INSTRUCTOR'S SOLUTIONS MANUAL (ONLINE ONLY)

MARK SCHERVISH
Carnegie Mellon University

PROBABILITY AND STATISTICS
FOURTH EDITION

Morris DeGroot
Carnegie Mellon University

Mark Schervish
Carnegie Mellon University

Addison-Wesley
is an imprint of
PEARSON

Reproduced by Pearson Addison-Wesley from electronic files supplied by the author.

Publishing as Pearson Addison-Wesley, 75 Arlington Street, Boston, MA 02116.

ISBN-13: 978-0-321-71597-5
ISBN-10: 0-321-71597-7

Addison-Wesley
is an imprint of

PEARSON

www.pearsonhighered.com

Contents

Preface

This manual contains solutions to all of the exercises in *Probability and Statistics*, 4th edition, by Morris DeGroot and myself. I have preserved most of the solutions to the exercises that existed in the 3rd edition. Certainly errors have been introduced, and I will post any errors brought to my attention on my web page `http://www.stat.cmu.edu/ mark/` along with errors in the text itself. Feel free to send me comments.

For instructors who are familiar with earlier editions, I hope that you will find the 4th edition at least as useful. Some new material has been added, and little has been removed. Assuming that you will be spending the same amount of time using the text as before, something will have to be skipped. I have tried to arrange the material so that instructors can choose what to cover and what not to cover based on the type of course they want. This manual contains commentary on specific sections right before the solutions for those sections. This commentary is intended to explain special features of those sections and help instructors decide which parts they want to require of their students. Special attention is given to more challenging material and how the remainder of the text does or does not depend upon it.

To teach a mathematical statistics course for students with a strong calculus background, one could safely cover all of the material for which one could find time. The Bayesian sections include 4.8, 7.2, 7.3, 7.4, 8.6, 9.8, and 11.4. One can choose to skip some or all of this material if one desires, but that would be ignoring one of the unique features of the text. The more challenging material in Sections 7.7–7.9, and 9.2–9.4 is really only suitable for a mathematical statistics course. One should try to make time for some of the material in Sections 12.1–12.3 even if it meant cutting back on some of the nonparametrics and two-way ANOVA. To teach a more modern statistics course, one could skip Sections 7.7–7.9, 9.2–9.4, 10.8, and 11.7–11.8. This would leave time to discuss robust estimation (Section 10.7) and simulation (Chapter 12). Section 3.10 on Markov chains is not actually necessary even if one wishes to introduce Markov chain Monte Carlo (Section 12.5), although it is helpful for understanding what this topic is about.

Using Statistical Software

The text was written without reference to any particular statistical or mathematical software. However, there are several places throughout the text where references are made to what general statistical software might be able to do. This is done for at least two reasons. One is that different instructors who wish to use statistical software while teaching will generally choose different programs. I didn't want the text to be tied to a particular program to the exclusion of others. A second reason is that there are still many instructors of mathematical probability and statistics courses who prefer not to use any software at all.

Given how pervasive computing is becoming in the use of statistics, the second reason above is becoming less compelling. Given the free and multiplatform availability and the versatility of the environment *R*, even the first reason is becoming less compelling. Throughout this manual, I have inserted pointers to which *R* functions will perform many of the calculations that would formerly have been done by hand when using this text. The software can be downloaded for Unix, Windows, or Mac OS from
`http://www.r-project.org/`
That site also has manuals for installation and use. Help is also available directly from within the *R* environment.

Many tutorials for getting started with *R* are available online. At the official *R* site there is the detailed manual: `http://cran.r-project.org/doc/manuals/R-intro.html`
that starts simple and has a good table of contents and lots of examples. However, reading it from start to finish is *not* an efficient way to get started. The sample sessions should be most helpful.

One major issue with using an environment like *R* is that it is essentially programming. That is, students who have never programmed seriously before are going to have a steep learning curve. Without going into the philosophy of whether students should learn statistics without programming, the field is moving in the direction of requiring programming skills. People who want only to understand what a statistical analysis is about can still learn that without being able to program. But anyone who actually wants to do statistics as part of their job will be seriously handicapped without programming ability. At the end of this manual is a series of heavily commented R programms that illustrate many of the features of R in the context of a specific example from the text.

Mark J. Schervish

Chapter 1

Introduction to Probability

1.2 Interpretations of Probability

Commentary

It is interesting to have the students determine some of their own subjective probabilities. For example, let X denote the temperature at noon tomorrow outside the building in which the class is being held. Have each student determine a number x_1 such that the student considers the following two possible outcomes to be equally likely: $X \le x_1$ and $X > x_1$. Also, have each student determine numbers x_2 and x_3 (with $x_2 < x_3$) such that the student considers the following three possible outcomes to be equally likely: $X \le x_2$, $x_2 < X < x_3$, and $X \ge x_3$. Determinations of more than three outcomes that are considered to be equally likely can also be made. The different values of x_1 determined by different members of the class should be discussed, and also the possibility of getting the class to agree on a common value of x_1.

Similar determinations of equally likely outcomes can be made by the students in the class for quantities such as the following ones which were found in the 1973 World Almanac and Book of Facts: the number of freight cars that were in use by American railways in 1960 (1,690,396), the number of banks in the United States which closed temporarily or permanently in 1931 on account of financial difficulties (2,294), and the total number of telephones which were in service in South America in 1971 (6,137,000).

1.4 Set Theory

Solutions to Exercises

1. Assume that $x \in B^c$. We need to show that $x \in A^c$. We shall show this indirectly. Assume, to the contrary, that $x \in A$. Then $x \in B$ because $A \subset B$. This contradicts $x \in B^c$. Hence $x \in A$ is false and $x \in A^c$.

2. First, show that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. That is, $x \in A$ and either $x \in B$ or $x \in C$ (or both). So either ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$) or both. That is, either $x \in A \cap B$ or $x \in A \cap C$. This is what it means to say that $x \in (A \cap B) \cup (A \cap C)$. Thus $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Basically, running these steps backwards shows that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

3. To prove the first result, let $x \in (A \cup B)^c$. This means that x is not in $A \cup B$. In other words, x is neither in A nor in B. Hence $x \in A^c$ and $x \in B^c$. So $x \in A^c \cap B^c$. This proves that $(A \cup B)^c \subset A^c \cap B^c$. Next, suppose that $x \in A^c \cap B^c$. Then $x \in A^c$ and $x \in B^c$. So x is neither in A nor in B, so it can't be in $A \cup B$. Hence $x \in (A \cup B)^c$. This shows that $A^c \cap B^c \subset (A \cup B)^c$. The second result follows from the first by applying the first result to A^c and B^c and then taking complements of both sides.

4. To see that $A \cap B$ and $A \cap B^c$ are disjoint, let $x \in A \cap B$. Then $x \in B$, hence $x \notin B^c$ and so $x \notin A \cap B^c$. So no element of $A \cap B$ is in $A \cap B^c$, hence the two events are disjoint. To prove that $A = (A \cap B) \cup (A \cap B^c)$, we shall show that each side is a subset of the other side. First, let $x \in A$. Either $x \in B$ or $x \in B^c$. If $x \in B$, then $x \in A \cap B$. If $x \in B^c$, then $x \in A \cap B^c$. Either way, $x \in (A \cap B) \cup (A \cap B^c)$. So every element of A is an element of $(A \cap B) \cup (A \cap B^c)$ and we conclude that $A \subset (A \cap B) \cup (A \cap B^c)$. Finally, let $x \in (A \cap B) \cup (A \cap B^c)$. Then either $x \in A \cap B$, in which case $x \in A$, or $x \in A \cap B^c$, in which case $x \in A$. Either way $x \in A$, so every element of $(A \cap B) \cup (A \cap B^c)$ is also an element of A and $(A \cap B) \cup (A \cap B^c) \subset A$.

5. To prove the first result, let $x \in (\cup_i A_i)^c$. This means that x is not in $\cup_i A_i$. In other words, for every $i \in I$, x is not in A_i. Hence for every $i \in I$, $x \in A_i^c$. So $x \in \cap_i A_i^c$. This proves that $(\cup_i A_i)^c \subset \cap_i A_i^c$. Next, suppose that $x \in \cap_i A_i^c$. Then $x \in A_i^c$ for every $i \in I$. So for every $i \in I$, x is not in A_i. So x can't be in $\cup_i A_i$. Hence $x \in (\cup_i A_i)^c$. This shows that $\cap_i A_i^c \subset (\cup_i A_i)^c$. The second result follows from the first by applying the first result to A_i^c for $i \in I$ and then taking complements of both sides.

6. (a) Blue card numbered 2 or 4.
   (b) Blue card numbered 5, 6, 7, 8, 9, or 10.
   (c) Any blue card or a red card numbered 1, 2, 3, 4, 6, 8, or 10.
   (d) Blue card numbered 2, 4, 6, 8, or 10, or red card numbered 2 or 4.
   (e) Red card numbered 5, 7, or 9.

7. (a) These are the points not in A, hence they must be either below 1 or above 5. That is $A^c = \{x : x < 1 \text{ or } x > 5\}$.
   (b) These are the points in either A or B or both. So they must be between 1 and 5 or between 3 and 7. That is, $A \cup B = \{x : 1 \le x \le 7\}$.
   (c) These are the points in B but not in C. That is $BC^c = \{x : 3 < x \le 7\}$. (Note that $B \subset C^c$.)
   (d) These are the points in none of the three sets, namely $A^cB^cC^c = \{x : 0 < x < 1 \text{ or } x > 7\}$.
   (e) These are the points in the answer to part (b) and in C. There are no such values and $(A \cup B)C = \emptyset$.

8. Blood type A reacts only with anti-A, so type A blood corresponds to $A \cap B^c$. Type B blood reacts only with anti-B, so type B blood corresponds to A^cB. Type AB blood reacts with both, so $A \cap B$ characterizes type AB blood. Finally, type O reacts with neither antigen, so type O blood corresponds to the event A^cB^c.

9. (a) For each n, $B_n = B_{n+1} \cup A_n$, hence $B_n \supset B_{n+1}$ for all n. For each n, $C_{n+1} \cap A_n = C_n$, so $C_n \subset C_{n+1}$.
   (b) Suppose that $x \in \cap_{n=1}^{\infty} B_n$. Then $x \in B_n$ for all n. That is, $x \in \cup_{i=n}^{\infty} A_i$ for all n. For $n = 1$, there exists $i \ge n$ such that $x \in A_i$. Assume to the contrary that there are at most finitely many i such that $x \in A_i$. Let m be the largest such i. For $n = m + 1$, we know that there is $i \ge n$ such that $x \in A_i$. This contradicts m being the largest i such that $x \in A_i$. Hence, x is in infinitely many A_i. For the other direction, assume that x is in infinitely many A_i. Then, for every n, there is a value of $j > n$ such that $x \in A_j$, hence $x \in \cup_{i=n}^{\infty} A_i = B_n$ for every n and $x \in \cap_{n=1}^{\infty} B_n$.
   (c) Suppose that $x \in \cup_{n=1}^{\infty} C_n$. That is, there exists n such that $x \in C_n = \cap_{i=n}^{\infty} A_i$, so $x \in A_i$ for all $i \ge n$. So, there at most finitely many i (a subset of $1, \ldots, n-1$) such that $x \notin A_i$. Finally, suppose that $x \in A_i$ for all but finitely many i. Let k be the last i such that $x \notin A_i$. Then $x \in A_i$ for all $i \ge k + 1$, hence $x \in \cap_{i=k+1}^{\infty} A_i = C_{k+1}$. Hence $x \in \cup_{n=1}^{\infty} C_n$.

10. (a) All three dice show even numbers if and only if all three of A, B, and C occur. So, the event is $A \cap B \cap C$.

    (b) None of the three dice show even numbers if and only if all three of A^c, B^c, and C^c occur. So, the event is $A^c \cap B^c \cap C^c$.

    (c) At least one die shows an odd number if and only if at least one of A^c, B^c, and C^c occur. So, the event is $A^c \cup B^c \cup C^c$.

    (d) At most two dice show odd numbers if and only if at least one die shows an even number, so the event is $A \cup B \cup C$. This can also be expressed as the union of the three events of the form $A \cap B \cap C^c$ where exactly one die shows odd together with the three events of the form $A \cap B^c \cap C^c$ where exactly two dice show odd together with the even $A \cap B \cap C$ where no dice show odd.

    (e) We can enumerate all the sums that are no greater than 5: $1+1+1$, $2+1+1$, $1+2+1$, $1+1+2$, $2+2+1$, $2+1+2$, and $1+2+2$. The first of these corresponds to the event $A_1 \cap B_1 \cap C_1$, the second to $A_2 \cap B_1 \cap C_1$, etc. The union of the seven such events is what is requested, namely

$$(A_1 \cap B_1 \cap C_1) \cup (A_2 \cap B_1 \cap C_1) \cup (A_1 \cap B_2 \cap C_1) \cup (A_1 \cap B_1 \cap C_2) \cup (A_2 \cap B_2 \cap C_1) \cup (A_2 \cap B_1 \cap C_2) \cup (A_1 \cap B_2 \cap C_2).$$

11. (a) All of the events mentioned can be determined by knowing the voltages of the two subcells. Hence the following set can serve as a sample space

$$S = \{(x, y) : 0 \le x \le 5 \text{ and } 0 \le y \le 5\},$$

    where the first coordinate is the voltage of the first subcell and the second coordinate is the voltage of the second subcell. Any more complicated set from which these two voltages can be determined could serve as the sample space, so long as each outcome could at least hypothetically be learned.

    (b) The power cell is functional if and only if the sum of the voltages is at least 6. Hence, $A = \{(x, y) \in S : x + y \ge 6\}$. It is clear that $B = \{(x, y) \in S : x = y\}$ and $C = \{(x, y) \in S : x > y\}$. The powercell is not functional if and only if the sum of the voltages is less than 6. It needs less than one volt to be functional if and only if the sum of the voltages is greater than 5. The intersection of these two is the event $D = \{(x, y) \in S : 5 < x + y < 6\}$. The restriction "$\in S$" that appears in each of these descriptions guarantees that the set is a subset of S. One could leave off this restriction and add the two restrictions $0 \le x \le 5$ and $0 \le y \le 5$ to each set.

    (c) The description can be worded as "the power cell is not functional, and needs at least one more volt to be functional, and both subcells have the same voltage." This is the intersection of A^c, D^c, and B. That is, $A^c \cap D^c \cap B$. The part of D^c in which $x + y \ge 6$ is not part of this set because of the intersection with A^c.

    (d) We need the intersection of A^c (not functional) with C^c (second subcell at least as big as first) and with B^c (subcells are not the same). In particular, $C^c \cap B^c$ is the event that the second subcell is strictly higher than the first. So, the event is $A^c \cap B^c \cap C^c$.

1.5 The Definition of Probability

Solutions to Exercises

1. Define the following events:

$$\begin{aligned} A &= \{\text{the selected ball is red}\}, \\ B &= \{\text{the selected ball is white}\}, \\ C &= \{\text{the selected ball is either blue, yellow, or green}\}. \end{aligned}$$

We are asked to find $\Pr(C)$. The three events A, B, and C are disjoint and $A \cup B \cup C = S$. So $1 = \Pr(A) + \Pr(B) + \Pr(C)$. We are told that $\Pr(A) = 1/5$ and $\Pr(B) = 2/5$. It follows that $\Pr(C) = 2/5$.

2. Let B be the event that a boy is selected, and let G be the event that a girl is selected. We are told that $B \cup G = S$, so $G = B^c$. Since $\Pr(B) = 0.3$, it follows that $\Pr(G) = 0.7$.

3. (a) If A and B are disjoint then $B \subset A^c$ and $BA^c = B$, so $\Pr(BA^c) = \Pr(B) = 1/2$.
   (b) If $A \subset B$, then $B = A \cup (BA^c)$ with A and BA^c disjoint. So $\Pr(B) = \Pr(A) + \Pr(BA^c)$. That is, $1/2 = 1/3 + \Pr(BA^c)$, so $\Pr(BA^c) = 1/6$.
   (c) According to Theorem 1.4.11, $B = (BA) \cup (BA^c)$. Also, BA and BA^c are disjoint so, $\Pr(B) = \Pr(BA) + \Pr(BA^c)$. That is, $1/2 = 1/8 + \Pr(BA^c)$, so $\Pr(BA^c) = 3/8$.

4. Let E_1 be the event that student A fails and let E_2 be the event that student B fails. We want $\Pr(E_1 \cup E_2)$. We are told that $\Pr(E_1) = 0.5$, $\Pr(E_2) = 0.2$, and $\Pr(E_1E_2) = 0.1$. According to Theorem 1.5.7, $\Pr(E_1 \cup E_2) = 0.5 + 0.2 - 0.1 = 0.6$.

5. Using the same notation as in Exercise 4, we now want $\Pr(E_1^c \cap E_2^c)$. According to Theorems 1.4.9 and 1.5.3, this equals $1 - \Pr(E_1 \cup E_2) = 0.4$.

6. Using the same notation as in Exercise 4, we now want $\Pr([E_1 \cap E_2^c] \cup [E_1^c \cap E_2])$. These two events are disjoint, so

$$\Pr([E_1 \cap E_2^c] \cup [E_1^c \cap E_2]) = \Pr(E_1 \cap E_2^c) + \Pr(E_1^c \cap E_2).$$

Use the reasoning from part (c) of Exercise 3 above to conclude that

$$\begin{aligned} \Pr(E_1 \cap E_2^c) &= \Pr(E_1) - \Pr(E_1 \cap E_2) = 0.4, \\ \Pr(E_1^c \cap E_2) &= \Pr(E_2) - \Pr(E_1 \cap E_2) = 0.1. \end{aligned}$$

It follows that the probability we want is 0.5.

7. Rearranging terms in Eq. (1.5.1) of the text, we get

$$\Pr(A \cap B) = \Pr(A) + \Pr(B) - \Pr(A \cup B) = 0.4 + 0.7 - \Pr(A \cup B) = 1.1 - \Pr(A \cup B).$$

So $\Pr(A \cap B)$ is largest when $\Pr(A \cup B)$ is smallest and vice-versa. The smallest possible value for $\Pr(A \cup B)$ occurs when one of the events is a subset of the other. In the present exercise this could only happen if $A \subset B$, in which case $\Pr(A \cup B) = \Pr(B) = 0.7$, and $\Pr(A \cap B) = 0.4$. The largest possible value of $\Pr(A \cup B)$ occurs when either A and B are disjoint or when $A \cup B = S$. The former is not possible since the probabilities are too large, but the latter is possible. In this case $\Pr(A \cup B) = 1$ and $\Pr(A \cap B) = 0.1$.

8. Let A be the event that a randomly selected family subscribes to the morning paper, and let B be the event that a randomly selected family subscribes to the afternoon paper. We are told that $\Pr(A) = 0.5$, $\Pr(B) = 0.65$, and $\Pr(A \cup B) = 0.85$. We are asked to find $\Pr(A \cap B)$. Using Theorem 1.5.7 in the text we obtain

$$\Pr(A \cap B) = \Pr(A) + \Pr(B) - \Pr(A \cup B) = 0.5 + 0.65 - 0.85 = 0.3.$$

9. The required probability is

$$\begin{aligned}\Pr(A \cap B^C) + \Pr(A^C B) &= [\Pr(A) - \Pr(A \cap B)] + [\Pr(B) - \Pr(A \cap B)] \\ &= \Pr(A) + \Pr(B) - 2\Pr(A \cap B).\end{aligned}$$

10. Theorem 1.4.11 says that $A = (A \cap B) \cup (A \cap B^c)$. Clearly the two events $A \cap B$ and $A \cap B^c$ are disjoint. It follows from Theorem 1.5.6 that $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c)$.

11. (a) The set of points for which $(x - 1/2)^2 + (y - 1/2)^2 < 1/4$ is the interior of a circle that is contained in the unit square. (Its center is $(1/2, 1/2)$ and its radius is $1/2$.) The area of this circle is $\pi/4$, so the area of the remaining region (what we want) is $1 - \pi/4$.

    (b) We need the area of the region between the two lines $y = 1/2 - x$ and $y = 3/2 - x$. The remaining area is the union of two right triangles with base and height both equal to $1/2$. Each triangle has area $1/8$, so the region between the two lines has area $1 - 2/8 = 3/4$.

    (c) We can use calculus to do this. We want the area under the curve $y = 1 - x^2$ between $x = 0$ and $x = 1$. This equals

    $$\int_0^1 (1 - x^2)dx = x - \left.\frac{x^3}{3}\right|_{x=0}^1 = \frac{2}{3}.$$

    (d) The area of a line is 0, so the probability of a line segment is 0.

12. The events $B_1, B_2, \ldots$ are disjoint, because the event B_1 contains the points in A_1, the event B_2 contains the points in A_2 but not in A_1, the event B_3 contains the points in A_3 but not in A_1 or A_2, etc. By this same reasoning, it is seen that $\cup_{i=1}^n A_i = \cup_{i=1}^n B_i$ and $\cup_{i=1}^\infty A_i = \cup_{i=1}^\infty B_i$. Therefore,

$$\Pr\left(\bigcup_{i=1}^n A_i\right) = \Pr\left(\bigcup_{i=1}^n B_i\right)$$

and

$$\Pr\left(\bigcup_{i=1}^\infty A_i\right) = \Pr\left(\bigcup_{i=1}^\infty B_i\right).$$

However, since the events $B_1, B_2, \ldots$ are disjoint,

$$\Pr\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n Pr(B_i)$$

and

$$\Pr\left(\bigcup_{i=1}^\infty B_i\right) = \sum_{i=1}^\infty \Pr(B_i).$$

13. We know from Exercise 12 that

$$\Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \Pr(B_i).$$

Furthermore, from the definition of the events $B_1, \ldots, B_n$ it is seen that $B_i \subset A_i$ for $i = 1, \ldots, n$. Therefore, by Theorem 1.5.4, $\Pr(B_i) \le \Pr(A_i)$ for $i = 1, \ldots, n$. It now follows that

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \Pr(A_i).$$

(Of course, if the events $A_1, \ldots, A_n$ are disjoint, there is equality in this relation.)

For the second part, apply the first part with A_i replaced by A_i^c for $i = 1, \ldots, n$. We get

$$\Pr\left(\bigcup A_i^c\right) \le \sum_{i=1}^{n} \Pr(A_i^c). \tag{S.1.1}$$

Exercise 5 in Sec. 1.4 says that the left side of (S.1.1) is $\Pr\left(\left[\bigcap A_i\right]^c\right)$. Theorem 1.5.3 says that this last probability is $1 - \Pr\left(\bigcap A_i\right)$. Hence, we can rewrite (S.1.1) as

$$1 - \Pr\left(\bigcap A_i\right) \le \sum_{i=1}^{n} \Pr(A_i^c).$$

Finally take one minus both sides of the above inequality (which reverses the inequality) and produces the desired result.

14. First, note that the probability of type AB blood is $1 - (0.5 + 0.34 + 0.12) = 0.04$ by using Theorems 1.5.2 and 1.5.3.

    (a) The probability of blood reacting to anti-A is the probability that the blood is either type A or type AB. Since these are disjoint events, the probability is the sum of the two probabilities, namely $0.34 + 0.04 = 0.38$. Similarly, the probability of reacting with anti-B is the probability of being either type B or type AB, $0.12 + 0.04 = 0.16$.

    (b) The probability that both antigens react is the probability of type AB blood, namely 0.04.

1.6 Finite Sample Spaces

Solutions to Exercises

1. The safe way to obtain the answer at this stage of our development is to count that 18 of the 36 outcomes in the sample space yield an odd sum. Another way to solve the problem is to note that regardless of what number appears on the first die, there are three numbers on the second die that will yield an odd sum and three numbers that will yield an even sum. Either way the probability is 1/2.

2. The event whose probability we want is the complement of the event in Exercise 1, so the probability is also 1/2.

3. The only differences greater than or equal to 3 that are available are 3, 4 and 5. These large difference only occur for the six outcomes in the upper right and the six outcomes in the lower left of the array in Example 1.6.5 of the text. So the probability we want is $1 - 12/36 = 2/3$.

4. Let x be the proportion of the school in grade 3 (the same as grades 2–6). Then $2x$ is the proportion in grade 1 and $1 = 2x + 5x = 7x$. So $x = 1/7$, which is the probability that a randomly selected student will be in grade 3.

5. The probability of being in an odd-numbered grade is $2x + x + x = 4x = 4/7$.

6. Assume that all eight possible combinations of faces are equally likely. Only two of those combinations have all three faces the same, so the probability is 1/4.

7. The possible genotypes of the offspring are aa and Aa, since one parent will definitely contribute an a, while the other can contribute either A or a. Since the parent who is Aa contributes each possible allele with probability 1/2 each, the probabilities of the two possible offspring are each 1/2 as well.

8. (a) The sample space contains 12 outcomes: (Head, 1), (Tail, 1), (Head, 2), (Tail, 2), etc.
   (b) Assume that all 12 outcomes are equally likely. Three of the outcomes have Head and an odd number, so the probability is 1/4.

1.7 Counting Methods

Commentary

If you wish to stress computer evaluation of probabilities, then there are programs for computing factorials and log-factorials. For example, in the statistical software *R*, there are functions `factorial` and `lfactorial` that compute these. If you cover Stirling's formula (Theorem 1.7.5), you can use these functions to illustrate the closeness of the approximation.

Solutions to Exercises

1. Each pair of starting day and leap year/no leap year designation determines a calendar, and each calendar correspond to exactly one such pair. Since there are seven days and two designations, there are a total of $7 \times 2 = 14$ different calendars.

2. There are 20 ways to choose the student from the first class, and no matter which is chosen, there are 18 ways to choose the student from the second class. No matter which two students are chosen from the first two classes, there are 25 ways to choose the student from the third class. The multiplication rule can be applied to conclude that the total number of ways to choose the three members is $20 \times 18 \times 25 = 9000$.

3. This is a simple matter of permutations of five distinct items, so there are $5! = 120$ ways.

4. There are six different possible shirts, and no matter what shirt is picked, there are four different slacks. So there are 24 different combinations.

5. Let the sample space consist of all four-tuples of dice rolls. There are $6^4 = 1296$ possible outcomes. The outcomes with all four rolls different consist of all of the permutations of six items taken four at a time. There are $P_{6,4} = 360$ of these outcomes. So the probability we want is $360/1296 = 5/18$.

6. With six rolls, there are $6^6 = 46656$ possible outcomes. The outcomes with all different rolls are the permutations of six distinct items. There are $6! = 720$ outcomes in the event of interest, so the probability is $720/46656 = 0.01543$.

7. There are 20^{12} possible outcomes in the sample space. If the 12 balls are to be thrown into different boxes, the first ball can be thrown into any one of the 20 boxes, the second ball can then be thrown into any one of the other 19 boxes, etc. Thus, there are $20 \cdot 19 \cdot 18 \cdots 9$ possible outcomes in the event. So the probability is $20!/[8!20^{12}]$.

8. There are 7^5 possible outcomes in the sample space. If the five passengers are to get off at different floors, the first passenger can get off at any one of the seven floors, the second passenger can then get off at any one of the other six floors, etc. Thus, the probability is

$$\frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{7^5} = \frac{360}{2401}.$$

9. There are 6! possible arrangements in which the six runners can finish the race. If the three runners from team A finish in the first three positions, there are 3! arrangements of these three runners among these three positions and there are also 3! arrangements of the three runners from team B among the last three positions. Therefore, there are $3! \times 3!$ arrangements in which the runners from team A finish in the first three positions and the runners from team B finish in the last three positions. Thus, the probability is $(3!3!)/6! = 1/20$.

10. We can imagine that the 100 balls are randomly ordered in a list, and then drawn in that order. Thus, the required probability in part (a), (b), or (c) of this exercise is simply the probability that the first, fiftieth, or last ball in the list is red. Each of these probabilities is the same $\frac{r}{100}$, because of the random order of the list.

11. In terms of factorials, $P_{n,k} = n!/[k!(n-k)!]$. Since we are assuming that n and $n = k$ are large, we can use Stirling's formula to approximate both of them. The approximation to $n!$ is $(2\pi)^{1/2} n^{n+1/2} e^{-n}$, and the approximation to $(n-k)!$ is $(2\pi)^{1/2}(n-k)^{n-k+1/2}e^{-n+k}$. The approximation to the ratio is the ratio of the approximations because the ratio of each approximation to its corresponding factorial converges to 1. That is,

$$\frac{n!}{k!(n-k)!} \approx \frac{(2\pi)^{1/2} n^{n+1/2} e^{-n}}{k!(2\pi)^{1/2}(n-k)^{n-k+1/2}e^{-n+k}} = \frac{e^{-k}n^k}{k!}\left(1 - \frac{k}{n}\right)^{-n-k-1/2}.$$

Further simplification is available if one assumes that k is small compared to n, that is $k/n \approx 0$. In this case, the last factor is approximately e^k, and the whole approximation simplifies to $n^k/k!$. This makes sense because, if $n/(n-k)$ is essentially 1, then the product of the k largest factors in $n!$ is essentially n^k.

1.8 Combinatorial Methods

Commentary

This section ends with an extended example called "The Tennis Tournament". This is an application of combinatorics that uses a slightly subtle line of reasoning.

Solutions to Exercises

1. We have to assign 10 houses to one pollster, and the other pollster will get to canvas the other 10 houses. Hence, the number of assignments is the number of combinations of 20 items taken 10 at a time,

$$\binom{20}{10} = 184{,}756.$$

2. The ratio of $\binom{93}{30}$ to $\binom{93}{31}$ is $31/63 < 1$, so $\binom{93}{31}$ is larger.

3. Since $93 = 63 + 30$, the two numbers are the same.

4. Let the sample space consist of all subsets (not ordered tuples) of the 24 bulbs in the box. There are $\binom{24}{4} = 10626$ such subsets. There is only one subset that has all four defectives, so the probability we want is $1/10626$.

5. The number is $\dfrac{4251!}{(97!4154!)} = \dbinom{4251}{97}$, an integer.

6. There are $\binom{n}{2}$ possible pairs of seats that A and B can occupy. Of these pairs, $n-1$ pairs comprise two adjacent seats. Therefore, the probability is $\dfrac{n-1}{\binom{n}{2}} = \dfrac{2}{n}$.

7. There are $\binom{n}{k}$ possible sets of k seats to be occupied, and they are all equally likely. There are $n-k+1$ sets of k adjacent seats, so the probability we want is

$$\frac{n-k+1}{\binom{n}{k}} = \frac{(n-k+1)!k!}{n!}.$$

8. There are $\binom{n}{k}$ possible sets of k seats to be occupied, and they are all equally likely. Because the circle has no start or end, there are n sets of k adjacent seats, so the probability we want is

$$\frac{n}{\binom{n}{k}} = \frac{(n-k)!k!}{(n-1)!}.$$

9. This problem is slightly tricky. The total number of ways of choosing the n seats that will be occupied by the n people is $\binom{2n}{n}$. Offhand, it would seem that there are only two ways of choosing these seats so that no two adjacent seats are occupied, namely:

$$\text{X0X0}\ldots\text{0} \quad \text{and} \quad \text{0X0X}\ldots\text{0X}$$

Upon further consideration, however, $n-1$ more ways can be found, namely:

$$\text{X00X0X}\ldots\text{0X}, \quad \text{X0X00X0X}\ldots\text{0X, etc.}$$

Therefore, the total number of ways of choosing the seats so that no two adjacent seats are occupied is $n+1$. The probability is $(n+1)/\binom{2n}{n}$.

10. We shall let the sample space consist of all subsets (unordered) of 10 out of the 24 light bulbs in the box. There are $\binom{24}{10}$ such subsets. The number of subsets that contain the two defective bulbs is the number of subsets of size 8 out of the other 22 bulbs, $\binom{22}{8}$, so the probability we want is

$$\frac{\binom{22}{8}}{\binom{24}{10}} = \frac{10 \times 9}{24 \times 23} = 0.1630.$$

11. This exercise is similar to Exercise 10. Let the sample space consist of all subsets (unordered) of 12 out of the 100 people in the group. There are $\binom{100}{12}$ such subsets. The number of subsets that contain A and B is the number of subsets of size 10 out of the other 98 people, $\binom{98}{10}$, so the probability we want is

$$\frac{\binom{98}{10}}{\binom{100}{12}} = \frac{12 \times 11}{100 \times 99} = 0.01333.$$

12. There are $\binom{35}{10}$ ways of dividing the group into the two teams. As in Exercise 11, the number of ways of choosing the 10 players for the first team so as to include both A and B is $\binom{33}{8}$. The number of ways of choosing the 10 players for this team so as not to include either A or B (A and B will then be together on the other team) is $\binom{33}{10}$. The probability we want is then

$$\frac{\binom{33}{8} + \binom{33}{10}}{\binom{35}{10}} = \frac{10 \times 9 + 25 \times 24}{35 \times 34} = 0.5798.$$

13. This exercise is similar to Exercise 12. Here, we want four designated bulbs to be in the same group. The probability is

$$\frac{\binom{20}{6} + \binom{20}{10}}{\binom{24}{10}} = 0.1140.$$

14.

$$\begin{aligned}\binom{n}{k}+\binom{n}{k-1} &= \frac{n!}{k!(n-k)!}+\frac{n!}{(k-1)!(n-k+1)!}\\ &= \frac{n!}{(k-1)!(n-k)!}\left(\frac{1}{k}+\frac{1}{n-k+1}\right)\\ &= \frac{n!}{(k-1)!(n-k)!}\cdot\frac{n+1}{k(n-k+1)}\\ &= \frac{(n+1)!}{k!(n-k+1)!}=\binom{n+1}{k}.\end{aligned}$$

15. (a) If we express 2^n as $(1+1)^n$ and expand $(1+1)^n$ by the binomial theorem, we obtain the desired result.

(b) If we express 0 as $(1-1)^n$ and expand $(1-1)^n$ by the binomial theorem, we obtain the desired result.

16. (a) It is easier to calculate first the probability that the committee will not contain either of the two senators from the designated state. This probability is $\binom{98}{8}\Big/\binom{100}{8}$. Thus, the final answer is

$$1-\frac{\binom{98}{8}}{\binom{100}{8}}\approx 1-.08546=0.1543.$$

(b) There are $\binom{100}{50}$ combinations that might be chosen. If the group is to contain one senator from each state, then there are two possible choices for each of the fifty states. Hence, the number of possible combinations containing one senator from each state is 2^{50}.

17. Call the four players A, B, C, and D. The number of ways of choosing the positions in the deck that will be occupied by the four aces is $\binom{52}{4}$. Since player A will receive 13 cards, the number of ways of choosing the positions in the deck for the four aces so that all of them will be received by player A is $\binom{13}{4}$. Similarly, since player B will receive 13 other cards, the number of ways of choosing the positions for the four aces so that all of them will be received by player B is $\binom{13}{4}$. A similar result is true for each of the other players. Therefore, the total number of ways of choosing the positions in the deck for the four aces so that all of them will be received by the same player is $4\binom{13}{4}$. Thus, the final probability is $4\binom{13}{4}\Big/\binom{52}{4}$.

18. There are $\binom{100}{10}$ ways of choosing ten mathematics students. There are $\binom{20}{2}$ ways of choosing two

students from a given class of 20 students. Therefore, there are $\binom{20}{2}^5$ ways of choosing two students from each of the five classes. So, the final answer is $\binom{20}{2}^5 / \binom{100}{10} \approx 0.0143$.

19. From the description of what counts as a collection of customer choices, we see that each collection consists of a tuple $(m_1, \ldots, m_n)$, where m_i is the number of customers who choose item i for $i = 1, \ldots, n$. Each m_i must be between 0 and k and $m_1 + \cdots + m_n = k$. Each such tuple is equivalent to a sequence of $n + k - 1$ 0's and 1's as follows. The first m_1 terms are 0 followed by a 1. The next m_2 terms are 0 followed by a 1, and so on up to m_{n-1} 0's followed by a 1 and finally m_n 0's. Since $m_1 + \cdots + m_n = k$ and since we are putting exactly $n - 1$ 1's into the sequence, each such sequence has exactly $n + k - 1$ terms. Also, it is clear that each such sequence corresponds to exactly one tuple of customer choices. The numbers of 0's between successive 1's give the numbers of customers who choose that item, and the 1's indicate where we switch from one item to the next. So, the number of combinations of choices is the number of such sequences: $\binom{n+k-1}{k}$.

20. We shall use induction. For $n = 1$, we must prove that

$$x + y = \binom{1}{0} x^0 y^1 + \binom{1}{1} x^1 y^0.$$

Since the right side of this equation is $x + y$, the theorem is true for $n = 1$. Now assume that the theorem is true for each $n = 1, \ldots, n_0$ for $n_0 \geq 1$. For $n = n_0 + 1$, the theorem says

$$(x + y)^{n_0+1} = \sum_{k=0}^{n_0+1} \binom{n_0+1}{k} x^k y^{n_0+1-k}. \tag{S.1.2}$$

Since we have assumed that the theorem is true for $n = n_0$, we know that

$$(x + y)^{n_0} = \sum_{k=0}^{n_0} \binom{n_0}{k} x^k y^{n_0-k}. \tag{S.1.3}$$

We shall multiply both sides of (S.1.3) by $x + y$. We then need to prove that $x + y$ times the right side of (S.1.3) equals the right side of (S.1.2).

$$\begin{aligned}
(x+y)(x+y)^{n_0} &= (x+y)\sum_{k=0}^{n_0} \binom{n_0}{k} x^k y^{n_0-k} \\
&= \sum_{k=0}^{n_0} \binom{n_0}{k} x^{k+1} y^{n_0-k} + \sum_{k=0}^{n_0} \binom{n_0}{k} x^k y^{n_0+1-k} \\
&= \sum_{k=1}^{n_0+1} \binom{n_0}{k-1} x^k y^{n_0+1-k} + \sum_{k=0}^{n_0} \binom{n_0}{k} x^k y^{n_0+1-k} \\
&= y^{n_0+1} + \sum_{k=1}^{n_0} \left[\binom{n_0}{k-1} + \binom{n_0}{k} \right] x^k y^{n_0+1-k} + x^{n_0+1}.
\end{aligned}$$

Now, apply the result in Exercise 14 to conclude that

$$\binom{n_0}{k-1} + \binom{n_0}{k} = \binom{n_0+1}{k}.$$

This makes the final summation above equal to the right side of (S.1.2).

21. We are asked for the number of unordered samples with replacement, as constructed in Exercise 19. Here, $n = 365$, so there are $\binom{365+k}{k}$ different unordered sets of k birthdays chosen with replacement from $1, \dots, 365$.

22. The approximation to $n!$ is $(2\pi)^{1/2} n^{n+1/2} e^{-n}$, and the approximation to $(n/2)!$ is $(2\pi)^{1/2}(n/2)^{n/2+1/2}e^{-n/2}$. Then

$$\frac{n!}{(n/2)!^2} \approx \frac{(2\pi)^{1/2} n^{n+1/2} e^{-n}}{[(2\pi)^{1/2}(n/2)^{n/2+1/2}e^{-n/2}]^2} = (2\pi)^{-1/2} 2^{n+1} n^{-1/2}.$$

With $n = 500$, the approximation is $e^{343.24}$, too large to represent on a calculator with only two-digit exponents. The actual number is about 1/20 of 1% larger.

1.9 Multinomial Coefficients

Commentary

Multinomial coefficients are useful as a counting method, and they are needed for the definition of the multinomial distribution in Sec. 5.9. They are not used much elsewhere in the text. Although this section does not have an asterisk, it could be skipped (together with Sec. 5.9) if one were not interested in the multinomial distribution or the types of counting arguments that rely on multinomial coefficients.

Solutions to Exercises

1. We have three types of elements that need to be assigned to 21 houses so that exactly seven of each type are assigned. The number of ways to do this is the multinomial coefficient

$$\binom{21}{7,7,7} = 399{,}072{,}960.$$

2. We are asked for the number of arrangements of four distinct types of objects with 18 or one type, 12 of the next, 8 of the next and 12 of the last. This is the multinomial coefficient $\binom{50}{18,12,8,12}$.

3. We need to divide the 300 members of the organization into three subsets: the 5 in one committee, the 8 in the second committee, and the 287 in neither committee. There are $\binom{300}{5,8,287}$ ways to do this.

4. There are $\binom{10}{3,3,2,1,1}$ arrangements of the 10 letters of four distinct types. All of them are equally likely, and only one spells statistics. So, the probability is $1/\binom{10}{3,3,2,1,1} = 1/50400$.

5. There are $\binom{n}{n_1,n_2,n_3,n_4,n_5,n_6}$ many ways to arrange n_j j's (for $j = 1, \dots, 6$) among the n rolls. The number of possible equally likely rolls is 6^n. So, the probability we want is $\frac{1}{6^n}\binom{n}{n_1,n_2,n_3,n_4,n_5,n_6}$.

6. There are 6^7 possible outcomes for the seven dice. If each of the six numbers is to appear at least once among the seven dice, then one number must appear twice and each of the other five numbers must appear once. Suppose first that the number 1 appears twice and each of the other numbers appears once. The number of outcomes of this type in the sample space is equal to the number of different arrangements of the symbols 1, 1, 2, 3, 4, 5, 6, which is $\frac{7!}{2!(1!)^5} = \frac{7!}{2}$. There is an equal number of outcomes for each of the other five numbers which might appear twice among the seven dice. Therefore, the total number of outcomes in which each number appears at least once is $\frac{6(7!)}{2}$, and the probability of this event is

$$\frac{6(7!)}{(2)6^7} = \frac{7!}{2(6^6)}.$$

7. There are $\binom{25}{10,8,7}$ ways of distributing the 25 cards to the three players. There are $\binom{12}{6,2,4}$ ways of distributing the 12 red cards to the players so that each receives the designated number of red cards. There are then $\binom{13}{4,6,3}$ ways of distributing the other 13 cards to the players, so that each receives the designated total number of cards. The product of these last two numbers of ways is, therefore, the number of ways of distributing the 25 cards to the players so that each receives the designated number of red cards and the designated total number of cards. So, the final probability is $\binom{12}{6,2,4}\binom{13}{4,6,3}\Big/\binom{25}{10,8,7}$.

8. There are $\binom{52}{13,13,13,13}$ ways of distributing the cards to the four players. There are $\binom{12}{3,3,3,3}$ ways of distributing the 12 picture cards so that each player gets three. No matter which of these ways we choose, there are $\binom{40}{10,10,10,10}$ ways to distribute the remaining 40 nonpicture cards so that each player gets 10. So, the probability we need is

$$\frac{\binom{12}{3,3,3,3}\binom{40}{10,10,10,10}}{\binom{52}{13,13,13,13}} = \frac{\frac{12!}{(3!)^4}\frac{40!}{(10!)^4}}{\frac{52!}{(13!)^4}} \approx 0.0324.$$

9. There are $\binom{52}{13,13,13,13}$ ways of distributing the cards to the four players. Call these four players A, B, C, and D. There is only one way of distributing the cards so that player A receives all red cards, player B receives all yellow cards, player C receives all blue cards, and player D receives all green cards. However, there are 4! ways of assigning the four colors to the four players and therefore there are 4! ways of distributing the cards so that each player receives 13 cards of the same color. So, the probability we need is

$$\frac{4!}{\binom{52}{13,13,13,13}} = \frac{4!(13!)^4}{52!} \approx 4.474 \times 10^{-28}.$$

10. If we do not distinguish among boys with the same last name, then there are $\binom{9}{2,3,4}$ possible arrangements of the nine boys. We are interested in the probability of a particular one of these arrangements. So, the probability we need is

$$\frac{1}{\binom{9}{2,3,4}} = \frac{2!3!4!}{9!} \approx 7.937 \times 10^{-4}.$$

11. We shall use induction. Since we have already proven the binomial theorem, we know that the conclusion to the multinomial theorem is true for every n if $k = 2$. We shall use induction again, but this time using k instead of n. For $k = 2$, we already know the result is true. Suppose that the result is true for all $k \le k_0$ and for all n. For $k = k_0 + 1$ and arbitrary n we must show that

$$(x_1 + \cdots + x_{k_0+1})^n = \sum \binom{n}{n_1, \ldots, n_{k_0+1}} x_1^{n_1} \cdots x_{k_0+1}^{n_{k_0+1}}, \tag{S.1.4}$$

where the summation is over all $n_1, \ldots, n_{k_0+1}$ such that $n_1 + \cdots + n_{k_0+1} = n$. Let $y_i = x_i$ for $i = 1, \ldots, k_0 - 1$ and let $y_{k_0} = x_{k_0} + x_{k_0+1}$. We then have

$$(x_1 + \cdots + x_{k_0+1})^n = (y_1 + \cdots + y_{k_0})^n.$$

Since we have assumed that the theorem is true for $k = k_0$, we know that

$$(y_1 + \cdots + y_{k_0})^n = \sum \binom{n}{m_1, \ldots, m_{k_0}} y_1^{m_1} \cdots y_{k_0}^{m_{k_0}}, \tag{S.1.5}$$

where the summation is over all $m_1, \ldots, m_{k_0}$ such that $m_1 + \cdots + m_{k_0} = n$. On the right side of (S.1.5), substitute $x_{k_0} + x_{k_0+1}$ for y_{k_0} and apply the binomial theorem to obtain

$$\sum \binom{n}{m_1, \ldots, m_{k_0}} y_1^{m_1} \cdots y_{k_0-1}^{m_{k_0-1}} \sum_{i=0}^{m_{k_0}} \binom{m_{k_0}}{i} x_{k_0}^i x_{k_0+1}^{m_{k_0}-i}. \tag{S.1.6}$$

In (S.1.6), let $n_i = m_i$ for $i = 1, \ldots, k_0 - 1$, let $n_{k_0} = i$, and let $n_{k_0+1} = m_{k_0} - i$. Then, in the summation in (S.1.6), $n_1 + \cdots + n_{k_0+1} = n$ if and only if $m_1 + \cdots + m_{k_0} = n$. Also, note that

$$\binom{n}{m_1, \ldots, m_{k_0}} \binom{m_{k_0}}{i} = \binom{n}{n_1, \ldots, n_{k_0+1}}.$$

So, (S.1.6) becomes

$$\sum \binom{n}{n_1, \ldots, n_{k_0+1}} x_1^{n_1} \cdots x_{k_0+1}^{n_{k_0+1}},$$

where this last sum is over all $n_1, \ldots, n_{k_0+1}$ such that $n_1 + \cdots + n_{k_0+1} = n$.

12. For each element s' of S', the elements of S that lead to boxful s' are all the different sequences of elements of s'. That is, think of each s' as an unordered set of 12 numbers chosen with replacement from 1 to 7. For example, $\{1,1,2,3,3,3,5,6,7,7,7,7\}$ is one such set. The following are some of the elements of S lead to the same set s': $(1,1,2,3,3,3,5,6,7,7,7,7)$, $(1,2,3,5,6,7,1,3,7,3,7,7)$, $(7,1,7,2,3,5,7,1,6,3,7,3)$. This problem is pretty much the same as that which leads to the definition of multinomial coefficients. We are looking for the number of orderings of 12 digits chosen from the numbers 1 to 7 that have two of 1, one of 2, three of 3, none of 4, one of 5, one of 6, and four of 7. This is just $\binom{12}{1,1,3,0,1,1,4}$. For a general s', for $i = 1, \ldots, 7$, let $n_i(s')$ be the number of i's in the box s'. Then $n_1(s') + \cdots + n_7(s') = 12$, and the number of orderings of these numbers is

$$N(s') = \binom{12}{n_1(s'), n_2(s'), \ldots, n_7(s')}.$$

The multinomial theorem tells us that

$$\sum_{\text{All } s'} N(s') = \sum \binom{12}{n_1, n_2, \ldots, n_7} 1^{n_1} \cdots 1^{n_7} = 7^{12},$$

where the sum is over all possible combinations of nonnegative integers $n_1, \ldots, n_7$ that add to 12. This matches the number of outcomes in S.

1.10 The Probability of a Union of Events

Commentary

This section ends with an example of the matching problem. This is an application of the formula for the probability of a union of an arbitrary number of events. It requires a long line of argument and contains an interesting limiting result. The example will be interesting to students with good mathematics backgrounds, but it might be too challenging for students who have struggled to master combinatorics. One can use statistical software, such as R, to help illustrate how close the approximation is. The formula (1.10.10) can be computed as

```
ints=1:n
sum(exp(-lfactorial(ints))*(-1)^(ints+1)),
```

where `n` has previously been assigned the value of n for which one wishes to compute p_n.

Solutions to Exercises

1. Let A_i be the event that person i receives exactly two aces for $i = 1, 2, 3$. We want $\Pr(\cup_{i=1}^{3} A_i)$. We shall apply Theorem 1.10.1 directly. Let the sample space consist of all permutations of the 52 cards where the first five cards are dealt to person 1, the second five to person 2, and the third five to person 3. A permutation of 52 cards that leads to the occurrence of event A_i can be constructed as follows. First, choose which of person i's five locations will receive the two aces. There are $C_{5,2}$ ways to do this. Next, for each such choice, choose the two aces that will go in these locations, distinguishing the order in which they are placed. There are $P_{4,2}$ ways to do this. Next, for each of the preceding choices, choose the locations for the other two aces from among the 47 locations that are not dealt to person i, distinguishing order. There are $P_{47,2}$ ways to do this. Finally, for each of the preceding choices, choose a permutation of the remaining 48 cards among the remaining 48 locations. There are 48! ways to do this. Since there are 52! equally likely permutations in the sample space, we have

$$\Pr(A_i) = \frac{C_{5,2} P_{4,2} P_{47,2} 48!}{52!} = \frac{5!4!47!48!}{2!3!2!45!52!} \approx 0.0399.$$

Careful examination of the expression for $\Pr(A_i)$ reveals that it can also be expressed as

$$\Pr(A_i) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}.$$

This expression corresponds to a different, but equally correct, way of describing the sample space in terms of equally likely outcomes. In particular, the sample space would consist of the different possible five-card sets that person i could receive without regard to order.

Next, compute $\Pr(A_iA_j)$ for $i \neq j$. There are still $C_{5,2}$ ways to choose the locations for person i's aces amongst the five cards and for each such choice, there are $P_{4,2}$ ways to choose the two aces in order. For each of the preceding choices, there are $C_{5,2}$ ways to choose the locations for person j's aces and 2 ways to order the remaining two aces amongst the two locations. For each combination of the preceding choices, there are 48! ways to arrange the remaining 48 cards in the 48 unassigned locations. Then, $\Pr(A_iA_j)$ is

$$\Pr(A_iA_j) = \frac{2C_{5,2}^2 P_{4,2} 48!}{52!} = \frac{2(5!)^2 4! 48!}{(2!)^3(3!)^2 52!} \approx 3.694 \times 10^{-4}.$$

Once again, we can rewrite the expression for $\Pr(A_iA_j)$ as

$$\Pr(A_iA_j) = \frac{\binom{4}{2}\binom{48}{3,3,42}}{\binom{52}{5,5,42}}.$$

This corresponds to treating the sample space as the set of all pairs of five-card subsets.

Next, notice that it is impossible for all three players to receive two aces, so $\Pr(A_1A_2A_3) = 0$. Applying Theorem 1.10.1, we obtain

$$\Pr\left(\cup_{i=1}^3 A_i\right) = 3 \times 0.0399 - 3 \times 3.694 \times 10^{-4} = 0.1186.$$

2. Let A, B, and C stand for the events that a randomly selected family subscribes to the newspaper with the same name. Then $\Pr(A \cup B \cup C)$ is the proportion of families that subscribe to at least one newspaper. According to Theorem 1.10.1, we can express this probability as

$$\Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(AC) - \Pr(BC) + \Pr(A \cap BC).$$

The probabilities in this expression are the proportions of families that subscribe to the various combinations. These proportions are all stated in the exercise, so the formula yields

$$\Pr(A \cup B \cup C) = 0.6 + 0.4 + 0.3 - 0.2 - 0.1 - 0.2 + 0.05 = 0.85.$$

3. As seen from Fig. S.1.1, the required percentage is $P_1 + P_2 + P_3$. From the given values, we have, in percentages,

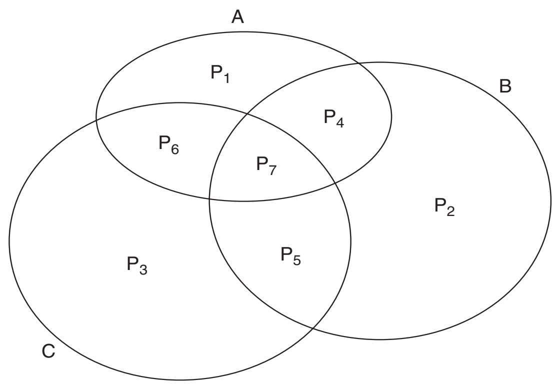

Figure S.1.1: Figure for Exercise 3 of Sec. 1.10.

$$\begin{aligned}
P_7 &= 5,\\
P_4 &= 20 - P_7 = 15,\\
P_5 &= 20 - P_7 = 15,\\
P_6 &= 10 - P_7 = 5,\\
P_1 &= 60 - P_4 - P_6 - P_7 = 35,\\
P_2 &= 40 - P_4 - P_5 - P_7 = 5,\\
P_3 &= 30 - P_5 - P_6 - P_7 = 5.
\end{aligned}$$

Therefore, $P_1 + P_2 + P_3 = 45$.

4. This is a case of the matching problem with $n = 3$. We are asked to find p_3. By Eq. (1.10.10) in the text, this equals

$$p_3 = 1 - \frac{1}{2} + \frac{1}{6} = \frac{2}{3}.$$

5. Determine first the probability that at least one guest will receive the proper hat. This probability is the value p_n specified in the matching problem, with $n = 4$, namely

$$p_4 = 1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24} = \frac{5}{8}.$$

So, the probability that no guest receives the proper hat is $1 - 5/8 = 3/8$.

6. Let A_1 denote the event that no red balls are selected, let A_2 denote the event that no white balls are selected, and let A_3 denote the event that no blue balls are selected. The desired probability is $\Pr(A_1 \cup A_2 \cup A_3)$ and we shall apply Theorem 1.10.1. The event A_1 will occur if and only if the ten selected balls are either white or blue. Since there are 60 white and blue balls, out of a total of 90 balls, we have $\Pr(A_1) = \binom{60}{10} \Big/ \binom{90}{10}$. Similarly, $\Pr(A_2)$ and $\Pr(A_3)$ have the same value. The event A_1A_2 will occur if and only if all ten selected balls are blue. Therefore, $\Pr(A_1A_2) = \binom{30}{10} \Big/ \binom{90}{10}$. Similarly, $\Pr(A_2A_3)$ and $\Pr(A_1A_3)$ have the same value. Finally, the event $A_1A_2A_3$ will occur if and only if all three colors are missing, which is obviously impossible. Therefore, $\Pr(A_1A_2A_3) = 0$. When these values

are substituted into Eq. (1.10.1), we obtain the desired probability,

$$\Pr(A_1 \cup A_2 \cup A_3) = 3\frac{\binom{60}{10}}{\binom{90}{10}} - 3\frac{\binom{30}{10}}{\binom{90}{10}}.$$

7. Let A_1 denote the event that no student from the freshman class is selected, and let A_2, A_3, and A_4 denote the corresponding events for the sophomore, junior, and senior classes, respectively. The probability that at least one student will be selected from each of the four classes is equal to $1 - \Pr(A_1 \cup A_2 \cup A_3 \cup A_4)$. We shall evaluate $\Pr(A_1 \cup A_2 \cup A_3 \cup A_4)$ by applying Theorem 1.10.2. The event A_1 will occur if and only if the 15 selected students are sophomores, juniors, or seniors. Since there are 90 such students out of a total of 100 students, we have $\Pr(A_1) = \binom{90}{15}/\binom{100}{15}$. The values of $\Pr(A_i)$ for $i = 2, 3, 4$ can be obtained in a similar fashion. Next, the event A_1A_2 will occur if and only if the 15 selected students are juniors or seniors. Since there are a total of 70 juniors and seniors, we have $\Pr(A_1A_2) = \binom{70}{15}/\binom{100}{15}$. The probability of each of the six events of the form A_iA_j for $i < j$ can be obtained in this way. Next the event $A_1A_2A_3$ will occur if and only if all 15 selected students are seniors. Therefore, $\Pr(A_1A_2A_3) = \binom{40}{15}/\binom{100}{15}$. The probabilities of the events $A_1A_2A_4$ and $A_1A_3A_4$ can also be obtained in this way. It should be noted, however, that $\Pr(A_2A_3A_4) = 0$ since it is impossible that all 15 selected students will be freshmen. Finally, the event $A_1A_2A_3A_4$ is also obviously impossible, so $\Pr(A_1A_2A_3A_4) = 0$. So, the probability we want is

$$1 - \left[\frac{\binom{90}{15}}{\binom{100}{15}} + \frac{\binom{80}{15}}{\binom{100}{15}} + \frac{\binom{70}{15}}{\binom{100}{15}} + \frac{\binom{60}{15}}{\binom{100}{15}} - \frac{\binom{70}{15}}{\binom{100}{15}} - \frac{\binom{60}{15}}{\binom{100}{15}} - \frac{\binom{50}{15}}{\binom{100}{15}} - \frac{\binom{50}{15}}{\binom{100}{15}} - \frac{\binom{40}{15}}{\binom{100}{15}} - \frac{\binom{30}{15}}{\binom{100}{15}} + \frac{\binom{40}{15}}{\binom{100}{15}} + \frac{\binom{30}{15}}{\binom{100}{15}} + \frac{\binom{20}{15}}{\binom{100}{15}}\right].$$

8. It is impossible to place exactly $n-1$ letters in the correct envelopes, because if $n-1$ letters are placed correctly, then the nth letter must also be placed correctly.

9. Let $p_n = 1 - q_n$. As discussed in the text, $p_{10} < p_{300} < 0.63212 < p_{53} < p_{21}$. Since p_n is smallest for $n = 10$, then q_n is largest for $n = 10$.

10. There is exactly one outcome in which only letter 1 is placed in the correct envelope, namely the outcome in which letter 1 is correctly placed, letter 2 is placed in envelope 3, and letter 3 is placed in envelope 2. Similarly there is exactly one outcome in which only letter 2 is placed correctly, and one in which only letter 3 is placed correctly. Hence, of the $3! = 6$ possible outcomes, 3 outcomes yield the result that exactly one letter is placed correctly. So, the probability is $3/6 = 1/2$.

11. Consider choosing 5 envelopes at random into which the 5 red letters will be placed. If there are exactly r red envelopes among the five selected envelopes ($r = 0, 1, \ldots, 5$), then exactly $x = 2r$ envelopes will

contain a card with a matching color. Hence, the only possible values of x are 0, 2, 4..., 10. Thus, for $x = 0, 2, \ldots, 10$ and $r = x/2$, the desired probability is the probability that there are exactly r red envelopes among the five selected envelopes, which is $\dfrac{\binom{5}{r}\binom{5}{5-r}}{\binom{10}{5}}$.

12. It was shown in the solution of Exercise 12 of Sec. 1.5. that

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \Pr(B_i) = \lim_{n\to\infty} \Pr\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \Pr\left(\bigcup_{i=1}^{n} A_i\right).$$

However, since $A_1 \subset A_2 \subset \ldots \subset A_n$, it follows that $\bigcup_{i=1}^{n} A_i = A_n$. Hence,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \Pr(A_n).$$

13. We know that

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c.$$

Hence,

$$\Pr\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \Pr\left(\bigcup_{i=1}^{\infty} A_i^c\right).$$

However, since $A_1 \supset A_2 \supset \ldots$, then $A_1^c \subset A_2^c \subset \ldots$. Therefore, by Exercise 12,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i^c\right) = \lim_{n\to\infty} \Pr(A_n^c) = \lim_{n\to\infty} [1 - \Pr(A_n)] = 1 - \lim_{n\to\infty} Pr(A_n).$$

It now follows that

$$\Pr\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \Pr(A_n).$$

1.12 Supplementary Exercises

Solutions to Exercises

1. No, since both A and B might occur.

2. $\Pr(A^c \cap B^c \cap D^c) = \Pr[(A \cup B \cup D)^c] = 0.3$.

3. $\dfrac{\binom{250}{18} \cdot \binom{100}{12}}{\binom{350}{30}}$.

4. There are $\binom{20}{10}$ ways of choosing 10 cards from the deck. For $j = 1, \ldots, 5$, there $\binom{4}{2}$ ways of choosing two cards with the number j. Hence, the answer is

$$\frac{\binom{4}{2} \cdots \binom{4}{2}}{\binom{20}{10}} = \frac{6^5}{\binom{20}{10}} \approx 0.0421.$$

5. The region where total utility demand is at least 215 is shaded in Fig. S.1.2. The area of the shaded

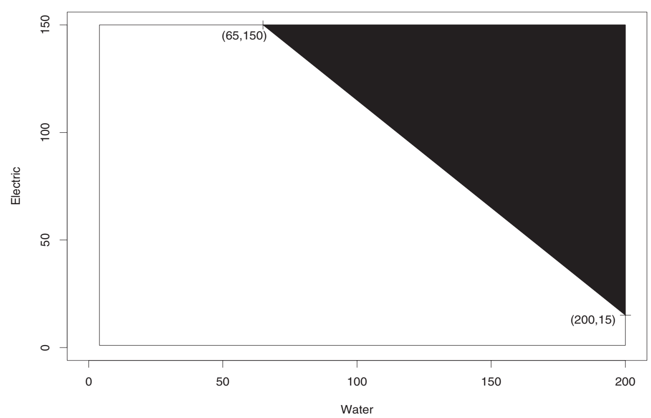

Figure S.1.2: Region where total utility demand is at least 215 in Exercise 5 of Sec. 1.12.

region is

$$\frac{1}{2} \times 135 \times 135 = 9112.5$$

The probability is then $9112.5/29204 = 0.3120$.

6. (a) There are $\binom{r+w}{r}$ possible positions that the red balls could occupy in the ordering as they are drawn. Therefore, the probability that they will be in the first r positions is $1/\binom{r+w}{r}$.

   (b) There are $\binom{r+1}{r}$ ways that the red balls can occupy the first $r+1$ positions in the ordering. Therefore, the probability is $\binom{r+1}{r}/\binom{r+w}{r} = (r+1)/\binom{r+w}{r}$.

7. The presence of the blue balls is irrelevant in this problem, since whenever a blue ball is drawn it is ignored. Hence, the answer is the same as in part (a) of Exercise 6.

8. There are $\binom{10}{7}$ ways of choosing the seven envelopes into which the red cards will be placed. There

are $\binom{7}{j}\binom{3}{7-j}$ ways of choosing exactly j red envelopes and $7-j$ green envelopes. Therefore, the probability that exactly j red envelopes will contain red cards is

$$\binom{7}{j}\binom{3}{7-j}\Big/\binom{10}{7} \qquad \text{for} \qquad j = 4, 5, 6, 7.$$

But if j red envelopes contain red cards, then $j-4$ green envelopes must also contain green cards. Hence, this is also the probability of exactly $k = j + (j-4) = 2j-4$ matches.

9. There are $\binom{10}{5}$ ways of choosing the five envelopes into which the red cards will be placed. There are $\binom{7}{j}\binom{3}{5-j}$ ways of choosing exactly j red envelopes and $5-j$ green envelopes. Therefore the probability that exactly j red envelopes will contain red cards is

$$\binom{7}{j}\binom{3}{5-j}\Big/\binom{10}{5} \qquad \text{for } j = 2, 3, 4, 5.$$

But if j red envelopes contain red cards, then $j-2$ green envelopes must also contain green cards. Hence, this is also the probability of exactly $k = j + (j-2) = 2j-2$ matches.

10. If there is a point x that belongs to neither A nor B, then x belongs to both A^c and B^c. Hence, A^c and B^c are not disjoint. Therefore, A^c and B^c will be disjoint if and only if $A \cup B = S$.

11. We can use Fig. S.1.1 by relabeling the events A, B, and C in the figure as A_1, A_2, and A_3 respectively. It is now easy to see that the probability that exactly one of the three events occurs is $p_1 + p_2 + p_3$. Also,

$$\begin{aligned} \Pr(A_1) &= p_1 + p_4 + p_6 + p_7, \\ \Pr(A_1 \cap A_2) &= p_4 + p_7, \text{ etc.} \end{aligned}$$

By breaking down each probability in the given expression in this way, we obtain the desired result.

12. The proof can be done in a manner similar to that of Theorem 1.10.2. Here is an alternative argument. Consider first a point that belongs to exactly one of the events $A_1, \dots, A_n$. Then this point will be counted in exactly one of the $\Pr(A_i)$ terms in the given expression, and in none of the intersections. Hence, it will be counted exactly once in the given expression, as required. Now consider a point that belongs to exactly r of the events $A_1, \dots, A_n (r \geq 2)$. Then it will be counted in exactly r of the $\Pr(A_i)$ terms, exactly $\binom{r}{2}$ of the $\Pr(A_iA_j)$ terms, exactly $\binom{r}{3}$ of the $\Pr(A_iA_jA_k)$ terms, etc. Hence, in the given expression it will be counted the following number of times:

$$\begin{aligned} r &- 2\binom{r}{2} + 3\binom{r}{3} - \cdots \pm r\binom{r}{r} \\ &= r\left[\binom{r-1}{0} - \binom{r-1}{1} + \binom{r-1}{2} - \cdots \pm \binom{r-1}{r-1}\right] = 0, \end{aligned}$$

by Exercise b of Sec. 1.8. Hence, a point will be counted in the given expression if and only if it belongs to exactly one of the events $A_1, \dots, A_n$, and then it will be counted exactly once.

13. (a) In order for the winning combination to have no consecutive numbers, between every pair of numbers in the winning combination there must be at least one number not in the winning combination. That is, there must be at least $k-1$ numbers not in the winning combination to be in between the pairs of numbers in the winning combination. Since there are k numbers in the winning combination, there must be at least $k+k-1=2k-1$ numbers available in order for it to be possible to have no consecutive numbers in the winning combination. So, n must be at least $2k-1$ to allow consecutive numbers.

(b) Let $i_1,\ldots,i_k$ and $j_1,\ldots,j_k$ be as described in the problem. For one direction, suppose that $i_1,\ldots,i_k$ contains at least one pair of consecutive integers, say $i_{a+1}=i_a+1$. Then

$$j_{a+1}=i_{a+1}-a=i_a+1-a=i_a-(a-1)=j_a.$$

So, $j_1,\ldots,j_k$ contains repeats. For the other direction, suppose that $j_1,\ldots,j_k$ contains repeats, say $j_{a+1}=j_a$. Then

$$i_{a+1}=j_{a+1}+a=j_a+a=i_a+1.$$

So $i_1,\ldots,i_k$ contains a pair of consecutive numbers.

(c) Since $i_1<i_2<\cdots<i_k$, we know that $i_a+1\le i_{a+1}$, so that $j_a=i_a-a+1\le i_{a+1}-a=j_{a+1}$ for each $a=1,\ldots,k-1$. Since $i_k\le n$, $j_k=i_k-k+1\le n-k+1$. The set of all $(j_1,\ldots,j_k)$ with $1\le j_1<\cdots<j_k\le n-k+1$ is just the number of combinations of $n-k+1$ items taken k at a time, that is $\binom{n-k+1}{k}$.

(d) By part (b), there are no pairs of consecutive integers in the winning combination $(i_1,\ldots,i_k)$ if and only if $(j_1,\ldots,j_k)$ has no repeats. The total number of winning combinations is $\binom{n}{k}$. In part (c), we computed the number of winning combinations with no repeats among $(j_1,\ldots,j_k)$ to be $\binom{n-k+1}{k}$. So, the probability of no consecutive integers is

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

(e) The probability of at least one pair of consecutive integers is one minus the answer to part (d).

Chapter 2

Conditional Probability

2.1 The Definition of Conditional Probability

Commentary

It is useful to stress the point raised in the note on page 59. That is, conditional probabilities behave just like probabilities. This will come up again in Sec. 3.6 where conditional distributions are introduced.

This section ends with an extended example called "The Game of Craps". This example helps to reinforce a subtle line of reasoning about conditional probability that was introduced in Example 2.1.5. In particular, it uses the idea that conditional probabilities given an event B can be calculated as if we knew ahead of time that B had to occur.

Solutions to Exercises

1. If $A \subset B$, then $A \cap B = A$ and $\Pr(A \cap B) = \Pr(A)$. So $\Pr(A|B) = \Pr(A)/\Pr(B)$.

2. Since $A \cap B = \emptyset$, it follows that $\Pr(A \cap B) = 0$. Therefore, $\Pr(A \mid B) = 0$.

3. Since $A \cap S = A$ and $\Pr(S) = 1$, it follows that $\Pr(A \mid S) = \Pr(A)$.

4. Let A_i stand for the event that the shopper purchases brand A on his ith purchase, for $i = 1, 2, \ldots$. Similarly, let B_i be the event that he purchases brand B on the ith purchase. Then

$$\begin{aligned}
\Pr(A_1) &= \frac{1}{2}, \\
\Pr(A_2 \mid A_1) &= \frac{1}{3}, \\
\Pr(B_3 \mid A_1 \cap A_2) &= \frac{2}{3}, \\
\Pr(B_4 \mid A_1 \cap A_2 \cap B_3) &= \frac{1}{3}.
\end{aligned}$$

   The desired probability is the product of these four probabilities, namely $1/27$.

5. Let R_i be the event that a red ball is drawn on the ith draw, and let B_i be the event that a blue ball is drawn on the ith draw for $i = 1, \ldots, 4$. Then

$$\Pr(R_1) = \frac{r}{r+b},$$

$$\begin{aligned}
\Pr(R_2 \mid R_1) &= \frac{r+k}{r+b+k}, \\
\Pr(R_3 \mid R_1 \cap R_2) &= \frac{r+2k}{r+b+2k}, \\
\Pr(B_4 \mid R_1 \cap R_2 \cap R_3) &= \frac{b}{r+b+3k}.
\end{aligned}$$

The desired probability is the product of these four probabilities, namely

$$\frac{r(r+k)(r+2k)b}{(r+b)(r+b+k)(r+b+2k)(r+b+3k)}.$$

6. This problem illustrates the importance of relying on the rules of conditional probability rather than on intuition to obtain the answer. Intuitively, but incorrectly, it might seem that since the observed side is green, and since the other side might be either red or green, the probability that it will be green is 1/2. The correct analysis is as follows: Let A be the event that the selected card is green on both sides, and let B be the event that the observed side is green. Since each of the three cards is equally likely to be selected, $\Pr(A) = \Pr(A \cap B) = 1/3$. Also, $\Pr(B) = 1/2$. The desired probability is $\Pr(A \mid B) = \left(\frac{1}{3}\right) / \left(\frac{1}{2}\right) = \frac{2}{3}$.

7. We know that $\Pr(A) = 0.6$ and $\Pr(A \cap B) = 0.2$. Therefore, $\Pr(B \mid A) = \frac{0.2}{0.6} = \frac{1}{3}$.

8. In Exercise 2 in Sec. 1.10 it was found that $\Pr(A \cup B \cup C) = 0.85$. Since $\Pr(A) = 0.6$, it follows that $\Pr(A \mid A \cup B \cup C) = \frac{0.60}{0.85} = \frac{12}{17}$.

9. (a) If card A has been selected, each of the other four cards is equally likely to be the other selected card. Since three of these four cards are red, the required probability is 3/4.

   (b) We know, without being told, that at least one red card must be selected, so this information does not affect the probabilities of any events. We have

$$\Pr(\text{both cards red}) = \Pr(R_1)\Pr(R_2 \mid R_1) = \frac{4}{5} \cdot \frac{3}{4} = \frac{3}{5}.$$

10. As in the text, let π_0 stand for the probability that the sum on the first roll is either 7 or 11, and let π_i be the probability that the sum on the first roll is i for $i = 2, \ldots, 12$. In this version of the game of craps, we have

$$\begin{aligned}
\pi_0 &= \frac{2}{9}, \\
\pi_4 &= \pi_{10} = \frac{3}{36} \cdot \frac{\frac{3}{36}}{\frac{3}{36} + \frac{6}{36} + \frac{2}{36}} = \frac{1}{44}, \\
\pi_5 &= \pi_9 = \frac{4}{36} \cdot \frac{\frac{4}{36}}{\frac{4}{36} + \frac{6}{36} + \frac{2}{36}} = \frac{1}{27}, \\
\pi_6 &= \pi_8 = \frac{5}{36} \cdot \frac{\frac{5}{36}}{\frac{5}{36} + \frac{6}{36} + \frac{2}{36}} = \frac{25}{468}.
\end{aligned}$$

The probability of winning, which is the sum of these probabilities, is 0.448.

11. This is the conditional version of Theorem 1.5.3. From the definition of conditional probability, we have

$$\begin{aligned}
\Pr(A^c|B) &= \frac{\Pr(A^c \cap B)}{\Pr(B)}, \\
1 - \Pr(A|B) &= 1 - \frac{\Pr(A \cap B)}{\Pr(B)}, \\
&= \frac{\Pr(B) - \Pr(A \cap B)}{\Pr(B)}. \qquad \text{(S.2.1)}
\end{aligned}$$

According to Theorem 1.5.6 (switching the names A and B), $\Pr(B) - \Pr(A \cap B) = \Pr(A^c \cap B)$. Combining this with (S.2.1) yields $1 - \Pr(A|B) = \Pr(A^c|B)$.

12. This is the conditional version of Theorem 1.5.7. Let $A_1 = A \cap D$ and $A_2 = B \cap D$. Then $A_1 \cup A_2 = (A \cup B) \cap D$ and $A_1 \cap A_2 = A \cap B \cap D$. Now apply Theorem 1.5.7 to determine $\Pr(A_1 \cup A_2)$.

$$\begin{aligned}
\Pr([A \cup B] \cap D) &= \Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2) = \Pr(A \cap D) \\
&\quad + \Pr(B \cap D) - \Pr(A \cap B \cap D).
\end{aligned}$$

Now, divide the extreme left and right ends of this string of equalities by $\Pr(D)$ to obtain

$$\begin{aligned}
\Pr(A \cup B|D) &= \frac{\Pr([A \cup B] \cap D)}{\Pr(D)} = \frac{\Pr(A \cap D) + \Pr(B \cap D) - \Pr(A \cap B \cap D)}{\Pr(D)} \\
&= \Pr(A|D) + \Pr(B|D) - \Pr(A \cap B|D).
\end{aligned}$$

13. Let A_1 denote the event that the selected coin has a head on each side, let A_2 denote the event that it has a tail on each side, let A_3 denote the event that it is fair, and let B denote the event that a head in obtained. Then

$$\begin{aligned}
\Pr(A_1) &= \frac{3}{9}, \quad \Pr(A_2) = \frac{4}{9}, \quad \Pr(A_3) = \frac{2}{9}, \\
\Pr(B \mid A_1) &= 1, \quad \Pr(B \mid A_2) = 0, \quad \Pr(B \mid A_3) = \frac{1}{2}.
\end{aligned}$$

Hence,

$$\Pr(B) = \sum_{i=1}^{3} \Pr(A_i)\Pr(B \mid A_i) = \frac{4}{9}.$$

14. We partition the space of possibilities into three events B_1, B_2, B_3 as follows. Let B_1 be the event that the machine is in good working order. Let B_2 be the event that the machine is wearing down. Let B_3 be the event that it needs maintenance. We are told that $\Pr(B_1) = 0.8$ and $\Pr(B_2) = \Pr(B_3) = 0.1$. Let A be the event that a part is defective. We are asked to find $\Pr(A)$. We are told that $\Pr(A|B_1) = 0.02$, $\Pr(A|B_2) = 0.1$, and $\Pr(A|B_3) = 0.3$. The law of total probability allows us to compute $\Pr(A)$ as follows

$$\Pr(A) = \sum_{j=1}^{3} \Pr(B_j)\Pr(A|B_j) = 0.8 \times 0.02 + 0.1 \times 0.1 + 0.1 \times 0.3 = 0.056.$$

15. The analysis is similar to that given in the previous exercise, and the probability is 0.47.

16. In the usual notation, we have

$$\begin{aligned}\Pr(B_2) &= \Pr(A_1 \cap B_2) + \Pr(B_1 \cap B_2) = \Pr(A_1)\Pr(B_2 \mid A_1) + \Pr(B_1)\Pr(B_2 \mid B_1) \\ &= \frac{1}{4} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{3} = \frac{5}{12}.\end{aligned}$$

17. Clearly, we must assume that $\Pr(B_j \cap C) > 0$ for all j, otherwise (2.1.5) is undefined. By applying the definition of conditional probability to each term, the right side of (2.1.5) can be rewritten as

$$\sum_{i=1}^{k} \frac{\Pr(B_j \cap C)}{\Pr(C)} \frac{\Pr(A \cap B_j \cap C)}{\Pr(B_j \cap C)} = \frac{1}{\Pr(C)} \sum_{i=1}^{k} \Pr(A \cap B_j \cap C).$$

According to the law of total probability, the last sum above is $\Pr(A \cap C)$, hence the ratio is $\Pr(A|C)$.

2.2 Independent Events

Commentary

Near the end of this section, we introduce conditionally independent events. This is a prelude to conditionally independent and conditionally i.i.d. random variables that are introduced in Sec. 3.7. Conditional independence has become more popular in statistical modeling with the introduction of latent-variable models and expert systems. Although these models are not introduced in this text, students who will encounter them in the future would do well to study conditional independence early and often.

Conditional independence is also useful for illustrating how learning data can change the distribution of an unknown value. The first examples of this come in Sec. 2.3 after Bayes' theorem. The assumption that a sample of random variables is conditionally i.i.d. given an unknown parameter is the analog in Bayesian inference to the assumption that the random sample is i.i.d. marginally. Instructors who are not going to cover Bayesian topics might wish to bypass this material, even though it can also be useful in its own right. If you decide to not discuss conditional independence, then there is some material later in the book that you might wish to bypass as well:

- Exercise 23 in this section.
- The discussion of conditionally independent events on pages 81–84 in Sec. 2.3.
- Exercises 12, 14 and 15 in Sec. 2.3.
- The discussion of conditionally independent random variables that starts on page 163.
- Exercises 13 and 14 in Sec. 3.7.
- Virtually all of the Bayesian material.

This section ends with an extended example called "The Collector's Problem". This example combines methods from Chapters 1 and 2 to solve an easily stated but challenging problem.

Solutions to Exercises

1. If $\Pr(B) < 1$, then $\Pr(B^c) = 1 - \Pr(B) > 0$. We then compute

$$\begin{aligned}
\Pr(A^c|B^c) &= \frac{\Pr(A^c \cap B^c)}{\Pr(B^c)} \\
&= \frac{1 - \Pr(A \cup B)}{1 - \Pr(B)} \\
&= \frac{1 - \Pr(A) - \Pr(B) + \Pr(A \cap B)}{1 - \Pr(B)} \\
&= \frac{1 - \Pr(A) - \Pr(B) + \Pr(A)\Pr(B)}{1 - \Pr(B)} \\
&= \frac{[1 - \Pr(A)][1 - \Pr(B)]}{1 - \Pr(B)} \\
&= 1 - \Pr(A) = \Pr(A^c).
\end{aligned}$$

2.

$$\begin{aligned}
\Pr(A^cB^c) &= \Pr[(A \cup B)^c] = 1 - \Pr(A \cup B) \\
&= 1 - [\Pr(A) + \Pr(B) - \Pr(A \cap B)] \\
&= 1 - \Pr(A) - \Pr(B) + \Pr(A)\Pr(B)] \\
&= [1 - \Pr(A)][1 - \Pr(B)] \\
&= \Pr(A^c)\Pr(B^c).
\end{aligned}$$

3. Since the event $A \cap B$ is a subset of the event A, and $\Pr(A) = 0$, it follows that $\Pr(A \cap B) = 0$. Hence, $\Pr(A \cap B) = 0 = \Pr(A)\Pr(B)$.

4. The probability that the sum will be seven on any given roll of the dice is $1/6$. The probability that this event will occur on three successive rolls is therefore $(1/6)^3$.

5. The probability that both systems will malfunction is $(0.001)^2 = 10^{-6}$. The probability that at least one of the systems will function is therefore $1 - 10^{-6}$.

6. The probability that the man will win the first lottery is $100/10000 = 0.01$, and the probability that he will win the second lottery is $100/5000 = 0.02$. The probability that he will win at least one lottery is, therefore,

$$0.01 + 0.02 - (0.01)(0.02) = 0.0298.$$

7. Let E_1 be the event that A is in class, and let E_2 be the event that B is in class. Let C be the event that at least one of the students is in class. That is, $C = E_1 \cup E_2$.

   (a) We want $\Pr(C)$. We shall use Theorem 1.5.7 to compute the probability. Since E_1 and E_2 are independent, we have $\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2)$. Hence

   $$\Pr(C) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2) = 0.8 + 0.6 - 0.8 \times 0.6 = 0.92.$$

   (b) We want $\Pr(E_1|C)$. We computed $\Pr(C) = 0.92$ in part (a). Since $E_1 \subset C$, $\Pr(E_1 \cap C) = \Pr(E_1) = 0.8$. So, $\Pr(E_1|C) = 0.8/0.92 = 0.8696$.

8. The probability that all three numbers will be equal to a specified value is $1/6^3$. Therefore, the probability that all three numbers will be equal to any one of the six possible values is $6/6^3 = 1/36$.

9. The probability that exactly n tosses will be required on a given performance is $1/2^n$. Therefore, the probability that exactly n tosses will be required on all three performances is $(1/2^n)^3 = 1/8^n$. The probability that the same number of tosses will be required on all three performances is $\sum_{n=1}^{\infty} \frac{1}{8^n} = \frac{1}{7}$.

10. The probability p_j that exactly j children will have blue eyes is

$$p_j = \binom{5}{j} \left(\frac{1}{4}\right)^j \left(\frac{3}{4}\right)^{5-j} \quad for \quad j = 0, 1, \ldots, 5.$$

The desired probability is

$$\frac{p_3 + p_4 + p_5}{p_1 + p_2 + p_3 + p_4 + p_5}.$$

11. (a) We must determine the probability that at least two of the four oldest children will have blue eyes. The probability p_j that exactly j of these four children will have blue eyes is

$$p_j = \binom{4}{j} \left(\frac{1}{4}\right)^j \left(\frac{3}{4}\right)^{4-j}.$$

The desired probability is therefore $p_2 + p_3 + p_4$.

(b) The two different types of information provided in Exercise 10 and part (a) are similar to the two different types of information provided in part (a) and part (b) of Exercise 9 of Sec. 2.1.

12. (a) $\Pr(A^c \cap B^c \cap C^c) = \Pr(A^c)\Pr(B^c)\Pr(C^c) = \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{4}$.

(b) The desired probability is

$$\begin{aligned}\Pr(A \cap B^c \cap C^c) &+ \Pr(A^c \cap B \cap C^c) + \Pr(A^c \cap B^c \cap C) = \frac{1}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \\ &+ \frac{3}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{11}{24}.\end{aligned}$$

13. The probability of obtaining a particular sequence of ten particles in which one particle penetrates the shield and nine particles do not is $(0.01)(0.99)^9$. Since there are 10 such sequences in the sample space, the desired probability is $10(0.01)(0.99)^9$.

14. The probability that none of the ten particles will penetrate the shield is $(0.99)^{10}$. Therefore, the probability that at least one particle will penetrate the shield is $1 - (0.99)^{10}$.

15. If n particles are emitted, the probability that at least one particle will penetrate the shield is $1-(0.99)^n$. In order for this value to be at least 0.8 we must have

$$\begin{aligned}1 - (0.99)^n &\geq 0.8 \\ (0.99)^n &\leq 0.2 \\ n\log(0.99) &\leq \log(0.2).\end{aligned}$$

Since $\log(0.99)$ is negative, this final relation is equivalent to the relation

$$n \geq \frac{\log(0.2)}{\log(0.99)} \approx 160.1.$$

So 161 or more particles are needed.

16. To determine the probability that team A will win the World Series, we shall calculate the probabilities that A will win in exactly four, five, six, and seven games, and then sum these probabilities. The probability that A will win four straight game is $(1/3)^4$. The probability that A will win in five games is equal to the probability that the fourth victory of team A will occur in the fifth game. As explained in Example 2.2.8, this probability is $\binom{4}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)$. Similarly, the probabilities that A will win in six games and in seven games are $\binom{5}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2$ and $\binom{6}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^3$, respectively. By summing these probabilities, we obtain the result $\sum_{i=3}^{6}\binom{i}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{i-3}$, which equals 379/2187.

A second way to solve this problem is to pretend that all seven games are going to be played, regardless of whether one team has won four games before the seventh game. From this point of view, of the seven games that are played, the team that wins the World Series might win four, five, six, or seven games. Therefore, the probability that team A will win the series can be determined by calculating the probabilities that team A will win exactly four, five, six, and seven games, and then summing these probabilities. In this way, we obtain the result

$$\sum_{i=4}^{7}\binom{7}{i}\left(\frac{1}{3}\right)^i\left(\frac{2}{3}\right)^{7-i}.$$

It can be shown that this answer is equal to the answer that we obtained first.

17. In order for the target to be hit for the first time on the third throw of boy A, all five of the following independent events must occur: (1) A misses on his first throw, (2) B misses on his first throw, (3) A misses on his second throw, (4) B misses on his second throw, (5) A hits on his third throw. The probability of all five events occurring is $\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{1}{3}=\frac{1}{12}$.

18. Let E denote the event that boy A hits the target before boy B. There are two methods of solving this problem. The first method is to note that the event E can occur in two different ways: (i) If A hits the target on the first throw. This event occurs with probability $\frac{1}{3}$. (ii) If both A and B miss the target on their first throws, and then subsequently A hits the target before B. The probability that A and B will both miss on their first throws is $\frac{2}{3}\cdot\frac{3}{4}=\frac{1}{2}$. When they do miss, the conditions of the game become exactly the same as they were at the beginning of the game. In effect, it is as if the boys were starting a new game all over again, and so the probability that A will subsequently hit the target before B is again $\Pr(E)$. Therefore, by considering these two ways in which the event E can occur, we obtain the relation

$$\Pr(E) = \frac{1}{3} + \frac{1}{2}\Pr(E).$$

The solution is $\Pr(E) = \frac{2}{3}$.

The second method of solving the problem is to calculate the probabilities that the target will be hit for the first time on boy A's first throw, on his second throw, on his third throw, etc., and then to sum these probabilities. For the target to be hit for the first time on his jth throw, both A and B must miss on each of their first $j-1$ throws, and then A must hit on his next throw. The probability of this event is

$$\left(\frac{2}{3}\right)^{j-1}\left(\frac{3}{4}\right)^{j-1}\left(\frac{1}{3}\right)=\left(\frac{1}{2}\right)^{j-1}\left(\frac{1}{3}\right).$$

Hence,

$$\Pr(E)=\frac{1}{3}\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{j-1}=\frac{1}{3}(2)=\frac{2}{3}.$$

19. Let A_1 denote the event that no red balls are selected, let A_2 denote the event that no white balls are selected, and let A_3 denote the event that no blue balls are selected. We must determine the value of $\Pr(A_1 \cup A_2 \cup A_3)$. We shall apply Theorem 1.10.1. The event A_1 will occur if and only if all ten selected balls are white or blue. Since there is probability 0.8 that any given selected ball will be white or blue, we have $\Pr(A_1) = (0.8)^{10}$. Similarly, $\Pr(A_2) = (0.7)^{10}$ and $\Pr(A_3) = (0.5)^{10}$. The event $A_1 \cap A_2$ will occur if and only if all ten selected balls are blue. Therefore $\Pr(A_1 \cap A_2) = (0.5)^{10}$. Similarly, $\Pr(A_2 \cap A_3) = (0.2)^{10}$ and $\Pr(A_1 \cap A_3) = (0.3)^{10}$. Finally, the event $A_1 \cap A_2 \cap A_3$ cannot possibly occur, so $\Pr(A_1 \cap A_2 \cap A_3) = 0$. So, the desired probability is

$$(0.8)^{10} + (0.7)^{10} + (0.5)^{10} - (0.5)^{10} - (0.2)^{10} - (0.3)^{10} \approx 0.1356.$$

20. To prove that $B_1, \ldots, B_k$ are independent events, we must prove that for every subset of r of these events ($r = 1, \ldots, k$), we have

$$\Pr(B_{i_1} \cap \ldots \cap B_{i_r}) = \Pr(B_{i_1}) \cdots \Pr(B_{i_r}).$$

We shall simplify the notation by writing simply $B_1, \ldots, B_r$ instead of $B_{i_1}, \ldots, B_{i_r}$. Hence, we must show that

$$\Pr(B_1 \cap \ldots \cap B_r) = \Pr(B_1) \cdots \Pr(B_r). \tag{S.2.2}$$

Suppose that the relation (S.2.2) is satisfied whenever $B_j = A_j^c$ for m or fewer values of j and $B_j = A_j$ for the other $k-m$ or more values of j. We shall show that (S.2.2) is also satisfied whenever $B_j = A_j^c$ for $m+1$ values of j. Without loss of generality, we shall assume that $j = r$ is one of these $m+1$ values, so that $B_r = A_r^c$. It is always true that

$$\Pr(B_1 \cap \ldots \cap B_r) = \Pr(B_1 \cap \ldots \cap B_{r-1}) - \Pr(B_1 \cap \ldots \cap B_{r-1} \cap B_r^c).$$

Since among the events $B_1, \ldots, B_{r-1}$ there are m or fewer values of j such that $B_j = A_j^c$, it follows from the induction hypothesis that

$$\Pr(B_1 \cap \ldots \cap B_{r-1}) = \Pr(B_1) \cdots \Pr(B_{r-1}).$$

Furthermore, since $B_r^c = A_r$, the same induction hypothesis implies that

$$\Pr(B_1 \cap \ldots \cap B_{r-1} B_r^c) = \Pr(B_1) \cdots \Pr(B_{r-1}) \Pr\left(B_r^c\right).$$

It now follows that

$$\Pr(B_1 \cap \ldots \cap B_r) = \Pr(B_1) \cdots \Pr(B_{r-1})[1 - \Pr(B_r^c)] = \Pr(B_1) \ldots Pr(B_r).$$

Thus, we have shown that if the events $B_1, \ldots, B_k$ are independent whenever there are m or fewer values of j such that $B_j = A_j^c$, then the events $B_1, \ldots, B_k$ are also independent whenever there are $m+1$ values of j such that $B_j = A_j^c$. Since $B_1, \ldots, B_k$ are obviously independent whenever there are zero values of j such that $B_j = A_j^c$ (i.e., whenever $B_j = A_j$ for $j = 1, \ldots, k$), the induction argument is complete. Therefore, the events $B_1, \ldots, B_k$ are independent regardless of whether $B_j = A_j$ or $B_j = A_j^c$ for each value of j.

21. For the "only if" direction, we need to prove that if $A_1, \ldots, A_k$ are independent then

$$\Pr(A_{i_1} \cap \cdots \cap A_{i_m} | A_{j_1} \cap \cdots \cap A_{j_\ell}) = \Pr(A_{i_1} \cap \cdots \cap A_{i_m}),$$

for all disjoint subsets $\{i_1, \ldots, i_m\}$ and $\{j_1, \ldots, j_\ell\}$ of $\{1, \ldots, k\}$. If $A_1, \ldots, A_k$ are independent, then

$$\Pr(A_{i_1} \cap \cdots \cap A_{i_m} \cap A_{j_1} \cap \cdots \cap A_{j_\ell}) = \Pr(A_{i_1} \cap \cdots \cap A_{i_m}) \Pr(A_{j_1} \cap \cdots \cap A_{j_\ell}),$$

hence it follows that

$$\Pr(A_{i_1} \cap \cdots \cap A_{i_m} | A_{j_1} \cap \cdots \cap A_{j_\ell}) = \frac{\Pr(A_{i_1} \cap \cdots \cap A_{i_m} \cap A_{j_1} \cap \cdots \cap A_{j_\ell})}{\Pr(A_{j_1} \cap \cdots \cap A_{j_\ell})} = \Pr(A_{i_1} \cap \cdots \cap A_{i_m}).$$

For the "if" direction, assume that $\Pr(A_{i_1} \cap \cdots \cap A_{i_m} | A_{j_1} \cap \cdots \cap A_{j_\ell}) = \Pr(A_{i_1} \cap \cdots \cap A_{i_m})$ for all disjoint subsets $\{i_1, \ldots, i_m\}$ and $\{j_1, \ldots, j_\ell\}$ of $\{1, \ldots, k\}$. We must prove that $A_1, \ldots, A_k$ are independent. That is, we must prove that for every subset $\{s_1, \ldots, s_n\}$ of $\{1, \ldots, k\}$, $\Pr(A_{s_1} \cap \cdots \cap A_{s_n}) = \Pr(A_{s_1}) \cdots \Pr(A_{s_n})$. We shall do this by induction on n. For $n = 1$, we have that $\Pr(A_{s_1}) = \Pr(A_{s_1})$ for each subset $\{s_1\}$ of $\{1, \ldots, k\}$. Now, assume that for all $n \le n_0$ and for all subsets $\{s_1, \ldots, s_n\}$ of $\{1, \ldots, k\}$ it is true that $\Pr(A_{s_1} \cap \cdots \cap A_{s_n}) = \Pr(A_{s_1}) \cdots \Pr(A_{s_n})$. We need to prove that for every subset $\{t_1, \ldots, t_{n_0+1}\}$ of $\{1, \ldots, k\}$

$$\Pr(A_{t_1} \cap \cdots \cap A_{t_{n_0+1}}) = \Pr(A_{t_1}) \cdots \Pr(A_{t_{n_0+1}}). \tag{S.2.3}$$

It is clear that

$$\Pr(A_{t_1} \cap \cdots \cap A_{t_{n_0+1}}) = \Pr(A_{t_1} \cap \cdots \cap A_{t_{n_0}} | A_{t_{n_0+1}}) \Pr(A_{t_{n_0+1}}). \tag{S.2.4}$$

We have assumed that $\Pr(A_{t_1} \cap \cdots \cap A_{t_{n_0}} | A_{t_{n_0+1}}) = \Pr(A_{t_1} \cap \cdots \cap A_{t_{n_0}})$ for all disjoint subsets $\{t_1, \ldots, t_{n_0}\}$ and $\{t_{n_0+1}\}$ of $\{1, \ldots, k\}$. Since the right side of this last equation is the probability of the intersection of only n_0 events, then we know that

$$\Pr(A_{t_1} \cap \cdots \cap A_{t_{n_0}}) = \Pr(A_{t_1}) \cdots \Pr(A_{t_{n_0}}).$$

Combining this with Eq. (S.2.4) implies that (S.2.3) holds.

22. For the "only if" direction, we assume that A_1 and A_2 are conditionally independent given B and we must prove that $\Pr(A_2 | A_1 \cap B) = \Pr(A_2 | B)$. Since A_1 and A_2 are conditionally independent given B, $\Pr(A_1 \cap A_2 | B) = \Pr(A_1 | B) \Pr(A_2 | B)$. This implies that

$$\Pr(A_2 | B) = \frac{\Pr(A_1 \cap A_2 | B)}{\Pr(A_1 | B)}.$$

Also,

$$\Pr(A_2|A_1 \cap B) = \frac{\Pr(A_1 \cap A_2 \cap B)}{\Pr(A_1 \cap B)} = \frac{\Pr(A_1 \cap A_2 \cap B)/\Pr(B)}{\Pr(A_1 \cap B)/\Pr(B)} = \frac{\Pr(A_1 \cap A_2|B)}{\Pr(A_1|B)}.$$

Hence, $\Pr(A_2|A_1 \cap B) = \Pr(A_2|B)$.

For the "if" direction, we assume that $\Pr(A_2|A_1 \cap B) = \Pr(A_2|B)$, and we must prove that A_1 and A_2 are conditionally independent given B. That is, we must prove that $\Pr(A_1 \cap A_2|B) = \Pr(A_1|B)\Pr(A_2|B)$. We know that

$$\Pr(A_1 \cap A_2|B)\Pr(B) = \Pr(A_2|A_1 \cap B)\Pr(A_1 \cap B),$$

since both sides are equal to $\Pr(A_1 \cap A_2 \cap B)$. Divide both sides of this equation by $\Pr(B)$ and use the assumption $\Pr(A_2|A_1B) = \Pr(A_2|B)$ together with $\Pr(A_1 \cap B)/\Pr(B) = \Pr(A_1|B)$ to obtain

$$\Pr(A_1 \cap A_2|B) = \Pr(A_2|B)\Pr(A_1|B).$$

23. (a) Conditional on B the events $A_1, \ldots, A_{11}$ are independent with probability 0.8 each. The conditional probability that a particular collection of eight programs out of the 11 will compile is $0.8^8 0.2^3 = 0.001342$. There are $\binom{11}{8} = 165$ different such collections of eight programs out of the 11, so the probability of exactly eight programs will compile is $165 \times 0.001342 = 0.2215$.

 (b) Conditional on B^c the events $A_1, \ldots, A_{11}$ are independent with probability 0.4 each. The conditional probability that a particular collection of eight programs out of the 11 will compile is $0.4^8 0.6^3 = 0.0001416$. There are $\binom{11}{8} = 165$ different such collections of eight programs out of the 11, so the probability of exactly eight programs will compile is $165 \times 0.0001416 = 0.02335$.

24. Let $n > 1$, and assume that $A_1, \ldots, A_n$ are mutually exclusive. For the "if" direction, assume that at most one of the events has strictly positive probability. Then, the intersection of every collection of size 2 or more has probability 0. Also, the product of every collection of 2 or more probabilities is 0, so the events satisfy Definition 2.2.2 and are mutually independent. For the "only if" direction, assume that the events are mutually independent. The intersection of every collection of size 2 or more is empty and must have probability 0. Hence the product of the probabilities of every collection of size 2 or more must be 0. This means that at least one factor from every product of at least 2 probabilities must itself be 0. Hence there can be no more than one of the probabilities greater than 0, otherwise the product of the two nonzero probabilities would be nonzero.

2.3 Bayes' Theorem

Commentary

This section ends with two extended discussions on how Bayes' theorem is applied. The first involves a sequence of simple updates to the probability of a specific event. It illustrates how conditional independence allows one to use posterior probabilities after observing some events as prior probabilities before observing later events. This idea is subtle, but very useful in Bayesian inference. The second discussion builds upon this idea and illustrates the type of reasoning that can be used in real inference problems. Examples 2.3.7 and 2.3.8 are particularly useful in this regard. They show how data can bring very different prior probabilities into closer posterior agreement. Exercise 12 illustrates the effect of the size of a sample on the degree to which the data can reduce differences in subjective probabilities.

Statistical software like R can be used to facilitate calculations like those that occur in the above-mentioned examples. For example, suppose that the 11 prior probabilities are assigned to the vector `prior` and that the data consist of `s` successes and `f` failures. Then the posterior probabilities can be computed by

```
ints=1:11
post=prior*((ints-1)/10)^s*(1-(ints-1)/10)^f
post=post/sum(post)
```

Solutions to Exercises

1. It must be true that $\sum_{i=1}^{k} \Pr(B_i) = 1$ and $\sum_{i=1}^{k} \Pr(B_i \mid A) = 1$. However, if $\Pr(B_1 \mid A) < \Pr(B_1)$ and $\Pr(B_i \mid A) \le \Pr(B_i)$ for $i = 2, \ldots, k$, we would have $\sum_{i=1}^{k} \Pr(B_i \mid A) < \sum_{i=1}^{k} \Pr(B_i)$, a contradiction. Therefore, it must be true that $\Pr(B_i \mid A) > \Pr(B_i)$ for at least one value of i $(i = 2, \ldots, k)$.

2. It was shown in the text that $\Pr(A_2 \mid B) = 0.26 < \Pr(A_2) = 0.3$. Similarly,

$$\Pr(A_1 \mid B) = \frac{(0.2)(0.01)}{(0.2)(0.01) + (0.3)(0.02) + (0.5)(0.03)} = 0.09.$$

Since $\Pr(A_1) = 0.2$, we have $\Pr(A_1 \mid B) < \Pr(A_1)$. Furthermore,

$$\Pr(A_3 \mid B) = \frac{(0.5)(0.03)}{(0.2)(0.01) + (0.3)(0.02) + (0.5)(0.03)} = 0.65.$$

Since $\Pr(A_3) = 0.5$, we have $\Pr(A_3 \mid B) > \Pr(A_3)$.

3. Let C denote the event that the selected item is nondefective. Then

$$\Pr(A_2 \mid C) = \frac{(0.3)(0.98)}{(0.2)(0.99) + (0.3)(0.98) + (0.5)(0.97)} = 0.301.$$

Commentary: It should be noted that if the selected item is observed to be defective, the probability that the item was produced by machine M_2 is decreased from the prior value of 0.3 to the posterior value of 0.26. However, if the selected item is observed to be nondefective, this probability changes very little, from a prior value of 0.3 to a posterior value of 0.301. In this example, therefore, obtaining a defective is more informative than obtaining a nondefective, but it is much more probable that a nondefective will be obtained.

4. The desired probability Pr(Cancer | Positive) can be calculated as follows:

$$\frac{\Pr(\text{Cancer})\Pr(\text{Positive} \mid \text{Cancer})}{\Pr(\text{Cancer})\Pr(\text{Positive} \mid \text{Cancer}) + \Pr(\text{No Cancer})\Pr(\text{Positive} \mid \text{No Cancer})}$$
$$= \frac{(0.00001)(0.95)}{(0.00001)(0.95) + (0.99999)(0.05)} = 0.00019.$$

Commentary: It should be noted that even though this test provides a correct diagnosis 95 percent of the time, the probability that a person has this type of cancer, given that he has a positive reaction to the test, is *not* 0.95. In fact, as this exercise shows, even though a person has a positive reaction, the probability that she has this type of cancer is still only 0.00019. In other words, the probability that the person has this type of cancer is 19 times larger than it was before he took this test $\left(\frac{0.00019}{0.000001} = 19\right)$, but it is still very small because the disease is so rare in the population.

5. The desired probability Pr(Lib.|NoVote) can be calculated as follows:

$$\frac{\Pr(\text{Lib.})\Pr(\text{NoVote}|\text{Lib.})}{\Pr(\text{Cons.})\Pr(\text{NoVote}|\text{Cons.}) + \Pr(\text{Lib.})\Pr(\text{NoVote}|\text{Lib.}) + \Pr(\text{Ind.})\Pr(\text{NoVote}|\text{Ind.})}$$
$$= \frac{(0.5)(0.18)}{(0.3)(0.35) + (0.5)(0.18) + (0.2)(0.50)} = \frac{18}{59}.$$

6. (a) Let A_1 denote the event that the machine is adjusted properly, let A_2 denote the event that it is adjusted improperly, and let B be the event that four of the five inspected items are of high quality. Then

$$\Pr(A_1 \mid B) = \frac{\Pr(A_1)\Pr(B \mid A_1)}{\Pr(A_1)\Pr(B \mid A_1) + \Pr(A_2)\Pr(B \mid A_2)}$$
$$= \frac{(0.9)\binom{5}{4}(0.5)^5}{(0.9)\binom{5}{4}(0.5)^5 + (0.1)\binom{5}{4}(0.25)^4(0.75)} = \frac{96}{97}.$$

(b) The prior probabilities before this additional item is observed are the values found in part (a): $\Pr(A_1) = 96/97$ and $\Pr(A_2) = 1/97$. Let C denote the event that the additional item is of medium quality. Then

$$\Pr(A_1 \mid C) = \frac{\frac{96}{97}\cdot\frac{1}{2}}{\frac{96}{97}\cdot\frac{1}{2} + \frac{1}{97}\cdot\frac{3}{4}} = \frac{64}{65}.$$

7. (a) Let π_i denote the posterior probability that coin i was selected. The prior probability of each coin is 1/5. Therefore

$$\pi_i = \frac{\frac{1}{5}p_i}{\sum_{j=1}^{5}\frac{1}{5}p_j} \quad \text{for} \quad i = 1, \ldots, 5.$$

The five values are $\pi_1 = 0$, $\pi_2 = 0.1$, $\pi_3 = 0.2$, $\pi_4 = 0.3$, and $\pi_5 = 0.4$.

(b) The probability of obtaining another head is equal to

$$\sum_{i=1}^{5}\Pr(\text{Coin } i)\Pr(\text{Head} \mid \text{ Coin } i) = \sum_{i=1}^{5}\pi_i p_i = \frac{3}{4}.$$

(c) The posterior probability π_i of coin i would now be

$$\pi_i = \frac{\frac{1}{5}(1-p_i)}{\sum_{j=1}^{5}\frac{1}{5}(1-p_j)} \quad \text{for } i = 1, \ldots, 5.$$

Thus, $\pi_1 = 0.4, \pi_2 = 0.3, \pi_3 = 0.2, \pi_4 = 0.1$, and $\pi_5 = 0$. The probability of obtaining a head on the next toss is therefore $\sum_{i=1}^{5} \pi_i p_i = \frac{1}{4}$.

8. (a) If coin i is selected, the probability that the first head will be obtained on the fourth toss is $(1-p_i)^3 p_i$. Therefore, the posterior probability that coin i was selected is

$$\pi_i = \frac{\frac{1}{5}(1-p_i)^3 p_i}{\sum_{j=1}^{5} \frac{1}{5}(1-p_j)^3 p_j} \quad \text{for } i = 1, \dots, 5.$$

The five values are $\pi_1 = 0$, $\pi_2 = 0.5870$, $\pi_3 = 0.3478$, $\pi_4 = 0.0652$, and $\pi_5 = 0$.

(b) If coin i is used, the probability that exactly three additional tosses will be required to obtain another head is $(1-p_i)^2 p_i$. Therefore, the desired probability is

$$\sum_{i=1}^{5} \pi_i (1-p_i)^2 p_i = 0.1291.$$

9. We shall continue to use the notation from the solution to Exercise 14 in Sec. 2.1. Let C be the event that exactly one out of seven observed parts is defective. We are asked to find $\Pr(B_j|C)$ for $j = 1, 2, 3$. We need $\Pr(C|B_j)$ for each j. Let A_i be the event that the ith part is defective. For all i, $\Pr(A_i|B_1) = 0.02$, $\Pr(A_i|B_2) = 0.1$, and $\Pr(A_i|B_3) = 0.3$. Since the seven parts are conditionally independent given each state of the machine, the probability of each possible sequence of seven parts with one defective is $\Pr(A_i|B_j)[1 - \Pr(A_i|B_j)]^6$. There are seven distinct such sequences, so

$$\begin{aligned}
\Pr(C|B_1) &= 7 \times 0.02 \times 0.98^6 = 0.1240, \\
\Pr(C|B_2) &= 7 \times 0.1 \times 0.9^6 = 0.3720, \\
\Pr(C|B_3) &= 7 \times 0.3 \times 0.7^6 = 0.2471.
\end{aligned}$$

The expression in the denominator of Bayes' theorem is

$$\Pr(C) = 0.8 \times 0.1240 + 0.1 \times 0.3720 + 0.1 \times 0.2471 = 0.1611.$$

Bayes' theorem now says

$$\begin{aligned}
\Pr(B_1|C) &= \frac{0.8 \times 0.1240}{0.1611} = 0.6157, \\
\Pr(B_2|C) &= \frac{0.1 \times 0.3720}{0.1611} = 0.2309, \\
\Pr(B_3|C) &= \frac{0.1 \times 0.2471}{0.1611} = 0.1534.
\end{aligned}$$

10. Bayes' theorem says that the posterior probability of each B_i is $\Pr(B_i|E) = \Pr(B_i)\Pr(E|B_i)/\Pr(E)$. So $\Pr(B_i|E) < \Pr(B_i)$ if and only if $\Pr(E|B_i) < \Pr(E)$. Since $\Pr(E) = 3/4$, we need to find those i for which $\Pr(E|B_i) < 3/4$. These are $i = 5, 6$.

11. This time, we want $\Pr(B_4|E^c)$. We know that $\Pr(E^c) = 1 - \Pr(E) = 1/4$ and $\Pr(E^c|B_4) = 1 - \Pr(E|B_4) = 1/4$. This means that E^c and B_4 are independent so that $\Pr(B_4|E^c) = \Pr(B_4) = 1/4$.

12. We are doing the same calculations as in Examples 2.3.7 and 2.3.8 except that we only have five patients and three successes. So, in particular

$$\Pr(B_j \mid A) = \frac{\Pr(B_j)\binom{5}{3}([j-1]/10)^3(1-[j-1]/10)^2}{\sum_{i=1}^{11}\Pr(B_i)\binom{5}{3}([i-1]/10)^3(1-[i-1]/10)^2}. \tag{S.2.5}$$

In one case, $\Pr(B_i) = 1/11$ for all i, and in the other case, the prior probabilities are given in the table in Example 2.3.8 of the text. The numbers that show up in both calculations are

i	$\left(\frac{i-1}{10}\right)^3\left(1-\frac{i-1}{10}\right)^2$	i	$\left(\frac{i-1}{10}\right)^3\left(1-\frac{i-1}{10}\right)^2$
1	0	7	0.0346
2	0.0008	8	0.0309
3	0.0051	9	0.0205
4	0.0132	10	0.0073
5	0.0230	11	0
6	0.0313		

We can use these with the two sets of prior probabilities to compute the posterior probabilities according to Eq. (S.2.5).

i	Example 2.3.7	Example 2.3.8	i	Example 2.3.7	Example 2.3.8
1	0	0	7	0.2074	0.1641
2	0.0049	0.0300	8	0.1852	0.0879
3	0.0307	0.0972	9	0.1229	0.0389
4	0.07939	0.1633	10	0.0437	0.0138
5	0.1383	0.1969	11	0	0
6	0.1875	0.2077			

These numbers are not nearly so close as those in the examples in the text because we do not have as much information in the small sample of five patients.

13. (a) Let B_1 be the event that the coin is fair, and let B_2 be the event that the coin has two heads. Let H_i be the event that we obtain a head on the ith toss for $i = 1, 2, 3, 4$. We shall apply Bayes' theorem conditional on H_1H_2.

$$\begin{aligned}
&\Pr(B_1|H_1 \cap H_2 \cap H_3) \\
&= \frac{\Pr(B_1|H_1 \cap H_2)\Pr(H_3|B_1 \cap H_1 \cap H_2)}{\Pr(B_1|H_1 \cap H_2)\Pr(H_3|B_1 \cap H_1 \cap H_2) + \Pr(B_2|H_1 \cap H_2)\Pr(H_3|B_2 \cap H_1 \cap H_2)} \\
&= \frac{(1/5)\times(1/2)}{(1/5)\times(1/2)+(4/5)\times 1} = \frac{1}{9}.
\end{aligned}$$

(b) If the coin ever shows a tail, it can't have two heads. Hence the posterior probability of B_1 becomes 1 after we observe a tail.

14. In Exercise 23 of Sec. 2.2, B is the event that the programming task was easy. In that exercise, we computed $\Pr(A|B) = 0.2215$ and $\Pr(A|B^c) = 0.02335$. We are also told that $\Pr(B) = 0.4$. Bayes'

theorem tells us that

$$\begin{aligned}\Pr(B|A) &= \frac{\Pr(B)\Pr(A|B)}{\Pr(B)\Pr(A|B)+\Pr(B^c)\Pr(A|B^c)} = \frac{0.4\times 0.2215}{0.4\times 0.2215+(1-0.4)0.02335} \\ &= 0.8635.\end{aligned}$$

15. The law of total probability tells us how to compute $\Pr(E_1)$.

$$\Pr(E_1) = \sum_{i=1}^{11}\Pr(B_i)\frac{i-1}{10}.$$

Using the numbers in Example 2.3.8 for $\Pr(B_i)$ we obtain 0.274. This is smaller than the value 0.5 computed in Example 2.3.7 because the prior probabilities in Example 2.3.8 are much higher for the B_i with low values of i, particularly $i = 2, 3, 4$, and they are much smaller for those B_i with large values of i. Since $\Pr(E_1)$ is a weighted average of the values $(i-1)/10$ with the weights being $\Pr(B_i)$ for $i = 1, \ldots 11$, the more weight we give to small values of $(i-1)/10$, the smaller the weighted average will be.

16. (a) From the description of the problem $\Pr(D_i|B) = 0.01$ for all i. If we can show that $\Pr(D_i|B^c) = 0.01$ for all i, then $\Pr(D_i) = 0.01$ for all i. We will prove this by induction. We have assumed that D_1 is independent of B and hence it is independent of B^c. This makes $\Pr(D_1|B^c) = 0.01$. Now, assume that $\Pr(D_i|B^c) = 0.01$ for all $i \le j$. Write

$$\Pr(D_{j+1}|B^c) = \Pr(D_{j+1}|D_j \cap B^c)\Pr(D_j|B^c) + \Pr(D_{j+1}|D_j^c \cap B^c)\Pr(D_j^c|B^c).$$

The induction hypothesis says that $\Pr(D_j|B^c) = 0.01$ and $\Pr(D_j^c|B^c) = 0.99$. In the problem description, we have $\Pr(D_{j+1}|D_j \cap B^c) = 2/5$ and $\Pr(D_{j+1}|D_j^c \cap B^c) = 1/165$. Plugging these into (16a) gives

$$\Pr(D_{j+1}|B^c) = \frac{2}{5}\times 0.01 + \frac{1}{165}\times 0.99 = 0.01.$$

This completes the proof.

(b) It is straightforward to compute

$$\Pr(E|B) = 0.99\times 0.99\times 0.01\times 0.01\times 0.99\times 0.99 = 0.00009606.$$

By the conditional independence assumption stated in the problem description, we have

$$\Pr(E|B^c) = \Pr(D_1^c|B^c)\Pr(D_2^c|D_1^c\cap B^c)\Pr(D_3|D_2^c\cap B^c)\Pr(D_4|D_3\cap B^c)\Pr(D_5^c|D_4\cap B^c)\Pr(D_6^c|D_5^c\cap B^c).$$

The six factors on the right side of this equation are respectively 0.99, 164/165, 1/165, 2/5, 3/5, and 164/165. The product is 0.001423. It follows that

$$\begin{aligned}\Pr(B|E) &= \frac{\Pr(E|B)\Pr(B)}{\Pr(E|B)\Pr(B)+\Pr(E|B^c)\Pr(B^c)} \\ &= \frac{0.00009606\times(2/3)}{0.00009606\times(2/3)+0.001423\times(1/3)} = 0.1190.\end{aligned}$$

2.4 The Gambler's Ruin Problem

Commentary

This section is independent of the rest of the book. Instructors can discuss this section at any time that they find convenient or they can omit it entirely.

If Sec. 3.10 on Markov chains has been discussed before this section is discussed, it is helpful to point out that the game considered here forms a Markov chain with stationary transition probabilities. The state of the chain at any time is the fortune of gambler A at that time. Therefore, the possible states of the chain are the $k+1$ integers $0, 1, \dots, k$. If the chain is in state i ($i = 1, \dots, k-1$) at any time n, then at time $n+1$ it will move to state $i+1$ with probability p and it will move to state $i-1$ with probability $1-p$. It is assumed that if the chain is either in the state 0 or the state k at any time, then it will remain in that same state at every future time. (These are absorbing states.) Therefore, the $(k+1)\times(k+1)$ transition matrix $\boldsymbol{P}$ of the chain is as follows:

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1-p & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

Solutions to Exercises

1. Clearly a_i in Eq. (2.4.9) is an increasing function of i. Hence, if $a_{98} < 1/2$, then $a_i < 1/2$ for all $i \le 98$. For $i = 98$, Eq. (2.4.9) yields almost exactly 4/9, which is less that 1/2.

2. The probability of winning a fair game is just the ratio of the initial fortune to the total funds available. This ratio is the same in all three cases.

3. If the initial fortune of gambler A is i dollars, then for conditions (a), (b), and (c), the initial fortune of gambler B is $i/2$ dollars. Hence, $k = 3i/2$. If we let $r = (1-p)/p > 1$, then it follows from Eq. (2.4.8) that the probability that A will win under conditions (a), (b), or (c) is

$$\frac{r^i - 1}{r^{3i/2} - 1} = \frac{1 - (1/r_i)}{r^{i/2} - (1/r_i)}.$$

If i and j are positive integers with $i < j$, it now follows that

$$\frac{1-(1/r_j)}{r^{j/2} - (1/r_j)} < \frac{1-(1/r_j)}{r^{i/2} - (1/r_j)} < \frac{1-(1/r_i)}{r^{i/2} - (1/r_i)}.$$

Thus the larger the initial fortune of gambler A is, the smaller is his probability of winning. Therefore, he has the largest probability of winning under condition (a).

4. If we consider this problem from the point of view of gambler B, then each play of the game is unfavorable to her. Hence, by a procedure similar to that described in the solution to Exercise 3, it follows that she has the smallest probability of winning when her initial fortune is largest. Therefore, gambler A has the largest probability of winning when her initial fortune is largest, which corresponds to condition (c).

5. In this exercise, $p = 1/2$ and $k = i + 2$. Therefore $a_i = i/(i + 2)$. In order to make $a_i \geq 0.99$, we must have $i \geq 198$.

6. In this exercise $p = 2/3$, and $k = i + 2$. Therefore, by Eq. (2.4.9),

$$a_i = \frac{\left(\frac{1}{2}\right)^i - 1}{\left(\frac{1}{2}\right)^{i+2} - 1}.$$

It follows that $a_i \geq 0.99$ if and only if

$$2^i \geq 75.25.$$

Therefore, we must have $i \geq 7$.

7. In this exercise $p = 1/3$ and $k = i + 2$. Therefore, by Eq. (2.4.9)

$$a_i = \frac{2^i - 1}{2^{i+2} - 1} = \frac{1 - (1/2^i)}{4 - (1/2^i)}.$$

But for every number x ($0 < x < 1$), we have $\frac{1 - x}{4 - x} < \frac{1}{4}$. Hence, $a_i < 1/4$ for every positive integer i.

8. This problem can be expressed as a gambler's ruin problem. Suppose that the initial fortunes of both gambler A and gambler B are 3 dollars, that gambler A will win one dollar from gambler B whenever a head is obtained on the coin, and gambler B will win one dollar from gambler A whenever a tail is obtained on the coin. Then the condition that $X_n = Y_n + 3$ means that A has won all of B's fortune, and the condition that $Y_n = X_n + 3$ means that A is ruined. Therefore, if $p = 1/2$, the required probability is given by Eq. (2.4.6) with $i = 3$ and $k = 6$, and the answer is $1/2$. If $p \neq 1/2$, the required probability is given by Eq. (2.4.9) with $i = 3$ and $k = 6$. In either case, the answer can be expressed in the form

$$\frac{1}{\left(\frac{1-p}{p}\right)^3 + 1}.$$

9. This problem can be expressed as a gambler's ruin problem. We consider the initial fortune of gambler A to be five dollars and the initial fortune of gambler B to be ten dollars. Gambler A wins one dollar from gambler B each time that box B is selected, and gambler B wins one dollar from gambler A each time that box A is selected. Since $i = 5, k = 15$, and $p = 1/2$, it follows from Eq. (2.4.6) that the probability that gambler A will win (and box B will become empty) is $1/3$. Therefore, the probability that box A will become empty first is $2/3$.

2.5 Supplementary Exercises

Solutions to Exercises

1. Let $\Pr(D) = p > 0$. Then

$$\begin{aligned}\Pr(A) &= p\Pr(A \mid D) + (1-p)\Pr(A \mid D^c) \\ &\geq p\Pr(B \mid D) + (1-p)\Pr(B \mid D^c) = \Pr(B).\end{aligned}$$

2. (a) Sample space:

$$\begin{array}{ll} HT & TH \\ HHT & TTH \\ HHHT & TTTH \\ \ldots & \ldots \end{array}$$

(b) $\Pr(HHT \text{ or } TTH) = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right) + \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right) = \frac{1}{4}$.

3. Since $\Pr(A \mid B) = \dfrac{\Pr(A\cap B)}{\Pr(B)}$ and $\Pr(B \mid A) = \dfrac{\Pr(A\cap B)}{\Pr(A)}$, we have $\dfrac{\Pr(A\cap B)}{(1/5)} + \dfrac{\Pr(A\cap B)}{(1/3)} = \dfrac{2}{3}$. Hence $\Pr(A\cap B) = 1/12$, and $\Pr(A^c \cup B^c) = 1 - \Pr(A\cap B) = 11/12$.

4. $\Pr(A \cup B^c \mid B) = \Pr(A \mid B) + \Pr(B^c \mid B) - \Pr(A \cap B^c \mid B) = \Pr(A) + 0 - 0 = \Pr(A)$.

5. The probability of obtaining the number 6 exactly three times in ten rolls is $a = \binom{10}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$. Hence, the probability of obtaining the number 6 on the first three rolls and on none of the subsequent rolls is $b = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$. Hence, the required probability is $\frac{b}{a} = 1/\binom{10}{3}$.

6. $\Pr(A\cap B) = \dfrac{\Pr(A\cap B\cap D)}{\Pr(D \mid A\cap B)} = \dfrac{0.04}{0.25} = 0.16$. But also, by independence,

$$\Pr(A\cap B) = \Pr(A)\Pr(B) = 4[\Pr\ (A)]^2.$$

Hence, $4[\Pr(A)]^2 = 0.16$, so $\Pr(A) = 0.2$. It now follows that

$$\Pr(A\cup B) = \Pr(A) + \Pr(B) - \Pr(A\cap B) = (0.2) + 4(0.2) - (0.16) = 0.84.$$

7. The three events are always independent under the stated conditions. The proof is a straightforward generalization of the proof of Exercise 2 in Sec. 2.2.

8. No, since $\Pr(A\cap B) = 0$ but $\Pr(A)\Pr(B) > 0$. This also follows from Theorem 2.2.3.

9. Let $\Pr(A) = p$. Then $\Pr(A\cap B) = \Pr(A\cap B\cap C) = 0$, $\Pr(A\cap C) = 4p^2$, $\Pr(B\cap C) = 8p^2$. Therefore, by Theorem 1.10.1, $5p = p + 2p + 4p - [0 + 4p^2 + 8p^2] + 0$, and $p = 1/6$.

10. $\Pr(\text{Sum} = 7) = 2\Pr[(1,6)] + 2\Pr[(2,5)] + 2\Pr[(3,4)] = 2(0.1)(0.1) + 2(0.1)(0.1) + 2(0.3)(0.3) = 0.22$.

11. $1 - \Pr(\text{losing 50 times}) = 1 - \left(\frac{49}{50}\right)^{50}$.

12. The event will occur when (X_1, X_2, X_3) has the following values:

$$\begin{array}{llllll} (6,5,1) & (6,4,1) & (6,3,1) & (5,4,1) & (5,3,1) & (4,3,1) \\ (6,5,2) & (6,4,2) & (6,3,2) & (5,4,2) & (5,3,2) & (4,3,2) \\ (6,5,3) & (6,4,3) & (6,2,1) & (5,4,3) & (5,2,1) & (4,2,1) \\ (6,5,4) & & & & & (3,2,1). \end{array}$$

Each of these 20 points has probability $1/6^3$, so the answer is $20/216 = 5/54$.

13. Let A, B, and C stand for the events that each of the students is in class on a particular day.

   (a) We want $\Pr(A \cup B \cup C)$. We can use Theorem 1.10.1. Independence makes it easy to compute the probabilities of the various intersections.

$$\Pr(A \cup B \cup C) = 0.3 + 0.5 + 0.8 - [0.3 \times 0.5 + 0.3 \times 0.8 + 0.5 \times 0.8] + 0.3 \times 0.5 \times 0.8 = 0.93.$$

   (b) Once again, use independence to calculate probabilities of intersections.

$$\begin{aligned} &\Pr(A \cap B^c \cap C^c) + \Pr(A^c \cap B \cap C^c) + \Pr(A^c \cap B^c \cap C) \\ = &\quad (0.3)(0.5)(0.2) + (0.7)(0.5)(0.2) + (0.7)(0.5)(0.8) = 0.38. \end{aligned}$$

14. Seven games will be required if and only if team A wins exactly three of the first six games. This probability is $\binom{6}{3} p^3(1-p)^3$, following the model calculation in Example 2.2.5.

15. $\Pr(\text{Each box contains one red ball}) = \dfrac{3!}{3^3} = \dfrac{2}{9} = \Pr(\text{Each box contains one white ball})$.
So $\Pr(\text{Each box contains both colors}) = \left(\dfrac{2}{9}\right)^2$.

16. Let A_i be the event that box i has at least three balls. Then

$$\Pr(A_i) = \sum_{j=3}^{5} \Pr(\text{Box } i \text{ has exactly } j \text{ balls}) = \frac{\binom{5}{3}(n-1)^2}{n^5} + \frac{\binom{5}{4}(n-1)}{n^5} + \frac{1}{n^5} = p, \text{ say.}$$

Since there are only five balls, it is impossible for two boxes to have at least three balls at the same time. Therefore, the events A_i are disjoint, and the probability that at least one of the events A_i occurs is np. Hence, the probability that no box contains more than two balls is $1 - np$.

17. $\Pr(U + V = j)$ is as follows, for $j = 0, 1, \ldots, 18$:

j	Prob.	j	Prob.
0	0.01	10	0.09
1	0.02	11	0.08
2	0.03	12	0.07
3	0.04	13	0.06
4	0.05	14	0.05
5	0.06	15	0.04
6	0.07	16	0.03
7	0.08	17	0.02
8	0.09	18	0.01
9	0.10		

Thus

$$\begin{aligned} \Pr(U + V = W + X) &= \sum_{j=0}^{18} \Pr(U + V = j)\Pr(W + X = j) \\ &= (0.01)^2 + (0.02)^2 + \cdots + (0.01)^2 = 0.067\,. \end{aligned}$$

18. Let A_i denote the event that member i does not serve on any of the three committees ($i = 1, \ldots, 8$). Then

$$\begin{aligned}
\Pr(A_i) &= \frac{\binom{7}{3}}{\binom{8}{3}} \cdot \frac{\binom{7}{4}}{\binom{8}{4}} \cdot \frac{\binom{7}{5}}{\binom{8}{5}} = \frac{5}{8} \cdot \frac{4}{8} \cdot \frac{3}{8} = a \\
\Pr(A_i \cap A_j) &= \frac{\binom{6}{3}}{\binom{8}{3}} \cdot \frac{\binom{6}{4}}{\binom{8}{4}} \cdot \frac{\binom{6}{5}}{\binom{8}{5}} = \left(\frac{5}{8} \cdot \frac{4}{7}\right)\left(\frac{4}{8} \cdot \frac{3}{7}\right)\left(\frac{3}{8} \cdot \frac{2}{7}\right) = \text{b} \text{ for } \text{i} < j, \\
\Pr(A_i \cap A_j \cap A_k) &= \frac{\binom{5}{3}}{\binom{8}{3}} \cdot \frac{\binom{5}{4}}{\binom{8}{4}} \cdot \frac{\binom{5}{5}}{\binom{8}{5}} = \left(\frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6}\right)\left(\frac{4}{8} \cdot \frac{3}{7} \cdot \frac{2}{6}\right)\left(\frac{3}{8} \cdot \frac{2}{7} \cdot \frac{1}{6}\right) \\
&= c \text{ for } i < j < k, \\
\Pr(A_i \cap A_j \cap A_k \cap A_\ell) &= 0, \; i < j < k < \ell.
\end{aligned}$$

Hence, by Theorem 1.10.2,

$$\Pr\left(\bigcup_{i=1}^{8} A_i\right) = 8a - \binom{8}{2} b + \binom{8}{3} c + 0$$

Therefore, the required probability is $1 - .7207 = .2793$.

19. Let E_i be the event that A and B are both selected for committee i ($i = 1, 2, 3$) and let $\Pr(E_i) = p_i$. Then

$$p_1 = \frac{\binom{6}{1}}{\binom{8}{3}} \approx 0.1071, \quad p_2 = \frac{\binom{6}{2}}{\binom{8}{4}} \approx 0.2143, \quad p_3 = \frac{\binom{6}{3}}{\binom{8}{5}} \approx 0.3571.$$

Since $E_1, E_2,$ *and* E_3 are independent, it follows from Theorem 1.10.1 that the required probability is

$$\begin{aligned}
\Pr(E_1 \cup E_2 \cup E_3) &= p_1 + p_2 + p_3 - p_1 p_2 - p_2 p_3 - p_1 p_3 + p_1 p_2 p_3 \\
&\approx 0.5490.
\end{aligned}$$

20. Let E denote the event that B wins. B will win if A misses on her first turn and B wins on her first turn, which has probability $\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)$, or if both players miss on their first turn and B then goes on to subsequently win, which has probability $\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\Pr(E)$. (See Exercise 17, Sec. 2.2.) Hence, $\Pr(E) = \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\Pr(E)$, and $\Pr(E) = \frac{5}{11}$. This problem could also be solved by summing the infinite series $\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^5\left(\frac{1}{6}\right) + \cdots$

21. A will win if he wins on his first toss (probability 1/2) or if all three players miss on their first tosses (probability 1/8) and then A subsequently wins. Hence,

$$\Pr(A \text{ wins}) = \frac{1}{2} + \frac{1}{8}\Pr(A \text{ wins}),$$

and $\Pr(A \text{ wins}) = 4/7$.

Similarly, B will win if A misses on his first toss and B wins on his first toss, or if all three players miss on their first tosses and then B subsequently wins. Hence,

$$\Pr(B \text{ wins}) = \frac{1}{4} + \frac{1}{8}\Pr(B \text{ wins}),$$

and $\Pr(B \text{ wins}) = 2/7$. Thus, $\Pr(C \text{ wins}) = 1 - 4/7 - 2/7 = 1/7$.

22. Let A_j denote the outcome of the jth roll. Then

$$\begin{aligned} & \Pr(X = x) = \Pr(A_2 \neq A_1, A_3 \neq A_2, \ldots, A_{x-1} \neq A_{x-2}, A_x = A_{x=1}). \\ = \; & \Pr(A_2 \neq A_1)\Pr(A_3 \neq A_2 \mid A_2 \neq A_1)\cdots\Pr(A_x = A_{x-1} \mid A_{x-1} \neq A_{x-2}, etc.). \\ = \; & \underbrace{\left(\frac{5}{6}\right)\cdots\left(\frac{5}{6}\right)}_{x-2 \text{ factors}}\left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^{x-2}\left(\frac{1}{6}\right). \end{aligned}$$

23. Let A be the event that the person you meet is a statistician, and let B be the event that he is shy. Then

$$\Pr(A \mid B) = \frac{(0.8)(0.1)}{(0.8)(0.1) + (0.15)(0.9)} = 0.372.$$

24. $\Pr(A \mid \text{ lemon}) = \dfrac{(0.05)(0.2)}{(0.05)(0.2) + (0.02)(0.5) + (0.1)(0.3)} = \frac{1}{5}.$

25. (a) $$\Pr(\text{Defective} \mid \text{Removed}) = \frac{(0.9)(0.3)}{(0.9)(0.3) + (0.2)(0.7)} = \frac{27}{41} = 0.659.$$

(b) $$\Pr(\text{Defective} \mid \text{Not Removed}) = \frac{(0.1)(0.3)}{(0.1)(0.3) + (0.8)(0.7)} = \frac{3}{59} = 0.051.$$

26. Let X and Y denote the number of tosses required on the first experiment and second experiment, respectively. Then $X = n$ if and only if the first $n-1$ tosses of the first experiment are tails and the nth toss is a head, which has probability $1/2^n$. Furthermore, $Y > n$ if and only if the first n tosses of the second experiment are all tails, which also has probability $1/2^n$.
Hence

$$\begin{aligned} \Pr(Y > X) &= \sum_{n=1}^{\infty} \Pr(Y > n \mid X = n)\Pr(X = n) \\ &= \sum_{n=1}^{\infty} \frac{1}{2^n}\cdot\frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{3}. \end{aligned}$$

27. Let A denote the event that the family has at least one boy, and B the event that it has at least one girl. Then

$$\begin{aligned}\Pr(B) &= 1-(1/2)^n,\\ \Pr(A\cap B) &= 1-\Pr(\text{All girls})-\Pr(\text{All boys}) = 1-(1/2)^n-(1/2)^n.\end{aligned}$$

Hence,

$$\Pr(A \mid B) = \frac{\Pr(A\cap B)}{\Pr(B)} = \frac{1-(1/2)^{n-1}}{1-(1/2)^n}$$

28. (a) Let X denote the number of heads, Then

$$\Pr(X=n-1 \mid X \ge n-2) = \frac{\Pr(X=n-1)}{\Pr(X\ge n-2)}$$

$$= \frac{\binom{n}{n-1}(1/2)^n}{\left[\binom{n}{n-2}+\binom{n}{n-1}+\binom{n}{n}\right](1/2)^n} = \frac{n}{\frac{n(n-1)}{2}+n+1} = \frac{2n}{n^2+n+2}.$$

(b) The required probability is the probability of obtaining exactly one head on the last two tosses, namely 1/2.

29. (a) Let X denote the number of aces selected.
Then

$$\begin{aligned}\Pr(X=i) &= \frac{\binom{4}{i}\binom{48}{13-i}}{\binom{52}{13}}, \quad i=0,1,2,3,4.\\ \Pr(X\ge 2 \mid X\ge 1) &= \frac{1-\Pr(X=0)-\Pr(X=1)}{1-\Pr(X=0)}\\ &\approx \frac{1-0.3038-0.4388}{1-0.3038} = 0.3697.\end{aligned}$$

(b) Let A denote the event that the ace of hearts and no other aces are obtained, and let H denote the event that the ace of hearts is obtained.
Then

$$\Pr(A) = \frac{\binom{48}{12}}{\binom{52}{13}} \approx 0.1097, \quad \Pr(H) = \frac{13}{52} = 0.25.$$

The required probability is

$$\frac{\Pr(H)-\Pr(A)}{\Pr(H)} \approx \frac{0.25-0.1097}{0.25} = 0.5612.$$

30. The probability that a particular letter, say letter A, will be placed in the correct envelope is $1/n$. The probability that none of the other $n-1$ letters will then be placed in the correct envelope is $q_{n-1} = 1 - p_{n-1}$. Therefore, the probability that only letter A, and no other letter, will be placed in the correct envelope is q_{n-1}/n. It follows that the probability that exactly one of the n letters will be placed in the correct envelope, without specifying which letter will be correctly placed is $nq_{n-1}/n = q_{n-1}$.

31. The probability that two specified letters will be placed in the correct envelopes is $1[n(n-1)]$. The probability that none of the other $n-2$ letters will then be placed in the correct envelopes is q_{n-2}. Therefore, the probability that only the two specified letters, and no other letters, will be placed in the correct envelopes is $\frac{1}{n(n-1)}q_{n-2}$. It follows that the probability that exactly two of the n letters will be placed in the correct envelopes, without specifying which pair will be correctly placed, is $\binom{n}{2}\frac{1}{n(n-1)}q_{n-2} = \frac{1}{2}q_{n-2}$.

32. The probability that exactly one student will be in class is

$$\Pr(A)\Pr(B^c) + \Pr(A^c)\Pr(B) = (0.8)(0.4) + (0.2)(0.6) = 0.44.$$

The probability that exactly one student will be in class and that student will be A is

$$\Pr(A)\Pr(B^c) = 0.32.$$

Hence, the required probability is $\frac{32}{44} = \frac{8}{11}$.

33. By Exercise 3 of Sec. 1.10, the probability that a family subscribes to exactly one of the three newspapers is 0.45. As can be seen from the solution to that exercise, the probability that a family subscribes only to newspaper A is 0.35. Hence, the required probability is $35/45 = 7/9$.

34. A more reasonable analysis by prisoner A might proceed as follows: The pair to be executed is equally likely to be (A, B), (A, C), or (B, C). If it is (A, B) or (A, C), the jailer will surely respond B or C, respectively. If it is (B, C), the jailer is equally likely to respond B or C. Hence, if the jailer responds B, the conditional probability that the pair to be executed is (A, B) is

$$\begin{aligned}
\Pr[(A, B) \mid \text{response}] &= \frac{1 \cdot \Pr(A, B)}{1 \cdot \Pr(A, B) + 0 \cdot \Pr(A, C) + \frac{1}{2}\Pr(B, C)} \\
&= \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3}.
\end{aligned}$$

Thus, the probability that A will be executed is the same $2/3$ as it was before he questioned the jailer. This answer will change if the probability that the jailer will respond B, given (B, C), is assumed to be some value other than $1/2$.

35. The second situation, with stakes of two dollars, is equivalent to the situation in which A and B have initial fortunes of 25 dollars and bet one dollar on each play. In the notation of Sec. 2.4, we have $i = 50$ and $k = 100$ in the first situation and $i = 25$ and $k = 50$ in the second situation. Hence, if $p = 1/2$, it follows from Eq. (2.4.6) that gambler A has the same probability $1/2$ of ruining gambler B in either

situation. If $p \neq 1/2$, then it follows from Eq. (2.4.9) that the probabilities α_1 and α_2 of winning in the two situations equal the values

$$\begin{aligned}
\alpha_1 &= \frac{([1-p]/p)^{50}-1}{([1-p]/p)^{100}-1} = \frac{1}{([1-p]/p)^{50}+1}, \\
\alpha_2 &= \frac{([1-p]/p)^{25}-1}{([1-p]/p)^{50}-1} = \frac{1}{([1-p]/p)^{25}+1}.
\end{aligned}$$

Hence, if $p < 1/2$, then $([1-p]/p) > 1$ and $\alpha_2 > \alpha_1$. If $p > 1/2$, then $([1-p]/p) < 1$ and $\alpha_1 > \alpha_2$.

36. (a) Since each candidate is equally likely to appear at each point in the sequence, the one who happens to be the best out of the first i has probability r/i of appearing in the first r interviews when $i > r$.

(b) If $i \leq r$, then A and B_i are disjoint and $\Pr(A \cap B_i) = 0$ because we cannot hire any of the first r candidates. So $\Pr(A|B_i) = \Pr(A \cap B_i)/\Pr(B_i) = 0$. Next, let $i > r$ and assume that B_i occurs. Let C_i denote the event that we keep interviewing until we see candidate i. If C_i also occurs, then we shall rank candidate i higher than any of the ones previously seen and the algorithm tells us to stop and hire candidate i. In this case A occurs. This means that $B_i \cap C_i \subset A$. However, if C_i fails, then we shall hire someone before we get to interview candidate i and A will not occur. This means that $B_i \cap C_i^c \cap A = \emptyset$. Since $B_i \cap A = (B_i \cap C_i \cap A) \cup (B_i \cap C_i^c \cap A)$, we have $B_i \cap A = B_i \cap C_i$ and $\Pr(B_i \cap A) = \Pr(B_i \cap C_i)$. So $\Pr(A|B_i) = \Pr(C_i|B_i)$. Conditional on B_i, C_i occurs if and only if the best of the first $i-1$ candidates appears in the first r positions. The conditional probability of C_i given B_i is then $r/(i-1)$.

(c) If we use the value $r > 0$ to determine our algorithm, then we can compute

$$p_r = \Pr(A) = \sum_{i=1}^{n} \Pr(B_i)\Pr(A|B_i) = \sum_{i=r+1}^{n} \frac{1}{n}\frac{r}{i-1} = \frac{r}{n}\sum_{i=r+1}^{n}\frac{1}{i-1}.$$

For $r = 0$, if we take $r/r = 1$, then only the first term in the sum produces a nonzero result and $p_0 = 1/n$. This is indeed the probability that the first candidate will be the best one seen so far when the first interview occurs.

(d) Using the formula for p_r with $r > 0$, we have

$$q_r = p_r - p_{r-1} = \frac{1}{n}\left[\sum_{i=r+1}^{n}\frac{1}{i-1} - 1\right],$$

which clearly decreases as r increases because the terms in the sum are the same for all r, but there are fewer terms when r is larger. Since all the terms are positive, q_r is strictly decreasing.

(e) Since $p_r = q_r + p_{r-1}$ for $r \geq 1$, we have that $p_r = p_0 + q_1 + \cdots + q_r$. If there exists r such that $q_r \leq 0$, then $q_j < 0$ for all $j > r$ and $p_j \leq p_{r-1}$ for all $j \geq r$. On the other hand, for each r such that $q_r > 0$, $p_r > p_{r-1}$. Hence, we should choose r to be the last value such that $q_r > 0$.

(f) For $n = 10$, the first few values of q_r are

r	1	2	3	4
q_r	0.1829	0.0829	0.0390	−0.0004

So, we should use $r = 3$. We can then compute $p_3 = 0.3987$.

Chapter 3

Random Variables and Distributions

3.1 Random Variables and Discrete Distributions

Solutions to Exercises

1. Each of the 11 integers from 10 to 20 has the same probability of being the value of X. Six of the 11 integers are even, so the probability that X is even is 6/11.

2. The sum of the values of $f(x)$ must be equal to 1. Since $\sum_{x=1}^{5} f(x) = 15c$, we must have $c = 1/15$.

3. By looking over the 36 possible outcomes enumerated in Example 1.6.5, we find that $X = 0$ for 6 outcomes, $X = 1$ for 10 outcomes, $X = 2$ for 8 outcomes, $X = 3$ for 6 outcomes, $X = 4$ for 4 outcomes, and $X = 5$ for 2 outcomes. Hence, the p.f. $f(x)$ is as follows:

x	0	1	2	3	4	5
$f(x)$	3/18	5/18	4/18	3/18	2/18	1/18

4. For $x = 0, 1, \ldots, 10$, the probability of obtaining exactly x heads is $\binom{10}{x}\left(\frac{1}{2}\right)^{10}$.

5. For $x = 2, 3, 4, 5$, the probability of obtaining exactly x red balls is $\binom{7}{x}\binom{3}{5-x}\Big/\binom{10}{5}$.

6. The desired probability is the sum of the entries for $k = 0$, 1, 2, 3, 4, and 5 in that part of the table of binomial probabilities given in the back of the book corresponding to $n = 15$ and $p = 0.5$. The sum is 0.1509.

7. Suppose that a machine produces a defective item with probability 0.7 and produces a nondefective item with probability 0.3. If X denotes the number of defective items that are obtained when 8 items are inspected, then the random variable X will have the binomial distribution with parameters $n = 8$ and $p = 0.7$. By the same reasoning, however, if Y denotes the number of nondefective items that are obtained, then Y will have the binomial distribution with parameters $n = 8$ and $p = 0.3$. Furthermore, $Y = 8 - X$. Therefore, $X \geq 5$ if and only if $Y \leq 3$ and it follows that $\Pr(X \geq 5) = \Pr(Y \leq 3)$. Probabilities for the binomial distribution with $n = 8$ and $p = 0.3$ are given in the table in the back of the book. The value of $\Pr(Y \leq 3)$ will be the sum of the entries for $k = 0, 1, 2$, and 3.

8. The number of red balls obtained will have the binomial distribution with parameters $n = 20$ and $p = 0.1$. The required probability can be found from the table of binomial probabilities in the back of the book. Add up the numbers in the $n = 20$ and $p = 0.1$ section from $k = 4$ to $k = 20$. Or add up the numbers from $k = 0$ to $k = 3$ and subtract the sum from 1. The answer is 0.1330.

9. We need $\sum_{x=0}^{\infty} f(x) = 1$, which means that $c = 1/\sum_{x=0}^{\infty} 2^{-x}$. The last sum is known from Calculus to equal $1/(1-1/2) = 2$, so $c = 1/2$.

10. (a) The p.f. of X is $f(x) = c(x+1)(8-x)$ for $x = 0, \ldots, 7$ where c is chosen so that $\sum_{x=0}^{7} f(x) = 1$. So, c is one over $\sum_{x=0}^{7}(x+1)(8-x)$, which sum equals 120, so $c = 1/120$. That is $f(x) = (x+1)(8-x)/120$ for $x = 0, \ldots, 7$.
   (b) $\Pr(X \geq 5) = [(5+1)(8-5) + (6+1)(8-6) + (7+1)(8-7)]/120 = 1/3$.

11. In order for the specified function to be a p.f., it must be the case that $\sum_{x=1}^{\infty} \frac{c}{x} = 1$ or equivalently $\sum_{x=1}^{\infty} \frac{1}{x} = \frac{1}{c}$. But $\sum_{x=1}^{\infty} \frac{1}{x} = \infty$, so there cannot be such a constant c.

3.2 Continuous Distributions

Commentary

This section ends with a brief discussion of probability distributions that are neither discrete nor continuous. Although such distributions have great theoretical interest and occasionally arise in practice, students can go a long way without actually concerning themselves about these distributions.

Solutions to Exercises

1. We compute $\Pr(X \leq 8/27)$ by integrating the p.d.f. from 0 to 8/27.

$$\Pr\left(X \leq \frac{8}{27}\right) = \int_0^{8/27} \frac{2}{3} x^{-1/3} dx = x^{2/3}\Big|_0^{8/27} = \frac{4}{9}.$$

2. The p.d.f. has the appearance of Fig. S.3.1.

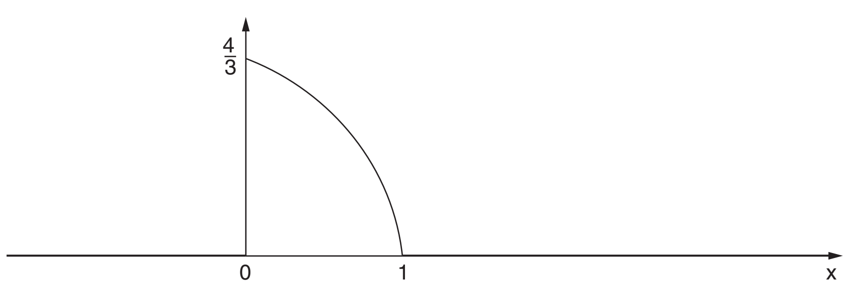

Figure S.3.1: Figure for Exercise 2 of Sec. 3.2.

   (a) $\Pr\left(X < \frac{1}{2}\right) = \int_0^{1/2} 4(1-x^3)dx/3 = 0.6458.$

(b) $\Pr\left(\frac{1}{4} < X < \frac{3}{4}\right) = \int_{1/4}^{3/4} 4(1-x^3)dx/3 = 0.5625.$

(c) $\Pr\left(X > \frac{1}{3}\right) = \int_{1/3}^{1} 4(1-x^3)dx/3 = 0.5597.$

3. The p.d.f. has the appearance of Fig. S.3.2.

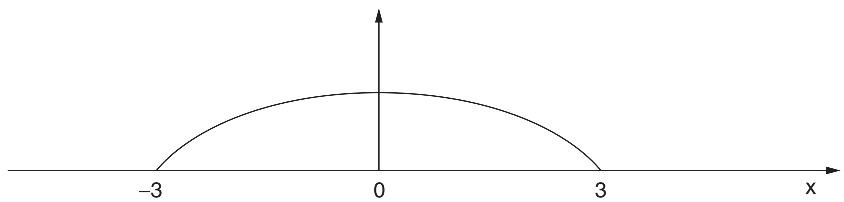

Figure S.3.2: Figure for Exercise 3 of Sec. 3.2.

(a) $\Pr(X < 0) = \frac{1}{36}\int_{-3}^{0}(9-x^2)dx = 0.5.$

(b) $\Pr(-1 < X < 1) = \frac{1}{36}\int_{-1}^{1}(9-x^2)dx = 0.4815.$

(c) $\Pr(X > 2) = \frac{1}{36}\int_{2}^{3}(9-x^2)dx = 0.07407.$

The answer in part (a) could also be obtained directly from the fact that the p.d.f. is symmetric about the point $x = 0$. Therefore, the probability to the left of $x = 0$ and the probability to the right of $x = 0$ must each be equal to 1/2.

4. (a) We must have

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{1}^{2} cx^2 dx = \frac{7}{3}c = 1.$$

Therefore, $c = 3/7$. This p.d.f. has the appearance of Fig. S.3.3.

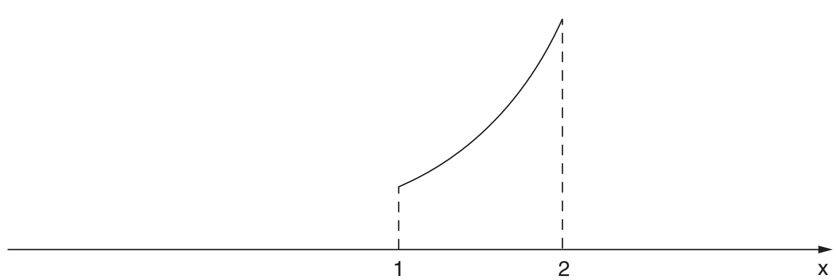

Figure S.3.3: Figure for Exercise 4a of Sec. 3.2.

(b) $\int_{3/2}^{2} f(x)dx = 37/56.$

5. (a) $\int_{0}^{t} \frac{1}{8}x\,dx = 1/4$, or $t^2/16 = 1/4$. Hence, $t = 2$.

(b) $\int_{t}^{4} (x/8)\,dx = 1/2$, or $1 - t^2/16 = 1/2$. Hence, $t = \sqrt{8}$.

6. The value of X must be between 0 and 4. We will have

$$Y = \begin{cases} 0 & \text{if } 0 \le X < 1/2, \\ 1 & \text{if } 1/2 < X < 3/2, \\ 2 & \text{if } 3/2 < X < 5/2, \\ 3 & \text{if } 5/2 < X < 7/2, \\ 4 & \text{if } 7/2 < X \le 4. \end{cases}$$

We need not worry about how to define Y if $X \in \{1/2, 3/2, 5/2, 7/2\}$, because the probability that X will be equal to one of these four values is 0. It now follows that

$$\begin{aligned} \Pr(Y = 0) &= \int_0^{1/2} f(x)dx = \frac{1}{64}, \\ \Pr(Y = 1) &= \int_{1/2}^{3/2} f(x)dx = \frac{1}{8}, \\ \Pr(Y = 2) &= \int_{3/2}^{5/2} f(x)dx = \frac{1}{4}, \\ \Pr(Y = 3) &= \int_{5/2}^{7/2} f(x)dx = \frac{3}{8}, \\ \Pr(Y = 4) &= \int_{7/2}^{4} f(x)dx = \frac{15}{64}. \end{aligned}$$

7. Since the uniform distribution extends over an interval of length 10 units, the value of the p.d.f. must be 1/10 throughout the interval. Hence,

$$\int_0^7 f(x)\,dx = \frac{7}{10}.$$

8. (a) We must have

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} c\exp(-2x)\,dx = \frac{1}{2}c = 1.$$

Therefore, $c = 2$. This p.d.f. has the appearance of Fig. S.3.4.

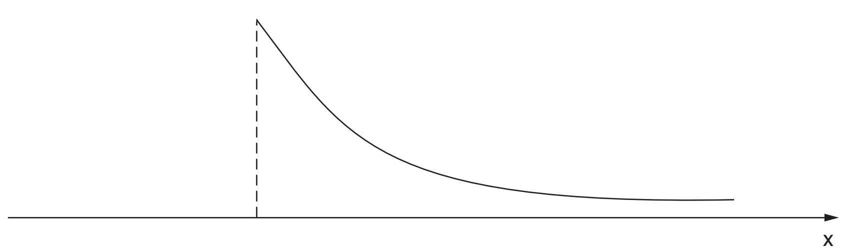

Figure S.3.4: Figure for Exercise 8a of Sec. 3.2.

(b) $\int_1^2 f(x)\,dx = \exp(-2) - \exp(-4)$.

9. Since $\int_0^{\infty} 1/(1+x)\,dx = \infty$, there is no constant c such that $\int_0^{\infty} f(x)\,dx = 1$.

10. (a) We must have

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 \frac{c}{(1-x)^{1/2}}\,dx = 2c = 1.$$

Therefore $c = 1/2$. This p.d.f. has the appearance of Fig. S.3.5.

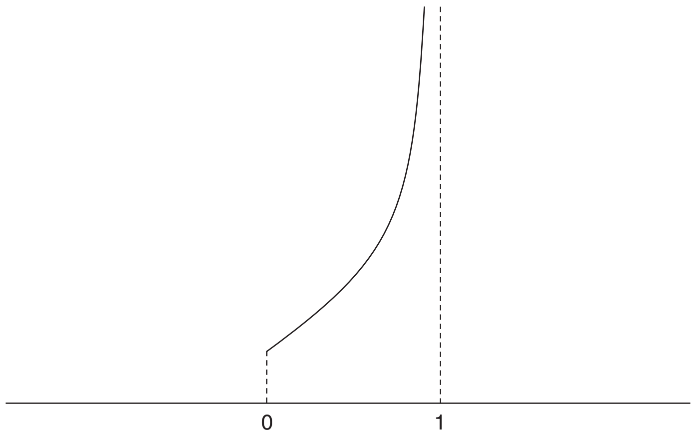

Figure S.3.5: Figure for Exercise 10a of Sec. 3.2.

It should be noted that although the values of $f(x)$ become arbitrarily large in the neighborhood of $x = 1$, the total area under the curve is equal to 1. It is seen, therefore, that the values of a p.d.f. can be greater than 1 and, in fact, can be arbitrarily large.

(b) $\int_0^{1/2} f(x)\,dx = 1 - (1/2)^{1/2}$.

11. Since $\int_0^1 (1/x)dx = \infty$, there is no constant c such that $\int_0^1 f(x)\,dx = 1$.

12. We shall find the c.d.f. of Y and evaluate it at 50. The c.d.f. of a random variable Y is $F(y) = \Pr(Y \le y)$. In Fig. 3.1, on page 94 of the text, the event $\{Y \le y\}$ has area $(y-1)\times(200-4) = 196(y-1)$ if $1 \le y \le 150$. We need to divide this by the area of the entire rectangle, 29,204. The c.d.f. of Y is then

$$F(y) = \begin{cases} 0 & \text{for } y < 1, \\ \dfrac{196(y-1)}{29204} & \text{for } 1 \le y \le 150, \\ 1 & \text{for } y > 150. \end{cases}$$

So, in particular, $\Pr(Y \le 50) = 0.3289$.

13. We find $\Pr(X < 20) = \int_0^{20} cx dx = 200c$. Setting this equal to 0.9 yields $c = 0.0045$.

3.3 The Cumulative Distribution Function

Commentary

This section includes a discussion of quantile functions. These arise repeatedly in the construction of hypothesis tests and confidence intervals later in the book.

Solutions to Exercises

1. The c.d.f. $F(x)$ of X is 0 for $x < 0$. It jumps to $0.3 = \Pr(X = 0)$ at $x = 0$, and it jumps to 1 and stays there at $x = 1$. The c.d.f. is sketched in Fig. S.3.6.

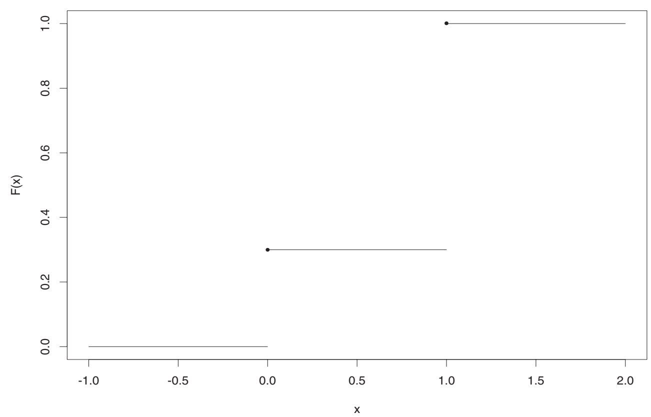

Figure S.3.6: C.d.f. of X in Exercise 1 of Sec. 3.3.

2. The c.d.f. must have the appearance of Fig. S.3.7.

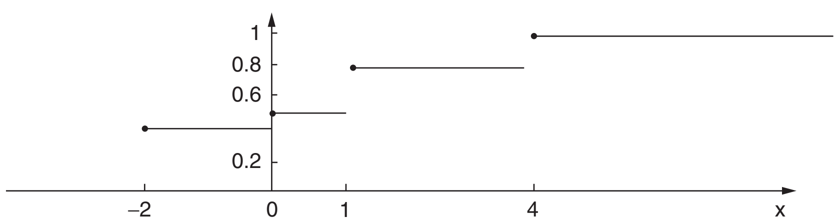

Figure S.3.7: C.d.f. for Exercise 2 of Sec. 3.3.

3. Here $\Pr(X = n) = 1/2^n$ for $n = 1, 2, \ldots$. Therefore, the c.d.f. must have the appearance of Fig. S.3.8.

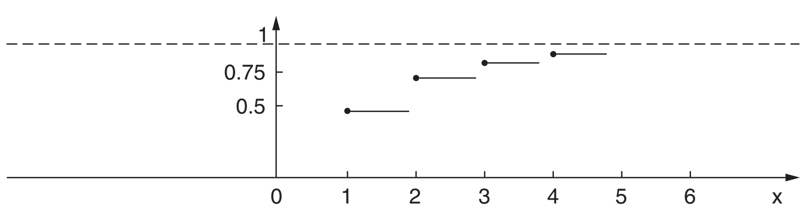

Figure S.3.8: C.d.f. for Exercise 3 of Sec. 3.3.

4. The numbers can be read off of the figure or found by subtracting two numbers off of the figure.

   (a) The jump at $x = -1$ is $F(-1) - F(-1^-) = 0.1$.

   (b) The c.d.f. to the left of $x = 0$ is $F(0^-) = 0.1$.

   (c) The c.d.f. at $x = 0$ is $F(0) = 0.2$.

(d) There is no jump at $x = 1$, so $\Pr(X = 1) = 0$.

(e) $F(3) - F(0) = 0.6$.

(f) $F(3^-) - F(0) = 0.4$.

(g) $F(3) - F(0^-) = 0.7$.

(h) $F(2) - F(1) = 0$.

(i) $F(2) - F(1^-) = 0$.

(j) $1 - F(5) = 0$.

(k) $1 - F(5^-) = 0$.

(l) $F(4) - F(3^-) = 0.2$.

5. $$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & \text{for } x < 0, \\ \frac{2}{9}x & \text{for } 0 < x < 3, \\ 0 & \text{for } x > 3. \end{cases}$$

The value of $f(x)$ at $x = 0$ and $x = 3$ is irrelevant. This p.d.f. has the appearance of Fig. S.3.9.

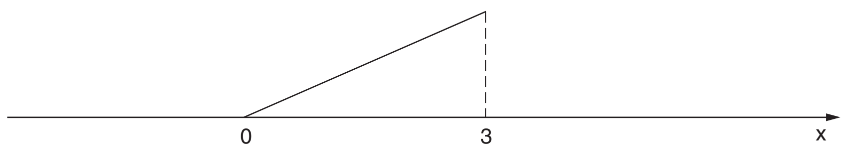

Figure S.3.9: Figure for Exercise 5 of Sec. 3.3.

6. $$f(x) = \frac{dF(x)}{dx} = \begin{cases} \exp(x - 3) & \text{for } x < 3, \\ 0 & \text{for } x > 3. \end{cases}$$

The value of $f(x)$ at $x = 3$ is irrelevant. This p.d.f. has the appearance of Fig. S.3.10.

