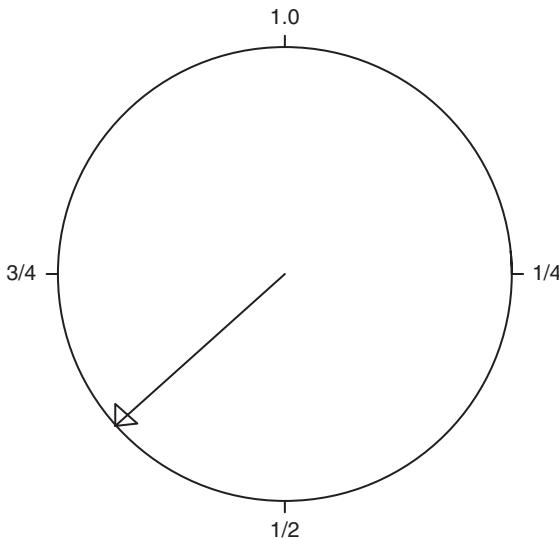
Now suppose that the wheel is loaded and that it is three times as likely that the arrow lands in the left-hand half of the wheel than in the right-hand half. We suppose that

$$P\left(X \geq \frac{1}{2}\right) = 3 \cdot P\left(X \leq \frac{1}{2}\right).$$

Again we ask, “What probability should be put on each of the points in S ? ” Again, since there is still an uncountably infinite number of points in S , the answer is

$$P(X = x) = 0.$$

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

**Figure 3.1** The spinner.

Definition If a random variable X takes values on an interval or intervals, then X is said to be a *continuous* random variable.

Of course, $P(X = x) = 0$ for *any* continuous random variable X .

So the probability distribution function is not informative in the continuous case, since, for example, we cannot distinguish between fair and loaded wheels! The fault, however, lies not in the answer, but in the question. Perhaps we can devise a question whose answer carries more information for us.

Consider now a function, $f(x)$, which we will call a *probability density function*. The abbreviation is again *pdf* (the same abbreviation used for probability distribution function), but the word *density* connotes a continuous distribution. Here are the properties we desire of the new function $f(x)$:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. If $a \leq b$, then $\int_a^b f(x) dx = P(a \leq X \leq b)$

These properties are quite analogous to those for a discrete random variable. Property (3) indicates that *areas* under $f(x)$ are probabilities, $f(x)$ must be nonnegative, else we encounter negative probabilities, so property (1) must hold. Property (2) indicates that the total probability on the sample space is 1.

What is $f(x)$ for the fair wheel? Since the circumference of the wheel contains the interval $[0, 1/4]$, and since the wheel is a fair one, we would like $P(0 \leq X \leq \frac{1}{4})$ to be $\frac{1}{4}$

so we must have $\int_0^{\frac{1}{4}} f(x) dx = \frac{1}{4}$. Many functions have this property. But we would like *any* interval of length $\frac{1}{4}$ to have probability $\frac{1}{4}$. In addition, we would like an interval of length a , say, to have probability a for $0 \leq a \leq 1$. The only function that has this property, in addition, to satisfying the above-mentioned properties (1) and (2) is a *uniform* probability density function:

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

For the loaded wheel, where we want $P(X \geq \frac{1}{2}) = 3P(X \leq \frac{1}{2})$, consider (among many other choices) the function

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then $P(X \geq \frac{1}{2}) = \int_{\frac{1}{2}}^1 2x \, dx = \frac{3}{4}$, so that $P(X \leq \frac{1}{2}) = \frac{1}{4}$ and so $P(X \geq \frac{1}{2}) = 3P(X \leq \frac{1}{2})$.

A graph of $f(x)$ is shown in Figure 3.2.

It is also easy to verify that $f(x)$ also satisfies properties (1) and (2) for a probability density function.

We see that $f(x)$, the probability *density* function, distinguishes continuous random variables in an informative way while the probability *distribution* function (which is useful for discrete random variables) does not.

To illustrate this point further, suppose the wheel has been rigged so that it is impossible for the pointer to stop between 0 and $\frac{1}{4}$, while it is still fair for the remainder of the circumference of the wheel. It follows then that

$$f(x) = \begin{cases} \frac{4}{3}, & \frac{1}{4} \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

This function satisfies all three properties for a probability density function. Its graph is shown in Figure 3.3. For this rigged wheel,

$$P\left(X \geq \frac{1}{2}\right) = \int_{\frac{1}{2}}^1 \frac{4}{3} \, dx = \frac{2}{3}.$$

It is also useful to define a *cumulative distribution function* (often abbreviated to *distribution function*), which is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) \, dx.$$

We used $F(x)$ in Chapter 2.

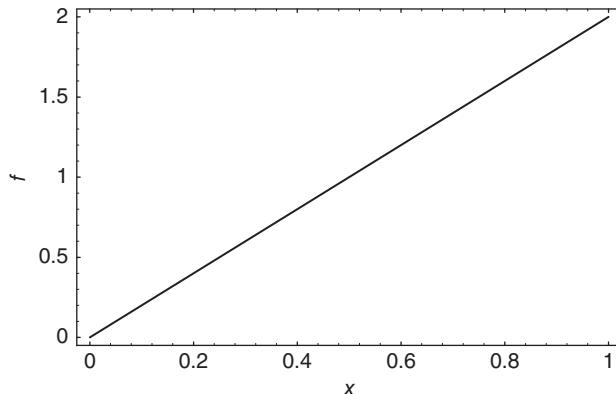


Figure 3.2 Probability density function for the loaded wheel.

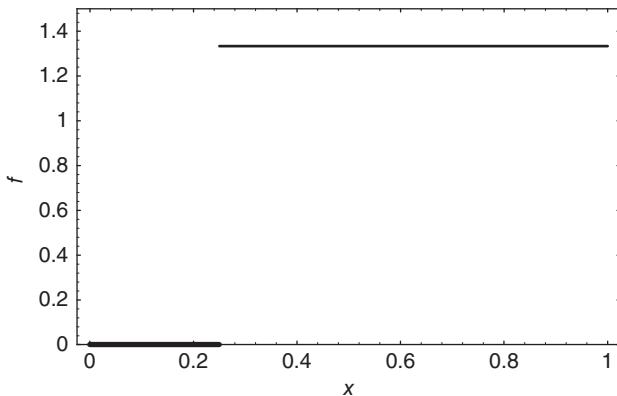


Figure 3.3 Probability density function for the rigged wheel.

The function $F(x)$ accumulates probabilities for a probability density function in exactly the same way as $F(x)$ accumulated probabilities in the discrete case.

As an example, if

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

then, being careful to distinguish the various regions in which x can be found, we find that

$$F(x) = \begin{cases} \int_{-\infty}^x 0 \, dx = 0, & x \leq 0 \\ \int_{-\infty}^0 0 \, dx + \int_0^x 1 \, dx = x, & 0 \leq x \leq 1 \\ \int_{-\infty}^0 0 \, dx + \int_0^1 1 \, dx + \int_1^\infty 0 \, dx = 1, & x \geq 1. \end{cases}$$

A graph of $F(x)$ is shown in Figure 3.4.

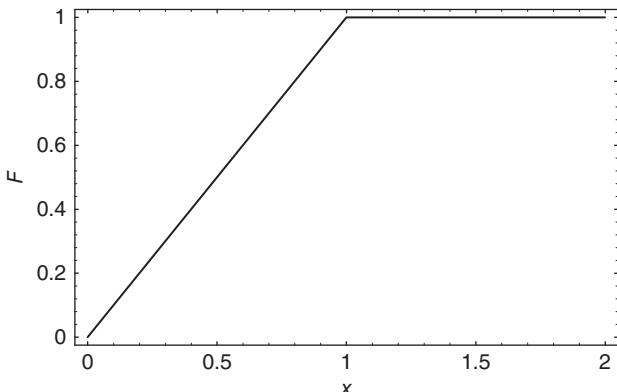


Figure 3.4 Cumulative distribution function for the fair wheel.

It is easy to see that $F(x)$, for any probability density function, $f(x)$, has the following properties:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} F(x) = 1,$$

If $a \leq b$, then $F(a) \leq F(b)$,

$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$, and

$$\frac{d[F(x)]}{dx} = f(x), \text{ provided the derivative exists.}$$

Mean and Variance

In analogy with the discrete case, the mean and variance of a continuous random variable with probability density function $f(x)$ are defined as

$$E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x) dx \text{ and} \tag{3.1}$$

$$\text{Var}(X) = \sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx \tag{3.2}$$

provided that the integrals converge.

These definitions are similar to those used for discrete random variables where the values of the random variable were weighted with their probabilities and the results added. Now it is natural to integrate so that the definitions for the mean and the variance in the continuous case appear to be analogous to their counterparts in the discrete case.

We can expand the definition of $\text{Var}(X)$ to find that

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

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

This is the same result we obtained for a discrete random variable.

Other properties of the mean and variance are as follows:

$$E(aX + b) = aE(X) + b$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

To show these properties, first consider $E(aX + b)$. By definition,

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) \cdot f(x) dx.$$

Expanding and simplifying the integrals, we find that

$$E(aX + b) = a \int_{-\infty}^{\infty} xf(x) dx + b \int_{-\infty}^{\infty} f(x) dx,$$

$$\text{so } E(aX + b) = aE(X) + b$$

or

$$E(aX + b) = a \cdot \mu + b,$$

establishing the first property.

Now

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b) - (a\mu + b)]^2 \\ &= E[a^2(X - \mu)^2] = a^2 E[(X - \mu)^2], \\ \text{so } \text{Var}(aX + b) &= a^2 \text{Var}(X), \end{aligned}$$

establishing the second property.

The definitions of the mean and variance are dependent on the convergence of the integrals involved. To show that this does not always happen, consider the density

$$f(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty.$$

The fact that

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi} \text{Arc tan}(x)|_{-\infty}^{\infty} = \frac{1}{\pi} \left[\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right] = 1$$

together with the fact that $f(x) \geq 0$ establishes $f(x)$ as a probability density function. However,

$$E(X) = \int_{-\infty}^{\infty} \frac{x \cdot dx}{\pi(1+x^2)} = \frac{1}{2\pi} \ln|x||_{-\infty}^{\infty} \text{ which does not exist.}$$

The random variable X in this case has no variance as well; in fact $E[X^k]$ does not exist for any k . The probability density is called the *Cauchy* density.

We now turn to an example of a better behaved probability density function.

Example 3.1.1

Given the loaded wheel for which

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

we find that

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x \geq 1. \end{cases}$$

If we want to calculate $P\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right)$, we could proceed in two different ways. First,

$$P\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right) = \int_{\frac{1}{2}}^{\frac{3}{4}} 2x \, dx = \frac{5}{16},$$

where $f(x)$ was used in the calculation. We could as easily have used $F(x)$:

$$P\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right) = F\left(\frac{3}{4}\right) - F\left(\frac{1}{2}\right) = \frac{5}{16},$$

giving the same result.

It would appear from this example that $F(x)$ is superfluous, since any probability can be found from a knowledge of $f(x)$ alone (and in fact $f(x)$ is needed to determine $F(x)!$). While this is true, it happens that there are other important uses to which $F(x)$ will be put later and so we introduce the function now. To pique the reader's interest, we pose the following question: the loaded wheel above is spun, X being the result. The player then wins $\$3X^2$. If the owner of the wheel wishes to make, on average, \$0.50 per play of the game, what is a fair price to charge to play the game? We will answer this question later, making use of $F(x)$, although the reader may be able to answer it now. The function $F(x)$ also plays a leading role in reliability theory, which is considered later in this chapter.

Example 3.1.2

A random variable X has probability density function

$$f(x) = \begin{cases} k \cdot (2 - x), & 0 \leq x \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

The constant k , of course, is a special value that makes the total area under the curve 1.

It follows that

$$\int_0^2 k \cdot (2 - x) \, dx = 1.$$

It follows from this that $k = 1/2$.

Now if we wish to find a conditional probability, for example, $P(X \geq 1 | X \geq \frac{1}{2})$, first note that the set of values where $X \geq \frac{1}{2}$ does not have area 1, so, as in the discrete case,

$$P\left(X \geq 1 | X \geq \frac{1}{2}\right) = \frac{P\left(X \geq 1 \text{ and } X \geq \frac{1}{2}\right)}{P\left(X \geq \frac{1}{2}\right)}.$$

This becomes

$$P\left(X \geq 1 | X \geq \frac{1}{2}\right) = \frac{P(X \geq 1)}{P\left(X \geq \frac{1}{2}\right)}.$$

We calculate this conditional probability as 4/9.

Before turning to the exercises, we note that there are many important special probability density functions of great interest since they arise in interesting and practical situations. We will consider some of these in detail in the remainder of this chapter.

A Word on Words

We considered, for a discrete random variable, the probability *distribution* function as well as the cumulative *distribution* function. For continuous random variables, the terms probability *density* function and cumulative *distribution* function are terms in common usage.

We will continue to make the distinction here between discrete and continuous random variables by making a distinction in the language we use to refer to them. In part, this is because the mathematics useful for discrete random variables is quite different from that for continuous random variables; the language serves to alert us to these distinctions. One would not want to integrate a discrete function nor try to sum a continuous one!

While we will be consistent about this, we will also refer to random variables, either discrete or continuous as *following* or *having* a certain probability *distribution* function. So we will refer to a random variable as following a binomial distribution or another random variable as following a Cauchy distribution although one is discrete and the other is continuous.

EXERCISES 3.1

1. A loaded wheel has probability density function

$$f(x) = 3x^2, \quad 0 \leq x \leq 1.$$

(a) Show that $f(x)$ is a probability density function.

(b) Find $P\left(\frac{1}{2} \leq X \leq \frac{3}{4}\right)$.

(c) Find $P\left(X \geq \frac{2}{3}\right)$.

(d) Find c so that $P(X \geq c) = \frac{2}{3}$.

2. A random variable X has probability density function

$$f(x) = \begin{cases} k(x^2 - x^3), & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find k .

(b) Find $P\left(X \geq \frac{3}{4}\right)$.

(c) Calculate $P\left(X \geq \frac{3}{4} \mid X \geq \frac{1}{2}\right)$.

3. If

$$f(x) = \begin{cases} k \sin x, & 0 \leq x \leq \pi, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that $k = 1/2$.

(b) Calculate $P\left(X \leq \frac{\pi}{3}\right)$.

(c) Find b so that $P(X \leq b) = \frac{1}{3}$.

4. A random variable X has probability density function

$$f(x) = \begin{cases} 4x^3, & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the mean, μ , and the variance, σ^2 , for X .
 (b) Calculate exactly $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$ and compare your answer with the result given by Tchebycheff's inequality.
 5. The length of life, X , in days, of a heavily used electric motor has probability density function

$$f(x) = \begin{cases} 3e^{-3x}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the probability the motor lasts at least 1/2 of a day, given that it has lasted 1/4 of a day.
 (b) Find the mean and variance for X .

6. A random variable X has probability density function

$$f(x) = \begin{cases} kx^2 e^{-x}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find k .
 (b) Graph $f(x)$
 (c) Find μ and σ^2 .

7. The distribution function for a random variable, X , is

$$F(x) = \begin{cases} 0, & x < -4 \\ \frac{1}{8}, & -4 \leq x < -3 \\ \frac{3}{8}, & -3 \leq x < 2 \\ \frac{3}{4}, & 2 \leq x < 5 \\ 1, & x \geq 5 \end{cases}$$

- (a) Find $P(X = 2)$.
 (b) Find $P(-3 \leq x < 2)$.
 8. A continuous random variable, X , has probability density function

$$f(x) = \begin{cases} k(x - x^2), & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find k .
 (b) Find the mean and variance of X .

- (c) Four independent observations of X are made. What is the probability that exactly two of these are greater than $3/4$?

9. Let $f(x) = \begin{cases} \frac{1}{2} + \frac{x}{4}, & -2 < x < 0 \\ \frac{1}{2} - \frac{x}{4}, & 0 < x < 2 \\ 0, & \text{otherwise.} \end{cases}$

- (a) Show that $f(x)$ is a probability density function.
 (b) Find $P[|X| < 1]$.
 (c) Find μ and σ^2 .

10. A random variable, X , has probability density function

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

- (a) Find $E[X]$.
 (b) Find $\text{Var}[X]$.
 (c) Find $F(x)$, being sure to specify this for all values of x .
 (d) What is the probability that at least two of three independent observations on X are greater than $1/2$?

11. The length of time, Y , in hours, a student takes to complete an examination is a random variable with

$$g(y) = \begin{cases} cy^2 + y, & 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find c .
 (b) Find the cumulative distribution function, $G(y)$.
 (c) Find an expression for $P(Y > y)$ for any value of y .

12. As a measure of intelligence, mice are timed when going through a maze to reach a reward of food. The time (in seconds) required for any mouse is a random variable Y with probability density function

$$f(y) = \begin{cases} \frac{10}{y^2} & y \geq 10 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Show that $f(y)$ has the properties of a probability density function.
 (b) Find $P(9 \leq Y \leq 99)$.
 (c) Find the probability a mouse requires at least 15 seconds to traverse the maze if it is known that the mouse requires at least 12 seconds.

13. A continuous random variable has probability density function

$$f(x) = \begin{cases} cx^2, & -3 \leq x \leq 3 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the mean and variance of X .
 (b) Verify Tchebycheff's inequality for the case $k = \sqrt{\frac{5}{3}}$.
14. Suppose the distance X between a point target and a shot aimed at the point in a video game is a continuous random variable with probability density function

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2), & -1 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Find the mean and variance of X .
 (b) Use Tchebycheff's inequality to give a bound for $P\left[|X| < \frac{1}{2}\right]$.
15. If the loaded wheel with $f(x) = 2x$, $0 \leq x \leq 1$, is spun three times, it can be shown that the probability density function for Y , the smallest of the three values obtained, is $g(y) = 6y(1-y^2)^2$, $0 \leq y \leq 1$. Find the mean and variance for Y .
16. Show that $g(y) = \begin{cases} 1260 y^4(1-y)^5, & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$ is a probability density function.
 Then find the mean and variance for Y .
17. Use your computer algebra system to draw a random sample of 100 observations from the distribution

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

The random variable X here is said to follow a *uniform distribution* on the interval $[0, 1]$.

- (a) Enumerate the observations in each of the categories $0 \leq x < 0.1$, $0.1 \leq x < 0.2$, and so on. Do the observations appear to be uniform?
 (b) We will show in Chapter 4 that if X is uniform on the interval $[0, 1]$ and if $Y = X^2$, then the probability density function for Y is $g(y) = 2y$, $0 \leq y \leq 1$. So the sample in part (a) can be used to simulate a random sample from the loaded wheel discussed in Section 3.1. Show a sample from the loaded wheel. Graph the sample values and decide whether or not the sample appears to have been selected from the loaded wheel.
18. Show that $g(y) = \begin{cases} \binom{n}{r} \cdot r \cdot y^{r-1} \cdot (1-y)^{n-r}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$ is a probability density function for n a positive integer and $r = 1, 2, \dots, n$.

19. Given

$$f(x) = \begin{cases} \frac{x^2}{2}, & 0 \leq x \leq 1 \\ \frac{3}{4} - \left(x - \frac{3}{2}\right)^2, & 1 \leq x \leq 2 \\ \frac{(x-3)^2}{2}, & 2 \leq x \leq 3 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Sketch $f(x)$ and show that it is a probability density function.
 (b) Find the mean and variance of X .

- 20.** Suppose that X is a random variable with probability distribution function $F(x)$ whose domain is $x \geq 0$.
- Show that $\int_0^\infty [1 - F(x)] dx = E(X)$. [Hint: Write the integral as a double integral and then change the order of integration.]
 - Write an integral involving $F(x)$ whose value is $E(X^2)$.
- 21.** Prove formulas 3.1 and 3.2.

3.2 UNIFORM DISTRIBUTION

The fair wheel, where $f(x) = 1$, $0 \leq x \leq 1$, is an example of a *uniform* probability density function. In general, if

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

then X is said to have a *uniform* probability distribution. This is the continuous analogy of the discrete uniform distribution considered in Chapter 2. A graph is shown in Figure 3.5.

The mean and variance are calculated as follows:

$$\begin{aligned} E(X) &= \int_a^b \frac{x}{b-a} dx = \frac{b+a}{2} \quad \text{and} \\ \text{Var}(X) &= E(X^2) - (E(X))^2 \\ \text{Var}(X) &= \int_a^b \frac{x^2}{b-a} dx - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 \\ \text{Var}(X) &= \frac{(b-a)^2}{12}. \end{aligned}$$

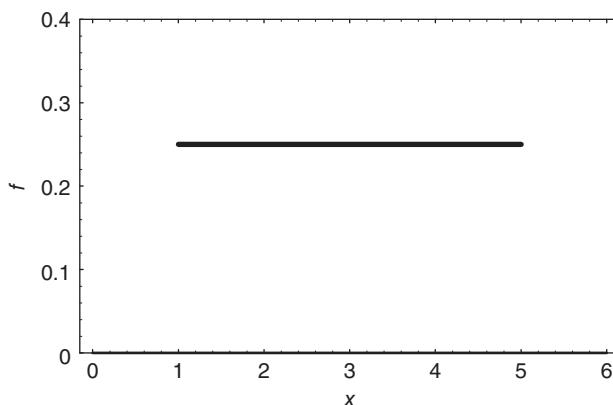


Figure 3.5 Uniform distribution on the interval $[a, b]$.

Example 3.2.1

Suppose X is uniform on the interval $[1,5]$. Then

$$f(x) = \frac{1}{4}, \quad 1 \leq x \leq 5.$$

Suppose also that we have an observation that is at least 2. What is the probability the observation is at least 3?

We need

$$P(X \geq 3 | X \geq 2) = \frac{P(X \geq 3)}{P(X \geq 2)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}.$$

Example 3.2.2

The wheel in Example 3.2.1 is spun again, the result being X . Now we spin the wheel until an observation greater than the value of X is found, say Y . What is the expected value for Y ?

Since the wheel is a fair one, we suppose that Y is uniformly distributed on $(x, 5)$ so that

$$g(y) = \frac{1}{5-x}, \quad 1 < x < 5.$$

Then

$$\begin{aligned} E(Y) &= \int_x^5 \frac{y}{5-x} dy \\ &= \frac{5^2 - x^2}{2(5-x)} \\ &= \frac{5+x}{2}, \end{aligned}$$

a natural result since the central value on the interval $(x, 5)$ is $\frac{5+x}{2}$.

EXERCISES 3.2

1. The arrival times of customers at an automobile repair shop is uniformly distributed over the interval from 8 a.m. to 9 a.m. If a customer has not arrived by 8:30 a.m., what is the probability he will arrive after 8:45 a.m.?
2. A traffic light is red for 60 seconds, yellow for 10 seconds, and green for 90 seconds. Assuming that arrival times at the light are uniformly distributed, what is the probability a car stops at the light for at most 30 seconds?
3. A crude Geiger counter records the number of radioactive particles a substance emits, but often errs in the number of particles recorded. If the error is uniformly distributed on the interval, what is the probability the counter will underrecord the number of particles emitted?

4. Suppose that X is a random variable uniformly distributed on the interval $(-2, 2)$.
- Find $P\left(\frac{1}{X} < 2\right)$.
 - Find $P\left(\frac{1}{X^2} < 2\right)$.
5. Let X be uniformly distributed over the intervals $(0,1)$ and $(2, 3)$.
If $P(0 \leq X \leq 1) = 2 \cdot P(2 \leq X \leq 3)$, find $P(\mu - \sigma \leq X \leq \mu + \sigma)$.
6. The termination of a chemical reaction occurs at a random time T between 6 and 7.5 hours after the start of the experiment. The time follows a uniform distribution.
- What is the probability the reaction lasts at least 6.5 hours and no more than 6.75 hours?
 - If the reaction is run four independent times, what is the probability that in exactly one of the four replications of the experiment the reaction will last no more than 6.5 hours?
7. Suppose that X is uniformly distributed on $0 < x < 12$. Use Tchebycheff's inequality to establish a bound on $P\left(\mu - \frac{3\sqrt{3}}{5} \cdot \sigma < X < \mu + \frac{3\sqrt{3}}{5} \cdot \sigma\right)$ and then verify that the bound is correct.
8. Let X be a uniform random variable on the interval $1 < x < b$. Determine b so that
- $$\sigma_x^2 = 3\mu_x.$$
9. A random variable X is uniformly distributed on the interval $-1 < x < 1$. Find $P\left(\frac{1}{4} \leq X^2 \leq \frac{3}{4}\right)$.
10. Find the probability that at least two of four random observations of a uniform random variable on the interval $[0,10]$ are greater than 7.
11. Suppose X is a uniform random variable on the interval $[a, a+2]$. Find a if $P[e^X < 1.765] = \frac{1}{2}$.

3.3 EXPONENTIAL DISTRIBUTION

Example 3.3.1

Customers in a checkout line at a supermarket find that the times the checkers take in the checkout process follows the probability density function

$$f(x) = e^{-x}, \quad x \geq 0$$

where X is measured in minutes. We see that $f(x) \geq 0$ and that $\int_0^\infty f(x) dx = 1$, so $f(x)$ defines a probability density function. Here, $f(x)$ is an example of an *exponential probability density function*. What is the probability a customer's checkout time is at least k minutes?

This is

$$P(X \geq k) = \int_k^\infty e^{-x} dx = e^{-k}. \quad (3.3)$$

Another calculation yields a somewhat surprising result. Suppose the checkout time has been at least s minutes. What is the probability it will be at least $s+t$ minutes? Using the

formula for conditional probability and formula (3.3) earlier, we have

$$\begin{aligned} P(X \geq s+t | X \geq s) &= \frac{P(X \geq s+t)}{P(X \geq s)} \\ &= \frac{e^{-(s+t)}}{e^{-s}} = e^{-t} \text{ so that} \\ P(X \geq s+t | X \geq s) &= P(X \geq t). \end{aligned}$$

Consequently, the probability the customer waits t minutes more, given that the waiting time has been at least s minutes, is the same as the probability the waiting time is *initially* t minutes! The fact that the customer has been waiting s minutes appears not to affect the future waiting time at all. We call this property of the exponential probability distribution the *memoryless* property. (It can be shown that the exponential probability density function is the only probability density function for which the memoryless property holds. Among discrete distributions, the geometric probability distribution is the only probability distribution function for which the property holds.)

A more general form of the exponential probability density function is

$$f(x) = \lambda e^{-\lambda(x-a)}, \quad x \geq a, \quad \lambda > 0.$$

Our checkout time example is a special case where $\lambda = 1$ and $a = 0$. The graph is shown in Figure 3.6 where a has been taken to be 2.

We note that and that $\int_a^\infty f(x) dx = \int_a^\infty \lambda e^{-\lambda(x-a)} dx = 1$, so $f(x)$ satisfies the properties of a probability density function.

Mean and Variance

For $f(x) = \lambda e^{-\lambda(x-a)}$, $x \geq a$, $\lambda > 0$, direct calculation shows that

$$E(X) = \int_a^\infty x \cdot f(x) dx = a + \frac{1}{\lambda}, \quad \text{and that}$$

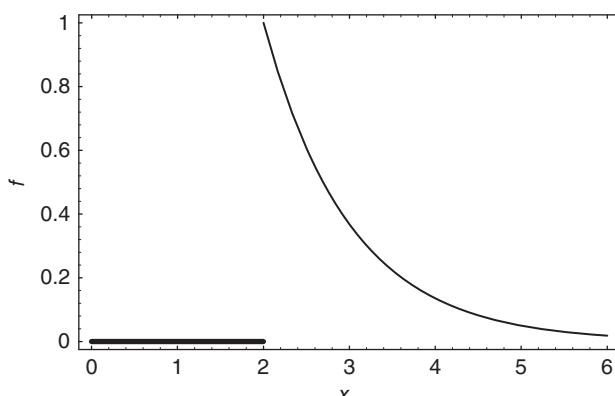


Figure 3.6 An exponential distribution.

$$E(X^2) = \int_a^\infty x^2 \cdot f(x) dx = a^2 + \frac{2a}{\lambda} + \frac{2}{\lambda^2}, \text{ so that}$$

$$\text{Var}(X) = \left(a^2 + \frac{2a}{\lambda} + \frac{2}{\lambda^2} \right) - \left(a + \frac{1}{\lambda} \right)^2$$

$$\text{where } \text{Var}(X) = \frac{1}{\lambda^2}.$$

Distribution Function

For the exponential density $f(x) = \lambda e^{-\lambda(x-a)}$, $x \geq a$, $\lambda > 0$.

$$F(x) = \int_a^x f(x) dx = 1 - e^{-\lambda(x-a)}.$$

Example 3.3.2

Refer again to the checkout line and the waiting time density

$$f(x) = e^{-x}, \quad x \geq 0.$$

Assume that customers' waiting times are independent. What is the probability that, of the next 5 customers, at least 3 will have waiting times in excess of 2 minutes?

There are two random variables here. Let X be the waiting time for an individual customer, and Y be the number of customers who wait at least 2 minutes. Here X is exponential; since the waiting times are independent and $P(X \geq 2)$ is the same for every customer, Y is binomial. Note that while X is continuous, Y is a discrete random variable.

It is easiest to start with X , where $P(X \geq 2)$ determines p in the binomial distribution.

$$P(X \geq 2) = \int_2^\infty e^{-x} dx = e^{-2}.$$

$$\text{Then } P(Y \geq 3) = \sum_{y=3}^5 \binom{5}{y} \cdot (e^{-2})^y \cdot (1 - e^{-2})^{5-y}.$$

The value of this expression is 0.020028, so the event is not very likely.

Example 3.3.3

A radioactive source is emitting particles according to a Poisson distribution with 14 particles expected to be emitted per minute. The source is observed until the first particle is emitted. What is the probability density function for this random variable?

Again we have two variables in the problem. If X denotes the number of particles emitted in 1 minute, then X is Poisson with parameter 14. However, we do not know the time

interval until the first particle is emitted. This is also a random variable, which we call Y . Note that Y is a continuous random variable. If y minutes pass before the first particle is emitted, then there must be no emissions in the first y minutes. Since the number of emissions in y minutes is Poisson with parameter $14y$, it follows that

$$P(Y \geq y) = e^{-14y}.$$

We conclude that

$$F(y) = P(Y \leq y) = 1 - e^{-14y}$$

and so

$$f(y) = \frac{dF(y)}{dy} = 14 \cdot e^{-14y}, \quad y \geq 0.$$

This is an exponential density. In this example, note that X is discrete while Y is continuous.

Example 3.3.4

As a final example of the exponential density $f(x) = \lambda e^{-\lambda(x-a)}$, $x \geq a$, we check the memoryless property for the more general form of the exponential density.

$$\begin{aligned} P(X \geq s+t | X \geq s) &= \frac{P(X \geq s+t)}{P(X \geq s)} \\ &= \frac{e^{-\lambda(s+t-a)}}{e^{-\lambda(s-a)}} = e^{-\lambda t} \end{aligned}$$

so that $P(X \geq s+t | X \geq s) = P(X \geq a+t)$.

So the memoryless property depends on the value for a .

3.4 RELIABILITY

The *reliability* of a system or of a component in a system refers to the lack of frequency with which failures of the system or component occur. Reliable systems or components fail less frequently than less reliable systems or components. Suppose, for example, that T , the time to failure of a light bulb, has an exponential distribution with expected value 10,000 hours. This gives the probability density function as

$$f(t) = \left(\frac{1}{10000} \right) e^{-\frac{t}{10000}}, \quad t \geq 0.$$

The reliability, $R(t)$, is defined as

$$R(t) = P(T > t)$$

that is, $R(t)$ gives the probability that the bulb lasts more than t hours. We assume that $R(0) = 1$ and we see that

$$R(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

and that

$$-R'(t) = f(t).$$

Since in this case

$$F(t) = \int_0^t \left(\frac{1}{10000} \right) e^{-\frac{t}{10000}} dt = 1 - e^{-\frac{t}{10000}}$$

$$\text{It follows that } R(t) = P(T > t) = e^{-\frac{t}{10000}}.$$

What is the probability that such a bulb lasts at least 2500 hours? This is $R(2500) = e^{-\frac{1}{4}} = 0.7788$. So although the mean time to failure is 10000 hours, only about 78% of these bulbs last longer than 1/4 of the mean lifetime.

Were it crucial that a bulb last 2500 hours, say that this happens with probability 0.95, what should be mean time to failure be? Let this mean time to failure be m . Then

$$e^{-\frac{2500}{m}} = 0.95$$

$$\text{so } m = 48,740 \text{ hours.}$$

Hazard Rate

The *hazard rate* of an item refers to the probability, per unit of time, that an item that has lasted t units of time will last Δt more units of time. We will denote the hazard rate by $H(t)$, so

$$\begin{aligned} H(t) &= \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} \\ &= \frac{F(t + \Delta t) - F(t)}{\Delta t \cdot P(T > t)}. \end{aligned}$$

As $\Delta t \rightarrow 0$, $H(t)$ approaches

$$H(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)} = -\frac{R'(t)}{R(t)}.$$

In actuarial work, the hazard rate is called the *force of mortality*. The hazard rate also occurs in econometrics as well as in other fields. In this section, we investigate the consequences of a constant hazard rate, λ . Consequences of a nonconstant hazard rate will be considered later in this chapter.

Suppose then that

$$H(t) = \frac{f(t)}{1 - F(t)} = \lambda, \text{ where } \lambda \text{ is a constant.}$$

Since $\frac{f(t)}{1 - F(t)} = -\frac{R'(t)}{R(t)}$ we have that

$$-\frac{R'(t)}{R(t)} = \lambda.$$

It follows that

$$-\ln[R(t)] = \lambda t + k \text{ so that}$$

$$R(t) = c \cdot e^{-\lambda t}.$$

Now and if we suppose that our components begin life at $T = 0$, then $R(0) = 1$ and

$$R(t) = e^{-\lambda t}.$$

Since $f(t) = -R'(t)$, it follows that $f(t) = \lambda e^{-\lambda t}$, $t \geq 0$.

A constant hazard rate then produces an exponential failure law. It is easy to show that an exponential failure law produces a constant hazard rate.

From Example 3.3.3, we conclude that failures occurring according to a Poisson process will also produce an exponential time to failure and hence a constant hazard rate.

Typically, the hazard rate is not constant for components. There is generally a “burn-in” period where the hazard rate may be declining. The hazard rate then usually becomes constant, or nearly so, after which it increases. This produces the “bathtub” function, as shown in Figure 3.7.

Different hazard rates, although constant, can have surprisingly different consequences. Suppose, for example, that component I has constant hazard rate λ while component II has hazard rate $k \cdot \lambda$ where $k > 0$. Then the corresponding reliability functions are

$$R_I(t) = e^{-\lambda t} \text{ while}$$

$$R_{II}(t) = e^{-k\lambda t} = (e^{-\lambda t})^k = [R_I(t)]^k.$$

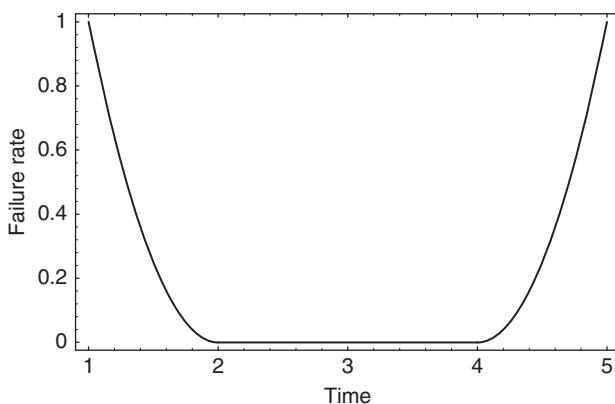


Figure 3.7 A “bathtub” hazard rate function.

So the probability component II that lasts t units or more is the k th power of the probability component I that lasts the same time. Since positive powers of probabilities become smaller as the power k increases, component II may rapidly become useless.

EXERCISES 3.4

1. An exponential distribution has $f(x) = 4e^{-4(x-2)}$ for $x \geq 2$. Find $E[X]$ and $\text{Var}[X]$.
2. In Exercise 1, suppose this is a waiting time density. Find the probability of the next 6 values, at most 3 are ≤ 2 .
3. Let X be an exponential random variable with mean 6.
 - (a) Find $P(X \geq 4)$.
 - (b) Find $P(X \geq 4 | X \geq 2)$.
4. The *median* of a probability distribution is the value that is exceeded 1/2 of the time.
 - (a) Find the median of an exponential distribution with mean λ .
 - (b) Find the probability an observation exceeds λ .
5. Snowfall in Indiana follows an exponential distribution with mean $15''$ per winter season.
 - (a) Find the probability the snowfall will exceed $17''$ next winter.
 - (b) Find the probability that in 4 out of the next 5 winters the snowfall will be less than the mean.
6. The length, X , of an international telephone call from a local business follows an exponential distribution with mean 2 minutes. In dollars, the cost of a call of X minutes is $3X^2 - 6X + 2$. Find the expected cost of a telephone call.
7. The lengths of life of batteries in transistor radios follow exponential probability distributions. Radio A takes 2 batteries, each of which has an expected life of 200 hours; radio B uses 4 batteries, but the expected life of each is 400 hours. Radio A works if at least one of its batteries operates; radio B works only if at least three of its batteries operate. An expedition needs a radio that will function at least 500 hours. Which radio should be taken, or does it matter?
8. Accidents at a busy intersection follow a Poisson distribution with three accidents expected in a week.
 - (a) What is the probability that at least 10 days pass between accidents?
 - (b) It has been 9 days since the last accident. What is the probability that it will be 5 days or more until the next accident?
9. The diameter of a manufactured part, X , is a random variable whose probability density function is

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

If $X < 1$, the manufacturer realizes a profit of \$3. If $X > 1$, the part must be discarded at a net loss of \$1. The machinery manufacturing the part may be set so that $\lambda = \frac{1}{4}$ or $\lambda = \frac{1}{2}$. Which setting will maximize the manufacturer's expected profit?

10. If X is a random selection from a uniform variable on the interval $(0, 1)$, then the transformation $Y = -\lambda \ln(1 - X)$ is known to produce random selections from an exponential density with mean λ .

- (a) Use a uniform random number generator to draw a sample of 200 observations from an exponential density with mean 7.
- (b) Draw a histogram of your sample and compare it graphically with the expected exponential density.
11. The hazard rate of an essential component in a rocket engine is 0.05. Find its reliability at time 125.
12. An exponential process has $R(200) = 0.85$. When is $R = 0.95$?
13. A Poisson process has mean μ . Show that the waiting time for the second occurrence is not exponentially distributed.
14. Find the probability an item fails before 200 units of time if its hazard rate is 0.008.
15. Suppose that the life length of an automobile is exponential with mean 72,000 miles. What is the expected length of life of automobiles that have lasted 50,000 miles?
16. An electronic device costs $\$K$ to produce. Its length of life, X , has probability density function

$$f(x) = 0.01e^{-0.01x}, \quad x \geq 0.$$

If the device lasts less than 3 units of time, the item is scrapped and has no value. If the life length is between 3 and 6, the item is sold for $\$S$; if the life length is greater than 6, the item is sold for $\$V$. Let Y be the net profit per item. Find the probability density for Y .

17. Suppose X is a random variable with probability density function

$$f(x) = 3e^{-3(x-a)}, \quad x \geq 2.$$

- (a) Show that $a = 2$.
- (b) Find the cumulative distribution function, $F(x)$.
- (c) Find $P(X > 5|X > 3)$.
- (d) If 8 independent observations are made, what is the probability that exactly 6 of them are less than 4?
18. A lamp contains 3 bulbs, each of which has life length that is exponentially distributed with mean 1000 hours. If the bulbs fail independently, what is the probability that some light emanates from the lamp for at least 1200 hours?
19. According to a kinetic theory, the distance, X , that a molecule travels before colliding with another molecule is described by the probability density function

$$f(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}, \quad x > 0, \quad \lambda > 0.$$

- (a) What is the average distance between collisions?
- (b) Find $P(X > 6|X > 4)$.

3.5 NORMAL DISTRIBUTION

We come now to the most important continuous probability density function and perhaps the most important probability distribution of any sort, the normal distribution. On several

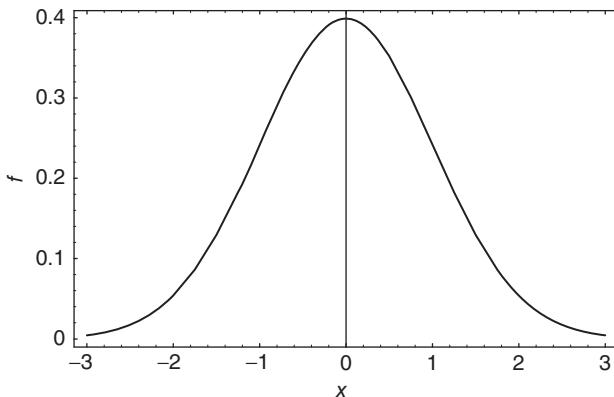


Figure 3.8 Standard normal probability density function.

occasions, we have observed its occurrence in graphs from, apparently, widely differing sources: the sums when three or more dice are thrown; the binomial distribution for large values of n ; and in the hypergeometric distribution. There are many other examples as well and several reasons, which will appear here, to call this distribution “normal.”

If

$$f(x) = \frac{1}{b \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2 \cdot b^2}(x-a)^2}, \quad -\infty < x < \infty, \quad -\infty < a < \infty, \quad b > 0, \quad (3.4)$$

we say that X has a *normal* probability distribution. A graph of a normal distribution, where we have chosen $a = 0$ and $b = 1$, appears in Figure 3.8.

The shape of a normal curve is highly dependent on the standard deviation. Figure 3.9 shows some normal curves, each with mean 0, but with different standard deviations. We will show presently that a is the mean value and b is the standard deviation of the normal curve.

We now establish some facts regarding $f(x)$ as defined earlier.

1. $f(x)$ defines a probability density function.

Proof $f(x) \geq 0$ and so we must show that $\int_{-\infty}^{\infty} f(x) dx = 1$. To do this, let $Z = \frac{X-a}{b}$ in (3.4).

We have

$$\int_{-\infty}^{\infty} \frac{1}{b \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2 \cdot b^2}(x-a)^2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2}z^2} dz.$$

Consider the curve $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty < x < \infty$, as shown in Figure 3.10.

Let $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. If the curve is revolved around the y -axis, the surface generated is

$$f(x, z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2+x^2)}, \quad -\infty < x < \infty, \quad -\infty < z < \infty$$

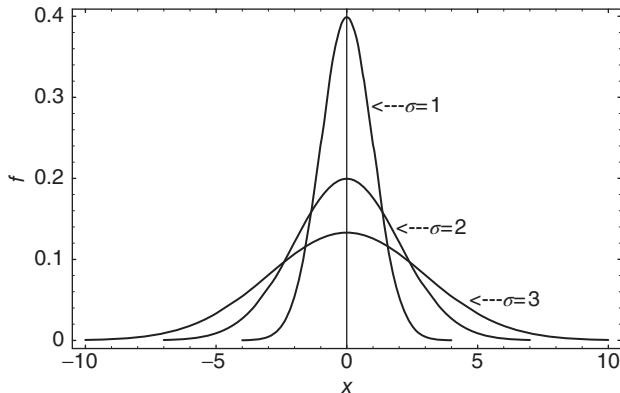


Figure 3.9 Some normal probability density functions.

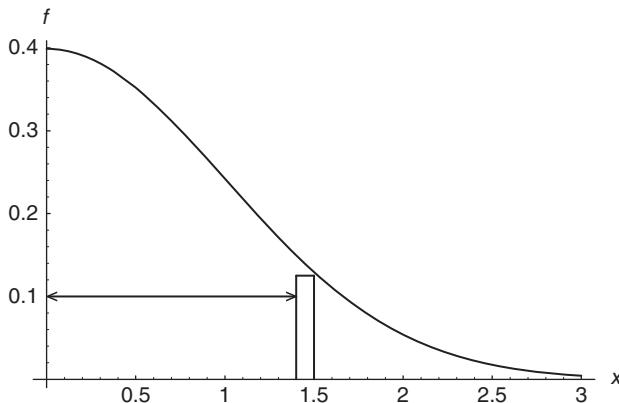


Figure 3.10 Revolving the standard normal curve around the y-axis.

since this surface has circular cross sections and the proper traces in the coordinate planes. The volume generated is then

$$\begin{aligned} V &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2+x^2)} dz dx \\ &= \frac{1}{\sqrt{2\pi}} (\sqrt{2\pi} I)^2 = \sqrt{2\pi} I^2. \end{aligned}$$

On the other hand, V can be found using cylindrical shells as

$$V = \frac{2\pi}{\sqrt{2\pi}} \int_0^{\infty} z \cdot e^{-\frac{1}{2}z^2} dz = \sqrt{2\pi}.$$

So $I^2 = 1$ and since $I > 0, I = 1$.

So $f(x)$ is a probability density function for $-\infty < a < \infty$ and $b > 0$.

2. We now find the mean and variance for X .

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{b \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2 \cdot b^2}(x-a)^2} dx$$

Let $z = \frac{x-a}{b}$ in the integral so that $x = a + bz$, and then

$$E(X) = a \int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2}z^2} dz + \frac{b}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} z \cdot e^{-\frac{1}{2}z^2} dz,$$

so, since the first integral is 1 and the second integral is 0, $E(X) = a$.

To find the variance, first calculate

$$E(X^2) = \int_{-\infty}^{\infty} \frac{x^2}{b \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2 \cdot b^2}(x-a)^2} dx.$$

Again let $z = \frac{x-a}{b}$. Then

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} \frac{(bz+a)^2}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2}z^2} dz \\ &= b^2 \int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2}z^2} dz + 2ab \int_{-\infty}^{\infty} \frac{z}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2}z^2} dz + a^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2}z^2} dz \end{aligned}$$

Since $\frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} z \cdot e^{-\frac{1}{2}z^2} dz = 0$, it follows that $E(X^2) = \frac{b^2}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz + a^2$, which simplifies to

$$E(X^2) = b^2 + a^2.$$

We conclude that

$$\begin{aligned} \mu &= a \text{ and} \\ \sigma^2 &= b^2. \end{aligned}$$

The probability density function for the normal curve is then usually written as

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

We will abbreviate this as

$$X \sim N(\mu, \sigma),$$

where the symbol \sim is read “is distributed as.”

This is our first example of a probability density whose formula involves the standard deviation. The implications of this will be encountered in examples and problems.

3. Finally, we show that if $X \sim N(\mu, \sigma)$ and if $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0, 1)$. We call the $N(0, 1)$ curve the *standard or unit normal curve*. The statement above indicates that the transformation $Z = \frac{X-\mu}{\sigma}$ can be used on an arbitrary normal curve to produce a standard normal curve.

To show this consider the cumulative distribution function for Z , $G(z)$, assuming that $F(x)$ is the cumulative distribution function for X . Then by definition,

$$\begin{aligned} G(z) &= P(Z \leq z) \\ &= P\left(\frac{X-\mu}{\sigma} \leq z\right) \\ &= P(X \leq \mu + \sigma \cdot z) = F(\mu + \sigma \cdot z). \end{aligned}$$

We now use the fact that

$$g(z) = \frac{dG(z)}{dz}$$

to see that

$$g(z) = \frac{dG(z)}{dz} = \frac{dF(\mu + \sigma \cdot z)}{dz} = \frac{dF(\mu + \sigma \cdot z)}{d(\mu + \sigma \cdot z)} \cdot \frac{d(\mu + \sigma \cdot z)}{dz}$$

by the chain rule. It follows that

$$\begin{aligned} g(z) &= \sigma \cdot f(\mu + \sigma z) \\ g(z) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty \end{aligned}$$

which is the standard normal distribution. This indicates that problems involving arbitrary values of μ and σ can be solved using a single, standard, and normal curve.

We will return to the process by which $g(z)$ was established in Chapter 4 and apply the same technique to other problems involving functions of random variables. For now, we consider some examples of the normal distribution and problems using it.

Example 3.5.1

Mathematics aptitude scores, X , on the Scholastic Aptitude Test (SAT) are $N(500, 100)$.

Find (a) the probability an individual's score exceeds 600 and

(b) the probability an individual's score exceeds 600, given that it exceeds 500.

(a) Many computer algebra systems will calculate $P(X > 600)$ directly. This will be found to be 0.158655. If a computer algebra system is not available, a table of the standard normal distribution may be used as follows:

The Z transformation here is $Z = \frac{X-500}{100}$, so $P(X > 600) = P(Z > 1) = 0.158655$ using Table 1 in the Appendix.

(b) Here, we need $P(X > 600 | X > 500) = \frac{P(X > 600)}{P(X > 500)} = \frac{0.158655}{0.500000} = 0.317310$.

Example 3.5.2

What Mathematics SAT score, or greater, can we expect to occur with probability 0.90?

Here, we know that $X \sim N(500, 100)$ and we want to find x so that

$$P(X \geq x) = 0.90. \text{ So, if } Z = \frac{X - 500}{100}, \text{ then}$$

$$P\left(Z \geq \frac{x - 500}{100}\right) = 0.90, \text{ but}$$

$$P(Z \geq -1.287266) = 0.90 \text{ so}$$

$$\frac{x - 500}{100} = -1.287266 \text{ giving}$$

$$x = 371.$$

Example 3.5.3

From Tchebycheff's inequality, we conclude that the standard deviation is in fact a measure of dispersion for a distribution, since the probability the interval from $\mu - k \cdot \sigma$ to $\mu + k \cdot \sigma$ is at least $1 - \frac{1}{k^2}$, a probability that increases as k increases. When the distribution is known, this probability can be determined exactly by integration. We do this now for a standard normal density. Again let $Z = \frac{X - \mu}{\sigma}$.

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) = 0.6826894921$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) = 0.9544997361$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) = 0.9973002039$$

Tchebycheff's inequality indicates that these probabilities are at least 0, 3/4, and 8/9, respectively.

The earlier results, sometimes called the "2/3, 95%, 99% Rule" can be very useful in estimating probabilities using the normal curve.

For example, to refer again to the Mathematics SAT scores that are $N(500, 100)$, an estimate for the probability a student's score is between 400 and 650 may be found by estimating the probability the corresponding z -score is between -1 and 1.50 . We know that $2/3$ of the area under the curve is between -1 and 1 , and we need to estimate the probability from 1 to 1.50 . This can be estimated at $1/2$ of the difference between 0.95 and $2/3$, giving a total estimate of $2/3 + (1/2)(0.95 - 2/3) = 0.81$. The exact probability is 0.775 . It is a rarity that the answers to probability problems can be estimated in advance of an exact solution.

The occurrence of the normal distribution throughout probability theory is striking. In the next section, we explain why the graphs of binomial distributions, considered in Chapter 2, become normal in appearance.

EXERCISES 3.5

1. Mathematics SAT scores are $N(500, 100)$.
 - (a) Find the probability an individual's score is between 350 and 650.
 - (b) Find the probability that one's score is less than 350, given that it is less than 400.

2. In exercise 1, find an SAT score a , such that
 - (a) $P(\text{Score} \leq a) = 0.95$.
 - (b) $P(a \leq \text{Score} \leq 650) = 0.30$.
3. IQ scores are known to be normally distributed with mean 100 and standard deviation 10.
 - (a) Find the probability an IQ score exceeds 128.
 - (b) Find the probability an IQ score is between 90 and 110.
4. The size of a boring in a metal block is normally distributed with mean 3 cm and standard deviation 0.01 cm.
 - (a) What proportion of the borings have sizes between 2.97 cm and 3.01 cm?
 - (b) For the borings exceeding 3.005 cm, what proportion exceeds 3.010 cm?
5. Brads, which are labeled $3/4''$ are actually normally distributed. Manufacturer I produces brads with mean $3/4''$ and standard deviation $0.002''$; manufacturer II produces brads with mean $0.749''$ and standard deviation $0.0018''$; brads from manufacturer III have mean $0.751''$ and standard deviation $0.0015''$. A builder requires brads in the range $3/4 \pm 0.005''$. From which manufacturer should the brads be purchased?
6. A soft drink machine dispenses cups of a soft drink whose volume is actually a normal random variable with mean 12 oz and standard deviation 0.1 oz.
 - (a) Find the probability a cup of the soft drink contains more than 12.2 oz.
 - (b) Find a volume, v , such that 99% of the time the cups contain at least v oz.
7. Resistors used in an electric circuit have resistances that are normally distributed with mean 0.21 ohms and standard deviation 0.045 ohms. A resistor is acceptable in the circuit if its resistance is at most 0.232 ohms. What percentage of the resistors are acceptable?
8. On May 5, in Colorado Springs, temperatures have been found to be normally distributed with mean 80° and standard deviation 8° . The record temperature on that day is 90° .
 - (a) What is the probability the record of 90° will be broken on next May 5?
 - (b) What is the probability the record of 90° will be broken at least three times during the next 5 years on May 5?
9. Sales in a fast food restaurant are normally distributed with mean \$42,000 and standard deviation \$2000 during a given sales period. During a recent sales period, sales were reported to a local taxing authority to be \$37,600. Should the taxing authority be suspicious?
10. Suppose that $X \sim N(\mu, \sigma)$. Find a in terms of μ and σ if
 - (a) $P(X > a) = 0.90$.
 - (b) $P(X > a) = \frac{1}{3}P(X \leq a)$.
11. The size of a manufactured part is a normal random variable with mean 100 and variance 25.

If the size is between 95 and 110, the parts can be sold at a profit of \$50 each. If the size exceeds 110, the part must be reworked and a net profit of \$10 is made per part. A part whose size is less than 95 must be scrapped at a loss of \$20. What is the expected profit for this process?

12. Rivets are useful in a device if their diameters are between $0.25''$ and $0.38''$. These limits are often called *upper and lower specification limits*. A manufacturer produces rivets that are normally distributed with mean $0.30''$ and standard deviation $0.03''$.
- What proportion of the rivets meet specifications?
 - Suppose the mean of the manufacturing process could be changed, but the manufacturing process is such that the standard deviation cannot be altered. What should the mean of the manufactured rivets be so as to maximize the proportion that meet specifications?
13. Refer to problem 10. Suppose that $X \sim N(\mu, \sigma)$ and that upper and lower specification limits are U and L , respectively. Show that if σ must be held fixed, then the value of μ that maximizes $P(U \leq X \leq L)$ is $\frac{U+L}{2}$.
14. Manufacturing processes that produce normally distributed output are often compared by calculating their *process capability indices*. The process capability index for a process with upper and lower specification limits U and L , respectively, is

$$C_p = \frac{U - L}{6\sigma}$$

where the variable X is distributed $N(\mu, \sigma)$.

What can be said about the process under each of the following conditions?

- $C_p = 1$.
 - $C_p < 1$.
 - $C_p > 1$.
15. Upper and lower *warning limits* are often established for measurements on manufactured products. Usually, if $X \sim N(\mu, \sigma)$, these are set at $\mu \pm 1.96\sigma$ so that 5% of the product is outside the warning limits. Discuss the proportion of the product outside the warning limits if the mean of the process increases by one standard deviation.
16. Suppose that $X \sim N(\mu, \sigma)$. Find μ and σ if $P(X > 2) = \frac{2}{3}$ and $P(X > 3) = \frac{1}{3}$.
17. “40 lb” bags of cement have weights that are actually $N(39.1, \sqrt{9.4})$.
- Find the probability that two of five randomly selected bags weigh less than 40 lbs.
 - How many bags must be purchased so that the probability that at least 1/2 of the bags weigh at most 40 lb is at least 0.95?
18. Suppose $X \sim N(0, 1)$. Find
- $P(|X| < 1.5)$.
 - $P(X^2 > 1)$.
19. Signals that are either 0's or 1's are sent in a noisy communication circuit. The signal received is the signal sent plus a random variable, ϵ , that is, $N\left(0, \frac{1}{3}\right)$. If a 0 is sent, the receiver will record a 0 when the signal received is at most a value, v ; otherwise a 1 is recorded. Find v if the probability that a 1 is recorded when a 0 is actually sent is 0.90.
20. The diameter of a ball bearing is a normally distributed random variable with mean 6 and standard deviation $\frac{1}{2}$.
- What is the probability a randomly selected ball bearing has a diameter between 5 and 7?

- (b) If a diameter is between 5 and 7, the bearing can be sold for a profit of \$1. If the diameter is greater than 7, the bearing may be reworked and sold at a profit of \$0.50; otherwise, the bearing must be discarded at a loss of \$2. Find the expected value for the profit.
21. Capacitors from a manufacturer are normally distributed with mean $5 \mu\text{f}$ and standard deviation $0.4 \mu\text{f}$. An application requires four capacitors between $4.3 \mu\text{f}$ and $5.9 \mu\text{f}$. If the manufacturer ships 5 randomly selected capacitors, what is the probability that a sufficient number of capacitors will be within specifications?
22. The height, X , a college high jumper will clear each time she jumps is a normal random variable with mean 6 feet and variance 5.76 in^2 .
- What is the probability the jumper will clear $6'4''$ on a single jump?
 - What is the greatest height jumped with probability 0.95?
 - Assuming the jumps are independent, what is the probability that $6'4''$ will be cleared on exactly three of the next four jumps?
23. A Chamber of Commerce advertises that about 16% of the motels in town charge \$120 or more for a room and that the average price of a room is \$90. Assuming that room rates are approximately normally distributed, what is the variance in the room rates?
24. A commuting student has discovered that her commuting time to school is normally distributed; she has two possible routes for her trip. The travel time by Route A has mean 55 minutes and standard deviation 9 minutes while the travel time by route B has mean 60 minutes and standard deviation 3 minutes. If the student has at most 63 minutes for the trip, which route should she take?
25. The diameter of an electric cable is normally distributed with mean $0.8''$ and standard deviation $0.02''$.
- What is the probability the diameter will exceed $0.81''$?
 - The cable is considered defective if the diameter differs from the mean by more than $0.025''$. What is the probability a randomly selected cable is defective?
 - Suppose now that the manufacturing process can be altered and that the standard deviation can be changed while keeping the mean at 0.8. If the criterion in part (b) is used, but we want only 10% of the cables to be defective, what value of σ must be met in the manufacturing process?
26. A cathode ray tube for a computer graphics terminal has a fine mesh screen behind the viewing surface, which is under tension produced in manufacturing. The tension readings follow an $N(275, 40)$ distribution, where measurements are in units of mV.
- The minimum acceptable tension is 200 mV. What proportion of tubes exceed this limit?
 - Tension above 375 mV will tear the mesh. Of the acceptable screens, what proportion have tensions at most 375 mV?
 - Refer to part (a). Suppose it is desired to have 99.5% acceptable screens, and that a new quality control manager thinks he can reduce σ^2 to an acceptable level. What value of σ^2 must be attained?
27. The life lengths of two electronic devices, at D_1 and D_2 , have distributions $N(40, 6)$ and $N(45, 3)$, respectively. If the device is to be used for a 48-hour period, which device should be selected?

28. “One pound” packages of cheese are marketed by a major manufacturer, but the actual weight in pounds of a randomly selected package is a normally distributed random variable with standard deviation 0.02 lb. The packaging machine has a setting allowing the mean value to be varied.
- Federal regulations allow for a maximum of 5% short weights (weights below the claim on the label). What should the setting on the machine be?
 - A package labeled “one pound” sells for \$1.50, but costs only \$1 to produce. If short weight packages are not sold and if the machine’s mean setting is that in part (a), what is the expected profit on 1000 packages of cheese?

3.6 NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

Example 3.6.1

A component used in the construction of an electric motor is produced in a factory assembly line. In the past, about 10% of the components have proven unsatisfactory for use in the motor. The situation may then be modeled by a binomial process in which p , denoting the probability of an unsatisfactory component, is 0.10. The assembly line produces 500 components per day. If X denotes the number of unsatisfactory components, then the probability distribution function is

$$P(X = x) = \binom{500}{x} (0.10)^x (0.90)^{500-x}, \quad x = 0, 1, \dots, 500.$$

A graph of the distribution is shown in Figure 3.11.

Figure 3.11 is centered on the mean value, $500 \cdot (0.10) = 50$. Note that the possible values of X are from $X = 0$ to $X = 500$ but that the probabilities decrease rapidly, so we show only a small portion of the curve.

The graph in Figure 3.11 certainly appears to be normal. We note, however, that, although the eye may see a normal curve, there are in reality no points on the graph between the X values that are integers, since X can only be an integer. In Figure 3.12, we have used

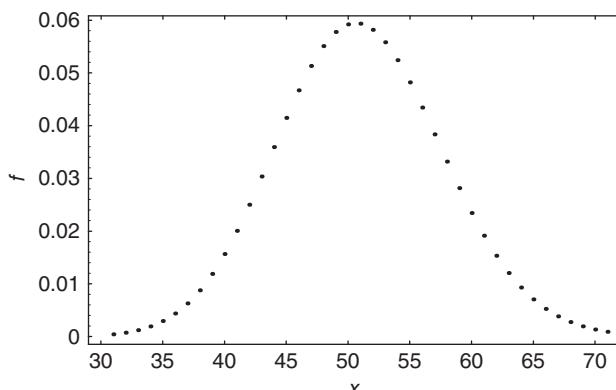


Figure 3.11 Binomial distribution, $n = 500$, $p = 0.10$.

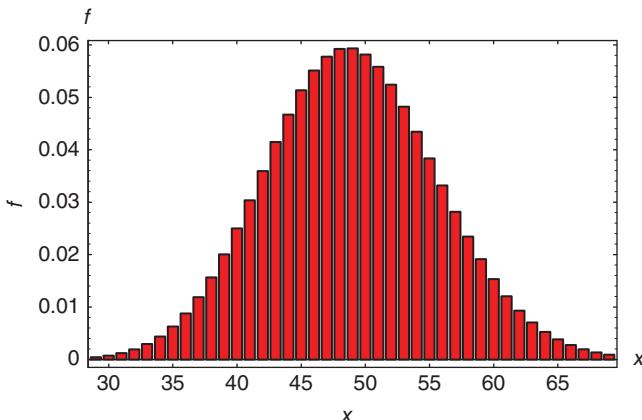


Figure 3.12 A histogram for the binomial distribution, $n = 500, p = 0.10$.

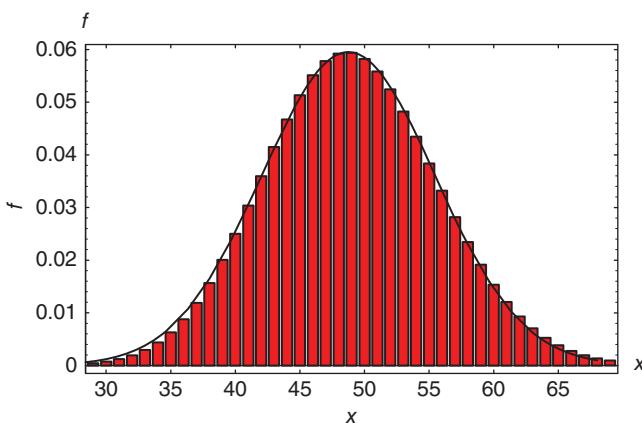


Figure 3.13 Normal curve approximation for the binomial, $n = 500, p = 0.10$.

the heights of the binomial curve in Figure 3.11 to produce a histogram. If we consider a particular value of X , say $X = 53$, notice that the base of the bar runs from 52.5 to 53.5 (both impossible values for X !) and that the height of the bar is $P(X = 53)$. Thus, the area of the bar at $X = 53$, since the base is of length 1, is $P(X = 53)$. This is the key that allows us to estimate binomial probabilities by the normal curve.

Figure 3.13 shows a normal curve imposed on the histogram of Figure 3.12. What normal curve should be used? It is natural to use a normal curve with mean and variance equal to the mean and variance of the binomial distribution which is being estimated, so we have used $N(500 \cdot 0.10, \sqrt{500 \cdot (0.10) \cdot (0.90)}) = N(50, \sqrt{45})$.

To estimate $P(X = 53)$, we find $P(52.5 \leq X \leq 53.5)$ using the approximation $N(50, \sqrt{45})$; this gives 0.0537716. The exact probability is 0.0524484.

As a final example, consider the probability the assembly line produces between 36 and 42 unsatisfactory components. This is estimated by $P(35.5 \leq X \leq 42.5)$ where $X \sim N(50, \sqrt{45})$. This is 0.116449. The exact probability is 0.118181.

When the sum of a large number of binomial probabilities is needed, a computer algebra system might be used to calculate the result exactly, although the computation might

well be lengthy. The same computer algebra system would also, more quickly and easily, calculate the relevant normal probabilities. In any event, whether the approximation is used or not, the approximation of the binomial distribution by the normal distribution is a striking fact. We will justify the approximation more thoroughly when we consider sums of random variables in Chapter 4. We note now that the approximation works well for moderate or large values of n , the quality of the approximation depending somewhat on the value of p .

In using a normal approximation to a binomial distribution, it is well to check the tail probabilities, hoping that these are small so that the approximation is an appropriate one. For example, if we want to approximate $P(9 \leq X \leq 31)$, we should check that the z -score for 31.5 exceeds 2.50 and that the z -score for 8.5 is less than -2.50 since these scores have about 2% of the curve in each tail.

EXERCISES 3.6

In solving these problems, find both the exact answer using a binomial distribution and the result given by the normal approximation.

1. A loaded coin comes up heads with probability 0.6. In 50 tosses, find the probability of between 28 and 32 heads.
2. Given X a binomial random variable with $n = 200$ and $p = 0.4$. Find $P(X = 80)$.
3. In 100 tosses of a fair coin, show that 50 heads and 50 tails is the most probable outcome, but that this event has probability of only about 0.08. Show that this compares favorably to the occurrence of at least 58 heads.
4. A manufacturer of components for electric motors has found that about 10% of the production will not meet customer specifications. Find the probability that in a lot of 500 components,
 - (a) exactly 53 do not meet customer specifications.
 - (b) between 36 and 42 (inclusive) components do not meet customer specifications.
5. A system of 50 components functions if at least 90% of the components function properly.
 - (a) Find the probability the system operates if the probability a component operates properly is 0.85.
 - (b) Suppose now that the probability a component operates properly is p . Find p if the probability the system operates properly is 0.95.
6. An acceptance sampling plan accepts a lot if at most 3% of a sample randomly chosen from a very large lot of items does not meet customer specifications. In the past, 2% of the items do not meet customer specifications. Find the probability the lot is accepted if the sample size is
 - (a) 10
 - (b) 100
 - (c) 1000.
7. A fair coin is tossed 1000 times. Let X denote the number of heads that occur. Find k so that $P(500 - k \leq X \leq 500 + k) = 0.90$.
8. An airline finds that, for a certain flight, 3% of the ticketed passengers do not appear for the flight. The plane holds 125 people. How many tickets should be sold if the airline wants to carry all the passengers who show up with probability 0.99?

9. Sam and Joe operate competing minibuses for travel from a central point in a city to the airport. Passengers appear and are equally likely to choose either minibus. During a given time period, 40 passengers appear. How many seats should each minibus have if Sam and Joe each want to accommodate all the passengers who show up for their minibus with probability 0.95?
10. A candidate in an election knows that 52% of the voters will vote for her. What is the probability that, out of 200 voters, she receives at least 50% of the votes?
11. A fair die is rolled 1200 times. Find the probability that at least 210 sixes appear.
12. In 10,000 tosses of a coin, 5150 heads appear. Is the coin loaded?
13. The length of life of a fluorescent fixture has an exponential distribution with expected life length 10,000 hours. Seventy of these bulbs operate in a factory. Find the probability that at most 40 of them last at least 8000 hours.
14. Suppose that X is uniformly distributed on $[0,10]$.
 - (a) Find $P(X > 7)$.
 - (b) Among 4 randomly chosen observations of X , what is the probability that at least 2 of these are greater than 7?
 - (c) What is the probability that, of 1000 observations, at least 730 are greater than 7?
15. Two percent of the production of an industrial process is not acceptable for sale. Suppose the company produces 1000 items a day. What is the probability a day's production contains between 1.4% and 2.2% nonacceptable items?

3.7 GAMMA AND CHI-SQUARED DISTRIBUTIONS

In Section 3.3 of this chapter, we considered the waiting time until the first Poisson event occurred and found that the waiting time followed an exponential distribution. We now want to consider the waiting time for the second Poisson event.

To make matters specific, suppose that the Poisson random variable has parameter λ and that Y is the waiting time for the second event. In y units of time, we expect $\lambda \cdot y$ events. Now if $Y \geq y$ there is at most 1 event in λy units of time, so

$$P(Y \geq y) = P(X = 0 \text{ or } 1) = \sum_{x=0}^1 \frac{e^{-\lambda \cdot y} (\lambda \cdot y)^x}{x!}.$$

It follows that

$$\begin{aligned} F(y) &= P(Y \leq y) = 1 - \sum_{x=0}^1 \frac{e^{-\lambda \cdot y} (\lambda \cdot y)^x}{x!} \\ F(y) &= 1 - e^{-\lambda \cdot y} - \lambda y e^{-\lambda \cdot y}, \text{ so} \\ f(y) &= \frac{dF(y)}{dy} = \lambda^2 y e^{-\lambda \cdot y}, \quad y \geq 0. \end{aligned}$$

A graph of $f(y)$ is shown in Figure 3.14.

Here, $f(y)$ is an example of a more general distribution, called the *gamma* distribution.

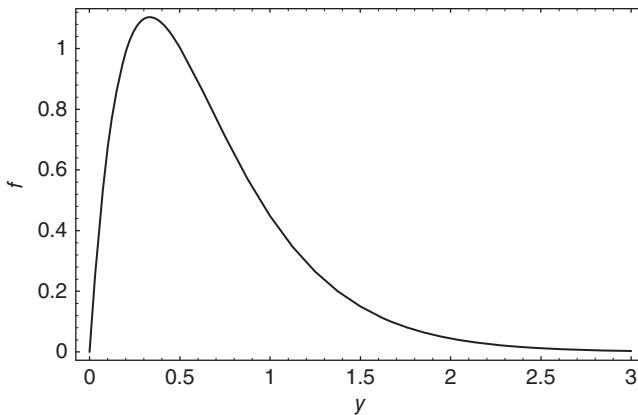


Figure 3.14 Waiting time for the second Poisson event.

Consider now waiting for the r th Poisson event from a Poisson distribution with parameter λ , and let Y denote the waiting time. Then at most $r - 1$ events must occur in y units of time, so

$$1 - F(y) = P(Y \geq y) = \sum_{x=0}^{r-1} \frac{e^{-\lambda \cdot y} \cdot (\lambda y)^x}{x!}.$$

It follows that

$$f(y) = \frac{dF(y)}{dy} = e^{-\lambda \cdot y} \sum_{x=0}^{r-1} \frac{\lambda^{x+1} y^x}{x!} - e^{-\lambda \cdot y} \sum_{x=1}^{r-1} \frac{\lambda^x y^{x-1}}{(x-1)!}.$$

This sum collapses leaving

$$f(y) = \frac{e^{-\lambda y} \lambda^r y^{r-1}}{(r-1)!}, \quad y \geq 0.$$

Here, $f(y)$ defines what we call a *gamma distribution*. The exponential distribution is a special case of $f(y)$ when $r = 1$.

Since $f(y)$ must be a probability density function, it follows that

$$\int_0^\infty \frac{e^{-\lambda y} \lambda^r y^{r-1}}{(r-1)!} dy = 1.$$

Now, letting $x = \lambda \cdot y$, it follows that

$$\int_0^\infty e^{-x} x^{r-1} dx = (r-1)! \text{ if } r \text{ is a positive integer.}$$

This integral is commonly denoted by $\Gamma(r)$. So

$$\Gamma(r) = \int_0^\infty e^{-x} x^{r-1} dx = (r-1)! \text{ if } r \text{ is a positive integer.}$$

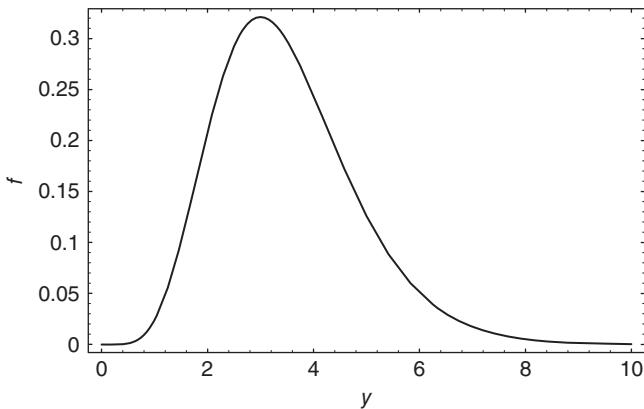


Figure 3.15 A gamma distribution.

Now consider the expected value:

$$\begin{aligned} E(Y) &= \int_0^\infty y \cdot \frac{e^{-\lambda y} \lambda^r y^{r-1}}{(r-1)!} dy \\ &= \frac{r}{\lambda} \int_0^\infty \frac{e^{-\lambda y} \lambda^r y^r}{r!} dy \\ \text{so, since } &\int_0^\infty e^{-\lambda y} (\lambda y)^r dy = r! \\ E(Y) &= \frac{r}{\lambda}. \end{aligned}$$

It can also be shown that

$$\text{Var}(Y) = \frac{r}{\lambda^2}.$$

Graphs of $f(y)$ are also easy to produce using a computer algebra system. Figure 3.15 shows $f(y)$ for $r = 7$ and $\lambda = 2$.

Again, the normal-like appearance of the graph is striking and so we consider a numerical example to investigate this phenomenon. It will follow from considerations given in Chapter 4 that $f(y)$ does indeed approach a normal distribution.

Suppose the Poisson process has $\lambda = 2$ and we consider Y , the waiting time for the seventh occurrence. Then

$$f(y) = \frac{e^{-2y} 2^7 y^6}{6!}, \quad y \geq 0.$$

It follows that

$$P(2 \leq Y \leq 5) = \int_2^5 \frac{e^{-2y} 2^7 y^6}{6!} dy = 0.759185.$$

Using earlier formulas, $E(Y) = \frac{7}{2}$ and $\text{Var}(Y) = \frac{7}{4}$, so the normal curve approximation uses the normal curve $N(\frac{7}{2}, \frac{\sqrt{7}}{2})$. This gives an approximation of 0.743161. The normal curve approximates this gamma distribution fairly well here.

We return now to the gamma function. We see that if r is a positive integer, $\Gamma(r) = (r - 1)!$ so the graph of $\Gamma(r)$ passes through the points $(r, r!)$. But $\Gamma(r)$ has values when r is not a positive integer. For example,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-y} y^{-1/2} dy.$$

Letting $y = \frac{z^2}{2}$ in this integral as well as inserting factors of $\sqrt{2\pi}$ results in

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{2} \cdot \sqrt{2\pi} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

The integral is $1/2$ the area under a standard normal curve and so is $1/2$. So

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Consequently, the gamma distribution is then often written as

$$f(y) = \frac{e^{-\lambda y} \lambda^r y^{r-1}}{\Gamma(r)}, \quad y \geq 0.$$

A special case of the gamma distribution occurs when $\lambda = 1/2$ and $r = n/2$. The distribution then takes the form

$$f(x) = \frac{e^{-\frac{x}{2}} \cdot x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, \quad x \geq 0.$$

Here X is said to follow a *Chi-squared distribution with n degrees of freedom*, which we denote by χ_n^2 . The exponent 2 has no particular significance; it is simply part of the notation which is in general use. We will discuss this distribution in greater detail in Chapter 4, but, since it is a special case of the gamma distribution, and since it has a large variety of practical applications, we show an example of its use now.

First, let us look at some graphs of χ_n^2 for some specific values of n . These graphs, which can be produced by a computer algebra system, are shown in Figure 3.16.

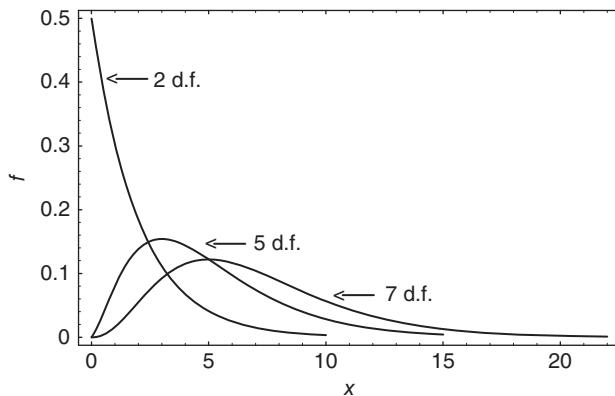


Figure 3.16 Some χ^2 distributions.

Again we note the approach to normality as n increases. This fact will be established in Chapter 4.

Example 3.7.1

A production line produces items that can be classified as “Good,” “Bad,” or “Rework,” the latter category indicating items that are not satisfactory on first production but which could be subject to further work and sold as good items. The line, in the past, has been producing 85%, 5%, and 10% in the three categories, respectively. 800 items are produced in 1 day of which 665 are good, 30 are bad, and 95 need to be reworked. These numbers, of course, are not *exactly* those expected, but the increase in items to be reworked worries the plant management. Has the process in fact changed or is the sample simply the result of random variation?

This is another instance of statistical inference since we use the sample to draw a conclusion regarding the population, or universe, from which it is selected. We lack of course an obvious random variable to use in this case. We might begin by computing the expected numbers in the three categories, which are 680, 40, and 80. Let the observed number in the i th category be O_i and the expected number in the i th category be E_i . It can then be shown, although not at all easily, that

$$\sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

follows a χ^2_{n-1} distribution where n is the number of categories.

In this case, we calculate

$$\chi^2_2 = \frac{(665 - 680)^2}{680} + \frac{(30 - 40)^2}{40} + \frac{(95 - 80)^2}{80} = 5.643382353.$$

The χ^2_2 curve is an exponential distribution,

$$f(x) = \left(\frac{1}{2}\right) e^{-\frac{x}{2}}, \quad x \geq 0.$$

This point is quite far out in the right-hand tail of the χ^2_2 distribution as Figure 3.17 indicates.

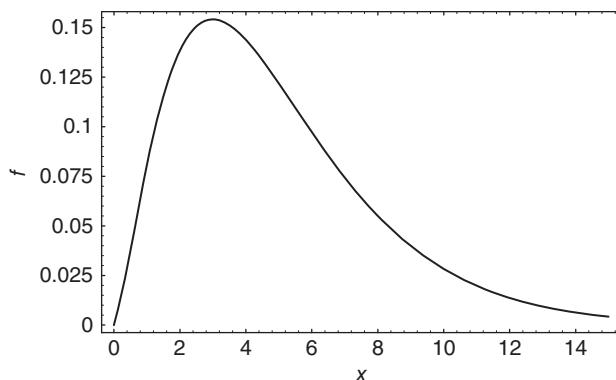


Figure 3.17 χ^2_2 distribution.

It is easy to find that $P(\chi_2^2 > 5.643382353) = 0.0595052$. So we are faced with a decision: if the process has in fact not changed at all, the value of χ_2^2 will exceed that of our sample only about 6% of the time. That is, simple random variation will produce this value of χ_2^2 , or an even greater value, about 6% of the time. Since this is fairly small, we would probably conclude that the sample is not simply a consequence of random variation and that the production process had changed.

EXERCISES 3.7

1. A Poisson distribution has parameter 2.
 - (a) Find the probability distribution for the waiting time for the third event.
 - (b) Find the probability waiting time of six events.
2. Grades in a statistics course are A, 12; B, 20; and C, 8. Is the professor correct in saying that the respective probabilities are A, 15%; B, 60%; and C, 25%?
3. A book publisher finds that the yearly sales, X , of a textbook (in thousands of books) follows a gamma distribution with $\lambda = 10$ and $r = 5$.
 - (a) Find the mean and variance for the yearly sales.
 - (b) Find the probability that the number of books sold in 1 year is between 200 and 600.
 - (c) Sketch the probability density function for X .
4. Particles are emitted from a radioactive source with three particles expected per minute.
 - (a) Find the probability density function for the waiting time for the fourth particle to be emitted.
 - (b) Find the mean and variance for the waiting time for the fourth particle.
 - (c) Find the probability that at least 20 seconds elapse before the fourth particle is emitted.
5. Weekly sales, S , in thousands of dollars, for a small shop follow a gamma distribution with $\lambda = 1$ and $r = 2$
 - (a) Sketch the probability density function for S .
 - (b) Find $P[S > 2 \cdot E(S)]$.
 - (c) Find $P(S > 1.5 | S > 1)$.
6. Yearly snowfall, S , in inches, in Southern Colorado follows a gamma distribution with $\lambda = 2$ and $r = 3$.
 - (a) Find the probability at least 8 inches of snow will fall in a given year.
 - (b) If 6 inches of snow have fallen in a given year, what is the probability of at least two more inches of snow?
 - (c) Find $P(\mu - \sigma \leq S \leq \mu + \sigma)$.
7. Show, using integration by parts, that $\Gamma(n) = (n - 1)!$ if n is a positive integer.
8. Show that $\Gamma\left(n + \frac{1}{2}\right) = \frac{\Gamma(n+1)\sqrt{\pi}}{2^n \Gamma\left(\frac{n+1}{2}\right)}$.
9. Show that $\binom{-a}{k} = \frac{(-1)^k \Gamma(a+k)}{\Gamma(k+1) \Gamma(a)}$.

- 10.** If X is a standard normal variable, then it is known that X^2 follows a χ_1^2 distribution. Calculate $P(X^2 < 2)$ in two different ways.
- 11.** A die is tossed 60 times with the following results:

Face	1	2	3	4	5	6
Observations	8	12	9	8	10	13

Is the die fair?

- 12.** Show that $E(\chi_n^2) = n$ and that $\text{Var}(\chi_n^2) = 2n$ by direct integration.
- 13.** Phone calls come into a switchboard according to a Poisson process at the rate of 5 calls per hour. Let Y denote the waiting time for the first call to arrive.
- (a) Find $P(Y > y)$.
- (b) Find the probability density function for Y .
- (c) Find $P(Y \geq 10)$.

3.8 WEIBULL DISTRIBUTION

We considered the reliability of a product and the hazard rate in Section 3.3. We showed there that a constant hazard rate produced an exponential time-to-failure law. Now let us consider nonconstant hazard rates. A variety of time-to-failure laws is used to produce non-constant hazard rates. As an example, we consider a Weibull distribution here since it provides such a variable hazard rate. In addition to providing a variable hazard rate, a Weibull distribution can be shown to hold when the performance of a system is governed by the least reliable of its components, which is not an uncommon occurrence.

We use the phrase “a Weibull distribution” to point out the fact that the distributions about to be described vary widely in appearance and properties and in fact define an entire family of related distributions.

We recall some facts from Section 3.4 first. Recall that if $f(t)$ defines a time-to-failure probability distribution, then the reliability function is

$$R(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t).$$

The hazard rate is

$$h(t) = \frac{f(t)}{R(t)} = -\frac{R'(t)}{R(t)}. \quad (3.5)$$

Now suppose that

$$h(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1}, \quad \alpha > 0, \quad \beta > 0, \quad t \geq 0.$$

Formula (3.5) indicates that

$$-\frac{R'(t)}{R(t)} = \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \text{ from which we find}$$

$$R(t) = e^{-\left(\frac{t}{\beta}\right)^\alpha} \quad \text{since } t \geq 0.$$

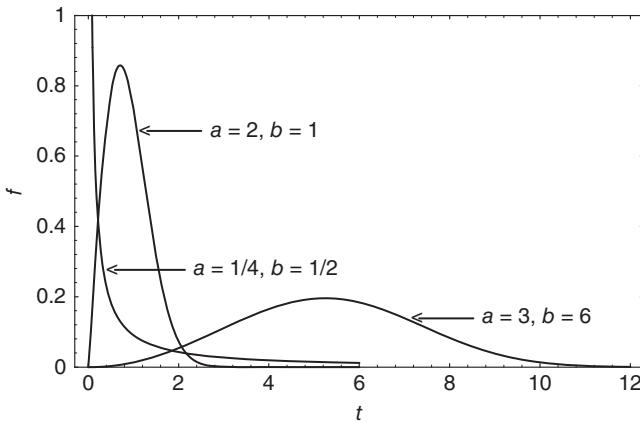


Figure 3.18 Some Weibull distributions.

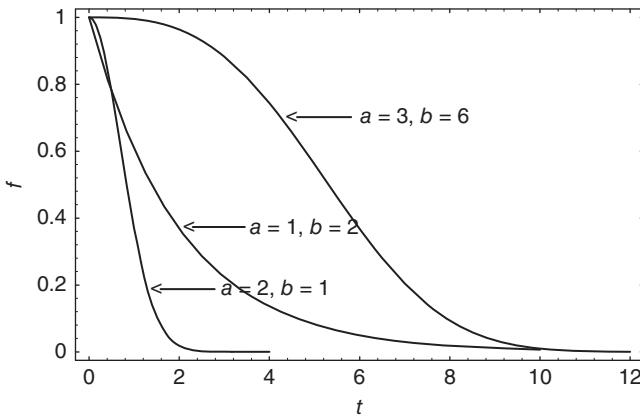


Figure 3.19 Some reliability functions.

We also find that

$$f(t) = -\frac{dR(t)}{dt} = \frac{\alpha}{\beta^\alpha} t^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha}, \quad t \geq 0.$$

$f(t)$ describes the *Weibull* family of probability distributions. Varying α and β produces graphs of different shapes as shown in Figure 3.18.

The reliability functions, $R(t)$, also differ widely as shown in Figure 3.19.