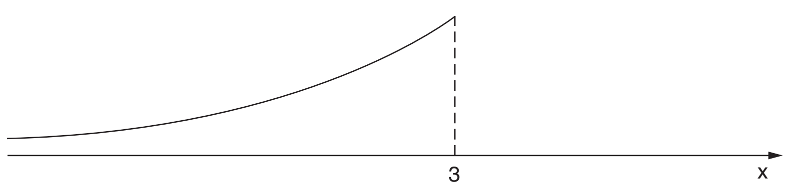

Figure S.3.10: Figure for Exercise 6 of Sec. 3.3.

It should be noted that although this p.d.f. is positive over the unbounded interval where $x < 3$, the total area under the curve is finite and is equal to 1.

7. The c.d.f. equals 0 for $x < -2$ and it equals 1 for $x > 8$. For $-2 \le x \le 8$, the c.d.f. equals

$$F(x) = \int_{-2}^{x} \frac{dy}{10} = \frac{x+2}{10}.$$

The c.d.f. has the appearance of Fig. S.3.11.

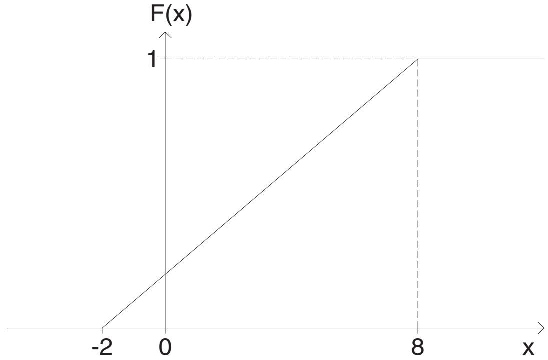

Figure S.3.11: Figure for Exercise 7 of Sec. 3.3.

8. $\Pr(Z \le z)$ is the probability that Z lies within a circle of radius z centered at the origin. This probability is

$$\frac{\text{Area of circle of radius } z}{\text{Area of circle of radius } 1} = z^2, \quad \text{for } 0 \le z \le 1.$$

The c.d.f. is plotted in Fig. S.3.12.

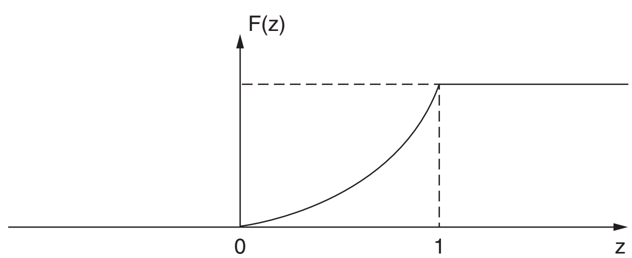

Figure S.3.12: C.d.f. for Exercise 8 of Sec. 3.3.

9. $\Pr(Y = 0) = \Pr(X \le 1) = 1/5$ and $\Pr(Y = 5) = \Pr(X \ge 3) = 2/5$. Also, Y is distributed uniformly between $Y = 1$ and $Y = 3$, with a total probability of $2/5$. Therefore, over this interval $F(y)$ will be linear with a total increase of $2/5$. The c.d.f. is plotted in Fig. S.3.13.

10. To find the quantile function $F^{-1}(p)$ when we know the c.d.f., we can set $F(x) = p$ and solve for x.

$$\frac{x}{1+x} = p; \quad x = p + px; \quad x(1-p) = p; \quad x = \frac{p}{1-p}.$$

The quantile function is $F^{-1}(p) = p/(1-p)$.

11. As in Exercise 10, we set $F(x) = p$ and solve for x.

$$\frac{1}{9}x^2 = p; \quad x^2 = 9p; \quad x = 3p^{1/2}.$$

The quantile function of X is $F^{-1}(p) = 3p^{1/2}$.

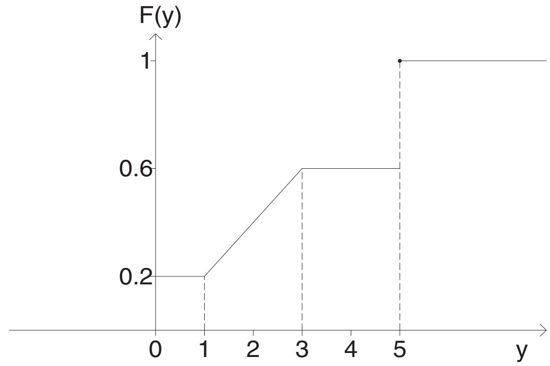

Figure S.3.13: C.d.f. for Exercise 9 of Sec. 3.3.

12. Once again, we set $F(x) = p$ and solve for x.

$$\exp(x-3) = p; \quad x - 3 = \log(p); \quad x = 3 + \log(p).$$

The quantile function of X is $F^{-1}(p) = 3 + \log(p)$.

13. VaR at probability level 0.95 is the negative of the 0.05 quantile. Using the result from Example 3.3.8, the 0.05 quantile of the uniform distribution on the interval $[-12, 24]$ is $0.05 \times 24 - 0.95 \times 12 = -10.2$. So, VaR at probability level 0.95 is 10.2

14. Using the table of binomial probabilities in the back of the book, we can compute the c.d.f. F of the binomial distribution with parameters 10 and 0.2. We then find the first values of x such that $F(x) \geq 0.25$, $F(x) \geq 0.5$, and $F(x) \geq 0.75$. The first few distinct values of the c.d.f. are

x	0	1	2	3
$F(x)$	0.0174	0.3758	0.6778	0.8791

So, the quartiles are 1 and 3, while the median is 2.

15. Since $f(x) = 0$ for $x \leq 0$ and for $x \geq 1$, the c.d.f. $F(x)$ will be flat (0) for $x \leq 0$ and flat (1) for $x \geq 1$. Between 0 and 1, we compute $F(x)$ by integrating the p.d.f. For $0 < x < 1$,

$$F(x) = \int_0^x 2y dy = x^2.$$

The requested plot is identical to Fig. S.3.12 for Exercise 8 in this section.

16. For each $0 < p < 1$, we solve for x in the equation $F(x) = p$, with F specified in (3.3.2):

$$\begin{aligned} p &= 1 - \frac{1}{1+x}, \\ \frac{1}{1-p} &= 1 + x, \\ \frac{1}{1-p} - 1 &= x. \end{aligned}$$

The quantile function is $F^{-1}(p) = 1/(1-p) - 1$ for $0 < p < 1$.

17. (a) Let $0 < p_1 < p_2 < 1$. Define $A_i = \{x : F(x) \geq p_i\}$ for $i = 1, 2$. Since $p_1 < p_2$ and F is nondecreasing, it follows that $A_2 \subset A_1$. Hence, the smallest number in A_1 (which equals $F^{-1}(p_1)$ by definition) is no greater than the smallest number in A_2 (which equals $F^{-1}(p_2)$ by definition. That is, $F^{-1}(p_1) \leq F^{-1}(p_2)$, and the quantile function is nondecreasing.

    (b) Let $x_0 = \lim_{\substack{p \to 0 \\ p > 0}} F^{-1}(p)$. We are asked to prove that x_0 is the greatest lower bound of the set $C = \{c : F(c) > 0\}$. First, we show that no $x > x_0$ is a lower bound on C. Let $x > x_0$ and $x_1 = (x + x_0)/2$. Then $x_0 < x_1 < x$. Because $F^{-1}(p)$ is nondecreasing, it follows that there exists $p > 0$ such that $F^{-1}(p) < x_1$, which in turn implies that $p \leq F(x_1)$, and $F(x_1) > 0$. Hence $x_1 \in C$, and x is not a lower bound on C. Next, we prove that x_0 is a lower bound on C. Let $x \in C$. We need only prove that $x_0 \leq x$. Because $F^{-1}(p)$ is nondecreasing, we must have $\lim_{\substack{p \to 0 \\ p > 0}} F^{-1}(p) \leq F^{-1}(q)$ for all $q > 0$. Hence, $x_0 \leq F^{-1}(p)$ for all $q > 0$. Because $x \in C$, we have $F(x) > 0$. Let $q = F(x)$ so that $q > 0$. Then $x_0 \leq F^{-1}(q) \leq x$. The proof that x_1 is the least upper bound on the set of all d such that $F(d) < 1$ is very similar.

    (c) Let $0 < p < 1$. Because F^{-1} is nondecreasing, $F^{-1}(p^-)$ is the least upper bound on the set $C = \{F^{-1}(q) : q < p\}$. We need to show that $F^{-1}(p)$ is also that least upper bound. Clearly, $F^{-1}(p)$ is an upper bound, because F^{-1} is nondecreasing and $p > q$ for all $q < p$. To see that $F^{-1}(p)$ is the least upper bound, let y be an upper bound. We need to show $F^{-1}(p) \leq y$. By definition, $F^{-1}(p)$ is the greatest lower bound on the set $D = \{x : F(x) \geq p\}$. Because y is an upper bound on C, it follows that $F^{-1}(q) \leq y$ for all $q < p$. Hence, $F(y) \geq q$ for all $q < p$. Because F is nondecreasing, we have $F(y) \geq p$, hence $y \in D$, and $F^{-1}(p) \leq y$.

18. We know that $\Pr(X = c) = F(c) - F(c^-)$. We will prove that $p_1 = F*(c)$ and $p_0 = F(c^-)$. For each $p \in (0, 1)$ define
$$C_p = \{x : F(x) \geq p\}.$$
Condition (i) says that, for every $p \in (p_0, p_1)$, c is the greatest lower bound on the set C_p. Hence $F(c) \geq p$ for all $p < p_1$ and $F(c) \geq p_1$. If $F(c) > p_1$, then for $p = (p_1 + F(c))/2$, $F^{-1}(p) \leq c$, and condition (iii) rules this out. So $F(c) = p_1$. The rest of the proof is broken into two cases. First, if $p_0 = 0$, then for every $\epsilon > 0$, c is the greatest lower bound on the set C_ϵ. This means that $F(x) < \epsilon$ for all $x < c$. Since this is true for all $\epsilon > 0$, $F(x) = 0$ for all $x < c$, and $F(c^-) = 0 = p_0$. For the second case, assume $p_0 > 0$. Condition (ii) says $F^{-1}(p_0) < c$. Since $F^{-1}(p_0)$ is the greatest lower bound on the set C_{p_0}, we have $F(x) < p_0$ for all $x < c$. Hence, $p_0 \geq F(c^-)$. Also, for all $p < p_0$, $p \leq F(c^-)$, hence $p_0 \leq F(c^-)$. Together, the last two inequalities imply $p_0 = F(c^-)$.

19. First, we show that $F^{-1}(F(x)) \leq x$. By definition $F^{-1}(F(x))$ is the smallest y such that $F(y) \geq F(x)$. Clearly $F(x) \geq F(x)$, hence $F^{-1}(F(x)) \leq x$. Next, we show that, if $p > F(x)$, then $F^{-1}(p) > x$. Let $p > F(x)$. By Exercise 17, we know that $F^{-1}(p) \geq x$. By definition, $F^{-1}(p)$ is the greatest lower bound on the set $C_p = \{y : F(y) \geq p\}$. All $y \in C_p$ satisfy $F(y) > (p + F(x))/2$. Since F is continuous from the right, $F(F^{-1}(p)) \geq (p + F(x))/2$. But $F(x) < (p + F(x))/2$, so $x \neq F^{-1}(p)$, hence $F^{-1}(p) > x$.

20. Figure S.3.14 has the plotted c.d.f., which equals $0.0045x^2/2$ for $0 < x < 20$. On the plot, we see that $F(10) = 0.225$.

3.4 Bivariate Distributions

Commentary

The bivariate distribution function is mentioned at the end of this section. The only part of this discussion that is used later in the text is the fact that the joint p.d.f. is the second mixed partial derivative of the

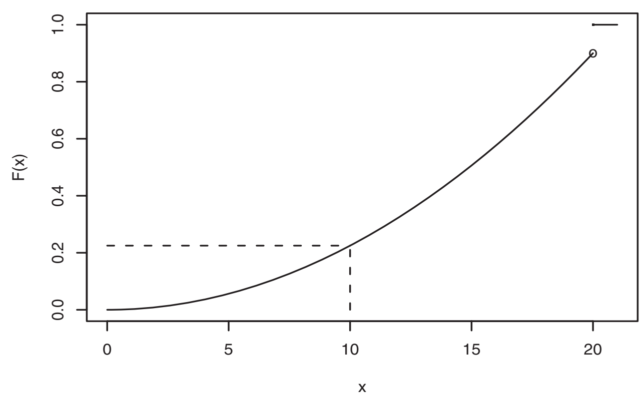

Figure S.3.14: C.d.f. for Exercise 20 of Sec. 3.3.

bivariate p.d.f. (in the discussion of functions of two or more random variables is Sec. 3.9.) If an instructor prefers not to discuss how to calculate probabilities of rectangles and is not going to cover functions of two or more random variables, there will be no loss of continuity.

Solutions to Exercises

1. (a) Let the constant value of the p.d.f. on the rectangle be c. The area of the rectangle is 2. So, the integral of the p.d.f. is $2c = 1$, hence $c = 1/2$.

   (b) $\Pr(X \geq Y)$ is the integral of the p.d.f. over that part of the rectangle where $x \geq y$. This region is shaded in Fig. S.3.15. The region is a trapezoid with area $1 \times (1 + 2)/2 = 1.5$. The integral of the

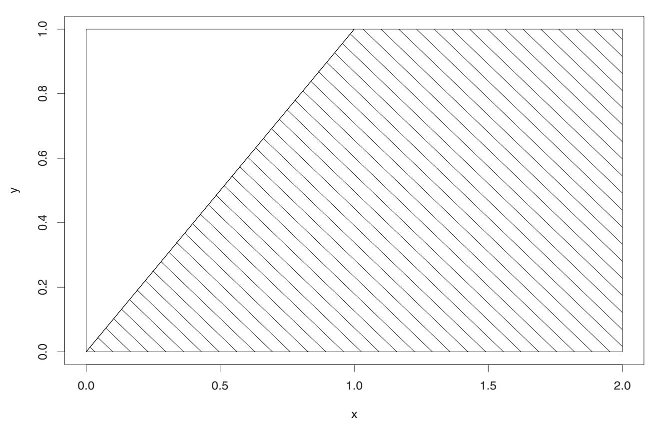

Figure S.3.15: Region where $x \geq y$ in Exercise 1b of Sec. 3.4.

   constant 1/2 over this region is then $0.75 = \Pr(X \geq Y)$.

2. The answers are found by summing the following entries in the table:

   (a) The entries in the third row of the table: 0.27.

(b) The last three columns of the table: 0.53.

(c) The nine entries in the upper left corner of the table: 0.69

(d) (0, 0), (1, 1), (2, 2), and (3, 3): 0.3.

(e) (1, 0), (2, 0), (3, 0), (2, 1), (3, 1), (3, 2): 0.25.

3. (a) If we sum $f(x, y)$ over the 25 possible pairs of values (x, y), we obtain $40c$. Since this sum must be equal to 1, it follows that $c = 1/40$.

(b) $f(0, -2) = (1/40) \cdot 2 = 1/20$.

(c) $\Pr(X = 1) = \sum_{y=-2}^{2} f(1, y) = 7/40$.

(d) The answer is found by summing $f(x, y)$ over the following pairs: $(-2, -2)$, $(-2, -1)$, $(-1, -2)$, $(-1, -1)$, $(-1, 0)$, $(0, -1)$, (0, 0), (0, 1), (1, 0), (1, 1), (1, 2), (2, 1), and (2, 2). The sum is 0.7.

4. (a) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = \int_0^1\int_0^2 cy^2\,dx\,dy = 2c/3$. Since the value of this integral must be 1, it follows that $c = 3/2$.

(b) The region over which to integrate is shaded in Fig. S.3.16.

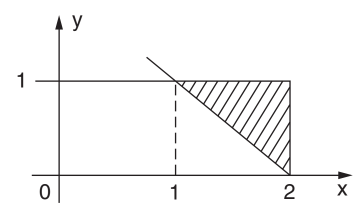

Figure S.3.16: Region of integration for Exercise 4b of Sec. 3.4.

$$\Pr(X + Y > 2) = \int\int_{\substack{\text{shaded}\\\text{region}}} f(x, y)\,dx\,dy = \int_1^2\int_{2-x}^1 \frac{3}{2}y^2\,dy\,dx = \frac{3}{8}.$$

(c) The region over which to integrate is shaded in Fig. S.3.17.

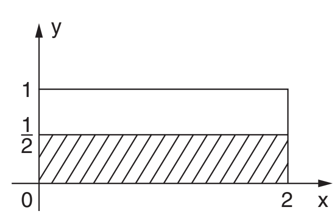

Figure S.3.17: Region of integration for Exercise 4c of Sec. 3.4.

$$\Pr\left(Y < \frac{1}{2}\right) = \int_0^2\int_0^{1/2} \frac{3}{2}y^2\,dy\,dx = \frac{1}{8}.$$

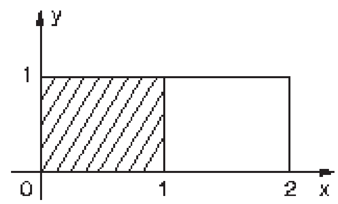

Figure S.3.18: Region of integration for Exercise 4d of Sec. 3.4.

(d) The region over which to integrate is shaded in Fig. S.3.18.

$$\Pr(X \leq 1) = \int_0^1 \int_0^1 \frac{3}{2} y^2 \, dy \, dx = \frac{1}{2}.$$

(e) The probability that (X, Y) will lie on the line $x = 3y$ is 0 for every continuous joint distribution.

5. (a) By sketching the curve $y = 1 - x^2$, we find that $y \leq 1 - x^2$ for all points on or below this curve. Also, $y \geq 0$ for all points on or above the x-axis. Therefore, $0 \leq y \leq 1 - x^2$ only for points in the shaded region in Fig. S.3.19.

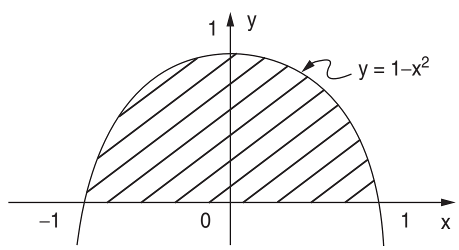

Figure S.3.19: Figure for Exercise 5a of Sec. 3.4.

Hence,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = \int_{-1}^{1} \int_0^{1-x^2} c\,(x^2 + y) \, dy \, dx = \frac{4}{5} c \cdot$$

Therefore, $c = 5/4$.

(b) Integration is done over the shaded region in Fig. S.3.20.

$$\Pr\left(0 \leq X \leq \frac{1}{2}\right) = \underset{\substack{\text{shaded}\\\text{region}}}{\int \int} f(x, y) \, dx \, dy = \int_0^{\frac{1}{2}} \int_0^{1-x^2} \frac{5}{4}(x^2 + y) \, dy \, dx = \frac{79}{256}.$$

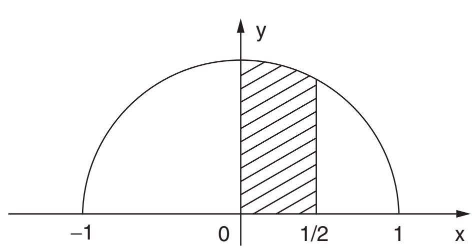

Figure S.3.20: Region of integration for Exercise 5b of Sec. 3.4.

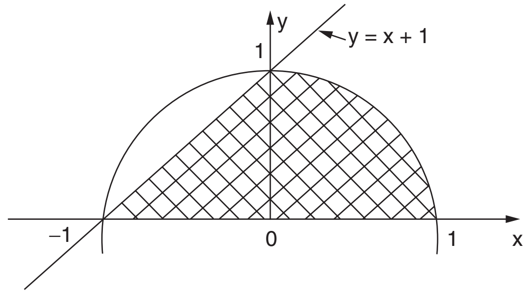

Figure S.3.21: Region of integration for Exercise 5c of Sec. 3.4.

(c) The region over which to integrate is shaded in Fig. S.3.21.

$$\begin{aligned}\Pr(Y \le X+1) &= \underset{\substack{\text{shaded}\\\text{region}}}{\int\int} f(x,y)\,dx\,dy = 1 - \underset{\substack{\text{unshaded}\\\text{region}}}{\int\int} f(x,y)\,dx\,dy \\ &= 1 - \int_{-1}^{0}\int_{x+1}^{1-x^2} \frac{5}{4}(x^2+y)\,dy\,dx = \frac{13}{16}.\end{aligned}$$

(d) The probability that (X, Y) will lie on the curve $y = x^2$ is 0 for every continuous joint distribution.

6. (a) The region S is the shaded region in Fig. S.3.22. Since the area of S is 2, and the joint p.d.f. is to

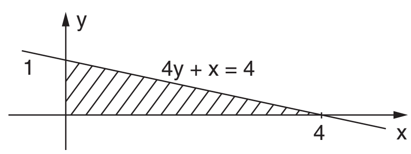

Figure S.3.22: Figure for Exercise 6a of Sec. 3.4.

be constant over S, then the value of the constant must be 1/2.

(b) The probability that (X, Y) will belong to any subset S_0 is proportional to the area of that subset. Therefore,

$$\Pr[(X,Y) \in S_0] = \int_{S_0}\int \frac{1}{2}\,dx\,dy = \frac{1}{2}(\text{area of } S_0) = \frac{\alpha}{2}.$$

7. (a) $\Pr(X \le 1/4)$ will be equal to the sum of the probabilities of the corners (0, 0) and (0, 1) and the probability that the point is an interior point of the square and lies in the shaded region in Fig. S.3.23. The probability that the point will be an interior point of the square rather than one

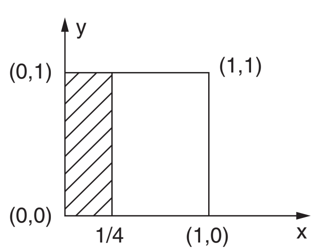

Figure S.3.23: Figure for Exercise 7a of Sec. 3.4.

of the four corners is $1-(0.1+0.2+0.4+0.1)=0.2$. The probability that it will lie in the shaded region, given that it is an interior point is 1/4. Therefore,

$$\Pr\left(X \le \frac{1}{4}\right) = 0.1+0.4+(0.2)\left(\frac{1}{4}\right) = 0.55.$$

(b) The region over which to integrate is shaded in Fig. S.3.24.

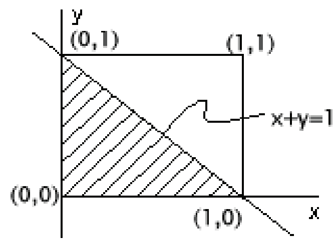

Figure S.3.24: Figure for Exercise 7b of Sec. 3.4.

$$\Pr(X+Y\le 1) = 0.1+0.2+0.4+(0.2)\left(\frac{1}{2}\right) = 0.8.$$

8. (a) Since the joint c.d.f. is continuous and is twice-differentiable in the given rectangle, the joint distribution of X and Y is continuous. Therefore,

$$\begin{aligned}\Pr(1\le X\le 2 \text{ and } 1\le Y\le 2) &= \Pr(1< X\le 2 \text{ and } 1< Y\le 2) = \\ F(2,2)-F(1,2)-F(2,1)+F(1,1) &= \frac{24}{156}-\frac{6}{156}-\frac{10}{156}+\frac{2}{156}=\frac{5}{78}\end{aligned}$$

(b)

$$\begin{aligned}\Pr(2\le X\le 4 \text{ and } 2\le Y\le 4) &= \Pr(2\le X\le 3 \text{ and } 2\le Y\le 4) \\ &= F(3,4)-F(2,4)-F(3,2)+F(2,2) \\ &= 1-\frac{64}{156}-\frac{66}{156}+\frac{24}{156}=\frac{25}{78}.\end{aligned}$$

(c) Since y must lie in the interval $0\le y\le 4$, $F_2(y)=0$ for $y<0$ and $F_2(y)=1$ for $y>4$. For $0\le y\le 4$,

$$F_2(y) = \lim_{x\to\infty} F(x,y) = \lim_{x\to 3}\frac{1}{156}xy(x^2+y) = \frac{1}{52}y\,(9+y).$$

(d) We have $f(x,y)=0$ unless $0\le x\le 3$ and $0\le y\le 4$. In this rectangle we have

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\partial y} = \frac{1}{156}(3x^2+2y).$$

(e) The region over which to integrate is shaded in Fig. S.3.25.

$$Pr(Y\le X) = \int\!\!\int_{\substack{\text{shaded}\\ \text{region}}} f(x,y)\,dx\,dy = \int_0^3\int_0^x \frac{1}{156}(3x^2+2y)\,dy\,dx = \frac{93}{208}.$$

9. The joint p.d.f. of water demand X and electricy demand Y is in (3.4.2), and is repeated here:

$$f(x,y) = \begin{cases} 1/29204 & \text{if } 4\le x\le 200 \text{ and } 1\le y\le 150,\\ 0 & \text{otherwise.}\end{cases}$$

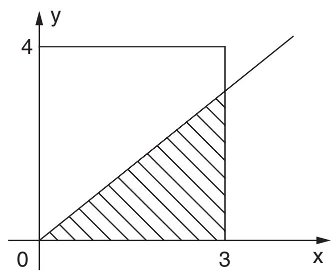

Figure S.3.25: Figure for Exercise 8e of Sec. 3.4.

We need to integrate this function over the set where $x > y$. That region can be written as $\{(x, y) : 4 < x < 200, 1 < y < \min\{x, 150\}\}$. The reason for the complicated upper limit on y is that we require both $y < x$ and $y < 150$.

$$\begin{aligned}
\int_4^{200}\int_1^{\min\{x,150\}} \frac{1}{29204}dydx &= \int_4^{200}\frac{\min\{x-1,149\}}{29204}dx \\
&= \int_4^{150}\frac{x-1}{29204}dx + \int_{150}^{200}\frac{149}{29204}dx \\
&= \left.\frac{(x-1)^2}{2\times 29204}\right|_{x=4}^{150} + \frac{50\times 149}{29204} \\
&= \frac{149^2-3^2}{58408} + \frac{7450}{29204} = 0.63505.
\end{aligned}$$

10. (a) The sum of $f(x, y)$ over all x for each fixed y is

$$\exp(-3y)\sum_{x=0}^{\infty}\frac{(2y)^2}{x!} = \exp(-3y)\exp(2y) = \exp(-y),$$

where the first equality follows from the power series expansion of $\exp(2y)$. The integral of the resulting sum is easily calculated to be 1.

(b) We can compute $\Pr(X = 0)$ by integrating $f(0, y)$ over all y:

$$\Pr(X=0) = \int_0^{\infty}\frac{(2y)^0}{0!}\exp(-3y)dy = \frac{1}{3}.$$

11. Let $f(x, y)$ stand for the joint p.f. in Table 3.3 in the text for $x = 0, 1$ and $y = 1, 2, 3, 4$.

(a) We are asked for the probability for the set $\{Y \in \{2, 3\}\} \cap \{X = 1\}$, which is $f(1, 2) + f(1, 3) = 0.166 + 0.107 = 0.273$.

(b) This time, we want $\Pr(X = 0) = f(0, 1) + f(0, 2) + f(0, 3) + f(0, 4) = 0.513$.

3.5 Marginal Distributions

Commentary

Students can get confused when solving problems like Exercises 7 and 8 in this section. They notice that the functional form of $f(x, y)$ factors into $g_1(x)g_2(y)$ for those (x, y) pairs such that $f(x, y) > 0$, but they don't understand that the factorization needs to hold even for those (x, y) pairs such that $f(x, y) = 0$. When the two marginal p.d.f.'s are both strictly positive on intervals, then the set of (x, y) pairs where $f_1(x)f_2(y) > 0$ must be a rectangle (with sides parallel to the coordinate axes), even if the rectangle is infinite in one or more directions. Hence, it is a *necessary* condition for independence that the set of (x, y) pairs such that $f(x, y) > 0$ be a rectangle with sides parallel to the coordinate axes. Of course, it is also necessary that $f(x, y) = g_1(x)g_2(y)$ for those (x, y) such that $f(x, y) > 0$. The two necessary conditions together are sufficient to insure independence, but neither is sufficient alone. See the solution to Exercise 8 below for an illustration of how to illustrate that point.

Solutions to Exercises

1. The joint p.d.f. is constant over a rectangle with sides parallel to the coordinate axes. So, for each x, the integral over y will equal the constant times the length of the interval of y values, namely $d - c$. Similarly, for each y, the integral over x will equal the constant times the length of the interval of x values, namely $b - a$. Of course the constant k must equal one over the area of the rectangle. So $k = 1/[(b - a)(d - c)]$. So the marginal p.d.f.'s of X and Y are

$$f_1(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{otherwise,} \end{cases}$$

$$f_2(y) = \begin{cases} \dfrac{1}{d-c} & \text{for } c \le y \le d, \\ 0 & \text{otherwise.} \end{cases}$$

2. (a) For $x = 0, 1, 2$, we have

$$f_1(x) = \sum_{y=0}^{3} f(x, y) = \frac{1}{30}(4x + 6) = \frac{1}{15}(2x + 3).$$

Similarly, for $y = 0, 1, 2, 3$, we have

$$f_2(y) = \sum_{x=0}^{2} f(x, y) = \frac{1}{30}(3 + 3y) = \frac{1}{10}(1 + y).$$

(b) X and Y are not independent because it is not true that $f(x, y) = f_1(x)f_2(y)$ for all possible values of x and y.

3. (a) For $0 \le x \le 2$, we have

$$f_1(x) = \int_0^1 f(x, y)\, dy = \frac{1}{2}.$$

Also, $f_1(x) = 0$ for x outside the interval $0 \le x \le 2$. Similarly, for $0 \le y \le 1$,

$$f_2(y) = \int_0^2 f(x, y)\, dx = 3y^2.$$

Also, $f_2(y) = 0$ for y outside the interval $0 \le y \le 1$.

(b) X and Y are independent because $f(x, y) = f_1(x)\ f_2(y)$ for $-\infty < x < \infty$ and $-\infty < y < \infty$.

(c) We have

$$\Pr\left(X < 1 \text{ and } Y \ge \frac{1}{2}\right) = \int_0^1 \int_{1/2}^1 f(x, y)\, dx\, dy$$

$$= \int_0^1 \int_{1/2}^1 f_1(x) f_2(y)\, dx\, dy$$
$$= \int_0^1 f_1(x)\, dx \int_{1/2}^1 f_2(y)\, dy = \Pr(X < 1) \Pr\left(Y > \frac{1}{2}\right).$$

Therefore, by the definition of the independence of two events (Definition 2.2.1), the two given events are independent.

We can also reach this answer, without carrying out the above calculation, by reasoning as follows: Since the random variables X and Y are independent, and since the occurrence or nonoccurence of the event $\{X < 1\}$ depends on the value of X only while the occurrence or nonoccurence of the event $\{Y \geq 1/2\}$ depends on the value of Y only, it follows that these two events must be independent.

4. (a) The region where $f(x, y)$ is non-zero is the shaded region in Fig. S.3.26. It can be seen that the

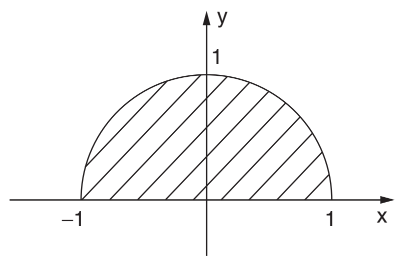

Figure S.3.26: Figure for Exercise 4a of Sec. 3.5.

possible values of X are confined to the interval $-1 \leq X \leq 1$. Hence, $f_1(x) = 0$ for values of x outside this interval. For $-1 \leq x \leq 1$, we have

$$f_1(x) = \int_0^{1-x^2} f(x, y)\, dy = \frac{15}{4} x^2(1 - x^2).$$

Similarly, it can be seen from the sketch that the possible values of Y are confined to the interval $0 \leq Y \leq 1$. Hence, $f_2(y) = 0$ for values of y outside this interval. For $0 \leq y \leq 1$, we have

$$f_2(y) = \int_{-(1-y)^{1/2}}^{(1-y)^{1/2}} f(x, y)\, dx = \frac{5}{2}(1 - y)^{3/2}.$$

(b) X and Y are not independent because $f(x, y) \neq f_1(x) f_2(y)$.

5. (a) Since X and Y are independent,

$$f(x, y) = \Pr(X = x \text{ and } Y = y) = \Pr(X = x) \Pr(Y = y) = p_x p_y.$$

(b) $\Pr(X = Y) = \sum_{i=0}^{3} f(i, i) = \sum_{i=0}^{3} p_i^2 = 0.3.$

(c) $\Pr(X > Y) = f(1, 0) + f(2, 0) + f(3, 0) + f(2, 1) + f(3, 1) + f(3, 2) = 0.35.$

6. (a) Since X and Y are independent

$$f(x, y) = f_1(x)\, f_2(y) = g(x) g(y) = \begin{cases} \frac{9}{64} x^2 y^2 & \text{for } 0 \leq x \leq 2,\ 0 \leq y \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

(b) Since X and Y have a continuous joint distribution, $\Pr(X = Y) = 0$.

(c) Since X and Y are independent random variables with the same probability distribution, it must be true that $\Pr(X > Y) = \Pr(Y > X)$. Since $\Pr(X = Y) = 0$, it therefore follows that $\Pr(X > Y) = 1/2$.

(d) $\Pr(X + Y \leq 1) = \Pr(\text{shaded region in sketch})$

$$= \int_0^1 \int_0^{1-y} f(x, y)\, dx\, dy = \frac{1}{1280}.$$

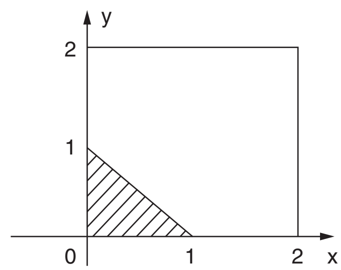

Figure S.3.27: Figure for Exercise 6d of Sec. 3.5.

7. Since $f(x, y) = 0$ outside a rectangle and $f(x, y)$ can be factored as in Eq. (3.5.7) inside the rectangle (use $h_1(x) = 2x$ and $h_2(y) = \exp(-y)$), it follows that X and Y are independent.

8. Although $f(x, y)$ can be factored as in Eq. (3.5.7) inside the triangle where $f(x, y) > 0$, the fact that $f(x, y) > 0$ inside a triangle, rather than a rectangle, implies that X and Y cannot be independent. (Note that $y \geq 0$ should have appeared as part of the condition for $f(x, y) > 0$ in the statement of the exercise.) For example, to factor $f(x, y)$ as in Eq. (3.5.7) we write $f(x, y) = g_1(x)g_2(y)$. Since $f(1/3, 1/4) = 2$ and $f(1/6, 3/4) = 3$, it must be that $g_1(1/3) > 0$ and $g_2(3/4) > 0$. However, since $f(1/3, 3/4) = 0$, it must be that either $g_1(1/3) = 0$ or $g_2(3/4) = 0$. These facts contradict each other, hence f cannot have a factorization as in (3.5.7).

9. (a) Since $f(x, y)$ is constant over the rectangle S and the area of S is 6 units, it follows that $f(x, y) = 1/6$ inside S and $f(x, y) = 0$ outside S. Next, for $0 \leq x \leq 2$,

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_1^4 \frac{1}{6} dy = \frac{1}{2}.$$

Also, $f_1(x) = 0$ otherwise. Similarly, for $1 \leq y \leq 4$,

$$f_2(y) = \int_0^2 \frac{1}{6}\, dx = \frac{1}{3}.$$

Also, $f_2(y) = 0$ otherwise. Thus, the marginal distribution of both X and Y are uniform distributions.

(b) Since $f(x, y) = f_1(x)\, f_2(y)$ for all values of x and y, it follows that X and Y are independent.

10. (a) $f(x, y)$ is constant over the circle S in Fig. S.3.28. The area of S is π units, and it follows that $f(x, y) = 1/\pi$ inside S and $f(x, y) = 0$ outside S. Next, the possible values of x range from -1 to 1. For any value of x in this interval, $f(x, y) > 0$ only for values of y between $-(1 - x^2)^{1/2}$ and $(1 - x^2)^{1/2}$. Hence, for $-1 \leq x \leq 1$,

$$f_1(x) = \int_{-(1-x^2)^{1/2}}^{(1-x^2)^{1/2}} \frac{1}{\pi} dy = \frac{2}{\pi}(1 - x^2)^{1/2}.$$

Also, $f_1(x) = 0$ otherwise. By symmetry, the random variable Y will have the same marginal p.d.f. as X.

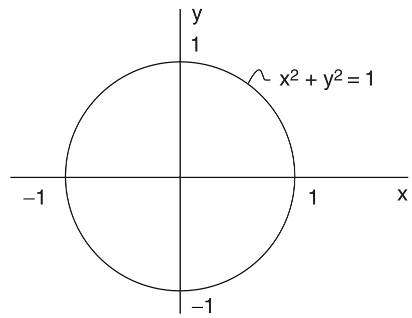

Figure S.3.28: Figure for Exercise 10 of Sec. 3.5.

(b) Since $f(x, y) \neq f_1(x)f_2(y)$, X and Y are not independent.
The conclusions found in this exercise in which X and Y have a uniform distribution over a circle should be contrasted with the conclusions found in Exercise 9, in which X and Y had a uniform distribution over a rectangle with sides parallel to the axes.

11. Let X and Y denote the arrival times of the two persons, measured in terms of the number of minutes after 5 P.M. Then X and Y each have the uniform distribution on the interval (0, 60) and they are independent. Therefore, the joint p.d.f. of X and Y is

$$f(x, y) = \begin{cases} \dfrac{1}{3600} & \text{for } 0 < x < 60,\ 0 < y < 60, \\ 0 & \text{otherwise.} \end{cases}$$

We must calculate $\Pr(|X - Y| < 10)$, which is equal to the probability that the point (X, Y) lies in the shaded region in Fig. S.3.29. Since the joint p.d.f. of X and Y is constant over the entire square, this

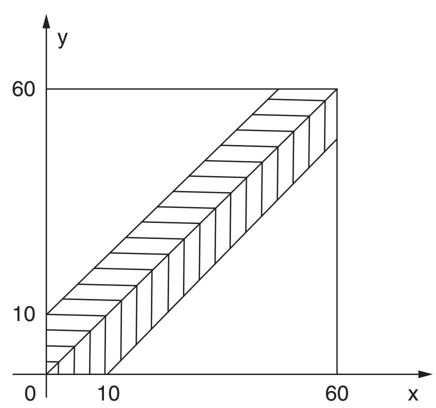

Figure S.3.29: Figure for Exercise 11 of Sec. 3.5.

probability is equal to (area of shaded region)/3600. The area of the shaded region is 1100. Therefore, the required probability is 1100/3600 = 11/36.

12. Let the rectangular region be $R = \{(x, y) : x_0 < x < x_1, y_0 < y < y_1\}$ with x_0 and/or y_0 possibly $-\infty$ and x_1 and/or y_1 possibly ∞. For the "if" direction, assume that $f(x, y) = h_1(x)h_2(y)$ for all (x, y) that satisfy $f(x, y) > 0$. Then define

$$h_1^*(x) = \begin{cases} h_1(x) & \text{if } x_0 < x < x_1, \\ 0 & \text{otherwise.} \end{cases}$$

$$h_2^*(y) = \begin{cases} h_2(y) & \text{if } y_0 < y < y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $h_1^*(x)h_2^*(y) = h_1(x)h_2(y) = f(x,y)$ for all $(x,y) \in R$ and $h_1^*(x)h_2^*(y) = 0 = f(x,y)$ for all $(x,y) \notin R$. Hence $f(x,y) = h_1^*(x)h_2^*(y)$ for all (x,y), and X and Y are independent.

For the "only if" direction, assume that X and Y are independent. According to Theorem 3.5.5, $f(x,y) = h_1(x)h_2(y)$ for all (x,y). Then $f(x,y) = h_1(x)h_2(y)$ for all $(x,y) \in R$.

13. Since $f(x,y) = f(y,x)$ for all (x,y), it follows that the marginal p.d.f.'s will be the same. Each of those marginals will equal the integral of $f(x,y)$ over the other variable. For example, to find $f_1(x)$, note that for each x, the values of y such that $f(x,y) > 0$ form the interval $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. Then, for $-1 \le x \le 1$,

$$\begin{aligned} f_1(x) &= \int f(x,y)dy \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} kx^2y^2dy \\ &= kx^2 \left. \frac{y^3}{3} \right|_{y=-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \\ &= 2kx^2(1-x^2)^{3/2}/3. \end{aligned}$$

14. The set in Fig. 3.12 is not rectangular, so X and Y are *not* independent.

15. (a) Figure S.3.30 shows the region where $f(x,y) > 0$ as the union of two shaded rectangles. Although the region is not a rectangle, it is a *product set*. That is, it has the form $\{(x,y) : x \in A, y \in B\}$ for two sets A and B of real numbers.

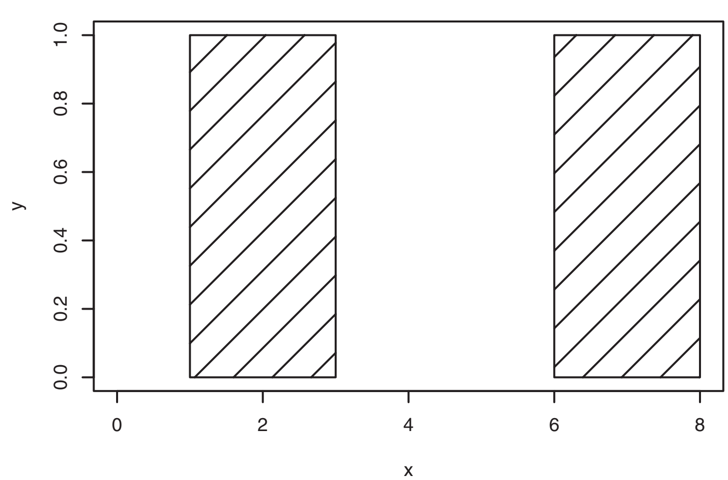

Figure S.3.30: Region of positive p.d.f. for Exercise 15a of Sec. 3.5.

(b) The marginal p.d.f. of X is

$$f_1(x) = \int_0^1 f(x,y)dy = \begin{cases} \frac{1}{3} & \text{if } 1 < x < 3, \\ \frac{1}{6} & \text{if } 6 < x < 8. \end{cases}$$

The marginal p.d.f. of Y is

$$f_2(y) = \int_1^3 \frac{1}{3}dx + \int_6^8 \frac{1}{6}dx = 1,$$

for $0 < y < 1$. The distribution of Y is the uniform distribution on the interval $[0,1]$.

(c) The product of the two marginal p.d.f.'s is

$$f_1(x)f_2(y) = \begin{cases} \frac{1}{3} & \text{if } 1 < x < 3 \text{ and } 0 < y < 1, \\ \frac{1}{6} & \text{if } 6 < x < 8 \text{ and } 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

which is the same as $f(x,y)$, hence the two random variables are independent. Although the region where $f(x,y) > 0$ is not a rectangle, it is a product set as we saw in part (a). Although it is sufficient in Theorem 3.5.6 for the region where $f(x,y) > 0$ to be a rectangle, it is necessary that the region be a product set. Technically, it is necessary that there is a version of the p.d.f. that is strictly positive on a product set. For continuous joint distributions, one can set the p.d.f. to arbitrary values on arbitrary one-dimensional curves without changing it's being a joint p.d.f.

3.6 Conditional Distributions

Commentary

When introducing conditional distributions given continuous random variables, it is important to stress that we are not conditioning on a set of 0 probability, even if the popular notation makes it appear that way. The note on page 146 can be helpful for students who understand two-variable calculus. Also, Exercise 25 in Sec. 3.11 can provide additional motivation for the idea that the conditional distribution of X given $Y = y$ is really a surrogate for the conditional distribution of X given that Y is close to y, but we don't wish to say precisely how close. Exercise 26 in Sec. 3.11 (the Borel paradox) brings home the point that conditional distributions really are not conditional on the probability 0 events such as $\{Y = y\}$.

Also, it is useful to stress that conditional distributions behave just like distributions. In particular, conditional probabilities can be calculated from conditional p.f.'s and conditional p.d.f.'s in the same way that probabilities are calculated from p.f.'s and p.d.f.'s. Also, be sure to advertise that all future concepts and theorems will have conditional versions that behave just like the marginal versions.

Solutions to Exercises

1. We begin by finding the marginal p.d.f. of Y. The set of x values for which $f(x,y) > 0$ is the interval $[-(1-y^2)^{1/2}, (1-y^2)^{1/2}]$. So, the marginal p.d.f. of Y is, for $-1 \le y \le 1$,

$$f_2(y) = \int_{-(1-y^2)^{1/2}}^{(1-y^2)^{1/2}} kx^2y^2dx = \frac{ky^2}{3}x^3\Big|_{x=-(1-y^2)^{1/2}}^{(1-y^2)^{1/2}} = \frac{2k}{3}y^2(1-y^2)^{3/2},$$

and 0 otherwise. The conditional p.d.f. of X given $Y = y$ is the ratio of the joint p.d.f. to the marginal p.d.f. just found.

$$g_1(x|y) = \begin{cases} \dfrac{3x^2}{2(1-y^2)^{3/2}} & \text{for } -(1-y^2)^{1/2} \le x \le (1-y^2)^{1/2}, \\ 0 & \text{otherwise.} \end{cases}$$

2. (a) We have Pr(Junior) $= 0.04 + 0.20 + 0.09 = 0.33$. Therefore,

$$\Pr(\text{Never}|\text{Junior}) = \frac{\Pr(\text{Junior and Never})}{\Pr(\text{Junior})} = \frac{0.04}{0.33} = \frac{4}{33}.$$

(b) The only way we can use the fact that a student visited the museum three times is to classify the student as having visited more than once. We have
Pr(More than once) $= 0.04 + 0.04 + 0.09 + 0.10 = 0.27$.
Therefore,

$$\begin{aligned}\Pr(\text{Senior}|\text{More than once}) &= \frac{\Pr(\text{Senior and More than once})}{\Pr(\text{More than once})} \\ &= \frac{0.10}{0.27} = \frac{10}{27}.\end{aligned}$$

3. The joint p.d.f. of X and Y is positive for all points inside the circle S shown in the sketch. Since the area of S is 9π and the joint p.d.f. of X and Y is constant over S, this joint p.d.f. must have the form:

$$f(x, y) = \begin{cases} \dfrac{1}{9\pi} & \text{for } (x, y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

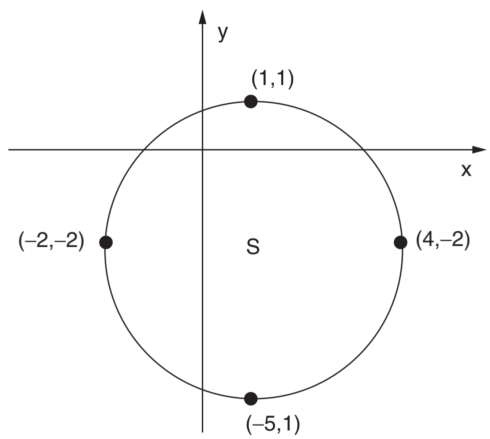

Figure S.3.31: Figure for Exercise 3 of Sec. 3.6.

It can be seen from Fig. S.3.31 that the possible values of X lie between -2 and 4. Therefore, for $-2 < x < 4$,

$$f_1(x) = \int_{-2-[9-(x-1)^2]^{1/2}}^{-2+[9-(x-1)^2]^{1/2}} \frac{1}{9\pi}\, dy = \frac{2}{9\pi}[9 - (x-1)^2]^{1/2}.$$

(a) It follows that for $-2 < x < 4$ and $-2 - [9 - (x-1)^2]^{1/2} < y < -2 + [9 - (x-1)^2]^{1/2}$,

$$g_2(y\,|\,x) = \frac{f(x, y)}{f_1(x)} = \frac{1}{2}[9 - (x-1)^2]^{-1/2}.$$

(b) When $X = 2$, it follows from part (a) that

$$g_2(y\,|\,x = 2) = \begin{cases} \dfrac{1}{2\sqrt{8}} & \text{for } -2 - \sqrt{8} < y < -2 + \sqrt{8} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\Pr(Y > 0 \mid X = 2) = \int_0^{-2+\sqrt{8}} g_2(y \mid x = 2)\, dy = \frac{-2+\sqrt{8}}{2\sqrt{8}} = \frac{2-\sqrt{2}}{4}.$$

4. (a) For $0 \le y \le 1$, the marginal p.d.f. of y is

$$f_2(y) = \int_0^1 f(x, y)\, dx = c\left(\frac{1}{2} + y^2\right).$$

Therefore, for $0 \le x \le 1$ and $0 \le y \le 1$, the conditional p.d.f. of X given that $Y = y$ is

$$g_1(x \mid y) = \frac{f(x, y)}{f_2(y)} = \frac{x + y^2}{\frac{1}{2} + y^2}.$$

It should be noted that it was not necessary to evaluate the constant c in order to determine this conditional p.d.f.

(b) When $Y = 1/2$, it follows from part (a) that

$$g_1\left(x \,\middle|\, y = \frac{1}{2}\right) = \begin{cases} \frac{4}{3}\left(x + \frac{1}{4}\right) & \text{for } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\Pr\left(X < \frac{1}{2} \,\middle|\, Y = \frac{1}{2}\right) = \int_0^{\frac{1}{2}} g_1\left(x \,\middle|\, y = \frac{1}{2}\right) dx = \frac{1}{3}.$$

5. (a) The joint p.d.f. $f(x, y)$ is given by Eq. (3.6.15) and the marginal p.d.f. $f_2(y)$ was also given in Example 3.6.10. Hence, for $0 < y < 1$ and $0 < x < y$, we have

$$g_1(x \mid y) = \frac{f(x, y)}{f_2(y)} = \frac{-1}{(1 - x)\ \log(1 - y)}.$$

(b) When $Y = 3/4$, it follows from part (a) that

$$g_1\left(x \,\middle|\, y = \frac{3}{4}\right) = \begin{cases} \frac{1}{(1 - x)\log 4} & \text{for } 0 < x < \frac{3}{4}, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\Pr\left(X > \frac{1}{2} \,\middle|\, Y = \frac{3}{4}\right) = \int_{1/2}^{3/4} g_1\left(x \,\middle|\, y = \frac{3}{4}\right) dx = \frac{\log 4 - \log 2}{\log 4} = \frac{1}{2}.$$

6. Since $f(x, y) = 0$ outside a rectangle with sides parallel to the x and y axes and since $f(x, y)$ can be factored as in Eq. (3.5.7), with $g_1(x) = c\sin(x)$ and $g_2(y) = 1$, it follows that X and Y are independent random variables. Furthermore, for $0 \le y \le 3$, the marginal p.d.f. $f_2(y)$ must be proportional to $g_2(y)$. In other words, $f_2(y)$ must be constant for $0 \le y \le 3$. Hence, Y has the uniform distribution on the interval $[0, 3]$ and

$$f_2(y) = \begin{cases} \frac{1}{3} & \text{for } 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Since X and Y are independent, the conditional p.d.f. of Y for any given value of X is the same as the marginal p.d.f. $f_2(y)$.

(b) Since X and Y are independent,

$$\Pr(1 < Y < 2 \mid X = 0.73) = \Pr(1 < Y < 2) = \int_1^2 f_2(y)\,dy = \frac{1}{3}.$$

7. The joint p.d.f. of X and Y is positive inside the triangle S shown in Fig. S.3.32. It is seen from Fig. S.3.32 that the possible values of X lie between 0 and 2. Hence, for $0 < x < 2$,

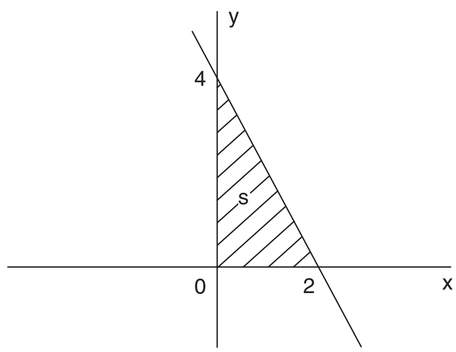

Figure S.3.32: Figure for Exercise 7 of Sec. 3.6.

$$f_1(x) = \int_0^{4-2x} f(x, y)\,dy = \frac{3}{8}(x - 2)^2.$$

(a) It follows that for $0 < x < 2$ and $0 < y < 4 - 2x$,

$$g_2(y \mid x) = \frac{f(x, y)}{f_1(x)} = \frac{4 - 2x - y}{2(x - 2)^2}.$$

(b) When $X = 1/2$, it follows from part (a) that

$$g_2\left(y \mid x = \frac{1}{2}\right) = \begin{cases} \frac{2}{9}(3 - y) & \text{for } 0 < y < 3, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\Pr\left(Y \geq 2 \mid X = \frac{1}{2}\right) = \int_2^3 g_2\left(y \mid x = \frac{1}{2}\right) dy = \frac{1}{9}.$$

8. (a) The answer is

$$\int_0^1 \int_{0.8}^1 f(x, y)\,dx\,dy = 0.264.$$

(b) For $0 \leq y \leq 1$, the marginal p.d.f. of Y is

$$f_2(y) = \int_0^1 f(x, y)\,dx = \frac{2}{5}(1 + 3y).$$

Hence, for $0 \leq x \leq 1$ and $0 \leq y \leq 1$,

$$g_1(x \mid y) = \frac{2x + 3y}{1 + 3y}.$$

When $Y = 0.3$, it follows that

$$g_1(x \mid y = 0.3) = \frac{2x + 0.9}{1.9} \quad \text{for } 0 \leq x \leq 1.$$

Hence,

$$\Pr(X > 0.8 \,|\, Y = 0.3) = \int_{0.8}^{1} g_1(x \,|\, y = 0.3) dx = 0.284.$$

(c) For $0 \le x \le 1$, the marginal p.d.f. of X is

$$f_1(x) = \int_0^1 f(x, y)\, dy = \frac{2}{5}\left(2x + \frac{3}{2}\right).$$

Hence, for $0 \le x \le 1$ and $0 \le y \le 1$,

$$g_2(y \,|\, x) = \frac{2x + 3y}{2x + \frac{3}{2}}.$$

When $X = 0.3$, it follows that

$$g_2(y \,|\, x = 0.3) = \frac{0.6 + 3y}{2.1} \quad \text{for } 0 \le y \le 1.$$

Hence,

$$\Pr(Y > 0.8 \,|\, X = 0.3) = \int_{0.8}^{1} g_2(y \,|\, x = 0.3)\, dy = 0.314.$$

9. Let Y denote the instrument that is chosen. Then $\Pr(Y = 1) = \Pr(Y = 2) = 1/2$. In this exercise the distribution of X is continuous and the distribution of Y is discrete. Hence, the joint distribution of X and Y is a mixed distribution, as described in Sec. 3.4. In this case, the joint p.f./p.d.f. of X and Y is as follows:

$$f(x, y) = \begin{cases} \frac{1}{2} \cdot 2x = x & \text{for } y = 1 \text{ and } 0 < x < 1, \\ \frac{1}{2} \cdot 3x^2 = \frac{3}{2}x^2 & \text{for } y = 2 \text{ and } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) It follows that for $0 < x < 1$,

$$f_1(x) = \sum_{y=1}^{2} f(x, y) = x + \frac{3}{2}x^2,$$

and $f_1(x) = 0$ otherwise.

(b) For $y = 1, 2$ and $0 < x < 1$, we have

$$\Pr(Y = y \,|\, X = x) = g_2(y \,|\, x) = \frac{f(x, y)}{f_1(x)}.$$

Hence,

$$\Pr\left(Y = 1 \,|\, X = \frac{1}{4}\right) = \frac{f\left(\frac{1}{4}, 1\right)}{f_1\left(\frac{1}{4}\right)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{3}{2} \cdot \frac{1}{16}} = \frac{8}{11}.$$

10. Let $Y = 1$ if a head is obtained when the coin is tossed and let $Y = 0$ if a tail is obtained. Then $\Pr(Y = 1 \,|\, X = x) = x$ and $\Pr(Y = 0 \,|\, X = x) = 1 - x$. In this exercise, the distribution of X is

continuous and the distribution of Y is discrete. Hence, the joint distribution of X and Y is a mixed distribution as described in Sec. 3.4. The conditional p.f. of Y given that $X = x$ is

$$g_2(y \,|\, x) = \begin{cases} x & \text{for } \; y = 1, \\ 1 - x & \text{for } \; y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The marginal p.d.f. $f_1(x)$ of X is given in the exercise, and the joint p.f./p.d.f. of X and Y is $f(x, y) = f_1(x)g_2(y \,|\, x)$. Thus, we have

$$f(x, y) = \begin{cases} 6x^2(1 - x) & \text{for } \; 0 < x < 1 \text{ and } \; y = 1, \\ 6x(1 - x)^2 & \text{for } \; 0 < x < 1 \text{ and } \; y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, for $y = 0, 1$,

$$\Pr(Y = y) = f_2(y) = \int_0^1 f(x, y)\, dx.$$

Hence,

$$\Pr(Y = 1) = \int_0^1 6x^2(1 - x)\, dx = \int_0^1 (6x^2 - 6x^3)\, dx = \frac{1}{2}.$$

(This result could also have been derived by noting that the p.d.f. $f_1(x)$ is symmetric about the point $x = 1/2$.)

It now follows that the conditional p.d.f. of X given that $Y = 1$ is, for $0 < x < 1$,

$$g_1(x \,|\, y = 1) = \frac{f(x, 1)}{\Pr(Y = 1)} = \frac{6x^2(1 - x)}{1/2} = 12x^2(1 - x).$$

11. Let F_2 be the c.d.f. of Y. Since f_2 is continuous at both y_0 and y_1, we can write, for $i = 0, 1$,

$$\Pr(Y \in A_i) = F_2(y_i + \epsilon) - F_2(y_i - \epsilon) = 2\epsilon f_2(y_i'),$$

where y_i' is within ϵ of y_i. This last equation follows from the mean value theorem of calculus. So

$$\frac{\Pr(Y \in A_0)}{\Pr(Y \in A_1)} = \frac{f_2(y_0')}{f_2(y_1')}. \tag{S.3.1}$$

Since f_2 is continuous, $\lim_{\epsilon \to 0} f_2(y_i') = f_2(y_i)$, and the limit of (S.3.1) is $0/f_2(y_1) = 0$.

12. (a) The joint p.f./p.d.f. of X and Y is the product $f_2(y)g_1(x|y)$.

$$f(x, y) = \begin{cases} (2y)^x \exp(-3y)/x! & \text{if } y > 0 \text{ and } x = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

The marginal p.f. of X is obtained by integrating over y.

$$f_1(x) = \int_0^\infty \frac{(2y)^x}{x!} \exp(-3y) dy = \frac{1}{3}\left(\frac{2}{3}\right)^x,$$

for $x = 0, 1, \ldots$.

(b) The conditional p.d.f. of Y given $X = 0$ is the ratio of the joint p.f./p.d.f. to $f_1(0)$.

$$g_2(y|0) = \frac{(2y)^0 \exp(-3y)/0!}{(1/3)(2/3)^0} = 3\exp(-3y),$$

for $y > 0$.

(c) The conditional p.d.f. of Y given $X = 1$ is the ratio of the joint p.f./p.d.f. to $f_1(1)$.

$$g_2(y|1) = \frac{(2y)^1 \exp(-3y)/1!}{(1/3)(2/3)^1} = 9y\exp(-3y),$$

for $y > 0$.

(d) The ratio of the two conditional p.d.f.'s is

$$\frac{g_2(y|1)}{g_2(y|0)} = \frac{9y\exp(-3y)}{3\exp(-3y)} = 3y.$$

The ratio is greater than 1 if $y > 1/3$. This corresponds to the intuition that if we observe more calls, then we should think the rate is higher.

13. There are four different treatments on which we are asked to condition. The marginal p.f. of treatment Y is given in the bottom row of Table 3.6 in the text. The conditional p.f. of response given each treatment is the ratio of the two rows above that to the bottom row:

$$g_1(x|1) = \begin{cases} \frac{0.120}{0.267} = 0.4494 & \text{if } x = 0, \\ \frac{0.147}{0.267} = 0.5506 & \text{if } x = 1. \end{cases}$$

$$g_1(x|2) = \begin{cases} \frac{0.087}{0.253} = 0.3439 & \text{if } x = 0, \\ \frac{0.166}{0.253} = 0.6561 & \text{if } x = 1. \end{cases}$$

$$g_1(x|3) = \begin{cases} \frac{0.146}{0.253} = 0.5771 & \text{if } x = 0, \\ \frac{0.107}{0.253} = 0.4229 & \text{if } x = 1. \end{cases}$$

$$g_1(x|4) = \begin{cases} \frac{0.160}{0.227} = 0.7048 & \text{if } x = 0, \\ \frac{0.067}{0.227} = 0.2952 & \text{if } x = 1. \end{cases}$$

The fourth one looks quite different from the others, especially from the second.

3.7 Multivariate Distributions

Commentary

The material around Definition 3.7.8 and Example 3.7.8 reintroduces the concept of conditionally independent random variables. This concept is important in Bayesian inference, but outside of Bayesian inference, it generally appears only in more advanced applications such as expert systems and latent variable models. If an instructor is going to forego all discussion of Bayesian inference then this material (and Exercises 13 and 14) could be skipped.

Solutions to Exercises

1. (a) We have

$$\int_0^1 \int_0^1 \int_0^1 f(x_1, d_2, x_3)\, dx_1\, dx_2\, dx_3 = 3c.$$

Since the value of this integral must be equal to 1, it follows that $c = 1/3$.

(b) For $0 \le x_1 \le 1$ and $0 \le x_3 \le 1$,

$$f_{13}(x_1, x_3) = \int_0^1 f(x_1, x_2, x_3)\, dx_2 = \frac{1}{3}(x_1 + 1 + 3x_3).$$

(c) The conditional p.d.f. of x_3 given that $x_1 = 1/4$ and $x_2 = 3/4$ is, for $0 \le x_3 \le 1$,

$$g_3\left(x_3 \,\middle|\, x_1 = \frac{1}{4}, x_2 = \frac{3}{4}\right) = \frac{f\left(\frac{1}{4}, \frac{3}{4}, x_3\right)}{f_{12}\left(\frac{1}{4}, \frac{3}{4}\right)} = \frac{7}{13} + \frac{12}{13}x_3.$$

Therefore,

$$\Pr\left(X_3 < \frac{1}{2} \,\middle|\, X_1 = \frac{1}{4}, X_2 = \frac{3}{4}\right) = \int_0^{\frac{1}{2}} \left(\frac{7}{13} + \frac{12}{13}x_3\right) dx_3 = \frac{5}{13}.$$

2. (a) First, integrate over x_1. We need to compute $\int_0^1 c x_1^{1+x_2+x_3}(1-x_1)^{3-x_2-x_3} dx_1$. The two exponents always add to 4 and each is always at least 1. So the possible pairs of exponents are $(1,3)$, $(2,2)$, and $(3,1)$. By the symmetry of the function, the first and last will give the same value of the integral. In this case, the values are

$$\int_0^1 c[x_1^3 - x_1^4]dx_1 = \frac{c}{4} - \frac{c}{5} = \frac{c}{20}. \tag{S.3.2}$$

In the other case, the integral is

$$\int_0^1 c[x_1^2 - 2x_1^3 + x_1^4]dx_1 = \frac{c}{3} - \frac{2c}{4} + \frac{c}{5} = \frac{c}{30}. \tag{S.3.3}$$

Finally, sum over the possible (x_2, x_3) pairs. The mapping between (x_2, x_3) values and the exponents in the integral is as follows:

(x_2, x_3)	Exponents
$(0,0)$	$(1,3)$
$(0,1)$	$(2,2)$
$(1,0)$	$(2,2)$
$(1,1)$	$(3,1)$

Summing over the four possible (x_2, x_3) pairs gives the sum of $c/6$, so $c = 6$.

(b) The marginal joint p.f. of (X_2, X_3) is given by setting $c = 6$ in (S.3.2) and (S.3.3) and using the above table.

$$f_{23}(x_2, x_3) = \begin{cases} 0.3 & \text{if } (x_2, x_3) \in \{(0,0), (1,1)\}, \\ 0.2 & \text{if } (x_2, x_3) \in \{(1,0), (0,1)\}. \end{cases}$$

(c) The conditional p.d.f. of X_1 given $X_2 = 1$ and $X_3 = 1$ is $1/0.3$ times the joint p.f./p.d.f. evaluated at $x_2 = x_3 = 1$:

$$g_1(x_1|1,1) = \begin{cases} 20x_1^3(1-x_1) & \text{if } 0 < x_1 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

3. The p.d.f. should be positive for all $x_i > 0$ not just for all $x_i > 1$ as stated in early printings. This will match the answers in the back of the text.

(a) We have

$$\int_0^\infty \int_0^\infty \int_0^\infty f(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 = \frac{1}{6}c.$$

Since the value of this integral must be equal to 1, it follows that c = 6. If one used $x_i > 1$ instead, then the integral would equal $\exp(-6)/6$, so that $c = 6\exp(6)$.

(b) For $x_1 > 0$, $x_3 > 0$,

$$f_{13}(x_1, x_3) \int_0^\infty f(x_1, x_2, x_3)\, dx_2 = 3\exp[-(x_1 + 3x_3)].$$

If one used $x_i > 1$ instead, then for $x_1 > 1$ and $x_3 > 1$, $f_{13}(x_1, x_3) = 3\exp(-x_1 - 3x_3 + 4)$.

(c) It is helpful at this stage to recognize that the random variables X_1, X_2, and X_3 are independent because their joint p.d.f. $f(x_1, x_2, x_3)$ can be factored as in Eq. (3.7.7); i.e., for $x_i > 0$ $(i = 1, 2, 3)$,

$$f(x_1, x_2, x_3) = (\exp(-x_1)) = (2\exp(-x_2))(3\exp(-x_3)).$$

It follows that

$$\Pr(X_1 < 1 \,|\, X_2 = 2, X_3 = 1) = Pr(X_1 < 1) = \int_0^1 f_1(x_1)\, dx_1 = \int_0^1 \exp(-x_1) dx_1 = 1 - \frac{1}{e}.$$

This answer could also be obtained without explicitly using the independence of X_1, X_2, and X_3 by calculating first the marginal joint p.d.f.

$$f_{23}(x_2, x_3) = \int_0^\infty f(x_1, x_2, x_3)\, dx_1,$$

then calculating the conditional p.d.f.

$$g_1(x_1 \,|\, x_2 = 2, x_3 = 1) = \frac{f(x_1, 2, 1)}{f_{2,3}(2, 1)},$$

and finally calculating the probability

$$\Pr(X_1 < 1 \,|\, X_2 = 2, X_3 = 1) = \int_0^1 g_1(x_1 \,|\, x_2 = 2, x_3 = 1) dx_1.$$

If one used $x_i > 1$ instead, then the probability in this part is 0.

4. The joint p.d.f. $f(x_1, x_2, x_3)$ is constant over the cube S. Since

$$\iiint_S dx_1\, dx_2\, dx_3 = \int_0^1 \int_0^1 \int_0^1 dx_1\, dx_2\, dx_3 = 1,$$

it follows that $f(x_1, x_2, x_3) = 1$ for $(x_1, x_2, x_3) \in S$. Hence, the probability of any subset of S will be equal to the volume of that subset.

(a) The set of points such that $(x_1 - 1/2)^2 + (x_2 - 1/2)^2 + (x_3 - 1/2)^2 \le 1/4$ is a sphere of radius 1/2 with center at the point $(1/2, 1/2, 1/2,)$. Hence, this sphere is entirely contained with in the cube S. Since the volume of any sphere is $4\pi r^3/3$, the volume of this sphere, and also its probability, is $4\pi(1/2)^3/3 = \pi/6$.

(b) The set of points such that $c_1^2 + x_2^2 + x_3^2 \le 1$ is a sphere of radius 1 with center at the origin (0, 0, 0). Hence, the volume of this sphere is $4\pi/3$. However, only one octant of this sphere, the octant in which all three coordinates are nonnegative, lies in S. Hence, the volume of the intersection of the sphere with the set S, and also its probability, is $\frac{1}{8} \cdot \frac{4}{3}\pi = \frac{1}{6}\pi$.

5. (a) The probability that all n independent components will function properly is the product of their individual probabilities and is therefore equal to $\prod_{i=1}^{n} p_i$.

   (b) The probability that all n independent components will not function properly is the product of their individual probabilities of not functioning properly and is therefore equal to $\prod_{i=1}^{n}(1-p_i)$. The probability that at least one component will function properly is $1 - \prod_{i=1}^{n}(1-p_i)$.

6. Since the n random variables $x_1, \dots, x_n$ are i.i.d. and each has the p.f. f, the probability that a particular variable X_i will be equal to a particular value x is $f(x)$, and the probability that all n variables will be equal to a particular value x is $[f(x)]^n$. Hence, the probability that all n variables will be equal, without any specification of their common value, is $\sum_x [f(x)]^n$.

7. The probability that a particular variable X_i will lie in the interval (a, b) is $p = \int_a^b f(x)\,dx$. Since the variables $X_1, \dots, X_n$ are independent, the probability that exactly i of these variables will lie in the interval (a, b) is $\binom{n}{i} p^i (1-p)^{n-i}$. Therefore, the required probability is

$$\sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

8. For any given value x of X, the random variables $Y_1, \dots, Y_n$ are i.i.d., each with the p.d.f. $g(y \mid x)$. Therefore, the conditional joint p.d.f. of $Y_1 \dots, Y_n$ given that $X = x$ is

$$h(y_1, \dots, y_n \mid x) = g(y_1 \mid x) \dots g(y_n \mid x) = \begin{cases} \dfrac{1}{x^n} & \text{for } 0 < y_i < x,\ i = 1, \dots, n, \\ 0 & \text{otherwise.} \end{cases}$$

The joint p.d.f. of X and $Y_1, \dots, Y_n$ is, therefore,

$$f(x)h(y_1, \dots, y_n \mid x) = \begin{cases} \dfrac{1}{n!}\exp(-x) & \text{for } \ 0 < y_i < x \quad (i = 1, \dots, n), \\ 0 & \text{otherwise.} \end{cases}$$

This joint p.d.f. is positive if and only if each $y_i > 0$ and x is greater than every y_i. In other words, x must be greater than $m = \max\{y_1, \dots, y_n\}$.

   (a) For $y_1 > 0$ $(i = 1, \dots, n)$, the marginal joint p.d.f. of $Y_1, \dots, Y_n$ is

$$g_0(y_1, \dots, y_n) = \int_{-\infty}^{\infty} f(x)h(y_1, \dots, y_n \mid x)\,dx = \int_m^{\infty} \frac{1}{n!}\exp(-x)\,dx = \frac{1}{n!}\exp(-m).$$

   (b) For $y_i > 0$ $(i = 1, \dots, n)$, the conditional p.d.f. of X given that $Y_i = y_i (i = 1, \dots, n)$ is

$$g_1(x \mid y_1, \dots, y_n) = \frac{f(x)h(y_1, \dots, y_n \mid x)}{g_0(y_1, \dots, y_n)} = \begin{cases} \exp(-(x-m)) & \text{for } x > m, \\ 0 & \text{otherwise.} \end{cases}$$

9. (a) Since $X_i = X$ for $i = 1, 2$, we know that X_i has the same distribution as X. Since X has a continuous distribution, then so does X_i for $i = 1, 2$.

   (b) We know that $\Pr(X_1 = X_2) = 1$. Let $A = \{(x_1, x_2) : x_1 = x_2\}$. Then $\Pr((X_1, X_2) \in A) = 1$. However, for every function f, $\int_A \int f(x_1, x_2) dx_1 dx_2 = 0$. So there is no possible joint p.d.f.

10. The marginal p.d.f. of Z is $2\exp(-2z)$, for $z > 0$. The coordinates of $\boldsymbol{X}$ are conditionally i.i.d. given $Z = z$ with p.d.f. $z\exp(-zx)$, for $x > 0$. This makes the joint p.d.f. of $(Z, \boldsymbol{X})$ equal to $2z^5\exp(-z[2 + x_1 + \cdots + x_5])$ for all variables positive. The marginal joint p.d.f. of $\boldsymbol{X}$ is obtained by integrating z out of this.

$$f_2(\boldsymbol{x}) = \int_0^\infty 2z^5\exp(-z[2 + x_1 + \cdots + x_5])dz = \frac{240}{(2 + x_1 + \cdots + x_5)^6}, \quad \text{for all } x_i > 0.$$

Here, we use the formula $\int_0^\infty y^k \exp(-y)dy = k!$ from Exercise 12 in Sec. 3.6. The conditional p.d.f. of Z given $\boldsymbol{X} = (x_1, \ldots, x_5)$ is then

$$g_1(z|\boldsymbol{x}) = \frac{(2 + x_1 + \cdots + x_5)^6}{120} z^5\exp(-z[2 + x_1 + \cdots + x_5]),$$

for $z > 0$.

11. Since $X_1, \ldots, X_n$ are independent, their joint p.f., p.d.f., or p.f./p.d.f. factors as

$$f(x_1, \ldots, x_n) = f_1(x_1)\cdots f_n(x_n),$$

where each f_i is a p.f. or p.d.f. If we sum or integrate over all x_j such that $j \notin \{i_1, \ldots, i_k\}$ we obtain the joint p.f., p.d.f., or p.f./p.d.f. of $X_{i_1}, \ldots, X_{i_k}$ equal to $f_{i_1}(x_{i_1})\cdots f_{i_k}(x_{i_k})$, which is factored in a way that makes it clear that $X_{i_1}, \ldots, X_{i_k}$ are independent.

12. Let $h(\boldsymbol{y}, \boldsymbol{w})$ be the marginal joint p.d.f. of $\boldsymbol{Y}$ and $\boldsymbol{W}$, and let $h_2(\boldsymbol{w})$ be the marginal p.d.f. of $\boldsymbol{w}$. Then

$$\begin{aligned}
h(\boldsymbol{y}, \boldsymbol{w}) &= \int f(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})d\boldsymbol{z}, \\
h_2(\boldsymbol{w}) &= \int\int f(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})d\boldsymbol{z}d\boldsymbol{y}, \\
g_1(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{w}) &= \frac{f(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})}{h_2(\boldsymbol{w})}, \\
g_2(\boldsymbol{y}|\boldsymbol{w}) &= \frac{h(\boldsymbol{y}, \boldsymbol{w})}{h_2(\boldsymbol{w})} = \frac{\int f(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})d\boldsymbol{z}}{h_2(\boldsymbol{w})} = \int g_1(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{w})d\boldsymbol{z}.
\end{aligned}$$

13. Let $f(x_1, x_2, x_3, z)$ be the joint p.d.f. of (X_1, X_2, X_3, Z). Let $f_{12}(x_1, x_2)$ be the marginal joint p.d.f. of (X_1, X_2). The the conditional p.d.f. of X_3 given $(X_1, X_2) = (x_1, x_2)$ is

$$\frac{\int f(x_1, x_2, x_3, z)dz}{f_{12}(x_1, x_2)} = \frac{\int g(x_1|z)g(x_2|z)g(x_3|z)f_0(z)dz}{f_{12}(x_1, x_2)} = \int g(x_3|z)\frac{g(x_1|z)g(x_2|z)f_0(z)}{f_{12}(x_1, x_2)}dz.$$

According to Bayes' theorem for random variables, the fraction in this last integral is $g_0(z|x_1, x_2)$. Using the specific formulas in the text, we can calculate the last integral as

$$\begin{aligned}
&\int_0^\infty z\exp(-zx_3)\frac{1}{2}(2 + x_1 + x_2)^3 z^2\exp(-z(2 + x_1 + x_2))dx \\
&= \frac{(2 + x_1 + x_2)^3}{2}\int_0^\infty z^3\exp(-z(2 + x_1 + x_2 + x_3))dz \\
&= \frac{(2 + x_1 + x_2)^3}{2}\frac{6}{(2 + x_1 + x_2 + x_3)^4} = \frac{3(2 + x_1 + x_2)^3}{(2 + x_1 + x_2 + x_3)^4}.
\end{aligned}$$

The joint p.d.f. of (X_1, X_2, X_3) can be computed in a manner similar to the joint p.d.f. of (X_1, X_2) and it is

$$f_{123}(x_1, x_2, x_3) = \frac{12}{(2 + x_1 + x_2 + x_3)^4}.$$

The ratio of $f_{123}(x_1, x_2, x_3)$ to $f_{12}(x_1, x_2)$ is the conditional p.d.f. calculated above.

14. (a) We can substitute $x_1 = 5$ and $x_2 = 7$ in the conditional p.d.f. computed in Exercise 13.

$$g_3(x_3|5, 7) = \frac{3(2 + 5 + 7)^3}{(2 + 5 + 7 + x_3)^4} = \frac{8232}{(14 + x_3)^4},$$

for $x_3 > 0$.

(b) The conditional probability we want is the integral of the p.d.f. above from 3 to ∞.

$$\int_3^\infty \frac{8232}{(14 + x_3)^4} dx_3 = -\frac{2744}{(14 + x_3)^3}\bigg|_{x_3=3}^\infty = 0.5585.$$

In Example 3.7.9, we computed the marginal probability $\Pr(X_3 > 3) = 0.4$. Now that we have observed two service times that are both longer than 3, namely 5 and 7, we think that the probability of $X_3 > 3$ should be larger.

15. Let A be an arbitrary n-dimensional set. Because $\Pr(W = c) = 1$, we have

$$\Pr((X_1, \ldots, X_n) \in A, W = w) = \begin{cases} \Pr(X_1, \ldots, X_n) \in A) & \text{if } w = c, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\Pr((X_1, \ldots, X_n) \in A|W = w) = \begin{cases} \Pr(X_1, \ldots, X_n) \in A) & \text{if } w = c, \\ 0 & \text{otherwise.} \end{cases}$$

Hence the conditional joint distribution of $X_1, \ldots, X_n$ given W is the same as the unconditional joint distribution of $X_1, \ldots, X_n$, which is the distribution of independent random variables.

3.8 Functions of a Random Variable

Commentary

A brief discussion of simulation appears at the end of this section. This can be considered a teaser for the more detailed treatment in Chapter 12. Simulation is becoming a very important tool in statistics and applied probability. Even those instructors who prefer not to cover Chapter 12 have the option of introducing the topic here for the benefit of students who will need to study simulation in more detail in another course.

If you wish to use the statistical software R, then the function `runif` will be most useful. For the purposes of this section, `runif(n)` will return `n` pseudo-uniform random numbers on the interval $[0, 1]$. Of course, either `n` must be assigned a value before expecting R to understand `runif(n)`, or one must put an explicit value of `n` into the function. The following two options both produce 1,000 pseudo-uniform random numbers and store them in an object called `unumbs`:

- `unumbs=runif(1000)`
- ```
  n=1000
  unumbs=runif(n)
  ```
  ```

Solutions to Exercises

1. The inverse transformation is $x = (1-y)^{1/2}$, whose derivative is $-(1-y)^{-1/2}/2$. The p.d.f. of Y is then

$$g(y) = f([1-y]^{1/2})(1-y)^{-1/2}/2 = \frac{3}{2}(1-y)^{1/2},$$

for $0 < y < 1$.

2. For each possible value of x, we have the following value of $y = x^2 - x$:

x	y
-3	12
-2	6
-1	2
0	0
1	0
2	2
3	6

Since the probability of each value of X is 1/7, it follows that the p.f. $g(y)$ is as follows:

y	$g(y)$
0	$\frac{2}{7}$
2	$\frac{2}{7}$
6	$\frac{2}{7}$
12	$\frac{1}{7}$

3. It is seen from Fig. S.3.33 that as x varies over the interval $0 < x < 2, y$ varies over the interval

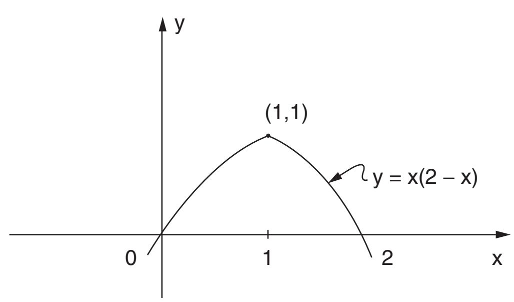

Figure S.3.33: Figure for Exercise 3 of Sec. 3.8.

$0 < y \leq 1$. Therefore, for $0 < y \leq 1$,

$$\begin{aligned} G(y) &= \Pr(Y \leq y) = \Pr\left[X(2-X) \leq y\right] = \Pr(X^2 - 2X \geq -y) \\ &= \Pr(X^2 - 2X + 1 \geq 1 - y) = \Pr[(X-1)^2 \geq 1 - y] \\ &= \Pr(X - 1 \leq -\sqrt{1-y}) + \Pr(X - 1 \geq \sqrt{1-y}) \\ &= \Pr(X \leq 1 - \sqrt{1-y}) + \Pr(X \geq 1 + \sqrt{1-y}) \\ &= \int_0^{1-\sqrt{1-y}} \frac{1}{2}x\,dx + \int_{1+\sqrt{1-y}}^{2} \frac{1}{2}x\,dx \end{aligned}$$

$$= \; 1 - \sqrt{1-y}.$$

It follows that, for $0 < y < 1$,

$$g(y) = \frac{dG(y)}{dy} = \frac{1}{2(1-y)^{1/2}}.$$

4. The function $y = 4 - x^3$ is strictly decreasing for $0 < x < 2$. When $x = 0$, we have $y = 4$, and when $x = 2$ we have $y = -4$. Therefore, as x varies over the interval $0 < x < 2$, y varies over the interval $-4 < y < 4$. The inverse function is $x = (4 - y)^{1/3}$ and

$$\frac{dx}{dy} = -\frac{1}{3}(4-y)^{-2/3}.$$

Therefore, for $-4 < y < 4$,

$$g(y) = f[(4-y)^{1/3}].\left|\frac{dx}{dy}\right| = \frac{1}{2}(4-y)^{1/3}.\frac{1}{3}(4-y)^{-2/3} = \frac{1}{6(4-y)^{1/3}}.$$

5. If $y = ax + b$, the inverse function is $x = (y - b)/a$ and $dx/dy = 1/a$. Therefore,

$$g(y) = f\left[\frac{1}{a}(y-b)\right]\left|\frac{dx}{dy}\right| = \frac{1}{|\,a\,|} f\left(\frac{y-b}{a}\right).$$

6. X lies between 0 and 2 if and only if Y lies between 2 and 8. Therefore, it follows from Exercise 3 that for $2 < y < 8$,

$$g(y) = \frac{1}{3} f\left(\frac{y-2}{3}\right) = \frac{1}{3}\cdot\frac{1}{2}\cdot\frac{y-2}{3} = \frac{1}{18}(y-2).$$

7. (a) If $y = x^2$, then as x varies over the interval $(0, 1)$, y also varies over the interval $(0, 1)$. Also, $x = y^{1/2}$ and $dx/dy = y^{-1/2}/2$. Hence, for $0 < y < 1$,

$$g(y) = f(y^{1/2})\left|\frac{dx}{dy}\right| = 1\cdot\frac{1}{2}y^{-1/2} = \frac{1}{2}y^{-1/2}.$$

(b) If $y = -x^3$, then as x varies over the interval $(0, 1)$, y varies over the interval $(-1, 0)$. Also, $x = -y^{1/3}$ and $dx/dy = -y^{-2/3}/3$. Hence, for $-1 < y < 0$,

$$g(y) = f(-y^{1/3})\left|\frac{dx}{dy}\right| = \frac{1}{3}\,|\,y\,|^{-2/3}.$$

(c) If $y = x^{1/2}$, then as x varies over the interval $(0, 1)$, y also varies over the interval $(0, 1)$. Also, $x = y^2$ and $dx/dy = 2y$. Hence, for $0 < y < 1$, $g(y) = f(y^2)2y = 2y$.

8. As x varies over all positive values, y also varies over all positive values. Also, $x = y^2$ and $dx/dy = 2y$. Therefore, for $y > 0$,

$$g(y) = f(y^2)(2y) = 2y\exp(-y^2).$$

9. The c.d.f. $G(y)$ corresponding to the p.d.f. $g(y)$ is, for $0 < y < 2$,

$$G(y) = \int_0^y g(t)dt = \int_0^y \frac{3}{8}t^2 dt = \frac{1}{8}y^3.$$

We know that the c.d.f. of the random variable $Y = G^{-1}(X)$ will be G. We must therefore determine the inverse function G^{-1}. If $X = G(Y) = Y^3/8$ then $Y = G^{-1}(X) = 2X^{1/3}$. It follows that $Y = 2X^{1/3}$.

10. For $0 < x < 2$, the c.d.f. of X is

$$F(x) = \int_0^x f(t)dt = \int_0^x \frac{1}{2}t\,dt = \frac{1}{4}x^2.$$

Therefore, by the probability integral transformation, we know that $U = F(X) = X^2/4$ will have the uniform distribution on the interval $[0, 1]$. Since U has this uniform distribution, we know from Exercise 8 that $Y = 2U^{1/3}$ will have the required p.d.f. g. Therefore, the required transformation is $Y = 2U^{1/3} = 2(X^2/4)^{1/3} = (2X^2)^{1/3}$.

11. We can use the probability integral transformation if we can find the inverse of the c.d.f. The c.d.f. is, for $0 < y < 1$,

$$G(y) = \int_{-\infty}^y g(t)dt = \frac{1}{2}\int_0^y (2t+1)dt = \frac{1}{2}(y^2+y).$$

The inverse of this function can be found by setting $G(y) = p$ and solving for y.

$$\frac{1}{2}(y^2+y) = p; \quad y^2 + y - 2p = 0; \quad y = \frac{-1 + (1+8p)^{1/2}}{2}.$$

So, we should generate four independent uniform pseudo-random variables P_1, P_2, P_3, P_4 and let $Y_i = [-1 + (1+8P_i)^{1/2}]/2$ for $i = 1, 2, 3, 4$.

12. Let X have the uniform distribution on $[0, 1]$, and let F be a c.d.f. Let $F^{-1}(p)$ be defined as the smallest x such that $F(x) \ge p$. Define $Y = F^{-1}(X)$. We need to show that $\Pr(Y \le y) = F(y)$ for all y, First, suppose that y is the unique x such that $F(x) = F(y)$. Then $Y \le y$ if and only if $X \le F(y)$. Since X has a uniform distribution $\Pr(X \le F(y)) = F(y)$. Next, suppose that $F(x) = F(y)$ for all x in the interval $[a, b)$ of $[a, b]$ with $b > a$, and suppose that $F(x) < F(y)$ for all $x < a$. Then $F^{-1}(X) \le y$ if and only if $X \le F(a) = F(y)$. Once again $\Pr(X \le F(y)) = F(y)$.

13. The inverse transformation is $z = 1/t$ with derivative $-1/t^2$. the p.d.f. of T is

$$g(t) = f(1/t)/t^2 = 2\exp(-2/t)/t^2,$$

for $t > 0$.

14. Let $Y = cX + d$. The inverse transformation is $x = (y - d)/c$. Assumethat $c > 0$. The derivative of the inverse is $1/c$. The p.d.f. of Y is

$$g(y) = f([y-d]/c)/c = [c(b-a)]^{-1}, \text{ for } a \le (y-d)/c \le b.$$

It is easy to see that $a \le (y-d)/c \le b$ if and only if $ca + d \le y \le cb + d$, so g is the p.d.f. of the uniform distribution on the interval $[ca + d, cb + d]$. If $c < 0$, the distribution of Y would be uniform on the interval $[cb + d, ca + d]$. If $c = 0$, the distribution of Y is degenerate at the value d, i.e., $\Pr(Y = d) = 1$.

15. Let F be the c.d.f. of X. First, find the c.d.f. of Y, namely, for $y > 0$,

$$\Pr(Y \le y) = \Pr(X^2 \le y) = \Pr(-y^{1/2} \le X \le y^{1/2}) = F(y^{1/2} - F(-y^{1/2}).$$

Now, the p.d.f. of Y is the derivative of the above expression, namely,

$$g(y) = \frac{d}{dy}[F(y^{1/2}) - F(-y^{1/2})] = \frac{f(y^{1/2})}{2y^{1/2}} + \frac{f(-y^{1/2})}{2y^{1/2}}.$$

This equals the expression in the exercise.

16. Because $0 < X < 1$ with probability 1, squaring X produces smaller values. There are wide intervals of values of X that produce small values of X^2 but the values of X that produce large values of X^2 are more limited. For example, to get $Y \in [0.9, 1]$, you need $X \in [0.9487, 1]$, whereas to get $Y \in [0, 0.1]$ (an interval of the same length), you need $X \in [0, 0.3162]$, a much bigger set.

17. (a) According to the problem description, $Y = 0$ if $X \le 100$, $Y = X - 100$ if $100 < X \le 5100$, and $Y = 5000$ if $X > 5100$. So, $Y = r(X)$, where

$$r(x) = \begin{cases} 0 & \text{if } x \le 100, \\ x - 100 & \text{if } 100 < x \le 5100, \\ 5000 & \text{if } x > 5100. \end{cases}$$

(b) Let G be the c.d.f. of Y. Then $G(y) = 0$ for $y < 0$, and $G(y) = 1$ for $y \ge 5000$. For $0 \le y < 5000$,

$$\begin{aligned} \Pr(Y \le y) &= \Pr(r(X) \le y) \\ &= \Pr(X \le y + 100) \\ &= \int_0^{y+100} \frac{dx}{(1+x)^2} \\ &= 1 - \frac{1}{y+101}. \end{aligned}$$

In summary,

$$G(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 - \frac{1}{y+101} & \text{if } 0 \le y < 5000, \\ 1 & \text{if } y \ge 5000. \end{cases}$$

(c) There is positive probability that $Y = 5000$, but the rest of the distribution of Y is spread out in a continuous manner between 0 and 5000.

3.9 Functions of Two or More Random Variables

Commentary

The material in this section can be very difficult, even for students who have studied calculus. Many textbooks at this level avoid the topic of general bivariate and multivariate transformations altogether. If an instructor wishes to avoid discussion of Jacobians and multivariate transformations, it might still be useful to introduce convolution, and the extremes of a random sample. The text is organized so that these topics appear early in the section, before any discussion of Jacobians. In the remainder of the text, the method of Jacobians is used in the following places:

- The proof of Theorem 5.8.1, the derivation of the beta distribution p.d.f.
- The proof of Theorem 5.10.1, the derivation of the joint p.d.f. of the bivariate normal distribution.
- The proof of Theorem 8.3.1, the derivation of the joint distribution of the sample mean and sample variance from a random sample of normal random variables.
- The proof of Theorem 8.4.1, the derivation of the p.d.f. of the t distribution.

Solutions to Exercises

1. The joint p.d.f. of X_1 and X_2 is

$$f(x_1, x_2) = \begin{cases} 1 & \text{for } 0 < x_1 < 1,\ 0 < x_2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

By Eq. (3.9.5), the p.d.f. of Y is

$$g(y) = \int_{-\infty}^{\infty} f(y - z, z)dz.$$

The integrand is positive only for $0 < y - z < 1$ and $0 < z < 1$. Therefore, for $0 < y \le 1$ it is positive only for $0 < z < y$ and we have

$$g(y) = \int_0^y 1 \cdot dz = y.$$

For $1 < y < 2$. the integrand is positive only for $y - 1 < z < 1$ and we have

$$g(y) = \int_{y-1}^1 1 \cdot dz = 2 - y.$$

2. Let f be the p.d.f. of $Y = X_1 + X_2$ found in Exercise 1, and let $Z = Y/2$. The inverse of this transformation is $y = 2z$ with derivative 2. The p.d.f. of Z is

$$g(z) = 2f(2z) = \begin{cases} 4z & \text{for } 0 < z < 1/2, \\ 4(1-z) & \text{for } 1/2 < z < 1, \\ 0 & \text{otherwise.} \end{cases}$$

3. The inverse transformation is:

$$\begin{aligned} x_1 &= y_1, \\ x_2 &= y_2/y_1, \\ x_3 &= y_3/y_2. \end{aligned}$$

Furthermore, the set S where $0 < x_i < 1$ for $i = 1, 2, 3$ corresponds to the set T where $0 < y_3 < y_2 < y_1 < 1$. We also have

$$J = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 0 \\ -\frac{y_2}{y_1^2} & \frac{1}{y_1} & 0 \\ 0 & -\frac{y_3}{y_2^2} & \frac{1}{y_2} \end{bmatrix} = \frac{1}{y_1 y_2}.$$

Therefore, for $0 < y_3 < y_2 < y_1 < 1$, the joint p.d.f. of $Y_1, Y_2,$ *and* Y_3 is

$$\begin{aligned} g(y_1, y_2, y_3) &= f\left(y_1, \frac{y_2}{y_1}, \frac{y_3}{y_2}\right) |J| \\ &= 8y_1 \frac{y_2}{y_1} \frac{y_3}{y_2} \frac{1}{y_1 y_2} = \frac{8y_3}{y_1 y_2}. \end{aligned}$$

4. As a convenient device, let $Z = X_1$. Then the transformation from X_1 *and* X_2 to Y and Z is a one-to-one transformation between the set S where $0 < x_1 < 1$ *and* $0 < x_2 < 1$ and the set T where $0 < y < z < 1$. The inverse transformation is

$$\begin{aligned} x_1 &= z, \\ x_2 &= \frac{y}{z}. \end{aligned}$$

Therefore,

$$J = \det \begin{bmatrix} \dfrac{\partial x_1}{\partial y} & \dfrac{\partial x_1}{\partial z} \\ \dfrac{\partial x_2}{\partial y} & \dfrac{\partial x_2}{\partial z} \end{bmatrix} = \det \begin{bmatrix} 0 & 1 \\ \dfrac{1}{z} & -\dfrac{y}{z^2} \end{bmatrix} = -\frac{1}{z}.$$

For $0 < y < z < 1$, the joint p.d.f. of Y and Z is

$$g(y, z) = f\left(z, \frac{y}{z}\right) |\, J \,| = \left(z + \frac{y}{z}\right)\left(\frac{1}{z}\right).$$

It follows that for $0 < y < 1$, the marginal p.d.f. of Y is

$$g_1(y) = \int_y^1 g(y, z)dz = 2(1 - y).$$

5. As a convenient device let $Y = X_2$. Then the transformation from X_1 *and* X_2 to Y and Z is a one-to-one transformation between the set S where $0 < x_1 < 1$ *and* $0 < x_2 < 1$ and the set T where $0 < y < 1$ *and* $0 < yz < 1$. The inverse transformation is

$$\begin{aligned} x_1 &= yz, \\ x_2 &= y. \end{aligned}$$

Therefore,

$$J = \det \begin{bmatrix} z & y \\ 1 & 0 \end{bmatrix} = -y.$$

The region where the p.d.f. of (Z, Y) is positive is in Fig. S.3.34. For $0 < y < 1$ *and* $0 < yz < 1$, the

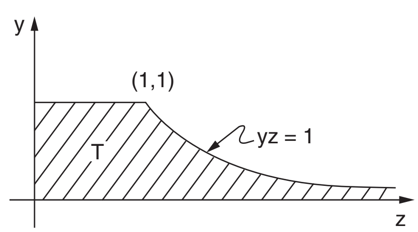

Figure S.3.34: The region where the p.d.f. of (Z, Y) is positive in Exercise 5 of Sec. 3.9.

joint p.d.f. of Y and Z is

$$g(y, z) = f(yz, y)\,|\, J \,| = (yz + y)(y).$$

It follows that for $0 < z \le 1$, the marginal p.d.f. of Z is

$$g_2(z) = \int_0^1 g(y,z)\,dy = \frac{1}{3}(z+1).$$

Also, for $z > 1$,

$$g_2(z) = \int_0^{\frac{1}{z}} g(y,z)\,dy = \frac{1}{3z^3}(z+1).$$

6. By Eq. (3.9.5) (with a change in notation),

$$g(z) = \int_{-\infty}^{\infty} f(z-t,t)\,dt \quad \text{for } -\infty < z < \infty.$$

However, the integrand is positive only for $0 \le z - t \le t \le 1$. Therefore, for $0 \le z \le 1$, it is positive only for $z/2 \le t \le z$ and we have

$$g(z) = \int_{z/2}^{z} 2z\,dt = z^2.$$

For $1 < z < 2$, the integrand is positive only for $z/2 \le t \le 1$ and we have

$$g(z) = \int_{z/2}^{z} 2z\,dt = z(2-z).$$

7. Let $Z = -X_2$. Then the p.d.f. of Z is

$$f_2(z) = \begin{cases} \exp(z) & \text{for } z < 0, \\ 0 & \text{for } z \ge 0. \end{cases}$$

Since X_1 and Z are independent, the joint p.d.f. of X_1 and Z is

$$f(x_1,z) = \begin{cases} \exp(-(x-z)) & \text{for } x > 0,\ z < 0, \\ 0 & \text{otherwise.} \end{cases}$$

It now follows from Eq. (3.9.5) that the p.d.f. of $Y = X_1 - X_2 = X_1 + Z$ is

$$g(y) = \int_{-\infty}^{\infty} f(y-z,z)\,dz.$$

The integrand is positive only for $y - z > 0$ and $z < 0$. Therefore, for $y \le 0$,

$$g(y) = \int_{-\infty}^{y} \exp(-(y-2z))\,dz = \frac{1}{2}\exp(y).$$

Also, for $y > 0$,

$$g(y) = \int_{-\infty}^{0} \exp(-(y-2z))\,dz = \frac{1}{2}\exp(-y).$$

8. We have

$$\begin{aligned}\Pr(Y_n \geq 0.99) &= 1 - \Pr(Y_n < 0.99)\\ &= 1 - \Pr(\text{All } n \text{ observations} < 0.99)\\ &= 1 - (0.99)^n.\end{aligned}$$

Next, $1 - (0.99)^n \geq 0.95$ if and only if

$$(0.99)^n \leq 0.05 \quad \text{or} \quad n \quad \log(0.99) \leq \log(0.05)$$
$$\text{or} \quad n \geq \frac{\log(0.05)}{\log(0.99)} \approx 298.1.$$

So, $n \geq 299$ is needed.

9. It was shown in this section that the joint c.d.f. of Y_1 and Y_n is, for $-\infty < y_1 < y_n < \infty$,

$$G(y_1, y_n) = [F(y_n)]^n - [F(y_n) - F(y_1)]^n.$$

Since $F(y) = y$ for the given uniform distribution, we have

$$\Pr(Y_1 \leq 0.1, Y_n \leq 0.8) = G(0.1, 0.8) = (0.8)^n - (0.7)^n.$$

10. $\Pr(Y_1 \leq 0.1 \text{ and } Y_n \geq 0.8)$

$$= \Pr(Y_1 \leq 0.1) - \Pr(Y_1 \leq 0.1 \text{ and } Y_n \leq 0.8).$$

It was shown in this section that the p.d.f. of Y_1 is

$$G_1(y) = 1 - [1 - F(y)]^n.$$

Therefore, $\Pr(Y_1 \leq 0.1) = G_1(0.1) = 1 - (0.9)^n$. Also, by Exercise 9,

$$\Pr(Y_1 \leq 0.1 \text{ and } Y_n \leq 0.8) = (0.8)^n - (0.7)^n.$$

Therefore,

$$\Pr(Y_1 \leq 0.1 \text{ and } Y_n \geq 0.8) = 1 - (0.9)^n - (0.8)^n + (0.7)^n.$$

11. The required probability is equal to

$$\Pr\left(\text{All } n \text{ observations } < \frac{1}{3}\right) + \Pr\left(\text{All } n \text{ observations } > \frac{1}{3}\right) = \left(\frac{1}{3}\right)^n + \left(\frac{2}{3}\right)^n.$$

This exercise could also be solved by using techniques similar to those used in Exercise 10.

12. The p.d.f. $h_1(w)$ of W was derived in Example 3.9.8. Therefore,

$$\begin{aligned}\Pr(W > 0.9) &= \int_{0.9}^{1} h_1(w)dw = \int_{0.9}^{1} n(n-1)w^{n-2}(1-w)dw\\ &= 1 - n(0.9)^{n-1} + (n-1)(0.9)^n.\end{aligned}$$

13. If X has the uniform distribution on the interval $[0, 1]$, then $aX + b$ $(a > 0)$ has the uniform distribution on the interval $[b, a + b]$. Therefore, $8X - 3$ has the uniform distribution on the interval $[-3, 5]$. It follows that if $X_1, \ldots, X_n$ form a random sample from the uniform distribution on the interval $[0, 1]$, then the n random variables $8X_1 - 3, \ldots, 8X_n - 3$ will have the same joint distribution as a random sample from the uniform distribution on the interval $[-3, 5]$.

Next, it follows that if the range of the sample $X_1, \ldots, X_n$ is W, then the range of the sample $8X_1 - 3, \ldots, 8X_n - 3$ will be $8W$. Therefore, if W is the range of a random sample from the uniform distribution on the interval $[0, 1]$, then $Z = 8W$ will have the same distribution as the range of a random sample from the uniform distribution on the interval $[-3, 5]$.

The p.d.f. $h(w)$ of W was given in Example 3.9.8. Therefore, the p.d.f. $f(z)$ of $Z = 8W$ is

$$g(z) = h\left(\frac{z}{8}\right) \cdot \frac{1}{8} = \frac{n(n-1)}{8} \left(\frac{z}{8}\right)^{n-2} \left(1 - \frac{z}{8}\right),$$

for $-3 < z < 5$.

This p.d.f. $g(z)$ could also have been derived from first principles as in Example 3.9.8.

14. Following the hint given in this exercise, we have

$$\begin{aligned} G(y) &= \Pr(\text{At least } n-1 \text{ observations are } \leq y) \\ &= \Pr(\text{Exactly } n-1 \text{ observations are } \leq y) + \Pr(\text{All } n \text{ observations are } \leq y) \\ &= ny^{n-1}(1-y) + y^n = ny^{n-1} - (n-1)y^n. \end{aligned}$$

Therefore, for $0 < y < 1$,

$$\begin{aligned} g(y) &= n\,(n-1)y^{n-2} - n(n-1)y^{n-1} \\ &= n(n-1)y^{n-2}(1-y). \end{aligned}$$

It is a curious result that for this uniform distribution, the p.d.f. of Y is the same as the p.d.f. of the range W, as given in Example 3.9.8. There actually is intuition to support those two distributions being the same.

15. For any n sets of real numbers $A_1, \ldots, A_n$, we have

$$\begin{aligned} \Pr(Y_1 \in A_1, \ldots, Y_n \in A_n) &= \Pr\left[r_1(X_1) \in A_1, \ldots, r_n(X_n) \in A_n\right] \\ &= \Pr\left[r_1(X_1) \in A_1\right] \ldots \Pr\left[r_n(X_n) \in A_n\right] \\ &= \Pr(Y_1 \in A_1) \ldots \Pr(Y_n \in A_n). \end{aligned}$$

Therefore, $Y_1, \ldots, Y_n$ are independent by Definition 3.5.2.

16. If f factors in the form given in this exercise, then there must exist a constant $c > 0$ such that the marginal joint p.d.f. of X_1 and X_2 is

$$f_{12}(x_1, x_2) = cg(x_1, x_2) \quad \text{for} \quad (x_1, x_2) \in R^2,$$

the marginal joint p.d.f. of X_3, X_4, and X_5 is

$$f_{345}(x_3, x_4, x_5) = \frac{1}{c} h(x_3, x_4, x_5) \quad \text{for} \quad (x_3, x_4, x_5) \in R^3,$$

and, therefore, for every point $(x_1 \ldots, x_5) \in R^5$ we have

$$f(x_1, \ldots, x_5) = f_{12}(x_1, x_2) f_{345}(x_3, x_4, x_5).$$

It now follows that for any sets of real numbers A_1 and A_2,

$$\begin{aligned}
\Pr(Y_1 \in A_1 \text{ and } Y_2 \in A_2) &= \int \cdots \int_{\substack{r_1(x_1,x_2)\in A_1 \text{ and} \\ r_2(x_3,x_4,x_5)\in A_2}} f(x_1, \ldots, x_5)\, dx_1 \ldots dx_5 \\
&= \int\int_{r_1(x_1,x_2)\in A_1} f_{12}(x_1, x_2)\, dx_1\, dx_2 \int\int\int_{r_2(x_3,x_4,x_5)\in A_2} f_{345}(x_3, x_4, x_5)\, dx_3\, dx_4\, dx_5 \\
&= \Pr(Y_1 \in A_1) Pr(Y_2 \in A_2).
\end{aligned}$$

Therefore, by definition, Y_1 and Y_2 are independent.

17. We need to transform (X, Y) to (Z, W), where $Z = XY$ and $W = Y$. The joint p.d.f. of (X, Y) is

$$f(x, y) = \begin{cases} y \exp(-xy) f_2(y) & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The inverse transformation is $x = z/w$ and $y = w$. The Jacobian is

$$J = \det \begin{pmatrix} 1/w & -z/w^2 \\ 0 & 1 \end{pmatrix} = \frac{1}{w}.$$

The joint p.d.f. of (Z, W) is

$$g(z, w) = f(z/w, w)/w = w \exp(-z) f_2(w)/w = \exp(-z) f_2(w), \text{ for } z > 0.$$

This is clearly factored in the appropriate way to show that Z and W are independent. Indeed, if we integrate $g(z, w)$ over w, we obtain the marginal p.d.f. of Z, namely $g_1(z) = \exp(-z)$, for $z > 0$. This is the same as the function in (3.9.18).

18. We need to transform (X, Y) to (Z, W), where $Z = X/Y$ and $W = Y$. The joint p.d.f. of (X, Y) is

$$f(x, y) = \begin{cases} 3x^2 f_2(y)/y^3 & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The inverse transformation is $x = zw$ and $y = w$. The Jacobian is

$$J = \det \begin{pmatrix} w & z \\ 0 & 1 \end{pmatrix} = w.$$

The joint p.d.f. of (Z, W) is

$$g(z, w) = f(zw, w)w = 3z^2 w^2 f_2(w) w/w^3 = 3z^2 f_2(w), \text{ for } 0 < x < 1.$$

This is clearly factored in the appropriate way to show that Z and W are independent. Indeed, if we integrate $g(z, w)$ over w, we obtain the marginal p.d.f. of Z, namely $g_1(z) = 3z^2$, for $0 < z < 1$.

19. This is a convolution. Let g be the p.d.f. of Y. By (3.9.5) we have, for $y > 0$,

$$\begin{aligned} g(y) &= \int f(y-z)f(z)dx \\ &= \int_0^y e^{z-y}e^{-z}dz \\ &= ye^{-y}. \end{aligned}$$

Clearly, $g(y) = 0$ for $y < 0$, so the p.d.f.øf Y is

$$g(y) = \begin{cases} ye^{-y} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

20. Let f_1 stand for the marginal p.d.f. of X_1, namely $f_1(x) = \int f(x, x_2)dx_2$. With $a_2 = 0$ and $a_1 = a$ in (3.9.2) we get

$$\begin{aligned} g(y) &= \int_{-\infty}^{\infty} f\left(\frac{y-b}{a}, x_2\right)\frac{1}{|a|}dx_2 \\ &= \frac{1}{|a|}f_1\left(\frac{y-b}{a}\right), \end{aligned}$$

which is the same as (3.8.1).

21. Transforming to $Z_1 = X_1/X_2$ and $Z_2 = X_1$ has the inverse $X_1 = Z_2$ and $X_2 = Z_2/Z_1$. The set of values where the joint p.d.f. of Z_1 and Z_2 is positive is where $0 < z_2 < 1$ and $0 < z_2/z_1 < 1$. This can be written as $0 < z_2 < \min\{1, z_1\}$. The Jacobian is the determinant of the matrix

$$\begin{pmatrix} 0 & 1 \\ -z_2/z_1^2 & 1/z_1 \end{pmatrix},$$

which is $|z_2/z_1^2|$. The joint p.d.f. of Z_1 and Z_2 is then

$$g(z_1, z_2) = \left|\frac{z_2}{z_1^2}\right| 4z_2\frac{z_2}{z_1} = 4z_2^3 z_1^3,$$

for $0 < z_2 < \min\{1, z_1\}$. Integrating z_2 out of this yields, for $z_1 > 0$,

$$\begin{aligned} g_1(z_1) &= \int_0^{\min\{1,z_1\}} 4\frac{z_2^3}{z_1^3}dz_2 \\ &= \frac{\min\{z_1, 1\}^4}{z_1^3} \\ &= \begin{cases} z_1 & \text{if } z_1 < 1, \\ z_1^{-3} & \text{if } z_1 \geq 1. \end{cases} \end{aligned}$$

This is the same thing we got in Example 3.9.11.

3.10 Markov Chains

Commentary

Instructors can discuss this section at any time that they find convenient or they can omit it entirely. Instructors who wish to cover Sec. 12.5 (Markov chain Monte Carlo) and who wish to give some theoretical justification for the methodology will want to discuss some of this material before covering Sec. 12.5. On the other hand, one could cover Sec. 12.5 and skip the justification for the methodology without introducing Markov chains at all.

Students may notice the following property, which is exhibited in some of the exercises at the end of this section: Suppose that the Markov chain is in a given state s_i at time n. Then the probability of being in a particular state s_j a few periods later, say at time $n+3$ or $n+4$, is approximately the same for each possible given state s_i at time n. For example, in Exercise 2, the probability that it will be sunny on Saturday is approximately the same regardless of whether it is sunny or cloudy on the preceding Wednesday, three days earlier. In Exercise 5, for given probabilities on Wednesday, the probability that it will be cloudy on Friday is approximately the same as the probability that it will be cloudy on Saturday. In Exercise 7, the probability that the student will be on time on the fourth day of class is approximately the same regardless of whether he was late or on time on the first day of class. In Exercise 10, the probabilities for $n = 3$ and $n = 4$, are generally similar. In Exercise 11, the answers in part (a) and part (b) are almost identical.

This property is a reflection of the fact that for many Markov chains, the nth power of the transition matrix $\boldsymbol{P}^n$ will converge, as $n \to \infty$, to a matrix for which all the elements in any given column are equal. For example, in Exercise 2, the matrix $\boldsymbol{P}^n$ converges to the following matrix:

$$\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

This type of convergence is an example of Theorem 3.10.4. This theorem, and analogs for more complicated Markov chains, provide the justification of the Markov chain Monte Carlo method introduced in Sec. 12.5.

Solutions to Exercises

1. The transition matrix for this Markov chain is

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

   (a) If we multiply the initial probability vector by this matrix we get

$$\boldsymbol{v}\boldsymbol{P} = \left(\frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{2}{3}, \frac{1}{2}\frac{2}{3} + \frac{1}{2}\frac{1}{3}\right) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

   (b) The two-step transition matrix is $\boldsymbol{P}^2$, namely

$$\begin{bmatrix} \frac{1}{3}\frac{1}{3} + \frac{2}{3}\frac{2}{3} & \frac{1}{3}\frac{2}{3} + \frac{2}{3}\frac{1}{3} \\ \frac{2}{3}\frac{1}{3} + \frac{1}{3}\frac{2}{3} & \frac{2}{3}\frac{2}{3} + \frac{1}{3}\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{5}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{5}{9} \end{bmatrix}.$$

2. (a) 0.4, the lower right corner of the matrix.

(b) $(0.7)(0.7) = 0.49$.

(c) The probability that it will be cloudy on the next three days is $(0.4)(0.4)(0.4) = 0.064$. The desired probability is $1 - 0.064 = 0.936$.

3. Saturday is three days after Wednesday, so we first compute

$$\boldsymbol{P}^3 = \begin{bmatrix} 0.667 & 0.333 \\ 0.666 & 0.334 \end{bmatrix}.$$

Therefore, the answers are (a) 0.667 and (b) 0.666.

4. (a) From Exercise 3, the probability that it will be sunny on Saturday is 0.667. Therefore, the answer is $(0.667)(0.7) = 0.4669$.

(b) From Exercise 3, the probability that it will be sunny on Saturday is 0.666. Therefore, the answer is $(0.666)(0.7) = 0.4662$.

5. Let $\boldsymbol{v} = (0.2, 0.8)$.

(a) The answer will be the second component of the vector $\boldsymbol{vP}$. We easily compute $\boldsymbol{vP} = (0.62, 0.38)$, so the probability is 0.38.

(b) The answer will be the second component of $\boldsymbol{vP}^2$. We can compute $\boldsymbol{vP}^2$ by multiplying $\boldsymbol{vP}$ by $\boldsymbol{P}$ to get $(0.662, 0.338)$, so the probability is 0.338.

(c) The answer will be the second component of of $\boldsymbol{vP}^3$. Since $\boldsymbol{vP}^3 = (0.6662, 0.3338)$, the answer is 0.3338.

6. In this exercise (and the next two) the transition matrix $\boldsymbol{P}$ is

	Late	On time
Late	0.2	0.8
On time	0.5	0.5

(a) $(0.8)(0.5)(0.5) = 0.2$

(b) $(0.5)(0.2)(0.2) = 0.02$.

7. Using the matrix in Exercise 6, it is found that

$$\boldsymbol{P}^3 = \begin{bmatrix} 0.368 & 0.632 \\ 0.395 & 0.605 \end{bmatrix}.$$

Therefore, the answers are (a) 0.632 and (b) 0.605.

8. Let $\boldsymbol{v} = (0.7, 0.3)$.

(a) The answer will be the first component of the vector $\boldsymbol{vP}$. We can easily compute $\boldsymbol{vP} = (0.29, 0.71)$, so the answer is 0.29.

(b) The answer will be the second component of the vector $\boldsymbol{vP}^3$. We compute $\boldsymbol{vP}^3 = (0.3761, 0.6239)$, so the answer is 0.6239.

9. (a) It is found that

$$P^2 = \begin{bmatrix} \frac{3}{16} & \frac{7}{16} & \frac{2}{16} & \frac{4}{16} \\ 0 & 1 & 0 & 0 \\ \frac{3}{8} & \frac{1}{8} & \frac{2}{8} & \frac{2}{8} \\ \frac{4}{16} & \frac{6}{16} & \frac{3}{16} & \frac{3}{16} \end{bmatrix}.$$

The answer is given by the element in the third row and second column.

(b) The answer is the element in the first row and third column of $\boldsymbol{P}^3$, namely 0.125.

10. Let $\boldsymbol{v} = \left(\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}\right)$.

(a) The probabilities for s_1, s_2, s_3, and s_4 will be the four components of the vector $\boldsymbol{vP}$.

(b) The required probabilities will be the four components of $\boldsymbol{vP}^2$.

(c) The required probabilities will be the four components of $\boldsymbol{vP}^3$.

11. The transition matrix for the states A and B is

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

It is found that

$$\boldsymbol{P}^4 = \begin{bmatrix} \frac{41}{81} & \frac{40}{81} \\ \frac{40}{81} & \frac{41}{81} \end{bmatrix}.$$

Therefore, the answers are (a) $\frac{40}{81}$ and (b) $\frac{41}{81}$.

12. (a) Using the transition probabilities stated in the exercise, we construct

$$\boldsymbol{P} = \begin{bmatrix} 0.0 & 0.2 & 0.8 \\ 0.6 & 0.0 & 0.4 \\ 0.5 & 0.5 & 0.0 \end{bmatrix}.$$

(b) It is found that

$$\boldsymbol{P}^2 = \begin{bmatrix} 0.52 & 0.40 & 0.08 \\ 0.20 & 0.32 & 0.48 \\ 0.30 & 0.10 & 0.60 \end{bmatrix}.$$

Let $\boldsymbol{v} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. The probabilities that A, B, and C will have the ball are equal to the three components of $\boldsymbol{vP}^2$. Since the third component is largest, it is most likely that C will have the ball at time $n + 2$.

13. The states are triples of possible outcomes: (HHH), (HHT), (HTH), etc. There are a total of eight such triples. The conditional probabilities of the possible values of the outcomes on trials $(n-1, n, n+1)$ given all trials up to time n depend only on the trials $(n-2, n-1, n)$ and not on n itself, hence we have a Markov chain with stationary transition probabilities. Every row of the transition matrix has the following form except the two corresponding to (HHH) and (TTT). Let a, b, c stand for three arbitrary elements of $\{H, T\}$, not all equal. The row for (abc) has 0 in every column except for the two columns (abH) and (abT), which have 1/2 in each. In the (HHH) row, every column has 0 except the (HHT) column, which has 1. In the (TTT) row, every column has 0 except the (TTH) column which has 1.

14. Since we switch a pair of balls during each operation, there are always three balls in box A during this process. There are a total of nine red balls available, so there are four possible states of the proposed Markov chain, 0, 1, 2, 3, each state giving the number of red balls in box A. The possible compositions of box A after the nth operation clearly depend only on the composition after the $n-1$st operation, so we have a Markov chain. Also, balls are drawn at random during each operation, so the probabilities of transition depend only on the current state. Hence, the transition probabilities are stationary. If there are currently 0 red balls in box A, then we shall certainly remove a green ball. The probability that we get a red ball from box B is 9/10, otherwise we stay in state 0. So, the first row of $\boldsymbol{P}$ is $(1/10, 9/10, 0, 0)$. If we start with 1 red ball, then we remove that ball with probability 1/3. We replace whatever we draw with a red ball with probability 8/10. So we can either go to state 0 (probability $1/3 \times 2/10$), stay in state 1 (probability $1/3 \times 8/10 + 2/3 \times 2/10$), or go to state 2 (probability $2/3 \times 8/10$). The second row of $\boldsymbol{P}$ is $(1/15, 2/5, 8/15, 0)$. If we start with 2 red balls, we remove one with probability 2/3 and we replace it with red with probability 7/10. So, the third row of $\boldsymbol{P}$ is $(0, 1/5, 17/30, 7/30)$. If we start with 3 red balls, we certainly remove one and we replace it by red with probability 6/10, so the fourth row of $\boldsymbol{P}$ is $(0, 0, 2/5, 3/5)$.

15. We are asked to verify the numbers in the second and fifth rows of the matrix in Example 3.10.6. For the second row, the parents have genotypes AA and Aa, so that the only possible offspring are AA and Aa. Each of these occurs with probability 1/2 because they are determined by which allele comes from the Aa parent. Since the two offspring in the second generation are independent, we will get $\{AA, AA\}$ with probability $(1/2)^2 = 1/4$ and we will get $\{Aa, Aa\}$ with probability 1/4 also. The remaining probability, 1/2, is the probability of $\{AA, Aa\}$. For the fifth row, the parent have genotypes Aa and aa. The only possible offspring are Aa and aa. Indeed, the situation is identical to the second row with a and A switched. The resulting probabilities are also the same after this same switch.

16. We have to multiply the initial probability vector into the transition matrix and do the arithmetic. For the first coordinate, we obtain

$$\frac{1}{16} \times 1 + \frac{1}{4} \times 0.25 + \frac{1}{4} \times 0.0625 = \frac{9}{64}.$$

The other five elements are calculated in a similar fashion. The resulting vector is

$$\left(\frac{9}{64}, \frac{3}{16}, \frac{1}{32}, \frac{5}{16}, \frac{3}{16}, \frac{9}{64}\right).$$

17. (a) We are asked to find the conditional distribution of X_n given $X_{n-1} = \{Aa, aa\}$ and $X_{n+1} = \{AA, aa\}$. For each possible state x_n, we can find

$$\begin{aligned} &\Pr(X_n = x_n | X_{n-1} = \{Aa, aa\}, X_{n+1} = \{AA, aa\}) \\ &= \frac{\Pr(X_n = x_n, X_{n+1} = \{AA, aa\} | X_{n-1} = \{Aa, aa\})}{\Pr(X_{n+1} = \{AA, aa\} | X_{n-1} = \{Aa, aa\})}. \end{aligned} \qquad \text{(S.3.4)}$$

The denominator is 0.0313 from the 2-step transition matrix in Example 3.10.9. The numerator is the product of two terms from the 1-step transition matrix: one from $\{Aa, aa\}$ to x_n and the other from x_n to $\{AA, aa\}$. These products are as follows:

x_n: $\{AA,AA\}$	$\{AA,Aa\}$	$\{AA,aa\}$	$\{Aa,Aa\}$	$\{Aa,aa\}$	$\{aa,aa\}$
0	0	0	0.25×0.125	0	0

Plugging these into (S.3.4) gives

$$\Pr(X_n = \{Aa, Aa\}|X_{n+1} = \{Aa, aa\}, X_{n+1} = \{AA, aa\}) = 1,$$

and all other states have probability 0.

(b) This time, we want

$$\begin{aligned}&\Pr(X_n = x_n|X_{n-1} = \{Aa, aa\}, X_{n+1} = \{aa, aa\})\\ &= \frac{\Pr(X_n = x_n, X_{n+1} = \{aa, aa\}|X_{n-1} = \{Aa, aa\})}{\Pr(X_{n+1} = \{aa, aa\}|X_{n-1} = \{Aa, aa\})}.\end{aligned}$$

The denominator is 0.3906. The numerator products and their ratios to the denominator are:

x_n	$\{AA,AA\}$	$\{AA,Aa\}$	$\{AA,aa\}$	$\{Aa,Aa\}$	$\{Aa,aa\}$	$\{aa,aa\}$
Numerator	0	0	0	0.25×0.0625	0.5×0.25	0.25×1
Ratio	0	0	0	0.0400	0.3200	0.6400

This time, we get

$$\Pr(X_n = x_n|X_{n-1} = \{Aa, aa\}, X_{n+1} = \{Aa, Aa\}) = \begin{cases} 0.04 & \text{if } x_n = \{Aa, Aa\}, \\ 0.32 & \text{if } x_n = \{Aa, aa\}, \\ 0.64 & \text{if } x_n = \{aa, aa\}, \end{cases},$$

and all others are 0.

18. We can see from the 2-step transition matrix that it is possible to get from every non-absorbing state into each of the absorbing states in two steps. So, no matter what non-absorbing state we start in, the probability is one that we will eventually end up in one of absorbing states. Hence, no distribution with positive probability on any non-absorbing state can be a stationary distribution.

19. The matrix $\boldsymbol{G}$ and its inverse are

$$\begin{aligned}\boldsymbol{G} &= \begin{pmatrix} -0.3 & 1 \\ 0.6 & 1 \end{pmatrix},\\ \boldsymbol{G}^{-1} &= -\frac{10}{9}\begin{pmatrix} 1 & -1 \\ -0.6 & -0.3 \end{pmatrix}.\end{aligned}$$

The bottom row of $\boldsymbol{G}^{-1}$ is $(2/3, 1/3)$, the unique stationary distribution.

20. The argument is essentially the same as in Exercise 18. All probability in non-absorbing states eventually moves into the absorbing states after sufficiently many transitions.

3.11 Supplementary Exercises

Solution to Exercises

1. We can calculate the c.d.f. of Z directly.

$$\begin{aligned}F(z) &= \Pr(Z \le z) = \Pr(Z = X)\Pr(X \le z) + \Pr(Z = Y)\Pr(Y \le z)\\ &= \frac{1}{2}\Pr(X \le z) + \frac{1}{2}\Pr(Y \le z)\end{aligned}$$

The graph is in Fig. S.3.35.

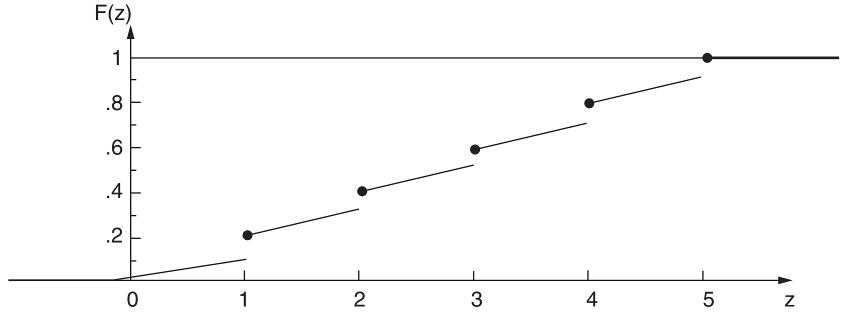

Figure S.3.35: Graph of c.d.f. for Exercise 1 of Sec. 3.11.

2. Let $x_1, \ldots, x_k$ be the finitely many values for which $f_1(x) > 0$. Since X and Y are independent, the conditional distribution of $Z = X + Y$ given $X = x$ is the same as the distribution of $x + Y$, which has the p.d.f. $f_2(z - x)$, and the c.d.f. $F_2(z - x)$. By the law of total probability the c.d.f. of Z is $\sum_{i=1}^{k} F_2(z - x_i)f_1(x_i)$. Notice that this is a weighted average of continuous functions of z, $F_2(z - x_i)$ for $i = 1, \ldots, k$, hence it is a continuous function. The p.d.f. of Z can easily be found by differentiating the c.d.f. to obtain $\sum_{i=1}^{k} f_2(z - x_i)f_1(x_i)$.

3. Since $F(x)$ is continuous and differentiable everywhere except at the points $x = 0$, 1, and 2,

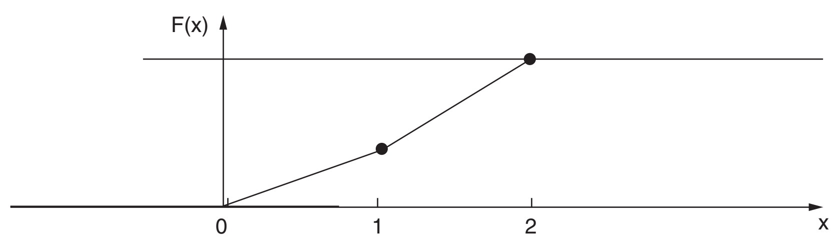

Figure S.3.36: Graph of c.d.f. for Exercise 3 of Sec. 3.11.

$$f(x) = \frac{dF(x)}{dx}\begin{cases} \dfrac{2}{5} & \text{for } 0 < x < 1, \\ \dfrac{3}{5} & \text{for } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

4. Since $f(x)$ is symmetric with respect to $x = 0$, $F(0) = Pr(X \leq 0) = 0.5$. Hence,

$$\int_0^{x_0} f(x)\,dx = \frac{1}{2}\int_0^{x_0} \exp(-x)\,dx = .4.$$

It follows that $\exp(-x_0) = .2$ and $x_0 = \log 5$.

5. X_1 and X_2 have the uniform distribution over the square, which has area 1. The area of the quarter circle in Fig. S.3.37, which is the required probability, is $\pi/4$.

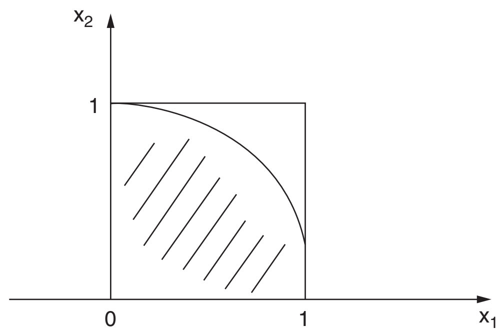

Figure S.3.37: Region for Exercise 5 of Sec. 3.11.

6. (a) $$\Pr(X \text{ divisible by } n) = f(n) + f(2n) + f(3n) + \cdots = \sum_{x=1}^{\infty} \frac{1}{c(p) \quad (nx)^p} = \frac{1}{n^p}.$$

   (b) By part (a), $\Pr(X \text{ even}) = 1/2^p$. Therefore, $\Pr(X \text{ odd}) = 1 - 1/2^p$.

7. $$\begin{aligned}\Pr(X + X_2 \text{ even}) &= \Pr(X_1 \text{ even})\Pr(X_2 \text{ even}) + \Pr(X_1 \text{ odd})\Pr(X_2 \text{ odd}) \\ &= \left(\frac{1}{2^p}\right)\left(\frac{1}{2^p}\right) + \left(1 - \frac{1}{2^p}\right)\left(1 - \frac{1}{2^p}\right) \\ &= 1 - \frac{1}{2^{p-1}} + \frac{1}{2^{2p-1}}.\end{aligned}$$

8. Let $G(x)$ devote the c.d.f. of the time until the system fails, let A denote the event that component 1 is still operating at time x, and let B denote the event that at least one of the other three components is still operating at time x. Then

   $$1 - G(x) = \Pr(\text{System still operating at time } x) = \Pr(A \cap B) = \Pr(A)\Pr(B) = [1 - F(x)][1 - F^3(x)].$$

   Hence, $G(x) = F(x)\,[1 + F^2(x) - F^3(x)]$.

9. Let A denote the event that the tack will land with its point up on all three tosses. Then $\Pr(A \mid X = x) = x^3$. Hence,

   $$\Pr(A) = \int_0^1 x^3 f(x)\,dx = \frac{1}{10}.$$

10. Let Y denote the area of the circle. Then $Y = \pi X^2$, so the inverse transformation is

    $$x = (y/\pi)^{1/2} \quad \text{and} \quad \frac{dx}{dy} = \frac{1}{2(\pi y)^{1/2}}.$$

    Also, if $0 < x < 2$, then $0 < y < 4\pi$. Thus,

    $$g(y) = \frac{1}{16(\pi y)^{1/2}}\left[3\left(\frac{y}{\pi}\right)^{1/2} + 1\right] \quad \text{for} \quad 0 < y < 4\pi$$

    and $g(y) = 0$ otherwise.

11. $F(x) = 1 - \exp(-2x)$ for $x > 0$. Therefore, by the probability integral transformation, $F(X)$ will have the uniform distribution on the interval $[0, 1]$. Therefore,

$$Y = 5F(X) = 5(1 - \exp(-2X))$$

will have the uniform distribution on the interval $[0, 5]$.

It might be noted that if Z has the uniform distribution on the interval $[0, 1]$, then $1 - Z$ has the same uniform distribution. Therefore,

$$Y = 5[1 - F(X)] = 5\exp(-2X)$$

will also have the uniform distribution on the interval $[0, 5]$.

12. This exercise, in different words, is exactly the same as Exercise 7 of Sec. 1.7 and, therefore, the solution is the same.

13. Only in (c) and (d) is the joint p.d.f. of X and Y positive over a rectangle with sides parallel to the axes, so only in (c) and (d) is there the possibility of X and Y being independent. Since the uniform density is constant, it can be regarded as being factored in the form of Eq. (3.5.7). Hence, X and Y are independent in (c) and (d).

14. The required probability p is the probability of the shaded area in Fig. S.3.38. Therefore,

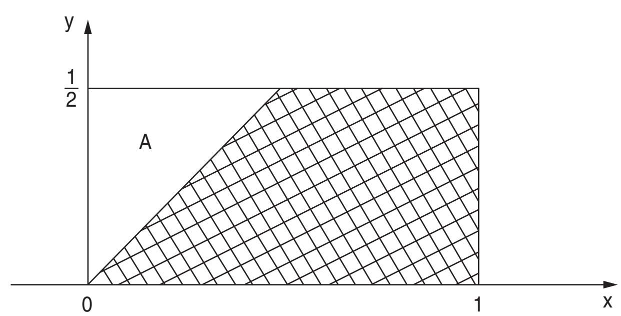

Figure S.3.38: Figure for Exercise 14 of Sec. 3.11.

$$p = 1 - \Pr(A) = 1 - \int_0^{1/2} \int_0^y f_1(x)\, f_2(y)\, dx\, dy = 1 - 1/3 = 2/3.$$

15. This problem is similar to Exercise 11 of Sec. 3.5, but now we have Fig. S.3.39. The area of the shaded region is now $550 + 787.5 = 1337.5$. Hence, the required probability is $\frac{1337.5}{3600} = .3715$

16. For $0 < x < 1$,

$$f_1(x) = \int_x^1 2(x + y)dy = 1 + 2x - 3x^2.$$

Therefore, $\Pr\left(X < \frac{1}{2}\right) = \int_0^{1/2} f_1(x)\, dx = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{5}{8}.$

Finally, for $0 < x, y < 1$,

$$g_2(y \mid x) = \frac{f(x, y)}{f_1(x)} = \frac{2(x + y)}{1 + 2x - 3x^2}.$$

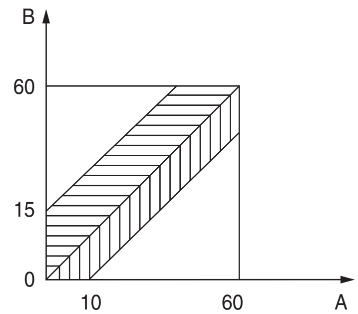

Figure S.3.39: Figure for Exercise 15 of Sec. 3.11.

17. $f(x, y) = f(x)\, g(y \mid x) = \dfrac{9y^2}{x}$ for $0 < y < x < 1$.

Hence,

$$f_2(y) = \int_y^1 f(x, y)\, dx = -9y^2 \log(y) \quad \text{for} \quad 0 < y < 1$$

and

$$g_1(x \mid y) = \frac{f(x, y)}{f_2(y)} = -\frac{1}{x \log(y)} \qquad \text{for} \quad 0 < y < x < 1.$$

18. X and Y have the uniform distribution over the region shown in Fig. S.3.40. The area of this region is

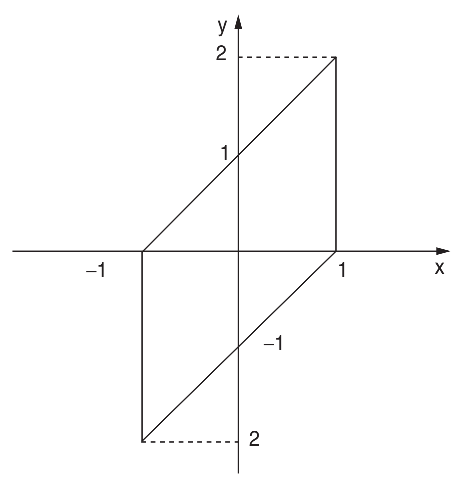

Figure S.3.40: Region for Exercise 18 of Sec. 3.11.

4. The area in the second plus the fourth quadrants is 1. Therefore, the area in the first plus the third quadrants is 3, and

$$\Pr(XY > 0) = \frac{3}{4}.$$

Furthermore, for any value of x $(-1 < x < 1)$, the conditional distribution of Y given that $X = x$ will be a uniform distribution over the interval $[x - 1, x + 1]$. Hence,

$$g_2(y \mid x) = \begin{cases} \frac{1}{2} & \text{for } x - 1 < y < x + 1, \\ 0 & \text{otherwise.} \end{cases}$$

19.

$$\begin{aligned} f_1(x) &= \int_x^1 \int_y^1 6 \, dz \, dy = 3 - 6x + 3x^2 = 3(1 - x)^2 \quad \text{for } 0 < x < 1, \\ f_2(y) &= \int_0^y \int_y^1 6 \, dz \, dx = 6y(1 - y) \quad \text{for } 0 < y < 1. \\ f_3(z) &= \int_0^z \int_0^y 6 \, dx \, dy = 3z^2 \quad \text{for } 0 < z < 1. \end{aligned}$$

20. Since $f(x, y, z)$ can be factored in the form $g(x, y)h(z)$ it follows that Z is independent of the random variables X and Y. Hence, the required conditional probability is the same as the unconditional probability $\Pr(3X > Y)$. Furthermore, it follows from the factorization just given that the marginal p.d.f. $h(z)$ is constant for $0 < z < 1$. Thus, this constant must be 1 and the marginal joint p.d.f. of X and Y must be simply $g(x, y) = 2$, for $0 < x < y < 1$. Therefore,

$$\Pr(3X > Y) = \int_0^1 \int_{y/3}^y 2 dx dy = \frac{2}{3}.$$

The range of integration is illustrated in Fig. S.3.41.

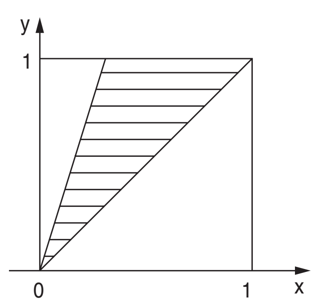

Figure S.3.41: Range of integration for Exercise 20 of Sec. 3.11.

21. (a)
$$f(x, y) = \begin{cases} \exp(-(x + y)) & \text{for } x > 0, \, y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Also, $x = uv$ and $y = (1 - u)v$, so

$$J = \begin{vmatrix} v & u \\ -v & 1 - u \end{vmatrix} = v > 0.$$

Therefore,

$$g(u, v) = f(uv, [1 - u]v) \, | \, J \, | = \begin{cases} v \exp(-v) & \text{for } 0 < u < 1, \, v > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Because g can be appropriately factored (the factor involving u is constant) and it is positive over an appropriate rectangle, it follows that U and V are independent.

22. Here, $x = uv$ and $y = v$, so

$$J = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v > 0.$$

Therefore,

$$g(u, v) = f(uv, v)\,|\,J\,| = \begin{cases} 8uv^3 & \text{for } 0 \le u \le 1,\ 0 \le v \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

X and Y are not independent because $f(x, y) > 0$ over a triangle, not a rectangle. However, it can be seen that U and V are independent.

23. Here, $f(x) = \dfrac{dF(x)}{dx} = \exp(-x)$ for $x > 0$. It follows from the results in Sec. 3.9 that

$$g(y_1, y_n) = n(n-1)(\exp(-y_1) - \exp(-y_n))^{n-2}\ \exp(-(y_1 + y_n))$$

for $0 < y_1 < y_n$. Also, the marginal p.d.f. of Y_n is

$$g_n(y_n) = n(1 - \exp(-y_n))^{n-1} \exp(-y_n) \qquad \text{for } y_n > 0.$$

Hence,

$$h(y_1 \,|\, y_n) = \frac{(n-1)(\exp(-y_1) - \exp(-y_n))^{n-2} \exp(-y_1)}{(1 - \exp(-yn))^{n-1}} \qquad \text{for } \ 0 < \ y_1 < y_n.$$

24. As in Example 3.9.7, let $W = Y_n - Y_1$ and $Z = Y_1$. The joint p.d.f. $g(w, z)$ of (W, Z) is, for $0 < w < 1$ and $0 < z < 1 - w$,

$$g(w, z) = 24[(w + z)^2 - z^2]\ z\ (w + z) = 24\ w\ (2z^3 + 3wz^2 + w^2 z),$$

and 0 otherwise. Hence, the p.d.f. of the range is, for $0 < w < 1$,

$$h(w) = \int_0^{1-w} g(w, z)\, dz = 12w(1 - w)^2.$$

25. (a) Let f_2 be the marginal p.d.f. of Y. We approximate

$$\Pr(y - \epsilon < Y \le y + \epsilon) = \int_{y-\epsilon}^{y+\epsilon} f_2(t)dt \approx 2\epsilon f_2(y).$$

(b) For each s, we approximate

$$\int_{y-\epsilon}^{y+\epsilon} f(s, t)dt \approx 2\epsilon f(s, y).$$

Using this, we can approximate

$$\Pr(X \le x, y - \epsilon < Y \le y + \epsilon) = \int_{-\infty}^{x} \int_{y-\epsilon}^{y+\epsilon} f(s, t)dtds \approx 2\epsilon \int_{-\infty}^{x} f(s, y)ds.$$

(c) Taking the ratio of the approximation in part (b) to the approximation in part (a), we obtain

$$\begin{aligned}\Pr(X \le x | y-\epsilon < Y \le y+\epsilon) &= \frac{\Pr(X \le x, y-\epsilon < Y \le y+\epsilon)}{\Pr(y-\epsilon < Y \le y+\epsilon)} \\ &\approx \frac{\int_{-\infty}^{x} f(s,y)ds}{f_2(y)} \\ &= \int_{-\infty}^{x} g_1(s|y)ds.\end{aligned}$$

26. (a) Let $Y = X_1$. The transformation is $Y = X_1$ and $Z = X_1 - X_2$. The inverse is $x_1 = y$ and $x_2 = y - z$. The Jacobian has absolute value 1. The joint p.d.f. of (Y, Z) is

$$g(y,z) = \exp(-y-(y-z)) = \exp(-2y+z),$$

for $y > 0$ and $z < y$.

(b) The marginal p.d.f. of Z is

$$\int_{\max\{0,z\}}^{\infty} \exp(-2y+z)dy = \frac{1}{2}\exp(z)\exp(-2\max\{0,z\}) = \frac{1}{2}\begin{cases} \exp(-z) & \text{if } z \ge 0, \\ \exp(z) & \text{if } z < 0. \end{cases}$$

The conditional p.d.f. of $Y = X_1$ given $Z = 0$ is the ratio of these two with $z = 0$, namely

$$g_1(x_1|0) = 2\exp(-2x_1), \text{ for } x_1 > 0.$$

(c) Let $Y = X_1$. The transformation is now $Y = X_1$ and $W = X_1/X_2$. The inverse is $x_1 = y$ and $x_2 = y/w$. The Jacobian is

$$J = \det\begin{pmatrix} 1 & 0 \\ 1/w & -y/w^2 \end{pmatrix} = -\frac{y}{w^2}.$$

The joint p.d.f. of (Y, W) is

$$g(y,w) = \exp(-y-y/w)y/w^2 = y\exp(-y(1+1/w))/w^2,$$

for $y, w > 0$.

(d) The marginal p.d.f. of W is

$$\int_0^{\infty} y\exp(-y(1+1/w))/w^2 dy = \frac{1}{w^2(1+1/w)^2} = \frac{1}{(1+w)^2},$$

for $w > 0$. The conditional p.d.f. of $Y = X_1$ given $W = 1$ is the ratio of these two with $w = 1$, namely

$$h_1(x_1|1) = 4x_1\exp(-2x_1), \text{ for } x_1 > 0.$$

(e) The conditional p.d.f. g_1 in part (b) is supposed to be the conditional p.d.f. of X_1 given that Z is close to 0, that is, that $|X_1 - X_2|$ is small. The conditional p.d.f. h_1 in part (d) is supposed to be the conditional p.d.f. of X_1 given that W is close to 1, that is, that $|X_1/X_2 - 1|$ is small. The sets of (x_1, x_2) values such that $|x_1 - x_2|$ is small and that $|x_1/x_2 - 1|$ is small are drawn in Fig. S.3.42. One can see how the two sets, although close, are different enough to account for different conditional distributions.

27. The transition matrix is as follows:

		Players in game $n+1$		
		(A,B)	(A,C)	(B,C)
Players in	(A,B)	0	0.3	0.7
game n	(A,C)	0.6	0	0.4
	(B,C)	0.8	0.2	0

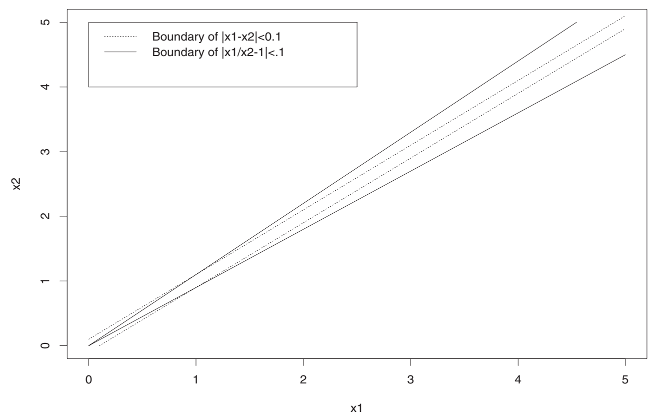

Figure S.3.42: Boundaries of the two regions where $|x_1 - x_2| < 0.1$ and $|x_1/x_2 - 1| < 0.1$ in Exercise 26e of Sec. 3.11.

28. If A and B play in the first game, then there are the following two sequences of outcomes which result in their playing in the fourth game:
i) A beats B in the first game, C beats A in the second game, B beats C in the third game;
ii) B beats A in the first game, C beats B in the second game, A beats C in the third game.

The probability of the first sequence is (0.3) (0.4) (0.8) = 0.096. The probability of the second sequence is (0.7) (0.2) (0.6) = 0.084. Therefore, the overall probability that A and B will play again in the fourth game is 0.18. The same sort of calcuations show that this answer will be the same if A and C play in the first game or if B and C play in the first game.

29. The matrix $\boldsymbol{G}$ and its inverse are

$$\boldsymbol{G} = \begin{pmatrix} -1.0 & 0.3 & 1.0 \\ 0.6 & -1.0 & 1.0 \\ 0.8 & 0.2 & 1.0 \end{pmatrix},$$

$$\boldsymbol{G}^{-1} = \begin{pmatrix} -0.5505 & -0.4587 & 0.5963 \\ 0.0917 & -0.8257 & 0.7339 \\ 0.4220 & 0.2018 & 0.3761 \end{pmatrix}.$$

The bottom row of $\boldsymbol{G}^{-1}$ is the unique stationary distribution, $(0.4220, 0.2018, 0.3761)$.

Chapter 4

Expectation

4.1 The Expectation of a Random Variable

Commentary

It is useful to stress the fact that the expectation of a random variable depends only on the distribution of the random variable. Every two random variables with the same distribution will have the same mean. This also applies to variance (Sec. 4.3), other moments and m.g.f. (Sec. 4.4), and median (Sec. 4.5). For this reason, one often refers to means, variance, quantiles, etc. of a distribution rather than of a random variable. One need not even have a random variable in mind in order to calculate the mean of a distribution.

Solutions to Exercises

1. The mean of X is

$$E(X) = \int xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

2. $E(X) = \dfrac{1}{100}(1+2+\cdots+100) = \dfrac{1}{100}\dfrac{(100)(101)}{2} = 50.5.$

3. The total number of students is 50. Therefore,

$$E(X) = 18\left(\frac{20}{50}\right) + 19\left(\frac{22}{50}\right) + 20\left(\frac{4}{50}\right) + 21\left(\frac{3}{50}\right) + 25\left(\frac{1}{50}\right) = 18.92.$$

4. There are eight words in the sentence and they are each equally probable. Therefore, the possible values of X and their probabilities are as follows:

x	$f(x)$
2	1/8
3	5/8
4	1/8
9	1/8

It follows that $E(X) = 2\left(\frac{1}{8}\right) + 3\left(\frac{5}{8}\right) + 4\left(\frac{1}{8}\right) + 9\left(\frac{1}{8}\right) = 3.75.$

5. There are 30 letters and they are each equally probable:

   2 letters appear in the only two-letter word;

   15 letters appear in three-letter words;

   4 letters appear in the only four-letter word;

   9 letters appear in the only nine-letter word.

   Therefore, the possible values of Y and their probabilities are as follows:

y	$g(y)$
2	2/30
3	15/30
4	4/30
9	9/30

$$E(Y) = 2\left(\frac{2}{30}\right) + 3\left(\frac{15}{30}\right) + 4\left(\frac{4}{30}\right) + 9\left(\frac{9}{30}\right) = 4.867.$$

6. $E\left(\frac{1}{X}\right) = \int_0^1 \frac{1}{x} \cdot 2xdx = 2.$

7. $E\left(\frac{1}{X}\right) = \int_0^1 \frac{1}{x}\,\mathrm{dx} = -\lim_{x\to 0}\log(x) = \infty$. Since the integral is not finite, $E\left(\frac{1}{X}\right)$ does not exist.

8. $E(XY) = \int_0^1 \int_0^x xy \cdot 12y^2 dydx = \frac{1}{2}.$

9. If X denotes the point at which the stick is broken, then X has the uniform distribution on the interval $[0,1]$. If Y denotes the length of the longer piece, then $Y = \max\{X, 1-X\}$. Therefore,

$$E(Y) = \int_0^1 \max(x, 1-x)dx = \int_0^{1/2}(1-x)dx + \int_{1/2}^1 xdx = \frac{3}{4}.$$

10. Since α has the uniform distribution on the interval $[-\pi/2, \pi/2]$, the p.d.f. of α is

$$f(\alpha) = \begin{cases} \frac{1}{\pi} & \text{for } -\frac{\pi}{2} < \alpha < \frac{\pi}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

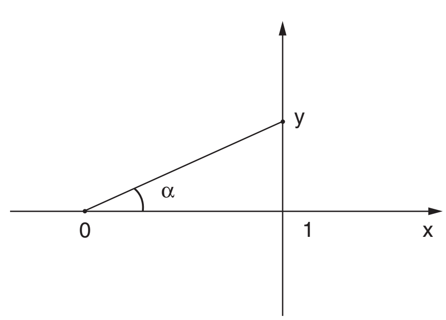

Figure S.4.1: Figure for Exercise 10 of Sec. 4.1.

Also, $Y = \tan(\alpha)$. Therefore, the inverse transformation is $\alpha = \tan^{-1} Y$ and $d\alpha/dy = 1/(1+y^2)$. As α varies over the interval $(-\pi/2, \pi/2)$, Y varies over the entire real line. Therefore, for $-\infty < y < \infty$, the p.d.f. of Y is

$$g(y) = f(\tan^{-1} y)\frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}.$$

11. The p.d.f.'s of Y_1 and Y_n were found in Sec. 3.9. For the given uniform distribution, the p.d.f. of Y_1 is

$$g_1(y) = \begin{cases} n(1-y)^{n-1} & \text{for} \quad 0 < y < 1, \\ 0 & \quad\quad\text{otherwise.} \end{cases}$$

Therefore,

$$E(Y_1) = \int_0^1 yn(1-y)^{n-1} dy = \frac{1}{n+1}.$$

The p.d.f. of Y_n is

$$g_n(y) = \begin{cases} ny^{n-1} & \text{for} \quad 0 < y < 1, \\ 0 & \quad\quad\text{otherwise.} \end{cases}$$

Therefore,

$$E(Y_n) = \int_0^1 y \cdot ny^{n-1} dy = \frac{n}{n+1}.$$

12. It follows from the probability integral transformation that the joint distribution of $F(X_1), \ldots, F(X_n)$ is the same as the joint distribution of a random sample from the uniform distribution on the interval $[0, 1]$. Since $F(Y_1)$ is the smallest of these values and $F(Y_n)$ is the largest, the distributions of these variables will be the same as the distributions of the smallest and the largest values in a random sample from the uniform distribution on the interval $[0, 1]$. Therefore, $E[F(Y_1)]$ and $E[F(Y_n)]$ will be equal to the values found in Exercise 11.

13. Let $p = \Pr(X = 300)$. Then $E(X) = 300p + 100(1-p) = 200p + 100$. For risk-neutrality, we need $E(X) = 110 * (1.058) = 116.38$. Setting $200p + 100 = 116.38$ yields $p = 0.0819$. The option has a value of 150 if $X = 300$ and it has a value of 0 if $X = 100$, so the mean of the option value is $150p = 12.285$. The present value of this amount is $12.285/1.058 = 11.61$, the risk-neutral price of the option.

14. For convenience, we shall not use dollar signs in these calculations.

    (a) We need to check the investor's net worth at the end of the year in four situations:

        i. $X = 180$ and she makes the transactions
        ii. $X = 180$ and she doesn't make the transactions
        iii. $X = 260$ and she makes the transactions
        iv. $X = 260$ and she doesn't make the transactions

Since we don't know the investor's entire net worth, we shall only calculate it relative to all other investments. This means that we only need to pretend as if the investor had one share of the stock worth 200. We don't care what else she has. We need to show that cases (i) and (ii) lead to the same net worth and that cases (iii) and (iv) lead to the same net worth. In case (ii), her net worth will change by -20. In case (iv), her net worth will change by 60. In case (i), nobody will exercise the options. So she will sell the three extra shares for 180 each (total 540) and pay the loan of 519.24 plus interest 20.77 for a net 0.01 loss. Plus her one original share of stock has lost 20 and her net worth has changed a total of -20.01, which is the same as case (i) except for the accumulated rounding error. In case (iii), the options will be exercised, and she will receive 800 for four shares of the stock. She will have to pay back the loan of 519.24 plus 20.77 in interest for a net gain of 259.99. But she no longer has the one share of stock that was worth 200, so her change in net worth is 59.99, the same as case (iv) to within the same one cent of rounding.

(b) If the option price is $x < 20.19$, then the investor only receives $4x$ for selling the options, but still needs to pay 600 for the three shares, so she must borrow $600 - 4x$. The rest of the calculations proceed just as in part (a) but we must replace 519.24 by $600 - 4x$, and the interest 20.77 must be replaced by $0.04(600 - 4x)$. That is, to pay back the loan with interest, she must pay $624 - 4.16x$ instead of 540.01. So she pays an additional $83.99 - 4.16x$ relative to the situation in case (a) regardless of what happens to the stock price.

(c) This situation is the same as part (b) except now the value of $83.99 - 4.16x$ is negative instead of positive, so the investor pays back less and hence makes additional profit rather than suffers additional loss.

15. The value of the option is 0 if $X = 260$ and it is 40 if $X = 180$, so the expected value of the option is $40(1 - p) = 40 \times 0.65 = 26$. The present value of this amount is $26/1.04 = 25$.

16. If f is the p.f. of X, and $Y = |X|$, then for $y \geq 0$, $\Pr(Y = y) = \Pr(X = y) + \Pr(X = -y)$. In Example 4.1.4, $\Pr(X = y) = \Pr(X = -y) = 1/[2y(y + 1)]$, and this makes $\Pr(Y = y)$ the p.f. in Example 4.1.5.

4.2 Properties of Expectations

Commentary

Be sure to stress the fact that Theorem 4.2.6 on the expected value of a product of random variables has the condition that the random variables are independent. This section ends with a derivation for the expectation of a nonnegative discrete random variable. Although this method has theoretical interest, it is not central to the rest of the text.

Solutions to Exercises

1. The random variable Y is equal to $10(R - 1.5)$ in dollars. The mean of Y is $10[E(R) - 1.5]$. From Exercise 1 in Sec. 4.1, we know that $E(R) = (-3 + 7)/2 = 2$, so $E(Y) = 5$.

2. $E(2X_1 - 3X_2 + X_3 - 4) = 2E(X_1) - 3E(X_2) + E(X_3) - 4 = 2(5) - 3(5) + 5 - 4 = -4$.

3.

$$\begin{aligned} E[(X_1 - 2X_2 + X_3)^2] &= E(X_1^2 + 4X_2^2 + X_3^2 - 4X_1X_2 + 2X_1X_3 - 4X_2X_3) \\ &= E(X_1^2) + 4E(X_2^2) + E(X_3^2) - 4E(X_1X_2) \\ &\quad + 2E(X_1X_3) - 4E(X_2X_3). \end{aligned}$$

Since X_1, X_2, and X_3 are independent,

$$E(X_iX_j) = E(X_i)E(X_j) \qquad \text{for } i \neq j.$$

Therefore, the above expectation can be written in the form:

$$E(X_1^2) + 4E(X_2^2) + E(X_3^2) - 4E(X_1)E(X_2) + 2E(X_1)E(X_3) - 4E(X_2)E(X_3).$$

Also, since each X_i has the uniform distribution on the interval $[0, 1]$, then $E(X_i) = \frac{1}{2}$ and

$$E(X_i^2) = \int_0^1 x^2 dx = \frac{1}{3}.$$

Hence, the desired expectation has the value 1/2.

4. The area of the rectangle is XY. Since X and Y are independent, $E(XY) = E(X)E(Y)$. Also, $E(X) = 1/2$ and $E(Y) = 7$. Therefore, $E(XY) = 7/2$.

5. For $i = 1, \ldots, n$, let $Y_i = 1$ if the observation X_i falls within the interval (a, b), and let $Y_i = 0$ otherwise. Then $E(Y_i) = \Pr(Y_i = 1) = \int_a^b f(x)dx$. The total number of observations that fall within the interval (a, b) is $Y_1 + \cdots + Y_n$, and

$$E(Y_1 + \cdots + Y_n) = E(Y_1) + \cdots + E(Y_n) = n \int_a^b f(x)dx.$$

6. Let $X_i = 1$ if the ith jump of the particle is one unit to the right and let $X_i = -1$ if the ith jump is one unit to the left. Then, for $i = 1, \ldots, n$,

$$E(X_i) = (-1)p + (1)(1-p) = 1 - 2p.$$

The position of the particle after n jumps is $X_1 + \cdots + X_n$, and

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n(1 - 2p).$$

7. For $i = 1, \ldots, n$, let $X_i = 2$ if the gambler's fortune is doubled on the ith play of the game and let $X_i = 1/2$ if his fortune is cut in half on the ith play. Then

$$E(X_i) = 2\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{5}{4}.$$

After the first play of the game, the gambler's fortune will be cX_1, after the second play it will be $(cX_1)X_2$, and by continuing in this way it is seen that after n plays the gambler's fortune will be $cX_1X_2 \ldots X_n$. Since $X_1, \ldots, X_n$ are independent,

$$E(cX_1 \ldots X_n) = cE(X_1) \ldots E(X_n) = c\left(\frac{5}{4}\right)^n.$$

8. It follows from Example 4.2.4 that

$$E(X) = 8\left(\frac{10}{25}\right) = \frac{16}{5}.$$

Since $Y = 8 - X$, $E(Y) = 8 - E(X) = \dfrac{24}{5}$. Finally, $E(X - Y) = E(X) - E(Y) = -\dfrac{8}{5}$.

9. We know that $E(X) = np$. Since $Y = n - X$, $E(X - Y) = E(2X - n) = 2E(X) - n = n(2p - 1)$.

10. (a) Since the probability of success on any trial is $p = 1/2$, it follows from the material presented at the end of this section that the expected number of tosses is $1/p = 2$.

    (b) The number of tails that will be obtained is equal to the total number of tosses minus one (the final head). Therefore, the expected number of tails is $2 - 1 = 1$.

11. We shall use the notation presented in the hint for this exercise. It follows from part (a) of Exercise 10 that $E(X_i) = 2$ for $i = 1, \dots, k$. Therefore,

$$E(X) = E(X_1) + \cdots + E(X_k) = 2k.$$

12. (a) We need the p.d.f. of $X = 54R_1 + 110R_2$ where R_1 has the uniform distribution on the interval $[-10, 20]$ and R_2 has the uniform distribution on the interval $[-4.5, 10]$. We can rewrite X as $X_1 + X_2$ where $X_1 = 54R_1$ has the uniform distribution on the interval $[-540, 1080]$ and $X_2 = 110R_2$ has the uniform distribution on the interval $[-495, 1100]$. Let f_i be the p.d.f. of X_i for $i = 1, 2$, and use the same technique as in Example 3.9.5. First, compute

$$f_1(z)f_2(x - z) = \begin{cases} 3.87 \times 10^{-7} & \text{for } -540 \le z \le 1080 \text{ and } -495 \le x - z \le 1100, \\ 0 & \text{otherwise.} \end{cases}$$

    We need to integrate this over z for each fixed x. The set of x for which the function above is ever positive is the interval $[-1035, 2180]$. For $-1035 \le x < 560$, we must integrate z from -540 to $x + 495$. For $560 \le x < 585$, we must integrate z from $x - 1100$ to $x + 495$. For $585 \le x \le 2180$, we must integrate z from $x - 1100$ to 1080. The resulting integral is

$$g(x) = \begin{cases} 3.87 \times 10^{-7}x + 4.01 \times 10^{-4} & \text{for } -1035 \le x < 560, \\ 6.17 \times 10^{-4} & \text{for } 560 \le x < 585, \\ 8.44 \times 10^{-4} - 3.87 \times 10^{-7}x & \text{for } 585 \le x \le 2180. \end{cases}$$

    (b) We need the negative of the 0.03 quantile. For $-1035 \le x \le 560$, the c.d.f. of X is

$$F(x) = \frac{3.87 \times 10^{-7}(x^2 - 1035^2)}{2} + 4.01 \times 10^{-4}(x + 1035).$$

    This function is a second degree polynomial in x. To be sure that the 0.03 quantile is between -1035 and 560, we compute $F(-1035) = 0$ and $F(560) = 0.493$, which assures us that the 0.03 quantile is in this range. Setting $F(x) = 0.03$ and solving for x using the quadratic formula yields $x = -642.4$, so VaR is 642.4.

13. Use Taylor's theorem with remainder to write

$$g(X) = g(\mu) + (X - \mu)g'(\mu) + \frac{(X - \mu)^2}{2}g''(Y), \tag{S.4.1}$$

where $\mu = E(X)$ and Y is between X and μ. Take the mean of both sides of (S.4.1). We get

$$E[g(X)] = g(\mu) + 0 + E\left(\frac{(X - \mu)^2}{2}g''(Y)\right).$$

The random variable whose mean is on the far right is nonnegative, hence the mean is nonnegative and $E[g(X)] \ge g(\mu)$.

4.3 Variance

Commentary

Be sure to stress the fact that Theorem 4.3.5 on the variance of a sum of random variables has the condition that the random variables are independent.

Solutions to Exercises

1. We found in Exercise 1 of Sec. 4.1 that $E(X) = (0 + 1)/2 = 1/2$. We can find

$$E(X^2) = \int_0^1 x^2 dx = \frac{1}{3}.$$

So $\mathrm{Var}(X) = 1/3 - (1/2)^2 = 1/12$.

2. The p.f. of X and the value of $E(X)$ were determined in Exercise 4 of Sec. 4.1. From the p.f. given there we also find that

$$E(X^2) = 2^2\left(\frac{1}{8}\right) + 3^2\left(\frac{5}{8}\right) + 4^2\left(\frac{1}{8}\right) + 9^2\left(\frac{1}{8}\right) = \frac{73}{4}.$$

Therefore, $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{73}{4} - \left(\frac{15}{4}\right)^2 = \frac{67}{16}$.

3. The p.d.f. of this distribution is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $E(X) = \frac{b+a}{2}$ and

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{1}{3}(b^2 + ab + a^2).$$

It follows that $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{12}(b-a)^2$.

4. $E[X(X-1)] = E(X^2 - X) = E(X^2) - \mu = \mathrm{Var}(X) + [E(X)]^2 - \mu = \sigma^2 + \mu^2 - \mu$.

5. $E[(X-c)^2] = E(X^2) - 2cE(X) + c^2 = \mathrm{Var}(X) + [E(X)]^2 - 2c\mu + c^2 = \sigma^2 + \mu^2 - 2c\mu + c^2 = \sigma^2 + (\mu - c)^2$.

6. Since $E(X) = E(Y), E(X - Y) = 0$. Therefore,

$$E[(X-Y)^2] = \mathrm{Var}(X - Y) = \mathrm{Var}[X + (-Y)].$$

Since X and $-Y$ are independent, it follows that

$$E[(X-Y)^2] = \mathrm{Var}(X) + \mathrm{Var}(-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

7. (a) Since X and Y are independent, $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 3 + 3 = 6$.
   (b) $\mathrm{Var}(2X - 3Y + 1) = 2^2\,\mathrm{Var}(X) + 3^2\,\mathrm{Var}(Y) = 4(3) + 9(3) = 39$.

8. Consider a p.d.f. of the form

$$f(x) = \begin{cases} c\dfrac{1}{x^3} & \text{for } x \geq 1, \\ 0 & \text{for } x < 1. \end{cases}$$

Here,

$$E(X) = \int_1^\infty c\frac{1}{x^2}dx < \infty$$

but

$$E(X^2) = \int_1^\infty c\frac{1}{x}dx \quad \text{is not finite.}$$

Therefore, $E(X)$ is finite but $E(X^2)$ is not. Therefore, $\text{Var}(X)$ is not finite.

9. The mean of X is $(n+1)/2$, and the mean of X^2 is $\sum_{k=1}^{n} k^2/n = (n+1)(2n+1)/6$. So,

$$\text{Var}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.$$

10. The example efficient portfolio has $s_1 = 524.7$, $s_2 = 609.7$, and $s_3 = 39250$.

   (a) We know that R_1 has a mean of 6 and a variance of 55, while R_2 has a mean of 4 and a variance of 28. Since we are assuming that R_i has the uniform distribution on the interval $[a_i, b_i]$ for $i = 1, 2$, we know that

$$E(R_i) = \frac{a_i + b_i}{2},$$
$$\text{Var}(R_i) = \frac{(b_i - a_i)^2}{12}.$$

   (See Exercise 1 of this section for the variance of a uniform distribution.) For $i = 1$, we set $(a_1 + b_1)/2 = 6$ and $(b_1 - a_1)^2/12 = 55$. The solution is $a_1 = -6.845$ and $b_1 = 18.845$. For $i = 2$, we set $(a_2 + b_2)/2 = 4$ and $(b_2 - a_2)^2/12 = 28$. The solution is $a_2 = -5.165$ and $b_2 = 13.165$.

   (b) Let $X_1 = s_1R_1$ and $X_2 = s_2R_2$. Then the distribution of X_1 is the uniform distribution on the interval $[-3591.6, 9888.0]$, and X_2 has the uniform distribution on the interval $[-3149.1, 8026.7]$. The value of the return on the portfolio is $Y = X_1 + X_2 + 1413$. We need to find the 0.03 quantile of Y. As in Exercise 12 of Sec. 4.2, the p.d.f. of Y will be linear for the lowest values of y. Those values are $-5327.7 \leq y < 5848.1$. The line is $g(y) = 6.638 \times 10^{-9}y + 3.537 \times 10^{-5}$. In this range, the c.d.f. is

$$G(y) = \frac{6.638 \times 10^{-9}}{2}(y^2 - 5327.7^2) + 3.537 \times 10^{-5}(y + 5327.7).$$

   Since $G(-5327.7) = 0$ and $G(5848.1) = 0.4146$, we know that the 0.03 quantile is in this range. Setting $G(y) = 0.03$, we find $y = -2321.9$. So VaR is 2321.9.

11. The quantile function of X can be found from Example 3.3.8 with $a = 0$ and $b = 1$. It is $F^{-1}(p) = p$. So, the IQR is $0.75 - 0.25 = 0.5$.

12. The c.d.f. is $F(x) = 1 - \exp(-x)$, for $x > 0$ and $F(x) = 0$ for $x \le 0$. The quantile function is $F^{-1}(p) = -\log(1-p)$. So, the 0.25 and 0.75 quantiles are respectively $-\log(0.75) = 0.2877$ and $-\log(0.25) = 1.3863$. The IQR is then $1.3863 - 0.2877 = 1.0986$.

13. From Table 3.1, we find the 0.25 and 0.75 quantiles of the distribution of X to be 1 and 2 respectively. This makes the IQR equal to $2 - 1 = 1$.

14. The result will follow from the following general result: If x is the p quantile of X and $a > 0$, then ax is the p quantile of $Y = aX$. To prove this, let F be the c.d.f. of X. Note that x is the greatest lower bound on the set $C_p = \{z : F(z) \ge p\}$. Let G be the c.d.f. of Y, then $G(z) = F(z/a)$ because $Y \le z$ if and only if $aX \le z$ if and only if $X \le z/a$. The p quantile of Y is the greatest lower bound on the set

$$D_p = \{y : G(y) \ge p\} = \{y : F(y/a) \ge p\} = \{az : F(z) \ge p\} = aC_p,$$

where the third equality follows from the fact that $F(y/a) \ge p$ if and only if $y = za$ where $F(z) \ge p$. The greatest lower bound on aC_p is a times the greatest lower bound on C_p because $a > 0$.

4.4 Moments

Commentary

The moment generating function (m.g.f.) is a challenging topic that is introduced in this section. The m.g.f. is used later in the text to outline a proof of the central limit theorem (Sec. 6.3). It is also used in a few places to show that certain sums of random variables have particular distributions (e.g., Poisson, Bernoulli, exponential). If students are not going to study the proofs of these results, one could skip the material on moment generating functions.

Solutions to Exercises

1. Since the uniform p.d.f. is symmetric with respect to its mean $\mu = (a+b)/2$, it follows that $E[(X-\mu)^5] = 0$.

2. The mean of X is $(b+a)/2$, so the $2k$th central moment of X is the mean of $(X - [b+a]/2)^{2k}$. Note that $Y = X - [b+a]/2$ has the uniform distribution on the interval $[-(b-a)/2, (b-a)/2]$. Also, $Z = 2Y/(b-a)$ has the uniform distribution on the interval $[-1, 1]$. So $E(Y^{2k}) = [(b-a)/2]^{2k} E(Z^{2k})$.

$$E(Z^{2k}) = \int_{-1}^{1} \frac{z^{2k}}{2} dz = \frac{1}{2k+1}.$$

So, the $2k$th central moment of X is $[(b-a)/2]^{2k}/(2k+1)$.

3. $E[(X-\mu)^3] = E[(X-1)^3] = E(X^3 - 3X^2 + 3X - 1) = 5 - 3(2) + 3(1) - 1 = 1$.

4. Since $\text{Var}(X) \ge 0$ and $\text{Var}(X) = E(X^2) - [E(X)]^2$, it follows that $E(X^2) \ge [E(X)]^2$. The second part of the exercise follows from Theorem 4.3.3.

5. Let $Y = (X-\mu)^2$. Then by Exercise 4,

$$E(Y^2) = E[(X-\mu)^4] \ge [E(Y)]^2 = [\text{Var}(X)]^2 = \sigma^4.$$

6. Since

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for} \quad a < x < b, \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\psi(t) = \int_a^b \exp(tx) \frac{1}{b-a} dx.$$

Therefore, for $t \neq 0$,

$$\psi(t) = \frac{\exp(tb) - \exp(ta)}{t\ (b-a)}.$$

As always, $\psi(0) = 1$.

7. $\psi'(t) = \dfrac{1}{4}(3\exp(t) - \exp(-t))$ and $\psi''(t) = \dfrac{1}{4}(3\exp(t) + \exp(-t))$. Therefore, $\mu = \psi'(0) = 1/2$ and $\sigma^2 = \psi''(0) - \mu^2 = 1 - \left(\dfrac{1}{2}\right)^2 = \dfrac{3}{4}$.

8. $\psi'(t) = (2t+3)\exp(t^2 + 3t)$ and $\psi''(t) = (2t+3)^2\exp(t^2 + 3t) + 2\exp(t^2 + 3t)$. Therefore, $\mu = \psi'(0) = 3$ and $\sigma^2 = \psi''(0) - \mu^2 = 11 - (3)^2 = 2$.

9. $\psi_2'(t) = c\psi_1'(t)\exp(c[\psi_1(t) - 1])$ and $\psi''{}_2(t) = \{[c\psi_1'(t)]^2 + c\psi_1''(t)\}\exp(c[\psi_1(t) - 1])$. We know that

$$\psi_1(0) = 1, \psi_1'(0) = \mu, \text{ and } \psi_1''(0) = \sigma^2 + \mu^2.$$

Therefore, $E(Y) = \psi_2'(0) = c\mu$ and

$$\text{Var}(Y) = \psi_2''(0) - [E(Y)]^2 = \{(c\mu)^2 + c(\sigma^2 + \mu^2)\} - (c\mu)^2 = c(\sigma^2 + \mu^2).$$

10. The m.g.f. of Z is

$$\begin{aligned} \psi_1(t) &= E(\exp(tZ)) = E[\exp(t(2X - 3Y + 4))] \\ &= \exp(4t)E(\exp(2tX)\exp(-3ty)) \\ &= \exp(4t)E(\exp(2tX))E(\exp(-3tY)) \qquad \text{(since } X \text{ and } Y \text{ are independent)} \\ &= \exp(4t)\psi(2t)\psi(-3t) \\ &= \exp(4t)\exp(4t^2 + 6t)\exp(9t^2 - 9t) \\ &= \exp(13t^2 + t). \end{aligned}$$

11. If X can take only a finite number of values $x_1, \ldots, x_k$ with probabilities $p_1, \ldots, p_k$, respectively, then the m.g.f. of X will be

$$\psi(t) = p_1\exp(tx_1) + p_2\exp(tx_2) + \cdots + p_k\exp(tx_k).$$

By matching this expression for $\psi(t)$ with the expression given in the exercise, it can be seen that X can take only the three values 1, 4, and 8, and that $f(1) = 1/5$, $f(4) = 2/5$, and $f(8) = 2/5$.

12. We shall first rewrite $\psi(t)$ as follows:

$$\psi(t) = \frac{4}{6}\exp(0) + \frac{1}{6}\exp(t) + \frac{1}{6}\exp(-t).$$

By reasoning as in Exercise 11, it can now be seen that X can take only the three value 0, 1, and -1, and that $f(0) = 4/6$, $f(1) = 1/6$, and $f(-1) = 1/6$.

13. The m.g.f. of a Cauchy random variable would be

$$\psi(t) = \int_{-\infty}^{\infty} \frac{\exp(tx)}{\pi(1+x^2)} dx. \tag{S.4.2}$$

If $t > 0$, $\lim_{x\to\infty} \exp(tx)/(1+x^2) = \infty$, so the integral in Eq. (S.4.2) is infinite. Similarly, if $t < 0$, $\lim_{x\to-\infty} \exp(tx)/(1+x^2) = \infty$, so the integral is still infinite. Only for $t = 0$ is the integral finite, and that value is $\psi(0) = 1$ as it is for every random variable.

14. The m.g.f. is

$$\psi(t) = \int_{1}^{\infty} \frac{\exp(tx)}{x^2} dx.$$

If $t \leq 0$, $\exp(tx)$ is bounded, so the integral is finite. If $t > 0$, then $\lim_{x\to\infty} \exp(tx)/x^2 = \infty$, and the integral is infinite.

15. Let X have a discrete distribution with p.f. $f(x)$. Assume that $E(|X|^a) < \infty$ for some $a > 0$. Let $0 < b < a$. Then

$$\begin{aligned} E(|X|^b) &= \sum_x |x|^b f(x) = \sum_{|x|\leq 1} |x|^b f(x) + \sum_{|x|>1} |x|^b f(x) \\ &\leq 1 + \sum_{|x|>1} |x|^a f(x) \leq 1 + E(|X|^a) < \infty, \end{aligned}$$

where the first inequality follows from the fact that $0 \leq |x|^b \leq 1$ for all $|x| \leq 1$ and $|x|^b < |x|^a$ for all $|x| > 1$. The next-to-last inequality follows from the fact that the final sum is only part of the sum that makes up $E(|X|^a)$.

16. Let $Z = n - X$. It is easy to see that Z has the same distribution as Y since, if X is the number of successes in n independent Bernoulli trials with probability of success p, then Z is the number of failures and the probability of failure is $1 - p$. It is known from Theorem 4.3.5 that $\text{Var}(Z) = \text{Var}(X)$, which also equals $\text{Var}(Y)$. Also $E(Z) = n - E(X)$, so $Z - E(Z) = n - X - n + E(X) = E(X) - X$. Hence the third central moment of Z is the negative of the third central moment of X and the skewnesses are negatives of each other.

17. We already computed the mean $\mu = 1$ and variance $\sigma^2 = 1$ in Example 4.4.3. Using the m.g.f., the third moment is computed from the third derivative:

$$\psi'''(t) = \frac{6}{(1-t)^4}.$$

The third moment is 6. The third central moment is

$$E([X-1]^3) = E(X^3) - 3E(X^2) + 3E(X) - 1 = 6 - 6 + 3 - 1 = 2.$$

The skewness is then $2/1 = 2$.

4.5 The Mean and the Median

Solutions to Exercises

1. The 1/2 quantile defined in Definition 3.3.2 applies to a continuous random variable whose c.d.f. is one-to-one. The 1/2 quantile is then $x_0 = F^{-1}(1/2)$. That is, $F(x_0) = 1/2$. In order for a number m to be a median as define in this section, it must be that $\Pr(X \le m) \ge 1/2$ and $\Pr(X \ge m) \ge 1/2$. If X has a continuous distribution, then $\Pr(X \le m) = F(m)$ and $\Pr(X \ge m) = 1 - F(m)$. Since $F(x_0) = 1/2$, $m = x_0$ is a median.

2. In this example, $\sum_{x=1}^{6} f(x) = 21c$. Therefore, $c = 1/21$. Since $\Pr(X \le 5) = 15/21$ and $\Pr(X \ge 5) = 11/21$, it follows that 5 is a median and it can be verified that it is the unique median.

3. A median m must satisfy the equation

$$\int_0^m \exp(-x)dx = \frac{1}{2}.$$

Therefore, $1 - \exp(-m) = 1/2$. It follows that $m = \log\ 2$ is the unique median of this distribution.

4. Let X denote the number of children per family. Then

$$\Pr(X \le 2) = \frac{21 + 40 + 42}{153} \ge \frac{1}{2}$$

and

$$\Pr(X \ge 2) = \frac{42 + 27 + 23}{153} \ge \frac{1}{2}.$$

Therefore, 2 is the unique median. Since all families with 4 or more children are in the upper half of the distribution no matter how many children they have (so long as it is at least 4), it doesn't matter how they are distributed among the values 4, 5, ... Next, let $Y = \min\{X, 4\}$, that is Y is the number of children per family if we assume that all families with more than 4 children have exactly 4. We can compute the mean of Y as

$$E(Y) = \frac{1}{153}(0 \times 21 + 1 \times 40 + 2 \times 42 + 3 \times 27 + 4 \times 23) = 1.941.$$

5. The p.d.f. of X will be $h(x) = [f(x) + g(x)]/2$ for $-\infty < x < \infty$. Therefore,

$$E(X) = \frac{1}{2}\int_{-\infty}^{\infty} x[f(x) + g(x)]dx = \frac{1}{2}(\mu_f + \mu_g).$$

Since $\int_{-\infty}^{1} h(X)dx = \int_0^1 \frac{1}{2}f(x)dx = \frac{1}{2}$ and $\int_2^{\infty} h(x)dx = \int_2^4 \frac{1}{2}g(x)dx = \frac{1}{2}$, it follows that every value of m in the interval $1 \le m \le 2$ will be a median.

6. (a) The required value is the mean $E(X)$, and

$$E(X) = \int_0^1 x \cdot 2x\, dx = \frac{2}{3}.$$

(b) The required value is the median m, where

$$\int_0^m 2x\,dx = \frac{1}{2}.$$

Therefore, $m = 1/\sqrt{2}$.

7. (a) The required value is $E(X)$, and

$$E(X) = \int_0^1 x\left(x + \frac{1}{2}\right)dx = \frac{7}{12}.$$

(b) The required value is the median m, where

$$\int_0^m \left(x + \frac{1}{2}\right)dx = \frac{1}{2}.$$

Therefore, $m = (\sqrt{5} - 1)/2$.

8. $E[(X-d)^4] = E(X^4) - 4E(X^3)d + 6E(X^2)d^2 - 4E(X)d^3 + d^4$. Since the distribution of X is symmetric with respect to 0,

$$E(X) = E(X^3) = 0.$$

Therefore,

$$E[(X-d)^4] = E(X^4) + 6E(X^2)d^2 + d^4.$$

For any given nonnegative values of $E(X^4)$ and $E(X^2)$, this is a polynomial of fourth degree in d and it is a minimum when $d = 0$.

9. (a) The required point is the mean $E(X)$, and

$$E(X) = (0.2)(-3) + (0.1)(-1) + (0.1)(0) + (0.4)(1) + (0.2)(2) = 0.1.$$

(b) The required point is the median m. Since $\Pr(X \le 1) = 0.8$ and $\Pr(X \ge 1) = 0.6$, the point 1 is the unique median.

10. Let $x_1 < \cdots < x_n$ denote the locations of the n houses and let d denote the location of the store. We must choose d to minimize $\sum_{i=1}^{n} |x_i - d|$ or equivalently to minimize $\sum_{i=1}^{n} |x_i - d|/n$. This sum can be interpreted as the M.A.E. of d for a discrete distribution in which each of the n points $x_1, \ldots, x_n$ has probability $1/n$. Therefore, d should be chosen equal to a median of this distribution. If n is odd, then the middle value among $x_1, \ldots, x_n$ is the unique median. If n is even, then any point between the two middle values among $x_1, \ldots, x_n$ (including the two middle values themselves) will be a median.

11. The M.S.E. of any prediction is a minimum when the prediction is equal to the mean of the variable being predicted, and the minimum value of the M.S.E. is then the variance of the variable. It was shown in the derivation of Eq. (4.3.3) that the variance of the binomial distribution with parameters n and p is $np(1-p)$. Therefore, the minimum M.S.E. that can be attained when predicting X is $\text{Var}(X) = 7(1/4)(3/4) = 21/16$ and the minimum M.S.E. that can be attained when predicting Y is $\text{Var}(Y) = 5(1/2)(1/2) = 5/4 = 20/16$. Thus, Y can be predicted with the smaller M.S.E.

12. (a) The required value is the mean $E(X)$. The random variable X will have the binomial distribution with parameters $n = 15$ and $p = 0.3$. Therefore, $E(X) = np = 4.5$.

(b) The required value is the median of the binomial distribution with parameters $n = 15$ and $p = 0.3$. From the table of this distribution given in the back of the book, it is found that 4 is the unique median.

13. To say that the distribution of X is symmetric around m, means that X and $2m - X$ have the same distribution. That is, $\Pr(X \leq x) = \Pr(2m - X \leq x)$ for all x. This can be rewritten as $\Pr(X \leq x) = \Pr(X \geq 2m - x)$. With $x = m$, we see that $\Pr(X \leq m) = \Pr(X \geq m)$. If $\Pr(X \leq m) < 1/2$, then $\Pr(X \leq m) + \Pr(X > m) < 1$, which is impossible. Hence $\Pr(X \leq m) \geq 1/2$ and $\Pr(X \geq m) \geq 1/2$, and m is a median.

14. The Cauchy distribution is symmetric around 0, so 0 is a median by Exercise 13. Since the p.d.f. of the Cauchy distribution is strictly positive everywhere, the c.d.f. will be one-to-one and 0 is the unique median.

15. (a) Since a is assumed to be a median, $F(a) = \Pr(X \leq a) \geq 1/2$. Since $b > a$ is assumed to be a median $\Pr(X \geq b) \geq 1/2$. If $\Pr(X \leq a) > 1/2$, then $\Pr(X \leq a) + \Pr(X \geq b) > 1$. But $\{X \leq a\}$ and $\{X \geq b\}$ are disjoint events, so the sum of their probabilities can't be greater than 1. This means that $F(a) > 1/2$ is impossible, so $F(a) = 1/2$.

(b) The c.d.f. F is nondecreasing, so $A = \{x : F(x) = 1/2\}$ is an interval. Since F is continuous from the right, the lower endpoint c of the interval A must also be in A. For every x, $\Pr(X \leq x) + \Pr(X \geq x) \geq 1$. For every $x \in A$, $\Pr(X \leq x) = 1/2$, hence it must be that $\Pr(X \geq x) \geq 1/2$ and x is a median. Let d be the upper endpoint of the interval A. We need to show that d is also a median. Since F is not necessarily continuous from the left, $F(d) > 1/2$ is possible. If $F(d) = 1/2$, then $d \in A$ and d is a median by the argument just given. If $F(d) > 1/2$, then $\Pr(X = d) = F(d) - 1/2$. This makes

$$\Pr(X \geq d) = \Pr(X > d) + \Pr(X = d) = 1 - F(d) + F(d) - 1/2 = 1/2.$$

Hence d is also a median

(c) If X has a discrete distribution, then clearly F must be discontinuous at d otherwise $F(x) = 1/2$ even for some $x > d$ and d would not be the right endpoint of A.

16. We know that $1 = \Pr(X < m) + \Pr(X = m) + \Pr(X > m)$. Since $\Pr(X < m) = \Pr(X > m)$, both $\Pr(X < m) \leq 1/2$ and $\Pr(X > m) \leq 1/2$, otherwise their sum would be more than 1. Since $\Pr(X < m) \leq 1/2$, $\Pr(X \geq m) = 1 - \Pr(X < m) \geq 1/2$. Similarly, $\Pr(X \leq m) = 1 - \Pr(X > m) \geq 1/2$. Hence m is a median.

17. As in the previous problem, $1 = \Pr(X < m) + \Pr(X = m) + \Pr(X > m)$. Since $\Pr(X < m) < 1/2$ and $\Pr(X > m) < 1/2$, we have $\Pr(X \geq m) = 1 - \Pr(X < m) > 1/2$ and $\Pr(X \leq m) = 1 - \Pr(X > m) > 1/2$. Hence m is a median. Let $k > m$. Then $\Pr(X \geq k) \leq \Pr(X > m) < 1/2$, and k is not a median. Similarly, if $k < m$, then $\Pr(X \leq k) \leq \Pr(X < m) < 1/2$, and k is not a median. So, m is the unique median.

18. Let m be the p quantile of X, and let r be strictly increasing. Let $Y = r(X)$ and let $G(y)$ be the c.d.f. of Y while $F(x)$ is the c.d.f. of X. Since $Y \leq y$ if and only if $r(X) \leq y$ if and only if $X \leq r^{-1}(y)$, we have $G(y) = F(r^{-1}(y))$. The p quantile of Y is the smallest element of the set

$$C_p = \{y : G(y) \geq p\} = \{y : F(r^{-1}(y)) \geq p\} = \{r(x) : F(x) \geq p\}.$$

Also, m is the smallest x such that $F(x) \geq p$. Because r is strictly increasing, the smallest $r(x)$ such that $F(x) \geq p$ is $r(m)$. Hence, $r(m)$ is the smallest number in C_p.

4.6 Covariance and Correlation

Solutions to Exercises

1. The location of the circle makes no difference since it only affects the means of X and Y. So, we shall assume that the circle is centered at (0,0). As in Example 4.6.5, $\text{Cov}(X, Y) = 0$. It follows that $\rho(X, Y) = 0$ also.

2. We shall follow the hint given in this exercise. The relation $[(X - \mu_X) + (Y - \mu_Y)]^2 \geq 0$ implies that

$$(X - \mu_X)(Y - \mu_Y) \leq \frac{1}{2}[(X - \mu_X)^2 + (Y - \mu_Y)^2].$$

Similarly, the relation $[(X - \mu_X) - (Y - \mu_Y)]^2 \geq 0$ implies that

$$-(X - \mu_X)(Y - \mu_Y) \leq \frac{1}{2}[(X - \mu_X)^2 + (Y - \mu_Y)^2].$$

Hence, it follows that

$$|(X - \mu_X)(Y - \mu_Y)| \leq \frac{1}{2}[(X - \mu_X)^2 + (Y - \mu_Y)^2].$$

By taking expectations on both sides of this relation, we find that

$$E[|(X - \mu_X)(Y - \mu_Y)|] \leq \frac{1}{2}[\text{Var}(X) + \text{Var}(Y)] < \infty.$$

Since the expectation on the left side of this relation is finite, it follows that

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

exists and is finite.

3. Since the p.d.f. of X is symmetric with respect to 0, it follows that $E(X) = 0$ and that $E(X^k) = 0$ for every odd positive integer k. Therefore, $E(XY) = E(X^7) = 0$. Since $E(XY) = 0$ and $E(X)E(Y) = 0$, it follows that $\text{Cov}(X, Y) = 0$ and $\rho(X, Y) = 0$.

4. It follows from the assumption that $0 < E(X^4) < \infty$, that $0 < \sigma_X^2 < \infty$ and $0 < \sigma_Y^2 < \infty$. Hence, $\rho(X, Y)$ is well defined. Since the distribution of X is symmetric with respect to $0, E(X) = 0$ and $E(X^3) = 0$. Therefore, $E(XY) = E(X^3) = 0$. It now follows that $\text{Cov}(X, Y) = 0$ and $\rho(X, Y) = 0$.

5. We have $E(aX + b) = a\mu_X + b$ and $E(cY + d) = c\mu_Y + d$. Therefore,

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= E[(aX + b - a\mu_X - b)(cY + d - c\mu_Y - d)] \\ &= E[ac(X - \mu_X)(Y - \mu_Y)] = ac\,\text{Cov}(X, Y). \end{aligned}$$

6. By Exercise 5, $\text{Cov}(U, V) = ac\,\text{Cov}(X, Y)$. Also, $\text{Var}(U) = a^2\sigma_X^2$ and $\text{Var}(V) = c^2\sigma_Y^2$. Hence

$$p(U, V) = \frac{ac\,\text{Cov}(X, Y)}{|a|\sigma_X \cdot |c|\sigma_Y} = \begin{cases} \rho(X, Y) & \text{if} \quad ac > 0, \\ -\rho(X, Y) & \text{if} \quad ac < 0. \end{cases}$$

7. We have $E(aX+bY+c) = a\mu_X + b\mu_Y + c$. Therefore,

$$\begin{aligned}
\mathrm{Cov}(aX+bY+c, Z) &= E[(aX+bY+c-a\mu_X-b\mu_Y-c)(Z-\mu_Z)] \\
&= E\{[a(X-\mu_X)+b(Y-\mu_Y)](Z-\mu_Z)\} \\
&= aE[(X-\mu_X)(Z-\mu_Z)] + bE[(Y-\mu_Y)(Z-\mu_Z)] \\
&= a\,\mathrm{Cov}(X,Z) + b\,\mathrm{Cov}(Y,Z).
\end{aligned}$$

8. We have

$$\begin{aligned}
\mathrm{Cov}\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right) &= E\left[\sum_{i=1}^m a_i(X_i-\mu_{X_i})\sum_{j=1}^n b_j(Y_j-\mu_{Y_j})\right] \\
&= E\left[\sum_{i=1}^m \sum_{j=1}^n a_i b_j (X_i-\mu_{X_i})(Y_j-\mu_{Y_j})\right] \\
&= \sum_{i=1}^m \sum_{j=1}^n a_i b_j E\left[(X_i-\mu_{X_i})(Y_j-\mu_{Y_j})\right] \\
&= \sum_{i=1}^m \sum_{j=1}^n a_i b_j \,\mathrm{Cov}(X_i, Y_j).
\end{aligned}$$

9. Let $U = X+Y$ and $V = X-Y$. Then

$$E(UV) = E[(X+Y)(X-Y)] = E(X^2 - Y^2) = E(X^2) - E(Y^2).$$

Also,

$$E(U)E(V) = E(X+Y)E(X-Y) = (\mu_X+\mu_Y)(\mu_X-\mu_Y) = \mu_X^2 - \mu_Y^2.$$

Therefore,

$$\begin{aligned}
\mathrm{Cov}(U,V) &= E(UV) - E(U)E(V) = [E(X^2)-\mu_X^2] - [E(Y^2)-\mu_Y^2] \\
&= \mathrm{Var}(X) - \mathrm{Var}(Y) = 0.
\end{aligned}$$

It follows that $\rho(U,V) = 0$.

10. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$ and $\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X,Y)$. Since $\mathrm{Cov}(X,Y) < 0$, it follows that

$$\mathrm{Var}(X+Y) < \mathrm{Var}(X-Y).$$

11. For the given values,

$$\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - [E(X)]^2 = 10 - 9 = 1, \\
\mathrm{Var}(Y) &= E(Y^2) - [E(Y)]^2 = 29 - 4 = 25, \\
\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) = 0 - 6 = -6.
\end{aligned}$$

Therefore,

$$\rho(X,Y) = \frac{-6}{(1)(5)} = -\frac{6}{5}, \text{ which is impossible.}$$

12.

$$\begin{aligned}
E(X) &= \int_0^1 \int_0^2 x \cdot \frac{1}{3}(x+y)dy\,dx = \frac{5}{9}, \\
E(Y) &= \int_0^1 \int_0^2 y \cdot \frac{1}{3}(x+y)dy\,dx = \frac{11}{9}, \\
E(X^2) &= \int_0^1 \int_0^2 x^2 \cdot \frac{1}{3}(x+y)dy\,dx = \frac{7}{18}, \\
E(Y^2) &= \int_0^1 \int_0^2 y^2 \cdot \frac{1}{3}(x+y)dy\,dx = \frac{16}{9}, \\
E(XY) &= \int_0^1 \int_0^2 xy \cdot \frac{1}{3}(x+y)dy\,dx = \frac{2}{3}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{Var}(X) &= \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{13}{162}, \\
\text{Var}(Y) &= \frac{16}{9} - \left(\frac{11}{9}\right)^2 = \frac{23}{81}, \\
\text{Cov}(XY) &= \frac{2}{3} - \left(\frac{5}{9}\right)\left(\frac{11}{9}\right) = -\frac{1}{81}.
\end{aligned}$$

It now follows that

$$\begin{aligned}
\text{Var}(2X - 3Y + 8) &= 4\text{Var}(X) + 9\,\text{Var}(Y) - (2)(2)(3)\,\text{Cov}(X,Y) \\
&= \frac{245}{81}.
\end{aligned}$$

13. $\text{Cov}(X,Y) = \rho(X,Y)\sigma_X\sigma_Y = -\frac{1}{6}(3)(2) = -1.$

(a) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y) = 11.$

(b) $\text{Var}(X - 3Y + 4) = \text{Var}(X) + 9\,\text{Var}(Y) - (2)(3)\,\text{Cov}(X,Y) = 51.$

14. (a) $\text{Var}(X+Y+Z) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z) + 2\,\text{Cov}(X,Y) + 2\,\text{Cov}(X,Z) + 2\,\text{Cov}(Y,Z) = 17.$

(b) $\text{Var}(3X - Y - 2Z + 1) = 9\,\text{Var}(X) + \text{Var}(Y) + 4\,\text{Var}(Z) - 6\,\text{Cov}(X,Y) - 12\,\text{Cov}(X,Z) + 4\,\text{Cov}(Y,Z) = 59.$

15. Since each variance is equal to 1 and each covariance is equal to 1/4,

$$\begin{aligned}
\text{Var}(X_1 + \cdots + X_n) &= \sum_i \text{Var}(X_i) + 2\sum_{i<j}\sum \text{Cov}(X_i, X_j) \\
&= n(1) + 2.\frac{n(n-1)}{2}\left(\frac{1}{4}\right) = n + \frac{n(n-1)}{4}.
\end{aligned}$$

16. We need the cost to be 6000 dollars, so that $50s_1 + 30s_2 = 6000$. We also need the variance to be 0. The variance of $s_1R_1 + s_2R_2$ is

$$s_1^2\,\text{Var}(R_1) + s_2^2\,\text{Var}(R_2) + 2s_1s_2\,\text{Cov}(R_1, R_2).$$

The variances of R_1 and R_2 are $\text{Var}(R_1) = 75$ and $\text{Var}(R_2) = 17.52$. Since the correlation between R_1 and R_2 is -1, their covariance is $-1(75)^{1/2}(17.52)^{1/2} = -36.25$. To make the variance 0, we need $75s_1^2 + 17.52s_2^2 - 36.25s_1s_2 = 0$. This equation can be rewritten $(75^{1/2}s_1 - 17.52^{1/2}s_2)^2 = 0$. So, we need to solve the two equations

$$75^{1/2}s_1 - 17.52^{1/2}s_2 = 0, \quad \text{and} \quad 50s_1 + 30s_2 = 6000.$$

The solution is $s_1 = 53.54$ and $s_2 = 110.77$. The reason that such a portfolio is unrealistic is that it has positive mean (1126.2) but zero variance, that is one can earn money with no risk. Such a "money pump" would surely dry up the moment anyone recognized it.

17. Let $\mu_X = E(X)$ and $\mu_Y = E(Y)$. Apply Theorem 4.6.2 with $U = X - \mu_X$ and $V = Y - \mu_Y$. Then (4.6.4) becomes

$$\text{Cov}(X,Y)^2 \le \text{Var}(X)\,\text{Var}(Y). \tag{S.4.3}$$

Now $|\rho(X,Y)| = 1$ is equivalent to equality in (S.4.3). According to Theorem 4.6.2, we get equality in (4.6.4) and (S.4.3) if and only if there exist constants a and b such that $aU + bV = 0$, that is $a(X - \mu_X) + b(Y - \mu_Y) = 0$, with probability 1. So $|\rho(X,Y)| = 1$ implies $aX + bY = a\mu_X = b\mu_Y$ with probability 1.

18. The means of X and Y are the same since $f(x,y) = f(y,x)$ for all x and y. The mean of X (and the mean of Y) is

$$E(X) = \int_0^1 \int_0^1 x(x+y)dxdy = \int_0^1 \left(\frac{1}{3} + \frac{y}{2}\right) dy = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

Also,

$$E(XY) = \int_0^1 \int_0^1 xy(x+y)dxdy = \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2}\right) dy = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

So,

$$\text{Cov}(X,Y) = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = -0.00695.$$

4.7 Conditional Expectation

Solutions to Exercises

1. The M.S.E. after observing $X = 18$ is $\text{Var}(P|18) = 19 \times (41 - 18)/[42^2 \times 43] = 0.00576$. This is about seven percent of the marginal M.S.E.

2. If X denotes the score of the selected student, then

$$E(X) = E[E(X \mid \text{School})] = (0.2)(80) + (0.3)(76) + (0.5)(84) = 80.8.$$

3. Since $E(X \mid Y) = c$, then $E(X) = E[E(X \mid Y)] = c$ and $E(XY) = E[E(XY \mid Y)] = E(YE(X \mid Y)] = E(cY) = cE(Y)$.

   Therefore,

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = cE(Y) - cE(Y) = 0.$$

4. Since X is symmetric with respect to 0, $E(X^k) = 0$ for every odd integer k. Therefore,

$$E(X^{2m}Y) = E[E(X^{2m}Y \mid X)] = E[X^{2m}E(Y \mid X)] = E[aX^{2m+1} + bX^{2m}] = bE(X^{2m}).$$

Also,

$$E(Y) = E[E(Y \mid X)] = E(aX + b) = b.$$

It follows that

$$\text{Cov}(X^{2m}, Y) = E(X^{2m}Y) - E(X^{2m})E(Y) = bE(X^{2m}) - bE(X^{2m}) = 0.$$

5. For any given value x_{n-1} of X_{n-1}, $E(X_n \mid x_{n-1})$ will be the midpoint of the interval $(x_{n-1}, 1)$. Therefore, $E(X_n \mid X_{n-1}) = \dfrac{1 + X_{n-1}}{2}$.
It follows that

$$E(X_n) = E[E(X_n \mid X_{n-1})] = \frac{1}{2} + \frac{1}{2}E(X_{n-1}).$$

Similarly, $E(X_{n-1}) = \dfrac{1}{2} + \dfrac{1}{2}E(X_{n-2})$, etc. Since $E(X_1) = \dfrac{1}{2}$, we obtain

$$E(X_n) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n} = 1 - \frac{1}{2^n}.$$

6. The joint p.d.f. of X and Y is

$$f(x, y) = \begin{cases} c & \text{for} \quad x^2 + y^2 < 1, \\ 0 & \quad\quad \text{otherwise.} \end{cases}$$

Therefore, for any given value of y in the interval $-1 < y < 1$, the conditional p.d.f. of X given that $Y = y$ will be of the form

$$g(x \mid y) = \frac{f(x, y)}{f_2(y)} = \begin{cases} \dfrac{c}{f_2(y)} & \text{for } -\sqrt{1 - y^2} < x < \sqrt{1 - y^2}, \\ 0 & \text{otherwise.} \end{cases}$$

For each given value of y, this conditional p.d.f. is a constant over an interval of values of x symmetric with respect to $x = 0$. Therefore, $E(X \mid y) = 0$ for each value of y.

7. The marginal p.d.f. of X is

$$f_1(x) = \int_0^1 (x + y)dy = x + \frac{1}{2} \qquad \text{for} \quad 0 \le x \le 1.$$

Therefore, for $0 \le x \le 1$, the conditional p.d.f. of Y given that $X = x$ is

$$g(y \mid x) = \frac{f(x, y)}{f_1(x)} = \frac{2(x + y)}{2x + 1} \qquad \text{for } 0 \le y \le 1.$$

Hence,

$$\begin{aligned} E(Y \mid x) &= \int_0^1 \frac{2(xy+y^2)}{2x+1} dy = \frac{3x+2}{3(2x+1)}, \\ E(Y^2 \mid x) &= \int_0^1 \frac{2(xy^2+y^3)}{2x+1} dy = \frac{4x+3}{6(2x+1)}, \end{aligned}$$

and

$$\text{Var}(Y \mid x) = \frac{4x+3}{6(2x+1)} - \left[\frac{3x+2}{3(2x+1)}\right]^2 = \frac{1}{36}\left[3 - \frac{1}{(2x+1)^2}\right].$$

8. The prediction is $E\left(Y \mid X = \frac{1}{2}\right) = \frac{7}{12}$ and the M.S.E. is $\text{Var}\left(Y \mid X = \frac{1}{2}\right) = \frac{11}{144}$.

9. The overall M.S.E. is

$$E[\text{Var}(Y \mid X)] = \int_0^1 \frac{1}{36}\left[3 - \frac{1}{(2x+1)^2}\right] f_1(x)dx.$$

It was found in the solution of Exercise 7 that

$$f_1(x) = x + \frac{1}{2} \qquad \text{for } 0 \le x \le 1.$$

Therefore, it can be found that $E[\text{Var}(Y \mid X)] = \frac{1}{12} - \frac{\log\ 3}{144}$.

10. It was found in Exercise 9 that when Y is predicted from X, the overall M.S.E. is $\frac{1}{12} - \frac{\log\ 3}{144}$. Therefore, the total loss would be

$$\frac{1}{12} - \frac{\log\ 3}{144} + c.$$

If Y is predicted without using X, the M.S.E. is $\text{Var}(Y)$. It is found that

$$E(Y) = \int_0^1 \int_0^1 y(x+y)dx\,dy = \frac{7}{12}$$

and

$$E(Y^2) = \int_0^1 \int_0^1 y^2(x+y)dx\,dy = \frac{5}{12}.$$

Hence, $\text{Var}(Y) = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$. The total loss when X is used for predicting Y will be less than $\text{Var}(Y)$ if and only if $c < \frac{\log\ 3 - 1}{144}$.

11. Let $E(Y) = \mu_Y$. Then

$$\begin{aligned}\text{Var}(Y) &= E[(Y-\mu_Y)^2] = E\{[(Y - E(Y \mid X)) + (E(Y \mid X) - \mu_Y)]^2\} \\ &= E\{[Y - E(Y \mid X)]^2\} + 2E\{[Y - E(Y \mid X)][E(Y \mid X) - \mu_Y]\} \\ &\quad + E\{[E(Y \mid X) - \mu_Y]^2\}.\end{aligned}$$

We shall now consider further each of the three expectations in the final sum. First,

$$E\{[Y - E(Y \mid X)]^2\} = E(E\{[Y - E(Y \mid X]^2 \mid X\}) = E[\text{Var}(Y \mid X)].$$

Next,

$$\begin{aligned}E\{[Y - E(Y \mid X)][E(Y \mid X) - \mu_Y]\} &= E(E\{[Y - E(Y \mid X)][E(Y \mid X) - \mu_Y] \mid X\}) \\ &= E([E(Y \mid X) - \mu_Y]E\{Y - E(Y \mid X) \mid X\}) \\ &= E([E(Y \mid X) - \mu_Y] \cdot 0) \\ &= 0.\end{aligned}$$

Finally, since the mean of $E(Y \mid X)$ is $E[E(Y \mid X)] = \mu_Y$, we have

$$E\{[E(Y \mid X) - \mu_Y]^2\} = \text{Var}[E(Y \mid X)].$$

It now follows that

$$\text{Var}(Y) = E[\text{Var}(Y \mid X)] + \text{Var}[E(Y \mid X)].$$

12. Since $E(Y) = E[E(Y \mid X)]$, then

$$E(Y) = aE(X) + b.$$

Also, as found in Example 4.7.7,

$$E(XY) = aE(X^2) + bE(X).$$

By solving these two equations simultaneously for a and b we obtain,

$$a = \frac{E(XY) - E(X)E(Y)}{E(X^2) - [E(X)]^2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

and

$$b = E(Y) - aE(X).$$

13. (a) The prediction is the mean of Y:

$$E(Y) = \int_0^1 \int_0^1 y \cdot \frac{2}{5}(2x + 3y)dx\,dy = \frac{3}{5}.$$

(b) The prediction is the median m of X. The marginal p.d.f. of X is

$$f_1(x) = \int_0^1 \frac{2}{5}(2x+3y)dy = \frac{1}{5}(4x+3) \qquad \text{for } 0 \le x \le 1.$$

We must have

$$\int_0^m \frac{1}{5}(4x+3)dx = \frac{1}{2}.$$

Therefore, $4m^2 + 6m - 5 = 0$ and $m = \dfrac{\sqrt{29}-3}{4}$.

14. First,

$$E(XY) = \int_0^1 \int_0^1 xy \cdot \frac{2}{5}(2x+3y)dx\, dy = \frac{1}{3}.$$

Next, the marginal p.d.f. f_1 of X was found in Exercise 13(b). Therefore,

$$E(X) = \int_0^1 x f_1(x)dx = \frac{17}{30}.$$

Furthermore, it was found in Exercise 13(a) that $E(Y) = 3/5$. It follows that $\text{Cov}(X,Y) = 1/3 - (17/30)(3/5) = 17/51 - 17/50 < 0$. Therefore, X and Y are negatively correlated.

15. (a) For $0 \le x \le 1$ and $0 \le y \le 1$, the conditional p.d.f. of Y given that $X = x$ is

$$g(y \mid x) = \frac{f(x,y)}{f_1(x)} = \frac{2(2x+3y)}{4x+3}.$$

When $X = 0.8$, the prediction of Y is

$$E(Y \mid X = 0.8) = \int_0^1 y g(y \mid x = 0.8)dy = \int_0^1 \frac{y(1.6+3y)}{3.1}dy = \frac{18}{31}.$$

(b) The marginal p.d.f. of Y is

$$f_2(y) = \int_0^1 \frac{2}{5}(2x+3y)dx = \frac{2}{5}(1+3y) \qquad \text{for } 0 \le y \le 1.$$

Therefore, for $0 \le x \le 1$ and $0 \le y \le 1$, the conditional p.d.f. of X given that $Y = y$ is

$$h(x \mid y) = \frac{f(x,y)}{f_2(y)} = \frac{2x+3y}{1+3y}.$$

When $Y = 1/3$, the prediction of X is the median m of the conditional p.d.f. $h(x \mid y = 1/3)$. We must have

$$\int_0^m \frac{2x+1}{2}\, dx = \frac{1}{2}.$$

Hence, $m^2 + m = 1$ and $m = (\sqrt{5}-1)/2$.

16. Rather than repeat the entire proof of Theorem 4.7.3 with the necessary changes, we shall merely point out what changes need to be made. Let $d(X)$ be a conditional median of Y given X. Replace all squared differences by absolute differences. For example $[Y - d(X)]^2$ becomes $|Y - d(X)|$, $[Y - d^*(x)]^2$ becomes $|Y - d^*(x)|$, and so on. When we refer to Sec. 4.5 near the end of the proof, replace each "M.S.E." by "M.A.E." and replace the word "mean" by "median" each time it appears in the last four sentences.

17. Let $Z = r(X, Y)$, and let (X, Y) have joint p.f. $f(x, y)$. Also, let $W = r(x_0, Y)$, for some possible value x_0 of X. We need to show that the conditional p.f. of Z given $X = x_0$ is the same as the conditional p.f. of W given $X = x_0$ for all x_0.

    Let $f_1(x)$ be the marginal p.f. of X. For each possible value (z, x) of (Z, X), define $B_{(z,x)} = \{y : r(x, y) = z\}$. Then, $(Z, X) = (z, x)$ if and only if $X = x$ and $Y \in B_{(z,x)}$. The joint p.f. of (Z, X) is then

$$g(z, x) = \sum_{y \in B_{(z,x)}} f(x, y).$$

    The conditional p.f. of Z given $X = x_0$ is $g_1(z|x_0) = g(z, x_0)/f_1(x_0)$ for all z and all x_0.

    Next, notice that $(W, X) = (w, x)$ if and only if $X = x$ and $w \in B_{(w,x_0)}$. The joint p.f. of (W, X) is then

$$h(w, x) = \sum_{y \in B_{(w,x_0)}} f(x, y).$$

    The conditional p.f. of W given $X = x$ is $h_1(w|x) = h(w, x)/f_1(x)$. Now, for $x = x_0$, we get $h_1(w|x_0) = h(w, x_0)/f_1(x_0)$. But $h(w, x_0) = g(w, x_0)$ for all w and all x_0. Hence $h_1(w|x_0) = g_1(w|x_0)$ for all w and all x_0. This is the desired conclusion.

4.8 Utility

Commentary

It is interesting to have the students in the class determine their own utility functions for any possible gain between, say, 0 dollars and 100 dollars; in the other words, to have each student determine their own function $U(x)$ for $0 \le x \le 100$. One method for determining various points on a person's utility function is as follows:

First, notice that if $U(x)$ is a person's utility function, then the function $V(x) = aU(x) + b$, where a and b are constants with $a > 0$, could also be used as the person's utility function, because for any two gambles X and Y, we will have $E[U(X)] > E[U(Y)]$ if and only if $E[V(X)] = aE[U(X)] + b > E[V(Y)] = aE[U(Y)]+b$. Therefore, the function V reflects exactly the same preferences as U. The effect of being able to transform a person's utility function in this way by choosing any constants $a > 0$ and b is that we can arbitrarily fix the values of the person's utility function at the two points $x = 0$ and $x = 100$, as long as we use values such that $U(0) < U(100)$. For convenience, we shall assume that $U(0) = 0$ and $U(100) = 100$.

Now determine a value x_1 such that the person is indifferent between accepting a gamble from which the gain will be either 100 dollars with probability 1/2 or 0 dollars with probability 1/2 and accepting x_1 dollars as a sure thing. For this value x_1, we must have

$$U(X_1) = \frac{1}{2}U(0) + \frac{1}{2}U(100) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 100 = 50.$$

Hence, $U(x_1) = 50$.

Next, we might determine a value x_2 such that the person is indifferent between accepting a gamble from which the gain will be either x_1 dollars with probability 1/2 or 0 dollars with probability 1/2 and accepting x_2 dollars as a sure thing. For this value x_2, we must have

$$U(x_2) = \frac{1}{2}U(x_1) + \frac{1}{2}U(0) = \frac{1}{2} \cdot 50 + \frac{1}{2} \cdot 0 = 25.$$

Hence, $U(x_2) = 25$.

Similarly, we can determine a value x_3 such that the person is indifferent between accepting a gamble from which the gain will be either x_1 dollars with probability 1/2 or 100 dollars with probability 1/2 and accepting x_3 dollars as a sure thing. For this value x_3, we must have

$$U(x_3) = \frac{1}{2}U(x_1) + \frac{1}{2}U(100) = \frac{1}{2} \cdot 50 + \frac{1}{2} \cdot 100 = 75.$$

Hence, $U(x_3) = 75$.

By continuing in this way, arbitrarily many points on a person's utility function can be determined and the curve U(x) for $0 \le x \le 100$ can then be sketched. The difficulty is in having the person determine the values of x_1, x_2, x_3, etc., honestly and accurately in a hypothetical situation where he will not actually have to gamble. For this reason, it is necessary to check and recheck the values that are determined. For example, since

$$U(x_1) = 50 = \frac{1}{2}U(x_2) + \frac{1}{2}U(x_3),$$

the person should be indifferent between accepting x_1 dollars as a sure thing and accepting a gamble from which the gain will be either x_2 dollars with probability 1/2 or x_3 dollars with probability 1/2. By repeatedly carrying out checks of this type and allowing the person to adjust his answers, a reasonably accurate representation of a person's utility function can usually be obtained.

Solutions to Exercises

1. The utility of not buying the ticket is $U(0) = 0$. If the decision maker buys the ticket, the utility is $U(499)$ if the ticket is a winner and and $U(-1)$ if the ticket is a loser. That is the utility is 499^α with probability 0.001 and it is -1 with probability 0.999. The expected utility is then $0.001 \times 499^\alpha - 0.999$. The decision maker prefers buying the ticket if this expected utility is greater than 0. Setting the expected utility greater than 0 means $499^\alpha > 999$. Taking logarithms of both sides yields $\alpha > 1.11$.

2.

$$E[U(X)] = \frac{1}{2} \cdot 5^2 + \frac{1}{2} \cdot 25^2 = 325,$$
$$E[U(Y)] = \frac{1}{2} \cdot 10^2 + \frac{1}{2} \cdot 20^2 = 250,$$
$$E[U(Z)] = 15^2 = 225.$$

Hence, X is preferred.

3.

$$E[U(X)] = \frac{1}{2}\sqrt{5} + \frac{1}{2}\sqrt{25} = 3.618,$$
$$E[U(Y)] = \frac{1}{2}\sqrt{10} + \frac{1}{2}\sqrt{20} = 3.817,$$
$$E[U(Z)] = \sqrt{15} = 3.873.$$

Hence, Z is preferred.

4. For any gamble X, $E[U(X)] = aE(X) + b$. Therefore, among any set of gambles, the one for which the expected gain is largest will be preferred. We have

$$\begin{aligned} E(X) &= \frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 25 = 15, \\ E(Y) &= \frac{1}{2} \cdot 10 + \frac{1}{2} \cdot 20 = 15, \\ E(Z) &= 15. \end{aligned}$$

Hence, all three gambles are equally preferred.

5. Since the person is indifferent between the gamble and the sure thing,

$$U(50) = \frac{1}{3}U(0) + \frac{2}{3}U(100) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}.$$

6. Since the person is indifferent between X and Y, $E[U(X)] = E[U(Y)]$. Therefore,

$$(0.6)U(-1) + (0.2)U(0) + (0.2)U(2) = (0.9)U(0) + (0.1)U(1).$$

It follows from the given values that $U(-1) = 23/6$.

7. For any given values of a,

$$E[U(X)] = p \log\ a + (1-p)\log(1-a).$$

The maximum of this expected utility can be found by elementary differentiation. We have

$$\frac{\partial E[U(X)]}{\partial a} = \frac{p}{a} - \frac{1-p}{1-a}.$$

When this derivative is set equal to 0, we find that $a = p$. Since

$$\frac{\partial^2 E[U(X)]}{\partial a^2} = -\frac{p}{a^2} - \frac{1-p}{(1-a)^2} < 0,$$

It follows that E[U(X)] is a maximum when $a = p$.

8. For any given value of a,

$$E[U(X)] = pa^{1/2} + (1-p)(1-a)^{1/2}.$$

Therefore,

$$\frac{\partial E[U(X)]}{\partial a} = \frac{p}{2a^{1/2}} - \frac{1-p}{2(1-a)^{1/2}}.$$

When this derivative is set equal to 0, we find that

$$a = \frac{p^2}{p^2 + (1-p)^2}.$$

Since $\dfrac{\partial^2\ E[U(X)]}{\partial a^2} < 0$, it follows that $E[U(X)]$ is a maximum at this value of a.

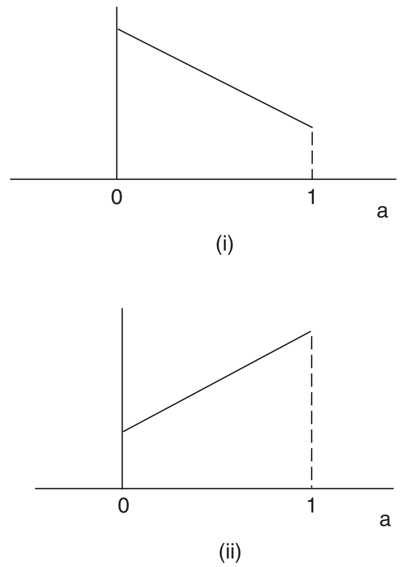

Figure S.4.2: Figure for Exercise 9 of Sec. 4.8.

9. For any given value of a,

$$E[U(X)] = pa + (1-p)(1-a).$$

This is a linear function of a. If $p < 1/2$, it has the form shown in sketch (i) of Fig. S.4.2. Therefore, $E[U(X)]$ is a maximum when $a = 0$. If $p > 1/2$, it has the form shown in sketch (ii) of Fig. S.4.2. Therefore, $E[U(X)]$ is a maximum when $a = 1$. If $p = 1/2$, then $E[U(X)] = 1/2$ for all values of a.

10. The person will prefer X_3 to X_4 if and only if

$$\begin{aligned} E[U(X_3)] &= (0.3)U(0) + (0.3)U(1) + (0.4)U(2) > E[U(X_4)] \\ &= (0.5)U(0) + (0.5)U(2). \end{aligned}$$

Therefore, the person will prefer X_3 to X_4 if and only if

$$(0.2)U(0) - (0.3)U(1) + (0.1)U(2) < 0.$$

Since the person prefers X_1 to X_2, we know that

$$\begin{aligned} E[U(X_1)] &= (0.2)U(0) + (0.5)U(1) + (0.3)U(2) > E[U(X_2)] \\ &= (0.4)U(0) + (0.2)U(1) + (0.4)U(2), \end{aligned}$$

which implies that

$$(0.2)U(0) - (0.3)U(1) + (0.1)U(2) < 0.$$

This is precisely the inequality which was needed to conclude that the person will prefer X_3 to X_4.

11. For any given value of b,

$$E[U(X)] = p\log(A+b) + (1-p)\log(A-b).$$

Therefore,

$$\frac{\partial E[U(X)]}{\partial b} = \frac{p}{A+b} - \frac{1-p}{A-b}.$$

When this derivative is set equal to 0, we find that

$$b = (2p-1)A.$$

Since $\frac{\partial^2 E[U(X)]}{\partial b^2} < 0$, this value of b does yield a maximum value of $E[U(X)]$. If $p \geq 1/2$, this value of b lies between 0 and A as required. However, if $p < 1/2$, this value of b is negative and not permissible. In this case, it can be shown that the maximum value of $E[U(X)]$ for $0 \leq b \leq A$ occurs when $b = 0$; that is, when the person does not bet at all.

12. For any given value of b,

$$E[U(X)] = p(A+b)^{1/2} + (1-p)(A-b)^{1/2}.$$

Therefore,

$$\frac{\partial E[U(X)]}{\partial b} = \frac{p}{2(A+b)^{1/2}} - \frac{1-p}{2(A-b)^{1/2}}.$$

When this derivative is set equal to 0, we find that

$$b = \frac{p^2 - (1-p)^2}{p^2 + (1-p)^2}A.$$

As in Exercise 11, if $p \geq 1/2$, then this value of b lies in the interval $0 \leq b \leq A$ and will maximize $E[U(X)]$. However, if $p < 1/2$, the value of b in the interval $0 \leq b \leq A$ for which $E[U(X)]$ is a maximum is $b = 0$.

13. For any given value of b,

$$E[U(X)] = p(A+b) + (1-p)(A-b).$$

This is a linear function of b. If $p > 1/2$, it has the form shown in sketch (i) of Fig. S.4.3 and $b = A$ is best. If $p < 1/2$, it has the form shown in sketch (ii) of Fig. S.4.3 and $b = 0$ is best. If $p = 1/2$, $E[U(X)] = A$ for all values of b.

14. For any given value of b,

$$E[U(X)] = p(A+b)^2 + (1-p)(A-b)^2.$$

This is a parabola in b. If $p \geq 1/2$, it has the shape shown in sketch (i) of Fig. S.4.4. Therefore, $E[U(X)]$ is a maximum for $b = A$. If $1/4 < p \leq 1/2$, it has the shape shown in sketch (ii) of Fig. S.4.4. Therefore, $E[U(X)]$ is again a maximum for $b = A$. If $0 \leq p < 1/4$, it has the shape shown in sketch (iii) of Fig. S.4.4. Therefore, $E[U(X)]$ is a maximum for $b = 0$. Finally, if $p = 1/4$, then it is symmetric with respect to the point $b = A/2$, as shown in sketch (iv) of Fig. S.4.4. Therefore, $E[U(X)]$ is a maximum for $b = 0$ and $b = A$.

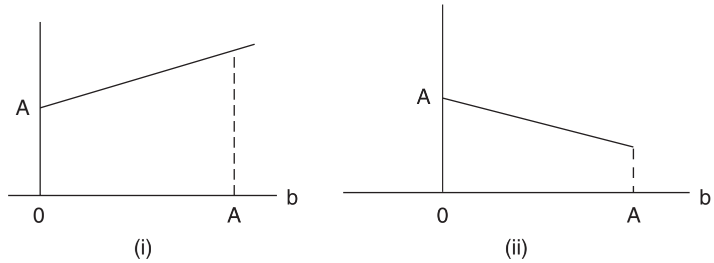

Figure S.4.3: Figure for Exercise 13 of Sec. 4.8.

15. The expected utility for the lottery ticket is

$$E[U(X)] = \int_0^4 x^\alpha \frac{1}{4} dx = \frac{4^\alpha}{\alpha+1}.$$

The utility of accepting x_0 dollars instead of the lottery ticket is $U(x_0) = x_0^\alpha$. Therefore, the person will prefer to sell the lottery ticket for x_0 dollars if

$$x_0^\alpha > \frac{4^\alpha}{\alpha+1} \quad \text{or if} \quad x_0 > \frac{4}{(\alpha+1)^{1/\alpha}}.$$

It can be shown that the right-hand side of this last inequality is an increasing function of α.

16. The expected utility from choosing the prediction d is

$$E[U(-[Y-d]^2) = E(|Y-d|).$$

We already saw (in Sec. 4.5) that d equal to a median of the distribution of Y minimizes this expectation.

17. The gain is 10^6 if $P > 1/2$ and -10^6 if $P \le 1/2$. The utility of continuing to promote is then $10^{5.4}$ if $P > 1/2$ and -10^6 if $P \le 1/2$. To find the expected utility, we need $\Pr(P \le 1/2)$. Using the stated p.d.f. for P, we get $\Pr(P \le 1/2) = \int_0^{1/2} 56p^6(1-p)dp = 0.03516$. The expected utility is then $10^{5.4} \times (1 - 0.03516) - 10^6 \times 0.03516 = 207197$. This is greater than 0, so we would continue to promote the treatment.

4.9 Supplementary Exercises

Solutions to Exercises

1. If $u \ge 0$,

$$\int_u^\infty xf(x)dx \ge u\int_u^\infty f(x)dx = u[1-F(u)].$$

Since

$$\lim_{u\to\infty}\int_{-\infty}^u xf(x)dx = E(X) = \int_{-\infty}^\infty xf(x)dx < \infty,$$

it follows that

$$\lim_{u\to\infty}\left[E(X) - \int_{-\infty}^u xf(x)dx\right] = \lim_{u\to\infty}\int_u^\infty xf(x)dx = 0.$$

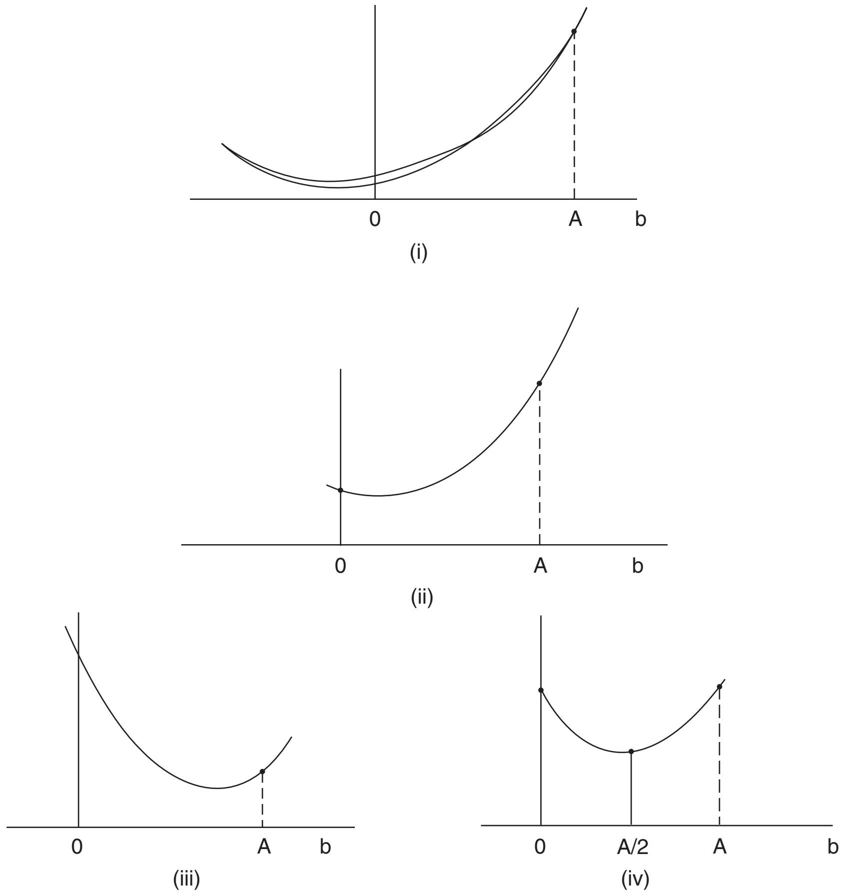

Figure S.4.4: Figure for Exercise 14 of Sec. 4.8.

2. We use integration by parts. Let $u = 1 - F(x)$ and $dv = dx$. Then $du = -f(x)dx$ and $v = x$, and the integral given in this exercise becomes

$$[uv]_0^\infty - \int_0^\infty v du = \int_0^\infty x f(x) dx = E(X).$$

3. Let $x_1, x_2, \dots$ denote the possible values of X. Since $F(X)$ is a step function, the integral given in Exercise 1 becomes the following sum:

$$\begin{aligned} &(x_1 - 0) + [1 - f(x_1)](x_2 - x_1) + [1 - f(x_1) - f(x_2)](x_3 - x_2) + \cdots \\ &= x_1 f(x_1) + x_2 \ f(x_2) + x_3 f(x_3) + \cdots \\ &= E(X). \end{aligned}$$

4. If X, Y, and Z each had the required uniform distribution, then

$$E(X + Y + Z) = E(X) + E(Y) + E(Z) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}.$$

But since $X + Y + Z \le 1.3$, this is impossible.

5. We need $E(Y) = a\mu + b = 0$ and

$$\text{Var}(Y) = a^2\sigma^2 = 1.$$

Therefore, $a = \pm \frac{1}{\sigma}$ and $b = -a\mu$.

6. The p.d.f. $h_1(w)$ of the range W is given at the end of Sec. 3.9. Therefore,

$$E(W) = n(n-1)\int_0^1 w^{n-1}(1-w)dw = \frac{n-1}{n+1}.$$

7. The dealer's expected gain is

$$E(Y - X) = \frac{1}{36}\int_0^6 \int_0^y (y-x)x\,dx\,dy = \frac{3}{2}.$$

8. It follows from Sec. 3.9 that the p.d.f. of Y_n is

$$g_n(y) = n[F(y)]^{n-1}\ f(y).$$

Here, $F(y) = \int_0^y 2xdx = y^2$, so

$$g_n(y) = 2ny^{2n-1} \quad \text{for} \quad 0 < y < 1.$$

Hence, $E(Y_n) = \int_0^1 y\ g_n(y)dy = \dfrac{2n}{2n+1}$.

9. Suppose first that $r(X)$ is nondecreasing. Then

$$\Pr[Y \geq r(m)] = \Pr[r(X) \geq r(m)] \geq \Pr(X \geq m) \geq \frac{1}{2},$$

and

$$\Pr[Y \leq r(m)] = \Pr[r(X) \leq r(m)] \geq \Pr(X \leq m) \geq \frac{1}{2}.$$

Hence, $r(m)$ is a median of the distribution of Y. If $r(X)$ is nonincreasing, then

$$\Pr[Y \geq r(m)] \geq \Pr(X \leq m) \geq \frac{1}{2}$$

and

$$\Pr[Y \leq r(m)] \geq \Pr(X \geq m) \geq \frac{1}{2}.$$

10. Since m is the median of a continuous distribution,

$$\begin{aligned} \Pr(X < m) = \Pr(X > m) &= \frac{1}{2}. \quad \text{Hence,} \\ \Pr(Y_n > m) &= 1 - \Pr(\text{All} \quad X_i'\ s < m) \\ &= 1 - \frac{1}{2^n}. \end{aligned}$$

11. Suppose that you order s liters. If the demand is $x < s$, you will make a profit of gx cents on the x liters sold and suffer a loss of $c(s-x)$ cents on the $s-x$ liters that you do not sell. Therefore, your net profit will be $gx - c(s-x) = (g+c)x - cs$. If the demand is $x \geq s$, then you will sell all s liters and make a profit of gs cents. Hence, your expected net gain is

$$\begin{aligned} E &= \int_0^s [(g+c)x - cs] f(x)dx + gs\int_s^\infty f(x)dx \\ &= \int_0^s (g+c)x\ f(x)dx - csF(s) + gs[1-F(s)]. \end{aligned}$$

To find the value of s that maximizes E, we find, after some calculations, that

$$\frac{dE}{ds} = g - (g+c)\quad F(s).$$

Thus, $\frac{dE}{ds} = 0$ and E is maximized when s is chosen so that $F(s) = g/(g+c)$.

12. Suppose that you return at time t. If the machine has failed at time $x \leq t$, then your cost is $c(t-x)$. If the machine has net yet failed $(x > t)$, then your cost is b. Therefore, your expected cost is

$$E = \int_0^t c(t-x)f(x)dx + b\int_t^\infty f(x)dx = ctF(t) - c\int_0^t xf(x)dx + b[1-F(t)].$$

Hence,

$$\frac{dE}{dt} = cF(t) - bf(t).$$

and E will be maximized at a time t such that $cF(t) = bf(t)$.

13. $E(Z) = 5(3) - 1 + 15 = 29$ in all three parts of this exercise. Also,

$$\text{Var}(Z) = 25\,\text{Var}(X) + \text{Var}(Y) - 10\,\text{Cov}(X,Y) = 109 - 10\,\text{Cov}(X,Y).$$

Hence, $\text{Var}(Z) = 109$ in parts (a) and (b). In part (c),

$$\text{Cov}(X,Y) = \rho\sigma_X\sigma_Y = (.25)(2)(3) = 1.5$$

so $\quad \text{Var}(Z) = 94.$

14. In this exercise, $\sum_{j=1}^{n} y_j = x_n - x_0$. Therefore,

$$\text{Var}(\overline{Y}_n) = \frac{1}{n^2}\text{Var}(X_n - X_0).$$

Since X_n and X_0 are independent,

$$\text{Var}(X_n - X_0) = \text{Var}(X_n) + \text{Var}(X_0).$$

Hence, $\text{Var}(\overline{Y}_n) = \dfrac{2\sigma^2}{n^2}$.

15. Let $v^2 = \text{Var}(X_1 + \cdots + X_n) = \sum_i \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$. In this problem $\text{Var}(X_i) = \sigma^2$ for all i and $\text{Cov}(X_i, X_j) = \rho\sigma^2$ for all $i \neq j$. Therefore,

$$v^2 = n\sigma^2 + n(n-1)\rho\sigma^2.$$

Since $v^2 \geq 0$, it follows that $\rho \geq -1/(n-1)$.

16. Since the correlation is unaffected by a translation of the distribution of X and Y in the xy-plane, we can assume without loss of generality that the origin is at the center of the rectangle. Hence, by symmetry, $E(X) = E(Y) = 0$. But it also follows from symmetry that $E(XY) = 0$ because, for any positive value of XY in the first or third quadrant, there is a corresponding negative value in the second or fourth quadrant with the same constant density. Thus, $\text{Cov}(X, Y) = 0$ and $\rho(X, Y) = 0$.

More directly, one can argue that the joint p.d.f. of (X, Y) factors into constants times the indicator functions of the two intervals that define the sides of the rectangles, hence X and Y are independent and uncorrelated.

17. For $i = 1, \ldots, n$, let $X_i = 1$ if the ith letter is placed in the correct envelope and let $X_i = 0$ otherwise. Then $E(X_i) = 1/n$ and, for $i \neq j$,

$$E(X_iX_j) = \Pr(X_iX_j = 1) = \Pr(X_i = 1 \text{ and } X_j = 1) = \frac{1}{n(n-1)}.$$

Also, $E(X_i^2) = E(X_i) = 1/n$. Hence,

$$\text{Var}(X_i) = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}$$

and $\text{Cov}(X_i, X_j) = \dfrac{1}{n(n-1)} - \dfrac{1}{n^2} = \dfrac{1}{n^2(n-1)}$. The total number of correct matches is $X = \sum_{i=1}^n X_i$. Therefore,

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j) = n \cdot \frac{n-1}{n^2} + n(n-1) \cdot \frac{1}{n^2(n-1)} = 1.$$

18.

$$\begin{aligned} E[(X - \mu)^3] &= E(X^3) - 3\mu E(X^2) + 3\mu^2 E(X) - \mu^3 \\ &= E(X^3) - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 - \mu^3 \\ &= E(X^3) - 3\mu\sigma^2 - \mu^3. \end{aligned}$$

19. $$c'(t) = \frac{\psi'(t)}{\psi(t)} \quad \text{and} \quad c''(t) = \frac{\psi(t)\psi''(t) - [\psi'(t)]^2}{[\psi(t)]^2}$$

Since $\psi(0) = 1$, $\psi'(0) = \mu$, and $\psi''(0) = E(X^2) = \sigma^2 + \mu^2$, it follows that $c'(0) = \mu$ and $c''(0) = \sigma^2$.

20. It was shown in Exercise 12 of Sec. 4.7 that if $E(Y \mid X) = aX + b$, then

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\rho\sigma_y}{\sigma_x}$$

and $b = \mu_Y - a\mu_X$. The desired result now follows immediately.

21. Since the coefficient of X in $E(Y \mid X)$ is negative, it follows from Exercise 20 that $\rho < 0$. Furthermore, it follows from Exercise 20 that the product of the coefficients of X and Y in $E(Y \mid X)$ and $E(X \mid Y)$ must be ρ^2. Hence, $\rho^2 = 1/4$ and, since $\rho < 0$, $\rho = -1/2$.

22. Let X and Y denote the lengths of the longer and shorter pieces, respectively. Since $Y = 3 - X$ with probability 1, it follows that $\rho = -1$.

23.

$$\begin{aligned}
\mathrm{Cov}(X, X+bY) &= \mathrm{Var}(X) + b\,\mathrm{Cov}(X,Y) \\
&= 1 + b\rho. \\
\mathrm{Var}(X) &= 1, \mathrm{Var}(X+bY) = 1 + b^2 + 2b\rho.
\end{aligned}$$

Hence,

$$\rho(X, X+bY) = \frac{1+b\rho}{(1+b^2+2b\rho)^{1/2}}.$$

If we set this quantity equal to ρ, square both sides, and solve for b, we obtain $b = -1/(2\rho)$.

24. The p.f. of the distribution of employees is

$$f(0) = .1,\ f(1) = .2,\ f(3) = .3,\ \text{and}\ \ f(5) = .4 \quad .$$

(a) The unique median of this distribution is 3, so the new office should be located at the point 3.

(b) The mean of this distribution is $(.1)(0) + (.2)(1) + (.3)(3) + (.4)(5) = 3.1$, so the new office should be located at the point 3.1.

25. (a) The marginal p.d.f. of X is

$$f_1(x) = \int_0^x 8xy\, dy = 4x^3 \qquad \text{for } 0 < x < 1.$$

Therefore, the conditional p.d.f. of Y given that $X = .2$ is

$$g_1(y \mid X = .2) = \frac{f(.2, y)}{f_1(.2)} = 50y \quad \text{for} \quad 0 < y < .2\ .$$

The mean of this distribution is

$$E(Y \mid X = .2) = \frac{2}{15} = .1333.$$

(b) The median of $g_1(y \mid X = .2)$ is $m = \left(\frac{1}{50}\right)^{1/2} = .1414$.

26.

$$\begin{aligned}
&\mathrm{Cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] \\
&= E\{[X - E(X \mid Z) + E(X \mid Z) - \mu_X] \cdot [Y - E(Y \mid Z) + E(Y \mid Z) - \mu_Y]\} \\
&= E\{[X - E(X \mid Z)][Y - E(Y \mid Z)]\} + E\{[X - E(X \mid Z)][E(Y \mid Z) - \mu_Y]\} \\
&\quad + E\{[E(X \mid Z) - \mu_X][Y - E(Y \mid Z)]\} + E\{[E(X \mid Z) - \mu_X][E(Y \mid Z) - \mu_Y]\}.
\end{aligned}$$

Consider these final four expectations. In the first one, if we first calculate the conditional expectation given Z and then take the expectation over Z we obtain $E[\mathrm{Cov}(X, Y \mid Z)]$. In the second and third expectations, we obtain the value 0 when we take the conditional expectation given Z. The fourth expectation is $\mathrm{Cov}[E(X \mid Z), E(Y \mid Z)]$.

27. Let N be the number of balls in the box. Since the proportion of red balls is p, there are Np red balls in the box. (Clearly, p must be an integer multiple of $1/N$.) There are $N(1-p)$ blue balls in the box. Let $K = Np$ so that there are $N - K$ blue balls and K red balls. If $n > K$, then $\Pr(Y = n) = 0$ since there are not enough red balls. Since $\Pr(X = n) > 0$ for all n, the result is true if $n > K$. For $n \le K$, let $X_i = 1$ if the ith ball is red for $i = 1, \ldots, n$. For sampling without replacement,

$$\Pr(Y = n) = \Pr(X_1 = 1) \prod_{i=2}^{n} \Pr(X_i = 1 | X_1 = 1, \ldots, X_{i-1} = 1) = \frac{K}{N}\frac{K-1}{N-1} \cdots \frac{K-n+1}{N-n+1}. \quad \text{(S.4.4)}$$

For sampling with replacement, the X_i's are independent, so

$$\Pr(X = n) = \prod_{i=1}^{n} \Pr(X_i = 1) = \left(\frac{K}{N}\right)^n. \quad \text{(S.4.5)}$$

For $j = 1, \ldots, n-1$, $KN - jN < KN - jK$, so $(K-j)/(N-j) < K/N$. Hence the product in (S.4.4) is smaller than the product in (S.4.5). This argument makes sense only if N is finite. If N is infinite, then sampling with and without replacement are equivalent.

28. The expected utility from the gamble X is $E[U(X)] = E(X^2)$. The utility of receiving $E(X)$ is $U[E(X)] = [E(X)]^2$. We know from Theorem 4.3.1 that $E(X^2) \ge [E(X)]^2$ for any gamble X, and from Theorem 4.3.3 that there is strict inequality unless X is actually constant with probability 1.

29. The expected utility from allocating the amounts a and $m - a$ is

$$\begin{aligned} E &= p\log(g_1 a) + (1-p)\log[g_2(m-a)] \\ &= p\log\ a + (1-p)\log(m-a) \\ &+ p\log\ g_1 + (1-p)\log\ g_2. \end{aligned}$$

The maximum over all values of a can now be found by elementary differentiation, as in Exercise 7 of Sec.4.8, and we obtain $a = pm$.

Chapter 5

Special Distributions

5.2 The Bernoulli and Binomial Distributions

Commentary

If one is using the statistical software R, then the functions `dbinom`, `pbinom`, and `qbinom` give the p.f., the c.d.f., and the quantile function of binomial distributions. The syntax is that the first argument is the argument of the function, and the next two are n and p respectively. The function `rbinom` gives a random sample of binomial random variables. The first argument is how many you want, and the next two are n and p. All of the solutions that require the calculation of binomial probabilites can be done using these functions instead of tables.

Solutions to Exercises

1. Since $E(X^k)$ has the same value for every positive integer k, we might try to find a random variable X such that X, X^2, X^3, X^4,... all have the same distribution. If X can take only the values 0 and 1, then $X^k = X$ for every positive integer k since $0^k = 0$ and $1^k = 1$. If $\Pr(X = 1) = p = 1 - \Pr(X = 0)$, then in order for $E(X^k) = 1/3$, as required, we must have $p = 1/3$. Therefore, a random variable X such that $\Pr(X = 1) = 1/3$ and $\Pr(X = 0) = 2/3$ satisfies the required conditions.

2. We wish to express $f(x)$ in the form $p^{\alpha(x)}(1-p)^{\beta(x)}$, where $\alpha(x) = 1$ and $\beta(x) = 0$ and $x = a$ and $\alpha(x) = 0$ and $\beta(x) = 1$ for $x = b$. If we choose $\alpha(x)$ and $\beta(x)$ to be linear functions of the form $\alpha(x) = \alpha_1 + \alpha_2 x$ and $\beta(x) = \beta_1 + \beta_2 x$, then the following two pairs of equations must be satisfied:

$$\begin{aligned} \alpha_1 + \alpha_2 a &= 1 \\ \alpha_1 + \alpha_2 b &= 0, \end{aligned}$$

and

$$\begin{aligned} \beta_1 + \beta_2 a &= 0 \\ \beta_1 + \beta_2 b &= 1. \end{aligned}$$

Hence,

$$\begin{aligned} \alpha_1 &= -\frac{b}{a-b}, \qquad \alpha_2 = \frac{1}{a-b} \\ \beta_1 &= -\frac{a}{b-a}, \qquad \beta_2 = \frac{1}{b-a}. \end{aligned}$$

3. Let X be the number of heads obtained. Then strictly more heads than tails are obtained if $X \in \{6, 7, 8, 9, 10\}$. The probability of this event is the sum of the numbers in the binomial table corresponding to $p = 0.5$ and $n = 10$ for $k = 6, \dots, 10$. By the symmetry of this binomial distribution, we can also compute the sum as $(1 - \Pr(X = 5))/2 = (1 - 0.2461)/2 = 0.37695$.

4. It is found from a table of the binomial distribution with parameters $n = 15$ and $p = 0.4$ that

$$\begin{aligned}\Pr(6 \le X \le 9) &= \Pr(X = 6) + \Pr(X = 7) + \Pr(X = 8) + \Pr(X = 9)\\ &= .2066 + .1771 + .1181 + .0612 = .5630.\end{aligned}$$

5. The tables do not include the value $p = 0.6$, so we must use the trick described in Exercise 7 of Sec. 3.1. The number of *tails* X will have the binomial distribution with parameters $n = 9$ and $p = 0.4$. Therefore,

$$\begin{aligned}\Pr(\text{Even number of heads}) &= \Pr(\text{Odd number of tails})\\ &= \Pr(X = 1) + \Pr(X = 3) + \Pr(X = 5) + \Pr(X = 7) + \Pr(X = 9)\\ &= .0605 + .2508 + .1672 + .0212 + .0003\\ &= .5000.\end{aligned}$$

6. Let N_A, N_B, and N_C denote the number of times each man hits the target. Then

$$\begin{aligned}E(N_A + N_B + N_C) &= E(N_A) + E(N_B) + E(N_C)\\ &= 3 \cdot \frac{1}{8} + 5 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} = \frac{21}{8}.\end{aligned}$$

7. If we assume that N_A, N_B, and N_C are independent, then

$$\begin{aligned}\text{Var}(N_A + N_B + N_C) &= \text{Var}(N_A) + \text{Var}(N_B) + \text{Var}(N_C)\\ &= 3 \cdot \frac{1}{8} \cdot \frac{7}{8} + 5 \cdot \frac{1}{4} \cdot \frac{3}{4} + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{113}{64}.\end{aligned}$$

8. The number X of components that fail will have the binomial distribution with parameters $n = 10$ and $p = 0.2$. Therefore,

$$\begin{aligned}\Pr(X \ge 2 \,|\, X \ge 1) &= \frac{\Pr(X \ge 2)}{\Pr(X \ge 1)} = \frac{1 - \Pr(X = 0) - \Pr(X = 1)}{1 - \Pr(X = 0)}\\ &= \frac{1 - .1074 - .2684}{1 - .1074} = \frac{.6242}{.8926} = .6993.\end{aligned}$$

9. $$\Pr\left(X_1 = 1 \middle| \sum_{i=1}^{n} X_i = k\right) = \frac{\Pr\left(X_1 = 1 \text{ and } \sum_{i=1}^{n} X_i = k\right)}{\Pr\left(\sum_{i=1}^{n} X_i = k\right)} = \frac{\Pr\left(X_1 = 1 \text{ and } \sum_{i=2}^{n} X_i = k - 1\right)}{\Pr\left(\sum_{i=1}^{n} X_i = k\right)}.$$

Since the random variables $X_1, \dots, X_n$ are independent, it follows that X_1 and $\sum_{i=2}^n X_i$ are independent. Therefore, the final expression can be rewritten as

$$\frac{\Pr(X_1 = 1)\Pr\left(\sum_{i=2}^n X_i = k-1\right)}{\Pr\left(\sum_{i=1}^n X_i = k\right)}.$$

The sum $\sum_{i=2}^n X_i$ has the binomial distribution with parameters $n-1$ and p, and the sum $\sum_{i=1}^n X_i$ has the binomial distribution with parameters n and p. Therefore,

$$\Pr\left(\sum_{i=2}^n X_i = k-1\right) = \binom{n-1}{k-1} p^{k-1}(1-p)^{(n-1)-(k-1)} = \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k},$$

and

$$\Pr\left(\sum_{i=1}^n X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Also, $\Pr(X_1 = 1) = p$. It now follows that

$$\Pr\left(X_1 = 1 \middle| \sum_{i=1}^n X_i = k\right) = \frac{\binom{n-1}{k-1} p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{k}{n}.$$

10. The number of children X in the family who will inherit the disease has the binomial distribution with parameters n and p. Let $f(x|n,p)$ denote the p.f. of this distribution. Then

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - f(0|n,p) = 1 - (1-p)^n.$$

For $x = 1, 2, \dots, n$,

$$\Pr(X = x | X \geq 1) = \frac{\Pr(X = x)}{\Pr(X \geq 1)} = \frac{f(x|n,p)}{1-(1-p)^n}.$$

Therefore, the conditional p.f. of X given that $X \geq 1$ is $f(x|n,p)/(1-[1-p]^n)$ for $x = 1, 2, \dots, n$. The required expectation $\mathrm{E}(X \mid X \geq 1)$ is the mean of this conditional distribution. Therefore,

$$E(X|X \geq 1) = \sum_{x=1}^n x \frac{f(x \mid n,p)}{1-(1-p)^n} = \frac{1}{1-(1-p)^n} \sum_{x=1}^n x f(x|n,p).$$

However, we know that the mean of the binomial distribution is np; i.e.,

$$E(X) = \sum_{x=0}^n x f(x \mid n,p) = np.$$

Furthermore, we can drop the term corresponding to $x = 0$ from this summation without affecting the value of the summation, because the value of that term is 0. Hence, $\sum_{x=1}^n x f(x|n,p) = np$. It now follows that $E(X|X \geq 1) = np/(1-[1-p]^n)$.

11. Since the value of the term being summed here will be 0 for $x = 0$ and for $x = 1$, we may change the lower limit of the summation from $x = 2$ to $x = 0$, without affecting the value of the sum. The summation can then be rewritten as

$$\sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} - \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}.$$

If X has the binomial distribution with parameters n and p, then the first summation is simply $E(X^2)$ and the second summation is simply $E(X)$. Finally,

$$E(X^2) - E(X) = \mathrm{Var}(X) + [E(X)]^2 - E(X) = np(1-p) + (np)^2 - np = n(n-1)p^2.$$

12. Assuming that p is not 0 or 1,

$$\frac{f(x+1 \mid n,p)}{f(x \mid n,p)} = \frac{\binom{n}{x+1} p^{x+1}(1-p)^{n-x-1}}{\binom{n}{x} p^x (1-p)^{n-x}} = \frac{n-x}{x+1} = \frac{p}{1-p}.$$

Therefore,

$$\frac{f(x+1 \mid n,p)}{f(x \mid n,p)} \geq 1 \quad \text{if and only if } x \leq (n+1)p - 1.$$

It follows from this relation that the values of $f(x \mid n,p)$ will increase as x increases from 0 up to the greatest integer less than $(n+1)p$, and will then decrease as x continues increasing up to n. Therefore, if $(n+1)p$ is not an integer, the unique mode will be the greatest integer less than $(n+1)p$. If $(n+1)p$ is an integer, then both $(n+1)p$ and $(n+1)p - 1$ are modes. If $p = 0$, the mode is 0 and if $p = 1$, the mode is n.

13. Let X be the number of successes in the group with probability 0.5 of success. Let Y be the number of successes in the group with probability 0.6 of success. We want $\Pr(X \geq Y)$. Both X and Y have discrete (binomial) distributions with possible values $0, \ldots, 5$. There are 36 possible (X, Y) pairs and we need the sum of the probabilities of the 21 of them for which $X \geq Y$. To save time, we shall calculate the probabilities of the 15 other ones and subtract the total from 1. Since X and Y are independent, we can write $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$, and find each of the factors in the binomial table in the back of the book. For example, for $x = 1$ and $y = 2$, we get $0.1562 \times 0.2304 = 0.03599$. Adding up all 15 of these and subtracting from 1 we get 0.4957.

14. Before we prove the three facts, we shall show that they imply the desired result. According to (c), every distribution with the specified moments must take only the values 0 and 1. The mean of such a distribution is $\Pr(X = 1)$. This number, $\Pr(X = 1)$, uniquely determines every distribution that can only take the two values 0 an 1.

    (a) Suppose that $\Pr(|X| > 1) > 0$. Then there exists $\epsilon > 0$ such that $\Pr(|X| > 1 + \epsilon) > 0$. Then

    $$E(X^{2k}) \geq (1+\epsilon)^{2k} \Pr(|X| > 1 + \epsilon).$$

    Since the right side of this equation goes to ∞ as $k \to \infty$, it cannot be the case that $E(X^{2k}) = 1/3$ for all k. This contradiction means that our assumption that $\Pr(|X| > 1) > 0$ must be false. That is, $\Pr(|X| \leq 1) = 1$.

(b) Since $X^4 < X^2$ whenever $|X| \le 1$ and $X^2 \notin \{0,1\}$, it follows that $E(X^4) < E(X^2)$ whenever $\Pr(|X| \le 1) = 1$ and $\Pr(X^2 \notin \{0,1\}) > 0$. Since we know that $E(X^4) = E(X^2)$ and $\Pr(|X| \le 1) = 1$, it must be that $\Pr(X^2 \notin \{0,1\}) = 0$. That is, $\Pr(X^2 \in \{0,1\}) = 1$.

(c) From (b) we know that $\Pr(X \in \{-1,0,1\}) = 1$. We also know that

$$\begin{aligned} E(X) &= \Pr(X=1) - \Pr(X=-1) \\ E(X^2) &= \Pr(X=1) + \Pr(X=-1). \end{aligned}$$

Since these two are equal, it follows that $\Pr(X = -1) = 0$.

15. We need the maximum number of tests if and only if every first-stage and second-stage subgroup has at least one positive result. In that case, we would need $10 + 100 + 1000 = 1110$ total tests. The probability that we have to run this many tests is the probability that every $Y_{2,i,k} = 1$, which in turn is the probability that every $Z_{2,i,k} > 0$. The $Z_{2,i,k}$'s are independent binomial random variables with parameters 10 and 0.002, and there are 100 of them altogether. The probability that each is positive is 0.0198, as computed in Example 5.2.7. The probability that they are all positive is $(0.0198)^{100} = 4.64 \times 10^{-171}$.

16. We use notation like that in Example 5.2.7 with one extra stage. For $i = 1,\dots,5$, let $Z_{1,i}$ be the number of people in group i who test positive. Let $Y_{1,i} = 1$ if $Z_{1,i} > 0$ and $Y_{1,i} = 0$ if not. Then $Z_{1,i}$ has the binomial distribution with parameters 200 and 0.002, while $Y_{1,i}$ has the Bernoulli distribution with parameter $1 - 0.998^{200} = 0.3299$. Let $Z_{2,i,k}$ be the number of people who test positive in the kth subgroup of group i for $k = 1,\dots,5$. Let $Y_{2,i,k} = 1$ if $Z_{2,i,k} > 0$ and $Y_{2,i,k} = 0$ if not. Each $Z_{2,i,k}$ has the binomial distribution with parameters 40 and 0.002, while $Y_{2,i,k}$ has the Bernoulli distribution with parameter $1 - 0.998^{40} = 0.0770$. Finally, let $Z_{3,i,k,j}$ be the number of people who test positive in the jth sub-subgroup of the kth subgroup of the ith group. Let $Y_{3,i,k,j} = 1$ if $Z_{3,i,k,j} > 0$ and $Y_{3,i,k,j} = 0$ otherwise. Then $Z_{3,i,k,j}$ has the binomial distribution with parameters 8 and 0.002, while $Y_{3,i,k,j}$ has the Bernoulli distribution with parameter $1 - 0.998^8 = 0.0159$.

The maximum number of tests is needed if and only if there is at least one positive amongst every one of the 125 sub-subgroups of size 8. In that case, we need to make $1000+125+25+5 = 1155$ total tests. Let $Y_1 = \sum_{i=1}^5 Y_{1,i}$, which is the number of groups that need further attention. Let $Y_2 = \sum_{i=1}^5 \sum_{k=1}^5 Y_{2,i,k}$, which is the number of subgroups that need further attention. Let $Y_3 = \sum_{i=1}^5 \sum_{k=1}^5 \sum_{j=1}^5 Y_{3,i,k,j}$, which is the number of sub-subgroups that need all 8 members tested. The actual number of tests needed is $Y = 5+5Y_1+5Y_2+8Y_3$. The mean of Y_1 is $5 \times 0.3299 = 1.6497$. The mean of Y_2 is $25 \times 0.0770 = 1.9239$. The mean of Y_3 is $125 \times 0.0159 = 1.9861$. The mean of Y is then

$$E(Y) = 5 + 5 \times 1.6497 + 5 \times 1.9239 + 8 \times 1.9861 = 38.7569.$$

5.3 The Hypergeometric Distributions

Commentary

The hypergeometric distribution arises in finite population sampling and in some theoretical calculations. It actually does not figure in the remainder of this text, and this section could be omitted despite the fact that it is not marked with an asterisk. The section ends with a discussion of how to extend the definition of binomial coefficients in order to make certain formulas easier to write. This discussion is not central to the rest of the text. It does arise again in a theoretical discussion at the end of Sec. 5.5.

If one is using the statistical software R, then the functions `dhyper`, `phyper`, and `qhyper` give the p.f., the c.d.f., and the quantile function of hypergeometric distributions. The syntax is that the first argument is the argument of the function, and the next three are A, B, and n in the notation of the text. The function `rhyper` gives a random sample of hypergeometric random variables. The first argument is how many you want, and the next three are A, B, and n. All of the solutions that require the calculation of hypergeometric probabilites can be done using these functions.

Solutions to Exercises

1. Using Eq. (5.3.1) with the parameters $A = 10$, $B = 24$, and $n = 11$, we obtain the desired probability

$$\Pr(X = 10) = \frac{\binom{10}{10}\binom{24}{1}}{\binom{34}{11}} = 8.389 \times 10^{-8}.$$

2. Let X denote the number of red balls that are obtained. Then X has the hypergeometric distribution with parameters $A = 5, B = 10$, and $n = 7$. The maximum value of X is $\min\{n, A\} = 5$, hence,

$$\Pr(X \geq 3) = \sum_{x=3}^{5} \frac{\binom{5}{x}\binom{10}{7-x}}{\binom{15}{7}} = \frac{2745}{6435} \approx 0.4266.$$

3. As in Exercise 2, let X denote the number of red balls in the sample. Then, by Eqs. (5.3.3) and (5.3.4),

$$E(X) = \frac{nA}{A+B} = \frac{7}{3} \quad \text{and} \quad \text{Var}(X) = \frac{nAB}{(A+B)^2} \cdot \frac{A+B-n}{A+B-1} = \frac{8}{9}.$$

Since $\overline{X} = X/n$,

$$E(\overline{X}) = \frac{1}{n}E(X) = \frac{1}{3} \quad \text{and} \quad \text{Var}(\overline{X}) = \frac{1}{n^2}\text{Var}(X) = \frac{8}{441}.$$

4. By Eq. (5.3.4),

$$\text{Var}(X) = \frac{(8)(20)}{(28)^2(27)}\, n(28-n).$$

The quadratic function $n(28 - n)$ is a maximum when $n = 14$.

5. By Eq. (5.3.4),

$$\text{Var}(X) = \frac{A(T-A)}{T^2(T-1)}\, n(T-n).$$

If T is an even integer, then the quadratic function $n(T - n)$ is a maximum when $n = T/2$. If T is an odd integer, then the maximum value of $n(T - n)$, for $n = 0, 1, 2, \ldots, T$, occurs at the two integers $(T - 1)/2$ and $(T + 1)/2$.

6. For $x = 0, 1, \dots, k$,

$$\Pr(X_1 = x | X_1 + X_2 = k) = \frac{\Pr(X_1 = x \text{ and } X_1 + X_2 = k)}{\Pr(X_1 + X_2 = k)} = \frac{\Pr(X_1 = x \text{ and } X_2 = k - x)}{\Pr(X_1 + X_2 = k)}.$$

Since X_1 and X_2 are independent,

$$\Pr(X_1 = x \text{ and } X_2 = k - x) = \Pr(X_1 = x)\Pr(X_2 = k - x).$$

Furthermore, it follows from a result given in Sec. 5.2 that $X_1 + X_2$ will have the binomial distribution with parameters $n_1 + n_2$ and p. Therefore,

$$\begin{aligned}\Pr(X_1 = x) &= \binom{n_1}{x} p^x (1-p)^{n_1 - x},\\ \Pr(X_2 = k - x) &= \binom{n_2}{k-x} p^{k-x}(1-p)^{n_2 - k + x},\\ \Pr(X_1 + X_2 = k) &= \binom{n_1 + n_2}{k} p^k (1-p)^{n_1 + n_2 - k}.\end{aligned}$$

By substituting these values into the expression given earlier, we find that for $x = 0, 1, \dots, k$,

$$\Pr(X_1 = x | X_1 + X_2 = k) = \frac{\binom{n_1}{x}\binom{n_2}{k-x}}{\binom{n_1+n_2}{k}}.$$

It can be seen that this conditional distribution is a hypergeometric distribution with parameters n_1, n_2, and k.

7. (a) The probability of obtaining exactly x defective items is

$$\frac{\binom{0.3T}{x}\binom{0.7T}{10-x}}{\binom{T}{10}}.$$

Therefore, the probability of obtaining not more than one defective item is the sum of these probabilities for $x = 0$ and $x = 1$.
Since

$$\binom{0.3T}{0} = 1 \quad \text{and} \quad \binom{0.3T}{1} = 0.3T,$$

this sum is equal to

$$\frac{\binom{0.7T}{10} + 0.3T\binom{0.7T}{9}}{\binom{T}{10}}.$$

(b) The probability of obtaining exactly x defectives according to the binomial distribution, is

$$\binom{10}{x}(0.3)^x(0.7)^{10-x}.$$

The desired probability is the sum of these probabilities for $x = 0$ and $x = 1$, which is

$$(0.7)^{10} + 10(0.3)(0.7)^9.$$

For a large value of T, the answers in (a) and (b) will be very close to each other, although this fact is not obvious from the different forms of the two answers.

8. If we let X_i denote the height of the ith person selected, for $i = 1, \dots, n$, then $X = X_1 + \cdots + X_n$. Furthermore, since X_i is equally likely to have any one of the T values $a_1, \dots, a_T$, then

$$E(X_i) = \frac{1}{T}\sum_{i=1}^{T} a_i = \mu$$

and

$$\text{Var}(X_i) = \frac{1}{T}\sum_{i=1}^{T}(a_i - \mu)^2 = \sigma^2.$$

It follows that $E(X) = n\mu$. Furthermore, by Theorem 4.6.7,

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum\sum_{i<j} \text{Cov}(X_i, X_j).$$

Because of the symmetry among the variables $X_1, \dots, X_n$, it follows that

$$\text{Var}(X) = n\sigma^2 + n(n-1)\,\text{Cov}(X_1, X_2).$$

We know that $\text{Var}(X) = 0$ for $n = T$. Therefore,

$$\text{Cov}(X_1, X_2) = -\frac{1}{T-1}\sigma^2.$$

It now follows that

$$\text{Var}(X) = n\sigma^2 - \frac{n(n-1)}{T-1}\sigma^2 = n\sigma^2\left(\frac{T-n}{T-1}\right).$$

9. By Eq. (5.3.14),

$$\binom{3/2}{4} = \frac{(3/2)(1/2)(-1/2)(-3/2)}{4!} = \frac{3}{128}.$$

10. By Eq. (5.3.14),

$$\binom{-n}{k} = \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} = \frac{(-1)^k(n)(n+1)\dots(n+k-1)}{k!}.$$

If we reverse the order of the factors in the numerator, we can rewrite this relation as follows:

$$\binom{-n}{k} = \frac{(-1)^k(n+k-1)(n+k-2)\cdots(n)}{k!} = (-1)^k\binom{n+k-1}{k}.$$

11. Write $(1+a_n)^{c_n}e^{-a_nc_n} = \exp[c_n \log(1+a_n) - a_nc_n]$. The result is proven if we can show that

$$\lim_{n\to\infty}[c_n \log(1+a_n) - a_nc_n] = 0. \tag{S.5.1}$$

Use Taylor's theorem with remainder to write

$$\log(1+a_n) = a_n - \frac{a_n^2}{2(1+y_n)^2},$$

where y_n is between 0 and a_n. It follows that

$$c_n \log(1+a_n) - a_nc_n = c_na_n - \frac{c_na_n^2}{2(1+y_n)^2} - a_nc_n = -\frac{c_na_n^2}{2(1+y_n)^2}.$$

We have assumed that $c_na_n^2$ goes to 0. Since y_n is between 0 and a_n, and a_n goes to 0, we have $1/[2(1+y_n)^2]$ goes to 0. This establishes (S.5.1).

5.4 The Poisson Distributions

Commentary

This section ends with a more theoretical look at the assumptions underlying the Poisson process. This material is designed for the more mathematically inclined students who might wish to see a derivation of the Poisson distribution from those assumptions. Such a derivation is outlined in Exercise 16 in this section.

If one is using the statistical software *R*, then the functions `dpois`, `ppois`, and `qpois` give the p.f., the c.d.f., and the quantile function of Poisson distributions. The syntax is that the first argument is the argument of the function, and the second is the mean. The function `rpois` gives a random sample of Poisson random variables. The first argument is how many you want, and the second is the mean. All of the solutions that require the calculation of Poisson probabilites can be done using these functions instead of tables.

Solutions to Exercises

1. The number of oocysts X in $t = 100$ liters of water has the Poisson distribution with mean $0.2 \times 0.1 \times 100 = 2$. Using the Poisson distribution table in the back of the book, we find

$$\Pr(X \geq 2) = 1 - \Pr(X \leq 1) = 1 - 0.1353 - 0.2707 = 0.594.$$

2. From the table of the Poisson distribution in the back of the book it is found that

$$\Pr(X \geq 3) = .0284 + .0050 + .0007 + .0001 + .0000 = .0342.$$

3. Since the number of defects on each bolt has the Poisson distribution with mean 0.4, and the observations for the five bolts are independent, the sum for the numbers of defects on five bolts will have the Poisson distribution with mean $5(0.4) = 2$. It is found from the table of the Poisson distribution that

$$\Pr(X \geq 6) = .0120 + .0034 + .0009 + .0002 + .0000 = .0165.$$

There is some rounding error in this, and 0.0166 is closer.

4. If $f(x \mid \lambda)$ is the p.f. of the Poisson distribution with mean λ, then

$$\Pr(X = 0) = f(0 \mid \lambda) = \exp(-\lambda).$$

5. Let Y denote the number of misprints on a given page. Then the probability p that a given page will contain more than k misprints is

$$p = \Pr(Y > k) = \sum_{i=k+1}^{\infty} f(i \mid \lambda) = \sum_{i=k+1}^{\infty} \frac{\exp(-\lambda)\lambda^i}{i!}.$$

Therefore,

$$1 - p = \sum_{i=0}^{k} f(i \mid \lambda) = \sum_{i=0}^{k} \frac{\exp(-\lambda)\lambda^i}{i!}.$$

Now let X denote the number of pages, among the n pages in the book, on which there are more than k misprints. Then for $x = 0, 1, \ldots, n$,

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

and

$$\Pr(X \geq m) = \sum_{x=m}^{n} \binom{n}{x} p^x (1-p)^{n-x}.$$

6. We shall assume that defects occur in accordance with a Poisson process. Then the number of defects in 1200 feet of tape will have the Poisson distribution with mean $\mu = 3(1.2) = 3.6$. Therefore, the probability that there will be no defects is $\exp(-\mu) = \exp(-3.6)$.

7. We shall assume that customers are served in accordance with a Poisson process. Then the number of customers served in a two-hour period will have the Poisson distribution with mean $\mu = 2(15) = 30$. Therefore, the probability that more than 20 customers will be served is

$$\Pr(X > 20) = \sum_{x=21}^{\infty} \frac{\exp(-30)(30)^x}{x!}.$$

8. For $x = 0, 1, \ldots, k$,

$$\Pr(X_1 = x \mid X_1 + X_2 = k) = \frac{\Pr(X_1 = x \text{ and } X_1 + X_2 = k)}{\Pr(X_1 + X_2 = k)} = \frac{\Pr(X_1 = x \text{ and } X_2 = k - x)}{\Pr(X_1 + X_2 = k)}.$$

Since X_1 and X_2 are independent,

$$\Pr(X_1 = x \text{ and } X_2 = k - x) = \Pr(X_1 = x)\Pr(X_2 = k - x).$$

Also, by Theorem 5.4.4 the sum $X_1 + X_2$ will have the Poisson distribution with mean $\lambda_1 + \lambda_2$. Hence,

$$\begin{aligned}
\Pr(X_1 = x) &= \frac{\exp(-\lambda_1)\lambda_1^x}{x!} \\
\Pr(X_2 = k - x) &= \frac{\exp(-\lambda_2)\lambda_2^{k-x}}{(k-x)!} \\
\Pr(X_1 + X_2 = k) &= \frac{\exp(-(\lambda_1 + \lambda_2))(\lambda_1 + \lambda_2)^k}{k!}
\end{aligned}$$

It now follows that

$$\Pr(X_1 = x \mid X_1 + X_2 = k) = \frac{k!}{x!(k-x)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{k-x} = \binom{k}{x} p^x (1-p)^{k-x},$$

where $p = \lambda_1/(\lambda_1+\lambda_2)$. It can now be seen that this conditional distribution is a binomial distribution with parameters k and $p = \lambda_1/(\lambda_1+\lambda_2)$.

9. Let N denote the total number of items produced by the machine and let X denote the number of defective items produced by the machine. Then, for $x = 0, 1, \ldots,$

$$\Pr(X = x) = \sum_{n=0}^{\infty} \Pr(X = x \mid N = n) \Pr(N = n).$$

Clearly, it must be true that $X \leq N$. Therefore, the terms in this summation for $n < x$ will be 0, and we may write

$$\Pr(X = x) = \sum_{n=x} \Pr(X = x \mid N = n) \Pr(N = n).$$

Clearly, $\Pr(X = 0 \mid N = 0) = 1$. Also, given that $N = n > 0$, the conditional distribution of X will be a binomial distribution with parameters n and p. Therefore,

$$\Pr(X = x \mid N = n) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

Also, since N has the Poisson distribution with mean λ,

$$\Pr(N = n) = \frac{\exp(-\lambda)\lambda^n}{n!}.$$

Hence,

$$\Pr(X = x) = \sum_{n=x}^{\infty} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \frac{\exp(-\lambda)\lambda^n}{n!} = \frac{1}{x!} p^x \exp(-\lambda) \sum_{n=x}^{\infty} \frac{1}{(n-x)!}(1-p)^{n-x}\lambda^n.$$

If we let $t = n - x$, then

$$\begin{aligned}
\Pr(X = x) &= \frac{1}{x!} p^x \exp(-\lambda) \sum_{t=0}^{\infty} \frac{1}{t!}(1-p)^t \lambda^{t+x} \\
&= \frac{1}{x!}(\lambda p)^x \exp(-\lambda) \sum_{t=0}^{\infty} \frac{[\lambda(1-p)]^t}{t!} \\
&= \frac{1}{x!}(\lambda p)^x \exp(-\lambda)\exp(\lambda(1-p)) = \frac{\exp(-\lambda p)(\lambda p)^x}{x!}.
\end{aligned}$$

It can be seen that this final term is the value of the p.f. of the Poisson distribution with mean λp.

10. It must be true that $X + Y = N$. Therefore, for any nonnegative integers x and y,

$$\begin{aligned}
\Pr(X = x \text{ and } Y = y) &= \Pr(X = x \text{ and } N = x + y) \\
&= \Pr(X = x \,|\, N = x + y)\Pr(N = x + y) \\
&= \frac{(x+y)!}{x!y!} p^x (1-p)^y \frac{\exp(-\lambda)\lambda^{x+y}}{(x+y)!} \\
&= \exp(-\lambda)\frac{(\lambda p)^x}{x!} \cdot \frac{[\lambda(1-p)]^y}{y!}.
\end{aligned}$$

The fact that we have factored $\Pr(X = x \text{ and } Y = y)$ into the product of a function of x and a function of y is sufficient for us to be able to conclude that X and Y are independent. However, if we continue further and write

$$\exp(-\lambda) = \exp(-\lambda p)\exp(-\lambda(1-p))$$

then we can obtain the factorization

$$\Pr(X = x \text{ and } Y = y) = \frac{\exp(-\lambda)p(\lambda p)^x}{x!} \cdot \frac{\exp(-\lambda(1-p))[\lambda(1-p)]^y}{y!} = \Pr(X = x)\Pr(Y = y).$$

11. If $f(x \,|\, \lambda)$ denotes the p.f. of the Poisson distribution with mean λ, then

$$\frac{f(x+1 \,|\, \lambda)}{f(x \,|\, \lambda)} = \frac{\lambda}{x+1}.$$

Therefore, $f(x \,|\, \lambda) < f(x+1 \,|\, \lambda)$ if and only if $x + 1 < \lambda$. It follows that if λ is not an integer, then the mode of this distribution will be the largest integer x that is less than λ or, equivalently, the smallest integer x such that $x + 1 > \lambda$. If λ is an integer, then both the values $\lambda - 1$ and λ will be modes.

12. It can be assumed that the exact distribution of the number of colorblind people in the group is a binomial distribution with parameters $n = 600$ and $p = 0.005$. Therefore, this distribution can be approximated by a Poisson distribution with mean $600(0.005) = 3$. It is found from the table of the Poisson distribution that

$$\Pr(X \le 1) = .0498 + .1494 = .1992.$$

13. It can be assumed that the exact number of sets of triplets in this hospital is a binomial distribution with parameters $n = 700$ and $p = 0.001$. Therefore, this distribution can be approximated by a Poisson distribution with mean $700(0.001) = 0.7$. It is found from the table of the Poisson distribution that

$$\Pr(X = 1) = 0.3476.$$

14. Let X denote the number of people who do not appear for the flight. Then everyone who does appear will have a seat if and only if $X \ge 2$. It can be assumed that the exact distribution of X is a binomial distribution with parameters $n = 200$ and $p = 0.01$. Therefore, this distribution can be approximated by a Poisson distribution with mean $200(0.01) = 2$. It is found from the table of the Poisson distribution that

$$\Pr(X \ge 2) = 1 - \Pr(X \le 1) = 1 - .1353 - .2707 = .5940.$$

15. The joint p.f./p.d.f. of X and λ is the Poisson p.f. with parameter λ times $f(\lambda)$ which equals

$$\exp(-\lambda)\frac{\lambda^x}{x!}2\exp(-2\lambda) = 2\exp(-3\lambda)\frac{\lambda^x}{x!}. \tag{S.5.2}$$

We need to compute the marginal p.f. of X at $x = 1$ and divide that into (S.5.2) to get the conditional p.d.f. of λ given $X = 1$. The marginal p.f. of X at $x = 1$ is the integral of (S.5.2) over λ when $x = 1$ is plugged in.

$$f_1(1) = \int_0^\infty 2\lambda\exp(-3\lambda)d\lambda = \frac{2}{9}.$$

This makes the conditional p.d.f. of λ equal to $9\lambda\exp(-3\lambda)$ for $\lambda > 0$.

16. (a) Let $A = \cup_{i=1}^n A_i$. Then

$$\{X = k\} = (\{X = k\} \cap A) \cup (\{X = k\} \cap A^c).$$

The second event on the right side of this equation is $\{W_n = k\}$. Call the first event on the right side of this equation B. Then $B \subset A$. Since B and $\{W_n = k\}$ are disjoint, $\Pr(X = k) = \Pr(W_n = k) + \Pr(B)$.

(b) Since the subintervals are disjoint, the events $A_1, \ldots, A_n$ are independent. Since the subintervals all have the same length t/n, each A_i has the same probability. It follows that

$$\Pr\left(\cap_{i=1}^n A_i^c\right) = [1 - \Pr(A_1)]^n.$$

By assumption, $\Pr(A_i) = o(1/n)$, so

$$\Pr(A) = 1 - \Pr\left(\cap_{i=1}^n A_i^c\right) = 1 - [1 - o(1/n)]^n.$$

So,

$$\lim_{n\to\infty} \Pr(A) = 1 - \lim_{n\to\infty}[1 - o(1/n)]^n = 1,$$

according to Eq. (5.4.9).

(c) Since the Y_i are i.i.d. Bernoulli random variables with parameter $p_n = \lambda t/n + o(1/n)$, we know that W_n has the binomial distribution with parameters n and p_n. Hence

$$\Pr(W_n = k) = \binom{n}{k} p_n^k(1-p_n)^{n-k} = \frac{n!}{(n-k)!k!}\left[\frac{\lambda t}{n} + o(1/n)\right]^k \left[1 - \frac{\lambda t}{n} - o(1/n)\right]^{n-k}.$$

For fixed k,

$$\lim_{n\to\infty} n^k\left[\frac{\lambda t}{n} + o(1/n)\right]^k = (\lambda t)^k.$$

Also, using the formula stated in the exercise,

$$\lim_{n\to\infty} n^k\left[1 - \frac{\lambda t}{n} - o(1/n)\right]^{n-k} = \exp(-\lambda t).$$

It follows that

$$\lim_{n\to\infty} \Pr(W_n = k) = \exp(-\lambda t)\frac{(\lambda t)^k}{k!}\lim_{n\to\infty}\frac{n!}{n^k(n-k)!}.$$

We can write

$$\frac{n!}{n^k(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{n\cdot n\cdots n}.$$

For fixed k, the limit of this ratio is 1 as $n \to \infty$.

(d) We have established that

$$\Pr(X = k) = \Pr(W_n = k) + \Pr(B).$$

Since the left side of this equation does not depend on n, we can write

$$\Pr(X = k) = \lim_{n\to\infty} \Pr(W_n = k) + \lim_{n\to\infty} \Pr(B).$$

In earlier parts of this exercise we showed that the two limits on the right are $\exp(-\lambda t)(\lambda t)^k/k!$ and 0 respectively. So, X has the Poisson distribution with mean λt.

17. Because $n_T A_T/(A_T + B_T)$ converges to λ, $n_T/(A_T + B_T)$ goes to 0. Hence, B_T eventually gets larger than n_T. Once B_T is larger than $n_T + x$ and A_T is larger than x, we have

$$\Pr(X_T = x) = \frac{\binom{A_T}{x}\binom{B_T}{n_T - x}}{\binom{A_T + B_T}{n_T}} = \frac{A_T! B_T! n_T! (A_T + B_T - n_T)!}{x!(A_T - x)!(n_T - x)!(B_T - n_T + x)!(A_T + B_T)!}.$$

Apply Stirling's formula to each of the factorials in the above expression except $x!$. A little manipulation gives that

$$\begin{aligned} &\lim_{T\to\infty} \frac{A_T^{A_T+1/2} B_T^{B_T+1/2} n_T^{n_T+1/2} (A_T + B_T - n_T)^{A_T+B_T-n_T+1/2} e^x}{\Pr(X_T = x)x!(A_T - x)^{A_T-x+1/2}(n_T - x)^{n_T-x+1/2}(B_T - n + x)^{B_T-n+x+1/2}(A_T + B_T)^{A_T+B_T+1/2}} \\ &= 1. \end{aligned} \tag{S.5.3}$$

Each of the following limits follows from Theorem 5.3.3:

$$\begin{aligned} \lim_{T\to\infty} \left(\frac{A_T}{A_T - x}\right)^{A_T-x+1/2} &= e^x, \\ \lim_{T\to\infty} \left(\frac{B_T}{B_T - n_T + x}\right)^{B_T-n_T+x+1/2} e^{-n_T} &= e^{-x}, \\ \lim_{T\to\infty} \left(\frac{A_T + B_T - n_T}{A_T + B_T}\right)^{A_T+B_T-n+1/2} e^{n_T} &= 1, \\ \lim_{T\to\infty} \left(\frac{n_T}{n_T - x}\right)^{n_T-x+1/2} &= e^{-x}, \\ \lim_{T\to\infty} \left(\frac{B_T}{A_T + B_T}\right)^{n_T-x} &= e^{-\lambda}, \end{aligned}$$

Inserting these limits in (S.5.3) yields

$$\lim_{T\to\infty} \frac{A_T^x e^{-\lambda} n_T^x}{\Pr(X_T = x)x!(A_T + B_T)^x} = 1. \tag{S.5.4}$$

Since $n_T A_T/(A_T + B_T)$ converges to λ, we have

$$\lim_{T\to\infty} \frac{A_T^x n_T^x}{(A_T + B_T)^x} = \lambda^x. \tag{S.5.5}$$

Together (S.5.4) and (S.5.5) imply that

$$\lim_{T\to\infty} \frac{\lambda^x e^{-\lambda}/x!}{\Pr(X_T = x)} = 1.$$

The numerator of this last expression is $\Pr(Y = x)$, which completes the proof.

18. First write

$$\frac{n_T A_T}{B_T} - \frac{n_T A_T}{A_T + B_T} = \frac{n_T A_T^2}{B_T(A_T + B_T)} = n_T \frac{A_T}{B_T} \frac{A_T}{A_T + B_T}. \tag{S.5.6}$$

For the "if" part, assume that $n_T A_T/B_T$ converges to λ. Since n_T goes to ∞, then A_T/B_T goes to 0, which implies that $A_T/(A_T + B_T)$ (which is smaller) goes to 0. In the final expression in (S.5.6), the product of the first two factors goes to λ by assumption, and the third factor goes to 0, so $n_T A_T/(A_T + B_T)$ converges to the same thing as $n_T A_T/B_T$, namely λ. For the "only if" part, assume that $n_T A_T/(A_T + B_T)$ converges to λ. It follows that $A_T/(A_T + B_T) = 1/(1 + B_T/A_T)$ goes to 0, hence A_T/B_T goes to 0. In the last expression in (S.5.6), the product of the first and third factors goes to λ by assumption, and the second factor goes to 0, hence $n_T A_T/B_T$ converges to the same thing as $n_T A_T/(A_T + B_T)$, namely λ.

5.5 The Negative Binomial Distributions

Commentary

This section ends with a discussion of how to extend the definition of negative binomial distribution by making use of the extended definition of binomial coefficients from Sec. 5.3.

If one is using the statistical software R, then the functions `dnbinom`, `pnbinom`, and `qnbinom` give the p.f., the c.d.f., and the quantile function of negative binomial distributions. The syntax is that the first argument is the argument of the function, and the next two are r and p in the notation of the text. The function `rnbinom` gives a random sample of binomial random variables. The first argument is how many you want, and the next two are r and p. All of the solutions that require the calculation of negative binomial probabilites can be done using these functions. There are also functions `dgeom`, `pgeom`, `qgeom`, and `rgeom` that compute similar features of geometric distributions. Just remove the "r" argument.

Solutions to Exercises

1. (a) Two particular days in a row have independent draws, and each draw has probability 0.01 of producing triples. So, the probability that two particular days in a row will both have triples is 10^{-4}.

   (b) Since a particular day and the next day are independent, the conditional probability of triples on the next day is 0.01 conditional on whatever happens on the first day.

2. (a) The number of tails will have the negative binomial distribution with parameters $r = 5$ and $p = 1/30$. By Eq. (5.5.7),

   $$E(X) = \frac{r(1-p)}{p} = 5(29) = 145.$$

   (b) By Eq. (5.5.7), $\text{Var}(X) = \dfrac{r(1-p)}{p^2} = 4350.$

3. (a) Let X denote the number of tails that are obtained before five heads are obtained, and let Y denote the total number of tosses that are required. Then $Y = X + 5$. Therefore, $E(Y) = E(X) + 5$. It follows from Exercise 2(a) that $E(Y) = 150$.

   (b) Suppose $Y = X + 5$, then $\text{Var}(Y) = \text{Var}(X)$. Therefore, it follows from Exercise 2(b) that $\text{Var}(Y) = 4350$.

4. (a) The number of failures X_A obtained by player A before he obtains r successes will have the negative binomial distribution with parameters r and p. The total number of throws required by player A will be $Y_A = X_A + r$. Therefore,

$$E(Y_A) = E(X_A) + r = r\frac{1-p}{p} + r = \frac{r}{p}.$$

The number of failures X_B obtained by player B before he obtains mr successes will have the negative binomial distribution with parameters mr and mp. The total number of throws required by player B will be $Y_B = X_B + mr$. Therefore,

$$E(Y_B) = E(X_B) + mr = (mr)\frac{(1-mp)}{mp} + mr = \frac{r}{p}.$$

(b)

$$\begin{aligned}\text{Var}(Y_A) &= \text{Var}(X_A) = \frac{r(1-p)}{p^2} = \frac{r}{p^2}(1-p) \text{ and}\\ \text{Var}(Y_B) &= \text{Var}(X_B) = \frac{(mr)(1-mp)}{(mp)^2} = \frac{r}{p^2}\left(\frac{1}{m} - p\right).\end{aligned}$$

Therefore, $\text{Var}(Y_B) < \text{Var}(Y_A)$.

5. By Eq. (5.5.6), the m.g.f. of X_i is

$$\psi_i(t) = \left(\frac{p}{1-(1-p)\exp(t)}\right)^{r_i} \quad \text{for} \quad t < \log\left(\frac{1}{1-p}\right).$$

Therefore, the m.g.f. of $X_1 + \cdots + X_k$ is

$$\psi(t) = \prod_{i=1}^{k} \psi_i(t) = \left(\frac{p}{1-(1-p)\exp(t)}\right)^{r_1+\cdots+r_k} \quad \text{for} \quad t < \log\left(\frac{1}{1-p}\right).$$

Since $\psi(t)$ is the m.g.f. of the negative binomial distribution with parameters $r_1 + \cdots + r_k$ and p, that must be the distribution of $X_1 + \cdots + X_k$.

6. For $x = 0, 1, 2, \ldots,$

$$\Pr(X = x) = p(1-p)^x.$$

If we let $x = 2i$, then as i runs through all the integers $0, 1, 2, \ldots,$ the value of $2i$ will run through all the even integers $0, 2, 4, \ldots$. Therefore,

$$\Pr(X \text{ is an even integer}) = \sum_{i=0}^{\infty} p(1-p)^{2i} = p\sum_{i=0}^{\infty}([1-p]^2)^i = p\frac{1}{1-(1-p)^2}.$$

7. $\Pr(X \geq k) = \sum_{x=j}^{\infty} p(1-p)^x = p(1-p)^k \sum_{x=j}^{\infty}(1-p)^{x-k}$. If we let $i = x - k$, then

$$\Pr(X \geq k) = p(1-p)^k \sum_{x=j}^{\infty}(1-p)^i = p(1-p)^k \frac{1}{1-[1-p]} = (1-p)^k.$$

8. $\Pr(X = k + t \mid X \geq k) = \dfrac{\Pr(X = k + t \text{ and } X \geq k)}{\Pr(X \geq k)} = \dfrac{\Pr(X = k + t)}{\Pr(X \geq k)}$.

By Eq. (5.5.3), $\Pr(X = k + t) = p(1 - p)^{k+t}$. By Exercise 7, $\Pr(X \geq k) = (1 - p)^k$. Therefore, $\Pr(X = k + t \mid X \geq k) = p(1 - p)^t = \Pr(X = t)$.

9. Since the components are connected in series, the system will function properly only as long as every component functions properly. Let X_i denote the number of periods that component i functions properly, for $i = 1, \ldots, n$, and let X denote the number of periods that system functions properly. Then for any nonnegative integer x,

$$\Pr(X \geq x) = \Pr(X_1 \geq x, \ldots, X_n \geq x) = \Pr(X_1 \geq x) \ldots \Pr(X_n \geq x),$$

because the n components are independent. By Exercise 7,

$$\Pr(X_i \geq x) = (1 - p_i)^x = (1 - p_i)^x .$$

Therefore, $\Pr(X \geq x) = \prod_{i=1}^n (1 - p_i)^x$. It follows that

$$\begin{aligned}
\Pr(X = x) &= \Pr(X \geq x) - \Pr(X \geq x + 1) = \prod_{i=1}^n (1 - p_i)^x - \prod_{i=1} (1 - p_i)^{x+1} \\
&= \left(1 - \prod_{i=1}^n (1 - p_i)\right) \left(\prod_{i=1}^n (1 - p_i)\right)^x .
\end{aligned}$$

It can be seen that this is the p.f. of the geometric distribution with $p = 1 - \prod_{i=1}^n (1 - p_i)$.

10. By the assumptions of the problem we have $p = 1 - \lambda/r$ and $r \to \infty$. To simplify some of the formulas, let $q = 1 - p$ so that $q = \lambda/r$. It now follows from Eq. (5.5.1) that

$$\begin{aligned}
f(x \mid r, p) &= \frac{(r + x - 1)(r + x - 2) \ldots r}{x!} \left(1 - \frac{\lambda}{r}\right)^r q^x \\
&= \frac{[q(r + x - 1)][q(r + x - 2)] \ldots (qr)}{x!} \left(1 - \frac{\lambda}{r}\right)^r \\
&= \frac{[\lambda + q(x - 1)][\lambda + q(x - 2)] \ldots (\lambda)}{x!} \left(1 - \frac{\lambda}{r}\right)^r .
\end{aligned}$$

As $r \to \infty$, $q \to 0$ and

$$\begin{aligned}
[\lambda + q(x - 1)][\lambda + q(x - 2)] \ldots (\lambda) &\to \lambda^x, \text{ and} \\
\left(1 - \frac{\lambda}{r}\right)^r &\to \exp(-\lambda).
\end{aligned}$$

Hence, $f(x \mid r, p) \to \dfrac{\lambda^x}{x!} \exp(-\lambda) = f(x \mid \lambda)$.

11. According to Exercise 10 in Sec. 5.3,

$$\binom{-r}{x} = (-1)^x \binom{r + x - 1}{x}.$$

This makes

$$\binom{-r}{x}p^r(-[1-p])^x = \binom{r+x-1}{x}p^r(1-p)^x,$$

which is the proper form of the negative binomial p.f. for $x = 0, 1, 2, \ldots$.

12. The joint p.f./p.d.f. of X and P is $f(p)$ times the geometric p.f. with parameter p, that is

$$p(1-p)^x 10(1-p)^9 = p(1-p)^{x+9}, \text{ for } x = 0, 1, \ldots \text{ and } 0 < p < 1. \quad \text{(S.5.7)}$$

The marginal p.f. of X at $x = 12$ is the integral of (S.5.7) over p with $x = 12$ substituted:

$$\int_0^1 p(1-p)^{21}dp = \int_0^1 p^{21}(1-p)dp = \frac{1}{22} - \frac{1}{23} = \frac{1}{506}.$$

The conditional p.d.f. of P given $X = 12$ is (S.5.7) divided by this last value

$$g(p|12) = 506p(1-p)^{21}, \text{ for } 0 < p < 1.$$

13. (a) The memoryless property says that, for all $k, t \geq 0$,

$$\frac{\Pr(X = k+t)}{1 - F(t-1)} = \Pr(X = k).$$

(The above version switches the use of k and t from Theorem 5.5.5.) If we sum both sides of this over $k = h, h+1, \ldots$, we get

$$\frac{1 - F(t+h-1)}{1 - F(t-1)} = 1 - F(h-1).$$

(b) $\ell(t+h) = \log[1 - F(t+h-1)]$. From part (a), we have

$$1 - F(t+h-1) = [1 - F(t-1)][1 - F(h-1)],$$

Hence

$$\ell(t+h) = \log([1 - F(t-1)] + \log[1 - F(h-1)] = \ell(t) + \ell(h).$$

(c) We prove this by induction. Clearly $\ell(1) = 1 \times \ell(1)$, so the result holds for $t = 1$. Assume that the result holds for all $t \leq t_0$. Then $\ell(t_0 + 1) = \ell(t_0) + \ell(1)$ by part (b). By the induction hypothesis, $\ell(t_0) = t_0\ell(1)$, hence $\ell(t_0 + 1) = (t_0 + 1)\ell(1)$, and the result holds for $t = t_0 + 1$.

(d) Since $\ell(1) = \log[1 - F(0)]$, we have $\ell(1) < 0$. Let $p = 1 - \exp[\ell(1)]$, which between 0 and 1. For every integer $x \geq 1$, we have, from part (c) and the definition of ℓ, that

$$F(x-1) = 1 - \exp[\ell(x)] = 1 - \exp[x\ell(1)] = 1 - (1-p)^x.$$

Setting $t = x - 1$ for $x \geq 1$, we get

$$F(t) = 1 - (1-p)^{t+1}, \text{ for } t = 0, 1, \ldots. \quad \text{(S.5.8)}$$

It is easy to verify that (S.5.8) is the c.d.f. of the geometric distribution with parameter p.

5.6 The Normal Distributions

Commentary

In addition to introducing the family of normal distributions, we also describe the family of lognormal distributions. These distributions arise frequently in engineering and financial applications. (Examples 5.6.9 and 5.6.10 give two such cases.) It is true that lognormal distributions are nothing more than simple transformations of normal distributions. However, at this point in their study, many students will not yet be sufficiently comfortable with transformations to be able to derive these distributions and their properties without a little help.

If one is using the statistical software *R*, then the functions `dnorm`, `pnorm`, and `qnorm` give the p.d.f., the c.d.f., and the quantile function of normal distributions. The syntax is that the first argument is the argument of the function, and the next two are the mean and standard deviation. The function `rnorm` gives a random sample of normal random variables. The first argument is how many you want, and the next two are the mean and standard deviation. All of the solutions that require the calculation of normal probabilites and quantiles can be done using these functions instead of tables. There are also functions `dlnorm`, `plnorm`, `qlnorm`, and `rlnorm` that compute similar features for lognormal distributions.

Solutions to Exercises

1. By the symmetry of the standard normal distribution around 0, the 0.5 quantile must be 0. The 0.75 quantile is found by locating 0.75 in the $\Phi(x)$ column of the standard normal table and interpolating in the x column. We find $\Phi(0.67) = 0.7486$ and $\Phi(0.68) = 0.7517$. Interpolating gives the 0.75 quantile as 0.6745. By symmetry, the 0.25 quantile is -0.6745. Similarly we find the 0.9 quantile by interpolation using $\Phi(1.28) = 0.8997$ and $\Phi(1.29) = 0.9015$. The 0.9 quantile is then 1.282 and the 0.1 quantile is -1.282.

2. Let $Z = (X - 1)/2$. Then Z has the standard normal distribution.

   (a) $\Pr(X \le 3) = \Pr(Z \le 1) = \Phi(1) = 0.8413$

   (b) $\Pr(X > 1.5) = \Pr(Z > 0.25) = 1 - \Phi(0.25) = 0.4013$.

   (c) $\Pr(X = 1) = 0$, because X has a continuous distribution.

   (d) $\Pr(2 < X < 5) = \Pr(0.5 < Z < 2) = \Phi(2) - \Phi(0.5) = 0.2858$.

   (e) $\Pr(X \ge 0) = \Pr(Z \ge -0.5) = \Pr(Z \le 0.5) = \Phi(0.5) = 0.6915$.

   (f) $\Pr(-1 < X < 0.5) = \Pr(-1 < Z < -0.25) = \Pr(0.25 < Z < 1) = \Phi(1) - \Phi(0.25) = 0.2426$.

   (g)

$$\begin{aligned}\Pr(\,|X|\, \le 2) &= \Pr(-2 \le X \le 2) = \Pr(-1.5 \le Z \le 0.5)\\ &= \Pr(Z \le 0.5) - \Pr(Z \le -1.5) = \Pr(Z \le 0.5)\\ &\quad -\Pr(Z \ge 1.5) = \Phi(0.5) - [1 - \Phi(1.5)] = 0.6247.\end{aligned}$$

   (h)

$$\begin{aligned}\Pr(1 \le -2X + 3 \le 8) &= \Pr(-2 \le -2X \le 5) = \Pr(-2.5 \le X \le 1)\\ &= \Pr(-1.75 \le Z \le 0) = \Pr(0 \le Z \le 1.75)\\ &= \Phi(1.75) - \Phi(0) = 0.4599.\end{aligned}$$

3. If X denotes the temperature in degrees Fahrenheit and Y denotes the temperature in degrees Celsius, then $Y = 5(X - 32)/9$. Since Y is a linear function of X, then Y will also have a normal distribution. Also,

$$E(Y) = \frac{5}{9}(68 - 32) = 20 \quad \text{and} \quad \text{Var}(Y) = \left(\frac{5}{9}\right)^2 (16) = \frac{400}{81}.$$

4. The q quantile of the temperature in degrees Fahrenheit is $68 + 4\Phi^{-1}(q)$. Using Exercise 1, we have $\Phi^{-1}(0.75) = 0.6745$ and $\Phi^{-1}(0.25) = -0.6745$. So, the 0.25 quantile is 65.302, and the 0.75 quantile is 70.698.

5. Let A_i be the event that chip i lasts at most 290 hours. We want the probability of $\cup_{i=1}^3 A_i^c$, whose probability is

$$1 - \Pr\left(\cap_{i=1}^3 A_i\right) = 1 - \prod_{i=1}^3 \Pr(A_i).$$

Since the lifetime of each chip has the normal distribution with mean 300 and standard deviation 10, each A_i has probability

$$\Phi([290 - 300]/10) = \Phi(-1) = 1 - 0.8413 = 0.1587.$$

So the probability we want is $1 - 0.1587^3 = 0.9960$.

6. By comparing the given m.g.f. with the m.g.f. of a normal distribution presented in Eq. (5.6.5), we can see that, for the given m.g.f., $\mu = 0$ and $\sigma^2 = 2$.

7. If X is a measurement having the specified normal distribution, and if $Z = (X - 120)/2$, then Z will have the standard normal distribution. Therefore, the probability that a particular measurement will lie in the given interval is

$$p = \Pr(116 < X < 118) = \Pr(-2 < Z < -1) = \Pr(1 < Z < 2) = \Phi(2) - \Phi(1) = 0.1360.$$

The probability that all three measurements will lie in the interval is p^3.

8. Except for a constant factor, this integrand has the form of the p.d.f. of a normal distribution for which $\mu = 0$ and $\sigma^2 = 1/6$. Therefore, if we multiply the integrand by

$$\frac{1}{(2\pi)^{1/2}\sigma} = \left(\frac{3}{\pi}\right)^{1/2},$$

we obtain the p.d.f. of a normal distribution and we know that the integral of this p.d.f. over the entire real line must be equal to 1. Therefore,

$$\int_{-\infty}^{\infty} \exp(-3x^2)dx = \left(\frac{\pi}{3}\right)^{1/2}.$$

Finally, since the integrand is symmetric with respect to $x = 0$, the integral over the positive half of the real line must be equal to the integral over the negative half of the real line. Hence,

$$\int_0^{\infty} \exp(-3x^2)dx = \frac{1}{2}\left(\frac{\pi}{3}\right)^{1/2}.$$

9. The total length of the rod is $X = A + B + C - 4$. Since X is a linear combination of A, B, and C, it will also have the normal distribution with

$$E(X) = 20 + 14 + 26 - 4 = 56$$

and $\text{Var}(X) = 0.04 + 0.01 + 0.04 = 0.09$. If we let $Z = (X - 56)/0.3$, then Z will have the standard normal distribution. Hence,

$$\Pr(55.7 < X < 56.3) = \Pr(-1 \le Z \le 1) = 2\Phi(1) - 1 = 0.6827.$$

10. We know that $E(\overline{X}_n) = \mu$ and $\text{Var}(\overline{X}_n) = \sigma^2/n = 4/25$. Hence, if we let $Z = (\overline{X}_n - \mu)/(2/5) = (5/2)(\overline{X}_n - \mu)$, then Z will have the standard normal distribution. Hence,

$$\Pr(\,|\overline{X}_n - \mu|\, \le 1) = \Pr(\,|Z|\, \le 2.5) = 2\Phi(2.5) - 1 = 0.9876.$$

11. If we let $Z = \sqrt{n}(\overline{X}_n - \mu)/2$, then Z will have the standard normal distribution. Therefore,

$$\Pr(\,|\overline{X}_n - \mu|\, < 0.1) = \Pr(\,|Z|\, < 0.05\sqrt{n}) = 2\Phi(0.05\sqrt{n}) - 1.$$

This value will be at least 0.9 if $2\Phi(0.05\sqrt{n}) - 1 \ge 0.9$ or $\Phi(0.05\sqrt{n}) \ge 0.95$. It is found from a table of the values of Φ that we must therefore have $0.05\sqrt{n} \ge 1.645$. The smallest integer n which satisfies this inequality is $n = 1083$.

12. (a) The general shape is as shown in Fig. S.5.1.

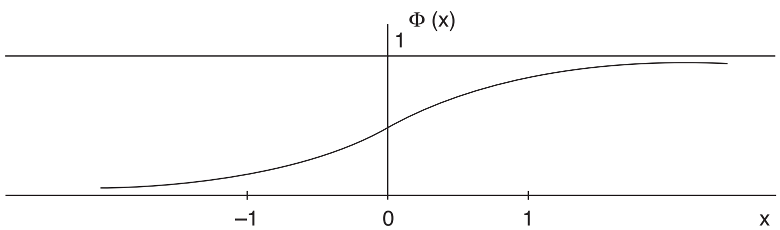

Figure S.5.1: Figure for Exercise 12 of Sec. 5.6.

(b) The sketch remains the same with the scale changed on the x-axis so that the points -1 and 0 become -5 and -2, respectively. It turns out that the point $x = 1$ remains fixed in this transformation.

13. Let X denote the diameter of the bolt and let Y denote the diameter of the nut. The $Y - X$ will have the normal distribution for which

$$E(Y - X) = 2.02 - 2 = 0.02$$

and

$$\text{Var}(Y - X) = 0.0016 + 0.0009 = 0.0025.$$

If we let $Z = (Y - X - 0.02)/0.05$, then Z will have the standard normal distribution. Therefore,

$$\Pr(0 < Y - X \le 0.05) = \Pr(-0.4 < Z \le 0.6) = \Phi(0.6) - [1 - \Phi(0.4)] = 0.3812.$$

14. Let $\overline{X}$ denote the average of the two scores from university A and let $\overline{Y}$ denote the average of the three scores from university B. Then $\overline{X}$ has the normal distribution for which

$$E(\overline{X}) = 625 \quad \text{and} \quad \text{Var}(\overline{X}) = \frac{100}{2} = 50.$$

Also, $\overline{Y}$ has the normal distribution for which

$$E(\overline{Y}) = 600 \quad \text{and} \quad \text{Var}(\overline{Y}) = \frac{150}{3} = 50.$$

Therefore $\overline{X} - \overline{Y}$ has the normal distribution for which

$$E(\overline{X} - \overline{Y}) = 625 - 600 = 25 \quad \text{and} \quad \text{Var}(\overline{X} - \overline{Y}) = 50 + 50 = 100.$$

It follows that if we let $Z = (\overline{X} - \overline{Y} - 25)/10$, then Z will have the standard normal distribution. Hence,

$$\Pr(\overline{X} - \overline{Y} > 0) = \Pr(Z > -2.5) = \Pr(Z < 2.5) = \Phi(2.5) = 0.9938.$$

15. Let $f_1(x)$ denote the p.d.f. of X if the person has glaucoma and let $f_2(x)$ denote the p.d.f. of X if the person does not have glaucoma. Furthermore, let A_1 denote the event that the person has glaucoma and let $A_2 = A_1^C$ denote the event that the person does not have glaucoma. Then

$$\begin{aligned}
\Pr(A_1) &= 0.1, \quad \Pr(A_2) = 0.9, \\
f_1(x) &= \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(x-25)^2\right\} \quad \text{for} \quad -\infty < x < \infty, \\
f_2(x) &= \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(x-20)^2\right\} \quad \text{for} \quad -\infty < x < \infty.
\end{aligned}$$

(a) $$\Pr(A_1 \mid X = x) = \frac{\Pr(A_1)f_1(x)}{\Pr(A_1)f_1(x) + \Pr(A_2)f_2(x)}$$

(b) The value found in part (a) will be greater than 1/2 if and only if

$$\Pr(A_1)f_1(x) > \Pr(A_2)f_2(x).$$

All of the following inequalities are equivalent to this one:
(i) $\exp\{-(x-25)^2/2\} > 9\exp\{-(x-20)^2/2\}$
(ii) $-(x-25)^2/2 > \log 9 - (x-20)^2/2$
(iii) $(x-20)^2 - (x-25)^2 > 2\log 9$
(iv) $10x - 225 > 2\log 9$
(v) $x > 22.5 + \log(9)/5$.

16. The given joint p.d.f. is the joint p.d.f. of two random variables that are independent and each of which has the standard normal distribution. Therefore, $X + Y$ has the normal distribution for which

$$E(X+Y) = 0 + 0 = 0 \quad \text{and} \quad \text{Var}(X+Y) = 1 + 1 = 2.$$

If we let $Z = (X+Y)/\sqrt{2}$, then Z will have the standard normal distribution. Hence,

$$\begin{aligned}
\Pr(-\sqrt{2} < X + Y < 2\sqrt{2}) &= \Pr(-1 < Z < 2) \\
&= \Pr(Z < 2) - \Pr(Z < -1) \\
&= \Phi(2) - [1 - \Phi(1)] = 0.8186.
\end{aligned}$$

17. If $Y = \log X$, then the p.d.f. of Y is

$$g(y) = \frac{1}{(2\pi)^{1/2}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \quad \text{for} \quad -\infty < y < \infty.$$

Since $\dfrac{dy}{dx} = \dfrac{1}{x}$, it now follows that the p.d.f. of X, for $x > 0$, is $f(x) = g(\log x)/x$.

18. Let $U = X/Y$ and, as a convenient device, let $V = Y$. If we exclude the possibility that $Y = 0$, the transformation from X and Y to U and V is then one-to-one. (Since $\Pr(Y = 0) = 0$, we can exclude this possibility.) The inverse transformation is

$$X = UV \quad \text{and} \quad Y = V.$$

Hence, the Jacobian is

$$J = \det\begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial u} \end{bmatrix} = \det\begin{bmatrix} v & u \\ 0 & 1 \end{bmatrix} = v.$$

Since X and Y are independent and each has the standard normal distribution, their joint p.d.f. $f(x, y)$ is as given in Exercise 16. Therefore, the joint p.d.f. $g(u, v)$ of U and V will be

$$g(u,v) = f(uv, v)\,|\,v\,| = \frac{|\,v\,|}{2\pi}\exp\left\{-\frac{1}{2}(u^2+1)v^2\right\}.$$

To find the marginal p.d.f. $g_1(u)$ of U, we can now integrate $g(u, v)$ over all values of v. (The fact that the single point $v = 0$ was excluded does not affect the value of the integral over the entire real line.) We have

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{\infty}\frac{|\,v\,|}{2\pi}\exp\left\{-\frac{1}{2}(u^2+1)v^2\right\}dv \\ &= \int_0^{\infty}\frac{1}{\pi}\exp\left\{-\frac{1}{2}(u^2+1)v^2\right\}vdv \\ &= \frac{1}{\pi(u^2+1)} \quad \text{for} \quad -\infty < u < \infty. \end{aligned}$$

It can now be seen that $g_1(u)$ is the p.d.f. of a Cauchy distribution as defined in Eq. (4.1.7).

19. The conditional p.d.f. of X given μ is

$$g_1(x|\mu) = \frac{1}{(2\pi)^{1/2}}\exp(-(x-\mu)^2/2),$$

while the marginal p.d.f. of μ is $f_2(\mu) = 0.1$ for $5 \le \mu \le 15$. We need the marginal p.d.f. of X, which we get by integrating μ out of the joint p.d.f.

$$g_1(x|\mu)f_2(\mu) = \frac{0.1}{(2\pi)^{1/2}}\exp(-(x-\mu)^2/2), \text{ for } 5 \le \mu \le 15.$$

The integral is

$$f_1(x) = \int_5^{15} \frac{0.1}{(2\pi)^{1/2}} \exp(-(x-\mu)^2/2)d\mu = 0.1[\Phi(15-x) - \Phi(5-x)].$$

With $x = 8$, the value is $0.1[\Phi(7) - \Phi(-3)] = 0.0999$. This makes the conditional p.d.f. of μ given $X = 8$

$$g_2(\mu|8) = \frac{1.0013}{(2\pi)^{1/2}} \exp(-(8-\mu)^2/2), \text{ for } 5 \le \mu \le 15.$$

20. This probability is the probability that $\log(X) \le \log(6.05) = 1.80$, which equals

$$\Phi([1.80 - 3]/1.44^{1/2}) = \Phi(-1) = 0.1587.$$

21. Note that $\log(XY) = \log(X) + \log(Y)$. Since X and Y are independent with normal distributions, we have that $\log(XY)$ has the normal distribution with the sum of the means (4.6) and sum of the variances (10.5). This means that XY has the lognormal distribution with parameters 4.6 and 10.5.

22. Since $\log(1/X) = -\log(X)$, we know that $-\log(X)$ has the normal distribution with mean $-\mu$ and variance σ^2. This mean that $1/X$ has the lognormal distribution with parameters $-\mu$ and σ^2.

23. Since $\log(3X^{1/2}) = \log(3) + \log(X)/2$, we know that $\log(3X^{1/2})$ has the normal distribution with mean $\log(3) + 4.1/2 = 3.149$ and variance $8/4 = 2$. This means that $3X^{1/2}$ has the lognormal distribution with parameters 3.149 and 2.

24. First expand the left side of the equation to get

$$\sum_{i=1}^n a_i(x-b_i)^2 + cx = cx + \sum_{i=1}^n [a_i x^2 - 2a_i b_i x + b_i^2]. \tag{S.5.9}$$

Now collect all the squared and linear terms in x. The coefficient of x^2 is $\sum_{i=1}^n a_i$. The coefficient of x is $c - 2\sum_{i=1}^n a_i b_i$. The constant term is $\sum_{i=1}^n a_i b_i^2$. This makes (S.5.9) equal to

$$x^2 \sum_{i=1}^n a_i + x\left[c - 2\sum_{i=1}^n a_i b_i\right] + \sum_{i=1}^n a_i b_i^2. \tag{S.5.10}$$

Next, expand each term on the right side of the original equation to produce

$$\left(\sum_{i=1}^n a_i\right)\left[x^2 - 2x\frac{\sum_{i=1}^n a_i b_i - c/2}{\sum_{i=1}^n a_i}\right] + \frac{\left(\sum_{i=1}^n a_i b_i\right)^2 - c\sum_{i=1}^n a_i b_i + c^2/4}{\sum_{i=1}^n a_i}$$
$$+\sum_{i=1}^n a_i b_i^2 - \frac{\left(\sum_{i=1}^n a_i b_i\right)^2}{\sum_{i=1}^n a_i} + \frac{c\sum_{i=1}^n a_i b_i - c^2/4}{\sum_{i=1}^n a_i}.$$

Combining like terms in this expression produces the same terms that are in (S.5.10).

25. Divide the time interval of u years into n intervals of length u/n each. At the end of n such intervals, the principal gets multiplied by $(1 + ru/n)^n$. The limit of this as $n \to \infty$ is $\exp(ru)$.

26. The integral that defines the mean is
$$E(X) = \int_{-\infty}^{\infty} \frac{x}{(2\pi)^{1/2}} \exp\left[-\frac{x^2}{2}\right] dx.$$
The integrand is a function f with the property that $f(-x) = -f(x)$. Since the range of integration is symmetric around 0, the integral is 0. The integral that defines the variance is then
$$\mathrm{Var}(X) = E(X^2) = \int_{-\infty}^{\infty} \frac{x^2}{(2\pi)^{1/2}} \exp\left[-\frac{x^2}{2}\right] dx.$$
In this integral, let $u = x$ and
$$dv = \frac{x}{(2\pi)^{1/2}} \exp\left[-\frac{x^2}{2}\right] dx.$$
It is easy to see that $du = dx$ and
$$v = -\frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{x^2}{2}\right].$$
Integration by parts yields
$$\mathrm{Var}(X) = -\frac{x}{(2\pi)^{1/2}} \exp\left[-\frac{x^2}{2}\right]\Bigg|_{x=-\infty}^{\infty} + \int \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{x^2}{2}\right] dx.$$
The term on the right above equals 0 at both ∞ and $-\infty$. The remaining integral is 1 because it is the integral of the standard normal p.d.f. So $\mathrm{Var}(X) = 1$.

5.7 The Gamma Distributions

Commentary

Gamma distributions are used in the derivation of the chi-square distribution in Sec. 8.2 and as conjugate prior distributions for various parameters. The gamma function arises in several integrals later in the text and is interesting in its own right as a generalization of the factorial function to noninteger arguments.

If one is using the statistical software *R*, then the function `gamma` computes the gamma function, and `lgamma` computes the logarithm of the gamma function. They take only one argument. The functions `dgamma`, `pgamma`, and `qgamma` give the p.d.f., the c.d.f., and the quantile function of gamma distributions. The syntax is that the first argument is the argument of the function, and the next two are α and β in the notation of the text. The function `rgamma` gives a random sample of gamma random variables. The first argument is how many you want, and the next two are α and β. All of the solutions that require the calculation of gamma probabilites and quantiles can be done using these functions. There are also functions `dexp`, `pexp`, `qexp`, and `rexp` that compute similar features for exponential distributions. Just remove the "α" parameter.

Solutions to Exercises

1. Let $f(x)$ denote the p.d.f. of X and let $Y = cX$. Then $X = Y/c$. Since $dx = dy/c$, then for $x > 0$,
$$g(y) = \frac{1}{c} f\left(\frac{y}{c}\right) = \frac{1}{c}\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{y}{c}\right)^{\alpha-1} \exp(-\beta(y/c)) = \frac{(\beta/c)^\alpha}{\Gamma(\alpha)} y^{\alpha-i} \exp(-(\beta/c)y).$$

2. The c.d.f. of the exponential distribution with parameter β is $F(x) = 1 - \exp(-\beta x)$ for $x > 0$. The inverse of this is the quantile function $F^{-1}(p) = -\log(1-p)/\beta$.

3. The three p.d.f.'s are in Fig. S.5.2.

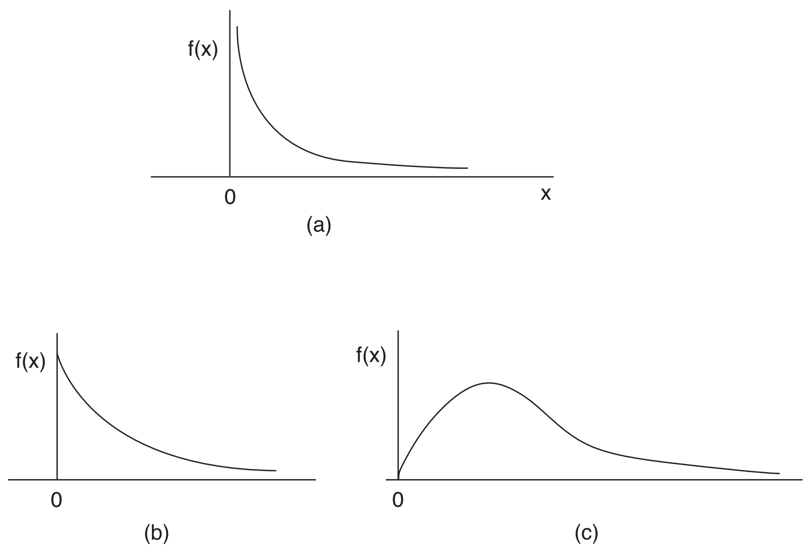

Figure S.5.2: Figure for Exercise 3 of Sec. 5.7.

4.

$$
\begin{aligned}
f(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x) && \text{for } x > 0. \\
f'(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} (\alpha - 1 - \beta x) x^{\alpha-2}\exp(-\beta x) && \text{for } x > 0.
\end{aligned}
$$

If $\alpha \le 1$, then $f'(x) < 0$ for $x > 0$. Therefore, the maximum value of $f(x)$ occurs at $x = 0$. If $\alpha > 1$, then $f'(x) = 0$ for $x = (\alpha - 1)/\beta$ and it can be verified that f(x) is actually a maximum at this value of x.

5. All three p.d.f.'s are in Fig. S.5.3.

6. Each X_i has the gamma distribution with parameters 1 and β. Therefore, by Theorem 5.7.7, the sum $\sum_{i=1}^{n} X_i$ has the gamma distribution with parameters n and β. Finally, by Exercise 1, $\overline{X}_n = \sum_{i=1}^{n} X_i/n$ has the gamma distribution with parameters n and $n\beta$.

7. Let $A_i = \{X_i > t\}$ for $i = 1, 2, 3$. The event that at least one X_i is greater than t is $\bigcup_{i=1}^3 A_i$. We could use the formula in Theorem 1.10.1, or we could use that $\Pr(\bigcup_{i=1}^3 A_i) = 1 - \Pr(\bigcap_{i=1}^3 A_i^c)$. The latter is easier because the X_i are mutually independent and identically distributed.

$$
\Pr\left(\bigcap_{i=1}^{3} A_i^c\right) = \Pr(A_1^c)^3 = [1 - \exp(-\beta t)]^3.
$$

So, the probability we want is $1 - [1 - \exp(-\beta t)]^3$.

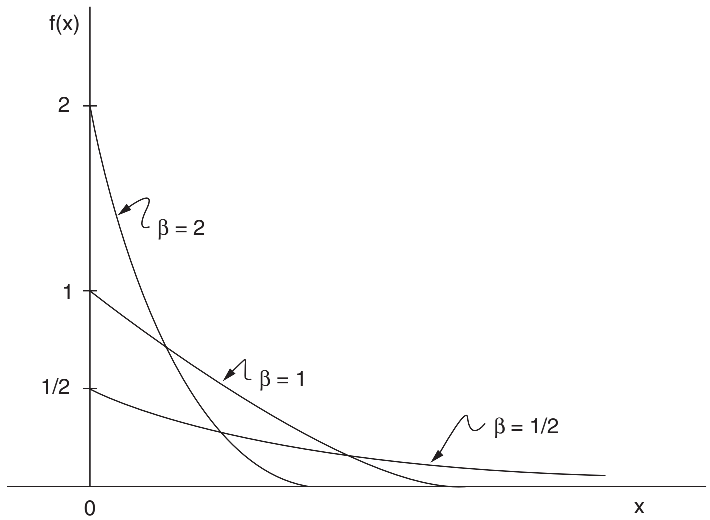

Figure S.5.3: Figure for Exercise 5 of Sec. 5.7.

8. For any number $y > 0$,

$$\begin{aligned}\Pr(Y > y) &= \Pr(X_1 > y, \ldots, X_k > y) = \Pr(X_1 > y) \ldots \Pr(X_k > y) \\ &= \exp(-\beta_1 y) \ldots \exp(-\beta_k y) = \exp(-(\beta_1 + \cdots + \beta_k)y),\end{aligned}$$

which is the probability that an exponential random variable with parameter $\beta_1 + \cdots + \beta_k$ is greater than y. Hence, Y has that exponential distribution.

9. Let Y denote the length of life of the system. Then by Exercise 8, Y has the exponential distribution with parameter $0.001 + 0.003 + 0.006 = 0.01$. Therefore,

$$\Pr(Y > 100) = \exp(-100\ (0.01)) = \frac{1}{e}.$$

10. Since the mean of the exponential distribution is μ, the parameter is $\beta = 1/\mu$. Therefore, the distribution of the time until the system fails is an exponential distribution with parameter $n\beta = n/\mu$. The mean of this distribution is $1/(n\beta) = \mu/n$ and the variance is $1/(n\beta)^2 = (\mu/n)^2$.

11. The length of time Y_1 until one component fails has the exponential distribution with parameter nβ. Therefore, $E(Y_1) = 1/(n\beta)$. The additional length of time Y_2 until a second component fails has the exponential distribution with parameter $(n-1)\beta$. Therefore, $E(Y_2) = 1/[(n-1)\beta]$. Similarly, $E(Y_3) = 1/[(n-2)\beta]$. The total time until three components fail is $Y_1 + Y_2 + Y_3$ and $E(Y_1 + Y_2 + Y_3) = \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2}\right)\frac{1}{\beta}$.

12. The length of time until the system fails will be $Y_1 + Y_2$, where these variables were defined in Exercise 11. Therefore, $E(Y_1 + Y_2) = \frac{1}{n\beta} + \frac{1}{(n-1)\beta} = \left(\frac{1}{n} + \frac{1}{n-1}\right)\mu$. Also, the variables Y_1 and Y_2 are independent, because the distribution of Y_2 is always the same exponential distribution regardless of the value of Y_1. Therefore,

$$\text{Var}(Y_1 + Y_2) = \text{Var}(Y_1) + \text{Var}(Y_2) = \frac{1}{(n\beta)^2} + \frac{1}{[(n-1)\beta]^2} = \left[\frac{1}{n^2} + \frac{1}{(n-1)^2}\right]\mu^2.$$

13. The time Y_1 until one of the students completes the examination has the exponential distribution with parameter $5\beta = 5/80 = 1/16$. Therefore,

$$\Pr(Y_1 < 40) = 1 - \exp(-40/16) = 1 - \exp(-5/2) = 0.9179.$$

14. The time Y_2 after one students completes the examination until a second student completes it has the exponential distribution with parameter $4\beta = 4/80 = 1/20$. Therefore,

$$\Pr(Y_2 < 35) = 1 - \exp(-35/20) = 1 - \exp(-7/4) = 0.8262.$$

15. No matter when the first student completes the examination, the second student to complete the examination will do so at least 10 minutes later than the first student if $Y_2 > 10$. Similarly, the third student to complete the examination will do so at least 10 minutes later than the second student if $Y_3 > 10$. Furthermore, the variables Y_1, Y_2, Y_3, Y_4, and Y_5 are independent. Therefore, the probability that no two students will complete the examination within 10 minutes of each other is

$$\begin{aligned}
\Pr(Y_2 > 10, \ldots, Y_5 > 10) &= \Pr(Y_2 > 10) \ldots \Pr(Y_5 > 10) \\
&= \exp(-(10)4\beta)\exp(-(10)3\beta)\exp(-(10)2\beta)\exp(-10\beta) \\
&= \exp(-40/80)\exp(-30/80)\exp(-20/80)\exp(-10/80) \\
&= \exp(-5/4) = 0.2865.
\end{aligned}$$

16. If $Y = \log(X/x_0)$, then $X = x_0 \exp(Y)$. Also, $dx = x_0 \exp(y)dy$ and $x > x_0$ if and only if $y > 0$. Therefore, for $y > 0$,

$$g(y) = f(x_0 \exp(y)|x_0, \alpha)x_0 \exp(y) = \alpha e y^{-\alpha y}.$$

17.

$$\begin{aligned}
E[(X-\mu)^{2n}] &= \int_{-\infty}^{\infty} (x-\mu)^{2n} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\
&= \frac{2}{(2\pi)^{1/2}\sigma} \int_{\mu}^{\infty} (x-\mu)^{2n} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx.