The mean and variance of a Weibull distribution are found using the gamma function. We find, for a Weibull distribution with parameters α and β , that

$$\begin{aligned} E(T) &= \beta \cdot \Gamma\left(\frac{1}{\alpha} + 1\right) \quad \text{and} \\ \text{Var}(T) &= \beta^2 \cdot \left[\Gamma\left(\frac{2}{\alpha} + 1\right) - \left\{ \Gamma\left(\frac{1}{\alpha} + 1\right) \right\}^2 \right]. \end{aligned}$$

EXERCISES 3.8

1. The lifetime of a part is a Weibull random variable with $\alpha = 2$ and $\beta = 10$ years.
 - (a) Sketch the probability density function.
 - (b) Find the probability the part lasts between 3 and 7 years.
 - (c) Find the probability a 3-year-old part lasts at least 7 years.
2. For the part described in problem 1,
 - (a) If the part carries a 15-year warranty, what percentage of the parts are still good at the end of the warranty period?
 - (b) What should the warranty be if it is desired to have 99% of the parts still good at the end of the warranty period?
3. The hazard rate for a generator is $10^{-4} t/\text{hour}$.
 - (a) Find $R(t)$, the reliability function.
 - (b) Find the expected length of life for the generator.
 - (c) Find the probability the generator lasts at least 150 hours.
4. A component's life length follows a Weibull distribution with $\alpha = 1/3$, $\beta = 1/27$.
 - (a) Plot the probability density function for the life length.
 - (b) Determine the hazard rate.
 - (c) Find the probability the component lasts at least 2 hours.
5. One component of a satellite has a hazard rate of $10^{-6}t^2/\text{hour}$.
 - (a) Plot $R(t)$, the reliability function.
 - (b) Find the probability the component fails within 100 hours.
6. How many of the components for the satellite in problem 5 must be used if we want the probability that at least one lasts at least 100 hours to be 0.99?
7. Find the median of a Weibull distribution with parameters α and β .
8. A Weibull random variable, X , has $\alpha = 4$ and $\beta = 30$. Compare the exact value of $P(20 < X < 30)$ with the normal approximation to that probability.
9. It has been noticed that 56% of a heavily used industrial bulb last at most 10,000 hours. Assuming that the life lengths of these bulbs follow a Weibull distribution with $\beta = 3$, what proportion of the bulbs will last at least 15,000 hours?

CHAPTER REVIEW

Random variables that can assume any value in an interval or intervals are called *continuous* random variables; they are the subject of this chapter.

It is clear that the probability distribution function, $f(x) = P(X = x)$, which was of primary importance in our work with discrete random variables is of little use when X is continuous, since, in that case, $P(X = x) = 0$ for all values of X , so this function carries no information whatsoever. It is possible, however, to distinguish different continuous random variables by a *probability density function*, $f(x)$, which has the following properties:

1. $f(x) \geq 0$

2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $\int_a^b f(x)dx = P(a \leq X \leq b)$

We study several of the most important probability density functions in this chapter.

The mean and variance of a continuous random variable can be calculated by

$$E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x) dx \text{ and}$$

$$\text{Var}(X) = \sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$$

provided, of course, that the integrals are convergent.

It is often useful to use the fact that

$$\sigma^2 = E(X^2) - [E(X)]^2$$

when calculating σ^2 .

The first distribution considered was the *uniform* distribution defined by

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

We found that

$$\mu = \frac{b+a}{2} \text{ and that } \sigma^2 = \frac{(b-a)^2}{12}.$$

The most general form of the *exponential* distribution is

$$f(x) = \lambda e^{-\lambda(x-a)}, \quad x \geq a \quad \text{where } \lambda > 0.$$

A computer algebra program or direct integration shows that

$$E(X) = \int_a^{\infty} x \cdot f(x) dx = a + \frac{1}{\lambda} \text{ and that}$$

$$V(X) = \frac{1}{\lambda^2}.$$

An interesting fact is that the waiting time for the first occurrence of a Poisson random variable is an exponential variable.

We then discussed *reliability* since this is an important modern application of probability theory. We defined the reliability as

$$R(t) = P(T > t)$$

where T is a random variable. The reliability then gives the probability a component whose lifetime is the random variable T that lasts at least t units of time.

The (instantaneous) *hazard rate* is the probability, per unit of time, that an item that has lasted t units of time will last Δt more units of time. We found that the hazard rate, $H(t)$, is

$$H(t) = \frac{f(t)}{1 - F(t)},$$

where $f(t)$ and $F(t)$ are the probability density and distribution functions for T , respectively.

The normal distribution, without doubt the most important continuous distribution of all, was considered next. We showed that its most general form is

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

If X has the normal distribution above, we write $X \sim N(\mu, \sigma)$.

An important fact is that if $X \sim N(\mu, \sigma)$ and if $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0, 1)$, a distribution that is referred to as the *standard* normal distribution. This fact allows a wide variety of normal curve calculations to be carried out using a single normal curve. This is a very unusual circumstance in probability theory, distributions often being highly dependent on sample size, for example, as we will see in later chapters.

The normal curve arises in a multitude of places; one of its most important uses is that it can be used to approximate a binomial distribution. We discussed the approximation of a binomial variable with parameters n and p by a $N(np, \sqrt{npq})$ curve.

Two distributions whose importance will be highlighted in later chapters are the *gamma* and *Chi-squared* distributions. The gamma distribution arises when we wait for the r th Poisson occurrence. Its probability density function is

$$f(y) = \frac{e^{-\lambda y} \lambda^r y^{r-1}}{(r-1)!}, \quad y \geq 0,$$

where λ is the parameter in the Poisson distribution.

The Chi-squared distribution arises when $\lambda = \frac{1}{2}$ and $r = \frac{n}{2}$.

Finally, we considered the Weibull family of distributions whose probability density functions are members of the family

$$f(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha}, \quad \alpha > 0, \quad \beta > 0, \quad t \geq 0.$$

It is fairly easy to show that

$$\begin{aligned} E(T) &= \beta \cdot \Gamma\left(\frac{1}{\alpha} + 1\right) \quad \text{and} \\ \text{Var}(T) &= \beta^2 \cdot \left[\Gamma\left(\frac{2}{\alpha} + 1\right) - \left\{ \Gamma\left(\frac{1}{\alpha} + 1\right) \right\}^2 \right]. \end{aligned}$$

The Weibull distribution is of importance in reliability theory; several examples were given in the chapter.

PROBLEMS FOR REVIEW

Exercises 3.1 # 2, 3, 4, 5, 7, 10, 14, 18

Exercises 3.2 # 1, 2, 4, 7, 9, 10

Exercises 3.4 # 1, 2, 5, 7, 9, 10, 16

Exercises 3.5 # 1, 3, 5, 8, 9, 10, 15, 16, 17, 19, 23, 26

Exercises 3.6 # 1, 2, 3, 7, 10, 12

Exercises 3.7 # 2, 3, 6, 8, 11

Exercises 3.8 # 1, 3, 4, 6, 7, 9

SUPPLEMENTARY EXERCISES FOR CHAPTER 3

1. A machining operation produces steel shafts having diameters that are normally distributed with mean 1.005 inches and standard deviation 0.01 inches. If specifications call for diameters to fall in the interval 1.000 ± 0.02 inches, what percentage of the steel shafts will fail to meet specifications?
2. Electric cable is made by two different manufacturers, each of whom claims that the diameters of their cables are normally distributed. The diameters, in inches, from Manufacturer I are $N(0.80, 0.02)$ while the diameters from Manufacturer II are $N(0.78, 0.03)$. A purchaser needs cable that has diameter less than 0.82 inches. Which manufacturer should be used?
3. A buyer requires a supplier to deliver parts that differ from 1.10 by no more than 0.05 units. The parts are distributed according to $N(1.12, 0.03)$. What proportion of the parts do not meet the buyer's specifications?
4. Manufactured parts have lifetimes in hours, X , that are distributed $N(1000, 100)$. If $800 \leq X \leq 1200$, the manufacturer makes a profit of \$50 per part. If $X > 1200$, the profit per part is \$75; otherwise, the manufacturer loses \$25 per part. What is the expected profit per part?
5. The annual rainfall (in inches) in a certain region is normally distributed with $\mu = 40$, $\sigma = 4$. Assuming rainfalls in different years are independent, what is the probability that in 2 of the next 4 years the rainfall will exceed 50 inches?
6. The weights of oranges in a good year are described by a normal distribution with $\mu = 16$ and $\sigma = 2$ (ounces).
 - (a) What is the probability that a randomly selected orange has weight in excess of 17 ounces?
 - (b) Three oranges are selected at random. What is the probability the weight of exactly one of them exceeds 17 ounces?
 - (c) How many oranges out of 10000 are expected to have weight between 15.4 and 17.3 ounces?
7. A sugar refinery has three processing plants, all receiving raw sugar in bulk. The amount of sugar in tons that each of the plants can process in a day has an exponential distribution with mean 4.
 - (a) Find the probability a given plant processes more than 4 tons in a day.
 - (b) Find the probability that at least two of the plants process more than 4 tons in a day.

8. The length of life in hours, X , of an electronic component has an exponential probability density function with mean 500 hours.
- Find the probability that a component lasts at least 900 hours.
 - Suppose a component has been in operation for 300 hours. What is the probability it will last for another 600 hours?
9. Students in an electrical engineering laboratory measure current in a circuit using an ammeter. Due to several random factors, the measurement, X , follows the probability density function

$$f(x) = 0.025 \cdot x + b, \quad 2 < x < 6.$$

- Show that $b = 0.15$.
 - Find the probability the measurement of the current exceeds 3 amps.
 - Find $E(X)$.
 - Find the probability that all three laboratory partners measure the current independently as less than 4 amps.
10. Let X be a random variable with probability density function

$$f(x) = \begin{cases} k & 1 \leq x \leq 2 \\ k(3-x) & 2 \leq x \leq 3. \end{cases}$$

- Find k .
 - Calculate $E(X)$.
 - Find the cumulative distribution function, $F(x)$.
11. The percentage, X , of antiknock additive in a particular gasoline, is a random variable with probability density function

$$f(x) = kx^3(1-x), \quad 0 < x < 1.$$

- Show that $k = 20$.
 - Evaluate $P[X < E(X)]$.
 - Find $F(x)$.
12. Suppose that $f(x) = 3x^2$, $0 < x < 1$, is the probability density function for some random variable X . Find $P\left(X \geq \frac{1}{2} \mid X \geq \frac{1}{4}\right)$.
13. A point B is chosen at random on a line segment AC of length 10. A right-angled triangle with sides AB and BC is constructed. Determine the probability that the area of the triangle is at least 7 square units.
14. A random variable, X , has probability density function

$$f(x) = \begin{cases} ax & 0 \leq x \leq 3 \\ 6a - ax & 3 \leq x \leq 6. \end{cases}$$

- Show that $a = \frac{1}{9}$.
- Find $P(X \geq 4)$.

15. Verify Tchebycheff's inequality for $k = \sqrt{2}$ for the probability density function

$$f(x) = \frac{1}{2}(x+1), \quad -1 < x < 1.$$

16. Suppose X has the distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ ax - \frac{1}{4}x^2 & 0 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

- (a) Find a .
 (b) Find $P(X \geq 1)$.

17. Find the mean and variance of the random variable X whose probability density function is

$$f(x) = \frac{3}{4}(1-x)(x-3), \quad 1 \leq x \leq 3.$$

18. A random variable X has probability density function

$$f(x) = \begin{cases} 1+x & -1 < x < 0 \\ 1-x & 0 \leq x < 1. \end{cases}$$

Find $E[X^2 - 2X + 2]$.

19. (a) Determine k so that $f(x) = kxe^{-x^2}$ is a probability density function for some non-negative random variable X .

- (b) Determine $F(x)$ and sketch it.

20. The time (in seconds) a car has to wait for a certain traffic light has probability density function

$$f(x) = \begin{cases} \frac{x}{2500} & 0 \leq x \leq 50 \\ \frac{1}{25} - \frac{x}{2500} & 50 \leq x \leq 100. \end{cases}$$

- (a) What is the probability that the waiting time is between 25 and 75 seconds?
 (b) If a car has waited 25 seconds, what is the probability it will wait at least 25 seconds more?

21. One hundred independent observations are made of the random variable X whose probability density function is

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2. \end{cases}$$

Find the probability that at least 20 of these observations exceed 1.5.

22. X is a random variable with distribution function

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{x+2}{4} & -2 \leq x \leq 2 \\ 1 & x > 2. \end{cases}$$

Identify the probability density function for X .

23. A random variable X has probability density function

$$f(x) = \begin{cases} 2x & 0 \leq x \leq \frac{1}{2} \\ 6 - 6x & \frac{1}{2} \leq x \leq 1. \end{cases}$$

Find $F(x)$, being sure to specify this for any value of x .

24. Find the constant c that makes $g(y) = \frac{c}{y}$, $1 \leq y \leq 2$, a probability density function.
 25. Find the mean and variance of the random variable X whose probability density function is

$$f(x) = \begin{cases} 1-x & 0 \leq x \leq 1 \\ x-1 & 1 \leq x \leq 2. \end{cases}$$

26. A random variable X has probability density function

$$f(x) = \frac{1}{2x}, \quad \frac{1}{e} < x < e.$$

Two independent observations are made on X . Find the probability that one observation is less than 1 and that the other observation is greater than 1.

27. A random variable X has distribution function

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{6} & -2 \leq x < -1 \\ \frac{1}{2} & -1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

Find $f(x)$.

28. The probability density function for X , the lifetime in hours of a certain type of electronic device, is given by

$$f(x) = \begin{cases} \frac{10}{x^2} & x > 10 \\ 0 & x \leq 10. \end{cases}$$

- (a) Find $P(X > 20)$.
(b) Find $F(x)$.
29. A random variable T has probability density function

$$g(t) = k(1+t)^{-2}, \quad t \geq 0.$$

Find $P(T \geq 2 | T \geq 1)$.

30. A player can win a solitaire card game with probability $1/12$. Find the probability that the player wins at least 10% of 500 games played.

Chapter 4

Functions of Random Variables; Generating Functions; Statistical Applications

4.1 INTRODUCTION

We now want to expand our applications of statistical inference first encountered in Chapter 2. In particular we want to consider tests of hypotheses and the construction of confidence intervals when continuous random variables are involved; we will also introduce simple linear regression. These considerations have direct bearing on problems of data analysis such as that encountered in the following situation.

A production process has been producing bearings with mean diameter 2.60 in.; the diameters exhibit some variability around this average value with the standard deviation of the diameters believed to be 0.03 in. A quality control inspector chooses a random sample of ten bearings and finds their average diameter to be 2.66 in. Has the process changed?

The quality control inspector here has a single observation, namely 2.66 in., the average of ten observations. This is most commonly the situation: only one sample is available; decisions must be made on the basis of that single sample. Nonetheless we can speculate on what would happen were the sampling to be repeated. In that case, another sample average will most likely occur. In order to decide whether 2.66 in. is unusual or not, we must know the probability distribution of these sample means so that the variation in the mean from sample to sample can be assessed. We can then base a test of the hypothesis that the process mean has not changed on that probability distribution. Confidence intervals can similarly be constructed, but again, the probability distribution of the sample mean must be known.

Determination of the probability distribution here is not particularly easy so we first need to make some mathematical considerations. This will not only enable us to consider the example at hand, but will also allow us to solve many other complex problems arising in the analysis of data. We also must investigate the distribution of the sample variance arising from samples drawn from a continuous distribution.

We begin by considering functions of random variables; sums and averages arising from samples are examples of complex functions of sample values. Special functions called *generating functions* provide a particularly powerful technique for solving these problems. While developing these techniques we will solve many interesting problems in probability.

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

Finally we will show several practical statistical problems and their solution, including a statistical process control chart.

4.2 SOME EXAMPLES OF FUNCTIONS OF RANDOM VARIABLES

In Chapter 3, the following problem was considered: an observation, X , was made from a uniform distribution on the interval $[0, 1]$ and then a square of side X was formed. What is the expected value of its area?

This problem is fairly easily solved. Since $E(X) = \frac{1}{2}$ and

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{12}, \text{ it follows that}$$

$$E(\text{Area}) = E(X^2) = \text{Var}(X) + [E(X)]^2 = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}.$$

Other problems of a similar nature, however, may not be quite so easy. As another example, suppose X is an exponential random variable with mean α and we seek $E(\sqrt{X})$. It would be unreasonable to think, for example, that $E(\sqrt{X}) = \sqrt{E(X)}$. This expectation is, after all, an integral and integrals rarely behave in such simple manner.

The reader is invited to calculate, or use a computer algebra system, in this example to find that

$$E(\sqrt{X}) = \int_0^\infty \frac{\sqrt{x}}{\alpha} e^{-\frac{x}{\alpha}} dx = \frac{1}{2} \sqrt{\pi\alpha}.$$

Another frequently used technique for evaluating the integral encountered earlier is to select a random sample of values from an exponential distribution and then calculate the average value of their square roots. This technique, widely used in problems that prove difficult for analytical techniques, is known as *simulation*. A computer program chose 1000 observations from an exponential distribution with $\alpha = 4$ and then calculated the mean of the square roots of these values. The observed value was 1.800, while the expected value is $\sqrt{\pi} = 1.7725$, so the simulation produced a value quite close to the expected value.

Expectations of many functions of random variables can be carried out by using the probability density function of X directly. In the first example (where X is uniformly distributed on $[0, 1]$ and denotes the length of the side of a square), suppose we wanted the probability that the area of the square was between $1/2$ and $3/4$. We can calculate

$$P\left(\frac{1}{2} \leq X^2 \leq \frac{3}{4}\right) = P\left(\frac{1}{\sqrt{2}} \leq X \leq \frac{\sqrt{3}}{2}\right) = \frac{\sqrt{3}}{2} - \frac{1}{\sqrt{2}} = 0.15892,$$

using the distribution of X directly.

In the second example, supposing X is a random observation from an exponential distribution with mean α , we calculate, for example,

$$\begin{aligned} P(1 \leq \sqrt{X} \leq 2) &= P(1 \leq X \leq 4) \\ &= \int_1^4 (1/\alpha) \cdot e^{-\frac{x}{\alpha}} \cdot dx = e^{-\frac{1}{\alpha}} - e^{-\frac{4}{\alpha}}, \end{aligned}$$

so often probabilities involving functions of random variables can be found from the distributions of the random variables themselves.

Now suppose that we have two independent observations of a random variable X and we consider the sum of these, $X_1 + X_2$. This is certainly a random variable. How can we calculate $P(X_1 + X_2 \leq 2)$ if X_1 and X_2 are, for example, independent observations from the exponential distribution? Clearly, this problem is not as simple as the preceding ones.

It is fortunate that there is another way to look at these problems. It turns out that this other view will solve these problems and has, in addition, considerable implications for the solutions of much more complex problems, solutions that are not easily found in any other way. Our approach will also explain why normality has occurred so frequently in our problems; the reason for this is not simple, as the reader might expect.

The expressions X^2 , \sqrt{X} , and $X_1 + X_2$ are *functions* of the random variable X . Since X is a random variable, so too are these functions of X ; then they have probability distributions. If these probability distributions could be determined, then the earlier problems, as well as many others, could be solved, so we now consider one method for determining the probability distribution of a function of the random variable X .

4.3 PROBABILITY DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES

We begin with an example discussed in Chapter 3.

Example 4.3.1

Suppose that a random variable X has a standard normal distribution, that is, $X \sim N(0, 1)$. Consider the random variable $Y = X^2$, so that Y is a quadratic function of the random variable X . What is the probability density function for Y ?

Our answer depends on the simple fact that when the derivative exists, and where $f(x)$ and $F(x)$ denote the probability density function and distribution function respectively, then

$$\frac{dF(x)}{dx} = f(x).$$

Let $g(y)$ and $G(y)$ denote the probability density function and the distribution function, respectively, for the random variable Y . Our basic strategy is to find $G(y)$ and then to differentiate it, using the property above, to produce $g(y)$.

Here

$$\begin{aligned} G(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}), \quad \text{so} \\ G(y) &= F(\sqrt{y}) - F(-\sqrt{y}), \end{aligned}$$

by a property of distribution functions. Now we differentiate throughout to find that

$$g(y) = \frac{dG(y)}{dy} = \frac{dF(\sqrt{y}) - dF(-\sqrt{y})}{dy}.$$

Now we must be careful because $\frac{dF(\sqrt{y})}{dy} \neq f(\sqrt{y})$. The problem lies in the fact that the variables in the numerator and in the denominator are not the same. However the chain rule comes to our rescue and we find that

$$g(y) = \frac{dF(\sqrt{y})}{d(\sqrt{y})} \cdot \frac{d(\sqrt{y})}{dy} - \frac{dF(-\sqrt{y})}{d(-\sqrt{y})} \cdot \frac{d(-\sqrt{y})}{dy}.$$

This becomes

$$g(y) = \frac{f(\sqrt{y})}{2\sqrt{y}} + \frac{f(-\sqrt{y})}{2\sqrt{y}}. \quad (4.1)$$

But

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, -\infty < y < \infty, \quad \text{and}$$

$$f(\sqrt{y}) = f(-\sqrt{y}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}, \quad y > 0, \quad \text{so}$$

$$g(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, \quad y > 0.$$

This is the χ_1^2 variable, first seen in Section 3.7. The domain of values for Y is established from that for X : since $-\infty < X < \infty$, then $-\infty < \sqrt{y} < \infty$ or $\sqrt{y} < \infty$, so $y \geq 0$. The same domain is correct for $-\sqrt{y}$ above.

The calculation that $\int_0^\infty g(y) dy = 1$, and the fact that $g(y) \geq 0$, checks our work and shows that $g(y)$ is a probability density function.

This process works well when the derivatives involved can be evaluated and that is often the case in the instances that interest us here.

In the previous example, Y is a quadratic function of X and the resulting distribution for Y bears little resemblance to that for X . We expect that a linear function would preserve the shape of the distribution in a sense. We consider a specific example first.

Example 4.3.2

Suppose X is uniform on $[3, 5]$ so that $f(x) = \frac{1}{2}$, $3 \leq x \leq 5$. Let $Y = \frac{X-2}{3}$, a linear function of X . Again we find the distribution function and differentiate it. Here

$$G(y) = P(Y \leq y) = P\left(\frac{X-2}{3} \leq y\right) = P(X \leq 3y+2) = F(3y+2).$$

Then

$$g(y) = \frac{dG(y)}{dy} = \frac{dF(3y+2)}{dy} = \frac{dF(3y+2)}{d(3y+2)} \cdot \frac{d(3y+2)}{dy}, \quad \text{so}$$

$$g(y) = f(3y+2) \cdot 3 = \frac{3}{2}.$$

To establish the domain for y , note that $f(3y + 2) = \frac{1}{2}$ if $3 \leq 3y + 2 \leq 5$ which simplifies to $\frac{1}{3} \leq y \leq 1$, producing the final result:

$$g(y) = \frac{3}{2}, \quad \frac{1}{3} \leq y \leq 1.$$

This is a probability density function since $g(y) \geq 0$ and $\int_{\frac{1}{3}}^1 g(y) dy = 1$.

We observe that the linear transformation $Y = \frac{X-2}{3}$ of X preserves the uniform distribution with which we began.

To consider the problem of a linear transformation in general, suppose that X has probability density function $f(x)$ and that $Y = aX + b$ for some constants a and b provided that $a \neq 0$. Then

$$G(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F\left(\frac{y-b}{a}\right) \text{ so}$$

$$g(y) = \frac{dG(y)}{dy} = \frac{dF\left(\frac{y-b}{a}\right)}{dy} = \frac{dF\left(\frac{y-b}{a}\right)}{d\left(\frac{y-b}{a}\right)} \cdot \frac{d\left(\frac{y-b}{a}\right)}{dy}$$

$$g(y) = f\left(\frac{y-b}{a}\right) \cdot \frac{1}{a},$$

showing that the shape of the distribution is preserved under the linear transformation.

If the reader has not already done so, please note that it is crucial that the variables, denoted by capital letters, must be clearly distinguished from their values, denoted by small letters; otherwise, confusion, and most likely errors, will abound.

Example 4.3.3

Consider, one more time, the fair wheel where $f(x) = 1$, for $0 \leq x \leq 1$. Now let us spin the wheel, say n times, obtaining the random observations X_1, X_2, \dots, X_n . We let Y denote the largest of these, so that

$$Y = \max\{X_1, X_2, \dots, X_n\}.$$

Y is clearly a nontrivial random variable. Again we seek the probability density function for Y , $g(y)$.

Note that if the maximum of the X 's is at most y , then *each* of the X 's must be at most y . So

$$\begin{aligned} G(y) &= P(Y \leq y) = P(\max\{X_1, X_2, \dots, X_n\} \leq y) \\ &= P(X_1 \leq y \text{ and } X_2 \leq y \text{ and } \dots \text{ and } X_n \leq y). \end{aligned}$$

Since the X 's are independent, it follows that

$$G(y) = [P(X_1 \leq y)] \cdot [P(X_2 \leq y)] \cdots [P(X_n \leq y)].$$

Now, since all the X 's have the same probability density function,

$$G(y) = [P(X \leq y)]^n = [F(y)]^n.$$

It follows that

$$g(y) = \frac{dG(y)}{dy} = n[F(y)]^{n-1} \cdot f(y).$$

Since the X 's all have the same uniform probability density function, in this case $F(y) = y$ so that $g(y) = ny^{n-1}$, for $0 \leq y \leq 1$.

In the general case, we note that the distribution for Y is dependent on $F(y)$. $F(y)$ is easy in this example, but it could prove intractable, as in the case of a normal variable which has no closed form for its distribution function. In fact, the probability distribution of the maximum observation from a random sample of observations from a normal distribution is unknown.

Expectation of a Function of X

In Chapters 1 and 2, we calculated expectations of functions of X using only the probability density function for X . Specifically, we let

$$E[H(X)] = \int_{-\infty}^{\infty} H(x) \cdot f(x) dx, \quad (4.2)$$

where $H(X)$ is some function of the random variable X and $f(x)$ is the probability density function for the random variable X . We took this as a matter of definition. For example, we wrote that

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx.$$

The reader may now well wonder a bit about this definition. The function $H(X)$ is also a random variable. To find its expectation, should not we find its probability density function first and then the expectation of the random variable using that probability density function? It would appear to be a strategy that is certain of success. Amazingly, it turns out not to be necessary, and formula ((4.2)) gives the correct result. Let's see why this is so.

To make matters simple, suppose $Y = H(X)$ and that $H(X)$ is a strictly increasing function of X . (A demonstration similar to that given here can be given for $H(X)$ strictly decreasing.) Then

$$G(y) = P(Y \leq y) = P[H(X) \leq y] = P[X \leq H^{-1}(y)],$$

since $H(X)$ is invertible. This means that

$$G(y) = F[H^{-1}(y)] \text{ or}$$

$$g(y) = \frac{dF[H^{-1}(y)]}{dy} \text{ so that}$$

$$g(y) = f[H^{-1}(y)] \frac{dH^{-1}(y)}{dy}, \text{ or}$$

$$g(y) = f(x) \cdot \frac{dx}{dy},$$

where x is expressed in terms of y . This formula can indeed be used in many of our change of variable formulas but the reader is warned that the function must be strictly increasing or strictly decreasing for this result to work. Now we calculate the expected value:

$$E(Y) = \int_{-\infty}^{\infty} y \cdot g(y) dy = \int_{-\infty}^{\infty} H(x) \cdot f(x) \cdot \frac{dx}{dy} dy = \int_{-\infty}^{\infty} H(x) \cdot f(x) dx,$$

showing that our definition of the expectation of a function of X was sound.

Example 4.3.4

Consider the probability density function

$$f(x) = k \cdot x^2, 0 < x < 2.$$

We calculate $E(X^2)$ in two ways: first, by finding the probability density function for X^2 , and second, without finding that probability density function.

k must be determined first. Since

$$\int_0^2 k \cdot x^2 dx = 1 \text{ it follows that}$$

$$k \cdot \frac{x^3}{3} \Big|_0^2 = k \cdot \frac{8}{3} = 1 \text{ so}$$

$$k = \frac{3}{8}.$$

Now consider the transformation $Y = X^2$.

$$G(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = F(\sqrt{y}),$$

since X takes on only non-negative values.

Now $g(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y})$, so

$$g(y) = \frac{1}{2\sqrt{y}} \cdot \frac{3}{8} \cdot y = \frac{3}{16} \cdot \sqrt{y}, \quad 0 < y < 4.$$

Then

$$E(Y) = \frac{3}{16} \cdot \int_0^4 y^{\frac{3}{2}} dy = \frac{12}{5}.$$

Now we use ((4.2)) directly:

$$E(Y) = \int_0^2 \frac{3}{8} \cdot x^4 dx = \frac{12}{5},$$

obtaining the previous result.

EXERCISES 4.3

1. Suppose that X is uniformly distributed on the interval $(2,5)$. Let $Y = 3X - 2$. Find $g(y)$, the probability density function for Y .
2. Suppose that the probability density function for a random variable X is $f(x) = \lambda \cdot e^{-\lambda x}$, $x \geq 0$, $\lambda > 0$. Let $Y = 3 - X$. Find $g(y)$, the probability distribution function for Y .
3. Let X have a uniform distribution on $(0,1)$. Find the probability density function for $Y = X^2$ and prove that the result is a probability density function.
4. The random variable X has the probability density function $f(x) = 2x$, $0 \leq x \leq 1$.
 - (a) Let $Y = X^2$ and find the probability density function for Y .
 - (b) Now suppose that X has the probability density function $f(x)$. What transformation, $Y = H(X)$, will result in Y having a uniform distribution? (Part (a) of this problem may help in discovering the answer.)
5. Suppose that $X \sim N(\mu, \sigma)$, and let $Y = e^X$.
 - (a) Find the mean and variance of Y .
 - (b) Find the probability density function for Y . The result is called the *lognormal* probability density function since the logarithm of the variable is $N(\mu, \sigma)$.
6. Random variable X has probability density function $f(x) = 4x(1 - x^2)$, $0 \leq x \leq 1$. Find $E(X^2)$ in two ways:
 - (a) Without finding the probability density function of Y .
 - (b) Using the probability density function of Y .
7. If X has a Weibull distribution with parameters α and β , show that the variable $Y = \left(\frac{X}{\alpha}\right)^\beta$ is an exponential variable with mean 1.
8. The *folded normal* distribution is the distribution of $|X|$ where $X \sim N(\mu, \sigma)$.
 - (a) Find the probability density function for a folded normal variable.
 - (b) Find $E(|X|)$.
9. Find the probability density function for $Y = X^3$ where X has an exponential distribution with mean value 1.
10. A circle is drawn by choosing a radius from the uniform distribution on the interval $(0, 1)$. Find the probability density function for the area of the circle.
11. Suppose that X is a uniform random variable on the interval $(-1, 1)$. Find the probability density function for the variable $Y = \sin(X)$.
12. Find the probability density function for $Y = e^X$ where X is uniformly distributed on $[0, 1]$.
13. Random variable X has probability density function

$$f(x) = \frac{1}{(1+x)^2}, \quad x \geq 0.$$

- (a) Find the probability density function for $Y = \sqrt{X}$.
- (b) Show that $P(0 \leq Y \leq b) = 1 - \frac{1}{1+b^2}$, where $b \geq 0$.

- 14.** A random variable X has the probability density function

$$f(x) = \frac{x+1}{2}, \quad -1 \leq x \leq 1.$$

Find $E(X^2)$,

- (a) by first finding the probability density function for $Y = X^2$.
- (b) without using the probability density function for $Y = X^2$.

- 15.** A fluctuating electric current, I , is a uniformly distributed random variable on the interval $[9, 11]$. If this current flows through a 2-ohm resistor, the power is $Y = 2I^2$. Find $E(Y)$ by first finding the probability density function for the power.
- 16.** A random variable X has the probability density function $f(x) = \frac{1}{x^2}$, $x \geq 1$. Find the probability density function for $Y = 1 - \frac{1}{X}$ and prove that your result is a probability density function.
- 17.** Independent observations $X_1, X_2, X_3, \dots, X_n$ are taken from the exponential distribution $f(x) = \lambda e^{-\lambda x}$ where $x > 0$ and $\lambda > 0$. Find the probability density function for $Y = \min(X_1, X_2, X_3, \dots, X_n)$.
- 18.** In triangle ABC , angle ABC is $\pi/2$, $|AB| = 1$ and angle BAC (in radians) is a random variable, uniformly distributed on the interval $[0, \pi/3]$. Find the expected length of side BC .

- 19.** Find the probability density function for $Y = X^2$ if X has the probability density function

$$f(x) = \begin{cases} \frac{x}{4} + \frac{1}{2}, & -2 \leq x \leq 0 \\ -\frac{x}{4} + \frac{1}{2}, & 0 \leq x \leq 2. \end{cases}$$

- 20.** Find the probability density function for $Y = -\ln X$ if X is uniformly distributed on $(0, 1)$.
- 21.** Is $E\left(\frac{1}{X}\right) = \frac{1}{E(X)}$ if $f(x) = e^{-x}$, $x \geq 0$?
- 22.** Find $g(y)$, the probability density function for $Y = X^2$ if X is uniformly distributed on $(-1, 2)$.
- 23.** Computers commonly produce random numbers that are uniform on the interval $(0, 1)$. These can often be used to simulate random selections from other probability distributions in the following way. Suppose we wish a function of the uniform variables to have a given probability distribution function, say $g(y)$. Then, if $G(y)$ is invertible, consider the transformation $Y = G^{-1}(X)$. Then,

$$P(Y \leq y) = P[G^{-1}(X) \leq y] = P[X \leq G(y)] = G(y)$$

since X is a uniform random variable, showing that Y has the required probability density function.

- (a) Find a function, Y , of a uniform $(0, 1)$ random variable so that Y is uniform on (a, b) .
 - (b) Find a function, Y , of a uniform $(0, 1)$ random variable so that Y has an exponential distribution with expected value $1/\lambda$.
- 24.** Show that $E[H(X)] = \int_{-\infty}^{\infty} H(x) \cdot f(x) \cdot \left|\frac{dx}{dy}\right| dy$, where $f(x)$ is the probability density function of X , if $H(X)$ is a strictly decreasing function of X .

25. Show, without using the probability density function of \sqrt{X} , that $E(\sqrt{X}) = \frac{1}{2}\sqrt{\alpha\pi}$ if X is an exponential random variable with mean α . [Hint: The variance of a $N(0, 1)$ variable is 1.]
26. Show that if $f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$, $-\infty < x < \infty$ and if $Y = \frac{1}{X}$, then $g(y) = \frac{1}{\pi} \cdot \frac{1}{1+y^2}$, $-\infty < y < \infty$.
27. An area is lighted by lamps whose length of life is exponential with mean 8000 hours. It is very important that some light be available in the area for 20,000 hours. How many lamps should be installed?
28. Random variable X has a Cauchy distribution, that is

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

Let $Y = \frac{1}{1+X^2}$.

- (a) Show that the probability density function of Y is

$$g(y) = \frac{1}{\pi\sqrt{y(1-y)}}, \quad -\infty < y < \infty.$$

- (b) Show that the distribution function for Y is

$$F(y) = \frac{2}{\pi} \arcsin(\sqrt{y}), \quad 0 < y < 1.$$

- (c) Find $E(Y)$ and $Var(Y)$.

29. $f(x) = 1/3$, $3 \leq x \leq 6$. Find the probability density function for $Y = \frac{1}{2} - \frac{1}{2}X$.

4.4 SUMS OF RANDOM VARIABLES I

Random variables can often be regarded as sums of other random variables. For example, if a coin is tossed and X , the number of heads that appear is recorded (X can only be 0 or 1), and subsequently the coin is tossed again and Y , the number of heads that appear is recorded, then clearly $X + Y$ denotes the total number of heads that appear. So the total number of heads when two coins are tossed can be regarded as a sum of two individual (and, in this case, independent) random variables. Clearly, we expect $X + Y$ to be a binomial random variable with $n = 2$. We can extend this argument to n tosses; the sum is then a binomial random variable. In Chapter 1, we encountered the random variable denoting the sum when two dice are thrown, so we have actually considered sums before.

Now we intend to study the behavior of sums of random variables primarily because the results are interesting and because the consequences have extensive implications to some problems in statistics. In this section, we start with some interesting results and examples.

Example 4.4.1

In the first example above, X is a random variable that takes on the values 1 or 0 with probabilities p and $1 - p$, respectively. Y has a similar distribution, and since X

and Y are independent, the distribution of $X + Y$ can be found by considering all the possibilities:

$$\begin{aligned} P(X + Y = 0) &= P(X = 0) \cdot P(Y = 0) = (1 - p)^2 \\ P(X + Y = 1) &= P(X = 0) \cdot P(Y = 1) + P(X = 1) \cdot P(Y = 0) \\ &= 2p(1 - p) \\ P(X + Y = 2) &= P(X = 1) \cdot P(Y = 1) = p^2. \end{aligned}$$

Recall that the individual variables X and Y are often called *Bernoulli* random variables; their sum, as the earlier calculation shows, is a binomial random variable with $n = 2$. Since the Bernoulli random variable can be regarded as a binomial variable with $n = 1$, we see that in this case the sum of two independent binomial variables is also binomial. This raises the question, "Is the sum of two independent binomial random variables in general always binomial?"

To answer this question, let us proceed with a calculation. Suppose X is binomial (n, p) and Y is binomial (m, p) and let $Z = X + Y$. The event $Z = z$ can arise in several mutually exclusive ways: $X = 0$ and $Y = z$; $X = 1$ and $Y = z - 1$, and so on, until $X = z$ and $Y = 0$. So, assuming also that X and Y are independent,

$$\begin{aligned} P(Z = z) &= \sum_{k=0}^z P(X = k) \cdot P(Y = z - k), \quad \text{or} \\ P(Z = z) &= \sum_{k=0}^z \binom{n}{k} p^k (1 - p)^{n-k} \cdot \binom{m}{z-k} p^{z-k} (1 - p)^{m-(z-k)}. \end{aligned}$$

This can be simplified to

$$P(Z = z) = p^z (1 - p)^{n+m-z} \sum_{k=0}^z \binom{n}{k} \binom{m}{z-k}.$$

But we recognize $\sum_{k=0}^z \binom{n}{k} \binom{m}{z-k}$ from the hypergeometric distribution as $\binom{n+m}{z}$. So

$$P(Z = z) = \binom{n+m}{z} p^z (1 - p)^{n+m-z}, \quad z = 0, 1, 2, \dots, n+m,$$

a binomial distribution with parameters $n + m$ and p . This establishes the fact that sums of independent binomial random variables are also binomial.

We note here, since $E(X) = np$ and $Var(X) = np(1 - p)$ (with similar results for Y), that $E(Z) = (n + m)p$ and $Var(Z) = (n + m)p(1 - p)$. We summarize these results as follows:

$$E(X + Y) = E(X) + E(Y) \quad \text{and}$$

$$Var(X + Y) = Var(X) + Var(Y),$$

since X and Y are independent. As we will see later, the assumption of independence is a crucial one here.

We note here that the sum of independent binomial random variables is again binomial. Occasionally, random variables are *reproductive* in the sense that their sums are distributed in the same way as the summands, but this is not always the case. In fact, it is not the case with binomials if the probability of success at any trial for the random variable X differs from the probability of success at any trial for the random variable Y . We turn now to

an example where the probability distribution of the sum is not of the same form as the summands.

Example 4.4.2

Suppose X and Y are each discrete uniform random variables, that is,

$$P(X = x) = \frac{1}{n}, \quad x = 0, 1, 2, \dots, n$$

with an identical distribution for Y . What happens if we add two randomly chosen observations? We investigate the probability distribution of the sum, $Z = X + Y$, assuming that X and Y are independent.

The special case $n = 4$ may be instructive. Then if we wanted to find, for example, $P(Z = 6)$ we could work out all the possibilities:

$$\begin{aligned} P(Z = 6) &= P(X = 2) \cdot P(Y = 4) + P(X = 3) \cdot P(Y = 3) + P(X = 4) \cdot P(Y = 2) \\ &= \left(\frac{1}{4}\right) \cdot \left(\frac{1}{4}\right) + \left(\frac{1}{4}\right) \cdot \left(\frac{1}{4}\right) + \left(\frac{1}{4}\right) \cdot \left(\frac{1}{4}\right) = \frac{3}{16}. \end{aligned}$$

Proceeding in a similar way for other values of z , we find

$$P(Z = z) = \begin{cases} \frac{1}{16}, & z = 2 \\ \frac{2}{16}, & z = 3 \\ \frac{3}{16}, & z = 4 \\ \frac{4}{16}, & z = 5 \\ \frac{3}{16}, & z = 6 \\ \frac{2}{16}, & z = 7 \\ \frac{1}{16}, & z = 8 \end{cases}$$

This result can also be summarized as

$$P(Z = z) = \begin{cases} \frac{z-1}{16}, & z = 2, 3, 4, 5 \\ \frac{9-z}{16}, & z = 6, 7, 8. \end{cases}$$

A graph of this is shown in Figure 4.1.

The sum is certainly not uniform. It is not clear what might happen when we increase the number of summands. We might conjecture that, as we add more independent uniform random variables, the sums become normal. This is in fact the case, but we need to develop some techniques before we can consider that circumstance and verify the normality. We will begin to do that in the next section.

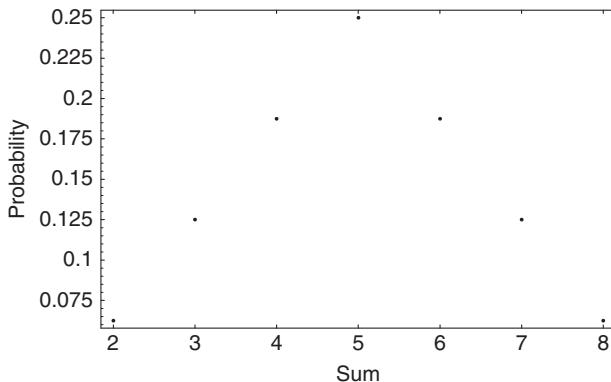


Figure 4.1 Sum of two independent discrete uniform variables, $n = 4$.

EXERCISES 4.4

1. Verify all the probabilities in Example 4.4.1
2. Random variable X has the probability density function given in the following table

x	1	2	3	4
$f(x)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$

 (a) Find the probability density function for two independent observations X_1 and X_2 .
 (b) Verify that $E[X_1 + X_2] = E[X_1] + E[X_2]$ and that $Var[X_1 + X_2] = Var[X_1] + Var[X_2]$.
3. Show that the sum of two independent Poisson variables with parameters λ_x and λ_y , respectively, has a Poisson distribution with parameter $\lambda_x + \lambda_y$.
4. Let X and Y be independent geometric random variables so that $P(X = x) = (1 - p)^{x-1} p$, $x = 1, 2, 3, \dots$ with a similar distribution for Y . Show, if X and Y are independent, that $X + Y$ has a negative binomial distribution.
5. Find the probability distribution for $X + Y + Z$ where X , Y , and Z each have a discrete uniform distribution on the integers 1, 2, 3, and 4.
6. Let X denote a Bernoulli random variable, that is, $P(X = 1) = p$ and $P(X = 0) = 1 - p$ and let Y be a binomial random variable with parameters n and p . Show that $X + Y$ is binomial with parameters $n + 1$ and p .
7. A coin, loaded so as to come up heads with probability $2/3$, is thrown until a head appears, then a fair coin is thrown until a head appears.
 - (a) Find the probability distribution for Z , the total number of tosses necessary.
 - (b) Find the expected value for Z from the probability distribution for Z .
8. Phone calls come into an office according to a Poisson distribution with four calls expected in an interval of 2 minutes. The calls are answered according to a binomial process with $p = 1/2$. Find the probability that exactly three calls are answered in a two-minute period.
9. Generalize problem 6: Consider Poisson events, in a given interval of time, with parameter λ , which are recorded according to a binomial process with parameter p . Show that the number of events recorded in the interval of time is Poisson with parameter λp .

4.5 GENERATING FUNCTIONS

At this point we have found the probability distribution functions of sums of random variables by working out the probabilities for each possible value of the sum. This technique of course cannot be carried out when the summands are continuous or when the number of summands is large.

We consider now another technique that will make some complex problems involving sums of either discrete or continuous random variables tractable. We start with the discrete case in this section.

Example 4.5.1

Consider throwing a fair die once, and this function:

$$G(t) = \frac{1}{6}(t + t^2 + t^3 + t^4 + t^5 + t^6).$$

If X is the random variable denoting the face showing on the die, then we observe that the coefficient of t^k in $G(t)$ is the probability that X equals k , $P(X = k)$. For example, $P(X = 3)$ is the coefficient of t^3 which is $1/6$. Since $G(t)$ has this property, it is called a *probability generating function*.

If X is a random variable taking values on the nonnegative integers, then any function of the form

$$\sum_{k=0}^{\infty} t^k \cdot P(X = k)$$

is called a probability generating function.

Note that in $G(t)$ we could easily load the die by altering the coefficients of the powers of t to reflect the different probabilities with which the faces appear. For example the function

$$H(t) = \frac{1}{10}t + \frac{1}{5}t^2 + \frac{1}{10}t^3 + \frac{1}{5}t^4 + \frac{1}{5}t^5 + \frac{1}{5}t^6$$

generates probabilities on a die loaded so that faces numbered 1 and 3 appear with probability $1/10$ while each of the other faces appears with probability $1/5$.

Probability generating functions are of great importance in probability; they provide neat summaries of probability distributions and have other remarkable properties as we will see.

Continuing our example, if we square $G(t)$ we see that

$$G^2(t) = \frac{1}{36}(t^2 + 2t^3 + 3t^4 + 4t^5 + 5t^6 + 6t^7 + 5t^8 + 4t^9 + 3t^{10} + 2t^{11} + t^{12}).$$

$G^2(t)$ is also a probability generating function—its coefficients are the probabilities of the sums when two dice are thrown.

In general, $G^n(t)$ generates probabilities

$$P(X_1 + X_2 + X_3 + \dots + X_n = k),$$

where X_i is the face showing on die i , $i = 1, 2, \dots, n$.

We may use this fact to find, for example, the probability that when four fair dice are thrown a sum of 17 is obtained. We would find this to be a very difficult problem if we were constrained to write out all the possibilities for which the sum is 17. Since $G(t)$ can be written as

$$G(t) = \frac{t(1-t^6)}{6(1-t)},$$

it follows that

$$G^4(t) = \frac{t^4(1-t^6)^4}{6^4(1-t)^4}.$$

This reduces our problem to finding the coefficient of t^{13} in $(1-t^6)^4(1-t)^{-4}$. Expanding this by the binomial theorem and ignoring the division by 6^4 for the moment, we see that the coefficient we seek is

$$\left[1 - \binom{4}{1} t^6 + \binom{4}{2} t^{12} + \dots\right] \left[1 + \binom{-4}{1} (-t) + \binom{-4}{2} (-t)^2 + \dots\right].$$

So the coefficient of t^{13} is

$$\begin{aligned} & \binom{-4}{13} (-1)^{13} - \binom{4}{1} \binom{-4}{7} (-1)^7 + \binom{4}{2} \binom{-4}{1} (-1) \\ &= \binom{16}{3} - 4 \binom{10}{3} + 6 \binom{4}{3} = 104. \end{aligned}$$

Therefore the probability we seek is $104/6^4 = 13/162$.

This process is certainly an improvement on that of counting all the possibilities, a technique that clearly becomes impossible when the number of dice is large.

A computer algebra system allows us to find

$$G^4(t) = \left[\frac{1}{6}(t + t^2 + t^3 + t^4 + t^5 + t^6) \right]^4$$

giving directly the following table of probabilities for the sums on four fair dice:

Sum	Probability	Sum	Probability
4	1/1296	15	35/324
5	1/324	16	125/1296
6	5/648	17	13/162
7	5/324	18	5/81
8	35/1296	19	7/162
9	7/162	20	35/1296
10	5/81	21	5/324
11	13/162	22	5/648
12	125/1296	23	1/324
13	35/324	24	1/1296
14	73/648		

While the computer can give us high powers of $G(t)$, it can also give us great insight into the problem as well. Consider a graph of the coefficients of $G(t)$ as shown in Figure 4.2.

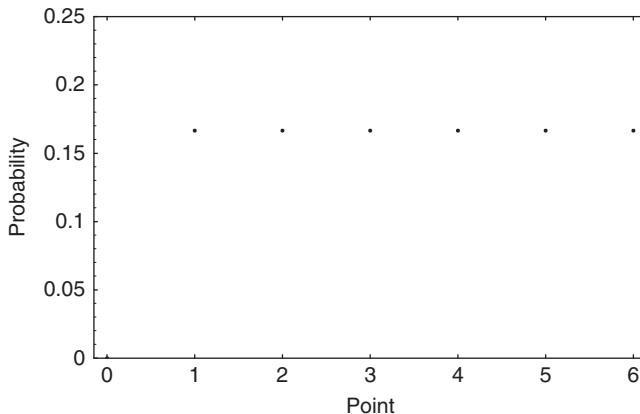


Figure 4.2 Probabilities for one fair die.

Now consider $G^2(t)$ whose coefficients are shown in Figure 4.3.

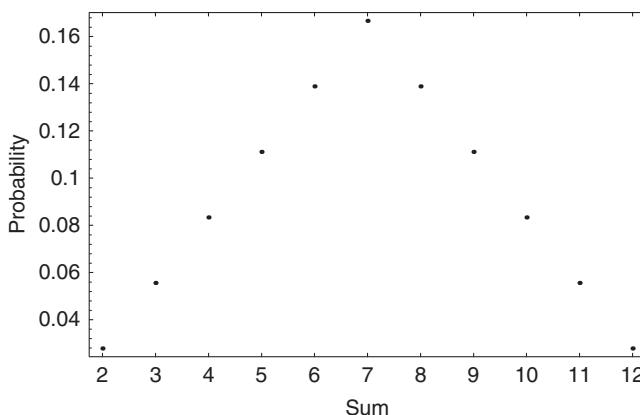


Figure 4.3 Probabilities for sums on two fair dice.

A graph of the coefficients in $G^4(t)$ is shown in Figure 4.4.

Finally, Figure 4.5 shows the probabilities for sums on 12 fair dice.

This is probably enough to convince the reader that normality, once again, is involved. The probability distribution for the sums on 12 fair dice is in fact remarkably close to a normal curve. We find, for example, that

$$P(36 \leq \text{Sum} \leq 48) = 0.724753, \text{ exactly},$$

while the normal curve gives 0.728101, a very good approximation.

Before the normality can be explained analytically we must consider some more characteristics of probability generating functions. We will consider these in the next section.

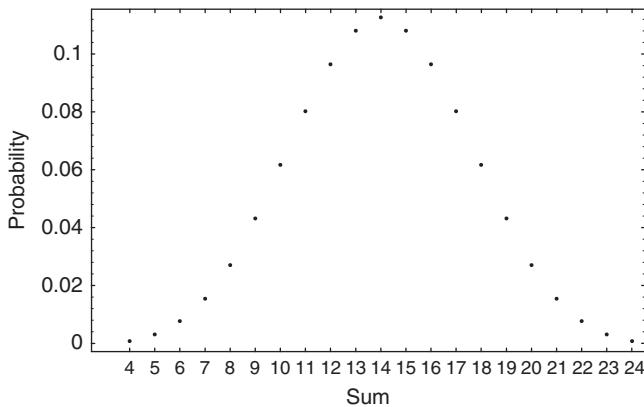


Figure 4.4 Probabilities for sums on 4 fair dice.

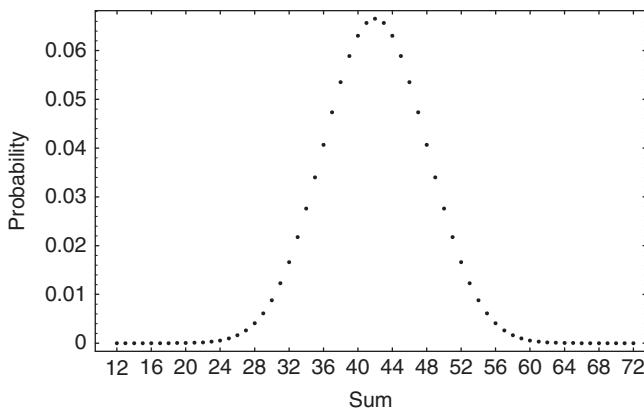


Figure 4.5 Probabilities for sums on 12 fair dice.

EXERCISES 4.5

1. Find all the probabilities when three fair dice are thrown.
2. (a) Find the generating function when a fair coin is tossed.
(b) Use part (a) to find the probability of no heads when a fair coin is tossed five times.
3. Show that the function $(1+t)^n$ generates the binomial coefficients, $\binom{n}{x}$, $x = 0, 1, 2, \dots, n$.
4. What sequence is generated by $(1-4t)^{-\frac{1}{2}}$?
5. Consider the set $\{a, b, c\}$. What does the function

$$(1+at)(1+bt)(1+ct) \text{ generate?}$$

6. Find a function that generates the sequence $0^2, 1^2, 2^2, 3^2, \dots$

7. Find a function that generates the sequence $\frac{1}{1 \cdot 2}, \frac{1}{2 \cdot 3}, \frac{1}{3 \cdot 4}, \frac{1}{4 \cdot 5}, \dots$
8. A die is loaded so that the probability a face appears is proportional to the face. If the die is thrown five times, find the probability the sum obtained is 17.
9. Suppose that the probability generating function for the random variable X is $P_X(t)$. Find an expression for the probability generating function for
 - (a) $X + k$, where k is a constant.
 - (b) $k \cdot X$, where k is a constant.
10. Verify in Example 4.3.1 that if X is the sum on 12 fair dice then $P(36 \leq X \leq 48) = 0.724753$.
11. A fair die and a die loaded are thrown so that the probability a face appears is proportional to the face are thrown. Find the probability distribution for the sums appearing on the uppermost faces.

4.6 SOME PROPERTIES OF GENERATING FUNCTIONS

Let us first explain why the products of generating functions generate probabilities associated with sums of random variables.

Suppose $A(t)$ and $B(t)$ are probability generating functions for random variables X and Y , respectively, where X and Y are defined on the set of nonnegative integers or some subset of them and let

$$\begin{aligned} A(t) &= a_0 + a_1 t + a_2 t^2 + \dots \text{ and} \\ B(t) &= b_0 + b_1 t + b_2 t^2 + \dots. \end{aligned}$$

Then

$$A(t) \cdot B(t) = a_0 b_0 + (a_0 b_1 + a_1 b_0)t + (a_0 b_2 + a_1 b_1 + a_2 b_0)t^2 + \dots,$$

so the coefficient of t^k is

$$\sum_{i=0}^k a_i b_{k-i} = \sum_{i=0}^k P(X = i) \cdot P(Y = k - i) = P(X + Y = k).$$

This explains why we could find powers of $G(t)$ in Example 4.5.1 and generate probabilities associated with throwing more than one die.

Since $E(t^X) = \sum_{i=0}^{\infty} t^i P(X = i)$, it follows that a probability generating function, say $P_X(t)$, can be regarded as an expectation:

$$P_X(t) = E(t^X) = \sum_{i=0}^{\infty} t^i P(X = i)$$

for a random variable X .

For example, $G(t) = \sum_{i=1}^6 t^i \cdot P(X = i) = \sum_{i=1}^6 t^i \cdot \frac{1}{6}$.
Note that if $t = 1$, then

$$P_X(1) = \sum_{i=0}^{\infty} P(X = i) = 1.$$

Also

$$\frac{dP_X(t)}{dt} = \frac{d}{dt} \sum_{i=0}^{\infty} t^i \cdot P(X = i) = \sum_{i=0}^{\infty} \frac{d(t^i)}{dt} P(X = i)$$

from which it follows that

$$P'_X(t) = \sum_{i=0}^{\infty} i \cdot t^{i-1} \cdot P(X = i).$$

So

$$P'_X(1) = \sum_{i=0}^{\infty} i \cdot P(X = i) = E(X).$$

In addition,

$$P''_X(t) = \sum_{i=0}^{\infty} i \cdot (i-1) \cdot t^{i-2} \cdot P(X = i).$$

$$\text{So } P''_X(1) = E[X \cdot (X - 1)],$$

with similar results holding for higher order derivatives.

Since $E(X^2) = E[X \cdot (X - 1)] + E(X)$, it follows that

$$\text{Var}(X) = E[X \cdot (X - 1)] + E(X) - [E(X)]^2 \quad \text{or}$$

$$\text{Var}(X) = P''_X(1) + P'_X(1) - [P'_X(1)]^2.$$

As an example, consider throwing a single die and let

$$G(t) = \frac{1}{6}(t + t^2 + t^3 + t^4 + t^5 + t^6). \text{ Then}$$

$$G'(t) = \frac{1}{6}(1 + 2t + 3t^2 + 4t^3 + 5t^4 + 6t^5). \text{ So}$$

$$G'(1) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} \text{ giving } E(X) \text{ and}$$

$$G''(t) = \frac{1}{6}(2 + 6t + 12t^2 + 20t^3 + 30t^4). \text{ So}$$

$$G''(1) = \frac{1}{6}(2 + 6 + 12 + 20 + 30) = \frac{70}{6}.$$

It follows that

$$\text{Var}(X) = P''_X(1) + P'_X(1) - [P'_X(1)]^2$$

$$\text{Var}(X) = \frac{70}{6} + \frac{7}{2} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

4.7 PROBABILITY GENERATING FUNCTIONS FOR SOME SPECIFIC PROBABILITY DISTRIBUTIONS

Probability generating functions for the binomial and geometric random variables are particularly useful so we derive their probability generating functions in this section.

Binomial Distribution

For the binomial distribution,

$$P_X(t) = E(t^X) = \sum_{x=0}^n t^x \binom{n}{x} p^x q^{n-x}.$$

This sum can be written as

$$P_X(t) = \sum_{x=0}^n \binom{n}{x} (tp)^x q^{n-x}.$$

Now the binomial theorem shows that $P_X(t) = (q + pt)^n$. It is easy to check that

$$P'_X(t) = np(q + pt)^{n-1} \quad \text{so that, since } p + q = 1,$$

$$P'_X(1) = np \quad \text{as expected.}$$

$$\text{Also, } P''_X(t) = n(n-1)p^2(q + pt)^{n-2},$$

$$\text{so } E(X^2) = P''_X(1) = n(n-1)p^2,$$

from which it follows that $\text{Var}(X) = n(n-1)p^2 + np - (np)^2 = np - np^2 = npq$.

Now we show, using probability generating functions, that the sum of independent binomial random variables is binomial. Suppose that X and Y are independent binomial variables with the probability generating functions $P_X(t) = (q + pt)^{n_x}$ and $P_Y(t) = (q + pt)^{n_y}$, respectively. If $Z = X + Y$, then the probability generating function for Z is

$$P_Z(t) = P_X(t) \cdot P_Y(t)$$

$$P_Z(t) = (q + pt)^{n_x} \cdot (q + pt)^{n_y} = (q + pt)^{n_x+n_y}.$$

Assuming that the probability generating functions are unique, that is, assuming that a probability generating function can arise from one and only one probability distribution function, this shows that Z is binomial with parameters $n_x + n_y$ and p .

The derivation above, done in one line, shows the power of the probability generating function technique; the reader can compare this with the derivation in Example 4.4.1.

It should be pointed out, however, as it may have occurred to the reader, that the fact that sums of binomials, with the same probabilities of success at any trial, is binomial is hardly surprising. If we have a series of n_x binomial trials and we record X , the number of successes, and follow this by a series of n_y trials recording Y successes, it is obvious,

since the trials are independent, that we have $X + Y$ successes in $n_x + n_y$ trials. The fact that we paused somewhere in the experiment to record the number of successes so far and then continued has nothing to do with the entire series of trials.

This raises the question of pausing in the series and changing the probability of success at that point. Now the resulting distribution is not at all obvious. Such trials are, confusingly perhaps, called *Poisson's trials*. This problem can be considered using generating functions.

Poisson's Trials

As an example, suppose we toss a fair coin 20 times followed by 10 tosses of a coin loaded so that the probability of a head is $1/3$. What is the probability of exactly 15 heads resulting?

Using probability generating functions, we see that we need the coefficient of t^{15} in $\left(\frac{1}{2} + \frac{1}{2}t\right)^{20} \cdot \left(\frac{2}{3} + \frac{1}{3}t\right)^{10}$. This is

$$\begin{aligned} & \sum_{x=5}^{15} \binom{20}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{20-x} \binom{10}{15-x} \left(\frac{2}{3}\right)^{x-5} \left(\frac{1}{3}\right)^{15-x} \\ & = \frac{156031933}{1289945088} = 0.12096. \end{aligned}$$

A computer algebra system will give this result as well as all the other coefficients in $\left(\frac{1}{2} + \frac{1}{2}t\right)^{20} \cdot \left(\frac{2}{3} + \frac{1}{3}t\right)^{10}$ directly, so it is of immense value in problems of this sort.

A graph of these coefficients is remarkably normal as shown in Figure 4.6.

If X is the number of heads in the first series and Y is the number of heads in the second series, it is still true that $E(X + Y) = E(X) + E(Y)$. In this example, $E(X + Y) = 20 \cdot \frac{1}{2} + 10 \cdot \frac{1}{3} = \frac{40}{3}$ and, since the tosses are independent,

$$\begin{aligned} Var(X + Y) &= Var(X) + Var(Y) \\ &= 20 \cdot \frac{1}{2} \cdot \frac{1}{2} + 10 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{65}{9}. \end{aligned}$$

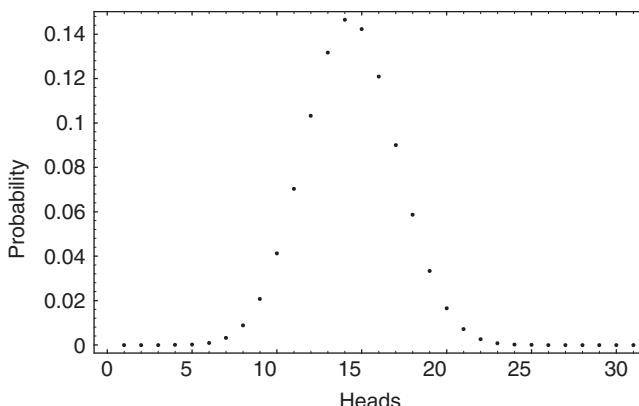


Figure 4.6 Probabilities for the total number of heads when a fair coin is tossed 20 times followed by 10 tosses of a loaded coin with $p = 1/3$.

We would expect a normal curve with these parameters to fit the distribution of $X + Y$ fairly well.

Example 4.7.1

A series of n binomial trials with probability p is conducted, and is followed by a series of m trials with probability x/n , where x is the number of successes in the first series of trials. Let Y denote the number of successes in the second series of trials. Then

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \\ &= n \cdot p + m \cdot \frac{E(X)}{n} = p \cdot (n + m). \end{aligned}$$

The variance of the sum is another matter, however, since the second series of trials is very clearly dependent on the first series and, because of this, $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$. General calculations of this sort will be considered in Chapter 5 when we discuss sample spaces with two or more random variables defined on them.

For now consider, as an example of this, a first series comprised of five trials with probability of success $1/2$, followed by a series of three trials. What is the probability of exactly four successes in the entire experiment? We find that

$$\begin{aligned} P(X + Y = 4) &= \sum_{x=1}^4 \binom{5}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} \binom{3}{4-x} \left(\frac{x}{5}\right)^{4-x} \left(1 - \frac{x}{5}\right)^{x-1}. \\ P(X + Y = 4) &= \frac{73}{400}. \end{aligned}$$

Again a computer algebra system is of great use in doing the calculations.

Geometric Distribution

The waiting time, X , for the first occurrence of a binomial random variable with parameter p has the probability distribution function

$$P(X = x) = q^{x-1}p, \quad x = 1, 2, \dots,$$

so $P_X(t) = \sum_{x=1}^{\infty} t^x \cdot q^{x-1}p = \frac{p}{q} \sum_{x=1}^{\infty} (tq)^x = \frac{pt}{1-tq}$, provided that $|qt| < 1$. Since $0 < q < 1$, and we are only interested when $t = 1$, the restriction is not important for us.

Using $P_X(t)$ we find that $P'_X(1) = E(X) = \frac{1}{p}$, and that $\text{Var}(X) = \frac{q}{p^2}$.

The variable X here denotes the waiting time for the first binomial success. When we wait for the r th success, say, the negative binomial distribution arises. Since a negative binomial variable is the sum of geometric variables, it follows, if X is now the waiting time for the r th binomial success, that

$$P_X(t) = \left(\frac{pt}{1-tq} \right)^r = \frac{p^r t^r}{(1-tq)^r}.$$

$P_X(t)$ can be used to show that the negative binomial distribution has mean r/p and variance rq/p^2 . This is left as an exercise for the reader.

Collecting Premiums in Cereal Boxes

Your favorite breakfast cereal, in an effort to urge you to buy more cereal, encloses a toy or a premium in each box. How many boxes must you buy in order to collect all the premiums? This problem is also often called the *coupon collector's problem* in the literature on probability theory. Of course we cannot be *certain* to collect all the premiums, given finite resources, but we could think about the average, or expected number, of boxes to be purchased.

To make matters specific, suppose there are 6 premiums. The first box gives us a premium we did not have before. The probability the next box will not duplicate the premium we already have is 5/6. This waiting time for the next premium not already collected is a geometric random variable, with probability 5/6. The expected waiting time for the second premium is then 1/(5/6). Now we have two premiums, so the probability the next box contains a new premium is 4/6. This is again a geometric variable and our waiting time for collecting the third premium is 1/(4/6). This process continues. Since the expectation of a sum is the sum of the expectations of the summands and if we let X denote the total number of boxes purchased in order to secure all the premiums, we conclude that

$$E(X) = 1 + \frac{1}{\frac{5}{6}} + \frac{1}{\frac{4}{6}} + \frac{1}{\frac{3}{6}} + \frac{1}{\frac{2}{6}} + \frac{1}{\frac{1}{6}}$$

$$E(X) = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1}$$

$$E(X) = 1 + 1.2 + 1.5 + 2 + 3 + 6 = 14.7 \text{ boxes.}$$

Clearly the cereal company knows what it is doing! An exercise will ask the reader to show that the variance of X is 38.99, so unlucky cereal eaters could be in for buying many more boxes than the expectation would indicate.

This is an example of a series of trials, analogous to Poisson's trials, in which the probabilities vary. Since the total number of trials, X , can be regarded as a sum of geometric variables (plus 1 for the first box), and since the probability generating function for a geometric variable is $\frac{qt}{1-pt}$, the probability generating function of X is

$$P_X(t) = \frac{\frac{5}{6}t}{1 - \frac{1}{6}t} \cdot \frac{\frac{4}{6}t}{1 - \frac{2}{6}t} \cdot \frac{\frac{3}{6}t}{1 - \frac{3}{6}t} \cdot \frac{\frac{2}{6}t}{1 - \frac{4}{6}t} \cdot \frac{\frac{1}{6}t}{1 - \frac{5}{6}t}.$$

This can be written as

$$P_X(t) = \frac{5!t^5}{(6-t)(6-2t)(6-3t)(6-4t)(6-5t)}.$$

The first few terms in a power series expansion of $P_X(t)$ are as follows:

$$P_X(t) = \frac{5t^5}{324} + \frac{25t^6}{648} + \frac{175t^7}{2916} + \frac{875t^8}{11664} + \frac{11585t^9}{139968} + \frac{875t^{10}}{10368} + \frac{616825t^{11}}{7558272} + \dots$$

Probabilities can be found from $P_X(t)$, but not at all easily without a computer algebra system. The series above shows that the probability it takes 9 boxes in total to collect all 6 premiums is $875/11664 = 0.075$.

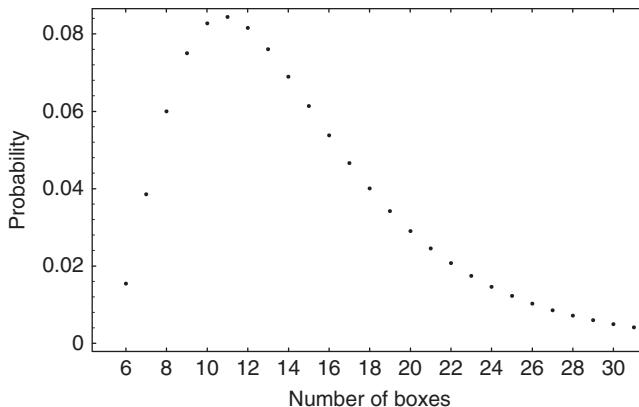


Figure 4.7 Probabilities for the cereal box problem.

A graph of the probability distribution function is shown in Figure 4.7. The probabilities shown there are the probabilities it takes n boxes to collect all 6 premiums. We will return to this problem and some of its variants in Chapter 7.

EXERCISES 4.7

1. Use the generating function for the binomial random variable with $p = 2/3$ to verify μ and σ^2 .
2. In the cereal box problem, find σ^2 using a generating function.
3. (a) Find the probability generating function for a Poisson random variable with parameter λ .
 (b) Use the generating function in part (a) to find the mean and variance of a Poisson random variable.
4. Use probability generating functions to show that the sum of independent Poisson variables, with parameters λ_x and λ_y , respectively, has a Poisson distribution with parameter $\lambda_x + \lambda_y$.
5. A discrete random variable, X , has probability distribution function $f(x) = k/2^x$, $x = 0, 1, 2, 3, 4$.
 - (a) Find k .
 - (b) Find $P_X(t)$, the probability generating function.
 - (c) Use $P_X(t)$ to find the mean and variance of X .
6. Use the probability generating function to find the mean and variance of a negative binomial variable with parameters r and p .
7. A fair coin is tossed eight times followed by 12 tosses of a coin loaded so as to come up heads with probability $3/4$. What is the probability that
 - (a) exactly 10 heads occur?
 - (b) at least 10 heads occur?
8. Use the probability generating function of a Bernoulli random variable to show that the sum of independent Bernoulli variables is a binomial random variable.

9. A random variable has probability distribution function

$$f(x) = \frac{1}{e \cdot x!}, \quad x = 0, 1, 2, 3, \dots$$

- (a) Find the probability generating function for X , $P_X(t)$.
- (b) Use $P_X(t)$ to find the mean and variance of X .
- 10. Suppose a series of 10 binomial trials with probability 1/2 of success is conducted, giving x successes. These trials are followed by 8 binomial trials with probability $x/10$ of success. Find the probability of exactly 6 successes in the entire series.
- 11. Verify that the variance of X is 38.99 in the cereal box problem.
- 12. Suppose X and Y are independent geometric variables with parameters p_1 and p_2 respectively.
 - (a) Find the probability generating function for $X + Y$.
 - (b) Use the probability generating function to find $P(X + Y = k)$ and then verify your result by calculating the probability directly.

4.8 MOMENT GENERATING FUNCTIONS

Another generating function that is commonly used in probability theory is the *moment generating function*. For a random variable X , this function generates the *moments*, $E(X^k)$, for the probability distribution of X . If $k = 1$, the moment becomes $E(X)$ or the mean of the distribution. If $k = 2$, then the moment is $E(X^2)$ which we use in calculating the variance.

The word *moment* has a physical connotation. If we think of the probability distribution as being a very thin piece of material of area 1, then $E(X)$ is the same as the center of gravity of the material and $E(X^2)$ is used in calculating the moment of inertia. Hence the name *moment* for these quantities which we use to describe probability distributions.

The extent to which we are successful in using the moments to describe probability distributions may be judged from certain considerations. If we were to specify $E(X)$ as a value for a probability distribution this would certainly constrain the set of random variables X under consideration, but we could still be considering an infinite set of variables. Were we to specify $E(X^2)$ as well, this would narrow the set of possible random variables. A value for $E(X^3)$ further narrows the set. For the examples we will consider, we ask the reader to accept the fact that, were *all* the moments specified, X would be determined uniquely.

Now let us see how this fact can be used. We begin with a definition of the moment generating function.

Definition The *moment generating function* of a random variable X is

$$M[X; t] = E[e^{tX}],$$

providing the expectation exists.

It follows that

$$M[X; t] = \sum_x e^{tx} \cdot P(X = x), \quad \text{if } X \text{ is discrete}$$

$$M[X; t] = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) dx, \quad \text{if } X \text{ is continuous}$$

provided the sum or integral exists and where $f(x)$ is the probability density function of X .

First we show that $M[X; t]$ does in fact generate moments. Consider the continuous case so that

$$M[X; t] = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \, dx.$$

Expanding e^{tx} in a power series, we have

$$M[X; t] = \int_{-\infty}^{\infty} \left(1 + t \cdot x + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots \right) \cdot f(x) \, dx.$$

Use the fact that the integral of a sum is the sum of the integrals and factor out all the powers of t to find that

$$\begin{aligned} M[X; t] &= \int_{-\infty}^{\infty} f(x) \, dx + t \int_{-\infty}^{\infty} x \cdot f(x) \, dx \\ &\quad + \frac{t^2}{2!} \cdot \int_{-\infty}^{\infty} x^2 f(x) \, dx + \frac{t^3}{3!} \cdot \int_{-\infty}^{\infty} x^3 f(x) \, dx + \dots \\ \text{so } M[X; t] &= 1 + t \cdot E(X) + \frac{t^2}{2!} \cdot E(X^2) + \frac{t^3}{3!} \cdot E(X^3) + \dots \end{aligned}$$

providing that the series converges.

$M[X; t]$ generates moments in the sense that the coefficient of $\frac{t^k}{k!}$ is $E(X^k)$.

We used the derivatives of the probability generating function, $P_X(t)$, to calculate $E(X)$, $E[X(X - 1)]$, $E[X(X - 1)(X - 2)]$, ..., quantities that are often called *factorial moments*. The moments defined above could be calculated from them. We did on several occasions to find the variance.

The derivatives of $M[X; t]$ also have some significance. Since

$$M[X; t] = 1 + t \cdot E(X) + \frac{t^2}{2!} \cdot E(X^2) + \frac{t^3}{3!} \cdot E(X^3) + \dots,$$

$$M'[X; t] = \frac{dM[X; t]}{dt} = E(X) + t \cdot E(X^2) + \frac{t^2}{2!} \cdot E(X^3) + \dots \text{ and}$$

$$M''[X; t] = \frac{d^2M[X; t]}{dt^2} = E(X^2) + t \cdot E(X^3) + \frac{t^2}{2!} \cdot E(X^4) + \dots$$

so it is evident that $M'[X; 0] = E(X)$ and $M''[X; 0] = E(X^2)$. There are then two methods for calculating moments—either a series expansion or by the derivatives of $M[X; t]$. There are in practice very few examples where each method is feasible; generally one method works well while the other method presents difficulties. We turn now to some examples.

Example 4.8.1

For the uniform random variable, $f(x) = 1$, for $0 \leq x \leq 1$. The moment generating function is then

$$M[X; t] = \int_0^1 1 \cdot e^{tx} \, dx = \frac{1}{t} e^{tx} \Big|_0^1 = \frac{1}{t} (e^t - 1).$$

In this instance it is easy to express $M[X; t]$ in a power series. Using the power series for e^t we find that

$$M[X; t] = 1 + \frac{t}{2!} + \frac{t^2}{3!} + \frac{t^3}{4!} + \dots$$

so $E(X^k) = \frac{1}{k+1}$.

However this is a fact that is much more easily found directly:

$$E(X^k) = \int_0^1 x^k \, dx = \frac{1}{k+1}.$$

The moment generating function does little here but provide a very difficult way in which to find the moments. This is almost always the case; moment generating functions are rarely used to generate moments; it is almost always easier to proceed by definition. What then is the use of the moment generating function? The answer to this question is that we use it almost exclusively to establish the distributions of functions of random variables and the distributions of sums of random variables, basing our conclusions on the fact that moment generating functions are unique, that is, only one distribution has a given moment generating function. We will return to this point later.

For now, continuing with the example, we found that

$$M[X; t] = \frac{1}{t}(e^t - 1).$$

If we differentiate this,

$$M'[X; t] = \frac{te^t - e^t + 1}{t^2}.$$

As $t \rightarrow 0$ we use L'Hospital's rule to find that

$$M'[X; t] \rightarrow \frac{1}{2},$$

so the process yields the correct result. This is without doubt the most difficult way in which to establish the fact that the mean of a uniform random variable on the interval $(0, 1)$ is $1/2$!

Clearly, we have other purposes in mind; the fact is that the moment generating function is an extremely powerful tool. Facts can be established easily using it that are very difficult to establish in other ways.

We continue with further examples since the generating functions themselves are of importance.

Example 4.8.2

Consider the exponential distribution $f(x) = e^{-x}$, $x \geq 0$. We calculate the moment generating function:

$$M[X; t] = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \, dx = \int_0^{\infty} e^{tx} \cdot e^{-x} \, dx.$$

This can be simplified to

$$M[X; t] = \int_0^{\infty} e^{-(1-t)x} \, dx = \frac{-1}{1-t} e^{-(1-t)x} \Big|_0^{\infty} = \frac{1}{1-t} \text{ if } t < 1.$$

Again the power series is easy to find:

$$M[X; t] = 1 + t + t^2 + t^3 + \dots$$

establishing the fact that $E[X^k] = k!$, for k a positive integer. This is a nice way to establish the fact that

$$\int_0^\infty x^k e^{-x} dx = k!$$

which arose earlier when the gamma distribution was considered.

The reader may also want to show that the moment generating function for $f(x) = \lambda e^{-\lambda x}$, $x > 0$, is

$$M(X; t) = \frac{\lambda}{\lambda - t}.$$

Example 4.8.3

The moment generating function for a normal random variable is by far our most important result, as will be seen later. We use here a standard normal distribution:

$$M[X; t] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \cdot e^{-\frac{x^2}{2}} dx.$$

The simplification of this integral takes some manipulation. Consider the exponent

$$tx - \frac{x^2}{2} = -\frac{1}{2}(x^2 - 2tx + t^2) + \frac{t^2}{2} = -\frac{1}{2}(x - t)^2 + \frac{t^2}{2}$$

by completing the square. This means that the generating function can be written as

$$M[X; t] = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-t)^2}{2}} dx.$$

The integral is 1 since it represents the area beneath a normal curve with mean t and variance 1.

It follows that

$$M[X; t] = e^{\frac{t^2}{2}}.$$

We can also find a power series for this generating function as

$$M[X; t] = 1 + \frac{t^2}{2} + \frac{1}{2!} \left(\frac{t^2}{2} \right)^2 + \frac{1}{3!} \left(\frac{t^2}{2} \right)^3 + \dots$$

It follows that

$$E(X^k) = 0 \text{ if } k \text{ is odd and}$$

$$E(X^{2k}) = \frac{(2k)!}{k! 2^k} \text{ for } k = 1, 2, 3, \dots$$

Moment generating functions for other commonly occurring distributions will be established in the exercises.

EXERCISES 4.8

1. Verify that the moment generating function for $f(x) = 3e^{-3x}$ is $\frac{3}{3-t}$.
2. Use the moment generating function in Exercise 1 to find μ and σ^2 .
3. Show that if $P_X(t)$ is the probability generating function for a random variable X , then $M[X; t] = P_X(e^t)$.
4. (a) Find the moment generating function for a binomial random variable with parameters n and p .
 (b) Take the case $n = 2$ and expand the moment generating function to find $E(X^2)$ from this expansion.
 (c) Use the moment generating function to find the mean and variance of a binomial random variable.
5. (a) Find the moment generating function for a Poisson random variable with parameter λ .
 (b) Use the moment generating function to find the mean and variance of a Poisson random variable.
6. Find the moment generating function for an exponential random variable with mean λ .
7. A random variable has moment generating function $M[X; t] = \frac{1}{6}e^{-t} + \frac{1}{2}e^{-2t} + \frac{1}{3}e^t$.
 - (a) Find the mean and variance of X .
 - (b) Find the first five terms in the power series expansion of $M[X; t]$.
8. A random variable, X , has probability distribution function $f(x) = k \cdot \left(\frac{1}{2}\right)^x$, $x = 1, 2, 3, \dots$
 - (a) Find k .
 - (b) Find the moment generating function and show the first five terms in its power series expansion.
 - (c) Find the mean and variance of X from the moment generating function in two ways.
9. (a) Find the moment generating function for a gamma distribution.
 (b) Use the moment generating function to find the mean and variance for the gamma distribution.
10. (a) Find the moment generating function for a χ_n^2 random variable.
 (b) Use the moment generating function to find the mean and variance of a χ_n^2 random variable.
11. A random variable X has the probability density function $f(x) = \begin{cases} \frac{1}{3} & -2 < x < -1 \\ \frac{1}{k} & 1 < x < 4. \end{cases}$
 Find the moment generating function for X .
12. Find $E[X^4]$ for a random variable whose moment generating function is $e^{\frac{t^2}{2}}$.
13. Random variable X has moment generating function $M[X; t] = e^{10t+2t^2}$.
 - (a) Find $P(7 \leq X \leq 12)$.
 - (b) Find the probability density function for $Y = 3X$.
14. A random variable X has the probability density function $f(x) = \begin{cases} \frac{1}{2}e^{-x} & x > 0 \\ \frac{1}{2}e^x & x < 0. \end{cases}$

- (a) Show that $M[X; t] = (1 - t^2)^{-1}$.
 (b) Find the mean and variance of X .
15. Suppose that X is a uniformly distributed random variable on $[2, 3]$.
 (a) Find the moment generating function for X .
 (b) Expand $M[X; t]$ in an infinite series and from this series find μ_x and σ_x^2 .
16. Find the moment generating function for X if $f(x) = 2x$, $0 < x < 1$. Then use the moment generating function to find μ_x and σ_x^2 .
17. The moment generating function for a random variable X is $\left(\frac{2}{3} + \frac{1}{3}e^t\right)^5$.
 (a) Find the mean and variance of X .
 (b) What is the probability distribution for X ?
18. The moment generating function for a random variable X is e^{t^2} . Find the mean and variance of X .

4.9 PROPERTIES OF MOMENT GENERATING FUNCTIONS

A primary use of the moment generating function is in determining the distributions of functions of random variables. It happens that the moment generating function for linear functions of X is easily related to the moment generating function for X .

Theorem:

- (a) $M[cX; t] = M[X; ct]$.
 (b) $M[X + c; t] = e^{ct}M[X; t]$, where c is a constant.

Proof

- (a) $M[cX; t] = E(e^{(cX)t}) = E(e^{X(ct)}) = M[X; ct]$.
 (b) $M[X + c; t] = E[e^{(X+c)t}] = E[e^{Xt} \cdot e^{ct}] = e^{ct}E[e^{Xt}] = e^{ct}M[X; t]$.

So multiplying the variable by a constant simply multiplies t by the constant in the generating function; adding a constant multiplies the generating function by e^{ct} .

Example 4.9.1

We use the earlier theorem to find the moment generating function for a $N(\mu, \sigma)$ random variable from the generating function for $N(0, 1)$ random variable. Let $Z = \frac{X-\mu}{\sigma}$. Since

$$M[Z; t] = e^{\frac{t^2}{2}}, \text{ it follows that}$$

$$M\left[\frac{X-\mu}{\sigma}; t\right] = M\left[X - \mu; \frac{t}{\sigma}\right] = e^{-\frac{\mu t}{\sigma}} M\left[X; \frac{t}{\sigma}\right],$$

from which it follows that

$$M\left[X; \frac{t}{\sigma}\right] = e^{\frac{t^2}{2} + \frac{\mu t}{\sigma}}. \text{ We conclude from this that}$$

$$M[X; t] = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

Therefore if a random variable X has, for example, $M[X; t] = e^{\frac{3t}{4} + \frac{t^2}{3}}$ then X is normal with mean $3/4$ and variance $2/3$. A remarkable fact, and one we use frequently here is that, if X and Y are independent, that

$$M[X + Y; t] = M[X; t] \cdot M[Y; t].$$

To indicate why this is true, we start with $M[X + Y; t] = E[e^{(X+Y)t}] = E[e^{tX} \cdot e^{tY}]$, which is the expectation of the product of two functions. If we can show that the expectation of the product is the product of the expectations, then

$$M[X + Y; t] = E[e^{tX} \cdot e^{tY}] = E[e^{tX}] \cdot E[e^{tY}] = M[X; t] \cdot M[Y; t].$$

As a partial explanation of the fact that the expectation of a product of independent random variables is the product of the expectations, consider X and Y as discrete independent random variables. Then

$$E(X \cdot Y) = \sum_x \sum_y x \cdot y \cdot P(X = x \text{ and } Y = y).$$

But if X and Y are independent, then

$$P(X = x \text{ and } Y = y) = P(X = x) \cdot P(Y = y)$$

and so

$$\begin{aligned} E[X \cdot Y] &= \sum_x \sum_y x \cdot y \cdot P(X = x \text{ and } Y = y) \\ &= \sum_x \sum_y x \cdot y \cdot P(X = x) \cdot P(Y = y) \\ &= \sum_x x \cdot P(X = x) \cdot \sum_y y \cdot P(Y = y) = E(X) \cdot E(Y). \end{aligned}$$

So it is plausible that the expectation of the product of independent random variables is the product of their expectations and we accept the fact that

$$M[X + Y; t] = M[X; t] \cdot M[Y; t]$$

if X and Y are independent.

We will return to this point in Chapter 5 when we consider bivariate probability distributions. For now we will make use of the result to establish some surprising results.

4.10 SUMS OF RANDOM VARIABLES – II

We have used the facts that $E(X + Y) = E(X) + E(Y)$ and, if X and Y are independent, that $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$, but these facts do not establish the distribution of $X + Y$. We

now turn to determining the distribution of the sums of two or more independent random variables; our solution here will show the power and usefulness of the moment generating function. We will return to this subject in Chapter 5 where we can demonstrate another procedure for finding the probability distribution of sums of random variables. Here we use the fact that the moment generating function for a sum of independently distributed random variables is the product of the individual generating functions.

Example 4.10.1 *Sums of Normal Random Variables*

It is probably not surprising to find that sums of independent normal variables are also normal. The proof of this is now easy: If X and Y are independent normal variables,

$$M[X + Y; t] = M[X; t] \cdot M[Y; t].$$

The exponent in the product on the right above is

$$\mu_x t + \frac{t^2 \sigma_x^2}{2} + \mu_y t + \frac{t^2 \sigma_y^2}{2}.$$

This can be rearranged as

$$(\mu_x + \mu_y)t + \frac{t^2(\sigma_x^2 + \sigma_y^2)}{2}$$

showing that $X + Y \sim N[\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2]$. Note that the mean and variance of the sum can be established in other ways. The argument above establishes the *normality* which otherwise would be very difficult to show.

However the big surprise is that sums of non-normal variables also become normal. We will explain this fully in Section 4.11, but the reader may note that this may explain the frequency with which we have seen the normal distribution up to this point. For the moment, we continue with another example.