\end{aligned}$$

Let $y = (x-\mu)^2$. Then $dx = dy/(2y^{1/2})$ and the above integral can be rewritten as

$$\frac{2}{(2\pi)^{1/2}\sigma} \int_0^{\infty} y^n \exp\left\{-\frac{y}{2\sigma^2}\right\} \frac{1}{2y^{1/2}} dy = \frac{1}{(2\pi)^{1/2}\sigma} \int_0^{\infty} y^{n-1/2} \exp\left\{-\frac{y}{2\sigma^2}\right\} dy.$$

The integrand in this integral is the p.d.f. of a gamma distribution with parameters $\alpha = n + 1/2$ and $\beta = 1/(2\sigma^2)$, except for the constant factor

$$\frac{\beta^\alpha}{\Gamma(\alpha)} = \frac{1}{(2\sigma^2)^{n+1/2}\Gamma(n+1/2)}.$$

Since the integral of the p.d.f. of the gamma distribution must be equal to 1, it follows that

$$\int_0^{\infty} y^{n-1/2} \exp\left\{-\frac{y}{2\sigma^2}\right\} dy = (2\sigma^2)^{n+1/2}\Gamma(n+1/2).$$

From Eqs. (5.7.6) and (5.7.9), $\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\left(\frac{1}{2}\right)\pi^{1/2}$. Therefore,

$$\begin{aligned}
E[(X-\mu)^{2n}] &= \frac{1}{(2\pi)^{1/2}\sigma}(2\sigma^2)^{n+1/2}\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\left(\frac{1}{2}\right)\pi^{1/2} \\
&= 2^n\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\left(\frac{1}{2}\right)\sigma^{2n} \\
&= (2n-1)(2n-3)\ldots(1)\sigma^{2n}.
\end{aligned}$$

18. For the exponential distribution with parameter β,

$$f(x) = \beta\exp(-\beta x)$$

and

$$1 - F(x) = \Pr(X > x) = \exp(-\beta x).$$

Therefore, $h(x) = \beta$ for $x > 0$.

19. Let $Y = X^b$. Then $X = Y^{1/b}$ and $dx = \frac{1}{b}y^{(1-b)/b}dy$. Therefore, for $y > 0$,

$$g(y) = f(y^{1/b}|a,b)\frac{1}{b}y^{(1-b)/b} = \frac{1}{a^b}\exp(-y/a^b).$$

20. If X has the Weibull distribution with parameters a and b, then the c.d.f. of X is

$$F(x) = \int_0^x \frac{b}{a^b}t^{b-1}\exp(-(t/a)^b)dt = [-\exp(-(t/a)^b)]_0^x = 1-\exp(-(x/a)^b).$$

Therefore,

$$h(x) = \frac{b}{a^b}x^{b-1}.$$

If $b > 1$, then $h(x)$ is an increasing function of x for $x > 0$, and if $b < 1$, then $h(x)$ is an decreasing function of x for $x > 0$.

21. (a) The mean of $1/X$ is

$$\int_0^\infty \frac{1}{x}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-2}\exp(-\beta x)dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} = \frac{\beta}{\alpha-1}.$$

(b) The mean of $1/X^2$ is

$$\begin{aligned}
\int_0^\infty \frac{1}{x^2}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)dx &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-3}\exp(-\beta x)dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha-2)}{\beta^{\alpha-2}} \\
&= \frac{\beta^2}{(\alpha-1)(\alpha-2)}.
\end{aligned}$$

This makes the variance of $1/X$ equal to

$$\frac{\beta^2}{(\alpha-1)(\alpha-2)} - \left(\frac{\beta}{\alpha-1}\right) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}.$$

22. The conditional p.d.f. of λ given $X = x$ can be obtained from Bayes' theorem for random variables (Theorem 3.6.4). We know

$$\begin{aligned} g_1(x|\lambda) &= \exp(-\lambda t)\frac{(\lambda t)^x}{x!}, \text{ for } x = 0, 1, \ldots, \\ f_2(\lambda) &= \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\lambda\beta), \text{ for } \lambda > 0. \end{aligned}$$

The marginal p.f. of X is

$$\begin{aligned} f_1(x) &= \frac{t^x\beta^\alpha}{x!\Gamma(\alpha)}\int_0^\infty \lambda^{\alpha+x-1}\exp(-\lambda[\beta+t])d\lambda \\ &= \frac{t^x\beta^\alpha\Gamma(\alpha+x)}{x!\Gamma(\alpha)(\beta+t)^{\alpha+x}}. \end{aligned}$$

So, the conditional p.d.f. of λ given $X = x$ is

$$g_2(\lambda|x) = \frac{(\beta+t)^{\alpha+x}}{\Gamma(\alpha+x)}\lambda^{\alpha+x-1}\exp(-\lambda[\beta+t]), \text{ for } \lambda > 0,$$

which is easily recognized as the p.d.f. of a gamma distribution with parameters $\alpha + x$ and $\beta + t$.

23. The memoryless property means that $\Pr(X > t + h|X > t) = \Pr(X > h)$.

(a) In terms of the c.d.f. the memoryless property means

$$\frac{1 - F(t+h)}{1 - F(t)} = 1 - F(h).$$

(b) From (a) we obtain $[1 - F(h)][1 - F(t)] = [1 - F(t + h)]$. Taking logarithms of both sides yields $\ell(h) + \ell(t) = \ell(t + h)$.

(c) Apply the result in part (b) with h and t both replaced by t/m. We obtain $\ell(2t/m) = 2\ell(t/m)$. Repeat with t replaced by $2t/m$ and $h = t/m$. The result is $\ell(3t/m) = 3\ell(t/m)$. After $k - 1$ such applications, we obtain

$$\ell(kt/m) = k\ell(t/m). \tag{S.5.11}$$

In particular, when $k = m$, we get $\ell(t) = m\ell(t/m)$ or $\ell(t/m) = \ell(t)/m$. Substituting this into (S.5.11) we obtain $\ell(kt/m) = (k/m)\ell(t)$.

(d) Let $c > 0$ and let $c_1, c_2, \ldots$ be a sequence of rational numbers that converges to c. Since ℓ is a continuous function, $\ell(c_n t) \to \ell(ct)$. But $\ell(c_n t) = c_n\ell(t)$ by part (c) since c_n is rational. It follows that $c_n\ell(t) \to \ell(ct)$. But, we know that $c_n\ell(t) \to c\ell(t)$. So, $c\ell(t) = \ell(ct)$.

(e) Apply part (d) with $c = 1/t$ to obtain $\ell(t)/t = \ell(1)$, a constant.

(f) Let $\beta = \ell(1)$. According to part (e), $\ell(t) = \beta t$ for all $t > 0$. Then $\log[1 - F(x)] = \beta x$ for $x > 0$. Solving for $F(x)$ gives $F(x) = 1 - \exp(-\beta x)$, which is the c.d.f. the exponential distribution with parameter $\beta = \ell(1)$.

24. Let ψ be the m.g.f. of W_u. The mean of S_u is

$$E(S_u) = S_0E(\exp(\mu u + W_u)) = S_0\exp(\mu u)\psi(1).$$

(a) Since W_u has the gamma distribution with parameters αu and $\beta > 1$, the m.g.f. is $\psi(t) = (\beta/[\beta - t])^{\alpha u}$. This makes the mean

$$E(S_u) = S_0 \exp(\mu u) \left(\frac{\beta}{\beta - 1}\right)^{\alpha u}.$$

So $\exp(-ru)E(S_u) = S_0$ if and only if

$$\exp((\mu - r)u) \left(\frac{\beta}{\beta - 1}\right)^{\alpha u} = 1.$$

Solving this equation for μ yields

$$\mu = r - \alpha \log\left(\frac{\beta}{\beta - 1}\right).$$

(b) Once again, we use the function

$$h(x) = \begin{cases} x - q & \text{if } x \geq q, \\ 0 & \text{if } x < q. \end{cases}$$

The value of the option at time u is $h(S_u)$. Notice that $S_u \geq q$ if and only if $W_u \geq \log(q/S_0) - \mu u = c$, as defined in the exercise. Then the present value of the option is

$$\begin{aligned}
\exp(-ru)E[h(S_u)] &= \exp(-ru) \int_c^\infty [S_0 \exp(\mu u + w) - q] \frac{\beta^{\alpha u}}{\Gamma(\alpha u)} w^{\alpha u - 1} \exp(-\beta w) dw \\
&= S_0 \exp([\mu - r]u) \frac{\beta^{\alpha u}}{\Gamma(\alpha u)} \int_c^\infty w^{\alpha u - 1} exp(-w[\beta - 1]) dw \\
&\quad - q \exp(-ru) \frac{\beta^{\alpha u}}{\Gamma(\alpha u)} \int_c^\infty w^{\alpha u - 1} \exp(-\beta w) dw \\
&= S_0 \frac{(\beta - 1)^{\alpha u}}{\Gamma(\alpha u)} \int_c^\infty w^{\alpha u - 1} \exp(-(\beta - 1)w) dw - q \exp(-ru) R(c\beta) \\
&= S_0 R(c[\beta - 1]) - q \exp(-ru) R(c\beta).
\end{aligned}$$

(c) We plug the values $u = 1$, $q = S_0$, $r = 0.06$, $\alpha = 1$, and $\beta = 10$ into the previous formula to get

$$\begin{aligned}
c &= \log(10/9) - 0.06 = 0.0454 \\
S_0 \left[R(0.0454 \times 9) - e^{-0.06} R(0.0454 \times 10)\right] &= 0.0665 S_0.
\end{aligned}$$

5.8 The Beta Distributions

Commentary

Beta distributions arise as conjugate priors for the parameters of the Bernoulli, binomial, geometric, and negative binomial distributions. They also appear in several exercises later in the text, either because of their relationship to the t and F distributions (Exercise 1 in Sec. 8.4 and Exercise 6 in Sec. 9.7) or as examples of numerical calculation of M.L.E.'s (Exercise 10 in Sec. 7.6) or calculation of sufficient statistics (Exercises 24(h), 24(i) in Sec. 7.3, Exercise 7 in Sec. 7.7, and Exercises 2 and 7(c) in Sec. 7.8). The derivation of the p.d.f. of the beta distribution relies on material from Sec. 3.9 (particularly Jacobians) which the instructor might have skipped earlier in the course.

If one is using the statistical software R, then the function `beta` computes the beta function, and `lbeta` computes the logarithm of the beta function. They take only the two necessary arguments. The functions `dbeta`, `pbeta`, and `qbeta` give the p.d.f., the c.d.f., and the quantile function of beta distributions. The syntax is that the first argument is the argument of the function, and the next two are α and β in the notation of the text. The function `rbeta` gives a random sample of beta random variables. The first argument is how many you want, and the next two are α and β. All of the solutions that require the calculation of beta probabilites and quantiles can be done using these functions.

Solutions to Exercises

1. The c.d.f. of the beta distribution with parameters $\alpha > 0$ and $\beta = 1$ is

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ x^\alpha & \text{for } 0 < x < 1, \\ 1 & \text{for } x \geq 1. \end{cases}$$

Setting this equal to p and solving for x yields $F^{-1}(p) = p^{1/\alpha}$.

2. $f'(x|\alpha, \beta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}[(\alpha-1)(1-x)-(\beta-1)x]x^{\alpha-2}(1-x)^{\beta-2}$. Therefore, $f'(x|\alpha,\beta) = 0$ and $x = (\alpha-1)/(\alpha+\beta-2)$. It can be verified that if $\alpha > 1$ and $\beta > 1$, then $f(x|\alpha,\beta)$ is actually a maximum for this value of x.

3. The vertical scale is to be chosen in each part of Fig. S.5.4 so that the area under the curve is 1. The figure in (h) is the mirror image of the figure in (g) with respect to $x = 1/2$.

4. Let $Y = 1 - X$. Then $X = 1 - Y$. Therefore, $|dx/dy| = 1$ and, $0 < y < 1$,

$$g(y) = f(1-y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-y)^{\alpha-1}y^{\beta-1}.$$

This is the p.d.f. of the beta distribution with the values α and β interchanged.

5.

$$\begin{aligned} E[X^r(1-X)^S] &= \int_0^1 x^r(1-x)^S \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+r-1}(1-x)^{\beta+s-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+r)\Gamma(\beta+s)}{\Gamma(\alpha+\beta+r+s)} \\ &= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta+s)}{\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r+s)} \\ &= \frac{[\alpha(\alpha+1)\cdots(\alpha+r-1)][\beta(\beta+1)\cdots(\beta+s-1)]}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+r+s-1)}. \end{aligned}$$

6. The joint p.d.f. of X and Y will be the product of their marginal p.d.f.'s Therefore, for $x > 0$ and $y > 0$,

$$\begin{aligned} f(x,y) &= \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)}x^{\alpha_1-1}\exp(-\beta x)\frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)}y^{\alpha_2-1}\exp(-\beta y) \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}x^{\alpha_1-1}y^{\alpha_2-1}\exp(-\beta(x+y)). \end{aligned}$$

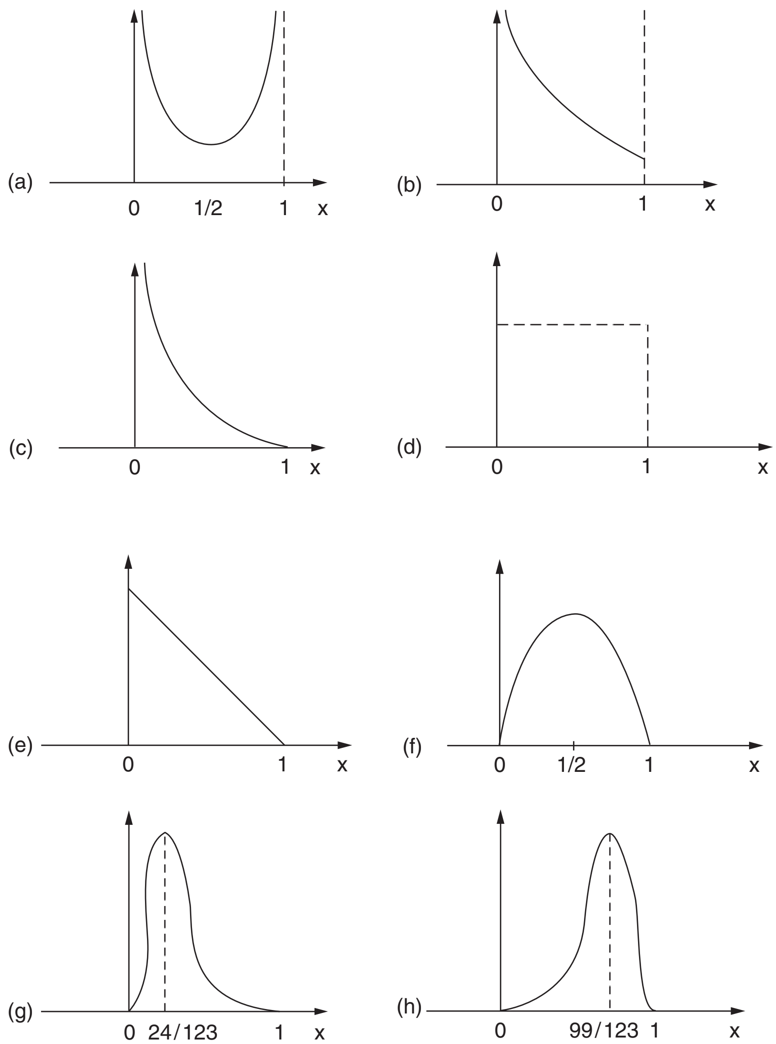

Figure S.5.4: Figure for Exercise 3 of Sec. 5.8.

Also, $X = UV$ and $Y = (1-U)V$. Therefore, the Jacobian is

$$J = \det \begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{bmatrix} = \det \begin{bmatrix} v & u \\ -v & 1-u \end{bmatrix} = v.$$

As x and y vary over all positive values, u will vary over the interval (0, 1) and v will vary over all possible values. Hence, for $0 < u < 1$ and $v > 0$, the joint p.d.f. of U and V will be

$$g(u,v) = f[uv,(1-u)v]v = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1-1}(1-u)^{\alpha_2-1} \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} v^{\alpha_1+\alpha_2-1}\exp(-\beta v).$$

It can be seen that this joint p.d.f. has been factored into the product of the p.d.f. of a beta distribution with parameters α_1 and α_2 and the p.d.f. of a gamma distribution with parameters $\alpha_1+\alpha_2$ and β. Therefore, U and V are independent, the distribution of U is the specified beta distribution, and the distribution of V is the specified gamma distribution.

7. Since X_1 and X_2 each have the gamma distribution with parameters $\alpha = 1$ and β, it follows from Exercise 6 that the distribution of $X_1/(X_1 + X_2)$ will be a beta distribution with parameters $\alpha = 1$ and $\beta = 1$. This beta distribution is the uniform distribution on the interval (0, 1).

8. (a) Let A denote the event that the item will be defective. Then

$$\Pr(A) = \int_0^1 \Pr(A \mid x) f(x)\, dx = \int_0^1 x f(x)\, dx = E(X) = \frac{\alpha}{\alpha + \beta}.$$

(b) Let B denote the event that both items will be defective. Then

$$\Pr(B) = \int_0^1 \Pr(B \mid x) f(x)\, dx = \int_0^1 x^2 f(x)\, dx = E(X^2) = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}.$$

9. Prior to observing the sample, the mean of P is $\alpha/(\alpha+\beta) = 0.05$, which means that $\alpha = \beta/19$. If we use the result in the note that follows Example 5.8.3, the distribution of P after finding 10 defectives in a sample of size 10 would be beta with parameters $\alpha+10$ and β, whose mean is $(\alpha+10)/(\alpha+\beta+10) = 0.9$. This means that $\alpha = 9\beta - 10$. So $9\beta - 10 = \beta/19$ and $\beta = 19/17$ so $\alpha = 1/17$. The distribution of P is then a beta distribution with parameters 1/17 and 19/17.

10. The distribution of P is a beta distribution with parameters 1 and 1. Applying the note after Example 5.8.3 with $n = 25$ and $x = 6$, the conditional distribution of P after observing the data is a beta distribution with parameters 7 and 20.

5.9 The Multinomial Distributions

Commentary

The family of multinomial distributions is the only named family of discrete multivariate distributions in the text. It arises in finite population sampling problems, but does not figure in the remainder of the text.

If one is using the statistical software R, then the function `dmultinom` gives the joint p.f. of a multinomial vector. The syntax is that the first argument is the argument of the function and must be a vector of the appropriate length with nonnegative integer coordinates. The next argument must be specified as `prob=` followed by the vector of probabilities, which must be a vector of the same length as the first argument. The function `rmultinom` gives a random sample of multinomial random vectors. The first argument is how many you want, the next argument specifies what the sum of the coordinates of every vector must be (n in the notation of the text), and the third argument is `prob` as above. All of the solutions that require the calculation of multinomial probabilites can be done using these functions.

Solutions to Exercises

1. Let $Y = X_1 + \cdots + X_\ell$. We shall show that Y has the binomial distribution with parameters n and $p_1 + \cdots + p_\ell$. Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with the p.f.

$$f(z) = \begin{cases} p_i & \text{for } z = i,\ i = 1, \ldots, k, \\ 0 & \text{otherwise.} \end{cases}$$

For each $i = 1, \ldots, k$ and each $j = 1, \ldots, n$, define

$$\begin{aligned} A_{ij} &= \{Z_j = i\}, \\ W_{ij} &= \begin{cases} 1 & \text{if } A_{ij} \text{ occurs,} \\ 0 & \text{if not.} \end{cases} \end{aligned}$$

Finally, define $V_i = \sum_{j=1}^{n} W_{ij}$ for $i = 1, \ldots, k$. It follows from the discussion in the text that $(X_1, \ldots, X_k)$ has the same distribution as $(V_1, \ldots, V_k)$. Hence Y has the same distribution as $U = V_1 + \cdots + V_\ell$. But

$$U = V_1 + \cdots + V_\ell = \sum_{i=1}^{\ell}\sum_{j=1}^{n} W_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{\ell} W_{ij}.$$

Define $U_j = \sum_{i=1}^{\ell} W_{ij}$. It is easy to see that $U_j = 1$ if $\cup_{i=1}^{\ell} A_{ij}$ occurs and $U_j = 0$ if not. Also $\Pr(\cup_{i=1}^{\ell} A_{ij}) = p_1 + \cdots + p_\ell$. Hence, $U_1, \ldots, U_n$ are i.i.d. random variables each having a Bernoulli distribution with parameter $p_1 + \cdots + p_\ell$. Since $U = \sum_{i=1}^{n} U_i$, we know that U has the binomial distribution with parameters n and $p_1 + \cdots + p_\ell$.

2. The probability that a given observed value will be less than α_1 is $p_1 = F(\alpha_1) = 0.3$, the probability that it will be between α_1 and α_2 is $p_2 = F(\alpha_2) - F(\alpha_1) = 0.5$, and the probability that it will be greater than α_2 is $p_3 = 1 - F(\alpha_2) = 0.2$. Therefore, the numbers of the 25 observations in each of these three intervals will have the multinomial distribution with parameters $n = 25$ and $\boldsymbol{p} = (p_1, p_2, p_3)$. Therefore, the required probability is

$$\frac{25!}{6!10!9!}(0.3)^6(0.5)^{10}(0.2)^9.$$

3. Let X_1 denote the number of times that the number 1 appears, let X_2 denote the number of times that the number 4 appears, and let X_3 denote the number of times that a number other than 1 or 4 appears. Then the vector (X_1, X_2, X_3) has the multinomial distribution with parameters $n = 5$ and $\boldsymbol{p} = (1/6, 1/6, 4/6)$. Therefore,

$$\begin{aligned}
\Pr(X_1 = X_2) &= \Pr(X_1 = 0,\ X_2 = 0,\ X_3 = 5) + \Pr(X_1 = 1,\ X_2 = 1,\ X_3 = 3) \\
&\quad + \Pr(X_1 = 2,\ X_2 = 2,\ X_3 = 1) \\
&= \left(\frac{4}{6}\right)^5 + \frac{5!}{1!1!3!}\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{4}{6}\right)^3 + \frac{5!}{2!2!1!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{4}{6}\right) \\
&= \frac{1024}{6^5} + \frac{1280}{6^5} + \frac{120}{6^5} = \frac{2424}{6^5}.
\end{aligned}$$

4. Let X_3 denote the number of rolls for which the number 5 appears. If $X_1 = 20$ and $X_2 = 15$, then it must also the true that $X_3 = 5$. The vector (X_1, X_2, X_3) has the multinomial distribution with parameters

$$\begin{aligned}
n &= 40, \\
q_1 &= p_2 + p_4 + p_6 = 0.30 + 0.05 + 0.07 = 0.42, \\
q_2 &= p_1 + p_3 = 0.11 + 0.22 = 0.33, \\
q_3 &= p_5 = 0.25.
\end{aligned}$$

Therefore,

$$\Pr(X_1 = 20 \text{ and } X_2 = 15) = \Pr(X_1 = 20, X_2 = 15, X_3 = 5) = \frac{40!}{20!15!5!}(0.42)^{20}(0.33)^{15}(0.25)^5.$$

5. The number X of freshman or sophomores selected will have the binomial distribution with parameters $n = 15$ and $p = 0.16 + 0.14 = 0.30$. Therefore, it is found from the table in the back of the book that

$$\Pr(X \geq 8) = .0348 + .0116 + .0030 + .0006 + .0001 = .0501.$$

6. By Eq. (5.9.3)

$$\begin{aligned} E(X_3) &= 15(0.38) = 5.7, \\ E(X_4) &= 15(0.32) = 4.8, \\ \text{Var}(X_3) &= 15(0.38)(0.62) = 3.534, \\ \text{Var}(X_4) &= 15(0.32)(0.68) = 3.264. \end{aligned}$$

By Eq. (5.9.3),

$$\text{Cov}(X_3, X_4) = -15(0.38)(0.32) = -1.824.$$

Hence,

$$E(X_3 - X_4) = 5.7 - 4.8 = 0.9$$

and

$$\text{Var}(X_3 - X_4) = 3.534 + 3.264 - 2(-1.824) = 10.446.$$

7. For any nonnegative integers $x_1, \ldots, x_k$ such that $\sum_{i=1}^k x_i = n$,

$$\Pr\left(X_1 = x_1, \ldots, X_k = x_k \,\middle|\, \sum_{i=1}^k X_i = n\right) = \frac{\Pr(X_1 = x_1, \ldots, X_k = x_k)}{\Pr\left(\sum_{i=1}^k X_i = n\right)}$$

Since $X_1, \ldots, X_k$ are independent,

$$\Pr(X_1 = x_1, \ldots, X_k = x_k) = \Pr(X_1 = x_1) \ldots \Pr(X_k = x_k).$$

Since X_i has the Poisson distribution with mean λ_i,

$$\Pr(X_i = x_i) = \frac{\exp(-\lambda_i)\lambda_i^{x_i}}{x_i!}.$$

Also, by Theorem 5.4.4, the distribution of $\sum_{i=1}^k X_i$ will be a Poisson distribution with mean $\lambda = \sum_{i=1}^k \lambda_i$. Therefore,

$$\Pr\left(\sum_{i=1}^k X_i = n\right) = \frac{\exp(-\lambda)\lambda^n}{n!}.$$

It follows that

$$\Pr\left(X_1 = x_1, \ldots, X_k = x_k \,\middle|\, \sum_{i=1}^k X_i = n\right) = \frac{n!}{x_i! \ldots x_k!} \prod_{i=1}^k \left(\frac{\lambda_i}{\lambda}\right)^{x_i}$$

8. Let the data be called $\boldsymbol{X} = (X_1, X_2, X_3)$, with X_1 being the number of working parts, X_2 being the number of impaired parts, and X_3 being the number of defective parts. The conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{p}$ is a multinomial distribution with parameters 10 and $\boldsymbol{p}$. So, the conditional p.f. of the observed data is

$$g(8,2,0|\boldsymbol{p}) = \binom{10}{8,2,0} p_1^8, p_2^2.$$

The joint p.f./p.d.f. of $\boldsymbol{X}$ and $\boldsymbol{p}$ is the product of this with the p.d.f. of $\boldsymbol{p}$:

$$12\binom{10}{8,2,0} p_1^{10} p_2^2 = 540 p_1^{10} p_2^2.$$

To find the conditional p.d.f. of $\boldsymbol{p}$ given $\boldsymbol{X} = (10, 2, 0)$, we need to divide this expression by the marginal p.f. of $\boldsymbol{X}$, which is the integral of this last expression over all (p_1, p_2) such that $p_i > 0$ and $p_1 + p_2 < 1$. This integral can be written as

$$\int_0^1 \int_0^{1-p_1} 540 p_1^{10} p_2^2 dp_2 dp_1 = \int_0^1 180 p_1^{10}(1-p_1)^3 = 180\frac{\Gamma(11)\Gamma(4)}{\Gamma(15)} = 0.0450.$$

For the second equality, we Theorem 5.8.1. So, the conditional p.d.f. of $\boldsymbol{p}$ given $\boldsymbol{X} = (10, 2, 0)$ is

$$\begin{cases} 12012 p_1^{10} p_2^2 & \text{if } 0 < p_1, p_2 < 1 \text{ and } p_1 + p_2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

5.10 The Bivariate Normal Distributions

Commentary

The joint distribution of the least squares estimators in a simple linear regression model (Sec. 11.3) is a bivariate normal distribution, as is the posterior distribution of the regression parameters in a Bayesian analysis of simple linear regression (Sec. 11.4). It also arises in the regression fallacy (Exercise 19 in Sec. 11.2 and Exercise 8 in Sec. 11.9) and as another theoretical avenue for introducing regression concepts (Exercises 2 and 3 in Sec. 11.9). The derivation of the bivariate normal p.d.f. relies on Jacobians from Sec. 3.9 which the instructor might have skipped earlier in the course.

Solutions to Exercises

1. The conditional distribution of the height of the wife given that the height of the husband is 72 inches is a normal distribution with mean $66.8 + 0.68 \times 2(72-70)/2 = 68.16$ and variance $(1-0.68^2)2^2 = 2.1504$. The 0.95 quantile of this distribution is

$$68.16 + 2.1504^{1/2}\Phi^{-1}(0.95) = 68.16 + 1.4664 \times 1.645 = 70.57.$$

2. Let X_1 denote the student's score on test A and let X_2 denote his score on test B. The conditional distribution of X_2 given that $X_1 = 80$ is a normal distribution with mean $90 + (0.8)(16)\left(\frac{80-85}{10}\right) = 83.6$ and variance $(1-0.64)(256) = 92.16$. Therefore, given that $X_1 = 80$, the random variable $Z = (X_2 - 83.6)/9.6$ will have the standard normal distribution. It follows that

$$\Pr(X_2 > 90 \,|\, X_1 = 80) = \Pr\left(Z > \frac{2}{3}\right) = 1 - \Phi\left(\frac{2}{3}\right) = 0.2524.$$

3. The sum $X_1 + X_2$ will have the normal distribution with mean $85 + 90 = 175$ and variance $(10)^2 + (16)^2 + 2(0.8)(10)(16) = 612$. Therefore, $Z = (X_1 + X_2 - 175)/24.7386$ will have the standard normal distribution. It follows that

$$\Pr(X_1 + X_2 > 200) = \Pr(Z > 1.0106) = 1 - \Phi(1.0106) = 0.1562.$$

4. The difference $X_1 - X_2$ will have the normal distribution with mean $85 - 90 = -5$ and variance $(10)^2 + (16)^2 - 2(0.8)(10)(16) = 100$. Therefore, $Z = (X_1 - X_2 + 5)/10$ will have the standard normal distribution. It follows that

$$\Pr(X_1 > X_2) = \Pr(X_1 - X_2 > 0) = \Pr(Z > 0.5) = 1 - \Phi(0.5) = 0.3085.$$

5. The predicted value should be the mean of the conditional distribution of X_1 given that $X_2 = 100$. This value is $85 + (0.8)(10)\left(\frac{100 - 90}{16}\right) = 90$. The M.S.E. for this prediction is the variance of the conditional distribution, which is $(1 - 0.64)100 = 36$.

6. $\text{Var}(X_1 + bX_2) = \sigma_1^2 + b^2\sigma_2^2 + 2b\rho\sigma_1\sigma_2$. This is a quadratic function of b. By differentiating with respect to b and setting the derivative equal to 0, we obtain the value $b = -\rho\sigma_1/\sigma_2$.

7. Since $E(X_1|X_2) = 3.7 - 0.15X_2$, it follows from Eq. (5.10.8) that

   (i) $\mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2 = 3.7$,

   (ii) $\rho\frac{\sigma_1}{\sigma_2} = -0.15$. Since $E(X_2|X_1) = 0.4 - 0.6X_1$, it follows from Eq. (5.10.6) that

   (iii) $\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 = 0.4$,

   (iv) $\rho\frac{\sigma_2}{\sigma_1} = -0.6$.

   Finally, since $\text{Var}(X_2|X_1) = 3.64$, it follows that

   (v) $(1 - \rho^2)\sigma_2^2 = 3.64$.

   By multiplying (ii) and (iv) we find that $\rho^2 = 0.09$. Therefore, $\rho = \pm 0.3$. Since the right side of (ii) is negative, ρ must be negative also. Hence, $\rho = -0.3$. It now follows from (v) that $\sigma_2^2 = 4$. Hence, $\sigma_2 = 2$ and it is found from (ii) that $\sigma_1 = 1$. By using the values we have obtained, we can rewrite (i) and (iii) as follows:

   (i) $\mu_1 + 0.15\mu_2 = 3.7$,

   (iii) $0.6\mu_1 + \mu_2 = 0.4$.

   By solving these two simultaneous linear equations, we find that $\mu_1 = 4$ and $\mu_2 = -2$.

8. The value of $f(x_1, x_2)$ will be a maximum when the exponent inside the curly braces is a maximum. In turn, this exponent will be a maximum when the expression inside the square brackets is a minimum. If we let

$$a_1 = \frac{x_1 - \mu_1}{\sigma_1} \quad \text{and} \quad a_2 = \frac{x_2 - \mu_2}{\sigma_2},$$

then this expression is

$$a_1^2 - 2\rho a_1 a_2 + a_2^2.$$

We shall now show that this expression must be nonnegative. We have

$$0 \le (|a_1| - |a_2|)^2 = a_1^2 + a_2^2 - 2|a_1 a_2| \le a_1^2 + a_2^2 - 2|\rho a_1 a_2|,$$

since $|\rho| < 1$. Furthermore, $|\rho a_1 a_2| \ge \rho a_1 a_2$. Hence,

$$0 \le a_1^2 + a_2^2 - 2\rho a_1 a_2.$$

The minimum possible value of $a_1^2 - 2\rho a_1 a_2 + a_2^2$ is therefore 0, and this value is attained when $a_1 = 0$ and $a_2 = 0$ or, equivalently, when $x_1 = \mu_1$ and $x_2 = \mu_2$.

9. Let a_1 and a_2 be as defined in Exercise 8. If $f(x_1, x_2) = k$, then $a_1^2 - 2\rho a_1 a_2 + a_2^2 = b^2$, where b is a particular positive constant. Suppose first that $\rho = 0$ and $\sigma_1 = \sigma_2 = \sigma$. Then this equation has the form

$$(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 = b^2\sigma^2.$$

This is the equation of a circle with center at (μ_1, μ_2) and radius bσ. Suppose next that $\rho = 0$ and $\sigma_1 \ne \sigma_2$. Then the equation has the form

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = b^2.$$

This is the equation of an ellipse for which the center is (μ_1, μ_2) and the major and minor axes are parallel to the x_1 and x_2 axes. Suppose finally that $\rho \ne 0$. It was shown in Exercise 8 that $a_1^2 - 2\rho a_1 a_2 + a_2^2 \ge 0$ for all values of a_1 and a_2. It therefore follows from the methods of analytic geometry and elementary calculus that the set of points which satisfy the equation

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)}{\sigma_1} \cdot \frac{(x_2 - \mu_2)}{\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = b^2.$$

will be an ellipse for which the center is (μ_1, μ_2) and for which the major and minor axes are rotated so that they are not parallel to the x_1 and x_2 axes.

10. Let $\Delta = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Since $\Delta \ne 0$, the transformation from X_1 and X_2 to Y_1 and Y_2 is a one-to-one transformation, for which the inverse transformation is:

$$\begin{aligned} X_1 &= \frac{1}{\Delta}[a_{22}(y_1 - b_1) - a_{12}(Y_2 - b_2)], \\ X_2 &= \frac{1}{\Delta}[-a_{21}(y_1 - b_1) + a_{22}(Y_2 - b_2)]. \end{aligned}$$

The joint p.d.f. of Y_1 and Y_2 can therefore be obtained by replacing x_1 and x_2 in $f(x_1, x_2)$ by their expressions in terms of y_1 and y_2, and then multiplying the result by the constant $1/|\Delta|$. After a great deal of algebra the exponent in this joint p.d.f. can be put into the following form:

$$-\frac{1}{2(1-r^2)}\left[\left(\frac{y_1 - m_1}{s_1}\right)^2 - 2r\left(\frac{y_1 - m_1}{s_1}\right)\left(\frac{y_2 - m_2}{s_2}\right) + \left(\frac{y_2 - m_2}{s_2}\right)^2\right],$$

where

$$\begin{aligned}
m_1 &= E(Y_1) = a_{11}\mu_1 + a_{12}\mu_2 + b_1,\\
m_2 &= E(Y_2) = a_{21}\mu_1 + a_{22}\mu_2 + b_2,\\
s_1^2 &= \text{Var}(Y_1) = a_{11}^2\sigma_1^2 + a_{12}^2\sigma_2^2 + 2a_{11}a_{12}\rho\sigma_1\sigma_2,\\
s_2^2 &= \text{Var}(Y_2) = a_{21}^2\sigma_1^2 + a_{22}^2\sigma_2^2 + 2a_{21}a_{22}\rho\sigma_1\sigma_2,\\
r &= \frac{\text{Cov}(Y_1,Y_2)}{s_1s_2} = \frac{1}{s_1s_2}[a_{11}a_{21}\sigma_1^2 + (a_{11}a_{22}+a_{12}a_{21})\rho\sigma_1\sigma_2 + a_{12}^2a_{22}^2\sigma_2^2.
\end{aligned}$$

It can then be concluded that this joint p.d.f. is the p.d.f. of a bivariate normal distribution for which the means are m_1 and m_2, the variances are s_1^2 and s_2^2, and the correlation is r.

11. By Exercise 10, the joint distribution of $X_1 + X_2$ and $X_1 - X_2$ is a bivariate normal distribution. By Exercise 9 of Sec. 4.6, these two variables are uncorrelated. Therefore, they are also independent.

12. (a) For the first species, the mean of $a_1X_1 + a_2X_2$ is $201a_1 + 118a_2$, while the variance is

$$15.2^2a_1^2 + 6.6^2a_2^2 + 2 \times 15.2 \times 6.6 \times 0.64a_1a_2.$$

The square-root of this is the standard deviation, $(231.04a_1^2 + 43.56a_2^2 + 128.41a_1a_2)^{1/2}$. For the second species, the mean is $187a_1 + 131a_2$. The standard deviation will be the same as for the first species because the values of σ_1, σ_2 and ρ are the same for both species.

(b) At first, it looks like we need a two-dimensional maximization. However, it is clear that the ratio in question, namely,

$$\frac{-14a_1 + 13a_2}{(231.04a_1^2 + 43.56a_2^2 + 128.41a_1a_2)^{1/2}} \qquad \text{(S.5.12)}$$

will have the same value if we multiply both a_1 and a_2 by the same positive constant. We could then assume that the pair (a_1, a_2) lies on a circle and hence reduce the maximization to a one-dimensional problem. Alternatively, we could assume that $a_1 + a_2 = 1$ and then find the maximum of the square of (S.5.12). (We would also have to check the one extra case in which $a_1 = -a_2$ to see if that produced a larger value.) We shall use this second approach. If we replace a_2 by $1 - a_1$, we need to find the maximum of

$$\frac{(13 - 27a_1)^2}{231.04a_1^2 + 43.56(1-a_1)^2 + 128.41a_1(1-a_1)}.$$

The derivative of this is the ratio of two polynomials, the denominator of which is always positive. So, the derivative is 0 when the numerator is 0. The numerator of the derivative is $13 - 27a_1$ times a linear function of a_1. The two roots of the numerator are 0.4815 and −0.5878. The first root produces the value 0 for (S.5.12), while the second produces the value 3.456. All pairs with $a_1 = -a_2$ lead to the values ±2.233. So $a_1 = -0.5878$ and $a_2 = 1.5878$ provide the maximum of (S.5.12).

13. The exponent of a bivariate normal p.d.f. can be expressed as $-[ax^2 + by^2 + cxy + ex + gy + h]$, where

$$\begin{aligned}
a &= \frac{1}{2\sigma_1^2(1-\rho^2)},\\
b &= \frac{1}{2\sigma_2^2(1-\rho^2)},\\
c &= -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)},
\end{aligned}$$

$$
\begin{aligned}
e &= -\frac{\mu_1}{\sigma_1^2(1-\rho^2)} + \frac{\mu_2\rho}{\sigma_1\sigma_2(1-\rho^2)}, \\
g &= -\frac{\mu_2}{\sigma_2^2(1-\rho^2)} + \frac{\mu_1\rho}{\sigma_1\sigma_2(1-\rho^2)},
\end{aligned}
$$

and h is irrelevant because $\exp(-h)$ just provides an additional constant factor that we are ignoring anyway. The only restrictions that the bivariate normal p.d.f. puts on the numbers a, b, c, e, and g are that $a, b > 0$ and whatever is equivalent to $|\rho| < 1$. It is easy to see that, so long as $a, b > 0$, we will have $|\rho| < 1$ if and only if $ab > (c/2)^2$. Hence, every set of numbers that satisfies these inequalities corresponds to a bivariate normal p.d.f. Assuming that these inequalities are satisfied, we can solve the above equations to find the parameters of the bivariate normal distribution.

$$
\begin{aligned}
\rho &= -\frac{c/2}{(ab)^{1/2}}, \\
\sigma_1^2 &= \frac{1}{2a - c^2/[2b]}, \\
\sigma_2^2 &= \frac{1}{2b - c^2/[2a]}, \\
\mu_1 &= \frac{cg - 2be}{4ab - c^2}, \\
\mu_2 &= \frac{ce - 2ag}{4ab - c^2}.
\end{aligned}
$$

14. The marginal p.d.f. of X is

$$
f_1(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}[x-\mu]^2\right),
$$

where μ and σ^2 are the mean and variance of X. The conditional p.d.f. of Y given $X = x$ is

$$
g_2(y|x) = \frac{1}{(2\pi\tau^2)^{1/2}} \exp\left(-\frac{1}{2\tau^2}[y - ax - b]^2\right).
$$

The joint p.d.f. of (X, Y) is the product of these two

$$
f(x, y) = \frac{1}{2\pi\sigma\tau} \exp\left(-[a'x^2 + b'y^2 + cxy + ex + gy + h]\right),
$$

where

$$
\begin{aligned}
a' &= \frac{1}{2\sigma^2} + \frac{a^2}{2\tau^2}, \\
b' &= \frac{1}{2\tau^2}, \\
c &= -\frac{a}{\tau^2}, \\
e &= -\frac{\mu}{\sigma^2} + \frac{ab}{\tau^2}, \\
g &= -\frac{b}{\tau^2},
\end{aligned}
$$

and h is irrelevant since we are going to apply the result from Exercise 13. Clearly a' and b' are positive. We only need to check that $a'b' > (c/2)^2$. Notice that

$$a'b' = \frac{1}{4\sigma^2\tau^2} + \frac{a^2}{4\tau^2} = (c/2)^2 + \frac{1}{4\sigma^2\tau^2},$$

so the conditions of Exercise 13 are met.

15. (a) Let $Y = \sum_{j \neq i} X_j$. Since $X_1, \ldots, X_n$ are independent, we know that Y is independent of X_i. Since Y is the sum of independent normal random variables it has a normal distribution. The mean and variance of Y are easily seen to be $(n-1)\mu$ and $(n-1)\sigma^2$ respectively. Since Y and X_i are independent, all pairs of linear combinations of them have a bivariate normal distribution. Now write

$$\begin{aligned} X_i &= 1X_i + 0Y, \\ \overline{X}_n &= \frac{1}{n}X_i + \frac{1}{n}Y. \end{aligned}$$

Clearly, both X_i and Y have mean μ, and we already know that X_i has variance σ^2 while Y has variance σ^2/n. The correlation can be computed from the covariance of the two linear combinations.

$$\text{Cov}\left(1X_i + 0Y, \frac{1}{n}X_i + \frac{1}{n}Y\right) = \frac{1}{n}\sigma^2.$$

The correlation is then $(\sigma^2/n)/[\sigma^2\sigma^2/n]^{1/2} = 1/n^{1/2}$.

(b) The conditional distribution of X_i given $\overline{X}_n = \overline{x}_n$ is a normal distribution with mean equal to

$$\mu + \frac{1}{n^{1/2}}\frac{\sigma}{\sigma/n^{1/2}}(\overline{x}_n - \mu) = \overline{x}_n.$$

The conditional variance is

$$\sigma^2 - \left(1 - \frac{1}{n}\right).$$

5.11 Supplementary Exercises

Solutions to Exercises

1. Let $g_1(x|p)$ be the conditional p.f. of X given $P = p$, which is the binomial p.f. with parameters n and p. Let $f_2(p)$ be the marginal p.d.f. of P, which is beta p.d.f. with parameters 1 and 1, also known as the uniform p.d.f. on the interval $[0,1]$. According to the law of total probability for random variables, the marginal p.f. of X is

$$f_1(x) = \int g_1(x|p)f_2(p)dp = \int_0^1 \binom{n}{x} p^x(1-p)^{n-x}dp = \binom{n}{x}\frac{x!(n-x)!}{(n+1)!} = \frac{1}{n+1},$$

for $x = 0, \ldots, n$. In the above, we used Theorem 5.8.1 and the fact that $\Gamma(k+1) = k!$ for each integer k.

2. The random variable $U = 3X + 2Y - 6Z$ has the normal distribution with mean 0 and variance $3^2 + 2^2 + 6^2 = 49$. Therefore, $Z = U/7$ has the standard normal distribution. The required probability is

$$\Pr(U < -7) = \Pr(Z < -1) = 1 - \Phi(1) = .1587.$$

3. Since $\text{Var}(X) = E(X) = \lambda_1$ and $\text{Var}(Y) = E(Y) = \lambda_2$, it follows that $\lambda_1 + \lambda_2 = 5$. Hence, $X + Y$ has the Poisson distribution with mean 5 and

$$\Pr(X + Y < 2) = \Pr(X + Y = 0) + \Pr(X + Y = 1) = \exp(-5) + 5\exp(-5) = .0067 + .0337 = .0404.$$

4. It can be found from the table of the standard normal distribution that 116 must be .84 standard deviations to the left of the mean and 328 must be 1.28 standard deviations to the right of the mean. Hence,

$$\begin{aligned} \mu - .84\sigma &= 116, \\ \mu + 1.28\sigma &= 328\,. \end{aligned}$$

Solving these equations, we obtain $\mu = 200$ and $\sigma = 100$, $\sigma^2 = 10,000$.

5. The event $\{\overline{X} < 1/2\}$ can occur only if all four observations are 0, which has probability $(\exp(-\lambda))^4$, or three of the observations are 0 and the other is 1, which has probability $4(\lambda\exp(-\lambda))(\exp(-\lambda))^3$. Hence, the total probability is as given in this exercise.

6. If X has the exponential distribution with parameter β, then

$$.25 = \Pr(X > 1000) = \exp(-(1000)\beta).$$

Hence, $\beta = \dfrac{1}{1000}\log 4$ and $E(X) = \dfrac{1}{\beta} = \dfrac{1000}{\log 4}$.

7. It follows from Exercise 18 of Sec. 4.9 that

$$E[(X - \mu)^3] = E(X^3) - 3\mu\sigma^2 - \mu^3.$$

Because of the symmetry of the normal distribution with respect to μ, the left side of this relation is 0. Hence,

$$E(X^3) = 3\mu\sigma^2 + \mu^3.$$

8. $\overline{X}$ and $\overline{Y}$ have independent normal distributions with the same mean μ, and $\text{Var}(\overline{X}) = 144/16 = 9$, $\text{Var}(\overline{Y}) = 400/25 = 16$. Hence, $\overline{X} - \overline{Y}$ has the normal distribution with mean 0 and variance $9 + 16 = 25$. Thus, $Z = (\overline{X} - \overline{Y})/5$ has the standard normal distribution. It follows that the required probability is $\Pr(|Z| < 1) = .6826$.

9. The number of men that arrive during the one-minute period has the Poisson distribution with mean 2. The number of women is independent of the number of men and has the Poisson distribution with mean 1. Therefore, the total number of people X that arrive has the Poisson distribution with mean 3. From the table in the back of the book it is found that

$$\Pr(X \leq 4) = .0498 + .1494 + .2240 + .2240 + .1680 = .8152.$$

10.

$$\begin{aligned}
\psi_Y(t) &= E(\exp(tY)) = E[\exp(tX_1 + \cdots + tX_N)] \\
&= E\{E[\exp(tX_1 + \cdots + tX_N)|N]\} \\
&= E\{[\psi(t)]^N\} \\
&= \sum_{x=0}^{\infty} [\psi(t)]^X \frac{\exp(-\lambda)\lambda^x}{x!} \\
&= \exp(-\lambda) \sum_{x=0}^{\infty} \frac{[\lambda\psi(t)]^x}{x!} \\
&= \exp(-\lambda)\exp(\lambda\psi(t)) = \exp\{\lambda[\psi(t)-1]\}.
\end{aligned}$$

11. The probability that at least one of the two children will be successful on a given Sunday is $(1/3) + (1/5) - (1/3)(1/5) = 7/15$. Therefore, from the geometric distribution, the expected number of Sundays until a successful launch is achieved is $15/7$.

12. For any positive integer n, the event $X > n$ will occur if and only if the first n tosses are either all heads or all tails. Therefore,

$$\Pr(X > n) = \left(\frac{1}{2}\right)^n + \left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^{n-1},$$

and, for $n = 2, 3, \ldots$.

$$\Pr(X = n) = \Pr(X > n-1) - \Pr(X > n) = \left(\frac{1}{2}\right)^{n-2} - \left(\frac{1}{2}\right)^{n-1} = \left(\frac{1}{2}\right)^{n-1}.$$

Hence,

$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^{x-1} & \text{for } x = 2, 3, 4, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

13. By the Poisson approximation, the distribution of X is approximately Poisson with mean $120(1/36) = 10/3$. The probability that such a Poisson random variable equals 3 is $\exp(-10/3)(10/3)^3/3! = 0.2202$. (The actual binomial probability is 0.2229.)

14. It was shown in Sec. 3.9 that the p.d.f.'s of Y_1, Y_n, and W, respectively, are as follows:

$$\begin{aligned}
g_1(y) &= n(1-y)^{n-1} && \text{for } 0 < y < 1, \\
g_n(y) &= ny^{n-1} && \text{for } 0 < y < 1, \\
h_1(w) &= n(n-1)w^{n-2}(1-w) && \text{for } 0 < w < 1.
\end{aligned}$$

Each of these is the p.d.f. of a beta distribution. For $g_1, \alpha = 1$ and $\beta = n$. For g_n, $\alpha = n$ and $\beta = 1$. For h_1, $\alpha = n-1$ and $\beta = 2$.

15. (a) $\Pr(T_1 > t) = \Pr(X = 0)$, where X is the number of occurrences between time 0 and time t. Since X has the Poisson distribution with mean $5t$, it follows that $\Pr(T_1 > t) = \exp(-5t)$. Hence, T_1 has the exponential distribution with parameter $\beta = 5$.

(b) T_k is the sum of k i.i.d. random variables, each of which has the exponential distribution given in part (a). Therefore, the distribution of T_k is a gamma distribution with parameters $\alpha = k$ and $\beta = 5$.

(c) Let X_i denote the time following the ith occurrence until the $(i+1)$st occurrence. Then the random variable $X_1, \ldots, X_{k-1}$ are i.i.d., each of which has the exponential distribution given in part (a). Since t is measured in hours in that distribution, the required probability is

$$\Pr\left(X_i > \frac{1}{3}, i = 1, \ldots, k-1\right) = (\exp(-5/3))^{k-1}.$$

16. We can express T_5 as $T_1 + V$, where V is the time required after one of the components has failed until the other four have failed. By the memoryless property of the exponential distribution, we know that T_1 and V are independent. Therefore,

$$\text{Cov}(T_1, T_5) = \text{Cov}(T_1, T_1 + V) = \text{Cov}(T_1, T_1) + \text{Cov}(T_1, V) = \text{Var}(T_1) + 0 = \frac{1}{25\beta^2},$$

since T_1 has the exponential distribution with parameter 5β.

17. $\Pr(X_1 > kX_2) = \int_0^\infty \Pr(X_1 > kX_2 | X_2 = x)(f_2 x)dx = \int_0^\infty \exp(-\beta_1 kx)\beta_2 \exp(-\beta_2 x)dx = \dfrac{\beta_2}{k\beta_1 + \beta_2}.$

18. Since the sample size is small relative to the size of the population, the distribution of the number X of people in the sample who are watching will have essentially the binomial distribution with $n = 200$ and $p = 15000/500000 = .03$, even if sampling is done without replacement. This binomial distribution is closely approximated by Poisson distribution with mean $\lambda = np = 6$. Hence, from the table in the back of the text,

$$\Pr(X < 4) = .0025 + .0149 + .0446 + .0892 = .1512.$$

19. It follows from Eq. (5.3.8) that

$$\text{Var}(\overline{X}) = \frac{1}{n^2}\text{Var}(X) = \frac{p(1-p)}{n} \cdot \frac{T-n}{T-1},$$

where T is the population size, p is the proportion of persons in the population who have the characteristic, and $n = 100$. Since $p(1-p) \le 1/4$ for $0 \le p \le 1$ and $(T-n)(T-1) \le 1$ for all values of T, it follows that

$$\text{Var}(\overline{X}_n) \le \frac{1}{400}.$$

Hence, the standard deviation is $\le 1/20 = .05$.

20. Consider the event that less than r successes are obtained in the first n Bernoulli trials. The left side represents the probability of this event in terms of the binomial distribution. But the event also means that more than n trials are going to be required in order to obtain r successes, which means that more than $n - r$ failures are going to be obtained before r successes are obtained. The right side expresses this probability in terms of the negative binomial distribution.

21. Consider the event that there are at least k occurrences between time 0 and time t. The number X of occurrences in this interval has the specified Poisson distribution, so the left side represents the probability of this event. But the event also means that the total waiting time Y until the kth event occurs is $\le t$. It follows from part (b) of Exercise 15 that Y has the specified gamma distribution. Hence, the right side also expresses the probability of this same event.

22. It follows from the definition of h(x) that

$$\int_0^x h(t)dt = \int_0^x \frac{f(t)}{1-F(t)}dt = -\log[1-F(x)].$$

Therefore,

$$\exp\left[-\int_0^x h(t)\,dt\right] = 1 - F(x).$$

23. (a) It follows from Theorem 5.9.2 that

$$\rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{[\text{Var}(X_i)\,\text{Var}(X_j)]^{1/2}} = \frac{-10p_ip_j}{10[p_i(1-p_i)p_j(1-p_j)]^{1/2}} = -\left(\frac{p_i}{1-p_i}\cdot\frac{p_j}{1-p_j}\right)^{1/2}$$

(b) $\rho(X_i, X_j)$ is most negative when p_i and p_j have their largest values; i.e., for $i = 1$ ($p_i = .4$) and $j = 2$ ($p_2 = .3$).

(c) $\rho(X_i, X_j)$ is closest to 0 when p_i and p_j have their smallest values; i.e., for $i = 3$ ($p_3 = .2$) and $j = 4$ ($p_4 = .1$).

24. It follows from Theorem 5.10.5 that $X_1 - 3X_2$ will have the normal distribution with mean $\mu_1 - 3\mu_2$ and variance $\sigma_1^2 + 9\sigma_2^2 - 6\rho\sigma_1\sigma_2$.

25. Since X has a normal distribution and the conditional distribution of Y given X is also normal with a mean that is a linear function of X and constant variance, it follows that X and Y jointly have a bivariate normal distribution. Hence, Y has a normal distribution. From Eq. (5.10.6), $2X - 3 = \mu_2 + \rho\sigma_2 X$. Hence, $\mu_2 = -3$ and $\rho\sigma_2 = 2$. Also, $(1-\rho^2)\sigma_2^2 = 12$. Therefore $\sigma_2^2 = 16$ and $\rho = 1/2$. Thus, Y has the normal distribution with mean -3 and variance 16, and $\rho(X, Y) = 1/2$.

26. We shall use the relation

$$E(X_1^2X_2) = E[E(X_1^2X_2 \mid X_2)] = E[X_2E(X_1^2 \mid X_2)].$$

But

$$E(X_1^2|X_2) = \text{Var}(X_1|X_2) + [E(X_1|X_2)]^2 = (1-\rho^2)\sigma_1^2 + \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}X_2\right)^2.$$

Hence,

$$X_2E(X_1^2|X_2) = (1-\rho^2)\sigma_1^2X_2 + \mu_1^2X_2 + 2\rho\mu_1\frac{\sigma_1}{\sigma_2}X_2^2 + \left(\rho\frac{\sigma_1}{\sigma_2}\right)^2 X_2^3.$$

The required value $E(X_1^2X_2)$ is the expectation of this quantity. But since X_2 has the normal distribution with $E(X_2) = 0$, it follows that $E(X_2^2) = \sigma_2^2$ and $E(X_2^3) = 0$.

Hence,

$$E(X_1^2X_2) = 2\rho\mu_1\sigma_1\sigma_2.$$

Chapter 6

Large Random Samples

6.1 Introduction

Solutions to Exercises

1. The p.d.f. of $Y = X_1 + X_2$ is

$$g(y) = \begin{cases} y & \text{if } 0 < y \le 1, \\ 2 - y & \text{if } 1 < y < 2, \\ 0 & \text{otherwise.} \end{cases}$$

It follows easily from the fact that $\overline{X}_2 = Y/2$ that the p.d.f. of $\overline{X}_2$ is

$$h(x) = \begin{cases} 4x & \text{if } 0 < x \le 1/2, \\ 4 - 4x & \text{if } 1/2 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

We easily compute

$$\begin{aligned} \Pr(|X_1 - 0.5| < 0.1) &= 0.6 - 0.4 = 0.2, \\ \Pr(|\overline{X}_2 - 0.5| < 0.1) &= \int_{0.4}^{0.5} 4x dx + \int_{0.5}^{0.6} (4 - 4x) dx \\ &= 2(0.5^2 - 0.4^2) + 4(0.6 - 0.5) - 2(0.6^2 - 0.5^2) = 0.36. \end{aligned}$$

The reason that $\overline{X}_2$ has higher probability of being close to 0.5 is that its p.d.f. is much higher near 0.5 than is the uniform p.d.f. of X_1 (twice as high right at 0.5).

2. The distribution of $\overline{X}_n$ is (by Corollary 5.6.2) the normal distribution with mean μ and variance σ^2/n. By Theorem 5.6.6,

$$\begin{aligned} \Pr(|\overline{X}_n - \mu| \le c) &= \Pr(\overline{X}_n \le c) - \Pr(\overline{X}_n \le -c) \\ &= \Phi\left(\frac{c}{\sigma/n^{1/2}}\right) - \Phi\left(\frac{-c}{\sigma/n^{1/2}}\right). \end{aligned} \quad \text{(S.6.1)}$$

As $n \to \infty$, $c/(\sigma/n^{1/2}) \to \infty$ and $-c/(\sigma/n^{1/2}) \to -\infty$. It follows from Property 3.3.2 of all c.d.f.'s that (S.6.1) goes to 1 as $n \to \infty$.

3. To do this by hand we would have to add all of the binomial probabilities corresponding to $W = 80, \dots, 120$. Most statistical software will do this calculation automatically. The result is 0.9964. It looks like the probability is increasing to 1.

6.2 The Law of Large Numbers

Commentary

The discussion of the strong law of large numbers at the end of the section might be suitable only for the more mathematically inclined students.

Solutions to Exercises

1. Let $\epsilon > 0$. We need to show that

$$\lim_{n\to\infty} \Pr(|X_n - 0| \geq \epsilon) = 0. \tag{S.6.2}$$

Since $X_n \geq 0$, we have $|X_n - 0| \geq \epsilon$ if and only if $X_n \geq \epsilon$. By the Markov inequality $\Pr(X_n \geq \epsilon) \leq \mu_n/\epsilon$. Since $\lim_{n\to\infty} \mu_n = 0$, Eq. (S.6.2) holds.

2. By the Markov inequality,

$$E(X) \geq 10\Pr(X \geq 10) = 2.$$

3. By the Chebyshev inequality,

$$\mathrm{Var}(X) \geq 9\Pr(|X - \mu| \geq 3) = 9\Pr(X \leq 7 \text{ or } X \geq 13) = 9(0.2 + 0.3) = \frac{9}{2}.$$

4. Consider a distribution which is concentrated on the three points μ, $\mu + 3\sigma$, and $\mu - 3\sigma$. Let $\Pr(X = \mu) = p_1$, $\Pr(X = \mu + 3\sigma) = p_2$, and $\Pr(X = \mu - 3\sigma) = p_3$. If we are to have $E(X) = \mu$, then we must have $p_2 = p_3$. Let p denote the common value of p_2 and p_3. Then $p_1 = 1 - 2p$, because $p_1 + p_2 + p_3 = 1$. Now

$$\mathrm{Var}(X) = E[(X - \mu)^2] = 9\sigma^2(p) + 9\sigma^2(p) + 0(1 - 2p) = 18\sigma^2 p.$$

Since we must have $\mathrm{Var}(X) = \sigma^2$, then we must choose $p = 1/18$. Therefore, the only distribution which is concentrated on the three points $\mu, \mu + 3\sigma$, and $\mu - 3\sigma$, and for which $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$, is the one with $p_1 = 8/9$ and $p_2 = p_3 = 1/18$. It can now be verified that for this distribution we have

$$\Pr(|X - \mu| \geq 3\sigma) = \Pr(X = \mu + 3\sigma) + \Pr(X = \mu - 3\sigma) = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}.$$

5. By the Chebyshev inequality,

$$\Pr(|X_n - \mu| \leq 2\sigma) \geq 1 - \frac{1}{4n}.$$

Therefore, we must have $1 - \frac{1}{4n} \geq 0.99.$ or $n \geq 25$.

6. By the Chebyshev inequality,

$$\Pr(6 \leq \overline{X}_n \leq 7) = \Pr\left(|\overline{X}_n - \mu| \leq \frac{1}{2}\right) \geq 1 - \frac{16}{n}.$$

Therefore, we must have $1 - \frac{16}{n} \geq 0.8$ or $n \geq 80$.

7. By the Markov inequality,

$$\Pr(|X-\mu|\geq t)=\Pr(|X-\mu|^4\geq t^4)\leq\frac{E(|X-\mu|^4)}{t^4}=\frac{\beta_4}{t^4}.$$

8. (a) In this example $E(Q_n)=0.3$ and $\text{Var}(Q_n)=(0.3)(0.7)/n=0.21/n$. Therefore,

$$\Pr(0.2\leq Q_n\leq 0.4)=\Pr(|Q_n-E(Q_n)|\leq 0.1)\geq 1-\frac{0.21}{n(0.01)}=1-\frac{21}{n}.$$

Therefore, we must have $1-\frac{21}{n}\geq 0.75$ or $n\geq 84$.

(b) Let X_n denote the total number of items in the sample that are of poor quality. Then $X_n=nQ_n$ and

$$\Pr(0.2\leq Q_n\leq 0.4)=\Pr(0.2n\leq X_n\leq 0.4n).$$

Since X_n has a binomial distribution with parameters n and $p=0.3$, the value of this probability can be determined for various values of n from the table of the binomial distribution given in the back of the book. For $n=15$, it is found that

$$\Pr(0.2n\leq X_n\leq 0.4n)=\Pr(3\leq X_n\leq 6)=0.7419.$$

For $n=20$, it is found that

$$\Pr(0.2n\leq X_n\leq 0.4n)=\Pr(4\leq X_n\leq 8)=0.7796.$$

Since this probability must be at least 0.75, we must have $n=20$, although it is possible that some value between $n=15$ and $n=20$ will also satisfy the required condition.

9. $E(Z_n)=n^2\cdot\frac{1}{n}+0\left(1-\frac{1}{n}\right)=n$. Hence, $\lim_{n\to\infty}E(Z_n)=\infty$. Also, for any given $\epsilon>0$,

$$\Pr(|Z_n|<\epsilon)=\Pr(Z_n=0)=1-\frac{1}{n}.$$

Hence, $\lim_{n\to\infty}\Pr(|Z_n|<\epsilon)=1$, which means that $Z_n\xrightarrow{p}0$.

10. By Exercise 5 of Sec. 4.3,

$$E[(Z_n-b)^2]=[E(Z_n)-b]^2+\text{Var}(Z_n).$$

Therefore, the limit of the left side will be 0 if and only if the limit of each of the two terms on the right side is 0. Moreover, $\lim_{n\to\infty}[E(Z_n)-b]^2=0$ if and only if $\lim_{n\to\infty}E(Z_n)=b$.

11. Suppose that the sequence $Z_1,Z_2,\ldots$ converges to b in the quadratic mean. Since

$$|Z_n-b|\leq|Z_n-E(Z_n)|+|E(Z_n)-b|\ ,$$

then for any value of $\epsilon>0$,

$$\begin{aligned}\Pr(|Z_n-b|<\epsilon) &\geq \Pr(|Z_n-E(Z_n)|+|E(Z_n)-b|<\epsilon)\\ &= \Pr(|Z_n-E(Z_n)|<\epsilon-|E(Z_n)-b|).\end{aligned}$$

By Exercise 10, we know that $\lim_{n\to\infty} E(Z_n) = b$. Therefore, for sufficiently large values of n, it will be true that $\epsilon - |E(Z_n) - b| > 0$. Hence, by the Chebyshev inequality, the final probability will be at least as large as

$$1 - \frac{\text{Var}(Z_n)}{[\epsilon - |E(Z_n) - b|]^2}.$$

Again, by Exercise 10,

$$\lim_{n\to\infty} \text{Var}(Z_n) = 0 \text{ and } \lim_{n\to\infty} [\epsilon - |E(Z_n) - b|]^2 = \epsilon^2.$$

Therefore,

$$\lim_{n\to\infty} \Pr(|Z_n - b| < \epsilon) \geq \lim_{n\to\infty} \left\{1 - \frac{\text{Var}(Z_n)}{[\epsilon - |E(Z_n) - b|]^2}\right\} = 1,$$

which means that $Z_n \xrightarrow{p} b$.

12. We know that $E(\overline{X}_n) = \mu$ and $\text{Var}(\overline{X}_n) = \sigma^2/n$. Therefore, $\lim_{n\to\infty} E(\overline{X}_n) = \mu$ and $\lim_{n\to\infty} \text{Var}(\overline{X}_n) = 0$. The desired result now follows from Exercise 10.

13. (a) For any value of n large enough so that $1/n < \epsilon$, we have

$$\Pr(|Z_n| < \epsilon) = \Pr\left(Z_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2}.$$

Therefore, $\lim_{n\to\infty} \Pr(|Z_n| < \epsilon) = 1$, which means that $Z_n \xrightarrow{p} 0$.

(b) $E(Z_n) = \frac{1}{n}\left(1 - \frac{1}{n^2}\right) + n\left(\frac{1}{n^2}\right) = \frac{2}{n} - \frac{1}{n^3}$. Therefore, $\lim_{n\to\infty} E(Z_n) = 0$. It follows from Exercise 10 that the only possible value for the constant c is $c = 0$, and there will be convergence to this value if and only if $\lim_{n\to\infty} \text{Var}(Z_n) = 0$. But

$$E(Z_n^2) = \frac{1}{n^2}\left(1 - \frac{1}{n^2}\right) + n^2 \cdot \frac{1}{n^2} = 1 + \frac{1}{n^2} - \frac{1}{n^4}.$$

Hence, $\text{Var}(Z_n) = 1 + \frac{1}{n^2} - \frac{1}{n^4} - \left(\frac{2}{n} - \frac{1}{n^3}\right)^2$ and $\lim_{n\to\infty} \text{Var}(Z_n) = 1$.

14. Let X have p.f. equal to f. Assume that $\text{Var}(X) > 0$ (otherwise it is surely less than 1/4). First, suppose that X has only two possible values, 0 and 1. Let $p = \Pr(X = 1)$. Then $E(X) = E(X^2) = p$ and $\text{Var}(X) = p - p^2$. The largest possible value of $p - p^2$ occurs when $p = 1/2$, and the value is 1/4. So $\text{Var}(X) \leq 1/4$ if X only has the two possible values 0 and 1. For the remainder of the proof, we shall show that if X has any possible values strictly between 0 and 1, then there is another random variable Y taking only the values 0 and 1 and with $\text{Var}(Y) \geq \text{Var}(X)$. So, assume that X takes at least one value strictly between 0 and 1. Without loss of generality, assume that one of those possible values is between 0 and μ. (Otherwise replace X by $1 - X$ which has the same variance.) Let $\mu = E(X)$, and let $x_1, x_2, \ldots$ be the values such that $x_i \leq \mu$ and $f(x_i) > 0$. Define a new random variable

$$X^* = \begin{cases} 0 & \text{if } X \leq \mu, \\ X & \text{if } X > \mu. \end{cases}$$

The p.f. of X^* is

$$f^*(x) = \begin{cases} f(x) & \text{for all } x > \mu, \\ \sum_i f(x_i) & \text{for } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The mean of X^* is $\mu^* = \mu - \sum_i x_i f(x_i)$. The mean of X^{*2} is $E(X^2) - \sum_i x_i^2 f(x_i)$. So, the variance of X^* is

$$\begin{aligned} \text{Var}(X^*) &= E(X^2) - \sum_i x_i^2 f(x_i) - \left[\mu - \sum_i x_i f(x_i)\right]^2 \\ &= \text{Var}(X) - \sum_i x_i^2 f(x_i) + 2\mu \sum_i x_i f(x_i) - \left[\sum_i x_i f(x_i)\right]^2. \end{aligned} \quad \text{(S.6.3)}$$

since $x_i \le \mu$ for each i, we have

$$-\sum_i x_i^2 f(x_i) + 2\mu \sum_i x_i f(x_i) \ge \sum_i x_i^2 f(x_i). \quad \text{(S.6.4)}$$

Let $t = \sum_i f(x_i) > 0$. Then

$$g(x) = \begin{cases} f(x)/t & \text{for } x \in \{x_1, x_2, \ldots\}, \\ 0 & \text{otherwise,} \end{cases}$$

is a p.f. Let Z be a random variable whose p.f. is g. Then

$$\begin{aligned} E(Z) &= \frac{1}{t} \sum_i x_i f(x_i), \\ \text{Var}(Z) &= \frac{1}{t} \sum_i x_i^2 f(x_i). \end{aligned}$$

Since $\text{Var}(Z) \ge 0$ and $t \le 1$, we have

$$\sum_i x_i^2 f(x_i) \ge \frac{1}{t} \left[\sum_i x_i f(x_i)\right]^2 \ge \left[\sum_i x_i f(x_i)\right]^2.$$

Combine this with (S.6.3) and (S.6.4) to see that $\text{Var}(X^*) \ge \text{Var}(X)$. If $f^*(x) > 0$ for some x strictly between 0 and 1, replace X^* by $1 - X^*$ and repeat the above process to produce the desired random variable Y.

15. We need to prove that, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|g(Z_n) - g(b)| < \epsilon) = 1.$$

Let $\epsilon > 0$. Since g is continuous at b, there exists δ such that $|z - b| < \delta$ implies that $|g(z) - g(b)| < \epsilon$. Also, since $Z_n \xrightarrow{P} b$, we know that

$$\lim_{n\to\infty} \Pr(|Z_n - b| < \delta) = 1.$$

But $\{|Z_n - b| < \delta\} \subset \{|g(Z_n) - g(b)| < \epsilon\}$. So

$$\Pr(|g(Z_n) - g(b)| < \epsilon) \geq \Pr(|Z_n - b| < \delta) \quad \text{(S.6.5)}$$

Since the right side of (S.6.5) goes to 1 as $n \to \infty$ so does the left side.

16. The argument here is similar to that given in Exercise 15. Let $\epsilon > 0$. Since g is continuous at (b, c), there exists δ such that $\sqrt{(z-b)^2 + (y-c)^2} < \delta$ implies $|g(z, y) - g(b, c)| < \epsilon$. Also, $|z - b| < \delta/\sqrt{2}$ and $|y - c| < \delta/\sqrt{2}$ together imply $\sqrt{(z-b)^2 + (y-c)^2} < \delta$. Let $B_n = \{|Z_n - b| < \delta/\sqrt{2}\}$ and $C_n = \{|Y_n - c| < \delta/\sqrt{2}\}$. It follows that

$$B_n \cap C_n \subset \{|g(Z_n, Y_n) - g(b, c)| < \delta\}. \quad \text{(S.6.6)}$$

We can write

$$\begin{aligned}\Pr(B_n \cap C_n) &= 1 - \Pr([B_n \cap C_n]^c) = 1 - \Pr(B_n^c \cup C_n^c) \geq 1 - \Pr(B_n^c) - \Pr(C_n^c) \\ &= \Pr(B_n) + \Pr(C_n) - 1.\end{aligned}$$

Combining this with (S.6.6), we get

$$\Pr(|g(Z_n, Y_n) - g(b, c)| < \delta) \geq \Pr(B_n) + \Pr(C_n) - 1.$$

Since $Z_n \xrightarrow{P} b$ and $Y_n \xrightarrow{P} c$, we know that both $\Pr(B_n)$ and $\Pr(C_n)$ go to 1 as $n \to \infty$. Hence $\Pr(|g(Z_n, Y_n) - g(b, c)| < \delta)$ goes to 1 as well.

17. (a) The mean of X is np, and the mean of Y is np/k. Since $Z = kY$, the mean of Z is $knp/k = np$.

    (b) The variance of X is $np(1-p)$, and the variance of Y is $n(p/k)(1-p/k)$. So, the variance $Z = kY$ is k^2 times the variance of Y, i.e.,

    $$\text{Var}(Z) = k^2 n(p/k)(1 - p/k) = knp(1 - p/k).$$

    If p is small, then both $1 - p$ and $1 - p/k$ will be close to 1, and $\text{Var}(Z)$ is approximately knp while the variance of X is approximately np.

    (c) In Fig. 6.1, each bar has height equal to 0.01 times a binomial random variable with parameters 100 and the probability that X_1 is in the interval under the bar. In Fig. 6.2, each bar has height equal to 0.02 times a binomial random variable with parameters 100 and probability that X_1 is in the interval under the bar. The bars in Fig. 6.2 have approximately one-half of the probability of the bars in Fig. 6.1, but their heights have been multiplied by 2. By part (b), we expect the heights in Fig. 6.2 to have approximately twice the variance of the heights in Fig. 6.1.

18. The result is trivial if the m.g.f. is infinite for all $s > 0$. So, assume that the m.g.f. is finite for at least some $s > 0$. For every t and every $s > 0$ such that the m.g.f. is finite, we can write

$$\Pr(X > t) = \Pr(\exp(sX) > \exp(st) \leq \frac{E(\exp(sX))}{\exp(st)} = \psi(s)\exp(-st),$$

where the second equality follows from the Markov inequality. Since $\Pr(X > t) \leq \psi(s)\exp(-st)$ for every s $\Pr(Y > t) \leq \min_s \psi(s)\exp(-st)$.

19. (a) First, insert s from (6.2.15) into the expression in (6.2.14). We get

$$n\left[\log(p)+\left(\frac{1-p}{p}+u\right)\log\left\{\frac{(1+u)p+1-p}{up+1-p}(1-p)\right\}-\log\left\{1-\frac{1-p}{\frac{(1+u)p+1-p}{up+1-p}(1-p)}\right\}\right].$$

The last term can be rewritten as

$$-\log\left\{1-\frac{up+1-p}{(1+u)p+1-p}\right\}=-\log(p)+\log\left\{(1+u)p+1-p\right\}.$$

The result is then

$$n\left[\left(\frac{1-p}{p}+u\right)\log\left\{\frac{(1+u)p+1-p}{up+1-p}(1-p)\right\}+\log\left\{(1+u)p+1-p\right\}\right].$$

This is easily recognized as n times the logarithm of (6.2.16).

(b) For all u, q is given by (6.2.16). For $u=0$, $q=(1-p)^{(1-p)/p}$. Since $0<1-p<1$ and $(1-p)/p>0$, we have $0<q<1$ when $u=0$. For general u, let $x=p(1+u)+1-p$ and rewrite

$$\log(q)=\log(p+x)+\frac{p+x}{p}\log\frac{(1-p)(p+x)}{x}.$$

Since x is a linear increasing function of u, if we show that $\log(q)$ is decreasing in x, then q is decreasing in u. The derivative of $\log(q)$ with respect to x is

$$-\frac{p}{x(p+x)}+\frac{1}{p}\log\frac{(1-p)(p+x)}{x}.$$

The first term is negative, and the second term is negative at $u=0$ ($x=1$). To be sure that the sum is always negative, examine the second term more closely. The derivative of the second term is

$$\frac{1}{p}\left(\frac{1}{p+x}-\frac{1}{x}\right)=\frac{-1}{x(p+x)}<0.$$

Hence, the derivative is always negative, and q is less than 1 for all u.

20. We already have the m.g.f. of Y in (6.2.9). We can multiply it by $e^{-sn/10}$ and minimize over $s>0$. Before minimizing, take the logarithm:

$$\log[\psi(s)e^{-sn/10}]=n\left[\log(1/2)+\log[exp(s)+1]-\frac{3s}{5}\right]. \tag{S.6.7}$$

The derivative of this logarithm is

$$n\left[\frac{\exp(s)}{\exp(s)+1}-\frac{3}{5}\right].$$

The derivative is 0 at $s=\log(3/2)$, and the second derivative is positive there, so $s=\log(3/2)$ provides the minimum. The minimum value of (S.6.7) is -0.02014, and the Chernoff bound is $\exp(-0.02014n)=(0.98)^n$ for $\Pr(Y>n/10)$. Similarly, for $\Pr(-Y>n/10)$, we need to minimize

$$\log[\psi(-s)e^{-sn/10}]=n\left[\log(1/2)+\log[exp(-s)+1]+\frac{2s}{5}\right]. \tag{S.6.8}$$

The derivative is

$$n\left[\frac{-\exp(-s)}{\exp(-s)+1}+\frac{2}{5}\right],$$

which equals 0 at $s = \log(3/2)$. The minimum value of (S.6.8) is again -0.02014, and the Chernoff bound for the entire probability is $2(0.98)^n$, a bit smaller than in the example.

21. (a) The m.g.f. of the exponential distribution with parameter 1 is $1/(1-s)$ for $s < 1$, hence the m.g.f. of Y_n is $1/(1-s)^n$ for $t < 1$. The Chernoff bound is the minimum (over $s > 0$) of $e^{-nus}/(1-s)^n$. The logarithm of this is $-n[us + \log(1-s)]$, which is minimized at $s = (u-1)/u$, which is positive if and only if $u > 1$. The Chernoff bound is $[u \exp(1-u)]^n$.

    (b) If $u < 1$, then the expression in Theorem 6.2.7 is minimized over $s > 0$ near $s = 0$, which provides a useless bound of 1 for $\Pr(Y_n > nu)$.

22. (a) The numbers $(k-1)k/2$ for $k = 1, 2, \ldots$ form a strictly increasing sequence starting at 0. Hence, every integer n falls between a unique pair of these numbers. So, k_n is the value of k such that n is larger than $(k-1)k/2$ but no larger than $k(k+1)/2$.

    (b) Clearly j_n is the excess of n over the lower bound in part (a), hence j_n runs from 1 up to the difference between the bounds, which is easily seen to be k_n.

    (c) The intervals where h_n equals 1 are defined to be disjoint for $j_n = 1, \ldots, k_n$, and they cover the whole interval $[0, 1)$. Hence, for each x $h_n(x) = 1$ for one and only one of these intervals, which correspond to n between the bounds in part (a).

    (d) For every $x \in [0, 1)$, $h_n(x) = 1$ for one n between the bounds in part (a). Since there are infintely many values of k_n, $h_n(x) = 1$ infintely often for every $x \in [0, 1)$, and $\Pr(X \in [0, 1)) = 1$.

    (e) For every $\epsilon > 0$ $|Z_n - 0| > \epsilon$ whenever $Z_n = 1$. Since $\Pr(Z_n = 1 \text{ infinitely often}) = 1$, the probability is 1 that Z_n fails to converge to 0. Hence, the probability is 0 that Z_n does converge to 0.

    (f) Notice that $h_n(x) = 1$ on an interval of length $1/k_n$. Hence, for each n, $\Pr(|Z_n - 0| > \epsilon) = 1/k_n$, which goes to 0. So, $Z_n \xrightarrow{P} 0$.

23. Each Z_n has the Bernoulli distribution with parameter $1/k_n$, hence $E[(Z_n - 0)^2] = 1/k_n$, which goes to 0.

24. (a) By construction, $\{Z_n \text{ converges to } 0\} = \{X > 0\}$. Since $\Pr(X > 0) = 1$, we have Z_n converges to 0 with probability 1.

    (b) $E[(Z_n - 0)^2] = E(Z_n^2) = n^4/n$, which does not go to 0.

6.3 The Central Limit Theorem

Commentary

The delta method is introduced as a practical application of the central limit theorem. The examples of the delta method given in this section are designed to help pave the way for some approximate confidence interval calculations that arise in Sec. 8.5. The delta method also helps in calculating the approximate distributions of some summaries of simulations that arise in Sec. 12.2. This section ends with two theoretical topics that might be of interest only to the more mathematically inclined students. The first is a central limit theorem for random variables that don't have identical distributions. The second is an outline of the proof of the i.i.d. central limit theorem that makes use of moment generating functions.

Solutions to Exercises

1. The length of rope produced in one hour X has a mean of $60 \times 4 = 240$ feet and a standard deviation of $60^{1/2} \times 5 = 38.73$ inches, which is 3.23 feet. The probability that $X \geq 250$ is approximately the probability that a normal random variable with mean 240 and standard deviation 3.23 is at least 250, namely $1 - \Phi([250 - 240]/3.23) = 1 - \Phi(3.1) = 0.001$.

2. The total number of people X from the suburbs attending the concert can be regarded as the sum of 1200 independent random variables, each of which has a Bernoulli distribution with parameter $p = 1/4$. Therefore, the distribution of X will be approximately a normal distribution with mean $1200(1/4) = 300$ and variance $1200(1/4)(3/4) = 225$. If we let $Z = (X - 300)/15$, then the distribution of Z will be approximately a standard normal distribution. Hence,

$$\Pr(X < 270) = \Pr(Z < -2) \simeq 1 - \Phi(2) = 0.0227.$$

3. Since the variance of a Poisson distribution is equal to the mean, the number of defects on any bolt has mean 5 and variance 5. Therefore, the distribution of the average number $\overline{X}_n$ on the 125 bolts will be approximately the normal distribution with mean 5 and variance $5/125 = 1/25$. If we let $Z = (\overline{X}_n - 5)/(1/5)$, then the distribution of Z will be approximately a standard normal distribution. Hence,

$$\Pr(\overline{X}_n < 5.5) = \Pr(Z < 2.5) \simeq \Phi(2.5) = 0.9938.$$

4. The distribution of $Z = \sqrt{n}(\overline{X}_n - \mu)/3$ will be approximately the standard normal distribution. Therefore,

$$\Pr(|\overline{X}_n - \mu| < 0.3) = \Pr(|Z| < 0.1\sqrt{n}) \simeq 2\Phi(0.1\sqrt{n}) - 1.$$

But $2\Phi(0.1\sqrt{n}) - 1 \geq 0.95$ if and only if $\Phi(0.1\sqrt{n}) \geq (1+0.95)/2 = 0.975$, and this inequality is satisfied if and only if $0.1\sqrt{n} \geq 1.96$ or, equivalently, $n \geq 384.16$. Hence, the smallest possible value of n is 385.

5. The distribution of the proportion $\overline{X}_n$ of defective items in the sample will be approximately the normal distribution with mean 0.1 and variance $(0.1)(0.9)/n = 0.09/n$. Therefore, the distribution of $Z = \sqrt{n}(\overline{X}_n - 0.1)/0.3$ will be approximately the standard normal distribution. It follows that

$$\Pr(\overline{X}_n < 0.13) = \Pr(Z < 0.1\sqrt{n}) \simeq \Phi(0.1\sqrt{n}).$$

For this value to be at least 0.99, we must have $0.1\sqrt{n} \geq 2.327$ or, equivalently, $n \geq 541.5$. Hence, the smallest possible value of n is 542.

6. The distribution of the total number of times X that the target is hit will be approximately the normal distribution with mean $10(0.3) + 15(0.2) + 20(0.1) = 8$ and variance $10(0.3)(0.7) + 15(0.2)(0.8) + 20(0.1)(0.9) = 6.3$. Therefore, the distribution of $Z = (X - 8)/\sqrt{6.3} = (X - 8)/2.51$ will be approximately a standard normal distribution. It follows that

$$\Pr(X \geq 12) = \Pr(Z \geq 1.5936) \simeq 1 - \Phi(1.5936) = 0.0555.$$

7. The mean of a random digit X is

$$E(X) = \frac{1}{10}(0 + 1 + \cdots + 9) = 4.5.$$

Also,

$$E(X^2) = \frac{1}{10}(0^2 + 1^2 + \cdots + 9^2) = \frac{1}{10} \cdot \frac{(9)(10)(19)}{6} = 28.5.$$

Therefore, $\text{Var}(X) = 28.5 - (4.5)^2 = 8.25$. The distribution of the average $\overline{X}_n$ of 16 random digits will therefore be approximately the normal distribution with mean 4.5 and variance $8.25/16 = 0.5156$. Hence, the distribution of

$$Z = \frac{\overline{X}_n - 4.5}{\sqrt{0.5156}} = \frac{\overline{X}_n - 4.5}{0.7181}$$

will be approximately a standard normal distribution. It follows that

$$\begin{aligned}\Pr(4 \le \overline{X}_n \le 6) &= \Pr(-0.6963 \le Z \le 2.0888) \\ &= \Phi(2.0888) - [1 - \Phi(0.6963)] \\ &= 0.9816 - 0.2431 = 0.7385.\end{aligned}$$

8. The distribution of the total amount X of 36 drinks will be approximately the normal distribution with mean $36(2) = 72$ and variance $36(1/4) = 9$. Therefore, the distribution of $Z = (X - 72)/3$ will be approximately a standard normal distribution. It follows that

$$\Pr(X < 63) = \Pr(Z < -3) = 1 - \Phi(3) = 0.0013.$$

9. (a) By Eq. (6.2.4),

$$\Pr\left(|\overline{X}_n - \mu| \ge \frac{\sigma}{4}\right) \le \frac{\sigma^2}{n} \cdot \frac{16}{\sigma^2} = \frac{16}{25}.$$

Therefore, $\Pr\left(|\overline{X}_n - \mu| \le \frac{\sigma}{4}\right) \ge 1 - \frac{16}{25} = 0.36.$

(b) The distribution of

$$Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{5}{\sigma}(\overline{X}_n - \mu)$$

will be approximately a standard normal distribution. Therefore,

$$\Pr\left(|\overline{X}_n - \mu| \le \frac{\sigma}{4}\right) = \Pr\left(|Z| \le \frac{5}{4}\right) \simeq 2\Phi(1.25) - 1 = 0.7887.$$

10. (a) As in part (a) of Exercise 9,

$$\Pr\left(|\overline{X}_n - \mu| \le \frac{\sigma}{4}\right) \ge 1 - \frac{16}{n}.$$

Now $1 - 16/n \ge 0.99$ if and only if $n \ge 1600$.

(b) As in part (b) of Exercise 9,

$$\Pr\left(|\overline{X}_n - \mu| \le \frac{\sigma}{4}\right) = \Pr\left(|Z| \le \frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1.$$

Now $2\Phi\left(\sqrt{n}/4\right) - 1 \ge 0.99$ if and only if $\Phi\left(\sqrt{n}/4\right) \ge 0.995$. This inequality will be satisfied if and only if $\sqrt{n}/4 \ge 2.567$ or, equivalently, $n \ge 105.4$. Therefore, the smallest possible sample size is 106.

11. For a student chosen at random, the number of parents X who will attend the graduation ceremony has mean $\mu = 0/3 + 1/3 + 2/3 = 1$ and variance $\sigma^2 = E[(X-\mu)^2] = (0-1)^2/3 + (1-1)^2/3 + (2-1)^2/3 = 2/3$. Therefore, the distribution of the total number of parents W who attend the ceremony will be approximately the normal distribution with mean $(600)(1) = 600$ and variance $600(2/3) = 400$. Therefore, the distribution of $Z = (W-600)(20)$ will be approximately a standard normal distribution. It follows that

$$\Pr(W \leq 650) = \Pr(Z \leq 2.5) = \Phi(2.5) = 0.9938.$$

12. The m.g.f. of the binomial distribution with parameters n and p_n is $\psi_n(t) = (p_n \exp(t) + 1 - p_n)^n$. If $np_n \to \lambda$,

$$\lim_{n\to\infty} \psi_n(t) = \lim_{n\to\infty} \left(1 + \frac{np_n}{n}[\exp(t) - 1]\right)^n.$$

This converges to $\exp(\lambda[e^t - 1])$, which is the m.g.f. of the Poisson distribution with mean λ.

13. We are asking for the asymptotic distribution of $g(\overline{X}_n)$, where $g(x) = x^3$. The distribution of $\overline{X}_n$ is normal with mean θ and variance σ^2/n. According to the delta method, the asymptotic distribution of $g(\overline{X}_n)$ should be the normal distribution with mean $g(\theta) = \theta^3$ and variance $(\sigma^2/n)[g'(\theta)]^2 = 9\theta^4\sigma^2/n$.

14. First, note that $Y_n = \sum_{i=1}^n X_i^2/n$ has asymptotically the normal distribution with mean σ^2 and variance $2\sigma^4/n$. Here, we have used the fact that $E(X_i^2) = \sigma^2$ and $E(X_i^4) = 2\sigma^4$.

 (a) Let $g(x) = 1/x$. Then $g'(x) = -1/x^2$. So, the asymptotic distribution of $g(Y_n)$ is the normal distribution with mean $1/\sigma^2$ and variance $(2\sigma^4/n)/\sigma^8 = 2/[n\sigma^4]$.

 (b) Let $h(\mu) = 2mu^2$. If the asymptotic mean of Y_n is μ the asymptotic variance of Y_n is $h(\mu)/n$. So, a variance stabilizing transformation is

$$\alpha(\mu) = \int_a^\mu \frac{dx}{2^{1/2}x} = \frac{1}{2^{1/2}} \log(\mu),$$

 where we have taken $a = 1$ to make the integral finite. So the asymptotic distribution of $\log(Y_n)/2^{1/2}$ is the normal distribution with mean $2\log(\sigma)/2^{1/2}$ and variance $1/n$.

15. (a) Clearly, $Y_n \leq y$ if and only if $X_i \leq y$ for $i = 1, \ldots, n$. Hence,

$$\Pr(Y_n \leq y) = \Pr(X_1 \leq y)^n = \begin{cases} (y/\theta)^n \text{if } 0 < y < \theta, & \\ 0 & \text{if } y \leq 0, \\ 1 & \text{if } y \geq \theta. \end{cases}$$

 (b) The c.d.f. of Z_n is, for $z < 0$,

$$\Pr(Z_n \leq z) = \Pr(Y_n \leq \theta + z/n) = (1 + z/[n\theta])^n. \quad \text{(S.6.9)}$$

 Since $Z_n \leq 0$, the c.d.f. is 1 for $z \geq 0$. According to Theorem 5.3.3, the expression in (S.6.9) converges to $\exp(z/\theta)$.

 (c) Let $\alpha(y) = y^2$. Then $\alpha'(y) = 2y$. We have $n(Y_n - \theta)$ converging in distribution to the c.d.f. in part (b). The delta method says that, for $\theta > 0$, $n(Y_n^2 - \theta^2)/[2\theta]$ converges in distribution to the same c.d.f.

6.4 The Correction for Continuity

Solutions to Exercises

1. The mean of X_i is 1 and the mean of X_i^2 is 1.5. So, the variance of X_i is 0.5. The central limit theorem says that $Y = X_1 + \cdots + X_{30}$ has approximately the normal distribution with mean 30 and variance 15. We want the probability that $Y \le 33$. Using the correction for continuity, we would assume that Y has the normal distribution with mean 30 and variance 15 and compute the probability that $Y \le 33.5$. This is $\Phi([33.5 - 30]/15^{1/2}) = \Phi(0.904) = 0.8169$.

2. (a) $E(X) = 15(.3) = 4.5$ and $\sigma_X = [(15)(.3)(.7)]^{1/2} = 1.775$. Therefore,

$$\begin{aligned}\Pr(X = 4) &= \Pr(3.5 \le X \le 4.5) = \Pr\left(\frac{3.5 - 4.5}{1.775} \le Z \le 0\right)\\ &= \Pr(-.5634 \le Z \le 0) \approx \Phi(.5634) - .5 \approx .214.\end{aligned}$$

   (b) The exact value is found from the table of binomial probabilities ($n = 15$, $p = 0.3, k = 4$) to be .2186.

3. In the notation of Example 2,

$$\Pr(H > 495) = \Pr(H \ge 495.5) = \Pr\left(Z \ge \frac{495.5 - 450}{15}\right) \approx 1 - \Phi(3.033) \approx .0012.$$

4. We follow the notation of the solution to Exercise 2 of Sec. 6.3:

$$\Pr(X < 270) = \Pr(X \le 269.5) = \Pr\left(Z \le \frac{269.5 - 300}{15}\right) \approx 1 - \Phi(2.033) \approx .0210.$$

5. Let X denote the total number of defects in the sample. Then X has a Poisson distribution with mean $5(125) = 625$, so σ_X is $(625)^{1/2} = 25$. Hence,

$$\Pr(\overline{X}_n < 5.5) = \Pr[X < 125(5.5)] = \Pr(X < 687.5).$$

   Since this final probability is just the value that would be used with the correction for continuity, the probability to be found here is the same as that originally found in Exercise 3 of Sec. 6.3.

6. We follow the notation of the solution to Exercise 6 of Sec. 6.3:

$$\Pr(X \ge 12) = \Pr(X \ge 11.5) = \Pr\left(Z \ge \frac{11.5 - 8}{2.51}\right) \approx 1 - \Phi(1.394) \approx .082.$$

7. Let S denote the sum of the 16 digits. Then

$$E(S) = 16(4.5) = 72 \quad \text{and} \quad \sigma_X = [16(8.25)]^{1/2} = 11.49.$$

   Hence,

$$\begin{aligned}\Pr(4 \le \overline{X}_n \le 6) &= \Pr(64 \le S \le 96) = \Pr(63.5 \le S \le 96.5)\\ &= \Pr\left(\frac{63.5 - 72}{11.49} \le Z \le \frac{96.5 - 72}{11.49}\right)\\ &\approx \Phi(2.132) - \Phi(-.740) \approx .9835 - .2296 = .7539.\end{aligned}$$

6.5 Supplementary Exercises

Solutions to Exercises

1. By the central limit theorem, the distribution of X is approximately normal with mean $(120)(1/6) = 20$ and standard deviation $[120(1/6)(5/6)]^{1/2} = 4.082$. Let $Z = (X - 20)/4.082$. Then from the table of the standard normal distribution we find that $\Pr(|Z| \leq 1.96) = .95$. Hence, $k = (1.96)(4.082) = 8.00$.

2. Because of the property of the Poisson distribution described in Theorem 5.4.4, the random variable X can be thought of as the sum of a large number of i.i.d. random variables, each of which has a Poisson distribution. Hence, the central limit theorem (Lindeberg and Lévy) implies the desired result. It can also be shown that the m.g.f. of X converges to the m.g.f. of the standard normal distribution.

3. By the previous exercise, X has approximately a normal distribution with mean 10 and standard deviation $(10)^{1/2} = 3.162$. Thus, without the correction for continuity,

$$\Pr(8 \leq X \leq 12) = \Pr\left(\frac{8-10}{3.162} \leq Z \leq \frac{12-10}{3.162}\right) \approx \Phi(.6325) - \Phi(-.6325) = .473.$$

With the correction for continuity, we find

$$\Pr(7.5 \leq X \leq 12.5) = \Pr\left(-\frac{2.5}{3.162} \leq Z \leq \frac{2.5}{3.162}\right) \approx \Phi(.7906) - \Phi(-.7906) = .571.$$

The exact probability is found from the Poisson table to be

$$(.1126) + (.1251) + (.1251) + (.1137) + (.0948) = .571.$$

Thus, the approximation with the correction for continuity is almost perfect.

4. If X has p.d.f. $f(x)$, then

$$E(X^k) = \int_0^\infty x^k f(x)dx \geq \int_t^\infty x^k f(x)dx \geq t^k \int_t^\infty f(x)dx = t^k \Pr(X \geq t).$$

A similar proof holds if X has a discrete distribution.

5. The central limit theorem says that $\overline{X}_n$ has approximately the normal distribution with mean p and variance $p(1-p)/n$. A variance stabilizing transformation will be

$$\alpha(x) = \int_a^x [p(1-p)]^{-1/2} dp.$$

To perform this integral, transform to $z = p^{1/2}$, that is, $p = z^2$. Then

$$\alpha(x) = \int_{a^{1/2}}^{x^{1/2}} \frac{dz}{(1-z^2)^{1/2}}.$$

Next, transform so that $z = \sin(w)$ or $w = \arcsin(z)$. Then $dz = \cos(w)dw$ and

$$\alpha(x) = \int_{\arcsin a^{1/2}}^{\arcsin x^{1/2}} dw = \arcsin x^{1/2},$$

where we have chosen $a = 0$. The variance stabilizing transformation is $\alpha(x) = \arcsin(x^{1/2})$.

6. According to the central limit theorem, $\overline{X}_n$ has approximately the normal distribution with mean θ and variance θ^2. A variance stabilizing transformation will be

$$\alpha(x) = \int_a^x \theta^{-1} d\theta = \log(x),$$

where we have used $a = 1$.

7. Let F_n be the c.d.f. of X_n. The most direct proof is to show that $\lim_{n\to\infty} F_n(x) = F(x)$ for every point at which F is continuous. Since F is the c.d.f. of an integer-valued distribution, the continuity points are all non-integer values of x together with those integer values of x to which F assigns probability 0. It is clear, that it suffices to prove that $\lim_{n\to\infty} F_n(x) = F(x)$ for every non-integer x, because continuity of F from the right and the fact that F is nondecreasing will take care of the integers with zero probability. For each non-integer x, let m_x be the largest integer such that $m < x$. Then

$$F_n(x) = \sum_{k=1}^{m} \Pr(X_n = k) \to \sum_{k=1}^{m} f(k) = F(m) = F(x),$$

where the convergence follows because the sums are finite.

8. We know that $\Pr(X_n = m) = \binom{k}{m} p_n^m (1 - p_n)^{k-m}$ for $m = 0, \ldots, k$ and all n. We also know that

$$\lim_{n\to\infty} \binom{k}{m} p_n^m (1 - p_n)^{k-m} = \binom{k}{m} p^m (1 - p)^{k-m},$$

for all m. By Exercise 7, X_n converges in distribution to the binomial distribution with parameters k and p.

9. Let $X_1, \ldots, X_{16}$ be the times required to serve the 16 customers. The parameter of the exponenital distribution is 1/3. According to Theorem 5.7.8, the mean and variance of each X_i are 3 and 9 respectively. Let $\sum_{k=1}^{16} X_k = Y$ be the total time. The central limit theorem approximation to the distribution of Y is the normal distribution with mean $16 \times 3 = 48$ and variance $16 \times 9 = 144$. The approximate probablity that $Y > 60$ is

$$1 - \Phi\left(\frac{60 - 48}{(144)^{1/2}}\right) = 1 - \Phi(1) = 0.1587.$$

The actual distribution of Y is the gamma distribution with parameters 16 and 1/3. Using the gamma c.d.f., the probability is 0.1565.

10. The number of defects in 2000 square-feet has the Poisson distribution with mean $2000 \times 0.01 = 20$. The central limit theorem approximation is the normal distribution with mean 20 and variance 20. Without correction for continuity, the approximate probability of at least 15 defects is

$$1 - \Phi\left(\frac{15 - 20}{(20)^{1/2}}\right) = 1 - \Phi(-1.1180) = 0.8682.$$

With the continuity correction, we get

$$1 - \Phi\left(\frac{14.5 - 20}{(20)^{1/2}}\right) = 1 - \Phi(-1.2298) = 0.8906.$$

The actual Poisson probability is 0.8951.

11. (a) The gamma distribution with parameters n and 3 is the distribution of the sum of n i.i.d. exponential random variables with parameter 3. If n is large, the central limit theorem should apply to approximate the distribution of the sum of n exponentials.

    (b) The mean and variance of each exponential random variable are 1/3 and 1/9 respectively. The distribution of the sum of n of these has approximately the normal distribution with mean $n/3$ and variance $n/9$.

12. (a) The exponential distribution with parameters n and 0.2 is the distribution of the sum of n i.i.d. geometric random variables with parameter 0.2. If n is large, the central limit theorem should apply to approximate the distribution of the sum of n geometrics.

    (b) The mean and variance of each geometric random variable are $0.8/0.2 = 4$ and $0.8/(0.2)^2 = 20$. The distribution of the sum of n of these has approximately the normal distribution with mean $4n$ and variance $20n$.

Chapter 7

Estimation

7.1 Statistical Inference

Commentary

Many students find statistical inference much more difficult to comprehend than elementary probability theory. For this reason, many examples of statistical inference problems have been introduced in the early chapters of this text. This will give instructors the opportunity to point back to relatively easy-to-understand examples that the students have already learned as a preview of what is to come. In addition to the examples mentioned in Sec. 7.1, some additional examples are Examples 2.3.3–2.3.5, 3.6.9, 3.7.14, 3.7.18, 4.8.9–4.8.10, and 5.8.1–5.8.2. In addition, the discussion of M.S.E. and M.A.E. in Sec. 4.5 and the discussion of the variance of the sample mean in Sec. 6.2 contain inferential ideas. Most of these are examples of Bayesian inference because the most common part of a Bayesian inference is the calculation of a conditional distribution or a conditional mean.

Solutions to Exercises

1. The random variables of interest are the observables $X_1, X_2, \ldots$ and the hypothetically observable (parameter) P. The X_i's are i.i.d. Bernoulli with parameter p given $P = p$.

2. The statistical inferences mentioned in Example 7.1.3 are computing the conditional distribution of P given observed data, computing the conditional mean of P given the data, and computing the M.S.E. of predictions of P both before and after observing data.

3. The random variables of interest are the observables $Z_1, Z_2, \ldots$, the times at which successive particles hit the target, and β, the hypothetically observable (parameter) rate of the Poisson process. The hit times occur according to a Poisson process with rate β conditional on β. Other random variables of interest are the observable inter-arrival times $Y_1 = Z_1$, and $Y_k = Z_k - Z_{k-1}$ for $k \geq 2$.

4. The random variables of interest are the observable heights $X_1, \ldots, X_n$, the hypothetically observable mean (parameter) μ, and the sample mean $\overline{X}_n$. The X_i's are modeled as normal random variables with mean μ and variance 9 given μ.

5. The statement that the interval $(\overline{X}_n - 0.98, \overline{X}_n + 0.98)$ has probability 0.95 of containing μ is an inference.

6. The random variables of interest are the observable number X of Mexican-American grand jurors and the hypothetically observable (parameter) P. The conditional distribution of X given $P = p$ is the

binomial distribution with parameters 220 and p. Also, P has the beta distribution with parameters α and β, which have not yet been specified.

7. The random variables of interest are Y, the hypothetically observable number of oocysts in t liters, the hypothetically observable indicators $X_1, X_2, \ldots$ of whether each oocyst is counted, X the observable count of oocysts, the probability (parameter) p of each oocyst being counted, and the (parameter) λ the rate of oocysts per liter. We model Y as a Poisson random variable with mean $t\lambda$ given λ. We model $X_1, \ldots, X_y$ as i.i.d. Bernoulli random variables with parameter p given p and given $Y = y$. We define $X = X_1 + \ldots + X_y$.

7.2 Prior and Posterior Distributions

Commentary

This section introduces some common terminology that is used in Bayesian inference. The concepts should all be familiar already to the students under other names. The prior distribution is just a marginal distribution while the posterior distribution is just a conditional distribution. The likelihood function might seem strange since it is a conditional density for the data given θ but thought of as a function of θ after the data have been observed.

Solutions to Exercises

1. We still have $y = 16178$, the sum of the five observed values. The posterior distribution of β is now the gamma distribution with parameters 6 and 21178. So,

$$\begin{aligned} f(x_6|\boldsymbol{x}) &= \int_0^\infty 7.518 \times 10^{23} \beta^5 \exp(-21178\beta)\beta \exp(-x_6\beta) d\beta \\ &= 7.518 \times 10^{23} \int_0^\infty \beta^6 \exp(-\beta[21178 + x_6]) d\beta \\ &= 7.518 \times 10^{23} \frac{\Gamma(7)}{(21178 + x_6)^7} = \frac{5.413 \times 10^{26}}{(21178 + x_6)^7}, \end{aligned}$$

for $x_6 > 0$. We can now compute $\Pr(X_6 > 3000|\boldsymbol{x})$ as

$$\Pr(X_6 > 3000|\boldsymbol{x}) = \int_{3000}^\infty \frac{5.413 \times 10^{26}}{(21178 + x_6)^7} dx_6 = \frac{5.413 \times 10^{26}}{6 \times 24178^6} = 0.4516.$$

2. The joint p.f. of the eight observations is given by Eq. (7.2.11). Since $n = 8$ and $y = 2$ in this exercise,

$$f_n(\boldsymbol{x} \mid \theta) = \theta^2(1-\theta)^6.$$

Therefore,

$$\begin{aligned} \xi(0.1 \mid \boldsymbol{x}) = \Pr(\theta = 0.1 \mid \boldsymbol{x}) &= \frac{\xi(0.1) f_n(\boldsymbol{x} \mid 0.1)}{\xi(0.1) f_n(\boldsymbol{x} \mid 0.1) + \xi(0.2) f_n(\boldsymbol{x} \mid 0.2)} \\ &= \frac{(0.7)(0.1)^2(0.9)^6}{(0.7)(0.1)^2(0.9)^6 + (0.3)(0.2)^2(0.8)^6} \\ &= 0.5418. \end{aligned}$$

It follows that $\xi(0.2 \mid \boldsymbol{x}) = 1 - \xi(0.1 \mid \boldsymbol{x}) = 0.4582$.

3. Let X denote the number of defects on the selected roll of tape. Then for any given value of λ, the p.f. of X is.

$$f(x \mid \lambda) = \frac{\exp(-\lambda)\lambda^x}{x!} \quad \text{for} \quad x = 0, 1, 2, \ldots.$$

Therefore,

$$\xi(1.0 \mid X = 3) = \Pr(\lambda = 1.0 \mid X = 3) = \frac{\xi(1.0)f(3 \mid 1.0)}{\xi(1.0)f(3 \mid 1.0) + \xi(1.5)f(3 \mid 1.5)}.$$

From the table of the Poisson distribution in the back of the book it is found that

$$f(3 \mid 1.0) = 0.0613 \quad \text{and} \quad f(3 \mid 1.5) = 0.1255.$$

Therefore, $\xi(1.0 \mid X = 3) = 0.2456$ and $\xi(1.5 \mid X = 3) = 1 - \xi(1.0 \mid X = 3) = 0.7544$.

4. If α and β denote the parameters of the gamma distribution, then we must have

$$\frac{\alpha}{\beta} = 10 \quad \text{and} \quad \frac{\alpha}{\beta^2} = 5.$$

Therefore, $\alpha = 20$ and $\beta = 2$. Hence, the prior p.d.f. of θ is as follows, for $\theta > 0$:

$$\xi(\theta) = \frac{2^{20}}{\Gamma(20)}\,\theta^{19}\exp(-2\theta).$$

5. If α and β denote the parameters of the beta distribution, then we must have

$$\frac{\alpha}{\alpha+\beta} = \frac{1}{3} \quad \text{and} \quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{2}{90}.$$

Since $\dfrac{\alpha}{\alpha+\beta} = \dfrac{1}{3}$, it follows that $\dfrac{\beta}{(\alpha+\beta)} = \dfrac{2}{3}$. Therefore,

$$\frac{\alpha\beta}{(\alpha+\beta)^2} = \frac{\alpha}{\alpha+\beta}\cdot\frac{\beta}{\alpha+\beta} = \frac{1}{3}\cdot\frac{2}{3} = \frac{2}{9}.$$

It now follows from the second equation that $\dfrac{2}{9(\alpha+\beta+1)} = \dfrac{2}{90}$ and, hence, that $\alpha+\beta+1 = 10$. Therefore, $\alpha+\beta = 9$ and it follows from the first equation that $\alpha = 3$ and $\beta = 6$. Hence, the prior p.d.f. of θ is as follows, for $0 < \theta < 1$:

$$\xi(\theta) = \frac{\Gamma(9)}{\Gamma(3)\Gamma(6)}\,\theta^2(1-\theta)^5.$$

6. The conditions of this exercise are precisely the conditions of Example 7.2.7 with $n = 8$ and $y = 3$. Therefore, the posterior distribution of θ is a beta distribution with parameters $\alpha = 4$ and $\beta = 6$.

7. Since $f_n(\boldsymbol{x} \mid \theta)$ is given by Eq. (7.2.11) with n = 8 and y = 3, then

$$f_n(\boldsymbol{x} \mid \theta)\xi(\theta) = 2\theta^3(1-\theta)^6.$$

When we compare this expression with Eq. (5.8.3), we see that it has the same form as the p.d.f. of a beta distribution with parameters $\alpha = 4$ and $\beta = 7$. Therefore, this beta distribution is the posterior distribution of θ.

8. By Eq. (7.2.14),

$$\xi(\theta \mid x_1) \propto f(x_1 \mid \theta)\xi(\theta),$$

and by Eq. (7.2.15),

$$\xi(\theta \mid x_1, x_2) \propto f(x_2 \mid \theta)\xi(\theta \mid x_1).$$

Hence,

$$\xi(\theta \mid x_1, x_2) \propto f(x_1 \mid \theta)f(x_2 \mid \theta)\xi(\theta).$$

By continuing in this way, we find that

$$\xi(\theta \mid x_1, x_2, x_3) \propto f(x_3 \mid \theta)\xi(\theta \mid x_1, x_2) \propto f(x_1 \mid \theta)f(x_2 \mid \theta)f(x_3 \mid \theta)\xi(\theta).$$

Ultimately, we will find that

$$\xi(\theta \mid x_1, \dots, x_n) \propto f(x_1 \,|\, \theta) \dots f(x_n \,|\, \theta)\xi(\theta).$$

From Eq. (7.2.4) it follows that, in vector notation, this relation can be written as

$$\xi(\theta \mid \boldsymbol{x}) \propto f_n(\boldsymbol{x} \mid \theta)\xi(\theta),$$

which is precisely the relation (7.2.10). Hence, when the appropriate factor is introduced on the right side of this relation so that the proportionality symbol can be replaced by an equality, $\xi(\theta \mid \boldsymbol{x})$ will be equal to the expression given in Eq. (7.2.7).

9. It follows from Exercise 8 that if the experiment yields a total of three defectives and five nondefectives, the posterior distribution will be the same regardless of whether the eight items were selected in one batch or one at a time in accordance with some stopping rule. Therefore, the posterior distribution in this exercise will be the same beta distribution as that obtained in Exercise 6.

10. In this exercise

$$f(x|\theta) = \begin{cases} 1 & \text{for } \theta - \dfrac{1}{2} < x < \theta + \dfrac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\xi(\theta) = \begin{cases} \dfrac{1}{10} & \text{for } 10 < \theta < 20, \\ 0 & \text{otherwise.} \end{cases}$$

The condition that $\theta - 1/2 < x < \theta + 1/2$ is the same as the condition that $x - 1/2 < \theta < x + 1/2$. Therefore, $f(x \mid \theta)\xi(\theta)$ will be positive only for values of θ which satisfy both the requirement that $x - 1/2 < \theta < x + 1/2$ and the requirement that $10 < \theta < 20$. Since $X = 12$ in this exercise, $f(x \mid \theta)\xi(\theta)$ is positive only for $11.5 < \theta < 12.5$. Furthermore, since $f(x \mid \theta)\xi(\theta)$ is constant over this interval, the posterior p.d.f. $\xi(\theta \mid x)$ will also be constant over this interval. In other words, the posterior distribution of θ must be a uniform distribution on this interval.

11. Let y_1 denote the smallest and let y_6 denote the largest of the six observations. Then the joint p.d.f. of the six observations is

$$f_n(\boldsymbol{x} \mid \theta) = \begin{cases} 1 & \text{for } \theta - \dfrac{1}{2} < y_1 < y_6 < \theta + \dfrac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

The condition that $\theta - \dfrac{1}{2} < y_1 < y_6 < \theta + \dfrac{1}{2}$ is the same as the condition that $y_6 - 1/2 < \theta < y_1 + 1/2$. Since $\xi(\theta)$ is again as given in Exercise 10, it follows that $f_n(\boldsymbol{x} \mid \theta)\xi(\theta)$ will be positive only for values of θ which satisfy both the requirement that $10 < \theta < 20$. Since $y_1 = 10.9$ and $y_6 = 11.7$ in this exercise, $f_n(\boldsymbol{x} \mid \theta)\xi(\theta)$ is positive only for $y_6 - 1/2 < \theta < y_1 + 1/2$ and the requirement that $10 < \theta < 20$. Since $y_1 = 10.9$ and $y_6 = 11.7$ in this exercise, $f_n(\boldsymbol{x} \mid \theta)\xi(\theta)$ is positive only for $11.2 < \theta < 11.4$. Furthermore, since $f_n(\boldsymbol{x} \mid \theta)\xi(\theta)$ is constant over this interval, the posterior p.d.f. $\xi(\theta \mid \boldsymbol{x})$ will also be constant over the interval. In other words, the posterior distribution of θ must be a uniform distribution on this interval.

7.3 Conjugate Prior Distributions

Commentary

This section introduces some convenient prior distributions that make Bayesian inferences mathematically tractable. The instructor can remind the student that numerical methods are available for performing Bayesian inferences even when other prior distributions are used. Mathematical tractability is useful when introducing a new concept so that attention can focus on the meaning and interpretation of the new concept rather than the numerical methods required to perform the calculations. Although conjugate priors for the parameter of the uniform distribution are not discussed in the body of the section, Exercises '17 and 18 illustrate how the general concept extends to these distributions.

Solutions to Exercises

1. The posterior mean of θ will be

$$\frac{100 \times 0 + 20v^2 \times 0.125}{100 + 20v^2} = 0.12.$$

We can solve this equation for v^2 by multiplying both sides by $100 + 20v^2$ and collecting terms. The result is $v^2 = 120$.

2. If we let $\gamma = (y + 1)(y + z + 2)$, then $1 - \gamma = (z + 1)(y + z + 2)$ and $V = \gamma(1 - \gamma)/(y + z + 3)$. The maximum value of $\gamma(1 - \gamma)$ is $1/4$, and is attained when $\gamma = 1/2$. Therefore, $V \leq 1/[4(y + z + 3)]$. It now follows that if $1/[4(y + z + 3)] \leq 0.01$, then $V \leq 0.01$. But the first inequality will be satisfied if $y + z \geq 22$. Since $y + z$ is the total number of items that have been selected, it follows that this number need not exceed 22.

3. Since the observed number of defective items is 3 and the observed number of nondefective items is 97, it follows from Theorem 7.3.1 that the posterior distribution of θ is a beta distribution with parameters $2 + 3 = 5$ and $200 + 97 = 297$.

4. Let α_1 and β_1 denote the parameters of the posterior beta distribution, and let $\gamma = \alpha_1/(\alpha_1+\beta_1)$. Then γ is the mean of the posterior distribution and we are told that $\gamma = 2/51$. The variance of the posterior distribution is

$$\begin{aligned}\frac{\alpha_1\beta_1}{(\alpha_1+\beta_1)^2(\alpha_1+\beta_1+1)} &= \frac{\alpha_1}{\alpha_1+\beta_1}\cdot\frac{\beta_1}{\alpha_1+\beta_1}\cdot\frac{1}{\alpha_1+\beta_1+1}\\ &= \gamma(1-\gamma)\,\frac{1}{\alpha_1+\beta_1+1}\\ &= \frac{2}{51}\cdot\frac{49}{51}\cdot\frac{1}{\alpha_1+\beta_1+1} = \frac{98}{(51)^2}\cdot\frac{1}{\alpha_1+\beta_1+1}.\end{aligned}$$

From the value of this variance given in the exercise it is now evident that $\alpha_1+\beta_1+1 = 103$. Hence, $\alpha_1+\beta_1 = 102$ and $\alpha_1 = \gamma(\alpha_1+\beta_1) = 2(102)/51 = 4$. In turn, it follows that $\beta_1 = 102-4 = 98$. Since the posterior distribution is a beta distribution, it follows from Theorem 7.3.1 that the prior distribution must have been a beta distribution with parameters α and β such that $\alpha+3 = \alpha_1$ and $\beta+97 = \beta_1$. Therefore, $\alpha = \beta = 1$. But the beta distribution for which $\alpha = \beta = 1$ is the uniform distribution on the interval [0,1].

5. By Theorem 7.3.2, the posterior distribution will be the gamma distribution for which the parameters are $3 + \sum_{i=1}^{n} x_i = 3 + 13 = 16$ and $1 + n = 1 + 5 = 6$.

6. The number of defects on a 1200-foot roll of tape has the same distribution as the total number of defects on twelve 100-foot rolls, and it is assumed that the number of defects on a 100-foot roll has the Poisson distribution with mean θ. By Theorem 7.3.2, the posterior distribution of θ is the gamma distribution for which the parameters are $2+4=6$ and $10+12=22$.

7. In the notation of Theorem 7.3.3, we have $\sigma^2 = 4,\ \mu = 68,\ \upsilon^2 = 1,\ n = 10$, and $\bar{x}_n = 69.5$. Therefore, the posterior distribution of θ is the normal distribution with mean $\mu_1 = 967/14$ and variance $\upsilon_1^2 = 2/7$.

8. Since the p.d.f. of a normal distribution attains its maximum value at the mean of the distribution and then drops off on each side of the mean, among all intervals of length 1 unit, the interval that is centered at the mean will contain the most probability. Therefore, the answer in part (a) is the interval centered at the mean of the prior distribution of θ and the answer in part (b) is the interval centered at the mean of the posterior distribution of θ. In part (c), if the distribution of θ is specified by its prior distribution, then $Z = \theta - 68$ will have a standard normal distribution. Therefore,

$$\Pr(67.5 \le \theta \le 68.5) = \Pr(-0.5 \le Z \le 0.5) = 2\Phi(0.5) = 1 = 0.3830.$$

Similarly, if the distribution of θ is specified by its posterior distribution, then $Z = (\theta - \mu_1)/\upsilon_1 = (\theta - 69.07)/0.5345$ will have a standard normal distribution. Therefore,

$$\begin{aligned}\Pr(68.57 \le \theta \le 69.57 \mid \theta) &= \Pr(-0.9355 \le Z \le 0.9355)\\ &= 2\Phi(0.9355) - 1 = 0.6506.\end{aligned}$$

9. Since the posterior distribution of θ is normal, the prior distribution of θ must also have been normal. Furthermore, from Eqs. (7.3.1) and (7.3.2), we obtain the relations:

$$8 = \frac{\mu + (20)(10)\upsilon^2}{1 + 20\upsilon^2}$$

and

$$\frac{1}{25} = \frac{\upsilon^2}{1+20\upsilon^2}.$$

It follows that $\upsilon^2 = 1/5$ and $\mu = 0$.

10. In this exercise, $\sigma^2 = 4$ and $\upsilon^2 = 1$. Therefore, by Eq. (7.3.2)

$$\upsilon_1^2 = \frac{4}{4+n}.$$

It follows that $\upsilon_1^2 \le 0.01$ if and only if $n \ge 396$.

11. In this exercise, $\sigma^2 = 4$ and $n = 100$. Therefore, by Eq. (7.3.2),

$$\upsilon_1^2 = \frac{4\upsilon^2}{4+100\upsilon^2} = \frac{1}{25+(1/\upsilon^2)} < \frac{1}{25}.$$

Since the variance of the posterior distribution is less than 1/25, the standard deviation must be less than 1/5.

12. Let α and β denote the parameters of the prior gamma distribution of θ. Then $\alpha/\beta = 0.2$ and $\alpha/\beta^2 = 1$. Therefore, $\beta = 0.2$ and $\alpha = 0.04$. Furthermore, the total time required to serve the sample of 20 customers is $y = 20(3.8) = 76$. Therefore, by Theorem 7.3.4, the posterior distribution of θ is the gamma distribution for which the parameters are $0.04 + 20 = 20.04$ and $0.2 + 76 = 76.2$.

13. The mean of the gamma distribution with parameters α and β is α/β and the standard deviation is $\alpha^{1/2}/\beta$. Therefore, the coefficient of variation is $\alpha^{-1/2}$. Since the coefficient of variation of the prior gamma distribution of θ is 2, it follows that $\alpha = 1/4$ in the prior distribution. Furthermore, it now follows from Theorem 7.3.4 that the coefficient of variation of the posterior gamma distribution of θ is $(\alpha+n)^{-1/2} = (n + {}^1/_4)^{-1/2}$. This value will be less than 0.1 if and only if $n \ge 99.75$.Thus, the required sample size is $n \ge 100$.

14. Consider a single observation X from a negative binomial distribution with parameters r and p, where the value of r is known and the value of p is unknown. Then the p.f. of X has the form $f(x \mid p) \propto p^r q^x$. If the prior distribution of p is the beta distribution with parameters α and β, then the prior p.d.f. $\xi(p)$ has the form $\xi(p) \propto p^{\alpha-1}q^{\beta-1}$. Therefore, the posterior p.d.f. $\xi(p \mid x)$ has the form

$$\xi(p \mid x) \propto \xi(p)f(p \mid x) \propto p^{\alpha+r-1}q^{\beta+x-1}.$$

This expression can be recognized as being, except for a constant factor, the p.d.f. of the beta distribution with parameters $\alpha + r$ and $\beta + x$. Since this distribution will be the prior distribution of p for future observations, it follows that the posterior distribution after any number of observations will also be a beta distribution.

15. (a) Let $y = 1/\theta$. Then $\theta = 1/y$ and $d\theta = -dy/y^2$. Hence,

$$\int_0^\infty \xi(\theta)d\theta = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\exp(-\beta y)dy = 1.$$

(b) If an observation X has a normal distribution with a known value of the mean μ and an unknown value of the variance θ, then the p.d.f. of X has the form

$$f(x \mid \theta) \propto \frac{1}{\theta^{1/2} \exp\left[-\dfrac{(x-\mu)^2}{2\theta}\right]}.$$

Also, the prior p.d.f. of θ has the form

$$\xi(\theta) \propto \theta^{-(\alpha+1)} \exp(-\beta/\theta).$$

Therefore, the posterior p.d.f. $\xi(\theta \mid x)$ has the form

$$\xi(\theta \mid x) \propto \xi(\theta) f(x \mid \theta) \propto \theta^{-(\alpha+3/2)} \exp\left\{-\left[\beta + \frac{1}{2}(x-\mu)^2\right] \cdot \frac{1}{\theta}\right\}.$$

Hence, the posterior p.d.f. of θ has the same form as $\xi(\theta)$ with α replaced by $\alpha + 1/2$ and β replaced by $\beta + 1/2(x-\mu)^2$. Since this distribution will be the prior distribution of θ for future observations, it follows that the posterior distribution after any number of observations will also belong to the same family of distributions.

16. If X has the normal distribution with a known value of the mean μ and an unknown value of the standard deviation σ, then the p.d.f. of X has the form

$$f(x \mid \sigma) \propto \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Therefore, if the prior p.d.f. $\xi(\sigma)$ has the form

$$\xi(\sigma) \propto \sigma^{-a} \exp(-b/\sigma^2),$$

then the posterior p.d.f. of σ will also have the same form, with a replaced by $a+1$ and b replaced by $b + (x-\mu)^2/2$. It remains to determine the precise form of $\xi(\sigma)$. If we let $y = 1/\sigma^2$, then $\sigma = y^{-1/2}$ and $d\sigma = -dy/(2y^{3/2})$. Therefore,

$$\int_0^\infty \sigma^{-a} \exp(-b/\sigma^2) d\sigma = \frac{1}{2} \int_0^\infty y^{(a-3)/2} \exp(-by) dy.$$

The integral will be finite if $a > 1$ and $b > 0$, and its value will be

$$\frac{\Gamma\left[\dfrac{1}{2}(a-1)\right]}{b^{(a-1)/2}}.$$

Hence, for $a > 1$ and $b > 0$, the following function will be a p.d.f. for $\sigma > 0$:

$$\xi(\sigma) = \frac{2b^{(a-1)/2}}{\Gamma\left[\dfrac{1}{2}(a-1)\right]} \sigma^{-a} \exp(-b/\sigma^2).$$

Finally, we can obtain a more standard form for this p.d.f. by replacing a and b by $\alpha = (a-1)/2$ and $\beta = b$. Then

$$\xi(\sigma) = \frac{2\beta^\alpha}{\Gamma(\alpha)} \sigma^{-(2\alpha+1)} \exp(-\beta/\sigma^2) \qquad \text{for } \sigma > 0.$$

The family of distributions for which the p.d.f. has this form, for all values of $\alpha > 0$ and $\beta > 0$, will be a conjugate family of prior distributions.

17. The joint p.d.f. of the three observations is

$$f(x_1, x_2, x_3 \mid \theta) = \begin{cases} 1/\theta^3 & \text{for } 0 < x_i < \theta \ (i = 1, 2, 3), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the posterior p.d.f. $\xi(\theta \mid x_1, x_2, x_3)$ will be positive only if $\theta \geq 4$, as required by the prior p.d.f., and also $\theta > 8$, the largest of the three observed values. Hence, for $\theta > 8$,

$$\xi(\theta \mid x_1, x_2, x_3) \propto \xi(\theta) f(x_1, x_2, x_3 \mid \theta) \propto 1/\theta^7.$$

Since

$$\int_8^\infty \frac{1}{\theta^7} d\theta = \frac{1}{6(8)^6},$$

it follows that

$$\xi(\theta \mid x_1, x_2, x_3) = \begin{cases} 6(8^6)/\theta^7 & \text{for } \theta > 8 \\ 0 & \text{for } \theta \leq 8. \end{cases}$$

18. Suppose that the prior distribution of θ is the Pareto distribution with parameters x_0 and α ($x_0 > 0$ and $\alpha > 0$). Then the prior p.d.f. $\xi(\theta)$ has the form

$$\xi(\theta) \propto 1/\theta^{\alpha+1} \qquad \text{for} \quad \theta \geq x_0.$$

If $X_1, \ldots, X_n$ form a random sample from a uniform distribution on the interval $[0, \theta]$, then

$$f_n(\boldsymbol{x} \mid \theta) \propto 1/\theta^n \qquad \text{for} \quad \theta > \max\{x_1, \ldots, x_n\}.$$

Hence, the posterior p.d.f. of θ has the form

$$\xi(\theta \mid \boldsymbol{x}) \propto \xi(\theta) f_n(\boldsymbol{x} \mid \theta) \propto 1/\theta^{\alpha+n+1},$$

for $\theta > \max\{x_0, x_1, \ldots, x_n\}$, and $\xi(\theta \mid \boldsymbol{x}) = 0$ for $\theta \leq \max\{x_0, x_1, \ldots, x_n\}$. This posterior p.d.f. can now be recognized as also being the Pareto distribution with parameters $\alpha + n$ and $\max\{x_0, x_1, \ldots, x_n\}$.

Commentary: Exercise 17 provides a numerical illustration of the general result presented in Exercise 18.

19. The joint p.d.f. of $X_1, \ldots, X_n$ has the following form, for $0 < x_i < 1 (i = 1, \ldots, n)$:

$$\begin{aligned} f_n(\boldsymbol{x} \mid \theta) &= \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta - 1} \propto \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta} \\ &= \theta^n \exp\left(\theta \sum_{i=1}^n \log\ x_i \right). \end{aligned}$$

The prior p.d.f. of θ has the form

$$\xi(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta).$$

Hence, the posterior p.d.f. of θ has the form

$$\xi(\theta \mid \boldsymbol{x}) \propto \xi(\theta) f_n(\boldsymbol{x} \mid \theta) \propto \theta^{\alpha+n-1} \exp\left[-\left(\beta - \sum_{i=1}^{n} \log\ x_i\right)\theta\right].$$

This expression can be recognized as being, except for a constant factor, the p.d.f. of the gamma distribution with parameters $\alpha_1 = \alpha + n$ and $\beta_1 = \beta - \sum_{i=1}^{n} \log x_i$. Therefore, the mean of the posterior distribution is α_1/β_1 and the variance is α_1/β_1^2.

20. The mean lifetime conditional on β is $1/\beta$. The mean lifetime is then the mean of $1/\beta$. Prior to observing the data, the distribution of β is the gamma distribution with parameters a and b, so the mean of $1/\beta$ is $b/(a-1)$ according to Exercise 21 in Sec. 5.7. After observing the data, the distribution of β is the gamma distribution with parameters $a+10$ and $b+60$, so the mean of $1/\beta$ is $(b+60)/(a+9)$. So, we must solve the following equations:

$$\begin{aligned} \frac{b}{a-1} &= 4, \\ \frac{b+60}{a+9} &= 5. \end{aligned}$$

These equations convert easily to the equations $b = 4a - 4$ and $b = 5a - 15$. So $a = 11$ and $b = 40$.

21. The posterior p.d.f. is proportional to the likelihood $\theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$ times $1/\theta$. This product can be written as $\theta^{n-1} \exp(-\theta n \overline{x}_n)$. As a function of θ this is recognizable as the p.d.f. of the gamma distribution with parameters n and $n\overline{x}_n$. The mean of this posterior distribution is then $n/[n\overline{x}_n] = 1/\overline{x}_n$.

22. The posterior p.d.f. is proportional to the likelihood since the prior "p.d.f." is constant. The likelihood is proportional to

$$\exp\left[-\frac{20}{2 \times 60}(\theta - (-0.95))^2\right],$$

using the same reasoning as in the proof of Theorem 7.3.3 of the text. As a function of θ this is easily recognized as being proportional to the p.d.f. of the normal distribution with men -0.95 and variance $60/20 = 3$. The posterior probability that $\theta > 1$ is then

$$1 - \Phi\left(\frac{1 - (-0.95)}{3^{1/2}}\right) = 1 - \Phi(1.126) = 0.1301.$$

23. (a) Let the prior p.d.f. of θ be $\xi_{\alpha,\beta}(\theta)$. Suppose that $X_1, \ldots, X_n$ are i.i.d. with conditional p.d.f. $f(x|\theta)$ given θ, where f is as stated in the exercise. The posterior p.d.f. after observing these data is

$$\xi(\theta|\boldsymbol{x}) = \frac{a(\theta)^{\alpha+n} \exp\left[c(\theta)\left\{\beta + \sum_{i=1}^{b} d(x_i)\right\}\right]}{\int_\Omega a(\theta)^{\alpha+n} \exp\left[c(\theta)\left\{\beta + \sum_{i=1}^{b} d(x_i)\right\}\right] d\theta}. \quad \text{(S.7.1)}$$

Eq. (S.7.1) is of the form of $\xi_{\alpha',\beta'}(\theta)$ with $\alpha' = \alpha + n$ and $\beta' = \beta + \sum_{i=1}^{n} d(x_i)$. The integral in the denominator of (S.7.1) must be finite with probability 1 (as a function of $x_1, \ldots, x_n$) because $\prod_{i=1}^{n} b(x_i)$ times this denominator is the marginal (joint) p.d.f. of $X_1, \ldots, X_n$.

(b) This was essentially the calculation done in part (a).

24. In each part of this exercise we shall first present the p.d.f. or the p.f. f, and then we shall identify the functions a, b, c, and d in the form for an exponential family given in Exercise 23.

(a) $f(x \mid p) = p^x(1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x$. Therefore, $a(p) = 1-p$, $b(x) = 1$, $c(p) = \log\left(\frac{p}{1-p}\right)$, $d(x) = x$.

(b) $f(x \mid \theta) = \frac{\exp(-\theta)\theta^x}{x!}$. Therefore, $a(\theta) = \exp(-\theta)$, $b(x) = 1/x!$, $c(\theta) = \log\ \theta$, $d(x) = x$.

(c) $f(x \mid p) = \binom{r+x-1}{x} p^r(1-p)^x$. Therefore, $a(p) = p^r$, $b(x) = \binom{r+x-1}{x}$, $c(p) = \log(1-p)$, $d(x) = x$.

(d)

$$
\begin{aligned}
f(x \mid \mu) &= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{\mu x}{\sigma^2}\right).
\end{aligned}
$$

Therefore, $a(\mu) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp(-\frac{\mu^2}{2\sigma^2})$, $b(x) = \exp(-\frac{x^2}{2\sigma^2})$, $c(\mu) = \frac{\mu}{\sigma^2}$, $d(x) = x$.

(e) $f(x \mid \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$. Therefore, $a(\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}}$, $b(x) = 1$, $c(\sigma^2) = -\frac{1}{2\sigma^2}$, $d(x) = (x-\mu)^2$.

(f) $f(x \mid \alpha) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x)$. Therefore, $a(\alpha) = \frac{\beta^\alpha}{\Gamma(\alpha)}$, $b(x) = \exp(-\beta x)$, $c(\alpha) = \alpha - 1$, $d(x) = \log\ x$.

(g) $f(x \mid \beta)$ in this part is the same as the p.d.f. in part (f). Therefore, $a(\beta) = \beta^\alpha$, $b(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}$, $c(\beta) = -\beta$, $d(x) = x$.

(h) $f(x \mid \alpha) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$. Therefore, $a(\alpha) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}$, $b(x) = \frac{(1-x)^{\beta-1}}{\Gamma(\beta)}$, $c(\alpha) = \alpha-1$, $d(x) = \log x$.

(i) $f(x \mid \beta)$ in this part is the same as the p.d.f. given in part (h). Therefore, $a(\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}$, $b(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}$, $c(\beta) = \beta - 1$, $d(x) = \log(1-x)$.

25. For every θ, the p.d.f. (or p.f.) $f(x|\theta)$ for an exponential family is strictly postive for all x such that $b(x) > 0$. That is, the set of x for which $f(x|\theta) > 0$ is the same for all θ. This is not true for uniform distributions where the set of x such that $f(x|\theta) > 0$ is $[0, \theta]$.

26. The same reasoning applies as in the previous exercise, for uniform distributions, the set of x such that $f(x|\theta) > 0$ depends on θ. For exponential families, the set of x such that $f(x|\theta) > 0$ is the same for all θ.

7.4 Bayes Estimators

Commentary

We introduce the fundamental concepts of Bayesian decision theory. The use of a loss function arises again in Bayesian hypothesis testing in Sec. 9.8. This section ends with foundational discussion of the limitations of Bayes estimators. This material is included for those instructors who want their students to have both a working and a critical understanding of the topic.

If you are using the statistical software *R*, the function mentioned in Example 7.4.5 to compute the median of a beta distribution is `qbeta` with the first argument equal to 0.5 and the next two equal to $\alpha + y$ and $\beta + n - y$, in the notation of the example.

Solutions to Exercises

1. The posterior distribution of θ would be the beta distribution with parameters 2 and 1. The mean of the posterior distribution is 2/3, which would be the Bayes estimate under squared error loss. The median of the posterior distribution would be the Bayes estimate under absolute error loss. To find the median, write the c.d.f. as

$$F(\theta) = \int_0^\theta 2t dt = \theta^2,$$

for $0 < \theta < 1$. The quantile function is then $F^{-1}(p) = p^{1/2}$, so the median is $(1/2)^{1/2} = 0.7071$.

2. The posterior distribution of θ is the beta distribution with parameters $5 + 1 = 6$ and $10 + 19 = 29$. The mean of this distribution is $6/(6 + 29) = 6/35$. Therefore, the Bayes estimate of θ is 6/35.

3. If y denotes the number of defective items in the sample, then the posterior distribution of θ will be the beta distribution with parameters $5 + y$ and $10 + 20 - y = 30 - y$. The variance V of this beta distribution is

$$V = \frac{(5 + y)(30 - y)}{(35)^2(36)}.$$

Since the Bayes estimate of θ is the mean μ of the posterior distribution, the mean squared error of this estimate is $E[(\theta - \mu)^2 \mid \boldsymbol{x}]$, which is the variance V of the posterior distribution.

   (a) V will attain its maximum at a value of y for which $(5+y)(30-y)$ is a maximum. By differentiating with respect to y and setting the derivative equal to 0, we find that the maximum is attained when $y = 12.5$. Since the number of defective items y must be an integer, the maximum of V will be attained for $y = 12$ or $y = 13$. When these values are substituted into $(5 + y)(30 - y)$, it is found that they both yield the same value.

   (b) Since $(5 + y)(30 - y)$ is a quadratic function of y and the coefficient of y^2 is negative, its minimum value over the interval $0 \le y \le 20$ will be attained at one of the endpoints of the interval. It is found that the value for $y = 0$ is smaller than the value for $y = 20$.

4. Suppose that the parameters of the prior beta distribution of θ are α and β. Then $\mu_0 = \alpha/(\alpha + \beta)$. As shown in Example 7.4.3, the mean of the posterior distribution of θ is

$$\frac{\alpha + \sum_{i=1}^n X_i}{\alpha + \beta + n} = \frac{\alpha + \beta}{\alpha + \beta + n}\mu_0 + \frac{n}{\alpha + \beta + n}\overline{X}_n.$$

Hence, $\gamma_n = n/(\alpha + \beta + n)$ and $\gamma_n \to 1 \; as \; n \to \infty$.

5. It was shown in Exercise 5 of Sec. 7.3 that the posterior distribution of θ is the gamma distribution with parameters $\alpha = 16$ and $\beta = 6$. The Bayes estimate of θ is the mean of this distribution and is equal to $16/6 = 8/3$.

6. Suppose that the parameters of the prior gamma distribution of θ are α and β. Then $\mu_0 = \alpha/\beta$. The posterior distribution of θ was given in Theorem 7.3.2. The mean of this posterior distribution is

$$\frac{\alpha + \sum_{i=1}^{n} X_i}{\beta + n} = \frac{\beta}{\beta + n}\,\mu_0 + \frac{n}{\beta + n}\overline{X}_n.$$

Hence, $\gamma_n = n/(\beta + n)$ and $\gamma_n \to 1$ *as* $n \to \infty$.

7. The Bayes estimator is the mean of the posterior distribution of θ, as given in Exercise 6. Since θ is the mean of the Poisson distribution, it follows from the law of large numbers that $\overline{X}_n$ converges to θ in probability as $n \to \infty$. It now follows from Exercise 6 that, since $\gamma_n \to 1$, the Bayes estimators will also converge to θ in probability as $n \to \infty$. Hence, the Bayes estimators form a consistent sequence of estimators of θ.

8. It was shown in Exercise 7 of Sec. 7.3 that the posterior distribution of θ is the normal distribution with mean 69.07 and variance 0.286.

   (a) The Bayes estimate is the mean of this distribution and is equal to 69.07.

   (b) The Bayes estimate is the median of the posterior distribution and is therefore again equal to 69.07.

9. For any given values in the random sample, the Bayes estimate of θ is the mean of the posterior distribution of θ. Therefore, the mean squared error of the estimate will be the variance of the posterior distribution of θ. It was shown in Exercise 10 of Sec. 7.3 that this variance will be 0.01 or less for $n \geq 396$.

10. It was shown in Exercise 12 of Sec. 7.3 that the posterior distribution of θ will be a gamma distribution with parameters $\alpha = 20.04$ and $\beta = 76.2$. The Bayes estimate is the mean of this distribution and is equal to $20.04/76.2 = 0.263$.

11. Let $X_1, \ldots, X_n$ denote the observations in the random sample, and let α and β denote the parameters of the prior gamma distribution of θ. It was shown in Theorem 7.3.4 that the posterior distribution of θ will be the gamma distribution with parameters $\alpha + n$ and $\beta + n\overline{X}_n$. The Bayes estimator, which is the mean of this posterior distribution is, therefore,

$$\frac{\alpha + n}{\beta + n\overline{X}_n} = \frac{1 + (\alpha/n)}{\overline{X}_n + (\beta/n)}.$$

Since the mean of the exponential distribution is $1/\theta$, it follows from the law of large numbers that $\overline{X}_n$ will converge in probability to $1/\theta$ as $n \to \infty$. It follows, therefore, that the Bayes estimators will converge in probability to θ as $n \to \infty$. Hence, the Bayes estimators form a consistent sequence of estimators of θ.

12. (a) A's prior distribution for θ is the beta distribution with parameters $\alpha = 2$ and $\beta = 1$. Therefore, A's posterior distribution for θ is the beta distribution with parameters $2+710 = 712$ and $1+290 = 291$. B's prior distribution for θ is a beta distribution with parameters $\alpha = 4$ and $\beta = 1$. Therefore, B's posterior distribution for θ is the beta distribution with parameters $4 + 710 = 714$ and $1 + 290 = 291$.

(b) A's Bayes estimate of θ is $712/(712+291) = 712/1003$. B's Bayes estimate of θ is $714/(714+291) = 714/1005$.

(c) If y denotes the number in the sample who were in favor of the proposition, then A's posterior distribution for θ will be the beta distribution with parameters $2+y$ and $1+1000-y = 1001-y$, and B's posterior distribution will be a beta distribution with parameters $4+y$ and $1+1000-y = 1001-y$. Therefore, A's Bayes estimate of θ will be $(2+y)/1003$ and B's Bayes estimate of θ will be $(4+y)/1005$. But

$$\left|\frac{4+y}{1005} - \frac{2+y}{1003}\right| = \frac{2(1001-y)}{(1005)(1003)}.$$

This difference is a maximum when $y = 0$, but even then its value is only

$$\frac{2(1001)}{(1005)(1003)} < \frac{2}{1000}.$$

13. If θ has the Pareto distribution with parameters $\alpha > 1$ and $x_0 > 0$, then

$$E(\theta) = \int_{x_0}^{\infty} \theta \cdot \frac{\alpha x_0^{\alpha}}{\theta^{\alpha+1}}\, d\theta = \frac{\alpha}{\alpha-1} x_0.$$

It was shown in Exercise 18 of Sec. 7.3 that the posterior distribution of θ will be a Pareto distribution with parameters $\alpha + n$ and $\max\{x_0, X_1, \ldots, X_n\}$. The Bayes estimator is the mean of this posterior distribution and is, therefore, equal to $(\alpha+n)\max\{x_0, X_1, \ldots, X_n\}/(\alpha+n-1)$.

14. Since $\psi = \theta^2$, the posterior distribution of ψ can be derived from the posterior distribution of θ. The Bayes estimator $\hat{\psi}$ will then be the mean $E(\psi)$ of the posterior distribution of ψ. But $E(\psi) = E(\theta)^2$, where the first expectation is calculated with respect to the posterior distribution of ψ and the second with respect to the posterior distribution of θ. Since $\hat{\theta}$ is the mean of the posterior distribution of θ, it is also true that $\hat{\theta} = E(\theta)$. Finally, since the posterior distribution of θ is a continuous distribution, it follows from the hint given in this exercise that

$$\hat{\psi} = E(\theta^2) > [E(\theta)]^2 = \hat{\theta}^2.$$

15. Let a_0 be a $1/(1+c)$ quantile of the posterior distribution, and let a_1 be some other value. Assume that $a_1 < a_0$. The proof for $a_1 > a_0$ is similar. Let $g(\theta|x)$ denote the posterior p.d.f. The posterior mean of the loss for action a is

$$h(a) = c\int_{-\infty}^{a} (a-\theta)g(\theta|x)d\theta + \int_{a}^{\infty} (\theta-a)g(\theta|x)d\theta.$$

We shall now show that $h(a_1) \geq h(a_0)$, with strict inequality if a_1 is not a $1/(1+c)$ quantile.

$$\begin{aligned} h(a_1) - h(a_0) &= c\int_{-\infty}^{a_0} (a_1 - a_0)g(\theta|x)d\theta + \int_{a_0}^{a_1} (ca_1 - (1+c)\theta + a_0)g(\theta|x)d\theta \\ &\quad + \int_{a_1}^{\infty} (a_0 - a_1)g(\theta|x)d\theta \end{aligned} \qquad \text{(S.7.2)}$$

The first integral in (S.7.2) equals $c(a_1 - a_0)/(1+c)$ because a_0 is a $1/(1+c)$ quantile of a the posterior distribution, and the posterior distribution is continuous. The second integral in (S.7.2) is at least as large as $(a_0 - a_1)\Pr(a_0 < \theta \leq a_1|x)$ since $-(1+c)\theta > -(1+c)a_1$ for all θ in that integral. In fact, the integral will be strictly larger than $(a_0 - a_1)\Pr(a_0 < \theta \le a_1|x)$ if this probability is positive. The last integral in (S.7.2) equals $(a_0 - a_1)\Pr(\theta > a_1|x)$. So

$$h(a_1) - h(a_0) \ge c\frac{a_1 - a_0}{1 + c} + (a_0 - a_1)\Pr(\theta > a_0|x) = 0. \tag{S.7.3}$$

The equality follows from the fact that $\Pr(\theta > a_0|x) = c/(1+c)$. The inequality in (S.7.3) will be strict if and only if $\Pr(a_0 < \theta \le a_1|x) > 0$, which occurs if and only if a_1 is not another $1/(1 + c)$ quantile.

7.5 Maximum Likelihood Estimators

Commentary

Although maximum likelihood is a popular method of estimation, it can be valuable for the more capable students to see some limitations that are described at the end of this section. These limitations arise only in more complicated situations than those that are typically encountered in practice. This material is probably not suitable for students with a limited mathematical background who are learning statistical inference for the first time.

Solutions to Exercises

1. We can easily compute

$$\begin{aligned} E(Y) &= \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n, \\ E(Y^2) &= \frac{1}{n}\sum_{i=1}^{n} x_i^2. \end{aligned}$$

   Then

$$\text{Var}(Y) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2.$$

2. It was shown in Example 7.5.4 that the M.L.E. is $\overline{x}_n$. In this exercise, $\overline{x}_n = 58/70 = 29/35$.

3. The likelihood function for the given sample is $p^{58}(1 - p)^{12}$. Among all values of p in the interval $1/2 \le p \le 2/3$, this function is a maximum when $p = 2/3$.

4. Let y denote the sum of the observations in the sample. Then the likelihood function is $p^y(1 - p)^{n-y}$. If $y = 0$, this function is a decreasing function of p. Since $p = 0$ is not a value in the parameter space, there is no M.L.E. Similarly, if $y = n$, then the likelihood function is an increasing function of p. Since $p = 1$ is not a value in the parameter space, there is no M.L.E.

5. Let y denote the sum of the observed values $x_1, \ldots, x_n$. Then the likelihood function is

$$f_n(\boldsymbol{x} \mid \theta) = \frac{\exp(-n\theta)\theta^y}{\prod_{i=1}^{n}(x_i!)}.$$

(a) If $y > 0$ and we let $L(\theta) = \log\ f_n(\boldsymbol{x} \mid \theta)$, then
$$\frac{\partial}{\partial\theta}L(\theta) = -n + \frac{y}{\theta}.$$
The maximum of $L(\theta)$ will be attained at the value of θ for which this derivative is equal to 0. In this way, we find that $\hat{\theta} = y/n = \overline{x}_n$.

(b) If $y = 0$, then $f_n(\boldsymbol{x} \mid \theta)$ is a decreasing function of θ. Since $\theta = 0$ is not a value in the parameter space, there is no M.L.E.

6. Let $\theta = \sigma^2$. Then the likelihood function is
$$f_n(\boldsymbol{x} \mid \theta) = \frac{1}{(2\pi\theta)^{n/2}} \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i - \mu)^2\right\}.$$
If we let $L(\theta) = \log\ f_n(\boldsymbol{x} \mid \theta)$, then
$$\frac{\partial}{\partial\theta}L(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$
The maximum of $L(\theta)$ will be attained at a value of θ for which this derivative is equal to 0. In this way, we find that
$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

7. Let y denote the sum of the observed values $x_1, \ldots, x_n$. Then the likelihood function is
$$f_n(\boldsymbol{x} \mid \beta) = \beta^n \exp(-\beta y).$$
If we let $L(\beta) = \log\ f_n(\boldsymbol{x} \mid \beta)$, then
$$\frac{\partial L(\beta)}{\partial\beta} = \frac{n}{\beta} - y.$$
The maximum of $L(\beta)$ will be attained at the value of β for which this derivative is equal to 0. Therefore, $\hat{\beta} = n/y = 1/\overline{x}_n$.

8. Let y denote the sum of the observed values $x_1, \ldots, x_n$. Then the likelihood function is
$$f_n(\boldsymbol{x} \mid \theta) = \begin{cases} \exp(n\theta - y) & \text{for } \min\{x_1, \ldots, x_n\} > \theta \\ 0 & \text{otherwise.} \end{cases}$$

(a) For each value of $\boldsymbol{x}$, $f_n(\boldsymbol{x} \mid \theta)$ will be a maximum when θ is made as large as possible subject to the strict inequality $\theta < \min\{x_1, \ldots, x_n\}$. Therefore, the value $\theta = \min\{x_1, \ldots, x_n\}$ cannot be used and there is no M.L.E.

(b) Suppose that the p.d.f. given in this exercise is replaced by the following equivalent p.d.f., in which strict and weak inequalities have been changed:
$$f(x \mid \theta) = \begin{cases} \exp(\theta - x) & \text{for } x \geq \theta, \\ 0 & \text{for } x < \theta. \end{cases}$$
Then the likelihood function $f_n(\boldsymbol{x} \mid \theta)$ will be nonzero for $\theta \leq \min\{x_1, \ldots, x_n\}$ and the M.L.E. will be $\hat{\theta} = \min\{x_1, \ldots, x_n\}$.

9. If $0 < x_i < 1$ for $i = 1, \dots, n$, then the likelihood function will be as follows:

$$f_n(\boldsymbol{x} \mid \theta) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta-1}.$$

If we let $L(\theta) = \log\ f_n(\boldsymbol{x} \mid \theta)$, then

$$\frac{\partial}{\partial\theta} L(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log\ x_i.$$

Therefore, $\hat{\theta} = -n / \sum_{i=1}^{n} \log\ x_i$. It should be noted that $\hat{\theta} > 0$.

10. The likelihood function is

$$f_n(\boldsymbol{x} \mid \theta) = \frac{1}{2^n} \exp\left\{ -\sum_{i=1}^{n} |x_i - \theta| \right\}.$$

Therefore, the M.L.E. of θ will be the value that minimizes $\sum_{i=1}^{n} |x_i - \theta|$. The solution to this minimization problem was given in the solution to Exercise 10 of Sec. 4.5.

11. The p.d.f. of each observation can be written as follows:

$$f(x \mid \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{for } \theta_1 \le x \le \theta_2, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the likelihood function is

$$f_n(\boldsymbol{x} \mid \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n}$$

for $\theta_1 \le \min\{x_1, \dots, x_n\} < \max\{x_1, \dots, x_n\} \le \theta_2$, and $f_n(\boldsymbol{x} \mid \theta_1, \theta_2) = 0$ otherwise. Hence, $f_n(\boldsymbol{x} \mid \theta_1, \theta_2)$ will be a maximum when $\theta_2 - \theta_1$ is made as small as possible. Since the smallest possible value of θ_2 is $\max\{x_1, \dots, x_n\}$ and the largest possible value of θ_1 is $\min\{x_1, \dots, x_n\}$, these values are the M.L.E.'s.

12. The likelihood function is

$$f_n(\boldsymbol{x} \mid \theta_1, \dots, \theta_k) = \theta_1^{n_1} \cdots \theta_k^{n_k}.$$

If we let $L(\theta_1, \dots, \theta_k) = \log\ f_n(\boldsymbol{x} \mid \theta_1, \dots, \theta_k)$ and let $\theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$, then

$$\frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_i} = \frac{n_i}{\theta_i} - \frac{n_k}{\theta_k} \qquad \text{for } i = 1, \dots, k-1.$$

If each of these derivatives is set equal to 0, we obtain the relations

$$\frac{\theta_1}{n_1} = \frac{\theta_2}{n_2} = \cdots = \frac{\theta_k}{n_k}.$$

If we let $\theta_i = \alpha n_i$ for $i = 1, \dots, k$, then

$$1 = \sum_{i=1}^{k} \theta_i = \alpha \sum_{i=1}^{k} n_i = \alpha n.$$

Hence $\alpha = 1/n$. It follows that $\hat{\theta}_i = n_i/n$ for $i = 1, \dots, k$.

13. It follows from Eq. (5.10.2) (with x_1 and x_2 now replaced by x and y) that the likelihood function is

$$f_n(\boldsymbol{x},\boldsymbol{y} \mid \mu_1,\mu_2) \propto \exp\left\{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^{n}\left[\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_i-\mu_1}{\sigma_1}\right)\left(\frac{y_i-\mu_2}{\sigma_2}\right)+\left(\frac{y_i-\mu_2}{\sigma_2}\right)^2\right]\right\}.$$

If we let $L(\mu_1,\mu_2) = \log\ f(\boldsymbol{x},\boldsymbol{y} \mid \mu_1,\mu_2)$, then

$$\frac{\partial L(\mu_1,\mu_2)}{\partial \mu_1} = \frac{1}{1-\rho^2}\left[\frac{1}{\sigma_1^2}\left(\sum_{i=1}^{n} x_i - n\mu_1\right) - \frac{\rho}{\sigma_1\sigma_2}\left(\sum_{i=1}^{n} y_i - n\mu_2\right)\right],$$
$$\frac{\partial L(\mu_1,\mu_2)}{\partial \mu_2} = \frac{1}{1-\rho^2}\left[\frac{1}{\sigma_2^2}\left(\sum_{i=1}^{n} y_i - n\mu_2\right) - \frac{\rho}{\sigma_1\sigma_2}\left(\sum_{i=1}^{n} x_i - n\mu_1\right)\right].$$

When these derivatives are set equal to 0, the unique solution is $\mu_1 = \overline{x}_n$ and $\mu_2 = \overline{y}_n$. Hence, these values are the M.L.E.'s.

7.6 Properties of Maximum Likelihood Estimators

Commentary

The material on sampling plans at the end of this section is a bit more subtle than the rest of the section, and should only be introduced to students who are capable of a deeper understanding of the material.

If you are using the software *R*, the digamma function mentioned in Example 7.6.4 can be computed with the function `digamma` which takes only one argument. The trigamma function mentioned in Example 7.6.6 can be computed with the function `trigamma` which takes only one argument. *R* also has several functions like `nlm` and `optim` for minimizing general functions. The required arguments to `nlm` are the name of another *R* function with a vector argument over which the minimization is done, and a starting value for the argument. If the function has additional arguments that remain fixed during the minimization, those can be listed after the starting vector, but they must be named explicitly. For `optim`, the first two arguments are reversed. Both functions have an optional argument `hessian` which, if set to `TRUE`, will tell the function to compute a matrix of numerical second partial derivatives. For example, if we want to minimize a function `f(x,y)` over `x` with `y` fixed at `c(3,1.2)` starting from `x=x0`, we could use
`optim(x0,f,y=c(3,1.2))`. If we wish to maximize a function g, we can define `f` to be $-g$ and pass that to either `optim` or `nlm`.

Solutions to Exercises

1. The M.L.E. of $\exp(-1/\theta)$ is $\exp(-1/\hat{\theta})$, where $\hat{\theta} = -n/\sum_{i=1}^{n}\log(x_i)$ is the M.L.E. of θ. That is, the M.L.E. of $\exp(-1/\theta)$ is

$$\exp\left(\sum_{i=1}^{n}\log(x_i)/n\right) = \exp\left(\log\left[\prod_{i=1}^{n} x_i\right]^{1/n}\right) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

2. The standard deviation of the Poisson distribution with mean θ is $\sigma = \theta^{1/2}$. Therefore, $\hat{\sigma} = \hat{\theta}^{1/2}$. It was found in Exercise 5 of Sec. 7.5 that $\hat{\theta} = \overline{X}_n$.

3. The median of an exponential distribution with parameter β is the number m such that

$$\int_0^m \beta \exp(-\beta x)dx = \frac{1}{2}.$$

Therefore, $m = (\log\ 2)/\beta$, and it follows that $\hat{m} = (\log 2)/\hat{\beta}$. It was shown in Exercise 7 of Sec. 7.5 that $\hat{\beta} = 1/\overline{X}_n$.

4. The probability that a given lamp will fail in a period of T hours is $p = 1 - \exp(-\beta T)$, and the probability that exactly x lamps will fail is $\binom{n}{x} p^x(1-p)^{n-x}$. It was shown in Example 7.5.4 that $\hat{p} = x/n$. Since $\beta = -\log(1-p)/T$, it follows that $\hat{\beta} = -\log(1 - x/n)/T$.

5. Since the mean of the uniform distribution is $\mu = (a+b)/2$, it follows that $\hat{\mu} = (\hat{a}+\hat{b})/2$. It was shown in Exercise 11 of Sec. 7.5 that $\hat{a} = \min\{X_1, \ldots, X_n\}$ and $\hat{b} = \max\{X_1, \ldots, X_n\}$.

6. The distribution of $Z = (X - \mu)/\sigma$ will be a standard normal distribution. Therefore,

$$0.95 = \Pr(X < \theta) = \Pr\left(Z < \frac{\theta - \mu}{\sigma}\right) = \Phi\left(\frac{\theta - \mu}{\sigma}\right).$$

Hence, from a table of the values of Φ it is found that $(\theta - \mu)/\sigma = 1.645$. Since $\theta = \mu + 1.645\sigma$, it follows that $\hat{\theta} = \hat{\mu} + 1.645\hat{\sigma}$. By example 6.5.4, we have

$$\hat{\mu} = \overline{X}_n \qquad \text{and} \qquad \hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right]^{1/2}.$$

7. $\nu = \Pr(X > 2) = \Pr\left(Z > \frac{2-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{2-\mu}{\sigma}\right) = \Phi\left(\frac{\mu - 2}{\sigma}\right).$

Therefore, $\hat{\nu} = \Phi((\hat{\mu} - 2)/\hat{\sigma})$.

8. Let $\theta = \Gamma'(\alpha)/\Gamma(\alpha)$. Then $\hat{\theta} = \Gamma'(\hat{\alpha})/\Gamma(\hat{\alpha})$. It follows from Eq. (7.6.5) that $\hat{\theta} = \sum_{i=1}^n (\log\ X_i)/n$.

9. If we let $y = \sum_{i=1}^n x_i$, then the likelihood function is

$$f_n(\boldsymbol{x} \mid \alpha, \beta) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n}\left(\prod_{i=1}^n x_i\right)^{\alpha-1} \exp(-\beta y).$$

If we now let $L(\alpha, \beta) = \log\ f_n(\boldsymbol{x} \mid \alpha, \beta)$, then

$$L(\alpha, \beta) = n\alpha\ \log\ \beta - n\ \log\ \Gamma(\alpha) + (\alpha - 1)\ \log\left(\prod_{i=1}^n x_i\right) - \beta y.$$

Hence,

$$\frac{\partial L(\alpha, \beta)}{\partial \beta} = \frac{n\alpha}{\beta} - y.$$

Since $\hat{\alpha}$ and $\hat{\beta}$ must satisfy the equation $\partial L(\alpha, \beta)/\partial\beta = 0$ [as well as the equation $\partial L(\alpha, \beta)/\partial\alpha = 0$], it follows that $\hat{\alpha}/\hat{\beta} = y/n = \overline{x}_n$.

10. The likelihood function is

$$f_n(\boldsymbol{x} \mid \alpha, \beta) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} \left[\prod_{i=1}^n (1-x_i)\right]^{\beta-1}.$$

If we let $L(\alpha, \beta) = \log\ f_n(\boldsymbol{x} \mid \alpha, \beta)$, then

$$\begin{aligned} L(\alpha,\beta) &= n\ \log\ \Gamma(\alpha+\beta) - n\ \log\ \Gamma(\alpha) - n\ \log\ \Gamma(\beta) \\ &\quad +(\alpha-1)\sum_{i=1}^n \log\ x_i + (\beta-1)\sum_{i=1}^n \log(1-x_i). \end{aligned}$$

Hence,

$$\frac{\partial L(\alpha,\beta)}{\partial \alpha} = n\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^n \log\ x_i$$

and

$$\frac{\partial L(\alpha,\beta)}{\partial \beta} = n\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - n\frac{\Gamma'(\beta)}{\Gamma(\beta)} + \sum_{i=1}^n \log(1-x_i).$$

The estimates $\hat{\alpha}$ and $\hat{\beta}$ must satisfy the equations $\partial L(\alpha,\beta)/\partial\alpha = 0$ and $\partial L(\alpha,\beta)/\partial\beta = 0$. Therefore, $\hat{\alpha}$ and $\hat{\beta}$ must also satisfy the equation $\partial L(\alpha,\beta)/\partial\alpha = \partial L(\alpha,\beta)/\partial\beta$. This equation reduces to the one given in the exercise.

11. Let $Y_n = \max\{X_1, \ldots, X_n\}$. It was shown in Example 7.5.7 that $\hat{\theta} = Y_n$. Therefore, for $\varepsilon > 0$,

$$\Pr(|\hat{\theta} - \theta| < \varepsilon) = \Pr(Y_n > \theta - \varepsilon) = 1 - \left(\frac{\theta - \varepsilon}{\theta}\right)^n.$$

It follows that $\lim_{n\to\infty} \Pr(|\hat{\theta} - \theta| < \varepsilon) = 1$. Therefore, $\hat{\theta} \xrightarrow{p} \theta$.

12. We know that $\hat{\beta} = 1/\overline{X}_n$. Also, since the mean of the exponential distribution is $\mu = 1/\beta$, it follows from the law of large numbers that $\overline{X}_n \xrightarrow{p} 1/\beta$. Hence, $\hat{\beta} \xrightarrow{p} \beta$.

13. Let $Z_i = -\log\ X_i$ for $i = 1, \ldots, n$. Then by Exercise 9 of Sec. 7.5, $\hat{\theta} = 1/\overline{Z}_n$. If X_i has the p.d.f. $f(x \mid \theta)$ specified in that exercise, then the p.d.f. $g(z \mid \theta)$ of Z_i will be as follows, for $z > 0$:

$$g(z \mid \theta) = f(\exp(-z) \mid \theta)\left|\frac{dx}{dz}\right| = \theta(\exp(-z))^{\theta-1}\exp(-z) = \theta\exp(-\theta z).$$

Therefore, Z_i has an exponential distribution with parameter θ. It follows that $E(Z_i) = 1/\theta$. Furthermore, since $X_1, \ldots, X_n$ form a random sample from a distribution for which the p.d.f. is $f(x \mid \theta)$, it follows that $Z_1, \ldots, Z_n$ will have the same joint distribution as a random sample from an exponential distribution with parameter θ. Therefore, by the law of large numbers, $\overline{Z}_n \xrightarrow{p} 1/\theta$. It follows that $\hat{\theta} \xrightarrow{p} \theta$.

14. The M.L.E. $\hat{p}$ is equal to the proportion of butterflies in the sample that have the special marking, regardless of the sampling plan. Therefore, (a)$\hat{p} = 5/43$ and (b) $\hat{p} = 3/58$.

15. As explained in this section, the likelihood function for the 21 observations is equal to the joint p.d.f. of the 20 observations for which the exact value is known, multiplied by the probability $\exp(-15/\mu)$ that the 21st observation is greater than 15. If we let y denote the sum of the first 20 observations, then the likelihood function is

$$\frac{1}{\mu^{20}}\exp(-y/\mu)\exp(-15/\mu).$$

Since $y = (20)(6) = 120$, this likelihood function reduces to

$$\frac{1}{\mu^{20}}\exp(-135/\mu).$$

The value of μ which maximizes this likelihood function is $\hat{\mu} = 6.75$.

16. The likelihood function determined by any observed value x of X is $\theta^3 x^2 \exp(-\theta x)/2$. The likelihood function determined by any observed value y of Y is $(2\theta)^y \exp(-2\theta)/y!$. Therefore, when $X = 2$ and $Y = 3$, each of these functions is proportional to $\theta^3 \exp(-2\theta)$. The M.L.E. obtained by either statistician will be the value of θ. which maximizes this expression. That value is $\hat{\theta} = 3/2$.

17. The likelihood function determined by any observed value x of X is $\binom{10}{x} p^x(1-p)^{10-x}$. By Eq. (5.5.1) the likelihood function determined by any observed value y of Y is $\binom{3+y}{y} p^4(1-p)^y$. Therefore, when $X = 4$ and $Y = 6$, each of these likelihood functions is proportional to $p^4(1-p)^6$. The M.L.E. obtained by either statistician will be the value of p which maximizes this expression. That value is $\hat{p} = 2/5$.

18. The mean of a Bernoulli random variable with parameter p is p. Hence, the method of moments estimator is the sample mean, which is also the M.L.E.

19. The mean of an exponential random variable with parameter β is $1/\beta$, so the method of moments estimator is one over the sample mean, which is also the M.L.E.

20. The mean of a Poisson random variable is θ, hence the method of moments estimator of θ is the sample mean, which is also the M.L.E.

21. The M.L.E. of the mean is the sample mean, which is the method of moments estimator. The M.L.E. of σ^2 is the mean of the X_i^2's minus the square of the sample mean, which is also the method of moments estimator of the variance.

22. The mean of X_i is $\theta/2$, so the method of moments estimator is $2\overline{X}_n$. The M.L.E. is the maximum of the X_i values.

23. (a) The means of X_i and X_i^2 are respectively $\alpha/(\alpha+\beta)$ and $\alpha(\alpha+1)/[(\alpha+\beta)(\alpha+\beta+1)$. We set these equal to the sample moments $\overline{x}_n$ and $\overline{x^2}_n$ and solve for α and β. After some tedious algebra, we get

$$\begin{aligned}\hat{\alpha} &= \frac{\overline{x}_n(\overline{x}_n - \overline{x^2}_n)}{\overline{x^2}_n - \overline{x}_n^2},\\ \hat{\beta} &= \frac{(1-\overline{x}_n)(\overline{x}_n - \overline{x^2}_n)}{\overline{x^2}_n - \overline{x}_n^2}.\end{aligned}$$

(b) The M.L.E. involves derivatives of the gamma function and the products $\prod_{i=1}^n x_i$ and $\prod_{i=1}^n (1-x_i)$.

24. The p.d.f. of each (X_i, Y_i) pair can be factored as

$$\frac{1}{(2\pi)^{1/2}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(x_i-\mu_1)^2\right) \frac{1}{(2\pi)^{1/2}\sigma_{2.1}} \exp\left(-\frac{1}{2\sigma_{2.1}^2}(y_i-\alpha-\beta x_i)^2\right), \tag{S.7.4}$$

where the new parameters are defined in the exercise. The product of n factors of the form (S.7.4) can be factored into the product of the n first factors times the product of the n second factors, each of which can be maximized separately because there are no parameters in common. The product of the first factors is the same as the likelihood of a sample of normal random variables, and the M.L.E.'s are $\hat{\mu}_1$ and $\widehat{\sigma_1^2}$ as stated in the exercise. The product of the second factors is slightly more complicated than the likelihood from a sample of normal random variables, but not much more so. Take the logarithm to get

$$-\frac{n}{2}[\log(2\pi)+\log(\sigma_{2.1}^2)] - \frac{1}{2\sigma_{2.1}^2}\sum_{i=1}^n (y_i-\alpha-\beta x_i)^2. \tag{S.7.5}$$

Taking the partial derivatives with respect to α and β yields

$$\begin{aligned} \frac{\partial}{\partial\alpha} &= \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^n (y_i-\alpha-\beta x_i), \\ \frac{\partial}{\partial\beta} &= \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^n x_i(y_i-\alpha-\beta x_i). \end{aligned}$$

Setting the first line equal to 0 and solving for α yields

$$\alpha = \overline{y}_n - \beta\overline{x}_n. \tag{S.7.6}$$

Plug (S.7.6) into the second of the partial derivatives to get (after a bit of algebra)

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i-\overline{x}_n)(y_i-\overline{y}_n)}{\sum_{i=1}^n (x_i-\overline{x}_n)^2}. \tag{S.7.7}$$

Substitute (S.7.7) back into (S.7.6) to get

$$\hat{\alpha} = \overline{y}_n - \hat{\beta}\hat{\mu}_1.$$

Next, take the partial derivative of (S.7.5) with respect to $\sigma_{2.1}^2$ to get

$$-\frac{n}{2\sigma_{2.1}^2} + \frac{1}{2\sigma_{2.1}^4}\sum_{i=1} (y_i-\alpha-\beta x_i)^2. \tag{S.7.8}$$

Now, substitue both $\hat{\alpha}$ and $\hat{\beta}$ into (S.7.8) and solve for $\sigma_{2.1}^2$. The result is

$$\widehat{\sigma_{2.1}^2} = \frac{1}{n}\sum_{i=1}^n (y_i-\hat{\alpha}-\hat{\beta}x_i)^2.$$

Finally, we can solve for the M.L.E.'s of the original parameters. We already have $\hat{\mu}_1$ and $\widehat{\sigma_1^2}$. The equation $\alpha = \mu_2 - \rho\sigma_2\mu_1/\sigma_1$ can be rewritten $\alpha = \mu_2 - \beta\mu_1$. It follows that

$$\hat{\mu}_2 = \hat{\alpha} + \hat{\beta}\hat{\mu}_1 = \overline{y}_n.$$

The equation $\beta = \rho\sigma_2/\sigma_1$ can be rewritten $\rho\sigma_2 = \beta\sigma_1$. Plugging this into $\sigma_{2.1}^2 = (1-\rho^2)\sigma_2^2$ yields $\sigma_{2.1}^2 = \sigma_2^2 - \beta^2\sigma_1^2$. Hence,

$$\begin{aligned}\widehat{\sigma_2^2} &= \widehat{\sigma_{2.1}^2} + \hat{\beta}^2\widehat{\sigma_1^2} \\ &= \frac{1}{m}\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \frac{\left[\sum_{i=1}^n (y_i - \overline{y}_n)(x_i - \overline{x}_n)\right]^2}{\sum_{i=1}^n (x_i - \overline{x}_n)^2} \\ &= \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y}_n)^2,\end{aligned}$$

where the final equality is tediuous but straightforward algebra. Finally,

$$\hat{\rho} = \frac{\hat{\beta}\widehat{\sigma_1}}{\widehat{\sigma_2}} = \frac{\sum_{i=1}^n (y_i - \overline{y}_n)(x_i - \overline{x}_n)}{\left[\sum_{i=1}^n (x_i - \overline{x}_n)^2\right]^{1/2}\left[\sum_{i=1}^n (y_i - \overline{y}_n)^2\right]^{1/2}}.$$

25. When we observe only the first $n-k$ Y_i's, the M.L.E.'s of μ_1 and σ_1^2 are not affected. The M.L.E.'s of α, β and $\sigma_{2.1}^2$ are just as in the previous exercise but with n replaced by $n-k$. The M.L.E.'s of μ_2, σ_2^2 and ρ are obtained by substituting $\hat{\alpha}$, $\hat{\beta}$ and $\widehat{\sigma_{2.1}^2}$ into the three equations Exercise 24:

$$\begin{aligned}\hat{\mu}_2 &= \hat{\alpha} + \hat{\beta}\hat{\mu}_1 \\ \widehat{\sigma_2^2} &= \widehat{\sigma_{2.1}^2} + \hat{\beta}^2\widehat{\sigma_1^2} \\ \hat{\rho} &= \frac{\hat{\beta}\widehat{\sigma_1}}{\widehat{\sigma_2}}.\end{aligned}$$

7.7 Sufficient Statistics

Commentary

The concept of sufficient statistics is fundamental to much of the traditional theory of statistical inference. However, it plays little or no role in the most common practice of statistics. For the most popular distributional models for real data, the most obvious data summaries are sufficient statistics. In Bayesian inference, the posterior distribution is automatically a function of every sufficient statistic, so one does not even have to think about sufficiency in Bayesian inference. For these reasons, the material in Secs. 7.7–7.9 should only be covered in courses that place a great deal of emphasis on the mathematical theory of statistics.

Solutions to Exercises

In Exercises 1–11, let t denote the value of the statistic T when the observed values of $X_1, \dots, X_n$ are $x_1, \dots, x_n$. In each exercise, we shall show that T is a sufficient statistic by showing that the joint p.f. or joint p.d.f. can be factored as in Eq. (7.7.1).

1. The joint p.f. is

$$f_n(\boldsymbol{x} \mid p) = p^t(1-p)^{n-t}.$$

2. The joint p.f. is

$$f_n(\boldsymbol{x} \mid p) = p^n(1-p)^t.$$

3. The joint p.f. is

$$f_n(\boldsymbol{x} \mid p) = \left[\prod_{i=1}^{n} \binom{r+x_i-1}{x_i}\right][p^{nr}(1-p)^t].$$

Since the expression inside the first set of square brackets does not depend on the parameter p, it follows that T is a sufficient statistic for p.

4. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{t}{2\sigma^2}\right\}.$$

5. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid \beta) = \left\{\frac{1}{[\Gamma(\alpha)]^n}\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1}\right\}\{\beta^{n\alpha}\exp(-n\beta t)\}.$$

6. The joint p.d.f. in this exercise is the same as that given in Exercise 5. However, since the unknown parameter is now α instead of β, the appropriate factorization is now as follows:

$$f_n(\boldsymbol{x} \mid \alpha) = \left\{\exp\left(-\beta\sum_{i=1}^{n} x_i\right)\right\}\left\{\frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n}t^{\alpha-1}\right\}.$$

7. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid \alpha) = \left\{\frac{1}{[\Gamma(\beta)]^n}\left[\prod_{i=1}^{n}(1-x_i)\right]^{\beta-1}\right\}\left\{\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\right]^n t^{\alpha-1}\right\}$$

8. The p.f. of an individual observation is

$$f(x \mid \theta) = \begin{cases} \dfrac{1}{\theta} & \text{for } x = 1, 2, \ldots, \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the joint p.f. is

$$f_n(\boldsymbol{x} \mid \theta) = \begin{cases} \dfrac{1}{\theta^n} & \text{for } t \le \theta, \\ 0 & \text{otherwise.} \end{cases}$$

If the function $h(t, \theta)$ is defined as in Example 7.7.5, with the values of t and θ now restricted to positive integers, then it follows that

$$f_n(\boldsymbol{x} \mid \theta) = \frac{1}{\theta^n} h(t, \theta).$$

9. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid b) = \frac{h(t,b)}{(b-a)^n},$$

where h is defined in Example 7.7.5.

10. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid a) = \frac{h(a,t)}{(b-a)^n},$$

where h is defined in Example 7.7.5.

11. The joint p.d.f. or joint p.f. is

$$f_n(\boldsymbol{x} \mid \theta) = \left\{ \prod_{i=1}^{n} b(x_i) \right\} \{[a(\theta)]^n \exp[c(\theta)t]\}.$$

12. The likelihood function is

$$\frac{\alpha^n x_0^{\alpha n}}{\prod_{i=1}^n x_i]^{\alpha+1}}, \tag{S.7.9}$$

for all $x_i \geq x_0$.

(a) If x_0 is known, α is the parameter, and (S.7.9) has the form $u(\boldsymbol{x})v[r(\boldsymbol{x}),\alpha]$, with $u(\boldsymbol{x}) = 1$ if all $x_i \geq x_0$ and 0 if not, $r(\boldsymbol{x}) = \prod_{i=1}^n x_i$, and $v[t,\alpha] = \alpha^n x_0^{\alpha n}/t^{\alpha+1}$. So $\prod_{i=1}^n X_i$ is a sufficient statistic.

(b) If α is known, x_0 is the parameter, and (S.7.9) has the form $u(\boldsymbol{x})v[r(\boldsymbol{x}),x_0]$, with $u(\boldsymbol{x}) = \alpha^n/[\prod_{i=1}^n x_i]^{\alpha+1}$, $r(\boldsymbol{x}) = \min\{x_1,\ldots,x_n\}$, and $v[t,x_0] = 1$ if $t \geq x_0$ and 0 if not. Hence $\min\{X_1,\ldots,X_n\}$ is a sufficient statistic.

13. The statistic T will be a sufficient statistic for θ if and only if $f_n(\boldsymbol{x} \mid \theta)$ can be factored as in Eq. (7.7.1). However, since $r(\boldsymbol{x})$ can be expressed as a function of $r'(\boldsymbol{x})$, and conversely, there will be a factorization of the form given in Eq. (7.7.1) if and only if there is a similar factorization in which the function v is a function of $r'(\boldsymbol{x})$ and θ. Therefore, T will be a sufficient statistic if and only if T' is a sufficient statistic.

14. This result follows from previous exercises in two different ways. First, by Exercise 6, the statistic $T' = \prod_{i=1}^n X_i$ is a sufficient statistic. Hence, by Exercise 13, $T = \log\, T'$ is also a sufficient statistic. A second way to establish the same result is to note that, by Exercise 24(g) of Sec. 7.3, the gamma distributions form an exponential family with $d(x) = \log\, x$. Therefore, by Exercise 11, the statistic $T = \sum_{i=1}^n d(X_i)$ is a sufficient statistic.

15. It follows from Exercise 11 and Exercise 24(i) of Sec. 7.3 that the statistic $T' = \sum_{i=1}^n \log(1 - X_i)$ is a sufficient statistic. Since T is a one-to-one function of T', it follows from Exercise 13 that T is also a sufficient statistic.

16. Let $f(\theta)$ be a prior p.d.f. for θ. The posterior p.d.f. of θ is, according to Bayes' theorem,

$$g(\theta|\boldsymbol{x}) = \frac{f_n(\boldsymbol{x}|\theta)f(\theta)}{\int f_n(\boldsymbol{x}|\psi)f(\psi)d\psi} = \frac{u(\boldsymbol{x})v[r(\boldsymbol{x}),\theta]f(\theta)}{\int u(\boldsymbol{x})v[r(\boldsymbol{x},\psi)]f(\psi)d\psi} = \frac{v[r(\boldsymbol{x}),\theta]f(\theta)}{\int v[r(\boldsymbol{x}),\psi]f(\psi)d\psi},$$

where the second equality uses the factorization criterion. One can see that this last expression depends on $\boldsymbol{x}$ only through $r(\boldsymbol{x})$.

17. First, suppose that T is sufficient. Then the likelihood function from observing $\boldsymbol{X} = \boldsymbol{x}$ is $u(\boldsymbol{x})v[r(\boldsymbol{x}), \theta]$, which is proportional to $v[r(\boldsymbol{x}), \theta]$. The likelihood from observing $T = t$ (when $t = r(\boldsymbol{x})$) is

$$\sum u(\boldsymbol{x})v[r(\boldsymbol{x}), \theta] = v[t, \theta] \sum u(\boldsymbol{x}), \tag{S.7.10}$$

where the sums in (S.7.10) are over all $\boldsymbol{x}$ such that $t = r(\boldsymbol{x})$. Notice that the right side of (S.7.10) is proportional to $v[t, \theta] = v[r(\boldsymbol{x}), \theta]$. So the two likelihoods are proportional. Next, suppose that the two likelihoods are proportional. That is, let $f(\boldsymbol{x}|\theta)$ be the p.f. of $\boldsymbol{X}$ and let $h(t|\theta)$ be the p.f. of T. If $t = r(\boldsymbol{x})$ then there exists $c(\boldsymbol{x})$ such that

$$f(\boldsymbol{x}|\theta) = u(\boldsymbol{x})h(t|\theta).$$

Let $v[t, \theta] = h(t|\theta)$ and apply the factorization criterion to see that T is sufficient.

7.8 Jointly Sufficient Statistics

Commentary

Even those instructors who wish to cover the concept of sufficient statistic in Sec. 7.7, may decide not to cover jointly sufficient statistics. This material is at a slightly more mathematical level than most of the text.

Solutions to Exercises

In Exercises 1–4, let t_1 and t_2 denote the values of T_1 and T_2 when the observed values of $X_1, \ldots, X_n$ are $x_1, \ldots, x_n$. In each exercise, we shall show that T_1 and T_2 are jointly sufficient statistics by showing that the joint p.d.f. of $X_1, \ldots, X_n$ can be factored as in Eq. (7.8.1).

1. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid \alpha, \beta) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} t_1^{\alpha-1} \exp(-\beta t_2).$$

2. The joint p.d.f. is

$$f_n(\boldsymbol{x} \mid \alpha, \beta) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]^n t_1^{\alpha-1} t_2^{\beta-1}.$$

3. Let the function h be as defined in Example 7.8.4. Then the joint p.d.f. can be written in the following form:

$$f_n(x \mid x_0, \alpha) = \frac{(\alpha x_0^{\alpha})^n}{t_2^{\alpha+1}} h(x_0, t_1).$$

4. Again let the function h be as defined in Example 7.8.4. Then the joint p.d.f. can be written as follows:

$$f_n(\boldsymbol{x} \mid \theta) = \frac{h(\theta, t_1)h(t_2, \theta+3)}{3^n}.$$

5. The joint p.d.f. of the vectors (X_i, Y_i), for $i = 1, \dots, n$, was given in Eq. (5.10.2). The following relations hold:

$$\sum_{i=1}^n (x_i - \mu_1)^2 = \sum_{i=1}^n x_i^2 - 2\mu_1 \sum_{i=1}^n x_i + n\mu_1^2,$$
$$\sum_{i=1}^n (y_i - \mu_2)^2 = \sum_{i=1}^n y_i^2 - 2\mu_2 \sum_{i=1}^n y_i + n\mu_2^2,$$
$$\sum_{i=1}^n (x_i - \mu_1)(y_i - \mu_2) = \sum_{i=1}^n x_i y_i - \mu_2 \sum_{i=1}^n x_i - \mu_1 \sum_{i=1}^n y_i + n\mu_1\mu_2.$$

Because of these relations, it can be seen that the joint p.d.f. depends on the observed values of the n vectors in the sample only through the values of the five statistics given in this exercise. Therefore, they are jointly sufficient statistics.

6. The joint p.d.f. or joint p.f. is

$$f_n(\boldsymbol{x} \mid \theta) = \left\{ \prod_{j=1}^n b(x_j) \right\} \left\{ [a(\theta)]^n \exp \left[\sum_{i=1}^k c_i(\theta) \sum_{j=1}^k d_i(x_j) \right] \right\}.$$

It follows from the factorization criterion that $T_1, \dots, T_k$ are jointly sufficient statistics for θ.

7. In each part of this exercise we shall first present the p.d.f. f, and then we shall identify the functions a, b, c_1, d_1, c_2, and d_2 in the form for a two-parameter exponential family given in Exercise 6.

   (a) Let $\theta = (\mu, \sigma^2)$. Then $f(x \mid \theta)$ is as given in the solution to Exercise 24(d) of Sec. 7.3. Therefore, $a(\theta) = \dfrac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\dfrac{\mu^2}{2\sigma^2}\right)$, $b(x) = 1$, $c_1(\theta) = -\dfrac{1}{2\sigma^2}$, $d_1(x) = x^2$, $c_2(\theta) = \dfrac{\mu}{\sigma^2}$, $d_2(x) = x$.

   (b) Let $\theta = (\alpha, \beta)$. Then $f(x \mid \theta)$ is as given in the solution to Exercise 24(f) of Sec. 7.3. Therefore, $a(\theta) = \dfrac{\beta^\alpha}{\Gamma(\alpha)}$, $b(x) = 1$, $c_1(\theta) = \alpha - 1$, $d_1(x) = \log\ x$, $c_2(\theta) = -\beta$, $d_2(x) = x$.

   (c) Let $\theta = (\alpha, \beta)$. Then $f(x \mid \theta)$ is as given in the solution to Exercise 24(h) of Sec. 7.3. Therefore, $a(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $b(x) = 1$, $c_1(\theta) = \alpha - 1$, $d_1(x) = \log x$, $c_2(\theta) = \beta - 1$, and $d_2(x) = \log(1 - x)$.

8. The M.L.E. of β is $n / \sum_{i=1}^n X_i$. (See Exercise 7 in Sec. 7.5.) This is a one-to-one function of the sufficient statistic found in Exercise 5 of Sec. 7.7. Hence, the M.L.E. is sufficient. This makes it minimal sufficient.

9. By Example 7.5.4, $\hat{p} = \overline{X}_n$. By Exercise 1 of Sec. 7.7, $\hat{p}$ is a sufficient statistic. Therefore, $\hat{p}$ is a minimal sufficient statistic.

10. By Example 7.5.7, $\hat{\theta} = \max\{X_1, \dots, X_n\}$. By Example 7.7.5, $\hat{\theta}$ is a sufficient statistic. Therefore, $\hat{\theta}$ is a minimal sufficient statistic.

11. By Example 7.8.5, the order statistics are minimal jointly sufficient statistics. Therefore, the M.L.E. of θ, all by itself, cannot be a sufficient statistic. (We know from Example 7.6.5 that there is no simple expression for this M.L.E., so we cannot solve this exercise by first deriving the M.L.E. and then checking to see whether it is a sufficient statistic.)

12. If we let $T = \max\{X_1, \ldots, X_n\}$, let t denote the observed value of T, and let the function h be as defined in Example 7.8.4, then the likelihood function can be written as follows:

$$f_n(\boldsymbol{x} \mid \theta) = \frac{2^n \left(\prod_{i=1}^{n} x_i\right)}{\theta^{2n}} h(t, \theta).$$

This function will be a maximum when θ is chosen as small as possible, subject to the constraint that $h(t, \theta) = 1$. Therefore, the M.L.E. of θ is $\hat{\theta} = t$. The median m of the distribution will be the value such that

$$\int_0^m f(x \mid \theta)\, dx = \frac{1}{2}.$$

Hence, $m = \theta/\sqrt{2}$. It follows from the invariance principle that the M.L.E. of m is $\hat{m} = \hat{\theta}/\sqrt{2} = t/\sqrt{2}$. By applying the factorization criterion to $f_n(\boldsymbol{x} \mid \theta)$, it can be seen that the statistic T is a sufficient statistic for θ. Therefore, the statistic $T/\sqrt{2}$ which is the M.L.E. of m, is also a sufficient statistic for θ.

13. By Exercise 11 of Sec. 7.5, $\hat{a} = \min\{X_1, \ldots, X_n\}$ and $\hat{b} = \max\{X_1, \ldots X_n\}$. By Example 7.8.4, $\hat{a}$ and $\hat{b}$ are jointly sufficient statistics. Therefore, $\hat{a}$ and $\hat{b}$ are minimal jointly sufficient statistics.

14. It can be shown that the values of the five M.L.E.'s given in Exercise 24 of Sec. 7.6 can be derived from the values of the five statistics given in Exercise 5 of this section by a one-to-one transformation. Since the five statistics in Exercise 5 are jointly sufficient statistics, the five M.L.E.'s are also jointly sufficient statistics. Hence, the M.L.E.'s will be minimal jointly sufficient statistics.

15. The Bayes estimator of p is given by Eq. (7.4.5). Since $\sum_{i=1}^n x_i$ is a sufficient statistic for p, the Bayes estimator is also a sufficient statistic for p. Hence, this estimator will be a minimal sufficient statistic.

16. It follows from Theorem 7.3.2 that the Bayes estimator of λ is $(\alpha + \sum_{i=1}^n X_i)/(\beta + n)$. Since $\sum_{i=1}^n X_i$ is a sufficient statistic for λ, the Bayes estimator is also a sufficient statistic for λ. Hence, this estimator will be a minimal sufficient statistic.

17. The Bayes estimator of μ is given by Eq. (7.4.6). Since $\overline{X}_n$ is a sufficient statistic for μ, the Bayes estimator is also a sufficient statistic. Hence, this estimator will be a minimal sufficient statistic.

7.9 Improving an Estimator

Commentary

If you decided to cover the material in Secs. 7.7 and 7.8, this section gives one valuable application of that material, Theorem 7.9.1 of Blackwell and Rao. This section ends with some foundational discussion of the use of sufficient statistics. This material is included for those instructors who want their students to have both a working and a critical understanding of the topic.

Solutions to Exercises

1. The statistic $Y_n = \sum_{i=1}^n X_i^2$ is a sufficient statistic for θ. Since the value of the estimator δ_1 cannot be determined from the value of Y_n alone, δ_1 is inadmissible.

2. A sufficient statistic in this example is $\max\{X_1, \dots, X_n\}$. Since $2\overline{X}_n$ is not a function of the sufficient statistic, it cannot be admissible.

3. The mean of the uniform distribution on the interval $[0, \theta]$ is $\theta/2$ and the variance is $\theta^2/12$. Therefore, $E_\theta(\overline{X}_n) = \theta/2$ and $\text{Var}_\theta(\overline{X}_n) = \theta^2/(12n)$. In turn, it now follows that

$$E_\theta(\delta_1) = \theta \text{ and } \text{Var}_\theta(\delta_1) = \frac{\theta^2}{3n}.$$

Hence, for $\theta > 0$,

$$R(\theta, \delta_1) = E_\theta[(\delta_1 - \theta)^2] = \text{Var}_\theta(\delta_1) = \frac{\theta^2}{3n}.$$

4. (a) It follows from the discussion given in Sec. 3.9 that the p.d.f. of Y_n is as follows:

$$g(y \mid \theta) = \begin{cases} \dfrac{ny^{n-1}}{\theta^n} & \text{for } 0 \le y \le \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, for $\theta > 0$,

$$R(\theta, \delta_2) = E_\theta[(Y_n - \theta)^2] = \int_0^\theta (y - \theta)^2 \frac{ny^{n-1}}{\theta^n} dy = \frac{2\theta^2}{(n+1)(n+2)}.$$

(b) if $n = 2, R(\theta, \delta_1) = R(\theta, \delta_2) = \theta^2/6$.

(c) Suppose $n \ge 3$. Then $R(\theta, \delta_2) < R(\theta, \delta_1)$ for any given value of $\theta > 0$ if and only if $2/[(n+1)(n+2)] < 1/(3n)$ or, equivalently, if and only if $6n < (n+1)(n+2) = n^2 + 3n + 2$. Hence, $R(\theta, \delta_2) < R(\theta, \delta_1)$ if and only if $n^2 - 3n + 2 > 0$ or, equivalently, if and only if $(n-2)(n-1) > 0$. Since this inequality is satisfied for all values of $n \ge 3$, it follows that $R(\theta, \delta_2) < R(\theta, \delta_1)$ for every value of $\theta > 0$. Hence, δ_2 dominates δ_1.

5. For any constant c,

$$\begin{aligned} R(\theta, cY_n) &= E_\theta[(cY_n - \theta)^2] = c^2 E_\theta(Y_n^2) - 2c\theta E_\theta(Y_n) + \theta^2 \\ &= \left(\frac{n}{n+2} c^2 - \frac{2n}{n+1} c + 1 \right) \theta^2. \end{aligned}$$

Hence, for any given value of n and any given value of $\theta > 0, R(\theta, cY_n)$ will be a minimum when c is chosen so that the coefficient of θ^2 in this expression is a minimum. By differentiating with respect to c, we find that the minimizing value of c is $c = (n+2)(n+1)$. Hence, the estimator $(n+2)Y_n/(n+1)$ dominates every other estimator of the form cY_n.

6. It was shown in Exercise 6 of Sec. 7.7 that $\prod_{i=1}^n X_i$ is a sufficient statistic in this problem. Since the value of $\overline{X}_n$ cannot be determined from the value of the sufficient statistic alone, $\overline{X}_n$ is inadmissible.

7. (a) Since the value of δ is always 3, $R(\beta, \delta) = (\beta - 3)^2$.

(b) Since $R(3, \delta) = 0$, no other estimator δ_1 can dominate δ unless it is also true that $R(3, \delta_1) = 0$. But the only way that the M.S.E. of an estimator δ_1 can be 0 is for the estimator δ_1 to be equal to 3 with probability 1. In other words, the estimator δ_1 must be the same as the estimator δ. Therefore, δ is not dominated by any other estimator and it is admissible.

In other words, the estimator that always estimates the value of β to be 3 is admissible because it is the best estimator to use if β happens to be equal to 3. Of course, it is a poor estimator to use if β happens to be different from 3.

8. It was shown in Example 7.7.2 that $\sum_{i=1}^{n} X_i$ is a sufficient statistic in this problem. Since the proportion $\hat{\beta}$ of observations that have the value 0 cannot be determined from the value of the sufficient statistic alone, $\hat{\beta}$ is inadmissible.

9. Suppose that X has a continuous distribution for which the p.d.f. is f. Then

$$E(X) = \int_{-\infty}^{0} x f(x)\, dx + \int_{0}^{\infty} x f(x)\, dx.$$

Suppose first that $E(X) \leq 0$. Then

$$\begin{aligned}|E(X)| = -E(X) &= \int_{-\infty}^{0} (-x) f(x)\, dx - \int_{0}^{\infty} x f(x)\, dx \\ &\leq \int_{-\infty}^{0} (-x) f(x)\, dx + \int_{0}^{\infty} x f(x)\, dx \\ &= \int_{-\infty}^{\infty} |x| f(x)\, dx = E(|X|).\end{aligned}$$

A similar proof can be given if X has a discrete distribution or a more general type of distribution, or if $E(X) > 0$.

Alternatively, the result is immediate from Jensen's inequality, Theorem 4.2.5.

10. We shall follow the steps of the proof of Theorem 7.9.1. It follows from Exercise 9 that

$$E_\theta(|\,\delta - \theta\,||T) \geq |\,E_\theta(\delta - \theta\,|\,T)| = |E_\theta(\delta\,|\,T) - \theta| = |\delta_0 - \theta|.$$

Therefore,

$$E_\theta(|\,\delta_0 - \theta\,|) \leq E_\theta[E_\theta(|\,\delta - \theta\,||T)] = E_\theta(|\,\delta - \theta\,|).$$

11. Since $\hat{\theta}$ is the M.L.E. of θ, we know from the discussion in Sec. 7.8 that $\hat{\theta}$ is a function of T alone. Since $\hat{\theta}$ is already a function of T, taking the conditional expectation $E(\hat{\theta}|T)$ will not affect $\hat{\theta}$. Hence, $\delta_0 = E(\hat{\theta}|T) = \hat{\theta}$.

12. Since X_1 must be either 0 or 1,

$$E(X_1\,|\,T = t) = \Pr(X_1 = 1\,|\,T = t).$$

If $t = 0$ then every X_i must be 0. Therefore,

$$E(X_1\,|\,T = 0) = 0.$$

Suppose now that $t > 0$. Then

$$\Pr(X_1 = 1\,|\,T = t) = \frac{\Pr(X_1 = 1 \text{ and } T = t)}{\Pr(T = t)} = \frac{\Pr\left(X_1 = 1 \text{ and } \sum_{i=2}^{n} X_i = t - 1\right)}{\Pr(T = t)}.$$

Since X_1 and $\sum_{i=2}^n X_i$ are independent,

$$\begin{aligned}
\Pr\left(X_1 = 1 \text{ and } \sum_{i=2}^n X_i = t-1\right) &= \Pr(X_1 = 1)\Pr\left(\sum_{i=2}^n X_i = t-1\right)\\
&= p \cdot \binom{n-1}{t-1} p^{t-1}(1-p)^{(n-1)-(t-1)}\\
&= \binom{n-1}{t-1} p^t (1-p)^{n-t}.
\end{aligned}$$

Also, $\Pr(T = t) = \binom{n}{t} p^t (1-p)^{n-t}$. It follows that

$$\Pr(X_1 = 1 | T = t) = \binom{n-1}{t-1} \Big/ \binom{n}{t} = \frac{t}{n}.$$

Therefore, for any value of T,

$$E(X_1|T) = T/n = \overline{X}_n.$$

A more elegant way to solve this exercise is as follows: By symmetry, $E(X_1|T) = E(X_2|T) = \ldots = E(X_n\,|\,T)$. Therefore, $nE(X_1\,|\,T) = \sum_{i=1}^n E(X_i\,|\,T)$. But

$$\sum_{i=1}^n E(X_i\,|\,T) = E\left(\sum_{i=1}^n X_i\,|\,T\right) = E(T\,|\,T) = T.$$

Hence, $E(X_1|T) = T/n$.

13. We shall carry out the analysis for Y_1. The analysis for every other value of i is similar. Since Y_1 must be 0 or 1,

$$\begin{aligned}
E(Y_1|T = t) &= \Pr(Y_1 = 1|T = t) = \Pr(X_1 = 0|T = t)\\
&= \frac{\Pr(X_1 = 0 \text{ and } T = t)}{\Pr(T = t)} = \frac{\Pr\left(X_1 = 0 \text{ and } \sum_{i=2}^n X_i = t\right)}{\Pr(T = t)}.
\end{aligned}$$

The random variables X_1 and $\sum_{i=2}^n X_i$ are independent, X_1 has a Poisson distribution with mean θ, and $\sum_{i=2}^n X_i$ has a Poisson distribution with mean $(n-1)\theta$. Therefore,

$$\Pr\left(X_1 = 0 \text{ and } \sum_{i=2}^n X_i = t\right) = \Pr(X_1 = 0)\Pr\left(\sum_{i=2}^n X_i = t\right) = \exp(-\theta) \cdot \frac{\exp(-(n-1)\theta)[(n-1)\theta]^t}{t!}.$$

Also, since T has a Poisson distribution with mean $n\theta$,

$$\Pr(T = t) = \frac{\exp(-n\theta)(n\theta)^t}{t!}$$

It now follows that $E(Y_1|T = t) = ([n-1]/n)^t$.

14. If Y_i is defined as in Exercise 12 for $i = 1, \dots, n$, then $\hat{\beta} = \sum_{i=1}^{n} Y_i/n$. Also, we know from Exercise 12 that $E(Y_i|T) = ([n-1]/n)^T$ for $i = 1, \dots, n$. Therefore,

$$E(\hat{\beta} \mid T) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \mid T\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i \mid T) = \frac{1}{n}\cdot n\left(\frac{n-1}{n}\right)^T = \left(\frac{n-1}{n}\right)^T.$$

15. Let $\hat{\theta}$ be the M.L.E. of θ. Then the M.L.E. of $\exp(\theta + 0.125)$ is $\exp(\hat{\theta} + 0.125)$. The M.L.E. of θ is $\overline{X}_n$, so the M.L.E. of $\exp(\theta+0.125)$ is $\exp(\overline{X}_n+0.125)$. The M.S.E. of an estimator of the form $\exp(\overline{X}_n+c)$ is

$$\begin{aligned}
&E\left[(\exp[\overline{X}_n + c] - \exp[\theta + 0.125])^2\right] \\
&= \mathrm{Var}(\exp[\overline{X}_n + c]) + \left[E(\exp[\overline{X}_n + c]) - \exp(\theta + 0.125)\right]^2 \\
&= \exp(2\theta + 0.25/n + 2c)[\exp(0.25/n) - 1] + [\exp(\theta + 0.125/n + c) - \exp(\theta + 0.125)]^2 \\
&= \exp(2\theta)\{\exp(0.25/n + 2c)[\exp(0.25/n) - 1] + \exp(0.25/n + 2c) - 2\exp(0.125[1 + 1/n] + c) \\
&\quad + \exp(0.5)\} \\
&= \exp(2\theta)\left[\exp(2c)\exp(0.5/n) - 2\exp(c)\exp(0.125[1 + 1/n]) + \exp(0.5)\right].
\end{aligned}$$

Let $a = \exp(c)$ in this last expression. Then we can minimize the M.S.E. simultaneously for all θ by minimizing

$$a^2\exp(0.5/n) - 2a\exp(0.125[1 + 1/n]) + \exp(0.5).$$

The minimum occurs at $a = \exp(0.125 - 0.375/n)$, so $c = 0.125 - 0.375/n$.

16. $p = \Pr(X_i = 1|\theta) = \exp(-\theta)\theta$. The M.L.E. of θ is the number of arrivals divided by the observation time, namely $\overline{X}_n$. So, the M.L.E. of p is $\exp(-\overline{X}_n)\overline{X}_n$. In Example 7.9.2, $T/n = \overline{X}_n$. If n is large, then T should also be large so that $(1 - 1/n)^T \approx \exp(-T/n)$ according to Theorem 5.3.3.

7.10 Supplementary Exercises

Solutions to Exercises

1. (a) The prior distribution of θ is the beta distribution with parameters 1 and 1, so the posterior distribution of θ is the beta distribution with parameters $1 + 10 = 11$ and $1 + 25 - 10 = 16$.

   (b) With squared error loss, the estimate to use is the posterior mean, which is $11/27$ in this case.

2. We know that the M.L.E. of $\theta = \overline{X}_n$. Hence, by the invariance property described in Sec. 7.6, the M.L.E. of θ^2 *is* $\overline{X}_n^2$.

3. The prior distribution of θ is the beta distribution with $\alpha = 3$ and $\beta = 4$, so it follows from Theorem 7.3.1 that the posterior distribution is the beta distribution with $\alpha = 3 + 3 = 6$ and $\beta = 4 + 7 = 11$. The Bayes estimate is the mean of this posterior distribution, namely $6/17$.

4. Since the joint p.d.f. of the observations is equal to $1/\theta^n$ provided that $\theta \le x_i \le 2\theta$ for $i = 1, \dots, n$, the M.L.E. will be the smallest value of θ that satisfies these restrictions. Since we can rewrite the restrictions in the form

$$\frac{1}{2}\max\{x_1, \dots, x_n\} \le \theta \le \min\{x_1, \dots, x_n\}$$

it follows that the smallest possible value of θ is

$$\hat{\theta} = \frac{1}{2}\max\{x_1, \ldots, x_n\}.$$

5. The joint p.d.f. of X_1 and X_2 is

$$\frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2\sigma_1^2}(x_1 - b_1\mu)^2 - \frac{1}{2\sigma_2^2}(x_2 - b_2\mu)^2\right].$$

If we let $L(\mu)$ denote the logarithm of this expression, and solve the equation $dL(\mu)/d\mu = 0$, we find that

$$\hat{\mu} = \frac{\sigma_2^2 b_1 x_1 + \sigma_1^2 b_2 x_2}{\sigma_2^2 b_1^2 + \sigma_1^2 b_2^2}.$$

6. Since $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, it follows that $\Gamma'(\alpha+1) = \alpha\Gamma'(\alpha) + \Gamma(\alpha)$. Hence,

$$\begin{aligned} \psi(\alpha+1) &= \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)} = \frac{\alpha\Gamma'(\alpha)}{\Gamma(\alpha+1)} + \frac{\Gamma(\alpha)}{\Gamma(\alpha+1)} \\ &= \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \frac{1}{\alpha} = \psi(\alpha) + \frac{1}{\alpha}. \end{aligned}$$

7. The joint p.d.f. of X_1, X_2, X_3 is

$$f(\boldsymbol{x}|\theta) = \frac{1}{\theta}\exp\left(-\frac{1}{\theta}x_1\right)\cdot\frac{1}{2\theta}\exp\left(-\frac{1}{2\theta}x_2\right)\cdot\frac{1}{3\theta}\exp\left(-\frac{1}{3\theta}x_3\right) = \frac{1}{6\theta^3}\exp\left[-\left(x_1 + \frac{x_2}{2} + \frac{x_3}{3}\right)\frac{1}{\theta}\right].$$

(a) By solving the equation $\partial\log(f)/\partial\theta = 0$, we find that

$$\hat{\theta} = \frac{1}{3}\left(X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3\right).$$

(b) In terms of ψ, the joint p.d.f. of X_1, X_2, X_3 is

$$f(\boldsymbol{x} \mid \psi) = \frac{\psi^3}{6}\exp\left[-\left(x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3\right)\psi\right].$$

Since the prior p.d.f. of ψ is

$$\xi(\psi) \propto \psi^{\alpha-1}\exp(-\beta\psi),$$

it follows that the posterior p.d.f. is

$$\xi(\psi \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \psi)\xi(\psi) \propto \psi^{\alpha+2}\exp\left[-\left(\beta + x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3\right)\psi\right].$$

Hence, the posterior distribution of ψ is the gamma distribution with parameters $\alpha+3$ and $\beta + x_1 + x_2/2 + x_3/3$.

8. The joint p.f. of $X_2, \ldots, X_{n+1}$ is the product of n factors. If $X_i = x_i$ and $X_{i+1} = x_{i+1}$, then the ith factor will be the transition probability of going from state x_i to state x_{i+1} $(i = 1, \ldots, n)$. Hence, each transition from s_1 to s_1 will introduce the factor θ, each transition from s_1 to s_2 will introduce the factor $1 - \theta$, and every other transition will introduce either the constant factor $3/4$ or the constant factor $1/4$. Hence, if A denotes the number of transitions from s_1 to s_1 among the n transitions and B denotes the number from s_1 to s_2, then the joint p.f. of $X_2, \ldots, X_{n+1}$ has the form (const.) $\theta^A(1-\theta)^B$. Therefore, this joint p.f., or likelihood function, is maximized when

$$\hat{\theta} = \frac{A}{A+B}.$$

9. The posterior p.d.f. of θ given $X = x$ satisfies the relation

$$\xi(\theta \mid \boldsymbol{x}) \propto f(x \mid \theta)\xi(\theta) \propto \exp(-\theta), \text{ for } \theta > x.$$

Hence,

$$\xi(\theta \mid \boldsymbol{x}) = \begin{cases} \exp(x - \theta) & \text{for } \theta > x, \\ 0 & \text{otherwise.} \end{cases}$$

(a) The Bayes estimator is the mean of this posterior distribution, $\hat{\theta} = x + 1$.

(b) The Bayes estimator is the median of this posterior distribution, $\hat{\theta} = x + \log\ 2$.

10. In this exercise, θ must lie in the interval $1/3 \le \theta \le 2/3$. Hence, as in Exercise 3 of Sec. 7.5, the M.L.E. of θ is

$$\hat{\theta} = \begin{cases} \overline{X}_n & \text{if } \frac{1}{3} \le \overline{X}_n \le \frac{2}{3}, \\ \frac{1}{3} & \text{if } \overline{X}_n < \frac{1}{3}, \\ \frac{2}{3} & \text{if } \overline{X}_n > \frac{2}{3}. \end{cases}$$

It then follows from Sec. 7.6 that $\hat{\beta} = 3\hat{\theta} - 1$.

11. Under these conditions, X has a binomial distribution with parameters n and $\theta = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2}p = \frac{1}{4} + \frac{1}{2}p$. Since $0 \le p \le 1$, it follows that $1/4 \le \theta \le 3/4$. Hence, as in Exercise 3 of Sec. 7.5, the M.L.E. of $\hat{\theta}$ is

$$\hat{\theta} = \begin{cases} \frac{X}{n} & \text{if } \frac{1}{4} \le \frac{X}{n} \le \frac{3}{4}, \\ \frac{1}{4} & \text{if } \frac{X}{n} < \frac{1}{4}, \\ \frac{3}{4} & \text{if } \frac{X}{n} > \frac{3}{4}. \end{cases}$$

It then follows from Theorem 7.6.1 that the M.L.E. of p is $\hat{p} = 2(\hat{\theta} - 1/4)$.

12. The prior distribution of θ is the Pareto distribution with parameters $x_0 = 1$ and $\alpha = 1$. Therefore, it follows from Exercise 18 of Sec. 7.3 that the posterior distribution of θ will be a Pareto distribution with parameters $\alpha + n$ and $\max\{x_0, x_1, \ldots, x_n\}$. In this exercise $n = 4$ and $\max\{x_0, x_1, \ldots, x_n\} = 1$. Hence, the posterior Pareto distribution has parameters $\alpha = 5$ and $x_0 = 1$. The Bayes estimate of θ will be the mean of this posterior distribution, namely

$$\hat{\theta} = \int_1^\infty \theta \quad \frac{5}{\theta^6}\, d\theta = \frac{5}{4}.$$

13. The Bayes estimate of θ will be the median of the posterior Pareto distribution. This will be the value m such that

$$\frac{1}{2} = \int_m^\infty \frac{5}{\theta^6}\, d\theta = \frac{1}{m^5}.$$

Hence, $\hat{\theta} = m = 2^{1/5}$.

14. The joint p.d.f. of $X_1, \ldots, X_n$ can be written in the form

$$f_n(\boldsymbol{x}|\beta, \theta) = \beta^n \exp\left(n\beta\theta - \beta\sum_{i=1}^n x_i\right)$$

for $\min\{x_1, \ldots, x_n\} \geq \theta$, and $f_n(\boldsymbol{x}|\beta, \theta) = 0$ otherwise. Hence, by the factorization criterion, $\sum_{i=1}^n X_i$ and $\min\{X_1, \ldots, X_n\}$ is a pair of jointly sufficient statistics and so is any other pair of statistics that is a one-to-one function of this pair.

15. The joint p.d.f. of the observations is $\alpha^n\, {x_0}^{n\alpha} / \left(\prod_{i=1}^n x_i\right)^{\alpha+1}$ for $\min\{x_1, \ldots, x_n\} \geq x_0$. This p.d.f. is maximized when x_0 is made as large as possible. Thus,

$$\hat{x}_0 = \min\{X_1, \ldots, X_n\}.$$

16. Since α is known in Exercise 15, it follows from the factorization criterion, by a technique similar to that used in Example 7.7.5 or Exercise 12 of Sec. 7.8, that $\min\{X_1, \ldots, X_n\}$ is a sufficient statistic. Thus, from Theorem 7.8.3, since the M.L.E. $\hat{x}_0$ is a sufficient statistic, it is a minimal sufficient statistic.

17. It follows from Exercise 15 that $\hat{x}_0 = \min\{X_1, \ldots, X_n\}$ will again be the M.L.E. of x_0 , since this value of x_0 maximizes the likelihood function regardless of the value of α. If we substitute $\hat{x}_0$ for x_0 and let $L(\alpha)$ denote the logarithm of the resulting likelihood function, which was given in Exercise 15, then

$$L(\alpha) = n\log\alpha + n\ \alpha\log\hat{x}_0 - (\alpha+1)\sum_{i=1}^n \log x_i$$

and

$$\frac{dL(\alpha)}{d\alpha} = \frac{n}{\alpha} + n\log\hat{x}_0 - \sum_{i=1}^n \log x_i.$$

Hence, by setting this expression equal to 0, we find that

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^n \log x_i - \log\hat{x}_0\right)^{-1}.$$

18. It can be shown that the pair of estimators $\hat{x}_0$ and $\hat{\alpha}$ found in Exercise 17 form a one-to-one transform of the pair of jointly sufficient statistics T_1 and T_2 given in Exercise 3 of Sec. 7.8. Hence, $\hat{x}_0$ and $\hat{\alpha}$ are themselves jointly sufficient statistics. It now follows from Sec. 7.8 that $\hat{x}_0$ and $\hat{\alpha}$ must be minimal jointly sufficient statistics.

19. The p.f. of X is

$$f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The M.L.E. of n will be the value that maximizes this expression for given values of x and p. The ratio given in the hint to this exercise reduces to

$$R = \frac{n+1}{n+1-x} \quad (1-p).$$

Since R is a decreasing function of n, it follows that $f(x|n,p)$ will be maximized at the smallest value of n for which $R < 1$. After some algebra, it is found that $R < 1$ if and only if $n > x/p - 1$. Hence, n will be the smallest integer greater than $x/p - 1$. If $x/p - 1$ is itself an integer, then $x/p - 1$ and x/p are both M.L.E.'s.

20. The joint p.d.f. of X_1 and X_2 is $1/(4\theta^2)$ provided that each of the observations lies in either the interval $(0,\theta)$ or the interval $(2\theta, 3\theta)$. Thus, the M.L.E. of θ will be the smallest value of θ for which these restrictions are satisfied.

   (a) If we take $3\hat{\theta} = 9$, or $\hat{\theta} = 3$, then $\hat{\theta}$ will be as small as possible, and the restrictions will be satisfied because both observed values will lie in the interval $(2\hat{\theta}, 3\hat{\theta})$.

   (b) It is not possible that both X_1 and X_2 lie in the interval $(2\theta, 3\theta)$, because for that to be true it is necessary that $X_2/X_1 \le 3/2$. Here, however, $X_2/X_1 = 9/4$. Therefore, if we take $\hat{\theta} = 4$, then $\hat{\theta}$ will be as small as possible and the restrictions will be satisfied because X_1 will lie in the interval $(0, \hat{\theta})$ and X_2 will lie in $(2\hat{\theta}, 3\hat{\theta})$.

   (c) It is not possible that both X_1 and X_2 lie in the interval $(2\theta, 3\theta)$ for the reason given in part (b). It is also not possible that X_1 lies in $(0,\theta)$ and X_2 lies in $(2\theta, 3\theta)$, because for that to be true it is necessary that $X_2/X_1 \ge 2$. Here, however, $X_2/X_1 = 9/5$. Hence, it must be true that both X_1 and X_2 lie in the interval $(0,\theta)$. Under this condition, the smallest possible value of θ is $\hat{\theta} = 9$.

21. The Bayes estimator of θ is the mean of the posterior distribution of θ, and the expected loss or M.S.E. of this estimator is the variance of the posterior distribution. This variance, as given by Eq. (7.3.2), is

$$\nu_1^2 = \frac{(100)(25)}{100 + 25n} = \frac{100}{n+4}.$$

Hence, n must be chosen to minimize

$$\frac{100}{n+4} + \frac{1}{4}\, n.$$

By setting the first derivative equal to 0, it is found that the minimum occurs when $n = 16$.

22. It was shown in Example 7.7.2 that $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic in this problem. Since the sample variance is not a function of T alone, it follows from Theorem 7.9.1 that it is inadmissible.

Chapter 8

Sampling Distributions of Estimators

8.1 The Sampling Distribution of a Statistic

Solutions to Exercises

1. The c.d.f. of $U = \max\{X_1, \ldots, X_n\}$ is

$$F(u) = \begin{cases} 0 & \text{for } u \le 0, \\ (u/\theta)^n & \text{for } 0 < u < \theta, \\ 1 & \text{for } u \ge \theta. \end{cases}$$

Since $U \le \theta$ with probability 1, the event that $|U - \theta| \le 0.1\theta$ is the same as the event that $U \ge 0.9\theta$. The probability of this is $1 - F(0.9\theta) = 1 - 0.9^n$. In order for this to be at least 0.95, we need $0.9^n \le 0.05$ or $n \ge \log(0.05)/\log(0.9) = 28.43$. So $n \ge 29$ is needed.

2. It is known that $\overline{X}_n$ has the normal distribution with mean θ and variance $4/n$. Therefore,

$$E_\theta(|\overline{X}_n - \theta|^2) = \text{Var}_\theta(\overline{X}_n) = 4/n,$$

and $4/n \le 0.1$ if and only if $n \ge 40$.

3. Once again, $\overline{X}_n$ has the normal distribution with mean θ and variance $4/n$. Hence, the random variable $Z = (\overline{X}_n - \theta)/(2/\sqrt{n})$ will has the standard normal distribution. Therefore,

$$\begin{aligned} E_\theta(|\overline{X}_n - \theta|) &= \frac{2}{\sqrt{n}} E_\theta(|Z|) = \frac{2}{\sqrt{n}} \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) dz = 2\sqrt{\frac{2}{n\pi}} \int_0^\infty z \exp(-z^2/2) dz \\ &= 2\sqrt{\frac{2}{n\pi}}. \end{aligned}$$

But $2\sqrt{2/(n\pi)} \le 0.1$ if and only if $n \ge 800/\pi = 254.6$. Hence, n must be least 255.

4. If Z is defined as in the solution of Exercise 3, then

$$\Pr(|\overline{X}_n - \theta| \le 0.1) = \Pr(|Z| \le 0.05\sqrt{n}) = 2\Phi(0.05\sqrt{n}) - 1.$$

Therefore, this value will be at least 0.95 if and only if $\Phi(0.05\sqrt{n}) \ge 0.975$. It is found from a table of values of Φ that we must have $0.05\sqrt{n} \ge 1.96$. Therefore, we must have $n \ge 1536.64$ or, since n must be an integer, $n \ge 1537$.

5. When $p = 0.2$, the random variable $Z_n = n\overline{X}_n$ will have a binomial distribution with parameters n and $p = 0.2$, and

$$\Pr(|\overline{X}_n - p| \le 0.1) = \Pr(0.1n \le Z_n \le 0.3n).$$

The value of n for which this probability will be at least 0.75 must be determined by trial and error from the table of the binomial distribution in the back of the book. For $n = 8$, the probability becomes

$$\Pr(0.8 \le Z_8 \le 2.4) = \Pr(Z_8 = 1) + \Pr(Z_8 = 2) = 0.3355 + 0.2936 = 0.6291.$$

For $n = 9$, we have

$$\Pr(0.9 \le Z_9 \le 2.7) = \Pr(Z_9 = 1) + \Pr(Z_9 = 2) = 0.3020 + 0.3020 = 0.6040.$$

For $n = 10$, we have

$$\Pr(1 \le Z_{10} \le 3) = \Pr(Z_{10} = 1) + \Pr(Z_{10} = 2) + \Pr(Z_{10} = 3) = 0.2684 + 0.3020 + 0.2013 = 0.7717.$$

Hence, $n = 10$ is sufficient.

It should be noted that although a sample size of $n = 10$ will meet the required conditions, a sample size of $n = 11$ will not meet the required conditions. For $n = 11$, we would have

$$\Pr(1.1 \le Z_{11} \le 3.3) = \Pr(Z_{11} = 2) + \Pr(Z_{11} = 3).$$

Thus, only two terms of the binomial distribution for $n = 11$ are included, whereas three terms of binomial distribution for $n = 10$ were included.

6. It is known that when $p = 0.2$, $E(\overline{X}_n) = p = 0.2$ and $\text{Var}(\overline{X}_n) = (0.2)(0.8)/n = 0.16/n$. Therefore, $Z = (\overline{X}_n - 0.2)/(0.4/\sqrt{n})$ will have approximately a standard normal distribution. It now follows that

$$\Pr(|\overline{X}_n - p| \le 0.1) = \Pr(|Z| \le 0.25\sqrt{n}) \approx 2\Phi(0.25\sqrt{n}) - 1.$$

Therefore, this value will be at least 0.95 if and only if $\Phi(0.25\sqrt{n}) \ge 0.975$ or, equivalently, if and only if $0.25\sqrt{n} \ge 1.96$. This final relation is satisfied if and only if $n \ge 61.5$. Therefore, the sample size must be $n \ge 62$.

7. It follows from the results given in the solution to Exercise 6 that, when $p = 0.2$,

$$E_p(|\overline{X}_n - p|^2) = \text{Var}(\overline{X}_n) = \frac{0.16}{n},$$

and $0.16/n \le 0.01$ if and only if $n \ge 16$.

8. For an arbitrary value of p,

$$E_p(|\overline{X}_n - p|^2) = \text{Var}(\overline{X}_n) = \frac{p(1-p)}{n}.$$

This variance will be a maximum when $p = 1/2$, at which point its value is $1/(4n)$. Therefore, this variance will be not greater than 0.01 for all values of $p (0 \le p \le 1)$ if and only if $1/(4n) \le 0.01$ or, equivalently, if and only if $n \ge 25$.

9. The M.L.E. is $\hat{\theta} = n/T$, where T was shown to have the gamma distribution with parameters n and θ. Let $G(\cdot)$ denote the c.d.f. of the sampling distribution of T. Let $H(\cdot)$ be the c.d.f. of the sampling distribution of $\hat{\theta}$. Then $H(t) = 0$ for $t \le 0$, and for $t > 0$,

$$H(t) = \Pr(\hat{\theta} \le t) = \Pr\left(\frac{n}{T} \le t\right) = \Pr\left(T \ge \frac{n}{t}\right) = 1 - G\left(\frac{n}{t}\right).$$

8.2 The Chi-Square Distributions

Commentary

If one is using the software R, then the functions `dchisq`, `pchisq`, and `qchisq` give the p.d.f., the c.d.f., and the quantile function of χ^2 distributions. The syntax is that the first argument is the argument of the function, and the second is the degrees of freedom. The function `rchisq` gives a random sample of χ^2 random variables. The first argument is how many you want, and the second is the degrees of freedom. All of the solutions that require the calculation of χ^2 probabilites or quantiles can be done using these functions instead of tables.

Solutions to Exercises

1. The distribution of $20T/0.09$ is the χ^2 distribution with 20 degrees of freedom. We can write $\Pr(T \leq c) = \Pr(20T/0.09 \leq 20c/0.09)$. In order for this probability to be 0.9, we need $20c/0.09$ to equal the 0.9 quantile of the χ^2 distribution with 20 degrees of freedom. That quantile is 28.41. Set $28.41 = 20c/0.09$ and solve for $c = 0.1278$.

2. The mode will be the value of x at which the p.d.f. $f(x)$ is a maximum or, equivalently, the value of x at which $\log f(x)$ is a maximum. We have

$$\log f(x) = (const.) + \left(\frac{m}{2} - 1\right) \log x - \frac{x}{2}.$$

   If $m = 1$, this function is strictly decreasing and increases without bound as $x \to 0$. If $m = 2$, this function is strictly decreasing and attains its maximum value when $x = 0$. If $m \geq 3$, the value of x at which the maximum is attained can be found by setting the derivative with respect to x equal to 0. In this way it is found that $x = m - 2$.

3. The median of each distribution is found from the table of the χ^2 distribution given at the end of the book.

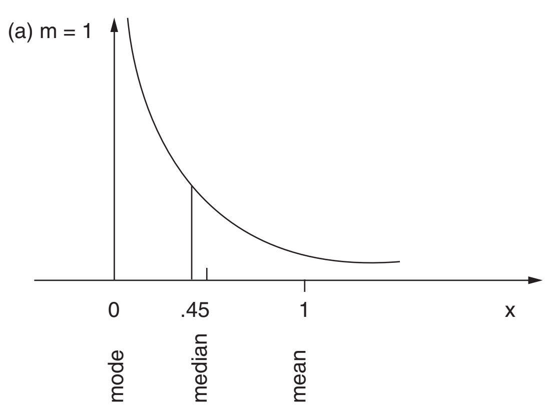

Figure S.8.1: First figure for Exercise 3 of Sec. 8.2.

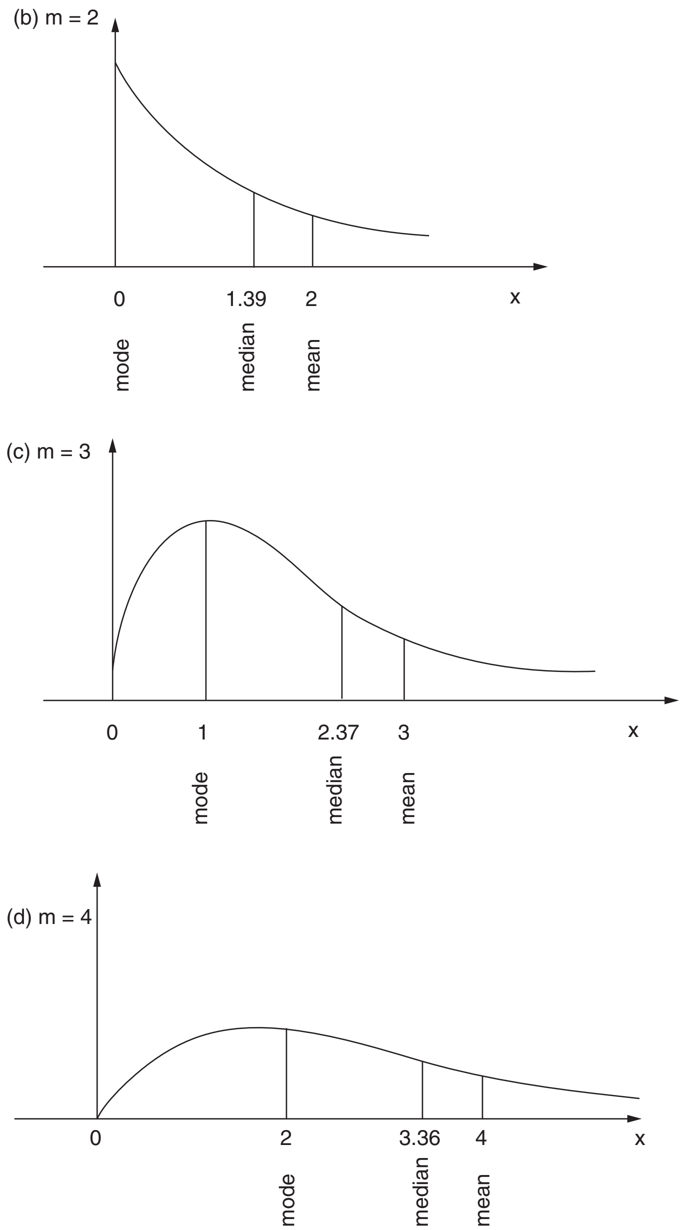

Figure S.8.2: Second figure for Exercise 3 of Sec. 8.2.

4. Let r denote the radius of the circle. The point (X, Y) will lie inside the circle if and only if $X^2 + Y^2 < r^2$. Also, $X^2 + Y^2$ has a χ^2 distribution with two degrees of freedom. It is found from the table at the end of the book that $\Pr(X^2 + Y^2 \leq 9.210) = 0.99$. Therefore, we must have $r^2 \geq 9.210$.

5. We must determine $\Pr(X^2 + Y^2 + Z^2 \leq 1)$. Since $X^2 + Y^2 + Z^2$ has the χ^2 distribution with three degrees of freedom, it is found from the table at the end of the book that the required probability is slightly less than 0.20.

6. We must determine the probability that, at time 2, $X^2 + Y^2 + Z^2 \leq 16\sigma^2$. At time 2, each of the independent variables X, Y, and Z will have a normal distribution with mean 0 and variance $2\sigma^2$. Therefore, each of the variables $X/\sqrt{2}\sigma$, $Y/\sqrt{2}\sigma$, and $Z/\sqrt{2}\sigma$ will have a standard normal distribution. Hence, $V = (X^2 + Y^2 + Z^2)/(2\sigma^2)$ will have a χ^2 distribution with three degrees of freedom. It now follows that

$$\Pr(X^2 + Y^2 + Z^2 \leq 16\sigma^2) = \Pr(V < 8).$$

It can now be found from the table at the end of the book that this probability is slightly greater than 0.95.

7. By the probability integral transformation, we know that $T_i = F_i(X_i)$ has a uniform distribution on the interval $[0, 1]$. Now let $Z_i = -\log T_i$. We shall determine the p.d.f. g of Z_i. The p.d.f. of T_i is

$$f(t) = \begin{cases} 1 & \text{for } 0 < t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since $T_i = \exp(-Z_i)$, we have $dt/dz = -\exp(-z)$. Therefore, for $z > 0$,

$$g(z) = f(\exp(-z))\left|\frac{dt}{dz}\right| = \exp(-z).$$

Thus, it is now seen that Z_i has the exponential distribution with parameter $\beta = 1$ or, in other words, the gamma distribution with parameters $\alpha = 1$ and $\beta = 1$. Therefore, by Exercise 1 of Sec. 5.7, $2Z_i$ has the gamma distribution with parameters $\alpha = 1$ and $\beta = 1/2$. Finally, by Theorem 5.7.7 $\sum_{i=1}^{n} 2Z_i$ will have the gamma distribution with parameters $\alpha = n$ and $\beta = 1/2$ or, equivalently, the χ^2 distribution with 2n degrees of freedom.

8. It was shown in Sec. 3.9 that the p.d.f. of W is as follows, for $0 < w < 1$:

$$h_1(w) = n(n-1)w^{n-2}(1-w).$$

Let $X = 2n(1 - W)$. Then $W = 1 - X/(2n)$ and $dw/dx = -1/(2n)$. Therefore, the p.d.f. $g_n(x)$ is as follows, for $0 < x < 2n$:

$$\begin{aligned} g_n(x) &= h_1\left(1 - \frac{x}{2n}\right)\left|\frac{dw}{dx}\right| = n(n-1)\left(1 - \frac{x}{2n}\right)^{n-2}\left(\frac{x}{2n}\right)\left(\frac{1}{2n}\right) \\ &= \left(\frac{1}{4}\right)\left(\frac{n-1}{n}\right)x\left(1 - \frac{x}{2n}\right)^{-2}\left(1 - \frac{x}{2n}\right)^{n}. \end{aligned}$$

Now, as $n \to \infty$,

$$\frac{n-1}{n} \to 1 \text{ and } \left(1 - \frac{x}{2n}\right)^{-2} \to 1.$$

Also, for any real number t, $(1 + t/n)^n \to \exp(t)$. Therefore, $(1 - x/(2n))^n \to \exp(-x/2)$. Hence, for $x > 0$,

$$g_n(x) \to \frac{1}{4} x \exp(-x/2).$$

This limit is the p.d.f. of the χ^2 distribution with four degrees of freedom.

9. It is known that $\overline{X}_n$ has the normal distribution with mean μ and variance σ^2/n. Therefore, $(\overline{X}_n - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution and the square of this variable has the χ^2 distribution with one degree of freedom.

10. Each of the variables $X_1 + X_2 + X_3$ and $X_4 + X_5 + X_6$ will have the normal distribution with mean 0 and variance 3. Therefore, if each of them is divided by $\sqrt{3}$, each will have a standard normal distribution. Therefore, the square of each will have the χ^2 distribution with one degree of freedom and the sum of these two squares will have the χ^2 distribution with two degrees of freedom. In other words, $Y/3$ will have the χ^2 distribution with two degrees of freedom.

11. The simplest way to determine the mean is to calculate $E(X^{1/2})$ directly, where X has the χ^2 distribution with n degrees of freedom. Thus,

$$\begin{aligned} E(X^{1/2}) &= \int_0^\infty x^{1/2} \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} \exp(-x/2) dx = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty x^{(n-1)/2} \exp(-x/2) dx \\ &= \frac{1}{2^{n/2}\Gamma(n/2)} \cdot 2^{(n+1)/2}\Gamma[(n+1)/2] = \frac{\sqrt{2}\Gamma[(n+1)/2]}{\Gamma(n/2)}. \end{aligned}$$

12. For general σ^2,

$$\Pr(Y \leq 0.09) = \Pr\left(W \leq \frac{10 \times 0.09}{\sigma^2}\right), \qquad \text{(S.8.1)}$$

where $W = 10Y/\sigma^2$ has the χ^2 distribution with 10 degrees of freedom. The probability in (S.8.1) is at least 0.9 if and only if $0.9/\sigma^2$ is at least the 0.9 quantile of the χ^2 distribution with 10 degrees of freedom. This quantile is 15.99, so $0.9/\sigma^2 \geq 15.99$ is equivalent to $\sigma^2 \leq 0.0563$.

13. We already found that the distribution of $W = n\widehat{\sigma^2}/\sigma^2$ is the χ^2 distribution with n degrees of freedom, which is also the gamma distribution with parameters $n/2$ and $1/2$. If we multiply a gamma random variable by a constant, we change its distribution to another gamma distribution with the same first parameter and the second parameter gets divided by the constant. (See Exercise 1 in Sec. 6.3.) Since $\widehat{\sigma^2} = (\sigma^2/n)W$, we see that the distribution of $\widehat{\sigma^2}$ is the gamma distribution with parameters $n/2$ and $n/(2\sigma^2)$.

8.3 Joint Distribution of the Sample Mean and Sample Variance

Commentary

This section contains some relatively mathematical results that rely on some matrix theory. We prove the statistical independence of the sample average and sample variance. We also derive the distribution of the sample variance. If your course does not focus on the mathematical details, then you can safely cite Theorem 8.3.1 and look at the examples without going through the orthogonal matrix results. The mathematical derivations rely on a calculation involving Jacobians (Sec. 3.9) which the instructor might have skipped earlier in the course.

Solutions to Exercises

1. We found that $U = n\widehat{\sigma^2}/\sigma^2$ has the χ^2 distribution with $n-1$ degrees of freedom, which is also the gamma distribution with parameters $(n-1)/2$ and $1/2$. If we multiply a gamma random variable by a number c, we change the second parameter by dividing it by c. So, with $c = \sigma^2/n$, we find that $cU = \hat{\sigma^2}$ has the gamma distribution with parameters $(n-1)/2$ and $n/(2\sigma^2)$.

2. It can be verified that the matrices in (a), (b), and (e) are orthogonal because in each case the sum of the squares of the elements in each row is 1 and the sum of the products of the corresponding terms in any two different rows is 0. The matrix in (c) is not orthogonal because the sum of squares for the bottom row is not 1. The matrix in (d) is not orthogonal because the sum of the products for rows 1 and 2 (or any other two rows) is not 0.

3. (a) Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ a_1 & a_2 \end{bmatrix}.$$

For $\boldsymbol{A}$ to be orthogonal, we must have $a_1^2 + a_2^2 = 1$ and $\frac{1}{\sqrt{2}}a_1 + \frac{1}{\sqrt{2}}a_2 = 0$. It follows from the second equation that $a_1 = -a_2$ and, in turn, from the first equation that $a_1^2 = 1/2$. Hence, either the pair of values $a_1 = 1/\sqrt{2}$ and $a_2 = -1/\sqrt{2}$ or the pair $a_1 = -1/\sqrt{2}$ and $a_2 = 1/\sqrt{2}$ will make $\boldsymbol{A}$ orthogonal.

(b) Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

For $\boldsymbol{A}$ to be orthogonal, we must have

$$a_1^2 + a_2^2 + a_3^2 = 1$$

and

$$\frac{1}{\sqrt{3}}a_1 + \frac{1}{\sqrt{3}}a_2 + \frac{1}{\sqrt{3}}a_3 = 0.$$

Therefore, $a_3 = -a_1 - a_2$ and it follows from the first equation that

$$a_1^2 + a_2^2 + (a_1 + a_2)^2 = 2a_1^2 + 2a_2^2 + 2a_1a_2 = 1.$$

Any values of a_1 and a_2 satisfying this equation can be chosen. We shall use $a_1 = 2/\sqrt{6}$ and $a_2 = -1/\sqrt{6}$. Then $a_3 = -1/\sqrt{6}$.

Finally, we must have $b_1^2 + b_2^2 + b_3^2 = 1$ as well as

$$\frac{1}{\sqrt{3}}b_1 + \frac{1}{\sqrt{3}}b_2 + \frac{1}{\sqrt{3}}b_3 = 0$$

and

$$\frac{2}{\sqrt{6}}b_1 - \frac{1}{\sqrt{6}}b_2 - \frac{1}{\sqrt{6}}b_3 = 0.$$

This final pair of equations can be rewritten as

$$b_2 + b_3 = -b_1 \quad \text{and} \quad b_2 + b_3 = 2b_1.$$

Therefore, $b_1 = 0$ and $b_2 = -b_3$. Since we must have $b_2^2 + b_3^2 = 1$, it follows that we can use either $b_2 = 1/\sqrt{2}$ and $b_3 = -1/\sqrt{2}$ or $b_2 = -1/\sqrt{2}$ and $b_3 = 1/\sqrt{2}$. Thus, one orthogonal matrix is

$$\boldsymbol{A} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

4. The 3×3 matrix of coefficients of this transformation is

$$A = \begin{bmatrix} 0.8 & 0.6 & 0 \\ (0.3)\sqrt{2} & -(0.4)\sqrt{2} & -(0.5)\sqrt{2} \\ (0.3)\sqrt{2} & -(0.4)\sqrt{2} & (0.5)\sqrt{2} \end{bmatrix}.$$

Since the matrix $\boldsymbol{A}$ is orthogonal, it follows from Theorem 8.3.4 that Y_1, Y_2, and Y_3 are independent and each has a standard normal distribution.

5. Let $Z_i = (X_i - \mu)/\sigma$ for $i = 1, 2$. Then Z_1 and Z_2 are independent and each has a standard normal distribution. Next, let $Y_1 = (Z_1 + Z_2)/\sqrt{2}$ and $Y_2 = (Z_1 - Z_2)/\sqrt{2}$. Then the 2×2 matrix of coefficients of this transformation is

$$\boldsymbol{A} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}.$$

Since the matrix $\boldsymbol{A}$ is orthogonal, it follows from Theorem 8.3.4 that Y_1 and Y_2 are also independent and each has a standard normal distribution. Finally, let $W_1 = X_1 + X_2$ and $W_2 = X_1 - X_2$. Then $W_1 = \sqrt{2}\sigma Y_1 + 2\mu$ and $W_2 = \sqrt{2}\sigma Y_2$. Since Y_1 and Y_2 are independent, it now follows from Exercise 15 of Sec. 3.9 that W_1 and W_2 are also independent.

6. (a) Since $(X_i - \mu)/\sigma$ has a standard normal distribution for $i = 1, \ldots, n$, then $W = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$ has the χ^2 distribution with n degrees of freedom. The required probability can be rewritten as follows:

$$\Pr\left(\frac{n}{2} \le W \le 2n\right).$$

Thus, when $n = 16$, we must evaluate $\Pr(8 \le W \le 32) = \Pr(W \le 32) - \Pr(W \le 8)$, Where W has the χ^2 distribution with 16 degrees of freedom. It is found from the table at the end of the book that $\Pr(W \le 32) = 0.99$ and $\Pr(W \le 8) = 0.05$.

(b) By Theorem 8.3.1, $V = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2}$ has the χ^2 distribution with $n-1$ degrees of freedom. The required probability can be rewritten as follows:

$$\Pr\left(\frac{n}{2} \leq V \leq 2n\right).$$

Thus, when $n=16$, we must evaluate $\Pr(8 \leq V \leq 32) = \Pr(V \leq 32) - \Pr(V \leq 8)$, Where V has the χ^2 distribution with 15 degrees of freedom. It is found from the table that $\Pr(V \leq 32) = 0.993$ and $\Pr(V \leq 8) = 0.079$.

7. (a) The random variable $V = n\hat{\sigma}^2/\sigma^2$ has a χ^2 distribution with $n-1$ degrees of freedom. The required probability can be written in the form $\Pr(V \leq 1.5n) \geq 0.95$. By trial and error, it is found that for $n = 20$, V has 19 degrees of freedom and $\Pr(V \leq 30) < 0.95$. However, for $n = 21$, V has 20 degrees of freedom and $\Pr(V \leq 31.5) > 0.95$.

   (b) The required probability can be written in the form

$$\Pr\left(\frac{n}{2} \leq V \leq \frac{3n}{2}\right) = \Pr\left(V \leq \frac{3n}{2}\right) - \Pr\left(V \leq \frac{n}{2}\right),$$

   where V again has the χ^2 distribution with $n-1$ degrees of freedom. By trial and error, it is found that for $n = 12$, V has 11 degrees of freedom and

$$\Pr(V \leq 18) - \Pr(V \leq 6) = 0.915 - 0.130 < 0.8.$$

   However, for $n = 13$, V has 12 degrees of freedom and

$$\Pr(V \leq 19.5) - \Pr(V \leq 6.5) = 0.919 - 0.113 > 0.8.$$

8. If X has the χ^2 distribution with 200 degrees of freedom, then it follows from Theorem 8.2.2 that X can be represented as the sum of 200 independent and identically distributed random variables, each of which has a χ^2 distribution with one degree of freedom. Since $E(X) = 200$ and $\text{Var}(X) = 400$, it follows from the central limit theorem that $Z = (X - 200)/20$ will have approximately a standard normal distribution. Therefore,

$$\Pr(160 < X < 240) = \Pr(-2 < Z < 2) \approx 2\Phi(2) - 1 = 0.9546.$$

9. The sample mean and the sample variance are independent. Therefore, the information that the sample variance is closer to σ^2 in one sample than it is in the other sample provides no information about which of the two sample means will be closer to μ. In other words, in either sample, the conditional distribution of $\overline{X}_n$, given the observed value of the sample variance, is still the normal distribution with mean μ and variance σ^2/n.

8.4 The t Distributions

Commentary

In this section, we derive the p.d.f. of the t distribution. That portion of the section (entitled "Derivation of the p.d.f.") can be skipped by instructors who do not wish to focus on mathematical details. Indeed, the derivation involves the use of Jacobians (Sec. 3.9) that the instructor might have skipped earlier in the course.

If one is using the software R, then the functions `dt`, `pt`, and `qt` give the p.d.f., the c.d.f., and the quantile function of t distributions. The syntax is that the first argument is the argument of the function, and the second is the degrees of freedom. The function `rt` gives a random sample of t random variables. The first argument is how many you want, and the second is the degrees of freedom. All of the solutions that require the calculation of t probabilites or quantiles can be done using these functions instead of tables.

Solutions to Exercises

1. $E(X^2) = c\int_{-\infty}^{\infty} x^2\left(1+\frac{x^2}{n}\right)^{-(n+1)/2} dx = 2c\int_0^{\infty} x^2\left(1+\frac{x^2}{n}\right)^{-(n+1)/2} dx,$

   where $c = \frac{\Gamma[(n+1)/2]}{(n\pi)^{1/2}\Gamma(n/2)}$. If y is defined as in the hint for this exercise, then $x = \left(\frac{ny}{1-y}\right)^{1/2}$ and $\frac{dx}{dy} = \frac{\sqrt{n}}{2}y^{-1/2}(1-y)^{-3/2}$. Therefore,

$$\begin{aligned} E(X^2) &= \sqrt{n}(const.)\int_0^1 \frac{ny}{1-y}\left(1+\frac{y}{1-y}\right)^{-(n+1)/2} y^{-1/2}(1-y)^{-3/2}dy \\ &= n^{3/2}(const.)\int_0^1 y^{1/2}(1-y)^{(n-4)/2}dy \\ &= n^{3/2}(const.)\frac{\Gamma(3/2)\Gamma[(n-2)/2]}{\Gamma[(n+1)/2]} = n\pi^{-1/2}\Gamma\left(\frac{3}{2}\right)\cdot\frac{\Gamma[(n-2)/2]}{\Gamma(n/2)} \\ &= n\pi^{-1/2}\left(\frac{1}{2}\sqrt{\pi}\right)\frac{1}{[(n-2)/2]} = \frac{n}{n-2}. \end{aligned}$$

   Since $E(X) = 0$, it now follows that $\text{Var}(X) = n/(n-2)$.

2. Since $\hat{\mu} = \overline{X}_n$ and $\hat{\sigma}^2 = S_n^2/n$, it follows from the definition of U in Eq. (8.4.4) that

$$\Pr(\hat{\mu} > \mu + k\hat{\sigma}) = \Pr\left(\frac{\overline{X}_n - \mu}{\hat{\sigma}} > k\right) = \Pr[U > k(n-1)^{1/2}].$$

   Since U has the t distribution with $n-1$ degrees of freedom and $n = 17$, we must choose k such that $\Pr(U > 4k) = 0.95$. It is found from a table of the t distribution with 16 degrees of freedom that $\Pr(U < 1.746) = 0.95$. Hence, by symmetry, $\Pr(U > -1.746) = 0.95$. It now follows that $4k = -1.746$ and $k = -0.4365$.

3. $X_1 + X_2$ has the normal distribution with mean 0 and variance 2. Therefore, $Y = (X_1+X_2)/\sqrt{2}$ has a standard normal distribution. Also, $Z = X_3^2 + X_4^2 + X_5^2$ has the χ^2 distribution with 3 degrees of freedom, and Y and Z are independent. Therefore, $U = \frac{Y}{(Z/3)^{1/2}}$ has the t distribution with 3 degrees of freedom. Thus, if we choose $c = \sqrt{3/2}$, the given random variable-will be equal to U.

4. Let $y = x/2$. Then

$$\begin{aligned} \int_{-\infty}^{2.5} \frac{dx}{(12+x^2)^2} &= \frac{1}{144}\int_{-\infty}^{2.5}\left(1+\frac{x^2}{12}\right)^{-2} dx \\ &= \frac{1}{72}\int_{-\infty}^{1.25}\left(1+\frac{y^2}{3}\right)^{-2} dy \\ &= \frac{\sqrt{3\pi}\Gamma(3/2)}{\Gamma(2)}\left(\frac{1}{72}\right)\int_{-\infty}^{1.25} g_3(y)dy, \end{aligned}$$

   where $g_3(y)$ is the p.d.f. of the t distribution with 3 degrees of freedom. It is found from the table of this distribution that the value of the integral is 0.85. Hence, the desired value is

$$\frac{\sqrt{3\pi}\Gamma\left(\frac{3}{2}\right)}{\Gamma(2)}\left(\frac{1}{72}\right)(0.85) = \frac{\sqrt{3\pi}\left(\frac{1}{2}\sqrt{\pi}\right)(0.85)}{72} = \frac{\sqrt{3}(0.85)\pi}{144}.$$

5. Let $\overline{X}_2 = (X_1 + X_2)/2$ and $S_2^2 = \sum_{i=1}^{2}(X_i - \overline{X}_2)^2$. Then

$$W = \frac{(X_1 + X_2)^2}{(X_1 - X_2)^2} = \frac{2\overline{X}_2^2}{S_2^2}.$$

It follows from Eq. (8.4.4) that $U = \sqrt{2}\,\overline{X}_2/\sqrt{S_2^2}$ has the t distribution with one degree of freedom. Since $W = U^2$, we have

$$\Pr(W < 4) = \Pr(-2 < U < 2) = 2\Pr(U < 2) - 1.$$

It can be found from a table of the t distribution with one degree of freedom that $\Pr(U < 2)$ is just slightly greater than 0.85. Hence, $\Pr(W < 4) = 0.70$.

6. The distribution of $U = (20)^{1/2}(\overline{X}_{20} - \mu)/\sigma'$ is a t distribution with 19 degrees of freedom. Let v be the 0.95 quantile of this t distribution, namely 1.729. Then

$$0.95 = \Pr(U \le 1.729) = \Pr(\overline{X}_{20} \le \mu + 1.729/(20)^{1/2}\sigma').$$

It follows that we want $c = 1.729/(20)^{1/2} = 0.3866$.

7. According to Theorem 5.7.4,

$$\begin{aligned}
\lim_{m\to\infty} \frac{(2\pi)^{1/2}(m+1/2)^m \exp(-m-1/2)}{\Gamma(m+1/2)} &= 1, \\
\lim_{m\to\infty} \frac{(2\pi)^{1/2}(m)^{m-1/2} \exp(-m)}{\Gamma(m)} &= 1.
\end{aligned}$$

Taking the ratio of the above and dividing by $m^{1/2}$, we get

$$\begin{aligned}
\lim_{m\to\infty} \frac{\Gamma(m+1/2)}{\Gamma(m)m^{1/2}} &= \lim_{m\to\infty} \frac{(2\pi)^{1/2}(m+1/2)^m \exp(-m-1/2)}{(2\pi)^{1/2}(m)^{m-1/2} \exp(-m)m^{1/2}} \\
&= \lim_{m\to\infty} \left(\frac{m+1/2}{m}\right)^m \exp(-1/2) \\
&= 1,
\end{aligned}$$

where the last equality follows from Theorem 5.3.3 applied to $(1 + 1/(2m))^m$.

8. Let f be the p.d.f. of X and let g be the p.d.f. of Y. Define

$$h(c) = \Pr(-c < X < c) - \Pr(-c < Y < c) = \int_{-c}^{c} [f(x) - g(x)]dx. \tag{S.8.2}$$

Suppose that c_0 can be chosen so that $f(x) > g(x)$ for all $-c_0 < x < c_0$ and $f(x) < g(x)$ for all $|x| > c_0$. It should now be clear that $h(c_0) = \max_c h(c)$. To prove this, first let $c > c_0$. Then

$$h(c) = h(c_0) + \int_{-c}^{-c_0} [f(x) - g(x)]dx + \int_{c_0}^{c} [f(x) - g(x)]dx.$$

Since $f(x) - g(x) < 0$ for all x in these last two integrals, $h(c) < h(c_0)$. Similarly, if $0 < c < c_0$,

$$h(c) = h(c_0) - \int_{-c_0}^{c} [f(x) - g(x)]dx - \int_{c}^{c_0} [f(x) - g(x)]dx.$$

Since $f(x) - g(x) > 0$ for all x in these last two integrals, $h(c) < h(c_0)$. Finally, notice that the standard normal p.d.f. is greater than the t p.d.f. with five degrees of freedom for all $-c < x < c$ if $c = 1.63$ and the normal p.d.f. is smaller that the t p.d.f. if $|x| > 1.63$.

8.5 Confidence Intervals

Commentary

This section ends with an extended discussion of shortcomings of confidence intervals. The first paragraph on interpretation is fairly straightforward. Students at this level should be able to understand what the confidence statement is and is not saying. The long Example 8.5.11 illustrates how additional information that is available can be ignored in the confidence statement. Instructors should gauge the mathematical abilities of their students before discussing this example in detail. Although there is nothing more complicated than what has appeared earlier in the text, it does make use of multivariable calculus and some subtle reasoning.

Many instructors will recognize the statistic Z in Example 8.5.11 as an ancillary. In many examples, conditioning on an ancillary is one way of making confidence levels (and significance levels) more representative of the amount of information available. The concept of ancillarity is beyond the scope of this text, and it is not pursued in the example. The example merely raises the issue that available information like Z is not necessarily taken into account in reporting a confidence coefficient. This makes the connection between the statistical meaning of confidence and the colloquial meaning more tenuous.

If one is using the software R, remember that `qnorm` and `qt` compute quantiles of normal and t distributions. These quantiles are ubiquitous in the construction of confidence intervals.

Solutions to Exercises

1. We need to show that

$$\Pr\left[\overline{X}_n - \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma}{n^{1/2}} < \mu < \overline{X}_n + \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma}{n^{1/2}}\right] = \gamma. \tag{S.8.3}$$

By subtracting $\overline{X}_n$ from all three sides of the above inequalities and then dividing all three sides by $\sigma/n^{1/2} > 0$, we can rewrite the probability in (S.8.3) as

$$\Pr\left[-\Phi^{-1}\left(\frac{1+\gamma}{2}\right) < \frac{\mu - \overline{X}_n}{\sigma/n^{1/2}} < \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right].$$

The random variable $(\mu - \overline{X}_n)/(\sigma/n^{1/2})$ has a standard normal distribution no matter what μ and σ^2 are. And the probability that a standard normal random variable is between $-\Phi^{-1}([1+\gamma]/2)$ and $\Phi^{-1}([1+\gamma]/2)$ is $(1+\gamma)/2 - [1 - (1+\gamma)/2] = \gamma$.

2. In this exercise, $\overline{X}_n = 3.0625$, $\sigma' = \left[\sum_{i=1}^{n}(X_i - \overline{X}_n)^2/(n-1)\right]^{1/2} = 0.5125$ and $\sigma'/n^{1/2} = 0.1812$. Therefore, the shortest confidence interval for μ will have the form $3.0625 - 0.1812c < \mu < 3.0625 + 0.1812c$.

If a confidence coefficient γ is to be used, then c must satisfy the relation $\Pr(-c < U < c) = \gamma$, where U has the t distribution with $n - 1 = 7$ degrees of freedom. By symmetry,

$$\Pr(-c < U < c) = \Pr(U < c) - \Pr(U < -c) = \Pr(U < c) - [1 - \Pr(U < c)] = 2\Pr(U < c) - 1.$$

As in the text, we find that c must be the $(1 + \gamma)/2$ quantile of the t distribution with 7 degrees of freedom.

(a) Here $\gamma = 0.90$, so $(1 + \gamma)/2 = 0.95$. It is found from a table of the t distribution with 7 degrees of freedom that $c = 1.895$. Therefore, the confidence interval for μ has endpoints $3.0625 - (0.1812)(1.895) = 2.719$ and $3.0625 + (0.1812)(1.895) = 3.406$.

(b) Here $\gamma = 0.95$, $(1 + \gamma)/2 = 0.975$, and $c = 2.365$. Therefore, the endpoints of the confidence interval for μ are $3.0625 - (0.1812)(2.365) = 2.634$ and $3.0625 + (0.1813)(2.365) = 3.491$.

(c) Here $\gamma = 0.99$, $(1 + \gamma)/2 = 0.995$, and $c = 3.499$. Therefore, the endpoints of the interval are 2.428 and 3.697.

One obvious feature of this exercise, that should be emphasized, is that the larger the confidence coefficient γ, the wider the confidence interval must be.

3. The endpoints of the confidence interval are $\overline{X}_n - c\sigma'/n^{1/2}$ and $\overline{X}_n + c\sigma'/n^{1/2}$. Therefore, $L = 2\sigma'/n^{1/2}$ and $L^2 = 4c^2\sigma'^2/n$. Since

$$W = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2}$$

has the χ^2 distribution with $n - 1$ degrees of freedom, $E(W) = n - 1$. Therefore, $E(\sigma'^2) = E(\sigma^2 W/[n - 1]) = \sigma^2$. It follows that $E(L^2) = 4c^2\sigma^2/n$. As in the text, c must be the $(1 + \gamma)/2$ quantile of the t distribution with $n - 1$ degrees of freedom.

(a) Here, $(1 + \gamma)/2 = 0.975$. Therefore, from a table of the t distribution with $n - 1 = 4$ degrees of freedom it is found that $c = 2.776$. Hence, $c^2 = 7.706$ and $E(L^2) = 4(7.706)\sigma^2/5 = 6.16\sigma^2$.

(b) For the t distribution with 9 degrees of freedom, $c = 2.262$. Hence, $E(L^2) = 2.05\sigma^2$.

(c) Here, $c = 2.045$ and $E(L^2) = 0.56\sigma^2$.

It should be noted from parts (a), (b), and (c) that for a fixed value of γ, $E(L^2)$ decreases as the sample size n increases.

(d) Here, $\gamma = 0.90$, so $(1 + \gamma)/2 = 0.95$. It is found that $c = 1.895$. Hence, $E(L^2) = 4(1.895)^2\sigma^2/8 = 1.80\sigma^2$.

(e) Here, $\gamma = 0.95$, so $(1 + \gamma)/2 = 0.975$ and $c = 2.365$. Hence, $E(L^2) = 2.80\sigma^2$.

(f) Here, $\gamma = 0.99$, so $(1 + \gamma)/2 = 0.995$ and $c = 3.499$. Hence, $E(L^2) = 6.12\sigma^2$.

It should be noted from parts (d), (e), and (f) that for a fixed sample size n, $E(L^2)$ increases as γ increases.

4. Since $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ has a standard normal distribution, $\Pr\left[-1.96 < \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} < 1.96\right] = 0.95.$

This relation can be rewritten in the form

$$\Pr\left(\overline{X}_n - \frac{1.96\sigma}{\sqrt{n}} < \mu < \overline{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right) = 0.95.$$

Therefore, the interval with endpoints $\overline{X}_n - 1.96\sigma/\sqrt{n}$ and $\overline{X}_n + 1.96\sigma/\sqrt{n}$ will be a confidence interval for μ with confidence coefficient 0.95. The length of this interval will be $3.92\sigma/\sqrt{n}$. It now follows that $3.92\sigma/\sqrt{n} < 0.01\sigma$ if and only if $\sqrt{n} > 392$. This means that $n > 153664$ of $n = 153665$ or more.

5. Since $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2/\sigma^2$ has a χ^2 distribution with $n-1$ degrees of freedom, it is possible to find constants c_1 and c_2 which satisfy the relation given in the hint for this exercise. (As explained in this section, there are an infinite number of different pairs of values of c_1 and c_2 that might be used.) The relation given in the hint can be rewritten in the form

$$\Pr\left[\frac{1}{c_2}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 < \sigma^2 < \frac{1}{c_1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right] = \gamma.$$

Therefore, the interval with endpoints equal to the observed values of $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2/c_2$ and $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2/c_1$ will be a confidence interval for σ^2 with confidence coefficient γ.

6. The exponential distribution with mean μ is the same as the gamma distribution with $\alpha = 1$ and $\beta = 1/\mu$. Therefore, by Theorem 5.7.7, $\sum_{i=1}^{n} X_i$ will have the gamma distribution with parameters $\alpha = n$ and $\beta = 1/\mu$. In turn, it follows from Exercise 1 of Sec. 5.7 that $\sum_{i=1}^{n} X_i/\mu$ has the gamma distribution with parameters $\alpha = n$ and $\beta = 1$. It follows from Definition 8.2.1 that $2\sum_{i=1}^{n} X_i/\mu$ has the χ^2 distribution with $2n$ degrees of freedom. Constants c_1 and c_2 which satisfy the relation given in the hint for this exercise will then each be 1/2 times some quantile of the χ^2 distribution with $2n$ degrees of freedom. There are an infinite number of pairs of values of such quantiles, one corresponding to each pair of numbers $q_1 \ge 0$ and $q_2 \ge 0$ such that $q_2 - q_1 = \gamma$. For example, with $q_1 = (1-\gamma)/2$ and $q_2 = (1+\gamma)/2$ we can let c_i be 1/2 times the q_i quantile of the χ^2 distribution with $2n$ degrees of freedom for $i = 1, 2$. It now follows that

$$\Pr\left(\frac{1}{c_2}\sum_{i=1}^{n} X_i < \mu < \frac{1}{c_1}\sum_{i=1}^{n} X_i\right) = \gamma.$$

Therefore, the interval with endpoints equal to the observed values of $\sum_{i=1}^{n} X_i/c_2$ and $\sum_{i=1}^{n} X_i/c_1$ will be a confidence interval for μ with confidence coefficient γ.

7. The average of the $n = 20$ values is $\overline{x}_n = 156.85$, and $\sigma' = 22.64$. The appropriate t distribution quantile is $T_{19}^{-1}(0.95) = 1.729$. The endpoints of the confidence interval are then $156.85 \pm 22.64 \times 1.729/20^{1/2}$. Completing the calculation, we get the interval $(148.1, 165.6)$.

8. According to (8.5.15), $\Pr(|\overline{X}_2 - \theta| < 0.3|Z = 0.9) = 1$, because $0.3 > (1-0.9)/2$. Since $Z = 0.9$, we know that the interval between X_1 and X_2 covers a length of 0.9 in the interval $[\theta - 1/2, \theta + 1/2]$. Hence $\overline{X}_2$ has to lie between

$$\frac{\theta - 1/2 + (\theta - 1/2 + .9)}{2} = \theta - 0.05 \quad \text{and} \quad \frac{\theta - 1/2 + 0.1 + \theta + 1/2}{2} = \theta + 0.05.$$

Hence $\overline{X}_2$ must be within 0.05 of θ, hence well within 0.3 of θ.

9. (a) The interval between the smaller an larger values is $(4.7, 5.3)$.
   (b) The values of θ consistent with the observed data are those between $5.3-0.5 = 4.8$ and $4.7+0.5 = 5.2$.
   (c) The interval in part (a) contains the set of all possible θ values, hence it is larger than the set of possible θ values.
   (d) The value of Z is $5.3 - 4.7 = 0.6$.
   (e) According to (8.5.15),

$$\Pr(|\overline{X}_2 - \theta| < 0.1|Z = 0.6) = \frac{2 \times 0.1}{1 - 0.6} = 0.5.$$

10. (a) The likelihood function is

$$f(\boldsymbol{x}|\theta) = \begin{cases} 1 & \text{if } 4.8 < \theta < 5.2, \\ 0 & \text{otherwise.} \end{cases}$$

(See the solution to Exercise 9(b) to see how the numbers 4.8 and 5.2 arise.) The posterior p.d.f. of θ is proportional to this likelihood times the prior p.d.f., hence the posterior p.d.f. is

$$\begin{cases} c\exp(-0.1\theta) & \text{if } 4.8 < \theta < 5.2, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant that makes this function into a p.d.f. The constant must satisfy

$$c\int_{4.8}^{5.2} \exp(-0.1\theta)d\theta = 1.$$

Since the integral above equals $10[\exp(-0.48)-\exp(-0.52)] = 0.2426$, we must have $c = 1/0.2426 = 4.122$.

   (b) The observed value of $\overline{X}_2$ is $\overline{x}_2 = 5$. So, the posterior probability that $|\theta - \overline{x}_2| < 0.1$ is

$$\int_{4.9}^{5.1} 4.122\exp(-0.1\theta)d\theta = 41.22[\exp(-0.49) - \exp(-0.51)] = 0.5.$$

   (c) Since the interval in part (a) of Exercise 9 contains the entire set of possible θ values, the posterior probability that θ lies in that interval is 1.
   (d) The posterior p.d.f. of θ is almost constant over the interval $(4.8, 5.2)$, hence the c.d.f. will be almost linear. The function in (8.5.15) is also linear. Indeed, for $c \le 0.2$, the posterior probability of $|\theta - 5| < c$ equals

$$\begin{aligned} \int_{5-c}^{5+c} 4.122\exp(-0.1\theta)d\theta &= 41.22\exp(-0.5)[\exp(0.1c) - \exp(-0.1c)] \\ &\approx 25 \times 2 \times 0.1c = 5c. \end{aligned}$$

Since $z = 0.6$ in this example, $5c = 2c/(1-z)$, the same as (8.5.15).

11. The variance stabilizing transformation is $\alpha(x) = \arcsin(x^{1/2})$, and the approximate distribution of $\alpha(\overline{X}_n)$ is the normal distribution with mean $\alpha(p)$ and variance $1/n$. So,

$$\Pr\left(\arcsin(\overline{X}_n^{1/2}) - \Phi^{-1}([1+\gamma]/2)n^{-1/2} < \arcsin p^{1/2} < \arcsin(\overline{X}_n^{1/2}) + \Phi^{-1}([1+\gamma]/2)n^{-1/2}\right) \approx \gamma.$$

This would make the interval with endpoints

$$\arcsin(\overline{x}_n^{1/2}) \pm \Phi^{-1}([1+\gamma]/2)n^{-1/2} \qquad \text{(S.8.4)}$$

an approximate coefficient γ confidence interval for $\arcsin(p^{1/2})$. The transformation $\alpha(x)$ has an inverse $\alpha^{-1}(y) = \sin^2(y)$ for $0 \le y \le \pi/2$. If the endpoints in (S.8.4) are between 0 and $\pi/2$, then the interval with endpoints

$$\sin^2\left(\arcsin(\overline{x}_n^{1/2}) \pm \Phi^{-1}([1+\gamma]/2)n^{-1/2}\right) \tag{S.8.5}$$

will be an approximate coefficient γ confidence interval for p. If the lower endpoint in (S.8.4) is negative replace the lower endpoint in (S.8.5) by 0. If the upper endpoint in (S.8.4) is greater than $\pi/2$, replace the upper endpoint in (S.8.5) by 1. With these modifications, the interval with the endpoints in (S.8.5) becomes an approximate coefficient γ confidence interval for p.

12. For this part of the proof, we define

$$\begin{aligned} A &= r\left(G^{-1}(\gamma_2), \boldsymbol{X}\right), \\ B &= r\left(G^{-1}(\gamma_1), \boldsymbol{X}\right). \end{aligned}$$

If $r(v, \boldsymbol{x})$ is strictly decreasing in v for each $\boldsymbol{x}$, we have

$$V(\boldsymbol{X}, \theta) < c \text{ if and only if } g(\theta) > r(c, \boldsymbol{X}). \tag{S.8.6}$$

Let $c = G^{-1}(\gamma_i)$ in Eq. (S.8.6) for each of $i = 1, 2$ to obtain

$$\Pr(g(\theta) > B) = \gamma_1, \quad \Pr(g(\theta) > A) = \gamma_2. \tag{S.8.7}$$

Because V has a continuous distribution and r is strictly decreasing,

$$\Pr(A = g(\theta)) = \Pr(V(\boldsymbol{X}, \theta) = G^{-1}(\gamma_2)) = 0,$$

and similarly $\Pr(B = g(\theta)) = 0$. The two equations in (S.8.7) combine to give $\Pr(A < g(\theta) < B) = \gamma$.

8.6 Bayesian Analysis of Samples from a Normal Distribution

Commentary

Obviously, this section should only be covered by those who are treating Bayesian topics. One might find it useful to discuss the interpretation of the prior hyperparameters in terms of amount of information and prior estimates. In this sense λ_0 and $2\alpha_0$ represent amounts of prior information about the mean and variance respectively, while μ_0 and β_0/α_0 are prior estimates of the and variance respectively. The corresponding posterior estimates are then weighted averages of the prior estimates and data-based estimates with weights equal to the amounts of information. The posterior estimate of variance, namely β_1/α_1 is the weighted average of β_0/α_0 (with weight $2\alpha_0$), σ'^2 (with weight $n-1$), and $n\lambda_0(\overline{x}_n - \mu_0)^2/(\lambda_0 + n)$ (with weight 1). This last term results from the fact that the prior distribution of the mean depends on variance (precision), hence how far $\overline{x}_n$ is from μ_0 tells us something about the variance also.

If one is using the software R, the functions `qt` and `pt` respectively compute the quantile function and c.d.f. of a t distribution. These functions can replace the use of tables for some of the calculations done in this section and in the exercises.

Solutions to Exercises

1. Since X has the normal distribution with mean μ and variance $1/\tau$, we know that Y has the normal distribution with mean $a\mu + b$ and variance a^2/τ. Therefore, the precision of Y is τ/a^2.

2. This exercise is merely a restatement of Theorem 7.3.3 with θ replaced by μ, σ^2 replaced by $1/\tau$, μ replaced by μ_0, and v^2 replaced by $1/\lambda_0$. The precision of the posterior distribution is the reciprocal of the variance of the posterior distribution given in that theorem.

3. The joint p.d.f. $f_n(\boldsymbol{x}|\tau)$ of $X_1, \ldots, X_n$ is given shortly after Definition 8.6.1 in the text, and the prior p.d.f. $\xi(\tau)$ is proportional to the expression $\xi_2(\tau)$ in the proof of Theorem 8.6.1. Therefore, the posterior p.d.f. of τ satisfies the following relation:

$$\begin{aligned}\xi(\tau \mid \boldsymbol{x}) &\propto f_n(\boldsymbol{x}|\, \tau)\xi(\tau) \propto \tau^{n/2}\exp\left\{\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]\tau\right\}\tau^{\alpha_0-1}\exp(-\beta_0\tau)\\ &= \tau^{\alpha_0+(n/2)-1}\exp\left\{-\left[\beta_0+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]\tau\right\}.\end{aligned}$$

It can now be seen that this posterior p.d.f. is, except for a constant factor, the p.d.f. of the gamma distribution specified in the exercise.

4. The posterior distribution of τ, after using the usual improper prior, is the gamma distribution with parameters $(n-1)/2$ and $s_n^2/2$. Now, V is a constant $(n-1)\sigma'^2$ times τ, so V has the gamma distribution with parameters $(n-1)/2$ and $(s_n^2/2)/[(n-1)\sigma'^2] = 1/2$. This last gamma distribution is also known as the χ^2 distribution with $n-1$ degrees of freedom.

5. Since $E(\tau) = \alpha_0/\beta_0 = 1/2$ and $\text{Var}(\tau) = \alpha_0/\beta_0^2 = 1/3$, then $\alpha_0 = 2$ and $\beta_0 = 4$. Also, $\mu_0 = E(\mu) = -5$. Finally, $\text{Var}(\mu) = \beta_0/[\lambda_0(\alpha_0 - 1)] = 1$. Therefore, $\lambda_0 = 4$.

6. Since $E(\tau) = \alpha_0/\beta_0 = 1/2$ and $\text{Var}(\tau) = \alpha_0/\beta_0^2 = 1/4$, then $\alpha_0 = 1$ and $\beta_0 = 2$. But $\text{Var}(\mu)$ is finite only if $\alpha_0 > -1$.

7. Since $E(\tau) = \alpha_0/\beta_0 = 1$ and $\text{Var}(\tau) = \alpha_0/\beta_0^2 = 4$, then $\alpha_0 = \beta_0 = 1/4$. But $E(\mu)$ exists only if $\alpha_0 > 1/2$.

8. It follows from Theorem 8.6.2 that the random variable $U = (\mu - 4)/4$ has the t distribution with $2\alpha_0 = 2$ degrees of freedom.

   (a) $\Pr(\mu > 0) = \Pr(Y > -1) = \Pr(Y < 1) = 0.79$.

   (b)

$$\begin{aligned}\Pr(0.736 < \mu < 15.680) &= \Pr(-0.816 < Y < 2.920)\\ &= \Pr(Y < 2.920) - \Pr(Y < -0.816)\\ &= \Pr(Y < 2.920) - [1 - \Pr(Y < 0.816)]\\ &= 0.95 - (1 - 0.75) = 0.70.\end{aligned}$$

9. (a) The posterior hyperparameters are computed in the example. The degrees of freedom are $2\alpha_1 = 22$, so the quantile from the t distribution is $T_{22}^{-1}([1 + .9]/2) = 1.717$, and the interval is

$$\mu_1 \pm 1.717\left(\frac{\beta_1}{\lambda_1\alpha_1}\right)^{1/2} = 183.95 \pm 1.717\left(\frac{50925.37}{20 \times 11}\right)^{1/2} = (157.83, 210.07).$$

(b) This interval has endpoints $182.17 \pm (88678.5/[17 \times 18])^{1/2} T_{17}^{-1}(0.95)$. With $T_{17}^{-1}(0.95) = 1.740$, we get the interval $(152.55, 211.79)$.

10. Since $E(\tau) = \alpha_0/\beta_0 = 2$ and

$$\text{Var}(\tau) = \frac{\alpha_0}{\beta_0^2} = E(\tau^2) - [E(\tau)]^2 = 1,$$

then $\alpha_0 = 4$ and $\beta_0 = 2$. Also $\mu_0 = E(\mu) = 0$. Therefore, by Eq. (14), $Y = (2\lambda_0)^{\frac{1}{2}}\mu$ has a t distribution with $2\alpha_0 = 8$ degrees of freedom. It is found from a table of the t distribution that $\Pr(|Y| < 0.706) = 0.5$. Therefore, $\Pr\left(|\mu| < \dfrac{0.706}{(2\lambda_0)^{1/2}}\right) = 0.5$. It now follows from the condition given in the exercise that $\dfrac{0.706}{(2\lambda_0)^{1/2}} = 1.412$. Hence, $\lambda_0 = 1/8$.

11. It follows from Theorem 8.6.1 that $\mu_1 = 80/81$, $\lambda_1 = 81/8$, $\alpha_1 = 9$, and $\beta_1 = 491/81$. Therefore, if Eq. (8.6.9) is applied to this posterior distribution, it is seen that the random variable $U = (3.877)(\mu - 0.988)$ has the t distribution with 18 degrees of freedom. Therefore, it is found from a table $\Pr(-2.101 < Y < 2.101) = 0.95$. An equivalent statement is $\Pr(0.446 < \mu < 1.530) = 0.95$. This interval will be the shortest one having probability 0.95 because the center of the interval is μ_1, the point where the p.d.f. of μ is a maximum. Since the p.d.f. of μ decreases as we move away from μ_1 in either direction, it follows that an interval having given length will have the maximum probability when it is centered at μ_1.

12. Since $E(\tau) = \alpha_0/\beta_0 = 1$ and $\text{Var}(\tau) = \alpha_0/\beta_0^2 = 1/3$, it follows that $\alpha_0 = \beta_0 = 3$. Also, since the distribution of μ is symmetric with respect to μ_0 and we are given that $\Pr(\mu > 3) = 0.5$, then $\mu_0 = 3$. Now, by Theorem 8.6.2, $U = \lambda_0^{1/2}(\mu - 3)$ has the t distribution with $2\alpha_0 = 6$ degrees of freedom. It is found from a table that $\Pr(Y < 1.440) = 0.90$. Therefore, $\Pr(Y > -1.440) = 0.90$ and it follows that $\Pr\left(\mu > 3 - \dfrac{1.440}{\lambda_0^{1/2}}\right) = 0.90$. It now follows from the condition given in the exercise that $3 - \dfrac{1.440}{\lambda_0^{1/2}} = 0.12$. Hence, $\lambda_0 = 1/4$.

13. It follows from Theorem 8.6.1 that $\mu_1 = 67/33$, $\lambda_1 = 33/4$, $\alpha_1 = 7$, and $\beta_1 = 367/33$. In calculating the value of β_1, we have used the relation

$$\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2.$$

If Theorem 8.6.2 is now applied to this posterior distribution, it is seen that the random variable $U = (2.279)(\mu - 2.030)$ has the t distribution with 14 degrees of freedom. Therefore, it is found from a table that $\Pr(-2.977 < Y < 2.977) = 0.99$. An equivalent statement is $\Pr(0.724 < \mu < 3.336) = 0.99$.

14. The interval should run between the values $\mu_1 \pm (\beta_1/[\lambda_1\alpha_1])^{1/2} T_{2\alpha_1}^{-1}(0.95)$. The values we need are available from the example or the table of the t distribution: $\mu_1 = 1.345$, $\beta_1 = 1.0484$, $\lambda_1 = 11$, $\alpha_1 = 5.5$, and $T_{11}^{-1}(0.95) = 1.796$. The resulting interval is $(1.109, 1.581)$. This interval is a bit wider than the confidence interval in Example 8.5.4. This is due mostly to the fact that $(\beta_1/\alpha_1)^{1/2}$ is somewhat larger than σ'.

15. (a) The posterior hyperparameters are

$$\begin{aligned}
\mu_1 &= \frac{2 \times 3.5 + 11 \times 7.2}{2 + 11} = 6.63, \\
\lambda_1 &= 2 + 11 = 13, \\
\alpha_1 &= 2 + \frac{11}{2} = 7.5, \\
\beta_1 &= 1 + \frac{1}{2}\left(20.3 + \frac{2 \times 11}{2 + 11}(7.2 - 3.5)^2\right) = 22.73.
\end{aligned}$$

(b) The interval should run between the values $\mu_1 \pm (\beta_1/[\lambda_1\alpha_1])^{1/2} T_{2\alpha_1}^{-1}(0.975)$. From the table of the t distribution in the book, we obtain $T_{15}^{-1}(0.975) = 2.131$. The interval is then $(5.601, 7.659)$.

16. (a) The average of all 30 observations is $\overline{x}_{30} = 1.442$ and $s_{30}^2 = 2.671$. Using the prior from Example 8.6.2, we obtain

$$\begin{aligned}
\mu_1 &= \frac{1 \times 1 + 30 \times 1.442}{1 + 30} = 1.428, \\
\lambda_1 &= 1 + 30 = 31, \\
\alpha_1 &= 0.5 + \frac{30}{2} = 15.5, \\
\beta_1 &= 0.5 + \frac{1}{2}\left(2.671 + \frac{1 \times 30}{1 + 30}(1.442 - 1)^2\right) = 1.930.
\end{aligned}$$

The posterior distribution of μ and τ is a joint normal-gamma distribution with the above hyperparameters.

(b) The average of the 20 new observations is $\overline{x}_{20} - 1.474$ and $s_{20}^2 = 1.645$. Using the posterior in Example 8.6.2 as the prior, we obtain the hyperparameters

$$\begin{aligned}
\mu_1 &= \frac{11 \times 1.345 + 20 \times 1.474}{11 + 20} = 1.428, \\
\lambda_1 &= 11 + 20 = 31, \\
\alpha_1 &= 5.5 + \frac{20}{2} = 15.5, \\
\beta_1 &= 1.0484 + \frac{1}{2}\left(1.645 + \frac{11 \times 20}{11 + 20}(1.474 - 1.345)^2\right) = 1.930.
\end{aligned}$$

The posterior hyperparameters are the same as those found in part (a). Indeed, one can prove that they must be the same when one updates sequentially or all at once.

17. Using just the first ten observations, we have $\overline{x}_n = 1.379$ and $s_n^2 = 0.9663$. This makes $\mu_1 = 1.379$, $\lambda_1 = 10$, $\alpha_1 = 4.5$, and $\beta_1 = 0.4831$. The posterior distribution of μ and τ is the normal-gamma distribution with these hyperparameters

18. Now, we use the hyperparameters found in Exercise 18 as prior hyperparameters and combine these with the last 20 observations. The average of the 20 new observations is $\overline{x}_{20} - 1.474$ and $s_{20}^2 = 1.645$. We then obtain

$$\begin{aligned}
\mu_1 &= \frac{10 \times 1.379 + 20 \times 1.474}{10 + 20} = 1.442, \\
\lambda_1 &= 10 + 20 = 30, \\
\alpha_1 &= 4.5 + \frac{20}{2} = 14.5, \\
\beta_1 &= 0.4831 \frac{1}{2}\left(1.645 + \frac{10 \times 20}{10 + 20}(1.474 - 1.379)^2\right) = 1.336.
\end{aligned}$$

Comparing two sets of hyperparameters is not as informative as comparing inferences. For example, a posterior probability interval will be centered at μ_1 and have half-width proportional to $(\beta_1/[\alpha_1\lambda_1])^{1/2}$. Since μ_1 is nearly the same in this case and in Exercise 16 part (b), the two intervals will be centered in about the same place. The values of $(\beta_1/[\alpha_1\lambda_1])^{1/2}$ for this exercise and for Exercise 16 part (b) are respectively 0.05542 and 0.06338. So we expect the intervals to be slightly shorter in this exercise than in Exercise 16. (However, the quantiles of the t distribution with 31 degrees of freedom are a bit larger in this exercise than the quantiles of the t distribution with 31 degrees of freedom in Exercise 16.)

19. (a) For the 20 observations given in Exercise 7 of Sec. 8.5, the data summaries are $\overline{x}_n = 156.85$ and $s_n^2 = 9740.55$. So, the posterior hyperparameters are

$$\begin{aligned}\mu_1 &= \frac{0.5 \times 150 + 20 \times 156.85}{0.5 + 20} = 156.68,\\ \lambda_1 &= 0.5 + 20 = 20.5,\\ \alpha_1 &= 1 + \frac{20}{2} = 11,\\ \beta_1 &= 4 + \frac{1}{2}\left(9740.55 + \frac{0.5 \times 20}{0.5 + 20}(156.85 - 150)^2\right) = 4885.7.\end{aligned}$$

The joint posterior of μ and τ is the normal-gamma distribution with these hyperparameters.

(b) The interval we want has endpoints $\mu_1 \pm (\beta_1/[\alpha_1\lambda_1])^{1/2}T_{2\alpha_1}^{-1}(0.95)$. The quantile we want is $T_{22}^{-1}(0.95) = 1.717$. Substituting the posterior hyperparameters gives the endpoints to be $a = 148.69$ and $b = 164.7$.

20. The data summaries in Example 7.3.10 are $n = 20$, $\overline{x}_{20} = 0.125$. Combine these with $s_{20}^2 = 2102.9$ to get the posterior hyperparameters:

$$\begin{aligned}\mu_1 &= \frac{1 \times 0 + 20 \times 0.125}{1 + 20} = 0.1190,\\ \lambda_1 &= 1 + 20 = 21,\\ \alpha_1 &= 1 + \frac{20}{2} = 11,\\ \beta_1 &= 60 + \frac{2102.9}{2} + \frac{20 \times 1 \times (0.125 - 0)^2}{2(1 + 20)} = 1111.5.\end{aligned}$$

(a) The posterior distribution of (μ, τ) is the normal-gamma distribution with the posterior hyperparameters given above.

(b) The posterior distribution of

$$\left(\frac{21 \times 11}{1111.5}\right)^{1/2}(\mu - 0.1190) = 0.4559(\mu - 0.1190) = T$$

is the t distribution with 22 degrees of freedom. So,

$$\Pr(\mu > 1|\boldsymbol{x}) = \Pr[0.4559(\mu - 0.1190) > 0.4559(1 - 0.1190)] = \Pr(T > 0.4016) = 0.3459,$$

where the final probability can be found by using statistical software or interpolating in the table of the t distribution.

8.7 Unbiased Estimators

Commentary

The subsection on limitations of unbiased estimators at the end of this section should be used selectively by instructors after gauging the ability of their students to understand examples with nonstandard structure.

Solutions to Exercises

1. (a) The variance of a Poisson random variable with mean θ is also θ. So the variance is $\sigma^2 = g(\theta) = \theta$.

   (b) The M.L.E. of $g(\theta) = \theta$ was found in Exercise 5 of Sec. 7.5, and it equals $\overline{X}_n$. The mean of $\overline{X}_n$ is the same as the mean of each X_i, namely θ, hence the M.L.E. is unbiased.

2. Let $E(X^k) = \beta_k$. Then

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\cdot n\beta_k = \beta_k.$$

3. By Exercise 2, $\delta_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ is an unbiased estimator of $E(X^2)$. Also, we know that $\delta_2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ is an unbiased estimator of $\mathrm{Var}(X)$. Therefore, it follows from the hint for this exercise that $\delta_1 - \delta_2$ will be an unbiased estimator of $[E(X)]^2$.

4. If X has the geometric distribution with parameter p, then it follows from Eq. (5.5.7) that $E(X) = (1-p)/p = 1/p - 1$. Therefore, $E(X+1) = 1/p$, which implies that $X+1$ is an unbiased estimator of $1/p$.

5. We shall follow the hint for this exercise. If $E[\delta(X)] = \exp(\lambda)$, then

$$\exp(\lambda) = E[\delta(X)] = \sum_{x=0}^{\infty}\delta(x)f(x \mid \lambda) = \sum_{x=0}^{\infty}\frac{\delta(x)\exp(-\lambda)\lambda^x}{x!}.$$

   Therefore,

$$\sum_{x=0}^{\infty}\frac{\delta(x)\lambda^x}{x!} = \exp(2\lambda) = \sum_{x=0}^{\infty}\frac{(2\lambda)^x}{x!} = \sum_{x=0}^{\infty}\frac{2^x\lambda^x}{x!}.$$

   Since two power series in λ can be equal only if the coefficients of λ^x are equal for $x = 0, 1, 2, \ldots$, if follows that $\delta(x) = 2^x$ for $x = 0, 1, 2, \ldots$. This argument also shows that this estimator $\delta(X)$ is the unique unbiased estimator of $\exp(\lambda)$ in this problem.

6. The M.S.E. of $\hat{\sigma}_0^2$ is given by Eq. (8.7.8) with $c = 1/n$ and it is, therefore, equal to $(2n-1)\sigma^4/n^2$. The M.S.E. of $\hat{\sigma}_1^2$ is given by Eq. (8.7.8) with $c = 1/(n-1)$ and it is, therefore, equal to $2\sigma^4/(n-1)$. Since $(2n-1)/n^2 < 2/(n-1)$ for every positive integer n, it follows that the M.S.E. of $\hat{\sigma}_0^2$ is smaller than the M.S.E. of $\hat{\sigma}_1^2$ for all values of μ and σ^2.

7. For any possible values $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$, let $y = \sum_{i=1}^{n} x_i$. Then

$$E[\delta(X_1, \ldots, X_n)] = \sum \delta(x_1, \ldots, x_n)p^y(1-p)^{n-y},$$

   where the summation extends over all possible values of $x_1, \ldots, x_n$. Since $p^y(1-p)^{n-y}$ is a polynomial in p of degree n, it follows that $E[\delta(X_1, \ldots, X_n)]$ is the sum of a finite number of terms, each of which is equal to a constant $\delta(x_1, \ldots, x_n)$ times a polynomial in p of degree n. Therefore, $E[\delta(X_1, \ldots, X_n)]$ must itself be a polynomial in p of degree n or less. The degree would actually be less than n if the sum of the terms of order p^n is 0.

8. If $E[\delta(X)] = p$, then

$$p = E[\delta(X)] = \sum_{x=0}^{\infty} \delta(x) p(1-p)^x.$$

Therefore, $\sum_{x=0}^{\infty} \delta(x)(1-p)^x = 1$. Since this relation must be satisfied for all values of $1-p$, it follows that the constant term $\delta(0)$ in the power series must be equal to 1, and the coefficient $\delta(x)$ of $(1-p)^x$ must be equal to 0 for $x = 1, 2, \ldots$.

9. If $E[\delta(X)] = \exp(-2\lambda)$, then

$$\sum_{x=0}^{\infty} \delta(x) \frac{\exp(-\lambda)\lambda^x}{x!} = \exp(-2\lambda).$$

Therefore,

$$\sum_{x=0}^{\infty} \frac{\delta(x)\lambda^x}{x!} = \exp(-\lambda) = \sum_{x=0}^{\infty} \frac{(-1)^x \lambda^x}{x!}.$$

Therefore, $\delta(X) = (-1)^x$ or, in other world, $\delta(X) = 1$ if x is even and $\delta(x) = -1$ if x is odd.

10. Let X denote the number of failures that are obtained before k successes have been obtained. Then X has the negative binomial distribution with parameters k and p, and $N = X + k$. Therefore, by Eq. (5.5.1),

$$\begin{aligned} E\left(\frac{k-1}{N-1}\right) = E\left(\frac{k-1}{X+k-1}\right) &= \sum_{x=0}^{\infty} \frac{k-1}{x+k-1} \binom{x+k-1}{x} p^k (1-p)^x \\ &= \sum_{x=0}^{\infty} \frac{(x+k-2)!}{x!(k-2)!} p^k (1-p)^x \\ &= p \sum_{x=0}^{\infty} \binom{x+k-2}{x} p^{k-1} (1-p)^x. \end{aligned}$$

But the final summation is the sum of the probabilities for a negative binomial distribution with parameters $k-1$ and p. Therefore, the value of this summation is 1, and $E([k-1]/[N-1]) = p$.

11. (a) $E(\hat{\theta}) = \alpha E(\overline{X}_m) + (1-\alpha)E(\overline{Y}_n) = \alpha\theta + (1-\alpha)\theta = \theta$. Hence, $\hat{\theta}$ is an unbiased estimator of θ for all values of α, m and n.

(b) Since the two samples are taken independently, $\overline{X}_m$ and $\overline{Y}_n$ are independent. Hence,

$$\text{Var}(\hat{\theta}) = \alpha^2 \text{Var}(\overline{X}_m) + (1-\alpha)^2 \text{Var}(\overline{Y}_n) = \alpha^2 \left(\frac{\sigma_A^2}{m}\right) + (1-\alpha)^2 \left(\frac{\sigma_B^2}{n}\right).$$

Since $\sigma_A^2 = 4\sigma_B^2$, it follows that

$$\text{Var}(\hat{\theta}) = \left[\frac{4\alpha^2}{m} + \frac{(1-\alpha)^2}{n}\right] \sigma_B^2.$$

By differentiating the coefficient of σ_B^2, it is found that $\text{Var}(\hat{\theta})$ is a minimum when $\alpha = m/(m+4n)$.

12. (a) Let X denote the value of the characteristic for a person chosen at random from the total population, and let A_i denote the event that the person belongs to stratum i $(i = 1, \ldots, k)$. Then

$$\mu = E(X) = \sum_{i=1}^{k} E(X \mid A_i) \Pr(A_i) = \sum_{i=1}^{k} \mu_i p_i.$$

Also,

$$E(\hat{\mu}) = \sum_{i=1}^{k} p_i E(\overline{X}_i) = \sum_{i=1}^{k} p_i \mu_i = \mu.$$

(b) Since the samples are taken independently of each other, the variables $\overline{X}_1, \ldots, \overline{X}_k$ are independent. Therefore,

$$\operatorname{Var}(\hat{\mu}) = \sum_{i=1}^{k} p_i^2 \operatorname{Var}(\overline{X}_i) = \sum_{i=1}^{k} \frac{p_i^2 \sigma_i^2}{n_i}.$$

Hence, the values of $n_1, \ldots, n_k$ must be chosen to minimize $v = \sum_{i=1}^{k} \frac{(p_i\sigma_i)^2}{n_i}$, subject to the constraint that $\sum_{i=1}^{k} n_i = n$. If we let $n_k = n - \sum_{i=1}^{k-1} n_i$, then

$$\frac{\partial v}{\partial n_i} = \frac{-(p_i\sigma_i)^2}{n_i^2} + \frac{(p_k\sigma_k)^2}{n_k^2} \quad for \;\; i = 1, \ldots, k-1.$$

When each of these partial derivatives is set equal to 0, it is found that $n_i/(p_i\sigma_i)$ has the same value for $i = 1, \ldots, k$. Therefore, $n_i = cp_i\sigma_i$ for some constant c. It follows that $n = \sum_{j=1}^{k} n_j = c\sum_{j=1}^{k} p_j\sigma_j$.

Hence, $c = n/\sum_{j=1}^{k} p_j\sigma_j$ and, in turn,

$$n_i = \frac{np_i\sigma_i}{\sum_{j=1}^{k} p_j\sigma_j}.$$

This analysis ignores the fact that the values of $n_1, \ldots, n_k$ must be integers. The integers $n_1, \ldots, n_k$ for which v is a minimum would presumably be near the minimizing values of $n_1, \ldots, n_k$ which have just been found.

13. (a) By Theorem 4.7.1,

$$E(\delta) = E[E(\delta|T)] = E(\delta_0).$$

Therefore, δ and δ_0 have the same expectation. Since δ is unbiased, $\mathrm{E}(\delta) = \theta$. Hence, $\mathrm{E}\,(\delta_0) = \theta$ also. In other words, δ_0 is also unbiased.

(b) Let $Y = \delta(\boldsymbol{X})$ and $X = T$ in Theorem 4.7.4. The result there implies that

$$\operatorname{Var}_\theta(\delta(\boldsymbol{X})) = \operatorname{Var}_\theta(\delta_0(\boldsymbol{X})) + E_\theta \operatorname{Var}(\delta(\boldsymbol{X})|T).$$

Since $\operatorname{Var}(\delta(\boldsymbol{X})|T) \geq 0$, so too is $E_\theta \operatorname{Var}(\delta(\boldsymbol{X})|T)$, so $\operatorname{Var}_\theta(\delta(\boldsymbol{X})) \geq \operatorname{Var}_\theta(\delta_0(\boldsymbol{X}))$.

14. For $0 < y < \theta$, the c.d.f. of Y_n is

$$F(y \mid \theta) = \Pr(Y \le y \mid \theta) = \Pr(X_1 \le y, \dots, X_n \le y \mid \theta) = \left(\frac{y}{\theta}\right)^n.$$

Therefore, for $0 < y < \theta$, the p.d.f. of Y_n is

$$f(y \mid \theta) = \frac{d}{dy} F(y \mid \theta) = \frac{n y^{n-1}}{\theta^n}.$$

It now follows that

$$E_\theta(Y_n) = \int_0^\theta y \frac{n y^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta.$$

Hence, $E_\theta([n+1]Y_n/n) = \theta$, which means that $(n+1)Y_n/n$ is an unbiased estimator of θ.

15. (a) $f(1 \mid \theta) + f(2 \mid \theta) = \theta^2[\theta + (1-\theta)] = \theta^2$,
$f(4 \mid \theta) + f(5 \mid \theta) = (1-\theta)^2[\theta + (1-\theta)] = (1-\theta)^2$,
$f(3 \mid \theta) = 2\theta(1-\theta)$.
The sum of the five probabilities on the left sides of these equations is equal to the sum the right sides, which is

$$\theta^2 + (1-\theta)^2 + 2\theta(1-\theta) = [\theta + (1-\theta)]^2 = 1.$$

(b) $E_\theta[\delta_c(X)] = \sum_{x=1}^{5} \delta_c(x) f(x \mid \theta) = 1 \cdot \theta^3 + (2-2c)\theta^2(1-\theta) + (c)2\theta(1-\theta) + (1-2c)\theta(1-\theta)^2 + 0.$
It will be found that the sum of the coefficients of θ^3 is 0, the sum of the coefficients of θ^2 is 0, the sum of the coefficients of θ is 1, and the constant term is 0. Hence, $E_\theta[\delta_c(X)] = \theta$.

(c) For every value of c,

$$\text{Var}_{\theta_0}(\delta_c) = E_{\theta_0}(\delta_c^2) - [E_{\theta_0}(\delta_c)]^2 = E_{\theta_0}(\delta_c^2) - \theta^2.$$

Hence, the value of c for which $\text{Var}_{\theta_0}(\delta_c)$ is a minimum will be the value of c for which $E_{\theta_0}(\delta_c^2)$ is a minimum. Now

$$\begin{aligned} E_{\theta_0}(\delta_c^2) &= (1)^2\theta_0^3 + (2-2c)^2\theta_0^2(1-\theta_0) + (c)^2 2\theta_0(1-\theta_0) \\ &\quad + (1-2c)^2\theta_0(1-\theta_0)^2 + 0 \\ &= 2c^2[2\theta_0^2(1-\theta_0) + \theta_0(1-\theta_0) + 2\theta_0(1-\theta_0)^2] \\ &\quad - 4c[2\theta_0^2(1-\theta_0) + \theta_0(1-\theta_0)^2] + \text{ terms not involving } c. \end{aligned}$$

After further simplification of the coefficients of c^2 and c, we obtain the relation

$$E_{\theta_0}(\delta_c^2) = 6\theta_0(1-\theta_0)c^2 + 4\theta_0(1-\theta_0^2)c + \text{ terms not involving } c.$$

By differentiating with respect to c and setting the derivative equal to 0, it is found that the value of c for which $E_{\theta_0}(\delta_c^2)$ is a minimum is $c = (1+\theta_0)/3$.

16. The unbiased estimator in Exercise 3 is

$$\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

For the observed values $X_1 = 2$ and $X_2 = -1$, we obtain the value -2 for the estimate. This is unacceptable. Because $[E(X)]^2 \ge 0$, we should demand an estimate that is also nonnegative.

8.8 Fisher Information

Commentary

Although this section is optional, it does contain the interesting theoretical result on asymptotic normality of maximum likelihood estimators. It also contains the Cramér-Rao inequality, which can be useful for finding minimum variance unbiased estimators. However, the material is really only suitable for a fairly mathematically oriented course.

Solutions to Exercises

1.

$$\begin{aligned} f(x \mid \mu) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \\ f'(x \mid \mu) &= \frac{1}{\sqrt{2\pi}\sigma}\frac{(x-\mu)}{\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} = \frac{x-\mu}{\sigma^2} f(x \mid \mu), \\ f''(x \mid \mu) &= \left[\frac{(x-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2}\right] f(x \mid \mu). \end{aligned}$$

Therefore,

$$\int_{-\infty}^{\infty} f'(x \mid \mu)dx = \frac{1}{\sigma^2}\int_{-\infty}^{\infty}(x-\mu)f(x \mid \mu)d\mu = \frac{1}{\sigma^2}E(X-\mu) = 0,$$

and

$$\int_{-\infty}^{\infty} f''(x \mid \mu)dx = \frac{E[(X-\mu)^2]}{\sigma^4} - \frac{1}{\sigma^2} = \frac{\sigma^2}{\sigma^4} - \frac{1}{\sigma^2} = 0.$$

2. The p.f. is

$$f(x|p) = p(1-p)^x, \text{ for } x = 0, 1, \ldots.$$

The logarithm of this is $\log(p) + x\log(1-p)$, and the derivative is

$$\frac{1}{p} - \frac{x}{1-p}.$$

According to Eq. (5.5.7) in the text, the variance of X is $(1-p)/p^2$, hence the Fisher information is

$$I(p) = \mathrm{Var}\left[\frac{1}{p} - \frac{X}{1-p}\right] = \frac{\mathrm{Var}(X)}{(1-p)^2} = \frac{1}{p^2(1-p)}.$$

3.

$$\begin{aligned} f(x \mid \theta) &= \frac{\exp(-\theta)\theta^x}{x!}, \\ \lambda(x \mid \theta) &= -\theta + x\log\theta - \log(x!), \\ \lambda'(x \mid \theta) &= -1 + \frac{x}{\theta}, \\ \lambda''(x \mid \theta) &= -\frac{x}{\theta^2}. \end{aligned}$$

Therefore, by Eq. (8.8.3),

$$I(\theta) = -E_\theta[\lambda''(X \mid \theta)] = \frac{E(X)}{\theta^2} = \frac{1}{\theta}.$$

4.

$$\begin{aligned}
f(x \mid \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \\
\lambda(x \mid \sigma) &= -\log \sigma - \frac{x^2}{2\sigma^2} + const. \\
\lambda'(x \mid \sigma) &= -\frac{1}{\sigma} + \frac{x^2}{\sigma^3}, \\
\lambda''(x \mid \sigma) &= \frac{1}{\sigma^2} - \frac{3x^2}{\sigma^4}.
\end{aligned}$$

Therefore,

$$I(\theta) = -E_\theta[\lambda''(X \mid \theta)] = -\frac{1}{\sigma^2} + \frac{3E(X^2)}{\sigma^4} = -\frac{1}{\sigma^2} + \frac{3}{\sigma^2} = \frac{2}{\sigma^2}.$$

5. Let $\nu = \sigma^2$. Then

$$\begin{aligned}
f(x \mid \nu) &= \frac{1}{\sqrt{2\pi\nu}} \exp\left\{-\frac{x^2}{2\nu}\right\}, \\
\lambda(x \mid \nu) &= -\frac{1}{2}\log \nu - \frac{x^2}{2\nu} + const., \\
\lambda'(x \mid \nu) &= -\frac{1}{2\nu} + \frac{x^2}{2\nu^2}, \\
\lambda''(x \mid \nu) &= \frac{1}{2\nu^2} - \frac{x^2}{\nu^3}.
\end{aligned}$$

Therefore,

$$I(\sigma^2) = I(\nu) = -E_\nu[\lambda''(X \mid \nu)] = -\frac{1}{2\nu^2} + \frac{\nu}{\nu^3} = \frac{1}{2\nu^2} = \frac{1}{2\sigma^4}.$$

6. Let $g(x \mid \mu)$ denote the p.d.f. or the p.f. of X when μ is regarded as the parameter. Then $g(x \mid \mu) = f[x \mid \psi(\mu)]$. Therefore,

$$\log g(x \mid \mu) = \log f[x \mid \psi(\mu)] = \lambda[x \mid \psi(\mu)],$$

and

$$\frac{\partial}{\partial \mu} \log g(x \mid \mu) = \lambda'[x \mid \psi(\mu)]\psi'(\mu).$$

It now follows that

$$I_1(\mu) = E_\mu\left\{\left[\frac{\partial}{\partial \mu} \log g(X \mid \mu)\right]^2\right\} = [\psi'(\mu)]^2 E_\mu(\{\lambda'[X \mid \psi(\mu)]\}^2) = [\psi'(\mu)]^2 I_0[\psi(\mu)].$$

7. We know that $E(\overline{X}_n) = p$ and $\text{Var}(\overline{X}_n) = p(1-p)/n$. It was shown in Example 8.8.2 that $I(P) = 1/[p(1-p)]$. Therefore, $\text{Var}(\overline{X}_n)$ is equal to the lower bound $1/[nI(p)]$ provided by the information inequality.

8. We know that $E(\overline{X}_n) = \mu$ and $\text{Var}\,(\overline{X}_n) = \frac{\sigma^2}{n}$. It was shown in Example 8.8.3 that $I(\mu) = 1/\sigma^2$. Therefore, $\text{Var}\,(\overline{X}_n)$ is equal to the lower bound $1/[nI(\mu)]$ provided by the information inequality.

9. We shall attack this exercise by trying to find an estimator of the form $c|X|$ that is unbiased. One approach is as follows: We know that X^2/σ^2 has the χ^2 distribution with one degree of freedom. Therefore, by Exercise 11 of Sec. 8.2, $|X|/\sigma$ has the χ distribution with one degree of freedom, and it was shown in that exercise that

$$E\left(\frac{|X|}{\sigma}\right) = \frac{\sqrt{2}\Gamma(1)}{\Gamma(1/2)} = \sqrt{\frac{2}{\pi}}.$$

Hence, $E(|X|) = \sigma\sqrt{2/\pi}$. It follows that $E(|X|\sqrt{\pi/2}) = \sigma$. Let $\delta = |X|\sqrt{\pi/2}$. Then

$$E(\delta^2) = \frac{\pi}{2}E(|X|^2) = \frac{\pi}{2}\sigma^2.$$

Hence,

$$\text{Var}\,\delta = E(\delta^2) - [E(\delta)]^2 = \frac{\pi}{2}\sigma^2 - \sigma^2 = \left(\frac{\pi}{2} - 1\right)\sigma^2.$$

Since $1/I(\sigma) = \sigma^2/2$, it follows that $\text{Var}(\delta) > 1/I(\sigma)$.

Another unbiased estimator is $\delta_1(X) = \sqrt{2\pi}\ X$ if $X \geq 0$ and $\delta_1(X) = 0$ if $X < 0$. However, it can be shown, using advanced methods, that the estimator δ found in this exercise is the *only* unbiased estimator of σ that depends on X only through $|X|$.

10. If $m(\sigma) = \log\sigma$, then $m'(\sigma) = 1/\sigma$ and $[m'(\sigma)]^2 = 1/\sigma^2$. Also, it was shown in Exercise 4 that $I(\sigma) = 2/\sigma^2$. Therefore, if T is an unbiased estimator of $\log\sigma$, it follows from the relation (8.8.14) that

$$\text{Var}(T) \geq \frac{1}{\sigma^2} \cdot \frac{\sigma^2}{2n} = \frac{1}{2n}.$$

11. If $f(x \mid \theta) = a(\theta)b(x)\exp[c(\theta)d(x)]$, then

$$\lambda(x \mid \theta) = \log a(\theta) + \log b(x) + c(\theta)d(x)$$

and

$$\lambda'(x \mid \theta) = \frac{a'(\theta)}{a(\theta)} + c'(\theta)d(x).$$

Therefore,

$$\lambda'_n(\boldsymbol{X} \mid \theta) = \sum_{i=1}^{n} \lambda'(X_i \mid \theta) = n\frac{a'(\theta)}{a(\theta)} + c'(\theta)\sum_{i=1}^{n} d(X_i).$$

If we choose

$$u(\theta) = \frac{1}{c'(\theta)} \quad \text{and} \quad v(\theta) = -\frac{na'(\theta)}{a(\theta)c'(\theta)},$$

then Eq. (8.8.14) will be satisfied with $T = \sum_{i=1}^{n} d(X_i)$. Hence, this statistic is an efficient estimator of its expectation.

12. Let $\theta = \sigma^2$ denote the unknown variance. Then

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2\theta}(x-\mu)^2\right\}.$$

This p.d.f. $f(x \mid \theta)$ has the form of an exponential family, as given in Exercise 12, with $d(x) = (x-\mu)^2$. Therefore, $T = \sum_{i=1}^{n}(X_i - \mu)^2$ will be an efficient estimator. Since $E[(X_i - \mu)^2] = \sigma^2$ for $i = 1, \ldots, n$, then $E(T) = n\sigma^2$. Also, by Exercise 17 of Sec. 5.7, $E[(X_i - \mu)^4] = 3\sigma^4$ for $i = 1, \ldots, n$. Therefore, $\text{Var}[(X_i - \mu)^2] = 3\sigma^4 - \sigma^4 = 2\sigma^4$, and it follows that $\text{Var}(T) = 2n\sigma^4$.

It should be emphasized that any linear function of T will also be an efficient estimator. In particular, T/n will be an efficient estimator of σ^2.

13. The incorrect part of the argument is at the beginning, because the information inequality cannot be applied to the uniform distribution. For each different value of θ, there is a different set of values of x for which $f(x \mid \theta) \geq 0$.

14.

$$\begin{aligned}
f(x \mid \alpha) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \\
\lambda(x \mid \alpha) &= \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1)\log x - \beta x, \\
\lambda'(x \mid \alpha) &= \log \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \log x, \\
\lambda''(x \mid \alpha) &= -\frac{\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2}{[\Gamma(\alpha)]^2}
\end{aligned}$$

Therefore,

$$I(\alpha) = \frac{\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2}{[\Gamma(\alpha)]^2}$$

The distribution of the M.L.E. of α will be approximately the normal distribution with mean α and variance $1/[nI(\alpha)]$.
It should be noted that we have determined this distribution without actually determining the M.L.E. itself.

15. We know that the M.L.E. of μ is $\hat{\mu} = \overline{x}_n$ and, from Example 8.8.3, that $I(\mu) = 1/\sigma^2$. The posterior distribution of μ will be approximately a normal distribution with mean $\hat{\mu}$ and variance $1/[nI(\hat{\mu})] = \sigma^2/n$.

16. We know that the M.L.E. of p is $\hat{p} = \overline{x}_n$ and, from Example 8.8.2, that $I(p) = 1/[p(1-p)]$. The posterior distribution of p will be approximately a normal distribution with mean $\hat{p}$ and variance $1/[nI(\hat{p})] = \overline{x}_n(1 - \overline{x}_n)/n$.

17. The derivative of the log-likelihood with respect to p is

$$\lambda'(x|p) = \frac{\partial}{\partial p}\left[\log\binom{n}{x} + x\log(p) + (n-x)\log(1-p)\right] = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}.$$

The mean of $\lambda'(X|p)$ is clearly 0, so its variance is

$$I(p) = \frac{\text{Var}(X)}{p^2(1-p)^2} = \frac{n}{p(1-p)}.$$

18. The derivative of the log-likelihood with respect to p is

$$\lambda'(x|p) = \frac{\partial}{\partial p}\left[\log\binom{r+x-1}{x} + r\log(p) + x\log(1-p)\right] = \frac{r}{p} - \frac{x}{1-p} = \frac{r-rp-xp}{p(1-p)}.$$

The mean of $\lambda'(X|p)$ is clearly 0, so its variance is

$$I(p) = \frac{p^2\,\text{Var}(X)}{p^2(1-p)^2} = \frac{r}{p^2(1-p)}.$$

8.9 Supplementary Exercises

Solutions to Exercises

1. According to Exercise 5 in Sec. 8.8, the Fisher information $I(\sigma^2)$ based on a sample of size 1 is $1/[2\sigma^4]$. According to the information inequality, the variance of an unbiased estimator of σ^2 must be at least $2\sigma^4/n$. The variance of $V = \sum_{i=1}^{n} X_i^2/n$ is $\text{Var}(X_1^2)/n$. Since X_1^2/σ^2 has a χ^2 distribution with 1 degree of freedom, its variance is 2. Hence $\text{Var}(X_1^2) = 2\sigma^4$ and $\text{Var}(V)$ equals the lower bound from the information inequality. $E(V) = E(X_1^2) = \sigma^2$, so V is unbiased.

2. The t distribution with one degree of freedom is the Cauchy distribution. Therefore, by Exercise 18 of Sec. 5.6, we can represent the random variable X in the form $X = U/V$, where U and V are independent and each has a standard normal distribution. But $1/X$ can then be represented as $1/X = V/U$. Since V/U is again the ratio of independent, standard normal variables, it follows that $1/X$ again has the Cauchy distribution.

3. It is known from Exercise 18 of Sec. 5.6 that U/V has a Cauchy distribution, which is the t distribution with one degree of freedom. Next, since $|V| = (V^2)^{1/2}$, it follows from Definition 8.4.1 that $U/|V|$ has the required t distribution. Hence, by the previous exercise in this section, $|V|/U$ will also have this t distribution. Since U and V are i.i.d., it now follows that $|U|/V$ must have the same distribution as $|V|/U$.

4. It is known from Exercise 5 of Sec. 8.3 that $X_1 + X_2$ and $X_1 - X_2$ are independent. Further, if we let

$$Y_1 = \frac{1}{\sqrt{2}\sigma}(X_1 + X_2) \quad \text{and} \quad Y_2 = \frac{1}{\sqrt{2}\sigma}(X_1 - X_2),$$

then Y_1 and Y_2 have standard normal distributions. It follows, therefore, from Exercise 18 of Sec. 5.6 that Y_1/Y_2 has a Cauchy distribution, which is the same as the t distribution with one degree of freedom. But

$$\frac{Y_1}{Y_2} = \frac{X_1 + X_2}{X_1 - X_2},$$

so the desired result has been established. This result could also have been established by a direct calculation of the required p.d.f.

5. Since X_i has the exponential distribution with parameter β, it follows that $2\beta\ X_i$ has the exponential distribution with parameter 1/2. But this exponential distribution is the χ^2 distribution with 2 degrees of freedom. Therefore, the sum of the i.i.d. random variables $2\beta X_i$ $(i = 1, \ldots, n)$ will have a χ^2 distribution with $2n$ degrees of freedom.

6. Let $\hat{\theta}_n$ be the proportion of the n observations that lie in the set A. Since each observation has probability θ of lying in A, the observations can be thought of as forming n Bernoulli trials, each with probability θ of success. Hence, $E(\hat{\theta}_n) = \theta$ and $\mathrm{Var}(\hat{\theta}_n) = \theta(1-\theta)/n$.

7. (a) $E(\alpha S_X^2 + \beta S_Y^2) = \alpha(m-1)\sigma^2 + \beta(n-1)2\sigma^2$.
Hence, this estimator will be unbiased if $\alpha(m-1) + 2\beta(n-1) = 1$.
(b) Since S_X^2 and S_Y^2 are independent,

$$\begin{aligned}\mathrm{Var}\,(\alpha S_X^2 + \beta S_Y^2) &= \alpha^2\,\mathrm{Var}\,(S_X^2) + \beta^2 var(S_Y^2)\\ &= \alpha^2[2(m-1)\sigma^4] + \beta^2[2(n-1)\cdot 4\sigma^4]\\ &= 2\sigma^4[(m-1)\alpha^2 + 4(n-1)\beta^2].\end{aligned}$$

Therefore, we must minimize

$$A = (m-1)\alpha^2 + 4(n-1)\beta^2$$

subject to the constraint $(m-1)\alpha + 2(n-1)\beta = 1$. If we solve this constraint for β in terms of α, and make this substitution for β in A, we can then minimize A over all values of α. The result is $\alpha = \dfrac{1}{m+n-2}$ and, hence, $\beta = \dfrac{1}{2(m+n-2)}$.

8. $X_{n+1} - \overline{X}_n$ has the normal distribution with mean 0 and variance $(1+1/n)\sigma^2$. Hence, the distribution of $(n/[n+1])^{1/2}(X_{n+1} - \overline{X}_n)/\sigma$ is a standard normal distribution. Also, nT_n^2/σ^2 has an independent χ^2 distribution with $n-1$ degrees of freedom. Thus, the following ratio will have the t distribution with $n-1$ degrees of freedom:

$$\frac{\left(\dfrac{n}{n+1}\right)^{1/2}(X_{n+1} - \overline{X}_n)/\sigma}{\left[\dfrac{nT_n^2}{(n-1)\sigma^2}\right]^{1/2}} = \left(\frac{n-1}{n+1}\right)^{1/2}\frac{X_{n+1} - \overline{X}_n}{T_n}.$$

It can now be seen that $k = ([n-1]/[n+1])^{1/2}$.

9. Under the given conditions, $Y/(2\sigma)$ has a standard normal distribution and S_n^2/σ^2 has an independent χ^2 distribution with $n-1$ degrees of freedom. Thus, the following random variable will have a t distribution with $n-1$ degrees of freedom:

$$\frac{Y/(2\sigma)}{\{S_n^2/[\sigma^2(n-1)]\}^{1/2}} = \frac{Y/2}{\sigma'},$$

where $\sigma' = [S_n^2/(n-1)]^{1/2}$.

10. As found in Exercise 3 of Sec. 8.5, the expected squared length of the confidence interval is $E(L^2) = 4c^2\sigma^2/n$, where c is found from the table of the t distribution with $n-1$ degrees of freedom in the back of the book under the .95 column (to give probability .90 between $-c$ and c). We must compute the value of $4c^2/n$ for various values of n and see when it is less than 1/2. For $n = 23$, it is found that $c_{22} = 1.717$ and the coefficient of σ^2 in $E(L^2)$ is $4(1.717)^2/23 = .512$. For $n = 24$, $c_{23} = 1.714$ and the coefficient of σ^2 is $4(1.714)^2/24 = .490$. Hence, $n = 24$ is the required value.

11. Let c denote the .99 quantile of the t distribution with $n-1$ degrees of freedom; i.e., $\Pr(U < c) = .99$ if U has the specified t distribution. Therefore, $\Pr\left[\dfrac{n^{1/2}(\overline{X}_n - \mu)}{\sigma'} < c\right] = .99$ or, equivalently, $\Pr\left[\mu > \overline{X}_n - \dfrac{c\sigma'}{n^{1/2}}\right] = .99$. Hence, $L = \overline{X}_n - c\sigma'/n^{1/2}$.

12. Let c denote the .01 quantile of the χ^2 distribution with $n-1$ degrees of freedom; i.e., $\Pr(V < c) = .01$ if V has the specified χ^2 distribution. Therefore,

$$\Pr\left(\frac{S_n^2}{\sigma^2} > c\right) = .99$$

or, equivalently,

$$\Pr\left(\sigma^2 < S_n^2/c\right) = .99.$$

Hence, $U = S_n^2/c$.

13. (a) The posterior distribution of θ is the normal distribution with mean μ_1 and variance ν_1^2, as given by (7.3.1) and (7.3.2). Therefore, under this distribution,

$$\Pr(\mu_1 - 1.96\nu_1 < \theta < \mu_1 + 1.96\nu_1) = .95.$$

This interval I is the shortest one that has the required probability because it is symmetrically placed around the mean μ_1 of the normal distribution.

(b) It follows from (7.3.1) that $\mu_1 \to \overline{x}_n$ as $\nu^2 \to \infty$ and from (7.3.2) that $\nu_1^2 \to \sigma^2/n$. Hence, the interval I converges to the interval

$$\overline{x}_n - \frac{1.96\sigma}{n^{1/2}} < \theta < \overline{x}_n + \frac{1.96\sigma}{n^{1/2}}.$$

It was shown in Exercise 4 of Sec. 8.5 that this interval is a confidence interval for θ with confidence coefficient .95.

14. (a) Since Y has a Poisson distribution with mean $n\theta$, it follows that

$$\begin{aligned} E(\exp(-cY)) &= \sum_{y=0}^{\infty} \frac{\exp(-cy)\exp(-n\theta)(n\theta)^y}{y!} = \exp(-n\theta)\sum_{y=0}^{\infty}\frac{(n\theta\exp(-c))^y}{y!} \\ &= \exp(-n\theta)\exp[n\theta\exp(-c)] = \exp(n\theta[\exp(-c)-1]). \end{aligned}$$

Since this expectation must be $\exp(-\theta)$, it follows that $n(\exp(-c) - 1) = -1$ or $c = \log[n/(n-1)]$.

(b) It was shown in Exercise 3 of Sec. 8.8 that $I(\theta) = 1/\theta$ in this problem. Since $m(\theta) = \exp(-\theta)$, $[m'(\theta)]^2 = \exp(-2\theta)$. Hence, from Eq. (8.8.14),

$$\text{Var}(\exp(-cY)) \geq \frac{\theta\exp(-2\theta)}{n}.$$

15. In the notation of Sec. 8.8,

$$\begin{aligned}\lambda(x \mid \theta) &= \log\theta + (\theta-1)\log x,\\ \lambda'(x \mid \theta) &= \frac{1}{\theta} + \log x,\\ \lambda''(x \mid \theta) &= -1/\theta^2.\end{aligned}$$

Hence, by Eq. (8.8.3), $I(\theta) = 1/\theta^2$ and it follows that the asymptotic distribution of

$$\frac{n^{1/2}}{\theta}(\hat{\theta}_n - \theta)$$

is standard normal.

16.

$$\begin{aligned}f(x \mid \theta) &= \theta^{-1}\exp(-x/\theta),\\ \lambda(x \mid \theta) &= -\log\theta - x/\theta,\\ \lambda'(x \mid \theta) &= -\frac{1}{\theta} + \frac{x}{\theta^2},\\ \lambda''(x \mid \theta) &= \frac{1}{\theta^2} - \frac{2x}{\theta^3},\end{aligned}$$

Therefore,

$$I(\theta) = -E_\theta[\lambda''(X|\theta)] = \frac{1}{\theta^2}.$$

17. If $m(p) = (1-p)^2$, then $m'(p) = -2(1-p)$ and $[m'(p)]^2 = 4(1-p)^2$. It was shown in Example 8.8.2 that $I(p) = 1/[p(1-p)]$. Therefore, if T is an unbiased estimator of $m(p)$, it follows from the relation (8.8.14) that

$$\text{Var}(T) \geq \frac{4(1-p)^2 p(1-p)}{n} = \frac{4p(1-p)^3}{n}.$$

18. $f(x|\beta) = \beta\exp(-\beta x)$. This p.d.f. has the form of an exponential family, as given in Exercise 11 of Sec. 8.8, with $d(x) = x$. Therefore, $T = \sum_{i=1}^{n} X_i$ will be an efficient estimator. We know that $E(X_i) = 1/\beta$ and $\text{Var}(X_i) = 1/\beta^2$. Hence, $E(T) = n/\beta$ and $\text{Var}(T) = n/\beta^2$.

Since any linear function of T will also be an efficient estimator, it follows that $\overline{X}_n = T/n$ will be an efficient estimator of $1/\beta$. As a check of this result, it can be verified directly that $\text{Var}(\overline{X}_n) = 1/[n\beta^2] = [m'(\beta)]^2/[nI(\beta)]$, where $m(\beta) = 1/\beta$ and $I(\beta)$ was obtained in Example 8.8.6.

19. It was shown in Example 8.8.6 that $I(\beta) = 1/\beta^2$. The distribution of the M.L.E. of β will be approximately the normal distribution with mean β and variance $1/[nI(\beta)]$.

20. (a) Let $\alpha(\beta) = 1/\beta$. Then $\alpha'(\beta) = -1/\beta^2$. By Exercise 19, it is known that $\hat{\beta}_n$ is approximately normal with mean β and variance β^2/n. Therefore, $1/\hat{\beta}_n$ will be approximately normal with mean $1/\beta$ and variance $[\alpha'(\beta)]^2(\beta^2/n) = 1/(n\beta^2)$. Equivalently, the asymptotic distribution of

$$(n\beta^2)^{1/2}(1/\hat{\beta}_n - 1/\beta)$$

is standard normal.

(b) Since the mean of the exponential distribution is $1/\beta$ and the variance is $1/\beta^2$, it follows directly from the central limit theorem that the asymptotic distribution of $\overline{X}_n = 1/\hat{\beta}_n$ is exactly that found in part (a).

21. (a) The distribution of Y is the Poisson distribution with mean $n\theta$. In order for $r(Y)$ to be an unbiased estimator of $1/\theta$, we need

$$\frac{1}{\theta} = E_\theta(r(Y)) = \sum_{y=0}^{\infty} r(y)\exp(-n\theta)\frac{(n\theta)^y}{y!}.$$

This equation can be rewritten as

$$\exp(n\theta) = \sum_{y=0}^{\infty} \frac{r(y)n^y}{y!}\theta^{y+1}. \tag{S.8.8}$$

The function on the left side of (S.8.8) has a unique power series representation, hence the right side of (S.8.8) must equal that power series. However, the limit as $\theta \to 0$ of the left side of (S.8.8) is 1, while the limit of the right side is 0, hence the power series on the right cannot represent the function on the left.

(b) $E(n/[Y+1]) = \sum_{y=0}^{\infty} n\exp(-n\theta)[n\theta]^y/(y+1)!$. By letting $u = y+1$ in this sum, we get $n[1 - \exp(-n\theta)]/[n\theta] = 1/\theta - \exp(-n\theta)/\theta$. So the bias is $\exp(-n\theta)/\theta$. Clearly $\exp(-n\theta)$ goes to 0 as $n \to \infty$.

(c) $n/(1+Y) = 1/(\overline{X}_n + 1/n)$. We know that $\overline{X}_n + 1/n$ has approximately the normal distribution with mean $\theta + 1/n$ and variance θ/n. We can ignore the $1/n$ added to θ in the mean since this will eventually be small relative to θ. Using the delta method, we find that $1/(\overline{X}_n + 1/n)$ has approximately the normal distribution with mean $1/\theta$ and variance $(1/\theta^2)^2\theta/n = (n\theta^3)^{-1}$.

22. (a) The p.d.f. of Y_n is

$$f(y|\theta) = \begin{cases} ny^{n-1}/\theta^n & \text{if } 0 \le y \le \theta, \\ 0 & \text{otherwise.} \end{cases}$$

This can be found using the method of Example 3.9.6. If $X = Y_n/\theta$, then the p.d.f. of X is

$$g(x|\theta) = f(x\theta|\theta)\theta = \begin{cases} nx^{n-1} & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that this does not depend on θ. The c.d.f. is then $G(x) = x^n$ for $0 < x < 1$. The quantile function is $G^{-1}(p) = p^{1/n}$.

(b) The bias of Y_n as an estimator of θ is

$$E_\theta(Y_n) - \theta = \int_0^\theta y\frac{ny^{n-1}}{\theta^n}dy - \theta = -\frac{\theta}{n+1}.$$

(c) The distribution of $Z = Y_n/\theta$ has p.d.f.

$$g(z) = \theta f(z\theta|\theta) = \begin{cases} nz^{n-1} & \text{for } 0 \le z \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $f(\cdot|\theta)$ comes from part (a). One can see that $g(z)$ does not depend on θ, hence the distribution of Z is the same for all θ.

(d) We would like to find two random variables $A(Y_n)$ and $B(Y_n)$ such that

$$\Pr(A(Y_n) \le \theta \le B(Y_n)) = \gamma, \text{ for all } \theta. \tag{S.8.9}$$

This can be arranged by using the fact that Y_n/θ has the c.d.f. $G(x) = x^n$ for all θ. This means that

$$\Pr\left(a \le \frac{Y_n}{\theta} \le b\right) = b^n - a^n,$$

for all θ. Let a and b be constants such that $b^n - a^n = \gamma$ (e.g., $b = ([1+\gamma]/2)^{1/n}$ and $a = ([1-\gamma]/2)^{1/n}$). Then set $A(Y_n) = Y_n/b$ and $B(Y_n) = Y_n/a$. It follows that (S.8.9) holds.

Chapter 9

Testing Hypotheses

9.1 Problems of Testing Hypotheses

Commentary

This section was augemented in the fourth edition. It now includes a general introduction to likelihood ratio tests and some foundational discussion of the terminology of hypothesis testing. After covering this section, one could skip directly to Sec. 9.5 and discuss the t test without using any of the material in Sec. 9.2–9.4. Indeed, unless your course is a rigorous mathematical statistics course, it might be highly advisable to skip ahead.

Solutions to Exercises

1. (a) Let δ be the test that rejects H_0 when $X \geq 1$.The power function of δ is

$$\pi(\beta|\delta) = \Pr(X \geq 1|\beta) = \exp(-\beta),$$

for $\beta > 0$.

(b) The size of the test δ is $\sup_{\beta \geq 1} \pi(\beta|\delta)$. Using the answer to part (a), we see that $\pi(\beta|\delta)$ is a decreasing function of β, hence the size of the test is $\pi(1|\delta) = \exp(-1)$.

2. (a) We know that if $0 < y < \theta$, then $\Pr(Y_n \leq y) = (y/\theta)^n$. Also, if $y \geq \theta$, then $\Pr(Y_n \leq y) = 1$. Therefore, if $\theta \leq 1.5$, then $\pi(\theta) = \Pr(Y_n \leq 1.5) = 1$. If $\theta > 1.5$, then $\pi(\theta) = \Pr(Y_n \leq 1.5) = (1.5/\theta)^n$.

(b) The size of the test is

$$\alpha = \sup_{\theta \geq 2} \pi(\theta) = \sup_{\theta \geq 2} \left(\frac{1.5}{\theta}\right)^n = \left(\frac{1.5}{2}\right)^n = \left(\frac{3}{4}\right)^n.$$

3. (a) For any given value of $p, \pi(p) = \Pr(Y \geq 7) + \Pr(Y \leq 1)$, where Y has a binomial distribution with parameters $n = 20$ and p. For $p = 0, \Pr(Y \geq 7) = 0$ and $\Pr(Y \leq 1) = 1$. Therefore, $\pi(0) = 1$. For $p = 0.1$, it is found from the table of the binomial distribution that

$$\Pr(Y \geq 7) = .0020 + .0003 + .0001 + .0000 = .0024$$

and $\Pr(Y \leq 1) = .1216 + .2701 = .3917$. Hence, $\pi(0.1) = 0.3941$. Similarly, for $p = 0.2$, it is found that

$$\Pr(Y \geq 7) = .0545 + .0222 + .0074 + .0020 + .0005 + .0001 = .0867$$

and $\Pr(Y \le 1) = .0115 + .0576 = .0691$. Hence, $\pi(0.2) = 0.1558$. By continuing to use the tables in this way, we can find the values of $\pi(0.3)$, $\pi(0.4)$, and $\pi(0.5)$. For $p = 0.6$, we must use the fact that if Y has a binomial distribution with parameters 20 and 0.6, then $Z = 20 - Y$ has a binomial distribution with parameters 20 and 0.4. Also, $\Pr(Y \ge 7) = \Pr(Z \le 13)$ and $\Pr(Y \le 1) = \Pr(Z \ge 19)$. It is found from the tables that $\Pr(Z \le 13) = .9935$ and $\Pr(Z \ge 19) = .0000$. Hence, $\pi(0.6) = .9935$. Similarly, if $p = 0.7$, then $Z = 20-Y$ will have a binomial distribution with parameters 20 and 0.3. In this case it is found that $\Pr(Z \le 13) = .9998$ and $\Pr(Z \ge 19) = .0000$. Hence, $\pi(0.7) = 0.9998$. By continuing in this way, the values of $\pi(0.8)$, $\pi(0.9)$, and $\pi(1.0) = 1$ can be obtained.

(b) Since H_0 is a simple hypothesis, the size α of the test is just the value of the power function at the point specified by H_0. Thus, $\alpha = \pi(0.2) = 0.1558$.

4. The null hypothesis H_0 is simple. Therefore, the size α of the test is $\alpha = \Pr(\text{Rejecting } H_0 \mid \mu = \mu_0)$. When $\mu = \mu_0$, the random variable $Z = n^{1/2}(\overline{X}_n - \mu_0)$ will have the standard normal distribution. Hence, since $n = 25$,

$$\alpha = \Pr(|\overline{X}_n - \mu_0| \ge c) = \Pr(|Z| \ge 5c) = 2[1 - \Phi(5c)].$$

Thus, $\alpha = 0.05$ if and only if $\Phi(5c) = 0.975$. It is found from a table of the standard normal distribution that $5c = 1.96$ and $c = 0.392$.

5. A hypothesis is simple if and only if it specifies a single value of both μ and σ. Therefore, only the hypothesis in (a) is simple. All the others are composite. In particular, although the hypothesis in (d) specifies the value of μ, it leaves the value of σ arbitrary.

6. If H_0 is true, then X will surely be smaller than 3.5. If H_1 is true, then X will surely be greater than 3.5. Therefore, the test procedure which rejects H_0 if and only if $X > 3.5$ will have probability 0 of leading to a wrong decision, no matter what the true value of θ is.

7. Let C be the critical region of Y_n values for the test δ, and let C^* be the critical region for δ^*. It is easy to see that $C^* \subset C$. Hence

$$\pi(\theta|\delta) - \pi(\theta|\delta^*) = \Pr\left(Y_n \in C \cap (C^*)^C \middle| \theta\right).$$

Here $C \cap (C^*)^C = [4, 4.5]$, so

$$\pi(\theta|\delta) - \pi(\theta|\delta^*) = \Pr(4 \le Y_n \le 4.5|\theta). \tag{S.9.1}$$

(a) For $\theta \le 4$ $\Pr(4 \le Y_n|\theta) = 0$, so the two power functions must be equal by (S.9.1).

(b) For $\theta > 4$,

$$\Pr(4 \le Y_n \le 4.5|\theta) = \frac{(\min\{\theta, 4.5\})^n - 4^n}{\theta^n} > 0.$$

Hence, $\pi(\theta|\delta) > \pi(\theta|\delta^*)$ by (S.9.1).

(c) The only places where the power functions differ are for $\theta > 4$. Since these values are all in Ω_1, it is better for a test to have higher power function for these values. Since δ has higher power function than δ^* for all of these values, δ is the better test.

8. (a) The distribution of Z given μ is the normal distribution with mean $n^{1/2}(\mu - \mu_0)$ and variance 1. We can write

$$\Pr(Z \ge c|\mu) = 1 - \Phi([c - n^{1/2}(\mu - \mu_0)]) = \Phi(n^{1/2}\mu - n^{1/2}\mu_0 - c).$$

Since Φ is an increasing function and $n^{1/2}\mu - n^{1/2}\mu_0 - c$ is an increasing function of μ, the power function is an increasing function of μ.

(b) The size of the test will be the power function at $\mu = \mu_0$, since μ_0 is the largest value in Ω_0 and the power function is increasing. Hence, the size is $\Phi(-c)$. If we set this equal to α_0, we can solve for $c = -\Phi^{-1}(\alpha_0)$.

9. A sensible test would be to reject H_0 if $\overline{X}_n < c'$. So, let $T = \mu_0 - \overline{X}_n$. Then the power function of the test δ that rejects H_0 when $T \ge c$ is

$$\begin{aligned}\pi(\mu|\delta) &= \Pr(T \ge c|\mu)\\ &= \Pr(\overline{X}_n \le \mu_0 - c|\mu)\\ &= \Phi(\sqrt{n}[\mu_0 - c - \mu]).\end{aligned}$$

Since Φ is an increasing function and $\sqrt{n}[\mu_0 - c - \mu]$ is a decreasing function of μ, it follows that $\Phi(\sqrt{n}[\mu_0 - c - \mu])$ is a decreasing function of μ.

10. When $Z = z$ is observed, the p-value is $\Pr(Z \ge z|\mu_0) = \Phi(n^{1/2}[\mu_0 - z])$.

11. (a) For $c_1 \ge 2$, $\Pr(Y \le c_1|p = 0.4) \ge 0.23$, hence $c_1 \le 1$. Also, for $c_2 \le 5$, $\Pr(Y \ge c_2|p = 0.4) \ge 0.26$, hence $c_2 \ge 6$. Here are some values of the desired probability for various (c_1, c_2) pairs

c_1	c_2	$\Pr(Y \le c_1\|p = 0.4) + \Pr(Y \ge c_2\|p = 0.4)$
1	6	0.1699
1	7	0.0956
0	6	0.1094
−1	6	0.0994

So, the closest we can get to 0.1 without going over is 0.0994, which is achieved when $c_1 < 0$ and $c_2 = 6$.

(b) The size of the test is 0.0994, as we calculated in part (a).

(c) The power function is plotted in Fig. S.9.1. Notice that the power function is too low for values of $p < 0.4$. This is due to the fact that the test only rejects H_0 when $Y \ge 6$ A better test might be one with $c_1 = 1$ and $c_2 = 7$. Even though the size is slightly smaller (as is the power for $p > 0.4$), its power is much greater for $p < 0.4$.

12. (a) The power function of δ_c is

$$\pi(\theta|\delta_c) = \Pr(X \ge c|\theta) = \int_c^\infty \frac{dx}{\pi[1 + (x - \theta)^2]} = \frac{1}{\pi}\left[\frac{\pi}{2} - \arctan(c - \theta)\right].$$

Since arctan is an increasing function and $c - \theta$ is a decreasing function of θ, the power function is increasing in θ.

(b) To make the size of the test 0.05, we need to solve

$$0.05 = \frac{1}{\pi}\left[\frac{\pi}{2} - \arctan(c - \theta_0)\right],$$

for c. We get

$$c = \theta_0 + \tan(0.45\pi) = \theta_0 + 6.314.$$

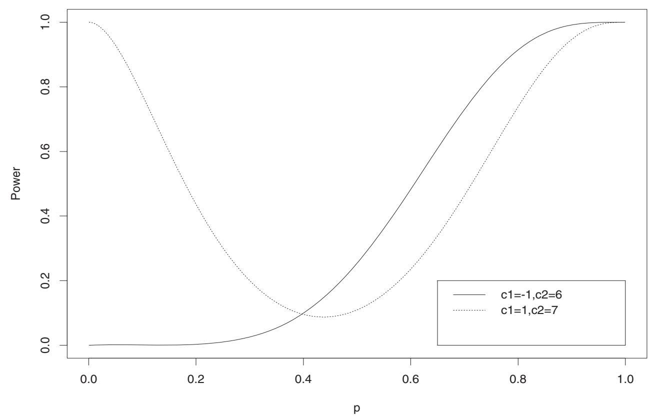

Figure S.9.1: Power function of test in Exercise 11c of Sec. 9.1.

(c) The p-value when $X = x$ is observed is

$$\Pr(X \geq x|\theta = \theta_0) = \frac{1}{\pi}\left[\frac{\pi}{2} - \arctan(x - \theta_0)\right].$$

13. For $c = 3$, $\Pr(X \geq c|\theta = 1) = 0.0803$, while for $c = 2$, the probability is 0.2642. Hence, we must use $c = 3$.

14. (a) The distribution of X is a gamma distribution with parameters n and θ and $Y = X\theta$ has a gamma distribution with parameters n and 1. Let G_n be the c.d.f. of the gamma distribution with parameters n and 1. The power function of δ_c is then

$$\pi(\theta|\delta_c) = \Pr(X \geq c|\theta) = \Pr(Y \geq c\theta|\theta) = 1 - G_n(c\theta).$$

Since $1 - G_n$ is an decreasing function and $c\theta$ is an increasing function of θ, $1 - G_n(c\theta)$ is a decreasing function of θ.

(b) We need $1 - G_n(c\theta_0) = \alpha_0$. This means that $c = G_n^{-1}(1 - \alpha_0)/\theta_0$.

(c) With $\alpha_0 = 0.1$, $n = 1$ and $\theta_0 = 2$, we find that $G_n(y) = 1 - \exp(-y)$ and $G_n^{-1}(p) = -\log(1 - p)$. So, $c = -\log(0.1)/2 = 1.151$. The power function is plotted in Fig. S.9.2.

15. The p-value when $X = x$ is observed is the size of the test that rejects H_0 when $X \geq x$, namely

$$\Pr(X \geq x|\theta = 1) = \begin{cases} 0 & \text{if } x \geq 1, \\ 1 - x & \text{if } 0 < x < 1. \end{cases}$$

16. The confidence interval is $(s_n^2/c_2, s_n^2/c_1)$, where $s_n^2 = \sum_{i=1}^{n}(x_i - \overline{x}_n)^2$ and c_1, c_2 are the $(1 - \gamma)/2$ and $(1 + \gamma)/2$ quantiles of the χ^2 distribution with $n - 1$ degrees of freedom. We create the test δ_c of $H_0 : \sigma^2 = c$ by rejecting H_0 if c is not in the interval. Let $T(\boldsymbol{x}) = s_n^2$ and notice that c is outside of the interval if and only if $T(\boldsymbol{x})$ is not in the interval (c_1c, c_2c).

17. We need $q(y)$ to have the property that $\Pr(q(Y) < p|p) \geq \gamma$ for all p. We shall prove that $q(y)$ equal to the smallest p_0 such that $\Pr(Y \geq y|p = p_0) \geq 1 - \gamma$ satisfies this property. For each p, let $A_p = \{y : q(y) < p\}$. We need to show that $\Pr(Y \in A_p|p) \geq \gamma$. First, notice that $q(y)$ is an increasing

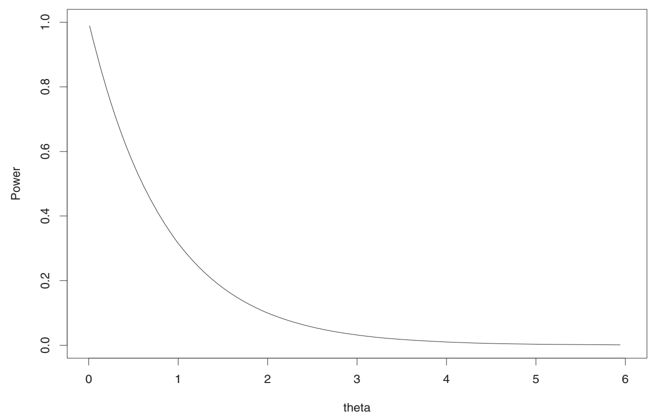

Figure S.9.2: Power function of test in Exercise 14c of Sec. 9.1.

function of y. This means that for each p there is y_p such that $A_p = \{0, \ldots, y_p\}$. So, we need to show that $\Pr(Y \le y_p|p) \ge \gamma$ for all p. Equivalently, we need to show that $\Pr(Y > y_p|p) \le 1-\gamma$. Notice that y_p is the largest value of y such that $q(y) < p$. That is, y_p is the largest value of y such that there exists $p_0 < p$ with $\Pr(Y \ge y|p_0) \ge 1-\gamma$. For each y, $\Pr(Y > y|p)$ is a continuous nondecreasing function of p. If $\Pr(Y > y_p|p) > 1-\gamma$, then there exists $p_0 < p$ such that

$$1-\gamma < \Pr(Y > y_p|p_0) = \Pr(Y \ge y_p + 1|p_0).$$

This contradicts the fact that y_p is the largest y such that there is $p_0 < p$ with $\Pr(Y \ge y|p_0) \ge 1-\gamma$. Hence $\Pr(Y > y_p|p) \le 1-\gamma$ and the proof is complete.

18. Our tests are all of the form "Reject H_0 if $T \ge c$." Let δ_c be this test, and define

$$\alpha(c) = \sup_{\theta \in \Omega_0} \Pr(T \ge c|\theta),$$

the size of the test δ_c. Then δ_c has level of significance α_0 if and only if $\alpha(c) \le \alpha_0$. Notice that $\alpha(c)$ is a decreasing function of c. When $T = t$ is observed, we reject H_0 at level of significance α_0 using δ_c if and only if $t \ge c$, which is equivalent to $\alpha(t) \le \alpha_0$. Hence $\alpha(t)$ is the smallest level of significance at which we can reject H_0 if $T = t$ is observed. Notice that $\alpha(t)$ is the expression in Eq. (9.1.12).

19. We want our test to reject H_0 if $\overline{X}_n \le Y$, where Y might be a random variable. We can write this as not rejecting H_0 if $\overline{X}_n > Y$. We want $\overline{X}_n > Y$ to be equivalent to μ_0 being inside of our interval. We need the test to have level α_0, so

$$\Pr(\overline{X}_n \le Y|\mu = \mu_0, \sigma^2) = \alpha_0 \qquad \text{(S.9.2)}$$

is necessary. We know that $n^{1/2}(\overline{X}_n - \mu_0)/\sigma'$ has the t distribution with $n-1$ degrees of freedom if $\mu = \mu_0$, hence Eq. (S.9.2) will hold if $Y = \mu_0 - n^{-1/2}\sigma' T_{n-1}^{-1}(1-\alpha_0)$. Now, $\overline{X}_n > Y$ if and only if $\mu_0 < \overline{X}_n + n^{-1/2}\sigma' T_{n-1}^{-1}(1-\alpha_0)$. This is equivalent to μ_0 in our interval if our interval is

$$\left(-\infty, \overline{X}_n + n^{-1/2}\sigma' T_{n-1}^{-1}(1-\alpha_0)\right).$$

20. Let $\theta_0 \in \Omega$, and let $g_0 = g(\theta_0)$. By construction, $g(\theta_0) \in \omega(\boldsymbol{X})$ if and only if δ_{g_0} does not reject $H_{0,g_0} : g(\theta) \le g_0$. Given $\theta = \theta_0$, the probability that δ_{g_0} does not reject H_{0,g_0} is at least γ because the null hypothesis is true and the level of the test is $\alpha_0 = 1 - \gamma$. Hence, (9.1.15) holds.

21. Let $U = n^{1/2}(\overline{X}_n - \mu_0)/\sigma'$.

    (a) We reject the null hypothesis in (9.1.22) if and only if

    $$U \ge T_{n-1}^{-1}(1-\alpha_0). \tag{S.9.3}$$

    We reject the null hypothesis in (9.1.27) if and only if

    $$U \le -T_{n-1}^{-1}(1-\alpha_0). \tag{S.9.4}$$

    With $\alpha_0 < 0.5$, $T_{n-1}^{-1}(1-\alpha_0) > 0$. So, (S.9.3) requires $U > 0$ while (S.9.4) requires $U < 0$. These cannot both occur.

    (b) Both (S.9.3) and (S.9.4) fail if and only if U is strictly between $-T_{n-1}^{-1}(1-\alpha_0)$ and $T_{n-1}^{-1}(1-\alpha_0)$. This can happen if $\overline{X}_n$ is sufficiently close to μ_0. This has probability $1 - 2\alpha_0 > 0$.

    (c) If $\alpha_0 > 0.5$, then $T_{n-1}^{-1}(1-\alpha_0) < 0$, and both null hypotheses would be rejected if U is between the numbers $T_{n-1}^{-1}(1-\alpha_0) < 0$ and $-T_{n-1}^{-1}(1-\alpha_0) > 0$. This has probability $2\alpha_0 - 1 > 0$.

9.2 Testing Simple Hypotheses

Commentary

This section, and the two following, contain some traditional optimality results concerning tests of hypotheses about one-dimensional parameters. In this section, we present the Neyman-Pearson lemma which gives optimal tests for simple null hypotheses against simple alternative hypotheses. It is recommended that one skip this section, and the two that follow, unless one is teaching a rigorous mathematical statistics course. This section ends with a brief discussion of randomized tests. Randomized tests are mainly of theoretical interest. They only show up in one additional place in the text, namely the proof of Theorem 9.3.1.

Solutions to Exercises

1. According to Theorem 9.2.1, we should reject H_0 if $f_1(x) > f_0(x)$, not reject H_0 if $f_1(x) < f_0(x)$ and do whatever we wish if $f_1(x) = f_0(x)$. Here

$$f_0(x) = \begin{cases} 0.3 & \text{if } x = 1, \\ 0.7 & \text{if } x = 0, \end{cases}$$

$$f_1(x) = \begin{cases} 0.6 & \text{if } x = 1, \\ 0.4 & \text{if } x = 0. \end{cases}$$

   We have $f_1(x) > f_0(x)$ if $x = 1$ and $f_1(x) < f_0(x)$ if $x = 0$. We never have $f_1(x) = f_0(x)$. So, the test is to reject H_0 if $X = 1$ and not reject H_0 if $X = 0$.

2. (a) Theorem 9.2.1 can be applied with $a = 1$ and $b = 2$. Therefore, H_0 should not be rejected if $f_1(x)/f_0(x) < 1/2$. Since $f_1(x)/f_0(x) = 2x$, the procedure is to not reject H_0 if $x < 1/4$ and to reject H_0 if $x > 1/4$.

(b) For this procedure,

$$\alpha(\delta) = \Pr(\text{Rej. } H_0 \,|\, f_0) = \int_{1/4}^{1} f_0(x)\,dx = \frac{3}{4}$$

and

$$\beta(\delta) = \Pr(\text{Acc. } H_0 \,|\, f_1) = \int_{0}^{1/4} 2x\,dx = \frac{1}{16}.$$

Therefore, $\alpha(\delta) + 2\beta(\delta) = 7/8$.

3. (a) Theorem 9.2.1 can be applied with $a = 3$ and $b = 1$. Therefore, H_0 should not be rejected if $f_1(x)/f_0(x) = 2x < 3$. Since all possible values of X lie in the interval (0,1), and since $2x < 3$ for all values in this interval, the optimal procedure is to not reject H_0 for every possible observed value.

   (b) Since H_0 is never rejecte, $\alpha(\delta) = 0$ and $\beta(\delta) = 1$. Therefore, $3\alpha(\delta) + \beta(\delta) = 1$.

4. (a) By the Neyman-Pearson lemma, H_0 should be rejected if $f_1(x)/f_0(x) = 2x > k$, where k is chosen so that $\Pr(2x > k \,|\, f_0) = 0.1$. For $0 < k < 2$,

$$\Pr(2X > k \,|\, f_0) = \Pr\left(X > \frac{k}{2}\middle|\, f_0\right) = 1 - \frac{k}{2}.$$

   If this value is to be equal to 0.1, then $k = 1.8$. Therefore, the optimal procedure is to reject H_0 if $2x > 1.8$ or, equivalently, if $x > 0.9$.

   (b) For this procedure,

$$\beta(\delta) = \Pr(\text{Acc. } H_0 \,|\, f_1) = \int_{0}^{0.9} f_1(x)dx = 0.81.$$

5. (a) The conditions here are different from those of the Neyman-Pearson lemma. Rather than fixing the value of $\alpha(\delta)$ and minimizing $\beta(\delta)$, we must here fix the value of $\beta(\delta)$ and minimize $\alpha(\delta)$. Nevertheless, the same proof as that given for the Neyman-Pearson lemma shows that the optimal procedure is again to reject H_0 if $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > k$, where k is now chosen so that

$$\beta(\delta) = \Pr(\text{Acc. } H_0 \,|\, H_1) = \Pr\left[\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} < k \,|\, H_1\right] = 0.05.$$

   In this exercise,

$$f_0(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - 3.5)^2\right]$$

   and

$$f_1(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - 5.0)^2\right].$$

   Therefore,

$$\begin{aligned}
\log\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} &= \frac{1}{2}\left[\sum_{i=1}^{n}(x_i - 3.5)^2 - \sum_{i=1}^{n}(x_i - 5.0)^2\right] \\
&= \frac{1}{2}\left[\sum_{i=1}^{n} x_i^2 - 7\sum_{i=1}^{n} x_i + 12.25n - \sum_{i=1}^{n} x_i^2 + 10\sum_{i=1}^{n} x_i - 25n\right] \\
&= \frac{3}{2} n\overline{x}_n - (const.).
\end{aligned}$$

It follows that the likelihood ratio $f_1(\boldsymbol{X})/f_0(\boldsymbol{X})$ will be greater than some specified constant k if and only if $\overline{x}_n$ is greater than some other constant k'. Therefore, the optimal procedure is to reject H_0 if $\overline{x}_n > k'$, where k' is chosen so that

$$\Pr(\overline{X}_n < k' \mid H_1) = 0.05.$$

We shall now determine the value of k'. If H_1 is true, then $\overline{X}_n$ will have a normal distribution with mean 5.0 and variance $1/n$. Therefore, $Z = \sqrt{n}(\overline{X}_n - 5.0)$ will have the standard normal distribution, and it follows that

$$\Pr(\overline{X}_n < k' \mid H_1) = \Pr[Z < \sqrt{n}(k' - 5.0)] = \Phi[\sqrt{n}(k' - 5.0)].$$

If this probability is to be equal to 0.05, then it can be found from a table of values of Φ that $\sqrt{n}(k' - 5.0) = -1.645$. Hence, $k' = 5.0 - 1.645n^{-1/2}$.

(b) For $n = 4$, the test procedure is to reject H_0 if $\overline{X}_n > 5.0 - 1.645/2 = 4.1775$. Therefore,

$$\alpha(\delta) = \Pr(\text{Rej. } H_0 \mid H_0) = \Pr(\overline{X}_n > 4.1775 \mid H_0).$$

When H_0 is true, $\overline{X}_n$has a normal distribution with mean 3.5 and variance $1/n = 1/4$. Therefore, $Z = 2(\overline{X}_n - 3.5)$ will have the standard normal distribution, and

$$\begin{aligned}\alpha(\delta) &= \Pr[Z > 2(4.1775 - 3.5)] = \Pr(Z > 1.355) \\ &= 1 - \Phi(1.355) = 0.0877.\end{aligned}$$

6. Theorem 9.2.1 can be applied with $a = b = 1$. Therefore, H_0 should be rejected if $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > 1$. If we let $y = \sum_{i=1}^{n} x_i$, then

$$f_1(\boldsymbol{X}) = p_1^y(1 - p_1)^{n-y}$$

and

$$f_0(\boldsymbol{X}) = p_0^y(1 - p_0)^{n-y}.$$

Hence,

$$\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} = \left[\frac{p_1(1 - p_0)}{p_0(1 - p_1)}\right]^y \left(\frac{1 - p_1}{1 - p_0}\right)^n.$$

But $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > 1$ if and only if $\log[f_1(\boldsymbol{X})/f_0(\boldsymbol{X})] > 0$, and this inequality will be satisfied if and only if

$$y \log\left[\frac{p_1(1 - p_0)}{p_0(1 - p_1)}\right] + n \log\left(\frac{1 - p_1}{1 - p_0}\right) > 0.$$

Since $p_1 < p_0$ and $1 - p_0 < 1 - p_1$, the first logarithm on the left side of this relation is negative. Finally, if we let $\overline{x}_n = y/n$, then this relation can be rewritten as follows:

$$\overline{x}_n < \frac{\log\left(\dfrac{1 - p_1}{1 - p_0}\right)}{\log\left[\dfrac{p_0(1 - p_1)}{p_1(1 - p_0)}\right]}.$$

The optimal procedure is to reject H_0 when this inequality is satisfied.

7. (a) By the Neyman-Pearson lemma, H_0 should be rejected if $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > k$. Here,

$$f_0(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}2^{n/2}} \exp\left[-\frac{1}{4}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

and

$$f_1(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}3^{n/2}} \exp\left[-\frac{1}{6}\sum_{i=1}^{n}(x_i-\mu)^2\right].$$

Therefore,

$$\log\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} = \frac{1}{12}\sum_{i=1}^{n}(x_i-\mu)^2 + (const.).$$

It follows that the likelihood ratio will be greater than a specified constant k if and only if $\sum_{i=1}^{n}(x_i-\mu)^2$ is greater than some other constant c. The constant c is to be chosen so that

$$\Pr\left[\sum_{i=1}^{n}(X_i-\mu)^2 > c \,\middle|\, H_0\right] = 0.05.$$

The value of c can be determined as follows. When H_0 is true, $W = \sum_{i=1}^{n}(X_i-\mu)^2/2$ will have χ^2 distribution with n degrees of freedom. Therefore,

$$\Pr\left[\sum_{i=1}^{n}(X_i-\mu)^2 > c \,\middle|\, H_0\right] = \Pr\left(W > \frac{c}{2}\right).$$

If this probability is to be equal to 0.05, then the value of $c/2$ can be determined from a table of the χ^2 distribution.

(b) For $n = 8$, it is found from a table of the χ^2 distribution with 8 degrees of freedom that $c/2 = 15.51$ and $c = 31.02$.

8. (a) The p.d.f.'s $f_0(x)$ and $f_1(x)$ are as sketched in Fig. S.9.3. Under H_0 it is impossible to obtain a value of X greater than 1, but such values are possible under H_1. Therefore, if a test procedure rejects H_0 only if $x > 1$, then it is impossible to make an error of type 1, and $\alpha(\delta) = 0$. Also,

$$\beta(\delta) = \Pr(X < 1 \,|\, H_1) = \frac{1}{2}.$$

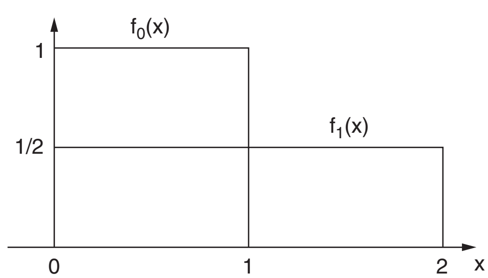

Figure S.9.3: Figure for Exercise 8a of Sec. 9.2.

(b) To have $\alpha(\delta) = 0$, we can include in the critical region only a set of points having probability 0 under H_0. Therefore, only points $x > 1$ can be considered. To minimize $\beta(\delta)$ we should choose this set to have maximum probability under H_1. Therefore, all points $x > 1$ should be used in the critical region.

9. As in Exercise 8, we should reject H_0 if at least one of the n observations is greater than 1. For this test, $\alpha(\delta) = 0$ and

$$\beta(\delta) = \Pr(\text{Acc. } H_0 \,|\, H_1) = \Pr(X_1 < 1, \ldots, X_n < 1 \,|\, H_1) = \left(\frac{1}{2}\right)^n.$$

10. (a) and (b). Theorem 9.2.1 can be applied with $a = b = 1$. The optimal procedure is to reject H_0 if $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > 1$. If we let $y = \sum_{i=1}^{n} x_i$, then for $i = 0, 1$,

$$f_i(\boldsymbol{X}) = \frac{\exp(-n\lambda_i)\lambda_i^y}{\prod_{i=1}^n (x_i!)}.$$

Therefore,

$$\log \frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} = y \log\left(\frac{\lambda_1}{\lambda_0}\right) - n(\lambda_1 - \lambda_0).$$

Since $\lambda_1 > \lambda_0$, it follows that $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > 1$ if and only if $\overline{x}_n = y/n > (\lambda_1 - \lambda_0)/(\log \lambda_1 - \log \lambda_0)$.

(c) If H_i is true, then Y will have a Poison distribution with mean $n\lambda_i$. For $\lambda_0 = 1/4$, $\lambda_1 = 1/2$, and $n = 20$,

$$\frac{n(\lambda_1 - \lambda_0)}{\log \lambda_1 - \log \lambda_0} = \frac{20(0.25)}{0.69314} = 7.214.$$

Therefore, it is found from a table of the Poison distribution with mean $20(1/4) = 5$ that

$$\alpha(\delta) = \Pr(Y > 7.214 \,|\, H_0) = \Pr(Y \geq 8 \,|\, H_0) = 0.1333.$$

Also, it is found from a table with mean $20(1/2) = 10$ that

$$\beta(\delta) = \Pr(Y \leq 7.214 \,|\, H_1) = \Pr(Y \leq 7 \,|\, H_1) = 0.2203.$$

Therefore, $\alpha(\delta) + \beta(\delta) = 0.3536$.

11. Theorem 9.2.1 can be applied with $a = b = 1$. The optimal procedure is to reject H_0 if $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > 1$. Here,

$$f_0(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}2^n} \exp\left[-\frac{1}{8}\sum_{i=1}^{n}(x_i + 1)^2\right]$$

and

$$f_1(\boldsymbol{X}) = \frac{1}{(2\pi)^{n/2}2^n} \exp\left[-\frac{1}{8}\sum_{i=1}^{n}(x_i - 1)^2\right].$$

After some algebraic reduction, it can be shown that $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > 1$ if and only if $\overline{x}_n > 0$. If H_0 is true, $\overline{X}_n$ will have the normal distribution with mean -1 and variance $4/n$. Therefore, $Z = \sqrt{n}(\overline{X}_n + 1)/2$ will have the standard normal distribution, and

$$\alpha(\delta) = \Pr(\overline{X}_n > 0 \,|\, H_0) = \Pr\left(Z > \frac{1}{2}\sqrt{n}\right) = 1 - \Phi\left(\frac{1}{2}\sqrt{n}\right).$$

Similarly, if H_1 is true, $\overline{X}_n$will have the normal distribution with mean 1 and variance $4/n$. Therefore, $Z' = \sqrt{n}(\overline{X}_n - 1)/2$ will have the standard normal distribution, and

$$\beta(\delta) = \Pr(\overline{X}_n < 0 \,|\, H_1) = \Pr\left(Z' < -\frac{1}{2}\sqrt{n}\right) = 1 - \Phi\left(\frac{1}{2}\sqrt{n}\right).$$

Hence, $\alpha(\delta) + \beta(\delta) = 2[1 - \Phi(\sqrt{n}/2)]$. We can now use a program that computes Φ to obtain the following results:

(a) If $n = 1$, $\alpha(\delta) + \beta(\delta) = 2(0.3085) = 0.6170$.

(b) If $n = 4$, $\alpha(\delta) + \beta(\delta) = 2(0.1587) = 0.3173$.

(c) If $n = 16$, $\alpha(\delta) + \beta(\delta) = 2(0.0228) = 0.0455$.

(d) If $n = 36$, $\alpha(\delta) + \beta(\delta) = 2(0.0013) = 0.0027$.

Slight discrepancies appear above due to rounding *after* multiplying by 2 rather than before.

12. In the notation of this section, $f_i(\boldsymbol{x}) = \theta_i^n \exp\left(-\theta_i \sum_{j=1}^n x_j\right)$ for $i = 0, 1$. The desired test has the following form: reject H_0 if $f_1(\boldsymbol{x})/f_0(\boldsymbol{x}) > k$ where k is chosen so that the probability of rejecting H_0 is α_0 given $\theta = \theta_0$. The ratio of f_1 to f_0 is

$$\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} = \frac{\theta_1^n}{\theta_0^n} \exp\left([\theta_0 - \theta_1] \sum_{i=1}^n x_i\right).$$

Since $\theta_0 < \theta_1$, the above ratio will be greater than k if and only if $\sum_{i=1}^n x_i$ is less than some other constant, c. That c is chosen so that $\Pr\left(\sum_{i=1}^n X_i < c \middle| \theta = \theta_0\right) = \alpha_0$. The distribution of $\sum_{i=1}^n X_i$ given $\theta = \theta_0$ is the gamma distribution with parameters n and θ_0. Hence, c must be the α_0 quantile of that distribution.

13. (a) The test rejects H_0 if $f_0(\boldsymbol{X}) < f_12(\boldsymbol{X})$. In this case, $f_0(\boldsymbol{x}) = \exp(-[x_1 + x_2]/2)/4$, and $f_1(\boldsymbol{x}) = 4/(2 + x_1 + x_2)^3$ for both $x_1 > 0$ and $x_2 > 0$. Let $T = X_1 + X_2$. Then we reject H_0 if

$$\exp(-T/2)/4 < 4/(2 + T)^3. \tag{S.9.5}$$

(b) If $X_1 = 4$ and $X_2 = 3$ are observed, then $T = 7$. The inequality in (S.9.5) is $\exp(-7/2)/4 < 4/9^3$ or $0.007549 < 0.00549$, which is false, so we do not reject H_0.

(c) If H_0 is true, then T is the sum of two independent exponential random variables with parameter 1/2. Hence, it has the gamma distribution with parameters 2 and 1/2 by Theorem 5.7.7.

(d) The test is to reject H_0 if $f_1(\boldsymbol{X})/f_0(\boldsymbol{X}) > c$, where c is chosen so that the probability is 0.1 that we reject H_0 given $\theta = \theta_0$. We can write

$$\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} = \frac{16\exp(T/2)}{(2 + T)^3}. \tag{S.9.6}$$

The function on the right side of (S.9.6) takes the value 2 at $T = 0$, decreases to the value 0.5473 at $T = 4$, and increases for $T > 4$. Let G be the c.d.f. of the gamma distribution with parameters 2 and 1/2 (also the χ^2 distribution with 4 degrees of freedom). The level 0.01 test will reject H_0 if $T < c_1$ or $T > c_2$ where c_1 and c_2 satisfy $G(c_1) + 1 - G(c_2) = 0.01$, and either $16\exp(c_1/2)/(2+c_1)^3 = 16\exp(c_2/2)/(2+c_2)^3$ or $c_1 = 0$ and $16\exp(c_2/2)/(2+c_2)^3 > 2$. It follows that $1 - G(c_2) \le 0.01$, that is, $c_2 \ge G^{-1}(0.99) = 13.28$. But

$$16\exp(13.28)/(2 + 13.28)^3 = 3.4 > 2.$$

It follows that $c_1 = 0$ and the test is to reject H_0 if $T > 13.28$.

(e) If $X_1 = 4$ and $X_2 = 3$, then $T = 7$ and we do not reject H_0.

9.3 Uniformly Most Powerful Tests

Commentary

This section introduces the concept of monotone likelihood ratio, which is used to provide conditions under which uniformly most powerful tests exist for one-sided hypotheses. One may safely skip this section if one is not teaching a rigorous mathematical statistics course. One step in the proof of Theorem 9.3.1 relies on randomized tests (Sec. 9.2), which the instructor might have skipped earlier.

Solutions to Exercises

1. Let $y = \sum_{i=1}^{n} x_i$. Then the joint p.f. is

$$f_n(\boldsymbol{X} \mid \lambda) = \frac{\exp(-n\lambda)\lambda^y}{\prod_{i=1}^{n}(x_i!)}$$

Therefore, for $0 < \lambda_1 < \lambda_2$,

$$\frac{f_n(\boldsymbol{X} \mid \lambda_2)}{f_n(\boldsymbol{X} \mid \lambda_1)} = \exp(-n(\lambda_2 - \lambda_1)) \left(\frac{\lambda_2}{\lambda_1}\right)^y,$$

which is an increasing function of y.

2. Let $y = \sum_{i=1}^{n}(x_1 - \mu)^2$. Then the joint p.d.f. is

$$f_n(\boldsymbol{X} \mid \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp(-\frac{y}{2\sigma^2}).$$

Therefore, for $0 < \sigma_1^2 < \sigma_2^2$,

$$\frac{f_n(\boldsymbol{X} \mid \sigma_2^2)}{f_n(\boldsymbol{X} \mid \sigma_1^2)} = \frac{\sigma_1^n}{\sigma_2^n} \exp\left\{\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) y\right\},$$

which is an increasing function of y.

3. Let $y = \prod_{i=1}^{n} x_i$ and let $z = \sum_{i=1}^{n} x_i$. Then the joint p.d.f. is

$$f_n(\boldsymbol{X} \mid \alpha) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} y^{\alpha-1} \exp(-\beta z).$$

Therefore, for $0 < \alpha_1 < \alpha_2$,

$$\frac{f_n(\boldsymbol{X} \mid \alpha_2)}{f_n(\boldsymbol{X} \mid \alpha_1)} = (const.) y^{\alpha_2 - \alpha_1},$$

which is an increasing function of y.

4. The joint p.d.f. $f_n(\boldsymbol{X} \mid \beta)$ in this exercise is the same as the joint p.d.f. $f_n(\boldsymbol{X} \mid \alpha)$ given in Exercise 3, except that the value of β is now unknown and the value of α is known. Since $z = n\overline{x}_n$, it follows that for $0 < \beta_1 < \beta_2$,

$$\frac{f_n(\boldsymbol{X}|\beta_2)}{f_n(\boldsymbol{X}|\beta_1)} = (const.)\exp([\beta_1 - \beta_2]n\overline{x}_n).$$

The expression on the right side of this relation is a decreasing function of $\overline{x}_n$, because $\beta_1 - \beta_2 < 0$. Therefore, this expression is an increasing function of $-\overline{x}_n$.

5. Let $y = \sum_{i=1}^{n} d(x_i)$. Then the joint p.d.f. or the joint p.f. is

$$f_n(\boldsymbol{X} \mid \theta) = [a(\theta)]^n \left[\prod_{i=1}^{n} b(x_i)\right] \exp[c(\theta)y].$$

Therefore, for $\theta_1 < \theta_2$,

$$\frac{f_n(\boldsymbol{X} \mid \theta_2)}{f_n(\boldsymbol{X} \mid \theta_1)} = \left[\frac{a(\theta_2)}{a(\theta_1)}\right]^n \exp\{[c(\theta_2) - c(\theta_1)]y\}.$$

Since $c(\theta_2) - c(\theta_1) > 0$, this expression is an increasing function of y.

6. Let $\theta_1 < \theta_2$. The range of possible values of $r(\boldsymbol{X}) = \max\{X_1, \ldots, X_n\}$ is the interval $[0, \theta_2]$ when comparing θ_1 and θ_2. The likelihood ratio for values of $r(\boldsymbol{x})$ in this interval is

$$\begin{cases} \dfrac{\theta_1^n}{\theta_2^n} & \text{if } 0 \le r(\boldsymbol{x}) \le \theta_1, \\ \infty & \text{if } \theta_1 < r(\boldsymbol{x}) \le \theta_2. \end{cases}$$

This is monotone increasing, even though it takes only two values. It does take the larger value ∞ when $r(\boldsymbol{x})$ is large and it takes the smaller value θ_1^n/θ_2^n when $r(\boldsymbol{x})$ is small.

7. No matter what the true value of θ is, the probability that H_0 will be rejected is 0.05. Therefore, the value of the power function at every value of θ is 0.05.

8. We know from Exercise 2 that the joint p.d.f. of $X_1, \ldots, X_n$ has a monotone likelihood ratio in the statistic $\sum_{i=1}^{n} X_i^2$. Therefore, by Theorem 9.3.1, a test which rejects H_0 when $\sum_{i=1}^{n} X_i^2 \ge c$ will be a UMP test. To achieve a specified level of significance α_0, the constant c should be chosen so that $\Pr\left(\sum_{i=1}^{n} X_i^2 \ge c/\sigma^2 = 2\right) = \alpha_0$. Since $\sum_{i=1}^{n} X_i^2$ has a continuous distribution and not a discrete distribution, there will be a value of c which satisfies this equation for any specified value of α_0 $(0 < \alpha_0 < 1)$.

9. The first part of this exercise was answered in Exercise 8. When n $= 10$ and $\sigma^2 = 2$, the distribution of $Y = \sum_{i=1}^{n} X_i^2/2$ will be the χ^2 distribution with 10 degrees of freedom, and it is found from a table of this distribution that $\Pr(Y \ge 18.31) = 0.05$. Also,

$$\Pr\left(\sum_{i=1}^{n} X_i^2 \ge c \mid \sigma^2 = 2\right) = \Pr\left(Y \ge \frac{c}{2}\right).$$

Therefore, if this probability is to be equal to 0.05, then $c/2 = 18.31$ or $c = 36.62$.

10. Let $Y = \sum_{i=1}^{n} X_i$. As in Example 9.3.7, a test which specifies rejecting H_0 if $Y \geq c$ is a UMP test. When $n = 20$ and $p = 1/2$, it is found from the table of the binomial distribution given at the end of the book that

$$\Pr(Y \geq 14) = .0370 + .0148 + .0046 + .0011 + .0002 = .0577.$$

Therefore, the level of significance of the UMP test which rejects H_0 when $Y \geq 14$ will be 0.0577. Similarly, the UMP test that rejects H_0 when $Y \geq 15$ has level of significance 0.0207.

11. It is known from Exercise 1 that the joint p.f. of $X_1, \ldots, X_n$ has a monotone likelihood ratio in he statistic $Y = \sum_{i=1}^{n} X_i$. Therefore, by Theorem 9.3.1, a test which rejects H_0 when $Y \geq c$ will be a UMP test. When $\lambda = 1$ and $n = 10$, Y will have a Poisson distribution with mean 10, and it is found from the table of the Poisson distribution given at the end of this book that

$$\Pr(Y \geq 18) = .0071 + .0037 + .0019 + .0009 + .0004 + .0002 + .0001 = .0143.$$

Therefore, the level of significance of the UMP test which rejects H_0 when $Y \geq 18$ will be 0.0143.

12. Change the parameter from $\theta\ to\ \zeta = -\theta$. In terms of the new parameter ζ, the hypotheses to be tested are:

$$\begin{aligned} H_0 &: \zeta \leq -\theta_0, \\ H_1 &: \zeta > -\theta_0. \end{aligned}$$

Let $g_n(\boldsymbol{X} \mid \zeta) = f_n(\boldsymbol{X} \mid -\zeta)$ denote the joint p.d.f. or the joint p.f. of $X_1, \ldots, X_n$ when ζ is regarded as the parameter. If $\zeta_1 < \zeta_2$, then $\theta_1 = -\zeta_1 > -\zeta_2 = \theta_2$. Therefore, the ratio $g_n/(\boldsymbol{X} \mid \zeta_2)/g_n(\boldsymbol{X} \mid \zeta_1)$ will be a decreasing function of $r(\boldsymbol{X})$. It follows that this ratio will be an increasing function of the statistic $s(\boldsymbol{X}) = -r(\boldsymbol{X})$.

Thus, in terms of ζ, the hypotheses have the same form as the hypotheses (9.3.8) and $g_n(\underset{\sim}{x} \mid \zeta)$ has a monotone likelihood ratio in the statistic $s(\boldsymbol{X})$. Therefore, by Theorem 9.3.1, a test which rejects H_0 when $s(\boldsymbol{X}) \geq c'$, for some constant c', will be a UMP test. But $s(\boldsymbol{X}) \geq c'$ if and only if $T = r(\boldsymbol{X}) \leq c$, where $c = -c'$. Therefore, the test which rejects H_0 when $T \leq c$ will be a UMP test. If c is chosen to satisfy the relation given in the exercise, then it follows from Theorem 9.3.1 that level of significance of this test will be α_0.

13. (a) By Exercise 12, the test which rejects H_0 when $\overline{X}_n \leq c$ will be a UMP test. For the level of significance to be 0.1, c should be chosen so that $\Pr(\overline{X}_n \leq c \mid \mu = 10) = 0.1$. In this exercise, $n = 4$. When $\mu = 10$, the random variable $z = 2(\overline{X}_n - 10)$ has a standard normal distribution and $\Pr(\overline{X}_n \leq c \mid \mu = 10) = \Pr[Z \leq 2(c - 10)]$. It is found from a table of the standard normal distribution that $\Pr(Z \leq -1.282) = 0.1$. Therefore, $2(c - 10) = -1.282$ or $c = 9.359$.

    (b) When $\mu = 9$, the random variable $2(\overline{X}_n - 9)$ has the standard normal distribution. Therefore, the power of the test is

    $$\Pr(\overline{X}_n \leq 9.359 \mid \mu = 9) = \Pr(Z \leq 0.718) = \Phi(0.718) = 0.7636,$$

    where we have interpolated in the table of the normal distribution between 0.71 and 0.72.

(c) When $\mu = 11$, the random variable $Z = 2(\overline{X}_n - 11)$ has the standard normal distribution. Therefore, the probability of rejecting H_0 is

$$\Pr(\overline{X}_n \geq 93359 \,|\, \mu = 11) = \Pr(Z \geq -3.282) = \Pr(Z \leq 3.282) = \Phi(3.282) = 0.9995.$$

14. By Exercise 12, a test which rejects H_0 when $\sum_{i=1}^{n} X_i \leq c$ will be a UMP test. When $n = 10$ and $\lambda = 1, \sum_{i=1}^{n} X_i$ will have a Poisson distribution with mean 10 and $\alpha_0 = \Pr\left(\sum_{i=1}^{n} X_i \leq c \,|\, \lambda = 1\right)$. From a table of the Poisson distribution, the following values of α_0 are obtained.

$$\begin{aligned}
c &= 0, \alpha_0 = .0000; \\
c &= 1, \alpha_0 = .0000 + .0005 = .0005; \\
c &= 2, \alpha_0 = .0000 + .0005 + .0023 = .0028; \\
c &= 3, \alpha_0 = .0000 + .0005 + .0023 + .0076 = .0104; \\
c &= 4, \alpha_0 = .0000 + .0005 + .0023 + .0076 + .0189 = .0293.
\end{aligned}$$

For larger values of $c, \alpha_0 > 0.03$.

15. By Exercise 4, the joint p.d.f. of $X_1, \ldots, X_n$ has a monotone likelihood ratio in the statistic $-\overline{X}_n$. Therefore, by Exercise 12, a test which rejects H_0 when $-\overline{X}_n \leq c'$, for some constant c', will be a UMP test. But this test is equivalent to a test which rejects H_0 when $\overline{X}_n \geq c$, where $c = -c'$. Since $\overline{X}_n$ has a continuous distribution, for any specified value of $\alpha_0 (0 < \alpha_0 < 1)$ there exists a value of c such that $\Pr(\overline{X}_n \geq c \,|\, \beta = 1/2) = \alpha_0$.

16. We must find a constant c such that when $n = 10, \Pr(\overline{X}_n \geq c \,|\, \beta = 1/2) = 0.05$. When $\beta = 1/2$, each observation X_i has an exponential distribution with $\beta = 1/2$, which is a gamma distribution with parameters $\alpha = 1$ and $\beta = 1/2$. Therefore, $\sum_{i=1}^{n} X_i$ has a gamma distribution with parameters $\alpha = n = 10$ and $\beta = 1/2$, which is a χ^2 distribution with $2n = 20$ degrees of freedom. But

$$\Pr\left(\overline{X}_n \geq c \,|\, \beta = \frac{1}{2}\right) = \Pr\left(\sum_{i=1}^{n} X_i \geq 10c \,|\, \beta = \frac{1}{2}\right).$$

It is found from a table of the χ^2 distribution with 20 degrees of freedom that $\Pr(\sum_{i=1}^{n} X_i \geq 31.41) = 0.05$. Therefore, $10c = 31.41$ and $c = 3.141$.

17. In this exercise, H_0 is a simple hypothesis. By the Neyman-Pearson lemma, the test which has maximum power at a particular alternative value $\theta_1 > 0$ will reject H_0 if $f(x \,|\, \theta = \theta_1)/f(x \,|\, \theta = 0) > c$, where c is chosen so that the probability that this inequality will be satisfied when $\theta = 0$ is α_0. Here,

$$\frac{f(x \,|\, \theta = \theta_1)}{f(x \,|\, \theta = 0)} > c$$

if and only if $(1 - c)x^2 + 2c\theta_1 x > c\theta_1^2 - (1 - c)$. For each value of θ_1, the value of c is to be chosen so that the set of points satisfying this inequality has probability α_0 when $\theta = 0$. For two different values of θ_1, these two sets will be different. Therefore, different test procedures will maximize the power at the two different values of θ_1. Hence, no single test is a UMP test.

18. The UMP test will reject H_0 when $\overline{X}_n \geq c$, where $\Pr(\overline{X}_n \geq c \,|\, \mu = 0) = \Pr(\sqrt{n}\,\overline{X}_n \geq \sqrt{n}c \,|\, \mu = 0) = 0.025$. However, when $\mu = 0$, $\sqrt{n}\,\overline{X}_n$ has the standard normal distribution. Therefore, $\Pr(\sqrt{n}\,\overline{X}_n \geq 1.96 \,|\, \mu = 0) = 0.025$. It follows that $\sqrt{n}c = 1.96$ and $c = 1.96n^{-1/2}$.

(a) When $\mu = 0.5$, the random variable $Z = \sqrt{n}(\overline{X}_n - 0.5)$ has the standard normal distribution. Therefore,

$$\begin{aligned}\pi(0.5 \,|\, \delta^*) &= \Pr(\overline{X}_n \geq 1.96n^{-1/2} \,|\, \mu = 0.5) = \Pr(Z \geq 1.96 - 0.5n^{1/2}) \\ &= \Pr(Z \leq 0.5n^{1/2} - 1.96) = \Phi(0.5n^{1/2} - 1.96).\end{aligned}$$

But $\Phi(1.282) = 0.9$. Therefore, $\pi(0.5 \,|\, \delta^*) \geq 0.9$ if and only if $0.5n^{1/2} - 1.96 \geq 1.282$, or, equivalently, if and only if $n \geq 42.042$. Thus, a sample of size $n = 43$ is required in order to have $\pi(0.5 \,|\, \delta^*) \geq 0.9$. Since the power function is a strictly increasing function of μ, it will then also be true that $\pi(0.5 \,|\, \delta^*) \geq 0.9$ for $\mu > 0.5$.

(b) When $\mu = -0.1$, the random variable $Z = \sqrt{n}(\overline{X}_n + 0.1)$ has the standard normal distribution. Therefore,

$$\begin{aligned}\pi(-0.1 \,|\, \delta^*) &= \Pr(\overline{X}_n \geq 1.96n^{-1/2} \,|\, \mu = -0.1) = \Pr(Z \geq 1.96 + 0.1n^{1/2}) \\ &= 1 - \Phi(1.96 + 0.1n^{1/2}).\end{aligned}$$

But $\Phi(3.10) = 0.999$. Therefore, $\pi(-0.1 \,|\, \delta^*) \leq 0.001$ if and only if $1.96 + 0.1n^{1/2} \geq 3.10$ or, equivalently, if and only if $n \geq 129.96$. Thus, a sample of size n = 130 is required in order to have $\pi(-0.1 \,|\, \delta^*) \leq 0.001$. Since the power function is a strictly increasing function of μ, it will then also be true that $\pi(\mu \,|\, \delta^*) \leq 0.001$ for $\mu < -0.1$.

19. (a) Let $f(\boldsymbol{x}|\mu)$ be the joint p.d.f. of $\boldsymbol{X}$ given μ. For each set A and $i = 0, 1$,

$$\Pr(\boldsymbol{X} \in A|\mu = \mu_i) = \int \underset{A}{\cdots} \int f(\boldsymbol{x}|\mu_i)d\boldsymbol{x}. \tag{S.9.7}$$

It is clear that $f(\boldsymbol{x}|\mu_0) > 0$ for all $\boldsymbol{x}$ and so is $f(\boldsymbol{x}|\mu_1) > 0$ for all $\boldsymbol{x}$. Hence (S.9.7) is strictly positive for $i = 0$ if and only if it is strictly positive for $i = 1$.

(b) Both δ and δ_1 are size α_0 tests of $H_0' : \mu = \mu_0$ versus $H_1' : \mu > \mu_0$. Let

$$\begin{aligned}A &= \{\boldsymbol{x} : \delta \text{ rejects but } \delta_1 \text{ does not reject}\}, \\ B &= \{\boldsymbol{x} : \delta \text{ does not reject but } \delta_1 \text{ rejects}\}, \\ C &= \{\boldsymbol{x} : \text{ both tests reject}\}.\end{aligned}$$

Because both tests have the same size, it must be the case that

$$\Pr(\boldsymbol{X} \in A|\mu = \mu_0) + \Pr(\boldsymbol{X} \in |\mu = \mu_0) = \alpha_0 = \Pr(\boldsymbol{X} \in B|\mu = \mu_0) + \Pr(\boldsymbol{X} \in |\mu = \mu_0).$$

Hence,

$$\Pr(\boldsymbol{X} \in A|\mu = \mu_0) = Pr(\boldsymbol{X} \in B|\mu = \mu_0). \tag{S.9.8}$$

Because of the MLR and the form of the test δ_1, we know that there is a constant c such that for every $\mu > \mu_0$ and every $\boldsymbol{x} \in B$ and every $\boldsymbol{y} \in A$,

$$\frac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu_0)} > c > \frac{f(\boldsymbol{y}|\mu)}{f(\boldsymbol{y}|\mu_0)}. \tag{S.9.9}$$

Now,

$$\pi(\mu|\delta) = \int \underset{A}{\cdots} \int f(\boldsymbol{x}|\mu)d\boldsymbol{x} + \int \underset{C}{\cdots} \int f(\boldsymbol{x}|\mu)d\boldsymbol{x}.$$

Also,

$$\pi(\mu|\delta_1) = \int \underset{B}{\cdots} \int f(\boldsymbol{x}|\mu)d\boldsymbol{x} + \int \underset{C}{\cdots} \int f(\boldsymbol{x}|\mu)d\boldsymbol{x}.$$

It follows that, for $\mu > \mu_0$,

$$\begin{aligned}
\pi(\mu|\delta_1) - \pi(\mu|\delta) &= \int \underset{B}{\cdots} \int f(\boldsymbol{x}|\mu)d\boldsymbol{x} - \int \underset{A}{\cdots} \int f(\boldsymbol{x}|\mu)d\boldsymbol{x} \\
&= \int \underset{B}{\cdots} \int \frac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu_0)} f(\boldsymbol{x}|\mu_0)d\boldsymbol{x} - \int \underset{A}{\cdots} \int \frac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu_0)} f(\boldsymbol{x}|\mu_0)d\boldsymbol{x} \\
&> \int \underset{B}{\cdots} \int c f(\boldsymbol{x}|\mu_0)d\boldsymbol{x} - \int \underset{A}{\cdots} \int c f(\boldsymbol{x}|\mu_0)d\boldsymbol{x} = 0,
\end{aligned}$$

where the inequality follows from (S.9.9), and the final equality follows from (S.9.8).

9.4 Two-Sided Alternatives

Commentary

This section considers tests for simple (and interval) null hypotheses against two-sided alternative hypotheses. The concept of unbiased tests is introduced in a subsection at the end. Even students in a mathematical statistics course may have trouble with the concept of unbiased test.

Solutions to Exercises

1. If $\pi(\mu \,|\, \delta)$ is to be symmetric with respect to the point $\mu = \mu_0$, then the constants c_1 and c_2 must be chosen to be symmetric with respect to the value μ_0. Let $c_1 = \mu_0 - k$ and $c_1 = \mu_0 + k$. When $\mu = \mu_0$, the random variable $Z = n^{1/2}(\overline{X}_n - \mu_0)$ has the standard normal distribution. Therefore,

$$\begin{aligned}
\pi(\mu_0 \,|\, \delta) &= \Pr(\overline{X}_n \le \mu_0 - k \,|\, \mu_0) + \Pr(\overline{X}_n \ge \mu_0 + k \,|\, \mu_0) \\
&= \Pr(Z \le -n^{1/2}k) + \Pr(Z \ge n^{1/2}k) \\
&= 2\Pr(Z \ge n^{1/2}k) = 2[1 - \Phi(n^{1/2}k)].
\end{aligned}$$

Since k must be chosen so that $\pi(\mu_0 \,|\, \delta) = 0.10$, it follows that $\Phi(n^{1/2}k) = 0.95$. Therefore, $n^{1/2}k = 1.645$ and $k = 1.645n^{-1/2}$.

2. When $\mu = \mu_0$, the random variable $Z = n^{1/2}(\overline{X}_n - \mu_0)$ has the standard normal distribution. Therefore,

$$\begin{aligned}
\pi(\mu_0 \,|\, \delta) &= \Pr(\overline{X}_n \le c_1 \,|\, \mu_0) + \Pr(\overline{X}_n \ge c_2 \,|\, \mu_0) \\
&= \Pr(Z \le -1.96) + \Pr\left[Z \ge n^{1/2}(c_2 - \mu_0)\right] \\
&= \Phi(-1.96) + 1 - \Phi[n^{1/2}(c_2 - \mu_0)] \\
&= 1.025 - \Phi[n^{1/2}(c_2 - \mu_0)].
\end{aligned}$$

If we are to have $\pi(\mu_0 \,|\, \delta) = 0.10$, then we must have $\Phi[n^{1/2}(c_2 - \mu_0)] = 0.925$. Therefore, $n^{1/2}(c_2 - \mu_0) = 1.439$ and $c_2 = \mu_0 + 1.439n^{-1/2}$.

3. From Exercise 1, we know that if $c_1 = \mu_0 - 1.645n^{-1/2}$ and $c_2 = \mu_0 + 1.645n^{-1/2}$, then $\pi(\mu_0 \,|\, \delta) = 0.10$ and, by symmetry, $\pi(\mu_0 + 1 \,|\, \delta) = \pi(\mu_0 - 1 \,|\, \delta)$. Also, when $\mu = \mu_0 + 1$, the random variable

$n^{1/2}(\overline{X}_n - \mu_0 - 1)$ has the standard normal distribution. Therefore,

$$\begin{aligned}\pi(\mu_0+1\,|\,\delta) &= \Pr(\overline{X}_n \le c_1\,|\,\mu_0+1) + \Pr(\overline{X}_n \ge c_2\,|\,\mu_0+1)\\ &= \Pr(Z \le -1.645 - n^{1/2}) + \Pr(Z \ge 1.645 - n^{1/2})\\ &= \Phi(-1.645 - n^{1/2}) + \Phi(n^{1/2} - 1.645).\end{aligned}$$

$$\begin{aligned}\text{For } n = 9, &\quad \pi(\mu_0+1\,|\,\delta) = \Phi(-4.645) + \Phi(1.355) < 0.95.\\ \text{For } n = 10, &\quad \pi(\mu_0+1\,|\,\delta) = \Phi(-4.807) + \Phi(1.517) < 0.95.\\ \text{For } n = 11, &\quad \pi(\mu_0+1\,|\,\delta) = \Phi(-4.962) + \Phi(1.672) > 0.95.\end{aligned}$$

4. If we choose c_1 and c_2 to be symmetric with respect to the value 0.15, then it will be true that $\pi(0.1\,|\,\delta) = \pi(0.2\,|\,\delta)$. Accordingly, let $c_1 = 0.15 - k$ and $c_2 = 0.15 + k$. When $\mu = 0.1$, the random variable $Z = 5(\overline{X}_n - 0.1)$ has a standard normal distribution. Therefore,

$$\begin{aligned}\pi(0.1\,|\,\delta) &= \Pr(\overline{X}_n \le c_1\,|\,0.1) + \Pr(\overline{X}_n \ge c_2\,|\,0.1)\\ &= \Pr(Z \le 0.25 - 5k) + \Pr(Z \ge 0.25 + 5k)\\ &= \Phi(0.25 - 5k) + \Phi(-0.25 - 5k).\end{aligned}$$

We must choose k so that $\pi(0.1\,|\,\delta) = 0.07$. By trial and error, using the table of the standard normal distribution, we find that when $5k = 1.867$,

$$\pi(0.1\,|\,\delta) = \Phi(-1.617) + \Phi(-2.117) = 0.0529 + 0.0171 = 0.07.$$

Hence, $k = 0.3734$.

5. As in Exercise 4,

$$\pi(0.1\,|\,\delta) = \Pr\left[Z \le 5(c_1 - 0.1)\right] + \Pr\left[Z \ge 5(c_2 - 0.1)\right] = \Phi(5c_1 - 0.5) + \Phi(0.5 - 5c_2).$$

Similarly,

$$\pi(0.2\,|\,\delta) = \Pr\left[Z \le 5(c_1 - 0.2)\right] + \Pr\left[Z \ge 5(c_2 - 0.2)\right] = \Phi(5c_1 - 1) + \Phi(1 - 5c_2).$$

Hence, the following two equations must be solved simultaneously:

$$\begin{aligned}\Phi(5c_1 - 0.5) + \Phi(0.5 - 5c_2) &= 0.02,\\ \Phi(5c_1 - 1) + \Phi(1 - 5c_2) &= 0.05.\end{aligned}$$

By trial and error, using the table of the standard normal distribution, it is found ultimately that if $5c_1 = -2.12$ and $5c_2 = 2.655$, then

$$\Phi(5c_1 - 0.5) + \Phi(0.5 - 5c_2) = \Phi(-2.62) + \Phi(-2.155) = 0.0044 + 0.0155 = 0.02.$$

and

$$\Phi(5c_1 - 0.1) + \Phi(1 - 5c_2) = \Phi(-3.12) + \Phi(-1.655) = 0.0009 + 0.0490 = 0.05.$$

6. Let $T = \max(X_1, \ldots, X_n)$. Then

$$f_n(\boldsymbol{X} \mid \theta) = \begin{cases} \dfrac{1}{\theta^n} & \text{for } 0 \le t \le \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, for $\theta_1 < \theta_2$,

$$\frac{f_n(\boldsymbol{X} \mid \theta_2)}{f_n(\boldsymbol{X} \mid \theta_1)} = \begin{cases} \left(\dfrac{\theta_1}{\theta_2}\right)^n & \text{for } 0 \le t \le \theta_1, \\ \infty & \text{for } \theta_1 < t \le \theta_2. \end{cases}$$

It can be seen from this relationship that $f_n(\boldsymbol{X} \mid \theta)$ has a monotone likelihood ratio in the statistic T (although we are being somewhat nonrigorous by treating ∞ as a number).

For any constant c $(0 < c \le 3)$, $\Pr(T \ge c \mid \theta = 3) = 1 - (c/3)^n$. Therefore, to achieve a given level of significance α_0, we should choose $c = 3(1-\alpha_0)^{1/n}$. It follows from Theorem 9.3.1 that the corresponding test will be a UMP test.

7. For $\theta > 0$, the power function is $\pi(\theta \mid \delta) = \Pr(T \ge c \mid \theta)$. Hence,

$$\pi(\theta \mid \delta) = \begin{cases} 0 & \text{for } \theta \le c, \\ 1 - \left(\dfrac{c}{\theta}\right)^n & \text{for } \theta > c. \end{cases}$$

The plot is in Fig. S.9.4.

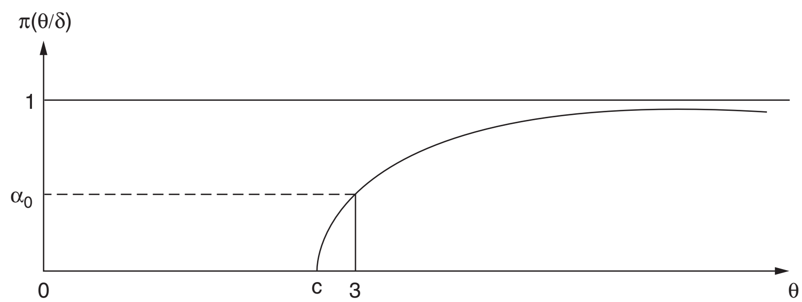

Figure S.9.4: Figure for Exercise 7 of Sec. 9.4.

8. (a) It follows from Exercise 8 of Sec. 9.3 that the specified test will be a UMP test.
   (b) For any given value of c $(0 < c < 3)$, $\Pr(T \le c \mid \theta = 3) = (c/3)^n$. Therefore, to achieve a given level of significance α_0, we should choose $c = 3\alpha_0^{1/n}$.

9. A sketch is given in Fig. S.9.5.

10. (a) Let $\alpha_0 = 0.05$ and let $c_1 = 3\alpha_0^{1/n}$ as in Exercise 8. Also, let $c_2 = 3$. Then

$$\pi(\theta \mid \delta) = \Pr(T \le 3\alpha_0^{1/n} \mid \theta) + \Pr(T \ge 3 \mid \theta).$$

Since $\Pr(T \ge 3 \mid \theta) = 0$ for $\theta \le 3$, the function $\pi(\theta \mid \delta)$ is as sketched in Exercise 10 for $\theta \le 3$. For $\theta > 3$,

$$\pi(\theta \mid \delta) = \left[\frac{3\alpha_0^{1/n}}{\theta}\right]^n + \left[1 - \left(\frac{3}{\theta}\right)^n\right] > \alpha_0. \tag{S.9.10}$$

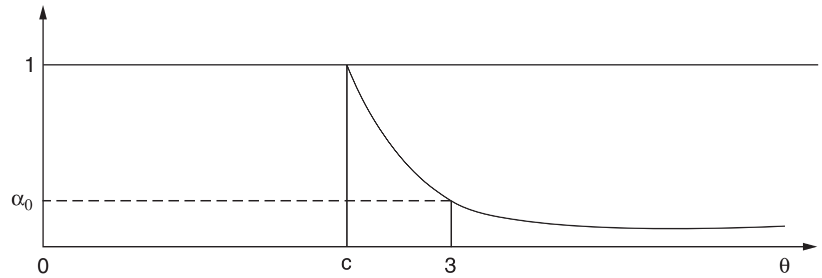

Figure S.9.5: Figure for Exercise 9 of Sec. 9.4.

(b) In order for a test δ to be UMP level α_0, for (9.4.15), it necessary and sufficient that the following three things happen:

- δ has the same power function as the test in Exercise 6 for $\theta > 3$.
- δ has the same power function as the test in Exercise 8 for $\theta < 3$.
- $\pi(3|\delta) \leq \alpha_0$.

Because $c_2 = 3$, we saw in part (a) that $\pi(\theta|\delta)$ is the same as the power function of the test in Exercise 8 for $\theta < 3$. We also saw in part (a) that $\pi(3|\delta) = 0.05 = \alpha_0$. For $\theta > 3$, the power function of the test in Exercise 6 is

$$\Pr(T \geq 3(1-\alpha_0)^{1/n}|\theta) = 1 - \left(\frac{3(1-\alpha_0)^{1/n}}{\theta}\right)^n = 1 - \left(\frac{3}{\theta}\right)^n (1-\alpha_0).$$

It is straightforward to see that this is the same as (S.9.10).

11. It can be verified that if c_1 and c_2 are chosen to be symmetric with respect to the value μ_0, then the power function $\pi(\mu \,|\, \delta)$ will be symmetric with respect to the point $\mu = \mu_0$ and will attain its minimum value at $\mu = \mu_0$. Therefore, if c_1 and c_2 are chosen as in Exercise 1, the required conditions will be satisfied.

12. The power function of the test δ described in this exercise is

$$\pi(\beta|\delta) = 1 - \exp(-c_1\beta) + \exp(-c2\beta).$$

(a) In order for δ to have level of significance α_0, we must have $\pi(1|\delta) \leq \alpha_0$. Indeed, the test will have size α_0 exactly if

$$\alpha_0 = 1 - \exp(-c_1) + \exp(-c_2).$$

(b) We can let $c_1 = -\log(1-\alpha_0/2)$ and $c_2 = -\log(\alpha_0/2)$ to solve this equation.

13. The first term on the right of (9.4.13) is

$$\frac{n}{\theta}\int_0^x \frac{\theta^n}{\Gamma(n)} t^{n-1}\exp(-t\theta)dt = \frac{n}{\theta}G(x; n, \theta).$$

The second term on the right of (9.4.13) is the negative of

$$\frac{n}{\theta}\int_0^x \frac{\theta^{n+1}}{\Gamma(n+1)} t^{n}\exp(-t\theta)dt = \frac{n}{\theta}G(x; n+1, \theta).$$

9.5 The t Test

Commentary

This section provides a natural continuation to Sec. 9.1 in a modern statistics course. We introduce the t test and its power function, defined in terms of the noncentral t distribution. The theoretical derivation of the t test as a likelihood ratio test is isolated at the end of the section and could easily be skipped without interrupting the flow of material. Indeed, that derivation should only be of interest in a fairly mathematical statistics course.

As with confidence intervals, computer software can replace tables for obtaining quantiles of the t distributions that are used in tests. The R function `qt` can compute these. For computing p-values, one can use `pt`. The precise use of `pt` depends on whether the alternatoive hypothesis is one-sided or two-sided. For testing $H_0 : \mu \le \mu_0$ versus $H_1 : \mu > \mu_0$ using the statistic U in Eq. (9.5.2), the p-value would be `1-pt(u,n-1)`, where `u` is the observed value of U. Fot the opposite one-sided hypotheses, the p-value would be `pt(u,n-1)`. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \ne \mu_0$, the p-value is `2*(1-pt(abs(u),n-1))`. The power function of a t test can be computed using the optional third parameter with `pt`, which is the noncentrality parameter (whose default value is 0). Similar considerations apply to the comparison of two means in Sec. 9.6.

Solutions to Exercises

1. We computed the summary statistics $\overline{x}_n = 1.379$ and $\sigma' = 0.3277$ in Example 8.5.4.

   (a) The test statistic is U from (9.5.2)
   $$U = 10^{1/2}\frac{1.379 - 1.2}{0.3277} = 1.727.$$
   We reject H_0 at level $\alpha_0 = 0.05$ if $U \ge 1.833$, the 0.95 quantile of the t distribution with 9 degrees of freedom. Since $1.727 \not\ge 1.833$, we do not reject H_0 at level 0.05.

   (b) We need to compute the probability that a t random variable with 9 degrees of freedom exceeds 1.727. This probability can be computed by most statistical software, and it equals 0.0591. Without a computer, one could interpolate in the table of the t distribution in the back of the book. That would yield 0.0618.

2. When $\mu_0 = 20$, the statistic U given by Eq. (9.5.2) has a t distribution with 8 degrees of freedom. The value of U in this exercise is 2.

   (a) We would reject H_0 if $U \ge 1.860$. Therefore, we reject H_0.

   (b) We would reject H_0 if $U \le -2.306$ or $U \ge 2.306$. Therefore, we don't reject H_0.

   (c) We should include in the confidence interval, all values of μ_0 for which the value of U given by Eq. (9.5.2) will lie between -2.306 and 2.306. These values form the interval $19.694 < \mu_0 < 24.306$.

3. It must be assumed that the miles per gallon obtained from the different tankfuls are independent and identically distributed, and that each has a normal distribution. When $\mu_0 = 20$, the statistic U given by Eq. (9.5.2) has the t distribution with 8 degrees of freedom. Here, we are testing the following hypotheses:
   $$H_0 : \mu \ge 20,$$
   $$H_1 : \mu < 20.$$
   We would reject H_0 if $U \le -1.860$. From the given value, it is found that $\overline{X}_n = 19$ and $S_n^2 = 22$. Hence, $U = -1.809$ and we do not reject H_0.

4. When $\mu_0 = 0$, the statistic U given by Eq. (9.5.2) has the t distribution with 7 degrees of freedom. Here

$$\overline{X}_n = \frac{-11.2}{8} = -1.4$$

and

$$\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = 43.7 - 8(1.4)^2 = 28.02.$$

The value of U can now be found to be -1.979. We should reject H_0 if $U \leq -1.895$ or $U \geq 1.895$. Therefore, we reject H_0.

5. It is found from the table of the t distribution with 7 degrees of freedom that $c_1 = -2.998$ and c_2 lies between 1.415 and 1.895. Since U $= -1.979$, we do not reject H_0.

6. Let U be given by Eq. (9.5.2) and suppose that c is chosen so that the level of significance of the test is α_0. Then

$$\pi(\mu, \sigma^2 \mid \delta) = \Pr(U \geq c \mid \mu, \sigma^2).$$

If we let $Y = n^{1/2}(\overline{X}_n - \mu)/\sigma$ and $Z = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2/\sigma^2$, then Y will have a standard normal distribution, Z will have a χ^2 distribution with $n-1$ degrees of freedom, and Y and Z will be independent. Also,

$$U = \frac{Y + n^{1/2}\left(\dfrac{\mu - \mu_0}{\sigma}\right)}{[Z/(n-1)]^{1/2}}.$$

It follows that all pairs (μ, σ^2) which yield the same value of $(\mu - \mu_0)/\sigma$ will yield the same value of $\pi(\mu, \sigma^2 \mid \delta)$.

7. The random variable $T = (X - \mu)/\sigma$ will have the standard normal distribution, the random variable $Z = \sum_{i=1}^{n} Y_i^2/\sigma^2$ will have a χ^2 distribution with n degrees of freedom, and T and Z will be independent. Therefore, when $\mu = \mu_0$, the following random variable U will have the t distribution with n degrees of freedom:

$$U = \frac{T}{[Z/n]^{1/2}} = \frac{n^{1/2}(X - \mu_0)}{\left[\sum_{i=1}^{n} Y_i^2\right]^{1/2}}.$$

The hypothesis H_0 would be rejected if $U \geq c$.

8. When $\sigma^2 = \sigma_0^2, S_n^2/\sigma_0^2$ has a χ^2 distribution with $n-1$ degrees of freedom. Choose c so that, when $\sigma^2 = \sigma_0^2, \Pr(S_n^2/\sigma_0^2 \geq c) = \alpha_0$, and reject H_0 if $S_n^2/\sigma_0^2 \geq c$. Then $\pi(\mu, \sigma^2 \mid \delta) = \alpha_0$ if $\sigma^2 = \sigma_0^2$. If $\sigma^2 \neq \sigma_0^2$, then $Z = S_n^2/\sigma^2$ has the χ^2 distribution with $n-1$ degrees for freedom, and $S_n^2/\sigma_0^2 = (\sigma^2/\sigma_0^2)T$. Therefore,

$$\pi(\mu, \sigma^2 \mid \delta) = \Pr(S_n^2/\sigma_0^2 \geq c \mid \mu, \sigma^2) = \Pr(T \geq c\sigma_0^2/\sigma^2).$$

If $\sigma_0^2/\sigma^2 > 1$, then $\pi(\mu, \sigma^2 \mid \delta) < \Pr(T \geq c) = \alpha_0$. If $\sigma_0^2/\sigma^2 < 1$, then $\pi(\mu, \sigma^2 \mid \delta) > \Pr(T \geq c) = \alpha_0$.

9. When $\sigma^2 = 4, S_n^2/4$ has the χ^2 distribution with 9 degrees of freedom. We would reject H_0 if $S_n^2/4 \ge 16.92$. Since $S_n^2/4 = 60/4 = 15$, we do not reject H_0.

10. When $\sigma^2 = 4, S_n^2/4$ has the χ^2 distribution with 9 degrees of freedom. Therefore, $\Pr(S_n^2/4 < 2.700) = \Pr(S_n^2/4 > 19.02) = 0.025$. It follows that $c_1 = 4(2.700) = 10.80$ and $c_2 = 4(19.02) = 76.08$.

11. U_1 has the distribution of X/Y where X has a normal distribution with mean ψ and variance 1, and Y is independent of X such that mY^2 has the χ^2 distribution with m degrees of freedom. Notice that $-X$ has a normal distribution with mean $-\psi$ and variance 1 and is independent of Y. So U_2 has the distribution of $-X/Y = -U_1$. So

$$\Pr(U_2 \le -c) = \Pr(-U_1 \le -c) = \Pr(U_1 \ge c).$$

12. The statistic U has the t distribution with 16 degrees of freedom. The calculated value is

$$U = \frac{\sqrt{17}(\overline{X}_n - 3)}{[S_n^2/16]^{1/2}} = \frac{0.2}{(0.09/16)^{1/2}} = \frac{8}{3}$$

and the corresponding tail area is $\Pr(U > 8/3)$.

13. The test statistic is $U = 169^{1/2}(3.2 - 3)/(0.09)^{1/2} = 8.667$. The p-value can be calculated using statistical software as $1 - T_{169}(8.667) = 1.776 \times 10^{-15}$.

14. The statistic U has the t distribution with 16 degrees of freedom. The calculated value of U is

$$U = \frac{0.1}{(0.09/16)^{1/2}} = \frac{4}{3}.$$

Because the alternative hypothesis is two-sided, the corresponding tail area is

$$\Pr\left(U \ge \frac{4}{3}\right) + \Pr\left(U \le -\frac{4}{3}\right) = 2\Pr\left(U \ge \frac{4}{3}\right).$$

15. The test statistic is $U = 169^{1/2}(3.2 - 3.1)/(0.09)^{1/2} = 4.333$. The p-value can be calculated using statistical software as $2[1 - T_{169}(4.333)] = 2.512 \times 10^{-5}$.

16. The calculated value of U is

$$U = \frac{-0.1}{(0.09/16)^{1/2}} = -\frac{4}{3}.$$

Since this value is the negative of the value found in Exercise 14, the corresponding tail area will be the same as in Exercise 14.

17. The denominator of $\Lambda(\boldsymbol{x})$ is still (9.5.11). The M.L.E. $(\hat{\mu}_0, \hat{\sigma}_0^2)$ is easier to calculate in this exercise, namely $\hat{\mu}_0 = \mu_0$ (the only possible value) and

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

These are the same values that lead to Eq. (9.5.12) in the text. Hence, $\Lambda(\boldsymbol{x})$ has the value given in Eq. (9.5.14). For $k < 1$, $\Lambda(\boldsymbol{x}) \le k$ if and only if

$$|U| \ge ((n-1)[k^{2/n} - 1])^{1/2} = c.$$

18. In this case $\Omega_0 = \{(\mu, \sigma^2) : \mu \geq \mu_0\}$, and $\Lambda(\boldsymbol{x}) = 1$ if $\overline{x}_n \geq \mu_0$. If $\overline{x}_n < \mu_0$, then the numerator of $\Lambda(\boldsymbol{x})$ is (9.5.12), and the formula for $\Lambda(\boldsymbol{x})$ is the same as (9.5.13) with the branch labels switched. This time, $\Lambda(\boldsymbol{x})$ is a non-decreasing function of u, the observed value of U. So for $k < 1$, $\Lambda(\boldsymbol{x}) \leq k$ if and only if $U \leq c$, for the same c as in Example 9.5.12.

9.6 Comparing the Means of Two Normal Distributions

Commentary

The two-sample t test is introduced for the case of equal variances. There is some material near the end of the section about the case of unequal variances. This is useful material, but is not traditionally covered and can be skipped. Also, the derivation of the two-sample t test as a likelihood ratio test is provided for mathematical interest at the end of the section.

Solutions to Exercises

1. In this example, $n = 5$, $m = 5$, $\overline{X}_m = 18.18$, $\overline{Y}_n = 17.32$, $S_X^2 = 12.61$, and $S_Y^2 = 11.01$. Then

$$U = \frac{(5+5-2)^{1/2}(18.18 - 17.32)}{\left(\frac{1}{5}+\frac{1}{5}\right)^{1/2}(11.01 + 12.61)^{1/2}} = 0.7913.$$

We see that $|U| = 0.7913$ is much smaller than the 0.975 quantile of the t distribution with 8 degrees of freedom.

2. In this exercise, $m = 8$, $n = 6$, $\overline{x}_m = 1.5125$, $\overline{y}_n = 1.6683$, $S_X^2 = 0.18075$, and $S_Y^2 = 0.16768$. When $\mu_1 = \mu_2$, the statistic U defined by Eq. (9.6.3) will have the t distribution with 12 degrees of freedom. The hypotheses are as follows:

$$\begin{aligned} H_0 : &\quad \mu_1 \geq \mu_2, \\ H_1 : &\quad \mu_1 < \mu_2. \end{aligned}$$

Since the inequalities are reversed from those in (9.6.1), the hypothesis H_0 should be rejected if $U < c$. It is found from a table that $c = -1.356$. The calculated value of U is -1.692. Therefore, H_0 is rejected.

3. The value $c = 1.782$ can be found from a table of the t distribution with 12 degrees of freedom. Since $U = -1.692$, H_0 is not rejected.

4. The random variable $\overline{X}_m - \overline{Y}_n$ has a normal distribution with mean 0 and variance $(\sigma_1^2/m) + (k\sigma_1^2/n)$. Therefore, the following random variable has the standard normal distribution:

$$Z_1 = \frac{\overline{X}_m - \overline{Y}_n}{\left(\dfrac{1}{m} + \dfrac{k}{n}\right)^{1/2} \sigma_1}.$$

The random variable S_X^2/σ_1^2 has a χ^2 distribution with m-1 degrees of freedom. The random variable $S_Y^2/(k\sigma_1^2)$ has a χ^2 distribution with $n-1$ degrees of freedom. These two random variables are independent. Therefore, $Z_2 = (1/\sigma_1^2)(S_X^2 + S_Y^2/k)$ has a χ^2 distribution with $m+n-2$ degrees of freedom. Since Z_1 and Z_2 are independent, it follows that $U = (m+n-2)^{1/2} Z_1/Z_2^{1/2}$ has the t distribution with $m+n-2$ degrees of freedom.

5. Again, H_0 should be rejected if $U < -1.356$. Since $U = -1.672, H_0$ is rejected.

6. If $\mu_1 - \mu_2 = \lambda$, the following statistic U will have the t distribution with $m+n-2$ degrees of freedom:

$$U = \frac{(m+n-2)^{1/2}(\overline{X}_m - \overline{Y}_n - \lambda)}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}(S_X^2 + S_Y^2)^{1/2}}.$$

The hypothesis H_0 should be rejected if either $U < c_1$ or $U > c_2$.

7. To test the hypotheses in Exercise 6, H_0 would not be rejected if $-1.782 < U < 1.782$. The set of all values of λ for which H_0 would not be rejected will form a confidence interval for $\mu_1 - \mu_2$ with confidence coefficient 0.90. The value of U, for an arbitrary value of λ, is found to be

$$U = \frac{\sqrt{12}(-0.1558 - \lambda)}{0.3188}.$$

It is found that $-1.782 < U < 1.782$ if and only if $-0.320 < \lambda < 0.008$.

8. The noncentrality parameter when $|\mu_1 - \mu_2| = \sigma$ is

$$\psi = \frac{1}{\left(\frac{1}{8} + \frac{1}{10}\right)^{1/2}} = 2.108.$$

The degrees of freedom are 16. Figure 9.14 in the text makes it look like the power is about 0.23. Using computer software, we can compute the noncentral t probability to be 0.248.

9. The p-value can be computed as the size of the test that rejects H_0 when $|U| \geq |u|$, where u is the observed value of the test statistic. Since U has the t distribution with $m+n-2$ degrees of freedom when H_0 is true, the size of the test that rejects H_0 when $|U| \geq |u|$ is the probability that a t random variable with $m+n-2$ degrees of freedom is either less than $-|u|$ or greater than $|u|$. This probability is

$$T_{m+n-2}(-|u|) + 1 - T_{m+n-2}(|u|) = 2[1 - T_{m+n-2}(|u|),$$

by the symmetry of t distributions.

10. Let X_i stand for an observation in the calcium supplement group and let Y_j stand for and observation in the placebo group. The summary statistics are

$$\begin{aligned} m &= 10, \\ n &= 11, \\ \overline{x}_m &= 109.9, \\ \overline{y}_n &= 113.9, \\ s_x^2 &= 546.9, \\ s_y^2 &= 1282.9. \end{aligned}$$

We would reject the null hypothesis if $U > T_{19}^{-1}(0.9) = 1.328$. The test statistic has the observed value $u = -0.9350$. Since $u < 1.328$, we do not reject the null hypothesis.

11. (a) The observed value of the test statistic U is

$$u = \frac{(43+35-2)^{1/2}(8.560-5.551)}{\left(\frac{1}{43}+\frac{1}{35}\right)^{1/2}(2745.7+783.9)^{1/2}} = 1.939.$$

We would reject the null hypothesis at level $\alpha_0 = 0.01$ if $U > 2.376$, the 0.99 quantile of the t distribution with 76 degrees of freedom. Since $u < 2.376$, we do not reject H_0 at level 0.01. (The answer in the back of the book is incorrect in early printings.)

(b) For Welch's test, the approximate degrees of freedom is

$$\nu = \frac{\left(\frac{2745.7}{43\times 42}+\frac{783.9}{35\times 34}\right)^2}{\frac{1}{42^3}\left(\frac{2745.7}{43}\right)^2+\frac{1}{34^3}\left(\frac{783.9}{35}\right)^2} = 70.04.$$

The corresponding t quantile is 2.381. The test statistic is

$$\frac{8.560-5.551}{\left(\frac{2745.7}{43\times 42}+\frac{783.9}{35\times 34}\right)^{1/2}} = 2.038.$$

Once again, we do not reject H_0. (The answer in the back of the book is incorrect in early printings.) book is incorrect.)

12. The W in (9.6.15) is the sum of two independent random variables, one having a gamma distribution with parameters $(m-1)/2$ and $m(m-1)/(2\sigma_1^2)$ and the other having a gamma distribution with parameters $(n-1)/2$ and $n(n-1)/(2\sigma_2^2)$. So, the mean and variance of W are

$$\begin{aligned} E(W) &= \frac{(m-1)/2}{m(m-1)/(2\sigma_1^2)}+\frac{(n-1)/2}{n(n-1)/(2\sigma_2^2)} = \frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}, \\ \mathrm{Var}(W) &= \frac{(m-1)/2}{m^2(m-1)^2/(4\sigma_1^4)}+\frac{(n-1)/2}{n^2(n-1)^2/(4\sigma_2^4)} = \frac{2\sigma_1^4}{m^2(m-1)}+\frac{2\sigma_2^4}{n^2(n-1)}. \end{aligned}$$

The gamma distribution with parameters α and β has the above mean and variance if $\alpha/\beta = E(W)$ and $\alpha/\beta^2 = \mathrm{Var}(W)$. In particular, $\alpha = E(W)^2/\mathrm{Var}(W)$, so

$$2\alpha = \frac{\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)^2}{\frac{\sigma_1^4}{m^2(m-1)}+\frac{\sigma_2^4}{n^2(n-1)}}.$$

This is easily seen to be the same as the expression in (9.6.16).

13. The likelihood ratio statistic for this case is

$$\Lambda(\boldsymbol{x},\boldsymbol{y}) = \frac{\sup_{\{(\mu_1,\mu_2,\sigma^2):\mu_1\neq\mu_2\}} g(\boldsymbol{x},\boldsymbol{y}\mid\mu_1,\mu_2,\sigma^2)}{\sup_{\{(\mu_1,\mu_2,\sigma^2):\mu_1=\mu_2\}} g(\boldsymbol{x},\boldsymbol{y}\mid\mu_1,\mu_2,\sigma^2)}. \qquad \text{(S.9.11)}$$

Maximizing the numerator of (S.9.11) is identical to maximizing the numerator of (9.6.10) when $\overline{x}_m \le \overline{y}_n$ because we need $\mu_1 = \mu_2$ in both cases. So the M.L.E.'s are

$$\begin{aligned} \hat{\mu}_1 &= \hat{\mu}_2 = \frac{m\overline{x}_m+n\overline{y}_n}{m+n}, \\ \hat{\sigma}^2 &= \frac{mn(\overline{x}_m-\overline{y}_n)^2/(m+n)+s_x^2+s_y^2}{m+n}. \end{aligned}$$

Maximizing the denominator of (S.9.11) is identical to the maximization of the denominator of (9.6.10) when $\overline{x}_m \le \overline{y}_n$. We use the overall M.L.E.'s

$$\hat{\mu}_1 = \overline{x}_m, \quad \hat{\mu}_2 = \overline{y}_n, \text{ and } \hat{\sigma}^2 = \frac{1}{m+n}(s_x^2 + s_y^2).$$

This makes $\Lambda(\boldsymbol{x}, \boldsymbol{y})$ equal to $(1+v^2)^{-(m+n)/2}$ where v is defined in (9.6.12). So $\Lambda(\boldsymbol{x}, \boldsymbol{y}) \ge k$ if and only if $v^2 \le k'$ for some other constant k'. This translates easily to $|U| \ge c$.

9.7 The F Distributions

Commentary

The F distributions are introduced along with the F test for equality of variances from normal samples. The power function of the F test is derived also. The derivation of the F test as a likelihood ratio test is provided for mathematical interest.

Those using the software R can make use of the functions `df`, `pf`, and `qf` which compute respectively the p.d.f., c.d.f., and quantile function of an arbitrary F distribution. The first argument is the argument of the function and the next two are the degrees of freedom. The function `rf` produces a random sample of F random variables.

Solutions to Exercises

1. The test statistic is $V = [2745.7/42]/[783.9/34] = 2.835$. We reject the null hypothesis if V is greater than the 0.95 quantile of the F distribution with 42 and 34 degrees of freedom, which is 1.737. So, we reject the null hypothesis at level 0.05.

2. Let $Y = 1/X$. Then Y has the F distribution with 8 and 3 degrees of freedom. Also

$$\Pr(X > c) = \Pr\left(Y < \frac{1}{c}\right) = 0.975.$$

It can be found from the table given at the end of the book that $\Pr(Y < 14.54) = 0.975$. Therefore, $1/c = 14.54$ and $c = 0.069$.

3. If Y has the t distribution with 8 degrees of freedom, then $X = Y^2$ will have the F distribution with 1 and 8 degrees of freedom. Also,

$$0.3 = \Pr(X > c) = \Pr(Y > \sqrt{c}) + \Pr(Y < -\sqrt{c}) = 2\Pr(Y > \sqrt{c}).$$

Therefore, $\Pr(Y > \sqrt{c}) = 0.15$. It can be found from the table given at the end of the book that $\Pr(Y > 1.108) = 0.15$. Hence, $\sqrt{c} = 1.108$ and $c = 1.228$.

4. Suppose that X is represented as in Eq. (9.7.1). Since Y and Z are independent,

$$E(X) = \frac{n}{m}E\left(\frac{Y}{Z}\right) = \frac{n}{m}E(Y)E\left(\frac{1}{z}\right).$$

Since Y has the χ^2 distribution with m degrees of freedom, $E(Y) = m$. Since Z has the χ^2 distribution with n degrees of freedom,

$$\begin{aligned} E\left(\frac{1}{Z}\right) &= \int_0^\infty \frac{1}{z} f(z)dz = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty z^{(n/2)-2}\exp(-z/2)dz \\ &= \frac{2^{(n/2)-1}\Gamma[(n/2)-1]}{2^{n/2}\Gamma(n/2)} = \frac{1}{2[(n/2)-1]} = \frac{1}{n-2}. \end{aligned}$$

Hence, $E(X) = n/(n-2)$.

5. By Eq. (9.7.1), X can be represented in the form $X = Y/Z$, where Y and Z are independent and have identical χ^2 distributions. Therefore, $\Pr(Y > Z) = \Pr(Y < Z) = 1/2$. Equivalently, $\Pr(X > 1) = \Pr(X < 1) = 1/2$. Therefore, the median of the distribution of X is 1.

6. Let $f(x)$ denote the p.d.f. of X, let $W = mX/(mX+n)$, and let $g(w)$ denote the p.d.f. of W. Then $X = nW/[m(1-W)]$ and $\dfrac{dx}{dw} = \dfrac{n}{m} \cdot \dfrac{1}{(1-w)^2}$. For $0 < w < 1$,

$$\begin{aligned} g(w) &= f\left[\frac{nw}{m(1-w)}\right] \cdot \left|\frac{dx}{dw}\right| \\ &= k\left(\frac{n}{m}\right)^{(m/2)-1} \frac{w^{(m/2)-1}}{(1-w)^{(m/2)-1}} \cdot \frac{(1-w)^{(m+n)/2}}{n^{(m+n)/2}} \left(\frac{n}{m}\right) \frac{1}{(1-w)^2} \\ &= k\frac{1}{m^{m/2}n^{n/2}} w^{(m/2)-1}(1-w)^{(n/2)-1}, \end{aligned}$$

where

$$k = \frac{\Gamma[(m+n)/2]m^{m/2}n^{n/2}}{\Gamma(m/2)\Gamma(n/2)}.$$

It can be seen $g(w)$ is the p.d.f. of the required beta distribution.

7. (a) Here, $\overline{X}_m = 84/16 = 5.25$ and $\overline{Y}_n = 18/10 = 1.8$. Therefore, $S_1^2 = \sum_{i=1}^{16} X_i^2 - 16(\overline{X}_m^2) = 122$ and $S_2^2 = \sum_{i=1}^{10} Y_i^2 - 10(\overline{Y}_n^2) = 39.6$. It follows that

$$\hat{\sigma}_1^2 = \frac{1}{16}S_1^2 = 7.625 \quad \text{and} \quad \hat{\sigma}_2^2 = \frac{1}{10}S_2^2 = 3.96.$$

If $\sigma_1^2 = \sigma_2^2$, the following statistic V will have the F distribution with 15 and 9 degrees of freedom:

$$V = \frac{S_1^2/15}{S_2^2/9}.$$

(b) If the test is to be carried out at the level of significance 0.05, then H_0 should be rejected if $V > 3.01$. It is found that $V = 1.848$. Therefore, we do not reject H_0.

8. For any values of σ_1^2 and σ_2^2, the random variable

$$\frac{S_1^2/(15\sigma_1^2)}{S_2^2/(9\sigma_2^2)}$$

has the F distribution with 15 and 9 degrees of freedom. Therefore, if $\sigma_1^2 = 3\sigma_2^2$, the following statistic V will have that F distribution:

$$V = \frac{S_1^2/45}{S_2^2/9}.$$

As before, H_0 should be rejected if $V > c$, where $c = 3.01$ if the desired level of significance is 0.05.

9. When $\sigma_1^2 = \sigma_2^2$, V has an F distribution with 15 and 9 degrees of freedom. Therefore, $\Pr(V > 3.77) = 0.025$, which implies that $c_2 = 3.77$. Also, $1/V$ has an F distribution with 9 and 15 degrees of freedom. Therefore, $\Pr(1/V > 3.12) = 0.025$. It follows that $\Pr(V < 1/(3.12)) = 0.025$, which means that $c_1 = 1/(3.12) = 0.321$.

10. Let V be as defined in Exercise 9. If $\sigma_1^2 = r\sigma_2^2$, then V/r has the F distribution with 15 and 9 degrees of freedom. Therefore, H_0 would be rejected if $V/r < c_1$ or $V/r > c_2$, where c_1 and c_2 have the values found in Exercise 9.

11. For any positive number r, the hypothesis H_0 in Exercise 9 will not be rejected if $c_1 < V/r < c_2$. The set of all values of r for which H_0 would not be rejected will form a confidence interval with confidence coefficient 0.95. But $c_1 < V/r < c_2$ if and only if $V/c_2 < r < V/c_1$. Therefore, the confidence interval will contain all values of r between $V/3.77 = 0.265V$ and $V/0.321 = 3.12V$.

12. If a random variable Z has the χ^2 distribution with n degrees of freedom, Z can be represented as the sum of n independent and identically distributed random variables $Z_1, \ldots, Z_n$, each of which has a χ^2 distribution with 1 degree of freedom. Therefore, $Z/n = \sum_{i=1}^n Z_i/n = \overline{Z}_n$. As $n \to \infty$, it follows from the law of large numbers that $\overline{Z}_n$ will converge in probability to the mean of each Z_i, which is 1. Therefore $Z/n \xrightarrow{P} 1$. It follows from Eq. (9.7.1) that if X has the F distribution with m_0 and n degrees of freedom, then as $n \to \infty$, the distribution of X will become the same as the distribution of Y/m_0.

13. Suppose that X has the F distribution with m and n degrees of freedom, and consider the representation of X in Eq. (9.7.1). Then $Y/m \xrightarrow{P} 1$. Therefore, as $m \to \infty$, the distribution of X will become the same as the distribution of n/Z, where Z has a χ^2 distribution with n degrees of freedom. Suppose that c is the 0.05 quantile of the χ^2 distribution with n degrees of freedom. Then $\Pr(n/Z < n/c) = 0.95$. Hence, $\Pr(X < n/c) = 0.95$, and the value n/c should be entered in the column of the F distribution with $m = \infty$.

14. The test rejects the null hypothesis if the F statistic is greater than the 0.95 quantile of the F distribution with 15 and 9 degrees of freedom, which is 3.01. The power of the test when $\sigma_1^2 = 2\sigma_2^2$ is $1 - G_{15,9}(3.01/2) = 0.2724$. This can be computed using a computer program that evaluates the c.d.f. of an arbitrary F distribution.

15. The p-value will be the value of α_0 such that the observed v is exactly equal to either c_1 or c_2. The problem is deciding wheter $v = c_1$ or $v = c_2$ since, we haven't constructed a specific test. Since c_1 and c_2 are assumed to be the $\alpha_0/2$ and $1 - \alpha_0/2$ quantiles of the F distribution with $m - 1$ and $n - 1$ degrees of freedom, we must have $c_1 < c_2$ and $G_{m=1,n-1}(c_1) < 1/2$ and $G_{m-1,n-1}(c_2) > 1/2$. These inequalities allow us to choose whether $v = c_1$ or $v = c_2$. Every $v > 0$ is *some* quantile of each F distribution, indeed the $G^{-1}_{m-1,n-1}(v)$ quantile. If $G_{m-1,n-1}(v) < 1/2$, then $v = c_1$ and $\alpha_0 = 2G_{m-1,n-1}(v)$. If $G_{m-1,n-1}(v) > 1/2$, then $v = c_2$, and $\alpha_0 = 2[1 - G_{m-1,n-1}(v)]$. (There is 0 probability that $G_{m-1,n-1}(v) = 1/2$.) Hence, α_0 is the smaller of the two numbers $2G_{m-1,n-1}(v)$ and $2[1 - G_{m-1,n-1}(v)]$.

In Example 9.7.4, $v = 0.9491$ was the observed value and $2G_{25,25}(0.9491) = 0.8971$, so this would be the p-value.

16. The denominator of the likelihood ratio is maximized when all parameters equal their M.L.E.'s. The numerator is maximized when $\sigma_1^2 = \sigma_2^2$. As in the text, the likelihood ratio then equals

$$\Lambda(\boldsymbol{x}, \boldsymbol{y}) = dw^{m/2}(1-w)^{n/2},$$

where w and d are defined in the text. In particular, w is a strictly increasing function of the observed value of V. Notice that $\Lambda(\boldsymbol{x}, \boldsymbol{y}) \le k$ when $w \le k_1$ or $w \ge k_2$. This corresponds to $V \le c_1$ or $V \ge c_2$. In order for the test to have level α_0, the values c_1 and c_2 have to satisfy $\Pr(V \le c_1) + \Pr(V \ge c_2) = \alpha_0$ when $\sigma_1^2 = \sigma_2^2$.

17. The test found in Exercise 9 uses the values $c_1 = 0.321$ and $c_2 = 3.77$. The likelihood ratio test rejects H_0 when $dw^8(1-w)^5) \le k$, which is equivalent to $w^8(1-w)^5 \le k/d$. If $V = v$, then $w = 15v/(15v+9)$. In order for a test to be a likelihood ratio test, the two values c_1 and c_2 must lead to the same value of the likelihood ratio. In particular, we must have

$$\left(\frac{15c_1}{15c_1+9}\right)^8 \left(1 - \frac{15c_1}{15c_1+9}\right)^5 = \left(\frac{15c_2}{15c_2+9}\right)^8 \left(1 - \frac{15c_2}{15c_2+9}\right)^5.$$

Plugging the values of c_1 and c_2 from Exercise 9 into this formula we get 2.555×10^{-5} on the left and 1.497×10^{-5} on the right.

18. Let V^* be defined as in (9.7.5) so that V^* has the F distribution with $m-1$ and $n-1$ degrees of freedom and The distribution of $V = (\sigma_1^2/\sigma_2^2)V^*$. It is straightforward to compute

$$\Pr(V \le c_1) = \Pr\left(\frac{\sigma_1^2}{\sigma_2^2}V^* \le c_1\right) = \Pr\left(V^* \le \frac{\sigma_2^2}{\sigma_1^2}c_1\right) = G_{m-1,n-1}\left(\frac{\sigma_2^2}{\sigma_1^2}c_1\right),$$

and similarly,

$$\Pr(V \ge c_2) = 1 - G_{m-1,n-1}\left(\frac{\sigma_2^2}{\sigma_1^2}c_2\right).$$

19. (a) Apply the result of Exercise 18 with $c_1 = G_{10,20}^{-1}(0.025) = 0.2952$ and $c_2 = G_{10,20}^{-1}(0.975) = 2.774$ and $\sigma_2^2/\sigma_1^2 = 1/1.01$. The result is

$$G_{10,20}(c_1/1.01) + 1 - G_{10,20}(c_2/1.01) = G_{10,20}(0.289625) + 1 - G_{10,20}(2.746209) = 0.0503.$$

(b) Apply the result of Exercise 18 with $c_1 = G_{10,20}^{-1}(0.025) = 0.2952$ and $c_2 = G_{10,20}^{-1}(0.975) = 2.774$ and $\sigma_2^2/\sigma_1^2 = 1.01$. The result is

$$G_{10,20}(1.01 \times c_1) + 1 - G_{10,20}(1.01 \times c_2) = G_{10,20}(0.2954475) + 1 - G_{10,20}(2.80148) = 0.0498.$$

(c) Since the answer to part (b) is less than 0.05 (the value of the power function for all parameters in the null hypothesis set), the test is not unbiased.

9.8 Bayes Test Procedures

Commentary

This section introduces Bayes tests for the situations described in the earlier sections of the chapter. It derives Bayes tests as solutions to decision problems in which the loss function takes only three values: 0 for correct decision, and one positive value for each type of error. The cases of simple and one-sided hypotheses are covered as are the various situations involving samples from normal distributions. The calculations are done with improper prior distributions so that attention can focus on the methodology and similarity with the non-Bayesian results.

Solutions to Exercises

1. In this exercise, $\xi_0 = 0.9$, $\xi_1 = 0.1$, $w_0 = 1000$, and $w_1 = 18{,}000$. Also,

$$f_0(x) = \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}(x-50)^2\right]$$

and

$$f_1(x) = \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}(x-52)^2\right].$$

By the results of this section, it should be decided that the process is out of control if

$$\frac{f_1(x)}{f_0(x)} > \frac{\xi_0 w_0}{\xi_1 w_1} = \frac{1}{2}.$$

This inequality can be reduced to the inequality $2x - 102 > -\log\ 2$ or, equivalently, $x > 50.653$.

2. In this exercise, $\xi_0 = 2/3$, $\xi_1 = 1/3$, $w_0 = 1$, and $w_1 = 4$. Therefore, by the results of this section, it should be decided that f_0 is correct if

$$\frac{f_1(x)}{f_0(x)} < \frac{\xi_0 w_0}{\xi_1 w_1} = \frac{1}{2}.$$

Since $f_1(x)/f_0(x) = 4x^3$, it should be decided that f_0 is correct if $4x^3 < 1/2$ or, equivalently, if $x < 1/2$.

3. In this exercise, $\xi_0 = 0.8$, $\xi_1 = 0.2$, $w_0 = 400$, and $w_1 = 2500$. Also, if we let $y = \sum_{i=1}^n x_i$, then

$$f_0(\boldsymbol{X}) = \frac{\exp(-3n)\, 3^y}{\prod_{i=1}^n (x_i!)}$$

and

$$f_1(\boldsymbol{X}) = \frac{\exp(-7n)\, 7^y}{\prod_{i=1}^n (x_i!)}.$$

By the results of this section, it should be decided that the failure was caused by a major defect if

$$\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} = \exp(-4n)\left(\frac{7}{3}\right)^y > \frac{\xi_0 w_0}{\xi_1 w_1} = 0.64$$

or, equivalently, if

$$y > \frac{4n + \log(0.64)}{\log\left(\frac{7}{3}\right)}.$$

4. In this exercise, $\xi_0 = 1/4$, $\xi_1 = 3/4$, and $w_0 = w_1 = 1$. Let $x_1, \ldots, x_n$ denote the observed values in the sample, and let $y = \sum_{i=1}^{n} x_i$. Then

$$f_0(\boldsymbol{X}) = (0.3)^y (0.7)^{n-y}$$

and

$$f_1(\boldsymbol{X}) = (0.4)^y (0.6)^{n-y}.$$

By the results of this section, H_0 should be rejected if

$$\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} > \frac{\xi_0 w_0}{\xi_1 w_1} = \frac{1}{3}.$$

But

$$\frac{f_1(\boldsymbol{X})}{f_0(\boldsymbol{X})} = \left(\frac{4}{3}\cdot\frac{7}{6}\right)^y \left(\frac{6}{7}\right)^n > \frac{1}{3}$$

if and only if

$$y \log\frac{14}{9} + n \log\frac{6}{7} > \log\frac{1}{3}$$

or, equivalently, if and only if

$$\overline{x}_n = \frac{y}{n} > \frac{\log\frac{7}{6} + \frac{1}{n}\log\frac{1}{3}}{\log\frac{14}{9}}.$$

5. (a) In the notation of this section $\xi_0 = \Pr(\theta = \theta_0)$ and f_i is the p.f. or p.d.f. of $\boldsymbol{X}$ given $\theta = \theta_i$. By the law of total probability, the marginal p.f. or p.d.f. of $\boldsymbol{X}$ is $\xi_0 f_0(\boldsymbol{x}) + \xi_1 f_1(\boldsymbol{x})$. Applying Bayes' theorem for random variables gives us that

$$\Pr(\theta = \theta_0|\boldsymbol{x}) = \frac{\xi_0 f_0(\boldsymbol{x})}{\xi_0 f_0(\boldsymbol{x}) + \xi_1 f_1(\boldsymbol{x})}.$$

(b) The posterior expected value of the loss given $\boldsymbol{X} = \boldsymbol{x}$ is

$$\begin{array}{ll} \dfrac{w_0 \xi_0 f_0(\boldsymbol{x})}{\xi_0 f_0(\boldsymbol{x}) + \xi_1 f_1(\boldsymbol{x})} & \text{if reject } H_0, \\ \dfrac{w_1 \xi_1 f_1(\boldsymbol{x})}{\xi_0 f_0(\boldsymbol{x}) + \xi_1 f_1(\boldsymbol{x})} & \text{if don't reject } H_0. \end{array}$$

The tests δ that minimize $r(\delta)$ have the form

$$\begin{aligned} \text{Don't Reject } H_0 \quad &\text{if} \quad \xi_0 w_0 f_0(\boldsymbol{x}) > \xi_1 w_1 f_1(\boldsymbol{x}), \\ \text{Reject } H_0 \quad &\text{if} \quad \xi_0 w_0 f_0(\boldsymbol{x}) < \xi_1 w_1 f_1(\boldsymbol{x}), \\ \text{Do either} \quad &\text{if} \quad \xi_0 w_0 f_0(\boldsymbol{x}) = \xi_1 w_1 f_1(\boldsymbol{x}). \end{aligned}$$

Notice that these tests choose the action that has smaller posterior expected loss. If neither action has smaller posterior expected loss, these tests can do either, but either would then minimize the posterior expected loss.

(c) The "reject H_0" condition in part (b) is $\xi_0 w_0 f_0(\boldsymbol{x}) < \xi_1 w_1 f_1(\boldsymbol{x})$. This is equivalent to $w_0 \Pr(\theta = \theta_0|\boldsymbol{x}) < w_1[1 - \Pr(\theta = \theta_0|\boldsymbol{x})]$. Simplifying this inequality yields $\Pr(\theta = \theta_0|\boldsymbol{x}) < w_1/(w_0 + w_1)$. Since we can do whatever we want when equality holds, and since "H_0 true" means $\theta = \theta_0$, we see that the test described in part (c) is one of the tests from part (b).

6. The proof is just as described in the hint. For example, (9.8.12) becomes

$$\int_{\theta_0}^{\infty} \int_{-\infty}^{\theta_0} w_1(\theta) w_0(\psi) \xi(\theta) \xi(\psi) [f_n(\boldsymbol{x}_1 \mid \theta) f_n(\boldsymbol{x}_2 \mid \psi) - f_n(\boldsymbol{x}_2 \mid \theta) f_n(\boldsymbol{x}_1 \mid \psi)] d\psi d\theta.$$

The steps after (9.8.12) are unchanged.

7. We shall argue indirectly. Suppose that there is $\boldsymbol{x}$ such that the p-value is not equal to the posterior probability that H_0 is true. First, suppose that the p-value is greater. Let α_0 be greater than the posterior probability and less than the p-value. Then the test that rejects H_0 when $\Pr(H_0 \text{ true}|\boldsymbol{x}) \le \alpha_0$ will reject H_0, but the level α_0 test will not reject H_0 because the p-value is greater than α_0. This contradicts the fact that the two tests are the same. The case in which the p-value is smaller is very similar.

8. (a) The joint p.d.f. of the data given the parameters is

$$(2\pi)^{-(m+n)/2} \tau^{(m+n)/2} \exp\left(-\frac{\tau}{2} \left[\sum_{i=1}^{m} (x_i - \mu_1)^2 + \sum_{j=1}^{n} (y_j - \mu_2)^2 \right] \right).$$

Use the following two identities to complete the proof of this part:

$$\begin{aligned} \sum_{i=1}^{m} (x_i - \mu_1)^2 &= \sum_{i=1}^{m} (x_i - \overline{x}_m)^2 + m(\overline{x}_m - \mu_1)^2, \\ \sum_{j=1}^{n} (y_j - \mu_2)^2 &= \sum_{j=1}^{n} (y_j - \overline{y}_n)^2 + n(\overline{y}_n - \mu_2)^2. \end{aligned}$$

(b) The prior p.d.f. is just $1/\tau$.

i. As a function of μ_1, the posterior p.d.f. is a constant times $\exp(-m\tau(\overline{x}_m - \mu_1)^2/2)$, which is just like the p.d.f. of a normal distribution with mean $\overline{x}_m$ and variance $1/(m\tau)$.

ii. The result for μ_2 is similar to that for μ_1.

iii. As a function of (μ_1, μ_2), the posterior p.d.f. looks like a constant times

$$\exp(-m\tau(\overline{x}_m - \mu_1)^2/2) \exp(-n\tau(\overline{y}_n - \mu_2)^2/2),$$

which is like the product of the two normal p.d.f.'s from parts (i) and (ii). Hence, the conditional posterior distribution of (μ_1, μ_2) given τ is that of two independent normal random variables with the two distributions from parts (i) and (ii).

iv. We can integrate μ_1 and μ_2 out of the joint p.d.f. Integrating $\exp(-m\tau(\overline{x}_m - \mu_1)^2)$ yields $(2\pi)^{1/2}\tau^{-1/2}$. Integrating $\exp(-n\tau(\overline{y}_n - \mu_2)^2)$ yields $(2\pi)^{1/2}\tau^{-1/2}$ also. So, the marginal posterior p.d.f. of τ is a constant times $\tau^{(m+n-2)/2}\exp(-0.5\tau(s_x^2 + s_y^2))$. This is the p.d.f. of a gamma distribution with parameters $(m+n-2)/2$ and $(s_x^2 + s_y^2)/2$, except for a constant factor.

(c) Since μ_1 and μ_2 are independent, conditional on τ, we have that $\mu_1 - \mu_2$ has a normal distribution conditional on τ with mean equal to the difference of the means and variance equal to the sum of the variances. That is, $\mu_1 - \mu_2$ has a normal distribution with mean $\overline{x}_m - \overline{y}_n$ and variance $\tau^{-1}(1/m + 1/n)$ given τ. If we subtract the mean and divide by the square-root of the variance, we get a standard normal distribution for the result, which is the Z stated in this part of the problem. Since the standard normal distribution is independent of τ, then Z is independent of τ and has the standard normal distribution marginally.

(d) Recall that τ has a gamma distribution with parameters $(m+n-2)/2$ and $(s_x^2 + s_y^2)/2$. If we multiply τ by $s_x^2 + s_y^2$, the result W has a gamma distribution with the same first parameter but with the second parameter divided by $s_x^2 + s_y^2$, namely 1/2.

(e) Since Z and W are independent with Z having the standard normal distribution and W having the χ^2 distribution with $m+n-2$ degrees of freedom, it follows from the definition of the t distribution that $Z/(W/[m+n-2])^{1/2}$ has the t distribution with $m+n-2$ degrees of freedom. It is easy to check that $Z/(W/[m+n-2])^{1/2}$ is the same as (9.8.17).

9. (a) The null hypothesis can be rewritten as $\tau_1 \geq \tau_2$, where $\tau_i = 1/\sigma_i^2$. This can be further rewritten as $\tau_1/\tau_2 \geq 1$. Using the usual improper prior for all parameters yields the posterior distribution of τ_1 and τ_2 to be that of independent gamma random variables with τ_1 having parameters $(m-1)/2$ and $s_x^2/2$ while τ_2 has parameters $(n-1)/2$ and $s_y^2/2$. Put another way, $\tau_1 s_x^2$ has the χ^2 distribution with $m-1$ degrees of freedom independent of $\tau_2^2 s_y^2$ which has the χ^2 distribution with $n-1$ degrees of freedom. This makes the distribution of

$$W = \frac{\tau_1 s_x^2/(m-1)}{\tau_2 s_y^2/(n-1)}$$

the F distribution with $m-1$ and $n-1$ degrees of freedom. The posterior probability that H_0 is true is

$$\Pr(\tau_1/\tau_2 \geq 1) = \Pr\left(W \geq \frac{s_x^2/(m-1)}{s_y^2/(n-1)}\right) = 1 - G_{m-1,n-1}\left(\frac{s_x^2/(m-1)}{s_y^2/(n-1)}\right).$$

The posterior probability is at most α_0 if and only if

$$\frac{s_x^2/(m-1)}{s_y^2/(n-1)} \geq F^{-1}_{m-1,n-1}(1-\alpha_0).$$

This is exactly the form of the rejection region for the level α_0 F test of H_0.

(b) This is a special case of Exercise 7.

10. (a) Using Theorem 9.8.2, the posterior distribution of

$$(26 + 26 - 2)^{1/2}\frac{\mu_1 - \mu_2 - [5.134 - 3.990]}{(1/26 + 1/26)^{1/2}(63.96 + 67.39)^{1/2}} = \frac{\mu_1 - \mu_2 - 1.144}{0.4495}$$

is the t distribution with 50 degrees of freedom.

(b) We can compute

$$\Pr(|\mu_1 - \mu_2| \leq d) = T_{50}\left(\frac{d - 1.144}{0.4495}\right) - T_{50}\left(\frac{-d - 1.144}{0.4495}\right).$$

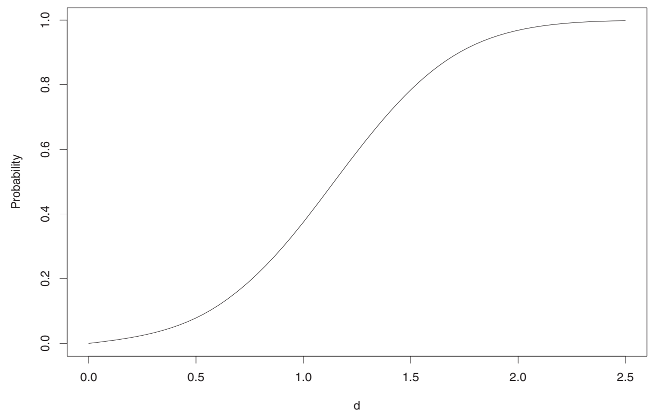

Figure S.9.6: Figure for Exercise 10b of Sec. 9.8.

A plot of this function is in Fig. S.9.6.

11. (a) First, let $H_0 : \theta \in \Omega'$ and $H_1 : \theta \in \Omega''$. Then $\Omega_0 = \Omega'$ and $\Omega_1 = \Omega''$. Since d_0 is the decision that H_0 is true we have $d_0 = d'$ and $d_1 = d''$. Since w_0 is the cost of type II error, and type I error is to choose $\theta \in \Omega''$ when $\theta \in \Omega'$, $w_0 = w'$, and $w_1 = w''$. It is straightforward to see that everything switches for the other case.

(b) The test procedure is to

$$\text{choose } d_1 \text{ if } \Pr(\theta \in \Omega_0|\boldsymbol{x}) < \frac{w_1}{w_0 + w_1}, \tag{S.9.12}$$

$$\text{and choose either action if the two sides are equal.} \tag{S.9.13}$$

In the first case, this translates to "choose d" if $\Pr(\theta \in \Omega'|\boldsymbol{x}) < w"/(w' + w")$, and choose either action if the two sides are equal." This is equivalent to "choose d' if $\Pr(\theta \in \Omega'|\boldsymbol{x}) > w"/(w' + w")$, and choose either action if the two sides are equal." This, in turn, is equivalent to "choose d' if $\Pr(\theta \in \Omega"|\boldsymbol{x}) < w'/(w' + w")$, and choose either action if the two sides are equal." This last statement, in the second case, translates to (S.9.12). Hence, the Bayes test produces the same action (d' or d'') regardless of which hypothesis you choose to call the null and which the alternative.

9.9 Foundational Issues

Commentary

This section discusses some subtle issues that arise when the foundations of hypothesis testing are examined closely . These issues are the relationship between sample size and the level of a test and the distinction between statistical significance and practical significance. The term "statistical significance" is not introduced in the text until this section, hence instructors who do not wish to discuss this issue can avoid it altogether.

Solutions to Exercises

1. (a) When $\mu = 0$, X has the standard normal distribution. Therefore, $c = 1.96$. Since H_0 should be rejected if $|X| > c$, then H_0 will be rejected when $X = 2$.

(b)
$$\frac{f(X\,|\,\mu=0)}{f(X\,|\,\mu=5)} = \frac{\exp\left(-\frac{1}{2}X^2\right)}{\exp[-\frac{1}{2}(X-5)^2]} = \exp\left[\frac{1}{2}(25-10X)\right].$$

When $X = 2$, this likelihood ratio has the value $\exp(5/2) = 12.2$. Also,

$$\frac{f(X\,|\,\mu=0)}{f(X\,|\,\mu=-5)} = \frac{\exp\left(-\frac{1}{2}X^2\right)}{\exp\left[-\frac{1}{2}(X+5)^2\right]} = \exp[\frac{1}{2}(25+10X)].$$

When $X = 2$, this likelihood ratio has the value $\exp(45/2) = 5.9 \times 10^9$.

2. When $\mu = 0$, $100\overline{X}_n$ has the standard normal distribution. Therefore, $\Pr(100\,|\,\overline{X}_n\,| > 1.96\,|\,\mu = 0) = 0.05$. It follows that $c = 1.96/100 = 0.0196$.

   (a) When $\mu = 0.01$, the random variable $Z = 100(\overline{X}_n - 0.01)$ has the standard normal distribution. Therefore,

$$\begin{aligned}\Pr(\,|\,\overline{X}_n\,| < c\,|\,\mu = 0.01) &= \Pr(-1.96 < 100\overline{X}_n < 1.96\,|\,\mu = 0.01)\\ &= \Pr(-2.96 < z < 0.96)\\ &= 0.8315 - 0.0015 = 0.8300.\end{aligned}$$

   It follows that $\Pr(\,|\,\overline{X}_n\,| \geq c\,|\,\mu = 0.01) = 1 - 0.8300 = 0.1700$.

   (b) When $\mu = 0.02$, the random variable $Z = 100(\overline{X}_n - 0.02)$ has the standard normal distribution. Therefore,

$$\begin{aligned}\Pr(\,|\,\overline{X}_n\,| < c\,|\,\mu = 0.02) &= \Pr(-3.96 < Z < -0.04)\\ &= \Pr(0.04 < Z < 3.96) = 1 - 0.5160.\end{aligned}$$

   It follows that $\Pr(\,|\,\overline{X}_n\,| < c\,|\,\mu = 0.02) = 0.5160$.

3. When $\mu = 0$, $100\overline{X}_n$ has the standard normal distribution. The calculated value of $100\overline{X}_n$ is $100(0.03) = 3$. The corresponding tail area is $\Pr(100\,\overline{X}_n > 3) = 0.0013$.

4. (a) According to Theorem 9.2.1, we reject H_0 if

$$\frac{19}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}x_i^2\right) < \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-0.5)^2\right).$$

   This inequality is equivalent to

$$\frac{2\log(19)}{n} + \frac{1}{4} < \overline{x}_n.$$

   That is, $c_n = 2\log(19)/n + 1/4$. For $n = 1, 100, 100000$, the values of c_n are 6.139, 0.3089, and 0.2506.

   (b) The size of the test is

$$\Pr(\overline{X}_n \geq c_n|\theta = 0) = 1 - \Phi(c_n \times n^{1/2}).$$

   For $n = 1, 100, 10000$, the sizes are 4.152×10^{-10}, 0.001, and 0.

5. (a) We want to choose c_n so that

$$19[1 - \Phi(\sqrt{n}c_n)] = \Phi(\sqrt{n}[c_n - 0.5]).$$

   Solving this equation must be done numerically. For $n = 1$, the equation is solved for $c_n = 1.681$. For $n = 100$, we need $c_n = 0.3021$. For $n = 10000$, we need $c_n = 0.25$ (both sides are essentially 0).

   (b) The size of the test is $1 - \Phi(c_n n^{1/2})$, which is 0.0464 for $n = 1$, 0.00126 for $n = 100$ and essentially 0 for $n = 10000$.

9.10 Supplementary Exercises

Solutions to Exercises

1. According to Theorem 9.2.1, we want to reject H_0 when

$$(1/2)^3 < (3/4)^x(1/4)^{3-x}.$$

We don't reject H_0 when the reverse inequality holds, and we can do either if equality holds. The inequality above can be simplified to $x > \log(8)/\log(3) = 1.892$. That is, we reject H_0 if X is 2 or 3, and we don't reject H_0 if X is 0 or 1. The probability of type I error is $3(1/2)^3 + (1/2)^3 = 1/2$ and the probability of type II error is $(1/4)^3 + 3(1/4)^2(3/4) = 5/32$.

2. The probability of an error of type 1 is

$$\alpha = \Pr(\text{Rej. } H_0 \,|\, H_0) = \Pr(X \le 5 \,|\, \theta = 0.1) = 1 - (.9)^5 = .41.$$

The probability of an error of type 2 is

$$\beta = \Pr(\text{Acc. } H_0 \,|\, H_1) = \Pr(X \ge 6 \,|\, \theta = 0.2) = (.8)^5 = .33.$$

3. It follows from Sec. 9.2 that the Bayes test procedure rejects H_0 when $f_1(x)/f_0(x) > 1$. In this problem,

$$f_1(x) = (.8)^{x-1}(.2) \quad \text{for} \quad x = 1, 2, \ldots,$$

and

$$f_0(x) = (.9)^{x-1}(.1) \quad \text{for} \quad x = 1, 2, \ldots.$$

Hence, H_0 should be rejected when $2(8/9)^{x-1} \ge 1$ or $x - 1 \le 5.885$. Thus, H_0 should be rejected for $X \le 6$.

4. It follows from Theorem 9.2.1 that the desired test will reject H_0 if

$$\frac{f_1(x)}{f_0(x)} = \frac{f(x \,|\, \theta = 0)}{f(x \,|\, \theta = 2)} > \frac{1}{2}.$$

In this exercise, the ratio on the left side reduces to $x/(1-x)$. Hence, the test specifies rejecting H_0 *if* $x > 1/3$. For this test,

$$\alpha(\delta) = \Pr\left(X > \frac{1}{3} \,|\, \theta = 2\right) = \frac{4}{9},$$
$$\beta(\delta) = \Pr\left(X < \frac{1}{3} \,|\, \theta = 0\right) = \frac{1}{9}.$$

Hence, $\alpha(\delta) + 2\beta(\delta) = 2/3$.

5. It follows from the previous exercise and the Neyman-Pearson lemma that the optimal procedure δ specifies rejecting H_0 when $x/(1-x) > k'$ or, equivalently, when $x > k$. The constant k must be chosen so that

$$\alpha = \Pr(X > k \,|\, \theta = 2) = \int_k^1 f(x \,|\, \theta = 2)dx = (1-k)^2.$$

Hence, $k = 1 - \alpha^{1/2}$ and

$$\beta(\delta) = \Pr(X < k \,|\, \theta = 0) = k^2 = (1 - \alpha^{1/2})^2.$$

6. (a) The power function is given by

$$\pi(\theta \,|\, \delta) = \Pr(X > 0.9 \,|\, \theta) = \int_{0.9}^1 f(x \,|\, \theta)dx = .19 - .09\theta.$$

(b) The size of δ is

$$\sup_{\theta \geq 1} \pi(\theta \,|\, \delta) = .10.$$

7. A direct calculation shows that for $\theta_1 < \theta_2$,

$$\frac{d}{dx}\left[\frac{f(x \,|\, \theta_2)}{f(x \,|\, \theta_1)}\right] = \frac{2(\theta_1 - \theta_2)}{[2(1-\theta_1)x + \theta_1]^2} < 0.$$

Hence, the ratio $f(x \,|\, \theta_2)/f(x \,|\, \theta_1)$ is a decreasing function of x or, equivalently, an increasing function of $r(x) = -x$. It follows from Theorem 9.3.1 that a UMP test of the given hypotheses will reject H_0 when $r(X) \geq c$ or , equivalently, when $X \leq k$. Hence, k must be chosen so that

$$.05 = \Pr\left(X \leq k \,|\, \theta = \frac{1}{2}\right) = \int_0^k f\left(x \,|\, \theta = \frac{1}{2}\right)dx = \frac{1}{2}(k^2 + k), \quad \text{or} \quad k = \frac{1}{2}(\sqrt{1.4} - 1).$$

8. Suppose that the proportions of red, brown, and blue chips are p_1, p_2, and p_3, respectively. It follows from the multinomial distribution that the probability of obtaining exactly one chip of each color is

$$\frac{3!}{1!1!1!}p_1p_2p_3 = 6p_1p_2p_3.$$

Hence, $\Pr(Rej.\ H_0 \,|\, p_1, p_2, p_3) = 1 - 6p_1p_2p_3$.

(a) The size of the test is

$$\alpha = \Pr\left(\text{Rej. } H_0 \,\middle|\, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{7}{9}.$$

(b) The power under the given alternative distribution is $\Pr(\text{Rej. } H_0 \,|1/7, 2/7, 4/7) = 295/343 = .860$.

9. Let $f_i(x)$ denote the p.d.f. of X under the hypotheses $H_i (i = 0, 1)$. Then

$$\frac{f_1(x)}{f_0(x)} = \begin{cases} \infty & \text{for } x \leq 0 \text{ or } x \geq 1, \\ \varphi(x) & \text{for } 0 < x < 1, \end{cases}$$

where $\varphi(x)$ is the standard normal p.d.f. The most powerful test δ of size 0.01 rejects H_0 when $f_1(x)/f_0(x) > k$. Since $\varphi(x)$ is strictly decreasing for $0 < x < 1$, it follows that δ will reject H_0 if

$X \le 0$, $X \ge 1$, or $0 < X < c$, where c is chosen so that $\Pr(0 < X < c \,|\, H_0) = .01$. Since X has a uniform distribution under H_0 , $c = .01$. Thus, δ specifies rejecting H_0 if $X \le .01$ or $X \ge 1$. The power of δ under H_1 is

$$\Pr(X \le .01 \,|\, H_1) + \Pr(X \ge 1 \,|\, H_1) = \Phi(.01) + [1 - \Phi(1)] = .5040 + .1587 = .6627.$$