Example 4.10.2 *Sums of Exponential Random Variables*

We begin with a decidedly non-normal random variable, namely an exponential variable where we take the mean to be 1. So

$$f(x) = e^{-x}, \quad x \geq 0.$$

We know that

$$M[X; t] = (1 - t)^{-1}.$$

It follows that the moment generating function of the sum of two independent exponential random variables is

$$M[X + Y; t] = (1 - t)^{-2}.$$

This, however, is the moment generating function of $f(x) = xe^{-x}$, $x \geq 0$. The graph of this distribution is shown in Figure 4.8.

Now consider the sum of three independent exponential random variables. The moment generating function is $M[X + Y + Z; t] = (1 - t)^{-3}$. A computer algebra system, or otherwise, shows that this is the moment generating function for $f(x) = \frac{x^2}{2}e^{-x}$, $x \geq 0$. Figure 4.9 shows a graph of this distribution.

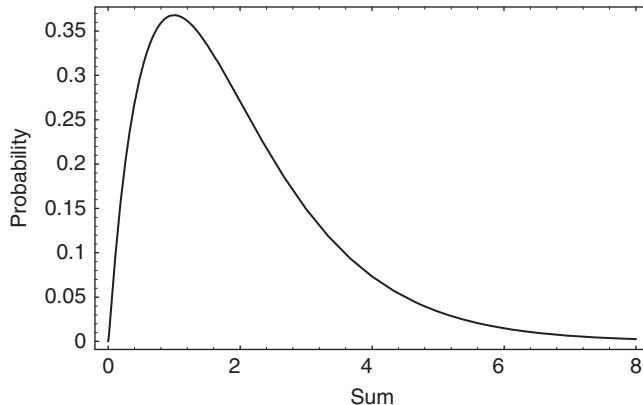


Figure 4.8 Sum of two independent exponential random variables.

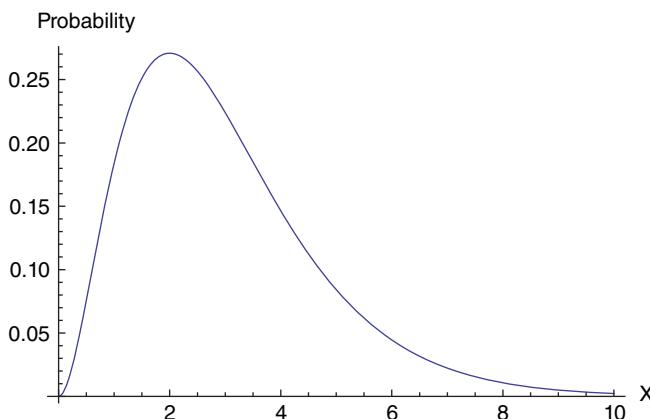


Figure 4.9 Sum of three independent exponential variables.

We now strongly suspect the occurrence of normality if we were to add more variables. We know that

$$M[(X_1 + X_2 + X_3 + \dots + X_n); t] = (1 - t)^{-n}.$$

This is the moment generating function for the gamma distribution, $f(x) = \frac{1}{\Gamma[n]} x^{n-1} e^{-x}$, $x \geq 0$. Since the mean and variance of each of the X_i 's above is 1, we can consider $X = \sum_{i=1}^n X_i$ and $Z = \frac{X-n}{\sqrt{n}}$. Then,

$$M[Z; t] = M\left[\frac{X-n}{\sqrt{n}}; t\right] = M\left[X-n; \frac{t}{\sqrt{n}}\right] = e^{-t\sqrt{n}} \left(1 - \frac{t}{\sqrt{n}}\right)^{-n}.$$

The behavior of this is most easily found using a computer algebra system. We expand $M[Z; t]$ and then let $n \rightarrow \infty$. We find that

$$M[Z; t] \rightarrow e^{\frac{t^2}{2}},$$

showing that Z approaches the standard normal distribution. Some details of this calculation are given in Appendix A. This establishes the fact that the sums of independent exponential variables approach a normal distribution.

In the beginning of this chapter we indicated that the sums showing on n dice—the sums of independent discrete uniform random variables—became normal, although we lacked the techniques for proving this at that time. The proof of this will not be shown here, but we note that the process followed in Example 4.10.2 will work in this case. Now we know that the distribution of sums of independent exponential variables also approaches the normal distribution. The fact that the distribution of the sum of widely different summands approaching the normal distribution is perhaps one of the most surprising facts in mathematics. The fact that normality should occur for a wide range of random variables is investigated in the next section.

EXERCISES 4.10

1. For the uniform random variable $f(x) = 1, 0 \leq x \leq 1$,
 - (a) Find the moment generating function.
 - (b) Find the mean and variance of the sum of 3 uniformly distributed random variables.
2. Expand the moment generating function in exercise 1 and verify the mean and variance.
3. The moment generating function for a random variable X is $M[X; t] = \frac{5}{5-t}$.
 - (a) Find the mean and variance of X .
 - (b) Identify the probability distribution for X .
4. Random variable X has $M[X; t] = \left(\frac{2}{3} + \frac{1}{3}e^t\right)^6$.
 - (a) Find the mean and variance of X .
 - (b) Identify the probability distribution for X .
5. Find the mean and variance for the random variable whose moment generating function is $M(Z; t) = (1 - 2t)^{-5}$.
6. Find the moment generating function for the exponential random variable whose probability density function is $f(x) = 2e^{-2x}, x \geq 0$.
7. Suppose $X \sim N(36, \sqrt{10})$ and $Y \sim N(15, \sqrt{6})$. If X and Y are independent, find $P(X + Y \geq 43)$.
8. A random variable X has probability density function $f(x) = xe^{-x}, x \geq 0$.
 - (a) Find the moment generating function, $M[X; t]$.
 - (b) Use $M[X; t]$ to find μ_x and σ_x^2 .
 - (c) Find a formula for $E(X^k)$.
9. Find the variance of a random variable whose moment generating function is $M[X; t] = (1 - t)^{-1}$.
10. Explain why the function $2 + \frac{1}{1-t}$ cannot be the moment generating function for any random variable.
11. What is the probability distribution for a random variable whose moment generating function is

$$M[X; t] = e^{-\lambda(1-e^t+t)} \quad ?$$

- 12.** Identify the random variable whose moment generating function is

$$M[X; t] = \left(\frac{1}{2}\right)^{16} \cdot e^{-4t} \cdot \left(1 + e^{\frac{t}{2}}\right)^{16}.$$

- 13.** Show that the sum of two independent binomial random variables with parameters n and p , and m and p , respectively, is binomial with parameters $n+m$ and p .
- 14.** Show that the sum of independent Poisson random variables with parameters λ_x and λ_y , respectively, is Poisson with parameter $\lambda_x + \lambda_y$.
- 15.** Show that the sum of independent χ^2 random variables is a χ^2 random variable.
- 16.** Show that a Poisson variable with parameter λ becomes normal as $\lambda \rightarrow \infty$. [Hint: Find the limit of $M\left[\frac{X-\lambda}{\sqrt{\lambda}}; t\right]$.]
- 17.** (a) If X is uniformly distributed on $[0, 1]$ show that $M[X; t] = \frac{1}{t}(e^t - 1)$.
 (b) Suppose that X_1 and X_2 are independent observations from the uniform distribution in part (a). Find $M[X_1 + X_2; t]$.
 (c) Let Y be a random variable with the probability density function

$$g(y) = \begin{cases} y, & 0 < y < 1 \\ 2 - y, & 1 < y < 2. \end{cases}$$

Find the moment generating function for Y .

- (d) What can be concluded from parts (b) and (c) above?
- 18.** Suppose that X_i , $i = 1, 2, 3, \dots, n$ is each exponentially distributed with means $1/\lambda_i$ respectively. Let $S = X_1 + X_2 + \dots + X_n$. Find the moment generating function for S .
- 19.** Suppose that X and Y are independent random variables with probability density functions

$$f(x) = 1, \quad 0 \leq x \leq 1 \quad \text{and}$$

$$g(y) = 1, \quad -1 \leq y \leq 0.$$

Find $M[X + Y; t]$.

- 20.** If $M[X; t] = e^{-\lambda(1-e^t)}$, what is $M[15 - 3X; t]$?
- 21.** The price asked for a security is normally distributed with a mean of \$50 and standard deviation of \$5. Buyers are willing to pay an amount that is also normally distributed with a mean of \$45 and a standard deviation of \$2.50. What is the probability a transaction will take place?
- 22.** A rod is made up of five sections. A study of the individual sections shows that the end sections have mean lengths of 1.001 in. and the three middle sections have mean lengths of 1.999 in. each. The standard deviation of the length of each section is 0.004 in. If random assembly is employed,
- (a) what will be the average length of the assembled rods?
 (b) what will be the standard deviation of the assembled lengths?
 (c) what is the probability the assembled rod will have length in excess of 8.002 in.?
- 23.** Show that the binomial random variable with parameters n and p becomes normal as $n \rightarrow \infty$.

24. Two independent observations, X and Y , are selected from a probability distribution with

$$f(x) = 2x, \quad 0 \leq x \leq 1.$$

(a) Find the moment generating function for the sum, $Z = X + Y$.

(b) Find $E(Z^4)$.

25. Let S denote the sum of r independent exponential random variables, each with expectation $1/\alpha$. Show that $2\alpha S$ has a χ_{2r}^2 distribution.

26. Show that the moment generating function for the gamma distribution

$$f(x) = \frac{1}{\Gamma[n]} x^{n-1} e^{-x}, \quad x \geq 0$$

is $(1-t)^{-n}$.

4.11 THE CENTRAL LIMIT THEOREM

We have had numerous examples of sums which approach normality as the number of summands increases. We now want to consider means, which are multiples of sums, of random variables and consider the limiting distribution of such averages.

Theorem 1: If \bar{X} denotes the mean of n observations of a random variable X with mean μ and variance σ^2 , then the limiting distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is $N(0, 1)$ provided X has a moment generating function.

The theorem indicates that the probability distribution of the random variable $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ approaches the $N(0, 1)$ probability distribution. The result is known as the *central limit theorem*. Actually there is a class of theorems known as central limit theorems in probability, but since this is the only one we will consider, we will refer to it uniquely and call it the central limit theorem. We now indicate a proof.

Since we presume that X has a moment generating function, let this moment generating function be

$$M[X; t] = 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \dots$$

where, for convenience, μ_k denotes $E(X^k)$. Now

$$M[\bar{X}; t] = M\left[\frac{\sum_{i=1}^n X_i}{n}; t\right] = \left(M\left[X; \frac{t}{n}\right]\right)^n \quad \text{so}$$

$$\log[M[\bar{X}; t]] = n \log M\left[X; \frac{t}{n}\right].$$

Using the series expansion $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$, where $|x| < 1$, we find

$$\begin{aligned}\log M\left[X; \frac{t}{n}\right] &= \left(\mu_1 \frac{t}{n} + \mu_2 \frac{t^2}{2!n^2} + \mu_3 \frac{t^3}{3!n^3} + \dots\right) \\ &\quad - \frac{1}{2} \left(\mu_1 \frac{t}{n} + \mu_2 \frac{t^2}{2!n^2} + \mu_3 \frac{t^3}{3!n^3} + \dots\right)^2 \\ &\quad + \frac{1}{3} \left(\mu_1 \frac{t}{n} + \mu_2 \frac{t^2}{2!n^2} + \mu_3 \frac{t^3}{3!n^3} + \dots\right)^3 - \dots.\end{aligned}$$

So $n \cdot \log M\left[X; \frac{t}{n}\right] = \log[M[\bar{X}; t]]$ simplifies to

$$\log[M[\bar{X}; t]] = \mu_1 t + \sigma^2 \frac{t^2}{2n} \text{ plus terms that approach 0 as } n \rightarrow \infty.$$

This shows that $M[\bar{X}; t] \rightarrow e^{\mu t + \frac{\sigma^2}{n} \frac{t^2}{2}}$, the moment generating function for a normal curve with mean μ and variance $\frac{\sigma^2}{n}$.

This explains many of the normal-like graphs we have encountered previously. If the variables are sums, or means, of variables with moment generating functions (as all of ours have been), we expect normality as the number of summands increases. This is exactly what has happened. This phenomenon was encountered in Section 4.10 where we examined sums of independent exponential random variables and found that these approach a normal probability distribution.

This also explains why the normal curve can be regarded as something that is “normal” in the sense that it is usual or expected. It can, in fact, be generated from almost any probability distribution by taking sums or forming averages.

We will show some very important statistical applications of this result in the remaining sections in this chapter.

Example 4.11.1

We noticed in Chapter 2 that the graphs of the binomial distributions we considered became normal-like as the number of trials increased; we also promised a full explanation of the fact that binomial curves with large values for n can be approximated by normal curves. We will do this by using the technique described earlier, namely by finding the limiting behavior of the moment generating function.

Let X be a binomial random variable with parameters n and p . Then the moment generating function of X is

$$M[X; t] = (q + pe^t)^n.$$

As usual we let $Z = \frac{X-\mu}{\sigma}$ so that

$$M[Z; t] = e^{-\frac{\mu t}{\sigma}} \cdot \left(q + pe^{\frac{t}{\sigma}}\right)^n.$$

So

$$\begin{aligned}
 \log(M[Z; t]) &= -\frac{\mu t}{\sigma} + n \log \left[q + p \left(1 + \frac{t}{\sigma} + \frac{t^2}{2!\sigma^2} + \frac{t^3}{3!\sigma^3} + \dots \right) \right] \\
 &= -\frac{\mu t}{\sigma} + n \log \left[1 + p \cdot \frac{t}{\sigma} + p \cdot \frac{t^2}{2!\sigma^2} + p \cdot \frac{t^3}{3!\sigma^3} + \dots \right] \\
 &= -\frac{\mu t}{\sigma} + n \left(p \cdot \frac{t}{\sigma} + p \cdot \frac{t^2}{2!\sigma^2} + p \cdot \frac{t^3}{3!\sigma^3} + \dots \right) \\
 &\quad - \frac{n}{2} \left(p \cdot \frac{t}{\sigma} + p \cdot \frac{t^2}{2!\sigma^2} + p \cdot \frac{t^3}{3!\sigma^3} + \dots \right)^2 + \dots
 \end{aligned}$$

Now, using the facts that $\mu = n \cdot p$ and that $\sigma^2 = n \cdot p \cdot q$, we find that

$$\log(M[Z; t]) = \frac{t^2}{2} + \text{ terms that approach 0 as } n \rightarrow \infty.$$

It follows that $M[Z; t]$ approaches the moment generating function of the standard normal random variable.

This justifies our use of the normal distribution in approximating the binomial distribution for large values of n , although computer algebra systems allow us to compute binomial probabilities exactly for values of n that occur in most practical cases.

The central limit theorem has wide application to the statistical analysis of data. We will show some of the statistical applications of this result in the remaining sections of this chapter.

EXERCISES 4.11

- Show that, if X is normal with mean μ and variance σ^2 , then \bar{X} is normal with mean μ and variance $\frac{\sigma^2}{n}$, where \bar{X} is the mean of n of the X 's.
- Use the central limit theorem to approximate the probability that the sum on 12 fair dice is 38 and then compare the approximation to the exact value.
- Approximate the probability that the sum of 8 observations taken from an exponential distribution with the mean 2 exceeds 5 by using the central limit theorem.
- Light fixtures in a warehouse contain bulbs whose life lengths are exponential with a mean of 720 hours. When a light burns out, it is immediately replaced with a new bulb.
 - What is the probability that three bulbs last at least 2000 hours?
 - If we want the probability that the bulbs on hand will last at least 3500 hours with the probability of 0.95, how many bulbs should be stocked?
- Suppose that X is a random variable with an unknown mean, μ , but its variance is known to be 100. How many observations of X must be taken so that the probability \bar{X} is within 2 units of μ is 0.99?
- The components in a system are known to have $R(1000 \text{ hours}) = 0.91$, where R denotes the reliability function.
 - Approximate the reliability of a system of 100 such components if at least 70 of the components must function at least 1000 hours.

- (b) How many components must be installed in the system if at least 90 components must last at least 1000 hours with probability 0.98?
7. Articles are shipped in lots of 1000 items. It is known that the probability an item is defective is 0.04. Presume that the production process follows the assumptions of the binomial distribution.
- (a) Approximate the probability that in 100 lots, the average number of defectives is less than 39.5.
- (b) Now suppose we would like the probability that the average number of defectives in 100 lots is less than 39.5 to be 0.10. What should the size of each lot be?
8. Traffic accidents at an intersection follow a Poisson distribution with 40 accidents expected per year. Approximate the probability of at most 55 accidents in a given year at that intersection.
9. An elevator can carry a maximum of 1575 lb. What is the probability that 10 people will overload the elevator if their weights are random selections from $N(150, 10)$?
10. A graduating class has 200 graduates. Assume each graduate invites two guests who attend, independently, with the probability of 0.8. How many seats for guests should be provided at commencement if they desire to be 99% confident of seating everyone?
11. A machine turns out precision bolts whose lengths may be regarded as a normal random variable with mean 6 and variance 0.0036. To check on whether or not the machine is in control, 36 bolts are randomly selected from each day's production. The machine is considered to be under control if the mean of these lengths falls between 5.970 and 6.015. What is the probability a sample will fail to meet this criterion, even though the machine is under control?
12. At a local discount store, service times at the checkout counter are observed to be normally distributed with mean 3.5 minutes and variance 1.44 min^2 .
- (a) Find the probability a customer takes more than 5 minutes to check out.
- (b) A customer has been checking out for 3 minutes. What is the probability it will take at least 5 minutes for the entire process?
- (c) What is the probability that the next 6 customers check out in a total of 20 minutes or less?
13. One hundred bolts are packed in a box. The weight of a bolt has mean 1 oz. and standard deviation 0.1 oz. Approximate the probability a box weighs more than 102 oz.
14. A candy maker produces mints that have a label weight of 20.4 g, but the actual distribution of the weights has $\mu = 21.37 \text{ g}$ and $\sigma^2 = 0.16 \text{ g}^2$. Let \bar{X} be the mean weight of a sample of 36 units. Find $P(21.21 \leq \bar{X} \leq 21.45)$.
15. Let X be a normal random variable with mean 1 and variance 16.
- (a) What is the probability an observation is within 2 units of the mean?
- (b) What is the probability that the mean of 4 observations is within 2 units of the mean?
16. Civil engineers believe that W , the weight (in units of 1000 lb) the span of a bridge can withstand without structural damage resulting is 78.5. Suppose that the weight (again in units of 1000 lb) is a random variable with a mean of 3 and a standard deviation of 0.3. How many cars can be allowed on the bridge span for the probability that structural damage will not occur to be 0.99?

17. In Example 2.3.1.2 we remarked that a deviation of more than 0.11 in the average result when a fair die is thrown 1000 times is highly unlikely. Show that this is true.

4.12 WEAK LAW OF LARGE NUMBERS

If we have a production process that is producing a defective item with probability p , we have an intuitive notion that we can discover the value of p , at least within a given range, if we observe the production process long enough. If we have a distribution with unknown mean, μ , we have a similar belief, namely that we can determine μ , again with a given accuracy, if we take a large enough sample and compute the mean of the sample. It is reasonable to believe that the mean of this sample ought to be close to μ . These ideas are actually correct and we examine mathematical demonstrations here of them.

We might even refer to these results as a Law of Averages, although the literature of probability generally refers to these results as the *Weak Law of Large Numbers*.

We consider the second problem first. Suppose that $X_1, X_2, X_3, \dots, X_n$ is a random sample from some distribution with finite mean and variance, say $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for $i = 1, 2, \dots, n$. By the central limit theorem $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/n$.

Now we can apply Tchebycheff's inequality to find that

$$P\left[|\bar{X} - \mu| \leq k \cdot \frac{\sigma}{\sqrt{n}}\right] \geq 1 - \frac{1}{k^2} \text{ for some } k > 0.$$

Now let $\epsilon = k \cdot \frac{\sigma}{\sqrt{n}}$ so that $k = \epsilon \cdot \frac{\sqrt{n}}{\sigma}$ then the inequality above becomes

$$P[|\bar{X} - \mu| \leq \epsilon] \geq 1 - \frac{\sigma^2}{n \cdot \epsilon^2}.$$

As $n \rightarrow \infty$, $P[|\bar{X} - \mu| \leq \epsilon] \rightarrow 1$ even when ϵ is arbitrarily small. So the probability that \bar{X} and μ are arbitrarily close approaches 1. This is a verification of our conjecture that a sample mean can be made arbitrarily close to the population mean as the sample size increases.

For the first conjecture, let p_s denote the sample proportion of defective items chosen from a production process that is producing defective items with probability p . We know that $E(p_s) = p$ and that $Var(p_s) = \frac{p \cdot q}{n}$ where n is the sample size. Again applying Tchebycheff's inequality, we find that

$$P\left[|p_s - p| \leq k \cdot \sqrt{\frac{p \cdot q}{n}}\right] \geq 1 - \frac{1}{k^2} \text{ for some } k > 0.$$

Now if $\epsilon = k \cdot \sqrt{\frac{p \cdot q}{n}}$, the inequality becomes

$$P[|p_s - p| \leq \epsilon] \geq 1 - \frac{p \cdot q}{n \cdot \epsilon^2}.$$

So p_s and p can be made arbitrarily close as n becomes large with probability approaching 1.

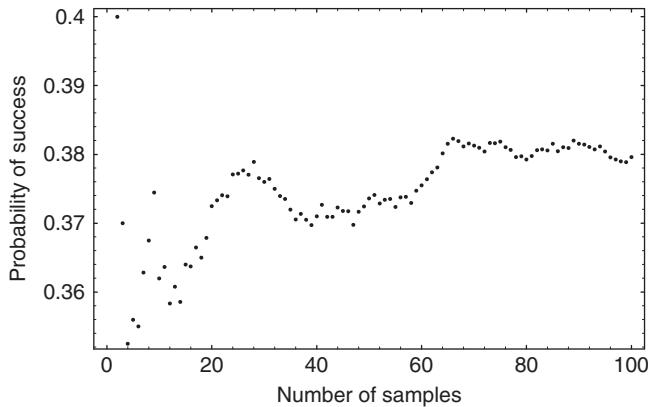


Figure 4.10 Simulation illustrating the weak law of large numbers.

A computer simulation provides some concrete evidence of the above statement. Figure 4.10 shows the result of 100 samples of size 100 each, drawn from a binomial population with $p = 0.38$.

The horizontal axis shows the number of samples while the vertical axis displays the cumulative ratio of successes to the total number of trials. While the initial values exhibit fairly large variation, the later ratios are very close to 0.38 as we expect.

The convergence in statements such as $P[|\bar{X} - \mu| \leq \epsilon] \rightarrow 1$, indicating that a sequence of means approaches a population value with probability 1, is referred to as *convergence in probability*. It differs from the convergence usually encountered in calculus where that convergence is normally pointwise.

4.13 SAMPLING DISTRIBUTION OF THE SAMPLE VARIANCE

The remainder of this chapter will be devoted to data analysis, so we now turn to some statistical applications of the theory presented to this point. In particular we want to investigate hypothesis tests, some confidence intervals and the analysis of data arising in many practical situations. We will also examine the theory of least squares as it applies to fitting a linear function to data.

The central limit theorem indicates that the probability distribution of sample means drawn from a variety of populations is approximately normal even for samples of moderate size. The probability distributions of other quantities calculated from samples (usually referred to as *statistics*) do not have such simple distributions and, in addition, are often seriously affected by the type of probability distribution from which the samples come.

In this section we determine the probability distribution of the sample variance. Other statistics will become of importance to us, and we will consider their probability distributions when they arise. It is worth considering the sample variance by itself first.

First we define the *sample variance* for a sample x_1, x_2, \dots, x_n as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ where } \bar{x} \text{ is the mean of the sample.}$$

The formula may also be written as

$$s^2 = \frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n-1)}.$$

Clearly there is some relationship between s^2 and σ^2 . The divisor of $n-1$ may be puzzling; but, as we will presently show, this is chosen so that $E(s^2) = \sigma^2$. Since $E(s^2) = \sigma^2$, s^2 is called an *unbiased estimator* of σ^2 . If a divisor of n had been chosen, the expected value of the sample variances thus calculated would not be the population value σ^2 .

Now let us consider a specific example. Consider all the possible samples of size 3, chosen without replacement, from the discrete uniform distribution on the set of integers $\{1, 2, \dots, 20\}$. We calculate the *sample variance* for each sample. Each sample variance is calculated using

$$s^2 = \frac{1}{2} \sum_{i=1}^3 (x_i - \bar{x})^2 \text{ where } \bar{x} = \frac{x_1 + x_2 + x_3}{3}.$$

The probability distribution of s^2 in part is as follows. Permutations of the samples have been ignored.

s^2	1	7/3	4	13/3	19/3	...	301/3	307/3	313/3	109	343/3
1140 · Prob.	18	34	16	32	30	...	2	4	2	2	2

The complete distribution is easy to work out with the aid of a computer algebra system. There are 83 possible values for s^2 . A graph of the distribution of these values is shown in Figure 4.11.

The graph indicates that large values of the variance are unusual. The graph also indicates that the probability distribution of s^2 is probably not normal. However the sample size is quite small, so we can't draw any definite conclusions here.

We do see that the probability distribution shown in Figure 4.12 strongly resembles that suggested by Figure 4.11.

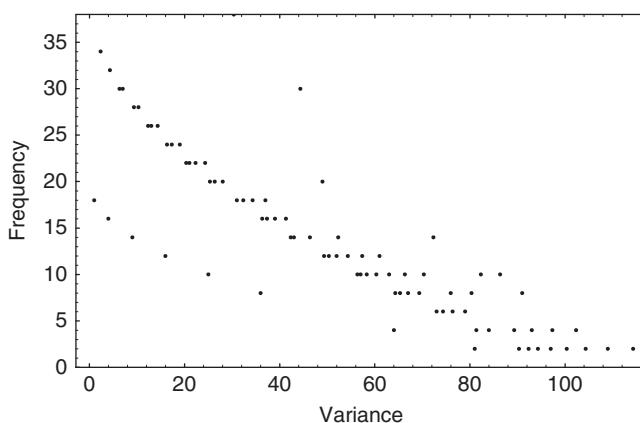


Figure 4.11 Sampling distribution for sample variances.

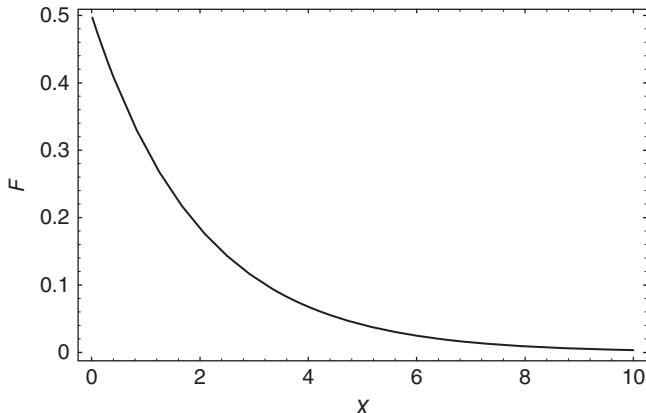


Figure 4.12 A probability distribution suggested by Figure 4.11.

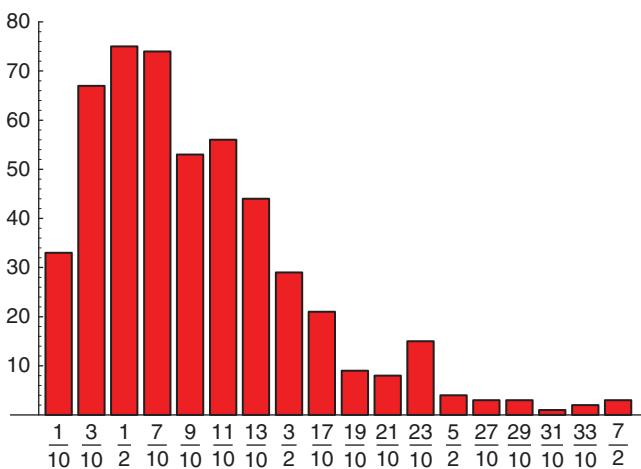


Figure 4.13 Distribution of sample variances chosen from a standard normal distribution.

As another example, 500 samples of size 5 each were selected from a standard normal distribution. The sample variance for each of these samples was computed; the results are shown in the histogram in Figure 4.13. We now see a distribution with a long tail that resembles the probability distribution shown in Figure 4.14. Figure 4.14 is in fact a graph of the probability distribution of a chi-squared distribution with 4 degrees of freedom. We now show that this is the probability distribution of a function of the sample variance s^2 .

The sample variance s^2 is a very complex random variable, since it involves \bar{x} , which, in addition to the sample values themselves, varies from sample to sample. To narrow our focus, suppose now that the sample comes from a normal distribution $N(\mu, \sigma^2)$. Note that no such distributional restriction was necessary in discussing the distribution of the sample mean. The restriction to normality is common among functions of the sample values other than the sample mean, and, although much is known when this restriction is lifted, we cannot discuss this in this book.

We now present a fairly plausible derivation of the distribution of a function of the sample variance provided that the sample is chosen from a $N(\mu, \sigma^2)$ distribution.

From the definition of the sample variance, we can write

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2}.$$

Now the sum in the numerator can be written as

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2.$$

This in turn simplifies to

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2,$$

so

$$\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2} = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} - \frac{(\bar{x} - \mu)^2}{\sigma^2/n},$$

or

$$\frac{(n-1)s^2}{\sigma^2} + \frac{(\bar{x} - \mu)^2}{\sigma^2/n} = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}.$$

It can be shown, in sampling from a normal population, that \bar{X} and s^2 are independent. This fact is far from being intuitively obvious; its proof is beyond the scope of this book but a proof can be found in Hogg and Craig [18]. Using this fact of independence it follows that

$$M\left[\frac{(n-1)s^2}{\sigma^2}; t\right] \cdot M\left[\frac{(\bar{x} - \mu)^2}{\sigma^2/n}; t\right] = M\left[\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}; t\right],$$

where $M[X; t]$ denotes the moment generating function. Now $\sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}$ is the sum of squares of $N(0, 1)$ variables and hence has a chi-squared distribution with n degrees of freedom. Also $\frac{(\bar{x} - \mu)^2}{\sigma^2/n} = \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)^2$ is the square of a single $N(0, 1)$ variable and so has a chi-squared distribution with 1 degree of freedom. Therefore, using the moment generating function for the chi-squared random variable, we have

$$\begin{aligned} M\left[\frac{(n-1)s^2}{\sigma^2}; t\right] \cdot (1-2t)^{-\frac{1}{2}} &= (1-2t)^{-\frac{n}{2}} \quad \text{or} \\ M\left[\frac{(n-1)s^2}{\sigma^2}; t\right] &= (1-2t)^{-\frac{n-1}{2}}, \end{aligned}$$

indicating that $\frac{(n-1)s^2}{\sigma^2}$ has a χ^2_{n-1} distribution.

Since it can be shown that $E(\chi_{n-1}^2) = n - 1$, it follows that $E\left[\frac{(n-1)s^2}{\sigma^2}\right] = n - 1$ from which it follows that

$$E(s^2) = \sigma^2,$$

showing that s^2 is an unbiased estimator for σ^2 .

It is also true that $\text{Var}(\chi_{n-1}^2) = 2(n - 1)$ so

$$\begin{aligned}\text{Var}\left[\frac{(n-1)s^2}{\sigma^2}\right] &= \frac{(n-1)^2}{\sigma^4} \text{Var}(s^2) = 2(n-1), \quad \text{or} \\ \text{Var}(s^2) &= \frac{2\sigma^4}{n-1}.\end{aligned}$$

This shows that the sample variance is very variable, a multiple of the fourth power of the standard deviation. The variability of the sample variance was noted in the early part of this section and this result verifies that observation.

Also in the early part of this section, we considered the sampling distribution of the sample variance when we took samples of size three from the uniform distribution on the integers $\{1, 2, 3, \dots, 20\}$. The graph in Figure 4.11 resembles that in Figure 4.12 which in reality is a chi-squared distribution with 2 degrees of freedom. Figure 4.11, while at first appearing to be somewhat chaotic, is in reality remarkable since the sampling is certainly not done from a normal distribution with mean 0 and variance 1. This indicates that the sampling distribution of the sample variance may be somewhat *robust*, that is, insensitive to deviations from the assumptions used to derive it.

Example 4.13.1

Samples of size 5 are drawn from a normal population with mean 20 and variance 300. A 95% confidence interval for the sample variance, s^2 , is found by using the χ_4^2 curve whose graph is shown in Figure 4.14. A table of values for some chi-squared distributions can be found in Appendix B.

The normal distribution has a point of symmetry and this is often used in calculations. The chi-squared distribution, however, has no point of symmetry and so tables

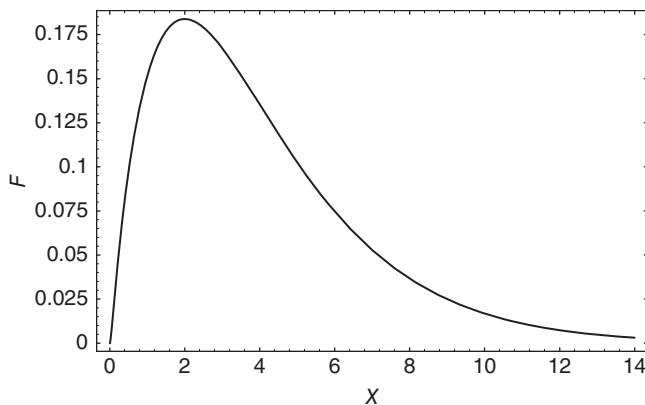


Figure 4.14 χ_4^2 distribution.

must be used to find both upper and lower significance points. We find, for example, that

$$P(0.2972011 \leq \chi_4^2 \leq 10.0255) = 0.95 \text{ so}$$

$$P\left(0.2972011 \leq \frac{4s^2}{300} \leq 10.0255\right) = 0.95 \text{ or}$$

$$P(22.290 \leq s^2 \leq 751.9125) = 0.95$$

a very large range for s^2 . It is approximately true that

$$P(4.721 \leq s \leq 27.42) = 0.95$$

by taking square roots in the confidence interval for s^2 . The exact distribution for s could be found by finding the probability distribution of the square root of a χ^2 distribution, but the above interval is a good approximation. We will consider the exact distribution in Section 4.17.

Other 95% confidence intervals are possible. Another example is

$$P(0.48442 \leq \chi_4^2 \leq 11.1433) = 0.95 \text{ which leads to the interval}$$

$$P(36.3315 \leq s^2 \leq 835.7475) = 0.95.$$

There are many other possibilities which can most easily be found with the aid of a computer algebra system since tables give very restricted choices for the chi-squared values needed. Note that the two 95% confidence intervals above have unequal lengths. This is due to the lack of symmetry of the chi-squared distribution.

EXERCISES 4.13

1. A sample of five “Six Hour” VCR tapes had actual lengths (in minutes) of 366, 339, 364, 356, and 379 minutes. Find a 95% confidence interval for σ^2 , assuming that the lengths are $N(\mu, \sigma^2)$.
2. It is crucial that the variance of a measurement of the length of a piston rod be no greater than 1 square unit. A sample gave the following lengths (which have been coded for convenience): -3, 6, -7, 8, 4, 0, 2, 12, -8. Find a one-sided 99% confidence interval for the true variance of the length measurements.
3. Suppose $X \sim N(\mu, \sigma^2)$ where μ is known. Find a 95% two-sided confidence interval for σ^2 based on a random sample of size n .
4. Suppose that $\{X_1, X_2, \dots, X_{2n}\}$ is a random sample from a distribution with $E[X] = 0$ and $Var[X] = \sigma^2$. Find k if

$$E[k \cdot \{(X_1 - X_2)^2 + (X_3 - X_4)^2 + (X_5 - X_6)^2 + \dots + (X_{2n-1} - X_{2n})^2\}] = \sigma^2.$$

5. A random sample of n observations from $N(\mu, \sigma^2)$ has $s^2 = 42$ and produced a two-sided 95% confidence interval for σ^2 of length 100. Find n .

6. Six readings on the amount of calcium in drinking water gave $s^2 = 0.0285$. Find a 90% confidence interval for σ^2 .
7. A random sample of 12 observations is taken from a normal population with variance 100. Find the probability that the sample variance is between 50 and 240.
8. A random sample of 12 shearing pins is taken in a study of the Rockwell hardness of the head of a pin. Measurements of the Rockwell hardness were made for each of the 12 giving a sample average of 50 with a sample standard deviation of 2. Find a 90% confidence interval for the true variance of the Rockwell hardness. What assumptions must be made for your analysis to be correct?
9. A study of the fracture toughness of base plate of 18% nickel maraging steel gave $s^2 = 5.04$ based on a sample of 22 observations. Assuming that the sample comes from a normal population, construct a 99% confidence interval for σ^2 , the true variance.

4.14 HYPOTHESIS TESTS AND CONFIDENCE INTERVALS FOR A SINGLE MEAN

We are now prepared to return to the structure of hypothesis testing considered in Chapter 2 and to show some applications of the preceding theory to the statistical analysis of data. Only the binomial distribution was available to us in Chapter 2. Now we have not only continuous distributions but also the central limit theorem, which is the basis for much of our analysis. We begin with an example.

Example 4.14.1

A manufacturer of steel has measured the hardness of the steel produced and has found that the hardness, X , has had in the past a mean value of 2200 lb. with a known standard deviation of 4591.84 lb. It is desired to detect any significant shift in the mean value, and for this purpose samples of 25 pieces of the steel are taken periodically and the mean strength of the sample, \bar{X} , is found. The manufacturer is willing to have the probability of a Type I error no greater than 0.05. When should the manufacturer decide that the steel no longer has mean hardness 2200 lb?

In this case, since it is desired to detect deviations either greater than or less than 2200 lb, we take as null and alternative hypotheses

$$H_o: \mu = 2200$$

$$H_a: \mu \neq 2200.$$

The central limit theorem tells us that

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ is approximately a } N(0, 1) \text{ variable.}$$

Since the alternative hypothesis is *two-sided*, that is, it comprises the two *one-sided* hypotheses $\mu > 2200$ and $\mu < 2200$, we take a two-sided rejection region, $\{\bar{X} > k\} \cup \{\bar{X} < h\}$. Since $\alpha = 0.05$, we find k and h such that

$$P[\bar{X} > k] = P[\bar{X} < h] = 0.025 \quad \text{so that}$$

$$\frac{k - 2200}{\sqrt{\frac{21085000}{25}}} = 1.96 \text{ and } \frac{h - 2200}{\sqrt{\frac{21085000}{25}}} = -1.96.$$

These equations give $k = 4000$ and $h = 400$ approximately. So H_0 is accepted if $400 \leq \bar{X} \leq 4000$.

The size of the Type II error, β , is a function of the specific alternative hypothesis. In this case if the alternative is $H_a: \mu = 2600$, for example, then

$$\beta = P[400 < \bar{X} < 4000 | \mu = 2600]$$

$$\begin{aligned} &= P\left[\frac{400 - 2600}{\sqrt{\frac{21085000}{25}}} \leq z \leq \frac{4000 - 2600}{\sqrt{\frac{21085000}{25}}}\right] \\ &= P[-2.39555 \leq z \leq 1.5244] \\ &= 0.927998, \end{aligned}$$

so the test is not particularly sensitive to this alternative.

Confidence Intervals, σ Known

Suppose that \bar{X} is the mean of a sample of n observations selected from a population with known standard deviation σ . By the central limit theorem for a given α we can find z so that

$$P\left(-z \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = 1 - \alpha.$$

These inequalities can in turn be solved for μ producing $(1 - \alpha)\%$ confidence intervals

$$\bar{X} - z \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z \cdot \frac{\sigma}{\sqrt{n}}.$$

Each of these confidence intervals has length $2 \cdot z \cdot \frac{\sigma}{\sqrt{n}}$.

Example 4.14.2

A sample of 10 observations from a normal distribution with $\sigma = 6$ gave a sample mean $\bar{X} = 28.45$. A 90% confidence interval for the unknown mean, μ , of the population is

$$\begin{aligned} 28.45 - 1.285 \cdot \frac{6}{\sqrt{10}} &\leq \mu \leq 28.45 + 1.285 \cdot \frac{6}{\sqrt{10}} \text{ or} \\ 26.0119 &\leq \mu \leq 30.8881. \end{aligned}$$

Example 4.14.3

How large a sample must be selected from a normal distribution with standard deviation 12 in order to estimate μ to within 2 units with probability 0.95?

Here $1/2$ the length of a 95% confidence interval is 2. So

$$\begin{aligned} z \cdot \frac{\sigma}{\sqrt{n}} &= 1.96 \cdot \frac{12}{\sqrt{n}} = 2 \text{ so} \\ n &= \left(\frac{1.96 \cdot 12}{2} \right)^2 \end{aligned}$$

Therefore a sample size of $n = 139$ is sufficient.

Student's t Distribution

In the previous example, it was assumed that σ^2 was known. What if this parameter is also unknown? This of course is the most commonly encountered situation in practice, that is, neither μ nor σ is known when the sampling is done. Although we will not prove it here, the following theorem is useful if the sampling is done from a normal population.

Theorem: The ratio of a standard normal random variable and the square root of an independent chi-squared random variable divided by n , its degrees of freedom, follows a Student's t distribution with $n - 1$ degrees of freedom. Symbolically,

$$\frac{N(0, 1)}{\sqrt{\chi_n^2/n}} = t_{n-1}.$$

A proof can be found in Hogg and Craig [18].

How is this of help here? We know that $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is approximately normal by the central limit theorem and we know from the previous section, if the sampling is done from a normal population, then $\frac{(n-1)s^2}{\sigma^2}$ is a chi-squared random variable with $n-1$ degrees of freedom. So

$$\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = t_{n-1}.$$

The sample then provides all the information we need to calculate t . The Student's t distribution (which was discovered by W. G. Gossett who wrote using the pseudonym "Student") becomes normal-like as the sample size increases but differs significantly from the normal distribution for small samples. Several t distributions are shown in Figure 4.15. A table of critical values for various t distributions can be found in Appendix B.

Now tests of hypotheses can be carried out and confidence intervals can be calculated if the sampling is from a normal distribution with unknown variance as the following example indicates.

Example 4.14.4

Tests on a ball bearing manufactured in a day's run in a plant show the following diameters (which have been coded for convenience): 8, 7, 3, 5, 9, 4, 10, 2, 6, 7.

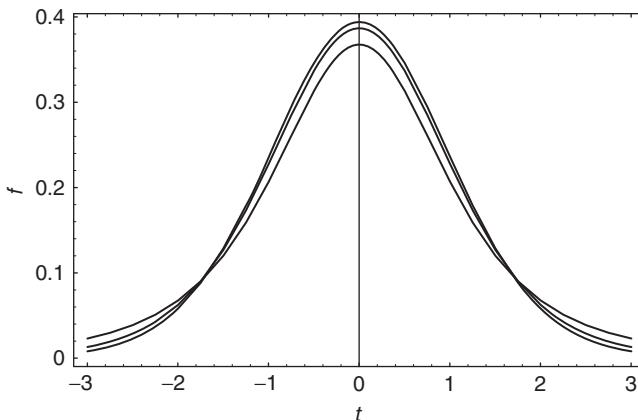


Figure 4.15 Student t distributions for 3, 8, and 20 degrees of freedom.

The sample gives $\bar{x} = 6.1$ and $s^2 = 203/30$.

If we wished to test $H_0: \mu = 7$ against the alternative $H_a: \mu \neq 7$ with $\alpha = 0.05$, we find that

$$t_9 = \frac{6.1 - 7}{\sqrt{(203/30)/10}} = -1.09.$$

A table of t values can be found in Appendix B. The critical values for t_9 are ± 2.26 , so the hypothesis is accepted.

Confidence intervals for μ can also be constructed. Using the sample data, we have

$$P\left[-2.26 \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq 2.26\right] = 0.95 \text{ so}$$

$$P\left[-2.26 \leq \frac{6.1 - \mu}{\sqrt{(203/30)/10}} \leq 2.26\right] = 0.95 \text{ which simplifies as}$$

$$P[4.5697 \leq \mu \leq 7.63023] = 0.95.$$

The 95% confidence interval is also the acceptance region for a hypothesis tested at the 5% level. Recall that, when σ is known, the confidence intervals arising from separate samples all have the same length. This was shown above. If, however, σ is unknown, then the confidence intervals will have varying widths as well as various central values. Some possible 95% confidence intervals are shown in Figure 4.16.

p Values

We have always given the α or *significance* value when constructing a test of a hypothesis. These values of α have an arbitrary appearance, to say the least. Who is to say that this significance level should be 0.05 or 0.01, or some other value? How does one decide what value to choose?

These are often troublesome questions for an experimenter. The acceptance or rejection of a hypothesis is of course completely dependent on the choice of the significance level.

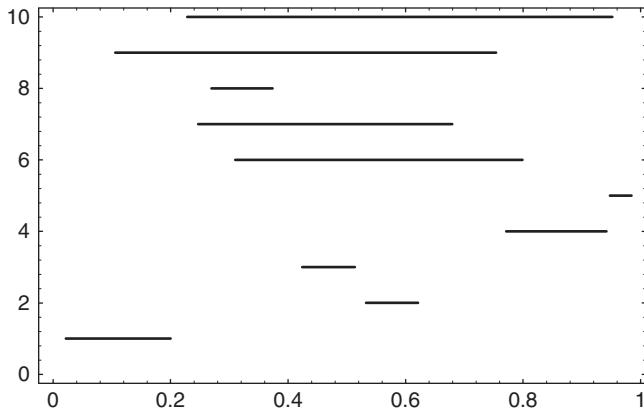


Figure 4.16 Some confidence intervals.

Another way to report the result of a test would be to report the *smallest value of α at which the test results would be significant*. This is called the *p value* for the test. We give some examples of this.

Example 4.14.5

In Example 4.14.4, we found that the sample of size 10 gave $\bar{x} = 6.1$ and $s^2 = 203/30$. This in turn produced $t_9 = -1.09$ and the hypothesis was accepted since α had been chosen as 5%.

However, we can use tables or a computer algebra system to find that

$$P(t_9 < -1.09) = 0.152018.$$

This means that the observed value for t would be in the rejection region if half the α value were less than 0.152018.

Since the test in this case is two sided, we report the *p value* as twice the above value, or 0.304036.

Now the person interpreting the test results can decide if this value suggests that the results are significant or not. Undoubtedly the decision here would be that the result is not significant although this *p value* would be of value and interest in many studies.

Example 4.14.6

Suppose we revise Example 4.14.1 as follows. Suppose the hypotheses are as follows:

$$H_o: \mu = 2200$$

$$H_a: \mu > 2200$$

and that a sample of size 25 gave a sample mean of $\bar{x} = 3945$.

Since we know in this case that $\sigma^2 = 21,085,000$ we find that

$$z = \frac{3945 - 2200}{\sqrt{\frac{21085000}{25}}} = 1.90011 \text{ and that}$$

$$P(Z > 1.90011) = 0.0287094,$$

so this is the p value for this test. If the significance level is greater than 0.0287094 then the result is significant; otherwise, it is not. Many computer statistical packages now report p values together with other test results.

EXERCISES 4.14

1. Test runs with an experimental engine showed they operated, respectively, for 24, 28, 21, 23, 32, and 22 minutes with 1 gallon of fuel.

- (a) Is this evidence at the 1% significance level that

$$H_o: \mu = 29 \text{ should be accepted against}$$

$$H_a: \mu < 29?$$

- (b) Find the p value for the test.

2. Machines used in producing a particular brand of yarn are given periodic checks to help insure stable quality. A machine has been set so that it is expected that strands of yarn it produces will have breaking strength $\mu = 19.50$ oz, with a standard deviation of 1.80 oz. A random sample of 12 pieces of yarn has a mean of 18.46 oz. Assuming that the standard deviation remains constant over a fairly wide range of values for μ ,

- (a) Test $H_o: \mu = 19.50$ against $H_a: \mu \neq 19.50$ at the 5% significance level. Find the p value for the test.

- (b) Now suppose that σ is also unknown and that the sample standard deviation is 1.80. Test the hypothesis in part a] again. Are any additional assumptions needed?

- (c) Under the conditions in part a], find β for the alternative $H_a: \mu = 19.70$.

3. “One quarter” inch rivets are produced by a machine which is checked periodically by taking a random sample of 10 rivets and measuring their diameters. It is feared that the wear-off factor in the machine will eventually cause the machine to produce rivets with diameters that are less than 1/4 inch. Assume that the variance of the diameters is known to be $(0.0015)^2$.

- (a) Describe the critical region, in terms of \bar{X} , for a test at the 1% level of significance for

$$H_o: \mu = 0.25 \text{ against the alternative}$$

$$H_a: \mu < 0.25.$$

- (b) What is the power of the test at $\mu = 0.2490$?

- (c) Now suppose we wish to test $H_o: \mu = 0.25$ against $H_a: \mu = 0.2490$ with $\alpha = 1\%$ and so that the power of the test is 0.99. What sample size is necessary to achieve this?

4. A manufacturer of light bulbs claims that the life of the bulbs is normally distributed with mean 800 hours and standard deviation 40 hours. Before buying a large lot, a buyer tests 30 of the bulbs and finds an average life of 789 hours.

- (a) Test the hypothesis $H_o: \mu = 800$ against the alternative $H_a: \mu < 800$ using a test of size 5%.
- (b) Find the probability of a Type II error for the alternative $H_a: \mu = 790$.
- (c) Find the p value for the test.
5. A sample of size 16 from a distribution whose variance is known to be 900 is used to test $H_o: \mu = 350$ against the alternative $H_a: \mu > 350$, using the critical region $\bar{X} > 365$.
- (a) What is α for this test?
- (b) Find β for the alternative $H_a: \mu = 372.50$.
6. A manufacturer of sports equipment has developed a new synthetic fishing line that he claims has a mean breaking strength of 8 kg. To test $H_o: \mu = 8$ against the alternative $H_a: \mu \neq 8$, a sample of 50 lines is tested; the sample has a mean breaking strength of 7.8 kg.
- (a) If σ is assumed to be 0.5 kg and $\alpha = 5\%$, is the manufacturer's claim supported by the sample?
- (b) Find β for the above test for the alternative $H_a: \mu = 7.7$
- (c) Find the p value for the test.
7. For a certain species of fish, a sample of measurements for DDT is 5, 10, 8, 7, 4, 9, and 13 parts per million.
- (a) Find a range of values of μ_o for which the hypothesis $H_a: \mu = \mu_o$ would be accepted at the 5% level.
- (b) Find a 95% confidence interval for σ^2 , the true variance of the measurements.
8. The time to repair breakdowns for an office copying machine is claimed by the manufacturer to have a mean of 93 minutes. To test this claim, 23 breakdowns of a model were observed, resulting in a mean repair time of 98.8 minutes and a standard deviation of 26.6 minutes.
- (a) Test $H_o: \mu = 93$ against the alternative $H_a: \mu > 93$ with $\alpha = 5\%$ and state your conclusions.
- (b) Supposing that $\sigma^2 = 625$, find β for the alternative $H_a: \mu = 95$.
- (c) Find the p value for the test.
9. A firm produces metal wheels. The mean diameter of these wheels should be 4 in. Because of other factors as well as chance variation, the diameters of the wheels vary with standard deviation 0.05 in. A test is conducted on 50 randomly selected wheels.
- (a) Find a test with $\alpha = 0.01$ for testing $H_o: \mu = 4$ against the alternative $H_a: \mu \neq 4$.
- (b) If the sample average is 3.97, what decision is made?
- (c) Calculate β for the alternative $H_a: \mu = 3.99$.
10. A tensile test was performed to determine the strength of a particular adhesive for a glass-to-glass assembly. The data are: 16, 14, 19, 18, 19, 20, 15, 18, 17, 18. Test $H_o: \mu = 19$ against the alternative $H_a: \mu < 19$,
- (a) if σ^2 is known to be 2.
- (b) if σ^2 is unknown.
11. The activation times for an automatic sprinkler system are a subject of study by the system's manufacturer. A sample of activation times is 27, 41, 22, 27, 23, 35, 30, 33, 24, 27, 28, 22, and 24 seconds. The design of the system calls for its activation in at most 25 seconds. Does the data contradict the validity of this design specification?

- 12.** The breaking strengths of cables produced by a manufacturer have mean 1800 lb. It is claimed that a new manufacturing process will increase the mean breaking strength of the cables. To test this hypothesis, a sample of 30 cables, manufactured using the new process, is tested giving $\bar{X} = 1850$ and $s = 100$.
- If $\alpha = 0.05$, what conclusion can be drawn regarding the new process?
 - Find the p value for the test.
- 13.** A sample of 80 observations is taken from a population with known standard deviation 56 to test $H_o: \mu \leq 300$ against the alternative $H_a: \mu > 300$ giving $\bar{X} = 310$.
- Find the minimum value of α so that H_o would be rejected by the sample.
 - Assuming that the critical region is $\bar{X} > 310$, find β for the alternative $\mu = 315$.
- 14.** A contractor must have cement with a compressive strength of at least 5000 kg/cm^2 . He knows that the standard deviation of the compressive strengths is 120. In order to test $H_o: \mu \geq 5000$ against the alternative $H_a: \mu < 5000$, a random sample of four pieces of cement is tested.
- If the average compressive strength of the sample is 4870 ksc., is the concrete acceptable? Use $\alpha = 0.01$.
 - The contractor must be 95% certain that the compressive strength is not less than 4800 ksc. How large a sample should be taken to insure this?
- 15.** The assembly time in a plant is a normal random variable with mean 18.5 seconds and standard deviation 2.4 seconds.
- A random sample of 10 assembly times gave $\bar{X} = 19.6$. Is this evidence that $H_o: \mu = 18.5$ should be rejected in favor of the alternative $H_a: \mu > 18.5$ if $\alpha = 5\%$?
 - Find the probability that H_o is accepted if $\mu = 19$.
 - It is very important that the assembly time not exceed 20 seconds. How large a sample is necessary to reject $H_o: \mu = 18.5$ with probability 0.95 if $\mu = 20$?
- 16.** A lot of rolls of paper is acceptable for making bags for grocery stores if its true mean breaking strength is not less than 40 lb. It is known from past experience that $\sigma = 2.5$ lb. A sample of 20 is chosen.
- Find the critical region for testing the hypothesis $H_o: \mu = 40$ against the alternative $H_a: \mu < 40$ at the 5% level of significance.
 - Find the probability of accepting H_o if in fact $\mu = 40.5$ lb.
 - If σ were unknown and a sample of 20 gave $\bar{X} = 39$ lb and $s = 2.4$ lb, would H_o be accepted with $\alpha = 5\%$?
- 17.** The drying time of a particular brand and type of paint is known to be normally distributed with $\mu = 75$ minutes and $\sigma = 9.4$ minutes. In an attempt to improve the drying time, a new additive has been developed. Use of the additive in 100 test samples of the paint gave an average drying time of 68.5 minutes. We wish to test $H_o: \mu = 75$ against the alternative $H_a: \mu < 75$.
- Find the critical region if $\alpha = 5\%$.
 - Does the experimental evidence indicate that the additive improves drying time?
 - What is the probability that H_o will be rejected if in fact $\mu = 72$ minutes?
 - Find the p value for the test.

18. The breaking strength of a fiber used in manufacturing cloth is required to be not less than 160 lb/in.² inch. Past evidence indicates that $\sigma = 3$ psi. A random sample of four specimens is tested and the average breaking strength is found to be 158 psi.
- Test $H_0: \mu = 160$ against a suitable alternative using $\alpha = 5\%$.
 - Find β for the alternative $\mu = 157$.
19. An engineer is investigating the wear characteristics of a particular type of radial automobile tire used by the company fleet of cars. A random sample of 16 tires is selected and each tire used until the wear bars appear. The sample gave $\bar{x} = 41,116$ and $s^2 = 1,814,786$.
- Find a so that $P(\mu > a) = 0.95$.
 - Find a 90% confidence interval for σ^2 .
 - Answer part (a) assuming that the sample size is 43 with \bar{x} and s^2 as before.
20. The diameter of steel rods produced by a sub-contractor is known to have standard deviation 2 cm., and, in order to meet specifications, must have $\mu = 12$.
- If the mean of a sample of size 5 is 13.3, is this sufficient to reject $H_0: \mu = 12$ in favor of the alternative $H_a: \mu > 12$? Use $\alpha = 0.05$.
 - The manufacturer wants to be fairly certain that H_0 is rejected if $\mu = 13$. How large a sample should be taken to make this probability 0.92?

4.15 HYPOTHESIS TESTS ON TWO SAMPLES

A basic scientific problem is that of comparing two samples, possibly one from a control group and the other from an experimental group. The investigator may want to decide whether or not the two populations from which the samples are drawn have the same mean value, or interest may center on the equality of the true variances of the populations. We begin with a comparison of population means.

Tests on Two Means

Example 4.15.1

Suppose an investigator is comparing two methods of teaching students to use a popular computer algebra program. One group (X) is taught by the conventional lecture-demonstration method while the second group (Y) is divided into small groups and uses cooperative learning. After some time of instruction, the groups are given the same examination with the following results:

$$\begin{aligned} n_x &= 13, \quad \bar{x} = 77, \quad s_x^2 = 193.7, \quad \text{and} \\ n_y &= 9, \quad \bar{y} = 84, \quad s_y^2 = 309.4. \end{aligned}$$

We wish to test the hypothesis

$$H_0: \mu_x = \mu_y \quad \text{against}$$

$$H_a: \mu_x < \mu_y.$$

Assume that the sampling is from normal distributions. We know that

$$E(\bar{X} - \bar{Y}) = \mu_x - \mu_y \quad \text{and that}$$

$$\text{Var}(\bar{X} - \bar{Y}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

so, from the central limit theorem,

$$z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \quad \text{is a } N(0, 1) \text{ variable.}$$

Now z can be used to test hypotheses or to construct confidence intervals if the variances are known. Consider for the moment that we know that the populations have equal variances, say $\sigma^2 = 289$. Then

$$z = \frac{(77 - 84) - 0}{\sqrt{\frac{289}{13} + \frac{289}{9}}} = -0.9496.$$

If the test had been at the 5% level then the null hypothesis would be accepted since $z > -1.645$.

We could also use z to construct a confidence interval. Here a one-sided interval is appropriate because of H_a . We have

$$P\left[\left(\bar{X} - \bar{Y}\right) - 1.645\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \leq \mu_x - \mu_y\right] = 0.95,$$

which becomes in this case the interval greater than -19.126 . Since 0 is in this interval, the hypothesis of equal means is then accepted.

Example 4.15.2

A situation more common than that in the previous example occurs when the population variances are unknown. There are then two possibilities: they are equal or they are not. We consider first the case where the variances are unknown, but they are known to be equal. Denote the common value for the variances by σ^2 . The variable

$$z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \quad \text{is a } N(0, 1) \text{ variable.}$$

Now

$$\frac{(n_x - 1)s_x^2}{\sigma^2} + \frac{(n_y - 1)s_y^2}{\sigma^2} \quad \text{is a } \chi^2 \text{ variable with } (n_x - 1) + (n_y - 1) = n_x + n_y - 2$$

degrees of freedom since each of the summands is a chi-squared variable. Since a t variable is the ratio of $N(0, 1)$ variable to the square root of a chi-squared variable divided by its number of degrees of freedom, it follows that

$$t_{n_x+n_y-2} = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}}}{\sqrt{\frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}}}.$$

This can be simplified to

$$t_{n_x+n_y-2} = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \text{ where } s_p^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2}.$$

s_p^2 is called the *pooled variance*.

Using the data in Example 4.15.1, we find that

$$s_p^2 = \frac{12(193.7) + 8(309.4)}{13+9-2} = 239.98 \quad \text{and}$$

$$t_{20} = \frac{(77-84)-0}{\sqrt{239.98 \left(\frac{1}{13} + \frac{1}{9} \right)}} = -1.04.$$

Since the one-sided test rejects H_0 if $t_{20} < -1.725$, the hypothesis is accepted if $\alpha = 0.05$.

Example 4.15.3

Finally, we consider the case where the population variances are unknown and cannot be assumed to be equal. (Later in this chapter, we will show how that hypothesis may be tested also.) Unfortunately, there is no exact solution to this problem, known in the statistical literature as the Behrens–Fisher problem. Several approximate solutions are known; we give one here due to Welch [36].

Welch's approximation is as follows:

The variable

$$T = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

is approximately a t variable with v degrees of freedom where

$$v = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^2}{\frac{\left(\frac{s_x^2}{n_x} \right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y} \right)^2}{n_y - 1}}.$$

Using the data in the previous examples, we find that $v = 14.6081$ so we must use a t variable with 14 degrees of freedom. This gives

$$T_{14} = -0.997178,$$

a result quite comparable to previous results. The critical t value is -1.761 . The Welch approximation will make a very significant difference if the population variances are quite disparate.

Tests on Two Variances

It is essential to determine whether or not the population variances are equal before testing the equality of population means. It is possible to test this using two samples from the populations.

If χ_a^2 and χ_b^2 are independent chi-squared variables then the random variable

$$\frac{\chi_a^2/a}{\chi_b^2/b} = F(a, b),$$

where $F(a, b)$ denotes the F random variable with a and b degrees of freedom respectively. A proof of this fact will not be given here. The reader is referred to Hogg and Craig [18] for a proof. A table of some critical values of the F distribution can be found in Appendix B.

The probability density function for $F(a, b)$ is

$$f(x) = \frac{\Gamma\left(\frac{a}{2}\right) \cdot \Gamma\left(\frac{b}{2}\right)}{\Gamma\left(\frac{a+b}{2}\right)} \cdot a^{\frac{a}{2}} \cdot b^{\frac{b}{2}} \cdot (ax + b)^{-\frac{a+b}{2}} \cdot x^{\frac{a}{2}-1}, \quad x \geq 0.$$

The F variable has two numbers of degrees of freedom; one is associated with the numerator and the other with the denominator. Due to the definition of F , it is clear that

$$\frac{1}{F(a, b)} = F(b, a).$$

So the reciprocal of an F variable is an F variable with the numbers of degrees of freedom interchanged.

Several F curves are shown in Figure 4.17.

The F distribution can be used in testing the equality of variances in the following way. If the sampling is from normal populations, then

$$\frac{(n_x - 1)s_x^2}{\sigma_x^2} \text{ and } \frac{(n_y - 1)s_y^2}{\sigma_y^2}$$

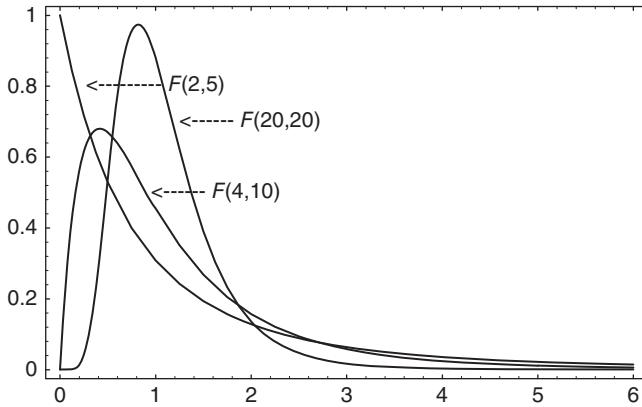


Figure 4.17 Some F distributions.

are independent chi-squared variables. It follows then that

$$\frac{\frac{(n_x-1)s_x^2}{\sigma_x^2(n_x-1)}}{\frac{(n_y-1)s_y^2}{\sigma_y^2(n_y-1)}}$$

is the ratio of two independent chi-squared random variables each divided by its number of degrees of freedom; it follows that this variable, simplified as

$$\frac{\frac{s_x^2}{\sigma_x^2}}{\frac{s_y^2}{\sigma_y^2}} = F(n_x - 1, n_y - 1).$$

Now consider the hypotheses

$$\begin{aligned} H_0: \sigma_x^2 &= \sigma_y^2 \\ H_a: \sigma_x^2 &\neq \sigma_y^2. \end{aligned}$$

If the null hypothesis is true, then the F variable becomes

$$F(n_x - 1, n_y - 1) = \frac{s_x^2}{s_y^2}.$$

This is used as the test statistic with a two-tailed critical region.

Example 4.15.4

As an example, consider the data used in the previous section where

$$n_x = 13, \quad s_x^2 = 193.7 \text{ and}$$

$$n_y = 9, \quad s_y^2 = 309.4.$$

Here $F(12, 8) = \frac{193.7}{309.4} = 0.626$. The critical values, choosing $\alpha = 0.05$, are 4.1995 and 0.2002 so the null hypothesis is accepted.

One-sided tests can also be performed with the F statistic; it is common not to worry if a variance is too small, but in many instances, care must be taken that the variance has not become too large. In this case, large variation may result in a production process producing too great a percentage of product that does not meet specifications. We will discuss this further in Section 4.17.

EXERCISES 4.15

1. To test the hypothesis that the resistance of wire can be reduced by at least 0.050 ohms by alloying, samples of 12 for each type of wire gave the following results:

	Mean	Standard Deviation
Alloyed wire	0.083	0.003
Standard wire	0.136	0.002

- (a) Test $H_0: \sigma_1^2 = \sigma_2^2$ using $\alpha = 0.05$.
 (b) Does the data substantiate the claim?
2. Two analysts took repeated readings on the hardness of city water with the following results:

Analyst A <i>x</i>	Analyst B <i>y</i>
0.46	0.82
0.62	0.61
0.37	0.89
0.40	0.51
0.44	0.33
0.58	0.48
0.48	0.23
0.53	0.25
	0.67
	0.88

- (a) Test $H_0: \mu_x = 0.55$ against the alternative $H_1: \mu_x \neq 0.55$ using $\alpha = 0.05$.
 (b) Test the hypothesis in part a] again assuming now that $\sigma_x^2 = 0.0081$.
 (c) Test $H_0: \mu_x = \mu_y$ against the alternative $H_a: \mu_x < \mu_y$ with $\alpha = 5\%$.

3. Over a long period of time, ten patients selected at random are each given two different treatments for arthritis. The results of standard tests are as follows:

Patient	Treatment 1	Treatment 2
1	47	52
2	38	35
3	50	52
4	33	35
5	47	46
6	23	27
7	40	45
8	42	41
9	15	17
10	36	41

Test $H_0: \mu_1 = \mu_2$ against $H_a: \mu_1 < \mu_2$ at the 1% level of significance, assuming that the population variances are equal.

4. An experiment compares two different processes for producing steel plate. The measurements represent the thickness of the plate. The samples gave

$$n_x = 8 \quad \bar{x} = 6.701 \quad s_x = 0.108$$

$$n_y = 6 \quad \bar{y} = 6.841 \quad s_y = 0.155$$

- (a) Using $\alpha = 0.02$, test $H_0: \sigma_x^2 = \sigma_y^2$ against the alternative $H_a: \sigma_x^2 \neq \sigma_y^2$.
 (b) Now test $H_0: \mu_x = \mu_y$ against the alternative $H_a: \mu_x \neq \mu_y$ using a 5% test.

5. Samples from normal populations gave

$$n_x = 6 \quad \bar{x} = 22.6 \quad s_x^2 = 102.4$$

$$n_y = 8 \quad \bar{y} = 31.9 \quad s_y^2 = 89.6$$

Find a 98% confidence interval for σ_x^2/σ_y^2 .

6. In comparing times to failure (in hours) of two different types of light bulbs, two samples gave

$$n_x = 13 \quad \bar{x} = 984 \quad s_x^2 = 8742$$

$$n_y = 15 \quad \bar{y} = 1121 \quad s_y^2 = 9411$$

Find a 95% confidence interval for the difference of the true population means, $\mu_x - \mu_y$,

- (a) assuming $\sigma_x^2 = \sigma_y^2$.
 (b) assuming $\sigma_x^2 = 9000$ and $\sigma_y^2 = 9500$.

7. In a batch chemical process, two catalysts are being compared for their effect on the output of the process reaction. A sample of 11 batches was prepared using catalyst 1

and a sample of 9 batches was prepared using catalyst 2. The sample results are as follows:

$$\begin{aligned} n_1 &= 11 \quad \bar{x}_1 = 85 \quad s_1^2 = 16 \\ n_2 &= 9 \quad \bar{x}_2 = 81 \quad s_2^2 = 25. \end{aligned}$$

Assuming that the true variances are equal, find a 95% confidence interval for the differences between the means, $\mu_1 - \mu_2$.

8. To determine yield strengths, a study of 10 pieces of cold-rolled steel (X) gave a sample mean of 29.8 kilograms per square inch (ksi.) and a sample variance of 4.2 ksi². A second sample of 13 pieces of galvanized steel (Y) gave a sample mean of 34.7 ksi and a sample variance of 4.9 ksi².
 - (a) Assuming the true variances are equal, find a 95% confidence interval for the difference between the true strengths, $\mu_x - \mu_y$.
 - (b) Repeat part (a) assuming the true variances are $\sigma_x^2 = 4$ and $\sigma_y^2 = 5$.
9. An experiment is conducted to compare the crash resistance of two different types of automobile bumpers. Type A bumpers were mounted on 12 cars and type B bumpers were mounted on 9 cars. The cars were driven into a concrete wall at 10 mph. and the resulting damage (in \$ to repair) was assessed. The results were as follows:

$$\begin{aligned} \bar{A} &= 235 \quad s_A^2 = 421 \\ \bar{B} &= 286 \quad s_B^2 = 511. \end{aligned}$$

- (a) Would the hypothesis $\sigma_A^2 = \sigma_B^2$, when tested against $\sigma_A^2 \neq \sigma_B^2$ with $\alpha = 2\%$, be accepted or rejected?
- (b) Find a 98% confidence interval for σ_A^2 / σ_B^2 .
- (c) Test $H_0: \mu_A = \mu_B$ against the alternative $H_a: \mu_A < \mu_B$ in a test of size 1%.
10. The vending machines in the student lounge and in the cafeteria should dispense the same amount of coffee. However, some students believe that the mean amount of coffee dispensed in the lounge (L) is less than that dispensed in the cafeteria (C). The following summary statistics were obtained from samples from each machine.

$$\begin{aligned} n_L &= 12 \quad \bar{x}_L = 10.1 \quad s_L^2 = 0.8 \\ n_C &= 10 \quad \bar{x}_C = 9.8 \quad s_C^2 = 1.4. \end{aligned}$$

Is there statistical evidence to support the student's claim? Assume that the amounts dispensed are approximately normal and use a test of size 5%.

11. A recent study of accident victims in a Boston hospital gave the following results:

	<i>n</i>	Mean	Standard Deviation
Seat Belts	15	565	220
No Seat Belts	12	1200	540

the data indicating the cost of hospitalization.

- (a) Assuming that σ_{SB} (the true standard deviation for seat belt wearers) is 220 and σ_{NSB} (the true standard deviation for nonseat belt wearers) is 540, find a 95% confidence interval for $\mu_{SB} - \mu_{NSB}$.
- (b) Answer part (a) assuming now that the true standard deviations are unknown, but equal.
- (c) Is it tenable to believe that the population variances are equal? State the smallest p value at which the data would reject the hypothesis of equal variances (against the alternative of unequal variances).
12. The following data represent the running times of films produced by two motion picture companies:

Company	Time (minutes)						
A	102	86	98	109	92		
B	81	165	97	134	92	87	114

- (a) Test $H_o: \sigma_A^2 = \sigma_B^2$ against the alternative $H_a: \sigma_A^2 \neq \sigma_B^2$ with $\alpha = 2\%$.
- (b) Test $H_o: \mu_A = \mu_B - 10$ against the alternative $\mu_A < \mu_B - 10$ with $\alpha = 1\%$.
13. Wire cable is manufactured by two processes. It is desired to determine if the process affects the mean breaking strength of the cable. Laboratory tests are performed by putting samples under tension and recording the load required to break the cable. Following is the sample data:

Sample	Size	Mean	Variance
X	6	8.2	2.0
Y	7	11.2	4.0

- (a) Test $H_o: \sigma_x^2 = \sigma_y^2$ against the alternative $H_a: \sigma_x^2 \neq \sigma_y^2$ with $\alpha = 2\%$.
- (b) Now test $H_o: \mu_x = \mu_y$ against the alternative $H_a: \mu_x \neq \mu_y$ using a 5% test.
14. Five samples of a ferrous-type substance are to be used to determine if there is a difference between a laboratory chemical analysis and an X-ray fluorescence analysis of iron content. Each sample was split into two sub-samples and the two types of analysis were applied with the following data, which represents per cent yield:

	Sample				
Analysis	1	2	3	4	5
X-Ray	11.0	2.0	8.3	3.1	2.4
Chemical	11.2	1.9	8.5	3.3	2.4.

Assuming the population of measurements to be normal, test whether or not the two methods of analysis give on average, the same result. Use $\alpha = 5\%$.

15. An automobile designer suggests that painting a racing car reduces its top speed. He selects 6 cars and tests them with and without paint. The results are as follows:

	Top Unpainted	Speed (mph) Painted
1	189	186
2	186	185
3	183	179
4	188	184
5	185	183
6	188	186

Use the data to decide whether or not that painting the cars reduces the top speed. Use 5%.

- 16.** The golf scores of two competitors, A and B, are recorded over a period of 10 days. Scores for the two golfers are recorded over a period of 10 different days on which weather conditions varied widely. Golfer A claims that his game is better than golfer B. Does the data support this claim?

Day	A	B
1	87	89
2	86	85
3	79	83
4	82	87
5	78	76
6	87	90
7	84	85
8	81	78
9	83	85
10	81	84

- 17.** A wire manufacturer alters the production process hoping to increase the resistance of the wire. Below are the results of samples taken from the old process and the new process. Has the resistance of the wire increased?

New Process	Old Process
0.140	0.135
0.138	0.140
0.143	0.136
0.142	0.142
0.144	0.138
0.137	0.140

- 18.** Two randomly selected groups of industrial trainees are taught a new assembly line operation by two different methods. Measurements were made on the time to complete the operation with the following results:

Group	Size	Mean	Standard Deviation
I	10	60.43	20.2
II	10	31.23	26.8

- (a) Assuming normality, test $H_o: \sigma_I^2 = \sigma_{II}^2$ against the alternative $H_a: \sigma_I^2 \neq \sigma_{II}^2$ at the 5% level of significance.
- (b) Test $H_o: \mu_I = \mu_{II}$ against the alternative $H_a: \mu_I \neq \mu_{II}$ at the 5% level of significance.
- (c) Determine the values of k for which $H_o: \mu_I = \mu_{II} + k$ would be accepted at the 5% level of significance when tested against the alternative $H_a: \mu_I \neq \mu_{II} + k$.
19. Two samples are drawn from normal populations, each with variance 100. Due to sampling costs, it is possible to select $2n$ items from population A , but only n items from population B . In testing $H_o: \mu_A = \mu_B$ against the alternative $H_a: \mu_A = \mu_B + 3$ it is desired to have $\alpha = \beta = 0.10$. Find n so that this is approximately so and then discuss the implications of rounding n to an integer.
20. Suppose samples of sizes n_X and n_Y are available from normal populations whose variances and means are unknown. It is suspected that $\sigma_X^2 = 2\sigma_Y^2$.
- (a) Explain, by establishing the distribution of $\frac{s_X^2}{2s_Y^2}$, how $\frac{s_X^2}{2s_Y^2}$ can be used to test the hypothesis that $\sigma_X^2 = 2\sigma_Y^2$.
- (b) Assuming that the hypothesis in part a] is accepted, show that a test of $H_o: \mu_X = \mu_Y$ can be based on

$$r = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{s_w \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}, \quad \text{where}$$

$$s_w^2 = \frac{(n_X - 1)s_X^2 + 2(n_Y - 1)s_Y^2}{2(n_X + n_Y - 2)}$$

by establishing the distributions of r and s_w^2 .

4.16 LEAST SQUARES LINEAR REGRESSION

The estimation of unknown parameters was considered in the beginning of this chapter; we return to that problem here and introduce a new principle of estimation, that of *least squares*. This is a commonly used principle when a straight line or other curve must be fitted to a set of data and when one wants the “best” fitting straight line or curve to the experimental data.

Example 4.16.1

Suppose we are given the data set $\{x_1, x_2, \dots, x_n\}$. We want to find a number, a , such that

$$S = \sum_{i=1}^n (x_i - a)^2 \quad \text{is as small as possible.}$$

That is, we want to minimize the sum of the squared deviations from a . Such estimates, when they exist, are known as *least squares estimates*. In this case letting the derivative of

S with respect to a equal 0 gives

$$\frac{dS}{da} = -2 \sum_{i=1}^n (x_i - a) = 0 \quad \text{with the solution}$$

$$\hat{a} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}, \quad \text{the mean of the data set.}$$

So the sum of the squared deviations, $\sum_{i=1}^n (x_i - a)^2$, is minimized when $a = \bar{x}$.

Example 4.16.2

A researcher suspects that the achievement score on a standard mathematics examination, Y , for a group of students is a linear function of the student's IQ score, X . In order to investigate this hypothesis, data are collected and a *model* of the situation is presumed. In this case the model is composed of a *linear* part, $a + bx_i$, reflecting the researcher's hypothesis, and a *random* part, ϵ_i , reflecting the fact that the relationship between Y and X may be subject to other factors that are not accounted for in the experiment. The random part ϵ_i is in fact an observation of a random variable. Often this is a normal random variable as we will see. The model chosen is

$$y_i = a + bx_i + \epsilon_i, i = 1, 2, \dots, n.$$

Here y_i is the i th observation of Y , and x_i is the i th observation of X .

There are two fundamental problems here: one is the estimation, from the data, of the unknowns a and b ; the other is, given a and b , does the line fit the data well or not?

To answer the first question, we will use the principle of least squares to estimate the parameters a and b . This principle chooses values of a and b that minimize a sum of squares, namely,

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

We take the partial derivatives of S with respect to a and b and equate each to 0:

$$\frac{\partial S}{\partial a} = 2 \sum_{i=1}^n (y_i - a - bx_i)(-1) = 0$$

$$\frac{\partial S}{\partial b} = 2 \sum_{i=1}^n (y_i - a - bx_i)(-x_i) = 0.$$

Summing and simplifying gives

$$\begin{aligned} \sum_{i=1}^n y_i &= n\hat{a} + \hat{b} \sum_{i=1}^n x_i \quad \text{and} \\ \sum_{i=1}^n x_i y_i &= \hat{a} \sum_{i=1}^n x_i + \hat{b} \sum_{i=1}^n x_i^2. \end{aligned} \tag{4.2}$$

Equations ((4.2)) are called *least squares equations*. Their simultaneous solution is

$$\bar{y} = \hat{a} + \hat{b}\bar{x}$$

and

$$\hat{b} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Usually \hat{b} is found and then \hat{a} is found from the equation $\bar{y} = \hat{a} + \hat{b}\bar{x}$.

Suppose now that the data collected is as follows:

Math. score	92	86	104	109	75	100	91	110	128
IQ	104	91	123	102	86	99	92	114	99

A scatter plot of the data points is shown in Figure 4.18. The data appears to be somewhat linear, with considerable variation.

Substituting in ((4.2)) we find that the least squares estimates for a and b are as follows:

$$\hat{a} = 63.0792 \text{ and } \hat{b} = 0.38244$$

so that the least squares line, which is called the *regression of Y on X*, is

$$\hat{y}_i = 63.0792 + 0.38244x_i.$$

Figure 4.19 shows this line plotted with the data points.

The line does not appear to predict the Y values very well, so we consider whether or not the line fits the data satisfactorily.

We begin with the values the line predicts for the X values in the data set. We show a table below of the data points, the predicted values, and the *residuals* (the observed Y values minus the predicted Y values).

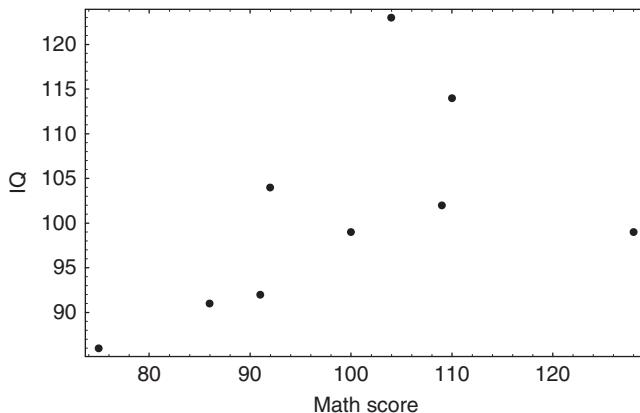


Figure 4.18 Scatter plot of data.

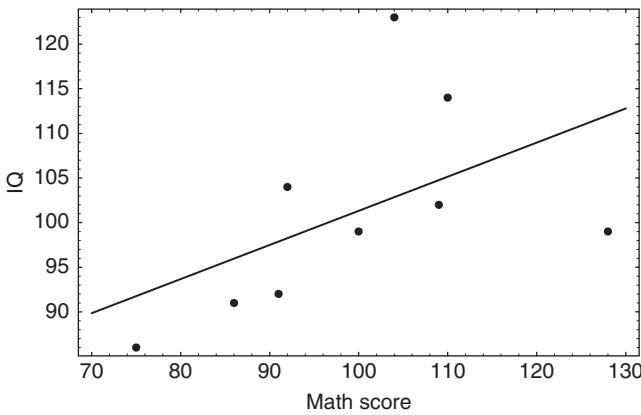


Figure 4.19 Regression line and data points.

MathScore (X)	92	86	104	109	75	100	91	110	128
IQ(Y)	104	91	123	102	86	99	92	114	99
Prediction	98.26	95.97	102.85	104.77	91.77	101.32	97.88	105.15	112.03
Residual	5.74	-4.97	20.15	-2.77	-5.76	-2.32	-5.88	8.85	-13.03

The absolute size of these residuals does not tell much except, as we previously noted, some of the residuals are quite large. To show a specific test of the hypothesis that the straight line fits the data well, consider the *total sum of squares of the residuals*:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ where } \hat{y}_i = \hat{a} + \hat{b}x_i.$$

Now we show a remarkable identity by adding and subtracting \bar{y} :

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \sum_{i=1}^n [(y_i - \bar{y}) - (\hat{y}_i - \bar{y})]^2. \\ \sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2 \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \end{aligned}$$

Using ((4.2)) the last two terms can be combined and we have

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{or} \\ \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \end{aligned} \tag{4.3}$$

This identity is an example of an *analysis of variance* identity. Such identities arise frequently in the analysis of experimental data. The terms have individual interpretations with respect to the regression problem:

$\sum_{i=1}^n (y_i - \bar{y})^2$ is the *total sum of squares*,

$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ is the *sum of squares due to regression*, and

$\sum_{i=1}^n (y_i - \hat{y}_i)^2$ is the *residual or error sum of squares*.

The identity ((4.3)) then partitions the total sum of squares into two parts: the sum of squares due to regression and the residual sum of squares. It is beyond the scope of this book, but, presuming the error term, ϵ_i , in the original model to be normally distributed with mean 0 and variance σ^2 , the sum of squares due to regression can be shown to have a chi-squared distribution; it can be shown that the error sum of squares, divided by $n - 2$, is also a chi-squared variable; moreover the two chi-squared variables are independent. It follows that the ratio of the chi-squared variables is a F variable. This information is usually exhibited in an *analysis of variance table*:

Source	Analysis of variance table			
	Sum of Squares	df	Mean Square	$F(1, n - 2)$
Regression	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	1	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / 1$	$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / 1}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}$
Error	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - 2$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - 2)$	
Total	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$		

If the data points are linearly related, then we expect some of the predicted values to differ significantly from the average of the y values, so we expect $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ to be large and we would expect that the error sum of squares, $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ to be small since the predicted values and the observed values should be close together. So, if the regression is truly linear, we expect the F ratio to be large. This leads to a one-tailed test of the hypothesis that the data follow a linear relationship.

The analysis of variance table for the data in this example is as follows:

Source	Analysis of variance table			
	Sum of Squares	Degrees of Freedom	Mean Square	$F(1, 7)$
Regression	284.368	1	284.368	2.5117
Error	792.521	7	113.217	
Total	1076.889	8		

The test is assessed by calculating $P[F(1, 7) \geq 2.5117] = 0.157022$. The test in this case can be shown to be a one-tailed test with the rejection region in the right-hand tail. We conclude that the regression is not significant in this case. The analysis shows that there is more random scatter in the data than there is a linear relationship; this might be suspected from the sums of squares above since the sum of squares for regression is 284.368 while the error sum of squares, 792.521, is much larger.

The ratio of the sum of squares due to regression to the total sum of squares is called the *coefficient of determination* or the square of the *correlation coefficient*, r :

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Since the numerator in r^2 is at most equal to the denominator, it follows that

$$0 \leq r^2 \leq 1 \quad \text{and so}$$

$$-1 \leq r \leq 1.$$

In this example, $r^2 = 0.264065$ or $r = 0.5139$. So only about 26% of the total variation in the y values is due to a linear relationship; 74% is due to randomness.

Computer algebra systems and statistical analysis packages make the calculation of the analysis of variance table easy; this should always be done since a least squares fit without a test for linearity is quite meaningless. The interpretation of a possible linear relationship based on the correlation coefficient alone is not advised. We will return to a study of least squares linear regression in Chapter 7.

EXERCISES 4.16

1. The following data represent the weight, x , in units of 1000 lb, and y , the fuel consumption in gallons per 100 miles, for six different brands of automobiles:

x	3.4	4.1	2.6	2.0	1.9	3.4
y	5.5	6.5	3.6	2.9	3.1	4.9

- (a) Make a scatter plot of the data.
 (b) Fit a least squares regression line to the data.
 (c) Show the analysis of variance table and state the conclusions that can be drawn from it.
2. Ohm's law can be written in the form of a regression as $I = \beta_0 + \beta_1 V$ where $\beta_0 = 0$ and $\beta_1 = 1/R$. Since V is set by the experimenter, this can be thought of as the independent variable while I is viewed as the dependent variable. Data from an experiment on one wire are as follows:

V	0.5	1.0	1.5	1.8	2.0
I	0.52	1.19	1.62	2.00	2.40

- (a) Plot the data. Do there appear to be any unusual points?
 (b) Find the least squares fit for the data.
 (c) Interpret the analysis of variance table and state conclusions this has for the experiment.
3. A recent paper reported a study on the relationship between applied stress (the independent variable, X , in kg/mm) and the time to fracture (the dependent variable, Y , in hours) for a type of stainless steel under uniaxial stress in a solution at a constant temperature. Ten different settings of applied stress were used and the following data resulted:

	X	Y
1	2.5	63
2	5.0	58
3	10.0	55
4	15.0	61
5	17.5	62
6	20.0	37
7	25.0	38
8	30.0	45
9	35.0	46
10	45.0	19

- (a) Plot the data. Do there appear to be any unusual points?
 (b) Find the least squares fit for the data.
 (c) Interpret the analysis of variance table and state any conclusions that can be drawn from it.
4. A small study on productivity in a factory compared hours worked (x) with parts assembled (y). The data are as follows:

x	y
1	2
2	3
3	5
4	6
5	4

- (a) Find the equation of the least squares regression line.
 (b) Show the analysis of variance table and state any conclusions that can be drawn from it.
 (c) Find the correlation coefficient.
5. The following data represent the total number of items produced by a manufacturing process (Y) and the total cost involved in production (X).