10. The usual t statistic U is defined by Eq. (9.5.2) with n = 12 and $\mu_0 = 3$. Because the one-sided hypotheses H_0 and H_1 are reversed from those in (9.5.1), we now want to reject H_0 if $U \le c$. If $\mu_0 = 3$, then U has the t distribution with 11 degrees of freedom. Under these conditions, we want $\Pr(U \le c) = 0.005$ or equivalently, by the symmetry of the t distribution, $\Pr(U \le -c) = 0.995$. It is found from the table of the t distribution that $-c = 3.106$. Hence, H_0 should be rejected if $U \le -3.106$.

11. It is known from Example 4 that the UMP test rejects H_0 if $\overline{X}_n \ge c$. Hence, c must be chosen so that

$$0.95 = \Pr(X_n \ge c \,|\, \theta = 1) = \Pr\left[Z \ge \sqrt{n}(c-1)\right],$$

where Z has the standard normal distribution. Hence, $\sqrt{n}(c-1) = -1.645$, and $c = 1 - (1.645)/n^{1/2}$. Since the power function of this test will be a strictly increasing function of θ, the size of the test will be

$$\begin{aligned}\alpha &= \sup_{\theta \le 0} \Pr(\text{Rej. } H_0 \,|\, \theta) = \Pr(\text{Rej. } H_0 \,|\, \theta = 0) = \Pr\left[\overline{X}_n \ge 1 - \frac{(1.645)}{n^{1/2}} \middle|\, \theta = 0\right] \\ &= \Pr(Z \ge n^{1/2} - 1.645),\end{aligned}$$

where Z again has the standard normal distribution. When $n = 16$,

$$\alpha = \Pr(Z \ge 2.355) = .0093.$$

12. For $\theta_1 < \theta_2$,

$$\frac{f_n(\boldsymbol{x} \,|\, \theta_2)}{f_n(\boldsymbol{x} \,|\, \theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^8 \left(\prod_{i=1}^{8} x_i\right)^{\theta_2 - \theta_1},$$

which is an increasing function of $T = \prod_{i=1}^{8} x_i$. Hence, the UMP test specifies rejecting H_0 when $T > c$ or, equivalently, when $-2\sum_{i=1}^{8} \log X_i < k$. The reason for expressing the test in this final form is that when $\theta = 1$, the observations $X_1, \ldots, X_8$ are i.i.d. and each has a uniform distribution on the interval (0,1). Under these conditions, $-2\sum_{i=1}^{8} \log X_i$ has a χ^2 distribution with $2n = 16$ degrees of freedom (see Exercise 7 of Sec. 8.2 and Exercise 5 of Sec. 8.9). Hence, in accordance with Theorem 9.3.1, H_0 should be rejected if $-2\sum_{i=1}^{8} \log X_i \le 7.962$, which is the 0.05 quantile of the χ^2 distribution with 16 degrees of freedom, or equivalently if $\sum_{i=1}^{8} \log x_i \ge -3.981$.

13. The χ^2 distribution with θ degrees of freedom is a gamma distribution with parameters $\alpha = \theta/2$ *and* $\beta = 1/2$. Hence, it follows from Exercise 3 of Sec. 9.3 that the joint p.d.f. of $X_i, \ldots, X_n$ has a monotone likelihood ratio in the statistic $T = \prod_{i=1}^{n} X_i$. Hence, there is a UMP test of the given hypotheses, and it specifies rejecting H_0 when $T \geq c$ or, equivalently, when $\log T = \sum_{i=1}^{n} \log X_i \geq k$.

14. Let $\overline{X}_1$ be the average of the four observations $X_1, \ldots, X_4$ and let $\overline{X}_2$ be the average of the six observations $X_5, \ldots, X_{10}$. Let $S_1^2 = \sum_{i=1}^{4}(X_i - \overline{X}_1)^2$ and $S_2^2 = \sum_{i=5}^{10}(X_i - \overline{X}_2)^2$. Then S_1^2/σ^2 and S_2^2/σ^2 have independent χ^2 distributions with 3 and 5 degrees of freedom, respectively. Hence, $(5S_1^2)/(3S_2^2)$ has the desired F distribution.

15. It was shown in Sec. 9.7 that the F test rejects H_0 *if* $V \geq 2.20$, where V is given by (9.7.4) and 2.20 is the 0.95 quantile of the F distribution with 15 and 20 degrees of freedom. For any values of σ_1^2 *and* σ_2^2, the random variable V^* given by (9.7.5) has the F distribution with 15 and 20 degrees of freedom. When $\sigma_1^2 = 2\sigma_2^2, V^* = V/2$. Hence, the power when $\sigma_1^2 = 2\sigma_2^2$ is

$$P^*(\text{Rej. } H_0) = P^*(V \geq 2.20) = P^*\left(\frac{1}{2}V \geq 1.10\right) = \Pr(V^* \geq 1.1),$$

where P^* denotes a probability calculated under the assumption that $\sigma_1^2 = 2\sigma_2^2$.

16. The ratio $V = S_X^2/S_Y^2$ has the F distribution with 8 and 8 degrees of freedom, and so does $1/V = S_Y^2/S_X^2$. Thus,

$$.05 = \Pr(T > c) = \Pr(V > c) + \Pr(1/V > c) = 2 \ \Pr(V > c).$$

It follows that c must be the .975 quantile of the distribution of V, which is found from the tables to be 4.43.

17. (a) Carrying out a test of size α on repeated independent samples is like performing a sequence of Bernoulli trials on each of which the probability of success is α. With probability 1, a success will ultimately be obtained. Thus, sooner or later, H_0 will ultimately be rejected. Therefore, the overall size of the test is 1.

    (b) As we know from the geometric distribution, the expected number samples, or trials, until a success is obtained is $1/\alpha$.

18. If U is defined as in Eq. (8.6.9), then the prior distribution of U is the t distribution with $2\alpha_0 = 2$ degrees of freedom. Since the t distribution is symmetric with respect to the origin, it follows that under the prior distribution, $\Pr(H_0) = \Pr(\mu \leq 3) = \Pr(U \leq 0) = 1/2$. It follows from (8.6.1) and (8.6.2) that under the posterior distribution,

$$\begin{aligned}
\mu_1 &= \frac{3 + (17)(3.2)}{1 + 17} = 3.189\,, \quad \lambda_1 = 18\,, \\
\alpha_1 &= 1 + \frac{17}{2} = 9.5\,, \\
\beta_1 &= 1 + \frac{1}{2}(17) + \frac{(17)(.04)}{2(18)} = 9.519.
\end{aligned}$$

If we now define Y to be the random variable in Eq. (8.6.12) then $Y = (4.24)(\mu - 3.19)$ and Y has the t distribution with $2\alpha_1 = 19$ degrees of freedom. Thus, under the posterior distribution,

$$\Pr(H_0) = \Pr(\mu \leq 3) = \Pr[Y \leq (4.24)(3 - 3.19)] = \Pr(Y \leq -.81) = \Pr(Y \geq .81).$$

It is found from the table of the t distribution with 19 degrees of freedom that this probability is approximately 0.21.

19. At each point $\theta \in \Omega_1, \pi(\theta \mid \delta)$ must be at least as large as it is at any point in Ω_0, because δ is unbiased. But $\sup_{\theta \in \Omega_0} \pi(\theta \mid \delta) = \alpha$, at every point $\theta \in \Omega_1$.

20. Since δ is unbiased and has size α, it follows from the previous exercise that $\pi(\theta \mid \delta) \leq \alpha$ for all θ inside the circle A and $\pi(\theta \mid \delta) \geq \alpha$ for all θ outside A. Since $\pi(\theta \mid \delta)$ is continuous, it must therefore be equal to α everywhere on the boundary of A. Note that this result is true regardless of whether all of any part of the boundary belongs to H_0 *or* H_1.

21. Since H_0 is simple and δ has size α, then $\pi(\theta_0 \mid \delta) = \alpha$. Since δ is unbiased, $\pi(\theta \mid \delta) \geq \alpha$ for all other values of θ. Therefore, $\pi(\theta \mid \delta)$ is a minimum at $\theta = \theta_0$. Since π is assumed to be differentiable, it follows that $\pi'(\theta_0 \mid \delta) = 0$.

22. (a) We want $\Pr(X > c_1 \mid H_0) = \Pr(Y > c_2 \mid H_0) = .05$. Under H_0, both X and Y/10 are standard normal. Therefore, $c_1 = 1.645$ and $c_2 = 16.45$.

    (b) The most powerful test of a size α_0, conditional on observing X with a variance of σ^2 is to reject H_0 if $X > \sigma\Phi^{-1}(1 - \alpha_0)$. In this problem we are asked to find two such tests: one with $\sigma = 1$ and $\alpha_0 = 2.0 \times 10^{-7}$ and the other with $\sigma = 10$ and $\alpha_0 = 0.0999998$. The resulting critical values are

$$\begin{aligned} \Phi^{-1}(1 - 2.0 \times 10^{-7}) &= 5.069, \\ 10\Phi^{-1}(1 - 0.0999998) &= 12.8155. \end{aligned}$$

    (c) The overall size of a test in this problem is the average of the two conditional sizes, since the two types of meteorological conditions have probability 1/2 each. In part (a), the two conditional sizes are both 0.05, so that is the average as well. In part (b), the average of the two sizes is $(2.0 \times 10^{-7} + 0.0999998)/2 = 0.05$ also. The powers are also the averages of the two conditional powers. The power of the conditional size α_0 test with variance σ^2 is

$$1 - \Phi(\sigma\Phi^{-1}(1 - \alpha_0) - 10).$$

    The results are tabulated below:

Part	Good	Poor	Average
(a)	1	0	0.5
(b)	0.9999996	0.002435	0.5012

23. (a) The data consist of both X and Y, where X is defined in Exercise 22 and $Y = 1$ if meteorological conditions are poor and $Y = 0$ if not. The joint p.f./p.d.f. of (X, Y) given $\Theta = \theta$ is

$$\frac{1}{2(2\pi)^{1/2}10^y} \exp\left(-\frac{1-y}{2}[x-\theta]^2 - \frac{y}{200}[x-\theta]^2\right).$$

    The Bayes test will choose H_0 when

$$\begin{aligned} & w_0\xi_0 \frac{1}{2(2\pi)^{1/2}10^y} \exp\left(-\frac{1-y}{2}x^2 - \frac{y}{200}x^2\right) \\ > \ & w_1\xi_1 \frac{1}{2(2\pi)^{1/2}10^y} \exp\left(-\frac{1-y}{2}[x-10]^2 - \frac{y}{200}[x-10]^2\right). \end{aligned}$$

It will choose H_1 when the reverse inequality holds, and it can do either when equality holds. This inequality can be rewritten by splitting according to the value of y. That is, choose H_0 if

$$\begin{cases} x < 5 + \log(w_0\xi_0/(w_1\xi_1))/10 & \text{if } y = 0, \\ x < 5 + 10\log(w_0\xi_0/(w_1\xi_1)) & \text{if } y = 1. \end{cases}$$

(b) In order for a test to be of the form of part (a), the two critical values c_0 and c_1 used for $y = 0$ and $y = 1$ respectively must satisfy $c_1 - 5 = 100(c_0 - 5)$. In part (a) of Exercise 22, the two critical values are $c_0 = 1.645$ and $c_1 = 16.45$. These do not even approximately satisfy $c_1 - 5 = 100(c_0 - 5)$.

(c) In part (b) of Exercise 22, the two critical values are $c_0 = 5.069$ and $c_1 = 12.8155$. These approximately satisfy $c_1 - 5 = 100(c_0 - 5)$.

24. (a) The Poisson distribution has M.L.R. in Y, so rejection H_0 when $Y \le c$ is a UMP test of its size. With $c = 0$, the size is $\Pr(Y = 0|\theta = 1) = \exp(-n)$.

(b) The power function of the test is $\Pr(Y = 0|\theta) = \exp(-n\theta)$.

25. Let I be the random interval that corresponds to the UMP test, and let J be a random interval that corresponds to some other level α_0 test. Translating UMP into what it says about the random interval I compared to J, we have for all $\theta > c$

$$\Pr(c \in I|\theta) \le \Pr(c \in J|\theta).$$

In other words, the observed value of I is a uniformly most accurate coefficient $1-\alpha_0$ confidence interval if, for every random interval J such that the observed value of J is a coefficient $1-\alpha_0$ confidence interval and for all $\theta_2 > \theta_1$,

$$\Pr(\theta_1 \in I|\theta = \theta_2) \le \Pr(\theta_1 \in J|\theta = \theta_2).$$

Chapter 10

Categorical Data and Nonparametric Methods

10.1 Tests of Goodness-of-Fit

Commentary

This section ends with a discussion of some issues related to the meaning of the χ^2 goodness-of-fit test for readers who want a deeper understanding of the procedure.

Solutions to Exercises

1. Let $Y = N_1$, the number of defective items, and let $\theta = p_1$, the probability that each item is defective. The level α_0 test requires us to choose c_1 and c_2 such that $\Pr(Y \le c_1|\theta = 0.1) + \Pr(Y \ge c_2|\theta = 0.1)$ is close to α_0. We can compute the probability that $Y = y$ for each $y = 0, \dots, 100$ and arrange the numbers from smallest to largest. The smallest values correspond to large values of y down to $y = 25$, then some values corresponding to small values of y start to appear in the list. The sum of the values reaches 0.0636 when $c_1 = 4$ and $c_2 = 16$. So $\alpha_0 = 0.0636$ is the smallest α_0 for which we would reject $H_0 : \theta = 0.1$ using such a test.

2.

$$\begin{aligned}
Q &= \sum_{i=1}^{k} \frac{(N_i - n/k)^2}{n/k} = \frac{k}{n}\sum_{i=1}^{k}\left(N_i^2 - 2\frac{n}{k}N_i + \frac{n^2}{k^2}\right) = \frac{k}{n}\left(\sum_{i=1}^{k} N_i^2 - 2\frac{n}{k}\sum_{i=1}^{k} N_i + \frac{n^2}{k}\right) \\
&= \frac{k}{n}\left(\sum_{i=1}^{k} N_i^2 - 2\frac{n^2}{k} + \frac{n^2}{k}\right) = \left(\frac{k}{n}\sum_{i=1}^{k} N_i^2\right) - n.
\end{aligned}$$

3. We obtain the following frequencies:

i	0	1	2	3	4	5	6	7	8	9
N_i	25	16	19	20	20	22	24	15	14	25

Since $P_i^0 = 1/10$ for every value of i, and $n = 200$, we find from Eq. (10.1.2) that $Q = 7.4$. If Q has the χ^2 distribution with 9 degrees of freedom, $\Pr(Q \ge 7.4) = 0.6$.

4. We obtain the following table:

	AA	Aa	aa
N_i	10	10	4
np_i^0	6	12	6

It is found from Eq. (10.1.2) that $Q = 11/3$. If Q has a χ^2 distribution with 2 degrees of freedom, then the value of $\Pr(Q \geq 11/3)$ is between 0.1 and 0.2.

5. (a) The number of successes is $n\overline{X}_n$ and the number of failures is $n(1-\overline{X}_n)$. Therefore,

$$\begin{aligned} Q &= \frac{(n\overline{X}_n - np_0)^2}{np_0} + \frac{[n(1-\overline{X}_n) - n(1-p_0)]^2}{n(1-p_0)} \\ &= n(\overline{X}_n - p_0)^2 \left(\frac{1}{p_0} + \frac{1}{1-p_0}\right) \\ &= \frac{n(\overline{X}_n - p_0)^2}{p_0(1-p_0)} \end{aligned}$$

(b) If $p = p_0$, then $E(\overline{X}_n) = p_0$ and $\text{Var}(\overline{X}_n) = p_0(1-p_0)/n$. Therefore, by the central limit theorem, the c.d.f. of

$$Z = \frac{\overline{X}_n - p_0}{[p_0(1-p_0)/n]^{1/2}}$$

converges to the c.d.f. of the standard normal distribution. Since $Q = Z^2$, the c.d.f. of Q will converge to the c.d.f. of the χ^2 distribution with 1 degree of freedom.

6. Here, $p_0 = 0.3$, $n = 50$, and $\overline{X}_n = 21/50$. By Exercise 5, $Q = 3.44$. If Q has a χ^2 distribution with 1 degree of freedom, then $\Pr(Q \geq 3.4)$ is slightly greater than 0.05.

7. We obtain the following table:

	$0 < x < 0.2$	$0.2 < x < 0.5$	$0.5 < x < 0.8$	$0.8 < x < 1.$
N_i	391	490	580	339
np_i^0	360	540	540	360

If Q has a χ^2 distribution with 3 degrees of freedom, then $\Pr(Q \geq 11.34) = 0.01$. Therefore, we should reject H_0 if $Q \geq 11.34$. It is found from Eq. (10.1.2) that Q = 11.5.

8. If Z denotes a random variable having a standard normal distribution and X denotes the height of a man selected at random from the city, then

$$\begin{aligned} \Pr(X < 66) &= \Pr(Z < -2) = 0.0227, \\ \Pr(66 < X < 67.5) &= \Pr(-2 < Z < -0.5) = 0.2858, \\ \Pr(67.5 < X < 68.5) &= \Pr(-0.5 < Z < 0.5) = 0.3830, \\ \Pr(68.5 < X < 70) &= \Pr(0.5 < Z < 2) = 0.2858, \\ \Pr(X > 70) &= \Pr(Z > 2) = 0.0227. \end{aligned}$$

Therefore, we obtain the following table:

	N_i	np_i^0
$x < 66$	18	11.35
$66 < x < 67.5$	177	142.9
$67.5 < x < 68.5$	198	191.5
$68.5 < x < 70$	102	142.9
$x > 70$	5	11.35

It is found from Eq. (10.1.2) that $Q = 27.5$. If Q has a χ^2 distribution with 4 degrees of freedom, then $\Pr(Q \geq 27.5)$ is much less than 0.005.

9. (a) The five intervals, each of which has probability 0.2, are as follows:

$$(-\infty, -0.842),\ (-0.842, -0.253),\ (-0.253, 0.253),\ (0.253, 0.842),\ (0.842, \infty).$$

We obtain the following table:

	N_i	np_i^0
$-\infty < x < -0.842$	15	10
$-0.842 < x < -0.253$	10	10
$-0.253 < x < 0.253$	7	10
$0.253 < x < 0.842$	12	10
$0.842 < x < \infty$	6	10

The calculated value of Q is 5.4. If Q has a χ^2 distribution with 4 degrees of freedom, then $\Pr(Q \geq 5.4) = 0.25$.

(b) The ten intervals, each of which has probability 0.1, are as given in the following table:

	N_i	np_i^0
$-\infty < x < -1.282$	8	5
$-1.282 < x < -0.842$	7	5
$-0.842 < x < -0.524$	3	5
$-0.524 < x < -0.253$	7	5
$-0.253 < x < 0$	5	5
$0 < x < 0.253$	2	5
$0.253 < x < 0.524$	5	5
$0.524 < x < 0.842$	7	5
$0.842 < x < 1.282$	2	5
$1.282 < x < \infty$	4	5

The calculated value of Q is 8.8. If Q has the χ^2 distribution with 9 degrees of freedom, then the value of $\Pr(Q \geq 8.8)$ is between 0.4 and 0.5.

10.2 Goodness-of-Fit for Composite Hypotheses

Commentary

The maximization of the log-likelihood in Eq. (10.2.5) could be performed numerically if one had appropriate software. The *R* functions `optim` and `nlm` can be used as described in the Commentary to Sec. 7.6 in this manual.

Solutions to Exercises.

1. There are many ways to perform a χ^2 test. For example, we could divide the real numbers into the intervals $(-\infty, 15]$, $(15, 30]$, $(30, 45]$, $(45, 60]$, $(60, 75]$, $(75, 90]$, $(90, \infty)$. The numbers of observations in these intervals are 14, 14, 4, 4, 3, 0, 2

   (a) The M.L.E.'s of the parameters of a normal distribution are $\hat{\mu} = 30.05$ and $\hat{\sigma^2} = 537.51$. Using the method of Chernoff and Lehmann, we compute two different p-values with 6 and 4 degrees of freedom. The probabilities for the seven intervals are 0.2581, 0.2410, 0.2413, 0.1613, 0.0719, 0.0214, 0.0049. The expected counts are 41 times each of these numbers. This makes $Q = 24.53$. The two p-values are both smaller than 0.0005.

   (b) The M.L.E.'s of the parameters of a lognormal distribution are $\hat{\mu} = 3.153$ and $\hat{\sigma^2} = 0.48111$. Using the method of Chernoff and Lehmann, we compute two different p-values with 6 and 4 degrees of freedom. The probabilities for the seven intervals are 0.2606, 0.3791, 0.1872, 0.0856, 0.0407, 0.0205, 0.0261. The expected counts are 41 times each of these numbers. This makes $Q = 5.714$. The two p-values are both larger than 0.2.

2. First, we must find the M.L.E. of Θ. From Eq. (10.2.5), ignoring the multinomial coefficient,

$$L(\theta) = \prod_{i=0}^{4} p_i^{N_i} = C\theta^{N_1+2N_2+3N_3+4N_4}(1-\theta)^{4N_0+3N_1+2N_2+N_3}, \text{ where } C = 4^{N_1}6^{N_2}4^{N_3}.$$

   Therefore,

$$\log L(\theta) = \log C + (N_1 + 2N_2 + 3N_3 + 4N_4)\log\theta + (4N_0 + 3N_1 + 2N_2 + N_3)\log(1-\theta).$$

   By solving the equation $\partial \log L(\theta)/\partial\theta = 0$, we obtain the result

$$\hat{\Theta} = \frac{N_1 + 2N_2 + 3N_3 + 4N_4}{4(N_0 + N_1 + N_2 + N_3 + N_4)} = \frac{N_1 + 2N_2 + 3N_3 + 4N_4}{4n}.$$

   It is found that $\hat{\Theta} = 0.4$. Therefore, we obtain the following table:

No. of Games	N_i	$N\pi_i(\hat{\Theta})$
0	33	25.92
1	67	69.12
2	66	69.12
3	15	30.72
4	19	5.12

   It is found from Eq. (10.2.4) that $Q = 47.81$. If Q has a χ^2 distribution with $5 - 1 - 1 = 3$ degrees of freedom, then $\Pr(Q \geq 47.81)$ is less than 0.005.

3. (a) It follows from Eqs. (10.2.2) and (10.2.6) that (aside from the multinomial coefficient)

$$\begin{aligned}\log L(\theta) &= (N_4 + N_5 + N_6)\log 2 + (2N_1 + N_4 + N_5)\log\theta_1 + (2N_2 + N_4 + N_6)\log\theta_2 \\ &\quad + (2N_3 + N_5 + N_6)\log(1 - \theta_1 - \theta_2).\end{aligned}$$

   By solving the equations

$$\frac{\partial \log L(\theta)}{\partial\theta_1} = 0 \quad \text{and} \quad \frac{\partial \log L(\theta)}{\partial\theta_2} = 0,$$

we obtain the results

$$\hat{\Theta}_1 = \frac{2N_1 + N_4 + N_5}{2n} \quad \text{and} \quad \hat{\Theta}_2 = \frac{2N_2 + N_4 + N_6}{2n},$$

where $n = \sum_{i=1}^{6} N_i$.

(b) For the given values, $n = 150, \hat{\Theta}_1 = 0.2,$ *and* $\hat{\Theta}_2 = 0.5$. Therefore, we obtain the following table:

i	N_i	$n\pi_i(\hat{\Theta})$
1	2	6
2	36	37.5
3	14	13.5
4	36	30
5	20	18
6	42	45

It is found from Eq. (10.2.4) that $Q = 4.37$. If Q has the χ^2 distribution with $6-1-2=3$ degrees of freedom, then the value of $\Pr(Q \geq 4.37)$ is approximately 0.226.

4. Suppose that X has the normal distribution with mean 67.6 and variance 1, and that Z has the standard normal distribution. Then:

$$\begin{aligned}
\pi_1(\hat{\Theta}) &= \Pr(X < 66) = \Pr(Z < -1.6) = 0.0548, \\
\pi_2(\hat{\Theta}) &= \Pr(66 < X < 67.5) = \Pr(-1.6 < Z < -0.1) = 0.4054, \\
\pi_3(\hat{\Theta}) &= \Pr(67.5 < \overline{X} < 68.5) = \Pr(-0.1 < Z < 0.9) = 0.3557, \\
\pi_4(\hat{\Theta}) &= \Pr(68.5 < X < 70) = \Pr(0.9 < Z < 2.4) = 0.1759, \\
\pi_5(\hat{\Theta}) &= \Pr(X > 70) = \Pr(Z > 2.4) = 0.0082.
\end{aligned}$$

Therefore, we obtain the following table:

i	N_i	$n\pi_i(\hat{\Theta})$
1	18	27.4
2	177	202.7
3	198	177.85
4	102	87.95
5	5	4.1

The value of Q is found from Eq. (10.2.4) to be 11.2. Since μ and σ^2 are estimated from the original observations rather than from the grouped data, the approximate distribution of Q when H_0 is true lies between the χ^2 distribution with 2 degrees of freedom and a χ^2 distribution with 4 degrees of freedom.

5. From the given observations, it is found that the M.L.E. of the mean Θ of the Poisson distribution is $\hat{\Theta} = \overline{X}_n = 1.5$. From the table of the Poisson distribution with $\Theta = 1.5$, we can obtain the values of $\pi_i(\hat{\Theta})$. In turn, we can then obtain the following table:

No. of tickets	N_i	$n\pi_i(\hat{\Theta})$
0	52	44.62
1	60	66.94
2	55	50.20
3	18	25.10
4	8	9.42
5 or more	7	3.70

It is found from Eq. (10.2.4) that $Q = 7.56$. Since $\hat{\Theta}$ is calculated from the original observations rather than from the grouped data, the approximate distribution of Q when H_0 is true lies between the χ^2 distribution with 4 degrees of freedom and the χ^2 distribution with 5 degrees of freedom. The two p-values for 4 and 5 degrees of freedom are 0.1091 and 0.1822.

6. The value of $\hat{\Theta} = \overline{X}_n$ can be found explicitly from the given data, and it equals 3.872. However, before carrying cut the χ^2 test, the observations in the bottom few rows of the table should be grouped together to obtain a single cell in which the expected number of observations is not too small. Reasonable choices would be to consider a single cell for the periods in which 11 or more particles were emitted (there would be 6 observations in that cell) or to consider a single cell for the periods in which 10 or more particles were emitted (there would be 16 observations in that cell). If the total number of cells after this grouping has been made is k, then under H_0 the statistic Q will have a distribution which lies between the χ^2 distribution with $k-2$ degrees of freedom and the χ^2 distribution with $k-1$ degrees of freedom. For example, with $k = 12$, the expected cell counts are

$$54.3, 210.3, 407.1, 525.3, 508.4, 393.7, 254.0, 140.5, 68.0, 29.2, 11.3, 5.8$$

The statistic Q is then 12.96. The two p-values for 10 and 11 degrees of freedom are 0.2258 and 0.2959.

7. There is no single correct answer to this problem. The M.L.E.'s $\hat{\mu} = \overline{X}_n$ and $\hat{\sigma}^2 = S_n^2/n$ should be calculated from the given observations. These observations should then be grouped into intervals and the observed number in each interval compared with the expected number in that interval if each of the 50 observations had the normal distribution with mean $\overline{X}_n$ and variance S_n^2/n. If the number of intervals is k, then when H_0 is true, the approximate distribution of the statistic Q will lie between the χ^2 distribution with $k-3$ degrees of freedom and the χ^2 distribution with $k-1$ degrees of freedom.

8. There is no single correct answer to this problem. The M.L.E. $\hat{\beta} = 1/\overline{X}_n$ of the parameter of the exponential distribution should be calculated from the given observations. These observations should then be grouped into intervals and the observed number in each interval compared with the expected number in that interval if each of the 50 observations had an exponential distribution with parameter $1/\overline{X}_n$. If the number of intervals is k, then when H_0 is true, the approximate distribution of the statistic Q will lie between a χ^2 distribution with $k-2$ degrees of freedom and the χ^2 distribution with $k-1$ degrees of freedom.

10.3 Contingency Tables

Solutions to Exercises.

1. Table S.10.1 contains the expected counts for this example. The value of the χ^2 statistic Q calculated

Table S.10.1: Expected cell counts for Exercise 1 of Sec. 10.3.

	Good grades	Athletic ability	Popularity
Boys	117.3	42.7	67.0
Girls	129.7	47.3	74.0

from these data is $Q = 21.5$. This should be compared to the χ^2 distribution with two degrees of freedom. The tail area can be calculated using statistical software as 2.2×10^{-5}.

2. $Q = \sum_{i=1}^{R}\sum_{j=1}^{C} \frac{(N_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} = \sum_{i=1}^{R}\sum_{j=1}^{C}\left(\frac{N_{ij}^2}{\hat{E}_{ij}} - 2N_{ij} + \hat{E}_{ij}\right) = \left(\sum_{i=1}^{R}\sum_{j=1}^{C}\frac{N_{ij}^2}{\hat{E}_{ij}}\right) - 2n + n$
$= \left(\sum_{i=1}^{R}\sum_{j=1}^{C}\frac{N_{ij}^2}{\hat{E}_{ij}}\right) - n.$

3. By Exercise 2,

$$Q = \sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i1}} + \sum_{i=1}^{R}\frac{N_{i2}^2}{\hat{E}_{i2}} - n.$$

But

$$\sum_{i=1}^{R}\frac{N_{i2}^2}{\hat{E}_{i2}} = \sum_{i=1}^{R}\frac{(N_{i+} - N_{i1})^2}{\hat{E}_{i2}} = \sum_{i=1}^{R}\frac{N_{i+}^2}{\hat{E}_{i2}} - 2\sum_{i=1}^{R}\frac{N_{i+}N_{i1}}{\hat{E}_{i2}} + \sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i2}}.$$

In the first two sums on the right, we let $\hat{E}_{i2} = N_{i+}N_{+2}/n$, and in the third sum we let $\hat{E}_{i2} = N_{+2}\hat{E}_{i1}/N_{+1}$. We then obtain

$$\sum_{i=1}^{R}\frac{N_{i2}^2}{\hat{E}_{i2}} = \frac{n}{N_{+2}}\sum_{i=1}^{R}N_{i+} - \frac{2n}{N_{+2}}\sum_{i=1}^{R}N_{i1} + \frac{N_{+1}}{N_{+2}}\sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i1}} = \frac{n^2}{N_{+2}} - 2n\frac{N_{+1}}{N_{+2}} + \frac{N_{+1}}{N_{+2}}\sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i1}}.$$

It follows that

$$Q = \left(1 + \frac{N_{+1}}{N_{+2}}\right)\sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i1}} + \frac{n}{N_{+2}}(n - 2N_{+1} - N_{+2}).$$

Since $n = N_{+1} + N_{+2}$,

$$Q = \frac{n}{N_{+2}}\sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i1}} - \frac{n}{N_{+2}}N_{+1}.$$

4. The values of $\hat{E}_{ij}$ are as given in the following table:

8	32
12	48

The value of Q is found from Eq. (10.3.4) to be 25/6. If Q has a χ^2 distribution with 1 degree of freedom, then $\Pr(Q \geq 25/6)$ lies between 0.025 and 0.05.

5. The values of $\hat{E}_{ij}$ are as given in the following table.

77.27	94.35	49.61	22.77
17.73	21.65	11.39	5.23

The value of Q is found from Eq. (10.3.4) to be 8.6. If Q has the χ^2 distribution with $(2-1)(4-1) = 3$ degrees of freedom, then $\Pr(Q \geq 8.6)$ lies between 0.025 and 0.05.

6. The values of $\hat{E}_{ij}$ are as given in the following table:

7.5	7.5
14.5	14.5

The value of Q is found from Eq. (10.3.4) to be 0.91. If Q has the χ^2 distribution with 1 degree of freedom, then $\Pr(Q \geq 0.91)$ lies between 0.3 and 0.4.

7. (a) The values of p_{i+} and p_{+j} are the marginal totals given in the following table:

			0.3
			0.3
			0.4
0.5	0.3	0.2	1.0

It can be verified that $p_{ij} = p_{i+}p_{+j}$ for each of the 9 entries in the table. It can be seen in advance that this relation will be satisfied for every entry in the table because it can be seen that the three rows of the table are proportional to each other or, equivalently, that the three columns are proportional to each other.

(b) Here is one example of a simulated data set

44	32	16	92
45	25	15	85
63	33	27	123
152	90	58	300

(c) The statistic Q calculated by any student from Eq. (10.3.4) will have the χ^2 distribution with $(3-1)(3-1) = 4$ degrees of freedom. For the data in part (b), the table of $\hat{E}_{ij}$ values is

46.6	27.6	17.8
43.1	25.5	16.4
62.3	36.9	23.8

The value of Q is then 2.105. The p-value 0.7165.

8. To test whether the values obtained by n different students form a random sample of size n from a χ^2 distribution with 4 degrees of freedom, follow these steps: (1) Partition the positive part of the real line into k intervals; (2) Determine the probabilities $p_1^0, \ldots, p_k^0$ of these intervals for the χ^2 distribution with 4 degrees of freedom; (3) Calculate the value of the statistic Q given by Eq. (10.1.2). If the hypothesis H_0 is true, this statistic Q will have approximately the χ^2 distribution with $k-1$ degrees of freedom.

9. Let N_{ijk} denote the number of observations in the random sample that fall into the (i, j, k) cell, and let

$$\begin{aligned} N_{i++} &= \sum_{j=1}^{C}\sum_{k=1}^{T} N_{ijk}, N_{+j+} = \sum_{i=1}^{R}\sum_{k=1}^{T} N_{ijk}, \\ N_{++k} &= \sum_{i=1}^{R}\sum_{j=1}^{C} N_{ijk}. \end{aligned}$$

Then the M.L.E.'s are

$$\hat{p}_{i++} = \frac{N_{i++}}{n}, \hat{p}_{+j+} = \frac{N_{+j+}}{n}, \hat{p}_{++k} = \frac{N_{++k}}{n}.$$

Therefore, when H_0 is true,

$$\hat{E}_{ijk} = n\hat{p}_{i++}\hat{p}_{+j+}\hat{p}_{++k} = \frac{N_{i++}N_{+j+}N_{++k}}{n^2}.$$

Since $\sum_{i=1}^{R} \hat{p}_{i++} = \sum_{j=1}^{C} \hat{p}_{+j+} = \sum_{k=1}^{T} \hat{p}_{++k} = 1$, the number of parameters that have been estimated is $(R-1)+(C-1)+(T-1) = R+C+T-3$. Therefore, when H_0 is true, the approximate distribution of

$$Q = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{T} \frac{(N_{ijk} - \hat{E}_{ijk})^2}{\hat{E}_{ijk}}$$

will be the χ^2 distribution for which the number of degrees of freedom is $RCT-1-(R+C+T-3) = RCT - R - C - T + 2$.

10. The M.L.E.'s are

$$\hat{p}_{ij+} = \frac{N_{ij+}}{n} \quad \text{and} \quad \hat{p}_{++k} = \frac{N_{++k}}{n}.$$

Therefore, when H_0 is true,

$$\hat{E}_{ijk} = n\hat{p}_{ij+}\,\hat{p}_{++k} = \frac{N_{ij+}N_{++k}}{n}.$$

Since $\sum_{i=1}^{R}\sum_{j=1}^{C} \hat{p}_{ij+} = \sum_{k=1}^{T} \hat{p}_{++k} = 1$, the number of parameters that have been estimated is $(RC-1) + (T-1) = RC + T - 2$. Therefore, when H_0 is true, the approximate distribution of

$$Q = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{T} \frac{(N_{ijk} - \hat{E}_{ijk})^2}{\hat{E}_{ijk}}.$$

will be the χ^2 distribution for which the number of degrees of freedom is $RCT - 1 - (RC + T - 2) = RCT - RC - T + 1$.

10.4 Tests of Homogeneity

Solutions to Exercises.

1. Table S.10.2 contains the expected cell counts. The value of the χ^2 statistic is $Q = 18.8$, which should

Table S.10.2: Expected cell counts for Exercise 1 of Sec. 10.4.

	Good grades	Athletic ability	Popularity
Rural	77.0	28.0	44.0
Suburban	78.0	28.4	44.5
Urban	92.0	33.5	52.5

be compared to the χ^2 distribution with four degrees of freedom. The tail area is 8.5×10^{-4}.

2. The value of the statistic Q given by Eqs. (10.4.3) and (10.4.4) is 7.57. If Q has a χ^2 distribution with $(2-1)(3-1) = 2$ degrees of freedom, then $\Pr(Q \geq 7.57) < 0.025$.

3. The value of the statistic Q given by Eqs. (10.4.3) and (10.4.4) is 18.9. If Q has the χ^2 distribution with $(4-1)(5-1) = 12$ degrees of freedom, then the value of $\Pr(Q \geq 18.9)$ lies between 0.1 and 0.05.

4. The table to be analyzed is as follows:

Person	Hits	Misses
1	8	9
2	4	12
3	7	3
4	13	11
5	10	6

The value of the statistic Q given by Eqs. (10.4.3) and (10.4.4) is 6.8. If Q has the χ^2 distribution with $(5-1)(2-1) = 4$ degrees of freedom, then the value of $\Pr(Q \geq 6.8)$ lies between 0.1 and 0.2.

5. The correct table to be analyzed is as follows:

Supplier	Defectives	Nondefectives
1	1	14
2	7	8
3	7	8

The value of Q found from this table is 7.2. If Q has the χ^2 distribution with $(3-1)(2-1) = 2$ degrees of freedom, then $\Pr(Q \geq 7.2) < 0.05$.

6. The proper table to be analyzed is as follows:

		After demonstration		
		Hit	Miss	
Before demonstration	Hit			27
	Miss			73
		35	65	

Although we are given the marginal totals, we are not given the entries in the table. If we were told the value in just a single cell, such as the number of students who hit the target both before and after the demonstration, we could fill in the rest of the table.

7. The proper table to be analyzed is as follows:

		After meeting		
		Favors A	Favors B	No preference
Before meeting	Favors A			
	Favors B			
	No preference			

Each person who attended the meeting can be classified in one of the nine cells of this table. If a speech was made on behalf of A at the meeting, we could evaluate the effectiveness of the speech by comparing the numbers of persons who switched from favoring B or having no preference before the meeting to favoring A after the meeting with the number who switched from favoring A before the meeting to one of the other positions after the meeting.

10.5 Simpson's Paradox

Solutions to Exercises

1. If population II has a relatively high proportion of men and population I has a relatively high proportion of women, then the indicated result will occur. For example, if 90 percent of population II are men and 10 percent are women, then the proportion of population II with the characteristic will be $(.9)(.6) + (.1)(.1) = .55$. If 10 percent of population I are men and 90 percent are women, then the proportion of population I with the characteristic will be only $(.1)(.8) + (.9)(.3) = .35$.

2. Each of these equalities holds if and only if A and B are independent events.

3. Assume that $\Pr(B|A) = \Pr(B|A^c)$. This means that A and B are independent. According to the law of total probability, we can write

$$\begin{aligned}\Pr(I|B) &= \Pr(I|A\cap B)\Pr(A|B) + \Pr(I|A^c\cap B)\Pr(A^c|B)\\ &= \Pr(I|A\cap B)\Pr(A) + \Pr(I|A^c\cap B)\Pr(A^c),\end{aligned}$$

where the last equality follows from the fact that A and B are independent. Similarly,

$$\Pr(I|B^c) = \Pr(I|A\cap B^c)\Pr(A) + \Pr(I|A^c\cap B^c)\Pr(A^c).$$

If the first two inequalities in (10.5.1) hold then the weighted average of the left sides of the inequalities must be larger than the same weighted average of the right sides. In particular,

$$\Pr(I|A\cap B)\Pr(A) + \Pr(I|A^c\cap B)\Pr(A^c) > \Pr(I|A\cap B^c)\Pr(A) + \Pr(I|A^c\cap B^c)\Pr(A^c).$$

But, we have just shown that this last equality is equivalent to $\Pr(I|B) > \Pr(I|B^c)$, which means that the third inequality cannot hold if the first two hold.

4. Define A to be the event if that a subject is a man, A^c the event that a subject is a woman, B the event that a subject receives treatment I, and B^c the event that a subject receives treatment II. Then the relation to be proved here is precisely the same as the relation that was proved in Exercise 2 in symbols.

5. Suppose that the first two inequalities in (10.5.1) hold, and that $\Pr(A|B) = \Pr(A|B^c)$, Then

$$\begin{aligned}\Pr(I|B) &= \Pr(I \mid A\cap B)\Pr(A \mid B) + \Pr(I \mid A^c\cap B)\Pr(A^c \mid B)\\ &> \Pr(I \mid A\cap B^c)\Pr(A \mid B) + \Pr(I \mid A^c\cap B^c)\Pr(A^c \mid B)\\ &= \Pr(I \mid A\cap B^c)\Pr(A \mid B^c) + \Pr(I \mid A^c\cap B^c)\Pr(A^c \mid B^c)\\ &= \Pr(I \mid B^c).\end{aligned}$$

Hence, the final inequality in (10.5.1) must be reversed.

6. This result can be obtained if the colleges that admit a relatively small proportion of their applicants receive a relatively large proportion of female applicants and the colleges that admit a relatively large proportion of their applicants receive a relatively small proportion of female applicants. As a specific example, suppose that the data are as given in the following table:

College	Proportion of total University applicants	Proportion male	Proportion female	Proportion of males admitted	Proportion of females admitted
1	.1	.9	.1	.32	.56
2	.1	.9	.1	.32	.56
3	.2	.8	.2	.32	.56
4	.2	.8	.2	.32	.56
5	.4	.1	.9	.05	.10

This table indicates, for example, that College 1 receives 10 percent of all the applications submitted to the university, that 90 percent of the applicants to College 1 are male and 10 percent are female, that 32 percent of the male applicants to College 1 are admitted, and that 56 percent of the female applicants are admitted. It can be seen from the last two columns of this table that in each college the proportion of females admitted is larger than the proportion of males admitted. However, in the whole university, the proportion of males admitted is

$$\frac{(.1)(.9)(.32) + (.1)(.9)(.32) + (.2)(.8)(.32) + (.2)(.8)(.32) + (.4)(.1)(.05)}{(.1)(.9) + (.1)(.9) + (.2)(.8) + (.2)(.8) + (.4)(.1)} = .3$$

and the proportion of females admitted is

$$\frac{(.1)(.1)(.56) + (.1)(.1)(.56) + (.2)(.2)(.56) + (.2)(.2)(.56) + (.4)(.9)(.10)}{(.1)(.1) + (.1)(.1) + (.2)(.2) + (.2)(.2) + (.4)(.9)} = .2\,.$$

7. (a) Table S.10.3 shows the proportions helped by each treatment in the four categories of subjects. The proportion helped by Treatment II is higher in each category.

Table S.10.3: Table for Exercise 7a in Sec. 10.5.

	Proportion helped	
Category	Treatment I	Treatment II
Older males	.200	.667
Younger males	.750	.800
Older females	.167	.286
Younger females	.500	.640

(b) Table S.10.4 shows the proportions helped by each treatment in the two aggregated categories. Treatment I helps a larger proportion in each of the two categories

Table S.10.4: Table for Exercise 7b in Sec. 10.5.

	Proportion helped	
Category	Treatment I	Treatment II
Older subjects	.433	.400
Younger subjects	.700	.667

(c) When all subjects are grouped together, the proportion helped by Treatment I is $200/400 = 0.5$, while the proportion helped by Treatment II is $240/400 = 0.6$.

10.6 Kolmogorov-Smirnov Tests

Commentary

This section is optional. However, some of the topics discussed here are useful in Chapter 12. In particular, the bootstrap in Sec. 12.6 makes much use of the sample c.d.f. Some of the plots done after Markov chain Monte Carlo also make use of the sample c.d.f. The crucial material is at the start of Sec. 10.6. The Glivenko-Cantelli lemma, together with the asymptotic distribution of the Kolmogorov-Smirnov test statistic in Table 10.32 are useful if one simulates sample c.d.f.'s and wishes to compute simulation standard errors for the entire sample c.d.f.

Empirical c.d.f.'s can be computed by the *R* function `ecdf`. The argument is a vector of data values. The result is an *R* function that computes the empirical c.d.f. at its argument. For example, if `x` has a sample of observations, then `empd.x=ecdf(x)` will create a function `empd.x` which can be used to compute values of the empirical c.d.f. For example, `empd.x(3)` will be the proportion of the sample with values at most 3. Kolmogorov-Smirnov tests can be performed using the *R* function `ks.test`. The first argument is a vector of data values. The second argument depends on whether one is doing a one-sample or two-sample test. In the two-sample case, the second argument is the second sample. In the one-sample case, the second argument is the name of a function that will compute the hypothesized c.d.f. If that function has any additional arguments, they can be provided next or named explicitly later in the argument list.

Solutions to Exercises.

1. $F_n(x) = 0$ for $x < y$, and $F_n(y_1) = 0.2$. Suppose first that $F(y_1) \geq 0.1$. Since F is continuous, the values of $F(x)$ will be arbitrarily close to $F(y_1)$ for x arbitrarily close to y_1. Therefore, $\sup_{x<y_1} |F_n(x) - F(x)| = F(y_1) \geq 0.1$, and it follows that $D_n \geq 0.1$. Suppose next that $F(y_1) \leq 0.1$. Since $F_n(y_1) = 0.2$, it follows that $| F_n(y_1) - F(y_1) | \geq 0.1$. Therefore, it is again true that $D_n \geq 0.1$. We can now conclude that it must always be true that $D_n \geq 0.1$. If the values of $F(y_i)$ are as specified in the second part of the exercise, for $i = 1, \dots, 5$, then:

$$| F_n(x) - F(x) | = \begin{cases} \ F(x) \leq 0.1 & \text{for } x < y_1, \\ \ 0.2 - 0.1 = 0.1 & \text{for } x = y_1, \\ | F(x) - 0.2 | \leq 0.1 & \text{for } y_1 < x < y_2, \\ \ 0.4 - 0.3 = 0.1 & \text{for } x = y_2, \\ | F(x) - 0.4 | \leq 0.1 & \text{for } y_2 < x < y_3, \\ \text{etc.} \end{cases}$$

Hence, $D_n = \sup_{-\infty < x < \infty} | F_n(x) - F(x) | = 0.1$.

2. $$F_n(x) = \begin{cases} 0 & \text{for } x < y_1, \\ 0.2 & \text{for } y_1 \leq x < y_2, \\ 0.4 & \text{for } y_2 \leq x < y_3, \\ 0.6 & \text{for } y_3 \leq x < y_4, \\ 0.8 & \text{for } y_4 \leq x < y_5, \\ 1 & \text{for } x \geq y_5. \end{cases}$$

If F satisfies the inequalities given in the exercise, then $| F_n(x) - F(x) | \leq 0.2$ for every value of x. Hence, $D_n \leq 0.2$. Conversely, if $F(y_i) > 0.2i$ for some value of i, then $F(x) - F_n(x) > 0.2$ for values of x approaching y_i from below. Hence, $D_n > 0.2$. Also, if $F(y_i) < 0.2(i - 1)$ for some value of i, then $F_n(y_i) - F(y_i) > 0.2$. Hence, again $D_n > 0.2$.

3. The largest value of the difference between the sample c.d.f. and the c.d.f. of the normal distribution with mean 3.912 and variance 0.25 occurs right before $x = 4.22$, the 12th observation. For x just below 4.22, the sample c.d.f. is $F_n(x) = 0.48$, while the normal c.d.f. is $\Phi([4.22 - 3.912]/0.5) = 0.73$. The difference is $D_n^* = 0.25$. The Kolmogorov-Smirnov test statistic is $23^{1/2} \times 0.25 = 1.2$. The tail area can be found from Table 10.32 as 0.11.

4. When the observations are ordered, we obtain Table S.10.5. The maximum value of $\mid F_n(x) - F(x) \mid$

Table S.10.5: Table for Exercise 4 in Sec. 10.6.

i	$y_i = F(y_i)$	$F_n(y_i)$	i	$y_i = F(y_i)$	$F_n(y_i)$
1	.01	.04	14	.41	.56
2	.06	.08	15	.42	.60
3	.08	.12	16	.48	.64
4	.09	.16	17	.57	.68
5	.11	.20	18	.66	.72
6	.16	.24	19	.71	.76
7	.22	.28	20	.75	.80
8	.23	.32	21	.78	.84
9	.29	.36	22	.79	.88
10	.30	.40	23	.82	.92
11	.35	.44	24	.88	.96
12	.38	.48	25	.90	1.00
13	.40	.52			

occurs at $x = y_{15}$ where its value is $0.60 - 0.42 = 0.18$. Since $n = 25, n^{1/2}D_n{}^* = 0.90$. From Table 10.32, $H(0.90) = 0.6073$. Therefore, the tail area corresponding to the observed value of $D_n{}^*$ is $1 - 0.6073 = 0.3927$.

5. Here,

$$F(x) = \begin{cases} \dfrac{3}{2}x & \text{for } 0 < x \leq 1/2, \\ \dfrac{1}{2}(1+x) & \text{for } \frac{1}{2} < x < 1. \end{cases}$$

Therefore, we obtain Table S.10.6. The supremum of $\mid F_n(x) - F(x) \mid$ occurs as $x \to y_{18}$ from below.

Table S.10.6: Table for Exercise 5 in Sec. 10.6.

i	y_i	$F(y_i)$	$F_n(y_i)$	i	y_i	$F(y_i)$	$F_n(y_i)$
1	.01	.015	.04	14	.41	.615	.56
2	.06	.09	.08	15	.42	.63	.60
3	.08	.12	.12	16	.48	.72	.64
4	.09	.135	.16	17	.57	.785	.68
5	.11	.165	.20	18	.66	.83	.72
6	.16	.24	.24	19	.71	.855	.76
7	.22	.33	.28	20	.75	.875	.80
8	.23	.345	.32	21	.78	.89	.84
9	.29	.435	.36	22	.79	.895	.88
10	.30	.45	.40	23	.82	.91	.92
11	.35	.525	.44	24	.88	.94	.96
12	.38	.57	.48	25	.90	.95	1.00
13	.40	.60	.52				

Here, $F(x) \to 0.83$ while $F_n(x)$ remains at 0.68. Therefore, $D_n{}^* = 0.83 - 0.68 = 0.15$. It follows that

$n^{1/2}D_n{}^* = 0.75$ and, from Table 10.32, $H(0.75) = 0.3728$. Therefore, the tail area corresponding to the observed value of $D_n{}^*$ is $1 - 0.3728 = 0.6272$.

6. Since the p.d.f. of the uniform distribution is identically equal to 1, the value of the joint p.d.f. of the 25 observations under the uniform distribution has the value $L_1 = 1$. Also, sixteen of the observations are less than 1/2 and nine are greater than 1/2. Therefore, the value of the joint p.d.f. of the observations under the other distribution is $L_2 = (3/2)^{16}(1/2)^9 = 1.2829$. The posterior probability that the observations came from a uniform distribution is

$$\frac{\frac{1}{2}L_1}{\frac{1}{2}L_1 + \frac{1}{2}L_2} = 0.438$$

7. We first replace each observed value x_i by the value $(x_i - 26)/2$. Then, under the null hypothesis, the transformed values will form a random sample from a standard normal distribution. When these transformed values are ordered, we obtain Table S.10.7. The maximum value of $\mid F_n(x) - \Phi(x) \mid$

Table S.10.7: Table for Exercise 7 in Sec. 10.6.

i	y_i	Φy_i	$F_n(y_i)$	i	y_i	$\Phi(y_i)$	$F_n(y_i)$
1	−2.2105	.0136	.02	26	−0.010	.4960	.52
2	−1.9265	.0270	.04	27	−0.002	.4992	.54
3	−1.492	.0675	.06	28	$^1/_4$0.010	.5040	.56
4	−1.3295	.0919	.08	29	$^1/_4$0.1515	.5602	.58
5	−1.309	.0953	.10	30	$^1/_4$0.258	.6018	.60
6	−1.2085	.1134	.12	31	$^1/_4$0.280	.6103	.62
7	−1.1995	.1152	.14	32	$^1/_4$0.3075	.6208	.64
8	−1.125	.1307	.16	33	$^1/_4$0.398	.6547	.66
9	−1.0775	.1417	.18	34	$^1/_4$0.4005	.6556	.68
10	−1.052	.1464	.20	35	$^1/_4$0.4245	.6645	.70
11	−0.961	.1682	.22	36	$^1/_4$0.482	.6851	.72
12	−0.8415	.2001	.24	37	$^1/_4$0.614	.7304	.74
13	−0.784	.2165	.26	38	$^1/_4$0.689	.7546	.76
14	−0.767	.2215	.28	39	$^1/_4$0.7165	.7631	.78
15	−0.678	.2482	.30	40	$^1/_4$0.7265	.7662	.80
16	−0.6285	.2648	.32	41	$^1/_4$0.9262	.8320	.82
17	−0.548	.2919	.34	42	$^1/_4$1.0645	.8564	.84
18	−0.456	.3242	.36	43	$^1/_4$1.120	.8686	.86
19	−0.4235	.3359	.38	44	$^1/_4$1.176	.8802	.88
20	−0.340	.3669	.40	45	$^1/_4$1.239	.8923	.90
21	−0.3245	.3728	.42	46	$^1/_4$1.4615	.9281	.92
22	−0.309	.3787	.44	47	$^1/_4$1.6315	.9487	.94
23	−0.266	.3951	.46	48	$^1/_4$1.7925	.9635	.96
24	−0.078	.4689	.48	49	$^1/_4$1.889	.9705	.98
25	−0.0535	.4787	.50	50	$^1/_4$2.216	.9866	1.00

is attained at $x = y_{23}$ and its value is 0.0649. Since $n = 50, n^{1/2}D_n{}^* = 0.453$. It follows from Table 10.32 that $H(0.453) = 0.02$. Therefore, the tail area corresponding to the observed value of $D_n{}^*$ is $1 - 0.02 = 0.98$.

8. We first replace each observed value x_i by the value $(x_i - 24)/2$. Then, under the null hypothesis, the transformed values will form a random sample from a standard normal distribution. Each of the transformed values will be one unit larger than the corresponding transformed value in Exercise 7. The ordered values are therefore omitted from the tabulation in Table S.10.8. The supremum of $\mid F_n(x) - \Phi(x) \mid$ occurs as $x \to y_{18}$ from below. Here, $\Phi(x) \to 0.7068$ while $F_n(x)$ remains at 0.34.

Table S.10.8: Table for Exercise 8 in Sec. 10.6.

i	$\Phi(y_i)$	$F_n(y_i)$	i	$\Phi(y_i)$	$F_n(y_i)$
1	.1130	.02	26	.8389	.52
2	.1779	.04	27	.8408	.54
3	.3114	.06	28	.8437	.56
4	.3710	.08	29	.8752	.58
5	.3787	.10	30	.8958	.60
6	.4174	.12	31	.8997	.62
7	.4209	.14	32	.9045	.64
8	.4502	.16	33	.9189	.66
9	.4691	.18	34	.9193	.68
10	.4793	.20	35	.9229	.70
11	.5136	.22	36	.9309	.72
12	.5630	.24	37	.9467	.74
13	.5856	.26	38	.9544	.76
14	.5921	.28	39	.9570	.78
15	.6263	.30	40	.9579	.80
16	.6449	.32	41	.9751	.82
17	.6743	.34	42	.9805	.84
18	.7068	.36	43	.9830	.86
19	.7178	.38	44	.9852	.88
20	.7454	.40	45	.9875	.90
21	.7503	.42	46	.9931	.92
22	.7552	.44	47	.9958	.94
23	.7685	.46	48	.9974	.96
24	.8217	.48	49	.9980	.98
25	.8280	.50	50	.9993	1.00

Therefore, ${D_n}^* = 0.7068 - 0.34 = 0.3668$. It follows that $n^{1/2}{D_n}^* = 2.593$ and, from Table 10.32, $H(2.593) = 1.0000$. Therefore, the tail area corresponding to the observed value of ${D_n}^*$ is 0.0000.

9. We shall denote the 25 ordered observations in the first sample by $x_1 < \cdots < x_{25}$ and shall denote the 20 ordered observations in the second sample by $y_1 < \cdots < y_{20}$. We obtain Table S.10.9. The maximum value of $\mid F_m(x) - G_n(x) \mid$ is attained at $x = -0.39$, where its value is $0.32 - 0.05 = 0.27$. Therefore, $D_{mn} = 0.27$ and, since $m = 25$ and $n = 20$, $(mn/[m+n])^{1/2}\, D_{mn} = 0.9$. From Table 10.32, $H(0.9) = 0.6073$. Hence, the tail area corresponding to the observed value of D_{mn} is $1 - 0.6073 = 0.3927$.

10. We shall add 2 units to each of the values in the first sample and then carry out the same procedure as in Exercise 9. We now obtain Table S.10.10. The maximum value of $\mid F_m(x) - G_n(x) \mid$ is attained at $x = 1.56$, where its value is $0.80 - 0.24 = 0.56$. Therefore, $D_{mn} = 0.56$ and $(mn/[m+n])^{1/2}\, D_{mn} = 1.8667$. From Table 10.32, $H(1.8667) = 0.998$. Therefore, the tail area corresponding to the observed value of D_{mn} is $1 - 0.998 = 0.002$.

11. We shall multiply each of the observations in the second sample by 3 and then carry out the same procedure as in Exercise 9. We now obtain Table S.10.11. The maximum value of $\mid F_m(x) - G_n(x) \mid$ is attained at $x = 1.06$, where its value is $0.80 - 0.30 = 0.50$. Therefore, $D_{mn} = 0.50$ and $(mn/[m+n])^{1/2}\, D_{mn} = 1.667$. From Table 10.32, $H(1.667) = 0.992$. Therefore, the tail area corresponding to the observed value of D_{mn} is $1 - 0.992 = 0.008$.

12. The maximum difference between the c.d.f. of the normal distribution with mean 3.912 and variance 0.25 and the empirical c.d.f. of the observed data is $D_n^* = 0.2528$ which occurs at the observation 4.22 where the empirical c.d.f. jumps from $11/23 = 0.4783$ to $12/23 = 0.5217$ and the normal c.d.f. equals $\Phi([4.22 - 3.912]/0.5) = 0.7311$. We now compare $(23)^{1/2} D_n^* = 1.2123$ to Table 10.32, where we find that $H(1.2123) \approx 0.89$. The tail area (p-value) is then about 0.11.

Table S.10.9: Table for Exercise 9 in Sec. 10.6.

x_i	y_j	$F_m(x)$	$G_n(x)$	x_i	y_j	$F_m(x)$	$G_n(x)$
−2.47		.04	0	0.51		.60	.45
−1.73		.08	0		0.52	.60	.50
−1.28		.12	0	0.59		.64	.50
−0.82		.16	0	0.61		.68	.50
−0.74		.20	0	0.64		.72	.50
	−0.71	.20	.05		0.66	.72	.55
−0.56		.24	.05		0.70	.72	.60
−0.40		.28	.05		0.96	.72	.65
−0.39		.32	.05	1.05		.76	.65
	−0.37	.32	.10	1.06		.80	.65
−0.32		.36	.10	1.09		.84	.65
	−0.30	.36	.15	1.31		.88	.65
	−0.27	.36	.20		1.38	.88	.70
−0.06		.40	.20		1.50	.88	.75
	0.00	.40	.25		1.56	.88	.80
0.05		.44	.25	1.64		.92	.80
0.06		.48	.25		1.66	.92	.85
	0.26	.48	.30	1.77		.96	.85
0.29		.52	.30		2.20	.96	.90
0.31		.56	.30		2.31	.96	.95
	0.36	.56	.35	2.36		1.00	.95
	0.38	.56	.40		3.29	1.00	1.00
	0.44	.56	.45				

Table S.10.10: Table for Exercise 10 in Sec. 10.6.

x_i	y_j	$F_m(x)$	$G_n(x)$	x_i	y_j	$F_m(x)$	$G_n(x)$
	−0.71	0	.05	1.61		.32	.80
−0.47		.04	.05		1.66	.32	.85
	−0.37	.04	.10	1.68		.36	.85
	−0.30	.04	.15	1.94		.40	.85
	−0.27	.04	.20	2.05		.44	.85
	0.00	.04	.25	2.06		.48	.85
	0.26	.04	.30		2.20	.48	.90
0.27		.08	.30	2.29		.52	.90
	0.36	.08	.35	2.31	2.31	.56	.95
	0.38	.08	.40	2.51		.60	.95
	0.44	.08	.45	2.59		.64	.95
	0.52	.08	.50	2.61		.68	.95
	0.66	.08	.55	2.64		.72	.95
	0.70	.08	.60	3.05		.76	.95
0.72		.12	.60	3.06		.80	.95
	0.96	.12	.65	3.09		.84	.95
1.18		.16	.65		3.29	.84	1.00
1.26		.20	.65	3.31		.88	1.00
	1.38	.20	.70	3.64		.92	1.00
1.44		.24	.70	3.77		.96	1.00
	1.50	.24	.75	4.36		1.00	1.00
	1.56	.24	.80				
1.60		.28	.80				

Table S.10.11: Table for Exercise 11 in Sec. 10.6

x_i	y_j	$F_m(x)$	$G_n(x)$	x_i	y_j	$F_m(x)$	$G_n(x)$
−2.47		.04	0		0.78	.72	.30
	−2.13	.04	.05	1.05		.76	.30
−1.73		.08	.05	1.06		.80	.30
−1.28		.12	.05		1.08	.80	.35
	−1.11	.12	.10	1.09		.84	.35
	−0.90	.12	.15		1.14	.84	.40
−0.82		.16	.15	1.31		.88	.40
	−0.81	.16	.20		1.32	.88	.45
−0.74		.20	.20		1.56	.88	.50
−0.56		.24	.20	1.64		.92	.50
−0.40		.28	.20	1.77		.96	.50
−0.39		.32	.20		1.98	.96	.55
−0.32		.36	.20		2.10	.96	.60
−0.06		.40	.20	2.36		1.00	.60
	0.00	.40	.25		2.88	1.00	.65
0.05		.44	.25		4.14	1.00	.70
0.06		.48	.25		4.50	1.00	.75
0.29		.52	.25		4.68	1.00	.80
0.31		.56	.25		4.98	1.00	.85
0.51		.60	.25		6.60	1.00	.90
0.59		.64	.25		6.93	1.00	.95
0.61		.68	.25		9.87	1.00	1.00
0.64		.72	.25				

10.7 Robust Estimation

Commentary

In recent years, interest has grown in the use of robust statistical methods. Although many robust methods are more suitable for advanced courses, this section introduces some robust methods that can be understood at the level of the rest of this text. This includes M-estimators of a location parameter.

The software R contains some functions that can be used for robust estimation. The function `quantile` computes sample quantiles. The first argument is a vector of observed values. The second argument is a vector of probabilities for the desired quantiles. For example `quantile(x,c(0.25,0.75))` computes the sample quartiles of the data `x`. The function `median` computes the sample median. The function `mad` computes the median absolute deviation of a sample. If you issue the command `library(MASS)`, some additional functions become available. One such function is `huber`, which computes M-estimators as on page 673 with $\hat{\sigma}$ equal to the median absolute deviation. The first argument is the vector of data values, and the second argument is k, in the notation of the text. To find the M-estimator with a general $\hat{\sigma}$, replace the second argument by $k\hat{\sigma}$ divided by the mean absolute deviation of the data.

Solutions to Exercises.

1. The observed values ordered from smallest to largest are 2.1, 2.2, 21.3, 21.5, 21.7, 21.7, 21.8, 22.1, 22.1, 22.2, 22.4, 22.5, 22.9, 23.0, 63.0.

   (a) The sample mean is the average of the numbers, 22.17.

   (b) The trimmed mean for a given value of k is found by dropping k values from each end of this ordered sequence and averaging the remaining values. In this problem we get

k	1	2	3	4
kth level trimmed mean	20.57	22.02	22	22

   (c) The sample median is the middle observation, 22.1.

   (d) The median absolute deviation is 0.4. Suppose that we start iterating with the sample average 22.17. The 7th and 8th iterations are both 22.

2. The observed values ordered from smallest to largest are $-2.40, -2.00$, -0.11, 0.00, 0.03, 0.10, 0.12, 0.23, 0.24, 0.24, 0.36, 0.69, 1.24, 1.78.

   (a) The sample mean is the average of these values, 0.0371.

   (b) The trimmed mean for a given value of k is found by dropping k values from each end of this ordered sequence and averaging the remaining values. In this problem we get

k	1	2	3	4
kth level trimmed mean	0.095	0.19	0.165	0.16

   (c) Since the number of observed values is even, the sample median is the average of the two middle values 0.12 and 0.23, which equals 0.175.

   (d) The median absolute deviation is 0.18. Suppose that we start iterating with the sample average 0.0371. The 9th and 10th iterations are both 0.165.

3. The distribution of $\tilde{\theta}_{.5,n}$ will be approximately normal with mean θ and standard deviation $1/[2n^{1/2}f(\theta)]$. In this exercise,

$$f(x) = \frac{1}{\sqrt{2\pi}} \quad \exp\left[-\frac{1}{2}(x-\theta)^2\right].$$

Hence, $f(\theta) = 1/\sqrt{2\pi}$. Since $n = 100$, the standard deviation of the approximate distribution of $\tilde{\theta}_{.5,n}$ is $\sqrt{2\pi}/20 = 0.1253$. It follows that the distribution of $Z = (\tilde{\theta}_{.5,n} - \theta)/0.1253$ will be approximately standard normal. Thus,

$$\Pr(|\tilde{\theta}_{.5,n} - \theta| \leq 0.1) = \Pr\left(|Z| \leq \frac{0.1}{0.1253}\right) = \Pr(|Z| \leq 0.798) = 2\Phi(0.798) - 1 = 0.575.$$

4. Here,

$$f(x) = \frac{1}{\pi[1 + (x - \theta)^2]}$$

Therefore, $f(\theta) = 1/\pi$ and, since $n = 100$, it follows that the distribution of $\tilde{\theta}_{.5,n}$ will be approximately normal with mean θ and standard deviation $\pi/20 = 0.1571$. Thus, the distribution of $Z = (\tilde{\theta}_{.5,n} - \theta)/0.1571$ will be approximately standard normal. Hence,

$$\Pr(|\tilde{\theta}_{.5,n} - \theta| \leq 0.1) = \Pr\left(|Z| \leq \frac{0.1}{0.1571}\right) = \Pr(|Z| \leq 0.637) = 2\Phi(0.637) - 1 = 0.476.$$

5. Let the first density on the right side of Eq. (10.7.1) be called h. Since both h and g are symmetric with respect to μ, so also is $f(x)$. Therefore, both the sample mean $\overline{X}_n$ and the sample median $\tilde{X}_n$ are unbiased estimators of μ. It follows that the M.S.E. of $\overline{X}_n$ is equal to $\text{Var}(\overline{X}_n)$ and that the M.S.E. of $\tilde{X}_n$ is equal to $\text{Var}(\tilde{X}_n)$. The variance of a single observation X is

$$\begin{aligned}\text{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\ &= \frac{1}{2}\int_{-\infty}^{\infty} (x-\mu)^2 h(x)dx + \frac{1}{2}\int_{-\infty}^{\infty} (x-\mu)^2 g(x)dx \\ &= \frac{1}{2}\,(1) + \frac{1}{2}\,(4) = \frac{5}{2}.\end{aligned}$$

Since $n = 100$, $\text{Var}(\overline{X}_n) = (1/100)(5/2) = 0.025$.
The variance of $\tilde{X}_n$ will be approximately $1/[4nh^2(\mu)]$. Since

$$h(x) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x-\mu)^2\right] \quad \text{and } g(x) = \frac{1}{2\sqrt{2\pi}}\exp\left[-\frac{1}{2(4)}(x-\mu)^2\right],$$

it follows that

$$f(\mu) = \frac{1}{2}h(\mu) + \frac{1}{2}g(\mu) = \frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}} + \frac{1}{2}\cdot\frac{1}{2\sqrt{2\pi}} = \frac{3}{4\sqrt{2\pi}}.$$

Therefore, $\text{Var}(\tilde{X}_n)$ is approximately $2\pi/225 = 0.028$.

6. Let $g_n(\boldsymbol{x})$ be the joint p.d.f. of the data given that they came from the uniform distribution, and let $f_n(\boldsymbol{x})$ be the joint p.d.f. given that they come from the p.d.f. in Exercise 5. According to Bayes' theorem, the posterior probability that they came from the uniform distribution is

$$\frac{\frac{1}{2}g_n(\boldsymbol{x})}{\frac{1}{2}g_n(\boldsymbol{x}) + \frac{1}{2}f_n(\boldsymbol{x})}.$$

It is easy to see that $g_n(\boldsymbol{x}) = 1$ for these data, while $f_n(\boldsymbol{x}) = (3/2)^{16}(1/2)^9 = 1.283$. This makes the posterior probability of the uniform distribution $1/2.283 = 0.4380$.

7. (a) The mean $\overline{X}_n$ is the mean of each X_i. Since $f(x)$ is a weighted average of two other p.d.f.'s, the $\int xf(x)dx$ is the same mixture of the means of the other two distributions. Since each of the distributions in the mixture has mean μ, so does the distribution with p.d.f. f.

   (b) The variance $\overline{X}_n$ is $1/n$ times the variance of X_i. The variance of X_i is $E(X_i^2) - \mu^2$. Since the p.d.f. of X_i is a weighted average of two other p.d.f.'s, the mean of X_i^2 is the same weighted average of the two means of X_i^2 from the two p.d.f.'s. The mean of X_i^2 from the first p.d.f. (the normal distribution with mean μ and variance σ^2) is $\mu^2 + \sigma^2$. The mean of X_i^2 from the second p.d.f. (the normal distribution with mean μ and variance $100\sigma^2$) is $\mu^2 + 100\sigma^2$. The weighted average is

$$(1-\epsilon)(\mu^2+\sigma^2)+\epsilon(\mu^2+100\sigma^2)=\mu^2+\sigma^2(1+99\epsilon).$$

   The variance of X_i is then $(1+99\epsilon)\sigma^2$, and the variance $\overline{X}_n$ is $(1+99\epsilon)\sigma^2/n$.

8. When $\epsilon = 1$, the distribution whose p.d.f. is in Eq. (10.7.2) is the normal distribution with mean μ and variance $100\sigma^2$. When $\epsilon = 0$, the distribution is the normal distribution with mean μ and variance σ^2. The ratio of the variances of the sample mean and sample median from a normal distribution does not depend on the variance of the normal distribution, hence the ratio will be the same whether the variance is σ^2 or $100\sigma^2$. The reason that the ratio doesn't depend on the variance of the specific normal distribution is that both the sample mean and the sample median have variances that equal the variance of the original distribution times constants that depend only on the sample size.

9. The likelihood function is

$$\frac{1}{2^n\sigma^n}\exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n}|x_i-\theta|\right).$$

It is easy to see that, no matter what σ equals, the M.L.E. of θ is the number that minimizes $\sum_{i=1}|x_i-\theta|$. This is the same as the number that minimizes $\sum_{i=1}|x_i-\theta|/n$. The value $\sum_{i=1}^{n}|x_i-\theta|/n$ is the mean of $|X-\theta|$ when the c.d.f. of X is the sample c.d.f. of $X_1,\dots,X_n$. The mean of $|X-\theta|$ is minimized by θ equal to a median of the distribution of X according to Theorem 4.5.3. The median of the sample distribution is the sample median.

10. The likelihood was given in Exercise 9. The logarithm of the likelihood equals

$$-n\log(2\sigma)-\frac{1}{\sigma}\sum_{i=1}^{n}|x_i-\theta|.$$

For convenience, assume that $x_1 < x_2 < \ldots < x_n$. Let θ be a given number between two consecutive x_i values. In particular, let $x_k < \theta < x_{k+1}$. For known σ, the likelihood can be written as a constant plus a constant times

$$\sum_{i=k+1}^{n}x_i-(n-k)\theta-\sum_{i=1}^{k}x_i+k\theta.$$

For θ between x_k and x_{k+1}, the derivative of this is $k-(n-k)$, the difference between the number of observations below θ and the number above θ.

11. Let x_q be the q quantile of X. The result will follow if we can prove that the q quantile of $aX + b$ is $ax_q + b$. Since

$$\Pr(aX + b \le ax_q + b) = \Pr(X \le x_q),$$

for all $a > 0$ and b and q, it follows that $ax_q + b$ is the q quantile of $aX + b$.

12. According to the solution to Exercise 11, the median of $aX + b$ is $am + b$, where m is the median of X. The median absolute deviation of X is the median of $|X - m|$, which equals σ. The median absolute deviation of $aX + b$ is the median of $|aX + b - (am + b)| = a|X - m|$. According to the solution to Exercise 11, the median of $a|X - m|$ is a times the median of $|X - m|$, that is, $a\sigma$.

13. The Cauchy distribution is symmetric around 0, so the median is 0, and the median absolute deviation is the median of $Y = |X|$. If F is the c.d.f. of X, then the c.d.f. of Y is

$$G(y) = \Pr(Y \le y) = \Pr(|X| \le y) = \Pr(-y \le X \le y) = F(y) - F(-y),$$

because X has a continuous distribution. Because X has a symmetric distribution around 0, $F(-y) = 1 - F(y)$, and $G(y) = 2F(y) - 1$. The median of Y is where $G(y) = 0.5$, that is $2F(y) - 1 = 0.5$ or $F(y) = 0.75$. So, the median of Y is the 0.75 quantile of X, namely $y = 1$.

14. (a) The c.d.f. of X is $F(x) = 1 - \exp(-x\lambda)$, so the quantile function is $F^{-1}(p) = -\log(1-p)/\lambda$. The IQR is

$$F^{-1}(0.75) - F^{-1}(0.25) = -\frac{\log(0.25)}{\lambda} + \frac{\log(0.75)}{\lambda} = \frac{\log(3)}{\lambda}.$$

(b) The median of X is $\log(2)/\lambda$, and the median absolute deviation is the median of $|X - \log(2)/\lambda|$. It is the value x such that $\Pr(\log(2)/\lambda - x \le X \le \log(2)/\lambda + x) = 0.5$. If we try letting $x = \log(3)/[2\lambda]$ (half of the IQR), then

$$\begin{aligned} Pr(\log(2)/\lambda - x \le X \le \log(2)/\lambda + x) &= [1 - \exp(-\log(2\sqrt{3}))] - [1 - \exp(-\log(2/\sqrt{3}))] \\ &= \frac{1}{2}[\sqrt{3} - 1/\sqrt{3}] = 0.5773. \end{aligned}$$

This is greater than 0.5, so the median absolute deviation is smaller than 1/2 of the IQR.

15. (a) The quantile function of the normal distribution with mean μ and variance σ^2 is the inverse of the c.d.f., $F(x) = \Phi([x - \mu]/\sigma)$. So,

$$F^{-1}(p) = \mu + \sigma\Phi^{-1}(p). \qquad \text{(S.10.1)}$$

The IQR is

$$F^{-1}(0.75) - F^{-1}(0.25) = \sigma[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)].$$

Since the standard normal distribution is symmetric around 0, $\Phi^{-1}(0.25) = -\Phi^{-1}(0.75)$, so the IQR is $2\sigma\Phi^{-1}(0.75)$.

(b) Let F be the c.d.f. of a distribution that is symmetric around its median μ. The median absolute deviation is then the value x such that $F(\mu + x) - F(\mu - x) = 0.5$. By symmetry around the median, we know that $F(\mu - x) = 1 - F(\mu + x)$, so x solves $2F(\mu + x) - 1 = 0.5$ or $F(\mu + x) = 0.75$. That is, $x = F^{-1}(0.75) - \mu$. For the case of normal random variables, use Eq. (S.10.1) to conclude that $x = \sigma\Phi^{-1}(0.75)$.

16. Here are the sorted values from smallest to largest:

$$-67,\ -48,\ 6,\ 8,\ 14,\ 16,\ 23,\ 24,\ 28,\ 29,\ 41,\ 49,\ 56,\ 60,\ 75.$$

(a) The average is 20.93.

(b) The trimmed means are

k	1	2	3	4
Trimmed mean	25.54	26.73	25.78	25

(c) The sample median is 24.

(d) The median absolute deviation divided by 0.6745 is 25.20385. Starting at $\theta = 0$ and iterating the procedure described on page 673 of the text, we get the following sequence of values for θ:

$$20.805,\ 24.017,\ 26.278,\ 24.342,\ 24.373,\ 24.376,\ 24.377,\ 24.377, \ldots$$

After 9 iterations, the value stays the same to 7 significant digits.

17. Let μ stand for the median of the distribution, and let $\mu + c$ be the 0.75 quantile. By symmetry, the 0.25 quantile is $\mu - c$. Also, $f(\mu + c) = f(\mu - c)$. The large sample joint distribution of the 0.25 and 0.75 sample quantiles is a bivariate normal distribution with means $\mu - c$ and $\mu + c$, variances both equal to $3/[16nf(\mu + c)^2]$, and covariance $1/[16nf(\mu + c)^2]$. The IQR is the difference between these two sample quantiles, so its large sample distribution is normal with mean $2c$ and variance

$$\frac{3}{16nf(\mu+c)^2} + \frac{3}{16nf(\mu+c)^2} - 2\frac{1}{16nf(\mu+c)^2} = \frac{1}{4nf(\mu+c)^2}.$$

10.8 Sign and Rank Tests

Commentary

This section ends with a derivation of the power function of the Wilcoxon-Mann-Whitney ranks test. This derivation is a bit more technical than the rest of the section and is perhaps suitable only for the more mathematically inclined reader.

If one is using the software *R*, the function `wilcox.test` performs the Wilcoxon-Mann-Whitney ranks test. The two arguments are the two samples whose distributions are being compared.

Solutions to Exercises.

1. Let W be the number of (X_i, Y_i) pairs with $X_i \le Y_i$. Then W has a binomial distribution with parameters n and p. To test H_0, we reject H_0 if W is too large. In particular, if c is chosen so that

$$\sum_{w=c+1}^{n} \binom{n}{w} \left(\frac{1}{2}\right)^n < \alpha_0 \le \sum_{w=c}^{n} \binom{n}{w} \left(\frac{1}{2}\right)^n,$$

then we can reject H_0 if $W \ge c$ for a level α_0 test.

2. The largest difference between the two sample c.d.f.'s occurs between 2.336 and 2.431 and equals $|0.8 - 0.125| = 0.675$. The test statistic is then

$$\left(\frac{8 \times 10}{8 + 10}\right)^{1/2} 0.675 = 1.423.$$

The tail area is between 0.0397 and 0.0298.

3. This test was performed in Example 9.6.5, and the tail area is 0.003.

4. By ordering all the observations, we obtain Table S.10.12. The sum of the ranks of $x_1, \ldots, x_{25}$ is

Table S.10.12: Table for Exercise 4 of Sec. 10.8.

Rank	Observed value	Sample	Rank	Observed value	Sample
1	0.04	x	21	1.01	y
2	0.13	x	22	1.07	y
3	0.16	x	23	1.12	x
4	0.28	x	24	1.15	x
5	0.35	x	25	1.20	x
6	0.39	x	26	1.25	y
7	0.40	x	27	1.26	y
8	0.44	x	28	1.31	y
9	0.49	x	29	1.38	x
10	0.58	x	30	1.48	y
11	0.68	y	31	1.50	x
12	0.71	x	32	1.54	x
13	0.72	x	33	1.59	y
14	0.75	x	34	1.63	y
15	0.77	x	35	1.64	x
16	0.83	x	36	1.73	x
17	0.86	y	37	1.78	y
18	0.89	y	38	1.81	y
19	0.90	x	39	1.82	y
20	0.91	x	40	1.95	y

$S = 399$. Since $m = 25$ and $n = 15$, it follows from Eqs. (10.8.3) and (10.8.4) that $E(S) = 512.5$, $\text{Var}(S) = 1281.25$, and $\sigma = (1281.25)^{1/2} = 35.7946$. Hence, $Z = (399 - 512.5)/35.7946 = -3.17$. It can be found from a table of the standard normal distribution that the corresponding two-sided tail area is 0.0015.

5. Since there are 25 observations in the first sample, $F_m(x)$ will jump by the amount 0.04 at each observed value. Since there are 15 observations in the second sample, $G_n(x)$ will jump by the amount 0.0667 at each observed value. From the table given in the solution to Exercise 4, we obtain Table S.10.13. It can be seen from this table that the maximum value of $|F_m(x) - G_n(x)|$ occurs when x is equal to the observed value of rank 16, and its value at this point is $.60 - .0667 = .5333$. Hence, $D_{mn} = 0.5333$ and $\left(\frac{mn}{m+n}\right)^{1/2} D_{mn} = \left(\frac{375}{40}\right)^{1/2} (0.5333) = 1.633$. It is found from Table 10.32 that the corresponding tail area is almost exactly 0.01.

6. It is found from the values given in Tables 10.44 and 10.45 that $\overline{x} = \sum_{i=1}^{25} x_i/25 = 0.8044$, $\overline{y} = \sum_{i=1}^{15} y_i/15 = 1.3593$, $S_x^2 = \sum_{i=1}^{25}(x_i - \overline{x})^2 = 5.8810$, and $S_y^2 = \sum_{i=1}^{15}(y_i - \overline{y})^2 = 2.2447$. Since $m = 25$ and $n = 15$, it follows from Eq. (9.6.3) that $U = -3.674$. It can be found from a table of the t distribution with $m + n - 2 = 38$ degrees of freedom that the corresponding two-sided tail area is less than 0.01.

7. We need to show that $F(\theta + G^{-1}(p)) = p$. Compute

$$F(\theta + G^{-1}(p)) = \int_{-\infty}^{\theta+G^{-1}(p)} f(x)dx = \int_{-\infty}^{\theta+G^{-1}(p)} g(x-\theta)dx = \int_{-\infty}^{G^{-1}(p)} g(y)dy = G(G^{-1}(p)) = p,$$

where the third equality follows by making the change of variables $y = x - \theta$.

Table S.10.13: Table for Exercise 5 of Sec. 10.8.

Rank of observations	$F_m(x)$	$G_n(x)$	Rank of observations	$F_m(x)$	$G_n(x)$
1	.04	0	21	.68	.2667
2	.08	0	22	.68	.3333
3	.12	0	23	.72	.3333
4	.16	0	24	.76	.3333
5	.20	0	25	.80	.3333
6	.24	0	26	.80	.4000
7	.28	0	27	.80	.4667
8	.32	0	28	.80	.5333
9	.36	0	29	.84	.5333
10	.40	0	30	.84	.6000
11	.40	.0667	31	.88	.6000
12	.44	.0667	32	.92	.6000
13	.48	.0667	33	.92	.6667
14	.52	.0667	34	.92	.7333
15	.56	.0667	35	.96	.7333
16	.60	.0667	36	1.00	.7333
17	.60	.1333	37	1.00	.8000
18	.60	.2000	38	1.00	.8667
19	.64	.2000	39	1.00	.9333
20	.68	.2000	40	1.00	1.0000

8. Since $Y+\theta$ and X have the same distribution, it follows that if $\theta > 0$ then the values in the first sample will tend to be larger than the values in the second sample. In other words, when $\theta > 0$, the sum S of the ranks in the first sample will tend to be larger than it would be if $\theta = 0$ or $\theta < 0$. Therefore, we will reject H_0 if $Z > c$, where Z is as defined in this section and c is an appropriate constant. If we want the test to have a specified level of significance α_0 $(0 < \alpha_0 < 1)$, then c should be chosen so that when Z has a standard normal distribution, $\Pr(Z > c) = \alpha_0$. It should be kept in mind that the level of significance of this test will only be approximately α_0 because for finite sample sizes, the distribution of Z will only be approximately a standard normal distribution when $\theta = 0$.

9. To test these hypotheses, add θ_0 to each observed value y_i in the second sample and then carry out the Wilcoxon-Mann-Whitney procedure on the original values in the first sample and the new values in the second sample.

10. For each value of θ_0, carry out a test of the hypotheses given in Exercise 7 at the level of significance $1-\alpha$. The confidence interval for θ will contain all values of θ_0 for which the null hypothesis H_0 would be accepted.

11. Let $r_1 < r_2 < \cdots < r_m$ denote the ranks of the observed values in the first sample, and let $X_{i_1} < X_{i_2} < \cdots < X_{i_m}$ denote the corresponding observed values. Then there are $r_1 - 1$ values of Y in the second sample that are smaller than X_{i_1}. Hence, there are $r_1 - 1$ pairs (X_{i_1}, Y_j) with $X_{i_1} > Y_j$. Similarly, there are $r_2 - 2$ values of Y in the second sample that are smaller than X_{i_2}. Hence, there are $r_2 - 2$ pairs (X_{i_2}, Y_j) with $X_{i_2} > Y_j$. By continuing in this way, we see that the number U is equal to

$$(r_1 - 1) + (r_2 - 2) + \cdots + (r_m - m) = \sum_{i=1}^{m} r_i - \sum_{i=1}^{m} i = S - \frac{1}{2}m(m+1).$$

12. Using the result in Exercise 11, we find that $E(S) = E(U) + m(m+1)/2$, where U is defined in Exercise 11 to be the number of (X_i, Y_j) pairs for which $X_i \geq Y_j$. So, we need to show that $E(U) = nm \Pr(X_1 \geq Y_1)$.

We can let $Z_{i,j} = 1$ if $X_i \geq Y_j$ and $Z_{i,j} = 0$ otherwise. Then

$$U = \sum_i \sum_j Z_{i,j}, \tag{S.10.2}$$

and

$$E(U) = \sum_{i=1}^{m} \sum_{j=1}^{n} E(Z_{i,j}).$$

Since all of the X_i are i.i.d. and all of the Y_j are i.i.d., it follows that $E(Z_{i,j}) = E(Z_{1,1})$ for all i and j. Of course $E(Z_{1,1}) = \Pr(X_1 \geq Y_1)$, so $E(U) = mn \Pr(X_1 \geq Y_1)$.

13. Since S and U differ by a constant, we need to show that $\text{Var}(U)$ is given by Eq. (10.8.6). Once again, write

$$U = \sum_i \sum_j Z_{i,j},$$

where $Z_{i,j} = 1$ if $X_i \geq Y_j$ and $Z_{i,j} = 0$ otherwise. Hence,

$$\text{Var}(U) = \sum_i \sum_j \text{Var}(Z_{i,j}) + \sum_{(i',j') \neq (i,j)} \text{Cov}(Z_{i,j}, Z_{i',j'}).$$

The first sum is $mn[\Pr(X_1 \geq Y_1) - \Pr(X_1 \geq Y_1)^2]$. The second sum can be broken into three parts:

- The terms with $i' = i$ but $j' \neq j$.
- The terms with $j' = j$ but $i' \neq i$.
- The terms with both $i' \neq i$ and $j' \neq j$.

For the last set of terms $\text{Cov}(Z_{i,j}, Z_{i',j'}) = 0$ since (X_i, Y_j) is independent of $(X_{i'}, Y_{j'})$. For each term in the first set

$$E(Z_{i,j} Z_{i,j'}) = \Pr(X_1 \geq Y_1, X_1 \geq Y_2),$$

so the covariances are

$$\text{Cov}(Z_{i,j}, Z_{i,j'}) = \Pr(X_1 \geq Y_1, X_1 \geq Y_2) - \Pr(X_1 \geq Y_1)^2.$$

There are $mn(n-1)$ terms of this sort. Similarly, for the second set of terms

$$\text{Cov}(Z_{i',j}, Z_{i,j}) = \Pr(X_1 \geq Y_1, X_2 \geq Y_1) - \Pr(X_1 \geq Y_1)^2.$$

There are $nm(m-1)$ of these terms. The variance is then

$$nm \left[\Pr(X_1 \geq Y_1) + (n-1) \Pr(X_1 \geq Y_1, X_1 \geq Y_2) \right.$$
$$\left. +(m-1) \Pr(X_1 \geq Y_1, X_2 \geq Y_1) - (m+n-1) \Pr(X_1 \geq Y_1)^2 \right].$$

14. When $F = G$, $\Pr(X_1 \geq Y_1) = 1/2$, so Eq. (10.8.5) yields

$$E(S) = \frac{mn}{2} + \frac{m(m+1)}{2} = \frac{m(m+n+1)}{2},$$

which is the same as (10.8.3). When $F = G$,

$$\Pr(X_1 \geq Y_1, X_1 \geq Y_2) = 1/3 = \Pr(X_1 \geq Y_1, X_2 \geq Y_1),$$

so the corrected version of (10.8.6) yields

$$\begin{aligned} nm\left[\frac{1}{2} - (m+n-1)\frac{1}{4} + (m+n-2)\frac{1}{3}\right] &= \frac{mn}{12}[6 - 3m - 3n + 3 + 4m + 4n - 8] \\ &= \frac{mn(m+n+1)}{12}, \end{aligned}$$

which is the same as (10.8.4).

15. (a) Arrange the observations so that $|D_1| \leq \cdots \leq |D_n|$. Then $D_i > 0$ if and only if $X_i > Y_i$ if and only if $W_i = 1$. Since rank i gets added into S_W if and only if $D_i > 0$, we see that $\sum_{i=1}^{n} iW_i$ adds just those ranks that correspond to positive D_i.

(b) Since the distribution of each D_i is symmetric around 0, the magnitude $|D_1|, \ldots, |D_n|$ are independent of the sign indicators $W_1, \ldots, W_n$. Using the result of part (a), if we assume that the $|D_i|$ are ordered from smallest to largest, $E(S_W) = \sum_{i=1}^{n} iE(W_i)$. Since the $|D_i|$ are independent of the W_i, we have $E(W_i) = 1/2$ even after we condition on the $|D_i|$ being arranged from smallest to largest. Since $\sum_{i=1}^{n} i = n(n+1)/2$, we have $E(S_W) = n(n+1)/4$.

(c) Since the W_i are independent before we condition on the $|D_i|$ and they are independent of the $|D_i|$, then the W_i are independent conditional on the $|D_i|$. Hence, $\text{Var}(S_W) = \sum_{i=1}^{n} i^2 \text{Var}(W_i)$. Since $\text{Var}(W_i) = 1/4$ for all i and $\sum_{i=1}^{n} i^2 = n(n+1)(2n+1)/6$, we have $\text{Var}(S_W) = n(n+1)(2n+1)/24$.

16. For $i = 1, \ldots, 15$, let

$$D_i = \text{(thickness for material A in pair i)} - \text{(thickness for material B in pair i)}.$$

(a) Of the 15 values of D_i, 10 are positive, 3 are negative, and 2 are zero. If we first regard the zeroes as positive, then there are 12 positive differences with $n = 15$, and it is found from the binomial tables that the corresponding tail area is 0.0176. If we next regard the zeroes as negative, then there are only 10 positive differences with $n = 15$, and it is found from the tables that the corresponding tail area is 0.1509. The results are not conclusive because of the zeroes present in the sample.

(b) For the Wilcoxon signed-ranks test, use Table S.10.14. Two different methods have been used. In Method (I), the differences that are equal to 0 are regarded as positive, and whenever two or more values of $| D_i |$ are tied, the positive differences D_i are assigned the largest possible ranks and the negative differences D_i are assigned the smallest ranks. In Method (II), the differences that are 0

Table S.10.14: Computation of Wilcoxon signed-ranks test statistic for Exercise 16b in Sec. 10.8.

Pair	D_i	Method (I) Rank of $\|D_i\|$	Method (I) Signed rank	Method (II) Rank of $\|D_i\|$	Method (II) Signed rank
1	−0.8	6	−6	7	−7
2	1.6	12	12	11	11
3	−0.5	5	−5	5	−5
4	0.2	3	3	3	3
5	−1.6	11	−11	13	−13
6	0.2	4	4	4	4
7	1.6	13	13	12	12
8	1.0	9	9	9	9
9	0.8	7	7	6	6
10	0.9	8	8	8	8
11	1.7	14	14	14	14
12	1.2	10	10	10	10
13	1.9	15	15	15	15
14	0	2	2	2	−2
15	0	1	1	1	−1

are regarded as negative, and among tied values of $| D_i |$, the negative differences are assigned the largest ranks and the positive differences are assigned the smallest ranks. Let S_n denote the sum of the positive ranks. Since $n = 15$, $E(S_n) = 60$ and Var $(S_n) = 310$. Hence, $\sigma_n = \sqrt{310} = 17.607$. For Method (I), $S_n = 98$. Therefore, $Z_n = (98 - 60)/17.607 = 2.158$ and it is found from a table of the standard normal distribution that the corresponding tail area is 0.0155. For Method (II), $S_n = 92$. Therefore, $Z_n = 1.817$ and it is found that the corresponding tail area is 0.0346. By either method of analysis, the null hypothesis would be rejected at the 0.05 level of significance, but not at the 0.01 level.

(c) The average of the pairwise differences (material A minus material B) is 0.5467. The value of σ' computed from the differences is 1.0197, so the t statistic is 2.076, and the p-value is 0.0284.

10.9 Supplementary Exercises

Solutions to Exercises

1. Here, $\alpha_0/2 = 0.025$. From a table of binomial probabilities we find that

$$\sum_{x=0}^{5} \binom{20}{x} 0.5^{20} = 0.021 \leq 0.025 \leq \sum_{x=0}^{6} \binom{20}{x} 0.5^{20} = 0.058.$$

So, the sign test would reject the null hypothesis that $\theta = \theta_0$ if the number W of observations with values at most θ_0 satisfies either $W \leq 5$ or $W \geq 20 - 5$. Equivalently, we would accept the null hypothesis if $6 \leq W \leq 14$. This, in turn, is true if and only if θ_0 is strictly between the sixth and fourteenth ordered values of the original data. These values are 141 and 175, so our 95 percent confidence interval is $(141, 175)$.

2. It follows from Eq. (10.1.2) that

$$Q = \sum_{i=1}^{5} \frac{(N_i - 80)^2}{80} = \frac{1}{80}\left[\sum_{i=1}^{5} N_i^2 - 2(80)(400) + 5(80)^2\right] = \frac{1}{80}\sum_{i=1}^{5} N_i^2 - 400.$$

It is found from the χ^2 distribution with 4 degrees of freedom that H_0 should be rejected for $Q > 13.28$ or, equivalently, for $\sum_{i=1}^{5} N_i^2 > 80(413.28) = 33,062.4$.

3. Under H_0, the proportion p_i^0 of families with i boys is as follows:

i	p_i^0	np_i^0
0	1/8	16
1	3/8	48
2	3/8	48
3	1/8	16

Hence, it follows from Eq. (10.1.2) that

$$Q = \frac{(26-16)^2}{16} + \frac{(32-48)^2}{48} + \frac{(40-48)^2}{48} + \frac{(30-16)^2}{16} = 25.1667.$$

Under H_0, Q has the χ^2 distribution with 3 degrees of freedom. Hence, the tail area corresponding to $Q = 25.1667$ is less than 0.005, the smallest probability in the table in the back of the book. It follows that H_0 should be rejected for any level of significance greater than this tail area.

4. The likelihood function of p based on the observed data is

$$(q^3)^{26}(3pq^2)^{32}(3p^2q)^{40}(p^3)^{30} = (const.)\ p^{202}q^{182},$$

where $q = 1 - p$. Hence, the M.L.E. of $\hat{p}$ based on these data is $\hat{p} = 202/384 = .526$. Under H_0, the estimated expected proportion $\hat{p}_i^0$ of families with i boys is as follows:

i	$\hat{p}_i^0$	$n\hat{p}_i^0$
0	$\hat{q}^3 = .1065$	13.632
1	$3\hat{p}\hat{q}^2 = .3545$	45.376
2	$3\hat{p}^2\hat{q} = .3935$	50.368
3	$\hat{p}^3 = .1455$	18.624

It follows from Eq. (10.2.4) that

$$Q = \frac{(26-13.632)^2}{13.632} + \frac{(32-45.376)^2}{45.376} + \frac{(40-50.368)^2}{50.368} + \frac{(30-18.624)^2}{18.624} = 24.247.$$

Under H_0, Q has the χ^2 distribution with 2 degrees of freedom. The tail area corresponding to $Q = 24.247$ is again less than 0.005. H_0 should be rejected for any level of significance greater than this tail area.

5. The expected numbers of observations in each cell, as specified by Eq. (10.4.4), are presented in the following table:

	A	B	AB	O
Group 1	19.4286	11	4.8571	14.7143
Group 2	38.8571	22	9.7143	29.4286
Group 3	77.7143	44	19.4286	58.8571

It is now found from Eq. (10.4.3) that $Q = 6.9526$. Under the hypothesis H_0 that the distribution is the same in all three groups, Q will have approximately the χ^2 distribution with $3 \times 2 = 6$ degrees of freedom. It is found from the tables that the 0.9 quantile of that distribution is 10.64, so H_0 should not be rejected.

6. If Table 10.47 is changed in such a way that the row and column totals remain unchanged, then the expected numbers given in the solution of Exercise 5 will remain unchanged. If we switch one person in group 1 from B to AB and one person in group 2 from AB to B, then all row and column totals will be unchanged and the new observed numbers in each of the four affected cells will be further from their expected values than before. Hence, the value of the χ^2 statistic Q as given by Eq. (10.4.3) is increased. Continuing to switch persons in this way will continue to increase Q. There are other similar switches that will also increase Q, such as switching one person in group 2 from O to A and one person in group 3 from A to O.

7.

$$
\begin{aligned}
(N_{11} - \hat{E}_{11})^2 &= \left(N_{11} - \frac{N_{1+}N_{+1}}{n}\right)^2 \\
&= \left[N_{11} - \frac{(N_{11} + N_{12})(N_{11} + N_{21})}{n}\right]^2 \\
&= \frac{1}{n^2}\left[nN_{11} - (N_{11} + N_{12})(N_{11} + N_{21})\right]^2 \\
&= \frac{1}{n^2}\left(N_{11}N_{22} - N_{12}N_{21}\right)^2,
\end{aligned}
$$

since $n = N_{11} + N_{21} + N_{21} + N_{22}$. Exactly the same value is obtained for $(N_{12} - \hat{E}_{12})^2, (N_{21} - \hat{E}_{21})^2$, and $(N_{22} - \hat{E}_{22})^2$.

8. It follows from Eq. (10.3.4) and Exercise 7 that

$$
Q = \frac{1}{n^2}(N_{11}N_{22} - N_{12}N_{21})^2 \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{1}{\hat{E}_{ij}}.
$$

But

$$
\begin{aligned}
\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{1}{\hat{E}_{ij}} &= \frac{n}{N_{1+}N_{+1}} + \frac{n}{N_{1+}N_{+2}} + \frac{n}{N_{2+}N_{+1}} + \frac{n}{N_{2+}N_{+2}} \\
&= \frac{n(N_{2+}N_{+2} + N_{2+}N_{+1} + N_{1+}N_{+2} + N_{1+}N_{+1})}{N_{1+}N_{2+}N_{+1}N_{+2}} \\
&= \frac{n^3}{N_{1+}N_{2+}N_{+1}N_{+2}},
\end{aligned}
$$

since $N_{1+} + N_{2+} = N_{+1} + N_{+2} = n$. Hence, Q has the specified form.

9. In this exercise, $N_{1+} = N_{2+} = N_{+1} = N_{+2} = 2n$ and $N_{11}N_{22} - N_{12}N_{21} = (n+a)^2 - (n-a)^2 = 4na$. It now follows from Exercise 8 (after we replace n by $4n$ in the expression for Q) that $Q = 4a^2/n$. Since H_0 should be rejected if $Q > 6.635$, it follows that H_0 should be rejected if $a > (6.635n)^{1/2}/2$ or $a < -(6.635n)^{1/2}/2$.

10. In this exercise $N_{1+} = N_{2+} = N_{+1} = N_{+2} = n$ and $N_{11}N_{22} - N_{12}N_{21} = (2\alpha - 1)n^2$. It now follows from Exercise 8 (after we replace n by 2n in the expression for Q) that $Q = 2n(2\alpha - 1)^2$. Since H_0 should be rejected if $Q > 3.841$, it follows that H_0 should be rejected if either

$$\alpha > \frac{1}{2}\left[1 + \left(\frac{3.841}{2n}\right)^{1/2}\right]$$

or

$$\alpha < \frac{1}{2}\left[1 - \left(\frac{3.841}{2n}\right)^{1/2}\right].$$

11. Results of this type are an example of Simpson's paradox. If there is a higher rate of respiratory diseases among older people than among younger people, and if city A has a higher proportion of older people than city B, then results of this type can very well occur.

12. Results of this type are another example of Simpson's paradox. If scores on the test tend to be higher for certain classes, such as seniors and juniors, and lower for the other classes, such as freshmen and sophomores, and if school B has a higher proportion of seniors and juniors than school A, then results of this type can very well occur.

13. The fundamental aspect of this exercise is that it is not possible to assess the effectiveness of the treatment without having any information about how the levels of depression of the patients would have changed over the three-month period if they had not received the treatment. In other words, without the presence of a control group of similar patients who received some other standard treatment or no treatment at all, there is little meaningful statistical analysis that can be carried out. We can compare the proportion of patients at various levels who showed improvement after the treatment with the proportion who remained the same or worsened, but without a control group we have no way of deciding whether these proportions are unusually large or small.

14. If $Y_1 < Y_2 < Y_3$ are the order statistics of the sample, then Y_2 is the sample median. For $0 < y < 1$,

$$\begin{aligned} G(y) \quad &= \Pr\,(Y_2 < y) \\ &= \Pr(\text{At least two obs.} < y) \\ &= \Pr(\text{Exactly two obs.} < y) + \Pr(\text{All three obs.} < y) \\ &= 3(y^\theta)^2(1 - y^\theta) + (y^\theta)^3 \\ &= 3y^{2\theta} - 2y^{3\theta}. \end{aligned}$$

Hence, for $0 < y < 1$. the p.d.f. of Y_2 is $g(y) = G'(y) = 6\theta(y^{2\theta-1} - y^{3\theta-1})$.

15. The c.d.f. of this distribution is $F(x) = x^\theta$, so the median of the distribution is the point m such that $m^\theta = 1/2$. Thus, $m = (1/2)^{1/\theta}$ and $f(m) = \theta 2^{1/\theta}/2$. It follows from Theorem 10.7.1 that the asymptotic distribution of the sample median will be normal with mean m and variance

$$\frac{1}{4nf^2(m)} = \frac{1}{n\theta^2 2^{2/\theta}}.$$

16. We know from Exercise 1 of Sec. 8.4 that the variance of the t distribution is finite only for $\alpha > 2$ and its value is $\alpha/(\alpha-2)$. Hence, it follows from the central limit theorem that for $\alpha > 2$, the asymptotic distribution of $\overline{X}_n$ will be normal with mean 0 and variance

$$\sigma_1^2 = \frac{\alpha}{n(\alpha-2)}.$$

Since the median of the t distribution is 0, it follows from Theorem 10.7.1 (with n replaced by α) that the asymptotic distribution of $\tilde{X}_n$ will be normal with mean 0 and variance

$$\sigma_2^2 = \frac{\alpha\pi\Gamma^2\left(\frac{\alpha}{2}\right)}{4n\Gamma^2\left(\frac{\alpha+1}{2}\right)}.$$

Thus, $\sigma_1^2 < \sigma_2^2$ if and only if

$$\frac{\pi(\alpha-2)\Gamma^2\left(\frac{\alpha}{2}\right)}{4\Gamma^2\left(\frac{\alpha+1}{2}\right)} > 1.$$

If we denote the left side of this inequality by L, then we obtain the following values:

α	$\Gamma\left(\frac{\alpha}{2}\right)$	$\Gamma\left(\frac{\alpha+1}{2}\right)$	L
3	$\frac{1}{2}\sqrt{\pi}$	1	$\pi^2/16$
4	1	$\frac{3}{2}\cdot\frac{1}{2}\cdot\sqrt{\pi}$	$8/9$
5	$\frac{3}{2}\cdot\frac{1}{2}\cdot\sqrt{\pi}$	2	$27\pi^2/(16)^2 = 1.04.$

Thus, $\sigma_1^2 < \sigma_2^2$ for $\alpha = 5, 6, 7, \ldots$.

17. As shown in Exercise 5 of Sec. 10.7, $E(\overline{X}_n) = E(\tilde{X}_n) = \theta$, so the M.S.E. of each of these estimators is equal to its variance. Furthermore, $\text{Var}(\overline{X}_n) = \frac{1}{n}[\alpha\cdot 1 + (1-\alpha)\sigma^2]$ and

$$\text{Var}(\tilde{X}_n) = \frac{1}{4n[h(\theta|\theta)]^2},$$

where

$$[h(\theta|\theta)]^2 = \frac{1}{2\pi}\left(\alpha + \frac{1-\alpha}{\sigma}\right)^2.$$

(a) For $\sigma^2 = 100$, $\text{Var}(\tilde{X}_n) < \text{Var}(\overline{X}_n)$ if and only if

$$\frac{50\pi}{[10\alpha + (1-\alpha)]^2} < \alpha + 100(1-\alpha).$$

Some numerical calculations show that this inequality is satisfied for $.031 < \alpha < .994$.

(b) For $\alpha = \frac{1}{2}$, $\text{Var}(\tilde{X}_n) < \text{Var}(\overline{X}_n)$ if and only if $\sigma < .447$ or $\sigma > 1/.447 = 2.237$.

18. The simplest and most intuitive way to establish this result is to note that for any fixed values $y_1 < y_2 < \cdots < y_n$,

$$g(y_1, \ldots, y_n)\triangle y_1 \cdots \triangle y_n \approx$$
$$\Pr(y_1 < Y_1 < y_1 + \triangle y_1, \ldots, y_n < Y_n < y_n + \triangle y_n) =$$
$$\Pr(\text{Exactly one observation in the interval } (y_j, y_j + \triangle y_j) \text{ for } j = 1, \ldots, n) =$$
$$n! \prod_{j=1}^{n} [F(y_j + \triangle y_j) - F(y_j)] \approx$$
$$n! \prod_{j=1}^{n} [f(y_j)\triangle y_j] = n!\, f(y_1) \cdots f(y_n)\triangle y_1 \cdots \triangle y_n,$$

where the factor $n!$ appears because there are $n!$ different arrangements of $X_1, \ldots, X_n$ such that exactly one of them is in each of the intervals $(y_j, y_j + \triangle y_j)$, $j = 1, \ldots, n$. Another, and somewhat more complicated, way to establish this result is to determine the general form of the joint c.d.f. $G(y_1, \ldots, y_n)$ of $Y_1, \ldots, Y_n$ for $y_1 < y_2 < \cdots < y_n$, and then to note that

$$g(y_1, \ldots, y_n) = \frac{\partial^n G(y_1, \ldots, y_n)}{\partial y_1 \cdots \partial y_n} = n!\, f(y_1) \cdots f(y_n).$$

19. It follows from Exercise 18 that the joint p.d.f. $g(y_1, y_2, y_3) = 3!$, a constant, for $0 < y_1 < y_2 < y_3 < 1$. Since the required conditional p.d.f. of Y_2 is proportional to $g(y_1, y_2, y_3)$, as a function of y_2 for fixed y_1 and y_3, it follows that this conditional p.d.f. is also constant. In other words, the required conditional distribution is uniform on the interval (y_1, y_3).

20. We have $Y_r < \theta < Y_{r+3}$ if and only if at least r observations and at most $r + 2$ observation are below θ Let X stand for the number of observations out of the sample of size 20 that are below θ. Then X has a binomial distribution with parameters 20 and 0.3. It follows that

$$\Pr(Y_r < \theta < Y_{r+3}) = \Pr(r \le X \le r + 2).$$

For each value of r, we can find this probability using a binomial distribution table or a computer. By searching through all values of r, we find that $r = 5$ yields a probability of 0.5348, which is the highest.

21. As shown in Exercise 10 of Sec. 10.8, we add θ_0 to each observation Y_j and then carry out the Wilcoxon-Mann-Whitney test on the sum S_{θ_0} of the ranks of the X_i's among these new values $Y_1 + \theta_0, \ldots, Y_n + \theta_0$. We accept H_0 if and only if

$$\frac{|\ S_{\theta_0} - E(S)\ |}{[\mathrm{Var}(S)]^{1/2}} < c\left(1 - \frac{\alpha}{2}\right),$$

where $E(S)$ and $\mathrm{Var}(S)$ are given by (10.8.3) and (10.8.4). However, by Exercise 11 of Sec. 10.8,

$$S_{\theta_0} = U_{\theta_0} + \frac{1}{2}m(m+1).$$

When we make this substitution for S_{θ_0} in the above inequality, we obtain the desired result.

22. We know from general principles that the set of all values θ_0 for which H_0 would be accepted in Exercise 21 will from a confidence interval with the required confidence coefficient $1-\alpha$. But if U_{θ_0}, the number of differences $X_i - Y_j$ that are greater than θ_0, is greater than the lower limit given in Exercise 21 then θ_0 must be less than B. Similarly, if U_{θ_0} is less than the upper limit given in Exercise 22, then θ_0 must be greater than A. Hence, $A < \theta < B$ is a confidence interval.

23. (a) We know that $\theta_p = b$ if and only if $\Pr(X \le b) = p$. So, let $Y_i = 1$ if $X_i \le b$ and $Y_i = 0$ if not. Then $Y_1, \ldots, Y_n$ are i.i.d. with a Bernoulli distribution with parameter p if and only if H_0 is true. Define $W = \sum_{i=1}^n Y_i$ To test H_0, reject H_0 if W is too big or too small. For an equal tailed level α_0 test, choose two numbers $c_1 < c_2$ such that

$$\sum_{w=0}^{c_1} \binom{n}{w} p^w (1-p)^{n-w} \le \frac{\alpha_0}{2} < \sum_{w=0}^{c_1+1} \binom{n}{w} p^w (1-p)^{n-w},$$
$$\sum_{w=c_2}^{n} \binom{n}{w} p^w (1-p)^{n-w} \le \frac{\alpha_0}{2} < \sum_{w=c_2-1}^{n} \binom{n}{w} p^w (1-p)^{n-w}.$$

Then a level α_0 test rejects H_0 if $W \le c_1$ or $W \ge c_2$.

(b) For each b, we have shown how to construct a test of $H_{0,b} : \theta_p = b$. For given observed data $X_1, \ldots, X_n$ find all values of b such that the test constructed in part (a) accepts $H_{0,b}$. The set of all such b forms our coefficient $1 - \alpha_0$ confidence interval. It is clear from the form of the test that, once we find three values $b_1 < b_2 < b_3$ such that H_{0,b_2} is accepted and H_{0,b_1} and H_{0,b_3} are rejected, we don't have to check any more values of $b < b_1$ or $b > b_3$ since all of those would be rejected also. Similarly, if we find $b_4 < b_5$ such that both H_{0,b_4} and H_{0,b_5} are accepted, the so are $H_{0,b}$ for all $b_4 < b < b_5$. This will save some time locating all of the necessary b values.

Chapter 11

Linear Statistical Models

11.1 The Method of Least Squares

Commentary

If one is using the software R, the functions `lsfit` and `lm` will perform least squares. While `lsfit` has simpler syntax, `lm` is more powerful. The first argument to `lsfit` is a matrix or vector with one row for each observation and one column for each x variable in the notation of the text (call this `x`). The second argument is a vector of the response values, one for each observation (call this `y`). By default, an intercept is fit. To prevent an intercept from being fit, use the optional argument `intercept=FALSE`. To perform the fit and store the result in `regfit`, use `regfit=lsfit(x,y)`. The result `regfit` is a "list" which contains (among other things) `coef`, the vector of coefficients $\beta_0, \ldots, \beta_k$ in the notation of the text, and `residuals` which are defined later in the text. To access the parts of `regfit`, use `regfit$coef`, etc. To use `lm`, `regfit=lm(y~x)` will perform least squares with an intercept and store the result in `regfit`. To prevent an intercept from being fit, use `regfit=lm(y~x-1)`. The result of `lm` also contains `coefficients` and `residuals` plus `fitted.values` which equals the original `y` minus `residuals`. The components of the output are accessed as above.

The `plot` function in R is useful for visualizing data in linear models. In the notation above, suppose that `x` has only one column. Then `plot(x,y)` will produce a scatterplot of `y` versus `x`. The least-squares line can be added to the plot by `lines(x,regfit$fitted.values)`. (If one used `lsfit`, one can create the fitted values by `regfit$fitted.values=y-regfit$residuals`.)

Solutions to Exercises

1. First write $c_1x_i + c_2 = c_1(x_i - \overline{x}_n) + (c_1\overline{x}_n + c_2)$ for every i. Then

$$(c_1x_i + c_2)^2 = c_1^2(x_i - \overline{x}_n)^2 + (c_1\overline{x}_n + c_2)^2 + 2c_1(x_i - \overline{x}_n)(c_1\overline{x}_n + c_2).$$

The sum over all i from 1 to n of the first two terms on the right produce the formula we desire. The sum of the last term over all i is 0 because $c_1(c_1\overline{x}_n + c_2)$ is the same for all i and $\sum_{i=1}^n (x_i - \overline{x}_n) = 0$.

2. (a) The result can be obtained from Eq. (11.1.1) and the following relations:

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) &= \sum_{i=1}^n (x_iy_i - \bar{x}_n y_i - \bar{y}_n x_i + \bar{x}_n\bar{y}_n) \\ &= \sum_{i=1}^n x_iy_i - \bar{x}_n \sum_{i=1}^n y_i - \bar{y}_n \sum_{i=1}^n x_i + n\bar{x}_n\bar{y}_n \\ &= \sum_{i=1}^n x_iy_i - n\bar{x}_n\bar{y}_n - n\bar{x}_n\bar{y}_n + n\bar{x}_n\bar{y}_n\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}_n\bar{y}_n, \text{ and} \\
\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 &= \sum_{i=1}^{n}(x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}_n \sum_{i=1}^{n} x_i + n\bar{x}_n^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}_n^2 + n\bar{x}_n^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2.
\end{aligned}
$$

(b) The result can be obtained from part (a) and the following relation:

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n) &= \sum_{i=1}^{n}(x_i - \bar{x}_n)y_i - \sum_{i=1}^{n}(\bar{x}_i - \bar{x}_n)\bar{y}_n \\
&= \sum_{i=1}^{n}(x_i - \bar{x}_n)y_i - \bar{y}_n \sum_{i=1}^{n}(x_i - \bar{x}_n) \\
&= \sum_{i=1}^{n}(x_i - \bar{x}_n)y_i, \quad \text{since} \quad \sum_{i=1}^{n}(x_i - \bar{x}_n) = 0.
\end{aligned}
$$

(c) This result can be obtained from part (a) and the following relation:

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n) &= \sum_{i=1}^{n} x_i(y_i - \bar{y}_n) - \bar{x}_n \sum_{i=1}^{n}(y_i - \bar{y}_n) \\
&= \sum_{i=1}^{n} x_i(y_i - \bar{y}_n), \quad \text{since} \quad \sum_{i=1}^{n}(y_i - \bar{y}_n) = 0.
\end{aligned}
$$

3. It must be shown that $\bar{y}_n = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_n$. But this result follows immediately from the expression for $\hat{\beta}_0$ given in Eq. (11.1.1).

4. Since the values of β_0 and β_1 to be chosen must satisfy the relation $\partial Q/\partial\beta_0 = 0$, it follows from Eq. (11.1.3) that they must satisfy the relation $\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$. Similarly, since they must also satisfy relation $\partial Q/\partial\beta_1 = 0$, it follows from Eq. (11.1.4) that $\sum_{i=1}^{n}(y_i - \hat{y}_i)x_i = 0$. These two relations are equivalent to the normal equations (11.1.5), for which $\hat{\beta}_0$ and $\hat{\beta}_1$ are the unique solution.

5. The least squares line will have the form $x = \gamma_0 + \gamma_1 y$, where γ_0 and γ_1 are defined similarly to $\hat{\beta}_0$ and $\hat{\beta}_1$ in Eq. (11.1.1) with the roles of x and y interchanged. Thus,

$$
\gamma_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}_n\bar{y}_n}{\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2}
$$

and $\hat{\gamma}_0 = \bar{x}_n - \hat{\gamma}_1\bar{y}_n$. It is found that $\hat{\gamma}_1 = 0.9394$ and $\hat{\gamma}_0 = 1.5691$. Hence, the least squares line is $x = 1.5691 + 0.9394\, y$ or, equivalently, $y = -1.6703 + 1.0645x$. This line and the line $y = -0.786 + 0.685x$ given in Fig. 11.4 can now be sketched on the same graph.

6. The sum of the squares of the deviations from the least squares parabola is the minimum possible value of $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$ as β_0, β_1, and β_2 vary over all possible real numbers. The sum of the squares of the deviations from the least squares line is the minimum possible value of this same sum when β_2 is restricted to have the value 0. This minimum cannot therefore be smaller than the first minimum.

7. (a) Here, $n = 8$, $\bar{x}_n = 2.25$, $\bar{y}_n = 42.125$, $\sum_{i=1}^{n} x_i y_i = 764$, and $\sum_{i=1}^{n} x_i^2 = 51$. Therefore, by Eq. (11.1.1), $\hat{\beta}_1 = 0.548$ and $\hat{\beta}_0 = 40.893$. Hence, the least squares line is $y = 40.893 + 0.548x$.

   (b) The normal equations (11.1.8) are found to be:

$$\begin{aligned} 8\beta_0 + 18\beta_1 + 51\beta_2 &= 337, \\ 18\beta_0 + 51\beta_1 + 162\beta_2 &= 764, \\ 51\beta_0 + 162\beta_1 + 548.25\beta_2 &= 2167.5. \end{aligned}$$

   Solving these three simultaneous linear equations, we obtain the solution:

$$\hat{\beta}_0 = 38.483, \quad \hat{\beta}_1 = 3.440, \quad \textit{and} \quad \hat{\beta}_2 = -0.643.$$

8. If the polynomial is to pass through the $k+1$ given points, then the following $k+1$ equations must be satisfied:

$$\begin{aligned} &\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_1^k = y_1, \\ &\beta_0 + \beta_1 x_2 + \cdots + \beta_k x_2^k = y_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\beta_0 + \beta_1 x_{k+1} + \cdots + \beta_k x_{k+1}^k = y_{k+1}. \end{aligned}$$

These equations form a set of $k+1$ simultaneous linear equations in $\beta_0, \ldots, \beta_k$. There will be a unique polynomial having the required properties if and only if these equations have a unique solution. These equations will have a unique solution if and only if the $(k+1)\times(k+1)$ matrix of coefficients of $\beta_0, \ldots, \beta_k$ is nonsingular (i.e., has a nonzero determinant). Thus, it must be shown that

$$\det \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{k+1} & x_{k+1}^2 & \cdots & x_{k+1}^k \end{bmatrix} \neq 0.$$

This determinant will be 0 if and only if the $k+1$ columns of the matrix are linearly dependent; i.e., if and only if there exist constants $a_1, \ldots, a_{k+1}$, not all of which are 0, such that

$$a_1 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} x_1 \\ \vdots \\ x_{k+1} \end{bmatrix} + a_3 \begin{bmatrix} x_1^2 \\ \vdots \\ x_{k+1}^2 \end{bmatrix} + \cdots + a_{k+1} \begin{bmatrix} x_1^k \\ \vdots \\ x_{k+1}^k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

But if such constants exist, then the $k+1$ distinct values $x_1, \ldots, x_{k+1}$ will all be roots of the equation

$$a_1 + a_2 x + a_3 x^2 + \cdots + a_{k+1} x^k.$$

It is impossible, however, for a polynomial of degree k or less to have $k+1$ distinct roots unless all the coefficients $a_1, \ldots, a_{k+1}$ are 0. It now follows that the determinant cannot be 0.

9. The normal equations (11.1.13) are found to be

$$\begin{aligned} 10\beta_0 + 1170\beta_1 + 18\beta_2 &= 1359, \\ 1170\beta_0 + 138,100\beta_1 + 2130\beta_2 &= 160,380, \\ 18\beta_0 + 2130\beta_1 + 38\beta_2 &= 2483. \end{aligned}$$

Solving these three simultaneous linear equations, we obtain the solution $\hat{\beta}_0 = 3.7148, \hat{\beta}_1 = 1.1013$, and $\hat{\beta}_2 = 1.8517$.

10. We begin by taking the partial derivative of the following sum with respect to β_0, β_1, and β_2, respectively:

$$\sum_{i=1}^{n} (y_i - \beta_0 x_{i1} - \beta_1 x_{i2} - \beta_2 x_{i2}^2)^2.$$

By setting each of these derivatives equal to 0, we obtain the following normal equations:

$$\begin{aligned} \beta_0 \sum_{i=1}^{n} x_{i1}^2 + \beta_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \beta_2 \sum_{i=1}^{n} x_{i1}x_{i2}^2 &= \sum_{i=1}^{n} x_{i1}y_i, \\ \beta_0 \sum_{i=1}^{n} x_{i1}x_{i2} + \beta_1 \sum_{i=1}^{n} x_{i2}^2 + \beta_2 \sum_{i=1}^{n} x_{i2}^3 &= \sum_{i=1}^{n} x_{i2}y_i, \\ \beta_0 \sum_{i=1}^{n} x_{i1}x_{i2}^2 + \beta_1 \sum_{i=1}^{n} x_{i2}^3 + \beta_2 \sum_{i=1}^{n} x_{i2}^4 &= \sum_{i=1}^{n} x_{i2}^2 y_i. \end{aligned}$$

When the given numerical data are used, these equations are found to be:

$$\begin{aligned} 138,100\beta_0 + 2130\beta_1 + 4550\beta_2 &= 160,380, \\ 2130\beta_0 + 38\beta_1 + 90\beta_2 &= 2483, \\ 4550\beta_0 + 90\beta_1 + 230\beta_2 &= 5305. \end{aligned}$$

Solving these three simultaneous linear equations, we obtain the solution $\hat{\beta}_0 = 1.0270, \hat{\beta}_1 = 17.2934$, and $\hat{\beta}_2 = -4.0186$.

11. In Exercise 9, it is found that

$$\sum_{i=1}^{10} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2 = 102.28.$$

In Exercise 10, it is found that

$$\sum_{i=1}^{10} (y_i - \hat{\beta}_0 x_{i1} - \hat{\beta}_1 x_{i2} - \hat{\beta}_2 x_{i2}^2)^2 = 42.72.$$

Therefore, a better fit is obtained in Exercise 10.

11.2 Regression

Commentary

The regression fallacy is an interesting issue that students ought to see. The description of the regression fallacy appears in Exercise 19 of this section. The discussion at the end of the section on "Design of the Experiment" is mostly of mathematical interest and could be skipped without disrupting the flow of material.

If one is using the software R, the variances and covariance of the least-squares estimators can be computed using the function `ls.diag` or the function `summary.lm`. The first takes as argument the result of `lsfit`, and the second takes as argument the result of `lm`. Both functions return a list containing a matrix that can be extracted via `$cov.unscaled`. For example, using the notation in the Commentary to Sec. 11.1 above, if we had used `lsfit`, then `morefit=ls.diag(regfit)` would contain the matrix `morefit$cov.unscaled`. This matrix, multiplied by the unknown parameter σ^2, would contain the variances of the least-squares estimators on its diagonal and the covaraínces between them in the off-diagonal locations. (If we had used `lm`, then `morefit=summary.lm(regfit)` would be used.)

Solutions to Exercises

1. After we have replaced β_0 and β_1 in (11.2.2) with $\hat{\beta}_0$ and $\hat{\beta}_1$, the maximization with respect to σ^2 is exactly the same as the maximization carried out in Example 7.5.6 in the text for finding the M.L.E. of σ^2.

2. Since $E(Y_i) = \beta_0 + \beta_1 x_i$, it follows from Eq. (11.2.7) that

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} = \frac{\beta_0 \sum_{i=1}^{n}(x_i - \bar{x}_n) + \beta_1 \sum_{i=1}^{n} x_i(x_i - \bar{x}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}.$$

But $\sum_{i=1}^{n}(x_i - \bar{x}_n) = 0$ and

$$\sum_{i=1}^{n} x_i(x_i - \bar{x}_n) = \sum_{i=1}^{n} x_i(x_i - \bar{x}_n) - \bar{x}_n \sum_{i=1}^{n}(x_i - \bar{x}_n) = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2.$$

It follows that $E(\hat{\beta}_1) = \beta_1$.

3. $E(\bar{Y}_n) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i) = \beta_0 + \beta_1 \bar{x}_n$.
Hence, as shown near the end of the proof of Theorem 11.2.2,

$$E(\hat{\beta}_0) = E(\bar{Y}_n) - \bar{x}_n E(\hat{\beta}_1) = (\beta_0 + \beta_1 \bar{x}_n) - \bar{x}_n \beta_1 = \beta_0.$$

4. Let $s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2$. Then

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i - \bar{x}_n \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)Y_i}{s_x^2} = \sum_{i=1}^{n}\left[\frac{1}{n} - \frac{\bar{x}_n}{s_x^2}(x_i - \bar{x}_n)\right] Y_i.$$

Since $Y_1, \ldots, Y_n$ are independent and each has variance σ^2,

$$\begin{aligned}
\operatorname{Var}(\hat{\beta}_0) &= \sum_{i=1}^{n} \left[\frac{1}{n} - \frac{\bar{x}_n}{s_x^2}(x_i - \bar{x}_n)\right]^2 \operatorname{Var}(Y_i) \\
&= \sigma^2 \sum_{i=1}^{n} \left[\frac{1}{n^2} + \frac{\bar{x}_n^2}{s_x^4}(x_i - \bar{x}_n)^2 - \frac{2\bar{x}_n}{ns_x^2}(x_i - \bar{x}_n)\right] \\
&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_x^2} - 0\right) = \frac{s_x^2 + n\bar{x}_n^2}{ns_x^2}\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{ns_x^2}\sigma^2,
\end{aligned}$$

as shown in part (a) of Exercise 2 of Sec. 11.1.

5. Since $\bar{Y}_n = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_n$, then

$$\operatorname{Var}(\bar{Y}_n) = \operatorname{Var}(\hat{\beta}_0) + \bar{x}_n^2 \operatorname{Var}(\hat{\beta}_1) + 2\bar{x}_n \operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1).$$

Therefore, if $\bar{x}_n \neq 0$,

$$\begin{aligned}
\operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= \frac{1}{2\bar{x}_n}\left[\operatorname{Var}(\bar{Y}_n) - \operatorname{Var}(\hat{\beta}_0) - \bar{x}_n^2 \operatorname{Var}(\hat{\beta}_1)\right] \\
&= \frac{1}{2\bar{x}_n}\left(\frac{\sigma^2}{n} - \frac{\sum_{i=1}^{n} x_i^2}{ns_x^2}\sigma^2 - \frac{\bar{x}_n^2}{s_x^2}\sigma^2\right) \\
&= \frac{\sigma^2}{2\bar{x}_n}\left(\frac{s_x^2 - \sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2}{ns_x^2}\right) \\
&= \frac{\sigma^2}{2\bar{x}_n}\left(\frac{-2n\bar{x}_n^2}{ns_x^2}\right) = \frac{-\bar{x}_n\sigma^2}{s_x^2}.
\end{aligned}$$

If $\bar{x}_n = 0$, then $\hat{\beta}_0 = \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$, and

$$\operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \operatorname{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i, \frac{1}{s_x^2}\sum_{j=1}^{n} x_j Y_j\right) = \frac{1}{ns_x^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_j \operatorname{Cov}(Y_i, Y_j),$$

by Exercise 8 of Sec. 4.6. Since $Y_1, \ldots, Y_n$ are independent and each has variance σ^2, then $\operatorname{Cov}(Y_i, Y_j) = 0$ for $i \neq j$ and $\operatorname{Cov}(Y_i, Y_j) = \sigma^2$ for $i = j$.

Hence,

$$\operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{\sigma^2}{ns_x^2}\sum_{i=1}^{n} x_i = 0.$$

6. Both $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear functions of the random variables $Y_1, \ldots, Y_n$ which are independent and have identical normal distributions. As stated in the text, it can therefore be shown that the joint distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ is a bivariate normal distribution. It follows from Eq. (11.2.6) that $\hat{\beta}_0$ and $\hat{\beta}_1$ are uncorrelated when $\bar{x}_n = 0$. Therefore, by the property of the bivariate normal distribution discussed in Sec. 5.10, $\hat{\beta}_0$ and $\hat{\beta}_1$ are independent when $\bar{x}_n = 0$.

7. (a) The M.L.E.'s $\hat{\beta}_0$ and $\hat{\beta}_1$ are the same as the least squares estimates found in Sec. 11.1 for Table 11.1. The value of $\hat{\sigma}^2$ can then be found from Eq. (11.2.3).

   (b) Also, $\text{Var}(\hat{\beta}_0) = 0.2505\sigma^2$ can be determined from Eq. (11.2.5) and $\text{Var}(\hat{\beta}_1) = 0.0277\sigma^2$ from Eq. (11.2.9).

   (c) It can be found from Eq. (11.2.6) that $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -0.0646\sigma^2$. By using the values of $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$ found in part (b), we obtain

$$\rho(\hat{\beta}_0, \hat{\beta}_1) = \frac{\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)}{[\text{Var}(\hat{\beta}_0)\,\text{Var}(\hat{\beta}_1)]^{1/2}} = -0.775.$$

8. $\hat{\theta} = 3\hat{\beta}_0 - 2\hat{\beta}_1 + 5 = 1.272$. Since $\hat{\theta}$ is an unbiased estimator of θ, the M.S.E. of $\hat{\theta}$ is the same as $\text{Var}(\hat{\theta})$ and

$$\text{Var}(\hat{\theta}) = 9\,\text{Var}(\hat{\beta}_0) + 4\,\text{Var}(\hat{\beta}_1) - 12\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = 3.140\sigma^2.$$

9. The unbiased estimator is $3\hat{\beta}_0 + c_1\hat{\beta}_1$. The M.S.E. of an unbiased estimator is its variance, and

$$\text{Var}(\hat{\theta}) = 9\,\text{Var}(\hat{\beta}_0) + 6c_1\,\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + c_1^2\,\text{Var}(\hat{\beta}_1).$$

   Using the values in Exercise 7, we get

$$\text{Var}(\hat{\theta}) = \sigma^2[9 \times 0.2505 - 6c_1(0.0646) + c_1^2 0.0277].$$

   We can minimize this by taking the derivative with respect to c_1 and setting the derivative equal to 0. We get $c_1 = 6.996$.

10. The prediction is $\hat{Y} = \hat{\beta}_0 + 2\hat{\beta}_1 = 0.584$. The M.S.E. of this prediction is

$$\text{Var}(\hat{Y}) + \text{Var}(Y) = \text{Var}(\hat{\beta}_0) + 4\,\text{Var}(\hat{\beta}_1) + 4\,\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \sigma^2 = 1.103\sigma^2.$$

   Alternatively, the M.S.E. of $\hat{Y}$ could be calculated from Eq. (11.2.11) with $x = 2$.

11. By Eq. (11.2.11), the M.S.E. is

$$\left[\frac{1}{ns_x^2}\sum_{i=1}^{n}(x_i - x)^2 + 1\right]\sigma^2.$$

   We know that $\sum_{i=1}^{n}(x_i - x)^2$ will be a minimum (and, hence, the M.S.E. will be a minimum) when $x = \bar{x}_n$.

12. The M.L.E.'s $\hat{\beta}_0$ and $\hat{\beta}_1$ have the same values as this least squares estimates found in part (a) of Exercise 7 of Sec. 11.1. The value of $\hat{\sigma}^2$ can then be found from Eq. (11.2.3). Also, $\text{Var}(\hat{\beta}_0)$ can be determined from Eq. (11.2.5) and $\text{Var}(\hat{\beta}_1)$ from Eq. (11.2.9).

13. It can be found from Eq. (11.2.6) that $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -0.214\sigma^2$. By using the values of $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$ found in Exercise 12, we obtain

$$\rho(\hat{\beta}_0, \hat{\beta}_1) = \frac{\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)}{[\text{Var}(\hat{\beta}_0)\,\text{Var}(\hat{\beta}_1)]^{1/2}} = -0.891.$$

14. $\hat{\theta} = 5 - 4\hat{\beta}_0 + \hat{\beta}_1 = -158.024$. Since $\hat{\theta}$ is an unbiased estimator of θ, the M.S.E. is the same as $\text{Var}(\hat{\theta})$ and

$$\text{Var}(\hat{\theta}) = 16\,\text{Var}(\hat{\beta}_0) + \text{Var}(\hat{\beta}_1) - 8\,\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = 11.524\sigma^2.$$

15. This exercise, is similar to Exercise 9. $\text{Var}(\hat{\theta})$ attains its minimum value when $c_1 = -\bar{x}_n$.

16. The prediction is $\hat{Y} = \hat{\beta}_0 + 3.25\hat{\beta}_1 = 42.673$. The M.S.E. of this prediction is

$$\text{Var}(\hat{Y}) + \text{Var}(Y) = \text{Var}(\hat{\beta}_0) + (3.25)^2\,\text{Var}(\hat{\beta}_1) + 6.50\,\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \sigma^2 = 1.220\sigma^2.$$

Alternatively, the M.S.E. of $\hat{Y}$ could be calculated from Eq. (11.2.11) with $x = 3.25$.

17. It was shown in Exercise 11, that the M.S.E. of $\hat{Y}$ will be a minimum when $x = \bar{x}_n = 2.25$.

18. (a) It is easiest to use a computer to find the least-squares coefficients. These are $\hat{\beta}_0 = -1.234$ and $\hat{\beta}_1 = 2.702$.

    (b) The predicted 1980 selling price for a species that sold for $x = 21.4$ in 1970 is

    $$\hat{\beta}_0 + \hat{\beta}_1 x = -1.234 + 2.702 \times 21.4 = 56.59.$$

    (c) The average of the x_i values is 41.1, and $s_x^2 = 18430$. Use Eq. (11.2.11) to compute the M.S.E. as

    $$\sigma^2\left[1 + \frac{1}{14} + \frac{(21.4 - 41.1)^2}{18430}\right] = 1.093\sigma^2.$$

19. The formula for $E(X_2|x_1)$ is Eq. (5.10.6), which we repeat here for the case in which $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$:

$$E(X_2|x_1) = \mu + \rho\sigma\left(\frac{x_1 - \mu}{\sigma}\right) = \mu + \rho(x_1 - \mu).$$

We are asked to show that $|E(X_2|x_1) - \mu| < |x_1 - \mu|$ for all x_1. Since $0 < \rho < 1$,

$$|E(X_2|x_1) - \mu| = |\mu + \rho(x_1 - \mu) - \mu| = \rho|x_1 - \mu| < |x_1 - \mu|.$$

11.3 Statistical Inference in Simple Linear Regression

Commentary

Computation and plotting of residuals is really only feasible with the help of a computer, except in problems that are so small that you can't learn much from residuals anyway. There is a subsection at the end of this section on joint inference about β_0 and β_1. This material is mathematically more challenging than the rest of the section and might be suitable only for special sets of students.

If one is using the software *R*, both `lm` and `lsfit` provide the residuals. These can then be plotted against any other available variables using `plot`. Normal quantile plots are done easily using `qqnorm` with one argument being the residuals. The function `qqline` (with the same argument) will add a straight line to the plot to help identify curvature and outliers.

Solutions to Exercises

1. It is found from Table 11.9 that $\bar{x}_n = 0.42$, $\bar{y}_n = 0.33$, $\sum_{i=1}^{n} x_i^2 = 10.16$, $\sum_{i=1}^{n} x_i y_i = 5.04$, $\hat{\beta}_1 = 0.435$ and $\hat{\beta}_0 = 0.147$ by Eq. (11.1.1), and $S^2 = 0.451$ by Eq. (11.3.9). Therefore, from Eq. (11.3.19) with $n = 10$ and $\beta_0^* = 0.7$, it is found that $U_0 = -6.695$. It is found from a table of the t distribution with $n-2 = 8$ degrees of freedom that to carry out a test at the 0.05 level of significance, H_0 should be rejected if $|U_0| > 2.306$. Therefore H_0 is rejected.

2. In this exercise, we must test the following hypotheses:

$$H_0 : \beta_0 = 0,$$
$$H_1 : \beta_0 \neq 0.$$

Hence, $\beta_0^* = 0$ and it is found from Eq. (11.3.19) that $U_0 = 1.783$. Since $|U_0| < 2.306$, the critical value found in Exercise 1, we should not reject H_0.

3. It follows from Eq. (11.3.22), with $\beta_1^* = 1$, that $U_1 = -6.894$. Since $|U_1| > 2.306$, the critical value found in Exercise 1, we should reject H_0.

4. In this exercise, we want to test the following hypotheses:

$$H_0 : \beta_1 = 0,$$
$$H_1 : \beta_1 \neq 0.$$

Hence, $\beta_1^* = 0$ and it is found from Eq. (11.3.22) that $U_1 = 5.313$. Since $|U_1| > 2.306$, we should reject H_0.

5. The hypotheses to be tested are:

$$H_0 : 5\beta_0 - \beta_1 = 0,$$
$$H_1 : 5\beta_0 - \beta_1 \neq 0.$$

Hence, in the notation of (11.3.13), $c_0 = 5, c_1 = -1$, and $c_* = 0$. It is found that $\sum_{i=1}^{n}(c_0 x_i - c_1)^2 = 306$ and, from Eq. (11.3.14), that $U_{01} = 0.664$. It is found from a table of the t distribution with $n-2 = 8$ degrees of freedom that to carry out a test at the 0.10 level of significance, H_0 should be rejected if $|U_{01}| > 1.860$. Therefore, H_0 is not rejected.

6. The hypotheses to be tested are:

$$H_0 : \beta_0 + \beta_1 = 1,$$
$$H_1 : \beta_0 + \beta_1 \neq 1.$$

Therefore, $c_0 = c_1 = c_* = 1$. It is found that $\sum_{i=1}^{n}(c_0 x_i - c_1)^2 = 11.76$ and, from Eq. (11.3.14), that $U_{01} = -4.701$. It is found from a table of the t distribution with $n-2 = 8$ degrees of freedom that to carry out a test at the 0.10 level of significance, H_0 should be rejected if $|U_{01}| > 3.355$. Therefore, H_0 is rejected.

7.

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}_1, D) &= \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0 + \hat{\beta}_1\bar{x}_n)\\ &= \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) + \bar{x}_n \,\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_1)\\ &= \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \bar{x}_n \,\mathrm{Var}(\hat{\beta}_1)\\ &= 0, \text{ by Eqs. (11.2.9) and (11.2.6).}\end{aligned}$$

Since $\hat{\beta}_0$ and $\hat{\beta}_1$ have a bivariate normal distribution, it follows from Exercise 10 of Sec. 5.10 that D and $\hat{\beta}_1$ will also have a bivariate normal distribution. Therefore, as discussed in Sec. 5.10, since D and $\hat{\beta}_1$ are uncorrelated they are also independent.

8. (a) We shall add $n\bar{x}_n^2(\hat{\beta}_1 - \hat{\beta}_1^*)^2$ and subtract the same amount to the right side of Q^2, as given by Eq. (11.3.30). The Q^2 can be rewritten as follows:

$$\begin{aligned}Q^2 &= \left(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2\right)(\hat{\beta}_1 - \beta_1^*)^2 + \ n[(\hat{\beta}_0 - \beta_0^*)^2 + 2\bar{x}_n(\hat{\beta}_0 - \beta_0^*)(\hat{\beta}_1 - \beta_1^*) + \bar{x}_n^2(\hat{\beta}_1 - \beta_1^*)^2]\\ &= \sigma^2\frac{(\hat{\beta}_1 - \beta_1^*)^2}{\mathrm{Var}(\hat{\beta}_1)} + n[(\hat{\beta}_0 - \beta_0^*) + \bar{x}_n(\hat{\beta}_1 - \beta_1^*)]^2.\end{aligned}$$

Hence,

$$\frac{Q^2}{\sigma^2} = \frac{(\hat{\beta}_1 - \beta_1^*)^2}{\mathrm{Var}(\hat{\beta}_1)} + \frac{n}{\sigma^2}(D - \beta_0^* - \beta_1^*\bar{x}_n)^2.$$

It remains to show that $\mathrm{Var}(D) = \frac{\sigma^2}{n}$. But

$$\mathrm{Var}(D) = \mathrm{Var}(\hat{\beta}_0) + \bar{x}_n^2 \,\mathrm{Var}(\hat{\beta}_1) + 2\bar{x}_n \,\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1).$$

The desired result can now be obtained from Eqs. (11.2.9), (11.2.5), and (11.2.6).

(b) It follows from Exercise 7 that the random variables $\hat{\beta}_1$ and D are independent and each has a normal distribution. When H_0 is true, $E(\hat{\beta}_1) = \beta_1^*$ and $E(D) = \beta_0^* + \beta_1^*\bar{x}_n$. Hence, H_0 is true, each of the two summands on the right side of the equation given in part (a) is the square of a random variable having a standard normal distribution.

9. Here, $\beta_0^* = 0$ and $\beta_1^* = 1$. It is found that $Q^2 = 2.759$, $S^2 = 0.451$, and $U^2 = 24.48$. It is found from a table of the F distribution with 2 and 8 degrees of freedom that to carry out a test at the 0.05 level of significance, H_0 should be rejected if $U^2 > 4.46$. Therefore, H_0 is rejected.

10. To attain a confidence coefficient of 0.95, it is found from a table of the t distribution with 8 degrees of freedom that the confidence interval will contain all values of β_0^* for which $|U_0| < 2.306$. When we use the numerical values found in Exercise 1, we find that this is the interval of all values of β_0^* such that $-2.306 < 12.111(0.147 - \beta_0^*) < 2.306$ or, equivalently, $-0.043 < \beta_0^* < 0.338$. This interval is, therefore, the confidence interval for β_0.

11. The solution here is analogous to the solution of Exercise 9. Since the confidence coefficient is again 0.95, the confidence interval will contain all values of β_1^* for which $|U_1| < 2.306$ or, equivalently, for which $-2.306 < 12.207(0.435 - \beta_1^*) < 2.306$. The interval is, therefore, found to be $0.246 < \beta_1 < 0.624$.

12. We shall first determine a confidence interval for $5\beta_0 - \beta_1$ with confidence coefficient 0.90. It is found from a table of the t distribution with 8 degrees of freedom (as in Exercise 5) that this confidence interval will contain all values of c_* for which $|U_{01}| < 1.860$ or, equivalently, for which $-1.860 <$

$2.207(0.301 - c_*) < 1.860$. This interval reduces to $-0.542 < c_* < 1.144$. Since this is a confidence interval for $5\beta_0 - \beta_1$, the corresponding confidence interval for $5\beta_0 - \beta_1 + 4$ is the interval with end points $(-0.542) + 4 = 3.458$ and $(1.144) + 4 = 5.144$.

13. We must determine a confidence interval for $y = \beta_0 + \beta_1$ with confidence coefficient 0.99. It is found from a table of the t distribution with 8 degrees of freedom (as in Exercise 6) that this confidence interval will contain all values of c_* for which $|U_{01}| < 3.355$ or, equivalently, for which $-3.355 < 11.257(0.582 - b) < 3.355$. This interval reduces to $0.284 < c_* < 0.880$. This interval is, therefore, the confidence interval for y.

14. We must determine a confidence interval for $y = \beta_0 + 0.42\beta_1$. Since the confidence coefficient is again 0.99, as in Exercise 13, this interval will again contain all values of c_* for which $|U_{01}| < 3.355$. Since $c_0 = 1$ and $c_1 = 0.42 = \bar{x}_n$ in this exercise, the value of $\sum_{i=1}^{n}(c_0 x_i - c_1)^2$, which is needed in determining U_{01}, is equal to $\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = 8.396$. Also, $c_0\hat{\beta}_0 + c_1\hat{\beta}_1 = \hat{\beta}_0 + \hat{\beta}_1$, $\bar{x}_n = \bar{y}_n = 0.33$. Hence it is found that the confidence interval for y contains all values of c_* for which $-3.355 < 13.322(0.33 - c_*) < 3.355$ or, equivalently, for which $0.078 < c_* < 0.582$.

15. Let q be the $1 - \alpha_0/2$ quantile of the t distribution $n-2$ degrees of freedom. A confidence interval for $\beta_0 + \beta_1 x$ contains all values of c_* for which $|U_{01}| < c$, where $c_0 = 1$ and $c_1 = x$ in Eq. (11.3.14). The inequality $|U_{01}| < q$ can be reduced to the following form

$$\hat{\beta}_0 + x\hat{\beta}_1 - q\left[\frac{S^2\sum_{i=1}^{n}(x_i - x)^2}{n(n-2)s_x^2}\right]^{1/2} < b < \hat{\beta}_0 + x\hat{\beta}_1 + q\left[\frac{S^2\sum_{i=1}^{n}(x_i - x)^2}{n(n-2)s_x^2}\right]^{1/2}$$

The length of this interval is

$$2q\left[\frac{S^2}{n(n-2)s_x^2}\quad\sum_{i=1}^{n}(x_i - x)^2\right]^{1/2}$$

The length will, therefore, be a minimum for the value of x which minimizes $\sum_{i=1}^{n}(x_i - x)^2$. We know that this quantity is a minimum when $x = \bar{x}_n$.

16. It is known from elementary calculus that the set of points (x, y) which satisfy an inequality of the form $Ax^2 + Bxy + Cy^2 < c^2$ will be an ellipse (with center at the origin) if and only if $B^2 - 4AC < 0$. It follows from Eqs. (11.3.30) and (11.3.32) that $U^2 < \gamma$ if and only if

$$n(\beta_0^* - \hat{\beta}_0)^2 + 2n\bar{x}_n(\beta_0^* - \hat{\beta}_0)(\beta_1^* - \hat{\beta}_1) + \left(\sum_{i=1}^{n} x_i^2\right)(\beta_1^* - \hat{\beta}_1)^2 < \gamma\frac{2}{n-2}S^2.$$

Hence, the set of points (β_0^*, β_1^*) which satisfy this inequality will be an ellipse [with center at $(\hat{\beta}_0, \hat{\beta}_1)$] if and only if

$$(2n\bar{x}_n)^2 - 4n\sum_{i=1}^{n} x_i^2 < 0$$

or, equivalently, if and only if

$$\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2 > 0.$$

Since the left side of this relation is equal to $\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$, it must be positive, assuming that the numbers $x_1, \dots, x_n$ are not all the same.

17. To attain a confidence coefficient of 0.95, it is found from a table of the F distribution with 2 and 8 degrees of freedom (as in Exercise 9) that the confidence ellipse for (β_0, β_1) will contain all points (β_0^*, β_1^*) for which $U^2 < 4.46$. Hence, it will contain all points for which

$$10(\beta_0^* - 0.147)^2 + 8.4(\beta_0^* - 0.147)(\beta_1^* - 0.435) + 10.16(\beta_1^* - 0.435)^2 < 0.503.$$

18. (a) The upper and lower limits of the confidence band are defined by (11.3.33). In this exercise, $n = 10$ and $(2\gamma)^{1/2} = 2.987$. The values of $\hat{\beta}_0, \hat{\beta}_1$, and S^2 have been found in Exercise 1.
Numerical computation yields the following points on the upper and lower limits of the confidence band:

x	Upper limit	Lower limit
-2	$-.090$	-1.356
-1	.124	$-.700$
0	.395	$-.101$
$\bar{x}_n = 0.42$	.554	.106
1	.848	.316
2	1.465	.569

The upper and lower limits containing these points are shown as the solid curves in Fig. S.11.1.

(b) The upper and lower limits are now given by (11.3.25), where $T_{n-2}(1 - \alpha_0/2) = 2.306$. The corresponding values of these upper and lower limits are as follows:

x	Upper limit	Lower limit
-2	$-.234$	-1.212
-1	.030	$-.606$
0	.338	$-.044$
$\bar{x}_n = 0.42$	.503	.157
1	.787	.377
2	1.363	.671

These upper and lower limits are shown as the dashed curves in Fig. S.11.1.

19. If S^2 is defined by Eq. (11.3.9), then S^2/σ^2 has a χ^2 distribution with $n - 2$ degrees of freedom. Therefore, $E(S^2/\sigma^2) = n - 2, E(S^2) = (n - 2)\sigma^2$, and $E(S^2/[n - 2]) = \sigma^2$.

20. (a) The prediction is $\hat{\beta}_0 + \hat{\beta}_1 X = 68.17 - 1.112 \times 24 = 41.482$.

(b) The 95% predicition interval is centered at the prediction from part (a) and has half-width equal to

$$T_{30}^{-1}(0.975)4.281\left[1 + \frac{1}{32} + \frac{(24 - 30.91)^2}{2054.8}\right]^{1/2} = 8.978.$$

So, the interval is $41.482 \pm 8.978 = [32.50, 50.46]$.

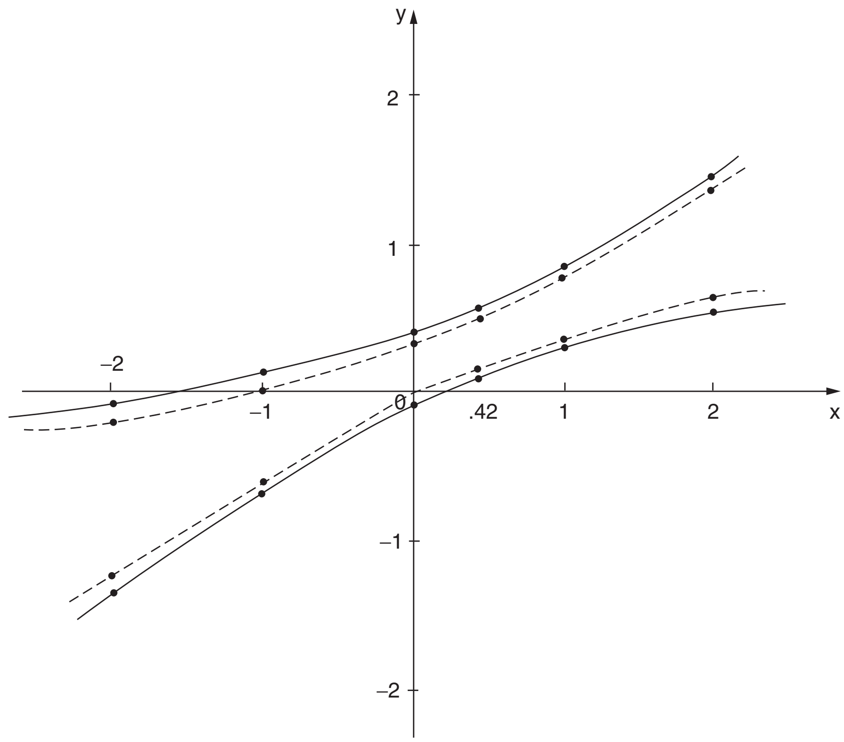

Figure S.11.1: Confidence bands and intervals for Exercise 18a in Sec. 11.3.

21. (a) A computer is useful to perform the regressions and plots for this exercise. The two plots for parts (a) and (b) are side-by-side in Fig. S.11.2. The plot for part (a) shows residuals that are more spread out for larger values of 1970 price than they are for smaller values. This suggests that the variance of Y is not constant as X changes.

    (b) The plot for part (b) in Fig. S.11.2 has more uniform spread in the residuals as 1970 price varies. However, there appear to be two points that are not fit very well.

22. In this problem we are asked to regress logarithm of 1980 fish price on the 1970 fish price. (It would have made more sense to regress on the logarithm of 1970 fish price, but the problem didn't ask for that.) The summary of the regression fit is $\hat{\beta}_0 = 3.099$, $\hat{\beta}_1 = 0.0266$, $\sigma' = 0.6641$, $\overline{x}_n = 41.1$, and $s_x^2 = 18430$.

    (a) The test statistic is given in Eq. (11.3.22),

$$U = s_x \frac{\hat{\beta}_1 - 2}{\sigma'} = 135.8 \frac{0.0266 - 2}{0.6641} = -403.5.$$

    We would reject H_0 at level 0.01 if U is greater than the 0.99 quantile of the t distribution with 12 degrees of freedom. We do not reject the null hypothesis at level 0.01.

    (b) A 90% confidence interval is centered at 0.0266 and has half-width equal to

$$T_{12}^{-1}(0.95) \frac{\sigma'}{s_x} = 1.782 \frac{0.6641}{135.8} = 0.00872.$$

    So, the interval is $0.0266 \pm 0.00872 = [0.0179, 0.0353]$.

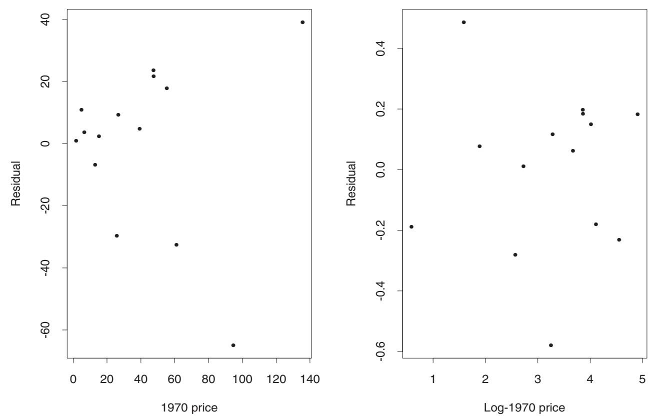

Figure S.11.2: Residual plots for Exercise 21a of Sec. 11.3. The plot on the left is for part (a) and the plot on the right is for part (b).

(c) A 90% prediction interval for the logarithm of 1980 price is centered at $3.099+0.0266\times 21.4 = 3.668$ and has half-width equal to

$$T_{12}^{-1}(0.95)\sigma'\left[1+\frac{1}{n}+\frac{(x-\overline{x}_n)^2}{s_x^2}\right]^{1/2} = 1.782\times 0.6641\left[1+\frac{1}{14}+\frac{(21.4-41.1)^2}{18430}\right]^{1/2} = 1.237.$$

So, the interval for the logarithm of 1980 price is $3.668 \pm 1.237 = [2.431, 4.905]$. To convert this to 1980 price take e to the power of both endpoints to get $[11.37, 134.96]$.

If we had been asked to regress on the logarithm of 1970 fish price, the summary results would have been $\hat{\beta}_0 = 1.132$, $\hat{\beta}_1 = 0.9547$, $\sigma' = 0.2776$, $\overline{x}_n = 3.206$, and $s_x^2 = 19.11$. The test statistic for part (a) would have been $4.371(0.9547-2)/0.2776 = -16.46$. Still, we would not reject the null hypothesis at level 0.01. The confidence interval for β_1 would have been $0.9547 \pm 1.782(0.2776/4.371) = [0.8415, 1.067]$. The prediction interval for the logarithm of price would have been

$$1.132+0.9547\log(21.4)\pm 0.2776\left(1+\frac{1}{14}+\frac{(\log(21.4)-3.206)^2}{19.11}\right)^{1/2} = [3.769, 4.344].$$

The interval for 1980 price would then be $[43.34, 77.02]$.

23. Define

$$W_{01} = \left[\frac{c_0^2}{n}+\frac{(c_0\overline{x}_n-c_1)^2}{s_x^2}\right]^{-1/2}\frac{c_0(\hat{\beta}_0-\beta_0)+c_2(\hat{\beta}_1-\beta_1)}{\sigma'},$$

which has the t distribution with $n-2$ degrees of freedom. Hence

$$\Pr(W_{01} \geq T_{n-2}^{-1}(1-\alpha_0)) = \alpha_0.$$

Suppose that $c_0\beta_0 + c_1\beta_1 < c_*$. Because $[(c_0^2/n)+(c_0\overline{x}_n-c_1)^2/s_x^2]/\sigma' > 0$, it follows that $W_{01} > U_{01}$. Finally, he probability of type I error is

$$\Pr(U_{01} \geq T_{n-2}^{-1}(1-\alpha_0)) \leq \Pr(W_{01} \geq T_{n-2}^{-1}(1-\alpha_0)) = \alpha_0,$$

where the first inequality follows from $W_{01} > U_{01}$.

24. (a) When $c_0 = 1$ and $c_1 = x > 0$, the smallest possible value of $\beta_0^* + x\beta_1^*$ occurs at the smallest values of β_0^* and β_1^* simultaneously. These values are

$$\hat{\beta}_0 - \sigma' \left[\tfrac{1}{n} + \tfrac{\overline{x}_n^2}{s_x^2}\right]^{1/2} T_{n-2}^{-1}\left(1 - \tfrac{\alpha_0}{4}\right),$$
$$\hat{\beta}_1 - \tfrac{\sigma'}{s_x} T_{n-2}^{-1}\left(1 - \tfrac{\alpha_0}{4}\right).$$

Similarly, the largest values occur when β_0^* and β_1^* both take their largest possible values, namely

$$\hat{\beta}_0 + \sigma' \left[\tfrac{1}{n} + \tfrac{\overline{x}_n^2}{s_x^2}\right]^{1/2} T_{n-2}^{-1}\left(1 - \tfrac{\alpha_0}{4}\right),$$
$$\hat{\beta}_1 + \tfrac{\sigma'}{s_x} T_{n-2}^{-1}\left(1 - \tfrac{\alpha_0}{4}\right).$$

The confidence interval is then

$$\left(\hat{\beta}_0 + \hat{\beta}_1 x - \sigma' \left\{\left[\frac{1}{n} + \frac{\overline{x}_n^2}{s_x^2}\right]^{1/2} + \frac{x}{s_x}\right\} T_{n-2}^{-1}\left(1 - \frac{\alpha_0}{4}\right),\right.$$
$$\left.\hat{\beta}_0 + \hat{\beta}_1 x + \sigma' \left\{\left[\frac{1}{n} + \frac{\overline{x}_n^2}{s_x^2}\right]^{1/2} + \frac{x}{s_x}\right\} T_{n-2}^{-1}\left(1 - \frac{\alpha_0}{4}\right)\right).$$

(b) When $c_0 = 1$ and $c_1 = x < 0$, the smallest possible value of $\beta_0^* + x\beta_1^*$ occurs when β_0^* takes its smallest possible value and β_1^* takes its largest possible value. Similarly, the largest possible value of $\beta_0^* + x\beta_1^*$ occurs when β_0^* takes its largest possible value and β_1^* takes its smallest possible value. All of these extreme values are given in part (a). The resulting interval is then

$$\left(\hat{\beta}_0 + \hat{\beta}_1 x - \sigma' \left\{\left[\frac{1}{n} + \frac{\overline{x}_n^2}{s_x^2}\right]^{1/2} - \frac{x}{s_x}\right\} T_{n-2}^{-1}\left(1 - \frac{\alpha_0}{4}\right),\right.$$
$$\left.\hat{\beta}_0 + \hat{\beta}_1 x + \sigma' \left\{\left[\frac{1}{n} + \frac{\overline{x}_n^2}{s_x^2}\right]^{1/2} - \frac{x}{s_x}\right\} T_{n-2}^{-1}\left(1 - \frac{\alpha_0}{4}\right)\right).$$

25. (a) The simultaneous intervals are the same as (11.3.33) with $[2F_{2,n-2}^{-1}(1-\alpha_0)]^{1/2}$ replaced by $T_{n-2}^{-1}(1-\alpha_0/4)$, namely for $i = 0, 1$,

$$\beta_0 + \beta_1 x_i \pm T_{n-2}^{-1}(1 - \alpha_0/4)\sigma' \left[\frac{1}{n} + \frac{(x_i - \overline{x}_n)^2}{s_x^2}\right]^{1/2}.$$

(b) Set $x = \alpha x_0 + (1 - \alpha)x_1$ and solve for α. The result is, by straightforward algebra,

$$\alpha(x) = \frac{x - x_1}{x_0 - x_1}.$$

(c) First, notice that for all x,

$$\beta_0 + \beta_1 x = \alpha(x)[\beta_0 + \beta_1 x_0] + [1 - \alpha(x)][\beta_0 + \beta_1 x_1]. \qquad \text{(S.11.1)}$$

That is, each parameter for which we want a confidence interval is a convex combination of the parameters for which we already have confidence intervals.

Suppose that C occurs. There are three cases that depend on where $\alpha(x)$ lies relative to the interval $[0, 1]$. The first case is when $0 \le \alpha(x) \le 1$. In this case, the smallest of the four numbers defining $L(x)$ and $U(x)$ is $L(x) = \alpha(x)A_0 + [1 - \alpha(x)]A_1$ and the largest is $U(x) = \alpha(x)B_0 + [1 - \alpha(x)]B_1$, because both $\alpha(x)$ and $1-\alpha(x)$ are nonnegative. For all such x, $A_0 < \beta_0+\beta_1 x_0 < B_0$ and $A_1 < \beta_0+\beta_1 x_1 < B_1$ together imply that

$$\alpha(x)A_0 + [1-\alpha(x)]A_1 < \alpha(x)[\beta_0+\beta_1 x_0] + [1-\alpha(x)][\beta_0+\beta_1 x_1] < \alpha(x)B_0 + [1-\alpha(x)]B_1.$$

Combining this with (S.11.1) and the formulas for $L(x)$ and $U(x)$ yields $L(x) < \beta_0+\beta_1 x < U(x)$ as desired. The other two cases are similar, so we shall do one only of them. If $\alpha(x) < 0$, then $1-\alpha(x) > 0$. In this case, the smallest of the four numbers defining $L(x)$ and $U(x)$ is $L(x) = \alpha(x)B_0 + [1-\alpha(x)]A_1$, and the largest is $U(x) = \alpha(x)A_0 + [1-\alpha(x)]B_1$. For all such x, $A_0 < \beta_0+\beta_1 x_0 < B_0$ and $A_1 < \beta_0+\beta_1 x_1 < B_1$ together imply that

$$\alpha(x)B_0 + [1-\alpha(x)]A_1 < \alpha(x)[\beta_0+\beta_1 x_0] + [1-\alpha(x)][\beta_0+\beta_1 x_1] < \alpha(x)A_0 + [1-\alpha(x)]B_1.$$

Combining this with (S.11.1) and the formulas for $L(x)$ and $U(x)$ yields $L(x) < \beta_0+\beta_1 x < U(x)$ as desired.

11.4 Bayesian Inference in Simple Linear Regression

Commentary

This section only discusses Bayesian analysis with improper priors. There are a couple of reasons for this. First, the posterior distribution that results from the improper prior makes many of the Bayesian inferences strikingly similar to their non-Bayesian counterparts. Second, the derivation of the posterior distribution from a proper prior is mathematically much more difficult than the derivation given here, and I felt that this would distract the reader from the real purpose of this section, namely to illustrate Bayesian posterior inference. This section describes some inferences that are similar to non-Bayesian inferences as well as some that are uniquely Bayesian.

Solutions to Exercises

1. The posterior distribution of β_1 is given as a special case of (11.4.1), namely that $U = s_x(\beta_1 - \hat{\beta}_1)/\sigma'$ has the t distribution with $n-2$ degrees of freedom. The coefficient $1-\alpha_0$ confidence interval from Sec. 11.3 has endpoints $\hat{\beta}_1 \pm T_{n-2}^{-1}(1-\alpha_0/2)\sigma'/s_x$. So, we can compute the posterior probability that β_1 is in the interval as follows:

$$\begin{aligned}&\Pr\left(\hat{\beta}_1 - T_{n-2}^{-1}(1-\alpha_0/2)\frac{\sigma'}{s_x} < \beta_1 < \hat{\beta}_1 + T_{n-2}^{-1}(1-\alpha_0/2)\frac{\sigma'}{s_x}\right)\\ &= \Pr\left(-T_{n-2}^{-1}(1-\alpha_0/2) < s_x\frac{\beta_1-\hat{\beta}_1}{\sigma'} < T_{n-2}^{-1}(1-\alpha_0/2)\right). \qquad \text{(S.11.2)}\end{aligned}$$

Since the t distributions are symmetric around 0, $-T_{n-2}^{-1}(1-\alpha_0/2) = T_{n-2}^{-1}(\alpha_0/2)$. Also, the random variable between the inequalities on the right side of (S.11.2) is U, which has the t distribution with $n-2$ degrees of freedom. Hence the right side of (S.11.2) equals

$$\Pr(U < T^{-1}(1-\alpha_0/2)) - \Pr(U \le T^{-1}(\alpha_0/2)) = 1-\alpha_0/2-\alpha_0/2 = 1-\alpha_0.$$

2. The posterior distribution of β_1 is given in (11.4.1), namely that

$$U = \left[\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_x^2}\right]^{-1/2} \frac{c_0\beta_0 + c_1\beta_1 - [c_0\hat{\beta}_0 + c_1\hat{\beta}_1]}{\sigma'}$$

has the t distribution with $n-2$ degrees of freedom. The coefficient $1-\alpha_0$ confidence interval from Sec. 11.3 has endpoints

$$c_0\hat{\beta}_0 + c_1\hat{\beta}_1 \pm T^{-1}_{n-2}(1-\alpha_0/2)\sigma' \left[\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_x^2}\right]^{1/2}.$$

So, we can compute the posterior probability that β_1 is in the interval as follows:

$$\begin{aligned}\Pr\Bigg(& c_0\hat{\beta}_0 + c_1\hat{\beta}_1 - T^{-1}_{n-2}(1-\alpha_0/2)\sigma' \left[\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_x^2}\right]^{1/2} < c_0\beta_0 + c_1\beta_1 \\ & < c_0\hat{\beta}_0 + c_1\hat{\beta}_1 - T^{-1}_{n-2}(1-\alpha_0/2)\sigma' \left[\frac{c_0^2}{n} + \frac{(c_0\overline{x}_n - c_1)^2}{s_x^2}\right]^{1/2}\Bigg) \\ = {} & \Pr\left(-T^{-1}_{n-2}(1-\alpha_0/2) < U < T^{-1}_{n-2}(1-\alpha_0/2)\right).\end{aligned}$$

As in the proof of Exercise 1, this equals $1-\alpha_0$.

3. The joint distribution of (β_0, β_1) given τ is a bivariate normal distribution as specified in Theorem 11.4.1. Using the means, variances, and correlation given in that theorem, we compute the mean of $\beta_0 + \beta_1 x$ as $\hat{\beta}_0 + \hat{\beta}_1 x = \hat{Y}$. The variance of $\beta_0 + \beta_1 x$ given τ is

$$\frac{1}{\tau}\left[\frac{1}{n} + \frac{\overline{x}_n^2}{s_x^2} + \frac{x^2}{s_x^2} - 2xn\frac{\overline{x}_n}{\left(n\sum_{i=1}^n x_i^2\right)^{1/2}}\right]\left(\frac{1}{n} + \frac{\overline{x}_n^2}{s_x^2}\right)^{1/2}\frac{1}{s_x}.$$

Use the fact that $1/n + \overline{x}_n^2/s_x^2 = \sum_{i=1}^n x_i^2/[ns_x^2]$ to simplify the above variance to the expression

$$\frac{1}{\tau}\left[\frac{1}{n} + \frac{(x-\overline{x}_n)^2}{s_x^2}\right]^{1/2}.$$

It follows that the conditional distribution of $\tau^{1/2}(\beta_0 - \beta_1 x - \hat{Y})$ is the normal distribution with mean 0 and variance as stated in the exercise.

4. The summaries from a simple linear regression are $\hat{\beta}_0 = 0.1472$, $\hat{\beta}_1 = 0.4352$, $\sigma' = 0.2374$, $\overline{x}_n = 0.42$, $n = 10$ and $s_x^2 = 8.396$.

   (a) The posterior distribution of the parameters is given in Theorem 11.4.1. With the numerical summaries above (recall that $\sum_{i=1}^n x_i^2 = s_x^2 + n\overline{x}_n^2 = 10.16$), we get the following posterior. Conditional on τ, (β_0, β_1) has a bivariate normal distribution with mean vector $(0.1472, 0.4352)$, correlation -0.4167, and variances $0.1210/\tau$ and $0.1191/\tau$. The distribution of τ is a gamma distribution with parameters 4 and 0.2254.

   (b) The interval is centered at 0.4352 with half-width equal to $T_8^{-1}(0.95)$ times $0.2374/8.396^{1/2} = 0.0819$. So, the interval is $[0.2828, 0.5876]$.

(c) The posterior distribution of β_0 is that $U = (\beta_0 - 0.1472)/0.1210^{1/2}$ has the t distribution with 8 degrees of freedom. So the probability that β_0 is between 0 and 2 is the probability that U is between $(0 - 0.1472)/0.3479 = -0.4232$ and $(2 - 0.1472)/0.3479 = 5.326$. The probability that a t random variable with 8 degrees of freedom is between these two numbers can be found using a computer program, and it equals 0.6580.

5. The summary data are in the solution to Exercise 4.

(a) According to Theorem 11.4.1, the posterior distribution of β_1 is that $2.898(\beta_1 - 0.4352)/0.2374$ has the t distribution with eight degrees of freedom.

(b) According to Theorem 11.4.1, the posterior distribution of $\beta_0 + \beta_1$ is that

$$\left(0.1 + \frac{(0.42 - 1)^2}{2.898^2}\right)^{-1/2} \frac{\beta_0 + \beta_1 - 0.5824}{0.2374}$$

has the t distribution with eight degrees of freedom.

6. The summary information from the regression is $\hat{\beta}_0 = 1.132$, $\hat{\beta}_1 = 0.9547$, $\sigma' = 0.2776$, $\overline{x}_n = 3.206$, $n = 14$, and $s_x^2 = 19.11$.

(a) The posterior distribution of β_1 is that $U = 19.11^{1/2}(\beta_1 - 0.9547)/0.2776$ has the t distribution with 12 degrees of freedom.

(b) The probability that $\beta_1 \le 2$ is the same as the probability that $U \le 19.11^{1/2}(2 - 0.9547)/0.2776 = 16.46$, which is essentially 1.

(c) The interval for log-price will be centered at $1.132 + 0.9547 \times \log(21.4) = 4.057$ and have half-width $T_{12}^{-1}(0.975)$ times $0.2776[1 + 1/14 + (3.206 - \log(21.4))^2/19.11]^{1/2} = 0.2875$. So, the interval for log-price is $[3.431, 4.683]$. The interval for 1980 price is e to the power of the endpoints, $[30.90, 108.1]$.

7. The conditional mean of β_0 given β_1 can be computed using results from Sec. 5.10. In particular,

$$E(\beta_0|\beta_1) = \hat{\beta}_0 - \frac{n\overline{x}_n s_x \left(\frac{1}{n} + \overline{x}_n^2/s_x^2\right)^{1/2}}{\left(n\sum_{i=1}^{n} x_i^2\right)^{1/2}}(\beta_1 - \hat{\beta}_1).$$

Now, use the fact that $\sum_{i=1}^{n} x_i^2 = s_x^2 + n\overline{x}_n^2$. The result is

$$E(\beta_0|\beta_1) = \hat{\beta}_0 + \overline{x}_n(\beta_1 - \hat{\beta}_1).$$

11.5 The General Linear Model and Multiple Regression

Commentary

If one is using the software R, the commands to fit multiple linear regression models are the same as those that fit simple linear regression as described in the Commentaries to Secs. 11.1–11.3. One need only put the additional predictor variables into additional columns of the `x` matrix.

Solutions to Exercises

1. After we have replaced $\beta_0, \dots, \beta_p$ in (11.5.4) by their M.L.E.'s $\hat{\beta}_0, \dots, \hat{\beta}_p$, the maximization with respect to σ^2 is exactly the same as the maximization carried out in Example 7.5.6 in the text for finding the M.L.E. of σ^2 or the maximization carried out in Exercise 1 of Sec. 11.2.

2. (The statement of the exercise should say that S^2/σ^2 has a χ^2 distribution.) According to Eq. (11.5.8),

$$\sigma'^2 = \frac{S^2}{n-p}.$$

Since we assume that S^2/σ^2 has a χ^2 distribution with $n-p$ degrees of freedom, the mean of S^2 is $\sigma^2(n-p)$, hence the mean of σ'^2 is σ^2, and σ'^2 is unbiased.

3. This problem is a special case of the general linear model with $p = 1$. The design matrix $\boldsymbol{Z}$ defined by Eq. (11.5.9) has dimension $n \times 1$ and is specified as follows:

$$\boldsymbol{Z} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Therefore, $\boldsymbol{Z}'\boldsymbol{Z} = \sum_{i=1}^{n} x_i^2$ and $(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \dfrac{1}{\sum_{i=1}^{n} x_i^2}$.

It follows from Eq. (11.5.10) that

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

4. From Theorem 11.5.3, $E(\hat{\beta}) = \beta$ and $\text{Var}(\hat{\beta}) = \sigma^2 / \sum_{i=1}^{n} x_i^2$.

5. It is found that $\sum_{i=1}^{n} x_i y_i = 342.4$ and $\sum_{i=1}^{n} x_i^2 = 66.8$. Therefore, from Exercises 3 and 4, $\hat{\beta} = 5.126$ and $\text{Var}(\hat{\beta}) = 0.0150\sigma^2$. Also, $S^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}x_i)^2 = 169.94$. Therefore, by Eq. (11.5.7), $\hat{\sigma}^2 = (169.94)/10 = 16.994$.

6. By Eq. (11.5.21), the following statistic will have the t distribution with 9 degrees of freedom when H_0 is true:

$$U = \left[\frac{9}{(0.0150)(169.94)}\right]^{1/2} (\hat{\beta} - 10) = -9.158.$$

The corresponding two-sided tail area is smaller than 0.01, the smallest two-sided tail area available from the table in the back of the book.

7. The values $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$ were determined in Example 11.1.3.

   By Eq. (11.5.7),

$$\hat{\sigma}^2 = \frac{1}{10}S^2 = \frac{1}{10}\sum_{i=1}^{10}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)^2 = \frac{1}{10}(9.37) = 0.937.$$

8. The design matrix $\boldsymbol{Z}$ has the following form:

$$\boldsymbol{Z} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdots & \cdots & \cdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

   Therefore, $\boldsymbol{Z}'\boldsymbol{Z}$ is the 3×3 matrix of coefficients on the left side of the three equations in (11.1.14):

$$\boldsymbol{Z}'\boldsymbol{Z} = \begin{bmatrix} 10 & 23.3 & 90.37 \\ 23.3 & 90.37 & 401 \\ 90.37 & 401 & 1892.7 \end{bmatrix}.$$

   It will now be found that

$$(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 0.400 & -0.307 & 0.046 \\ -0.307 & 0.421 & -0.074 \\ 0.046 & -0.074 & 0.014 \end{bmatrix}.$$

   The elements of $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$, multiplied by σ^2, are the variances and covariances of $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$.

9. By Eq. (11.5.21), the following statistic will have the t distribution with 7 degrees of freedom when H_0 is true:

$$U_2 = \left[\frac{7}{(0.014)(9.37)}\right]^{1/2} \hat{\beta}_2 = 0.095.$$

   The corresponding two-sided tail area is greater than 0.90. The null hypothesis would not be rejected at any reasonable level of significance.

10. By Eq. (11.5.21), the following statistic will have the t distribution with 7 degrees of freedom when H_0 is true:

$$U_1 = \left[\frac{7}{(0.421)(9.37)}\right]^{1/2} (\hat{\beta}_1 - 4) = -4.51.$$

   The corresponding two-sided tail area is less than 0.01.

11. It is found that $\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = 26.309$. Therefore,

$$R^2 = 1 - \frac{S^2}{26.309} = 0.644.$$

12. The values of $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ were determined in Example 11.1.5.
   By Eq. (11.5.7),

$$\hat{\sigma}^2 = \frac{1}{10}S^2 = \frac{1}{10}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2 = \frac{1}{10}(8.865) = 0.8865.$$

13. The design matrix $\boldsymbol{Z}$ has the following form:

$$\boldsymbol{Z} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}.$$

   Therefore, $\boldsymbol{Z}'\boldsymbol{Z}$ is the 3×3 matrix of coefficients on the left side of the three equations in (11.1.14):

$$\boldsymbol{Z}'\boldsymbol{Z} = \begin{bmatrix} 10 & 23.3 & 650 \\ 23.3 & 90.37 & 1563.6 \\ 650 & 1563.6 & 42,334 \end{bmatrix}.$$

   It will now be found that

$$(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 222.7 & 4.832 & -3.598 \\ 4.832 & 0.1355 & -0.0792 \\ -3.598 & -0.0792 & 0.0582 \end{bmatrix}.$$

   The elements of $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$, multiplied by σ^2, are the variances and covariances of $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$.

14. By Eq. (11.5.21), the following statistics will have the t distribution with 7 degrees of freedom when H_0 is true:

$$U_1 = \left[\frac{7}{(0.1355)(8.865)}\right]^{1/2} \hat{\beta}_1 = 1.087.$$

   The corresponding two-sided tail area is between 0.30 and 0.40.

15. By Eq. (11.5.21), the following statistic will have the t distribution with 7 degrees of freedom when H_0 is true:

$$U_2 = \left[\frac{7}{(0.0582)(8.865)}\right]^{1/2} (\hat{\beta}_2 + 1) = 4.319.$$

   The corresponding two-sided tail area is less than 0.01.

16. Just as in Exercise 11, $\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = 26.309$. Therefore,

$$R^2 = 1 - \frac{S^2}{26.309} = 0.663.$$

17.

$$
\begin{aligned}
\mathrm{Cov}(\hat{\beta}_j, A_{ij}) &= \mathrm{Cov}\left(\hat{\beta}_j, \hat{\beta}_i - \frac{\zeta_{ij}}{\zeta_{jj}}\hat{\beta}_j\right) \\
&= \mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_i) - \frac{\zeta_{ij}}{\zeta_{jj}}\,\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_j) \\
&= \mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) - \frac{\zeta_{ij}}{\zeta_{jj}}\,\mathrm{Var}(\hat{\beta}_j) \\
&= \zeta_{ij}\sigma^2 - \frac{\zeta_{ij}}{\zeta_{jj}}\zeta_{jj}\sigma^2 = 0.
\end{aligned}
$$

Just as in simple linear regression, it can be shown that the joint distribution of two estimators $\hat{\beta}_i$ and $\hat{\beta}_j$ will be a bivariate normal distribution. Since A_{ij} is a linear function of $\hat{\beta}_i$ and $\hat{\beta}_j$, the joint distribution of A_{ij} and $\hat{\beta}_j$ will also be a bivariate normal distribution. Therefore, since A_{ij} and $\hat{\beta}_j$ are uncorrelated, they are also independent.

18.

$$
\begin{aligned}
\mathrm{Var}(A_{ij}) &= \mathrm{Var}\left(\hat{\beta}_i - \frac{\zeta_{ij}}{\zeta_{jj}}\hat{\beta}_j\right) \\
&= \mathrm{Var}(\hat{\beta}_i) + \left(\frac{\zeta_{ij}}{\zeta_{jj}}\right)^2 \mathrm{Var}(\hat{\beta}_j) - 2\frac{\zeta_{ij}}{\zeta_{jj}}\,\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) \\
&= \zeta_{ii}\sigma^2 + \frac{\zeta_{ij}^2}{\zeta_{jj}}\sigma^2 - 2\frac{\zeta_{ij}^2}{\zeta_{jj}}\sigma^2 = \left(\zeta_{ii}^2 - \frac{\zeta_{ij}^2}{\zeta_{jj}}\right)\sigma^2.
\end{aligned}
$$

Now consider the right side of the equation given in the hint for this exercise.

$$
\begin{aligned}
[A_{ij} - E(A_{ij})]^2 &= [\hat{\beta}_i - \beta_i - \frac{\zeta_{ij}}{\zeta_{jj}}(\hat{\beta}_j - \beta_j)]^2 \\
&= (\hat{\beta}_i - \beta_i)^2 - \frac{2\zeta_{ij}}{\zeta_{jj}}(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j) + \frac{\zeta_{ij}^2}{\zeta_{jj}^2}(\hat{\beta}_j - \beta_j)^2.
\end{aligned}
$$

If each of the two terms on the right side of the equation given in the hint is put over the least common denominator $(\zeta_{ii}\zeta_{jj} - \zeta_{ij}^2)\sigma^2$, the right side can be reduced to the form given for W^2 in the text of the exercise. In the equation for W^2 given in the hint, W^2 has been represented as the sum of two independent random variables, each of which is the square of a variable having a standard normal distribution. Therefore, W^2 has a χ^2 distribution with 2 degrees of freedom.

19. (a) Since W^2 is a function only of $\hat{\beta}_i$ and $\hat{\beta}_j$, it follows that W^2 and S^2 are independent. Also, W^2 has a χ^2 distribution with 2 degrees of freedom and S^2/σ^2 has a χ^2 distribution with $n-p$ degrees of freedom. Therefore, $\dfrac{W^2/2}{S^2/[\sigma^2(n-p)]}$ has the F distribution with 2 and $n-p$ degrees of freedom.

(b) If we replace β_i and β_j in W^2 by their hypothesized values β_i^* and β_j^*, then the statistic given in part (a) will have the F distribution with 2 and $n-p$ degrees of freedom when H_0 is true and will tend to be larger when H_0 is not true. Therefore, we should reject H_0 if that statistic exceeds some constant C, where C can be chosen to obtain any specified level of significance $\alpha_0 (0 < \alpha_0 < 1)$.

20. In this problem $i = 2, j = 3, \beta_1^* = \beta_2^* = 0$ and, from the values found in Exercises 7 and 8,

$$W^2 = \frac{(0.014)(0.616)^2 + (0.421)(0.013)^2 + 2(0.074)(0.616)(0.13)}{[(0.421)(0.014) - (0.074)^2]\sigma^2} = \frac{16.7}{\sigma^2}.$$

Also, $S^2 = 9.37$, as found in the solution of Exercise 7. Hence, the value of the F statistic with 2 and 7 degrees of freedom is $(7/2)(16.7/9.37) = 6.23$. The corresponding tail area is between 0.025 and 0.05.

21. In this problem, $i = 2, j = 3, \beta_1^* = 1, \beta_2^* = 0$ and from the values found in Exercises 12 and 13,

$$\begin{aligned} W^2 &= \frac{(0.0582)(0.4503 - 1)^2 + (0.1355)(0.1725)^2 + 2(0.0792)(0.4503 - 1)(0.1725)}{[(0.1355)(0.0582) - (0.0792)^2]\sigma^2} \\ &= \frac{4.091}{\sigma^2}. \end{aligned}$$

Also, $S^2 = 8.865$, as found in the solution of Exercise 12. Hence, the value of the F statistic with 2 and 7 degrees of freedom is $(7/2)(4.091/8.865) = 1.615$. The corresponding tail area is greater than 0.05.

22. $S^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. Since $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 x_n$,

$$\begin{aligned} S^2 &= \sum_{i=1}^{n}[(y_i - \bar{y}_n) - \hat{\beta}_1(x_i - \bar{x}_n)]^2 \\ &= \sum_{i=1}^{n}(y_i - \bar{y}_n)^2 - \hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 - 2\hat{\beta}_1 \sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n). \end{aligned}$$

Since $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}$ and $R^2 = 1 - \dfrac{S^2}{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}$, the desired result can now be obtained.

23. We have the following relations:

$$\begin{aligned} E(\boldsymbol{X} + \boldsymbol{Y}) &= E\begin{bmatrix} X_1 + Y_1 \\ \vdots \\ X_n + Y_n \end{bmatrix} = \begin{bmatrix} E(X_1 + Y_1) \\ \vdots \\ E(X_n + Y_n) \end{bmatrix} = \begin{bmatrix} E(X_1) + E(Y_1) \\ \vdots \\ E(X_n) + E(Y_n) \end{bmatrix} \\ &= \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix} + \begin{bmatrix} E(Y_1) \\ \vdots \\ E(Y_n) \end{bmatrix} = E(\boldsymbol{X}) + E(\boldsymbol{Y}). \end{aligned}$$

24. The element in row i and column j of the $n \times n$ matrix $\text{Cov}(\boldsymbol{X} + \boldsymbol{Y})$ is $\text{Cov}(X_i + Y_i, X_j + Y_j) = \text{Cov}(X_i, X_j) + \text{Cov}(X_i, Y_j) + \text{Cov}(Y_i, X_j) + \text{Cov}(Y_i, Y_j)$. Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, $\text{Cov}(X_i, Y_j) = 0$ and $\text{Cov}(Y_i, X_j) = 0$. Therefore, this element reduces to $\text{Cov}(X_i, X_j) + \text{Cov}(Y_i, Y_j)$. But $\text{Cov}(X_i, X_j)$ is the element in row i and column j of $\text{Cov}(\boldsymbol{X})$, and $\text{Cov}(Y_i, Y_j)$ is the corresponding element in $\text{Cov}(\boldsymbol{Y})$. Hence, the sum of these two covariances is the element in row i and column j of $\text{Cov}(\boldsymbol{X}) + \text{Cov}(\boldsymbol{Y})$. Thus, we have shown that the element in row i and column j of $\text{Cov}(\boldsymbol{X} + \boldsymbol{Y})$ is equal to the corresponding element of $\text{Cov}(\boldsymbol{X}) + \text{Cov}(\boldsymbol{Y})$.

25. We know that $\text{Var}(3Y_1 + Y_2 - 2Y_3 + 8) = \text{Var}(3Y_1 + Y_2 - 2Y_3)$. By Theorem 11.5.2, with $p = 1$,

$$\text{Var}(3Y_1 + Y_2 - 2Y_3) = (3, 1, -2)\,\text{Cov}(\boldsymbol{Y})\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} = 87.$$

26. (a) We can see that $\sum_{i=0}^{p-1} c_j\hat{\beta}_j$ is equal to $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$, where $\boldsymbol{c}$ is defined in part (b) and $\hat{\boldsymbol{\beta}}$ is the least-squares regression coefficient vector. If $\boldsymbol{Y}$ is the vector in Eq. (11.5.13), then we can write $\boldsymbol{c}'\hat{\boldsymbol{\beta}} = \boldsymbol{a}'\boldsymbol{Y}$, where $\boldsymbol{a}' = \boldsymbol{c}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}$. It follows from Theorem 11.3.1 that $\boldsymbol{a}'\boldsymbol{Y}$ has a normal distribution, and it follows from Theorem 11.5.2 that the mean of $\boldsymbol{a}'\boldsymbol{Y}$ is $\boldsymbol{c}'\boldsymbol{\beta}$ and the variance is

$$\sigma^2\boldsymbol{a}'\boldsymbol{a} = \sigma^2\boldsymbol{c}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{c}. \tag{S.11.3}$$

(b) If H_0 is true, then $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ has the normal distribution with mean c_* and variance given by (S.11.3). It follows that the following random variable Z has a standard normal distribution:

$$Z = \frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - c_*}{\sigma(\boldsymbol{c}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{c})^{1/2}}.$$

Also, recall that $(n-p)\sigma'^2/\sigma^2$ has a χ^2 distribution with $n-p$ degrees of freedom and is independent of Z. So, if we divide Z by σ'/σ, we get a random variable that has a t distribution with $n-p$ degrees of freedom, which also happens to equal U.

(c) To test H_0 at level α_0, we can reject H_0 if $|U| > T_{n-p}^{-1}(1 - \alpha_0/2)$. If H_0 is true,

$$\Pr(|U| > T_{n-p}^{-1}(1 - \alpha_0/2)) = \alpha_0,$$

so this test will have level α_0.

27. In a simple linear regression, $\hat{Y}_i$ is the same linear function of X_i for all i. If $\hat{\beta}_1 > 0$, then every unit increase in X corresponds to an increase of $\hat{\beta}_1$ in $\hat{Y}$. So, a plot of residuals against $\hat{Y}$ will look the same as a plot of residuals against X except that the horizontal axis will be labeled differently. If $\hat{\beta}_1 < 0$, then a unit increase in X corresponds to a decrease of $-\hat{\beta}_1$ in $\hat{Y}$, so a plot of residuals against fitted values is a mirror image of a plot of residuals against X. (The plot is flipped horizontally around a vertical line.)

28. Since R^2 is a decreasing function of the residual sum of squares, we shall show that the residual sum of squares is at least as large when using $\boldsymbol{Z}'$ as when using $\boldsymbol{Z}$. Let $\boldsymbol{Z}$ have p columns and let $\boldsymbol{Z}'$ have $q < p$ columns. Let $\hat{\boldsymbol{\beta}}_*$ be the least-squares coefficients that we get when using design matrix $\boldsymbol{Z}'$. For each column that was deleted from $\boldsymbol{Z}$ to get $\boldsymbol{Z}'$, insert an additional coordinate equal to 0 into the q-dimensional vector $\hat{\boldsymbol{\beta}}_*$ to produce the p-dimensional vector $\tilde{\boldsymbol{\beta}}$. This vector $\tilde{\boldsymbol{\beta}}$ is one of the possible vectors in the solution of the minimization problem to find the least-squares estimates with the design matrix $\boldsymbol{Z}$. Furthermore, since $\tilde{\boldsymbol{\beta}}$ has 0's for all of the extra columns that are in $\boldsymbol{Z}$ but not in $\boldsymbol{Z}'$, it follows that the residual sum of squares when using $\tilde{\boldsymbol{\beta}}$ with design matrix $\boldsymbol{Z}$ is identical to the residual sum of squares when using $\hat{\boldsymbol{\beta}}_*$ with design matrix $\boldsymbol{Z}'$. Hence the minimum residual sum of squares available with design matrix $\boldsymbol{Z}$ must be no larger than the residual sum of squares using $\hat{\boldsymbol{\beta}}$ with design matrix $\boldsymbol{Z}'$.

29. In Example 11.5.5, we are told that $\sigma' = 352.9$, so the residual sum of squares is 2864383. We can calculate $\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ directly from the data in Table 11.13. It equals, 26844478. It follows that

$$R^2 = 1 - \frac{2864383}{26844478} = 0.893.$$

30. Use the notation in the solution of Exercise 26. Suppose that $\boldsymbol{c}'\boldsymbol{\beta} = d$. Then U has a noncentral t distribution with $n-p$ degrees of freedom and noncentrality parameter $(d-c)/[\sigma(\boldsymbol{c}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{c})^{1/2}]$. The argument is the same as any of those in the book that derived noncentral t distributions.

11.6 Analysis of Variance

Commmentary

If one is using the software *R*, there is a more direct way to fit an analysis of variance model than to construct the design matrix. Let `y` contain the observed responses arranged as in Eq. (11.6.1). Suppose also that `x` is a vector of the same dimension as `y`, each of whose values is the first subscript i of Y_{ij} in Eq. (11.6.1) so that the value of `x` identifies which of the p sample each observation comes from. Specifically, the first n_1 elements of `x` would be 1, the next n_2 would be 2, etc. Then `aovfit=lm(y~factor(x))` will fit the analysis of variance model. The function `factor` converts the vector of integers into category identifiers. Then `anova(aovfit)` will print the ANOVA table.

Solutions to Exercises

1. By analogy with Eq. (11.6.2),

$$\boldsymbol{Z}'\boldsymbol{Z} = \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & n_p \end{bmatrix} \text{ and } (\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 1/n_1 & 0 & \cdots & 0 \\ 0 & 1/n_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1/n_p \end{bmatrix}$$

Also,

$$\boldsymbol{Z}'\boldsymbol{Y} = \begin{bmatrix} \left(\sum_{j=1}^{n_1} Y_{1j}\right) \\ \vdots \\ \left(\sum_{j=1}^{n_p} Y_{pj}\right) \end{bmatrix} \text{ and } (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \begin{bmatrix} \bar{Y}_{1+} \\ \vdots \\ \bar{Y}_{p+} \end{bmatrix}.$$

2. Let $\boldsymbol{A}$ be the orthogonal matrix whose first row is $\boldsymbol{u}$ in the statement of the problem. Define

$$\boldsymbol{Y} = \begin{bmatrix} n_1^{1/2}\overline{Y}_{1+} \\ \cdots \\ n_p^{1/2}\overline{Y}_{p+} \end{bmatrix}.$$

Let $\boldsymbol{v}'$ be a vector that is orthogonal to $\boldsymbol{u}$ (like all the other rows of $\boldsymbol{A}$.) Then $\boldsymbol{v}'\boldsymbol{X} = \boldsymbol{v}'\boldsymbol{Y}/\sigma$. Define

$$\begin{aligned} \boldsymbol{U} &= \boldsymbol{A}\boldsymbol{X} = (U_1, \ldots, U_p)', \\ \boldsymbol{V} &= \boldsymbol{A}\boldsymbol{Y} = (V_1, \ldots, V_p)'. \end{aligned}$$

We just showed that $V_i/\sigma = U_i$ for $i = 2, \ldots, n$. Now,

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{X} &= (\boldsymbol{A}\boldsymbol{X})'(\boldsymbol{A}\boldsymbol{X}) = \sum_{i=1}^{p} U_i^2, \\ \boldsymbol{Y}'\boldsymbol{Y} &= (\boldsymbol{A}\boldsymbol{Y})'(\boldsymbol{A}\boldsymbol{Y}) = \sum_{i=1}^{p} V_i^2. \end{aligned}$$

Notice that $V_1 = \sum_{i=1}^{p} n_i \overline{Y}_{i+}/n^{1/2} = n^{1/2}\overline{Y}_{++}$, hence

$$\frac{S^2_{\text{Betw}}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{p} n_i \overline{Y}^2_{i+} - n\overline{Y}^2_{++} = \frac{1}{\sigma^2}\left(\boldsymbol{Y}'\boldsymbol{Y} - V_1^2\right) = \sum_{i=2}^{p}\left(\frac{V_i}{\sigma}\right)^2 = \sum_{i=2}^{p} U_i^2.$$

Now, the coordinates of $\boldsymbol{X}$ are i.i.d. standard normal random variables, so the coordinates of $\boldsymbol{U}$ are also i.i.d. standard normal random variables. Hence $\sum_{i=2}^{p} U_i^2 = S^2_{\text{Betw}}/\sigma^2$ has a χ^2 distribution with $p-1$ degrees of freedom.

3. Use the definitions of $\bar{Y}_{i+}$ and $\bar{Y}_{++}$ in the text to compute

$$\begin{aligned}
\sum_{i=1}^{p} n_i(\bar{Y}_{i+} - \bar{Y}_{++})^2 &= \sum_{i=1}^{p} n_i\bar{Y}^2_{i+} - n\bar{Y}^2_{++} - 2\bar{Y}_{++}\sum_{i=1}^{p} n_i\bar{Y}_{i+} \\
&= \sum_{i=1}^{p} n_i\bar{Y}^2_{i+} + n\bar{Y}^2_{++} - 2n\bar{Y}^2_{++}, \\
&= \sum_{i=1}^{p} n_i\bar{Y}^2_{i+} - n\bar{Y}^2_{++}.
\end{aligned}$$

4. (a) It is found that $\bar{Y}_{1+} = 6.6, \bar{Y}_{2+} = 9.0$, and $\bar{Y}_{3+} = 10.2$. Also,

$$\sum_{j=1}^{n_1}(Y_{1j} - \bar{Y}_{1+})^2 = 1.90, \sum_{j=1}^{n_2}(Y_{2j} - \bar{Y}_{2+})^2 = 17.8, \sum_{j=1}^{n_3}(Y_{3j} - \bar{Y}_{3+})^2 = 5.54.$$

Hence, by Eq. (11.6.4), $\hat{\sigma}^2 = (1.90 + 17.8 + 5.54)/13 = 1.942$.

(b) It is found that $\bar{Y}_{++} = 8.538$. Hence, by Eq. (11.6.9), $U^2 = 10(24.591)/[2(25.24)] = 4.871$. When H_0 is true, the statistic U^2 has the F distribution with 2 and 10 degrees of freedom. Therefore, the tail area corresponding to the value $U^2 = 4.871$ is between 0.025 and 0.05.

5. In this problem $n_i = 10$ for $i = 1, 2, 3, 4$, and $\bar{Y}_{1+} = 105.7, \bar{Y}_{2+} = 102.0, \bar{Y}_{3+} = 93.5, \bar{Y}_{4+} = 110.8$, and $\bar{Y}_{++} = 103$.

$$\begin{aligned}
\sum_{j=1}^{10}(Y_{1j} - \bar{Y}_{1+})^2 &= 303, \quad \sum_{j=1}^{10}(Y_{2j} - \bar{Y}_{2+})^2 = 544, \\
\sum_{j=1}^{10}(Y_{3j} - \bar{Y}_{3+})^2 &= 250, \quad \sum_{j=1}^{10}(Y_{4j} - \bar{Y}_{4+})^2 = 364.
\end{aligned}$$

Therefore, by Eq. (11.6.9),

$$U^2 = \frac{36(1593.8)}{3(1461)} = 13.09.$$

When H_0 is true, the statistic U^2 has the F distribution with 3 and 36 degrees of freedom. The tail area corresponding to the value $U^2 = 13.09$ is found to be less than 0.025.

6. The random variables $Q_1, \ldots, Q_p$ are independent, because each Q_i is a function of a different group of the observations in the sample. It is known that Q_i/σ^2 has a χ^2 distribution with $n_i - 1$ degrees of freedom. Therefore, $(Q_1 + \cdots + Q_p)/\sigma^2$ will have a χ^2 distribution with $\sum_{i=1}^{p}(n_i - 1) = n - p$ degrees of freedom. Since Q_1 and Q_p are independent, $\dfrac{Q_i/[\sigma^2(n_1-1)]}{Q_p/[\sigma^2(n_p-1)]}$ will have the F distribution with $n_1 - 1$ and $n_p - 1$ degrees of freedom. In other words, $(n_p - 1)Q_1/[(n_1 - 1)Q_p]$ will have that F distribution.

7. If U is defined by Eq. (9.6.3), then

$$U^2 = \frac{(m+n-2)(\bar{X}_m - \bar{Y}_n)^2}{\left(\frac{1}{m} + \frac{1}{n}\right)(S_x^2 + S_y^2)}.$$

The correspondence between the notation of Sec. 9.6 and our present notation is as follows:

Notation of Sec. 9.6	Present notation
m	n_1
n	n_2
$\bar{X}_m$	$\bar{Y}_{1+}$
$\bar{Y}_n$	$\bar{Y}_{2+}$
S_X^2	$\sum_{j=1}^{n_1}(Y_{1j} - \bar{Y}_{1+})^2$
S_Y^2	$\sum_{j=1}^{n_2}(Y_{2j} - \bar{Y}_{2+})^2$

Since $p = 2$, $\bar{Y}_{++} = n_1\bar{Y}_{1+}/n + n_2\bar{Y}_{2+}/n$. Therefore,

$$\begin{aligned} & n_1(\bar{Y}_{1+} - \bar{Y}_{++})^2 + n_2(\bar{Y}_{2+} - \bar{Y}_{++})^2 \\ &= n_1\left(\frac{n_2}{n}\right)^2(\bar{Y}_{1+} - \bar{Y}_{2+})^2 + n_2\left(\frac{n_1}{n}\right)^2(\bar{Y}_{1+} - \bar{Y}_{2+})^2 \\ &= \frac{n_1 n_2}{n}(\bar{Y}_{1+} - \bar{Y}_{2+})^2, \quad \text{since } n_1 + n_2 = n. \end{aligned}$$

Also, since $m + n$ in the notation of Sec. 9.6 is simply n in our present notation, we can now rewrite the expression for U^2 as follows:

$$U^2 = \frac{(n-2)\dfrac{n}{n_1 n_2}\displaystyle\sum_{i=1}^{2} n_i(\bar{Y}_{i+} - \bar{Y}_{++})^2}{\dfrac{n}{n_1 n_2}\displaystyle\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\bar{Y}_{ij} - \bar{Y}_{i+})^2}.$$

This expression reduces to the expression for U^2 given in Eq. (11.6.9), with $p = 2$.

8. $E\left[\dfrac{1}{n-p}\displaystyle\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\bar{Y}_{ij} - \bar{Y}_{i+})^2\right] = \dfrac{1}{n-p}\displaystyle\sum_{i=1}^{p} E\left[\sum_{j=1}^{n_i}(\bar{Y}_{ij} - \bar{Y}_{i+})^2\right]$

$$= \frac{1}{n-p}\sum_{i=1}^{p}(n_i - 1)\sigma^2, \text{ by the results of Sec. 8.7,}$$

$$= \frac{1}{n-p}(n-p)\sigma^2 = \sigma^2.$$

9. Each of the three given random variables is a linear function of the independent random variables $Y_{rs}(s = 1, \dots, n_r$ and $r = 1, \dots, p)$.

Let $W_1 = \sum_{r,s} a_{rs}Y_{rs}$, $W_2 = \sum_{r,s} b_{rs}Y_{rs}$, and $W_3 = \sum_{r,s} c_{rs}Y_{rs}$. We have $a_{ij} = 1 - \frac{1}{n_i}$, $a_{is} = -\frac{1}{n_i}$ for $s \neq j$, and $a_{rs} = 0$ for $r \neq i$. Also, $b_{i's} = \frac{1}{n_{i'}} - \frac{1}{n}$ and $b_{rs} = -\frac{1}{n}$ for $r \neq i'$. Finally, $c_{rs} = \frac{1}{n}$ for all values of r and s.

Now, $\text{Cov}(W_1, W_2) = \text{Cov}\left(\sum_{r,s} a_{rs}Y_{rs}, \sum_{r',s'} b_{r's'}Y_{r's'}\right) = \sum_{r,s}\sum_{r',s'} a_{rs}b_{r's'}\,\text{Cov}(Y_{rs}, Y_{r's'})$.

But $\text{Cov}(Y_{rs}, Y_{r's'}) = 0$ unless $r = r'$ and $s = s'$, since any two distinct Y's are independent. Also, $\text{Cov}(Y_{rs}, Y_{rs}) = \text{Var}(Y_{rs}) = \sigma^2$.

Therefore, $\text{Cov}(W_1, W_2) = \sigma^2 \sum_{r,s} a_{rs}b_{rs}$. If $i = i'$,

$$\sum_{r,s} a_{rs}b_{rs} = a_{ij}b_{ij} + \sum_{s\neq j} a_{is}b_{is} + 0 = \left(1 - \frac{1}{n_i}\right)\left(\frac{1}{n_i} - \frac{1}{n}\right) + (n_i - 1)\left(-\frac{1}{n_i}\right)\left(\frac{1}{n_i} - \frac{1}{n}\right) = 0.$$

If $i \neq i'$,

$$\sum_{r,s} a_{rs}b_{rs} = \left(1 - \frac{1}{n_i}\right)\left(-\frac{1}{n}\right) + (n_i - 1)\left(-\frac{1}{n_i}\right)\left(-\frac{1}{n}\right) = 0.$$

Similarly,

$$\text{Cov}(W_1, W_3) = \sigma^2 \sum_{r,s} a_{rs}c_{rs} = \sigma^2 \frac{1}{n}\left(a_{ij} + \sum_{s\neq j} a_{ij}\right) = 0$$

Finally,

$$\text{Cov}(W_2, W_3) = \sigma^2 \sum_{r,s} b_{rs}c_{rs} = \sigma^2 \frac{1}{n}\sum_{r,s} b_{rs} = \sigma^2 \frac{1}{n}\left[n_{i'}\left(\frac{1}{n_{i'}} - \frac{1}{n}\right) + (n - n_{i'})\left(-\frac{1}{n}\right)\right] = 0.$$

10. (a) The three group averages are 825.8, 845.0, and 775.0. The residual sum of squares is 1671. The ANOVA table is then

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between samples	2	15703	7851
Residuals	15	1671	111.4
Total	17	17374	

(b) The F statistic is $7851/111.4 = 70.48$. Comparing this to the F distribution with 2 and 15 degrees of freedom, we get a p-value of essentially 0.

11. Write $S^2_{\text{Betw}} = \sum_{i=1}^{p} \overline{Y}^2_{i+} - n\overline{Y}^2_{++}$. Recall, that the mean of the square of a normal random variable with mean μ and variance σ^2 is $\mu^2 + \sigma^2$. The distribution of $\overline{Y}_{i+}$ is the normal distribution with mean μ_i and variance σ^2/n_i, while the distribution of $\overline{Y}_{++}$ is the normal distribution with mean $\overline{\mu}$ and variance σ^2/n. Hence the mean of S^2_{Betw} is

$$\sum_{i=1}^{p} n_i(\mu_i^2 + \sigma^2/n_i) - n(\overline{\mu} + \sigma^2/n).$$

If we collect terms in this sum, we get $(p-1)\sigma^2 + \sum_{i=1}^{p} n_i\mu_i^2 - n\overline{\mu}^2$. This simplifies to the expression stated in the exercise.

12. If the null hypothesis is true, the likelihood is

$$(2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \mu)^2\right).$$

This is maximized by choosing $\mu = \overline{y}_{++}$ and

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{p}(y_{ij} - \overline{y}_{++})^2 = \frac{1}{n}S^2_{\text{Tot}}.$$

The resulting maximum value of the likelihood is

$$(2\pi)^{-n/2}\frac{n^{n/2}}{(S^2_{\text{Tot}})^{n/2}} \exp\left(-\frac{n}{2}\right). \qquad \text{(S.11.4)}$$

If the null hypothesis is false, the likelihood is

$$(2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2\right).$$

This is maximized by choosing $\mu_i = \overline{y}_{i+}$ and

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{p}(y_{ij} - \overline{y}_{i+})^2 = \frac{1}{n}S^2_{\text{Resid}}.$$

The resulting maximum value of the likelihood is

$$(2\pi)^{-n/2}\frac{n^{n/2}}{(S^2_{\text{Resid}})^{n/2}} \exp\left(-\frac{n}{2}\right). \qquad \text{(S.11.5)}$$

The ratio of (S.11.5) to (S.11.4) is

$$\left(\frac{S^2_{\text{Tot}}}{S^2_{\text{Resid}}}\right)^{n/2} = \left(1 + \frac{S^2_{\text{Betw}}}{S^2_{\text{Resid}}}\right)^{n/2}.$$

Rejecting H_0 when this ratio is greater than k is equivalent to rejecting when $U^2 > k'$ for some other constant k'. In order for the test to have level α_0, k' must equal the $1 - \alpha_0$ quantile of the distribution of U^2 when H_0 is true. We saw in the text that this distribution is the F distribution with $p - 1$ and $n - p$ degrees of freedom.

13. We know that $S^2_{\text{Tot}} = S^2_{\text{Betw}} + S^2_{\text{Resid}}$. We also know that $S^2_{\text{Betw}}/\sigma^2$ and $S^2_{\text{Resid}}/\sigma^2$ are independent. If H_0 is true then they both have χ^2 distributions, one with $p-1$ degrees of freedom (see Exercise 2) and the other with $n-p$ degrees of freedom as in the text. The sum of two independent χ^2 random variables has χ^2 distribution with the sum of the degrees of freedom, in this case $n-1$.

14. (a) (The exercise should have asked you to prove that $\sum_{i=1}^{p} n_i\alpha_i = 0$.) We see that $\sum_{i=1}^{p} n_i\alpha_i = \sum_{i=1}^{n} n_i\mu_i - n\mu = 0$.
    (b) The M.L.E. of μ_i is $\overline{Y}_{i+}$ and the M.L.E. of μ is $\overline{Y}_{++}$, so the M.L.E. of α_i is $\overline{Y}_{i+} - \overline{Y}_{++}$.
    (c) Notice that all of μ_i equal each other if and only if they all equal μ, if and only if all $\alpha_i = 0$.
    (d) This fact was proven in Exercise 11 with slightly different notation.

11.7 The Two-Way Layout

Commmentary

If one is using the software R, one can fit a two-way layout using `lm` with two `factor` variables. As in the Commentary to Sec. 11.6 above, let `y` contain the observed responses, and let `x1` and `x2` be two `factor` variables giving the levels of the two factors in the layout. Then `aovfit=lm(y~x1+x2)` will fit the model, and `anova(aovfit)` will print the ANOVA table.

Solutions to Exercises

1. Write $S^2_A = J\sum_{i=1}^{I}\overline{Y}^2_{i+} - IJ\overline{Y}^2_{++}$. Recall, that the mean of the square of a normal random variable with mean μ and variance σ^2 is $\mu^2+\sigma^2$. The distribution of $\overline{Y}_{i+}$ is the normal distribution with mean μ_i and variance σ^2/J, while the distribution of $\overline{Y}_{++}$ is the normal distribution with mean μ and variance σ^2/IJ. Hence the mean of S^2_A is $J\sum_{i=1}^{I}(\mu_i^2+\sigma^2/J) - IJ(\mu+\sigma^2/[IJ])$. If we collect terms in this sum, we get

$$(I-1)\sigma^2 + J\sum_{i=1}^{I}\mu_i^2 - IJ\mu^2 = (I-1)\sigma^2 + J\sum_{i=1}^{I}(\mu_i-\mu)^2 = (I-1)\sigma^2 + J\sum_{i=1}^{I}\alpha_i^2.$$

2. In each part of this exercise, let μ_{ij} denote the element in row i and column j of the given matrix.
    (a) The effects are not additive because $\mu_{21}-\mu_{11} = 5 \neq \mu_{22}-\mu_{12} = 7$.
    (b) The effects are additive because each element in the second row is 1 unit larger than the corresponding element in the first row. Alternatively, we could say that the effects are additive because each element in the second column is 3 units larger than the corresponding element in the first column.
    (c) The effects are additive because each element in the first row is 5 units smaller than the corresponding element in the second row and is 1 unit smaller than the corresponding element in the third row.
    (d) The effects are not additive because, for example, $\mu_{21}-\mu_{11} = 1 \neq \mu_{22}-\mu_{12} = 2$.

3. If the effects are additive, then there exist numbers Θ_i and Ψ_j such that Eq. (11.7.1) is satisfied for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Let $\bar{\Theta} = \frac{1}{I}\sum_{i=1}^{I}\Theta_i$ and $\bar{\Psi} = \frac{1}{J}\sum_{j=1}^{J}\Psi_j$, and define.

$$\begin{aligned}\mu &= \bar{\Theta} + \bar{\Psi},\\ \alpha_i &= \Theta_i - \bar{\Theta} \quad \text{for} \quad i = 1, \ldots, I,\\ \beta_j &= \Psi_j - \bar{\Psi} \quad \text{for} \quad j = 1, \ldots, J.\end{aligned}$$

Then it follows that Eqs. (11.7.2) and (11.7.3) will be satisfied.

It remains to be shown that no other set of values of μ, α_i, and β_j will satisfy Eqs. (11.7.2) and (11.7.3). Suppose that μ', α_i', and β_j' are another such set of values. Then $\mu + \alpha_i + \beta_j = \mu' + \alpha_i' + \beta_j'$ for all i and j. By summing both sides of this relation over i and j, we obtain the relation $IJ\mu = IJ\mu'$. Hence, $\mu = \mu'$. It follows, therefore, that $\alpha_i + \beta_j = \alpha_i' + \beta_j'$ for all i and j. By summing both sides of this relation over j, we obtain the result $\alpha_i = \alpha_i'$ for every value of i. Similarly, by summing both sides over i, we obtain the relation $\beta_j = \beta_j'$ for every value of j.

4. If the effects are additive, so that Eq. (11.7.1) is satisfied, and we denote the elements of the matrix by $\mu_{ij} = \Theta_i + \Psi_j$, then $\bar{\mu}_{++} = \bar{\Theta} + \bar{\Psi}, \bar{\mu}_{i+} = \Theta_i + \bar{\Psi}$, and $\bar{\mu}_{+j} = \bar{\Theta} + \Psi_j$. Therefore, $\mu_{ij} = \bar{\mu}_{++} + (\bar{\mu}_{i+} - \bar{\mu}_{++}) + (\bar{\mu}_{+j} - \bar{\mu}_{++})$. Hence, it can be verified that $\mu = \bar{\mu}_{++}, \alpha_i = \bar{\mu}_{i+} - \bar{\mu}_{++}$ and $\beta_j = \bar{\mu}_{+j} - \bar{\mu}_{++}$. In this exercise,

$$\begin{aligned}\bar{\mu}_{++} &= \frac{1}{4}(3+6+4+7) = 5,\\ \bar{\mu}_{1+} &= \frac{1}{2}(3+6) = 4.5, \bar{\mu}_{2+} = \frac{1}{2}(4+7) = 5.5,\\ \bar{\mu}_{+1} &= \frac{1}{2}(3+4) = 3.5, \bar{\mu}_{+2} = \frac{1}{2}(6+7) = 6.5.\end{aligned}$$

5. In this exercise,

$$\begin{aligned}\bar{\mu}_{++} &= \frac{1}{12}\sum_{i=1}^{3}\sum_{j=1}^{4}\mu_{ij} = \frac{1}{12}(39) = 3.25,\\ \bar{\mu}_{1+} &= \frac{5}{4} = 1.25, \bar{\mu}_{2+} = \frac{25}{4} = 6.35, \bar{\mu}_{3+} = \frac{9}{4} = 2.25,\\ \bar{\mu}_{+1} &= \frac{15}{3} = 5, \bar{\mu}_{+2} = \frac{3}{3} = 1, \bar{\mu}_{+3} = \frac{6}{3} = 2, \bar{\mu}_{+4} = \frac{15}{3} = 5.\end{aligned}$$

It follows that $\alpha_1 = 1.25 - 3.25 = -2$, $\alpha_2 = 6.25 - 3.25 = 3$, and $\alpha_3 = 2.25 - 3.25 = -1$. Also, $\beta_1 = 5 - 3.25 = 1.75$, $\beta_2 = 1 - 3.25 = -2.25$, $\beta_3 = 2 - 3.25 = -1.25$, and $\beta_4 = 5 - 3.25 = 1.75$.

6. $\sum_{i=1}^{I}\hat{\alpha}_i = \sum_{i=1}^{I}\bar{Y}_{i+} - I\bar{Y}_{++} = \frac{1}{J}\sum_{i=1}^{I}\sum_{j=1}^{J}Y_{ij} - \frac{1}{J}\sum_{i=1}^{I}\sum_{j=1}^{J}Y_{ij} = 0$. A similar argument show that $\sum_{j=1}^{J}\hat{\beta}_j = 0$. Also, since $E(Y_{ij}) = \mu + \alpha_i + \beta_j$,

$$\begin{aligned}E(\hat{\mu}) &= E(\bar{Y}_{++}) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}E(Y_{ij}) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(\mu + \alpha_i + \beta_j) = \frac{1}{IJ}(IJ\mu + 0 + 0) = \mu,\\ E(\hat{\alpha}_i) &= E(\bar{Y}_{i+} - \bar{Y}_{++}) = E\left(\frac{1}{J}\sum_{j=1}^{J}Y_{ij}\right) - \mu = \frac{1}{J}\sum_{j=1}^{J}(\mu + \alpha_i + \beta_j) - \mu = \alpha_i.\end{aligned}$$

A similar argument shows that $E(\hat{\beta}_j) = \beta_j$

7. $\text{Var}(\hat{\mu}) = \text{Var}\left(\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}\right) = \frac{1}{I^2J^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\text{Var}(Y_{ij}) = \frac{1}{I^2J^2}IJ\sigma^2 = \frac{1}{IJ}\sigma^2.$

   The estimator $\hat{\alpha}_i$ is a linear function of the IJ independent random variables Y_{rs} ($r = 1, \dots, I$ and $s = 1, \dots, J$). If we let $\hat{\alpha}_i = \sum_{r=1}^{I}\sum_{s=1}^{J} a_{rs}Y_{rs}$, then it can be found that $a_{is} = \frac{1}{J} - \frac{1}{IJ}$ for $s = 1, \dots, J$ and $a_{rs} = -\frac{1}{IJ}$ for $r \neq i$. Therefore,

$$\text{Var}(\hat{\alpha}_i) = \sum_{r=1}^{I}\sum_{s=1}^{J} a_{rs}^2\sigma^2 = \sigma^2\left[J\left(\frac{1}{J} - \frac{1}{IJ}\right)^2 + (I-1)J\left(\frac{1}{IJ}\right)^2\right] = \frac{I-1}{IJ}\sigma^2.$$

   The value of $\text{Var}(\hat{\beta}_j)$ can be found similarly.

8. If the square on the right of Eq. (11.7.9) is expanded and all the summations are then performed, we will obtain the right side of Eq. (11.7.8) provided that the summation of each of the cross product terms is 0. We shall now establish this result. Note that

$$\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++}) = J\bar{Y}_{i+} - J\bar{Y}_{i+} - J\bar{Y}_{++} + J\bar{Y}_{++} = 0.$$

   Similarly, $\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++}) = 0$. Therefore,

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++})(\bar{Y}_{i+} - \bar{Y}_{++}) = \sum_{i=1}^{I}(\bar{Y}_{i+} - \bar{Y}_{i+})\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++}) = 0,$$
$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++})(\bar{Y}_{+j} - \bar{Y}_{++}) = \sum_{j=1}^{J}(\bar{Y}_{+j} - \bar{Y}_{+j})\sum_{i=1}^{I}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++}) = 0.$$

   Finally,

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{Y}_{i+} - \bar{Y}_{++})(\bar{Y}_{+j} - \bar{Y}_{++}) = \sum_{i=1}^{I}(\bar{Y}_{i+} - \bar{Y}_{++})\sum_{j=1}^{J}(\bar{Y}_{+j} - \bar{Y}_{++}) = 0 \times 0.$$

9. Each of the four given random variables is a linear function of the independent random variables $Y_{rs}(r = 1, \dots, I$ and $s = 1, \dots, J)$. Let

$$W_1 = \sum_{r=1}^{I}\sum_{s=1}^{J} a_{rs}Y_{rs},$$
$$W_2 = \sum_{r=1}^{I}\sum_{s=1}^{J} b_{rs}Y_{rs},$$
$$W_3 = \sum_{r=1}^{I}\sum_{s=1}^{J} c_{rs}Y_{rs}.$$

Then

$$\begin{aligned} a_{ij} &= 1 - \frac{1}{J} - \frac{1}{I} + \frac{1}{IJ}, a_{rj} = -\frac{1}{I} + \frac{1}{IJ} \quad \text{for} \quad r \neq i, \\ a_{is} &= -\frac{1}{J} + \frac{1}{IJ} \quad \text{for} \quad s \neq j, \quad \text{and} \quad a_{rs} = \frac{1}{IJ} \quad \text{for} \quad r \neq i, \quad \text{and} \quad s \neq j. \end{aligned}$$

Also,

$$\begin{aligned} b_{i's} &= \frac{1}{J} - \frac{1}{IJ} \quad \text{and} \quad b_{rs} = -\frac{1}{IJ} \quad \text{for} \quad r \neq i', \\ c_{rj'} &= \frac{1}{I} - \frac{1}{IJ} \quad \text{and} \quad c_{rs} = -\frac{1}{IJ} \quad \text{for} \quad s \neq j'. \end{aligned}$$

As in Exercise 12 of Sec. 11.5, $\text{Cov}(W_1, W_2) = \sigma^2 \sum_{r,s} a_{rs} b_{rs}$. If $i = i'$,

$$\begin{aligned} \sum_{r,s} a_{rs} b_{rs} &= \left(1 - \frac{1}{J} - \frac{1}{I} + \frac{1}{IJ}\right)\left(\frac{1}{J} - \frac{1}{IJ}\right) + (J-1)\left(-\frac{1}{J} + \frac{1}{IJ}\right)\left(\frac{1}{J} - \frac{1}{IJ}\right) \\ &\quad + (I-1)\left(-\frac{1}{I} + \frac{1}{IJ}\right)\left(-\frac{1}{IJ}\right) + (I-1)(J-1)\left(\frac{1}{IJ}\right)\left(-\frac{1}{IJ}\right) = 0. \end{aligned}$$

If $i \neq i'$,

$$\begin{aligned} \sum_{r,s} a_{rs} b_{rs} &= \left(1 - \frac{1}{J} - \frac{1}{I} + \frac{1}{IJ}\right)\left(-\frac{1}{IJ}\right) && (i, j \text{ term}) \\ &\quad + (J-1)\left(-\frac{1}{J} + \frac{1}{IJ}\right)\left(-\frac{1}{IJ}\right) && (i, s \text{ terms for } s \neq j) \\ &\quad + \left(-\frac{1}{I} + \frac{1}{IJ}\right)\left(\frac{1}{J} - \frac{1}{IJ}\right) && (i', j \text{ term}) \\ &\quad + (J-1)\left(\frac{1}{IJ}\right)\left(\frac{1}{J} - \frac{1}{IJ}\right) && (i', j \text{ terms for } s \neq j) \\ &\quad + (I-2)\left(-\frac{1}{I} + \frac{1}{IJ}\right)\left(-\frac{1}{IJ}\right) && (r, j \text{ terms for } r \neq i, i') \\ &\quad + (I-2)(J-1)\left(\frac{1}{IJ}\right)\left(-\frac{1}{IJ}\right) && (r, s \text{ terms for } r \neq i, i' \text{ and } s \neq j). \\ &= 0. \end{aligned}$$

Similarly, the covariance between any other pair of the four variances can be shown to be 0.

10. $\sum_{i=1}^{I}(\bar{Y}_{i+} - \bar{Y}_{++})^2 = \sum_{i=1}^{I}\bar{Y}_{i+}^2 - 2\bar{Y}_{++}\sum_{i=1}^{I}\bar{Y}_{i+} - I\bar{Y}_{++}^2 = \sum_{i=1}^{I}\bar{Y}_{i+}^2 - 2I\bar{Y}_{++}^2 + I\bar{Y}_{++}^2 = \sum_{i=1}^{I}\bar{Y}_{i+}^2 - I\bar{Y}_{++}^2.$
The other part of this exercise is proved similarly.

11. $$\begin{aligned} \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++})^2 &= \sum_i\sum_j Y_{ij}^2 + J\sum_i \bar{Y}_{i+}^2 + I\sum_j \bar{Y}_{+j}^2 + IJ\bar{Y}_{++}^2 \\ &\quad -2\sum_i\sum_j Y_{ij}\bar{Y}_{i+} - 2\sum_j\sum_i Y_{ij}\bar{Y}_{+j} + 2\bar{Y}_{++}\sum_i\sum_j Y_{ij} \\ &\quad +2\sum_i\sum_j \bar{Y}_{i+}\bar{Y}_{+j} - 2J\bar{Y}_{++}\sum_i \bar{Y}_{i+} - 2I\bar{Y}_{++}\sum_j \bar{Y}_{+j} \end{aligned}$$

$$
\begin{aligned}
&= \sum_i \sum_j Y_{ij}^2 + J \sum_i \bar{Y}_{i+}^2 + I \sum_i \bar{Y}_{+j}^2 + IJ\bar{Y}_{++}^2 \\
&\quad -2J \sum_i \bar{Y}_{i+}^2 - 2I \sum_j \bar{Y}_{+j}^2 + 2IJ\bar{Y}_{++}^2 \\
&\quad +2\,IJ\bar{Y}_{++}^2 - 2\,IJ\bar{Y}_{++}^2 - 2\,IJ\bar{Y}_{++}^2 \\
&= \sum_i \sum_j Y_{ij}^2 - J \sum_i \bar{Y}_{i+}^2 - I \sum_j \bar{Y}_{+j}^2 + IJ\bar{Y}_{++}^2.
\end{aligned}
$$

12. It is found that $\bar{Y}_{1+} = 17.4, \bar{Y}_{2+} = 15.94, \bar{Y}_{3+} = 17.08, \bar{Y}_{+1} = 15.1, \bar{Y}_{+2} = 14.6, \bar{Y}_{+3} = 15.5333, \bar{Y}_{+4} = 19.5667, \bar{Y}_{+5} = 19.2333, \bar{Y}_{++} = 16.8097$. The values of $\hat{\mu}, \hat{\alpha}_i$ for $i = 1, 2, 3$, and $\hat{\beta}_j$ for $j = 1, \ldots, 5$, can now be obtained from Eq. (11.7.6).

13. The estimate of $E(Y_{ij})$ is $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$. From the values given in the solution of Exercise 12, we therefore obtain the following table of estimated expectations:

	1	2	3	4	5
1	15.6933	15.1933	16.1267	20.16	19.8267
2	14.2333	13.7333	14.6667	18.7	18.3667
3	15.3733	14.8733	15.8067	19.84	19.5067

Furthermore, Theorem 11.7.1 says that the M.L.E. of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{15}(29.470667) = 1.9647.$$

14. It is found from Eq. (11.7.12) that

$$U_A^2 = \frac{20(1.177865)}{(29.470667)} = 0.799.$$

When the null hypothesis is true, U_A^2 will have the F distribution with $I - 1 = 2$ and $(I - 1)(J - 1) = 8$ degrees of freedom. The tail area corresponding to the value just calculated is found to be greater than 0.05.

15. It is found from Eq. (11.7.13) that

$$U_B^2 = \frac{6(22.909769)}{(29.470667)} = 4.664.$$

When the null hypothesis is true, U_B^2 will have the F distribution with $J - 1 = 4$ and $(I - 1)(J - 1) = 8$ degrees of freedom. The tail area corresponding to the value just calculated is found to be between 0.025 and 0.05.

16. If the null hypothesis in (11.7.15) is true, then all Y_{ij} have the same mean μ. The random variables S_A^2/σ^2, S_B^2/σ^2, and S_{Resid}^2/σ^2 are independent, and their distributions are χ^2 with $I - 1$, $J - 1$, and $(I - 1)(J - 1)$ degrees of freedom respectively. Hence $S_A^2 + S_B^2$ has the χ^2 distribution with $I + J - 2$ degrees of freedom and is independent of σ'. The conclusion now follows directly from the definition of the F distribution.

11.8 The Two-Way Layout with Replications

Commentary

This section is optional. There is reference to some of the material in this section in Chapter 12. In particular, Example 12.3.4 shows how to use simulation to compute the size of the two-stage test procedure that is described in Sec. 11.8.

If one is using the software R, and if one has `factor` variables `x1` and `x2` (as in the Commentary to Sec. 11.7) giving the levels of the two factors in a two-way layout with replication, then the following commands will fit the model and print the ANOVA table:

```
aovfit=lm(y~x1*x2)
anova(aovfit)
```

Solutions to Exercises

1. Let $\mu = \bar{\Theta}_{++}, \alpha_i = \bar{\Theta}_{i+} - \bar{\Theta}_{++}, \beta_j = \bar{\Theta}_{+j} - \bar{\Theta}_{++}$, and $\gamma_{ij} = \Theta_{ij} - \bar{\Theta}_{i+} - \bar{\Theta}_{+j} + \bar{\Theta}_{++}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Then it can be verified that Eqs. (11.8.4) and (11.8.5) are satisfied. It remains to be shown that no other set of values of μ, α_i, β_j, and γ_{ij} will satisfy Eqs. (11.8.4) and (11.8.5).

   Suppose that $\mu', \alpha_i', \beta_j'$, and γ_{ij}' are another such set of values. Then, for all values of i and j,

   $$\mu + \alpha_i + \beta_j + \gamma_{ij} = \mu' + \alpha_i' + \beta_j' + \gamma_{ij}'$$

   By summing both sides of this equation over i and j, we obtain the relation $IJ\mu = IJ\mu'$. Hence, $\mu = \mu'$. It follows, therefore, that for all values of i and j,

   $$\alpha_i + \beta_j + \gamma_{ij} = \alpha_i' + \beta_j + \gamma_{ij}'.$$

   By summing both sides of this equation over j, we obtain the relation $J\alpha_i = J\alpha_i'$. By summing both sides of this equation over i, we obtain the relation $I\beta_j = I\beta_j'$. Hence, $\alpha_i = \alpha_i'$ and $\beta_j = \beta_j'$. It also follows, therefore, that $\gamma_{ij} = \gamma_{ij}'$.

2. Since $\overline{Y}_{ij+}$ is the M.L.E. of θ_{ij} for each i and j, it follows from the definitions in Exercise 1 that the M.L.E.'s of the various linear functions defined in Exercise 1 are the corresponding linear functions of $\overline{Y}_{ij+}$. These are precisely the values in (11.8.6) and (11.8.7).

3. The values of μ, α_i, β_j, and γ_{ij} can be determined in each part of this problem from the given values of Θ_{ij} by applying the definitions given in the solution of Exercise 1.

4. $$\sum_{i=1}^{I} \hat{\alpha}_i = \sum_{i=1}^{I} (\bar{Y}_{i++} - \bar{Y}_{+++}) = I\bar{Y}_{+++} - I\bar{Y}_{+++} = 0,$$

   $$\sum_{i=1}^{I} \hat{\gamma}_{ij} = \sum_{i=1}^{I} (\bar{Y}_{ij+} - \bar{Y}_{i++} - \bar{Y}_{+j+} + \bar{Y}_{+++}) = I\bar{Y}_{+j+} - I\bar{Y}_{+++} - I\bar{Y}_{+j+} + I\bar{Y}_{+++} = 0.$$

   The proofs that $\sum_{j=1}^{J} \hat{\beta}_j = 0$ and $\sum_{j=1}^{J} \hat{\gamma}_{ij} = 0$ are similar.

5. $E(\hat{\mu}) = \frac{1}{IJK}\sum_{i,j,k} E(Y_{ijk}) = \frac{1}{IJK}\sum_{i,j,k}\Theta_{ij} = \frac{1}{IJ}\sum_{i,j}\Theta_{ij} = \bar{\Theta}_{++} = \mu$, by Exercise 1;

$$\begin{aligned}
E(\hat{\alpha}_i) &= \frac{1}{JK}\sum_{j,k} E(Y_{ijk}) - E(\bar{Y}_{+++}) \\
&= \frac{1}{JK}\sum_{j,k}\Theta_{ij} - \bar{\Theta}_{++}, \text{ by the first part of this exercise} \\
&= \frac{1}{J}\sum_{j}\Theta_{ij} - \bar{\Theta}_{++} = \bar{\Theta}_{i+} - \bar{\Theta}_{++} = \alpha_i, \text{ by Exercise 1;} \\
E(\hat{\beta}_j) &= \beta_j, \text{ by a similar argument;} \\
E(\hat{\gamma}_{ij}) &= \frac{1}{K}\sum_{k} E(Y_{ijk}) - E(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j), \text{ by Eq. (11.8.7)} \\
&= \Theta_{ij} - \bar{\Theta}_{i+} - \bar{\Theta}_{+J} + \bar{\Theta}_{++}, \text{ by the previous parts of this exercise,} \\
&= \gamma_{ij}, \text{ by Exercise 1.}
\end{aligned}$$

6. The IJK random variables Y_{ijk} are independent and each has variance σ^2. Hence,

$$\text{Var}(\hat{\mu}) = \text{Var}\left(\frac{1}{IJK}\sum_{i,j,k} Y_{ijk}\right) = \frac{1}{(IJK)^2}\sum_{i,j,k}\text{Var}(Y_{ijk}) = \frac{1}{(IJK)^2} IJK\sigma^2 = \frac{\sigma^2}{IJK}.$$

he estimator $\hat{\alpha}_i$ is a linear function of the observations Y_{ijk} of the form $\hat{\alpha}_i = \sum_{r,s,t} a_{rst}Y_{rst}$, where

$$a_{ist} = \frac{1}{JK} - \frac{1}{IJK} = \frac{I-1}{IJK}, \quad \text{and} \quad a_{rst} = -\frac{1}{IJK},$$

for $r \neq i$. Hence,

$$\text{Var}(\hat{\alpha}_i) = \sum_{r,s,t} a_{rst}^2\sigma^2 = \left[JK\left(\frac{I-1}{IJK}\right)^2 + (I-1)JK\left(-\frac{1}{IJK}\right)^2\right]\sigma^2 = \frac{I-1}{IJK}\sigma^2.$$

$\text{Var}(\hat{\beta}_j)$ can be determined similarly. Finally, if we represent $\hat{\gamma}_{ij}$ in the form $\hat{\gamma}_{ij} = \sum_{r,s,t} c_{rst}Y_{rst}$, then it follows from Eq. (11.8.7) that

$$\begin{aligned}
c_{ijt} &= \frac{1}{K} - \frac{1}{JK} - \frac{1}{IK} + \frac{1}{IJK} \text{ for } t = 1, \dots, K, \\
c_{rjt} &= -\frac{1}{IK} + \frac{1}{IJK} \text{ for } r \neq i \text{ and } t = 1, \dots, K, \\
c_{ist} &= -\frac{1}{JK} + \frac{1}{IJK} \text{ for } s \neq j \text{ and } t = 1, \dots, K, \\
c_{rst} &= \frac{1}{IJK} \text{ for } r \neq i,\ s \neq j, \text{ and } t = 1, \dots, K.
\end{aligned}$$

Therefore,

$$\text{Var}(\hat{\gamma}_{ij}) = \sum_{r,s,t} c_{rst}^2\sigma^2$$

$$
\begin{aligned}
&= \left[K\left(\frac{1}{K}-\frac{1}{JK}-\frac{1}{IK}+\frac{1}{IJK}\right)^2+(I-1)K\left(-\frac{1}{IK}+\frac{1}{IJK}\right)^2\right.\\
&\quad \left.+(J-1)K\left(-\frac{1}{JK}+\frac{1}{IJK}\right)^2+(I-1)(J-1)K\left(\frac{1}{IJK}\right)^2\right]\sigma^2\\
&= \frac{(I-1)(J-1)}{IJK}\sigma^2.
\end{aligned}
$$

7. First, write S^2_{Tot} as

$$\sum_{i,j,k}[(Y_{ijk}-\bar{Y}_{ij+})+(\bar{Y}_{ij+}-\bar{Y}_{i++}-\bar{Y}_{+j+}+\bar{Y}_{+++})+(\bar{Y}_{i++}-\bar{Y}_{+++})+(\bar{Y}_{+j+}-\bar{Y}_{+++})]^2. \quad \text{(S.11.6)}$$

The sums of squares of the grouped terms in the summation in (S.11.6) are S^2_{Resid}, S^2_{Int}, S^2_A, and S^2_B. Hence, to verify Eq. (11.8.9), it must be shown that the sum of each of the pairs of cross-products of the grouped terms is 0. This is verified in a manner similar to what was done in the solution to Exercise 8 in Sec. 11.7. Each of the sums of $(Y_{ijk}-\bar{Y}_{ij+})$ times one of the other terms is 0 by summing over k first. The sum of the product of the last two grouped terms is 0 because the sum factors into sums of i and j that are each 0. The other two sums are similar, and we shall illustrate this one:

$$\sum_{i,j,k}(\bar{Y}_{ij+}-\bar{Y}_{i++}-\bar{Y}_{+j+}+\bar{Y}_{+++})(\bar{Y}_{i++}-\bar{Y}_{+++}).$$

Summing over j first produces 0 in this sum. For the other one, sum over i first.

8. Each of the five given random variables is a linear function of the IJK independent observations Y_{rst}. Let

$$
\begin{aligned}
\hat{\alpha}_{i_1} &= \sum_{r,s,t} a_{rst}Y_{rst},\\
\hat{\beta}_{j_1} &= \sum_{r,s,t} b_{rst}Y_{rst},\\
\hat{\gamma}_{i_2j_2} &= \sum_{r,s,t} c_{rst}Y_{rst}\\
Y_{ijk}-\bar{Y}_{ij+} &= \sum_{r,s,t} d_{rst}Y_{rst}.
\end{aligned}
$$

Of course, $\hat{\mu}=\sum_{r,s,t} Y_{rst}/[IJK]$. The value of a_{rst}, b_{rst},and c_{rst} were given in the solution of Exercise 6, and

$$
\begin{aligned}
d_{ijk} &= 1-\frac{1}{K},\\
d_{ijt} &= -\frac{1}{K} \quad \text{for} \quad t\neq k,\\
d_{rst} &= 0 \quad \text{otherwise.}
\end{aligned}
$$

To show that $\hat{\alpha}_{i_1}$ and $\hat{\beta}_{j_1}$ are uncorrelated, for example, it must be shown, as in Exercise 9 of Sec. 11.7, that $\sum_{r,s,t} a_{rst}b_{rst}=0$. We have

$$\sum_{r,s,t} a_{rst}b_{rst} = K\left(\frac{I-1}{IJK}\right)\left(\frac{J-1}{IJK}\right)+(J-1)K\left(\frac{I-1}{IJK}\right)\left(-\frac{1}{IJK}\right)$$

$$\begin{aligned} & +(I-1)K\left(-\frac{1}{IJK}\right)\left(\frac{J-1}{IJK}\right)+(I-1)(J-1)K\left(-\frac{1}{IJK}\right)\left(-\frac{1}{IJK}\right) \\ = \; & 0. \end{aligned}$$

Similarly, to show that $Y_{ijk}-\bar{Y}_{ij+}$ and $\hat{\gamma}_{i_2j_2}$ are uncorrelated, it must be shown that $\sum_{r,s,t} c_{rst}d_{rst}=0$. Suppose first that $i=i_2$ and $j=j_2$. Then

$$\begin{aligned} \sum_{r,s,t} c_{rst}d_{rst} &= \left(\frac{1}{K}-\frac{1}{JK}-\frac{1}{IK}+\frac{1}{IJK}\right)\left(1-\frac{1}{K}\right) \\ &\quad +(K-1)\left(\frac{1}{K}-\frac{1}{JK}-\frac{1}{IK}+\frac{1}{IJK}\right)\left(-\frac{1}{K}\right)=0. \end{aligned}$$

Suppose next that $i=i_2$ and $j\neq j_2$. Then

$$\sum_{r,s,t} c_{rst}d_{rst}=\left(-\frac{1}{JK}+\frac{1}{IJK}\right)\left(1-\frac{1}{K}\right)+(K-1)\left(-\frac{1}{JK}+\frac{1}{IJK}\right)\left(-\frac{1}{K}\right)=0.$$

The other cases can be treated similarly, and it can be shown that the five given random variables are uncorrelated with one another regardless of whether any of the values j, j_1 and j_2 are equal.

9. We shall first show that the numerators are equal:

$$\begin{aligned} \sum_{i,j}(\bar{Y}_{ij+}-\bar{Y}_{i++}-\bar{Y}_{+j+}+\bar{Y}_{+++})^2 &= \sum_{i,j}(\bar{Y}^2_{ij+}+\bar{Y}^2_{i++}+\bar{Y}^2_{+j+}+\bar{Y}^2_{+++}-2\bar{Y}_{ij+}\bar{Y}_{i++} \\ &\quad -2\bar{Y}_{ij+}\bar{Y}_{+j+}+2\bar{Y}_{ij+}\bar{Y}_{+++}+2\bar{Y}_{i++}\bar{Y}_{+j+} \\ &\quad -2\bar{Y}_{i++}\bar{Y}_{+++}-2\bar{Y}_{+j+}\bar{Y}_{+++}) \\ &= \sum_{i,j}\bar{Y}^2_{ij+}+J\sum_i \bar{Y}^2_{i++}+I\sum_j \bar{Y}^2_{+j+}+IJ\bar{Y}^2_{+++} \\ &\quad -2J\sum_i \bar{Y}_{i++}-2I\sum_j \bar{Y}^2_{+j+}+2IJ\bar{Y}^2_{+++} \\ &\quad +2IJ\bar{Y}^2_{+++}-2IJ\bar{Y}^2_{+++}-2IJ\bar{Y}^2_{+++} \\ &= \sum_{i,j}\bar{Y}^2_{ij+}-J\sum_i \bar{Y}^2_{i++}-I\sum_j \bar{Y}^2_{+j+}+IJ\bar{Y}^2_{+++}. \end{aligned}$$

Next, we shall show that the denominators are equal:

$$\begin{aligned} \sum_{i,j,k}(\bar{Y}_{ijk}-\bar{Y}_{ij+})^2 &= \sum_{i,j,k}(Y^2_{ijk}-2Y_{ijk}\bar{Y}_{ij+}+\bar{Y}^2_{ij+}) \\ &= \sum_{i,j}\left(\sum_k Y^2_{ijk}-2K\bar{Y}^2_{ij+}+K\bar{Y}^2_{ij+}\right) \\ &= \sum_{i,j,k} Y^2_{ijk}-K\sum_{i,j}\bar{Y}^2_{ij+}. \end{aligned}$$

10. In this problem, $I=3, J=4$, and $K=2$. The values of the estimates can be calculated directly from Eqs. (11.8.6), (11.8.7), and (11.8.3).

11. It is found from Eq. (11.8.12) that $U^2_{AB} = 0.7047$. When the hypothesis is true, U^2_{AB} has the F distribution with $(I-1)(J-1) = 6$ and $IJ(K-1) = 12$ degrees of freedom. The tail area corresponding to the value just calculated is found to be greater than 0.05.

12. Since the hypothesis in Exercise 11 was not rejected, we proceed to test the hypotheses (11.8.13). It is found from Eq. (11.8.14) that $U^2_A = 7.5245$. When the hypothesis is true, U^2_A has the F distribution with $(I-1)J = 8$ and 12 degrees of freedom. The tail area corresponding to the value just calculated is found to be less than 0.025.

13. It is found from Eq. (11.8.18) that $U^2_B = 9.0657$. When the hypothesis is true, U^2_B has the F distribution with $I(J-1) = 9$ and 12 degrees of freedom. The tail area corresponding to the value just calculated is found to be less than 0.025.

14. The estimator $\hat{\mu}$ has the normal distribution with mean μ and, by Exercise 6, variance $\sigma^2/24$. Also, $\sum_{i,j,k}(Y_{ijk} - \bar{Y}_{ij+})^2/\sigma^2$ has a χ^2 distribution with 12 degrees of freedom, and these two random variables are independent. Therefore, when H_0 is true, the following statistic V will have the t distribution with 12 degrees of freedom:

$$V = \frac{\sqrt{24}(\hat{\mu} - 8)}{\left[\frac{1}{12}\sum_{i,j,k}(Y_{ijk} - \bar{Y}_{ij+})^2\right]^{1/2}}.$$

We could test the given hypotheses by carrying out a two-sided t test using the statistic V. Equivalently, as described in Sec. 9.7, V^2 will have the F distribution with 1 and 12 degrees of freedom. It is found that

$$V^2 = \frac{24(0.7708)^2}{\frac{1}{12}(10.295)} = 16.6221.$$

The corresponding tail area is less than 0.025.

15. The estimator $\hat{\alpha}_2$ has the normal distribution with mean α_2 and, by Exercise 6, variance $\sigma^2/12$. Hence, as in the solution of Exercise 14, when $\alpha_2 = 1$, the following statistic V will have the t distribution with 12 degrees of freedom:

$$V = \frac{\sqrt{12}(\hat{\alpha}_2 - 1)}{\left[\frac{1}{12}\sum_{i,j,k}(Y_{ijk} - \bar{Y}_{ij+})^2\right]^{1/2}}$$

The null hypothesis H_0 should be rejected if $V \geq c$, where c is an appropriate constant. It is found that

$$V = \frac{\sqrt{12}(0.7667)}{\left[\frac{1}{12}(10.295)\right]^{1/2}} = 2.8673$$

The corresponding tail area is between 0.005 and 0.01.

16. Since $E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$, then $E(\bar{Y}_{ij+}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$. The desired results can now be obtained from Eq. (11.8.19) and Eq. (11.8.5).

17. $\sum_{i=1}^{I} \hat{\alpha}_i = \frac{1}{J} \sum_{i=1}^{I} \sum_{j=1}^{J} \bar{Y}_{ij+} - I\hat{\mu} = I\hat{\mu} - I\hat{\mu} = 0$ and

$$\sum_{i=1}^{I} \hat{\gamma}_{ij} = \left(\sum_{i=1}^{I} \bar{Y}_{ij+} - I\hat{\mu} \right) - \sum_{i=1}^{I} \hat{\alpha}_i - I\hat{\beta}_j = I\hat{\beta}_j - 0 - I\hat{\beta}_j = 0.$$

It can be shown similarly that

$$\sum_{j=1}^{J} \hat{\beta}_j = 0 \quad \text{and} \quad \sum_{j=1}^{J} \hat{\gamma}_{ij} = 0.$$

18. Both $\hat{\mu}$ and $\hat{\alpha}_i$ are linear functions of the $\sum_{i=1}^{I} \sum_{j=1}^{J} K_{ij}$ independent random variables Y_{ijk}. Let $\hat{\mu} = \sum_{r,s,t} m_{rst} Y_{rst}$ and $\hat{\alpha}_i = \sum_{r,s,t} a_{rst} Y_{rst}$. Then it is found from Eq. (11.8.19) that

$$M_{rst} = \frac{1}{IJK_{rs}} \qquad \text{for all values of } r, s, \text{ and } t,$$

and

$$\begin{aligned} a_{ist} &= \frac{1}{JK_{is}} - \frac{1}{IJK_{is}}, \\ a_{rst} &= -\frac{1}{IJK_{rs}} \quad \text{for } r \neq i. \end{aligned}$$

As in the solution of Exercise 8,

$$\begin{aligned} \text{Cov}(\hat{\mu}, \hat{\alpha}_i) &= \sigma^2 \sum_{r,s,t} m_{rst} a_{rst} \\ &= \sigma^2 \left[\sum_{s=1}^{J} \sum_{t=1}^{K_{is}} \frac{1}{IJK_{is}} \left(\frac{1}{JK_{is}} - \frac{1}{IJK_{is}} \right) + \sum_{r \neq i} \sum_{s=1}^{J} \sum_{t=1}^{K_{rs}} \left(\frac{1}{IJK_{rs}} \right) \left(-\frac{1}{IJK_{rs}} \right) \right] \\ &= \sigma^2 \left[\frac{I-1}{I^2 J^2} \sum_{s=1}^{J} \frac{1}{K_{is}} - \frac{1}{I^2 J^2} \sum_{r \neq i} \sum_{s=1}^{J} \frac{1}{K_{rs}} \right] \\ &= \frac{\sigma^2}{I^2 J^2} \left[(I-1) \sum_{s=1}^{J} \frac{1}{K_{is}} - \left(\sum_{r=1}^{I} \sum_{s=1}^{J} \frac{1}{K_{rs}} - \sum_{s=1}^{J} \frac{1}{K_{is}} \right) \right] \\ &= \frac{\sigma^2}{I^2 J^2} \left[I \sum_{s=1}^{J} \frac{1}{K_{is}} - \sum_{r=1}^{I} \sum_{s=1}^{J} \frac{1}{K_{rs}} \right]. \end{aligned}$$

19. Notice that we cannot reject the second null hypothesis unless we accept the first null hypothesis, since we don't even test the second hypothesis if we reject the first one. The probability that the two-stage procedure rejects at least one of the two hypotheses is then

$$\begin{aligned} &\Pr(\text{reject first null hypothesis}) \\ &\quad + \Pr(\text{reject second null hypothesis and accept first null hypothesis}). \end{aligned}$$

The first term above is α_0, and the second term can be rewritten as

$$\Pr(\text{reject second null hypothesis}|\text{accept first null hypothesis}) \\ \times \Pr(\text{accept first null hypothesis}).$$

This product equals $\beta_0(1-\alpha_0)$, hence the overall probability is $\alpha_0 + \beta_0(1-\alpha_0)$.

20. (a) The three additional cell averages are 822.5, 821.7, and 770. The ANOVA table for the combined samples is

Source of variation	Degrees of freedom	Sum of squares	Mean square
Main effects of filter	1	1003	1003
Main effects of size	2	25817	12908
Interactions	2	739	369.4
Residuals	30	1992	66.39
Total	35	29551	

(b) The F statistic for the test of no interaction is $369.4/66.39 = 5.56$. Comparing this to the F distribution with 2 and 30 degrees of freedom, we get a p-value of 0.009.

(c) If we use the one-stage test procedure in which both the main effects and interactions are hypothesized to be 0 together, we get an F statistic equal to $[(25817 + 739)/4]/66.39 = 100$ with 3 and 30 degrees of freedom. The p-value is essentially 0.

(d) If we use the one-stage test procedure in which both the main effects and interactions are hypothesized to be 0 together, we get an F statistic equal to $[(1003 + 739)/3]/66.39 = 8.75$ with 3 and 30 degrees of freedom. The p-value is 0.0003.

11.9 Supplementary Exercises

Solutions to Exercises

1. The necessary calculations were done in Example 11.3.6. The least-squares coefficients are $\hat{\beta}_0 = -0.9709$ and $\hat{\beta}_1 = 0.0206$, with $\sigma' = 8.730 \times 10^{-3}$, and $n = 17$. We also can compute $s_x^2 = 530.8$ and $\overline{x}_n = 203.0$.

(a) A 90% confidence interval for β_1 is $\hat{\beta}_1 \pm T_{n-2}^{-1}(0.95)\sigma'/s_x$. This becomes $(0.01996, 0.02129)$.

(b) Since 0 is not in the 90% interval in part (a), we would reject H_0 at level $\alpha_0 = 0.1$.

(c) A 90% prediction interval for log-pressure at boiling-point equal to x is

$$\hat{\beta}_0 + x\hat{\beta}_1 \pm T_{n-2}^{-1}(0.95)\sigma'\left(1 + \frac{1}{n} + \frac{(x-\overline{x}_n)^2}{s_x^2}\right)^{1/2}.$$

With the data we have, this gives $[3.233, 3, 264]$. Converting this to pressure gives $(25.35, 26.16)$.

2. This result follows directly from the expressions for $\hat{\rho}, \hat{\sigma}_1$, and $\hat{\sigma}_2$ given in Exercise 24 of Sec. 7.6 and the expression for $\hat{\beta}_1$ given in Exercise 2a of Sec. 11.1.

3. The conditional distribution of Y_i given $X_i = x_i$ has mean $\beta_0 + \beta_1 x_i$, where

$$\beta_0 = \mu_2 - \frac{\rho\sigma_2}{\sigma_1}\mu_1 \quad \text{and} \quad \beta_1 = \frac{\rho\sigma_2}{\sigma_1},$$

and variance $(1-\rho^2)\sigma_2^2$. Since $T = \hat{\beta}_1$, as given in Exercise 2b of Sec. 11.1, it follows that $E(T) = \beta_1 = \rho\sigma_2/\sigma_1$ and

$$\text{Var}(T) = \frac{(1-\rho^2)\sigma_2^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}.$$

4. The least squares estimates will be the values of θ_1, θ_2, and θ_3 that minimize $Q = \sum_{i=1}^{3}(y_i - \theta_i)^2$, where $\theta_3 = 180 - \theta_1 - \theta_2$. If we solve the equations $\partial Q/\partial\theta_1 = 0$ and $\partial Q/\partial\theta_2 = 0$, we obtain the relations

$$y_1 - \theta_1 = y_2 - \theta_2 = y_3 - \theta_3.$$

Since $\sum_{i=1}^{3} y_i = 186$ and $\sum_{i=1}^{3} \theta_i = 180$, it follows that $\hat{\theta}_i = y_i - 2$ for $i = 1, 2, 3$. Hence $\hat{\theta}_1 = 81$, $\hat{\theta}_2 = 45$, and $\hat{\theta}_3 = 54$.

5. This result can be established from the formulas for the least squares line given in Sec. 11.1 or directly from the following reasoning: Let $x_1 = a$ and $x_2 = b$. The data contain one observation (a, y_1) at $x = a$ and $n-1$ observations $(b, y_2), \ldots, (b, y_n)$ at $x = b$. Let u denote the average of the $n-1$ values $y_2, \ldots, y_n$, and let h_a and h_b denote the height of the least square line at $x = a$ and $x = b$, respectively. Then the value of Q, as given by Eq. (11.1.2), is

$$Q = (y_1 - h_a)^2 + \sum_{j=2}^{n}(y_i - h_b)^2.$$

The first term is minimized by taking $h_a = y_1$ and the summation is minimized by taking $h_b = u$. Hence, Q is minimized by passing the straight line through the two points (a, y_1) and (b, u). But (a, y_1) is the point (x_1, y_1).

6. The first line is the usual least squares line $y = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_1$ is given in Exercise 2a of Sec. 11.1. In the second line, the roles of x and y are interchanged, so it is $x = \hat{\alpha}_1 + \hat{\alpha}_2 y$, where

$$\hat{\alpha}_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}.$$

Both lines pass through the point $(\bar{x}_n, \bar{y}_n)$, so they will coincide if and only if they have the same slope; i.e., if and only if $\hat{\beta}_1 = 1/\hat{\alpha}_2$. This condition reduces to the condition that $\hat{\rho}^2 = 1$, where $\hat{\rho}$ is given in Exercise 24 of Sec. 7.6 and is the sample correlation coefficient. But $\hat{\rho}^2 = 1$ if and only if the n points lie exactly on a straight line. Hence, the two least squares lines will coincide if and only if all n points lie exactly on a straight line.

7. It is found from standard calculus texts that the sum of the squared distances from the points to the line is

$$Q = \frac{\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2}{1 + \beta_2^2}.$$

The equation $\partial Q/\partial\beta_1 = 0$ reduces to the relation $\beta_1 = \bar{y}_n - \beta_2\bar{x}_n$. If we replace β_1 in the equation $\partial Q/\partial\beta_2 = 0$ by this quantity, we obtain the relation:

$$(1+\beta_2^2) = \sum_{i=1}^{n}[(y_i-\bar{y}_n)-\beta_2(x_i-\bar{x}_n)]x_i + \beta_2\sum_{i=1}^{n}[(y_i-\bar{y}_n)-\beta_2(x_i-\bar{x}_n)]^2 = 0.$$

Note that we can replace the factor x_i in the first summation by $x_i - \bar{x}_n$ without changing the value of the summation. If we then let $x_i' = x_i - \bar{x}_n$ and $y_i' = y_i - \bar{y}_n$, and expand the final squared term, we obtain the following relation after some algebra:

$$(\beta_2^2 - 1)\sum_{i=1}^{n} x_i' y_i' + \beta_2 \sum_{i=1}^{n}(x_i'^2 - y_i'^2) = 0.$$

Hence

$$\beta_2 = \frac{\sum_{i=1}^{n}\left(y_i'^2 - x_i'^2\right) \pm \left[\left(\sum_{i=1}^{n}\left(y_i'^2 - x_i'^2\right)\right)^2 + 4\left(\sum_{i=1}^{n} x_i' y_i'\right)^2\right]^{1/2}}{2\sum_{i=1}^{n} x_i' y_i'}.$$

Either the plus sign or the minus sign should be used, depending on whether the optimal line has positive or negative slope.

8. This phenomenon was discussed in Exercise 19 of Sec. 11.2. The conditional expectation $E(X_2|X_1)$ of the sister's score X_2 given the first twin's score X_1 can be derived from Eq. (5.10.6) with $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$. Hence,

$$E(X_2|X_1) = \mu + \rho(X_1 - \mu) = (1-\rho)\mu + \rho X_1,$$

which is between μ and X_1. The same holds with subscripts 1 and 2 switched.

9.

$$\begin{aligned} v^2 &= \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+} + \bar{x}_{i+} - \bar{x}_{++})^2 \\ &= \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2 + \frac{1}{n}\sum_{i=1}^{k} n_i(\bar{x}_{i+} - \bar{x}_{++})^2 \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i[v_i^2 + (\bar{x}_{i+} - \bar{x}_{++})^2]. \end{aligned}$$

10. In the notation of Sec. 11.5, the design matrix is

$$\boldsymbol{Z} = \begin{bmatrix} w_1 & x_1 \\ \vdots & \vdots \\ w_n & x_n \end{bmatrix}.$$

For $t = w, x, Y$ and $u = w, x, Y$, let $S_{tu} = \sum_{i=1}^{n} t_i u_i$. Then

$$\begin{aligned}
\boldsymbol{Z}'\boldsymbol{Z} &= \begin{bmatrix} S_{ww} & S_{wx} \\ S_{wx} & S_{xx} \end{bmatrix}, \\
(\boldsymbol{Z}'\boldsymbol{Z})^{-1} &= \frac{1}{S_{ww}S_{xx} - S_{wx}^2} \begin{bmatrix} S_{xx} & -S_{wx} \\ -S_{wx} & S_{ww} \end{bmatrix}, \\
\boldsymbol{Z}'\boldsymbol{Y} &= \begin{bmatrix} S_{wY} \\ S_{xY} \end{bmatrix}.
\end{aligned}$$

Hence,

$$(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \frac{1}{S_{ww}S_{xx} - S_{wx}^2} \begin{bmatrix} S_{xx}S_{wY} - S_{wx}S_{xY} \\ S_{ww}S_{xY} - S_{wx}S_{wY} \end{bmatrix}.$$

The first component on the right side of this equation is $\hat{\beta}_0$ and the second is $\hat{\beta}_1$.

11. It was shown in Sec. 11.8 that the quantity $S^2_{\text{Resid}}/\sigma^2$ given in Eq. (11.8.10) has a χ^2 distribution with $IJ(K-1)$ degrees of freedom. Hence, the random variable $S^2_{\text{Resid}}/[IJ(K-1)]$ is an unbiased estimator of σ^2.

12. It follows from Table 11.23 that if $\alpha_i = \beta_j = 0$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, and $Q = S_A^2 + S_B^2$ then Q/σ^2 will have a χ^2 distribution with $(I-1) + (J-1) = I + J - 2$ degrees of freedom. Furthermore, regardless of the values of α_i and β_j, $R = S^2_{\text{Resid}}/\sigma^2$ will have a χ^2 distribution with $(I-1)(J-1)$ degrees of freedom, and Q and R will be independent. Hence, under H_0, the statistic

$$U = \frac{(I-1)(J-1)Q}{(I+J-2)R}$$

will have the F distribution with $I + J - 2$ and $(I-1)(J-1)$ degrees of freedom. The null hypothesis H_0 should be rejected if $U \geq c$.

13. Suppose that $\alpha_i = \beta_j = \gamma_{ij} = 0$ for all values of i and j. Then it follows from Table 11.28 that $(S_A^2 + S_B^2 + S^2_{\text{Int}})/\sigma^2$ will have a χ^2 distribution with $(I-1)+(J-1)+(I-1)(J-1) = IJ-1$ degrees of freedom. Furthermore, regardless of the values of α_i, β_j and γ_{ij}, $S^2_{\text{Resid}}/\sigma^2$ will have a χ^2 distribution with $IJ(K-1)$ degrees of freedom, and $S_A^2 + S_B^2 + S^2_{\text{Int}}$ and S^2_{Resid} will be independent. Hence, under H_0, the statistic

$$U = \frac{IJ(K-1)(S_A^2 + S_B^2 + S^2_{\text{Int}})}{(IJ-1)S^2_{\text{Resid}}}$$

will have the F distribution with $IJ - 1$ and $IJ(K-1)$ degrees of freedom. The null hypothesis H_0 should be rejected if $U \geq c$.

14. The design in this exercise is a two-way layout with two levels of each factor and K observations in each cell. The hypothesis H_0 is precisely the hypothesis H_0 given in (11.8.11) that the effects of the two factors are additive and all interactions are 0. Hence, H_0 should be rejected if $U^2_{AB} > c$, where U^2_{AB} is given by (11.8.12) with $I = J = 2$, and U^2_{AB} has an F distribution with 1 and $4(K-1)$ degrees of freedom when H_0 is true.

15. Let $Y_1 = W_1, Y_2 = W_2 - 5$, and $Y_3 = \frac{1}{2}W_3$. Then the random vector

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

satisfies the conditions of the general linear model as described in Sec. 11.5 with

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}.$$

Thus,

$$\boldsymbol{Z}'\boldsymbol{Z} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \qquad (\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{bmatrix},$$

and

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} = \begin{bmatrix} \frac{1}{4}Y_1 + \frac{1}{4}Y_2 + \frac{1}{2}Y_3 \\ \frac{1}{4}Y_1 + \frac{1}{4}Y_2 - \frac{1}{2}Y_3 \end{bmatrix}.$$

Also,

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{3}(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}) \\ &= \frac{1}{3}\left[(Y_1 - \hat{\theta}_1 - \hat{\theta}_2)^2 + (Y_2 - \hat{\theta}_1 - \hat{\theta}_2)^2 + (Y_3 - \hat{\theta}_1 + \hat{\theta}_2)^2\right]. \end{aligned}$$

The following distributional properties of these M.L.E.'s are known from Sec. 11.5: $(\hat{\theta}_1, \hat{\theta}_2)$ and $\hat{\sigma}^2$ are independent; $(\hat{\theta}_1, \hat{\theta}_2)$ has a bivariate normal distribution with mean vector (θ_1, θ_2) and covariance matrix

$$\sigma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \begin{bmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{bmatrix}\sigma^2;$$

$3\hat{\sigma}^2/\sigma^2$ has a χ^2 distribution with one degree of freedom.

16. Direct application of the theory of least squares would require choosing α and β to minimize

$$Q_1 = \sum_{i=1}^{n}(y_i - \alpha x_i^{\beta})^2.$$

This minimization must be carried out numerically since the solution cannot be found in closed form. However, if we express the required curve in the form $\log y = \log \alpha + \beta \log x$, and then apply the method of least squares, we must choose β_0 and β_1 to minimize

$$Q_2 = \sum_{i=1}^{n}(\log y_i - \beta_0 - \beta_1 \log x_i)^2,$$

where $\beta_0 = \log \alpha$ and $\beta_1 = \beta$. The least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ can now be found as in Sec. 11.1, based on the values of $\log y_i$ and $\log x_i$. Estimates of α and β can then be obtained from the relations $\log \hat{\alpha} = \hat{\beta}_0$ and $\hat{\beta} = \hat{\beta}_1$. It should be emphasized that these values will not be the same as the least squares estimates found by minimizing Q_1 directly.

The appropriateness of each of these methods depends on the appropriateness of minimizing Q_1 and Q_2. The first method is appropriate if $Y_i = \alpha x_i^\beta + \varepsilon_i$, where ε_i has a normal distribution with mean 0 and variance σ^2. The second method is appropriate if $Y_i = \alpha x_i^\beta \varepsilon_i$, where $\log \varepsilon_i$ has the normal distribution just described.

17. It follows from the expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$ given by Eqs. (11.1.1) and (11.2.7) that

$$\begin{aligned}
e_i &= Y_i - (\bar{Y}_n - \bar{x}_n\hat{\beta}_1) - \hat{\beta}_1 x_i \\
&= Y_i - \bar{Y}_n + \frac{(\bar{x}_n - x_i)\sum_{j=1}^{n}(x_j - \bar{x}_n)Y_j}{s_x^2} \\
&= Y_i\left[1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{s_x^2}\right] \\
&\quad -\sum_{j\neq i} Y_j\left[\frac{1}{n} + \frac{(x_i - \bar{x}_n)(x_j - \bar{x}_n)}{s_x^2}\right]
\end{aligned}$$

where $s_x^2 = \sum_{j=1}^{n}(x_j - \bar{x}_n)^2$. Since $Y_1, \ldots, Y_n$ are independent and each has variance σ^2, it follows that

$$\mathrm{Var}(e_i) = \sigma^2\left[1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{s_x^2}\right]^2 + \sigma^2\sum_{j\neq i}\left[\frac{1}{n} + \frac{(x_i - \bar{x}_n)(x_j - \bar{x}_n)}{s_x^2}\right]^2.$$

Let $Q_i = \frac{1}{n} + \frac{(x_i - \bar{x}_n)^2}{s_x^2}$. Then

$$\begin{aligned}
\mathrm{Var}(e_i) &= \sigma^2(1 - Q_i)^2 + \sigma^2\sum_{j=1}^{n}\left[\frac{1}{n} + \frac{(x_i - \bar{x}_n)(x_j - \bar{x}_n)}{s_x^2}\right]^2 - \sigma^2 Q_i^2 \\
&= \sigma^2[(1 - Q_i)^2 + Q_i - Q_i^2] \\
&= \sigma^2(1 - Q_i).
\end{aligned}$$

(This result could also have been obtained from the more general result to be obtained next in Exercise 18.) Since Q_i is an increasing function of $(x_i - \bar{x}_n)^2$, it follows that $\mathrm{Var}(e_i)$ is a decreasing function of $(x_i - \bar{x}_n)^2$ and, hence, of the distance between x_i and $\bar{x}_n$.

18. (a) Since $\hat{\boldsymbol{\beta}}$ has the form given in Eq. (11.5.10), it follows directly that $\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}$ has the specified form.

(b) Let $\boldsymbol{A} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. It can be verified directly that A is idempotent, i.e., $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$. Since $\boldsymbol{D} = \boldsymbol{I} - \boldsymbol{A}$, it now follows that

$$\boldsymbol{DD} = (\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} - \boldsymbol{A}) = \boldsymbol{II} - \boldsymbol{AI} - \boldsymbol{IA} + \boldsymbol{AA} = \boldsymbol{I} - \boldsymbol{A} - \boldsymbol{A} + \boldsymbol{A} = \boldsymbol{I} - \boldsymbol{A} = \boldsymbol{D}.$$

(c) As stated in Eq. (11.5.15), $\mathrm{Cov}(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}$. Hence, by Theorem 11.5.2, $\mathrm{Cov}(\boldsymbol{W}) = \mathrm{Cov}(\boldsymbol{D}\boldsymbol{Y}) = \boldsymbol{D}\,\mathrm{Cov}(\boldsymbol{Y})\boldsymbol{D}' = \boldsymbol{D}(\sigma^2\boldsymbol{I})\boldsymbol{D} = \sigma^2(\boldsymbol{D}\boldsymbol{D}) = \sigma^2\boldsymbol{D}$.

19. Let $\bar{\theta} = \sum_{i=1}^{I} v_i\theta_i/v_+$ and $\bar{\psi} = \sum_{j=1}^{J} w_j\psi_j/w_+$, and define $\mu = \bar{\theta} + \bar{\psi}, \alpha_i = \theta_i - \bar{\theta}$, and $\beta_j = \psi_j - \bar{\psi}$. Then $E(Y_{ij}) = \theta_i + \psi_j = \mu + \alpha_i + \beta_j$ and $\sum_{i=1}^{I} v_i\alpha_i = \sum_{j=1}^{J} w_j\beta_j = 0$. To establish uniqueness, suppose that μ', α_i', and β_j' are another set of values satisfying the required conditions. Then

$$\mu + \alpha_i + \beta_j = \mu' + \alpha_i' + \beta_j' \quad \text{for all} \quad i \quad \textit{and} \quad j.$$

If we multiply both sides by $v_i w_j$ and sum over i and j, we find that $\mu = \mu'$. Hence, $\alpha_i + \beta_j = \alpha_i' + \beta_j'$. If we now multiply both sides by v_i and sum over i, we find that $\beta_j = \beta_j'$. Similarly, if we multiply both sides by w_j and sum over j, we find that $\alpha_i = \alpha_i'$.

20. The value of μ, α_i, and β_j must be chosen to minimize

$$Q = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K_{ij}}(Y_{ijk} - \mu - \alpha_i - \beta_j)^2.$$

The equation $\partial Q/\partial\mu = 0$ reduces to

$$Y_{+++} - n\mu - \sum_{i=1}^{I} K_{i+}\alpha_i - \sum_{j=1}^{J} K_{+j}\beta_j = 0,$$

where $n = K_{++}$ is the total number of observations in the two-way layout. Next we shall calculate $\frac{\partial Q}{\partial \alpha_i}$ for $i = 1, \dots, I-1$, keeping in mind that $\sum_{i=1}^{I} K_{i+}\alpha_i = 0$. Hence, $\partial\alpha_I/\partial\alpha_i = -K_{i+}/K_{I+}$. It can be found that the equation $\partial Q/\partial\alpha_i = 0$ reduces to the following equation for $i = 1, \dots, I-1$:

$$Y_{i++} - K_{i+}\mu - K_{i+}\alpha_i - \sum_j K_{ij}\beta_j = \frac{K_{i+}}{K_{I+}}(Y_{I++} - K_{I+}\mu - K_{I+}\alpha_I - \sum_j K_{IJ}\beta_j).$$

In other words, the following quantity must have the same value for $i = 1, \dots, I$:

$$\frac{1}{K_{i+}}\left(Y_{i++} - K_{i+}\mu - K_{i+}\alpha_i - \sum_j K_{ij}\beta_j\right).$$

Similarly, the set of equations $\partial Q/\partial\beta_j = 0$ for $j = 1, \dots, J-1$ reduces to the requirement that the following quantity have the same value for $j = 1, \dots, J$:

$$\frac{1}{K_{+j}}\left(Y_{+j+} - K_{+j}\mu - \sum_i K_{ij}\alpha_i - K_{+j}\beta_j\right).$$

It can be verified by direct substitution that the values of $\hat{\mu}, \hat{\alpha}_i$, and $\hat{\beta}_j$ given in the exercise satisfy all these requirements and, hence, are the least squares estimators.

21. As in the solution of Exercise 18 of Sec. 11.8, let

$$\begin{aligned}
\hat{\mu} &= \sum_{r,s,t} m_{rst}Y_{rst}, \\
\hat{\alpha}_i &= \sum_{r,s,t} a_{rst}Y_{rst}, \\
\hat{\beta}_j &= \sum_{r,s,t} b_{rst}Y_{rst}.
\end{aligned}$$

To show that $\text{Cov}(\hat{\mu}, \hat{\alpha}_i) = 0$, we must show that $\sum_{r,s,t} m_{rst} a_{rst} = 0$. But $m_{rst} = \frac{1}{n}$ for all r, s, t, and

$$a_{rst} = \begin{cases} \dfrac{1}{K_{i+}} - \dfrac{1}{n} & \text{for} \quad r = i, \\ -\dfrac{1}{n} & \text{for} \quad r \neq i. \end{cases}$$

Hence, it is found that $\sum m_{rst} a_{rst} = 0$. Similarly,

$$b_{rst} = \begin{cases} \dfrac{1}{K_{+j}} - \dfrac{1}{n} & \text{for} \quad s = j, \\ -\dfrac{1}{n} & \text{for} \quad s \neq j. \end{cases}$$

and $\sum m_{rst} b_{rst} = 0$, so $\text{Cov}(\hat{\mu}, \hat{\beta}_j) = 0$.

22. We must show that $\sum a_{rst} b_{rst} = 0$, where a_{rst} and b_{rst} are given in the solution of Exercise 21:

$$\begin{aligned} \sum_{r,s,t} a_{rst} b_{rst} &= \sum_{k=1}^{K_{ij}} \left(\frac{1}{K_{i+}} - \frac{1}{n}\right)\left(\frac{1}{K_{+j}} - \frac{1}{n}\right) - \frac{1}{n}\sum_{s\neq j}\sum_{k=1}^{K_{is}}\left(\frac{1}{K_{i+}} - \frac{1}{n}\right) - \frac{1}{n}\sum_{r\neq i}\sum_{k=1}^{K_{rj}}\left(\frac{1}{K_{+j}} - \frac{1}{n}\right) \\ &\quad + \frac{1}{n^2}\sum_{r\neq i}\sum_{s\neq j} K_{rs}. \end{aligned}$$

Since $nK_{ij} = K_{i+}K_{+j}$, it can be verified that this sum is 0.

23. Consider the expression for θ_{ijk} given in this exercise. If we sum both sides of this expression over i, j, and k, then it follows from the constraints on the α's, β's, and γ's, that $\mu = \bar{\theta}_{+++}$. If we substitute this value for μ and sum both sides of the expression over j and k, we can solve the result for α_i^A. Similarly, α_j^B can be found by summing over i and k, and α_k^C by summing over i and j. After these values have been found, we can determine β_{ij}^{AB} by summing both sides over k, and determine β_{ik}^{AC} and β_{jk}^{BC} similarly. Finally, γ_{ijk} is determined by taking its value to be whatever is necessary to satisfy the required expression for θ_{ijk}. In this way, we obtain the following values:

$$\begin{aligned} \mu &= \bar{\theta}_{+++}, \\ \alpha_i^A &= \bar{\theta}_{i++} - \bar{\theta}_{+++}, \\ \alpha_j^B &= \bar{\theta}_{+j+} - \bar{\theta}_{+++}, \\ \alpha_k^C &= \bar{\theta}_{++k} - \bar{\theta}_{+++}, \\ \beta_{ij}^{AB} &= \bar{\theta}_{ij+} - \bar{\theta}_{i++} - \bar{\theta}_{+j+} + \bar{\theta}_{+++}, \\ \beta_{ik}^{AC} &= \bar{\theta}_{i+k} - \bar{\theta}_{i++} - \bar{\theta}_{++k} + \bar{\theta}_{+++}, \\ \beta_{jk}^{BC} &= \bar{\theta}_{+jk} - \bar{\theta}_{+j+} - \bar{\theta}_{++k} + \bar{\theta}_{+++}, \\ \gamma_{ijk} &= \theta_{ijk} - \bar{\theta}_{ij+} - \bar{\theta}_{i+k} - \bar{\theta}_{+jk} + \bar{\theta}_{i++} + \bar{\theta}_{+j+} + \bar{\theta}_{++k} - \bar{\theta}_{+++}. \end{aligned}$$

It can be verified that these quantities satisfy all the specified constraints. They are unique by the method of their construction, since they were derived as the only values that could possibly satisfy the constraints.

24. (a) The plot of Buchanan vote against total county vote is in Fig. S.11.3. Palm Beach county is plotted with the symbol P.

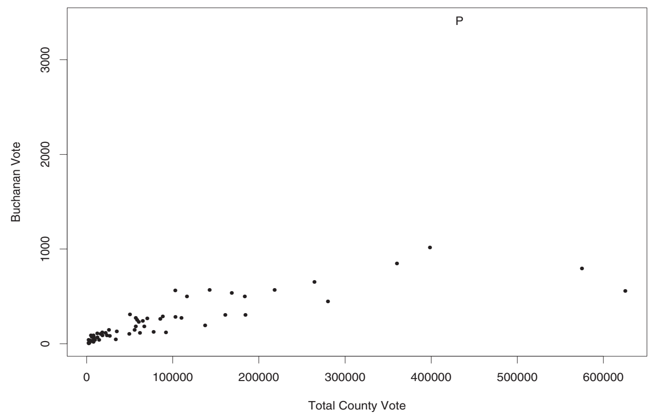

Figure S.11.3: Figure for Exercise 24a in Sec. 11.9.

(b) The summary of the fitted regression is $\hat{\beta}_0 = 83.69$, $\hat{\beta}_1 = 0.00153$, $\overline{x}_n = 8.254 \times 10^4$, $s_x^2 = 1.035 \times 10^{12}$, $n = 67$, and $\sigma' = 120.1$.

(c) The plot of residuals is in Fig. S.11.4. Notice that the residuals are much more spread out at the

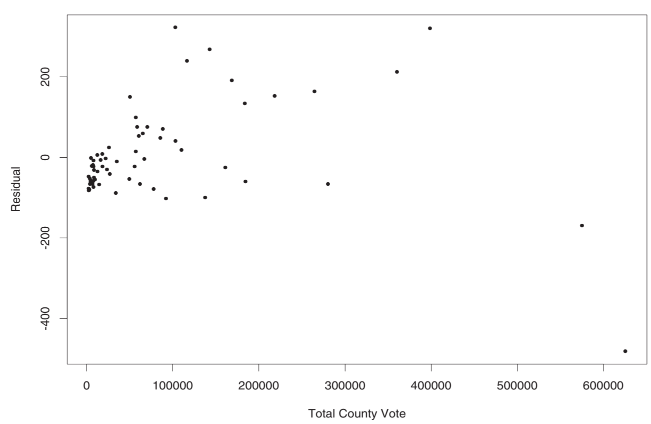

Figure S.11.4: Figure for Exercise 24c in Sec. 11.9.

right side of the plot than at the left. There also appears to be a bit of a curve to the plot.

(d) The summary of the fitted regression is $\hat{\beta}_0 = -2.746$, $\hat{\beta}_1 = 0.7263$, $\overline{x}_n = 10.32$, $s_x^2 = 151.5$, $n = 67$, and $\sigma' = 0.4647$.

(e) The new residual plot is in Fig. S.11.5. The spread is much more uniform from right to left and the curve is no longer evident.

(f) The quantile we need is $T_{65}^{-1}(0.995) = 2.654$. The logarithm of total vote for Palm Beach county is 12.98. prediction interval for logarithm of Buchanan vote when $X = 12.98$ is

$$
\begin{aligned}
& -2.746 + 12.98 \times 0.7263 \pm 2.654 \times 0.4647 \left(1 + \frac{1}{67} + \frac{(12.98 - 10.32)^2}{151.5}\right)^{1/2} \\
& = \quad [5.419, 7.942].
\end{aligned}
$$

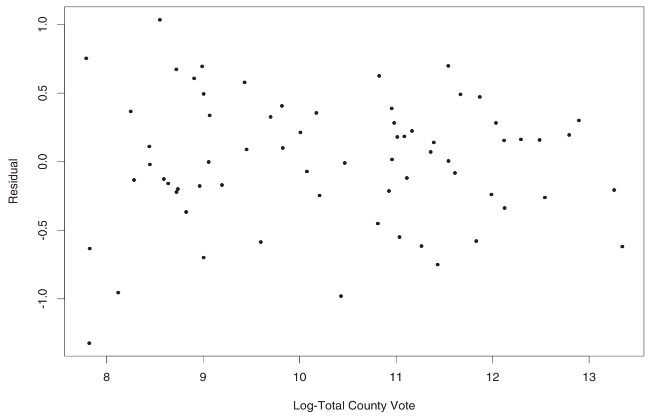

Figure S.11.5: Figure for Exercise 24e in Sec. 11.9.

Converting to Buchanan vote, we take e to the power of each endpoint and get the interval $[225.6, 2812]$.

(g) The official Gore total was 2912253, while the official Bush total was 2912790. Suppose that 2812 people in Palm Beach county had actually voted for Buchanan and the other $3411 - 2812 = 599$ had voted for Gore. Then the Gore total would have been 2912852, enough to make him the winner.

Chapter 12

Simulation

All exercises that involve simulation will produce different answers when run with repeatedly. Hence, one cannot expect numerical results to match perfectly with any answers given here.

For all exercises that require simulation, students will need access to software that will do some of the work for them. At a minimum they will need software to simulate uniform pseudo-random numbers on the interval $[0, 1]$. Some of the exercises require software to simulate all of the famous distributions and compute the c.d.f.'s and quantile functions of the famous distributions.

Some of the simulations require a nonnegligible programming effort. In particular, Markov chain Monte Carlo (MCMC) requires looping through all of the coordinates inside of the iteration loop. Assessing convergence and simulation standard error for a MCMC result requires running several chains in parallel. Students who do not have a lot of programming experience might need some help with these exercises.

If one is using the software *R*, the function `runif` will return uniform pseudo-random numbers, the argument is how many you want. For other named distributions, as mentioned earlier in this manual, one can use the functions `rbinom`, `rhyper`, `rpois`, `rnbinom`, `rgeom`, `rnorm`, `rlnorm`, `rgamma`, `rexp`, `rbeta`, and `rmultinom`.

Most simulations require calculation of averages and sample variances. The functions `mean`, `median`, and `var` compute the average, sample median, and sample variance respectively of their first argument. Each of these has an optional argument `na.rm`, which can be set either to `TRUE` or to `FALSE` (the default). If true, `na.rm` causes missing values to be ignored in the calculation. Missing values in simulations should be rare if calculations are being done correctly. Other useful functions for simulations are `sort` and `sort.list`. They both take a vector argument. The first returns its argument sorted algebraically from smallest to largest (or largest to smallest with optional argument `decreasing=TRUE`.) The second returns a list of integers giving the locations (in the vector) of the ordered values of its argument. The functions `min` and `max` give the smallest and largest values of their argument.

For looping, one can use