X	10	12	20	21	22	20	19
Y	10	15	20	20	25	30	30

- (a) Show a scatter plot of the data.
- (b) Find the equation of the least squares regression line predicting Y from X .
- (c) Find the equation of the least squares regression line predicting X from Y . Explain why the answer here is not equivalent to the answer in part (b).
6. A study was done to compare engine size (measured by cubic inches of displacement) and miles per gallon estimates for eight compact automobiles. The data are as follows:

$CID(X)$	121	120	97	98	122	97	85	122
$MPG(Y)$	30	31	34	27	29	34	38	32

- (a) Find the equation of the least squares regression line predicting Y from X .
- (b) Show the analysis of variance table and discuss any conclusions that can be drawn from it.
- (c) Find the correlation coefficient.
7. Raw material used in the production of a synthetic fiber is stored in a place that has no humidity control. Measurements of the relative humidity in the storage place and the moisture content of a sample of the raw material (both in percentages) on 12 days were as follows:

$Humidity(X)$	46	53	37	42	34	29	60	44	41	48	33	40
$Moisture(Y)$	12	14	11	13	10	8	17	12	10	15	9	13

- (a) Find the equation of the least squares regression line predicting Y from X .
- (b) Find the correlation coefficient. What is the interpretation of this number?
8. The yield of a chemical process is thought to be a function of the amount of catalyst added to the reaction. An experiment gave the following data:

$Yield(Y)$	60.54	63.86	63.76	60.15	66.66	71.66	70.81	65.72
$Catalyst(X)$	0.9	1.4	1.6	1.7	1.8	2.0	2.1	2.3

- (a) Find the least squares regression line predicting Y from X .
- (b) Find the correlation coefficient.
9. An experimenter has a data set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ and wishes to fit an equation of the form $y_i = \beta x_i^2$ to the data.
- (a) Use the principle of least squares to find a formula for $\hat{\beta}$, the least squares estimator for β .
- (b) Use the result in part (a) to calculate $\hat{\beta}$ for the data:

x	1	2	3	4	5	6	7
y	2	3	10	15	22	31	50

10. An experimenter wishes to fit an equation of the form $y_i = \alpha + \frac{\beta}{x_i}$ to the data set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. Show the least squares equations and show how to use these to estimate α and β .

- 11.** In a regression situation it is known that the regression line passes through the origin.
The model is as follows:

$$y_i = \beta x_i + e_i, \quad i = 1, 2, \dots, n,$$

where the e'_i 's are distributed independently and normally with mean 0 and variance σ^2 .

- (a) Find the least squares estimator of β , $\hat{\beta}$.
- (b) Let $a_i = \frac{x_i}{\sum_{i=1}^n x_i^2}$, a constant for each value of i . Find $E[\hat{\beta}]$ and $Var[\hat{\beta}]$.
- (c) In the usual regression situation $\sum_{i=1}^n (y_i - \hat{y}_i) = 0$. Show that this is not necessarily true in this case.
- (d) Show that:

$$1] \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n y_i^2 - \hat{\beta}^2 \sum_{i=1}^n x_i^2 \quad \text{and}$$

$$2] E \left[\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right] = (n-1)\sigma^2$$

- 12.** Suppose that an experimenter wishes to fit a straight line of the form $y_i = \alpha + mx_i$ to the data set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where m is a known constant.
- (a) Find the least squares estimate of α , $\hat{\alpha}$.
 - (b) Show that $E(\hat{\alpha}) = \alpha$.

4.17 QUALITY CONTROL CHART FOR \bar{X}

Manufacturers often monitor product quality through periodic sampling during a production process. The samples taken are usually small in size and are frequently reduced to simple statistics, such as the sample mean or range, for each sample. These statistics are then plotted in a graph indicating the time series of the measurements so that monitoring of the process can be done as time progresses. Such charts are called *quality control charts*. We will consider one type of quality control chart in this section and some of the mathematics behind the analysis of the data collected. We consider a specific example.

Example 4.17.1

A manufacturer of ball bearings takes periodic random samples of size 4 from the production line and measures the mean diameter of the ball bearings, \bar{X} , for each sample. The sample data (which has been coded for convenience) is shown below together with \bar{X} and the sample standard deviation, s for each sample.

Now the sample means are plotted in time order sequence in Figure 4.20.

Now what does the chart tell us about the process? First we seek a central value for the means. The sample size is small and generally nothing is known about the mean of the population from which the samples were drawn (in fact that population may be changing resulting in a change in the population mean which is one reason the control chart is

Data	\bar{X}	s
3, -1, -6, 4	0	4.54606
9, 0, 3, -2	2.5	4.79583
-12, 4, -9, -6	-5.75	6.94622
11, 9, 4, 1	6.25	4.57347
-1, -2, 4, -1	0	2.70801
8, 1, -2, 3	2.5	4.20317
-1, -4, -9, -3	-4.25	3.40343
-8, -3, -6, -4	-5.25	2.21736
1, -2, -2, 1	-0.5	1.73205
-2, -2, -3, -2	-2.25	0.50000
0, 4, 1, -2	0.75	2.50000

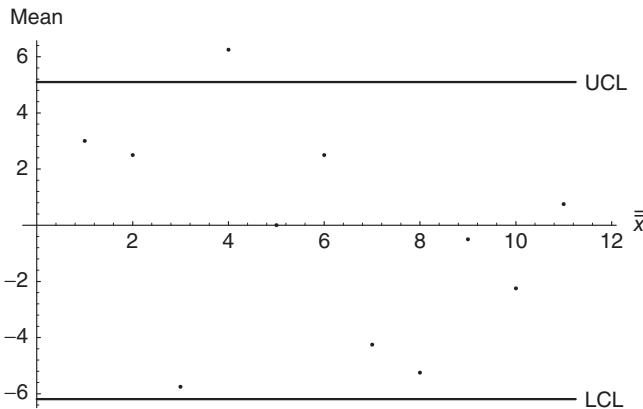


Figure 4.20 A control chart for sample means.

being used). Since the central limit theorem indicates that the means are approximately normal, this central value is taken as the mean of the sample means, $\bar{\bar{X}}$. Our second problem concerns the variation in the measurements. If σ^2 were known, then we could find a confidence interval for the true mean, μ , but we don't know this variance. Usually, an estimate of the confidence interval is found as

$$\bar{\bar{X}} \pm 3 \frac{\hat{\sigma}}{\sqrt{n}},$$

where $\hat{\sigma}$ is an estimate of the unknown standard deviation. The multiplier 3 is commonly used in industry and produces about 0.0027 of the data outside the limits. It is then fairly safe to assume that an observation outside the limits did not arise by chance, but is due to some alteration in the production process. The limits are called *upper and lower control limits*.

Now, how is σ estimated? There are many ways to do this; we show here a method based on the sample standard deviations.

We know that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2 \text{ and that } E(s^2) = \sigma^2.$$

and, with large samples, $E(s) \approx \sigma$, but this, unfortunately, is not true with small samples, so we will find an expression for $E(s)$ that will hold for any sample size. We find first the distribution of χ_{n-1} .

Suppose then that X is a χ_r^2 random variable so that

$$f(x) = \frac{1}{\Gamma\left[\frac{r}{2}\right] 2^{\frac{r}{2}}} x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}}, \quad x > 0 \text{ and let } Y = \sqrt{X}.$$

Then $G(y) = P(Y \leq y) = P(\sqrt{X} \leq y) = P(X \leq y^2) = F(y^2)$ so

$$g(y) = 2 \cdot y \cdot f(y^2) \text{ or}$$

$$g(y) = \frac{2}{\Gamma\left[\frac{r}{2}\right] 2^{\frac{r}{2}}} \cdot y^{r-1} \cdot e^{-\frac{y^2}{2}}, \quad y > 0 \text{ and}$$

$$E[Y] = \frac{2}{\Gamma\left[\frac{r}{2}\right] 2^{\frac{r}{2}}} \int_0^\infty y^r e^{-\frac{y^2}{2}} dy.$$

Letting $z = y^2/2$ in the integral,

$$E(Y) = \frac{\sqrt{2}}{\Gamma\left[\frac{r}{2}\right]} \int_0^\infty z^{\frac{r-1}{2}} e^{-z} dz = \frac{\sqrt{2}}{\Gamma\left[\frac{r}{2}\right]} \Gamma\left[\frac{r+1}{2}\right].$$

Now letting $r = n - 1$,

$$E(Y) = \frac{\sqrt{2}}{\Gamma\left[\frac{n-1}{2}\right]} \Gamma\left[\frac{n}{2}\right].$$

But $Y = \frac{\sqrt{n-1}s}{\sigma}$, so

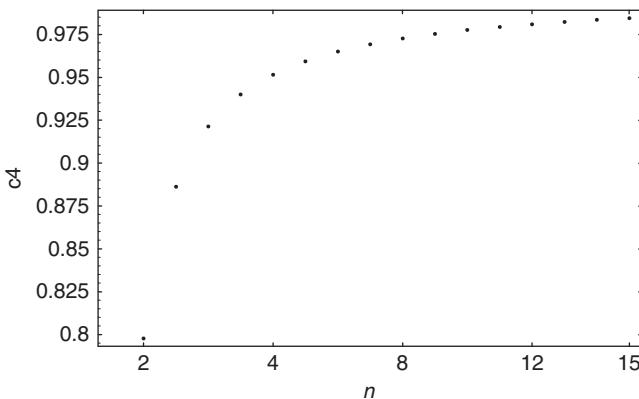
$$E(s) = \frac{\sqrt{2}}{\Gamma\left[\frac{n-1}{2}\right]} \Gamma\left[\frac{n}{2}\right] \cdot \frac{\sigma}{\sqrt{n-1}}.$$

The factor $\frac{\Gamma\left[\frac{n}{2}\right]}{\Gamma\left[\frac{n-1}{2}\right]} \cdot \frac{\sqrt{2}}{\sqrt{n-1}}$ is denoted by c_4 in quality control literature. A table of values of c_4 is shown in Table 1. While practical interest centers on small values for n , note that c_4 approaches 1 quite rapidly.

A graph of these values is shown in Figure 4.21.

Table 4.1

<i>n</i>	c4
2	0.797885
3	0.886227
4	0.921318
5	0.939986
6	0.951533
7	0.959369
8	0.965030
9	0.969311
10	0.972659
11	0.975350
12	0.977559
13	0.979406
14	0.980971
15	0.982316
16	0.983484
17	0.984506

**Figure 4.21** Factors c_4 for a quality control chart.

For the data in this example, $\bar{\bar{X}} = -0.54545$ and the average of the sample standard deviations is $\bar{s} = 3.46596$, so

$$\hat{\sigma} = \frac{3.46596}{0.921318} = 3.76196 \quad \text{giving control limits}$$

$$UCL = -0.5455 + 3 \cdot \frac{3.76196}{\sqrt{4}} = 5.0974 \quad \text{and}$$

$$LCL = -0.5455 - 3 \cdot \frac{3.76196}{\sqrt{4}} = -6.188.$$

These are the limits shown on the control chart in Figure 4.20. We see that the fourth sample has mean 6.25 which exceeds the upper control limit. Many manufacturers would investigate the production process at that point.

Another common method for estimating the process standard deviation is based on the mean range of the samples. One reason for using the range is that it is easily calculated on the production floor. Since the range ignores all of the sample except for two values, it is not surprising to find that it is not as efficient as the sample standard deviations for estimating σ .

We see that one value of the control chart is the glimpse it gives of the production process through the use of a sample, which usually is small. Many other types of control charts are used in industry. Interested readers are referred to Duncan [9] and Grant and Leavenworth [14] for more information.

EXERCISES 4.17

- 1.** Ten samples, shown below, were taken in order to establish control limits in an industrial process.

Sample	Values			
1	10.6	10.1	11.3	9.1
2	10.2	11.6	10.5	10.5
3	10.1	9.8	8.8	9.3
4	10.1	9.5	10.3	10.6
5	8.7	11.6	9.7	9.3
6	10.1	9.8	10.8	8.9
7	11.2	11.5	10.9	11.6
8	10.6	9.6	10.3	9.9
9	9.8	7.7	9.4	9.9
10	10.0	8.4	10.6	8.8

- (a) Calculate the sample mean and the sample standard deviation for each sample.
 (b) Calculate upper and lower control limits using the results in part (a).
 (c) Plot the control chart for the sample means. Are any of the data points unusual?
2. A new machine fills cereal boxes by weight. It is desired to start a control chart on the average weight of the boxes. Ten samples, each of size 5, are taken with the following results:

	Sample				
1	16.1	16.2	15.9	16.0	16.1
2	16.2	16.4	15.8	16.1	16.2
3	16.0	16.1	15.7	16.3	16.1
4	16.1	16.2	15.9	16.4	16.6
5	16.5	16.1	16.4	16.4	16.2
6	16.8	15.9	16.1	16.3	16.4
7	16.1	16.9	16.2	16.5	16.5
8	15.9	16.2	16.8	16.1	16.4
9	15.7	16.7	16.1	16.4	16.8
10	16.2	16.9	16.1	17.0	16.4

- (a) Calculate the sample mean and the sample standard deviation for each sample.
- (b) Calculate upper and lower control limits using the results in part (a).
- (c) Plot the control chart for the sample means. Are any of the data points unusual?

CHAPTER REVIEW

Drawing conclusions from samples—statistical inference—has been the subject of this chapter. Statistical inference is a central part of the scientific method since it involves the analysis of sample data gathered in the course of a scientific investigation.

Any quantity calculated from a sample is called a *statistic*. Statistics are thus random variables with probability distributions, means, and variances of their own. From the central limit theorem, we know that the mean of a sample, \bar{X} , has, approximately, a normal distribution with mean μ and variance σ^2/n , where n is the sample size and where μ and σ^2 are the true population values; this theorem is a primary tool in establishing tests of hypotheses on single means and on the difference between means from two populations.

The purpose of this chapter is to establish tests of hypotheses and confidence intervals for some common parameters of statistical distributions as well as to introduce the principle of least squares in fitting a linear function to a data set. In order to establish the probability distributions of some common statistics determined from samples, we first establish the probability distributions of some common *functions of random variables*. We begin with a procedure for finding the probability distribution function for a function of a random variable, say $Y = H(X)$. We used the fact that $g(y)$, the probability density function for Y , is

$$g(y) = \frac{dG(y)}{dy}$$

where $G(y) = P(Y \leq y) = P[H(X) \leq y]$.

We then solve for X and express $G(y)$ in terms of $F(x)$, the distribution function for X .

We established the fact that if $X \sim N(0, 1)$ then X^2 has a χ_1^2 distribution. We also established the distribution of the maximum of a set of uniformly distributed random variables.

The most important function of a random variable involves sums of independent random variables. In Section 4.4 we used the fact that

$$P(X + Y = z) = \sum_k P(X = x) \cdot P(Y = z - k).$$

This sum can be evaluated for a variety of probability distributions. Important facts here are that the sum of independent binomials with common value for p is binomial, and that the sum of independent Poissons is Poisson, where the parameter for the sum is the sum of the parameters of the individual Poissons. Binomial and Poisson variables are called *reproductive* since their sum has the same kind of distribution as the individual summands.

Variables are not commonly reproductive, however, and we calculated the probability distribution for the sum of independent uniform variables from some special cases. The sums appear to become normal; that is indeed the case, a fact that is established in subsequent parts of the chapter.

The probability distribution of a random variable can be summarized by the *probability generating function* for the random variable. If it exists, this function (which we denoted by $P_X(t)$) is the expected value of a special function of X , t^X , so

$$P_X(t) = E(t^X). \text{ It follows that}$$

$$P_X(t) = \sum_x t^x \cdot P(X = x)$$

since we used this function only for discrete random variables.

Sections 4.5 and 4.6 establish some properties of probability generating functions. These were supposing $A(t)$ and $B(t)$ to be probability generating functions for variables X and Y , respectively, that the coefficient of t^k in a power series expansion of $P_X(t)$ is $P(X = k)$. It is in this sense that $P_X(t)$ summarizes or characterizes the random variable since it is possible to find *all* the probabilities from $P_X(t) \cdot E(X) = P'_X(1)$, $Var(X) = P''_X(1) + P'_X(1) - [P'_X(1)]^2$. The coefficient of t^k in $A(t) \cdot B(t)$ is $P(X + Y = k)$.

This fact is of great importance in establishing probability distributions for sums.

Probability generating functions for some specific random variables are derived in Section 4.5. There it is found that

$$P_X(t) = (q + pt)^n \text{ for a binomial variable with parameters } n \text{ and } p, \text{ and}$$

$$P_X(t) = \frac{pt}{1 - qt} \text{ for a geometric random variable with parameter } p.$$

We then discussed a series of dependent binomial trials called *Poisson's trials* in which the probabilities in binomial trials vary.

The probability generating function is not often used for continuous random variables. Instead, a function we call the *moment generating function* is used. It too characterizes probability distributions. Its definition is

$$M[X; t] = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \, dx$$

provided of course that the integral exists. The moment generating function generates moments (although it is uncommon to use it for this purpose) in either of two ways: The coefficient of $t^k/k!$ in the power series expansion of $M[X; t]$ is $E(X^k)$. or

$$\frac{d^k M[X; t]}{dt^k} \Big|_{t=0} = E(X^k).$$

We calculated some specific moment generating functions:

$$\text{If } f(x) = 1, \quad 0 \leq x \leq 1, \quad \text{then } M[X; t] = \frac{1}{t}(e^t - 1).$$

$$\text{If } f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \text{then } M[X; t] = \frac{\lambda}{\lambda - t}.$$

$$\text{If } X \sim N(0, 1), \quad \text{then } M[X; t] = e^{\frac{t^2}{2}}.$$

This last fact takes on enormous importance since, if we can show that a moment generating function approaches that of the standard normal, we can conclude that the variable

approaches the standard normal distribution. In this sense the moment generating function is of great importance since it can be used to establish the distributions of functions of random variables, and, in particular, their sums.

These properties of moment generating functions are of importance:

$$1] M[c + X; t] = e^{ct} \cdot M[X; t] \quad \text{and}$$

$$2] M[cX; t] = M[X; ct].$$

These facts help establish the fact that if $X \sim N(\mu, \sigma)$ then $M[X; t] = e^{\mu t + \frac{t^2}{2} \cdot \sigma^2}$.

Sums of random variables are considered in Section 4.10; it is found that under quite general conditions these approach normality as the number of summands increases. Examples were shown that demonstrate that

Sums of normals are normal.

Sums of exponentials are normal.

Sums of binomials are normal.

These facts help explain the persistence of normality in many examples earlier in the book.

In Section 4.11 we state and demonstrate the *central limit theorem*: If X has mean μ and variance σ^2 , and if X has a moment generating function, then $\bar{X} \rightarrow N(\mu, \sigma/\sqrt{n})$.

Before examining confidence intervals and tests of hypotheses, we showed that the probability distribution of the sample variance can be determined from the fact that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

when the sample of size n is chosen from a normal distribution with known variance σ^2 .

The central limit theorem allows us to *test hypotheses* on single means. We also considered the distribution of the sample mean when the population variance is unknown and finally the distribution of the difference between sample means. The tests of various hypotheses considered are summarized here. Critical regions are presumed to be constructed so that the size of the test is α .

To test $H_0: \mu = \mu_o$ against $H_a: \mu \neq \mu_o$, the best critical region is $\bar{X} > a$ or $\bar{X} < b$. One-sided tests are used in testing one-sided alternatives. If the population standard deviation, σ , is known, then a and b can be determined using the fact that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is a normal variable.

To test $H_0: \mu = \mu_o$ against $H_a: \mu \neq \mu_o$ and the population standard deviation is unknown, then the best critical region is $\bar{X} > a$ or $\bar{X} < b$, where values for a and b can be determined using the fact that $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ follows a t_{n-1} distribution.

If two samples are drawn and both population variances are both known, and the hypothesis $H_0: \mu_x = \mu_y$ is to be tested against $H_a: \mu_x \neq \mu_y$ then the test statistic is

$$z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}},$$

where z is a normal random variable.

If the population variances are unknown but can be presumed to be equal and if the samples are chosen from normal populations then

$$\begin{aligned} t_v &= \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}, \quad \text{where} \\ s_p^2 &= \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} \quad \text{and} \\ v &= n_x + n_y - 2. \end{aligned}$$

If the population variances are unknown and known to be unequal then no exact test is known for the hypothesis that the population means are equal. An approximate test, due to Welch, is

$$\begin{aligned} T_v &= \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \quad \text{where} \\ v &= \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^2}{\frac{\left(\frac{s_x^2}{n_x} \right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y} \right)^2}{n_y - 1}}. \end{aligned}$$

A test of $H_o: \sigma_x^2 = \sigma_y^2$ and the alternative $H_a: \sigma_x^2 \neq \sigma_y^2$ is based on the fact that

$$\frac{\frac{s_x^2}{s_y^2}}{\frac{\sigma_x^2}{\sigma_y^2}} = F(n_x - 1, n_y - 1).$$

We then considered *simple linear regression*, or the fitting of data to a straight line of the form $y_i = a + bx_i, i = 1, 2, \dots, n$. The principle of least squares chooses those estimates that minimize

$$S = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

The result is a set of *least squares equations*:

$$\sum_{i=1}^n y_i = n\hat{a} + \hat{b} \sum_{i=1}^n x_i \quad \text{and}$$

$$\sum_{i=1}^n x_i y_i = \hat{a} \sum_{i=1}^n x_i + \hat{b} \sum_{i=1}^n x_i^2.$$

Their simultaneous solution is:

$$\begin{aligned}\hat{b} &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{and} \\ \bar{y} &= \hat{a} + \hat{b} \bar{x}.\end{aligned}$$

Finally in this chapter we considered *quality control charts* for sample means. This chart plots means calculated from periodic samples and establishes *upper and lower control limits*, indicating that the process may be out of control. These limits are $\bar{X} \pm 3 \frac{\hat{\sigma}}{\sqrt{n}}$ where $\hat{\sigma}$ is an estimate of the unknown population standard deviation, σ . Table 1 gives divisors of the average sample standard deviations which are used to find $\hat{\sigma}$.

PROBLEMS FOR REVIEW

Exercises 4.3 # 1, 3, 4, 8.

Exercises 4.4 # 1, 2, 4.

Exercises 4.5 # 1, 4.

Exercises 4.7 # 1, 2, 4, 6, 7, 10.

Exercises 4.8 # 2, 3, 4, 5, 6, 9.

Exercises 4.10 # 1, 2, 5, 9.

Exercises 4.11 # 1, 2, 4, 5

Exercises 4.13 # 1, 2, 5, 7

Exercises 4.14 # 1, 2, 5, 6, 8, 10, 17

Exercises 4.15 # 1, 2, 3, 6, 8, 9, 11, 14, 15, 19

Exercises 4.16 # 1, 3, 5, 9.

Exercises 4.17 # 1

SUPPLEMENTARY EXERCISES FOR CHAPTER 4

1. For the triangular distribution $f(x) = \frac{2}{a} \left(1 - \frac{x}{a}\right)$, $0 < x < a$,

(a) Find the moment generating function.

(b) Use the moment generating function to find the mean and variance of X and check these results by direct calculation.

2. Find the mean and variance of X where X has the Pareto distribution, $f(x) = a \cdot b^a \cdot x^{-(a+1)}$, $a > 0$, $b > 0$, $x > b$.
3. Consider the truncated exponential distribution, $f(x) = e^x$, $0 \leq x \leq \ln 2$.
 - (a) Find the moment generating function for X and expand it in a power series.
 - (b) From the series in part a], find the mean and variance of X .
4. Random variable X denotes the number of green marbles drawn when a sample of two is selected without replacement from a box containing 3 green and 7 yellow marbles.
 - (a) Find the moment generating function for X .
 - (b) Verify that $E(X^3) = 1$.
5. Find $E(X^k)$ if X is a Weibull random variable with parameters α and β .
6. Let $S = \sum_{i=1}^n X_i$, where X_i is a uniform random variable on the interval $(0,1)$. Find the moment generating function for $Z = \frac{S-\mu}{\sigma}$ where μ and σ are the mean and standard deviation, respectively, of S . Then show that this moment generating function approaches the moment generating function for a standard normal random variable as $n \rightarrow \infty$.
7. A fair quarter is tossed until it comes up heads; suppose X is the number of tosses necessary. If $X = x$, then x fair pennies are tossed; let Y denote the number of heads on the pennies. Find $P(Y = 3)$, simplifying the result as much as you can.
8. The coin loaded so as to come up heads 1/3 of the time is tossed until a head appears. This is followed by the toss of a coin loaded so as to come up heads with a probability of 1/4 until that coin comes up heads.
 - (a) Find the probability distribution of Z , the total number of tosses necessary.
 - (b) Find the mean and variance of Z .
9. Customers at a gasoline station buy regular or premium unleaded gasoline with probabilities p and $q = 1 - p$, respectively. The number of customers in a daily period is Poisson with mean μ . Find the probability distribution for the number of customers buying regular unleaded gasoline.
10. A company claims that the actual resistance of resistors are normally distributed with mean 200 ohms and variance $4 \cdot 10^{-4}$ ohms².
 - (a) What is the probability that a resistor drawn at random from this set of resistors will have resistance greater than 200.025 ohms?
 - (b) A sample of 25 resistors drawn at random from this set has an average resistance of 200.01 ohms. Would you conclude that the true population mean is still 200 ohms?
11. A sample of size n is drawn from a population about which nothing is known except that the variance is 4. How large a sample must be drawn so that the probability is at least 0.95 that the sample average, \bar{X} , is within 1 unit of the true population mean, μ ?
12. Suppose 12 fair dice are thrown. Let X denote the total number of spots showing on the 12 uppermost faces. Use the central limit theorem to estimate $P(25 \leq X \leq 40)$.
13. Mathematical and verbal SAT scores are, individually, $N(500, 100)$.
 - (a) Find the probability that the total mathematical plus verbal SAT score for an individual is at least 1100, assuming that the scores are independent.
 - (b) What is the probability that the average of five individual total scores is at least 1100?

- 14.** A student makes 100 check transactions in a period covering his bank statement. Rather than subtract the amount he spends exactly, he rounds each checkbook entry off to the nearest dollar. Assume that the errors are uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. What is the probability the total error is more than \$5?
- 15.** The time a construction crew takes to construct a building is normally distributed with mean 90 days and standard deviation 10 days. After construction, it takes additional time to install utilities and finish the interior. Assume the additional time is independent of the construction time, and is normally distributed with mean 30 days and standard deviation 5 days.
- Find the probability it takes at least 101 days for the construction of only a building.
 - Find the probability it takes an average of 101 days for the construction of only four buildings.
 - What is the probability that the total completion time for one building is at most 130 days?
 - What is the probability that the average additional completion time for five buildings is at least 35 days?
- 16.** A random variable X has the probability distribution function
- $$f(x) = \begin{cases} \frac{1}{3} & \text{for } x = -1, 0, \text{ or } 1. \\ 0 & \text{otherwise.} \end{cases}$$
- Find $M[X; t]$, the moment generating function for X .
 - If X_1 and X_2 are independent observations of X , find $M[X_1 + X_2; t]$ without first finding the probability distribution of $X_1 + X_2$.
 - Verify the result in part (b) by finding the probability distribution of $X_1 + X_2$.
- 17.** A random variable X has the probability density function $f(x) = 2(1-x)$, $0 < x < 1$.
- Find the moment generating function for X .
 - Use the moment generating function to find a formula for $E(X^k)$.
 - Let $Y = \frac{1}{2}(X+1)$. Find $M[Y; t]$.
- 18.** A random variable X has $M[X; t] = e^{-6t+32t^2}$. Find $P(-4 \leq X \leq 16)$.
- 19.** A discrete random variable X has the probability distribution function
- $$f(x) = \begin{cases} \frac{1}{2} & x = 1 \\ \frac{1}{3} & x = 2 \\ \frac{1}{6} & x = 3. \\ 0 & \text{otherwise.} \end{cases}$$
- Find μ_x and σ_x^2 .
 - Find $M[X; t]$.
 - Verify the results in part (a) using the moment generating function.
- 20.** Find the moment generating function for a random variable X with probability density function $f(x) = e - e^x$ if $0 \leq x \leq 1$.

21. A random variable has $M[X; t] = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t}$.
- What is the probability distribution function for X ?
 - Expand $M[X; t]$ in a power series and find μ_x and σ_x^2 .
22. Suppose that X_1 is the number of 6's in n_1 tosses of a fair die and that Y is the number of 3's in n_2 tosses of another fair die. Use moment generating functions to show that $S = X + Y$ has a binomial distribution with parameters $n = n_1 + n_2$ and $p = 1/6$.
23. A square law rectifier has the characteristic $Y = kX^2$, $x > 0$, where X and Y are the input and output voltages, respectively. If the input to the rectifier is noise with the probability density function

$$f(x) = \frac{2x}{\beta} e^{-x^2/\beta}, \quad x \geq 0, \quad \beta > 0,$$

find the probability density function of the output.

24. If $X \sim N(0, 1)$,
- find $P(X^2 \geq 5)$.
 - Suppose X_1, X_2, \dots, X_8 are independent observations of X . Find $P(X_1^2 + X_2^2 + \dots + X_8^2 > 10)$.
25. Suppose X has the probability density function

$$f(x) = \begin{cases} x+1, & -1 \leq x \leq 0 \\ 1-x, & 0 \leq x \leq 1. \end{cases}$$

- Find the probability density function of $Y = X^2$.
 - Show that the result in part (a) is a probability density function.
26. The resistance, R , of a resistor has the probability density function

$$f(r) = \frac{r}{200} - 1, \quad 200 < r < 220.$$

A fixed voltage of 5v is placed across the resistor.

- Using the fact that $V = I \cdot R$, find the probability density function of the current, I , through the resistor.
 - What is the expected value of the current?
27. If X is uniformly distributed on $[-1, 1]$, find the probability density function of $Y = \sqrt{1 - X^2}$.
28. Let X be uniformly distributed on $[0, 1]$.
- Find the probability density function for $Y = \frac{1}{X+1}$ and prove that your result is a probability density function.
 - Explain how values of X could be used to sample from the distribution $f(x) = 2x$, $0 < x < 1$.
29. Random variable X has the probability density function $f(x) = 2x$, $0 < x < 1$. Let $Y = \frac{1}{X}$. Find $E(Y)$ by
- first finding $g(y)$.
 - not using $g(y)$.

- 30.** Given $f(x) = 2e^{-2x}$, for $x > 0$. Find the probability density function for $Y = e^{-X}$ and, from it, find $E[e^{-X}]$.
- 31.** A random variable X has the probability density function $f(x) = \frac{2}{9}x(3-x)$, for $0 \leq x \leq 3$. Find the probability density function for $Y = X^2 - 1$.
- 32.** The moment generating function for a random variable Y is $M[Y; t] = \frac{e^{4t}-e^{2t}}{2t}$. Expand $M[Y; t]$ in a power series in t and find μ_y and σ_y^2 .
- 33.** Suppose that X is uniform on $[-1, 2]$. Find the probability density function for $Y = |X|$.
- 34.** The following data represent radiation readings, in milliroentgens per hour, taken from television display areas in different department stores: 0.40, 0.48, 0.60, 0.15, 0.50, 0.80, 0.50, 0.36, 0.16, and 0.89.
 - Find a 95% confidence interval for μ if it is known that $\sigma^2 = 1$.
 - Find a 95% confidence interval for μ if σ^2 is unknown.
- 35.** The variance of a normally distributed industrial measurement is known to be 225. If a random sample of 14 measurements is taken and the sample variance computed, what is the probability the sample variance is twice the true variance?
- 36.** A random sample of 21 observations is taken from a normal distribution with variance 100. What is the probability the sample variance exceeds 140?
- 37.** A machine that produces ball bearings is sampled periodically. The mean diameter of the ball bearings produced is known to be under control, but the variability of these diameters is of concern. If the machine is working properly, the variance is 0.50 mm^2 . If a sample of 31 measurements shows a sample variance of 0.94 mm^2 , should the operator of the machine be concerned that something is wrong with the machine? Use $\alpha = 0.05$.
- 38.** A manufacturer of piston rings for automobile engines assumes that piston ring diameter is approximately normally distributed. If a random sample of 15 rings has mean diameter 74.036 mm and sample standard deviation 0.008 mm, construct a 98% confidence interval for the true mean piston ring diameter.
- 39.** A commonly used method for determining the specific heat of iron has a standard deviation 0.0100. A new method of determination yielded a standard deviation of 0.0086 based on nine test runs. Assuming a normal distribution, is there evidence at the 10% level that the new method reduces the standard deviation?
- 40.**
 - A random sample of 10 electric light bulbs is selected from a normal population. The standard deviation of the lifetimes of these bulbs is 120 hours. Find 95% confidence limits for the variance of all such bulbs manufactured by the company.
 - Find 95% confidence limits for the standard deviation if the sample size is 100.
- 41.** A city draws a random sample of employees from its labor force of 5000 people. The number of years each employee has worked for the city is 8.2, 5.6, 4.7, 9.6, 7.8, 9.1, 6.4, 4.2, 9.1 and 5.6. Assume that the time employees have been employed is approximately normal. Calculate a 90% confidence interval for the average number of years an employee has worked for the city.
- 42.** The number of ounces of liquid a soft drink machine dispenses into a bottle is a normal random variable with unknown mean μ but known variance 0.25 oz^2 . A random sample of 75 bottles filled by this machine has mean 12.2 oz.

- (a) Determine a 95% two-sided confidence interval for μ .
 (b) It is desired to be 99% confident that the error in estimating the mean is less than 0.1 oz. What should the sample size be?
43. The maximum acceptable level for exposure to microwave radiation in the United States is an average of $10 \text{ } \mu\text{W}/\text{cm}^2$. It is feared that a large television transmitter may be polluting the air by exceeding a safe level of microwave radiation.
- (a) Test $H_0: \mu = 10$ against the alternative hypothesis $H_a: \mu > 10$ with $\alpha = 0.05$ if a sample of 36 readings gives a sample mean of $10.3 \text{ } \mu\text{W}$ and a sample standard deviation of $2.1 \text{ } \mu\text{W}$.
 (b) Find a 98% confidence interval for μ .
44. A machine producing washers is found to produce washers whose variance is 30 in^2 .
- (a) A sample of 36 washers is taken and the mean diameter, \bar{X} is found. Find the probability that \bar{X} is within 0.1 units of μ , the true mean diameter.
 (b) How large a sample is necessary so that the probability \bar{X} is within 0.2 units of μ is 0.90?
45. (a) To determine with 94% confidence the average hardness of a large number of selenium-alloy ball bearings, how many would have to be tested to obtain an estimate within 0.009 units of the true mean hardness if σ^2 is known to be 0.0016?
 (b) A small study of five bearings in part (a) gives $\bar{X} = 2.057$. What is the probability this differs from the true mean hardness by at least 0.009 units?
46. Heat transfer coefficients of 65, 63, 60, 68, and 72 were observed in a sample of heat exchangers made by a company. Find a 95% confidence interval for the true average heat transfer coefficient, μ , if
- (a) σ^2 is known to be 17.64.
 (b) σ^2 is unknown.
47. Find the probability that a random sample of 25 observations from a normal population with variance 6 will have a sample variance between 3.100 and 10.750.
48. The hardness (in degrees) of a certain rubber is claimed to be 65. A sample of 14 specimens gave $\bar{X} = 63.1$.
- (a) If σ^2 is known to be 12.25 degrees² for this rubber can $H_0: \mu = 65$ be accepted against the alternative hypothesis $H_a: \mu \neq 65$ if $\alpha = 5\%$?
 (b) Answer part (a) if the sample variance is 10.18 degrees².
49. A manufacturer of steel rods considers that the process is working properly if the mean length of the rods is 8.6 in. The standard deviation of these rods is approximately 0.3 in. Suppose that when 36 rods were tested, the sample mean was 8.45.
- (a) Test the hypothesis that the average length is 8.6 in. against the alternative that it is less than 8.6 in., using a 5% level of significance.
 (b) Since short rods must be scrapped, it is extremely important to know when the process began to produce rods of mean length less than 8.6. Find the probability of a Type II error when the alternative hypothesis is $H_a: \mu = 8.4$ in.
50. A coffee vending machine is supposed to dispense 6 oz per cup. The machine is tested nine times yielding an average fill, $\bar{X} = 6.1$ oz with standard deviation 0.15 oz.
- (a) Find a 90% confidence interval for μ , the true mean fill per cup.
 (b) Find a 90% confidence interval for σ^2 , the true variance of the fill per cup.

- 51.** A random sample of 22 freshman mathematics SAT scores at a large university has sample mean 680 and standard deviation 35. Find a 99% confidence interval for μ , the true population mean.
- 52.** A population has unknown mean μ but known standard deviation of 5. How large a sample is necessary so that we can be 95% confident that \bar{X} is within 1.5 units of the true mean?
- 53.** A fuel oil company claims that 20% of the homes in a city are heated by oil. Do we have reason to doubt this claim if 236 homes in a sample of 1000 homes are heated by oil? Use $\alpha = 1\%$.
- 54.** A brand of car battery claims that the standard deviation of the battery's lifetime is 0.9 years. If a random sample of 10 of these batteries has $s = 1.2$, test $H_o: \sigma^2 = 0.81$ against the alternative hypothesis, $H_a: \sigma^2 > 0.81$, if $\alpha = 0.05$.
- 55.** A researcher is studying the weights of male college students. She wishes to test $H_o: \mu = 68$ kg against the alternative hypothesis $H_a: \mu \neq 68$ kg. A sample of 64 students has $\bar{X} = 68.90$ and $s = 4$ kg.
- Is the hypothesis accepted or rejected?
 - Find β for the alternative $\mu = 69.3$ kg.
- 56.** Fractures in metals have been studied and it is thought that the rate at which fractures expand is normally distributed. A sample of 14 pieces of a particular steel gave $\bar{X} = 3205$ ft/s.
- Find a 95% confidence interval for μ , the true average rate of expansion, if σ is assumed to be 53 ft/s.
 - Now suppose σ is unknown. The sample variance is 6686.53 (ft/s)². Find a 95% confidence interval for μ .
- 57.** Engineers think that a design change will improve the gasoline mileage of a certain brand of automobile. Previously such cars averaged 18 mpg. under test conditions. A sample of 15 cars has $\bar{X} = 19.5$ mpg.
- Test $H_o: \mu = 18$ against the alternative hypothesis $H_a: \mu > 18$ assuming $\sigma^2 = 9$ and $\alpha = 5\%$.
 - Test the hypothesis in part a) at the 5% level if the sample variance is 7.4.
- 58.** One-hour carbon monoxide concentrations in 10 air samples from a city had mean 11.5 ppm and variance 40 (ppm)². After imposing smog control measures on a local industry, 12 air samples had mean 10 ppm and variance 43 (ppm)². Estimate the true difference in average carbon monoxide concentrations in a 98% confidence interval. What assumptions are necessary for your answer to be valid?
- 59.** Specifications for a certain type of ribbon call for a mean breaking strength of 185 lb. In order to monitor the process, a random sample of 30 pieces, selected from different rolls, is taken each hour and the sample mean used to decide if the mean breaking strength has shifted. The test then is of the hypothesis $H_o: \mu = 185$ against the alternative hypothesis $H_a: \mu < 185$ with $\alpha = 0.05$. Assuming $\sigma = 10$ lb,
- Find the critical region in terms of \bar{X} .
 - Find β for the alternative $\mu = 179.5$.

60. To test $H_0: \mu = 46$ against the alternative hypothesis $H_a: \mu > 46$, a random sample of 24 is taken. The critical region is $\bar{X} > 51.7$.
- Find α .
 - Find β for the alternative $\mu = 48$.
61. In 16 test runs the gasoline consumption of an experimental engine had sample standard deviation 2.2 gallons. Construct a 95% confidence interval for σ , the true standard deviation of gasoline consumption of the engine. What assumptions are necessary for your analysis to be valid?
62. A production supervisor wants to determine if changes in a production process reduce the amount of time necessary to complete a subassembly. Specifically she wishes to test $H_0: \mu = 30$ against the alternative hypothesis, $H_a: \mu < 30$, with $\alpha = 5\%$. The measurements are in minutes.
- Find the critical region for the test (in terms of \bar{X}) if a sample of four times is taken and the true variance is assumed to be 1.2.
 - Now suppose a sample gave $\bar{X} = 29.06$ and $s^2 = 1.44$. Is the hypothesis accepted or not?

Chapter 5

Bivariate Probability Distributions

5.1 INTRODUCTION

So far we have studied a single random variable defined on the points of a sample space. Scientific investigations, however, most commonly involve several random variables arising in the course of an investigation. A physicist, for example, may be interested in studying the effects of transmissions in a fiber optic cable when transmission rates and the composition of the cable are varied; sample surveys usually ask several questions of the respondents creating separate random variables for each question; educators studying grade point averages for college students find that these averages are dependent on intelligence, entrance examinations, rank in high school class, as well as many other factors that could be considered. Each of these examples suggests a sample space on which more than one random variable is defined.

While these variables could be considered individually as the *univariate* variables studied in the previous chapters, studies of the individual random variables will provide no information at all on how the variables behave together. Separate studies then offer no information on how the variables *interact* or are *correlated* with each other; this is often crucial information in scientific investigations, since the manner in which the variables act together may indicate the most important factors in explaining the outcome. Because of this, investigations involving only one factor at a time are becoming increasingly rare. The interactions revealed in studies are often of greater importance than the effects of the individual variables alone, but measuring them requires that we consider combinations of the variables together. In this chapter, we will study *jointly distributed random variables* and some of their characteristics. This is an essential prelude to the actual measurement of the influence of separate variables and interactions. Inferences from these measurements are statistical problems that are normally discussed in texts on statistics.

5.2 JOINT AND MARGINAL DISTRIBUTIONS

Example 5.2.1

In Example 2.9.1, we considered tossing two fair coins and recording X , the number of heads that occur. The coins that come up heads are put aside and only those that come up tails the first time are tossed again. Let Y denote the number of heads obtained in the second set of tosses. The variable Y is of primary interest here, but to investigate it we must

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

consider X as well. Although this might appear to be a purely theoretical exercise, the result is applicable when a number of components in a system fail according to a binomial model; interest centers on when all the components will fail, so our example is a generalization of this situation. We use five coins here and only two group tosses (so that it may be that not all the coins will turn up heads), but the extension to any number of coins is very similar to this special case.

Y is clearly dependent on X . In fact, since $5 - X$ coins came up tails the first time and were then tossed again by a binomial process, it follows that if $X = x$, then the conditional probability that $Y = y$ is given by a binomial probability:

$$P(Y = y|X = x) = \binom{5-x}{y} \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right)^{5-x-y}, \quad y = 0, 1, \dots, 5-x.$$

X itself is also a random variable and so the unconditional probability that $X = x$ is

$$P(X = x) = \binom{5}{x} \left(\frac{1}{2}\right)^5, \quad x = 0, 1, \dots, 5.$$

If we call

$$f(x, y) = P(X = x \text{ and } Y = y),$$

which we also denote as

$$f(x, y) = P(X = x, Y = y)$$

the *joint probability distribution* of X and Y , then

$$f(x, y) = P(X = x, Y = y) = P(X = x) \cdot P(Y = y|X = x),$$

where $P(Y = y|X = x)$ is the conditional probability that $Y = y$ if $X = x$.

In this example, the conditional probability $P(Y = y|X = x)$ is also binomial with $5 - x$ trials and probability of success at any trial $1/2$, as we have seen, so

$$f(x, y) = \binom{5}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} \cdot \binom{5-x}{y} \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right)^{5-x-y}, \quad x = 0, 1, \dots, 5; y = 0, 1, \dots, 5-x,$$

which can be simplified to

$$f(x, y) = \binom{5}{x} \binom{5-x}{y} \left(\frac{1}{2}\right)^{10-x}, \quad x = 0, 1, \dots, 5; y = 0, 1, \dots, 5-x.$$

These probabilities are exhibited in Table 5.1.

Table 5.1 Joint distribution for the coin tossing example

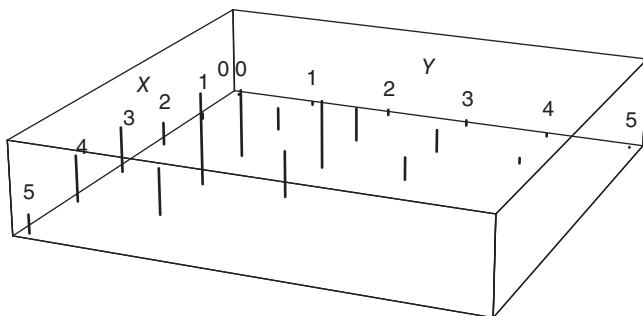
X	Y	0	1	2	3	4	5	$f(x)$
0		$\frac{1}{1024}$	$\frac{5}{1024}$	$\frac{10}{1024}$	$\frac{10}{1024}$	$\frac{5}{1024}$	$\frac{1}{1024}$	$\frac{32}{1024}$
1		$\frac{10}{1024}$	$\frac{40}{1024}$	$\frac{60}{1024}$	$\frac{40}{1024}$	$\frac{10}{1024}$	0	$\frac{160}{1024}$
2		$\frac{40}{1024}$	$\frac{120}{1024}$	$\frac{120}{1024}$	$\frac{40}{1024}$	0	0	$\frac{320}{1024}$
3		$\frac{80}{1024}$	$\frac{160}{1024}$	$\frac{80}{1024}$	0	0	0	$\frac{320}{1024}$
4		$\frac{80}{1024}$	$\frac{80}{1024}$	0	0	0	0	$\frac{160}{1024}$
5		$\frac{32}{1024}$	0	0	0	0	0	$\frac{32}{1024}$
	$g(y)$	$\frac{243}{1024}$	$\frac{405}{1024}$	$\frac{270}{1024}$	$\frac{90}{1024}$	$\frac{15}{1024}$	$\frac{1}{1024}$	1

Notice that the entries in the table must all be nonnegative (since they represent probabilities), and that the sum of these entries is 1.

Probabilities can be found from the table. For example,

$$P(X \geq 2, Y \geq 2) = 120/1024 + 80/1024 + 40/1024 = 15/64.$$

A scatter plot of the joint probability distribution is also useful (see Figure 5.1).

**Figure 5.1** Scatter plot for the coin tossing example.

Now suppose we want to recover information on the variables X and Y separately. These, individually, are random variables on their own. What are their probability distributions?

To find $P(X = 3)$, for example, since the three events $X = 3$ and $Y = 0$; $X = 3$ and $Y = 1$; and $X = 3$ and $Y = 2$ are mutually exclusive, we see that

$$\begin{aligned}
P(X = 3) &= P(X = 3, Y = 0) + P(X = 3, Y = 1) + P(X = 3, Y = 2) \\
&= \binom{5}{3} \binom{2}{0} \left(\frac{1}{2}\right)^7 + \binom{5}{3} \binom{2}{1} \left(\frac{1}{2}\right)^7 + \binom{5}{3} \binom{2}{2} \left(\frac{1}{2}\right)^7 \\
&= \binom{5}{3} \left(\frac{1}{2}\right)^7 \left\{ \binom{2}{0} + \binom{2}{1} + \binom{2}{2} \right\} \\
&= \binom{5}{3} \left(\frac{1}{2}\right)^7 2^2 = \binom{5}{3} \left(\frac{1}{2}\right)^5 = 10/32 = 5/16.
\end{aligned}$$

We find this probability entered in the side margin of the table at $X = 3$. It is found by adding the probabilities across the row, thus considering all the possible values of Y when $X = 3$.

Other values of $P(X = x)$ could be found in a similar manner and so we make the following definition.

Definition: The *marginal distribution* of X is given by

$$P(X = x) = \sum_y P(X = x, Y = y),$$

where the sum is over all possible values of y .

We denote $P(X = x)$ by $f(x)$. The term *marginal distribution of X* arises since the distribution occurs in the margin of the table.

So,

$$f(x) = P(X = x) = \sum_y P(X = x, Y = y)$$

where the sum is over all possible values of y .

To find $f(x)$ then in this example, we must calculate

$$\begin{aligned}
f(x) &= \sum_{y=0}^{5-x} f(x, y) = \sum_{y=0}^{5-x} \binom{5}{x} \binom{5-x}{y} \left(\frac{1}{2}\right)^{10-x} \\
&= \binom{5}{x} \left(\frac{1}{2}\right)^{10-x} \cdot \sum_{y=0}^{5-x} \binom{5-x}{y} = \binom{5}{x} \left(\frac{1}{2}\right)^{10-x} \cdot 2^{5-x} \\
\text{so } f(x) &= \binom{5}{x} \left(\frac{1}{2}\right)^5, x = 0, 1, 2, \dots, 5.
\end{aligned}$$

This verifies that X is binomial with $n = 5$ and $p = 1/2$.

If we denote the marginal distribution of Y by $g(y)$, then, reasoning in the same way as we did for $f(x)$, we conclude that

$$g(y) = P(Y = y) = \sum_x P(X = x, Y = y)$$

where the sum is over all values of x .

The functions $f(x)$ and $g(y)$ are given in the margins in Table 5.1.

In this case it is not so easy to see the pattern in the distribution of Y , but there is one. First, by the definition of $g(y)$,

$$P(Y = y) = g(y) = \sum_{x=0}^5 f(x, y) = \sum_{x=0}^5 \binom{5}{x} \binom{5-x}{y} \left(\frac{1}{2}\right)^{10-x},$$

and this can be written as

$$g(y) = \sum_{x=0}^5 \binom{5}{y} \binom{5-y}{x} \left(\frac{1}{2}\right)^{10-x}.$$

Now we remove common factors, rearrange, and insert the factor 1^{5-y-x} to find that we can write $g(y)$ as

$$\begin{aligned} g(y) &= \binom{5}{y} \left(\frac{1}{2}\right)^{10} \sum_{x=0}^5 \binom{5-y}{x} 2^x \cdot 1^{5-y-x} \\ &= \binom{5}{y} \left(\frac{1}{2}\right)^{10} (2+1)^{5-y} \end{aligned}$$

by the binomial theorem. It follows that

$$g(y) = \binom{5}{y} \cdot \left(\frac{1}{2}\right)^{10} \cdot 3^{5-y}, \quad y = 0, 1, \dots, 5.$$

Some characteristics of the distribution of Y may be of interest. A graph of its values is shown in Figure 5.2.

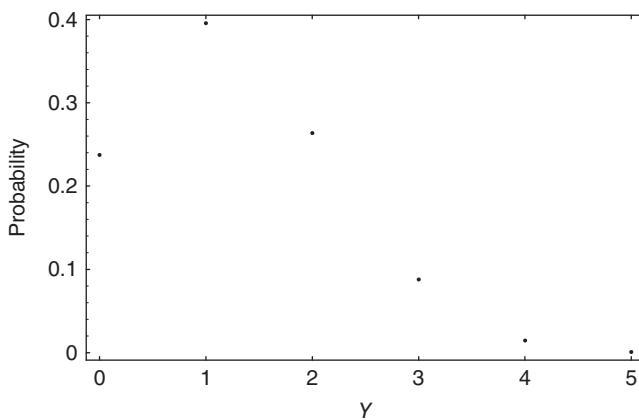


Figure 5.2 Marginal distribution for Y in Example 5.2.1.

Finally, we find $E(Y)$, the expected number of heads as a result of this experiment. A computer algebra system will evaluate $E(Y) = \sum_y y \cdot g(y) = 5/4$. This also has an intuitive interpretation. One can argue that as a result of the first set of tosses, $5/2$ coins are expected to be tails and of these $1/2$ can be expected to result in heads on the second set of tosses,

producing $5/4$ as $E(Y)$. Note that by this argument, $E(Y)$ was found without using the probability distribution for Y . It is often possible, and on occasion desirable, to do this. We will give this process more validity later in this chapter. Now we consider a continuous example.

Example 5.2.2

An investigator, intending to make a certain type of steel stronger, is examining the content of the steel. He considers adding carbon (X) and molybdenum (Y) to the steel and measuring the resulting strength. However, the carbon and molybdenum interact in a complex way in the steel being considered so the investigator takes some data by varying the values of X and Y (whose values have been coded here for convenience). He finds that the resulting strength of the steel can be approximated by the function

$$f(x, y) = x^2 + \left(\frac{8}{3}\right)xy \text{ for } 0 < x < 1 \text{ and } 0 < y < 1.$$

A graph of this surface is shown in Figure 5.3.

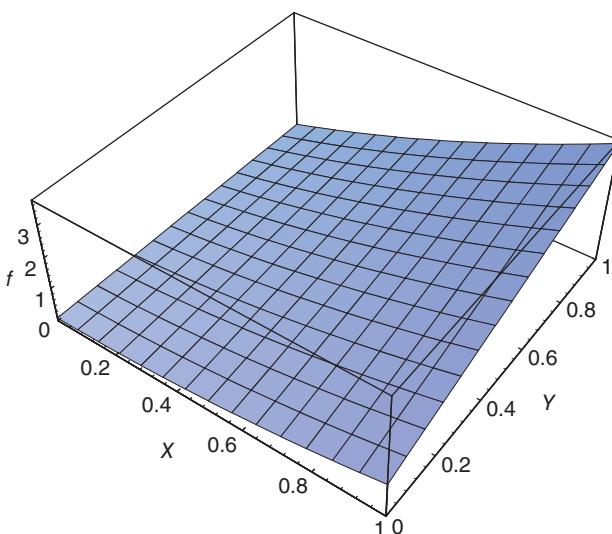


Figure 5.3 Surface for Example 5.2.2.

We find that

$$\int_0^1 \int_0^1 f(x, y) dy dx = 1 \text{ and that } f(x, y) \geq 0.$$

Because of these two facts, and in analogy with univariate probability densities, we call $f(x, y)$ a *continuous bivariate probability density function*. Note again the distinction between discrete probability *distributions* and continuous probability *densities*.

Rather than sum, as we did in the discrete example, we integrate to find the marginal densities.

We let

$$f(x) = \int_y f(x, y) dy$$

and $g(y) = \int_x f(x, y) dx$

providing, of course, that the integrals exist.

In this case,

$$f(x) = \int_0^1 (x^2 + (8/3)xy) dy = x^2 + \frac{4x}{3}, \quad 0 < x < 1$$

and

$$g(y) = \int_0^1 (x^2 + (8/3)xy) dx = \frac{4y+1}{3}, \quad 0 < y < 1.$$

Graphs of these probability densities are shown in Figures 5.4 and 5.5:

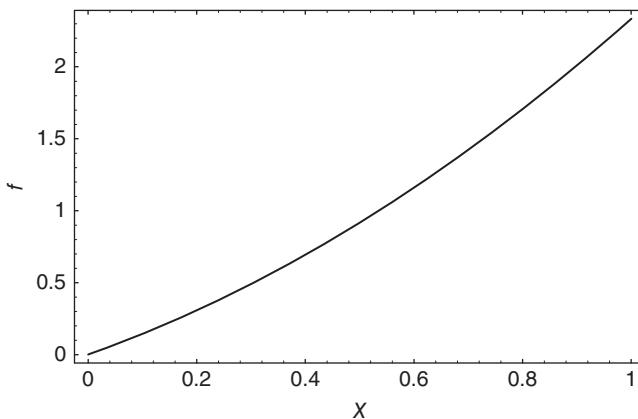


Figure 5.4 Marginal distribution for X , Example 5.2.2.

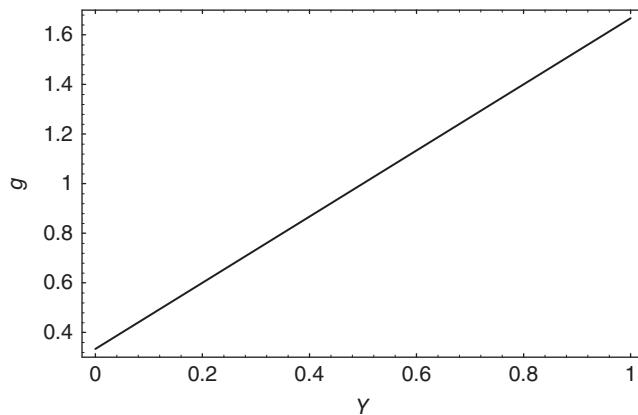


Figure 5.5 Marginal distribution for Y , Example 5.2.2.

We can verify that each of these is a univariate probability density. In Example 5.2.1, we found $E(Y)$ rather simply and without making use of the probability density for Y alone. In this case, it is easy to verify that $E(Y) = 11/18$, but it is not so easy to see how we would find this without finding $g(y)$ first. We will show how this can be done in Section 5.4.

Example 5.2.3

Since the volume under the bivariate probability density function is 1, parts of that volume represent probabilities. In this example, we find that

$$P\left(X > \frac{1}{2}, Y < \frac{2}{3}\right) = \int_{\frac{1}{2}}^1 \int_0^{\frac{2}{3}} \left(x^2 + \frac{8}{3}xy\right) dy dx = \frac{5}{12}.$$

We can also compute more complex probabilities, such as $P(X > Y)$. To calculate this we must integrate over the triangular region in the sample space where $X > Y$. This gives

$$\begin{aligned} P(X > Y) &= \int_0^1 \int_0^x x^2 + \frac{8}{3}xy \, dy \, dx \\ &= \int_0^1 x^2y + \frac{4xy^2}{3} \Big|_0^x \, dx \\ &= \int_0^1 \frac{7}{3}x^3 \, dx = \frac{7}{12}. \end{aligned}$$

EXERCISES 5.2

1. Verify $E[Y]$ in Example 5.2.2.
2. Find $\text{Var}[X]$ and $\text{Var}[Y]$ in Example 5.2.2.
3. An engineering college has made a study of the grade point averages of graduating engineers, denoted by the random variable Y . It is desired to study these as a function of high school grade point averages, denoted by the random variable X . The following table shows the joint probability distribution where the grade point averages have been combined into five categories for each variable.

		X				
		2.0	2.5	3.0	3.5	4.0
Y	2.0	0.05	0	0.01	0	0
	2.5	0.10	0.04	0	0.01	0
	3.0	0.02	0.10	0.05	0.10	0.01
	3.5	0	0	0.10	0.20	0.10
	4.0	0	0	0.05	0.02	0.05

- (a) Find the marginal distributions for X and Y .
- (b) Find $E(X)$ and $E(Y)$.
- (c) Find $P(X \geq 3, Y \geq 3)$.

4. A random sample of 6 items is drawn from a plant's daily production. Let the random variables X and Y denote the number of good items and the number of defective items chosen, respectively. If the production contains 40 good and 10 defective items, find
- the joint probability distribution function for X and Y .
 - the marginal distributions of X and Y .
5. Two cards are chosen without replacement from a deck of 52 cards. Let X denote the number of 3's and Y denote the number of Kings that are drawn.
- Find the joint probability distribution of X and Y .
 - Find the marginal distributions of X and Y .
6. Suppose that X and Y are continuous random variables with joint probability density function

$$f(x, y) = k, \quad 1 < x < 2, \quad 2 < y < 4,$$

where k is a constant. (The random variables X and Y are said to have a *joint uniform probability density*.)

- Find k .
 - Find the marginal densities for X and Y .
7. Suppose the joint (discrete) probability distribution function for discrete random variables X and Y is

$$P(X = x, Y = y) = k, \quad x = 1, 2, \dots, 10; \quad y = 10 - x, 11 - x, \dots, 10.$$

- Find k .
 - Find the marginal distributions for X and Y .
8. A researcher finds that two random variables of interest, X and Y , have joint probability density function
- $$f(x, y) = 24xy, \quad 0 < x < 1, \quad 0 < y < 1 - x.$$
- Show a graph of the joint probability density function.
 - Calculate the marginal densities.
 - Find $P(X > 1/2, Y < 1/4)$.
9. A researcher is conducting a sample survey and is interested in a particular question that respondents answer "yes" or "no". Suppose the probability a respondent answers "yes" is p and that respondents' answers are independent. Let X denote the number of yeses in the first n_1 trials and Y denote the number of yeses in the next n_2 trials.
- Show the joint probability distribution function.
 - Find the probability distribution of the random variable $X + Y$.
10. Refer to the previous problem. If, in the second set of trials, the probability of a "yes" response has become $p_1 \neq p$, find the joint probability distribution function and the marginal distributions. Explain why the variable $X + Y$ is *not* binomial.
11. The number of telephone calls, X , that come into an office during a certain period of the day is distributed as a Poisson random variable with $\lambda = 6$ per hour. The calls are answered according to a binomial process with $p = 3/4$. Let Y denote the number of calls answered.

- (a) Find the joint probability distribution of X and Y .
 (b) Express $P(Y = y)$ in simple form.
 (c) Find $E(Y)$ without using the result in part (b).
12. Suppose that random variables X and Y have joint probability density
- $$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad -\infty < x < \infty, \quad -\infty < y < \infty.$$
- (a) Show that the marginal densities are normal.
 (b) Find $P(X > Y)$.
13. Three students are randomly selected from a group of three freshmen, two sophomores, and two juniors. Let X denote the number of freshmen selected and Y denote the number of sophomores selected. Find the joint probability distribution of X and Y .
14. Random variables X and Y are jointly distributed random variables with $f(x, y) = k$, $x = 0, 1, 2, \dots$ and $y = 0, 1, 2, \dots, 3 - x$.
- (a) Find k .
 (b) Find the marginal densities for X and Y .
15. Suppose that random variables X and Y have joint probability density $f(x, y) = kxy$ on the region bounded by the curves $y = x^2$ and $y = x$ in the first quadrant.
- (a) Show that $k = 24$.
 (b) Find the marginal densities $f(x)$ and $g(y)$.
16. Let X and Y be random variables with joint probability density function $f(x, y) = \frac{k}{x}$, $0 < y < x$, $0 < x < 1$.
- (a) Show that $k = 1$.
 (b) Find $P\left(X > \frac{1}{2}, Y < \frac{1}{4}\right)$.
17. Random variables X and Y have joint probability density
- $$f(x, y) = kx, \quad x - 1 < y < 1 - x, \quad 0 < x < 1.$$
- (a) Find k .
 (b) Find $g(y)$, the marginal density for Y .
 (c) Find $\text{Var}(X)$.
18. Suppose that random variables X and Y have joint probability distribution function
- $$f(x, y) = \frac{1}{21}(x + y), \quad x = 1, 2, 3; \quad y = 1, 2.$$
- (a) Find the marginal densities for X and Y .
 (b) Find $P(X + Y \leq 3)$.
19. A fair coin is flipped three times. Let Y be the total number of heads on the first two tosses, and let W be the total number of heads on the last two tosses.
- (a) Determine the joint probability distribution of W and Y .
 (b) Find the marginal distributions.

- 20.** An environmental engineer measures the amount (by weight) of particulate pollution in air samples of a given volume collected over the smokestack of a coal-operated power plant. X denotes the amount of pollutant per sample collected when a cleaning device on the stack is not in operation, and Y denotes the same amount when the cleaning device is operating. It is known that the joint probability density function for X and Y is

$$f(x, y) = k, \quad 0 \leq x \leq 2, \quad 0 \leq y \leq 1, \quad x > 2y.$$

- (a) Find k .
 - (b) Find the marginal densities for X and Y .
 - (c) Find the probability that the amount of pollutant with the cleaning device in operation is at most $1/3$ of the amount without the cleaning device in operation.
- 21.** Random variables X and Y have joint probability density function

$$f(x, y) = k, \quad x \geq 0, \quad y \geq 0, \quad \frac{x}{4} + y \leq 1.$$

- (a) Show that $k = \frac{1}{2}$.
 - (b) Find $P(X \geq 2, Y \geq \frac{1}{4})$.
 - (c) Find $P(X \leq 1)$.
- 22.** A fair die is thrown once; let X denote the result. Then X fair coins are thrown; let Y denote the number of heads that occur.
- (a) Find an expression for $P(Y = y)$.
 - (b) Explain why $E(Y) = 7/4$.
- 23.** Suppose that X and Y are random variables whose joint probability density function is

$$f(x, y) = 3y, \quad 0 < x < y < 1.$$

- (a) Show that $f(x, y)$ is a joint probability density function.
 - (b) Find the marginal densities.
 - (c) Show that $E\left[\frac{X}{Y}\right] = \frac{E[X]}{E[Y]}$. Does $E\left[\frac{Y}{X}\right] = \frac{E[Y]}{E[X]}$?
- 24.** A coin is tossed until a head appears for the first time; denote the number of trials necessary by X . Then X of these coins are tossed; let Y denote the number of heads that appear.
- (a) Find the joint distribution of X and Y assuming that the coins are fair.
 - (b) Find the marginal distributions of X and Y . (X is geometric; to simplify the distribution of Y , consider the binomial expansion $(1-x)^{-n}$.)
 - (c) Show that $E(Y) = 1$ whether the coins are fair or not.
 - (d) Find the marginal distribution for Y assuming that the coins are loaded to come up heads with probability p .

5.3 CONDITIONAL DISTRIBUTIONS AND DENSITIES

In Example 5.2.1 we tossed five coins and recorded X , the number of heads that appeared. We then tossed the $5 - X$ coins that came up tails again and recorded Y , the number of

heads in the second set of tosses. The joint probability distribution function is shown in Table 5.1.

We might be interested in some conditional probabilities such as the probability that the second set of tosses showed at least two heads, given that one head appeared on the first toss or $P(Y \geq 2|X = 1)$.

We cannot look at the row for $X = 1$ and add the probabilities for $Y \geq 2$ since the probabilities in the row for $X = 1$ do not add up to 1, that is, the row for $X = 1$ is not a probability distribution. However, we know that

$$P(Y \geq 2|X = 1) = \frac{P(Y \geq 2, X = 1)}{P(X = 1)},$$

so a probability distribution can be created from the entries in the column for $X = 1$ by dividing each of them by $P(X = 1)$. If we do this, we find

$$P(Y \geq 2|X = 1) = \frac{\frac{60 + 40 + 10}{1024}}{\frac{160}{1024}} = \frac{11}{16}.$$

We conclude generally that

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} \text{ if } P(X = x) \neq 0.$$

This clearly holds for the case of discrete random variables. We proceed in the same way for continuous random variables, leading to the following definition.

Definition: The *conditional probability distributions* $f(y|X = x)$ and $f(x|Y = y)$, which we denote by $f(y|x)$ and $f(x|y)$ are defined as

$$\begin{aligned} f(y|X = x) &= f(y|x) = \frac{f(x, y)}{f(x)} \\ f(x|Y = y) &= f(x|y) = \frac{f(x, y)}{g(y)} \end{aligned}$$

where $f(x, y)$, $f(x)$, and $g(y)$ are the joint and marginal distributions for X and Y , respectively.

Example 5.3.1

In Example 5.2.2, we considered the joint probability density function of the continuous variables X and Y where

$$f(x, y) = x^2 + \left(\frac{8}{3}\right)x \cdot y, \quad 0 < x < 1, \quad 0 < y < 1.$$

The conditional densities can be seen geometrically as the intersections of the joint probability density surface and vertical or horizontal planes. These curves of intersection are in general not probability densities since they do not have area 1, so they must be divided by the marginal densities to achieve this.

Since the marginal densities are $f(x) = x^2 + \frac{4x}{3}$, $0 < x < 1$, and $g(y) = \frac{4y+1}{3}$, $0 < y < 1$, it follows that

$$f(x|y) = \frac{3x^2 + 8xy}{4y + 1}, \quad 0 < x < 1$$

and

$$f(y|x) = \frac{x^2 + \left(\frac{8}{3}\right)x \cdot y}{x^2 + \left(\frac{4}{3}\right)x} = \frac{3x^2 + 8xy}{3x^2 + 4x}, \quad 0 < y < 1.$$

The domain of each variable is denoted above; the remaining symbol is understood to be fixed.

That each of these is a probability density can be verified by calculating that $\int_0^1 f(x|y) dx = 1$ and $\int_0^1 f(y|x) dy = 1$ and noting that $f(x|y) \geq 0$ and $f(y|x) \geq 0$.

Areas under these conditional densities are probabilities; for example, if $Y = 3/4$, then $f(x|y = 3/4) = \frac{3}{4}(x^2 + 2x)$ so

$$P(X < 1/2 | Y = 3/4) = \frac{3}{4} \int_0^{\frac{1}{2}} x^2 + 2x \, dx = 7/32.$$

The mean values of the conditional densities are also of interest. These are denoted as $E(Y|X = x)$ and $E(X|Y = y)$, respectively. We see that

$$E(Y|X = x) = \int_y y \cdot f(y|x) \, dy \text{ and}$$

$$E(X|Y = y) = \int_x x \cdot f(x|y) \, dx.$$

In this example, $E(Y|X = x) = \int_0^1 y \cdot \frac{x^2 + \left(\frac{8}{3}\right)x \cdot y}{x^2 + \left(\frac{4}{3}\right)x} \, dy = \frac{9x+16}{18x+24}$ and

$$E(X|Y = y) = \int_0^1 x \cdot \frac{x^2 + \left(\frac{8}{3}\right)x \cdot y}{\frac{4}{3}y + \frac{1}{3}} \, dx = \frac{9 + 32y}{12 + 48y}.$$

We note that $E(X|Y = y)$ and $E(Y|X = x)$ are functions of y and x , respectively.

Example 5.3.2

Finally, in this section we consider a slightly more complex continuous example. Let

$$f(x, y) = 2e^{-x-y}, \quad x \geq 0, \quad y \geq x.$$

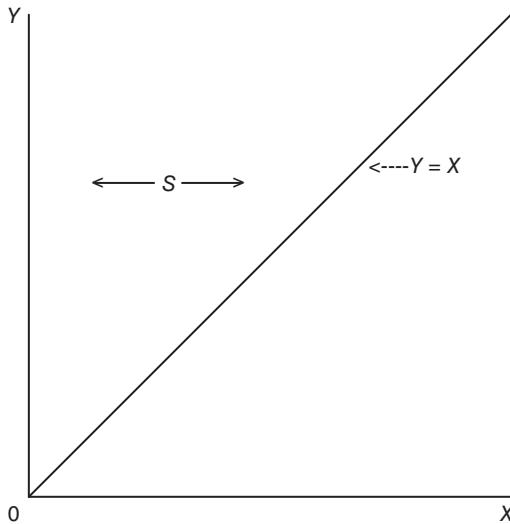


Figure 5.6 Sample space for $f(x,y) = 2e^{-x-y}$, $x \geq 0$, $y \geq x$.

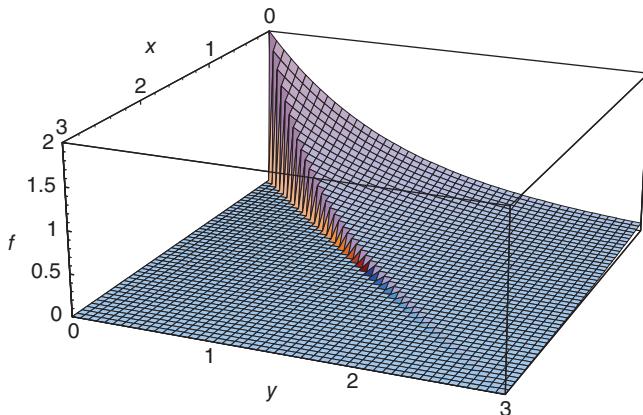


Figure 5.7 Probability surface for Example 5.3.2.

Here, we must be cautious in determining the limits of integration. A picture of the sample space is shown in Figure 5.6.

The bivariate function itself, shown in Figure 5.7, is also interesting:

The marginal densities are as follows:

$$f(x) = \int_0^{\infty} 2e^{-x-y} dy = 2e^{-2x}, \quad x \geq 0 \text{ and}$$

$$g(y) = \int_0^y 2e^{-x-y} dx = 2e^{-y}(1 - e^{-y}), \quad y \geq 0.$$

The conditional densities are then

$$f(x|y) = \frac{e^{-x}}{1 - e^{-y}}, \quad 0 \leq x \leq y \text{ and}$$

$$f(y|x) = e^{x-y}, \quad y \geq x.$$

The reader might verify that each of these is a probability density. The conditional expectations are then

$$E(X|Y = y) = \frac{e^y - y - 1}{e^y - 1} \text{ and } E(Y|X = x) = 1 + x.$$

EXERCISES 5.3

1. In Example 5.3.2, verify that $f(x|y)$ and $f(y|x)$ are each probability densities.
2. In Example 5.3.2, determine $E[X|Y = y]$ and $E[Y|X = x]$.
3. Suppose the joint probability density function of random variables X and Y is given by

$$f(x, y) = c \cdot (x + y), \quad 0 < x < 1, \quad 0 < y < 1.$$

- (a) Show that $c = 1$.
- (b) Find the marginal densities and the conditional densities, verifying in each case that these are probability densities.
- (c) Find $E(Y|X = x)$ and $E(X|Y = y)$.
4. Suppose X and Y are discrete random variables with

$$f(x, y) = \frac{x + y}{21}, \quad x = 1, 2, 3; \quad y = 1, 2.$$

- (a) Show that $f(x, y)$ is a probability distribution function.
- (b) Find the conditional distributions.
- (c) Find $E(Y|X = x)$ and $E(X|Y = y)$.
5. Let $f(x, y) = kxy$, $0 < x < 1$, $0 < y < x$ for random variables X and Y .
- (a) Show that $k = 8$.
- (b) Find the marginal densities and the conditional densities.
6. Random variables X and Y have joint probability density function

$$f(x, y) = c \cdot x \cdot (2 - x - y), \quad 0 < x < 1, \quad x < y < 1.$$

- (a) Show that $c = 8$.
- (b) Find the marginal densities.
- (c) Verify that $f(y|x)$ is a probability density.
7. Suppose that random variables X and Y have joint probability density $f(x, y) = k \cdot x \cdot y$ for $x \geq 0$ and $y \geq 0$ on the circle $x^2 + y^2 \leq 1$.
- (a) Find k .
- (b) Find $E(Y|X = x)$ and then find $E(Y)$ from this result.
8. Let X and Y have joint probability distribution: $f(0, 0) = 1/3$, $f(0, 1) = 1/2$, $f(1, 1) = 1/6$.
- (a) Show that $f(x, y)$ is a joint probability distribution.
- (b) Find the marginal and conditional distributions.

9. Random variables X and Y have $f(x, y) = k \cdot x \cdot y$, where the sample space is the finite area between $y = x^2$ and $y = x$.

(a) Find $E(Y|X = x)$ and $E(X|Y = y)$.

(b) Find $P(X > 1/2|Y = 1/3)$.

10. Suppose X and Y have joint probability density function

$$f(x, y) = \frac{3}{8}(x + y^2), \quad 0 < x < 2, \quad 0 < y < 1.$$

(a) Find $P\left(X > 1, Y < \frac{1}{2}\right)$.

(b) Find $f(y|x)$.

(c) Evaluate $P\left(Y > \frac{1}{2}|X = 1\right)$.

11. Let X and Y be random variables with joint probability density function

$$f(x, y) = 3x + 1, \quad 0 < x < 1, \quad 0 < y < 1 - x.$$

(a) Find the marginal densities.

(b) Find $f(y|x)$.

12. Random variables X and Y have joint probability density function

$$f(x, y) = k \cdot x \cdot y^2, \quad 0 < x < 1, \quad x < y < 1.$$

(a) Find k .

(b) Find the marginal densities $f(x)$ and $g(y)$.

(c) Find the conditional density $f(y|x)$.

5.4 EXPECTED VALUES AND THE CORRELATION COEFFICIENT

If random variables X and Y have a joint probability density $f(x, y)$, then, as we have seen, X and Y are univariate random variables and hence have means and variances. It is of course true that

$$\begin{aligned} E(X) &= \int_x x \cdot f(x) dx \text{ and} \\ E(Y) &= \int_y y \cdot g(y) dy, \end{aligned}$$

but these values can also be found from the joint probability density as

$$\begin{aligned} E(X) &= \int_x \int_y x \cdot f(x, y) dy dx \text{ and} \\ E(Y) &= \int_x \int_y y \cdot f(x, y) dy dx. \end{aligned} \tag{5.1}$$

That these relationships are true can be easily established. Consider the above expression for $E(X)$ and factor out x from the inner integral. This gives

$$E(X) = \int_x x \cdot \left[\int_y f(x, y) dy \right] dx = \int_x x \cdot f(x) dx,$$

so the formulas are equivalent. Formulas (5.1) show that the marginals are not needed, however, since the order of the integration can often be reversed and so the expectations can be found without finding the marginal densities.

Now we turn to measuring the degree of dependence of one random variable and the other. The idea of *independence* is easy to anticipate:

Definition: Random variables X and Y *independent* if and only if

$$f(x, y) = f(x) \cdot g(y) \text{ for all values of } x \text{ and } y,$$

where $f(x)$ and $g(y)$ are the marginal distributions or densities.

Usually, it is not necessary to consider the joint density of independent variables since probabilities can be calculated from the marginal densities. If X and Y are independent, then

$$\begin{aligned} P(a < X < b, c < Y < d) &= \int_a^b \int_c^d f(x, y) dy dx \\ &= \int_a^b \int_c^d f(x) \cdot g(y) dy dx \\ &= \int_a^b f(x) dx \cdot \int_c^d g(y) dy \text{ so} \\ P(a < X < b, c < Y < d) &= P(a < X < b) \cdot P(c < Y < d), \end{aligned}$$

showing that the joint density is not needed.

Referring to Example 5.3.2,

$$f(x, y) = 2e^{-x-y} \neq 2e^{-2x} \cdot 2e^{-y}(1 - e^{-y}) = f(x) \cdot g(y),$$

so X and Y are not independent. This raises the idea of measuring the extent of their dependence. In order to do this, we first define the *covariance* of random variables X and Y as follows:

Definition: The *covariance* of random variables X and Y is

$$\text{Covariance}(X, Y) = \text{Cov}(X, Y) = E[[X - E(X)][Y - E(Y)]] \quad (5.2)$$

As a special case, if $X = Y$, then $\text{Cov}(X, Y) = \text{Cov}(X, X)$ and the formula becomes the variance of X , $\text{Var}(X)$. But, unlike the variance, the covariance can be negative. Before calculating that, however, consider formula (5.2). By expanding it, we find that

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

a result that is often very useful.

In the example we are considering, $E(X \cdot Y) = 1$, $E(X) = 1/2$, and $E(Y) = 3/2$, so $\text{Cov}(X, Y) = 1/4$.

The covariance is also used to define the *correlation coefficient*, $\rho(x, y)$ as we do now.

Definition: The *correlation coefficient* of the random variables X and Y is

$$\rho(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y},$$

where σ_x and σ_y are the standard deviations of X and Y , respectively.

In this example, we find that $\sigma_x = 1/2$ and $\sigma_y = \sqrt{5}/2$, so $\text{Cov}(X, Y) = \frac{1}{\sqrt{5}} = 0.447214$.

Now consider jointly distributed random variables X and Y .

$$\begin{aligned}E(X + Y) &= \int_{x,y} (x + y) \cdot f(x, y) dy dx \\ &= \int_x \int_y x \cdot f(x, y) dy dx + \int_x \int_y y \cdot f(x, y) dy dx \text{ so} \\ E(X + Y) &= E(X) + E(Y), \text{ or}\end{aligned}$$

The expectation of a sum is the sum of the expectations.

It is also easy to see that if a and b are constants,

$$E(aX + bY) = aE(X) + bE(Y),$$

since the constants can be factored out of the integrals. The result can be easily generalized:

$$E(aX + bY + cZ + \dots) = aE(X) + bE(Y) + cE(Z) + \dots$$

As might be expected, variances of sums are a bit more complicated than expectations of sums. We begin with $\text{Var}(aX + bY)$. By definition this is

$$\begin{aligned}\text{Var}(aX + bY) &= E[aX + bY - E[aX + bY]]^2 \\ &= E[a[X - E(X)] + b[Y - E(Y)]]^2.\end{aligned}$$

Now squaring, factoring out the constants, and taking the expectation term by term, we find

$$\text{Var}(aX + bY) = a^2 E[X - E(X)]^2 + 2abE[(X - E(X))(Y - E(Y))] + b^2 E[Y - E(Y)]^2,$$

and we recognize the terms in this expression as

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + 2ab\text{Cov}(X, Y) + b^2\text{Var}(Y). \quad (5.3)$$

So we cannot say, as we did with expectations, that the variance of a sum is the sum of the variances, but this would be true if the covariance were zero. When does this occur? If X and Y are independent, then

$$\begin{aligned} E(X \cdot Y) &= \int_x \int_y xyf(x, y)dy dx \\ &= \int_x \int_y x \cdot y \cdot f(x) \cdot g(y) dy dx \\ &= \int_x x \cdot \left[\int_y y \cdot g(y) dy \right] \cdot f(x) dx \\ E(X \cdot Y) &= E(Y) \cdot \int_x xf(x)dx = E(X) \cdot E(Y). \end{aligned}$$

Since $E(X \cdot Y) = E(X) \cdot E(Y)$, $\text{Cov}(X, Y) = 0$. So we can say that if X and Y are independent, then

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y).$$

But the converse of this assertion is false: that is, if $\text{Cov}(X, Y) = 0$, then X and Y are not necessarily independent. An example will establish this. Consider the joint distribution of X and Y as given in the following table:

	Y	−1	0	1
X				
−1		a	b	a
0		b	0	b
1		a	b	a

We select a and b so that $4a + 4b = 1$. Take $a = 1/6$ and $b = 1/12$ as an example among many choices that could be made. The symmetry in the table shows that $E(X) = E(Y) = 0$ and that $E(X \cdot Y) = 0$. So X and Y have $\text{Cov}(X, Y) = 0$. But $P(X = −1, Y = −1) = 1/6 \neq (5/12) \cdot (5/12) = 25/144$, so X and Y are not independent. To take the more general case,

$$P(X = −1, Y = −1) = a \neq (2a + b)^2$$

so X and Y are not independent.

If $\text{Cov}(X, Y) = 0$, we call X and Y *uncorrelated*. We conclude that if X and Y are independent, then they are uncorrelated, but uncorrelated variables are not necessarily independent.

Finally, in this section we establish a useful fact, namely, that the correlation coefficient, ρ , is always in the interval from -1 to 1 :

$$-1 \leq \rho(x, y) \leq 1.$$

As a proof of this, consider variables X and Y that each have mean 0 and variance 1. (If this is not the case, transform the variables by subtracting their means and dividing by their standard deviations, producing X and Y .)

Since the variance of any variable is nonnegative,

$$\text{Var}(X - Y) \geq 0 \text{ or, by formula 5.3,}$$

$$\text{Var}(X - Y) = \text{Var}(X) - 2\text{Cov}(X, Y) + \text{Var}(Y).$$

But $\text{Var}(X) = \text{Var}(Y) = 1$ and $\text{Cov}(X, Y) = \rho$, so

$$1 - 2\rho + 1 \geq 0,$$

which implies that $\rho \leq 1$.

The other half of the inequality can be established in a similar way by noting that $\text{Var}(X + Y) \geq 0$.

The reader will be asked in problem 5 to show that the transformations done to insure that the variables having mean 0 and variance 1 do not affect the correlation coefficient.

Example 5.4.1

The fact that the expectation of a sum is the sum of the expectations and the fact that, if the summands are independent, then the variance of the sum is the sum of the variances, can be used to provide a neat derivation of the mean and variance of a binomial random variable.

Suppose that the random variable X represents the number of successes when there are n independent trials of a binomial random variable with probability p of success at any trial. Now define the variables

$$X_1 = \begin{cases} 1 & \text{if the first trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if the second trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

.

.

$$X_n = \begin{cases} 1 & \text{if the } n\text{th trial is a success} \\ 0 & \text{otherwise.} \end{cases}$$

The X_i 's are often called *indicator random variables*.

Since X_i is 1 only when a success occurs and is 0 when a failure occurs,

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^n X_i.$$

Now

$$E(X_i) = 1 \cdot p + 0 \cdot (1 - p) = p \text{ and}$$

$$E(X_i^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$$

so

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np.$$

Also,

$$\text{Var}(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1 - p) = pq,$$

so

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

We have again established the formulas for the mean and variance of a binomial random variable.

5.5 CONDITIONAL EXPECTATIONS

Recall Example 5.2.1 of this chapter where five fair coins were tossed, the coins showing heads being put aside and those showing tails tossed again. We called the random variable X the number of coins showing heads on the first toss and the random variable Y the number of coins showing heads on the second set of tosses. $E(Y)$ is of interest.

First note that $E(Y|X = x)$ and $E(X|Y = y)$ are functions of x and y , respectively. If we let

$$E(Y|X = x) = k(x),$$

then we could consider the function $k(X)$ of the random variable X . This is itself a random variable. We denote this random variable by $E(Y|X)$, so

$$E(Y|X) = k(X).$$

We now return to our example. Since there are $5 - x$ coins to be tossed the second time and the probability of success is $1/2$, it follows that $E(Y|X = x) = (5 - x)/2$. This conditional expectation $E(Y|X)$ is a function of X . It also has an expectation which is

$$E[E(Y|X)] = E\left[\frac{5 - X}{2}\right] = \frac{5}{2} - \frac{E(X)}{2} = \frac{5}{2} - \frac{5/2}{2} = \frac{5}{4}, \text{ which we found as } E(Y).$$

So we conjecture that $E[E(Y|X)] = E(Y)$ and that $E[E(X|Y)] = E(X)$.

We now justify this process of establishing *unconditional* expectations based on *conditional* expectations.

It is essential to note here that $E(Y|X)$ is a function of X ; its expectation is found using the marginal distribution of X . Similarly, $E(X|Y)$ is a function of Y , and its expectation is found using the marginal distribution of Y .

We will give a proof that $E(E(Y|X)) = E(Y)$ using a continuous bivariate distribution, say $f(x, y)$.

First we note that

$$\begin{aligned} E(E(Y|X)) &= \int_x E(Y|X) \cdot f(x) \, dx \\ &= \int_x \left[\int_y y \cdot \frac{f(x,y)}{f(x)} \, dy \right] \cdot f(x) \, dx \\ &= \int_x \int_y y \cdot f(x,y) \, dy \, dx = E(Y). \end{aligned}$$

The proof that $E(E(X|Y)) = E(X)$ is of course similar.

Example 5.5.1

We apply these results to Example 5.3.2. In this example, $f(x, y) = 2e^{-x-y}$, $x \geq 0, y \geq x$. We found that $f(y|x) = e^{x-y}$, $y \geq x$, and that $E(Y|X) = 1 + X$.

Now we calculate

$$E(Y) = \int_0^\infty 2 \cdot y \cdot e^{-y} \cdot (1 - e^{-y}) dy = 3/2$$

Also

$$E[E(Y|X)] = E(1 + X) = \int_0^\infty (1 + x) \cdot 2 \cdot e^{-2x} dx.$$

This integral, as the reader can easily check, is also $3/2$.

Example 5.5.2

An observation, X , is taken from a uniform density on $(0, 1)$, then observations are taken until the result exceeds X . Call this observation Y . What is the expected value of Y ?

It appears obvious that, on average, the first observation is $1/2$. Then Y can be considered to be a uniform variable on the interval $(1/2, 1)$, so its expectation is $3/4$. Let us make these calculations more formal so that the technique could be applied in a less obvious situation.

We have that X is uniform on $(0, 1)$ so that $f(x) = 1$, $0 < x < 1$ and Y is uniform on $(x, 1)$ so that

$$f(y|x) = \frac{1}{1-x}, x < y < 1.$$

The joint density is then

$$f(x, y) = f(x) \cdot f(y|x) = \frac{1}{1-x}, \quad x < y < 1, \quad 0 < x < 1$$

The marginal density for Y is then

$$g(y) = \int_0^y \frac{1}{1-x} dx = -\ln(1-y), \quad 0 < y < 1, \quad \text{and}$$

$$E(Y) = \int_0^1 -y \ln(1-y) dy = \frac{3}{4},$$

verifying our earlier informal result.

EXERCISES 5.5

1. In Example 5.3.2, find that $\text{Cov}(X, Y) = 1/\sqrt{5}$.
2. In Example 5.3.2, verify that $E[Y] = 3/8$.
3. Show that $\rho(X, Y)$ in Example 5.2.1 is $-1/\sqrt{3}$.
4. Find the correlation coefficient between X and Y for the probability density:

$$f(x, y) = x + y, \quad 0 < x < 1, \quad 0 < y < 1.$$

5. Show that $\text{Cov}(aX + bY, cX + dY) = ac\text{Var}(X) + (ad + bc)\text{Cov}(X, Y) + bd\text{Var}(Y)$.
6. Show that $\text{Cov}(X - Y, X + Y) = \sigma_x^2 - \sigma_y^2$.
7. Show that $\rho(aX + b, cY + d) = \rho(X, Y)$, provided that $a > 0$ and $b > 0$.
8. Let $f(x, y) = k, 0 \leq x \leq 2, 0 \leq y \leq 1, x \geq 2y$.
 - (a) Are X and Y independent?
 - (b) Find $P(Y < X/3)$.
9. Let $f(x, y) = \frac{3}{40} \left(x^2 + \frac{1}{2}xy^2 \right), \quad 0 < x < 2; -2 < y < 2$.
 - (a) Show that $f(x, y)$ is a joint probability density function.
 - (b) Show that X and Y are independent.
10. Let $f(x, y) = \left(\frac{3}{8} \right) (x^2 + y^2), \quad -1 < x < 1; -1 < y < 1$.
 - (a) Verify that $f(x, y)$ is a joint probability density function.
 - (b) Find the marginal densities, $f(x)$ and $g(y)$.
 - (c) Find the conditional densities, $f(x|y)$ and $f(y|x)$.
 - (d) Verify that $E[E(X|Y)] = E(X)$.
 - (e) Find $P\left(X > \frac{1}{2} | Y = \frac{1}{2}\right)$.
11. Let $f(x, y) = 1 + \frac{x}{2} - \frac{1}{4}xy^2, \quad -1/2 < x < 1/2; -1/2 < y < 1/2$.
 - (a) Show that $f(x, y)$ is a joint probability density function.
 - (b) Show that $f(x|y) = f(x, y)$.
12. Let $f(x, y) = k \cdot x^2 \cdot (8 - y), \quad x < y < 2x, \quad 0 < x < 2$.

(a) Find the marginal and conditional densities.

(b) Find $E(Y)$ from the marginal density of Y and then verify that $E[E(Y|X)] = E(Y)$.

13. Random variables X and Y have joint probability density function

$$f(x, y) = kx, \quad 0 < x < 2, \quad x^2 < y < 4.$$

(a) Show that $k = \frac{1}{4}$.

(b) Find $f(x|y)$.

(c) Find $E(X|Y = y)$.

14. Suppose that the joint probability density function for random variables X and Y is

$$f(x, y) = 2x + 2y - 4xy, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1.$$

Are X and Y independent? Why or why not?

15. The joint probability density function for random variables X and Y is

$$f(x, y) = k, \quad -1 < x < 1, \quad x^2 < y < 1.$$

Find the conditional densities $f(x|y)$ and $f(y|x)$ and show that each of these is a probability density function.

16. Suppose that random variables X and Y have joint probability density function $f(x, y) = kx(2 - x - y)$, $0 < y < 1$, $0 < x < y$. Find $f(x|y)$ and show that this is a probability density function.

17. Random variables X and Y have joint probability distribution function

$$f(x, y) = \frac{1}{6}, \quad x = 1, 2, 3 \text{ and } y = 1, 2, \dots, 4 - x.$$

(a) Find formulas for $f(x)$ and $g(y)$, the marginal densities.

(b) Find a formula for $f(y|x)$ and explain why the result is a probability distribution function.

18. Find the correlation coefficient between X and Y if their joint probability density function is

$$f(x, y) = k, \quad 0 \leq x \leq y, \quad 0 \leq y \leq 1.$$

19. Random variables X and Y are discrete with joint distribution given by

		Y	
		0	1
X	0	$\frac{1}{6}$	$\frac{1}{3}$
	1	$\frac{1}{3}$	$\frac{1}{6}$

Find the correlation coefficient between X and Y .

20. Random variables X and Y have joint probability density function $f(x, y) = \frac{1}{\pi}, \quad x^2 + y^2 \leq 1$. Are X and Y independent?

- 21.** Let random variables X and Y have joint probability density function $f(x, y) = kxy$ on the region bounded $0 \leq x \leq 2$, $0 < y < x$.
- Show that $k = 1/2$.
 - Find the marginal densities $f(x)$ and $g(y)$.
 - Are X and Y independent?
 - Find $P\left(X > \frac{3}{2} | Y = 1\right)$.
- 22.** Random variables X and Y are uniformly distributed on the region $x^2 \leq y \leq 1$, $0 \leq x \leq 1$. Verify that $E(Y|X) = E(Y)$.
- 23.** Suppose that X and Y are random variables with $\sigma_x^2 = 10$, $\sigma_y^2 = 20$, and $\rho = \frac{1}{2}$. Find $\text{Var}(2X - 3Y)$.
- 24.** Let X_1, X_2 , and X_3 be random variables with $E(X_i) = 0$ and $\text{Var}(X_i) = 1$, $i = 1, 2, 3$. Also, $\text{Cov}(X_i, X_j) = -\frac{1}{2}$ if $i \neq j$. Find $\text{Var}\left(\sum_{i=1}^3 iX_i\right)$.
- 25.** Let $X_1, X_2, X_3, \dots, X_n$ be uncorrelated random variables with common variance σ^2 . Show that $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $X_j - \bar{X}$ are uncorrelated, $j = 1, 2, \dots, n$.
- 26.** A box contains five red and two green marbles. Two marbles are drawn out, the first not being replaced before the second is drawn. Let X be 1 if the first marble is red and 0 otherwise; Y is 1 if the second marble is red and is 0 otherwise. Find the correlation coefficient between X and Y .
- 27.** In four tosses of a fair coin, let X be the number of heads and Y be the length of the longest *run* of heads (a sequence of tosses of heads). For example, in the sequence *HTHH*, $X = 3$ and $Y = 2$. Find the correlation coefficient between X and Y .
- 28.** An observation, X , is taken from the exponential distribution, $f(x) = e^{-x}$, $x \geq 0$. Sampling then continues until an observation, Y , is at most X . Show that $E(Y) = 2 - \frac{\pi^2}{6}$. [Hint: Expand the integral in a power series to show that $\int_0^\infty \frac{xe^{-2x}}{1-e^{-x}} dx = \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots$]
- 29.** Variances as well as expected values can be found by conditioning. Consider the formula

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)].$$

- (a)** Verify that the formula gives the correct result for the distribution in Example 5.3.2,
- $$f(x, y) = 2e^{-x-y}, \quad x \geq 0, \quad y \geq x.$$
- (b)** Prove that the formula is correct in general.
- 30.** The fair wheel is spun once and the result, X , is recorded. Then the wheel is spun again until the result is less than X . Call the second result Y .
- Find the joint probability density for X and Y .
 - Find $E(X|Y)$ and $E(Y|X)$.
 - $E[E(Y|X)] = \frac{1}{4}$.

5.6 BIVARIATE NORMAL DENSITIES

The bivariate extension of the normal density is a very important example of a bivariate density. We study this density and some of its applications in this section.

We say that X and Y have a *bivariate normal density* if

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \text{Exp}\left[-\frac{1}{2(1-\rho^2)}\right] \left\{ \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right\}$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$. Note that there are five parameters, the means and variances of each variable as well as ρ , and the correlation coefficient between X and Y .

A graph of a typical bivariate normal surface is shown in Figure 5.8. The surface is a *standard bivariate normal surface* since X and Y each have mean 0 and variance 1 and ρ has been taken to be 0.

As one might expect, the marginal and conditional densities (and indeed the intersection of the surface with any plane perpendicular to the X, Y plane) are normal densities. We now establish some of these facts. For convenience, and without any loss of generality, we consider a bivariate normal surface with $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$. We begin with a proof that the volume under the surface is 1.

The function we are considering is now

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}},$$

$-\infty < x < \infty$ and $-\infty < y < \infty$.

Completing the square in the exponent gives

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x-\rho y)^2}{2(1-\rho^2)} - \frac{1}{2}y^2}.$$

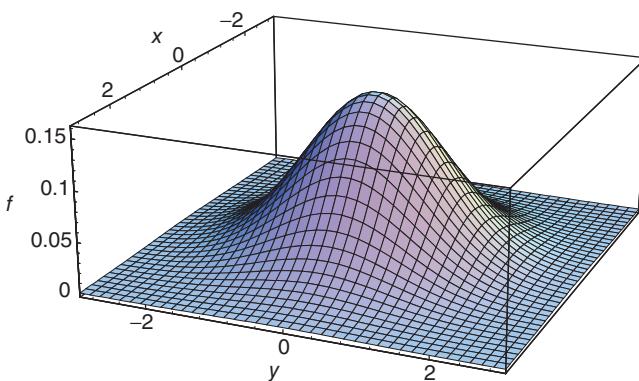


Figure 5.8 Normal probability surface, $\rho = 0$.

So

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dy \, dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} e^{-\frac{(x-\rho y)^2}{2(1-\rho^2)}} \, dx \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \, dy. \end{aligned}$$

The inner integral represents the area under a normal curve with mean ρy and variance $1 - \rho^2$ and so is 1. The outer integral is then the area under a standard normal curve and so is 1 also, showing that $f(x, y)$ has volume 1.

This also shows that the marginal density for Y is

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}, \quad -\infty < y < \infty,$$

with a similar result for $f(x)$.

Finding $\frac{f(x,y)}{g(y)}$ above gives

$$f(x|y) = \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} e^{-\frac{(x-\rho y)^2}{2(1-\rho^2)}},$$

which is a normal curve with mean ρy and standard deviation $\sqrt{1-\rho^2}$.

Now let us return to the general case where the density is not a standard bivariate normal density. It is easy to show that the marginals are now $N(\mu_x, \sigma_x)$ and $N(\mu_y, \sigma_y)$ while the conditional densities are

$$f(y|x) = N \left[\mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x), \sigma_y \sqrt{1 - \rho^2} \right]$$

and

$$f(x|y) = N \left[\mu_x + \rho \frac{\sigma_x}{\sigma_y} (y - \mu_y), \sigma_x \sqrt{1 - \rho^2} \right].$$

The expected value of Y given $X = x$, $E(Y|X = x)$, is called the *regression of Y on X* . Here, we find that

$$E(Y|X = x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x), \text{ a straight line.}$$

If $\rho = 0$ then note that $f(x, y) = f(x) \cdot g(y)$, so that X and Y are independent. In this case, it is probably easiest to use the individual marginal densities in finding probabilities. If $\rho \neq 0$, it is probably best to standardize the variables before calculating probabilities. Some computer algebra systems can calculate bivariate probabilities.

In Figures 5.9 and 5.10, we show two graphs to indicate the changes that result in the shape of the surface when the correlation coefficient varies.

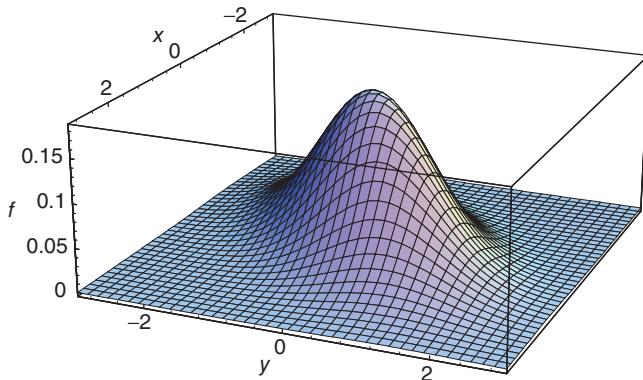


Figure 5.9 Normal bivariate surface, $\rho = 0.5$.

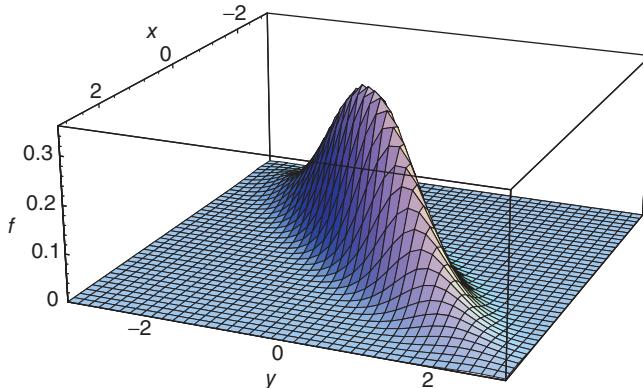


Figure 5.10 Normal bivariate surface, $\rho = 0.9$.

Contour Plots

Contours or *level curves* of a surface show the location of points for which the function, or height of the surface, takes on constant values. The contours are then slices of the surface with planes parallel to the X, Y plane.

If $\rho = 0$, we expect the contours to be circles, as shown in Figure 5.11.

If $\rho = 0.9$, however, the contours become ellipses as shown in Figure 5.12.

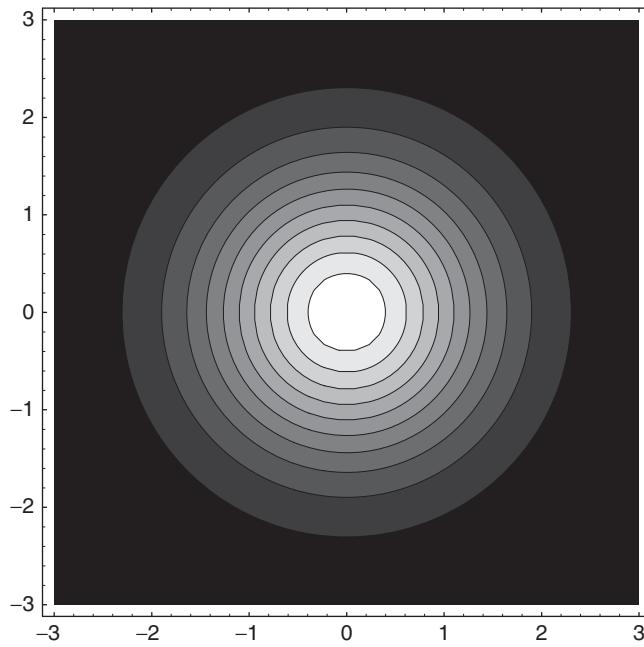


Figure 5.11 Circular contours for a normal probability surface, $\rho = 0$.

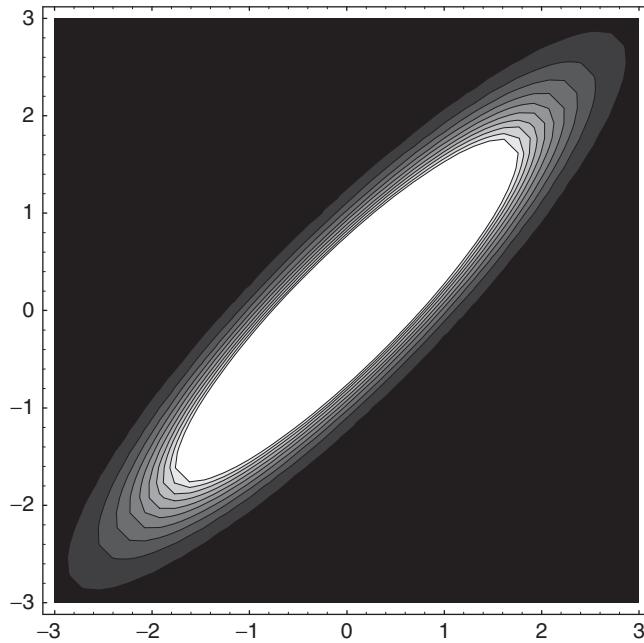


Figure 5.12 Elliptical contours for a normal probability surface, $\rho = 0.9$.

EXERCISES 5.6

1. Let X and Y have a standard bivariate normal density with $\rho = 0.6$.
 - Show that the marginal densities are normal.
 - Show that the conditional densities are normal.
 - Calculate $P(-2 < X < 1, 0 < Y < 2)$.
2. Height (X) and intelligence (Y) are presumed to be random variables that have a slight positive correlation coefficient. Suppose that these characteristics for a group of people are distributed according to a bivariate normal curve with $\mu_x = 67''$, $\sigma_x = 4''$, $\mu_y = 114$, $\sigma_y = 10$, and $\rho = 0.20$.
 - Find $P(66 < X < 70, 107 < Y < 123)$.
 - Find the probability that a person whose height is 5'7" has an intelligence quotient of at least 121.
 - Find the regression of Y on X .
3. Show that ρ is the correlation coefficient between X and Y when these variables have a bivariate normal density.
4. The guidance system for a missile is being tested. The aiming point of the missile (X, Y) is presumed to be a bivariate normal density with $\mu_x = 0$, $\sigma_x = 1$, $\mu_y = 0$, $\sigma_y = 6$, and $\rho = 0.42$. Find the probability the missile lands within 2 units of the origin.
5. Show that uncorrelated bivariate normal random variables are independent.

5.7 FUNCTIONS OF RANDOM VARIABLES

Joint distributions can be used to establish the probability densities of sums of random variables and can also be utilized to find the probability densities of products and quotients. We show how this is done through examples.

Example 5.7.1

Consider again two independent variables, X and Y , each uniformly distributed on $(0,1)$. The joint density is then $f(x,y) = 1$, $0 < x < 1, 0 < y < 1$.

If $Z = X + Y$, then the distribution function of Z , $G(z)$ can be found by calculating volumes beneath the joint density. The diagram in Figure 5.13 will help in doing this.

Computing volumes under the joint density function, we find that

$$(a) G(z) = P(X + Y \leq z) = \frac{1}{2}z^2 \text{ if } 0 < z < 1 \text{ and}$$

$$(b) G(z) = 1 - \frac{1}{2}[1 - (z - 1)]^2 \text{ if } 1 < z < 2.$$

It follows from this that

$$g(z) = \begin{cases} z & 0 < z < 1 \\ 2 - z & 1 < z < 2. \end{cases}$$

This gives the triangular density we have seen previously.

The technique is feasible for more difficult densities as the next example shows.

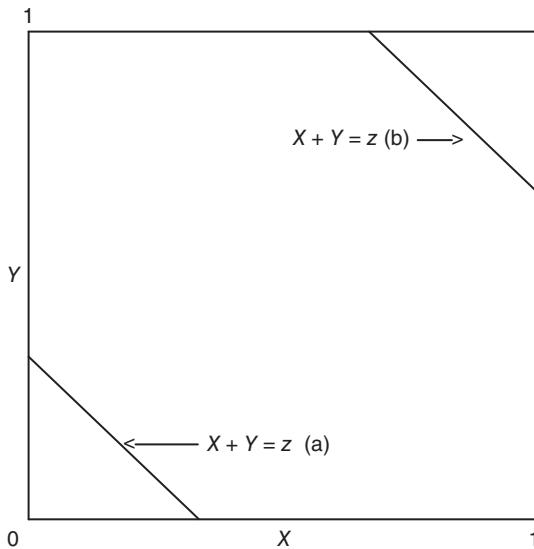


Figure 5.13 Jointly distributed uniform variables.

Example 5.7.2

Suppose X and Y are each independently exponentially distributed with $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, with a similar distribution for Y . The distribution function for $Z = X + Y$ can be found by considering first the sample space shown in Figure 5.14.

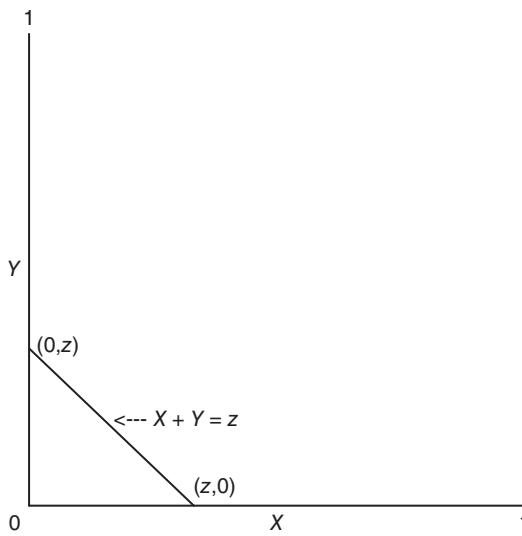


Figure 5.14 Sample space for two independent exponential variables.

Now

$$G(z) = \int_0^z \int_0^{z-x} \lambda^2 e^{-\lambda(x+y)} dy dx,$$

which can be found to be

$$G(z) = 1 - e^{-\lambda z} - \lambda z e^{-\lambda z},$$

so that

$$g(z) = \lambda^2 z e^{-\lambda z}, \quad z \geq 0.$$

This is the gamma density we have seen before.

This technique will work nicely on other random variables such as sums or quotients. An example of each follows.

Example 5.7.3

Let X and Y be independently distributed uniform random variables on $(0,1)$ and consider the product, $Z = X \cdot Y$. Figure 5.15 will help in seeing that the appropriate volume under the joint density gives

$$G(z) = z + \int_z^1 \int_0^{\frac{z}{x}} 1 \, dy \, dx,$$

which is found to be

$$G(z) = z - z \ln z$$

so that

$$g(z) = -\ln z, \quad 0 < z < 1.$$

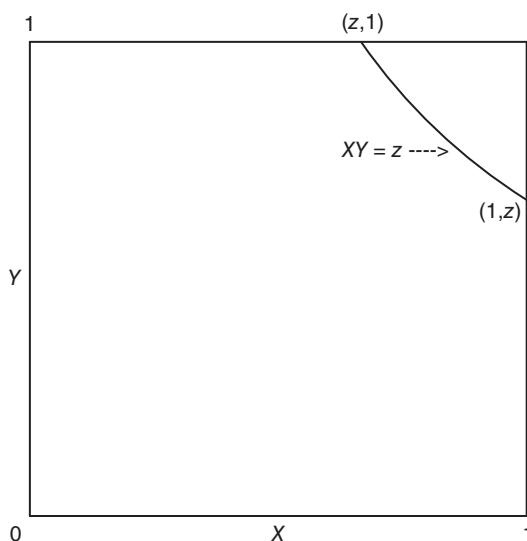


Figure 5.15 Sample space for the product of two uniform random variables.

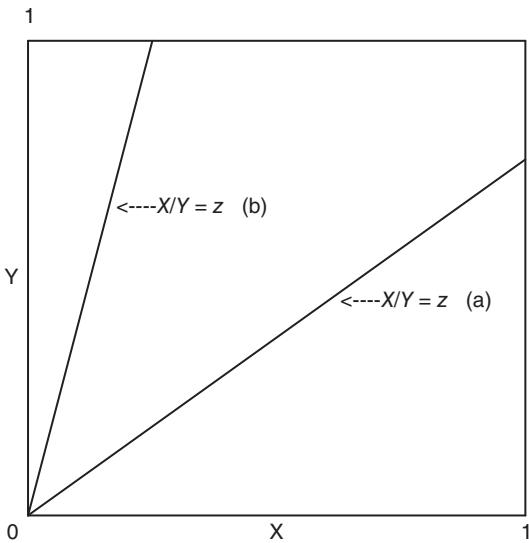


Figure 5.16 Sample space for the quotient of two uniform random variables.

Example 5.7.4

As a final example in this section, we consider the quotient of two independently distributed uniform variables, $Z = X/Y$. Figure 5.16 shows the relevant geometry.

Separating the two cases here and noting that $Z = \frac{X}{Y} \geq z$ above the line $Y = X/z$, we have

$$(a) G(z) = 1 - \int_0^1 \int_0^{\frac{x}{z}} 1 \, dy \, dx = 1 - \frac{1}{2z}, \quad 1 < z < \infty \text{ and}$$

$$(b) G(z) = \int_0^z \int_{\frac{x}{2}}^1 1 \, dy \, dx = \frac{z}{2}, \quad 0 \leq z \leq 1.$$

This gives the density of the quotient as

$$g(z) = \begin{cases} \frac{1}{2} & 0 < z < 1 \\ \frac{1}{2z^2} & 1 < z < \infty. \end{cases}$$

EXERCISES 5.7

- Find the probability density for $Z = X/Y$ if X has an exponential density with mean 1 and Y has an independent exponential density with mean $1/2$.
- Suppose X and Y are independent observations from $f(x) = 2x$, $0 < x < 1$. Find the probability density for $Z = X \cdot Y$.

3. Find the probability density for $Z = X + Y$ where X is an observation from $f(x) = 2x$, $0 < x < 1$, and Y is an independent observation from $g(y) = 2(1 - y)$, $0 < y < 1$.
4. Find the probability density for $Z = X \cdot Y$ where X and Y are independent observations from $f(x) = 3x^2$, $0 < x < 1$.

CHAPTER REVIEW

In this chapter, we study random variables that are defined on the same sample space and consider their *joint probability distributions* or *joint probability densities*

$$f(x, y) = P(X = x, Y = y) \text{ if } X \text{ and } Y \text{ are discrete and}$$

$$P(a < X < b, c < Y < d) = \int_c^d \int_a^b f(x, y) \, dy \, dx \text{ if } X \text{ and } Y \text{ are continuous.}$$

Our study has been confined to two random variables producing *bivariate* probability distributions or densities, although the techniques used in this chapter can often be extended to three or more random variables.

X and Y are random variables and their individual densities are called *marginal densities*. In the continuous case, denoting these by $f(x)$ and $g(y)$, respectively,

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

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

Appropriate sums are used in the discrete case.

It is best to think of $f(x, y)$ geometrically and to plot the surface. Slices of the surface for which X or Y are constant produce curves which, because their areas are not usually one, are rarely probability densities. *Conditional densities* can be produced, however, by dividing the conditional density by the area under the curve:

$$f(x|Y = y) = f(x|y) = \frac{f(x, y)}{f(x)} \text{ and } f(y|X = x) = f(y|x) = \frac{f(x, y)}{g(y)}.$$

Bivariate distributions or densities, like univariate distributions or densities, are summarized by means and variances. However, unlike univariate distributions or densities, means and variances for bivariate distributions or densities can be calculated in more than one way. For example,

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) \, dy \, dx$$

with a similar result for $E(Y)$.

Functions of random variables are of interest, sums being of particular importance. We noted that

$$E(X + Y) = E(X) + E(Y)$$

for any random variables X and Y . However, a direct calculation shows that

$$\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) \text{ where}$$

$\text{Cov}(X, Y) = E([X - E(X)][Y - E(Y)])$ denotes the covariance of X and Y .

If X and Y are independent, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. If X and Y are not independent, then they can be partially dependent on one another, or totally dependent on one another. The degree of dependence can be measured by the *correlation coefficient*, ρ .

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}.$$

It is known that $-1 \leq \rho \leq 1$.

We showed that for the random variables $E(X|Y)$ and $E(Y|X)$ that

$$E(E(X|Y)) = E(X) \text{ and}$$

$$E(E(Y|X)) = E(Y).$$

Normal bivariate densities comprise an important family of bivariate densities. Their probability densities in standard form can be written as

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}, \quad -\infty < x < \infty, \quad -\infty < y < \infty.$$

Finally, we considered products and quotients of random variables, finding the distribution functions of these functions of two random variables; by differentiating the result we found their respective probability density functions.

PROBLEMS FOR REVIEW

Exercises 5.2 - # 2, 4, 5, 7, 8

Exercises 5.3 - # 1, 2, 5, 7

Exercises 5.5 - # 2, 6, 9, 10, 13

Exercises 5.6 - # 1, 2

Exercises 5.7 - # 2, 3

SUPPLEMENTARY EXERCISES FOR CHAPTER 5

1. For the joint probability density function $f(x, y) = k(x^2 + y^2)$, $-2 < x < 2, -2 < y < 2$,
 - (a) Find k .

- (b) Are X and Y independent?
 (c) Find the marginal and conditional densities.
2. Given $f(x, y) = k \sin(x) \cos(y) e^{-x}$, $0 < x < \pi/2, -\pi/2 < y < \pi/2$.
- Find k .
 - Find the marginal and conditional densities.
3. Two fair dice are thrown. Let X denote the largest face that appears and Y denote the smallest face that appears.
- Find the joint probability distribution.
 - Find the conditional distribution $f(x|y)$ and show that $E(X|Y = y) = \frac{42-y^2}{13-2y}$, $y = 1, 2, \dots, 6$.
 Then use this result to find $E(X)$.
 - Show that $E(Y|X = x) = \frac{x^2}{2x-1}$, $x = 1, 2, \dots, 6$ and use this result to find $E(Y)$.
 - Explain why $E(X + Y) = 7$.
4. Given $g(x, y) = \left(\frac{1}{12}\right)(3x - 2y + 1)$, $1 < x < 3$, $0 < y < 1$.
- Find the marginal densities.
 - Find the conditional densities.
 - Verify that $E[E[X|Y]] = E[X]$ and that $E[E[Y|X]] = E[Y]$.
5. Suppose that $f(x) = 1$, $0 < x < 1$ and $g(y) = 2y$, $0 < y < 1$. Find $P(Y < X)$ if X and Y are independent.
6. Consider the joint probability density function $f(x, y) = kx$, $x - 1 < y < 1 - x$, $0 < x < 1$, for random variables X and Y .
- Show that $k = 3$.
 - Find the marginal density of Y .
 - Find the variance of X .
7. Suppose $f(x, y) = k$ is a joint probability density function on the area in the first quadrant between the curves $y = 4$ and $y = x^2$.
- Find k .
 - Find $E(X|Y = y)$.
8. Random variables X and Y have $g(y) = \lambda^2 y e^{-\lambda y}$, $y \geq 0$ and $f(x|y) = 1/y$, $0 < x < y$.
- Find the marginal density for X .
 - Find $E[Y|X = x]$.
9. X is an observation from $f(x) = 2x$, $0 < x < 1$ while Y is an observation from the same density but $Y \geq X$. Find $E(Y)$.
10. Show that random variables X and Y are independent of the joint probability density function

$$f(x, y) = k \left(4x - x^2 - 2xy^2 + \left(\frac{1}{2}\right)(xy)^2\right), \quad 0 < x < 2, \quad 0 < y < 1.$$

11. For what value of k is $f(x, y) = k e^{-\frac{x+y}{3}}$ a joint probability density function for $0 \leq x \leq 1$ and $0 \leq y \leq 1$?
12. X and Y have joint probability density function $f(x, y) = k$, $0 < x < 2$, $x^2 < y < 4$.

- (a) Show that $k = \frac{3}{16}$.
 (b) Find the marginal densities $f(x)$ and $g(y)$.
 (c) Find $f(x|y)$.
 (d) Find $E(X|Y = y)$.
13. Random variables X and Y have $g(y|X = x) = \frac{1}{2x}, 0 < y < 2x$ and $f(x) = 24x^2, 0 < x < \frac{1}{2}$.
 (a) Find the joint probability density function, $f(x, y)$.
 (b) Find $g(y)$, the marginal density for Y .
 (c) Are X and Y independent? Justify your answer.
14. Random variables X and Y have $E(X) = -5, E(Y) = 8, E(X^2) = 100, E(Y^2) = 364$, and $\text{Cov}(X, Y) = 100$. Find ρ .
15. The joint probability density function of X and Y is given by
- $$f(x, y) = kx(x - y), \quad 0 < x < 1, \quad -x < y < x.$$
- (a) Find k .
 (b) Are X and Y independent? Support your answer.
 (c) Find $P\left(X > \frac{1}{2} | Y = \frac{1}{4}\right)$.
16. A lot of television sets contain, unknown to the manufacturer, three with defective picture tubes, four with defective sound systems, and five that have no defective parts. Three of the sets are selected, without replacement. Let X denote the number in the sample with defective picture tubes and Y denote the number in the sample with defective sound systems.
 (a) Find the joint probability distribution of X and Y .
 (b) Find the marginal distributions.
 (c) Find the mean and variance of X .
 (d) Find $\text{Cov}(X, Y)$.
 (e) Find ρ .
 (f) Find $\text{Var}(X - Y)$.
17. Random variables X and Y have the following characteristics: $E(X) = 3, \text{Var}(X) = 10, E(Y) = 2, \text{Var}(Y) = 30, E(X \cdot Y) = 4$.
 (a) Find $E(X^2)$.
 (b) Find $\text{Cov}(X, Y)$.
 (c) Find $\rho_{X,Y}$.
 (d) Let $Z = 3X - 2$. Find $\rho_{X,Z}$.
 (e) Find $\text{Var}(5X + 1)$.
 (f) Find $\text{Var}(X - Y)$.
18. Suppose that a fair coin is tossed three times. Let X denote the total number of heads and let Y be the number of heads on just the first toss.
 (a) Find the joint probability distribution of X and Y .
 (b) Calculate $P(X \geq 1 | Y = 0)$.

- (c) Find the marginal distributions of X and Y .
 (d) Calculate the correlation coefficient.
19. Let R be the region bounded by $0 \leq x \leq 1$ and $0 \leq y \leq x^2$, and let the joint probability density function of X and Y be given by
- $$f(x,y) = \begin{cases} kxy & \text{if } (x,y) \text{ is in } R \\ 0 & \text{otherwise.} \end{cases}$$
- (a) Determine k .
 (b) Find the marginal densities, $f(x)$ and $g(y)$.
 (c) Are X and Y independent?
 (d) Calculate $P\left(X > \frac{1}{2}, Y > \frac{1}{4}\right)$.
20. Consider a box that has five white marbles and two yellow marbles. A marble is drawn out and not replaced, and then a second marble is chosen. Random variable X is 1 if the first marble is yellow and 0 otherwise. Random variable Y is 1 if the second marble is yellow and is 0 otherwise.
- (a) What is the joint probability distribution of X and Y ?
 (b) Find the correlation coefficient.
 (c) Are X and Y independent? Explain.
 (d) Find $\text{Var}(X + Y)$.
21. In four tosses of a fair coin let X denote the number of heads and Y denote the longest run of heads. (A run of heads is a sequence of successive heads.)
- (a) Show the joint probability distribution of X and Y in a table and determine the marginal distribution of Y .
 (b) Find $P(X = 2|Y < 2)$.
 (c) Find the expected value of X .
22. Suppose X and Y are independent random variables with marginal densities $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$ and $g(y) = \lambda e^{-\lambda y}$, $y \geq 0$.
- (a) Find the joint probability density function of X and Y .
 (b) Find $E(X \cdot Y)$.
 (c) Find $P(Y > 2X)$.
23. Let X and Y be random variables with probability density function

$$f(x,y) = \frac{k}{\sqrt{xy}}, \quad 0 < x < 1, \quad 0 < y < 1.$$

- (a) Find k .
 (b) Find the marginal densities.
 (c) Find $P\left(Y > \frac{X}{2}\right)$.

24. Let random variables X and Y denote the temperature and time in minutes it takes a diesel engine to start, respectively. The joint density for X and Y is

$$f(x, y) = c(4x + 2y + 1), \quad 0 \leq x \leq 4, \quad 0 \leq y \leq 2.$$

- (a) Find c .
- (b) Find the marginal densities and decide whether or not X and Y are independent.
- (c) Find $f(x|Y = 1)$.

Chapter 6

Recursions and Markov Chains

6.1 INTRODUCTION

In this chapter, we study two important parts of the theory of probability: recursions and Markov chains.

It happens that many interesting probability problems can be posed in a recursive manner; indeed, it is often most natural to consider some problems in this way. While we have seen several recursive functions in our previous work, we have not established their solution. Some of our examples will be recalled here and we will show the solution of some recursions. We will also show some new problems and solve them using recursive functions. In particular, we will devote some time to a class of probability problems known as *waiting time* problems.

Finally, we consider some of the theory of *Markov chains*, which arise in a number of practical situations. This theory is quite extensive, so we are able to give only a brief introduction in this book.

6.2 SOME RECURSIONS AND THEIR SOLUTIONS

We return now to problems involving recursions, considerably expanding our work in this area and considering more complex problems than we have previously.

Consider again the simple problem of counting the number of permutations of n distinct objects, letting P_n denote the number of permutations of these n distinct objects. P_n is, of course, a function of n . Now if we were to permute $n - 1$ of these objects, a new object, the n th, could be placed in any one of the $n - 2$ positions between the objects or in one of the two end positions, a total of n possible positions for the n th object. For a given permutation of the $n - 1$ objects, each one of the n choices for the n th object gives a distinct permutation. This reasoning shows that

$$P_n = n \cdot P_{n-1} \quad n \geq 1. \quad (6.1)$$

Formula (6.1) expresses one value of a function, P_n , in terms of another value of the same function, P_{n-1} . For this reason, (6.1) is called a *recursion* or *recurrence relation* or *difference equation*.

We have encountered recursions several times previously in this book. Recall that in Chapter 1, we observed that the number of combinations of n distinct objects taken r at a

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

time, $\binom{n}{r}$, could be characterized by the recursion

$$\binom{n}{r} = \frac{n-r+1}{r} \cdot \binom{n}{r-1}, \quad r = 1, 2, \dots, n. \quad (6.2)$$

We saw in Chapter 2 that values of the binomial probability distribution, where

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \dots, n,$$

are related by the recursion

$$P(X = x) = \frac{n-x}{x+1} \cdot P(X = x-1), \quad x = 1, 2, \dots, n.$$

This recursion was used to find the mean and the variance of a binomial random variable.

One of the primary values of a recursion is that, given a starting point, any value of the function can be calculated. In the example regarding the permutations of n distinct objects, it would be natural to let $P_1 = 1$. Then we find, by repeatedly applying the recursion, that

$$P_2 = 2P_1 = 2 \cdot 1 = 2,$$

$$P_3 = 3P_2 = 3 \cdot 2 \cdot 1 = 6,$$

$$P_4 = 4P_3 = 4 \cdot 3 \cdot 2 \cdot 1 = 24,$$

and so on. It is easy to conjecture that the general *solution* of the recursion (6.1) is $P_n = n!$. The conjecture can be proved to be correct by showing that it satisfies the original recursion. To do this, we check that

$$P_n = n \cdot P_{n-1} \text{ giving}$$

$$n! = n \cdot (n-1)!,$$

which is true. So we have found a solution for the recursion.

In this example, it is easy to see a pattern arising from some specific cases and this led us to the general solution; soon we will require a specific procedure for determining solutions for recursions where specific cases do not provide a hint of the general solution.

As another example, we saw in Chapter 1 that the solution for Equation (6.2) is $\binom{n}{r} = \frac{n!}{r!(n-r)!}$, where a starting point is $\binom{n}{0} = 1$.

Solutions for recursions are, however, not always so simple.

We want to abandon purely combinatorial examples now and turn our attention to recursions that arise in connection with problems in probability. It happens that many interesting problems can be described by recursions; we will show several of these, together with the solutions of these difference equations.

We begin with an example.

Example 6.2.1

A quality control inspector, thinking that he might make his work a bit easier, decides on the following inspection plan as, either good or nonconforming items come off his assembly

line: if an item is inspected, the next item is inspected with probability p ; if an item is not inspected, the next item will be inspected with probability $1 - p$. The inspector decides to make p small, hoping that he inspects only a few items this way. Does his plan work?

A model of the situation can be constructed by letting a_n denote the probability that the n th item is inspected.

$$\text{So } a_n = P(\text{nth item is inspected}).$$

The n th item will be subject to inspection in two mutually exclusive ways: either the $n - 1$ item is inspected, or it is not. Therefore,

$$a_n = pa_{n-1} + (1 - p) \cdot (1 - a_{n-1}) \text{ for } n \geq 2. \quad (6.3)$$

Letting $a_1 = 0$ (so the first item from the production line is not inspected), Equation (6.3) then gives the following values for some small values of n :

$$\begin{aligned} a_1 &= 0 \\ a_2 &= 1 - p \\ a_3 &= p(1 - p) + (1 - p)p = 2p(1 - p) \\ a_4 &= 2p^2(1 - p) + (1 - p)[1 - 2p(1 - p)] \\ &= 1 - 3p + 6p^2 - 4p^3. \end{aligned}$$

If further values are required, it is probably most sensible to proceed using a computer algebra system, which will calculate any number of these values. This is, in fact, one of the most useful features of a recursion – within reason, any of the values can be calculated from the problem itself. In the short term then, the question may be as valuable as the answer! In some cases, the question is *more* valuable than the answer since the answer may be very complex; in many cases the answer cannot be found at all, so we must be satisfied with a number of special cases.

To return to the problem, how then do the values of a_n behave as n increases? First, we note that if $p = 1/2$, then $a_2 = a_3 = a_4 = 1/2$, prompting us to look at (6.3) when $p = 1/2$. It is easy to see that in that case, $a_n = 1/2$, for all values of n so that if the inspector takes $p = 1/2$, then he will inspect $1/2$ of the items.

The inspector now searches for other values of p , hoping to find some that lead to less work.

A graph of a_{10} as a function of p , found using a computer algebra system, is shown in Figure 6.1.

This shows the inspector that, alas, any reasonable value for p leads to inspection about half the time! The graph indicates that a_{10} is very close to $1/2$ for $\frac{1}{4} \leq p \leq \frac{3}{4}$, but even values of p outside this range only effect the value of a_{10} in the early stages. Figure 6.2 shows a graph of a_n for $p = 1/4$, which indicates the very rapid convergence to $1/2$.

This evidence also suggests writing the solutions for (6.3) in terms of $(p - \frac{1}{2})$. Doing that we find:

$$\begin{aligned} a_1 &= 0 \\ a_2 &= 1 - p = \frac{1}{2} - \left(p - \frac{1}{2}\right) \end{aligned}$$

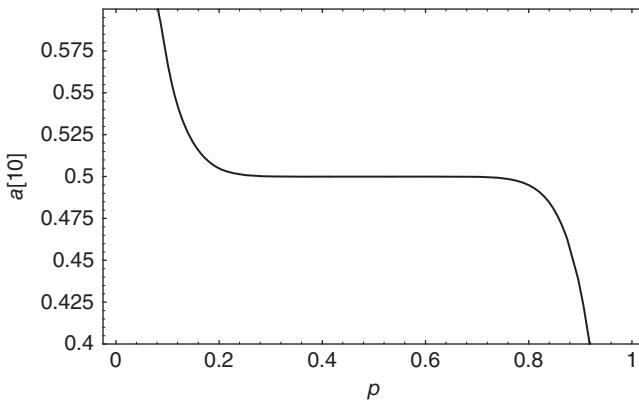


Figure 6.1 a_{10} for the quality control inspector.

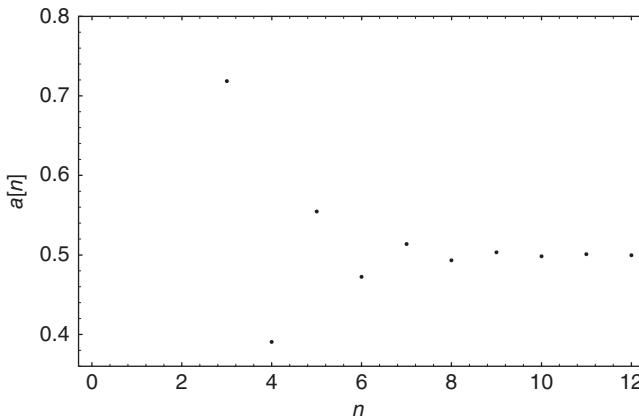


Figure 6.2 a_n for $p = 1/4$.

$$\begin{aligned} a_3 &= 2p(1-p) = \frac{1}{2} - 2\left(p - \frac{1}{2}\right)^2 \\ a_4 &= 1 - 3p + 6p^2 - 4p^3 = \frac{1}{2} - 2^2\left(p - \frac{1}{2}\right)^3. \end{aligned}$$

This strongly suggests that the general solution for (6.3) is

$$a_n = \frac{1}{2} - 2^{n-2}\left(p - \frac{1}{2}\right)^{n-1}.$$

Direct substitution in (6.3) will verify that this conjecture is, in fact, correct. Since $\left|p - \frac{1}{2}\right| < 1$, we see that $a_n \rightarrow \frac{1}{2}$ as $n \rightarrow \infty$. This could also have been predicted from (6.3), since, if $a_n \rightarrow L$, then $a_{n-1} \rightarrow L$ and so, from (6.3),

$$L = pL + (1-p)(1-L),$$

whose solution is $L = \frac{1}{2}$.

Our solution used the fact that the first item was not inspected, and so it may be thought that the long-range behavior of a_n may well be dependent on that assumption. This, however, is not true; in an exercise, the reader will be asked to show that $a_n \rightarrow \frac{1}{2}$ as $n \rightarrow \infty$ if $a_1 = 1$, that is, if the first item is inspected.

We used graphs and several specific values of a_n above to conjecture the solution of the recursion as well as its long-term behavior. While it is often possible to find exact solutions for recursions, often the behavior of the solution for large values of n is of most interest; this behavior can frequently be predicted when exact solutions are not available. However, an exact solution for (6.3) can be constructed, and since the solution is typical of that of many kinds of recursions, it is shown now.

Solution of the Recursion (6.3)

While many computer algebra systems solve recursions, we give here some indication of the algebraic steps involved in this example as well as in others in this chapter.

We begin with the recursion

$$a_n = pa_{n-1} + (1-p) \cdot (1 - a_{n-1}) \text{ for } n \geq 2$$

and write it as

$$a_n - (2p - 1)a_{n-1} = 1 - p, \quad n \geq 2.$$

Note first that the recursion has coefficients that are constants – they are not dependent on the variable n . Note also that the right-hand side of the equation is constant as well. We will consider here recursions having constant coefficients for the variables, but possibly having functions of n on the right-hand side.

The solution of these equations is known to be composed of two parts: the solution of the *homogeneous* equation, in this case

$$a_{n,h} - (2p - 1)a_{n-1,h} = 0,$$

and a *particular solution*, some specific solution of

$$a_{n,p} - (2p - 1)a_{n-1,p} = 1 - p.$$

The general solution is known to be the sum of $a_{n,h}$ and $a_{n,p}$:

$$a_n = a_{n,h} + a_{n,p}.$$

We now show how to determine these two components of the solution.

The homogeneous equation may be written as

$$a_{n,h} = (2p - 1)a_{n-1,h}.$$

This suggests that a value of the function, $a_{n,h}$, is a constant multiple of the previous value, $a_{n-1,h}$. Suppose then that

$$a_{n,h} = r^n \text{ for some constant } r.$$

It follows that

$$r^n - (2p - 1)r^{n-1} = 0,$$

from which we conclude that $r = 2p - 1$. Since the equation is homogeneous, a constant multiple of a solution is also a solution, so

$$a_{n,h} = c(2p - 1)^n \text{ where } c \text{ is some constant.}$$

The equation $r^n - (2p - 1)r^{n-1} = 0$ is often called the *characteristic equation*. The solutions for r are called *characteristics roots*.

The particular solution is some specific solution of

$$a_{n,p} - (2p - 1)a_{n-1,p} = 1 - p.$$

Since the right-hand side is a constant, we try a constant, say k , for a_n . The situation when the right-hand side is a function of n is considerably more complex; we will encounter some of these equations later in this chapter. Substituting the constant k into (6.3) gives

$$k - (2p - 1)k = 1 - p \text{ whose solution is } k = \frac{1}{2}.$$

The complete solution for the recursion is the sum of the homogeneous solution and a particular solution, so

$$a_n = \frac{1}{2} + c(2p - 1)^n,$$

Now $a_1 = 0$, giving $c = -\frac{1}{2(2p-1)}$. Writing $2p - 1 = 2(p - \frac{1}{2})$, we find that

$$a_n = \frac{1}{2} - 2^{n-2} \left(p - \frac{1}{2}\right)^{n-1}, \text{ for } n \geq 1,$$

which is the solution previously found.

The reader who wants to learn more about the solution of recursions is urged to read Grimaldi [16] or Goldberg [14]. We proceed now with some more difficult examples.

Example 6.2.2

We considered, in Chapter 1, the sample space when a loaded coin is flipped until two heads in a row occur. The reader may recall that if the event HH occurs for the first time at the n th toss that the number of points in the sample space for n tosses can be predicted from that of $n - 1$ tosses and $n - 2$ tosses by the Fibonacci sequence. We did not, at that time, discover a formula for the probability that the event will occur for the first time at the n th toss; we will do so now.

Let b_n denote the probability that HH appears for the first time at the n th trial. Now consider a sequence of n trials for which HH appears for the first time at the n th trial. Since such a sequence can begin with either a tail or a head followed immediately by a tail (so that HH does not appear on the second trial),

$$b_n = qb_{n-1} + pqb_{n-2}, n \geq 3 \tag{6.4}$$

is a recursion describing the problem.

We also take $b_1 = 0$ and $b_2 = p^2$.

This recursion gives the following values for some small values of n :

$$\begin{aligned} b_3 &= qp^2 \\ b_4 &= qp^2 \\ b_5 &= q^2 p^2 (1 + p) \\ b_6 &= q^2 p^2 (1 + qp) \\ b_7 &= p^2 q^3 (1 + pq + p + p^2). \end{aligned}$$

It is now very difficult to detect a pattern in the results (although there is one). The behavior of b_n can be seen in the graphs in Figures 6.3 and 6.4 where the values of a_n are shown for a fair coin and then a loaded coin.

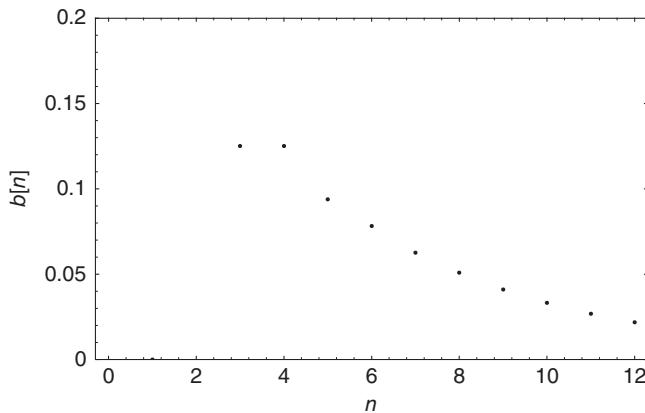


Figure 6.3 b_n for a fair coin.

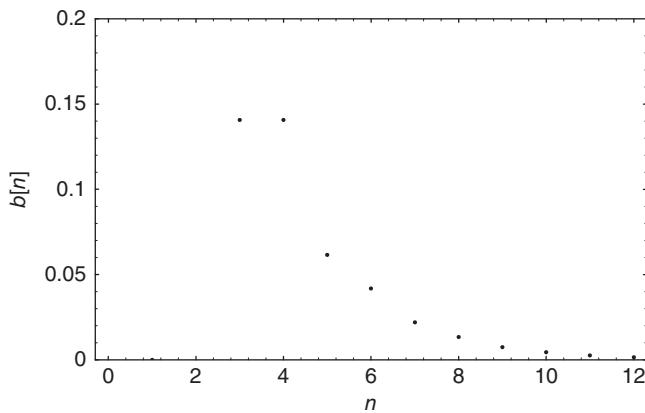


Figure 6.4 b_n for a coin with $p = 3/4$.

We proceed then with the solution of $b_n = qb_{n-1} + pqb_{n-2}$, $n \geq 3$, $b_1 = 0$, $b_2 = p^2$. Here, the equation is homogeneous and we write it as

$$b_n - qb_{n-1} - pqb_{n-2} = 0.$$

The solution in this case is similar to that of Example 6.2.1. If we presume that $b_n = r^n$, the characteristic equation becomes

$$r^2 - qr - pq = 0,$$

giving two distinct characteristic roots, $r = \frac{q \pm \sqrt{q^2 + 4pq}}{2}$. Since the sum of solutions for a linear homogeneous equation must also be a solution, and since a constant multiple of a solution is also a solution, the general solution is

$$b_n = c_1 \left(\frac{q + \sqrt{q^2 + 4pq}}{2} \right)^n + c_2 \left(\frac{q - \sqrt{q^2 + 4pq}}{2} \right)^n.$$

The constants c_1 and c_2 can be determined from the boundary conditions $b_1 = 0$, $b_2 = p^2$, giving

$$\begin{aligned} b_n &= \frac{2p^2}{q\sqrt{q^2 + 4pq} + q^2 + 4pq} \left(\frac{q + \sqrt{q^2 + 4pq}}{2} \right)^n \\ &\quad - \frac{2p^2}{q\sqrt{q^2 + 4pq} - q^2 - 4pq} \left(\frac{q - \sqrt{q^2 + 4pq}}{2} \right)^n. \end{aligned} \quad (6.5)$$

Computer algebra systems often solve recursions. Such a system solves the recursion (6.4) in the case where $p = q = 1/2$ as

$$b_n = \left(\frac{1 + \sqrt{5}}{4} \right)^n \cdot \left(\frac{\sqrt{5} - 5}{10} \right) - \left(\frac{1 - \sqrt{5}}{4} \right)^n \cdot \left(\frac{\sqrt{5} + 5}{10} \right).$$

This result can also be found by substituting $p = q = 1/2$ in (6.5). Other values for p and q make the solution much more complex.

Mean and Variance

The recursion $b_n = qb_{n-1} + pqb_{n-2}$, $n \geq 3$, can be used to determine the mean and variance of N , the number of tosses necessary to achieve two heads in a row. We use a technique that is very similar to the one we used in Chapter 2 to find means and variances of random variables. Multiplying the recursion through by n and summing from 3 to infinity gives

$$\sum_{n=3}^{\infty} nb_n = \sum_{n=3}^{\infty} qnb_{n-1} + \sum_{n=3}^{\infty} pqnb_{n-2}.$$

This becomes

$$\sum_{n=2}^{\infty} nb_n - 2b_2 = q \sum_{n=3}^{\infty} [(n-1)+1]b_{n-1} + pq \sum_{n=3}^{\infty} [(n-2)+2]b_{n-2}.$$

Expanding and simplifying gives

$$E[N] - 2b_2 = qE[N] + q + pqE[N] + 2pq,$$

from which it follows that

$$E[N] = \frac{1+p}{p^2}.$$

If $p = 1/2$, this gives an average waiting time of six tosses to achieve two heads in a row. The result differs a bit from $\frac{1}{p^2}$, a result that might be anticipated from the geometric distribution, but we note that the variable is *not* geometric here.

The variance is calculated in much the same way. We begin with

$$\begin{aligned} \sum_{n=3}^{\infty} n(n-1)b_n &= q \sum_{n=3}^{\infty} [(n-1)(n-2) + 2(n-1)]b_{n-1} \\ &\quad + pq \sum_{n=3}^{\infty} [(n-2)(n-3) + 4(n-2) + 2]b_{n-2}. \end{aligned}$$

Expanding and simplifying gives

$$\begin{aligned} E[N(N-1)] - 2b_2 &= qE[N(N-1)] + 2qE[N] + pqE[N(N-1)] \\ &\quad + 4pqE[N] + 2pq. \end{aligned}$$

It follows then that $\text{Var}(N) = \frac{1+2p-2p^2-p^3}{p^4}$. If $p = 1/2$, $\text{Var}(N) = 22$ tosses.

We turn now to other examples.

Example 6.2.3

Consider again a sequence of Bernoulli trials with p the probability of success at a single trial. In a series of n trials, what is the probability that the sequence SF never appears?

Here a few points in the sample space will assist in seeing a recursion for the probability:

$n = 2 \quad SS$

FS

FF

$n = 3 \quad SSS$

FSS

FFS

FFF

$n = 4 \quad SSSS$

$FSSS$

$FFSS$

$FFFS$

$FFFF.$

It is now evident that a sequence in which SF never appears can arise in one of two mutually exclusive ways: either the sequence is all F 's or, when an S appears, it must then be followed by all S 's. The latter sequences end in S and are preceded by a sequence of $n - 1$ trials in which no sequence SF appears. So, if u_n denotes the probability a sequence of n trials never contains the sequence SF , then

$$\begin{aligned} u_n &= q^n + pu_{n-1}, \quad n \geq 2, \quad u_1 = 1 \quad \text{or} \\ u_n - pu_{n-1} &= q^n, \quad n \geq 2, \quad u_1 = 1. \end{aligned} \quad (6.6)$$

A few values of u_n are as follows:

$$\begin{aligned} u_1 &= 1 \\ u_2 &= q^2 + p \\ u_3 &= q^3 + pq^2 + p^2. \end{aligned}$$

These can be rewritten as

$$\begin{aligned} u_1 &= \frac{q^2 - p^2}{q - p} \quad \text{if } p \neq q, \\ u_2 &= \frac{q^3 - p^3}{q - p} \quad \text{if } p \neq q, \\ u_3 &= \frac{q^4 - p^4}{q - p} \quad \text{if } p \neq q. \end{aligned}$$

This leads to the conjecture that $u_n = \frac{q^{n+1} - p^{n+1}}{q - p}$, $n \geq 1$, $p \neq q$. The validity of this can be seen by substituting u_n into the original recursion (6.6). Substituting in the right-hand side of (6.6), we find, provided that $p \neq q$,

$$q^n + p \cdot \frac{q^n - p^n}{q - p} \quad \text{and this simplifies to}$$

$$\frac{q^{n+1} - p^{n+1}}{q - p}, \quad \text{verifying the solution.}$$

The solution of the recursion (6.6) is also easy to construct directly. The characteristic equation is

$$r^n - p \cdot r^{n-1} = 0,$$

giving the characteristic root $r = p$. Therefore, the homogeneous solution is

$$u_{n,h} = c \cdot p^n.$$

Now we seek a particular solution.

Since the right-hand side of $u_n - pu_{n-1} = q^n$ is q^n , suppose that we try

$$u_{n,p} = k \cdot q^n.$$

Then, substituting in the recursion, we have

$$k \cdot q^n - k \cdot p \cdot q^{n-1} = q^n$$

and we find that

$$k = \frac{q}{q-p}, \text{ provided that } q \neq p.$$

So $u_{n,p} = \frac{q^{n+1}}{q-p}$, $q \neq p$, and the general solution is

$$u_n = cp^n + \frac{q^{n+1}}{q-p}, \text{ provided that } q \neq p.$$

By imposing the condition $u_1 = 1$ and simplifying, we find, as before, that

$$u_n = \frac{q^{n+1} - p^{n+1}}{q-p}, \quad n \geq 1, \quad q \neq p. \quad (6.7)$$

We now investigate the case when $p = q$. In that case, (6.6) becomes

$$u_n - \frac{1}{2}u_{n-1} = \left(\frac{1}{2}\right)^n. \quad (6.8)$$

Now the homogeneous solution is

$$u_{n,h} = c \cdot \left(\frac{1}{2}\right)^n,$$

so a particular solution $u_{n,p} = \left(\frac{1}{2}\right)^n$, a natural choice for $u_{n,p}$, will only produce 0 when substituted into the left-hand side of (6.8). We try the function

$$u_{n,p} = k \cdot n \cdot \left(\frac{1}{2}\right)^n$$

for the particular solution.

Then the left-hand side of (6.8) becomes

$$k \cdot n \cdot \left(\frac{1}{2}\right)^n - \frac{1}{2} \cdot k \cdot (n-1) \cdot \left(\frac{1}{2}\right)^{n-1}.$$

This simplifies to $k \cdot \left(\frac{1}{2}\right)^n$; so $k = 1$, and we have found then the general solution:

$$u_n = c \cdot \left(\frac{1}{2}\right)^n + n \cdot \left(\frac{1}{2}\right)^n.$$

The boundary condition $u_1 = 1$ gives $c = 1$. The general solution in this case is

$$u_n = (n+1) \left(\frac{1}{2}\right)^n.$$

This solution can also be found by using L'Hospital's rule in (6.7). In this case, we have

$$\lim_{p \rightarrow 1/2} \frac{q^{n+1} - p^{n+1}}{q-p} = \lim_{p \rightarrow 1/2} \frac{(1-p)^{n+1} - p^{n+1}}{1-2p}$$

$$\begin{aligned}
 &= \lim_{p \rightarrow 1/2} \frac{-(n+1)(1-p)^n - (n+1)p^n}{-2} \\
 &= (n+1) \left(\frac{1}{2}\right)^n.
 \end{aligned}$$

EXERCISES 6.2

- In Example 6.2.1, show that $a_n \rightarrow \frac{1}{2}$ as $n \rightarrow \infty$ if $a_1 = 1$ (the first item is inspected).
- Solve the recursion $L_n = L_{n-1} + L_{n-2}$ where $L_0 = 2$ and $L_1 = 1$. The result determines the Lucas numbers.

Exercises 3–5 refer to a sequence of Bernoulli trials, where p is the probability of an event and $p + q = 1$.

- Describe a problem for which the recursion $a_n = qa_{n-1}$, $n > 1$, where $a_1 = 1 - q$ is appropriate. Then solve the recursion verifying that it does in fact describe the problem.
- Let a_n denote the probability that a sequence of Bernoulli trials with probability of success p has an odd number of successes.
 - Show that $a_n = p(1 - a_{n-1}) + qa_{n-1}$, for $n \geq 1$ if $a_0 = 0$. [Hint: Condition on the result of the first toss.]
 - Solve the recursion in part (a).
- (a) Find a recursion for the probability, a_n , that at least two successes occur in n Bernoulli trials.
 - In part (a), let $p = 1/2$. Show that $a_n = 1 - \frac{f_{n+2}}{2^n}$, where f_n is the n th term in the Fibonacci sequence 1, 1, 2, 3, 5, 8, ...
- Two machines in a manufacturing plant produce items that are either good (g) or unacceptable (u). Machine 1 has produced g good and u unacceptable items, while the situation with machine 2 is exactly the reverse; it has produced u good items and g unacceptable items. An inspector is required to sample the production of the machines, and, to achieve a random order of items from each of the machines, the following plan is devised:
 - The first item is drawn from the output of machine 1.
 - Drawn items are returned to the output of the machine from which they were drawn.
 - If a sampled item is good, the next item is drawn from the first machine. If the sampled item is unacceptable, the next item is drawn from the second machine.

What is the probability that the n th sampled item is good?

- A basketball player makes a series of free throw attempts. If he makes a shot, he makes the next one also with probability p_1 . However, if he misses a shot, he makes the next one with probability p_2 where $p_1 \neq p_2$. If he makes the first shot, what is the probability he makes the n th shot?
- A coin loaded to come up heads with probability p is tossed until the sequence TH occurs for the first time. Let a_n denote the probability that the sequence TH occurs for the first time at the n th toss.
 - Show that $a_n = pa_{n-1} + pq^{n-1}$, if $n \geq 3$ where $a_1 = 0$ and $a_2 = pq$.
 - Show that the average waiting time for the first occurrence of the sequence TH is $1/pq$.
 - If $p = q = 1/2$, show that $a_n = \frac{n-1}{2^n}$, $n \geq 2$.

9. A party game starts with the host saying “yes” to the first person. This message is passed to the other guests in this way: if a person hears “yes,” he passes that on with probability 3/4; however, if a person hears “no”, that is passed on with probability 2/3.
- What is the probability the n th person hears “yes”?
 - Suppose we want the probability that the seventh person hears “yes” to be about 1/2. What should be the probability that a “no” response is correctly passed on?
10. A coin has a 1 marked on one side and a 2 on the other side. It is tossed repeatedly, and the cumulative sum that has occurred is recorded. Let p_n denote the probability that the sum is n at some time during the course of the game.
- Show a sample space for several possible sums. Explain why the number of points in the sample space follows the Fibonacci sequence.
 - Find an expression for p_n assuming the coin is fair.
 - Show that $p_n \rightarrow 2/3$ as $n \rightarrow \infty$.
 - How should the coin be loaded so as to make p_{17} , the probability that the sum becomes 17 at some time, as large as possible?
11. Find the mean for the waiting time for the pattern HHH in tossing a coin loaded to come up heads with probability p .

6.3 RANDOM WALK AND RUIN

We now show an interesting application of recursions and their solutions, namely, that of *random walk* problems.

The theory of random walk problems is applicable to problems in many fields. We begin with a gambling situation to illustrate one approach to the problem. (Another avenue of approach to the problem will be shown in the sections on Markov chains later in this chapter.)

Suppose a gambler is playing a game against an opponent where, at each trial, the gambler wins \$1 from his opponent or loses \$1 to his opponent. If in the course of play, the gambler loses all his money, then he is ruined; on the other hand, if he wins all his opponent’s money, he wins the game. We want to find then the probability, a_g , that the player (the *gambler*) wins the game with an initial fortune of $\$g$ while his opponent (the *house*) initially has $\$h$.

While the probability of winning at any particular trial is of obvious importance, we will see that, in addition to this, the probability the gambler wins the game is highly dependent on the amount of money with which he starts, as well as the amount of money the house has.

The game can be won with a fortune of $\$n$ under two mutually exclusive circumstances: the gambler wins the next trial (say with probability p) and goes on to win the game with a fortune of $\$(n + 1)$, or he loses the next trial with probability $1 - p = q$ and subsequently wins the game with a fortune of $\$(n - 1)$. This leads to the recursion

$$a_n = pa_{n+1} + qa_{n-1}, \quad a_0 = 0, \quad a_{g+h} = 1.$$

The characteristic equation is $pr^2 - r + q = 0$, which has roots of 1 and q/p .

Assuming that $p \neq q$, the solution is then of the form

$$a_n = A + B \cdot \left(\frac{q}{p}\right)^n.$$

Using the fact that $a_0 = 0$ gives $0 = A + B$, so $a_n = A \left[1 - \left(\frac{q}{p} \right)^n \right]$.

Using the fact that $a_{g+h} = 1$ produces the solution

$$a_n = \frac{1 - \left(\frac{q}{p} \right)^n}{1 - \left(\frac{q}{p} \right)^{(g+h)}}, \quad q \neq p.$$

So, in particular,

$$a_g = \frac{1 - \left(\frac{q}{p} \right)^g}{1 - \left(\frac{q}{p} \right)^{(g+h)}}, \quad q \neq p. \quad (6.9)$$

Formula (6.9) gives some interesting numerical results. Suppose that $p = 0.49$ so that the game is slightly unfavorable to the gambler, and that the gambler initially has $g = \$10$. The following table shows the probability the gambler wins the game for various fortunes ($\$h$) for his opponent:

$\$h$	Probability of winning
\$10	0.401300
\$14	0.305146
\$18	0.238174
\$22	0.189394
\$26	0.152694
\$30	0.124404.

One conclusion that can be drawn from the table is that the probability the gambler wins drops rapidly as his opponent's fortune increases. Figure 6.5 shows graphs of the probability of winning with $\$g$ as a function of the opponent's fortune, $\$h$.

Although the game is slightly (and only slightly) adverse for the gambler, the gambler still has, under some combinations of fortunes, a remarkably large probability of winning the game. If the opponent's fortune increases, however, that probability becomes very small

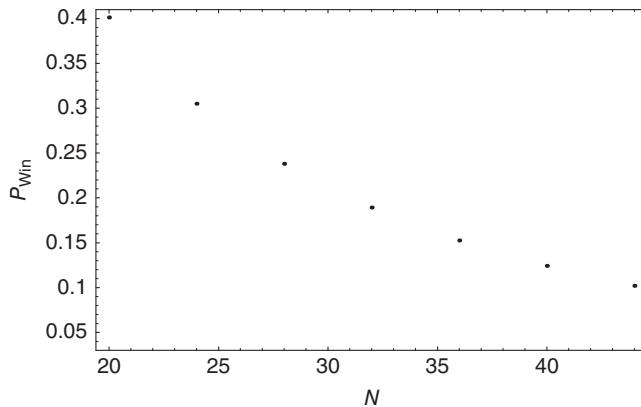


Figure 6.5 Probability of winning the gamblers' ruin with an initial fortune of \$10 against an opponent with an initial fortune of $\$N$ with $p = 0.49$.

very quickly. The following table shows, for $p = 0.49$, the probability that a gambler with an initial fortune of \$10 wins the game over an opponent with an initial fortune of $\$h$:

$\$h$	Probability of winning
\$90	0.00917265
\$98	0.00662658
\$106	0.00479399
\$114	0.00347174
\$122	0.00251603
\$130	0.00182438.

Now the gambler has little chance of winning the game, but his best chance occurs when his opponent has the least money or, equivalently, when the ratio of the gambler's fortune to that of his opponent is as large as possible. It is interesting to examine the surface generated by various initial fortunes and values of p . This surface, for $h = \$30$, is shown in Figure 6.6. The contours of this surface appear in Figure 6.7.

The chance of winning the game then is very slim, but that is because the gambler must exhaust his opponent's fortune and he has little hope of doing this. While the game is slightly unfavorable to the gambler, the real problem lies in the fact that the ratio of the gambler's fortune to that of his opponent is small. When this ratio increases, so does the gambler's chance of winning. These observations suggest two courses of action:

1. the gambler revises his plans and quits playing the game when he has achieved a certain fortune (not necessarily that of his opponent) or
2. the gambler bets larger amounts on each play of the game.

Either strategy will increase the player's chance of meeting his goals.

For example, given the game where $p = 0.49$ and an initial fortune of \$10, the gambler's chance of doubling his money to \$20 is 0.4013, regardless of the fortune of his opponent.

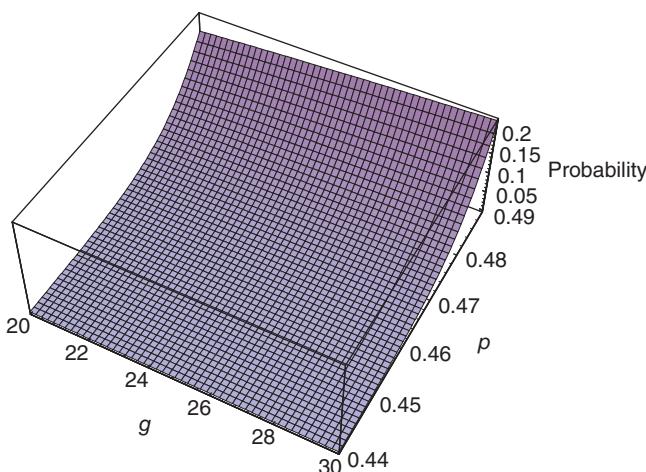


Figure 6.6 Probability of winning the gambler's ruin. The player has an initial fortune of $\$g$. p is the probability an individual game is won. The opponent initially has \$30.

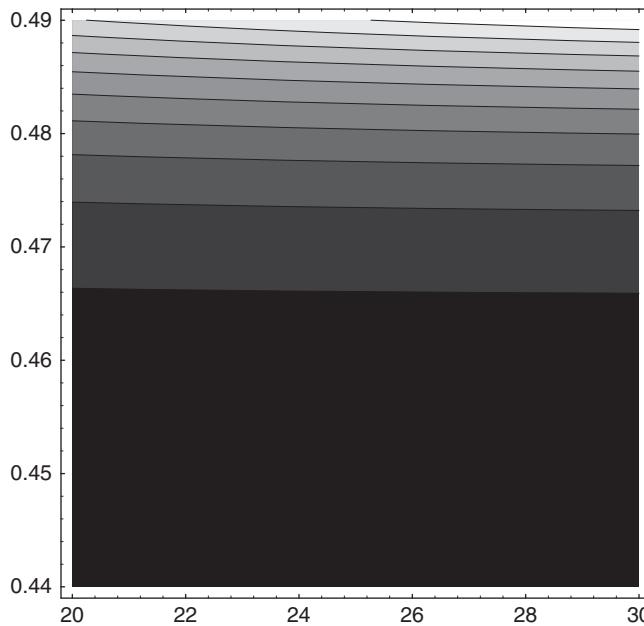


Figure 6.7 A contour plot of the gambler's ruin problem.

The probability that a player with \$50 will reach \$60 in a game where he has probability 0.49 of winning on any play is about 0.64. Clearly, the lesson here is that modest goals have a fair chance of being achieved, but the gambler must stop playing when he has reached them.

The following table shows the probability of winning a game against an opponent with initial fortune of \$100 when the bet on each play is \$25. The player's initial fortune is \$ g . Again, $p = 0.49$.

\$ g	Probability of winning
\$25	0.184326
\$50	0.307047
\$75	0.394564

Betting as much as you can in gambling games unfavorable to the gambler can then be combined with alternative strategies as the gambler's fortune increases (if, in fact, it does!) to increase the gambler's chance of winning the game.

We turn now to the expected duration of the game.

Expected Duration of the Game

Let E_n denote the expected duration of the game if the gambler has \$ n . Since winning or losing the next trial increases E_n by 1,

$$E_n = pE_{n+1} + qE_{n-1} + 1, \quad E_0 = 0, \quad E_{g+h} = 0.$$

This recursion is very similar to that for a_n , differing only in the boundary conditions and in the appearance of the constant 1. The characteristic roots are 1 and q/p again, and so the

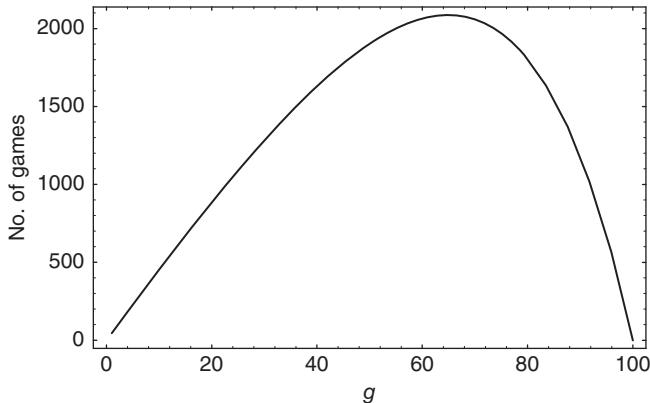


Figure 6.8 Expected duration of the gambler's ruin when the gambler initially has \$ g and the house has \$(100 - g).

solution is of the form

$$E_n = A + B \left(\frac{q}{p} \right)^n + C \cdot n, \quad q \neq p.$$

Here, the term $C \cdot n$ represents the particular solution since a constant is a solution to the homogeneous equation.

The constant C must satisfy the equation $Cn - pC(n+1) - qC(n-1) = 1$, so $C = \frac{1}{q-p}$. The boundary conditions are then imposed giving the result

$$E_n = \frac{n}{q-p} - \frac{g+h}{q-p} \frac{1 - \left(\frac{q}{p} \right)^n}{1 - \left(\frac{q}{p} \right)^{g+h}}, \quad q \neq p.$$

In particular, since the gambler starts with \$ g ,

$$E_g = \frac{g}{q-p} - \frac{g+h}{q-p} \frac{1 - \left(\frac{q}{p} \right)^g}{1 - \left(\frac{q}{p} \right)^{g+h}}, \quad q \neq p.$$

Some particular results from this formula may be of interest. Assume again that the game is slightly unfavorable to the gambler, so that $p = 0.49$ and assume that $N = \$100$. The game then has expected length 454 trials if the gambler starts with \$10.

With how much money should the gambler start in order to maximize the expected length of the game? If the gambler and the house have a combined fortune of \$100, a computer algebra system shows that the maximum expected length of the series is about 2088 games, occurring when $g = \$65$ (and $h = \$35$). A graph of the expected duration of the game if $g + h = \$100$ is shown in Figure 6.8.

EXERCISES 6.3

- Find the solution to the random walk and ruin problem if $p = 1/2$.
- Find the expected duration of the gambler's ruin game if $p = 1/2$.
- Show a graph of the probability of winning the gambler's ruin game if the game is favorable to the gambler with $p = 0.51$.

4. Show a graph of the expected duration of the gambler's ruin game when the game is favorable to the gambler, say $p = 0.51$.
5. Show that if $p = 0.49$ and $h = \$116$, the probability the gambler will be ruined is at least 0.99, regardless of the amount of money the gambler has.

6.4 WAITING TIMES FOR PATTERNS IN BERNOULLI TRIALS

In Chapter 1, we considered a game in which a fair coin is thrown until one of two patterns occurs: TH or HH . We concluded that the game is unfair since TH occurs before HH with probability $3/4$. We also considered other problems in Chapter 2, which we call *waiting time problems* such as waiting for a single Bernoulli success. In Section 6.2, we considered waiting for two successes in a row. We now want to consider more general waiting time problems, seeking probabilities as well as average waiting times.

Some of these problems may well reveal solutions that are counterintuitive; we have noted that the average waiting time for HH with a fair coin is 6 tosses; the average waiting time for TH is 4 tosses, results that can easily be verified by simulation. This is a very surprising result for a fair coin; many people suspect that the average waiting time for these patterns for a fair coin should be the same, but they actually differ.

We now show how to determine probability generating functions from recursions. Since the direct construction of recursions for first-time occurrences involving complex patterns is difficult, we show how to arrive at the recursions by first creating a generating function for occurrence times. Then we use this generating function to find a recursion for first-occurrence times. The general technique will be illustrated by an example.

Example 6.4.1

A fair coin is thrown until the pattern THT occurs for the first time. On average, how many throws will this take?

First, let us look at the sample space. Let n be the number of tosses necessary to observe the pattern THT for the first time.

$$n = 3 \quad THT$$

$$n = 4 \quad TTHT$$

$$HTHT$$

$$n = 5 \quad TTTHT$$

$$HHTHT$$

$$HTTHT$$

$$n = 6 \quad TTTTHT$$

$$HHHTHT$$

$$HHTTHT$$

$$THHTHT$$

$$HTTTHT$$

For $n = 7$, there are 9 sequences, while there are 16 sequences if $n = 10$. The pattern in the sequence 1, 2, 3, 5, 9, 16, ... may prove difficult for the reader to discover; we will find it later in this section. In any event, the task of identifying points for $n = 28$, say, is very difficult to say the least (there are 1,221,537 points to enumerate!).

Before solving the problem, let us first consider a definition of when the pattern *THT occurs*. In a sequence of throws, we examine the sequence from the beginning and note when we see the pattern *THT*; we say the pattern *occurs*; the examination of the sequence then begins again starting with the next throw. For example, in the sequence

$$HHTHTHTHHTHT,$$

the pattern *THT* occurs on the 5th and 12th throws; it does not occur on the 7th throw. This agreement, which might well strike the reader as strange, is necessary for some simple results that follow.

Let us suppose then that the pattern *THT occurs* on the n th throw. (This is not necessarily the *first* occurrence of the pattern.) Let u_n denote the probability that the pattern *THT* occurs at the n th throw; consider a sequence in which *THT* occurs at the n th throw. *THT* then either occurs at the n th trial or it occurs at the $n - 2$ trial, followed by *HT*. Since any sequence ending in *THT* has probability pq^2 , it follows that

$$u_n + pq \cdot u_{n-2} = pq^2, \quad n \geq 3. \quad (6.10)$$

We take $u_0 = 1$ (again so that a subsequent result is simple) and of course, $u_1 = 0$ and $u_2 = 0$.

Results from the recursion are interesting. We find, for example, that

$$\begin{aligned} u_{15} &= pq^2(1 - pq + p^2q^2 - p^3q^3 + p^4q^4 - p^5q^5 + p^6q^6) \\ &= \frac{pq^2[1 + (pq)^7]}{1 + pq}. \end{aligned}$$

In general, an examination of special values using a computer algebra system leads to the conjecture that

$$u_{2n} = u_{2n-1} = \frac{pq^2[1 - (-pq)^{n-1}]}{1 + pq}, \quad n = 1, 2, 3, \dots$$

This in fact will satisfy (6.10) and is its general solution.

It is easy to see from this that

$$u_{2n} \rightarrow \frac{pq^2}{1 + pq},$$

since $pq < 1$ and so $\lim_{n \rightarrow \infty} (pq)^{n-1} = 0$.

This result can also easily be found from (6.10) since, if $u_n \rightarrow L$, say then $u_{n-2} \rightarrow L$ as well. So in that case, we have

$$L + p \cdot q \cdot L = p \cdot q^2$$

whose solution is

$$L = \frac{pq^2}{1 + pq}.$$

Generating Functions

Now we seek a generating function for the sequence of u'_n 's.

Let $U(s) = \sum_{n=0}^{\infty} u_n s^n$ be the generating function for the u_n 's. Multiplying both sides of (6.10) by s^n and summing, we have

$$\sum_{n=3}^{\infty} u_n s^n + \sum_{n=3}^{\infty} u_{n-2} pq s^n = pq^2 \sum_{n=3}^{\infty} s^n.$$

Expanding and simplifying gives

$$U(s) - u_2 s^2 - u_1 s - u_0 + s^2 pq [U(s) - u_0] = \frac{pq^2 s^3}{1-s},$$

from which it follows that

$$U(s) = 1 + \frac{pq^2 s^3}{(1-s)(1+pq s^2)}.$$

Now let $F(s)$ denote a generating function for f_n , the probability that *THT* occurs for the first time at the n th trial. It can be shown (see Feller [7]) that

$$F(s) = 1 - \frac{1}{U(s)}.$$

In this case, we have

$$F(s) = \frac{pq^2 s^3}{1-s+pq s^2-p^2 q s^3}.$$

A power series expansion of $F(s)$, found using a computer algebra system, gives the following:

$$\begin{aligned} f_3 &= pq^2 \\ f_4 &= pq^2 \\ f_5 &= pq^2(1-pq) \\ f_6 &= pq^2(1-2pq+qp^2). \end{aligned}$$

Now we have a generating function whose coefficients give the probabilities of first-occurrence times. One could continue to find probabilities from this generating function using a computer algebra system, but we show now how to construct a recurrence from the generating function for $F(s)$.

$$\text{Let } F(s) = \frac{pq^2 s^3}{1-s+pq s^2-p^2 q s^3} = f_0 + f_1 s + f_2 s^2 + f_3 s^3 + f_4 s^4 + \dots$$

It follows that

$$pq^2 s^3 = (1-s+pq s^2-p^2 q s^3)(f_0 + f_1 s + f_2 s^2 + f_3 s^3 + f_4 s^4 + \dots).$$

By equating coefficients, we have

$$\begin{aligned}f_0 &= 0 \\f_1 &= 0 \\f_2 &= 0 \\f_3 &= pq^2\end{aligned}$$

and for $n \geq 4$, $f_n = f_{n-1} - pqf_{n-2} + p^2qf_{n-3}$. (6.11)

We have succeeded in finding a recursion for first-occurrence times. The reader with access to a computer algebra system will discover some interesting patterns in the formulas for the f_n . Numerical results are also easy to find. For the case $p = q$, f_{20} , the probability that *THT* will be seen for the first time at the 20th trial is only 0.01295.

Finally, we find the pattern in the number of points in the sample space for first-time occurrences. Let w_n denote the number of ways *THT* can occur for the first time in n trials. Since, when $p = q = 1/2$,

$$f_n = \frac{w_n}{2^n},$$

then

$$w_n = 2w_{n-1} - w_{n-2} + w_{n-3}, \quad n \geq 6,$$

where $w_3 = 1$, $w_4 = 2$, and $w_5 = 3$, using (6.11).

Average Waiting Times

The recurrence (6.11) can be used to find the mean waiting time for the first occurrence of the pattern *THT*. By multiplying through by n and summing, we find that

$$\begin{aligned}\sum_{n=4}^{\infty} n f_n &= \sum_{n=4}^{\infty} [(n-1)+1]f_{n-1} - pq \sum_{n=4}^{\infty} [(n-2)+2]f_{n-2} \\&\quad + p^2q \sum_{n=4}^{\infty} [(n-3)+3]f_{n-3}.\end{aligned}$$

This can be simplified as

$$E(N) - 3pq^2 = E(N) + 1 - pq[E(N) + 2] + p^2q[E(N) + 3],$$

from which we find that

$$E(N) = \frac{1+pq}{pq^2}.$$

For a fair coin, the average waiting time for *THT* to occur is 10 trials.

Means and Variances by Generating Functions

The technique used earlier to find the average waiting time can also be used to find the variance of the waiting times. This was illustrated in Example 6.2.2. Now, however, we have the probability generating function for first-occurrence times, so it can be used to determine means and variances.

A computer algebra system will show that $F'(1) = \frac{1+pq}{pq^2}$ and that

$$F''(1) = \frac{2(q + 4p^2 - 2p^3 - 2p^4 + p^5)}{p^2 q^4},$$

and since

$$\sigma^2 = F''(1) + F'(1) - [F'(1)]^2,$$

we find that

$$\sigma^2 = \frac{1 + 2pq - 5pq^2 + p^2q^2 - p^2q^3}{p^2q^4}.$$

With a fair coin then the average number of throws to see *THT* is 10 throws with variance 58.

EXERCISES 6.4

1. In Bernoulli trials with p the probability of success at any trial, consider the event *SSS*. Let u_n denote the probability that *SSS* occurs at the n th trial.
 - (a) Show that $u_n + pu_{n-1} + p^2u_{n-2} = p^3$, $n \geq 4$, and establish the boundary conditions.
 - (b) Find the generating function, $U(s)$, from the recursion in part (a). Use $U(s)$ to determine the probability that *SSS* occurs at the 20th trial if $p = 1/3$.
 - (c) Find $F(s)$, the generating function for first-occurrence times of the pattern *SSS*. Again, if $p = 1/3$, find the probability that *SSS* occurs for the first time at the 20th trial.
 - (d) Establish that w_n , the number of ways the pattern *SSS* can occur in n trials for the first time, is given by

$$w_n = w_{n-1} + w_{n-2} + w_{n-3},$$

for an appropriate range of values of n . This would apparently establish a “super-Fibonacci” sequence.

2. In waiting for the first occurrence of the pattern *HTH* in tossing a fair coin, we wish to create a fair game. We would like the probability the event occurs for the first time in n or fewer trials to be $1/2$. Find n .
3. Find the variance of N , the waiting time for the first occurrence of *THT*, with a loaded coin.
4. Suppose we wait for the pattern *TTHTH* in Bernoulli trials.
 - (a) Find a recursion for the probability of the occurrence of the pattern at the n th trial.
 - (b) Find a generating function for occurrence times and, from that, a recurrence for first-occurrence times.
 - (c) Find the mean and variance of first-occurrence times of the pattern.

6.5 MARKOV CHAINS

In practical discrete probability situations, Bernoulli trials probably occur with the greatest frequency. Bernoulli trials are assumed to be independent of constant probability of success from trial to trial; these assumptions lead to the binomial random variable. Can the binomial be generalized in any way? One way to do this is to relax the assumption of independence, so that the outcome of a particular trial is dependent on the outcomes of the previous trials. The simplest situation assumes that the outcome of a particular trial is dependent only on the outcome of the immediately preceding trial. Such trials form what are called, in honor of the Russian mathematician, a *Markov chain*.

While relaxing the assumption of independence and, in addition, making only the outcome of the previous trial an influence on the next trial, seem simple enough, they lead to a very complex, although beautiful, theory. We will consider only some of the simpler elements of that theory here and will frequently make statements without proof, although we will make them plausible.

Example 6.5.1

A gambler plays on one of four slot machines each of which has probability 1/10 of paying off with some reward. If the player wins on a particular machine, she continues to play on that machine; however, if she loses on any machine, she chooses one of the other machines with equal probability.

For convenience, number the machines 1, 2, 3, and 4. Here, we are interested in the machine being played at the moment, and that is a function of whether or not the player won on the last play. Now let p_{ij} denote the probability that machine j is played immediately following playing machine i . We call this a *transition probability* since it gives the probability of going from machine i to machine j .

Now we find some of these transition probabilities. For example, $p_{12} = \frac{9}{10} \cdot \frac{1}{3} = \frac{3}{10}$ since the player must lose with machine 1 and then switch to machine 2 with probability 1/3.

Also $p_{33} = 1/10$ since the player must win with machine 3 and then stay with that machine, while $p_{42} = 3/10$, the calculation being exactly the same as that for p_{12} .

The remaining transition probabilities are equally easy to calculate in this case. It is most convenient to display these transition probabilities as a matrix, T , whose entries are p_{ij} :

$$T = T[p_{ij}] = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} \frac{1}{10} & \frac{3}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{3}{10} & \frac{1}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{3}{10} & \frac{3}{10} & \frac{1}{10} \end{pmatrix} \end{matrix}$$

This transition matrix is called *stochastic* since the row sums are each 1. Of course that has to be true since the player either stays with the machine currently being played or moves to some other machine for the next play. Actually, T is *doubly stochastic* since its column sums are 1 as well, but such matrices will not be of great interest to us. Matrices describing Markov chains, however, must be stochastic.

The course of the play is of interest, so we might ask, “What is the probability the player moves from machine 2 to machine 4 after two plays?” We denote this probability by $p_{24}^{(2)}$.

Since the transition from machine 2 to machine 4 involves two plays, the player either goes to machine 2 from machine 1 or machine 2 or machine 3 or machine 4, and on the second play moves to machine 4, we see that

$$\begin{aligned} p_{24}^{(2)} &= p_{21}p_{14} + p_{22}p_{24} + p_{23}p_{34} + p_{24}p_{44} \\ &= \frac{3}{10}\frac{3}{10} + \frac{1}{10}\frac{3}{10} + \frac{3}{10}\frac{3}{10} + \frac{3}{10}\frac{1}{10} = \frac{6}{25}. \end{aligned}$$

But this product is simply the dot product of the second row of T with the fourth column of T – an entry of the matrix product of T with itself. Hence, $T^2 = [p_{ij}^{(2)}]$ where T^2 denotes the usual matrix product of T with itself. The reader should check that the remaining entries of T^2 give the proper two-step transition probabilities. We have

$$T^2 = \begin{pmatrix} \frac{7}{25} & \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{6}{25} & \frac{7}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{6}{25} & \frac{6}{25} & \frac{7}{25} & \frac{6}{25} \\ \frac{6}{25} & \frac{6}{25} & \frac{6}{25} & \frac{7}{25} \end{pmatrix}.$$

The entries of T^n then represent transition probabilities in n steps. A computer algebra system is handy in finding these powers. In this case, we find that

$$T^4 = \begin{pmatrix} \frac{157}{625} & \frac{156}{625} & \frac{156}{625} & \frac{156}{625} \\ \frac{156}{625} & \frac{157}{625} & \frac{156}{625} & \frac{156}{625} \\ \frac{156}{625} & \frac{156}{625} & \frac{157}{625} & \frac{156}{625} \\ \frac{156}{625} & \frac{156}{625} & \frac{156}{625} & \frac{157}{625} \end{pmatrix}.$$

Each of the entries in T^4 is now very close to $1/4$, and we conjecture that

$$T^n \rightarrow \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

This shows, provided that our conjecture is correct, that the player will play each of the machines about $1/4$ of the time as time goes on.

The vector $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ is called a *fixed vector* for the matrix T since

$$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \begin{pmatrix} \frac{1}{10} & \frac{3}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{3}{10} & \frac{1}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{3}{10} & \frac{3}{10} & \frac{1}{10} \end{pmatrix} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

We say that a nonzero vector, w , is a *fixed vector* for the matrix T if

$$wT = w.$$

Many (but not all) transition matrices have fixed vectors, and when they do have fixed vectors the components are rarely equal, unlike the case with the matrix T . The fixed vector, if there is one, shows the *steady state* of the process under consideration.

Is the constant vector with each entry $1/4$ a function of the probability, $1/10$, of staying with a winning machine? To answer this, suppose p is the probability the player stays with a winning machine and switches then with probability $\frac{1-p}{3}$ to each of the other machines. The transition matrix is then

$$P = \begin{pmatrix} p & \frac{1-p}{3} & \frac{1-p}{3} & \frac{1-p}{3} \\ \frac{1-p}{3} & p & \frac{1-p}{3} & \frac{1-p}{3} \\ \frac{1-p}{3} & \frac{1-p}{3} & p & \frac{1-p}{3} \\ \frac{1-p}{3} & \frac{1-p}{3} & \frac{1-p}{3} & p \end{pmatrix}.$$

We find that the vector $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ is a fixed point for the matrix P for any $0 < p < 1$, so the player will use each of the machines about $1/4$ of the time, regardless of the value for p !

It is common to refer to the possible positions in a Markov chain as *states*. In this example, the machines are the states and the transition probabilities are probabilities of moving from state to state. The states in Markov chains often represent the outcomes of the experiment under consideration.

Example 6.5.2

Consider the transition matrix with three states

$$R = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Calculation will show that powers of R approach the matrix

$$\begin{pmatrix} \frac{6}{17} & \frac{8}{17} & \frac{3}{17} \\ \frac{6}{17} & \frac{8}{17} & \frac{3}{17} \\ \frac{6}{17} & \frac{8}{17} & \frac{3}{17} \end{pmatrix}.$$

It will also be found that the solution of

$$(a, b, c) \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} = (a, b, c)$$

with the restriction that $a + b + c = 1$ has the solution $\left(\frac{6}{17}, \frac{8}{17}, \frac{3}{17}\right)$, so R has a fixed vector also.

The examples above illustrate the remarkable fact that the powers of some matrices approach a matrix with equal rows. We can be more specific now about the conditions under which this happens.

Definition We call a matrix T *regular* if, for some n , the entries of T^n are all positive (no zeroes are allowed).

In the earlier examples, T and R are regular. We now state a theorem without proof.

Theorem: If T is a regular transition matrix, then the powers of T , $T^n, n \geq 1$, each approach a matrix all of whose rows are the same probability vector w .

Readers interested in the proof of this are referred to Isaacson and Madsen [21] or Kemeny et al. [24].

Example 6.5.3

The matrix $K = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$ is not regular, since the center row remains fixed, regardless of the power of K being considered. But

$$(0 \ 1 \ 0) \cdot \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0 \ 1 \ 0)$$

showing that the vector $(0, 1, 0)$ is, however, a fixed vector for K .

Example 6.5.4

The matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ a & 0 & 1-a \\ 0 & 0 & 1 \end{pmatrix}$ for $0 < a < 1$ is not regular since $A^n = A$, for $n \geq 2$, and each row of A is a fixed vector. This shows that a nonregular matrix can have more than one fixed vector. If, however, a regular matrix has a unique fixed vector, then each row of the matrix T^n approaches that fixed vector.

To see this, suppose that the vector $w = (w_1, w_2, w_3)$ is a fixed probability vector for some 3 by 3 matrix T . Then $wT = w$ so $wT^2 = (wT)T = wT = w$ and so on. But if $T^n \rightarrow K$, where K is a matrix with identical fixed rows, say

$$K = \begin{pmatrix} a & b & c \\ a & b & c \\ a & b & c \end{pmatrix},$$

then $wK = w$, or

$$(w_1, w_2, w_3) \begin{pmatrix} a & b & c \\ a & b & c \\ a & b & c \end{pmatrix} = (w_1, w_2, w_3).$$

So $w_1a + w_2a + w_3a = w$, and since $w_1 + w_2 + w_3 = 1$, $w_1 = a$. A similar argument shows that $w_2 = b$ and $w_3 = c$, establishing the result. It is easy to reproduce the argument for any size matrix T .

What influence does the initial position in the chain have? We might conjecture that after a large number of transitions, the initial state has little, if any, influence on the long-term result. To see that this is indeed the case, suppose that there is a probability vector, $P^0 = (p_1^0, p_2^0, p_3^0)$ whose components give the probability of the process starting in any of the three states and where $\sum_{i=1}^3 p_i^0 = 1$. (We again argue, without any loss of generality, for a 3 by 3 matrix T .)

$P^0 T$ is a vector, say P^1 , whose components give the probability that the process is in any of the three states after one step.

Then $P^1 T = P^0 T \cdot T = P^0 T^2$ and so on. But if $T^n \rightarrow K$, and if the fixed vector for the matrix K is say (k_1, k_2, k_3) , then

$$P^0 K = (k_1, k_2, k_3)$$

since the components of P^0 add up to 1, showing the probability that the process is in any of the given states after a large number of transitions is independent of P^0 .

Now we discuss a number of Markov chains and some of their properties.

Example 6.5.5

The random walk and ruin problem of Section 6.2 can be considered to be a Markov chain, the states representing the fortunes of the player. For simplicity, suppose that \$1 is exchanged at each play, that the player has probability p of winning each game (and losing each game with probability $1 - p$), and that the player's fortune is $\$n$, while the opponent begins with \$4; the boundary conditions are $P_0 = 0$ and $P_4 = 1$. There are five states representing fortunes of \$0, \$1, \$2, \$3, and \$4. The transition matrix is

$$T = \begin{pmatrix} & 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & q & 0 & p & 0 & 0 \\ 2 & 0 & q & 0 & p & 0 \\ 3 & 0 & 0 & q & 0 & p \\ 4 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

reflecting the facts that the game is over when the states $n = 0$ or $n = 4$ are reached. These states are called *absorbing states* since, once entered, they are impossible to leave. We call a Markov chain *absorbing* if it has at least one absorbing state and if it is possible to move to some absorbing state from any nonabsorbing state in a finite number of moves. The matrix T describes an absorbing Markov chain.

It will be useful to reorder the states in an absorbing chain so that the absorbing states come first and the nonabsorbing states follow. If we do this for the matrix T , we find that

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ q & 0 & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & p & 0 & q & 0 \end{pmatrix}.$$

We can also write the matrix in block form as

$$T = \begin{pmatrix} I & O \\ R & Q \end{pmatrix}$$

where I is an identity matrix, O is a matrix each of whose entries is 0, and Q is a matrix whose entries are the transition probabilities from one nonabsorbing state to another. Any transition matrix with absorbing states can be written in the above-mentioned block form.

In addition, matrix multiplication shows that

$$T^n = \begin{pmatrix} I & O \\ R^n & Q^n \end{pmatrix}.$$

The entries of Q^n then give the probabilities of going from one nonabsorbing state to another in n steps.

It is a fact that if a chain has absorbing states, then eventually one of the absorbing states will be reached. The central reason for this is that any path avoiding the absorbing states has a probability that tends to 0 as the number of steps in the path increases.

The possible paths taken in a Markov chain are of some interest and one might consider, on average, how many times a nonabsorbing state is reached. Consider a particular nonabsorbing state, say state j .

The entries of Q give the probabilities of reaching j from any other nonabsorbing state, say i , in one step. The entries of Q^2 give the probabilities of reaching state j from state i in two steps, and, in general, the entries of Q^n give the probabilities of reaching state j from state i in n steps. Now define a sequence of indicator random variables:

$$X_k = \begin{cases} 1 & \text{if the chain is in state } j \text{ in } k \text{ steps} \\ 0 & \text{otherwise} \end{cases}.$$

Then X , the total number of times the process is in state j , is

$$X = \sum_k X_k,$$

and so the expected value of X , the expected number of times the chain is in state j , is

$$E(X) = I_j + \sum_i 1 \cdot q_{ij}^n,$$

where q_{ij}^n is the (i,j) entry in the matrix Q^n and where $I_j = 1$ or 0, depending on whether or not the chain starts in state j .

This shows that $E(X) = I + Q + Q^2 + Q^3 + \dots$ where I is the n by n identity matrix.

It can be shown in this circumstance that $(I - Q)^{-1} = I + Q + Q^2 + Q^3 + \dots$, and so the entries of $(I - Q)^{-1}$ give the expected number of times the process is in state j , given that it starts in state i .

The matrix $(I - Q)^{-1}$ is called the *fundamental matrix* for the absorbing Markov chain.

Example 6.5.6

Consider the transition matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \end{pmatrix}$$

representing a random walk. The matrix Q here is

$$Q = \begin{pmatrix} 0 & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & 0 \end{pmatrix} \text{ and } I - Q = \begin{pmatrix} 1 & -\frac{3}{4} & 0 \\ -\frac{1}{4} & 1 & -\frac{3}{4} \\ 0 & -\frac{1}{4} & 1 \end{pmatrix},$$

so

$$(I - Q)^{-1} = \begin{pmatrix} \frac{13}{10} & \frac{6}{5} & \frac{9}{10} \\ \frac{2}{5} & \frac{8}{5} & \frac{6}{5} \\ \frac{1}{10} & \frac{2}{5} & \frac{13}{10} \end{pmatrix}$$

If the chain starts in state 1, it spends on average $13/10$ times in state 1, $6/5$ times in state 2, and $13/10$ times in state 3. This means that, starting in state 1, the total number of times in various states is $\frac{13}{10} + \frac{6}{5} + \frac{13}{10} = \frac{19}{5}$. So the average number of turns before absorption must be $19/5$ if the process begins in state 1.

Similar calculations can be made for the other beginning states. If we let V be a column vector each of whose entries is 1, then $(I - Q)^{-1}V$ represents the average number of times the process is in each state before being absorbed. Here,

$$(I - Q)^{-1}V = \begin{pmatrix} \frac{13}{10} & \frac{6}{5} & \frac{13}{10} \\ \frac{2}{5} & \frac{8}{5} & \frac{6}{5} \\ \frac{1}{10} & \frac{2}{5} & \frac{13}{10} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{19}{5} \\ \frac{16}{5} \\ \frac{9}{5} \end{pmatrix}.$$

We continue with further examples of Markov chains.

Example 6.5.7

In Example 6.5.5, we considered a game in which \$1 was exchanged at each play and where the game ended if either player was ruined. Now consider players who prefer that the game never end. They agree that, if either player is ruined, the other player gives the ruined player \$1 so that the game can continue. This creates a random walk with *reflecting barriers*.

Here is an example of such a random walk. p is the probability a player wins at any play, $q = 1 - p$ and there are four possible states. The transition matrix is

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

It is probably not surprising to learn that M is not a regular transition matrix. Powers of M do, however, approach a matrix having, in this case, two sets of identical rows. We find, for example, that if $p = 2/3$, then

$$M^n \rightarrow \begin{pmatrix} 0 & \frac{3}{7} & 0 & \frac{4}{7} \\ \frac{1}{7} & 0 & \frac{6}{7} & 0 \\ 0 & \frac{3}{7} & 0 & \frac{4}{7} \\ \frac{1}{7} & 0 & \frac{6}{7} & 0 \end{pmatrix}.$$

Example 6.5.8

A private grade school offers instruction in grades K , 1, 2, and 3. At the end of each academic year, a student can be promoted (with probability p), asked to repeat the grade (with probability r), or asked to return to the previous grade (with probability $1 - p - r = q$). The transition matrix is

$$G = \begin{matrix} & K & 1 & 2 & 3 \\ \begin{matrix} K \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 1-p & p & 0 & 0 \\ q & r & p & 0 \\ 0 & q & r & p \\ 0 & 0 & q & 1-q \end{pmatrix} \end{matrix}.$$

For the particular matrix,

$$G = \begin{matrix} & K & 1 & 2 & 3 \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.1 & 0.2 & 0.7 & 0 \\ 0 & 0.1 & 0.2 & 0.7 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix} \end{matrix}$$

we find the fixed vector to be $(0.0025, 0.0175, 0.1225, 0.8575)$.

EXERCISES 6.5

1. Show that a n by n doubly stochastic transition matrix has a fixed vector each of whose entries is $1/n$.
2. A family on vacation goes either camping or to a theme park. If the family camped 1 year, it goes camping again the next year with probability 0.7; if it went to a theme park 1 year it goes camping the next year with probability 0.4. Show that the process is a Markov chain. In the long run how often does the family go camping?
3. Voters often change their party affiliation in subsequent elections. In a certain district, Republicans remain Republicans for the next election with probability 0.8. Democrats stay with their party with probability 0.9. Show that the process is a Markov chain and find the fixed vector.
4. A small manufacturing company has two boxes of parts. Box I has five good parts in it while box II has 6 good parts in it. There is one defective part, which initially is in the first box. A part is drawn out from box I and put into box II; on the second draw, a part is drawn from box II and put into box I. After five draws, what is the probability that the defective part is in the first box?
5. Electrical usage during a summer month can be classified as “normal,” “high,” or “low.” Weather conditions often make this level of usage change according to the following matrix:

$$\begin{array}{ccc} & N & H & L \\ \begin{matrix} N \\ H \\ L \end{matrix} & \left(\begin{array}{ccc} \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{2}{5} & \frac{1}{3} & \frac{4}{15} \\ \frac{1}{2} & \frac{2}{5} & \frac{1}{10} \end{array} \right) \end{array}$$

Find the fixed vector for this Markov chain and interpret its meaning.

6. A local stock either gains value (+), remains the same (0), or loses value (-) during a trading day according to the following matrix:

$$\begin{array}{ccc} & + & 0 & - \\ \begin{matrix} + \\ 0 \\ - \end{matrix} & \left(\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{array} \right) \end{array}$$

If you were to bet on the stock’s performance tomorrow, how would you bet?

7. Show that the fixed vector for the transition matrix

$$\begin{pmatrix} p & 1-p \\ r & 1-r \end{pmatrix}$$

where $0 < p < 1$, $0 < r < 1$, and $1 - p + r \neq 0$ is

$$\left(\frac{r}{1-p+r}, \frac{1-p}{1-p+r} \right).$$

8. Alter the gambler's ruin situation in Example 7.4.5 as follows. Suppose that if a gambler is ruined, the opposing player returns \$2 to him so that the game can go on (in fact, it can now go on forever!). Show the transition matrix if the probability of a gambler winning a game is $2/3$. If the matrix has a fixed point, find it.
9. The states in a Markov chain that is called *cyclical* are 0, 1, 2, 3, 4, and 5. If the chain is in state 0, the next state can be 5 or 2, and if the chain is in state 5 it can go to state 0 or 4. Show the transition matrix for this chain with probability $1/2$ of moving from one state to another possible state. If the matrix approaches a steady state, find it.

CHAPTER REVIEW

This chapter considers two primary topics, recursions and Markov chains.

Recursions are used when it is possible to express one probability, as a function of some variable, say n , in terms of other probabilities as functions of that same variable, n . In Example 6.1.2, we tossed a loaded coin until it came up heads twice in a row. If a_n represents the probability that HH occurs for the first time at the n th toss, then $a_n = qa_{n-1} + pqa_{n-2}$, $n \geq 3$, with $a_1 = 0$ and $a_2 = p^2$. Values of a_n can easily be found using a computer algebra systems. Frequently, such systems will also solve recursions, producing formulas for the variable as a function of n . We showed an algebraic technique for solving recursions involving a *characteristic equation*, and *homogeneous* and *particular* solutions.

Generating functions associated with a recursion, such as $G(s) = \sum_{n=0}^{\infty} a_n \cdot s^n$, were also considered. These are often of use when recursions are not easily found directly. We illustrated how to find a recursion for the event “*THT* occurs at the n th trial” and a generating function, $U(s)$, for the probability that *THT* occurs at the n th trial. The generating function for first-time occurrences, $F(s)$, is simply related to that of $U(s)$:

$$F(s) = 1 - \frac{1}{U(s)}.$$

We then showed how to find a recursion for first-time occurrences, given $F(s)$.

When the events in question form a probability distribution, means and variances can be determined from recursions. For example, if the recursion is

$$a_n = f \cdot a_{n-1} + g \cdot a_{n-2}, \quad n \geq 2$$

with initial values a_0 and a_1 , and where f and g are constants, then

$$\sum_{n=2}^{\infty} na_n = f \cdot \sum_{n=2}^{\infty} [(n-1)+1]a_{n-1} + g \cdot \sum_{n=2}^{\infty} [(n-2)+2]a_{n-2}$$

from which it follows that

$$E[N] = \frac{a_1 + f(1 - a_0) + 2g}{1 - f - g}.$$

Variances can also be determined from the recursion.

Markov chains arise when a process, consisting of a series of trials, can be regarded at any time as being in a particular *state*. Usually, the matrix $T = [p_{ij}]$ where the p_{ij} represent the probability that the process goes from state i to state j is called a *transition matrix*. The transition matrix clearly gives all the information needed about the process.

Entries of 1 in T indicate *absorbing states*, that is, states which once entered cannot be left. It is possible to partition the transition matrix for an absorbing chain as

$$T = \begin{pmatrix} I & O \\ R & Q \end{pmatrix}.$$

The matrix $I - Q$ is called the *fundamental matrix* for the Markov chain. The entries in $(I - Q)^{-1}$ give the average number of times the process is in state j , given that it started in state i .

PROBLEMS FOR REVIEW

Exercises 6.2 # 2, 3, 4, 6, 9

Exercises 6.3 #1, 2

Exercises 6.4 #1, 3

Exercises 6.5 #2, 4, 6

SUPPLEMENTARY EXERCISES FOR CHAPTER 6

1. Consider a sequence of Bernoulli trials with a probability p of success.
 - (a) Find a recursion giving the probability u_n that the number of successes in n trials is divisible by 3.
 - (b) Find a recursion giving the probability that when the number of successes in n trials is divided by 3, the remainder is 1. [Hint: Write a system of three recursions involving u_n , the probability that the number of successes is divisible by 3; v_n , the probability that the number of successes leaves a remainder of 1 when divided by 3; and w_n , the probability that the number of successes leaves a remainder of 2 when divided by 3.]
2. Find a recursion for the probability q_n that there is no run of three successes in n Bernoulli trials where the probability of success at any trial is 1/2.
3. Find the probability of an even number of successes in n Bernoulli trials where p is the probability of success at a single trial.
4. Find the probability that no two successive heads occur when a coin, loaded to come up heads with probability p , is tossed 12 times.
5. A loaded coin, whose probability of coming up heads at a single toss is p , is tossed and a running count of the heads and tails is kept. Show that if $u_n = P(\text{heads and tails count is equal at toss } 2n)$, then $u_n = \binom{2n}{n} p^n q^n$. Then find the probability that the heads and tails count is equal for the first time at trial $2n$. (The binomial expansion of $(1 - 4pq)^{-\frac{1}{2}}$ will help.)

6. Automobile buyers of brands A , B , and C stay or change brands according to the following matrix:

$$\begin{array}{ccc} & A & B & C \\ \begin{matrix} A \\ B \\ C \end{matrix} & \left(\begin{array}{ccc} \frac{4}{5} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \end{array} \right) \end{array}$$

After several years, what share of the market does brand C have?

7. A baseball pitcher throws curves (C), sliders (S), and fast balls (F). He changes pitches with the following probabilities:

$$\begin{array}{ccc} & C & S & F \\ \begin{matrix} C \\ S \\ F \end{matrix} & \left(\begin{array}{ccc} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{3}{10} & \frac{3}{5} \end{array} \right) \end{array}$$

The next batter hits fast balls well. Should he be replaced in the line up?

8. A small town has two supermarkets, K and C. A shopper who last shopped at K is as likely as not to return there on the next shopping trip. However, if a shopper shopped at C, the probability is $2/3$ that K will be chosen for the next shopping trip. What proportion of the time does the shopper shop at K?
9. Two players, A and B, play chess according to the following rule: the winner of a game plays the white pieces on the next game. If the probability of winning with the white pieces is p for either player and if A plays the white pieces on the first game, what is the probability that A wins the n th game?

Chapter 7

Some Challenging Problems

Five problems, or groups of problems, are introduced here. The intention is for the reader to investigate, verify, and add to or extend these problems. In many cases, results are stated without proof, and in these cases proofs may be difficult. Mathematica has been used widely and this tool, or a computer equivalent, will be very useful in achieving these goals.

7.1 My Socks and $\sqrt{\pi}$

I have 7 pairs of socks in a drawer. I do not sort them by pairs, so they are randomly distributed in the drawer. I select socks at random until I have a pair. The probability I get a pair in 2 drawings is $\frac{14}{14} \cdot \frac{1}{13} = \frac{1}{13}$, since the first sock can be any one of 14 socks and the second sock must be the other sock in the pair represented by the first sock.

The probability it takes 3 draws to get a pair is $\frac{14}{14} \cdot \frac{12}{13} \cdot \frac{2}{12} = \frac{1}{13}$, since the first sock can be any one of the 14 socks in the drawer, the second sock must not match the first, and the third sock can match either of the first two socks drawn.

In a similar way, we can find the probability distribution of the random variable X , the number of draws it takes to get a pair. The probability distribution is shown in the following table:

X	2	3	4	5	6	7	8
Probability	$\frac{1}{13}$	$\frac{2}{13}$	$\frac{30}{143}$	$\frac{32}{143}$	$\frac{80}{429}$	$\frac{16}{143}$	$\frac{16}{429}$

The sum of the probabilities is $\frac{1}{13} + \frac{2}{13} + \frac{30}{143} + \frac{32}{143} + \frac{80}{429} + \frac{16}{143} + \frac{16}{429} = 1$ as it should be.

We can also compute

$$\begin{aligned} E[X] &= 2 \cdot \frac{1}{13} + 3 \cdot \frac{2}{13} + 4 \cdot \frac{30}{143} + 5 \cdot \frac{32}{143} + 6 \cdot \frac{80}{429} + 7 \cdot \frac{16}{143} + 8 \cdot \frac{16}{429} = \frac{2048}{429} \\ &= 4.7739 \end{aligned}$$

and

$$\begin{aligned} E[X^2] &= 2^2 \cdot \frac{1}{13} + 3^2 \cdot \frac{2}{13} + 4^2 \cdot \frac{30}{143} + 5^2 \cdot \frac{32}{143} + 6^2 \cdot \frac{80}{429} + 7^2 \cdot \frac{16}{143} + 8^2 \cdot \frac{16}{429} \\ &= \frac{10822}{429} \end{aligned}$$

so that

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{10822}{429} - \left(\frac{2048}{429}\right)^2 = \frac{448334}{184041} = 2.4361$$

We also note here that $E[X] = \frac{2^{14}}{\binom{14}{7}} = \frac{2048}{429}$ and that $E[X] = \frac{\sqrt{\pi}\Gamma[8]}{\Gamma[7+1/2]}$, where Γ refers to the Euler gamma function. In general, $\Gamma[n+1/2] = \frac{(2n)!\sqrt{\pi}}{n!2^{2n}}$ and $\Gamma[n] = (n-1)!$.

Now we generalize the problem to n pairs of socks in the drawer. It is fairly easy to see that

$$P(X = x) = \frac{2n(2n-2)(2n-4)\cdots(2n-2)[2n-2(x-2)](x-1)}{2n(2n-1)(2n-2)\cdots2n-(x-1)}$$

and with some simplification this becomes

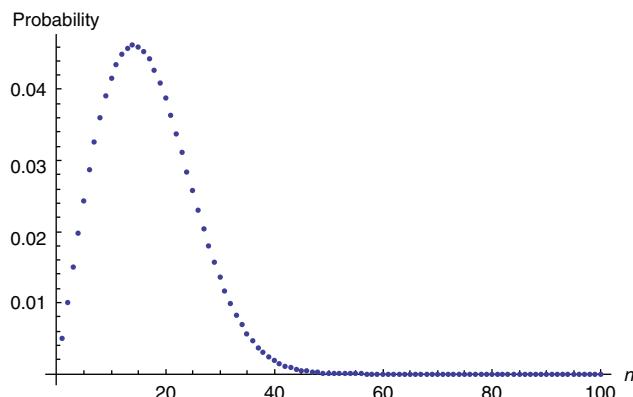
$$P(X = x) = \frac{2^{x-1} \cdot (x-1) \binom{2n-x}{n} (n-x)!}{\binom{2n}{n} \cdot (n-x+1)!} \text{ for } x = 2, 3, \dots, n+1.$$

The factorials are purposely not simplified so that the formula will avoid a division by zero for $x = n+1$. If the factorials are simplified, then it is easy to see that

$$P(X = n+1) = \frac{2^n}{\binom{2n}{n}}$$

This is a probability distribution since $\sum_{x=2}^{n+1} \frac{2^{x-1} \cdot (x-1) \binom{2n-x}{n} (n-x)!}{\binom{2n}{n} \cdot (n-x+1)!} = 1$.

If a computer algebra system, such as Mathematica, is available, then computation for any value of n is easy and graphs can be drawn. Here, for example, is the graph of $P(X = x)$ for $n = 100$.



It appears that the maximum probability is near 15 or so. Here are values of x and $P(X = x)$ for values near 15:

$$\begin{aligned} &\{12., 0.043533\}, \{13., 0.0449645\}, \{14., 0.0458461\}, \\ &\{15., 0.0461874\}, \{16., 0.0460091\}, \{17., 0.0453423\}, \end{aligned}$$

so the maximum is indeed at 15.

To determine the maximum in general, it is easiest to use a recursion. After some simplification, we find that

$$P(X = x + 1)/P(X = x) = \frac{2x(n + 1 - x)}{(2n - x)(x - 1)}.$$

Solving

$$\frac{2x(n + 1 - x)}{(2n - x)(x - 1)} = 1$$

we find that the maximum occurs near $\frac{1}{2}(1 + \sqrt{1 + 8n})$, which gives 14.651 when $n = 100$.

7.2 EXPECTED VALUE

The expected value $\sum_{x=2}^{n+1} x \cdot P(X = x)$ does not simplify easily. Mathematica simplifies this as $1 + \text{Hypergeometric2F1}[1, 1 - n, 1 - 2n, 2]$. Hypergeometric functions are related to probabilities encountered with the hypergeometric probability distribution and are generally quite difficult to deal with.

Fortunately, it is possible to expand the hypergeometric function above to the series

$$\begin{aligned} &1 + \frac{(2 - 2n)}{1 - 2n} + \frac{4(1 - n)(2 - n)}{(1 - 2n)(2 - 2n)} + \frac{8(1 - n)(2 - n)(3 - n)}{(1 - 2n)(2 - 2n)(3 - 2n)} \\ &+ \frac{16(1 - n)(2 - n)(3 - n)(4 - n)}{(1 - 2n)(2 - 2n)(3 - 2n)(4 - 2n)} + \frac{32(1 - n)(2 - n)(3 - n)(4 - n)(5 - n)}{(1 - 2n)(2 - 2n)(3 - 2n)(4 - 2n)(5 - 2n)} + \dots \end{aligned}$$

Interestingly, this can be expressed in terms of Pochhammer functions where

$$\text{Pochhammer}[a, n] = a(a + 1)(a + 2) \cdots (a + n - 1).$$

We then see that $1 + \text{Hypergeometric2F1}[1, 1 - n, 1 - 2n, 2] = 1 + \sum_{k=1}^{\infty} \frac{2^k \text{Pochhammer}[1-n, k]}{\text{Pochhammer}[1-2n, k]}$, which may make the expression appear easier, but in fact is just as complicated as the original. One should be cautioned though that in computation, enough terms be taken so that an infinite number of terms are 0. For example, if $n = 7$, then 7 terms must be used. The result is then always a rational number.

Here are some values of n followed by the expected values:

$$\begin{aligned} &\{2, 8/3\}, \{3, 16/5\}, \{4, 128/35\}, \{5, 256/63\}, \{6, 1024/231\}, \\ &\{7, 2048/429\}, \{8, 32768/6435\}, \{9, 65536/12155\}, \{10, 262144/46189\}, \\ &\{11, 524288/88179\}, \{12, 4194304/676039\}. \end{aligned}$$

One notices that the numerators of the expected values are each powers of 2, while the denominators are all odd. This suggests that some factors of 2 have been divided into the numerators, resulting in their simplification. In addition, the denominators almost always can be factored into prime factors that occur only once, suggesting that they arose from a binomial coefficient. Restoring those factors of 2 gives a surprising result. To take as an example, for $n = 10$, the expected value is

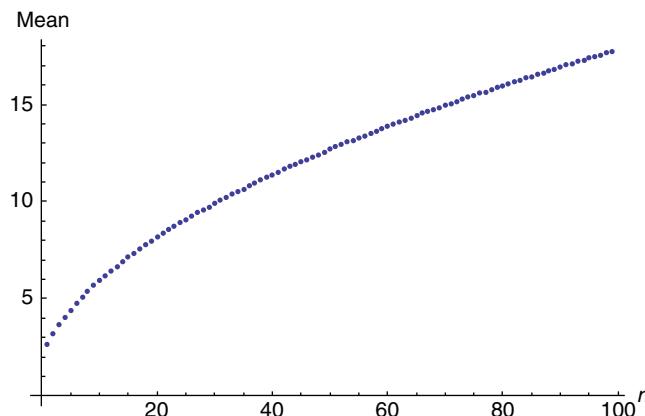
$$\frac{262\,144}{46\,189} = \frac{2^{18}}{46\,189} = \frac{2^{20}}{184\,756} = \frac{2^{20}}{\binom{20}{10}}.$$

Other expected values follow a similar pattern and we see, for n pairs of socks in the drawer, that

$$E[X] = \frac{2^{2n}}{\binom{2n}{n}}.$$

For $n = 100$, the expected value is $\frac{2^{200}}{\binom{200}{100}} = 17.747$.

Here is a graph of the expected values for $n = 2, 3, \dots, 100$:



While these means appear to be curvilinear, a straight line approximation gives a p -value of the order 10^{-74} , so the fit is fantastically good. For example, for $n = 20$ the expected value is

$$\frac{2^{40}}{\binom{40}{20}} = \frac{274\,877\,906\,944}{34\,461\,632\,205} = 7.9763$$

while the straight line fit gives $5.16173 + 0.13763(20) = 7.9143$.

Since $\binom{2n}{n}$ can be expressed as $\frac{2^{2n}\Gamma(n+1/2)}{\sqrt{\pi}\Gamma(n+1)}$, it follows that

$$E[X] = \frac{\sqrt{\pi}\Gamma(n+1)}{\Gamma(n+1/2)}.$$

7.3 VARIANCE

The variance is equally difficult.

Mathematica simplifies $E[X^2]$ as

$$\frac{4}{2n-1} \text{HypergeometricPFQ}[\{3, 3, 1-n\}, \{2, 2-2n\}, 2],$$

which can be expanded in a power series as

$$\begin{aligned} 1 + \frac{9(1-n)}{2-2n} + \frac{48(1-n)(2-n)}{(2-2n)(3-2n)} + \frac{200(1-n)(2-n)(3-n)}{(2-2n)(3-2n)(4-2n)} \\ + \frac{720(1-n)(2-n)(3-n)(4-n)}{(2-2n)(3-2n)(4-2n)(5-2n)} + \frac{2352(1-n)(2-n)(3-n)(4-n)(5-n)}{(2-2n)(3-2n)(4-2n)(5-2n)(6-2n)} + \dots \end{aligned}$$

and so the variance can be written as

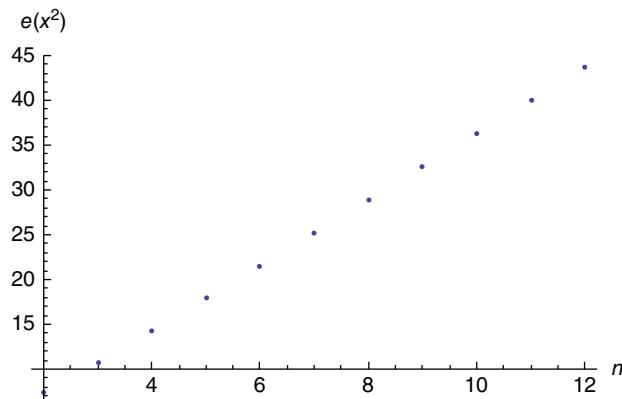
$$1 + \sum_{k=1} \frac{(\text{Pochhammer}[3, k])^2 * \text{Pochhammer}[1-n, k] * 2^k}{k! \text{Pochhammer}[2, k] * \text{Pochhammer}[2-2n, k]} - \left[\frac{2^{2n}}{\binom{2n}{n}} \right]^2.$$

Some of the variances along with values of n are as follows:

$$\{2, 2/9\}, \{3, 14/25\}, \{4, 1186/1225\}, \{5, 5654/3969\}, \dots,$$

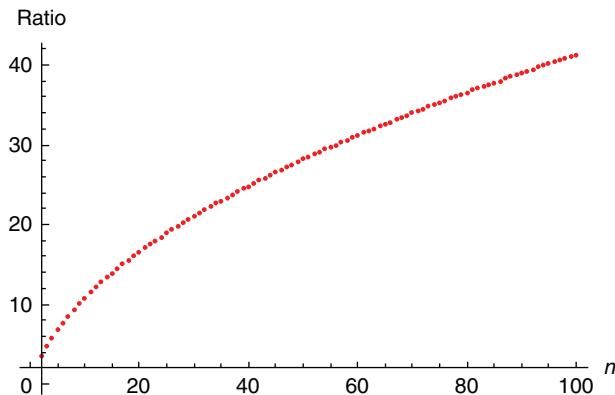
$$\{20, 1235293067078217235394/1187604094232693162025\}, \dots$$

Here is a graph of a few of the values of $E[X]^2$:



It is interesting that values of $E[X^2]$ are related, but not simply, to the values of $E[X]$.

Here is a graph of the ratio of $E[X^2]$ to $E[X]$.



A nonlinear least squares fit gives a p -value of the order 10^{-117} so the fit is excellent.

7.4 OTHER “SOCKS” PROBLEMS

One mathematical problem always leads to others. Here are some related problems for the reader:

1. There are some red socks and some blue socks in a drawer. How many socks of each color must be in the drawer to make the probability of drawing a matching pair on the first two drawings $1/2$? There are some interesting relationships between the values of the red socks as the size of the drawer increases.
2. If the drawer contains n pairs of socks and suppose that k socks have been drawn. Show that the expected number of pairs drawn is $\binom{k}{2} / (2n - 1)$.
3. Suppose 7 pairs of socks are in the drawer, but 2 of the pairs are identical yellow socks while the other 5 pairs are of different colors. It might be thought that this would reduce the expected value of X by 1, but this is not so. Show that $E[X] = 12482/3003 = 4.15651$.

7.5 COUPON COLLECTION AND RELATED PROBLEMS

A fast food restaurant offers three prizes, one with each purchase. On average, how many visits must one make to collect all three prizes?

How many tosses on average must a fair coin be thrown in order that both heads and tails appear for the first time?

What is the expected number of throws with a fair die so that each of the faces appears at least once?

How many random integers must be produced on average so that for the first time, each of the integers $0, 1, 2, \dots, 9$ has appeared?

These are all variants of what is known as the Coupon Collector’s Problem. [3], [11], [26].

We explore this problem here using Mathematica, which sheds considerable light on the problem, especially when the number of events to be seen is large.

Three Prizes

Let us begin with a modest number of prizes to be received at a fast food restaurant, say 3. There are two approaches to the problem; we will show both of them in this case.

Permutations

First consider writing down some of the permutations indicating the order in which the prizes are collected. Let the random variable N denote the number of visits necessary to collect all the prizes and R the number of prizes to be collected.

$$r = 3$$

If $r = 3$, then the three prizes are collected in three visits to the restaurant. There are six orders in which the prizes may be collected: ABC, ACB, BAC, BCA, CAB, and CBA.

If the prizes are equally likely to appear, then the probability that $N = 3$

$$P(N = 3) = 6/3^3 = 2/9$$

Now suppose that $N = 4$. Mathematica can be used to create all the permutations. One of the prizes must occur twice, the other once, and the last prize to be collected must occur only once.

To produce the permutations with $N = 4$, start with one of the 6 permutations of A, B, and C, say BAC. Now we preserve the order BAC and add one symbol – B or A – (since C must occupy the last place). There are two choices for the place to add the extra symbol – following the B or following the A. B can be inserted in two places producing BBAC and BABC. Inserting A in these places produces only one order, namely, BAAC.

So each of the 6 orders of 3 symbols produces 3 orders of 4 symbols, or a total of 18 orders.

$$\text{So } P(N = 4) = 18/3^4 = 2/9.$$

The reader may be interested to show that there are 42 distinct permutations for $N = 5$ and 90 distinct permutations when $N = 6$, so

$$P(N = 5) = 42/243 = 14/81 \text{ and } P(N = 6) = 90/729 = 10/81.$$

Continuing to count permutations becomes increasingly difficult, since duplications must be avoided. It becomes very difficult to change the probabilities with which the prizes occur (typically the prizes do not occur with equal probabilities, one prize often being rarer than the others). It turns out that there is an alternative method that does not have the disadvantages of counting the permutations and allows us to alter the probabilities with which the prizes occur as well.

An Alternative Approach

We use the General Addition Law. Suppose that $N = 3$ and that each of the prizes occurs with probability $1/3$. It is easiest to calculate the probability that not all the prizes occur in N trials and subtract this from 1. The probability that in n trials, at most two of the three prizes occur is $3 * ((2/3)^n)$ but we must subtract from this the probability that only one of the three prizes occur, $3 * (1/3)^n$, so the probability that all three prizes occur in n trials is $1 - 3 * (2/3)^n + 3 * (1/3)^n$.

If we make the function $f(n) = 1 - 3 * (2/3)^n + 3 * (1/3)^n$, then we may calculate $f(n)$ for various values of n to find these probabilities associated with values of $n : \{(3, 2/9), (4, 4/9), (5, 50/81), (6, 20/27), (7, 602/729), (8, 644/729), (9, 6050/6561), (10, 6220/6561)\}$.

But this table gives the probabilities that all three prizes occur in n trials or fewer. To find the individual probabilities, we must subtract one entry from the previous entry to find this table after prepending the entry for three trials:

$$\{(3, 2/9), (4, 2/9), (5, 14/81), (6, 10/81), (7, 62/729), (8, 14/243), (9, 254/6561), (10, 170/6561)\}.$$

The sum of these probabilities is 0.948026 so about 95% of the time the three prizes will occur within 10 trials.

Altering the Probabilities

It is very easy to alter the probabilities with which the prizes occur with this approach. Suppose $P(A) = 0.5$, $P(B) = 0.2$, and $P(C) = 0.3$.

Then, in general, let

$$\text{genprob}(n, pa, pb, pc) = (pa + pb)^n + (pa + pc)^n + (pb + pc)^n - pa^n - pb^n - pc^n.$$

If the three prizes are equally likely, a table of these values is

$$\{(3, 2/9), (4, 4/9), (5, 50/81), (6, 20/27), (7, 602/729), (8, 644/729), (9, 6050/6561), (10, 6220/6561)\},$$

which is the result we saw earlier. But now, we can alter the probabilities letting $pa = 0.5$, $pb = 0.2$, and $pc = 0.3$ to find

$$\{(3, 0.18), (4, 0.36), (5, 0.507), (6, 0.621), (7, 0.708162), (8, 0.774648), (9, 0.825449), (10, 0.864384)\}.$$

To find the probabilities that all the prizes are won in exactly $n + 1$ trials, we subtract the probability the event occurs in n trials from the probability the event occurs in $n + 1$ trials.

A General Result

It is evident, using the General Addition Law, that the probability all the r prizes are all collected in n trials is

$$\begin{aligned} p(n, r) &= 1 - \binom{r}{r-1} \left(\frac{r-1}{r}\right)^n + \binom{r}{r-2} \left(\frac{r-2}{r}\right)^n - \dots \pm \binom{r}{r} \left(\frac{1}{r}\right)^n \\ &= \sum_{i=1}^{r-1} (-1)^i \binom{r}{i} \left(\frac{r-i}{r}\right)^n. \end{aligned}$$

Our reasoning is similar to the reasoning we used in the case $r = 3$. $\binom{r}{r-1} \left(\frac{r-1}{r}\right)^n$ is the probability that at most $r - 1$ prizes appear in n trials; $\binom{r}{r-2} \left(\frac{r-2}{r}\right)^n$ is the probability that at most $r - 2$ prizes appear in n trials, and so on. These probabilities must be added or subtracted in turn so that the result is exactly r prizes in n trials.

Then, it is fairly easy to show that $p(n, r) - p(n - 1, r) = \sum_{i=1}^{r-1} (-1)^i \binom{r}{r-i} \left(\frac{r-i}{r}\right)^n \left(\frac{-i}{r}\right)$, giving the probabilities for individual values of n . The probability that $n = r$ is $r! / r^r$, so a complete table if we let $r = 3$ is

$$\{(3, 2/9), (4, 2/9), (5, 14/81), (6, 10/81), (7, 62/729), (8, 14/243), (9, 254/6561), \\ (10, 170/6561), (11, 1022/59049)\},$$

which checks our previous result.

Here are some results for small values of r :

$$r = 4: \{(4, 3/32), (5, 9/64), (6, 75/512), (7, 135/1024), (8, 903/8192), (9, 1449/16384),$$

$$(10, 9075/131072), (11, 13995/262144), (12, 85503/2097152)\}$$

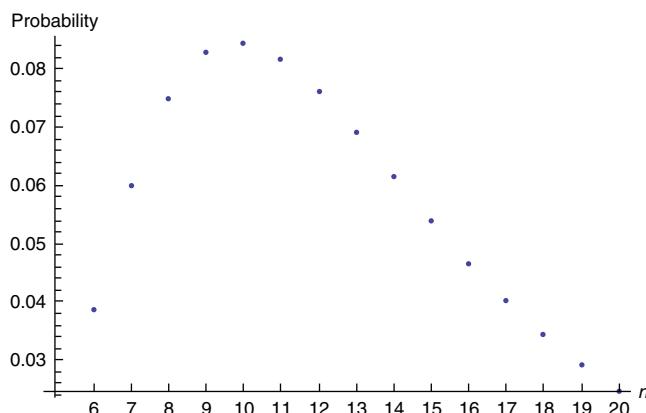
$$r = 5: \{(5, 24/625), (6, 48/625), (7, 312/3125), (8, 336/3125), (9, 40824/390625),$$

$$(10, 37296/390625), (11, 163704/1953125), (12, 27984/390625)\}$$

$$r = 6: \{(6, 5/324), (7, 25/648), (8, 175/2916), (9, 875/11664), (10, 11585/139968),$$

$$(11, 875/10368), (12, 616825/7558272)\}.$$

Here is a graph for $r = 6$:



Mathematica tells us that the probability that 100 equally likely premiums are collected in 200 trials is $4.311 \cdot 10^{-9}$.

Expectations and Variances

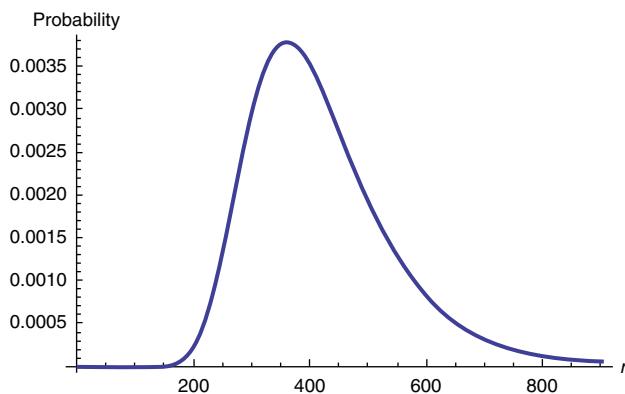
It is easy to calculate means and variances for various values of r . Since the sums used here are infinite, we approximate them with a finite number of terms. Exact results will also be given later.

Here are the means for $r = 3$ and $r = 4$, taken to 40 terms:

$$\sum_{n=3}^{40} n * p(n, 3) = 5.49999 \text{ and } \sum_{n=4}^{40} n * p(n, 4) = 8.33156$$

Mathematica allows us to calculate large values of n . Here is the approximate expected number of trials to collect 100 premiums (each equally likely), followed by a graph of the probabilities:

$$\sum_{n=100}^{1000} n * p(n, 100) = 518.738.$$



The following table shows that the maximum occurs when $n = 460$.

$\{(457., 0.00378522), (458., 0.00378608), (459., 0.00378652), (460., 0.00378656),$
$(461., 0.0037862), (462., 0.00378544)\}$

Geometric Distribution

The coupon collector's problems are all instances of the geometric probability distribution where $P(X = x) = pq^{x-1}$, $x = 1, 2, \dots$, where p is the probability of the event we await.

It can be shown that $E[X] = 1/p$ and $Var[X] = q/p^2$.

So in the case of r prizes, the first prize is collected on the first visit; the probability the next prize is found is $(r - 1)/r$, so the expected waiting time to collect the next prize is $r/(r - 1)$. The next prize occurs with probability $(r - 2)/r$, and so the expected waiting time to collect all the prizes is

$$\exp(r) = 1 + \sum_{i=1}^{r-1} \frac{r}{r-i}.$$

And here is a table of some results for small values of r . Note that our approximations for $r = 3$ and $r = 4$ are quite good:

$$\{(2., 3.), (3., 5.5), (4., 8.33333), (5., 11.4167), (6., 14.7), \\(7., 18.15), (8., 21.7429), (9., 25.4607), (10., 29.2897)\}.$$

So the expected waiting time for a fair coin to show both faces is 3, and the expected waiting time for a fair die to show all six faces is 14.7.

The expected waiting time for collecting 100 premiums is 518.738.

Variances

Variances are easily calculated by Mathematica using the function $\sum_{n=3}^{60} n^2 p(n, r) - (\sum_{n=3}^{60} np(n, r))^2$.

We find for $r = 3$ (using 60 terms in the sum) that the variance is 6.75 with standard deviation 2.59808, and we find for $r = 4$ that the variance is 14.4441 with standard deviation 3.80053.

The variances using the geometric distribution are calculated using the function

$$\text{variance}(r) = \sum_{i=1}^{r-1} \frac{r * i}{(r - i)^2}.$$

Here is a table of standard deviations:

$$\{(3, 2.59808), (4, 3.80058), (5, 5.01733), (6, 6.2442), \\(7, 7.47851), (8, 8.71849), (9, 9.96295), (10, 11.211)\}.$$

Waiting for Each of the Integers

Now we simulate looking for each of the digits 0, 1, ..., 9 for the first time. To investigate this, we first create random samples of these integers. We created 50 samples, each of size 60 (so that the failure of any digit to occur in 60 trials is very small).

Here is a typical random sample (this is the 23rd sample produced):

$$\{9, 3, 7, 0, 2, 9, 9, 7, 7, 4, 3, 5, 0, 8, 0, 4, 9, 4, 7, 0, 7, 1, 3, 7, 6, 7, 1, 3, 8, 6, \\8, 9, 6, 0, 3, 1, 0, 2, 3, 4, 9, 6, 7, 8, 6, 6, 7, 8, 8, 4, 1, 1, 0, 3, 1, 7, 2, 4, 7, 3.\}$$

The digits in order of appearance are 9, 3, 7, 0, 2, 4, 5, 8, 1, and 6 and these occurred in positions 1, 2, 3, 4, 5, 10, 12, 22, and 25, respectively, so we had to wait until the 25th sample to see all the integers.

Mathematica can find these positions:

$$\{\{4\}, \{13\}, \{15\}, \{20\}, \{34\}, \{37\}, \{53\}\}, \{\{22\}, \{27\}, \{36\}, \{51\}, \\ \{52\}, \{55\}\}, \{\{5\}, \{38\}, \{57\}\}, \{\{2\}, \{11\}, \{23\}, \{28\}, \{35\}, \{39\},$$

$$\{54\}, \{60\}, \{\{10\}, \{16\}, \{18\}, \{40\}, \{50\}, \{58\}\}, \{\{12\}\}, \{\{25\}, \\ \{30\}, \{33\}, \{42\}, \{45\}, \{46\}\}, \{\{3\}, \{8\}, \{9\}, \{19\}, \{21\}, \{24\}, \\ \{26\}, \{43\}, \{47\}, \{56\}, \{59\}\}, \{\{14\}, \{29\}, \{31\}, \{44\}, \{48\}, \{49\}\}, \\ \{\{1\}, \{6\}, \{7\}, \{17\}, \{32\}, \{41\}\}$$

where the table is read this way: we find that 0 occurred in positions 4, 13, 15, 20, 34, 37, and 53 while the integer 1 occurred in positions 22, 27, 36, 51, 52, and 55. But we are interested only in the smallest of the first position for any of the digits, and in this case this is 25, where 6 was the last digit to occur.

So for this sample, the waiting time was 25 observations before we saw all 10 digits.

Next, we found the positions of each integer in every one of the samples (but we suppress the output).

Since we have all the positions of each of the integers in each of the samples, we need the maximum of the positions of the first entries:

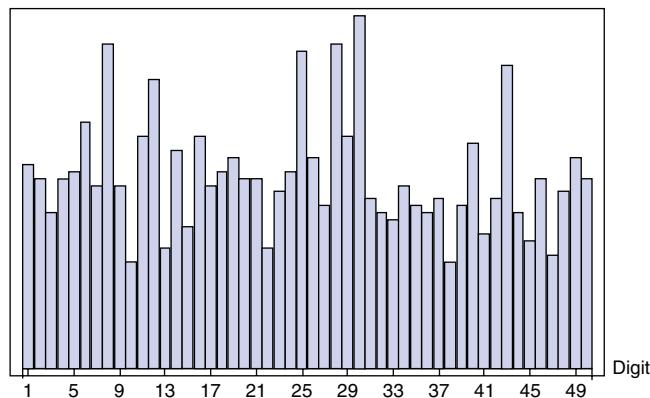
$$\{29, 27, 22, 27, 28, 35, 26, 46, 26, 15, 33, 41, 17, 31, 20, 33, 26, 28,$$

$$30, 27, 27, 17, 25, 28, 45, 30, 23, 46, 33, 50, 24, 22, 21, 26,$$

$$23, 22, 24, 15, 23, 32, 19, 24, 43, 22, 18, 27, 16, 25, 30, 27\}$$

The mean value of these positions is 27.48, which is very close to our theoretical value of 29.2897. The standard deviation of these values is 8.19218.

Here is a bar chart of the maximums of the first entries.



The simulation here sheds great light on a difficult problem.

Conditional Expectations

Suppose we have sampled n of the premiums in the coupon collector's problem. On average, how many of the distinct premiums do we have? Let us suppose there are three

premiums – A, B, and C and that they occur with equal frequency. Let us also assume that if all three premiums are collected, then this does not occur before the n th trial.

We take a specific example to see how the calculations are done. Let $n = 6$. Then the number of distinct items collected can be 1, 2, or 3. The number of orders in which the premiums are collected depend on the number of ways in which the integer 6 can be partitioned and then the number of permutations of A, B, and C determined by these partitions.

We find the partitions of 6 into at most three parts:

$$\{(6), (5, 1), (4, 2), (4, 1, 1), (3, 3), \{3, 2, 1\}, (2, 2, 2)\}$$

(6) is interpreted as all 6 premiums are the same. There are three ways in which this can occur: AAAAAB, BBBBAA, or CCCCCC.

(5, 1) means that two of the premiums occur, one five times and the other once. There are $\binom{3}{2}$ choices for the two premiums, $\binom{2}{1}$ choices for one premium to occur five times, and $\binom{6}{5}$ ways in which the two premiums can be permuted.

This gives us $\binom{3}{2} * \binom{2}{1} * \binom{6}{5} = 36$ ways in which this can occur.

Similarly, the partition (4, 2) produces

$$\binom{3}{2} * \binom{2}{1} * \binom{6}{4} = 90.$$

Finally, the partition {3, 3} produces

$$\binom{3}{2} * \binom{6}{3} = 60.$$

The partitions of six into three parts must be handled a bit differently since all three premiums are collected, but the last premium must complete the set.

The partition (4, 1, 1) means that the last premium must be one of 3 while the first 5 premiums can be collected in $\binom{3}{1} * \binom{2}{1} * \binom{5}{4} = 30$ ways in which this can occur. Similarly, the partition {3, 2, 1} gives $\binom{3}{1} * \binom{2}{1} * \binom{5}{3} = 60$ ways.

The partition (2, 2, 2) is impossible since the three premiums would be collected before the 6th trial.

So we have found that 1 premium can be collected in 3 ways, 2 premiums can be collected in $36 + 90 + 60 = 186$ ways, and three premiums can be collected in $30 + 60 = 90$ ways giving the expected number of premiums after six trials as $\frac{1*3+2*186+3*90}{3+186+90} = 2.31183$.

Other Expected Values

In an entirely similar way, expected values can be found for other values of n . Here are the results which the reader may care to verify:

<i>n</i>	<i>Expected Value</i>
3	2.1111
4	2.2381
5	2.2889
6	2.3118
7	2.4144

One notices that the expectation increases as n increases, but the increase in the expectation is modest when n is increased by 1.

The procedure here becomes increasingly difficult as n increases since the number of partitions of the integers increases rapidly.

Waiting for All the Sums on Two Dice

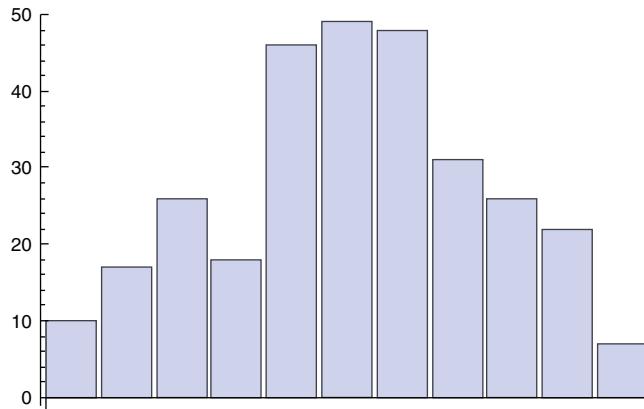
The procedure here is very similar to that of waiting for all the integers to appear, but the sums on two dice do not occur with equal frequency, so we must alter the sampling scheme somewhat. The probabilities of sums 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 are $1/36$, $2/36$, $3/36$, $4/36$, $5/36$, $6/36$, $5/36$, $4/36$, $3/36$, $2/36$, and $1/36$, so we take samples from the following set:

$$(2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, \\ 10, 10, 10, 11, 11, 12).$$

So we have taken, but will not show, 100 samples, each of size 300. This will minimize the chance that some of the sums will not appear. Here is a typical sample:

$$(6, 6, 8, 7, 5, 9, 4, 6, 10, 11, 11, 7, 8, 11, 6, 11, 7, 5, 2, 7, 10, 3, 9, 4, 5, 4, 9, 7, 10, 6, \\ 8, 9, 8, 8, 12, 8, 2, 8, 6, 6, 9, 7, 6, 6, 5, 5, 9, 2, 8, 5, 11, 10, 7, 3, 10, 8, 8, 7, 10, 7, \\ 9, 7, 9, 8, 12, 6, 6, 9, 10, 8, 9, 4, 7, 7, 7, 8, 7, 10, 7, 7, 8, 7, 8, 11, 5, 6, 7, 4, 3, 7, 9, \\ 9, 8, 4, 8, 8, 7, 6, 5, 6, 10, 11, 10, 7, 10, 6, 7, 6, 5, 8, 6, 7, 8, 7, 7, 4, 6, 4, 8, 8, 8, 9, \\ 11, 3, 9, 8, 9, 8, 6, 10, 4, 2, 6, 11, 6, 6, 8, 4, 9, 7, 6, 6, 3, 11, 8, 6, 5, 2, 9, 8, 3, 9, \\ 11, 4, 8, 7, 3, 2, 6, 8, 6, 3, 8, 3, 10, 5, 6, 11, 7, 4, 7, 10, 6, 12, 6, 9, 7, 7, 8, 6, 7, 6, \\ 12, 4, 9, 8, 6, 10, 9, 10, 7, 6, 6, 4, 6, 11, 7, 10, 4, 4, 9, 6, 8, 3, 7, 5, 6, 3, 7, 5, 3, 9, \\ 9, 8, 11, 8, 7, 4, 8, 10, 9, 11, 8, 7, 8, 7, 7, 8, 2, 3, 3, 6, 11, 8, 12, 7, 3, 7, 48, 10, \\ 6, 6, 8, 7, 10, 4, 9, 10, 10, 6, 4, 9, 7, 7, 4, 9, 12, 2, 7, 5, 2, 5, 4, 10, 8, 12, 11, 4, 8, \\ 10, 5, 6, 4, 8, 5, 11, 9, 11, 11, 7, 4, 10, 7, 2, 5, 11, 11, 7, 9, 6, 8, 8, 3, 6, 3, 6, 4, 9, 10)$$

Here are the frequencies with which the sums occurred in this sample: (10, 17, 26, 18, 46, 49, 48, 31, 26, 22, 7) producing the following bar chart.



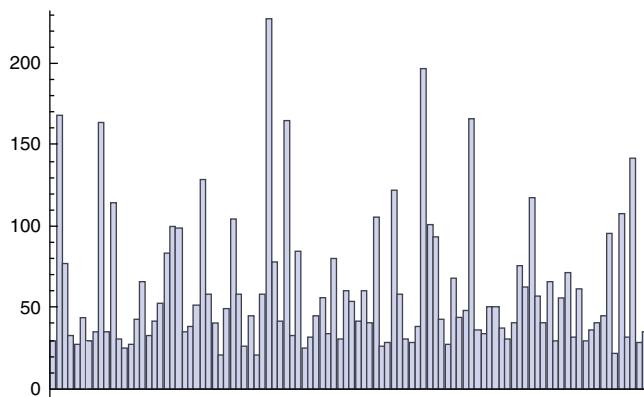
So the sample appears to reflect, very approximately, the frequency of each of the sums. We want to find out the average number of tosses to obtain each sum, so we looked at the positions of each of the sums and then found the maximums of each of the first occurrences of each of the sums, but none of the individual samples will not be shown here.

We need of course only the maximum of the first positions here to see how many dice tosses it took to see each of the sums which in this case is 35. Here are the maximums from all the samples:

(29, 168, 77, 33, 27, 44, 29, 35, 164, 35, 114, 31, 25, 27, 43, 66, 33, 42, 53, 83, 100, 99, 35, 38, 51, 128, 58, 40, 21, 49, 104, 58, 26, 45, 21, 58, 227, 78, 42, 165, 33, 85, 25, 32, 45, 56, 34, 80, 31, 60, 54, 42, 60, 41, 105, 26, 28, 122, 58, 31, 28, 38, 197, 101, 93, 43, 27, 68, 44, 48, 166, 36, 3450, 50, 37, 31, 40, 76, 62, 117, 57, 41, 66, 29, 56, 71, 32, 61, 30, 36, 40, 45, 96, 22, 108, 32, 142, 28, 35)

The mean of this data is 60.62 with a standard deviation of 41.1913.

Here is a bar chart of this data:



So the maximums are quite variable. In doing the above procedure 50 times, we found that the expected number of tosses to see all the sums is about 61.25.

7.6 CONCLUSION

The coupon collector's problem provides many mathematical challenges and interesting ways in which Mathematica can provide insight into the problem and its many facets. The reader is encouraged to consult the following references [3, 11, 26] for more information.

7.7 JACKKNIFED REGRESSION AND THE BOOTSTRAP

We consider here some statistical techniques that are relatively recent with respect to most standard statistical procedures and the techniques that are highly dependent on fast and capable computer programs.

The computations for the techniques discussed here were developed using Mathematica 10, although other computer programs may be capable of carrying out these procedures as well.

Jackknifed Regression

With some frequency, data sets show one or more data points that have significant influence over the usual least squares regression line and create least squares regression lines that are not as representative of the data as they might be.

Jackknifed regression is a technique that goes through the data set and eliminates exactly one data point from the data set at a time and computes the resultant least squares line.

We use an obviously created-for-the-purpose data set from Anscombe [1]:

x	4	5	6	7	8	9	10	11	12	13	14
y	5.39	5.73	6.08	6.42	6.77	7.11	7.46	7.81	8.15	12.74	8.84

The 10th data point (13, 12.74) appears to be an outlier. This is obvious from a graph of the data along with the least squares regression line shown in Figure 7.1.

The least squares regression line is $Y = 3.00245 + 0.499727X$. The analysis of variance is shown in Table 7.1.

The fit is very good, even with the (apparent) outlier included. We will soon show a test verifying that the point (13, 12.74) is indeed an outlier. For now, consider the jackknifed procedure where exactly one point is omitted from the data set at a time and the resulting regression lines are computed. While Mathematica will produce all 11 least squares lines and their analyses of variance we show only the lines and analyses when the first, 10th, and 11th points are omitted.

Omitting the first point, (4, 5.39) the least squares line is $y = 2.71745 + 0.525636x$ and the analysis of variance is shown in Table 7.2.

Omitting the 10th point, (13, 12.74), the strongly suspected outlier, the least squares line, is $y = 4.00565 + 0.34539x$ with analysis of variance shown in Table 7.3.

This is an astoundingly small p -value. Finally, we omit the last point (14, 8.84), giving the least squares line $y = 2.46176 + 0.57697x$ with corresponding analysis of variance shown in Table 7.4.

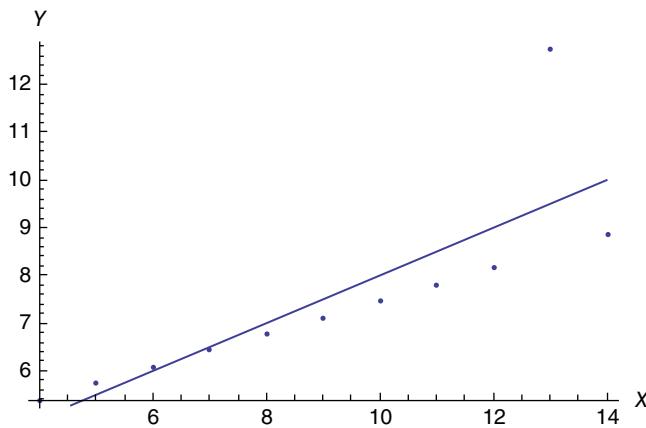


Figure 7.1 Data and least squares regression line.

Table 7.1

	DF	SS	MS	F-Statistic	P-Value
x	1	27.47	27.47	17.9723	0.00217631
Error	9	13.7562	1.52847		
Total	10	41.2262			

Table 7.2

	DF	SS	MS	F-Statistic	P-Value
x	1	22.7942	22.7942	13.4731	0.00630429
Error	8	13.5347	1.69183		
Total	9	36.3289			

Table 7.3

	DF	SS	MS	F-Statistic	P-Value
x	1	11.0228	11.0228	1.16069×10^6	6.17086×10^{-22}
Error	8	0.000075974	9.49675×10^{-6}		
Total	9	11.0228			

Table 7.4

	DF	SS	MS	F-Statistic	P-Value
x	1	27.4638	27.4638	18.6396	0.00255516
Error	8	11.7873	1.47341		
Total	9	39.251			

The 11 different lines and analyses of variance give the following p -values:

<i>PointOmitted</i>	(4,5.39)	(5,5.73)	(6,6.08)	(7,6.42)	(8,6.77)	(9,7.11)
<i>p – value</i>	0.0063	0.0056	0.0050	0.0045	0.0041	0.0038
<i>PointOmitted</i>	(10,7.46)	(11,7.81)	(12,8.15)	(13,12.74)	(14,8.84)	
<i>p – value</i>	0.0036	0.0034	0.0032	$6 \cdot 10^{-22}$	0.0026	

It is also interesting to compare the discrepancies between the observed y values and the predicted values. First, here are the discrepancies using the least squares line for all the points:

$$\{0.388642, 0.228915, 0.079188, -0.080539, -0.230266, -0.389993, \\ -0.53972, -0.689447, -0.849174, 3.2411, -1.15863\}.$$

The mean value here is $7 \cdot 10^{-6}$.

Here are the discrepancies using the least squares line when the point (13, 12.74) is omitted:

$$\{0.00279, -0.0026, 0.00201, -0.00338, 0.00123, -0.00416, 0.00045, 0.00506, \\ -0.00033, -0.00111\}.$$

The mean value here is $4 \cdot 10^{-6}$, less than that for the overall least squares line, due to the presence of the outlier.

7.8 COOK'S DISTANCE

We have been claiming that the point (13, 12.74) is an outlier, as indeed it appears to be from Figure 7.1, but we have not offered any substantial mathematical reason for this. R.D. Cook [5] has proposed a distance, commonly known as *Cook's d*, a quantity computed when each of the data points is omitted from the analysis one at a time. The computations go as follows:

For the i th data point, let e_i denote the residual at that point and let

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \quad \text{and} \quad d_i = \frac{e_i^2}{2 \cdot \text{MSE}} \cdot \left(\frac{h_i}{(1 - h_i)^2} \right)$$

for the i th data point, where MSE is the mean square for error in the least squares regression analysis of variance when the i th point is omitted.

For the 10th data point, (13, 12.74), we find that $h_{10} = \frac{1}{11} + \frac{(13-9)^2}{110} = \frac{13}{55}$ while

$$d_{10} = \frac{(12.74 - 9.4989)^2}{2 \cdot 1.52847} \cdot \frac{\frac{13}{55}}{\left(1 - \frac{13}{55}\right)^2} = 1.39285.$$

Generally, the value of d_i is regarded as significant if it exceeds the 50th percentile of the $F[2, n - 2]$ distribution which in this case is 0.743492, so the data point is significant.

A complete table of Cook's d for this data set is

$$\{0.0338173, 0.00694781, 0.000517641, 0.000354629, 0.00214148, 0.00547314, \\ 0.0117646, 0.0259839, 0.0595359, 1.39285, 0.300571\},$$

so our suspected influential point is, in fact, the only significant point.

The values of h_i are frequently used by themselves to detect influential points. It is easy to verify that $\sum_{i=1}^n h_i = 2$, so the mean value is $2/n$. Any value exceeding this is regarded as influential.

In this case, the critical value is then $2/11 = 0.18182$. A complete set of values of h_i is as follows:

$$\{0.318182, 0.236364, 0.172727, 0.127273, 0.1, 0.0909091, 0.1, 0.127273, 0.172727, \\ 0.236364, 0.318182\}.$$

Here one would conclude that several of the points are influential, a conclusion not supported by the values of Cook's d .

7.9 THE BOOTSTRAP

It is the common case that one has just one random sample to deal with. While the sample may be indicative of the general situation, it is not by itself useful in determining the sampling distribution of a statistic. It takes many random samples of size n , for example, from a normal distribution with variance σ^2 to determine that the sample mean has standard deviation $\frac{\sigma}{\sqrt{n}}$.

Efron [10] has developed a clever way to turn one random sample into many. He calls the procedure the Bootstrap, analogous to raising ones' self by his or her bootstraps, a physically impossible endeavor, but it turns out to be a very real mathematical one. Here is how it works and we take a specific example to illustrate it.

Example 7.9.1

Suppose we wish to discover the standard deviation of the median of samples chosen from a gamma distribution. This is not something that is commonly known! Figure 7.2 shows the population.

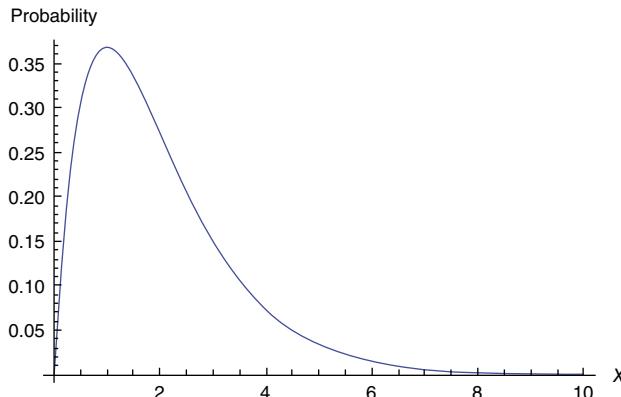


Figure 7.2 A gamma distribution.

The distribution is decidedly nonnormal. To do the bootstrap, we draw a sample of size 40 from the distribution. Here are the values in the sample:

```
{1.51239, 1.22365, 3.47943, 5.86536, 3.3359, 1.72314, 0.130842, 1.62945, 4.51917,
 1.09428, 0.736489, 0.511442, 0.926443, 3.99601, 1.98945, 5.31002, 0.597938,
 3.31717, 1.98591, 1.22595, 1.95226, 1.67722, 1.57653, 1.94384, 1.11182,
 0.273504, 0.27343, 0.306156, 3.70396, 3.32532, 2.51426, 2.87691, 1.42218,
 1.47679, 1.40976, 1.42221, 2.88933, 2.9803, 5.83437, 1.80322}.
```

The bootstrap procedure consists of drawing samples from this single sample, creating a number of samples – hence the bootstrap. We take 1000 samples of size 20 each (if these samples were of size 40, then sampling with replacement must be done), but we do sampling without replacement. We calculate the median of each sample and get a probability distribution for the median. A histogram of the results is shown in Figure 7.3.

We now find that the mean of these medians is 1.78865 with standard deviation 0.359114.

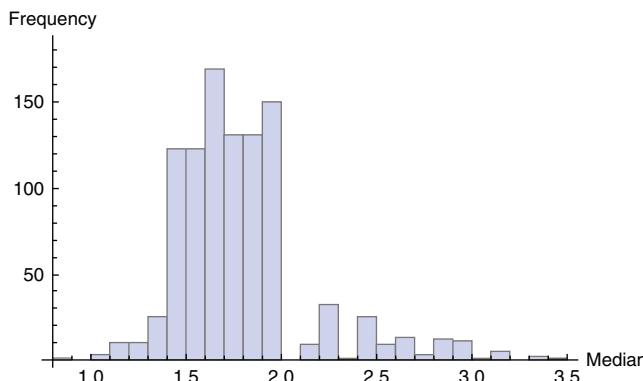


Figure 7.3 Medians of bootstrap samples.

It is very interesting that we can learn a bit about a (to say the least) statistic of rare interest by proceeding from a single sample!

We show one more example. The reader will no doubt find other examples of this technique.

Example 7.9.2

Suppose we wish to discover the expected value of the range in samples from the standard $N(0, 1)$ distribution.

As we proceeded in the median example, we first draw a sample of 100 from the $N(0, 1)$ distribution. We show a portion of this sample:

$$\begin{aligned} & \{0.353945, 0.949031, 0.930671, 0.868072, -1.64129, 1.03884, -0.229624, 0.261774, \\ & -0.203825, -1.61538, 1.04087, -0.476678, -0.763087, 1.00335, 2.51053, \\ & -0.340539, -1.14323, 0.159024, -1.62462, -1.08409, -0.450556, -1.89815, \\ & 0.618595, 1.23218, 0.96988, \dots \}. \end{aligned}$$

Then we selected 1000 samples, each of size 20, from this sample. Here is one of the samples:

$$\begin{aligned} & \{1.33482, -1.06213, -0.308785, 0.551384, -1.61538, 0.261774, -0.965737, 0.969888, \\ & -1.25885, 0.515682, 1.42721, -1.43946, -0.229624, -0.203825, \\ & -1.30396, 0.969888, 0.596313, -1.18455, -0.571001, 0.618595\}. \end{aligned}$$

The range of each sample was calculated. A histogram of these results is shown in Figure 7.4.

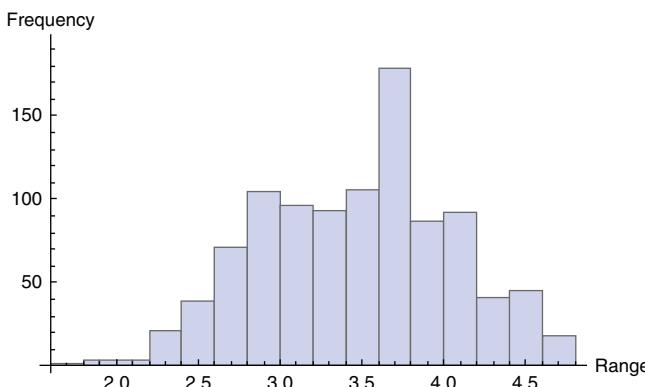


Figure 7.4 Range of bootstrap samples.

The expected value of the range is 3.36013 with standard deviation 0.575379.

7.10 ON WALDEGRAVE'S PROBLEM

Waldegrave, an English gentleman, in 1711 or so proposed the following problem:

$r + 1$ players play each other in a game in which either player playing has probability $1/2$ of winning an individual game. They play in order until one player has beaten each of the other players consecutively. Players who lose an individual game nevertheless retain their position in the playing order.

Three Players

Consider first the situation in which there are three players, say A , B , and C . Denote by X_Y that player X defeats player Y and consecutive symbols denoting subsequent games and their outcomes. One player must then defeat each of the others consecutively.

There are only two ways in which the game can end in two trials: $A_B A_C$ or $B_A B_C$. There are only two ways in which the game can end in three trials: $A_B C_A C_B$ or $B_A C_B C_A$. In fact, there are only two ways in which the game can end in n trials.

Let the random variable N denote the length of the series until a winner is established. Suppose the game ends on the n th game. The previous sequence of games can have no two successive games won by the same player, so the winner must alternate until the winner of game $n - 1$ also wins the n th game. Since the first game must be A_B or B_A , there are only two ways in which the game can end on the n th game.

Then $P(N = n) = 2 \cdot \left(\frac{1}{2}\right)^n = \frac{1}{2^{n-1}}$ for $n = 2, 3, \dots$ and the geometric series

$$\sum_{n=2}^{\infty} \frac{1}{2^{n-1}} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1 \text{ as it should be.}$$

In this case, it is easy to find the expected length of the game, $E(N)$, since

$$E(N) = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + \dots$$

and so

$$\frac{1}{2}E(N) = 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \dots$$

giving

$$E(N) - \frac{1}{2}E(N) = 2 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + \dots = 1 + \frac{\frac{1}{4}}{1 - \frac{1}{2}} = \frac{3}{2} \text{ and so } E(N) = 3.$$

7.11 PROBABILITIES OF WINNING

For three players, it is not difficult to compute the probabilities that each of the players will win the game. (We will show another way to do this later.)

Consider how player A can win the game.

For the sequence $A_B A_C$, the game lasts two plays. If the sequence $A_B C_A B_C A_B A_C$ occurs, A wins in 5 plays. But the initial sequence $A_B C_A B_C$ can occur any number of times before being followed by $A_B A_C$, and so A can win the game this way in 5, 8, 11, 14, ... games.

A final possibility for A to win the game is that the sequence $B_A C_B A_C A_B$ occurs and A wins in four games. And, similar to the previous possibility, the sequence $B_A C_B A_C$ can occur any number of times before A beats B for the final game. So A can also win the game in 4, 7, 10, ... games.

Putting all this together with the geometric series involved, we see that the probability A wins the game is

$$P(A \text{ wins}) = \left(\frac{1}{2}\right)^2 + \frac{\left(\frac{1}{2}\right)^5}{1 - \left(\frac{1}{2}\right)^3} + \frac{\left(\frac{1}{2}\right)^4}{1 - \left(\frac{1}{2}\right)^3} = \frac{5}{14}.$$

Now consider the ways in which B can win the series. B could win in two games by the series $B_A B_C$. Another possibility is that the sequence $B_A C_B A_C B_A B_C$ occurs but the sequence $B_A C_B A_C$ can occur any number of times before $B_A B_C$ occurs, so B can win in 5, 8, 11, ... games.

A final possibility is that the sequence $A_B C_A B_C$ occurs followed by B_A , but again the sequence $A_B C_A B_C$ can occur any number of times before B beats A and so B can win in 4, 7, 10, ... games.

We conclude that the probability B wins the series is exactly the same as the probability A wins the series. But this is probably evident since each player has probability 1/2 of winning the first game.

So the probability C wins the series is $1 - \frac{5}{14} - \frac{5}{14} = \frac{2}{7}$. The sample points for which C wins the game are easily found as well. In this case, the game ends in 3, 6, 9, ... trials. Note that there are two ways in which C can win the game in 3, 6, 9, ... trials.

In the case of three players, the probabilities that each wins the game are not that far apart since $\frac{5}{14} = 0.35714$ and $\frac{2}{7} = 0.28571$. This is a point we will return to, but for the moment consider adding a player to the game.

7.12 MORE THAN THREE PLAYERS

Adding even a single player makes the game considerably more difficult. However, some interesting patterns evolve which we will soon see. Consider the situation for four players.

Let us write out some sample points for various lengths of the game:

Length of the game	Sample points
$n = 3$	$A_B A_C A_D$ or $B_A B_C B_D$
$n = 4$	$A_B C_A C_B C_D$ or $B_A C_B C_D C_A$ $A_B C_A B_C B_D B_A$
$n = 5$	$A_B A_C D_A D_B D_C$ $B_A C_B D_C D_A D_B$ $B_A B_C D_B D_C D_A$

	Length of the game	Sample points
$n = 6$	$A_B C_A D_C A_D A_B A_C$	
	$A_B A_C D_A B_D B_A B_C$	
	$A_B C_A C_D A_C A_B A_B$	
	$B_A C_B D_C A_D A_B A_C$	
	$B_A B_C D_B C_D C_A C_B$	
	$B_A C_B C_D A_C A_D A_B$	

We could go on, but it is obvious that the situation is much more complex than that for three players. Yet there is a pattern here that will make us able to create the sample points for any value of N .

Consider the sample point $A_B A_C A_D$ for $n = 3$. Change the winner of the third game, and let that winner win the series producing the point $A_B A_C D_A D_B D_C$. Change the winner of the third game in the sample point $B_A B_C B_D$, and let that winner go on to win the series to produce the point $B_A B_C D_B D_C D_A$.

Consider the sample points for $n = 4$:

$$A_B C_A C_B C_D$$

$$B_A C_B C_D C_A$$

Change the winner of the third game, and let that player win the series to produce the points $A_B C_A B_C B_D B_A$ and $B_A C_B D_C D_A D_B$, obtaining all the ways the game can end in five trials.

Similarly, the points for $n = 4$ and $n = 5$ can be used to produce all the sample points for $n = 6$.

Finally, we show how the sample points for $n = 5$ and $n = 6$ can be used to produce the 10 sample points for $n = 7$. On the left, we show the sample points for $n = 5$ and $n = 6$. Change the winner of the fifth game, and let that winner go on to win the series to find the sample points in the right-hand column:

$A_B C_A D_C D_A D_B \rightarrow A_B C_A D_C D_A B_D B_A B_C$
$A_B A_C D_A D_B D_C \rightarrow A_B A_C D_A D_B C_D C_A C_B$
$B_A C_B D_C D_A D_B \rightarrow B_A C_B D_C D_A B_D B_A B_C$
$B_A B_C D_B D_C D_A \rightarrow B_A B_C D_B D_C A_D A_B A_C$
$A_B C_A D_C A_D A_B C \rightarrow A_B C_A D_C A_D B_A B_C B_D$
$A_B A_C D_A B_D B_A B_C \rightarrow A_B A_C D_A B_D A_B C_A C_D$
$A_B C_A C_D A_C A_D B \rightarrow A_B C_A C_D A_C D_A B_D B_C$
$B_A C_B D_C A_D A_B C \rightarrow B_A C_B D_C A_D B_A B_C B_D$
$B_A B_C D_B C_D C_A C_B \rightarrow B_A B_C D_B C_D A_C A_B A_D$
$B_A C_B C_D A_C A_D B \rightarrow B_A C_B C_D A_C D_A D_B D_C$

The reason the sample points for a series lasting n trials is dependent on the trials for $n - 1$ and $n - 2$ is fairly simple: if the series ends in some player winning the last three games, then either the winner of the $n - 1$ game also wins the n th game or the winner of the $n - 2$ game also wins the last two games.

So if we look at the ways the game can end, we find the series 2, 2, 4, 6, 10, 16, 26, 42,

Recall that the Fibonacci series is 1, 1, 2, 3, 5, 8, 13, 21, ... where successive terms are found by adding the previous two terms. Our series is exactly similar except that we begin with 2 and 2. We will return later to this point.

r + 1 Players

The argument above can be extended to the case of $r + 1$ players. To find the number of ways the game can end in n trials, consider the number of ways the game can end in $n - i$ trials. Change the winner of the $n - r + 1$ trial, and let that winner go on to win the next $r - i$ trials. So if $a(n)$ denotes the number of ways the game can end in n trials, then

$$a(n) = a(n - 1) + a(n - 2) + \cdots + a(n - r + 1).$$

If $p(n)$ denotes the probability the game ends in n trials, then

$$p(n) = a(n)/2^n \text{ and so } 2^n p(n) = 2^{n-1} p(n-1) + 2^{n-2} p(n-2) + \cdots + 2^{n-r+1} p(n-r+1) \text{ or}$$

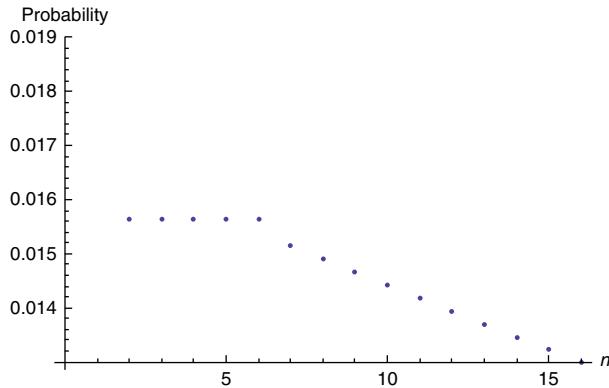
$$p(n) = (1/2)p(n-1) + (1/4)p(n-2) + \cdots + (1/2)^{n-r+1} p(n-r+1).$$

This recursion can be easily used with a system such as Mathematica to find the probabilities the game ends in any number of trials for any number of players.

For seven players, here are the number of trials in which the game ends followed by their probabilities:

n	Probability
5	$\frac{1}{32} = 0.03125$
6	$\frac{1}{64} = 0.015625$
7	$\frac{1}{64} = 0.015625$
8	$\frac{1}{64} = 0.015625$
9	$\frac{1}{64} = 0.015625$
10	$\frac{1}{64} = 0.015625$
11	$\frac{31}{2048} = 0.015137$
12	$\frac{61}{4096} = 0.014893$
13	$\frac{15}{1024} = 0.014648$
14	$\frac{59}{4096} = 0.014404$

A graph of these probabilities is as follows.



As r increases, one finds graphs very similar to that earlier.

Here are two probabilities that would be very difficult to do without a computer algebra system. The probability the game lasts n trials of course decreases as n increases. The probability the game lasts 150 trials is

$$\frac{1989179797398599794811479787771439644489521}{1427247692705959881058285969449495136382746624} = 0.00139372.$$

With 20 players, the probability the game ends in 30 trials is $\frac{1}{524288} = 1.90735 \times 10^{-6}$.

Probabilities of Each Player

We found that the probability A wins the series is equal to the probability B wins the series when there are three players. It is obvious that, since the game is fair, the probability A wins is the same as the probability B wins no matter the number of players. Finding the probability of a given player of winning the series is considerably more difficult than in the case of three players. Bernoulli proved, numbering the players now and letting p_i denote the probability that player i wins the series, that

$$p_1 = p_2 \text{ and } p_{i+1} = \frac{2^n}{1 + 2^n} p_i \text{ for } i = 2, 3, \dots, r \text{ where there are } r \text{ players in the game.}$$

We find then the following probabilities for some values of r :

r	p_A	p_S	p_C	p_D	p_E	p_F
2	$\frac{4}{15}$	$\frac{4}{15}$	$\frac{4}{14}$			
3	$\frac{81}{298}$	$\frac{81}{298}$	$\frac{72}{298}$	$\frac{64}{298}$		
4	$\frac{4913}{22898}$	$\frac{4913}{22898}$	$\frac{4624}{22898}$	$\frac{4352}{22898}$	$\frac{4096}{22898}$	
5	$\frac{1185921}{6766882}$	$\frac{1185921}{6766882}$	$\frac{1149984}{6766882}$	$\frac{1115136}{6766882}$	$\frac{1081344}{6766882}$	$\frac{1048576}{6766882}$

Since the fractions in the above-mentioned table are not simplified, it is clear that as the number of players increases, the probability that any one of them wins the series approaches $\frac{1}{r+1}$. That is also obvious since the factor relating successive probabilities, $\frac{2^n}{1+2^n} = \frac{1}{1+\left(\frac{1}{2}\right)^n}$, approaches 1 fairly rapidly. For 100 players, the probability that any one of them wins the series is very close to 0.01. For a proof of Bernoulli's result, see Hald [17], p. 378ff.

Expected Length of the Series

The expected length of the series for more than three players also becomes quite difficult. If we look at the expected length of the series for four players, we find that

$$E[N] = 3 \cdot 2 \cdot \left(\frac{1}{2}\right)^3 + 4 \cdot 2 \cdot \left(\frac{1}{2}\right)^4 + 5 \cdot 4 \cdot \left(\frac{1}{2}\right)^5 + 6 \cdot 6 \cdot \left(\frac{1}{2}\right)^6 + 7 \cdot 10 \cdot \left(\frac{1}{2}\right)^7 + 8 \cdot 16 \cdot \left(\frac{1}{2}\right)^8 + \dots, \text{ a challenging series to add to say the least.}$$

Mathematica, however, finds the sum to be 3 after using about 80 terms in the series, so the series is not only difficult, but it converges very slowly.

For five players, the series converges to 15, and for six players, the series converges to 31. We conjecture, and offer no proof of this, that

$$E[N] = 2^r - 1 \text{ for } r+1 \text{ players.}$$

Fibonacci Series

For four players, we found the Fibonacci-like series 2, 2, 4, 6, 10, 16, 26, ..., where after starting with 2, 2 we find successive terms by adding the previous two terms. We also found that the number of points in the sample space follows this sequence.

It is interesting to note that this is twice the usual Fibonacci series and is exactly the series one gets if we consider flipping a fair coin and waiting for two heads, *HH*, in a row.

<i>HH</i>
<i>THH</i>
<i>HTHH</i>
<i>TTHH</i>
<i>HTTHH</i>
<i>THTHH</i>
<i>TTTHH</i>

The sample space in that case is where the number of sample points in each successive sequence is found by adding the number of points in the previous two sequences, producing the usual Fibonacci series. The reason for this is that if the series is to end in *HH*, in *n* tosses, it must either be preceded by *HT* followed by *HH* in the remaining *n* – 1 tosses, or that it begins with *T* followed by *HH* in *n* – 2 tosses. This is very similar to our reasoning in the Waldegrave problem.

For more than four players in the Waldegrave problem, or for waiting for three or more heads in a row, we find “super Fibonacci” series where we start with 1, 1, 2 and add the previous three terms to arrive at the next term to find the number of sample points for a given length of the series.

7.13 CONCLUSION

The Waldegrave problem, for more than three players, presents some very nontrivial analytical mathematical questions, but we can do calculation with a computer system such as Mathematica.

7.14 ON HUYGEN'S FIRST PROBLEM

Christian Huygens proposed the following problem: player *A* wins if he casts a sum of 6 before his opponent, player *B*, casts a sum of 7 with two fair dice. The problem was considered by many mathematicians, including Fermat and James Bernoulli.

However, various scholars studying the problem presumed quite different orders of play, and these orders of play greatly influence the answer to the question. We will consider several different orders of play and the probability that *A* will win the game.

First, Huygens assumed that the order of play would be *ABBAABBA* . . . To generalize the problem a bit, suppose that the probability *A* wins at a particular trial is p_1 and the consequent probability that *A* loses at a particular trial is $1 - p_1 = q_1$ and the probability that *B* wins at a particular trial as p_2 with $q_2 = 1 - p_2$.

A can win the game in two mutually exclusive ways: *A* can win on trials 1, 4, 8, 12 . . . or *A* can win on trials 5, 9, 13,

The probability *A* wins on the first sequence is

$$P_1 = p_1 + q_1 q_2^2 p_1 + q_1 q_2^2 q_1^2 q_2^2 p_1 + q_1 q_2^2 q_1^2 q_2^2 q_1^2 q_2^2 p_1 + \dots, = p_1 \left[1 + \frac{q_1 q_2^2}{1 - q_1^2 q_2^2} \right] \text{ while}$$

the probability *A* wins on the second sequence is $P_2 = q_1 q_2^2 q_1 p_1 + q_1 q_2^2 q_1^2 q_2^2 q_1 p_1 + q_1 q_2^2 q_1^2 q_2^2 q_1^2 q_2^2 q_1 p_1 + \dots = p_1 \left[\frac{q_1^2 q_2^2}{1 - q_1^2 q_2^2} \right]$.

$$\text{Then } P(\text{A wins}) = P_1 + P_2 = p_1 \left[\frac{1 + q_1 q_2^2}{1 - q_1^2 q_2^2} \right].$$

$$\text{If we let } p_1 = 5/36 \text{ and } p_2 = 1/6, \text{ we find } P(\text{A wins}) = \frac{10355}{22631} = 0.45756.$$

If the game is fair so that $p_1 = p_2 = 1/2$, then $P(\text{A wins}) = \frac{1}{2} \frac{1+(1/2)^3}{1-(1/2)^4} = \frac{3}{5}$, giving, for these probabilities, the advantage clearly to *A*.

7.15 CHANGING THE SUMS FOR THE PLAYERS

It is possible to compute the probabilities that player *A* casts a sum of *a* before player *B* casts a sum of *b* for *a* or *b* = 2, 3, . . . , 12.

The following table gives all these probabilities and is to be read this way: if *a* = 4 and *b* = 6, the probability of *A* shooting a sum of 4 before *B* shoots a sum of 6 is $\frac{26123}{70343} = 0.37137$.

Note that the probabilities for any value of *b* is the same for values of *a* summing to 14 since the sum of the tops and the bottoms of two dice total 14.

Probabilities for Huygens' First Problem

$a = 2$ or 12	b	2	3	4	5	6	7
$a = 3$ or 11		$1261/2521$	$21779/65879$	$9419/38399$	$1289/6644$	$80291/502391$	$2171/16031$
$a = 4$ or 10		$44153/65879$	$307/613$	$4649/11687$	$1273/3874$	$7933/28435$	$1073/4439$
$a = 5$ or 9		$29027/3839$	$7067/11687$	$133/265$	$419/980$	$26123/70343$	$707/2159$
$a = 6$ or 8		$2683/3322$	$1307/1937$	$283/490$	$73/145$	$2419/5434$	$73/145$
$a = 7$		$42355/502391$	$20623/28435$	$446757/70343$	$6125/10868$	$1141/2257$	$10355/22631$
		$13901/16031$	$3389/4439$	$1469/2159$	$403/658$	$12581/22631$	$31/61$
$a = 2$ or 12	b	8	9	10	11	12	12
$a = 3$ or 11		$80291/50239$	$1289/6644$	$9419/38399$	$21779/65879$	$1261/2521$	
$a = 4$ or 10		$7933/28435$	$1273/3874$	$4649/11687$	$307/613$	$44153/65879$	
$a = 5$ or 9		$26123//70343$	$419/980$	$133/265$	$7067/11687$	$29027/38399$	
$a = 6$ or 8		$2419/5434$	$73/145$	$283/490$	$1307/1937$	$2683/3322$	
$a = 7$		$1141/2257$	$6125/10868$	$446757/70343$	$20623/28435$	$42355/502391$	
		$12581/22631$	$403/658$	$1469/2159$	$3389/4439$	$13901/16031$	

Notice that the probabilities for a and $14 - a$ are equal to that for $14 - a$ and a .

Decimal Equivalents

$a = 2$ or 12 : { $0.500198, 0.330591, 0.245293, 0.19401, 0.159818, 0.135425, 0.159818, 0.19401, 0.245293, 0.330591, 0.500198$ }

$a = 3$ or 11 : { $0.670214, 0.500816, 0.397792, 0.328601, 0.278987, 0.241721, 0.278987, 0.328601, 0.397792, 0.500816, 0.670214$ }

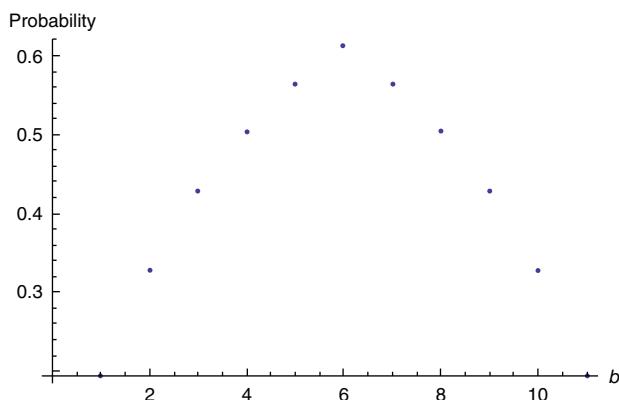
$a = 4$ or 10 : { $0.755931, 0.604689, 0.501887, 0.427551, 0.371366, 0.327466, 0.371366, 0.427551, 0.501887, 0.604689, 0.755931$ }

$a = 5$ or 9 : { $0.807646, 0.674755, 0.577551, 0.503448, 0.44516, 0.398176, 0.44516, 0.503448, 0.577551, 0.674755, 0.807646$ }

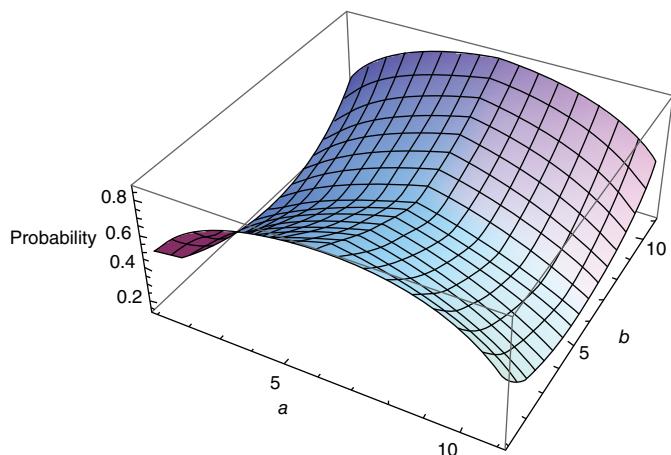
$a = 6$ or 8 : { $0.842282, 0.725268, 0.635102, 0.563581, 0.505538, 0.457558, 0.505538, 0.563581, 0.635102, 0.725268, 0.842282$ }

$a = 7$: { $0.867132, 0.76346, 0.680408, 0.612462, 0.555919, 0.508197, 0.555919, 0.612462, 0.680408, 0.76346, 0.867132$ }

Here is a plot of the probabilities for various values of b if $a = 5$.



A contour plot is also interesting:



Another order

If we change the order of play to $ABABAB \dots$, then

$$P(A \text{ wins}) = p_1 + q_1 q_2 p_1 + q_1 q_2 q_1 q_2 p_1 + \dots = \frac{p_1}{1 - q_1 q_2}.$$

If we let $p_1 = 5/36$ and $p_2 = 1/6$, we find $P(A \text{ wins}) = \frac{5/36}{1 - (31/36)(5/6)} = 0.491803$.

Were the game fair, then $P(A \text{ wins}) = \frac{1/2}{1 - 1/4} = \frac{2}{3}$.

One could of course calculate a table similar to that earlier for all possible values of a and b .

Bernoulli's Sequence

Bernoulli proposed the sequence $ABBAABBAAABBBBAAAABBBBB \dots$, which makes the problem much more difficult.

We look at some possibilities until we see a pattern.

First, A can win on trials 1,3,7,13,21, ... and this has probability

$$P_1 = p_1 + q_1 q_2 p_1 + q_1 q_2 q_1^2 q_2^2 p_1 + q_1 q_2 q_1^2 q_2^2 q_1^3 q_2^3 p_1 + \dots = \sum_{k=0}^{\infty} (q_1 q_2)^{\frac{k(k+1)}{2}} p_1$$

A could also win on trials 4, 8, 14, 22, ... and this has probability

$$P_2 = q_1 q_2 q_1 p_1 + q_1 q_2 q_1^2 q_2^2 q_1 p_1 + q_1 q_2 q_1^2 q_2^2 q_1^3 q_2^3 q_1 p_1 + \dots = \sum_{k=1}^{\infty} (q_1 q_2)^{\frac{k(k+1)}{2}} q_1 p_1$$

And the probability A wins on trials 9, 15, 23, ... is

$$P_3 = q_1 q_2 q_1^2 q_2^2 q_1^2 p_1 + q_1 q_2 q_1^2 q_2^2 q_1^3 q_2^3 q_1^2 p_1 + q_1 q_2 q_1^2 q_2^2 q_1^3 q_2^3 q_1^4 q_2^4 q_1^2 p_1 = \sum_{k=2}^{\infty} (q_1 q_2)^{\frac{k(k+1)}{2}} q_1^2 p_1$$

and one could go on but the pattern is evident.

We see then that $P(A \text{ wins}) = \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} (q_1 q_2)^{\frac{j(j+1)}{2}} q_1^j p_1$.

The series converges, but quite a lot of arithmetic is involved. Taking 10 terms in the series gives

$P(A \text{ wins}) = 0.490093$, and this is the same result as taking 30 terms in the series.

Were the game fair, then $P(A \text{ wins}) = \frac{3}{5}$, so the last two series give exactly the same probability that A wins the game.

Bibliography

WHERE TO LEARN MORE

There is now a vast literature on the theory of probability. A few of the following references are cited in the text; other titles that may be useful to the instructor or student are included here as well.

1. Anscombe, F. J., Graphs in Statistical Analysis, *The American Statistician*, 27, 17–21, 1979.
2. Bernoulli, Jacob, *Ars Conjectandi*, (1713), Johns Hopkins, 2006.
3. Blom, Gunnar, Lars Holst and Dennis Sandell, *Problems and Snapshots from the World of Probability* Springer-Verlag, 1994.
4. George E. P. Box, William G. Hunter and J. Stuart Hunter, *Statistics for Experimenters*, John Wiley & Sons, 1978.
5. Cook, R. D., Selection of Influential Observations in Linear Regression, *Technometrics*, 19, 15–18, 1972.
6. David, F. N. and D. E. Barton, *Combinatorial Chance*, Charles Griffin & Company Limited, 1962.
7. Drane, J. Wanzer, Suhua Cao, Lixia Wang and T. Postelnicu, Limiting Forms of Probability Mass Functions via Recurrence Formulas, *The American Statistician*, November 1993, Vol. 47, No. 4, p 269–274.
8. Draper, N.R. and H. Smith, *Applied Regression Analysis*, Second Edition, John Wiley & Sons, 1981.
9. Duncan, Acheson J., *Quality Control and Industrial Statistics*, Fifth Edition, Richard D. Irwin, Inc., 1986.
10. Efron, Bradley, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, CBMS-NSF Monography, Volume 38, 1982.
11. Feller, William, *An Introduction to Probability and Its Applications*, Volumes I and II, John Wiley & Sons, 1968.
12. Gnedenko, B. V., *The Theory of Probability*, Chelsea Publishing Company, Fifth Edition, 1989.
13. Goldberg, Samuel, *Probability, An Introduction*, Prentice-Hall, Inc., 1960.
14. Goldberg, Samuel *Introduction to Difference Equations*, Dover Publications, 1986.
15. Grant, Eugene L. and Richard S. Leavenworth, *Statistical Quality Control*, Sixth Edition, McGraw-Hill, 1988.
16. Grimaldi, Ralph P., *Discrete and Combinatorial Mathematics*, Fifth Edition, Addison-Wesley Publishing Co., Inc., 2004.
17. Hald, Anders, *A History of Probability and Statistics and Their Applications Before 1750* John Wiley & Sons, 1990.
18. Hogg, Robert V. and Allen T. Craig, *Introduction to Mathematical Statistics*, Fourth Edition, Macmillan Publishing Co, 1986.
19. Huff, Barthel W., *Another Definition of Independence*, Mathematics Magazine, September-October, 1971, pp. 196–197.
20. Hunter, Jeffery L., *Mathematical Techniques of Applied Probability*, Volumes 1 and 2, Academic Press, 1983.
21. Isaacson, Dean L. and Richard W. Madsen, *Markov Chains, Theory and Applications*, John Wiley & Sons, 1976.
22. Johnson, Norman L., Samuel Kotz and Adrienne W. Kemp, *Univariate Discrete Distributions*, Second Edition, John Wiley & Sons, 1992.

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

23. Johnson, Norman L, Samuel Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, volumes 1 and 2, second edition, John Wiley and Sons, 1994.
24. Kemeny, John G. and J. Laurie Snell, *Finite Markov Chains*, Springer-Verlag, 1976.
25. Kinney, John J., Tossing Coins Until All Are Heads, Mathematics Magazine, May, 1978, p184–186.
26. Kinney, John J., *A Probability and Statistics Companion*, John Wiley & Sons, 2009.
27. Kinney, John J., *Statistics for Science and Engineering*, Addison-Wesley Publishing Co., Inc., 2002.
28. Mosteller, Frederick, *Fifty Challenging Problems in Probability*, Addison-Wesley Publishing Co., Inc., 1965. Reprinted by Dover Publications.
29. Rao, C. Radhakrishna, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, 1973.
30. Riordan, John, *Combinatorial Identities*, John Wiley & Sons, 1968.
31. Ross, Sheldon, *A First Course in Probability*, Sixth Edition, Prentice Hall, 2002.
32. Salsburg, David, *The Lady Tasting Tea: How Science Revolutionized Science in the Twentieth Century*, W. H. Freeman and Company, 2001.
33. Silver, Nate, *The Signal and the Noise*, Penguin Books, 2013.
34. Thompson, W. A. Jr., On the Foundations of Reliability, *Technometrics*, February 1981, Vol. 23, No. 1, pp. 1–13.
35. Uspensky, J. V., *Introduction to Mathematical Probability*, McGraw-Hill Book Company, Inc., 1937.
36. Welch, B. L., The Significance of the Difference Between Means When the Population Variances are Unequal, *Biometrika*, 1938, Vol. 29, pp. 350–362.
37. When Aids Tests are Wrong, The New York Times, September 5, 1987, p. 22.
38. Whitworth, William Allen, *Choice and Chance*, Fifth Edition, Hafner Publishing Company, 1965.
39. Wolfram, Stephen, *Mathematica: A System for Doing Mathematics by Computer*, Addison-Wesley Publishing Co., 1991.

Appendix A

Use of Mathematica in Probability and Statistics

Reference is often made in the text to a computer algebra system. *Mathematica* was used for much of the work in this book, but other computer algebra systems are also capable of doing much of the work we do. We give here examples of the use of all the commands in *Mathematica* that have been used in the examples and graphs in the text, but we do not show the creation of every graph. In addition, no attempt is made to carry out our tasks in the most efficient manner; the reader will find that *Mathematica* offers many paths to the same result; the reader is encouraged to explore other ways to achieve the results shown here.

The material here is referred to by chapter in the text and by examples within that chapter. We often do not repeat the conditions of the examples, so the reader should read the text before studying the solutions. No attempt is made here to explain *Mathematica* syntax; the reader is directed to the extensive help pages for each of the commands we show here.

The text contains many graphs which are not reproduced here. Entries in *Mathematica* appear in bold-face type; the responses follow in ordinary type.

A simple calculation is necessary to load the *Mathematica* kernel. Then all the commands shown here will work as shown; no other knowledge in experience with *Mathematica* is necessary. This appendix is in actuality a *Mathematica* notebook and will run on a computer exactly as shown here with the exception of examples which use random samples; they will vary each time the program is run.

CHAPTER ONE

Section 1.1 Discrete Sample Spaces

The Fibonacci recursion is $a_n = a_{n-1} + a_{n-2}$, $a_1 = 1$; $a_2 = 1$. Values for the recursion can be found directly using the recursion.

```
In[1]= a[n_] := a[n] = a[n - 1] + a[n - 2]
In[2]= a[1] = 1
Out[2]= 1
In[3]= a[2] = 1
Out[3]= 1
In[4]= Table[a[n], {n, 1, 15}]
Out[4]= {1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610}
```

The 25th Fibonacci number is

```
In[5]= a[25]
Out[5]= 75 025
```

Section 1.4 Conditional Probability and Independence

Example 1.4.4

Figure 1.9 shows the graph of $P(A|T+)$ as a function of $P(A)$. It was drawn as follows:

```
In[6]= f[p_] := (20 000 p) / (1 + 19 999 p)
In[7]= Plot[f[p], {p, 0, 1}, Frame → True, FrameLabel → {"P(A)", "P(A|T+)"}, LabelStyle → FontFamily → "Helvetica-Bold"]
```

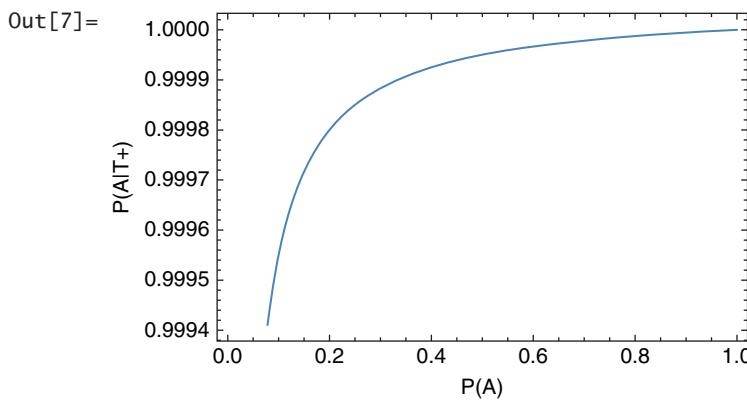
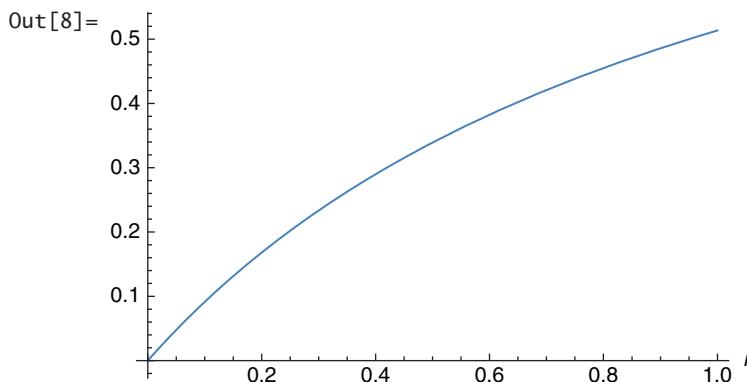


Figure 1.10 was drawn as follows:

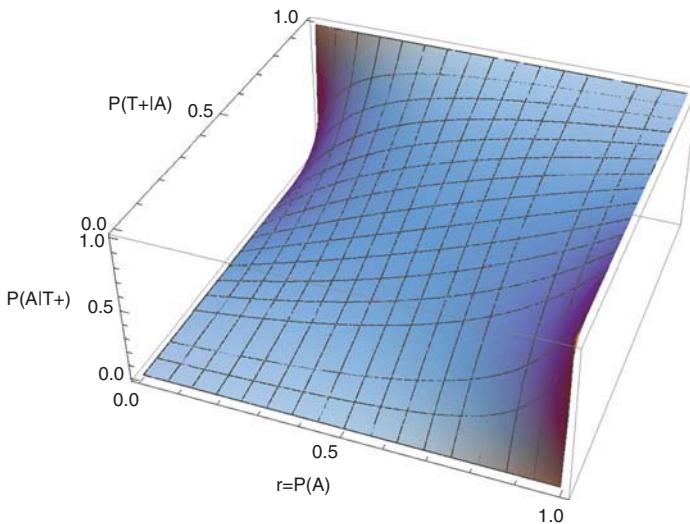
```
In[8]= Plot[(0.95 * r) / (0.95 + .90 * r), {r, 0, 1}, AxesLabel → {r=P(A)", "P(A|T+)", "P(T+|A)"}, LabelStyle → {FontFamily → "Helvetica-Bold"}]
```



This section also shows a three-dimensional graph of $P(A | T+)$ as a function of both the incidence rate of the disease, r , as well as $p = P(T+ | A)$. This was done as follows:

```
In[9]= f[r_, p_] := r * p / (r * p + (1 - r) * (1 - p))
In[10]= Plot3D[f[r, p], {r, 0, 1}, {p, 0, 1},
  AxesLabel -> {"r=P(A)", "P(T+|A)", "P(A|T+)"}, ViewPoint -> {0.965,
  -2.553, 2.000},
  LabelStyle -> {FontFamily -> "Helvetica-Bold"}, PlotPoints -> 50]
```

Out[10]=



Example 1.5.1 The Birthday Problem

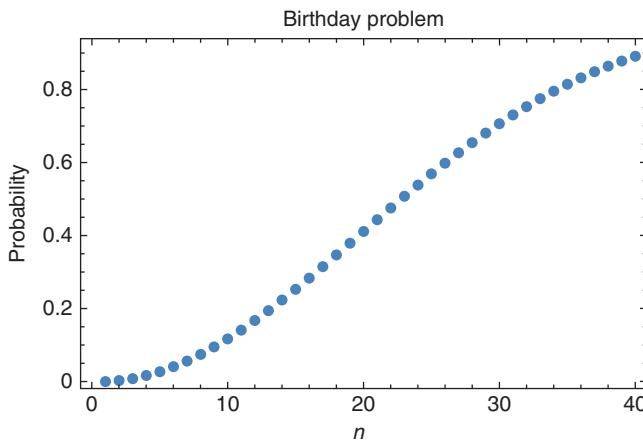
The table with exact values of $P(A)$ was constructed with this instruction:

```
In[11]= probs = Table[{i, 1 - Product[(366 - r) / 365, {r, i}]], {i, 1, 40}]//N
Out[11]= {{1., 0.}, {2., 0.00273973}, {3., 0.00820417}, {4., 0.0163559},
{5., 0.0271356}, {6., 0.0404625}, {7., 0.0562357}, {8., 0.0743353},
{9., 0.0946238}, {10., 0.116948}, {11., 0.141141}, {12., 0.167025},
{13., 0.19441}, {14., 0.223103}, {15., 0.252901}, {16., 0.283604},
{17., 0.315008}, {18., 0.346911}, {19., 0.379119}, {20., 0.411438},
{21., 0.443688}, {22., 0.475695}, {23., 0.507297}, {24., 0.538344},
{25., 0.5687}, {26., 0.598241}, {27., 0.626859}, {28., 0.654461},
{29., 0.680969}, {30., 0.706316}, {31., 0.730455}, {32., 0.753348},
{33., 0.774972}, {34., 0.795317}, {35., 0.814383}, {36., 0.832182},
{37., 0.848734}, {38., 0.864068}, {39., 0.87822}, {40., 0.891232}}
```

The graph in Figure 1.13 was drawn with these commands:

```
In[12]= values = Table[i, {i, 1, 40}]
Out[12]= {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38,
39, 40}
In[13]= ListPlot[probs, Frame -> True, FrameLabel -> {"n", "Probability"}, Ticks ->
{values, Automatic}, PlotLabel -> ("Birthday Problem"), LabelStyle ->
{FontFamily -> "Helvetica-Bold"}]
```

Out[13]=



Section 1.7 Counting Techniques

Mathematica does exact arithmetic:

```
In[14]= 52 !
Out[14]= 80 658 175 170 943 878 571 660 636 856 403 766 975 289 505 440 883 277
          824 000 000 000 000
```

This may convince the reader that a random order of a deck of cards is truly rare!

The number of permutations of r objects selected from n distinct objects is $n!/(n - r)!$. For example, the number of distinct arrangements of 30 objects chosen from 56 objects is $56!/26!$. Mathematica will evaluate this exactly.

```
In[15]= 56! / 26!
Out[15]= 1 762 989 441 479 047 465 097 977 043 769 075 758 530 560 000 000
```

Here are some examples of permutations.

```
In[16]= Permutations [{a, b, c}]
Out[16]= {{a, b, c}, {a, c, b}, {b, a, c}, {b, c, a}, {c, a, b}, {c, b, a}}
```

If some of the objects are alike, only the distinct permutations are returned:

```
In[17]= perms = Permutations[{a, a, b, b, c}]
Out[17]= {{a, a, b, b, c}, {a, a, b, c, b}, {a, a, c, b, b}, {a, b, a, b, c},
           {a, b, a, c, b}, {a, b, b, a, c}, {a, b, b, c, a}, {a, b, c, a, b},
           {a, b, c, b, a}, {a, c, a, b, b}, {a, c, b, a, b}, {a, c, b, b, a},
           {b, a, a, b, c}, {b, a, a, c, b}, {b, a, b, a, c}, {b, a, b, c, a},
           {b, a, c, a, b}, {b, a, c, b, a}, {b, b, a, a, c}, {b, b, a, c, a},
           {b, b, c, a, a}, {b, c, a, a, b}, {b, c, a, b, a}, {b, c, b, a, a},
           {c, a, a, b, b}, {c, a, b, a, b}, {c, a, b, b, a}, {c, b, a, a, b},
           {c, b, a, b, a}, {c, b, b, a, a}}
```

```
In[18]= Length[ perms]
Out[18]= 30
```

We can check that this is

```
In[19]= 5! / (2! 2! 1!)
Out[19]= 30
```

A sample of 3 random permutations of the letters in the set $\{a, a, b, b, c\}$ can be found as follows:

```
In[20]= RandomChoice[Permutations[{a, a, b, b, c}], 3]
Out[20]= {{a, b, a, b, c}, {a, c, b, b, a}, {a, c, b, b, a}}
```

The number of combinations are found using the Binomial $[n, r] = n!/(r!(n - r)!)$ function. The number of distinct poker hands is:

```
In[21]= Binomial[52, 5]
Out[21]= 2 598 960
```

Example 1.7.2

```
In[22]= Binomial [3, 1] * Binomial [8, 4] / Binomial [11, 5]
```

```
Out[22]= 5/11
```

Example 1.7.3 *The Matching Problem*

If there are 4 numbers to be permuted, we can generate all $4! = 24$ permutations of these 4 digits as follows:

```
In[23]= Permutations[{1, 2, 3, 4}]
Out[23]= {{1, 2, 3, 4}, {1, 2, 4, 3}, {1, 3, 2, 4}, {1, 3, 4, 2}, {1, 4, 2, 3},
{1, 4, 3, 2}, {2, 1, 3, 4}, {2, 1, 4, 3}, {2, 3, 1, 4}, {2, 3, 4, 1},
{2, 4, 1, 3}, {2, 4, 3, 1}, {3, 1, 2, 4}, {3, 1, 4, 2}, {3, 2, 1, 4},
{3, 2, 4, 1}, {3, 4, 1, 2}, {3, 4, 2, 1}, {4, 1, 2, 3}, {4, 1, 3, 2},
{4, 2, 1, 3}, {4, 2, 3, 1}, {4, 3, 1, 2}, {4, 3, 2, 1}}
```

If we would like all the permutations of 5 integers each of length 3 this can be done as follows:

```
In[24]= Permutations [{1, 2, 3, 4, 5}, {3}]
Out[24]= {{1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 2}, {1, 3, 4}, {1, 3, 5},
{1, 4, 2}, {1, 4, 3}, {1, 4, 5}, {1, 5, 2}, {1, 5, 3}, {1, 5, 4},
{2, 1, 3}, {2, 1, 4}, {2, 1, 5}, {2, 3, 1}, {2, 3, 4}, {2, 3, 5},
{2, 4, 1}, {2, 4, 3}, {2, 4, 5}, {2, 5, 1}, {2, 5, 3}, {2, 5, 4},
{3, 1, 2}, {3, 1, 4}, {3, 1, 5}, {3, 2, 1}, {3, 2, 4}, {3, 2, 5},
{3, 4, 1}, {3, 4, 2}, {3, 4, 5}, {3, 5, 1}, {3, 5, 2}, {3, 5, 4},
{4, 1, 2}, {4, 1, 3}, {4, 1, 5}, {4, 2, 1}, {4, 2, 3}, {4, 2, 5},
{4, 3, 1}, {4, 3, 2}, {4, 3, 5}, {4, 5, 1}, {4, 5, 2}, {4, 5, 3},
{5, 1, 2}, {5, 1, 3}, {5, 1, 4}, {5, 2, 1}, {5, 2, 3}, {5, 2, 4},
{5, 3, 1}, {5, 3, 2}, {5, 3, 4}, {5, 4, 1}, {5, 4, 2}, {5, 4, 3}}
In[25]= Length[%]
Out[25]= 60
```

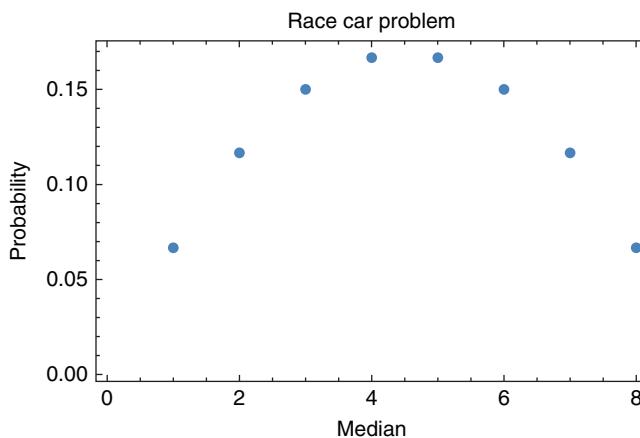
This verifies that this list of permutations should be of length $5 * 4 * 3 = 60$.

Example 1.7.5 Race Cars

Section 1.7 also shows graphs of the probability distribution of the median in the race car example. They were drawn this way:

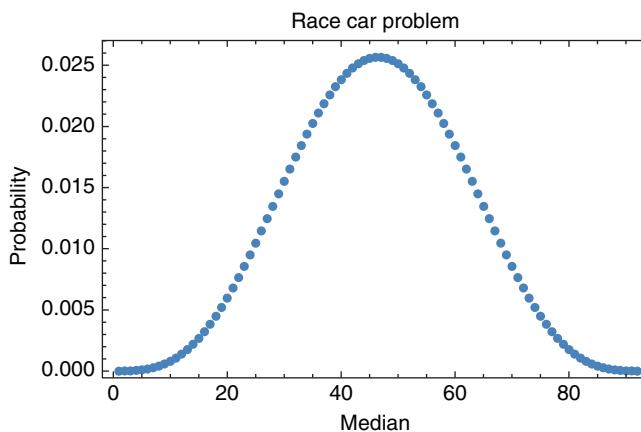
```
In[26]= med1[k_] := (k - 1) * (10 - k) / Binomial [10, 3]
In[27]= ListPlot [Table [med1 [k], {k, 2, 9}], Frame → True,
FrameLabel → {"Median", "Probability"}, PlotLabel → ("Race Car Problem"),
LabelStyle → (FontFamily → "Helvetica-Bold")]
```

Out[27]=



```
In[28]= med[k_] := Binomial[1, 1] * Binomial[k - 1, 4] * Binomial[100 - k, 4] /
Binomial [100, 9]
In[29]= ListPlot[Table [med[k], {k, 5, 96}], Frame → True,
FrameLabel → {"Median", "Probability"}, PlotLabel → ("Race Car Problem"),
LabelStyle → (FontFamily → "Helvetica-Bold")]
```

Out[29]=



CHAPTER TWO

Many discrete probability distributions, including all those in Chapter Two, are contained in Mathematica as defined functions. We give some examples of the use of these functions, of drawing graphs, and of drawing random samples from these distributions.

Section 2.1 Random Variables

Example 2.1.1

We show a random sample of 100 selected from the discrete uniform distribution $P(X = x) = 1/n$, $x = 1, 2, 3, \dots, n$. The starting and ending values for x must be specified. We take the range from 1 to 6 to simulate tosses of a fair die.

```
In[30]= data = Table[Random[DiscreteUniformDistribution[{1, 6}]], {100}]
Out[30]= {2, 5, 2, 4, 1, 4, 3, 6, 1, 2, 4, 6, 6, 2, 1, 6, 4, 5, 5, 3, 5, 3, 2,
3, 6, 6, 2, 4, 3, 4, 1, 1, 5, 3, 3, 4, 2, 4, 1, 3, 5, 4, 4, 2, 1, 5, 4,
1, 6, 6, 5, 3, 1, 3, 5, 2, 4, 1, 1, 4, 5, 6, 4, 6, 1, 4, 2, 4, 6, 1, 6,
4, 2, 4}
```

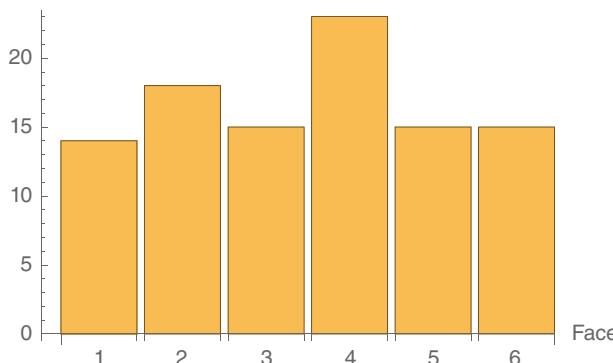
The data can then be organized by counting the frequency with which each integer occurs.

```
In[31]= freq = BinCounts[data, {1, 7, 1}]
Out[31]= {14, 18, 15, 23, 15, 15}
```

Now we can draw a histogram of the data:

```
In[32]= BarChart[{13, 20, 18, 18, 13, 18}, AxesLabel -> {"Face", "Frequency"},
ChartLabels -> {1, 2, 3, 4, 5, 6}]
```

```
Out[32]= Frequency
```



Example 2.1.2

Sampling from the die loaded so the probability that a face appears is proportional to the face is a bit more complex than sampling from a fair die. We sample from a discrete uniform distribution with values from 1 to 21; the value 1 becomes a one on the die; the next two

values, namely 2 or 3, become a 2 on the loaded die; the next three values are 4, 5, or 6 and they become a 3 on the loaded die, and so on.

```
In[33]= data1 = RandomVariate[ DiscreteUniformDistribution[{1, 21}], 200];
```

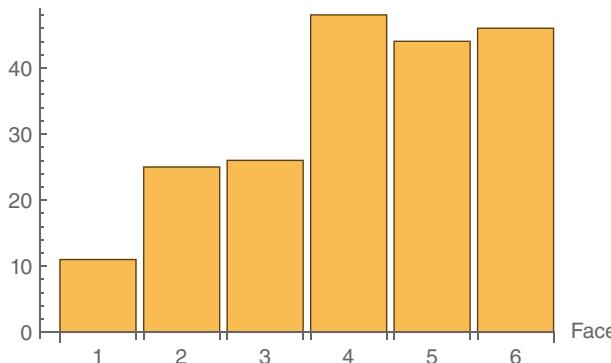
The semi-colon depresses printing the output.

```
In[34]= freq1 = BinCounts [data1, {1, 22, 1}]
Out[34]= {11, 14, 11, 8, 8, 10, 13, 16, 9, 10, 9, 8, 15, 5, 7, 10, 7, 5, 7, 7, 10}
```

Now we collate the data:

```
In[35]= orgdata = Table [Take [freq!, {(1/2) * (2 - m + m^2), m * (m + 1) / 2}], {m, 1, 6}]
Out[35]= {{11}, {14, 11}, {8, 8, 10}, {13, 16, 9, 10}, {9, 8, 15, 5, 7}, {10, 7, 5, 7, 7, 10}}
In[36]= orgfreq = Apply[Plus, orgdata, {1}]
Out[36]= {11, 25, 26, 48, 44, 46}
In[37]= BarChart[{11, 25, 26, 48, 44, 46},
AxesLabel → {"Face", "Frequency"}, ChartLabels → {1, 2, 3, 4, 5, 6}]
```

Out[37]= Frequency



Example 2.1.3

The probability distribution when two dice are thrown is

$$P(X = x) = P(X = 14 - x) = (x - 1)/36 \text{ for } x = 1, 2, 3, 4, 5, 6, 7.$$

A graph of this function is shown in Figure 2.2 and can be generated by the following commands

```
In[38]= prob1 = Table [(x - 1) / 36, {x, 2, 7}]
```

```
Out[38]= {1/36, 1/18, 1/12, 1/9, 5/36, 1/6}
```

In[39]= **prob2 = Table[(14 - x) / 36, {x, 9, 13}]**

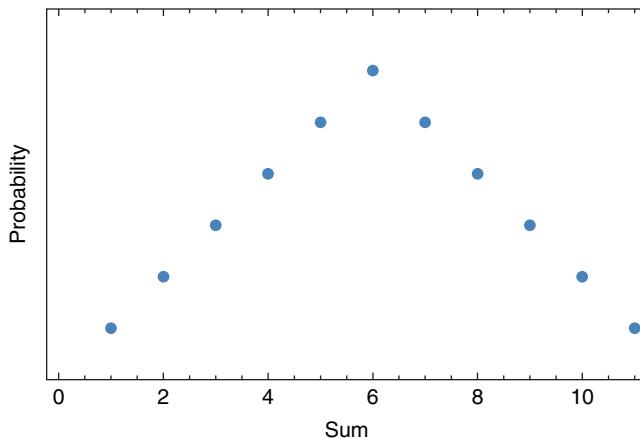
$$\text{Out}[39] = \left\{ \frac{5}{36}, \frac{1}{9}, \frac{1}{12}, \frac{1}{18}, \frac{1}{36} \right\}$$

In[40]= **ts = Table[{i, i + 1}, {i, 1, 11}]**

$$\text{Out}[40] = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 6\}, \{6, 7\}, \{7, 8\}, \{8, 9\}, \{9, 10\}, \{10, 11\}, \{11, 12\}\}$$

In[41]= **ListPlot[Flatten [{prob1, prob2}, 1], Frame → True, AxesOrigin → {0, 0}, FrameLabel → {"Sum", "Probability"}, PlotRange → {0, 0.20}, FrameTicks → {ts, Automatic}, LabelStyle → (FontFamily → "Helvetica-Bold")]**

Out[41]=



Another way to do this is to use a generating function:

In[42]= **g[t_]:= Sum[(1 / 6) t^i, {i, 1, 6}]**
g[t]

$$\text{Out}[42] = \frac{t}{6} + \frac{t^2}{6} + \frac{t^3}{6} + \frac{t^4}{6} + \frac{t^5}{6} + \frac{t^6}{6}$$

The coefficients of $g[t]$ give the probability that the die shows a particular face. The coefficients of $g[t]^2$ give the probabilities of the sum on two dice:

In[43]= **Expand[g[t]^2]**

$$\text{Out}[43] = \frac{t^2}{36} + \frac{t^3}{18} + \frac{t^4}{12} + \frac{t^5}{9} + \frac{5t^6}{36} + \frac{t^7}{6} + \frac{5t^8}{36} + \frac{t^9}{9} + \frac{t^{10}}{12} + \frac{t^{11}}{18} + \frac{t^{12}}{36}$$

In[44]= **ListPlot[Drop[CoefficientList[Expand[g[t]^2], t], 2], Frame → True, AxesOrigin → {0, 0}, FrameLabel → {"Sum", "Probability"}, PlotLabel → "Sums on Two Fair Dice", PlotRange → {0, 0.18}, LabelStyle → (FontFamily → "Helvetica-Bold")]**

Out[44]=

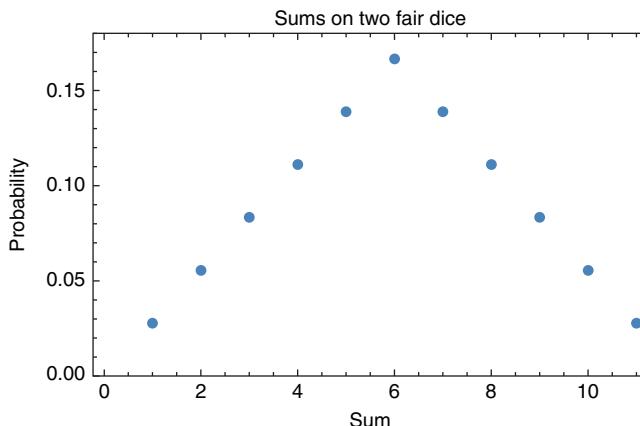
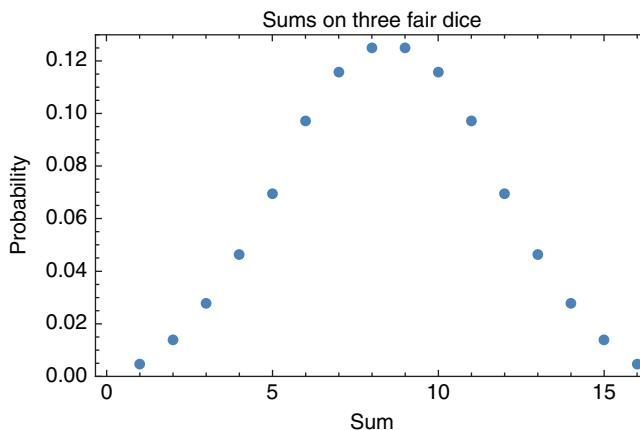


Figure 2.2

The coefficients of $g[t]^3$ give the probabilities of sums on three dice; here is a graph of the result:

```
In[45]= ListPlot[Drop[CoefficientList[Expand[g[t]^3], t], 3],
Frame → True, AxesOrigin → {1, 0}, FrameLabel → {"Sum",
"Probability"}, PlotLabel → "Sums on Three Fair Dice", PlotRange →
{0, 0.13}, LabelStyle → (FontFamily → "Helvetica-Bold")]
```

Out[45]=



Section 2.3 Expected Values of Discrete Random Variables

To find the mean and variance of the sums on three dice we proceed as follows:

```
In[46]= probsfor3 = Drop[CoefficientList[Expand[g[t]^3], t], 3]
```

```
Out[46]= {1/216, 1/72, 1/36, 5/108, 5/72, 7/216, 25/216, 1/8, 1/8, 25/216, 7/72, 5/108, 5/36, 1/72, 1/216}
```

```
In[47]= mean = Sum[i*probsfor3[[i - 2]], {i, 3, 18}]
```

```
Out[47]=  $\frac{21}{2}$ 
```

```
In[48]= variance = Sum[i^2*probsfor3[[i - 2]], {i, 3, 18}] - mean^2
```

```
Out[48]=  $\frac{35}{4}$ 
```

Section 2.4 Binomial Distribution

We show how to take a random sample of 1000 observations from a binomial distribution with $n = 20$ and $p = 3/4$:

```
In[49]= bidata = Table[Random[BinomialDistribution[20, 3 / 4]], {10}]
Out[49]= {14, 18, 15, 12, 15, 15, 13, 16, 14, 13}
In[50]= Histogram[bidata, 6, AxesLabel → {"Sum", "Probability"}, LabelStyle →
(FontFamily → "Helvetica-Bold")]
```

```
Out[50]= Probability
```

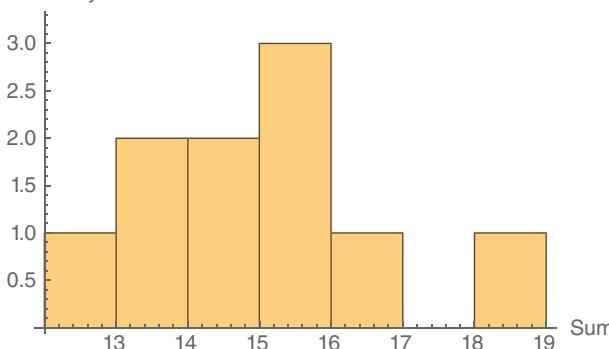


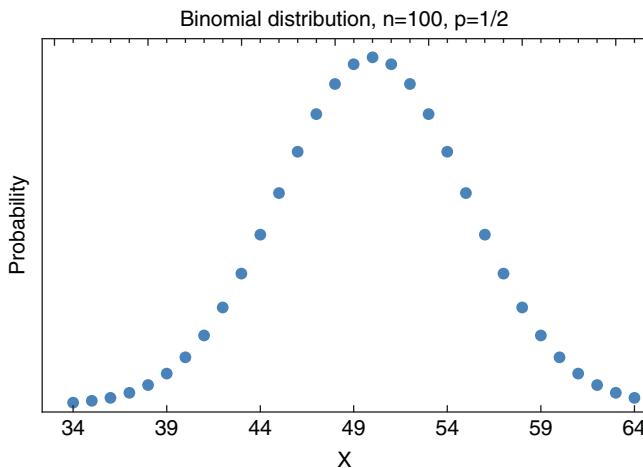
Figure 2.9 can be produced as follows:

```
In[51]= newxs = Table[{i, 33 + i}, {i, 1, 40, 5}]
```

```
Out[51]= {{1, 34}, {6, 39}, {11, 44}, {16, 49}, {21, 54}, {26, 59}, {31, 64}, {36, 69}}
```

```
In[52]= ListPlot[Table[PDF[BinomialDistribution[100, 1/2], x], {x, 34, 64}],
Frame → True, AxesOrigin → {0, 0}, FrameLabel → {"X",
"Probability"}, PlotLabel → "Binomial Distribution,n=100, p=1/2",
LabelStyle → (FontFamily → "Helvetica-Bold"), FrameTicks → {{None,
None}, {newxs, Automatic}}]
```

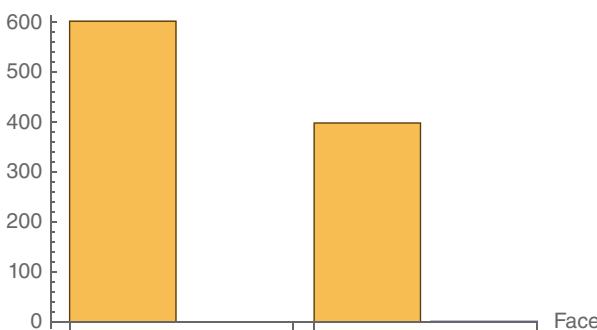
Out[52]=

**Example*****Flipping a Loaded Coin***

Here we simulate 1000 flips of a coin loaded to come up heads with probability $2/5$.

```
In[53]= biasdata = Table[Random[BinomialDistribution[1, 2/5]], {1000}];
In[54]= biasfreq = BinCounts[biasdata, {0, 2, 1}]
Out[54]= {602, 398}
In[55]= biasvalues = {0, 1}
Out[55]= {0, 1}
In[56]= BarChart[Transpose[{biasfreq, biasvalues}],
AxesLabel -> {"Face", "Frequency"}, LabelStyle -> (FontFamily ->
"Helvetica-Bold")]
```

Out[56]= Frequency



The bars here represent heads and tails.

Mathematica knows the mean and variance of the binomial distribution (as well as those moments for many other probability distributions):

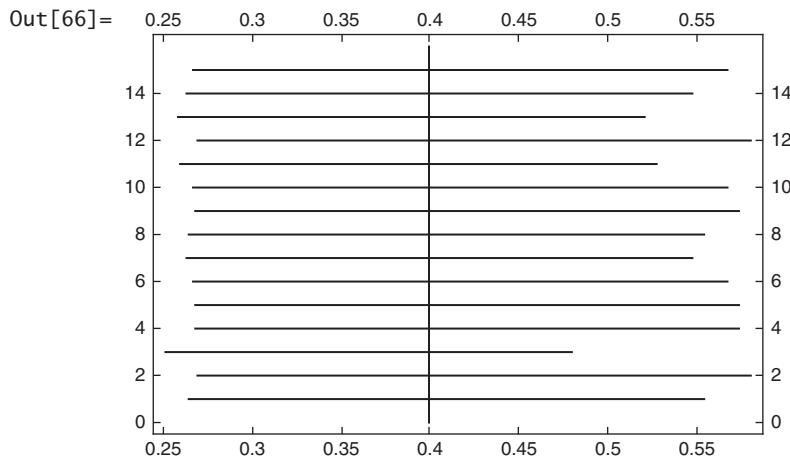
```
In[57]= Mean[BinomialDistribution[n, p] ]
Out[57]= n p
In[58]= Variance[BinomialDistribution[n, p] ]
Out[58]= n (1 - p)p
```

Section 2.6 Some Statistical Considerations

The confidence intervals in Figure 2.11 were generated and plotted as follows.

```
In[59]= soln = Simplify[Expand[Solve[100 p == X + 2 * Sqrt[100 p (1 - p)], p]]]
Out[59]= {{p → 1/520 (10 + 5x - √(100 + 100x - x^2))}, {p → 1/520 (10 + 5x + √(100 + 100x - x^2))}}
In[60]= leftend = p /. soln[[1]]
Out[60]= 1/520 (10 + 5X - √(100 + 100X - X^2))
In[61]= rightend = p /. soln[[2]]
Out[61]= 1/520 (10 + 5X + √(100 + 100X - X^2))
In[62]= chartdata = {40, 44, 29, 43, 43, 42, 39, 40, 43, 42, 36, 44, 35, 39, 42}
Out[62]= {40, 44, 29, 43, 43, 42, 39, 40, 43, 42, 36, 44, 35, 39, 42}
In[63]= endpts =
Table[{leftend, rightend} /. X → chartdata[[i]], {i, 1,
Length[chartdata]}]//N
Out[63]= {{0.307692, 0.5}, {0.344931, 0.539685}, {0.208721, 0.387433}, {0.335563,
0.529822}, {0.335563, 0.529822}, {0.326233, 0.519921}, {0.298482,
0.48998}, {0.307692, 0.5}, {0.335563, 0.529822}, {0.326233, 0.519921},
{0.271095, 0.459674}, {0.344931, 0.539685}, {0.26205, 0.449488},
{0.298482, 0.48998}, {0.326233, 0.519921}}
In[64]= vert = Show[Graphics[Line[{{13.3, 0}, {13.3, 16}}]];
In[65]= Show[Graphics[Table[Line[{{10*endpts[[i]][[1]], i},
{50*endpts[[i]][[2]], i}}], {i, 1, Length[endpts]}]], vert, Frame → True,
```

```
In[66]= FrameTicks → {{ {2, 0.25}, {5.8, 0.3}, {9.6, 0.35}, {13.3, 0.4}, {17.1, 0.45}, {20.9, 0.5}, {24.7, 0.55}}, [0, 2, 4, 6, 8, 10, 12, 14]}]
```



Section 2.7 Hypothesis Tests

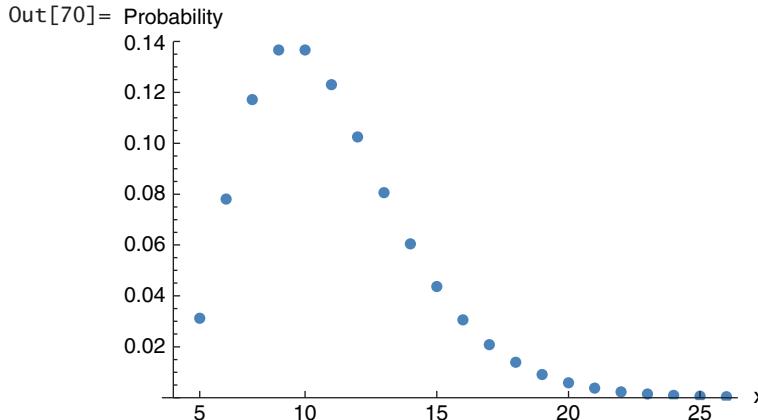
The alpha and beta errors in this section are sums of binomial probabilities.

```
In[67]= alpha = Sum[PDF[BinomialDistribution[20, 0.2], x], {x, 9, 20}]
Out[67]= 0.00998179
In[68]= beta = Sum[PDF[BinomialDistribution[20, 0.3], x], {x, 0, 8}]
Out[68]= 0.886669
```

Section 2.9 Geometric and Negative Binomial Distributions

We show how to draw Figure 2.13

```
In[69]= Negbin[x_, r_] := Binomial[x - 1, r - 1] * ((1 / 2)^r) * ((1 / 2)^(x - r))
In[70]= ListPlot[Table[Negbin[x, 5], {x, 5, 25}], AxesLabel → {"x", "Probability"}, PlotRange → {0, 0.14}, PlotLabel → "Negative Binomial Distribution, r=5, p=1/2", AxesOrigin → {0, 0}, LabelStyle → {FontFamily → "Helvetica-Bold"}, Ticks → {{1, 5}, {5, 10}, {10, 15}, {15, 20}, {20, 25}, {25, 30}}, Automatic}]
```



Section 2.10 Hypergeometric Distribution

Section 2.10.3

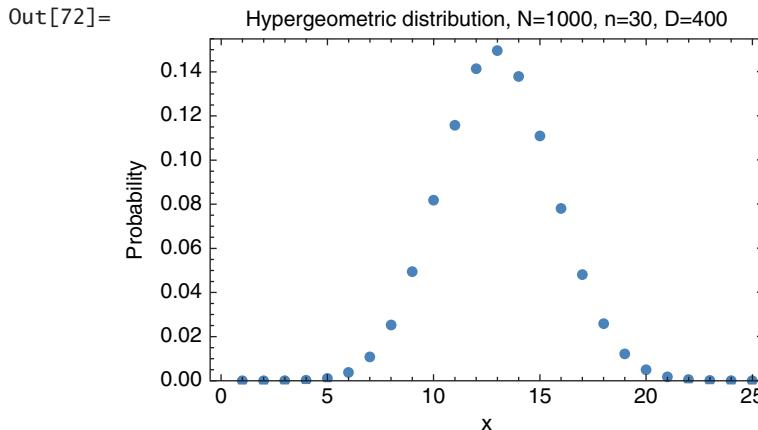
Figure 2.18 shows a hypergeometric distribution with $n = 30$, $D = 400$, and $N = 1000$. This distribution can be plotted with the following commands.

```
In[71]= hyperfcn = PDF[HypergeometricDistribution[30, 400, 1000], x]
```

Out[71]=
$$\begin{cases} \frac{\text{Binomial}[400,x]\text{Binomial}[600,30-x]}{242960819217374510327038983857675071930222606198631438800} & \text{True} \\ 0 & \text{False} \end{cases}$$

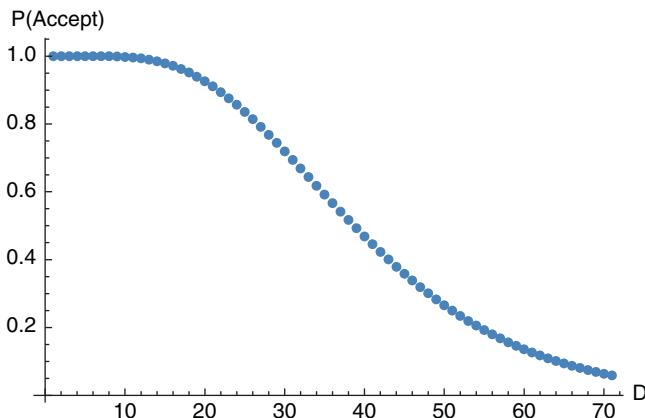
Now we generate a table of values of the function and then plot these values.

```
In[72]= ListPlot[Table[hyperfcn, {x, 0, 24}], Frame -> True, FrameLabel -> {"x", "Probability"}, PlotRange -> {0, 0.155} PlotLabel -> "Hypergeometric Distribution, N=1000, n=30, D=400", AxesOrigin -> {0,0}, LabelStyle -> {FontFamily -> "Helvetica-Bold"}]
```



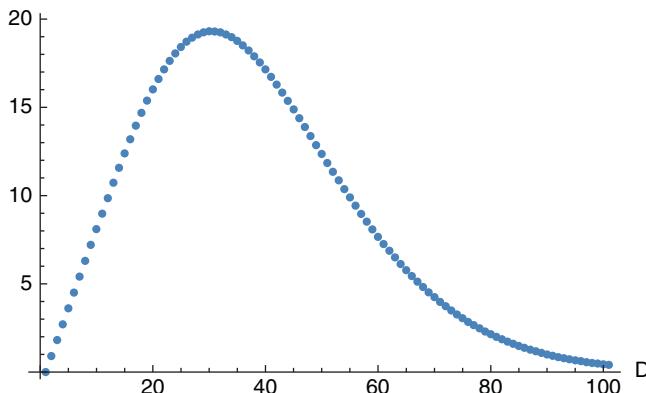
Section 2.11 Acceptance Sampling

We discussed a double sampling plan in this section. Here are the graphs showing the probability the lot is accepted (Figure 2.23) and the average outgoing quality (Figure 2.24).



```
In[73]= probacc = Sum[Binomial[40, x] * Binomial[460, 50 - x]/Binomial[500, 50],
{ {x, 0, 3} } + Sum[(Binomial[40, x] * Binomial[460, 50 - x]) * Binomial[410
+ x, 30]/(Binomial[500, 50] * Binomial[450, 30]), { {x, 4, 5} }]//N
Out[73]= 0.445334
In[74]= pacc[y_]:= 
Sum [Binomial [y, x] * Binomial [500 - y, 50 - x] / Binomial [500, 50],
{ {x, 0, 3} } + Sum [(Binomial [y, x] * Binomial [500 - y, 50 - x]) * Binomial [450 - y + x, 30] / (Binomial [500, 50] * Binomial [450, 30]),
{ {x, 4, 5} }]
In[75]= pacc[40]//N
Out[75]= 0.445334
In[76]= ListPlot[Table[pacc[y], {y, 0, 70}],
LabelStyle -> (FontFamily -> "Helvetica-Bold"), AxesLabel -> {"D",
"P(Accept)"}]
In[77]= aoq[y_]:= 
Sum[(y - x) * Binomial[y, x] * Binomial[500 - y, 50 - x]/Binomial[500,
50], { {x, 0, 3} } + Sum[(y - x) * (Binomial[y, x] * Binomial[500 - y,
50 - x]) *
Binomial[450 - y + x, 30]/(Binomial[500, 50] * Binomial[450, 30]),
{ {x, 4, 5} }]
In[78]= ListPlot[Table[aoq[y], {y, 0, 100}], LabelStyle -> (FontFamily ->
"Helvetica-Bold") , AxesLabel -> {"D" , "AOQ"}]
```

Out[78]= AOQ



In[79]= Max[Table [aoq[y], {y, 0, 100}]] //N

Out[79]= 19.3036

In[80]= Table[{y, aoq[y]}, {y, 27, 31}] //N

Out[80]= {{27., 19.1317}, {28., 19.2475}, {29., 19.3036}, {30., 19.3018}, {31., 19.2445}}

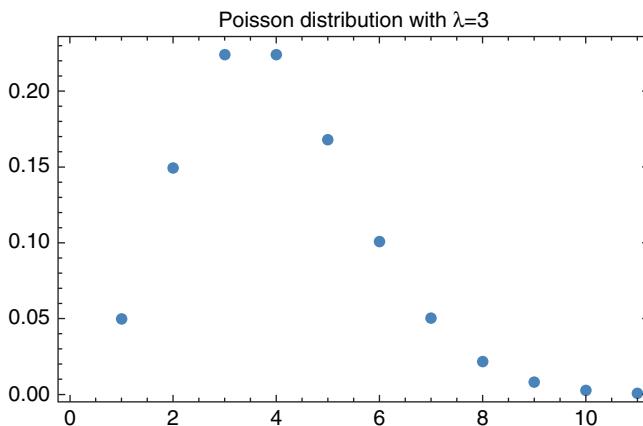
Showing the maximum to be at $D = 29$.

Section 2.13 Poisson Random Variable

Here is the way Figure 2.26 was drawn.

```
In[81]= ListPlot[Table[PDF[PoissonDistribution[3], x], {x, 0, 10}], PlotLabel -> "Poisson Distribution with  $\lambda=3$ ", Frame -> True, LabelStyle -> {FontFamily -> "Helvetica-Bold"} ]
```

Out[81]=



CHAPTER THREE

The standard probability density functions, such as the uniform, exponential, normal, chi-squared, gamma, Weibull distributions (as well as many, many others) are included in Mathematica. Some examples of their use is given here.

Section 3.1 Continuous Random Variables

Means and variances for continuous distributions are found directly by integration. Here we use the probability density function $3x^2$ for x in the interval $(0,1)$.

Example 3.1.1

```
In[82]= mean = Integrate[x * 3x^2, {x, 0, 1}]
Out[82]= 3/4
In[83]= variance = Integrate[(x^2) * 3x^2, {x, 0, 1}] - mean^2
Out[83]= 3/80
```

Section 3.3 Exponential Distribution

The probability density function for the exponential distribution can be found as follows. The value for λ must be specified.

```
In[84]= expdist = PDF[ExponentialDistribution[\lambda], x]
Out[84]=  $\begin{cases} e^{-x\lambda} & x > 0 \\ 0 & \text{True} \end{cases}$ 
```

Note that the mean is then $1/\lambda$. Probabilities are then found by integration.

Example 3.3.2

The probability that X exceeds 2 is given by the following.

```
In[85]= Integrate[PDF[ExponentialDistribution[1], x], {x, 2, Infinity}]
Out[85]= 1/e^2
```

Section 3.5 Normal Distribution

The mean and standard deviation must be specified to determine the normal distribution.

```
In[86]= normdisdt = PDF[NormalDistribution[a, b], x]
Out[86]= 
$$\frac{e^{-\frac{(-a+x)^2}{2b^2}}}{b\sqrt{2\pi}}$$

```

Example 3.5.1

Here we seek the conditional probability a score exceeds 600, given that it exceeds 500. There is no need to transform the scores or to consult a table.

```
In[87]= satdist = PDF[NormalDistribution[500, 100], x];
In[88]= Integrate[satdist, {x, 600, Infinity}]/Integrate[satdist, {x, 500, Infinity}]//N
Out[88]= 0.317311
```

Section 3.6 Normal Approximation to the Binomial

Direct comparisons with exact binomial probabilities and normal approximations are easily done. We use a binomial distribution with $n = 500$ and $p = 0.10$. To compare the exact value of $P(X = 53)$ with the normal curve we calculate as follows:

```
In[89]= PDF[BinomialDistribution[500, 0.10], 53]
Out[89]= 0.0524484
In[90]= Integrate[
  PDF[NormalDistribution[500 * 0.10, Sqrt[500 * 0.10 * 0.90]], x], {x, 52.5,
  53.5}]
Out[90]= 0.0537716
```

Section 3.7 Gamma and Chi-Squared Distributions

We show here the syntax for calling gamma and chi-squared distributions and we show two graphs. The number of degrees of freedom must be specified for the chi-squared distribution while the gamma distribution is characterized by two parameters, r and λ .

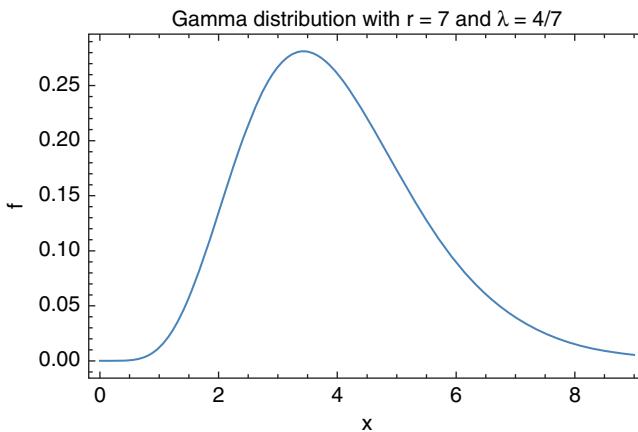
```
In[91]= gamdist = PDF[GammaDistribution[r, 1/\lambda], x]
Out[91]= 
$$\begin{cases} \frac{e^{-x/\lambda} x^{-1+r} \left(\frac{1}{\lambda}\right)^{-r}}{\text{Gamma}[r]} & x > 0 \\ 0 & \text{True} \end{cases}$$

```

Here r and λ are the parameters used. In the following graph, $r = 7$ and $\lambda = 4/7$.

```
In[92]= Plot[PDF[GammaDistribution[7, 4/7], x], {x, 0, 9}, FrameLabel -> {"x",
  "f"}, Frame -> True, PlotLabel -> "Gamma Distribution with r = 7 and
  \lambda = 4/7", LabelStyle -> (FontFamily -> "Helvetica-Bold")]
```

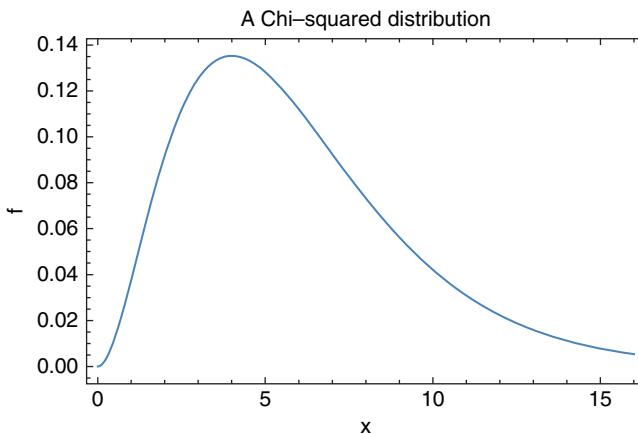
Out[92]=



Here is a chi-squared distribution as shown in Figure 3.17.

```
In[93]= Plot[PDF[ChiSquareDistribution[6], x], {x, 0, 16}, FrameLabel -> {"x", "f"}, Frame -> True, PlotLabel -> "A Chi-Squared Distribution", LabelStyle -> (FontFamily -> "Helvetica-Bold")]
```

Out[93]=



Section 3.7 Weibull Distribution

Parameters α and β must be specified. Mathematica returns the probability density function.

```
In[94]= weibdist = PDF[WeibullDistribution[a, b], x]
```

$$\text{Out[94]} = \begin{cases} \frac{a e^{-\left(\frac{x}{b}\right)^a} \left(\frac{x}{b}\right)^{-1+a}}{b} & x > 0 \\ 0 & \text{True} \end{cases}$$

Here are some graphs of Weibull Distributions:

```
In[95]= wplot = Plot [Evaluate@Table[{PDF[WeibullDistribution[2 , 3], x],
PDF[WeibullDistribution[1 / 4, 1 / 2], x], PDF[WeibullDistribution[3, 6],
x]}], {x, 0, 12}, Frame → True, FrameStyle → (FontFamily →
"Helvetica-Bold")];
In[96]= horiz1 = Graphics[Text["<--- {\alpha=1/4,\beta=1/2}", {1.7412,0.2974}]];
In[97]= horiz2 = Graphics[Text["<--- {\alpha=2,\beta=3}", {4.311,0.2418}]];
In[98]= horiz3 = Graphics [Text ["<--- {\alpha=3, \beta=6}", {8.166,0.1512}]];
In[99]= Show[wplot, horiz1, horiz2, horiz3]

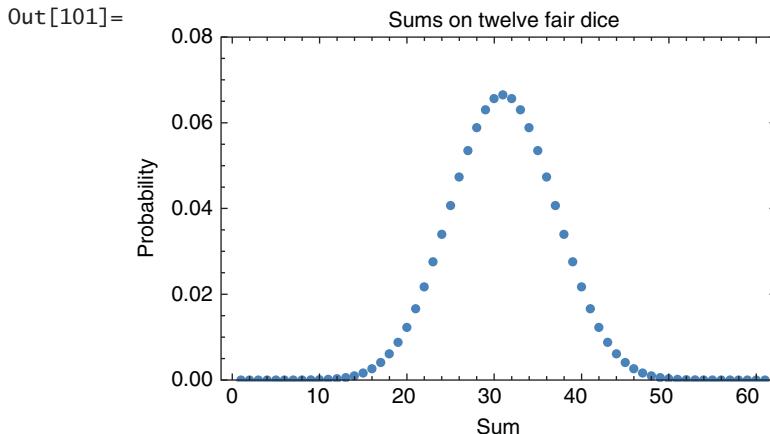
Out[99]= Show[wplot, horiz1, horiz2, horiz3]
```

CHAPTER FOUR

Section 4.5 Generating Functions

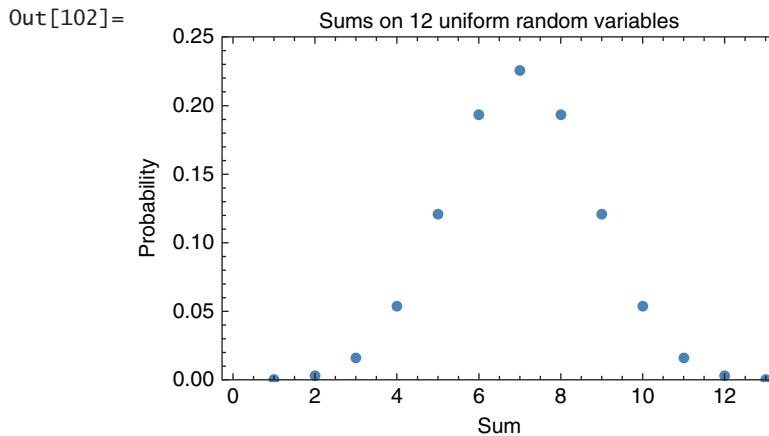
Figure 4.5 shows the probability distribution of the sum when 12 fair dice are thrown. The graph was done using $g[t]$, a probability generating function., which we used in Chapter Two.

```
In[100]= g[t_] := Sum[(1 / 6) t^i, {i, 1, 6}]
In[101]= ListPlot[Drop[CoefficientList[Expand[g[t]^12], t], 12], Frame → True, Axe-
sOrigin → {1, 0}, FrameLabel → {"Sum", "Probability"}, PlotLabel → "Sums
on Twelve Fair Dice", PlotRange → {0, 0.08}, LabelStyle → (FontFamily →
"Helvetica-Bold")]
```



We also show a graph of the sums when 12 uniform random variables are added. We first need the generating function.

```
In[102]= ListPlot[CoefficientList[((1 + t) / 2)^12, t], Frame -> True, LabelStyle -> {FontFamily -> "Helvetica-Bold"}, FrameLabel -> {"Sum", "Probability"}, PlotRange -> {0, 0.25}, PlotLabel -> "Sums on 12 Uniform Random Variables"]
```



Section 4.8 Moment Generating Functions

Moment Generating Functions can be found. We give an example of each.

The probability generating function for the binomial distribution is

```
In[103]= FunctionExpand[Sum[(t^Ax) * Binomial[n, x] * (p^x) * (1 - p)^(n - x), {x, 0, n}]]
```

Out[103]= $(1+p)(-1+t)^n$

This can also be written as

```
In[104]= gen[t_] := (q + p * t)^n
```

and we find the expected value as the first derivative of the probability generating function when $t = 1$:

```
In[105]= gen'[t] /. t -> 1
```

Out[105]= $np(p + q)^{-1+n}$

The moment generating function for the exponential distribution is

```
In[106]= expmomefcn = Integrate[Exp[tx] * Exp[-x], {x, 0, Infinity}]
```

```
Out[106]= Conditional Expression  $\left[ \frac{1}{1-t}, \operatorname{Re}[t] < 1 \right]$ 
```

and this can be expressed as a power series:

```
In[107]= Series[(1 - t)^{-1}, {t, 0, 10}]
```

```
Out[107]= 1 + t + t^2 + t^3 + t^4 + t^5 + t^6 + t^7 + t^8 + t^9 + t^{10} + 0[t]^11
```

Section 4.10 Sums of Random Variables II

Example 4.10.2

Sums of Exponential Random Variables

Here a limit must be calculated in order to show that the sum of exponential random variables becomes normal. We show the limit of $\log[M[Z:t]]$ as n becomes infinite.

```
In[108]= lmz = (-t * Sqrt[n]) - n * Log[1 - t /Sqrt[n]]
```

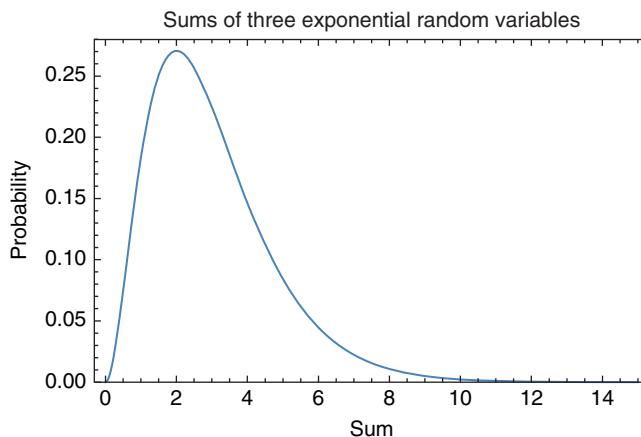
$$\text{Out}[108] = -\sqrt{n} \left(t - n \log \left[1 - \frac{t}{\sqrt{n}} \right] \right)$$

```
In[109]= Limit[lmz, n → Infinity]
```

$$\text{Out}[109] = \frac{t^2}{2}$$

Here is a graph of the sum of three independent exponential random variables. The probability density function is $f[x] = (1/2)x^2e^{-x}, x >= 0$:

```
In[110]= f2[x_] := (1/2) x^2 Exp[-x]
In[111]= Plot[f2[x], {x, 0, 15}, Frame → True, LabelStyle → {FontFamily → "Helvetica-Bold"}, FrameLabel → {"Sum", "Probability"}, PlotRange → {0, 0.28}, PlotLabel → "Sums on Three Exponential Random Variables"]
```



Section 4.11 The Central Limit Theorem

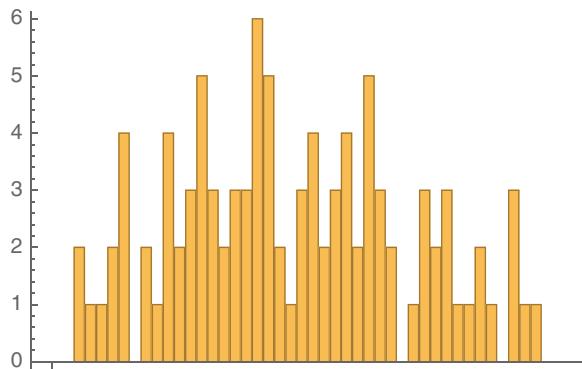
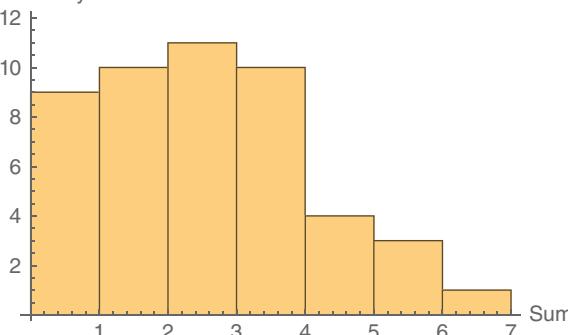
We discussed sampling from the uniform distribution in this appendix in Chapter Two. Here we want samples of size 3 from a uniform distribution on the integers from 1 to 20.

We show how to draw 100 such samples and compute the sample mean of each one. We show a histogram of the sample means as an illustration of the central limit theorem.

```
In[112]= Table[RandomInteger[{1, 20}, 3], {i, 1, 100}];
In[113]= totals = Apply[Plus, %, 1]
Out[113]= {44, 46, 33, 32, 52, 36, 39, 28, 31, 48, 25, 24, 38, 34, 28, 27, 47, 13,
36, 25, 41, 39, 27, 33, 52, 25, 44, 39, 31, 17, 41, 13, 26, 19, 16, 34,
50, 36, 29, 34, 14, 40, 17, 46, 24, 22, 45, 17, 39, 10, 24, 33, 30, 37,
21, 30, 46, 43, 35, 49, 21, 29, 15, 28, 54, 30, 23, 19, 34, 37, 29, 21,
24, 29, 37, 30, 21, 17, 26, 45, 49, 29, 40, 24, 37, 20, 29, 23, 22, 53,
38, 16, 44, 52, 27, 40, 23, 35, 39, 30}
In[114]= Length[totals]
Out[114]= 100
```

```
In[115]= Max[means]
Out[115]= 18
In[116]= Min[means]
Out[116]=  $\frac{10}{3}$ 
In[117]= freqs = BinCounts[means, {11/3, 59/3, 1/3}]
Out[117]= {0, 0, 2, 1, 1, 2, 4, 0, 2, 1, 4, 2, 3, 5, 3, 2, 3, 3, 6, 5, 2, 1, 3, 4,
2, 3, 4, 2, 5, 3, 2, 0, 1, 3, 2, 3, 1, 1, 2, 1, 0, 3, 1, 1, 0, 0, 0, 0}
In[118]= Length[freqs]
Out[118]= 48
In[119]= Histogram[freqs, 6, AxesLabel \rightarrow {"Sum", "Probability"}, LabelStyle \rightarrow
(FontFamily \rightarrow "Helvetica-Bold")]
```

In[120]= Probability



Section 4.12 Weak Law of Large Numbers

The following commands show the construction of Figure 4.10

```
In[121]= values = Table[RandomVariate[BinomialDistribution[100, 0.38]], {100}]
Out[121]= {32, 40, 42, 38, 35, 31, 41, 37, 36, 43, 33, 43, 37, 39, 27, 33, 45, 28,
38, 29, 38, 37, 31, 40, 44, 32, 41, 43, 44, 33, 45, 38, 43, 34, 35, 37,
36, 30, 39, 40, 39, 39, 37, 29, 40, 30, 30, 42, 37, 46, 37, 39, 38, 48,
38, 48, 33, 42, 44, 39, 21, 31, 37, 42, 21, 43, 36, 38,
46, 34, 41, 43, 41, 44, 47, 42, 33, 37, 51, 41, 42, 42, 36, 35, 40,
32, 37, 41, 34, 34, 36, 45, 37, 40, 37, 38, 32, 31, 42, 48}
```

Now we show how the partial sums of these values produce a mean value that approaches 0.38.

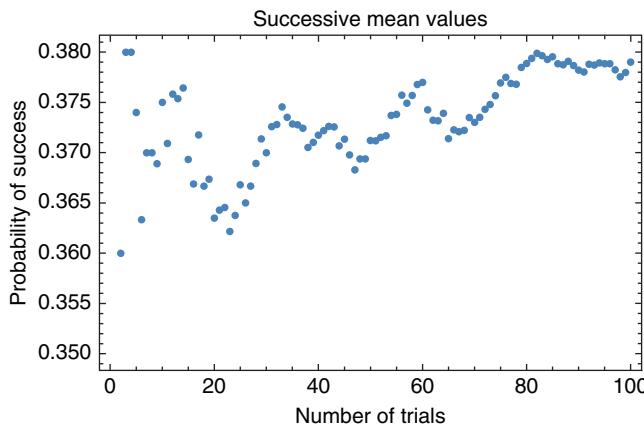
```
In[122]= newvalues = Table[Sum[values[[i]], {i, 1, j}]/(100 * j), {j, 1, 100}]

Out[122]= { $\frac{8}{25}, \frac{9}{25}, \frac{19}{50}, \frac{19}{50}, \frac{187}{500}, \frac{109}{300}, \frac{37}{100}, \frac{37}{100}, \frac{83}{225}, \frac{3}{8}, \frac{102}{275}, \frac{451}{1200}, \frac{122}{325}, \frac{527}{1400}, \frac{277}{750},$ 
 $\frac{587}{1600}, \frac{158}{425}, \frac{11}{30}, \frac{349}{950}, \frac{727}{2000}, \frac{51}{140}, \frac{401}{1100}, \frac{833}{2300}, \frac{291}{800}, \frac{917}{2500}, \frac{73}{200}, \frac{11}{30}, \frac{1033}{2800},$ 
 $\frac{1077}{2900}, \frac{37}{100}, \frac{231}{620}, \frac{1193}{3200}, \frac{103}{275}, \frac{127}{340}, \frac{261}{700}, \frac{671}{1800}, \frac{689}{1850}, \frac{176}{475}, \frac{1447}{3900}, \frac{1487}{4000},$ 
 $\frac{763}{2050}, \frac{313}{840}, \frac{801}{2150}, \frac{1631}{4400}, \frac{557}{1500}, \frac{1701}{4600}, \frac{1731}{4700}, \frac{591}{1600}, \frac{181}{490}, \frac{232}{625}, \frac{631}{1700}, \frac{483}{1300},$ 
 $\frac{197}{530}, \frac{1009}{2700}, \frac{514}{1375}, \frac{263}{700}, \frac{2137}{5700}, \frac{2179}{5800}, \frac{2223}{5900}, \frac{377}{1000}, \frac{2283}{6100}, \frac{1157}{3100}, \frac{2351}{6300}, \frac{2393}{6400},$ 
 $\frac{1207}{3250}, \frac{819}{2200}, \frac{2493}{6700}, \frac{2531}{6800}, \frac{859}{2300}, \frac{373}{1000}, \frac{663}{1775}, \frac{539}{1440}, \frac{684}{1825}, \frac{139}{370}, \frac{2827}{7500}, \frac{151}{400},$ 
 $\frac{1451}{3850}, \frac{2939}{7800}, \frac{299}{790}, \frac{3031}{8000}, \frac{3073}{8100}, \frac{623}{1640}, \frac{3151}{8300}, \frac{531}{1400}, \frac{1613}{4250}, \frac{1629}{4300}, \frac{659}{1740}, \frac{417}{1100},$ 
 $\frac{337}{890}, \frac{851}{2250}, \frac{172}{455}, \frac{697}{1840}, \frac{587}{1550}, \frac{1781}{4700}, \frac{3599}{9500}, \frac{3637}{9600}, \frac{3669}{9700}, \frac{37}{98}, \frac{1871}{4950}, \frac{379}{1000} \}$ 
```

Now $379/1000$ is very close to 0.38 and the graph of these means is interesting:

```
In[123]= ListPlot[newvalues, AxesOrigin -> {0, 0.35}, Frame -> True, FrameLabel -> {"Number of Trials", "Probability of Success"}, PlotLabel -> "Successive Mean Values", LabelStyle -> (FontFamily -> "Helvetica-Bold")]
```

Out[123]=

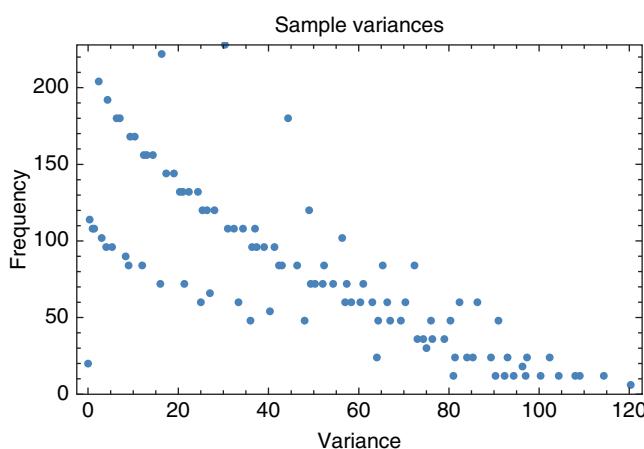


Section 4.14 Distribution of the Sample Variance

Figure 4.11 shows the variances of samples of size 3 drawn without replacement from a uniform distribution on the set $\{1, 2, \dots, 20\}$. The graph is drawn as follows.

```
In[124]= vardata = Flatten[Table[{i, j, k}, {i, 1, 20}, {j, 1, 20}, {k, 1, 20}], 2];
In[125]= variances = Table[Variance[vardata[[i]]], {i, 1, Length[vardata]}];
In[126]= sortvariances = Sort[variances];
In[127]= valuesvars = Union[sortvariances];
In[128]= freqvars = Table[Count[sortvariances, valuesvars[[i]]], {i, 1,
In[129]= Length[valuesvars]}];
In[130]= ticks = {{20, 20}, {40, 40}, {60, 60}, {80, 80}, {100, 100}};
Out[130]= {{20, 20}, {40, 40}, {60, 60}, {80, 80}, {100, 100}}
In[131]= ListPlot[Transpose[{valuesvars, freqvars}], Frame -> True, FrameLabel ->
  {"Variance", "Frequency"}, PlotRange -> {0, Max[freqvars]}, PlotLabel ->
  "Sample Variances", Ticks -> {ticks, Automatic}, LabelStyle -> (FontFamily -> "Helvetica-Bold")]
```

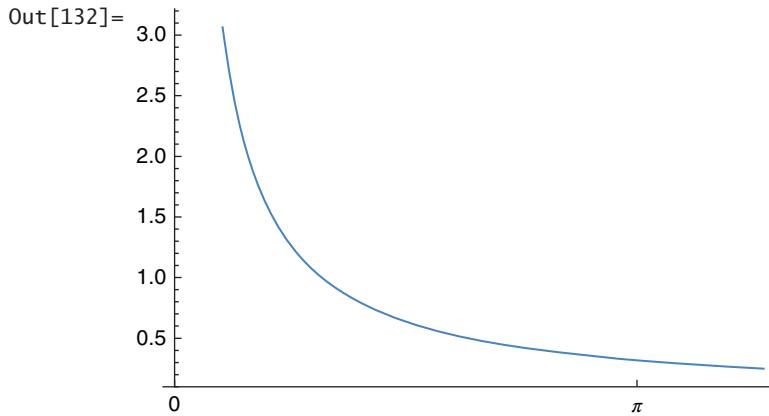
Out[131]=



416 Appendix A Use of Mathematica in Probability and Statistics

A probability density function suggested by Figure 4.11

```
In[132]= Plot[1 / x, {x, 0, 4}, Ticks → {{0, Pi}, Automatic}]
```



Section 4.14 Hypothesis Tests and Confidence Intervals for a Single Mean

Example 4.14.1

If the mean and standard deviation are known, which is the case in this example, a confidence can be found using the following.

```
In[133]= Needs["HypothesisTesting`"]
```

A 95% confidence interval for the population mean:

```
In[134]= NormalCI[2200, 4591/5]/N  
Out[134]= {400.361, 3999.64}
```

Note that we used the standard deviation of the mean based on a sample of size 25. The default confidence level is 95 %, but this can be changed to any probability.

```
In[135]= NormalCI[2200, 4591/5, ConfidenceLevel → 0.8436]/N  
Out[135]= {898.65, 3501.35}
```

Student *t* Distribution

The Student *t* distribution is included in Mathematica. Here is the Student *t* distribution with 5 degrees of freedom

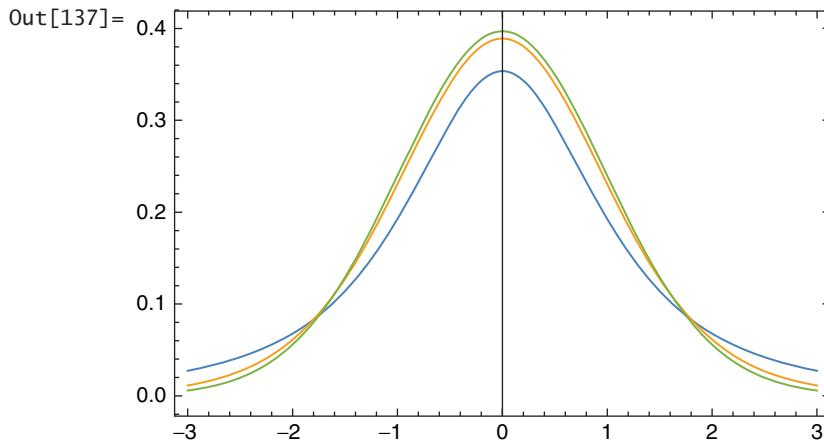
```
In[136]= tdist = PDF[StudentTDistribution[5], x]
```

```
Out[136]= 
$$\frac{200\sqrt{5}}{3\pi(5+x^2)^3}$$

```

Note that Mathematica returns the probability density function. Now here are some graphs of the Student t distribution.

```
In[137]= Plot[Evaluate@Table[{PDF[StudentTDistribution[2], x],
PDF[StudentTDistribution[10], x], PDF[StudentTDistribution[50], x]}],
{x, -3, 3}, Frame → True, FrameStyle → (FontFamily → "Helvetica-Bold")]
```



The Student t distributions in Figure 4.15 are for 5, 20, and 40 degrees of freedom. Confidence Intervals can be calculated.

Example 4.1.4

Suppose the data are as follows:

```
In[138]= tdata = {8, 7, 3, 5, 9, 4, 10, 2, 6, 7}
Out[138]= {8, 7, 3, 5, 9, 4, 10, 2, 6, 7}
In[139]= Variance[tdata]
Out[139]=  $\frac{203}{30}$ 
In[140]= Mean[tdata]
Out[140]=  $\frac{61}{10}$ 
In[141]= StudentTCI[61/10, Sqrt[(203/30)/10], 9]//N
Out[141]= {4.23916, 7.96084}
```

Section 4.15 Tests on Two Samples

Mathematica can do hypothesis testing and find confidence intervals directly from sample data. As an example, suppose the following are samples from two normal populations.

```
In[142]= data1 = {7, 4, 5, 6, 7, 3, 4, 2, 5, 8, 9, 12}
Out[142]= {7, 4, 5, 6, 7, 3, 4, 2, 5, 8, 9, 12}
In[143]= data2 = {8, 9, 5, 6, 7, 8, 3, 4, 5, 6, 3, 4, 5, 7, 6, 3, 4, 9}
{8, 9, 5, 6, 7, 8, 3, 4, 5, 6, 3, 4, 5, 7, 6, 3, 4, 9}
```

Note that the sample sizes are not the same. A 95 % confidence interval for the difference between the means can be found. We can assume the population variances are equal or not. The default assumption is that the variances are unequal.

```
In[144]= MeanDifferenceCI[data1, data2, EqualVariances → True]
Out[144]= {-1.45699, 2.12366}
```

Now assume the variances are not equal.

```
In[145]= MeanDifferenceCI[data1, data2]
Out[145]= {-1.62744, 2.2941}
```

These give quite different confidence intervals.

We can also find a confidence interval for the ratio of the variances. We must specify the sample ratio of the variances and then the number of degrees of freedom for both the numerator and the denominator.

```
In[146]= FRatioCI[Variance[data1]/Variance[data2], 11, 17]
Out[146]= {0.681112, 6.41411}
```

Since 1 is in this interval, we can presume the true variances are equal.

p values are given for hypothesis tests. We must specify the data and the difference between the means we wish to test. We test whether or not the means differ by 1. Again, the true variances are presumed to be unequal.

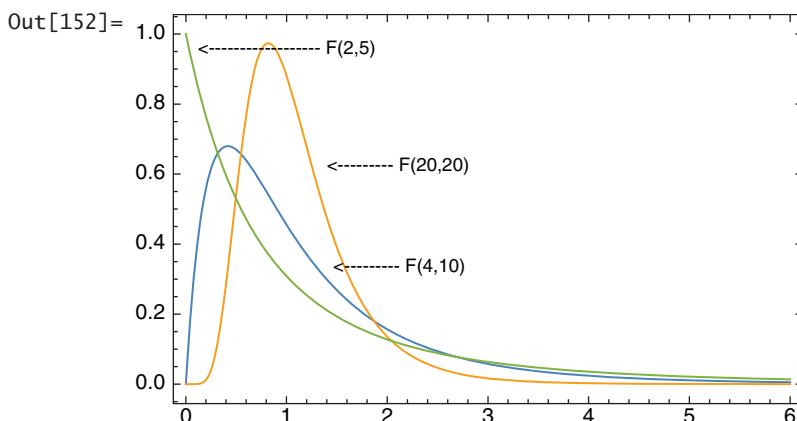
```
In[147]= LocationTest[{data1, data2}, 1]
Out[147]= 0.451979
```

LocationTest returns a *p* value of the null hypothesis that the true means are equal against the alternative that the true means are unequal. A small *p* value would indicate that it is unlikely that the null hypothesis is true, so here we would undoubtedly conclude that the true means are equal and do not differ by 1.

F Distribution

Here are graphs of some *F* distributions:

```
In[148]= fplots = Plot[Evaluate]
Table[{PDF[FRatioDistribution[4, 10], x], PDF[FRatioDistribution[20, 20], x], PDF[FRatioDistribution[2, 5], x]}, {x, 0, 6}, Frame → True];
In[149]= arrow1 = Graphics[Text["<-----F(2,5)", {0.9, 0.9579}]];
In[150]= arrow2 = Graphics[Text["<----F(4,10)", {2.0213, 0.3355}]];
In[151]= arrow3 = Graphics[Text["<---F(20,20)", {1.997, 0.6252}]];
In[152]= Show[fplots, arrow1, arrow2, arrow3]
```

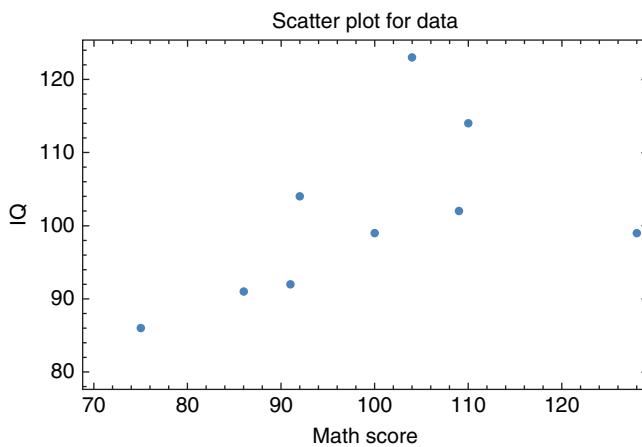


Section 4.16 Least Squares Linear Regression

We show the construction of Figure 4.18.

```
In[153]= rdata = {{92, 104}, {86, 91}, {104, 123},
  {{92, 104}, {86, 91}, {104, 123}, {109, 102},
  {75, 86}, {100, 99}, {91, 92}, {110, 114}, {128, 99}}
In[154]= scatter = ListPlot[rdata, Frame -> True, AxesOrigin -> {70, 80},
  PlotStyle -> {PointSize[0.015]}, PlotLabel -> "Scatter Plot for Data",
  FrameLabel -> {"Math Score", "IQ"}, LabelStyle -> (FontFamily ->
  "Helvetica-Bold")]
```

Out[154]=



To fit a straight line to the data we use the `LinearModelFit` command. The x 's indicate we want an intercept. Mathematica offers a wide range of functions that can be specified in addition to this simple linear function.

```
In[155]= lfit = LinearModelFit[rdata, x, x]
Out[155]= 
$$\text{FittedModel}[63.0792 + 0.382444 x]$$

In[156]= lfit["ANOVATable"]
```

	DF	SS	MS	F-Statistic	P-Value
x	1	284.368	284.368	2.5117	0.157022
Error	7	792.521	113.217		
Total	8	1076.89			

```
In[157]= lfit["ParameterTable"]
```

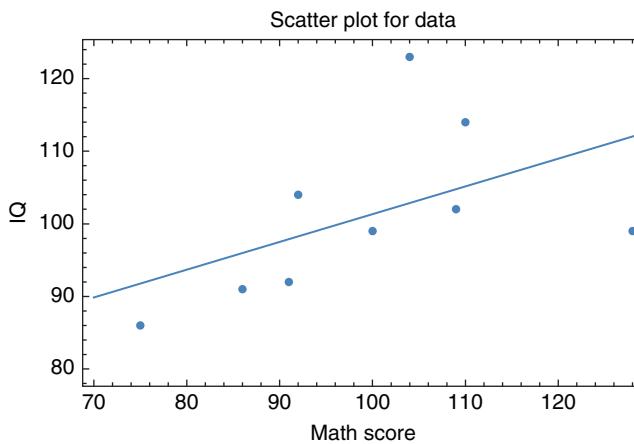
	Estimate	Standard Error	t-Statistic	P-Value
1	63.0792	24.2581	2.60034	0.0354077
x	0.382444	0.241314	1.58484	0.157022

There are many other options available in addition to the analysis of variance and the parameter table.

It is always useful to see the fit visually.

```
In[158]= t[x_] := 63.0792 + 0.38244 x
In[159]= stline = Plot[t[x], {x, 70, 130}];
In[160]= Show [scatter, stline]
```

Out[160]=

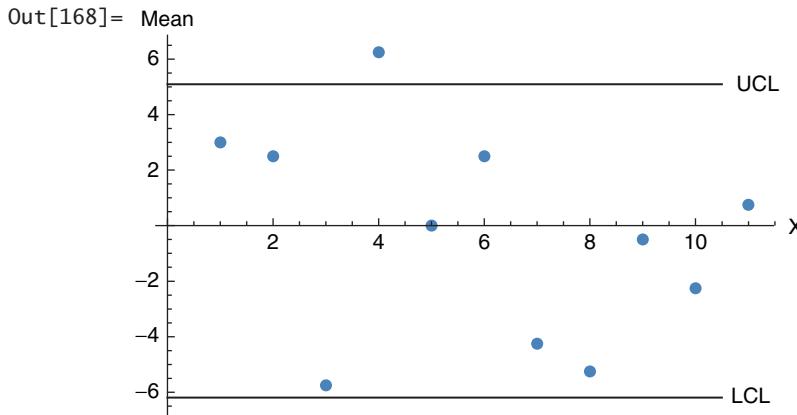


Section 4.17 Control Chart for X Bar

Example 4.17.1

We show a plot of a quality control chart as in Figure 4.20.

```
In[161]= controldata = {{3, -1, 6, 4}, {9, 0, 3, -2}, {-12, 4, -9, -6},
{11, 9, 4, 1}, {-1, -2, 4, -1}, {8, 1, -2, 3}, {-1, -4, -9, -3},
{-8, -3, -6, -4}, {1, -2, -2, 1}, {-2, -2, -3, -2}, {0, 4, 1, -2}},
{{3, -1, 6, 4}, {9, 0, 3, -2}, {-12, 4, -9, -6}, {11, 9, 4, 1},
{-1, -2, 4, -1}, {8, 1, -2, 3}, {-1, -4, -9, -3},
{-8, -3, -6, -4}, {1, -2, -2, 1}, {-2, -2, -3, -2}, {0, 4, 1, -2}}
In[162]= Length[controldata]
Out[162]= 11
In[163]= means = Table[Mean[controldata[[i]]], {i, 1, Length[controldata]}]
Out[163]= {3,  $\frac{5}{2}$ , - $\frac{23}{4}$ ,  $\frac{25}{4}$ , 0,  $\frac{5}{2}$ , - $\frac{17}{4}$ , - $\frac{21}{4}$ , - $\frac{1}{2}$ , - $\frac{9}{4}$ ,  $\frac{3}{4}$ }
In[164]= ucl = Graphics[Line[{{0, 5.0974}, {10.5, 5.0979}}]];
In[165]= lcl = Graphics[Line[{{0, -6.188}, {10.5, -6.188}}]];
In[166]= ltext = Graphics[Text["LCL", {11.25, -6.2}]];
In[167]= utext = Graphics[Text["UCL", {11.25, 5.0979}]];
In[168]= Show [ListPlot[means, AxesOrigin -> {0, 0}],
ucl, lcl, ltext, utext, AxesLabel -> {"X", "Mean"}, LabelStyle -> {FontFamily -> "Helvetica-Bold"}, PlotRange -> {-6.2, 6.2}]
```



CHAPTER FIVE

Section 5.2 Joint and Marginal Distributions

Example 5.2.1

Probabilities for the joint distribution of X and Y are found by the following :

```
In[169]= jprobs =
Table[Binomial [5, x] * Binomial [5 - x, y] * (1 / 2) ^ (10 - x), {x, 0, 5},
{y, 0, 5 - x}]
```

```
Out[169]= { { 1 / 1024, 5 / 1024, 5 / 512, 5 / 512, 5 / 1024, 1 / 1024 }, { 5 / 512, 5 / 128, 15 / 256, 5 / 128, 5 / 512 },
{ 5 / 128, 15 / 128, 15 / 128, 5 / 128 }, { 5 / 64, 5 / 32, 5 / 64 }, { 5 / 64, 5 / 64 }, { 1 / 32 } }
```

The marginal distribution for X can be found as follows:

```
In[170]= Apply[Plus, jprobs, 1]
```

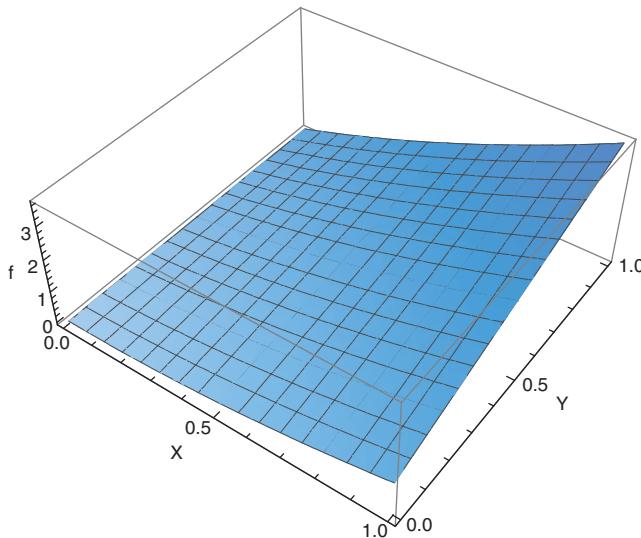
```
Out[170]= { 1 / 32, 5 / 32, 5 / 16, 5 / 16, 5 / 32, 1 / 32 }
```

Example 5.2.2

The joint probability distribution function is plotted as follows:

```
In[171]= f[x_,y_] := x ^ 2 + (8 / 3) * x * y
In[172]= Plot3D[f[x, y], {x, 0, 1}, {y, 0, 1}, Viewpoint -> {1.3, -1.8, 2}, AxesLabel -> {"X", "Y", "f"}, LabelStyle -> {FontFamily -> "Helvetica-Bold"}]
```

Out[172]=



Marginal distributions are easily found.

```
In[173]:= mf[x_] := Integrate[f[x, y], {y, 0, 1}]
In[174]:= mf[x]
```

$$\text{Out}[174] = \frac{4x}{3} + x^2$$

```
In[175]:= mg[y_] := Integrate[f[x, y], {x, 0, 1}]
In[176]:= mg[y]
```

$$\text{Out}[176] = \frac{1}{3} + \frac{4y}{3}$$

Probabilities are volumes.

```
In[177]:= Integrate[f[x, y], {x, 1/2, 1}, {y, 0, 2/3}]
Out[177] = \frac{5}{12}
```

Section 5.3 Conditional Distributions and Densities

Conditional distributions are ratios of the distributions previously calculated.

```
In[178]:= fxgivy[x_] := f[x, y] / mg[y]
In[179]:= fxgivy[x]

Out[179] = \frac{x^2 + \frac{8xy}{3}}{\frac{1}{3} + \frac{4y}{3}}
```

```
In[180]:= fygivx[y_] := f[x, y] / mf[x]
In[181]:= fygivx[y]
```

```
Out[181]= 
$$\frac{x^2 + \frac{8xy}{3}}{\frac{4x}{3} + x^2}$$

```

Section 5.4 Expected Values

Both conditional and unconditional expectations can now be found.

```
In[182]= eygivx = Integrate[y * fygivx[y], {y, 0, 1}]
```

```
Out[182]= 
$$\frac{8x}{9\left(\frac{4x}{3} + x^2\right)} + \frac{x^2}{2\left(\frac{4x}{3} + x^2\right)}$$

```

```
In[183]= FullSimplify[%]
```

```
Out[183]= 
$$\frac{1}{2} + \frac{2}{12 + 9x}$$

```

The integral of $E[Y|X = x] * f[x]$ gives the expected value of Y .

```
In[184]= ey = Integrate [eygivx * f[x], {x, 0, 1}]//N
Out[184]= 0.624065
```

Section 5.6 Bivariate Normal Densities

The bivariate normal density is defined as follows.

```
In[185]= k[r_] := 1 / (2 * Pi * Sqrt[1 - r^2])
In[186]= bivnormf[x_, y_] := k[r] * Exp[(-1 / (2(1 - r^2))) * (x^2 - 2 * r * x * y + y^2)]
```

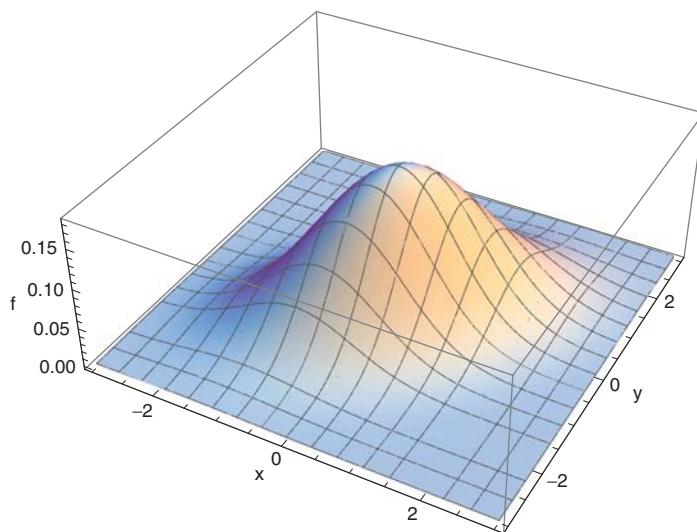
Let the correlation coefficient (r) be $1/2$:

```
In[187]= r=1/2
Out[187]=  $\frac{1}{2}$ 
```

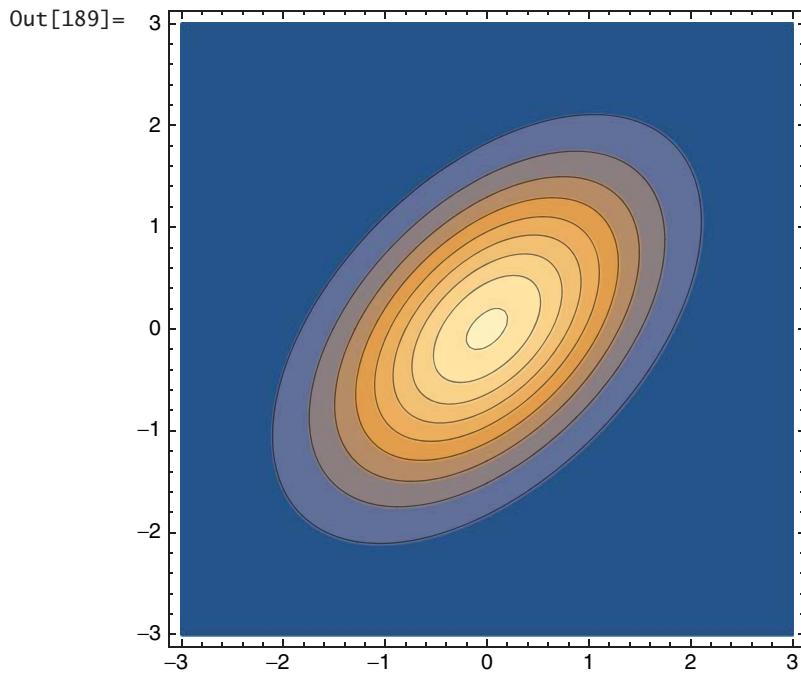
Here is a plot of the bivariate normal density with $\rho = 1/2$:

```
In[188]= Plot3D[bivnormf[x, y], {x, -3, 3}, {y, -3, 3}, AxesLabel -> {"x", "y", "f"}, PlotPoints -> 40, LabelStyle -> {FontFamily -> "Helvetica-Bold"}]
```

Out[188]=



Here is a contour plot.

In[189]= `ContourPlot [bivnormf[x, y], {x, -3, 3}, {y, -3, 3}, PlotPoints → 40]`

CHAPTER SIX

Recursions and Markov Chains are easily done in Mathematica.

Section 6.2 Some Recursions and their Solutions

Example 6.2.1

We saw a recursion for the Fibonacci numbers previously in this Appendix in Chapter 2.
We now define a new recursion and solve it.

```
In[190]= b[n_] := b[n] = p * b[n - 1] + (1 - p) * (1 - b[n - 1])
In[191]= b[1] = 0
Out[191]= 0
In[192]= b[2] = 1 - p
Out[192]= 1 - p
```

Now we compute some values for $b[n]$.

```
In[193]= b[1]
Out[193]= 0
In[194]= b[2]
Out[194]= 1 - p
In[195]= b[3]
Out[195]= 2 (1 - p) p
In[196]= Simplify [Expand [Table [b[n], {n, 2, 7}]]]
```

$$\{1 - p, -2(-1 + p)p, 1 - 3p + 6p^2 - 4p^3, 4p(1 - 3p + 4p^2 - 2p^3), \\ 1 - 5p + 20p^2 - 40p^3 + 40p^4 - 16p^5, 2p(3 - 15p + 40p^2 - 60p^3 + 48p^4 - 16p^5)\}$$

Now we look at a particular value, $b[4]$, and expand it about 1/2:

```
In[197]= exp = Normal [Series[b[4], {p, 1/2, 3}]]
Out[197]=  $\frac{1}{2} - 4 \left( -\frac{1}{2} + p \right)^3$ 
```

This suggests a simple form for $b[n]$

Mathematica will also solve the recursion; we choose the special case where $p = 1/3$.

```
In[198]= RSolve[{c[n] == (1/3) * c[n - 1] + (2/3) * (1 - c[n - 1]),
c[1] == 0, c[2] == 2/3}, c[n], n]
Out[198]= \{ \{ c[n] \rightarrow \frac{1}{2} 3^{-n} (3 (-1)^n + 3^n) \} \}
```

In[199]= **Expand[%]**

$$\text{Out}[199]= \left\{ \left\{ c[n] \rightarrow \frac{1}{2} + \frac{1}{2}(-1)^n 3^{1-n} \right\} \right\}$$

In[200]= **Limit** $\left[\frac{1}{2} + \frac{1}{2}(-1)^n 3^{1-n}, n \rightarrow \text{Infinity} \right]$

$$\text{Out}[200]= \frac{1}{2}$$

This shows that $c[n] \rightarrow 1/2$ as n becomes large.

Section 6.3 Random Walk and Ruin

The construction of Figure 6.6 is shown here. We first define the function $a[g]$ as used in the text. We'll suppose that $p = 0.49$ and $q = 0.51$.

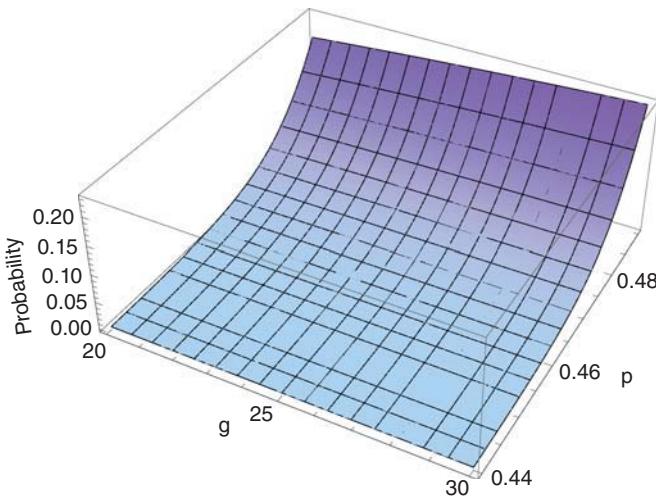
In[201]= **0.51 /0.49 //N**

$$\text{Out}[201]= 1.04082$$

In[202]= **a[g_, h_, p_] := (1 - ((1 - p) /p)^g) / (1 - ((1 - p) /p)^(g + h))**

In[203]= **Plot3D[a[g, 30, p], {g, 20, 30}, {p, 0.44, 0.49}, AxesLabel \rightarrow {"g", "p", "Probability"}, PlotPoints \rightarrow 40, LabelStyle \rightarrow (FontFamily \rightarrow "Helvetica-Bold")]**

Out[203]=



Section 6.5 Markov Chains

Example 6.5.1

The matrix in this example is entered as follows.

In[204]= **m = {{1 / 10, 3 / 10, 3 / 10, 3 / 10}, {3 / 10, 1 / 10, 3 / 10, 3 / 10}, {3 / 10, 3 / 10, 1 / 10, 3 / 10}, {3 / 10, 3 / 10, 3 / 10, 1 / 10}}**

Out[204]= $\left\{ \left\{ \frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{3}{10} \right\}, \left\{ \frac{3}{10}, \frac{1}{10}, \frac{3}{10}, \frac{3}{10} \right\}, \left\{ \frac{3}{10}, \frac{3}{10}, \frac{1}{10}, \frac{3}{10} \right\}, \left\{ \frac{3}{10}, \frac{3}{10}, \frac{3}{10}, \frac{1}{10} \right\} \right\}$

In[205]= **MatrixForm[m]**

Out[205]=
$$\begin{pmatrix} \frac{1}{10} & \frac{3}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} & \frac{3}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{3}{10} & \frac{1}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{3}{10} & \frac{3}{10} & \frac{1}{10} \end{pmatrix}$$

Powers of m can be computed, indicating the limiting form of this matrix

In[206]= **MatrixForm[m.m]**

Out[206]=
$$\begin{pmatrix} \frac{7}{25} & \frac{6}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{6}{25} & \frac{7}{25} & \frac{6}{25} & \frac{6}{25} \\ \frac{6}{25} & \frac{6}{25} & \frac{7}{25} & \frac{6}{25} \\ \frac{6}{25} & \frac{6}{25} & \frac{6}{25} & \frac{7}{25} \end{pmatrix}$$

For higher powers of m it is best to use the MatrixPower command. Here the 20 th power of m is written in the usual form for a matrix.

In[207]= **MatrixForm[MatrixPower[m, 20]]**

Out[207]=
$$\begin{pmatrix} 23 & 841 & 857 & 910 & 157 \\ 95 & 367 & 431 & 640 & 625 \\ 23 & 841 & 857 & 910 & 156 \\ 95 & 367 & 431 & 640 & 625 \\ 23 & 841 & 857 & 910 & 157 \\ 95 & 367 & 431 & 640 & 625 \\ 23 & 841 & 857 & 910 & 157 \\ 95 & 367 & 431 & 640 & 625 \\ 23 & 841 & 857 & 910 & 157 \\ 95 & 367 & 431 & 640 & 625 \end{pmatrix}^{20}$$

Each of the entries is equal and very close to 1/4. The fixed point for m can also be found.

In[208]= **Solve [{a, b, c, d}.m == {a, b, c, d}, {a, b, c, d}]**

Out[208]= $\{\{b \rightarrow a, c \rightarrow a, d \rightarrow a\}\}$

This indicates that all the entries for the fixed vector are equal

Example 6.5.6

In this example we need various matrix products and an inverse. These are done in the following way.

In[209]= **q = {{0, 3/4, 0}, {1/4, 0, 3/4}, {0, 1/4, 0}}**

428 Appendix A Use of Mathematica in Probability and Statistics

```
Out[209]= { {0, 3/4, 0}, {1/4, 0, 3/4}, {0, 1/4, 0} }

In[210]= iminusq = IdentityMatrix[3] - q

Out[210]= { {1, 3/4, 0}, {-1/4, 1, -3/4}, {0, -1/4, 1} }

In[211]= inv = Inverse[iminusq]

Out[211]= { {13/10, 6/5, 9/10}, {2/5, 8/5, 6/5}, {1/10, 2/5, 13/10} }

In[212]= inv.{1, 1, 1}

Out[212]= {17/5, 16/5, 9/5}
```

Appendix B

Answers for Odd-Numbered Exercises

CHAPTER 1

Exercises 1.1

1. A is included in samples ABC, ABD, ABE, ACD, ACE , and ADE and is not in samples BCD, BDE, BCE , and CDE .
3. $S = \{(x, y) | x \neq y, x, y \in \{1, 2, \dots, 9\}\}$, S has $9 \cdot 8 = 72$ points.
5. $S = \{(x_1, x_2, x_3, x_4, x_5) | x_i \in \{G, N\}, i = 1, 2, \dots, 5\}$, or $S = \{G, N\} \times \{G, N\} \times \{G, N\} \times \{G, N\} \times \{G, N\}$ where $A \times B$ denotes the cross product of the sets A and B . S has 32 points.
7. There are 15 sample points:

AAAAA	NHAAAAA
	NANAAA
NAAAAA	NAANAA
ANAAA	NAAAANA
AANAA	ANNAAA
AAANA	ANANAA
	ANAANA
	AANNAA
	AANANA
	AAANNA

9. $S = \{(x, y) | x, y \in \{1, 2, \dots, 12\}, x \neq y\}$. S contains $12 \cdot 11 = 132$ points.

11. There are 15 sample points:

32	62
43	63
42	64
52	65
53	72
54	73
	74
	75
	76

13. *HHHHH**HHHHT**HHHTH**HHTHT**HHHTT**HHTHH***Exercises 1.3**

- 3.** a) 5/16
b) 13/16

- 5.** a) $S = \{(x, y) | x, y \in \{1, 2, 3, 4, 5\}\}$.
b) 10/25
c) 6/25

7. $P(A \cap B) \leq \min\{P(A), P(B)\}$ so the given probabilities are impossible.

9. 0.7

11. Consider $P(B) = P[B \cap (A \cup \bar{A})] = P[(B \cap A) \cup (B \cap \bar{A})]$. Then use the addition law.

13. $P(\text{exactly one of } A, B) = P(A) + P(B) - 2P(A \cap B)$.

Exercises 1.4

- 3.** a) 58/135
b) 15/29

- 5.** a) 4.9%
b) 35/98

7. Use the addition law.

- 9.** a) 1/7
b) 1/10

11. 0.994

- 13.** a) 4
b) $n \geq \frac{\log(1-r)}{\log(1-p)}$

15. 1/4. [Consider $P(1|1 \text{ or even})$.]

- 17.** a) No
b) No
c) 1/6
d) 1/3

19. 2/3

- 21.** a) 17/70
b) 12/17

23. 1/2

- 25.** a) $p_1 = p, p_2 = q + r(p - q),$
 $p_3 = 2r(r - 1)(p - q) + p$ where $q = 1 - p$
 b) No
 c) $r = 1/2$
- 27.** a) $1/7$
 b) $2/5$
- 29.** The estimate of p is $2 \times$ the sample proportion answering “yes” – $1/2$.

Exercises 1.5

5. r	2	3	4	5	6	7	8	9	10	11	12	13
p_r	$\frac{1}{12}$	$\frac{17}{72}$	$\frac{41}{96}$	$\frac{89}{144}$	$\frac{1343}{1728}$	$\frac{3071}{3456}$	$\frac{39547}{41472}$	$\frac{122491}{124416}$	$\frac{495739}{497664}$	$\frac{2984059}{2985984}$	$\frac{35829883}{35831808}$	1

- 7.** a) $3/8$
 b) $4/9$
- 9.** a) $96/15625 = 0.006144$
 b) $198/15625 = 0.002146$

Exercises 1.6

- 1.** a) $1 - (1 - p_A)(1 - p_B)(1 - p_C)$
 b) $p_A p_B + p_A p_C + p_B p_C - 2p_A p_B p_C$
- 3.** $p_A p_C(2 - p_A)(2 - p_C)$

Exercises 1.7

- 1.** a) $4/9$
 a) $1/9$
 b) $2/9$
- 3.** There are more than $26^3 = 17,576$ people in Colorado Springs
- 5.** a) 280
 b) 431
- 7.** 34,650
- 9.** $7/24$
- 11.** a) At least one six in 4 rolls of a fair die.
- 13.** a) 0.4035
 b) $P(\text{at least 1 duplicate}) = \frac{1316998181}{1334062100} = 0.9872091.$
- 15.** $1/70$
- 17.** $2162/54145 = 0.0399298$
- 19.** 201,600
- 21.** 13,860

23. $(N - 2)/N$ **25.** 35

- 27.** **a)** $11/16$
b) $7/8$

Supplementary Exercises for Chapter 1

1. $77/969 = 0.0794634$ **3.** $1/2$ **5.** $3/4$ **7.** 4**9.** $5/33$

- 11.** **a)** $1/2$
b) $5/9$
c) $1/30$

- 13.** **a)** $2/5$
b) $5/6$
c) $2/3$

15. 20% A' 's, 50% B' 's, and 30% C' 's.**17.** $\frac{217006443}{318555566} = 0.68122$

- 19.** **a)** 7^{10}
b) 5^{10}
c) 207,446,400

21. $1 - \sqrt{2}/2$

- 25.** **a)** $S = \{(x, y) | x, y \in \{1, 2, \dots, 6\}, x < y\}.$
b) $3/5$
c) $13/15$

27. $\frac{100}{429}$ **29.**
$$\frac{4\binom{39}{13} - 6\binom{26}{13} + 4}{\binom{52}{13}} = \frac{1621364909}{31750677980} = 0.0510655$$
31. $\frac{2^{50}}{\binom{100}{50}} = 1.11595 \times 10^{-14}$ **33.** $4/5$

- 35.** **a)** 0.30
b) 0.10

- 39.** **a)** $1/208,012$
b) $12/13$

43. a) 0.936

b) 6

45. $\frac{(m^2-m+1)}{m^2(m-1)}$

47. 0.4

49. $1 - \frac{N!}{(N-n)!N^n}$

51. a) 7,694,644,696,200

b) 1,889,912,632,400

c) 70,671,744

d) 211,360,681,548

53. $2/n$

55. $21,349/22,407 = 0.952783$

CHAPTER 2

Exercises 2.2

1. a) $2/5$

b) $2/5$

3. a) $S = \{(x_1, x_2, x_3, x_4) | x_i \in \{H, T\}, i = 1, 2, 3, 4\}$

b) $3/16$

c) $7/15$

5. a) $k = 6/31$

b) $25/31$

c) $F(x) = \begin{cases} 0 & x \leq 1 \\ 6/31 & 1 < x < 2 \\ 18/31 & 2 < x < 3 \\ 27/31 & 3 < x < 4 \\ 1 & x \geq 4 \end{cases}$

Sum	2	3	4	5	6	7	8	9	10	11	12
$441 \times \text{Prob}$	36	60	73	76	70	56	35	20	10	4	1

11. $G(y) = \begin{cases} 0 & y < 2 \\ 1/4 & 2 \leq y < 3 \\ 1/2 & 3 \leq y < 4 \\ 3/4 & 4 \leq y < 5 \\ 1 & y \geq 5 \end{cases}$

13. $f(-1) = 1/3, f(0) = 1/2, f(2) = 1/6$

15. $F(x) = i/n, i \leq x < i + 1, i = 1, 2, \dots, n$

Exercises 2.3

1. $\mu = 13/3, \sigma^2 = 20/9$
3. a) $E(X) = 1/2, \text{Var}(X) = 0.45$
b) \$3,900
5. $E(X) = \$12,900, \text{Var}(X) = \$31,900,000, \sigma = \$1786.06$
7. $E(X) = 3/5, \text{Var}(X) = 72/175$
9. $E(X) = \frac{n+1}{2}, \text{Var}(X) = \frac{n^2-1}{12}$
11. 9/25
13. a) $P(X = x) = \begin{cases} 1/6 & \text{if } x = 1,2,3 \\ 1/8 & \text{if } x = 4,5,6 \\ 1/16 & \text{if } x = 7,8 \end{cases}$
b) \$1.00
15. $E(X)$

Exercises 2.5

1. a) 0.1091
b) 0.999437
3. a) 0.10292
b) 20
c) 0.3445
5. a) 0.854134
b) 0.0591414
7. a) 0.004629630
b) 0.0006766
c) 0.0791125
9. a) X is binomial with $n = 7$ and $p = 0.2$
b) 73/15625
c) 73/2313
11. -\$60,100
13.

N	-3	-2	-1	0	1	2
59049 \times Prob	29,161	10	40	80	80	29,678
15. a) 0.966634
b) 93
17. 5/36
19. 0.0973832
21. a) 0.128506
b) 0.0318415
23. 65/256

Exercises 2.6

- 1.** End points are $4/13$ and $1/2$
- 3.** 10 to 20
- 7.** 0.23598, 0.28551
- 9.** 0.330775, 0.473463
- 11.** $A : 0.102589, 0.151481$
 $J : 0.0275727, 0.0495844$
 No

Exercises 2.7

- 1.** $\alpha = 0.0378016, \beta = 0.0912053$
- 3.** **a)** 0.0866925
b) 0.60801
- 5.** $c = 4$ and $\beta = 0.99992$
- 7.** 16
- 9.** H_a
- 11.** Yes

Exercises 2.8

- 1.** 0.0731714, 0.162123
- 3.** 0.413143, 0.527647
- 5.** Yes
- 7.** 400
- 9.** 0.527975, 0.667947
- 11.** 0.419418, 0.580582

Exercises 2.9

- 1.** 0.0669796
- 3.** $3/4$
- 5.** **a)** 0.0064
b) 0.04096
- 7.** $8/65$
- 9.** 0.123959
- 11.** $11/32$
- 13.** 0.0313811
- 15.** **a)** $1/7$
b) 7

17. a) \$21

b) No

19. r/t **Exercises 2.10****1.** $1/4$ **3.** $677,007/832,370$ **5.** a)

X	3	4	5	6	7	8
$56 \times \text{Prob}$	1	3	6	10	15	21

b) $\mu = 27/4, \sigma^2 = 27/16$ **7.** a) 0.880527b) $79/3200$ **9.** a) 0.956094

b) 0.05629

11. 0.0573656**Exercises 2.11****1.** a) 0.416248

b) 0.00202524

3. $49/50$ **5.** 0.68122**7.** a) $27,683/32,340$ b) $97/2000$ **9.** 0.104399**11.** 10**13.** a) 0.999616

b) 0.00740206

15. a) 314.837**Exercises 2.14**

	X	1	2	3	4	5	6
Binomial	0.32768	0.4096	0.2048	0.0512	0.0064	0.00032	
Poisson	0.367879	0.367879	0.18394	0.0613132	0.0153283	0.00306566	

3. 0.104137**5.** a) 0.0916

b) 0.18045

c) 0.9380

7. a) 0.0469122

a) 0.0661276

9. The Poisson approximation gives 0.067086.

11. a) 0.227975

b) 30 seconds

13. 0.014388

15. a) 0.265026

b) 9.1774×10^{-5}

17. a) 0.04979

b) 0.03374

c) 1.535 ft³

19. 23

21. a) 0.180

b) 0.143

c)	X	0	1	2	3
	Prob	e^{-2}	$2e^{-2}$	$2e^{-2}$	$1 - 5e^{-2}$

d) 0.218

e) 4

23. Y 0 1 2 3 4 5 6 or more

Prob	e^{-4}	$4e^{-4}$	$8e^{-4}$	$\frac{32}{3}e^{-4}$	$\frac{32}{3}e^{-4}$	$\frac{128}{15}e^{-4}$	$1 - \frac{643}{15}e^{-4}$
------	----------	-----------	-----------	----------------------	----------------------	------------------------	----------------------------

Supplementary Problems for Chapter 2

1. 0.938031

3. a) 0.0111603

b) 0.0127952

c) 29

5. 4/5

7. 1/4

9. a) 2, 162/54, 145

b) 109

11. 5/273

13. 0.231639

15. a) $\sum_{y=1}^{\infty} \binom{2y-1}{1} p^2(1-p)^{2y-2}$

b) 1/3

17. 8/65

19. a) 0.013695

b) 0.33282

21. 0.865618

23. 20

25. $p_A^3 + 3p_A^2(1-p_A)(1-p_B^2) + 3p_A(1-p_A)^2(1-p_B)^2$

- 27.** 0.74286
29. 0.707143
31. 0.206051
33. 0.290228, 0.379419
35. Yes
37. 1692

39. a)

W	-1	0	1	2
216 × Prob	125	75	15	1

- b) $-1/2$
c) Make m as large as possible

- 41.** 0.112553
43. $1/12$

CHAPTER 3

Exercises 3.1

- 1.** b) $19/64$
c) $19/27$
d) 0.693361
- 3.** b) $1/4$
c) $\arccos(1/3)$
- 5.** a) $e^{-3/4}$
b) $\mu = 1/3, \sigma^2 = 1/9$
- 7.** a) $3/8$
b) $1/4$
- 9.** b) $3/4$
c) $\mu = 0, \sigma^2 = 2/3$
- 11.** a) $3/2$
b) $F(y) = \begin{cases} 0 & y \leq 0 \\ \frac{y^3}{2} + \frac{y^2}{2} & 0 \leq y \leq 1 \\ 1 & y \geq 1 \end{cases}$
c) $P(Y > y) = \begin{cases} 1 & y \leq 0 \\ 1 - \frac{y^3}{2} - \frac{y^2}{2} & 0 \leq y \leq 1 \\ 0 & y \geq 1 \end{cases}$
- 13.** a) $\mu = 0, \sigma^2 = 27/5$
15. $\mu = 16/35, \sigma^2 = 201/4,900$
19. $\mu = 3/2, \sigma^2 = 1/4$

Exercises 3.2

- 1.** $1/2$
- 3.** $1/3$
- 5.** 0.597126
- 7.** The exact probability is $3/5$.
- 9.** $\frac{\sqrt{3}-1}{3}$
- 11.** $\ln 1.765 - 1$

Exercises 3.4

- 1.** $E[X] = 9/4$, $\text{Var}[X] = 1/16$
- 3.** a) $e^{-2/3}$
b) $e^{-1/3}$
- 5.** a) $e^{-17/15}$
b) 0.293681
- 7.** A should be taken.
- 9.** $\lambda = 1/2$
- 11.** $e^{-6.25}$
- 15.** 122,000 miles
- 17.** b) $F(x) = 1 - e^{-3(x-2)}$, $x \geq 2$.
c) e^{-6}
d) 1.69495×10^{-4}
- 19.** a) λ
b) $e^{-2/\lambda}$

Exercises 3.5

- 1.** a) 0.866386
b) 0.421084
- 3.** a) 0.0026
b) 0.682689
- 5.** III
- 7.** 68.75%
- 9.** Yes
- 11.** \$37.98
- 15.** 17.01% will be outside warning limits
- 17.** a) 0.213485
b) 12 bags gives a probability of about 0.95.
- 19.** -0.427183855
- 21.** 0.975412

23. 910.059**25. a)** 0.308538**b)** 0.2113**c)** 0.0152**27.** D_2 **Exercises 3.6**

Answers given are the exact result followed by the normal approximation.

1. 0.529141, 0.529514

3. 0.0795892 and 0.0666053 are the exact probabilities. The normal approximations are 0.07965579 and 0.0668073.

5. a) 0.219353, 0.214145**b)** 0.934849**7.** 26, 26**9.** 26, 26**11.** 0.229707, 0.230906**13.** 0.9851, 0.9854**15.** 0.657949, 0.642834**Exercises 3.7**

$$\text{1. a) } f(y) = e^{-2y} * 2^3 * \frac{y^2}{2!}$$

$$\text{b) } g(y) = e^{-2y} * 2^6 * \frac{y^5}{5!}$$

$$\text{3. a) } \mu = 1/2, \sigma^2 = 1/20$$

$$\text{b) } 0.66229$$

$$\text{5. b) } 3e^{-2}$$

$$\text{c) } \frac{5}{4}\sqrt{e}$$

11. Yes

$$\text{13. a) } e^{\frac{-y}{12}}$$

$$\text{b) } \frac{1}{12}e^{\frac{-y}{12}}$$

$$\text{c) } 0.434598$$

Exercises 3.8

$$\text{1. b) } 0.301305$$

$$\text{c) } 0.67032$$

$$\text{3. a) } e^{-t^2/20,000}$$

$$\text{b) } 125.331 \text{ h}$$

$$\text{c) } 0.324652$$

5. b) 0.283468**7.** $\beta(\ln 2)^{\frac{1}{\alpha}}$ **9.** 0.443559**Supplementary Exercises for Chapter 3****1.** 0.0730169**3.** 0.1686**5.** 2.27789×10^{-4} **7. a)** 0.367879**b)** 0.306432**9. b)** 0.7875**c)** $62/15$ **d)** $729/8000$ **11. b)** $112/243$

$$\text{c)} \quad F(x) = \begin{cases} 0 & x < 0 \\ 5x^4 - 4x^5 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

13. $\sqrt{11}/5$ **17.** $\mu = 2, \sigma^2 = 1/5$ **19. a)** $k = 2$

$$\text{b)} \quad F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x^2} & x \geq 0 \end{cases}$$

21. 0.0220301

$$\text{23. } F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x < 1/2 \\ 6x - 3x^2 - 2 & \text{if } 1/2 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

25. $\mu = 1, \sigma^2 = 1/2$ **27.** $P(x = -2) = 1/6, P(x = -1) = 1/3, P(x = 2) = 1/2$ **29.** $2/3$ **CHAPTER 4****Exercises 4.3****1.** $g(y) = 1/9, 4 \leq y \leq 13$ **3.** $g(y) = 1/(2\sqrt{y}), 0 < y < 1$ **5. a)** $E(Y) = e^{\mu+1/2\sigma^2}, \text{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$

b) $g(y) = \frac{1}{\sigma y \sqrt{2\pi}} e^{-1/2 \left(\frac{\ln y - \nu}{\sigma} \right)^2}, y > 0$

9. $g(y) = \frac{1}{2} e^{-y^{1/3}}, y > 0$
 $3y^3$

11. $g(y) = \frac{1}{2\sqrt{1-y^2}}, \sin(-1) \leq y \leq \sin(1)$

13. a) $g(y) = \frac{2y}{(1-y^2)^2}, y \geq 0$

15. $602/3$

17. $g(y) = n\lambda e^{-n\lambda y}, y \geq 0$

19. $g(y) = \frac{1}{2\sqrt{y}} - \frac{1}{4}, 0 < y < 4$

21. No

23. a) $Y = a + (b - a)X$

b) $Y = -\frac{1}{\lambda} \ln X$

27. 35

29. $g(y) = 2/3, 1/2 \leq y < 2$

Exercises 4.4

5.

Sum	3	4	5	6	7	8	9	10	11	12
$64 \times \text{Prob}$	1	3	6	10	12	12	10	6	3	1

7. a) $P(X + Y = z) = \left(\frac{1}{2}\right)^{z-2} - 6\left(\frac{1}{3}\right)^z, z = 2, 3, 4, \dots$
 b) $7/2$

Exercises 4.5

1.

Sum	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$64 \times \text{Prob}$	1	3	6	10	15	21	25	27	27	25	21	15	10	3	6	1

5. All the subsets of $\{a, b, c\}$

7. $1 - \left(1 - \frac{1}{t}\right) \log(1 - t)$

9. a) $t^k P_X(t)$
 b) $P_X(t^k)$

11.

Sum	2	3	4	5	6	7	8	9	10	11	12
$126 \times \text{Prob}$	1	3	6	10	15	21	20	18	15	11	6

Exercises 4.7

3. a) $e^{-\lambda(1-t)}$
 5. a) $16/31$
 b) $\frac{1}{31}(16 + 8t + 4t^2 + 2t^3 + t^4)$
 c) $\mu = 26/31, \sigma^2 = 1,122/961$
 7. a) $\frac{71393157}{1073741824} = 0.0664901$
 b) $\frac{2046448125}{2147483648} = 0.952952$
 9. a) e^{t-1}
 b) $\mu = \sigma^2 = 1$

Exercises 4.8

5. a) $e^{-\lambda(1-e^t)}$
 7. a) $\mu = -5/6, \sigma^2 = 65/36$
 b) $1 - \frac{5}{6}t + \frac{5}{4}t^2 - \frac{23}{36}t^3 + \frac{17}{48}t^4 + \dots$
 9. a) $\left(\frac{\lambda}{\lambda-t}\right)^r$
 b) $\mu = r/\lambda, \sigma^2 = r/\lambda^2$
 11. $\frac{1}{3t}(e^{-t} - e^{-2t}) + \frac{2}{9t}(e^{4t} - e^t)$
 13. a) 0.774538
 b) $N(30, 6)$
 15. a) $\frac{1}{t}(e^{3t} - e^{2t})$
 b) $\mu = 5/2, \sigma^2 = 1/12$
 17. a) $\mu = 5/3, \sigma^2 = 10/9$
 b) X is binomial with $n = 5$ and $p = 1/3$

Exercises 4.10

1. a) $\frac{e^t-1}{t}$
 b) $\mu = 3/2, \sigma^2 = 1/4$
 3. a) $\mu = 1/5, \sigma^2 = 1/25$
 b) $f(x) = 5e^{-5x}, x \geq 0$
 5. $\mu = 10, \sigma^2 = 20$
 7. 0.97725
 9. $\mu = \sigma^2 = 1$
 11. M is the generating function for the variable $X - \lambda$, where X is Poisson with parameter λ .

- 17. d)** $X_1 + X_2$ has probability distribution given by $g(y)$
- 19.** $\frac{1}{t^2}(e^t + e^{-t} - 2)$
- 21.** 0.814453

Exercises 4.11

- 3.** 0.974085
- 5.** 166
- 7. a)** 0.3993
b) 1008
- 9.** 0.008853
- 11.** 0.0681571
- 13.** 0.0227501
- 15. a)** 0.382925
b) 0.682689
- 17.** The probability that \bar{X} is in the range 3.11 to 3.61 is 0.95833

Exercises 4.13

- 1.** 77.7867, 1789.36
- 3.** $\frac{(n-1)s^2}{\chi_U^2}, \frac{(n-1)s^2}{\chi_L^2}$ where χ_U^2 and χ_L^2 are upper and lower χ^2 values.
- 5.** 12
- 7.** 0.898904
- 9.** 2.55646, 13.1746

Exercises 4.14

- 1. a)** Yes
b) 0.03338
- 3. a)** $\bar{x} < 0.248897$
b) 0.41365
c) $n = 49$
- 5. a)** 0.0228
b) 0.158655
- 7. a)** 5.175, 10.825
b) 3.8756, 45.282
- 9. a)** Reject if $\bar{x} > 4.02816$ or if $\bar{x} < 3.98179$
b) Reject H_o
c) 0.877193
- 11.** Yes

- 13.** a) 0.055115
b) 0.21186
- 15.** a) No
b) 0.837945
- 17.** a) $\bar{x} < 73.4538$
b) Yes
c) 0.939007
d) 2.34×10^{-12}
- 19.** a) 40,800
b) 1,089,055 to 3,749,072
c) 40,778

Exercises 4.15

- 1.** a) Accept H_o
b) Yes
- 3.** Accept H_o
- 5.** 0.153189, 11.9492
- 7.** -0.223, 8.223
- 9.** a) Accept H_o
b) 0.141803, 4.16881
c) Reject H_o
- 11.** a) -960.19, -309.81
b) -949.55, -320.55
c) 0.00242706
- 13.** a) Accept H_o
b) Reject
- 15.** a) Reject the hypothesis that painting reduces the top speed.
- 17.** No, if $\alpha = 5\%$
- 19.** $n = 110$

Exercises 4.16

- 1.** b) $y = -0.273177 + 1.61719x$
c) $F(1,4) = 99.041$. $P[F(1,4) > 99.041] = 5.726 \times 10^{-4}$, so the fit is a very good one.
- 3.** b) $y = 66.198 - 0.862x$
c) $F(1,8) = 16.9988$, $P[F(1,8) > 16.9988] = 3.33 \times 10^{-3}$, so the fit is a very good one.
- 5.** b) $y = -0.192719 + 1.22056x$
c) $x = 7.31915 + 0.485106y$
- 7.** a) $y = 0.489784 + 0.272431x$
b) $r = 0.92$

9. a) $\hat{\beta} = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n x_i^4}$

b) $\hat{\beta} = 0.953807$

11. $\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$

Exercises 4.17

1. b) 8.8292, 11.2458

Supplementary Exercises for Chapter 4

1. a) $\frac{2}{a^2 t^2}(e^{at} - 1) - \frac{2}{at}$

b) $\mu = a/3, \sigma^2 = a^2/18$

3. a) $\frac{2^{1+t}-1}{1+t}$

b) $\mu = 0.3862, \sigma^2 = 0.03909$

5. $\beta^k \Gamma\left(1 + \frac{k}{\alpha}\right)$

7. 4/81

9. Poisson[λp]

11. 16

13. a) 0.23975

b) 0.0569231

15. a) 0.135666

b) 0.0139034

c) 0.8133

d) 0.0126737

17. a) $\frac{2}{t^2}(e^t - t - 1)$

b) $\frac{2}{(k+1)(k+2)}$

c) $\frac{8e^{t/2}}{t^2}(e^{t/2} - t/2 - 1)$

19. a) $\mu = 5/3, \sigma^2 = 5/9$

b) $\frac{e^t}{2} + \frac{e^{2t}}{3} + \frac{e^{3t}}{6}$

21. a) $P(X = 1) = 2/5, P(X = 2) = 1/5, P(X = 3) = 2/5$

b) $\mu = 2, \sigma^2 = 4/5$

23. $\frac{1}{k\beta} e^{-y/k\beta}, y > 0$

25. a) $g(y) = y^{\frac{-1}{2}}, -1 < y \leq 1$

27. $g(y) = \frac{y}{\sqrt{1-y^2}}, 0 < y < 1$

29. $g(y) = 2/y^3, y > 1. E(Y) = 2$

31. $g(y) = \frac{3-\sqrt{y+1}}{9}, -1 \leq y \leq 8$

33. $g(y) = \begin{cases} 2/3 & \text{if } 0 < y < 1 \\ 1/3 & \text{if } 1 \leq y \leq 2 \end{cases}$

35. 0.017008

37. Yes. $P(\chi_{30}^2 > 56.4) = 0.002456$

39. No

41. 5.88586, 8.17407

- 43.** a) Accept H_o
b) 9.44793, 11.1521

- 45.** a) 70
b) 0.617075

47. 0.965068

- 49.** a) Reject H_o
b) 0.00925637

51. 658.88, 701.13

53. Yes

- 55.** a) Accept H_o
b) 0.2610863

- 57.** a) Reject H_o
b) Reject H_o

- 59.** a) $\bar{x} < 181.997$
b) 0.857088

61. 1.48773, 3.97234

CHAPTER 5

Exercises 5.2

3. a)

x	2	2.5	3	3.5	4
$f(x)$	0.17	0.14	0.20	0.33	0.16
$g(y)$	0.05	0.15	0.28	0.40	0.12

- b)** $E(X) = 3.085, E(Y) = 3.195$
c) 0.68

5. a) $f(x, y) = \binom{4}{x} \binom{4}{y} \binom{44}{2-x-y} / \binom{52}{2}$, $x = 0, 1, 2; y = 0, 1, 2; x + y \leq 2$

b) $f(x) = \begin{cases} \frac{188}{221} & \text{if } x = 0 \\ \frac{32}{221} & \text{if } x = 1; \quad g(y) \text{ is similar.} \\ \frac{1}{221} & \text{if } x = 2 \end{cases}$

7. a) $1/65$

b) $f(x) = \frac{x+1}{65}, x = 1, 2, \dots, 10$

$$g(y) = \begin{cases} \frac{10}{65} & \text{if } y = 9, 10 \\ \frac{y+1}{65} & \text{if } y = 0, 1, \dots, 8 \end{cases}$$

9. a) $P(X = x, Y = y) = \binom{n_1}{x} \binom{n_2}{y} p^{x+y} q^{n_1+n_2-x-y}, x = 0, 1, \dots, n_1, y = 0, 1, \dots, n_2$

b) $X + Y$ is binomial with parameters $n_1 + n_2$ and p .

11. a) $f(x, y) = \frac{e^{-6} 6^x}{x!} \binom{x}{y} \left(\frac{3}{4}\right)^y \left(\frac{1}{4}\right)^{x-y}, x = 0, 1, \dots : y = 0, 1, \dots, x$

b) $g(y) = \frac{e^{-9/2}(9/2)^y}{y!}, y = 0, 1, \dots$

c) $E(Y) = \frac{3}{4} \cdot 6$

13. $P(X = x, Y = y) = \binom{3}{x} \binom{2}{y} \binom{2}{3-x-y} / \binom{7}{3}, x = 0, 1, 2, 3; y = 0, 1, 2; x + y \leq 3$

15. b) $f(x) = 12x^3(1-x^2), 0 < x < 1$

$g(y) = 12y^2(1-y), 0 < y < 1$

17. a) 3

b) $f(x) = 6x(1-x), 0 < x < 1$

$$g(y) = \begin{cases} \frac{3}{2}(1-y)^2 & 0 < y < 1 \\ \frac{3}{2}(1+y)^2 & -1 < y < 0 \end{cases}$$

19. a) W/Y

	0	1	2	$f(w)$
0	1/8	1/8	0	
1	1/8	2/8	1/8	
2	0	1/8	1/8	
$g(y)$	2/8	4/8	2/8	

b) $f(y) = \begin{cases} 1/4 & \text{if } y = 0 \text{ or } 2 \\ 1/2 & \text{if } y = 1 \end{cases}, g(w) \text{ is similar}$

21. b) $1/16$

c) $7/16$

23. **b)** $f(x) = \frac{3}{2}(1 - x^2), 0 < x < 1$
 $g(y) = 3y^2, 0 < y < 1$

Exercises 5.3

3. a) $f(x|y) = \frac{2(x+y)}{2y+1}, 0 \leq x \leq 1$
 $f(y|x) = \frac{2(x+y)}{2x+1}, 0 \leq y \leq 1$

b) $E(X|Y = y) = \frac{3y+2}{6y+3}$
 $E(Y|X = x) = \frac{3x+2}{6x+3}$

5. $f(x|y) = \frac{2x}{1-y^2}, y < x < 1$
 $f(y|x) = \frac{2y}{x^2}, 0 < y < x$

7. a) $k = 8$
b) $E(Y) = 8/15$

9. a) $E(Y|X) = \frac{2}{3}x \frac{1-x^3}{1-x^2}, 0 < x < 1; E(X|Y) = \frac{2}{3}y^{1/2} \frac{1-y^{3/2}}{1-y}, 0 < y < 1$
b) $3/8$

11. a) $f(x) = 1 + 2x - 3x^2, 0 < x < 1$
 $g(y) = \frac{1}{2}(5 - 8y + 3y^2), 0 < y < 1$
b) $f(y|x) = \frac{1}{1-x}, 0 < y < 1 - x$

Exercises 5.5

13. b) $f(x|y) = 2x/y, 0 < x < \sqrt{y}$

c) $\frac{2\sqrt{y}}{3}$

15. $f(x|y) = \frac{1}{2\sqrt{y}}, -\sqrt{y} < x < \sqrt{y}$
 $f(y|x) = \frac{1}{1-x^2}, x^2 < y < 1$

17. a) $f(x) = \frac{4-x}{6}, x = 1, 2, 3. g(y)$ is similar
b) $f(y|x) = \frac{1}{4-x}, y = 1, 2, \dots, 4 - x$

19. $-1/3$

21. b) $f(x) = x^3/4, 0 < x < 2$
c) No
d) $7/12$

23. $220 - 30\sqrt{2}$

27. $14/\sqrt{247}$

Exercises 5.6

1. c) 0.355435

Exercises 5.7

1. $g(z) = \frac{2}{(z+2)^2}, z > 0$

3. $g(z) = \begin{cases} \frac{2}{3}z^2(3-z), & z > 0 \\ \frac{2}{3}(z^3 - 3z^2 + 4) & 1 \leq z < 2 \end{cases}$

Supplementary Exercises for Chapter 5

1. a) 3/128

b) No

c) $f(x) = \frac{4+3x^2}{32}, -2 < x < 2$

$f(x) = \frac{4+3y^2}{32}, -2 < y < 2$

$f(x|y) = \frac{3(x^2+y^2)}{16+12y^2}, -2 < x < 2$

$f(y|x) = \frac{3(x^2+y^2)}{16+12x^2}, -2 < y < 2$

3. a) $f(x,y) = \begin{cases} 1/36 & x = y, x = 1,2, \dots, 6; y = 1,2, \dots, 6. \\ 2/36 & x > y, x = 1,2, \dots, 6; y = 1,2, \dots, 6. \end{cases}$

b) $E(X) = 161/36$

c) $E(Y) = 91/36$

d) $E(X + Y) = E(X) + E(Y)$

5. 1/3

7. a) 3/16

b) $\sqrt{y}/2, 0 < x < \sqrt{y}$

9. $\frac{16-12\log 2}{9} = 0.853582$

11. $\frac{1}{9(1-e^{-1/3})^2}$

13. a) $f(x,y) = 12x, 0 < y < 2x, 0 < x < 1/2$

b) $g(y) = \frac{3}{2}(1-y^2), 0 < y < 1$

c) No

15. a) 2

b) No

c) $76/81$ **17.** a) 19

b) -2

c) $-1/5\sqrt{3}$

d) 1

e) 250

f) 44

19. a) 12b) $f(x) = 6x^5, 0 < x < 1$ $g(y) = 6y(1-y), 0 < y < 1$

c) No

d) $27/32$

$Y \setminus X$	0	1	2	3	4	$g(y)$
0	1/16					1/16
1		4/16	3/16			7/16
2			3/16	2/16		5/16
3				2/16		2/16
4					1/16	1/16

b) $f(x) = 1/16, 4/16, 6/16, 4/16, 1/16$ **c)** 2**23.** a) $1/2$ b) $f(x) = 1, 0 < x < 1$

$$g(y) = \frac{1}{\sqrt{y}} - 1, 0 < y < 1$$

c) $1 - \sqrt{2}/2$

CHAPTER 6

Exercises 6.2

3. X is a geometric random variable. $a_n = pq^{n-1}, n \geq 1$ **5.** a) $a_n = p^2 + qa_{n-1} + pqa_{n-1}, n \geq 2, a_0 = a_1 = 0, a_2 = p^2$ **7.** $\frac{(1-p)(p_1-p_2)^{n-1}+p_2}{1-p_1+p_2}, n \geq 1$ **9.** a) $\frac{3}{7} \left(\frac{5}{12}\right)^n + \frac{4}{7}, n = 0, 1, 2, \dots$

b) 0.754238

11. $\frac{1+p+p^2}{p^3}$

Exercises 6.3

1. $p_n = \frac{n}{N}$

Exercises 6.4

1. a) $u_0 = 1, u_1 = u_2 = 0,$

b) $U(s) = 1 + \frac{(ps)^3}{(1-s)(1+ps+(ps)^2)}; \frac{29801576}{1162261467} = 0.025641$

c) $F(s) = \frac{(ps)^3}{(ps)^3 + (1-s)(1+ps+(ps)^2)}; \frac{58333184}{3486784401} = 0.0167298$

3. $\frac{1+2pq-5pq^2+p^2q^2-p^2q^3}{p^2q^4}$

Exercises 6.5

3. $(1/3, 2/3)$

5. $(444/699, 165/699, 90/699)$

9.
$$\begin{pmatrix} 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \end{pmatrix}$$

Supplementary Exercises for Chapter 6

1. a) $u_n = q^n + \sum_{k=3}^n \binom{k-1}{2} p^3 q^{k-3} u_{n-k}, n \geq 3, u_0 = 1, u_1 = q, u_2 = q^2, u_3 = p^3 + q^3$

b) $u_n = pw_{n-1} + qu_{n-1}; v_n = pu_{n-1} + qv_{n-1}; w_n = pv_{n-1} + qw_{n-1};$

$v_n = 3qv_{n-1} - 3q^2v_{n-2} + (p^3 + q^3)v_{n-3}$

$n \geq 4, v_0 = 0, v_1 = p, v_2 = 2pq, v_3 = 3pq^2$

3. $\frac{1}{2}[1 + (q + p)^n]$

5. $\binom{2n}{n} \frac{1}{2n-1} p^n q^n$

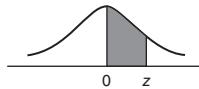
7. No

9. $p_n = \frac{1}{2}[1 + (2p - 1)^n], n \geq 1$

Appendix C

Standard Normal Distribution

The entries in this table give the areas under the standard normal curve from 0 to z .



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441

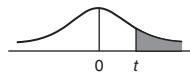
(continued)

(Continued)

<i>z</i>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

THE t DISTRIBUTION TABLE

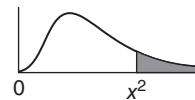
The entries in the table give the critical values of t for the specified number of degrees of freedom and areas in the right tail.



df	Area in the Right Tail under the t Distribution Curve					
	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
31	1.309	1.696	2.040	2.453	2.744	3.375
32	1.309	1.694	2.037	2.449	2.738	3.365
33	1.308	1.692	2.035	2.445	2.733	3.356
34	1.307	1.691	2.032	2.441	2.728	3.348
35	1.306	1.690	2.030	2.438	2.724	3.340
36	1.306	1.688	2.028	2.434	2.719	3.333
37	1.305	1.687	2.026	2.431	2.715	3.326
38	1.304	1.686	2.024	2.429	2.712	3.319
39	1.304	1.685	2.023	2.426	2.708	3.313
40	1.303	1.684	2.021	2.423	2.704	3.307
∞	1.282	1.645	1.960	2.326	2.576	3.090

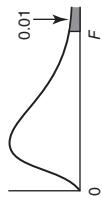
CHI-SQUARED DISTRIBUTION TABLE

The entries in this table give the critical values of X^2 for the specified number of degrees of freedom and areas in the right tail.



df	Area in the Right Tail under the Chi-square Distribution Curve									
	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Area in the Right Tail under the F Distribution Curve = 0.01.



Degrees of Freedom for the Numerator

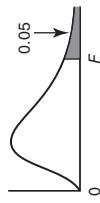
	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6157	6209	6240	6261	6287	6303	6334
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47	99.48	99.49	
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05	26.87	26.69	26.58	26.50	26.41	26.35	26.24
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.20	14.02	13.91	13.84	13.75	13.69	13.58
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.72	9.55	9.45	9.38	9.29	9.24	9.13
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.56	7.40	7.30	7.23	7.14	7.09	6.99
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.75
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.52	5.36	5.26	5.20	5.12	5.07	4.96
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.41
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.01
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.25	4.10	4.01	3.94	3.86	3.81	3.71
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.47
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.27
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.11
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.28	3.21	3.13	3.08	2.98

(continued)

(Continued)

Degrees of Freedom for the Numerator											
	1	2	3	4	5	6	7	8	9	10	11
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.37
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.36
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43

Area in the Right Tail under the F Distribution Curve = 0.05.



	Degrees of Freedom for the Numerator																		
	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	246.0	248.0	249.3	250.1	251.1	251.8	253.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.47	19.48	19.49		
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.41
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.94	3.87	3.83	3.81	3.77	3.75	3.71
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.27
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.22	3.15	3.11	3.08	3.04	3.02	2.97
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.76
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.59
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.72	2.65	2.60	2.57	2.53	2.51	2.46
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.35
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.53	2.46	2.41	2.38	2.34	2.31	2.26
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.46	2.39	2.34	2.31	2.27	2.24	2.19
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.12

(continued)

Continued)

Degrees of Freedom for the Numerator

	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.35	2.28	2.23	2.19	2.12	2.07	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.31	2.23	2.18	2.15	2.10	2.08	2.02
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.27	2.19	2.14	2.11	2.06	2.04	1.98
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.23	2.16	2.11	2.07	2.03	2.00	1.94
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.07	2.04	1.99	1.97	1.91
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.18	2.10	2.05	2.01	1.96	1.94	1.88
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.15	2.07	2.02	1.97	1.94	1.91	1.85
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.13	2.05	2.00	1.96	1.91	1.88	1.82
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.16	2.03	1.97	1.94	1.89	1.86	1.80
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.78
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.88	1.84	1.79	1.76	1.70
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.78	1.74	1.69	1.66	1.59
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.87	1.78	1.73	1.69	1.63	1.60	1.52
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.77	1.68	1.62	1.57	1.52	1.48	1.39

Index

- α , 94, 240
Acceptance Sampling, 111, 113, 119
Addition Theorem, 10
AIDS example, 18
All heads problem, 105
Analysis of variance, 262
Average outgoing quality, 113, 122
Average outgoing quality limit, 123
Average waiting time, 342
Axioms of probability, 7
- β , 94, 240
Banach match book (candy jars) problem, 107
Bayes' Theorem, 17
Behrens-Fisher problem, 250
Bernoulli random variables, 204
Bernoulli trials, 81, 330
 waiting time for patterns, 339
Binomial coefficients, 44 *see also:* combinations
Binomial distribution, 81, 213
 approximation to hypergeometric, 115
 examples, 116
 mean, 103
 normal approximation, 175, 330
 Poisson approximation, 132
 probability generating function, 207, 213
 recursion, 82
 sums, 204
 variance, 84, 103
Binomial theorem, 44
Birthday problem, 284
Bivariate normal distribution, 308
Bivariate random variable, 288
Bootstrap, 375
- Candy jars (Banach match book) problem, 107
Cauchy distribution, 153, 203
Central limit theorem, 229
- Cereal box problem, 216
Chi-squared distribution, 178, 187, 236
Combinations, 43 *see also:* binomial coefficients, 44
 Pascal's identity, 45
Conditional distributions, 293
Conditional expectation, 295, 303
Conditional probability, 14
 examples, 28, 152
Confidence coefficient, 100
Confidence interval
 for μ , 89, 90, 241, 243
 for σ^2 , 238
 for p , 90, 140
 for s , 239
Consumer's risk, 121
Continuous random variable, 146
Control chart, 266
Control limits, 267
Counting techniques, 39
Counting principles, 34
Coupon Collector's problem, 216, 362
Contour plots, 310
Cook's Distance, 374
Correlation coefficient, 298, 300
Covariance, 299
Critical region, 93
Cumulative distribution function, 68, 148, 161
 examples, 69, 144
 properties, 70, 150
- Derangements, 41, 49
Difference equation (recursion), 44
 solution, 322
Disjoint events, 16
Distribution function, 68, 149
 conditional, 294
 example, 69
 exponential, 161

- Distribution function (*continued*)
- F , 251
 - Marginal, 283
 - properties, 70
 - Student's t , 242
 - Weibull, 184
- Double sampling, 124
- Estimator, 89
- interval, 89, 90, 241, 243
 - least squares, 258
 - point, 89
 - unbiased, 98
- Events, 7, 12
- disjoint, 16
 - examples, 12
 - independent, 22, 23, 24
 - mutually exclusive, 9, 13
- Expected value
- bivariate, 298
 - conditional, 303
 - discrete, 72
 - examples, 72
 - of a function, 199
 - geometric, 102
 - hypergeometric, 111
 - Poisson, 300
 - of sums, 203
 - Uniform, 157
- Experiments, 2
- examples, 2
- Exponential distribution, 159
- mean, 160
 - memoryless property, 160
 - variance, 160
- F distribution, 251
- Fibonacci sequence, 5
- Fixed vector, 346
- Functions of random variables, 195
- products, 314
 - quotients, 315
 - sums, 203, 312
- Gamma distribution, 178
- mean, 180
 - moment generating function, 218
 - Variance, 180
- Generating functions, 207, 341
- binomial, 213
 - from recursions, 339, 341
 - geometric, 215
 - means and variances from, 343
- moment, 218
- probability, 213, 215
- properties, 211, 223
- Geometric random variable, 68, 74, 102, 366
- expected value, 74
 - probability generating function, 216
 - variance, 77
- General Addition Law, 46
- HIV example, 18
- Hazard Rate, 163
- Weibull, 184
- Huygens problem, 384
- Hypergeometric random variable, 111, 114, 128
- binomial approximation, 115
 - mean, 114
 - moment generating function, 218
 - variance, 114
- Hypothesis, 93, 240
- alternative, 93
 - composite, 95
 - null, 93
- Hypothesis testing
- binomial, 96
 - on μ , 240
 - on two means, 248
 - on two variances, 251
- Inclusion and exclusion, 12
- Independence, 299
- Indicator random variables, 302
- Independent events, 22, 23, 24
- examples, 24
- Intersection of sets, 9
- Interval estimator, *see* confidence interval
- Joint probability distributions, 283
- Ken–Ken, 50
- Law of total probability, 16
- Least squares, 258
- Linear model, 258
- Linear regression, 258
- Lottery, 128
- Marginal distributions, 286
- Markov chains, 344
- absorbing states, 349
- Matching problem, 39
- Matrix, 344
- fundamental, 344, 350

- stochastic, 345
- transition, 344
- Maximum, 48, 198
 - sampling distribution, 48
- Mean
 - binomial, 84
 - candy jars (Banach match book) problem, 107
 - control chart, 266
 - exponential, 160
 - gamma, 178
 - geometric, 215
 - hypergeometric, 111
 - negative binomial, 102
 - normal, 166
 - Poisson, 131
 - properties, 151
 - sampling distribution, 229
 - sample proportion, 98
 - uniform, 57
 - Weibull, 184
- Median, 48
 - race car example, 48
- Moment generating function, 218
 - exponential, 220
 - gamma, 225
 - normal, 225
 - properties, 218
 - sums, 224
 - uniform, 219
- Moments, 218
- Monte Hall example, 21
- Mowing the lawn, 31
- Multiplication theorem, 14
- Mutually exclusive events, 9, 23
- Negative binomial distribution, 102
 - mean, 102
 - variance, 102
- Normal distribution, 66, 166
 - approximation to binomial, 175
 - bivariate, 308
 - mean, 169
 - moment generating function, 223
 - sums, 225
 - variance, 169
- Operating characteristic curve, 120
- p* values, 243
- Pascal's identity, 45
- Patterns in Bernoulli trials, 339
- Permutations, 40
- Poisson process, 134
- Poisson random variable, 130
 - approximation to binomial, 132
 - mean, 131
 - variance, 131
- Poisson's trials, 214
- Probability
 - axioms, 8
 - conditional, 14
 - examples, 28
 - law of total, 16
 - theorems, 10
- Probability density function, 147
- Probability distribution function, 62
 - binomial, 81
 - conditional, 293
 - discrete uniform, 63
 - exponential, 159
 - F*, 251
 - gamma, 178
 - geometric, 68
 - joint, 283
 - marginal, 283
 - negative binomial, 102
 - normal, 166
 - Students' *t*, 242
 - uniform, 157
 - Weibull, 184
- Probability generating function, 213
 - binomial, 213
 - geometric, 215
 - properties, 211
 - sums, 203, 224
- Producer's risk, 121
- proportion, 98
 - sample, 98
 - confidence interval, 100
 - distribution, 99
 - variance, 99
- Quality control chart for sample means, 266
- Quality control inspector problem, 194
- Race car example, 48
- Random variable, 61
 - bivariate, 283
 - continuous, 146
 - discrete uniform, 63
 - expected value, 72
 - functions of, 195
 - geometric, 68
 - hypergeometric, 114
 - independent, 299
 - indicator, 302

- Random variable (*continued*)
 negative binomial, 102
 Poisson, 130
 sums, 203, 224
 variance, 75
- Random walk and ruin, 334
 expected duration, 337
 solution, 334
- Recursion, 49, 82, 108, 322
 binomial, 83
 candy jars (Banach match book) problem, 107
 generating functions from, 131, 341, 215
 hypergeometric, 111
 Poisson, 130
 solving, 326
 using to find means and variances, 329, 343
- Regression, 258, 309, 322
- Reliability, 34, 69, 162, 184
 Weibull distribution, 148
- Residuals, 258
- Risks
 consumer's, 161
 producer's, 121
- Sample proportion, 98
- Sample space, 2
 examples, 3
- Sampling
 double, 124
- Sampling distribution
 sample mean, 229, 273
 sample median, 48
 sample maximum, 198
 sample variance, 234
- Sampling inspection plans, 112
- Socks problem, 357
- Standard deviation, 76
 distribution, 267
- Stochastic matrix, 345
- Student's *t* distribution, 242
- Sums
 loaded dice, 66
 several dice, 208
 two dice, 67
- Sums of random variables, 203, 224
 expected value, 73, 224
 moment generating function, 224
 of binomials, 204
 of exponentials, 225
 of normals, 225
 variance, 224
- Sums of squares
 regression, 261
- residual, 261
 total, 261
- Systems
 parallel, 35
 series, 34
- Tchebycheff's inequality, 78
- Theorems
 addition, 10
 Bayes', 17
 probability, 10
- Transition matrix, 344
 regular, 344
- Type I error, 94, 240
- Type II error, 94, 240
- Unbiased estimator, 235
- Uniform probability distribution
 discrete, 63, 71
 examples, 63
 joint, 283
 mean, 157
 sums, 205
 variance, 157
- Union of sets, 9
- Variance, 75, 157, 234
 binomial, 84
 by conditioning, 307
 candy jars (Banach match book) problem, 107
 exponential, 160
 gamma, 178
 geometric, 71
 hypergeometric, 111
 negative binomial, 102
 normal, 166
 of sums, 224, 300
 Poisson, 131
 properties, 151
 sample proportion, 98
 sampling distribution, 234
 uniform, 151
 variance, 238
 Weibull, 184
- Vector
 fixed, 347
- Venn diagram, 10, 11
- Waiting times in Bernoulli trials, 330, 339
- Waldegrave problem, 378
- Weibull distribution, 184
 mean, 185
 variance, 185
- Weak law of large numbers, 233

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.