```
for(i in 1:n){ ... }
```

to perform all of the functions between `{` and `}` once for each value of `i` from `1` to `n`. For an indeterminate number of iterations, one can use

```
while(expression){ ... }
```

where *expression* stands for a logical expression that changes value from `TRUE` to `FALSE` at some point during the iterations.

A long series of examples appears at the end.

12.1 What is Simulation?

Solutions to Exercises

1. Simulate a large number of exponential random variables with paremeter 1, and take their average.

2. We would expect that every so often one of the simulated random variables would be much larger than the others and the sample average would go up significantly when that random variable got included in the average. The more simulations we do, the more such large observations we would expect, and the average should keep getting larger.

3. We would expect to get a lot of very large positive observations and a lot of very large negative observations. Each time we got one, the average would either jump up (when we get a positive one) or jump down (when we get a negative one). As we sampled more and more observations, the average should bounce up and down quite a bit and never settle anywhere.

4. We could count how many Bernoulli's we had to sample to get a success (1) and call that the first observation of a geometric random variable. Starting with the next Bernoulli, start counting again until the next 1, and call that the second geometric, etc. Average all the observed geometrics to approximate the mean.

5. (a) Simulate three exponentials at a time. Call the sum of the first two X and call the third one Y. For each triple, record whether $X < Y$ or not. The proportion of times that $X < Y$ in a large sample of triples approximates $\Pr(X < Y)$.

   (b) Let Z_1, Z_2, Z_3 be i.i.d. having the exponential distribution with parameter β, and let W_1, W_2, W_3 be i.i.d. having the exponential distribution with parameter 1. Then $Z_1 + Z_2 < Z_3$ if and only if $\beta Z_1 + \beta Z_2 < \beta Z_3$. But $(\beta Z_1, \beta Z_2, \beta Z_3)$ has precisely the same joint distribution as (W_1, W_2, W_3). So, the probability that $Z_1 + Z_2 < Z_3$ is the same as the probability that $W_1 + W_2 < W_3$, and it doesn't matter which parameter we use for the exponential distribution. All simulations will approximate the same quantity as we would approximate using parameter 1.

   (c) We know that X and Y are independent and that X has the gamma distribution with parameters 2 and 0.4. The joint p.d.f. is

   $$f(x, y) = 0.4^2 x \exp(-0.4x) 0.4 \exp(-0.4y), \text{ for } x, y > 0.$$

   The integral to compute the probability is

   $$\Pr(X < Y) = \int_0^\infty \int_x^\infty 0.4^3 x \exp(-0.4[x + y]) dy dx.$$

   There is also a version with the x integral on the inside.

   $$\Pr(X < Y) = \int_0^\infty \int_0^y 0.4^3 x \exp(-0.4[x + y]) dx dy.$$

12.2 Why Is Simulation Useful?

Commentary

This section introduces the fundamental concepts of simulation and illustrates the basic calculations that underlie almost all simulations. Instructors should stress the need for assessing the variability in a simulation result. For complicated simulations, it can be difficult to assess variability, but students need to be aware that a highly variable simulation may be no better than an educated guess.

The lengthy examples (12.2.13 and 12.2.14) at the end of the section and the exercises (15 and 16) that go with them are mainly illustrative of the power of simulation. These would only be covered in course that devoted a lot of time to simulation.

Solutions to Exercises

1. Since $E(Z) = \mu$, the Cheybshev inequality says that $\Pr(|Z-\mu| \le \epsilon) \ge \epsilon^2/\operatorname{Var}(Z)$. Since Z is the average of v independent random variables with variance σ^2, $\operatorname{Var}(Z) = \sigma^2/v$. It follows that

$$\Pr(|Z-\mu| \le \epsilon) \ge \frac{\epsilon^2 v}{\sigma^2}.$$

Now, suppose that $v \ge \sigma^2/[\epsilon^2(1-\gamma)$, then

$$\frac{\epsilon^2 v}{\sigma^2} \ge 1-\gamma.$$

2. In Example 12.2.11, we are approximating σ by 0.3892. According to Eq. (12.2.6), we need

$$v \ge \frac{0.3892^2}{0.01^2 \times .01} = 151476.64$$

So, v must be at least 151477.

3. We could simulate a lot (say v_0) standard normal random variables $W_1, \ldots, W_{v_0}$ and let $X_i = 7W_i + 2$. Then each X_i has the distribution of X. Let $W_i = \log(|X_i| + 1)$. We could then compute Z equal to the average of the W_i's as an estimate of $E(\log(|X| + 1))$. If we needed our estimate to be close to $E(\log(|X|+1))$ with high probability, we could estimate the variance of W_i by the sample variance and then use (12.2.5) to choose a possibly larger simulation size.

4. Simulate 15 random variables $U_1, \ldots, U_{15}$ with uniform distribution on the interval $[0, 1]$. For $i = 1, \ldots, 13$, let $X_i = 2(U_i - 0.5)$ and for $i = 14, 15$, let $X_i = 20(U_i - 0.5)$. Then $X_1, \ldots, X_{15}$ have the desired distribution. In most of my simulations, the median or the sample average was the closest to 0. The first simulation led to the following six values:

		Trimmed mean				
Estimator	Average	$k=1$	$k=2$	$k=3$	$k=4$	Median
Estimate	0.5634	0.3641	0.2205	0.2235	0.2359	0.1836

5. (a) In my ten samples, the sample median was closest to 0 nine times, and the $k = 3$ trimmed mean was closet to 0 one time.

   (b) Although the $k = 2$ trimmed mean was never closest to 0, it was also never very far from 0, and it had the smallest average squared distance from 0. The $k = 3$ trimmed mean was a close second. Here are the six values for my first 10 simulations:

		Trimmed mean				
Estimator	Average	$k=1$	$k=2$	$k=3$	$k=4$	Median
M.S.E.	0.4425	0.1354	0.0425	0.0450	0.0509	0.0508

   These rankings were also reflected in a much larger simulation.

6. (a) Simulate lots (say v_0) of random variables $X_1, \ldots, X_{v_0}$ and $Y_1, \ldots, Y_{v_0}$ with X_i have the beta distribution with parameters 3.5 and 2.7 while Y_i have the beta distribution with parameters 1.8 and 4.2. Let $Z_i = X_i/(X_i + Y_i)$. The sample average of $Z_1, \ldots, Z_{v_0}$ should be close to the mean of $X/(X+Y)$ if v_0 is large enough.

   (b) We could calculate the sample variance of $Z_1, \ldots, Z_{v_0}$ and use this as an estimate of σ^2 in Eq. (12.2.5) with $\gamma = 0.98$ and $\epsilon = 0.01$ to obtain a new simulation size.

7. (a) The distribution of X is the contaminated normal distribution with p.d.f. given in Eq. (10.7.2) with $\sigma = 1$, $\mu = 0$.

   (b) To calculate a number in Table 10.40, we should simulate lots of samples of size 20 from the distribution in part (a) with the desired ϵ (0.05 in this case). For each sample, compute the desired estimator (the sample median in this case). Then compute the average of the squares of the estimators (since $\mu = 0$ in our samples) and multiply by 20. As an example, we did two simulations of size 10000 each and got 1.617 and 1.621.

8. (a) The description is the same as in Exercise 7(b) with "sample median" replaced by "trimmed mean for $k = 2$" and 0.05 replaced by 0.1.

   (b) We did two simulations of size 10000 each and got 2.041 and 2.088. It would appear that this simulation is slightly more variable than the one in Exercise 7.

9. The marginal p.d.f. of X is

$$\int_0^\infty \frac{\mu^3}{2} \exp(-\mu(x+1)) d\mu = \frac{3}{(x+1)^4},$$

for $x > 0$. The c.d.f. of X is then

$$F(x) = \int_0^x \frac{3}{(t+1)^4} dx = 1 - \left(\frac{1}{x+1}\right)^3,$$

for $x > 0$, and $F(x) = 0$ for $x \le 0$. The median is that x such that $F(x) = 1/2$, which is easily seen to be $2^{1/3} - 1 = 0.2599$.

10. (a) The c.d.f. of each X_i is $F(x) = 1 - \exp(-\lambda x)$, for $x > 0$. The median is $\log(2)/\lambda$.

   (b) Let $Y_i = X_i \lambda$, and let M' be the sample median of $Y_1, \ldots, Y_{21}$. Then the Y_i's have the exponential distribution with parameter 1, the median of Y_i is $\log(2)$, and $M' = M\lambda$. The M.S.E. of M is then

$$\begin{aligned} E\left[\left(M - \frac{\log(2)}{\lambda}\right)^2\right] &= \frac{1}{\lambda^2} E[(M\lambda - \log(2))^2] \\ &= \frac{1}{\lambda^2} E[(M' - \log(2))^2] \\ &= \frac{\theta}{\lambda^2}. \end{aligned}$$

   (c) Simulate a lot (say $21v_0$) of random variables $X_1, \ldots, X_{21v_0}$ having the exponential distribution with parameter 1. For $i = 1, \ldots, v_0$, let M_i be the sample median of $X_{21(i-1)+1}, \ldots, X_{21i}$. Let $Y_i = (M_i - \log(2))^2$, and compute the sample average $Z = \frac{1}{v_0} \sum_{i=1}^{v_0} Y_i$ as an estimate of θ. If you want to see how good an estimate it is, compute the simulation standard error.

11. In Example 12.2.4, μ_x and μ_y are independent with $(\mu_x - \mu_{x1})/(\beta_{x1}/[\alpha_{x1}\lambda_{x1}])^{1/2}$ having the t distribution with $2\alpha_{x1}$ degrees of freedom and $(\mu_y - \mu_{y1})/(\beta_{y1}/[\alpha_{y1}\lambda_{y1}])^{1/2}$ having the t distribution with $2\alpha_{y1}$ degrees of freedom. We should simulate lots (say v) of t random variables $T_x^{(1)}, \ldots, T_x^{(v)}$ with $2\alpha_{x1}$ degrees of freedom and just as many t random variables $T_y^{(1)}, \ldots, T_y^{(v)}$ with $2\alpha_{y1}$ degrees of freedom. Then let

$$\begin{aligned} \mu_x^{(i)} &= \mu_{x1} + T_x^{(i)} \left(\frac{\beta_{x1}}{\alpha_{x1}\lambda_{x1}}\right)^{1/2}, \\ \mu_y^{(i)} &= \mu_{y1} + T_y^{(i)} \left(\frac{\beta_{y1}}{\alpha_{y1}\lambda_{y1}}\right)^{1/2}, \end{aligned}$$

for $i = 1, \ldots, v$. Then the values $\mu_x^{(i)} - \mu_y^{(i)}$ form a sample from the posterior distribution of $\mu_x - \mu_y$.

12. To the level of approximation in Eq. (12.2.7), we have

$$Z = g(E(Y), E(W)) + g_1(E(Y), E(W))[Y - E(Y)] + g_2(E(Y), E(W))[W - E(W)].$$

The variance of Z would then be

$$\begin{aligned} &g_1(E(Y), E(W))^2 \operatorname{Var}(Y) + g_2(E(Y), E(W))^2 \operatorname{Var}(W) \\ &\quad +2g_1(E(Y), E(W))g_2(E(Y), E(W)) \operatorname{Cov}(Y, W). \end{aligned} \tag{S.12.1}$$

Now substitute the entries of Σ for the variances and covariance.

13. The function g in this exercise is $g(y, w) = w - y^2$ with partial derivatives

$$\begin{aligned} g_1(y, w) &= 2y, \\ g_2(y, w) &= 1. \end{aligned}$$

In the formula for $\operatorname{Var}(Z)$ given in Exercise 12, make the following substitutions:

Exercise 12	This exercise
$E(Y)$	$\bar{Y}$
$E(W)$	$\bar{W}$
σ_{yy}	Z/v
σ_{ww}	V/v
σ_{yw}	C/v,

where Z, V, and C are defined in Example 12.2.10. The result is $[(2\bar{Y})^2 Z + V + 4\bar{Y}C]/v$, which simplifies to (12.2.3).

14. Let $Y_1, \ldots, Y_v$ be a large sample from the distribution of Y. Let $\bar{Y}$ be the sample average, and let V be the sample variance. For each i, define $W_i = (Y_i - \overline{Y})^3/V$. Estimate the skewness by the sample average of the W_i's. Use the sample variance to compute a simulation standard error to see if the simulation size is large enough.

15. (a) Since $S_u = S_0 \exp(\alpha u + W_u)$, we have that

$$E(S_u) = S_0 \exp(\alpha u) E\left(\exp(W_u)\right) = S_0 \exp(\alpha u)\psi(1).$$

In order for this mean to be $S_0 \exp(ru)$, it is necessary and sufficient that $\psi(1) = \exp(u[r - \alpha])$, or equivalently, $\alpha = r - \log(\psi(1))/u$.

(b) First, simulate lots (say v) of random variables $W^{(1)},\dots,W^{(v)}$ with the distribution of W_u. Define the function $h(s)$ as in Example 12.2.13. Define $Y^{(i)} = \exp(-ru)h(S_0\exp[\alpha u + W^{(i)}])$, where r is the risk free interest rate and α is the number found in part (a). The sample average of the $Y^{(i)}$'s would estimate the appropriate price for the option. One should compute a simulation standard error to see if the simulation size is large enough.

16. We can model our solution on Example 12.2.14. We should simulate a large number of operations of the queue up to time t. For each simulated operation of the queue, count how many customers are in the queue (including any being served). In order to simulation one instance of the queue operation up to time t, we can proceed as follows. Simulate interarrival times $X_1, X_2, \dots$ as exponential random variables with parameter λ. Define $T_j = \sum_{i=1}^{j} T_i$ for $j = 1, 2, \dots$. Stop simulating at the first k such that $T_k > t$. Start the queue with $W_0 = 0$, where W_j stands for the time that customer j leaves the queue. In what follows, $S_j \in \{1, 2\}$ will stand for which server serves customer j, and Z_j will stand for the time at which customer j begins being served.

For $j = 1,\dots,k-1$ and each $i < j$, define

$$U_{i,j} = \begin{cases} 1 & \text{if } W_i \ge T_j, \\ 0 & \text{otherwise.} \end{cases}$$

The number of customers in the queue when customer j arrives is $r = \sum_{i=0}^{j-1} U_{i,j}$.

- If $r = 0$, simulate U with a uniform distribution on the interval $[0, 1]$. Set $S_j = 1$ if $U < 1/2$ and $S_j = 2$ if $U \ge 1/2$. Set $Z_j = T_j$.
- If $r = 1$, find the value i such that $W_i \ge T_j$ and set $S_j = 2 - S_i$ so that customer j goes to the other server. Set $Z_j = T_j$.
- If $r \ge 2$, simulate U with a uniform distribution on the interval $[0, 1]$, and let customer j leave if $U < p_r$. If customer j leaves, set $W_j = T_j$. If customer j does not leave, find the second highest value $W_{i'}$ out of $W_1,\dots,W_{j-1}$ and set $S_j = S_{i'}$ and $Z_j = W_{i'}$.

For each customer that does not leave, simulate a service time Y_j having an exponential distribution with parameter μ_{S_j}, and set $W_j = Z_j + Y_j$. The number of customers in the queue at time t is the number of $j \in \{1,\dots,k-1\}$ such that $W_j \ge t$.

12.3 Simulating Specific Distributions

Commentary

This section is primarily of mathematical interest. Most distributions with which students are familiar can be simulated directly with existing statistical software. Instructors who wish to steer away from the theoretical side of simulation should look over the examples before skipping this section in case they contain some points that they would like to make. For example, a method is given for computing simulation standard error when the simulation result is an entire sample c.d.f. (see page 811). This relies on results from Sec. 10.6.

Solutions to Exercises

1. (a) Here we are being asked to perform the simulation outlined in the solution to Exercise 10 in Sec. 12.2 with $v_0 = 2000$ simulations. Each Y_i (in the notation of that solution) can be simulated

by taking a random variable U_i having uniform distribution on the interval $[0,1]$ and setting $Y_i = -\log(1-U_i)$. In addition to the run whose answers are in the back of the text, here are the results of two additional simulations: Approximation $= 0.0536$, sim. std. err. $= 0.0023$ and Approximation $= 0.0492$, sim. std. err. $= 0.0019$.

(b) For the two additional simulations in part (a), the value of v to achieve the desired goal are 706 and 459.

2. Let $V_i = a + (b-a)U_i$. Then the p.d.f. of V_i is easily seen to be $1/(b-a)$ for $a \le v \le b$, so V_i has the desired uniform distribution.

3. The c.d.f. corresponding to g_1 is

$$G_1(x) = \int_0^x \frac{1}{2t^{1/2}}dt = x^{1/2}, \text{ for } 0 \le x \le 1.$$

The quantile function is then $G_1^{-1}(p) = p^2$ for $0 < p < 1$. To simulate a random variable with the p.d.f. g_1, simulate U with a uniform distribution on the interval $[0,1]$ and let $X = U^2$. The c.d.f. corresponding to g_2 is

$$G_2(x) = \int_0^x \frac{1}{2(1-t)^{1/2}}dt = 1-(1-x)^{1/2}, \text{ for } 0 \le x \le 1.$$

The quantile function is then $G_2^{-1}(p) = 1-(1-p)^2$ for $0 < p < 1$. To simulate a random variable with the p.d.f. g_2, simulate U with a uniform distribution on the interval $[0,1]$ and let $X = 1-(1-U)^2$.

4. The c.d.f. of a Cauchy random variable is

$$F(x) = \int_{-\infty}^x \frac{dt}{\pi(1+t^2)} = \frac{1}{\pi}\left[\arctan(x) + \frac{\pi}{2}\right].$$

The quantile function is $F^{-1}(p) = \tan(\pi[p-1/2])$. So, if U has a uniform distribution on the interval $[0,1]$, then $\tan(\pi[U-1/2])$ has a Cauchy distribution.

5. The probability of acceptance on each attempt is $1/k$. Since the attempts (trials) are independent, the number of failures X until the first acceptance is a geometric random variable with parameter $1/k$. The number of iterations until the first acceptance is $X+1$. The mean of X is $(1-1/k)/(1/k) = k-1$, so the mean of $X+1$ is k.

6. (a) The c.d.f. of the Laplace distribution is

$$F(x) = \int_{-\infty}^x \frac{1}{2}\exp(-|t|)dt = \begin{cases} \frac{1}{2}\exp(x) & \text{if } x < 0, \\ 1-\frac{1}{2}\exp(-x) & \text{if } x \ge 0. \end{cases}$$

The quantile function is then

$$F^{-1}(p) = \begin{cases} \log(2p) & \text{if } 0 < p < 1/2, \\ -\log(2[1-p]) & \text{if } 1/2 \le p < 1. \end{cases}$$

Simulate a uniform random variable U on the interval $[0,1]$ and let $X = F^{-1}(U)$.

(b) Define

$$f(x) = \frac{1}{(2\pi)^{1/2}}\exp(-x^2/2),$$
$$g(x) = \frac{1}{2}\exp(-|x|).$$

We need to find a constant k such that $kg(x) \geq f(x)$ for all x. Equivalently, we need a constant c such that

$$c \geq \frac{\exp(-x^2/2)}{\exp(-|x|)}, \tag{S.12.2}$$

for all x. Then we can set $k = c(2/\pi)^{1/2}$. The smallest c that satisfies (S.12.2) is the supremum of $\exp(|x| - x^2/2)$. This function is symmetric around 0, so we can look for $\sup_{x\geq 0}\exp(x - x^2/2)$. To maximize this, we can maximize $x - x^2/2$ instead. The maximum of $x - x^2/2$ occurs at $x = 1$, so $c = \exp(1/2)$. Now, use acceptance/rejection with $k = \exp(1/2)(2/\pi)^{1/2} = 1.315$.

7. Simulate a random sample $X_1, \ldots, X_{11}$ from the standard normal distribution. Then $\sum_{i=1}^{4} X_i^2$ has the χ^2 distribution with 4 degrees of freedom and is independent of $\sum_{i=5}^{11} X_i^2$, which has the χ^2 distribution with 7 degrees of freedom. It follows that

$$\frac{7\sum_{i=1}^{4} X_i^2}{4\sum_{i=5}^{11} X_i^2}$$

the F distribution with 4 and 7 degrees of freedom.

8. (a) I did five simulations of the type requested and got the estimates 1.325, 1.385, 1.369, 1.306, and 1.329. There seems to be quite a bit of variability if we want three significant digits.
   (b) The five variance estimates were 1.333, 1.260, 1.217, 1.366, and 1.200.
   (c) The required sample sizes varied from 81000 to 91000, suggesting that we do not yet have a very precise estimate.

9. The simplest acceptance/rejection algorithm would use a uniform distribution on the interval $[0, 2]$. That is, let $g(x) = 0.5$ for $0 < x < 2$. Then $(4/3)g(x) \geq f(x)$ for all x, i.e. $k = 4/3$. We could simulate U and V both having a uniform distribution on the interval $[0, 1]$. Then let $X = 2V$ if $2f(2V) \geq (4/3)U$ and reject otherwise.

10. Using the prior distribution stated in the exercise, the posterior distributions for the probabilities of no relapse in the four treatment groups are

	Beta with parameters	
Group	α	β
Imipramine	23	19
Lithium	26	15
Combination	17	22
Placebo	11	25

We then simulate 5000 vectors of four beta random variables with the above parameters. Then we see what proportion of those 5000 vectors have the imipramine parameter the largest. We did five such simulations and got the proportions 0.1598, 0.1626, 0.1668, 0.1650, and 0.1650. The sample sizes required to achieve the desired accuracy are all around 5300.

11. The χ^2 distribution with m degrees of freedom is the same as the gamma distribution with parameters $m/2$ and $1/2$. So, we should simulate $Y^{(i)}$ having the χ^2 distribution with $n-p$ degrees of freedom and set $\tau^{(i)} = Y^{(i)}/S^2_{\text{Resid}}$.

12. We did a simulation of size $v = 2000$.

    (a) The plot of the sample c.d.f. of the $|\mu_x^{(i)} - \mu_y^{(i)}|$ values is in Fig. S.12.1.

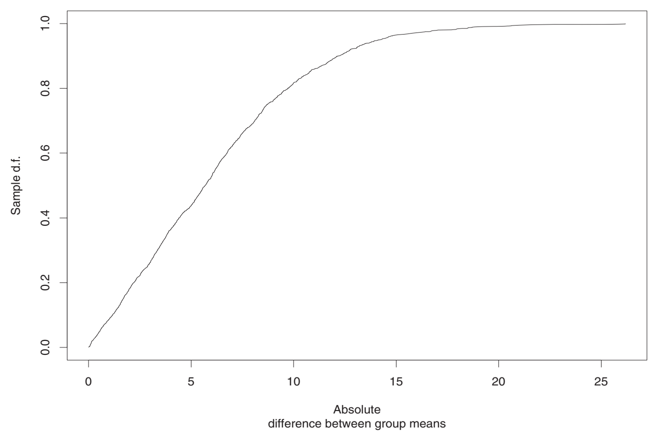

Figure S.12.1: Sample c.d.f. of $|\mu_x^{(i)} - \mu_y^{(i)}|$ values for Exercise 12a in Sec. 12.3.

    (b) The histogram of the ratios of calcium supplement precision to placebo precision is given in Fig. S.12.2. Only 12% of the simulated $\log(\tau_x^{(i)}/\tau_y^{(i)})$ were positive and 37% were less than -1.

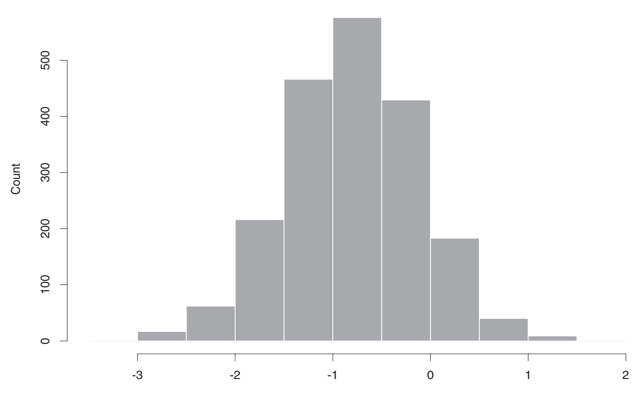

Figure S.12.2: Histogram of $\log(\tau_x^{(i)}/\tau_y^{(i)})$ values for Exercise 12b in Sec. 12.3.

There seems to be a sizeable probability that the two precisions (hence the variances) are unequal.

13. Let $X = F^{-1}(U)$, where F^{-1} is defined in Eq. (12.3.7) and U has a uniform distribution on the interval $[0, 1]$. Let G be the c.d.f. of X. We need to show that $G = F$, where F is defined in Eq. (12.3.6). Since F^{-1} only takes the values $t_1, \ldots, t_n$, it follows that G has jumps at those values and if flat everywhere else. Since F also has jumps at $t_1, \ldots, t_n$ and is flat everywhere else, we only need to show that $F(x) = G(x)$ for $x \in \{t_1, \ldots, t_n\}$. Let $q_n = 1$. Then $X \le t_i$ if and only if $U \le q_i$ for $i = 1, \ldots, n$. Since $\Pr(U \le q_i) = q_i$, it follows that $G(t_i) = q_i$ for $i = 1, \ldots, n$. That is, $G(x) = F(x)$ for $x \in \{t_1, \ldots, t_n\}$.

14. First, establish the Bonferroni inequality. Let $A_1, \ldots, A_k$ be events. Then

$$\Pr\left(\bigcap_{i=1}^k A_i\right) = 1 - \Pr\left(\bigcup_{i=1}^k A_i^c\right) \ge 1 - \sum_{i=1}^k \Pr(A_i^c) = 1 - \sum_{i=1}^k [1 - \Pr(A_i)].$$

Now, let $k = 3$ and

$$A_i = \{|G_{v,i}(x) - G_i(x)| \le 0.0082, \text{ for all } x\},$$

for $i = 1, 2, 3$. The event stated in the exercise is $\cap_{i=1}^3 A_i$. According to the arguments in Sec. 10.6,

$$\Pr\left(60000^{1/2}|G_{v,i}(x) - G(x)| \le 2, \text{ for all } x\right) \approx 0.9993.$$

Since $2/60000^{1/2} = 0.0082$, we have $\Pr(A_i) \approx 0.9993$ for $i = 1, 2, 3$. The Bonferroni inequality then says that $\Pr(\cap_{i=1}^3 A_i) \approx 0.9979$ or more.

15. The proof is exactly what the hint says. All joint p.d.f.'s should be considered joint p.f./p.d.f.'s and the p.d.f.'s of X and Y should be considered p.f.'s instead. The only integral over x in the proof is in the second displayed equation in the proof. The outer integral in that equation should be replaced by a sum over all possible x values. The rest of the proof is identical to the proof of Theorem 12.3.1.

16. Let $p_i = \exp(-\theta)\theta^i/(i!)$ for $i = 0, 1, \ldots$ and let $q_k = \sum_{i=1}^k p_i$. Let U have a uniform distribution on the interval $[0, 1]$. Let Y be the smallest k such that $U \le q_k$. Then Y has a Poisson distribution with mean θ.

17. Let $\{x_1, \ldots, x_m\}$ be the set of values that have positive probability under at least one of $g_1, \ldots, g_n$. That is, for each $j = 1, \ldots, m$ there is at least one i such that $g_i(x_j) > 0$ and for each $i = 1, \ldots, n$, $\sum_{j=1}^m g_i(x_j) = 1$. Then, the law of total probability says that

$$\Pr(X = x_j) = \sum_{i=1}^n \Pr(X = x_j | I = i) \Pr(I = i).$$

Since $\Pr(I = i) = 1/n$ for $i = 1, \ldots, n$ and $\Pr(X = x_j | I = i) = g_i(x_j)$, it follows that

$$\Pr(X = x_j) = \sum_{i=1}^n \frac{1}{n} g_i(x_j). \tag{S.12.3}$$

Since $x_1, \ldots, x_m$ are the only values that X can take, Eq. (S.12.3) specifies the entire p.f. of X and we see that Eq. (S.12.3) is the same as Eq. (12.3.8).

18. The Poisson probabilities with mean 5 from the table in the text are

x	0	1	2	3	4	5	6	7	8
$\Pr(X = x)$	.0067	.0337	.0842	.1404	.1755	.1755	.1462	.1044	.0653
x	9	10	11	12	13	14	15	16	
$\Pr(X = x)$	.0363	.0181	.0082	.0034	.0013	.0005	.0002	.0001	

where we have put the remainder probability under $x = 16$. In this case we have $n = 17$ different possible values. Since $1/17 = .0588$, we can use $x_1 = 0$ and $y_1 = 2$. Then $g_1(0) = .1139$ and $g_1(2) = .8861$. Then $f_1^*(2) = .8042 - (1 - .1139)/17 = .0321$. Next, take $x_2 = 1$ and $y_2 = 3$. Then $g_2(1) = .5729$ and $g_2(3) = .4271$. This makes $f_2^*(3) = .1153$. Next, take $x_3 = 2$ and $y_3 = 3$ so that $g_3(2) = .5453$, $g_3(3) = .4547$, and $f_3^*(3) = .0885$. Next, take $x_4 = 9$ and $y_4 = 3$, etc. The result of 16 such iterations is summarized in Table S.12.1.

Table S.12.1: Result of alias method in Exercise 18 of Sec. 12.3

i	x_i	$g_i(x_i)$	y_i	$g_i(y_i)$	i	x_i	$g_i(x_i)$	y_i	$g_i(y_i)$
1	0	.1139	2	.8861	10	13	.0221	5	.9779
2	1	.5729	3	.4271	11	14	.0085	5	.9915
3	2	.5453	3	.4547	12	5	.6246	6	.3754
4	9	.6171	3	.3829	13	15	.0034	6	.9966
5	10	.3077	3	.6923	14	16	.0017	6	.9983
6	3	.4298	4	.5702	15	6	.1151	7	.8849
7	11	.1394	4	.8606	16	7	.8899	8	.1101
8	12	.0578	4	.9422	17	8	1		
9	4	.6105	5	.3895					

The alias method is not unique. For example, we could have started with $x_1 = 1$ and $y_1 = 3$ or many other possible combinations. Each choice would lead to a different version of Table S.12.1.

19. For $k = 1, \ldots, n$, $I = k$ if and only if $k \le nY + 1 < k + 1$. Hence

$$\Pr(I = k) = \Pr\left(\frac{k-1}{n} \le Y \le \frac{k}{n}\right) = \frac{1}{n}.$$

The conditional c.d.f. of U given $I = k$ is

$$\begin{aligned}
\Pr(U \le t | I = k) &= \Pr(nY + 1 - I \le t | I = k) \\
&= \frac{\Pr(nY + 1 - k \le t, I = k)}{\Pr(I = k)} \\
&= \frac{\Pr\left(Y \le \frac{t+k-1}{n}, \frac{k-1}{n} \le Y \le \frac{k}{n}\right)}{1/n} \\
&= n \Pr\left(\frac{k-1}{n} \le Y < \frac{t+k-1}{n}\right) \\
&= t,
\end{aligned}$$

for $0 < t < 1$. So, the conditional distribution of U given $I = k$ is uniform on the interval $[0, 1]$ for all k. Since the conditional distribution is the same for all k, U and I are independent and the marginal distribution of U is uniform on the interval $[0, 1]$.

12.4 Importance Sampling

Commentary

Users of importance sampling might forget to check whether the importance function leads to a finite variance estimator. If the ratio of the function being integrated to the importance function is not bounded, one might have an infinite variance estimator. This doesn't happen in the examples in the text, but students should be made aware of the possibility. This section ends with an introduction to stratified importance sampling. This is an advanced topic that is quite useful, but might be skipped in a first pass. The last five exercises in this section introduce two additional variance reduction techniques, control variates and antithetic variates. These can be useful in many types of simulation problems, but those problems can be difficult to identify.

Solutions to Exercises

1. We want to approximate the integral $\int_a^b g(x)dx$. Suppose that we use importance sampling with f being the p.d.f. of the uniform distribution on the interval $[a, b]$. Then $g(x)/f(x) = (b-a)g(x)$ for $a < x < b$. Now, (12.4.1) is the same as (12.4.2).

2. First, we shall describe the second method in the exercise. We wish to approximate the integral $\int g(x)f(x)dx$ using importance sampling with importance function f. We should then simulate values $X^{(i)}$ with p.d.f. f and compute

$$Y^{(i)} = \frac{g(X^{(i)})f(X^{(i)})}{f(X^{(i)})} = g(X^{(i)}).$$

The importance sampling estimate is the average of the $Y^{(i)}$ values. Notice that this is precisely the same as the first method in the exercise.

3. (a) This is a distribution for which the quantile function is easy to compute. The c.d.f. is $F(x) = 1-(c/x)^{n/2}$ for $x > c$, so the quantile function is $F^{-1}(p) = c/(1-p)^{2/n}$. So, simulate U having a uniform distribution on the interval $[0, 1]$ and let $X = c/(1-U)^{2/n}$. Then X has the p.d.f. f.

(b) Let

$$a = \frac{\Gamma\left[\frac{1}{2}(m+n)\right] m^{m/2}n^{n/2}}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)}.$$

Then the p.d.f. of Y is $g(x) = ax^{(m/2)-1}/(mx+n)^{(m+n)/2}$, for $x > 0$. Hence,

$$\Pr(Y > c) = \int_c^\infty a\frac{x^{(m/2)-1}}{(mx+n)^{(m+n)/2}}dx.$$

We could approximate this by sampling lots of values $X^{(i)}$ with the p.d.f. f from part (a) and then averaging the values $g(X^{(i)})/f(X^{(i)})$.

(c) The ratio $g(x)/f(x)$ is, for $x > c$,

$$\frac{g(x)}{f(x)} = \frac{ax^{(m+n)/2}}{c^{n/2}(n/2)(mx+n)^{(m+n)/2}} = \frac{a}{c^{n/2}(n/2)(m+n/x)^{(m+n)/2}}.$$

This function is fairly flat for large x. Since we are only interested in $x > c$ in this exercise, importance sampling will have us averaging random variables $g(X^{(i)})/f(X^{(i)})$ that are nearly constant, hence the average should have small variance.

4. (a) If our 10000 exponentials are $X^{(1)}, \ldots, X^{(10000)}$, then our approximation is the average of the values $\log(1 + X^{(i)})$. In two example simulations, I got averages of 0.5960 and 0.5952 with simulation standard errors of 0.0042 both times.

   (b) Using importance sampling with the importance function being the gamma p.d.f. with parameters 1.5 and 1, I got estimates of 0.5965 and 0.5971 with simulation standard errors of 0.0012 both times.

   (c) The reason that the simulations in part (b) have smaller simulation standard error is that gamma importance function is a constant times $x^{1/2}\exp(-x)$. The ratio of the integrand to the importance function is a constant times $\log(1+x)x^{-1/2}$, which is nearly constant itself.

5. Let U have a uniform distribution on the interval $[0, 1]$, and let W be defined by Eq. (12.4.6). The inverse transformation is

$$u = \frac{\Phi\left(\dfrac{w-\mu_2}{\sigma_2}\right)}{\Phi\left(\dfrac{c_2-\mu_2}{\sigma_2}\right)}.$$

The derivative of the inverse transformation is

$$\frac{1}{(2\pi)^{1/2}\sigma_2\Phi\left(\dfrac{c_2-\mu_2}{\sigma_2}\right)}\exp\left(-\frac{1}{2\sigma_2^2}(w-\mu_2)^2\right). \tag{S.12.4}$$

Since the p.d.f. of U is constant, the p.d.f. of W is (S.12.4), which is the same as (12.4.5).

6. (a) We can simulate truncated normals as follows. If U has a uniform distribution on the interval $[0, 1]$, then $X = \Phi^{-1}(\Phi(1) + U[1 - \Phi(1)])$ has the truncated normal distribution in the exercise. If $X^{(1)}, \ldots, X^{(1000)}$ are our simulated values, then the estimate is the average of the $(1 - \Phi(1))X^{(i)2}$ values. Three simulations of size 1000 each produced the estimates 0.4095, 0.3878, and 0.4060.

   (b) If Y has an exponential distribution with parameter 0.5, and $X = (1 + Y)^{1/2}$, then we can find the p.d.f. of X. The inverse transformation is $y = x^2 - 1$ with derivative $2x$. The p.d.f. of X is then $2x0.5\exp(-0.5x^2 + 0.5)$. If $X^{(1)}, \ldots, X^{(1000)}$ are our simulated values, then the estimate is the average of the $X^{(i)}\exp(-0.5)/(2\pi)^{1/2}$ values. Three simulations of size 1000 each produced the estimates 0.3967, 0.3980, and 0.4016.

   (c) The simulation standard errors of the simulations in part (a) were close to 0.008, while those from part (b) were about 0.004, half as larger. The reason is that the random variables averaged in part (b) are closer to constant than those in part (a) since x is closer to constant than x^2.

7. (a) We can simulate bivariate normals by simulating one of the marginals first and then simulating the second coordinate conditional on the first one. For example, if we simulate $X_1^{(i)}$ $U^{(i)}$ as independent normal random variables with mean 0 and variance 1, we can simulate $X_2^{(i)} = 0.5X_1^{(i)} + U^{(i)}.75^{1/2}$. Three simulations of size 10000 each produced estimates of 0.8285, 0.8308, and 0.8316 with simulation standard errors of 0.0037 each time.

   (b) Using the method of Example 12.4.3, we did three simulations of size 10000 each and got estimates of 0.8386, 0.8387, and 0.8386 with simulation standard errors of about 3.4×10^{-5}, about 0.01 as large as those from part (a). Also, notice how much closer the three simulations are in part (b) compared to the three in part (a).

8. The random variables that are averaged to compute the importance sampling estimator are $Y^{(i)} = g(X^{(i)})/f(X^{(i)}$ where the $X^{(i)}$'s have the p.d.f. f. Since g/f is bounded, $Y^{(i)}$ has finite variance.

9. The inverse transformation is $v = F(x)$ with derivative $f(x)$. So, the p.d.f. of X is $f(x)/(b-a)$ for those x that can arise as values of $F^{-1}(V)$, namely $F^{-1}(a) < x < F^{-1}(b)$.

10. For part (a), the stratified importance samples can be found by replacing U in the formula used in Exercise 6(a) by $a + U(b-a)$ where (a, b) is one of the pairs $(0, .2)$, $(.2, .4)$, $(.4, .6)$, $(.6, .8)$, or $(.8, 1)$. For part (b), replace Y by $-\log(1-[a+U(b-a)])$ in the formula $X = (1+Y)^{1/2}$ using the same five (a, b) pairs. Three simulations using five intervals with 200 samples each produced estimates of 0.4018, 0.4029, and 0.2963 in part (a) and 0.4022, 0.4016, and 0.4012 in part (b). The simulation standard errors were about 0.0016 in part (a) and 0.0006 in part (b). Both parts have simulation standard errors about 1/5 or 1/6 the size of those in Exercise 6.

11. Since the conditional p.d.f. of X^* given $J = j$ is f_j, the marginal p.d.f. of X^* is

$$f^*(x) = \sum_{j=1}^{k} f_j(x) \Pr(J = j) = \frac{1}{k} \sum_{j=1}^{k} f_j(x).$$

Since $f_j(x) = kf(x)$ for $q_{j-1} \le x < q_j$, for each x there is one and only one $f_j(x) > 0$. Hence, $f^*(x) = f(x)$ for all x.

12. (a) The m.g.f. of a Laplace distribution with parameters 0 and σ is

$$\psi(t) = \int_{-\infty}^{\infty} \exp(tx) \frac{1}{2\sigma} \exp(-|x|/\sigma) dx.$$

The integral from $-\infty$ to 0 is finite if and only if $t > -1/\sigma$. The integral from 0 to ∞ is finite if and only if $t < 1/\sigma$. So the integral is finite if and only if $-1/\sigma < t < 1/\sigma$. The value of the integral is

$$\frac{1}{2\sigma}\left[\frac{1}{t+1/\sigma} + \frac{1}{-t+1/\sigma}\right] = \frac{1}{1-t^2\sigma^2}.$$

Plugging $\sigma^2 = u/100$ into this gives the expression in the exercise.

(b) With $u = 1$, $\psi(1) = 1/0.99$. With $r = 0.06$, we get $\alpha = 0.06 + \log(0.99) = 0.04995$. We ran three simulations of size 100000 each using the method described in the solution to Exercise 15 in Sec. 12.2. The estimated prices were S_0 times 0.0844, 0.0838, and 0.0843. The simulation standard errors were all about $3.6S_0 \times 10^{-4}$.

(c) $S_u > S_0$ if and only if $W_u > -\alpha u$, in this case $\alpha u = 0.04995$. The conditional c.d.f. of W_u given that $W_n > -0.04995$ is

$$F(w) = 1.4356 \begin{cases} 0.5[\exp(10w) - 0.6068] & \text{if } -0.04995 < w \le 0, \\ 1 - 0.5[\exp(-10w) + 0.6068] & \text{if } w > 0. \end{cases}$$

The quantile function is then

$$F^{-1}(p) = \begin{cases} 0.1 \log(1.3931p + 0.6068)) & \text{if } 0 < p < 0.2822, \\ -0.1 \log(2[1 - 0.6966p] - 0.6068) & \text{if } 0.2822 \le p < 1. \end{cases}$$

When we use samples from this conditional distribution, we need to divide the average by 1.4356, which is the ratio of the conditional p.d.f. to the unconditional p.d.f. We ran three more simulations of size 100000 each and got estimates of S_0 times 0.0845, 0.0846, and 0.0840 with simulation standard errors of about $2.66S_0 \times 10^{-4}$. The simulation standard error is only a little smaller than it was in part (b).

13. (a) $E(Z) = E(Y^{(i)}) + kc = E(W^{(i)}) - kE(V^{(i)}) + kc$. By the usual importance sampling argument, $E(W^{(i)}) = \int g(x)dx$ and $E(V^{(i)}) = c$, so $E(Z) = \int g(x)dx$.

(b) $\text{Var}(Z) = [\sigma_W^2 + k^2\sigma_V^2 - 2k\rho\sigma_W\sigma_V]$. This is a quadratic in k that is minimized when $k = \rho\sigma_W/\sigma_V$.

14. (a) We know that $\int_0^1 (1+x^2)^{-1}dx = \pi/4$. We shall use $f(x) = \exp(-x)/(1-\exp(-1))$ for $0 < x < 1$. We shall simulate $X^{(1)}, \ldots, X^{(10000)}$ with this p.d.f. and compute

$$\begin{aligned} W^{(i)} &= \frac{1-\exp(-1)}{1+X^{(i)2}}, \\ V^{(i)} &= \frac{\exp[X^{(i)}](1-\exp(-1))}{1+X^{(i)2}}. \end{aligned}$$

We ran three simulations of 10000 each and got estimates of the integral equal to 0.5248, 0.5262, and 0.5244 with simulation standard errors around 0.00135. This compares to 0.00097 in Example 12.4.1. We shall see what went wrong in part (b).

(b) We use the samples in our simulation to estimate σ_W at 0.0964, σ_Z at 0.0710, and ρ at -0.8683. Since the correlation appears to be negative, we should have used a negative value of k to multiply our control variate. Based on our estimates, we might use $k = -1.1789$. Additional simulations using this value of k produce simulation standard errors around 4.8×10^{-4}.

15. (a) Since $U^{(i)}$ and $1 - U^{(i)}$ both have uniform distributions on the interval $[0, 1]$, $X^{(i)} = F^{-1}(U^{(i)})$ and $T^{(i)} = F^{-1}(1 - U^{(i)})$ have the same distribution.

(b) Since $X^{(i)}$ and $T^{(i)}$ have the same distribution, so do $W^{(i)}$ and $V^{(i)}$, so the means of $W^{(i)}$ and $V^{(i)}$ are both the same and they are both $\int g(x)dx$, according to the importance sampling argument.

(c) Since F^{-1} is a monotone increasing function, we know that $X^{(i)}$ and $T^{(i)}$ are decreasing functions of each other. If $g(x)/f(x)$ is monotone, then $W^{(i)}$ and $V^{(i)}$ will also be decreasing functions of each other. As such they ought to be negatively correlated since one is small when the other is large.

(d) $\text{Var}(Z) = \text{Var}(Y^{(i)})/v$, and

$$\text{Var}(Y^{(i)}) = 0.25[\text{Var}(W^{(i)}) + \text{Var}(V^{(i)}) + 2\,\text{Cov}(W^{(i)}, V^{(i)})] = 0.5(1+\rho)\,\text{Var}(W^{(i)}).$$

Without antithetic variates, we get a variance of $\text{Var}(W^{(i)})/[2v]$. If $\rho < 0$, then $0.5(1+\rho) < 0.5$ and $\text{Var}(Z)$ is smaller than we get without antithetic variates.

16. Using the method outlined in Exercise 15, we did three simulations of size 5000 each and got estimates of 0.5250, 0.5247, and 0.5251 with estimates of $\text{Var}(Y^{(i)})^{1/2}$ of about 0.0238, approximately 1/4 of $\hat{\sigma}_3$ from Example 12.4.1.

17. In Exercise 3(c), $g(x)/f(x)$ is a monotone function of x, so antithetic variates should help. In Exercise4(b), we could use control variates with $h(x) = \exp(-x)$. In Exercises 6(a) and 6(b) the ratios $g(x)/f(x)$ are monotone, so antithetic variates should help. Control variates with $h(x) = x\exp(-x^2/2)$ could also help in Exercise 6(a). Exercise 10 involves the same function, so the same methods could also be used in the stratified versions.

12.5 Markov Chain Monte Carlo

Commentary

Markov chain Monte Carlo (MCMC) is primarily used to simulate parameters in a Bayesian analysis. Implementing Gibbs sampling in all but the simplest problems is generally a nontrivial programming task. Instructors should keep this in mind when assigning exercises. The less experience students have had with programming, the more help they will need in implementing Gibbs sampling. The theoretical justification given in the text relies on the material on Markov chains from Sec. 3.10, which might have been skipped earlier in the course. This material is not necessary for actually performing MCMC.

If one is using the software *R*, there is no substitute for old-fashioned programming. (There is a package called *BUGS*:
http://www.mrc-bsu.cam.ac.uk/bugs/ but I will not describe it here.) After the solutions, there is *R* code to do the calculations in Examples 12.5.6 and 12.5.7 in the text.

Solutions to Exercises

1. The conditional p.d.f. of X_2 given $X_2 = x_2$ is

$$g_1(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{cg(x_1, x_2)}{f_2(x_2)} = \frac{c}{f_2(x_2)}h_2(x_1).$$

Let $c_2 = c/f_2(x_2)$, which does not depend on x_1.

2. Let $f_2(x_2) = \int f(x_1, x_2)dx_1$ stand for the marginal p.d.f. of X_2, and let $g_1(x_1|x_2) = f(x_1, x_2)/f_2(x_2)$ stand for the conditional p.d.f. of $X_1^{(i)}$ given $X_2^{(i)} = x_2$. We are supposing that $X_2^{(i)}$ has the marginal distribution with p.d.f. f_2. In step 2 of the Gibbs sampling algorithm, after $X_2^{(i)} = x_2$ is observed, $X_1^{(i+1)}$ is sampled from the distribution with p.d.f. $g_a(x_1|x_2)$. Hence, the joint p.d.f. of $(X_1^{(i+1)}, X_2^{(i)})$ is $f_2(x_2)g_1(x_1, x_2) = f(x_1, x_2)$. In particular $X_1^{(i+1)}$ has the same marginal distribution as X_1, and the same argument we just gave (with subscripts 1 and 2 switched and applying step 3 instead of 2 in the Gibbs sampling algorithm) shows that $(X_1^{(i+1)}, X_2^{(i+1)})$ has the same joint distribution as $(X_1^{(i)}, X_2^{(i)})$.

3. Let $h(z)$ stand for the p.f. or p.d.f. of the stationary distribution and let $g(z|z')$ stand for the conditional p.d.f. or p.f. of Z_{i+1} given $Z_i = z'$, which is assumed to be the same for all i. Suppose that Z_i has the stationary distribution for some i, then (Z_i, Z_{i+1}) has the joint p.f. or p.d.f. $h(z_i)g(z_{i+1}|z_i)$. Since Z_1 does have the stationary distribution, (Z_1, Z_2) has the joint p.f. or p.d.f. $h(z_1)g(z_2|z_1)$. Hence, (Z_1, Z_2) has the same distribution as (Z_i, Z_{i+1}) whenever Z_i has the stationary distribution. The proof is complete if we can show that Z_i has the stationary distribution for every i. We shall show this by induction. We know that it is true for $i = 1$ (that is, Z_1 has the stationary distribution). Assume that each of $Z_1, \dots, Z_k$ has the stationary distribution, and prove that Z_{k+1} has the stationary distribution. Since h is the p.d.f. or p.f. of the stationary distribution, it follows that the marginal p.d.f. or p.f. of Z_{k+1} is $\int h(z_k)g(z_{k+1}|z_k)dz_k$ or $\sum_{\text{All } z_k} h(z_k)g(z_{k+1}|z_k)$, either of which is $h(z_{k+1})$ by the definition of stationary distribution. Hence Z_{k+1} also has the stationary distribution, and the induction proof is complete.

4. $\text{Var}(\overline{X}) = \sigma^2/n$ and

$$\text{Var}(\overline{Y}) = \frac{\sigma^2}{n} + \frac{1}{n^2}\sum\sum_{i \neq j} \text{Cov}(Y_i, Y_j).$$

Since $\mathrm{Cov}(Y_i, Y_j) > 0$, $\mathrm{Var}(\overline{Y}) > \sigma^2/n = \mathrm{Var}(\overline{X})$.

5. The sample average of all 30 observations is 1.442, and the value of s_n^2 is 2.671. The posterior hyperparameters are then

$$\alpha_1 = 15.5,\ \lambda_1 = 31,\ \mu_1 = 1.4277,\ \text{and}\ \beta_1 = 1.930.$$

The method described in Example 12.5.1 says to simulate values of μ having the normal distribution with mean 1.4277 and variance $(31\tau)^{-1}$ and to simulate values of τ having the gamma distribution with parameters 16 and $1.930+0.5(\mu-1.4277)^2$. In my particular simulation, I used five Markov chains with the following starting values for μ: 0.4, 1.0, 1.4, 1.8, and 2.2. The convergence criterion was met very quickly, but we did 100 burn-in anyway. The estimated mean of $(\sqrt{\tau}\mu)^{-1}$ was 0.2542 with simulation standard error 4.71×10^{-4}.

6. The data summaries that we need to follow the pattern of Example 12.5.4 are the following:

$$\begin{array}{lll} \overline{x}_1 = 12.5 & \overline{x}_2 = 47.89 & \overline{y} = 2341.4 \\ s_{11} = 5525 & s_{12} = 16737 & s_{22} = 61990.47 \\ s_{1y} = 927865 & s_{2y} = 3132934 & s_{yy} = 169378608, \end{array}$$

and $n = 26$.

(a) The histogram of $|\beta_1^{(\ell)}|$ values is in Fig. S.12.3.

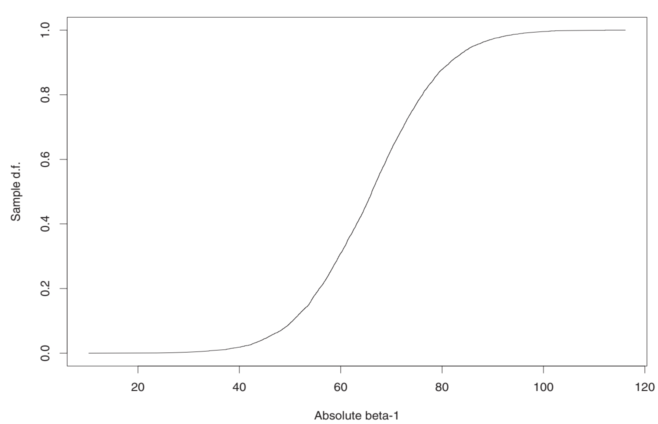

Figure S.12.3: Sample c.d.f. of $|\beta_1^{(\ell)}|$ values for Exercise 6a in Sec. 12.5.

(b) i. The histogram of $\beta_0^{(\ell)} + 26\beta_1^{(\ell)} + 67.2\beta_2^{(\ell)}$ values is in Fig. S.12.4.

ii. Let $\boldsymbol{z}' = (1, 26, 67.2)$ as in Example 11.5.7 of the text. To create the predictions, we take each of the values in the histogram in Fig. S.12.4 and add a pseudo-random normal variable to each with mean 0 and variance

$$\left[1 + \boldsymbol{z}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}\right]^{1/2} \tau^{(\ell)-1/2}.$$

We then use the sample 0.05 and 0.95 quantiles as the endpoints of our interval. In three separate simulations, I got the following intervals (3652, 5107), (3650, 5103), and (3666, 5131). These are all slightly wider than the interval in Example 11.5.7.

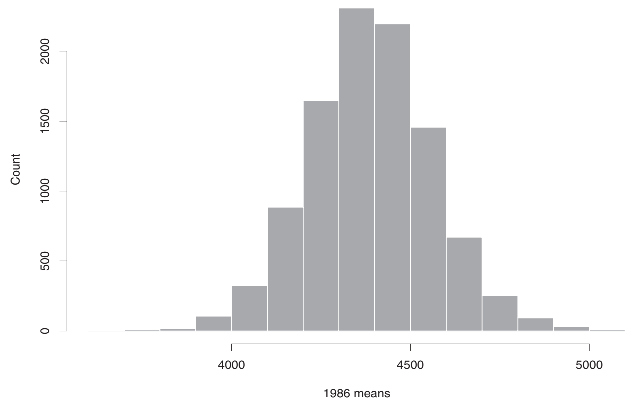

Figure S.12.4: Histogram of $\beta_0^{(\ell)} + 26\beta_1^{(\ell)} + 67.2\beta_2^{(\ell)}$ values for Exercise6(b)i in Sec. 12.5.

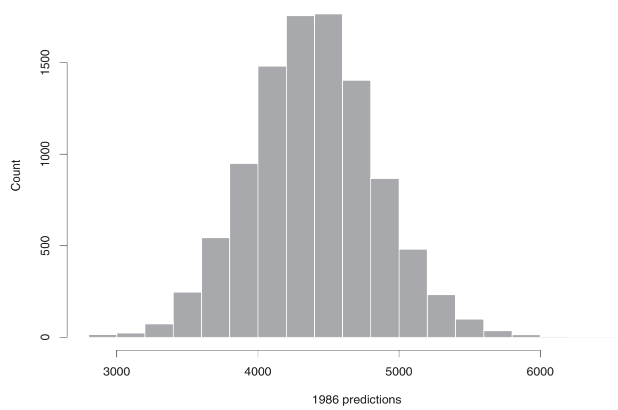

Figure S.12.5: Histogram of predicted values for 1986 sales in Exercise 6(b)iii in Sec. 12.5.

iii. The histogram of the sales figures used in Exercise 6(b)ii is in Fig. S.12.5. This histogram has more spread in it than the one in Fig. S.12.4 because the 1986 predictions equal the 1986 parameters plus independent random variables (as described in part (b)ii). The addition of the independent random variables increases the variance.

7. There are $n_i = 6$ observations in each of $p = 3$ groups. The sample averages are 825.83, 845.0, and 775.0. The w_i values are 570.83, 200.0, and 900.0. In three separate simulations of size 10000 each, I got the following three vectors of posterior mean estimates: (826.8, 843.2, 783.3), (826.8, 843.2, 783.1), and (826.8, 843.2, 783.2).

8. (a) To prove that the two models are the same, we need to prove that we get Model 1 when we integrate $\tau_1, \ldots, \tau_n$ out of Model 2. Since the τ_i's are independent, the Y_i's remain independent after integrating out the τ_i's. In Model 2, $[Y_i - (\beta_0 + \beta_1 x_i)]\tau_i^{1/2}$ has the standard normal distribution given τ_i, and is therefore independent of τ_i. Also, $\tau_i a\sigma^2$ has the χ^2 distribution with a degrees of

freedom, so

$$\frac{[Y_i - (\beta_0 + \beta_1 x_i)]\tau_i^{1/2}}{\tau_i^{1/2}\sigma}$$

has the t distribution with a degrees of freedom, which is the same as Model 1.

(b) The prior p.d.f. is a constant times

$$\eta^{b/2-1}\exp(-f\eta/2)\prod_{i=1}^{n}\left[\tau_i^{a/2-1}\eta^{a/2}\exp(-a\eta\tau_i/2)\right],$$

while the likelihood is

$$\prod_{i=1}^{n}\left[\tau_i^{1/2}\exp(-[y_i-\beta_0-\beta_1 x_i]^2\tau_i/2)\right].$$

The product of these two produces Eq. (12.5.4).

(c) As a function of η, we have η to the power $(na+b)/2-1$ times e to the power of $-\eta/2$ times $f + a\sum_{i=1}^{n}\tau_i$. This is, aside from a constant factor, the p.d.f. of the gamma distribution with parameters $(na+b)/2$ and $(f+a\sum_{i=1}^{n}\tau_i)/2$. As a function of τ_i, we have τ_i to the power $(a+1)/2-1$ times e to the power $-\tau_i[a\eta+(y_i-\beta_0-\beta_1 x_i)^2]/2$, which is (aside from a constant factor) the p.d.f. of the gamma distribution with parameters $(a+1)/2$ and $[a\eta+(y_i-\beta_0-\beta_1 x_i)^2]/2$. As a function of β_0, we have a constant times e to the power

$$-\sum_{i=1}^{n}\tau_i(\beta_0-[y_i-\beta_1 x_i])^2/2 = -\frac{1}{2}\sum_{i=1}^{n}\tau_i\left(\beta_0-\frac{\sum_{i=1}^{n}\tau_i(y_i-\beta_1 x_i)}{\sum_{i=1}^{n}\tau_i}\right)^2 + c,$$

where c does not depend on β_0. (Use the method of completing the square.) This is a constant times the p.d.f. of the normal distribution stated in the exercise. Completing the square as a function of β_1 produces the result stated for β_1 in the exercise.

9. In three separate simulations of size 10000 each I got posterior mean estimates for (β_0,β_1,η) of $(-0.9526, 0.02052, 1.124\times10^{-5})$, $(-0.9593, 0.02056, 1.143\times10^{-5})$, and $(-0.9491, 0.02050, 1.138\times10^{-5})$. It appears we need more than 10000 samples to get a good estimate of the posterior mean of β_0. The estimated posterior standard deviations from the three simulations were $(1.503\times10^{-2}, 7.412\times10^{-5}, 7.899\times10^{-6})$, $(2.388\times10^{-2}, 1.178\times10^{-4}, 5.799\times10^{-6})$, and $(2.287\times10^{-2}, 1.274\times10^{-4}, 6.858\times10^{-6})$.

10. Let the proper prior have hyperparameters μ_0, λ_0, α_0, and β_0. Conditional on the Y_i's, those X_i's that have $Y_i = 1$ are an i.i.d. sample of size $\sum_{i=1}^{n}Y_i$ from the normal distribution with mean μ and precision τ.

(a) The conditional distribution of μ given all else is the normal distribution with mean equal to $\dfrac{\mu_0+\sum_{i=1}^{n}Y_iX_i}{\lambda_0+\sum_{i=1}^{n}Y_i}$, and precision equal to $\tau\sum_{i=1}^{n}Y_i$.

(b) The conditional distribution of τ given all else is the gamma distribution with parameters $\alpha_0+\sum_{i=1}^{n}Y_i/2+1/2$ and

$$\beta_0+\frac{1}{2}\left[\lambda_0(\mu-\mu_0)^2+\sum_{i=1}^{n}Y_i(X_i-\mu)^2\right].$$

(c) Given everything except Y_i,

$$\Pr(Y_i = 1) = \frac{\tau^{1/2}\exp\left(-\frac{\tau}{2}[X_i-\mu]^2\right)}{\tau^{1/2}\exp\left(-\frac{\tau}{2}[X_i-\mu]^2\right)+\exp\left(-\frac{1}{2}X_i^2\right)}.$$

(d) To use Gibbs sampling, we need starting values for all but one of the unknowns. For example, we could randomly assign the data values to the two distributions with probabilities 1/2 each or randomly split the data into two equal-sized subsets. Given starting values for the Y_i's, we could start μ and τ at their posterior means given the observations that came from the distribution with unknown parameters. We would then cycle through simulating random variables with the distributions in parts (a)–(c). After burn-in and a large simulation run, estimate the the posterior means by the averages of the sample parameters in the large simulation run.

(e) The posterior mean of Y_i is the posterior probability that $Y_i = 1$. Since $Y_i = 1$ is the same as the event that X_i came from the distribution with unknown mean and variance, the posterior mean of Y_i is the posterior probability that X_i came from the distribution with unknown mean and variance.

11. For this exercise, I ran five Markov chains for 10000 iterations each. For each iteration, I obtain a vector of 10 Y_i values. Our estimated probability that X_i came from the distribution with unknown mean and variance equals the average of the 50000 Y_i values for each $i = 1, \dots, 10$. The ten estimated probabilities for each of my three runs are listed below:

Run	Estimated Probabilities									
1	0.291	0.292	0.302	0.339	0.370	0.281	0.651	0.374	0.943	0.816
2	0.285	0.286	0.302	0.339	0.375	0.280	0.656	0.371	0.945	0.819
3	0.283	0.286	0.301	0.340	0.373	0.280	0.651	0.370	0.945	0.820

12. Note that γ_0 should be the precision rather than the variance of the prior distribution of μ.

(a) The prior p.d.f. times the likelihood equals a constant times

$$\tau^{n/2}\exp\left(-\frac{\tau}{2}\left\{n[\overline{x}_n-\mu]^2+s_n^2\right\}\right)\exp\left(-\frac{\gamma_0}{2}[\mu-\mu_0]^2\right)\tau^{\alpha_0-1}\exp\left(-\frac{\tau\beta_0}{2}\right),$$

where $s_n^2 = \sum_{i=1}^n (x_i - \overline{x}_n)^2$. As a function of τ this looks like the p.d.f. of the gamma distribution with parameters $\alpha_0 + n/2$ and $[n(\overline{x}_n - \mu)^2 + s_n^2 + \beta_0]/2$. As a function of μ, (by completing the square) it looks like the p.d.f. of the normal distribution with mean $(n\tau\overline{x}_n + \mu_0\gamma_0)/(n\tau + \gamma_0)$ and variance $1/(n\tau + \gamma_0)$.

(b) The data summaries are $n = 18$, $\overline{x}_n = 182.17$, and $s_n^2 = 88678.5$. I ran five chains of length 10000 each for three separate simulations. For each simulation, I obtained 50000 parameter pairs. To obtain the interval, I sorted the 50000 μ values and chose the 1250th and 48750th values. For the three simulations, I got the intervals $(154.2, 216.2)$, $(154.6, 216.5)$, and $(154.7, 216.2)$.

13. In part (a), the exponent in the displayed formula should have been $-1/2$.

(a) The conditional distribution of $(\mu - \mu_0)\gamma^{1/2}$ given γ is standard normal, hence it is independent of γ. Also, the distribution of $2b_0\gamma$ is the χ^2 distribution with $2a_0$ degrees of freedom. It follows that $(\mu - \mu_0)/(b_0/a_0)^{1/2}$ has the t distribution with $2a_0$ degrees of freedom.

(b) The marginal prior distributions of τ are in the same form with the same hyperparameters in Exercise 12 and in Sec. 8.6. The marginal prior distributions of μ are in the same form also, but the hyperparameters are not identical. We need $a_0 = \alpha_0$ to make the degrees of freedom match, and we need $b_0 = \beta_0/\lambda_0$ in order to make the scale factor match.

(c) The prior p.d.f. times the likelihood equals a constant times

$$\tau^{\alpha_0+n/2-1}\gamma^{a_0+1/2-1}\exp\left(-\frac{\tau}{2}\left\{n[\overline{x}_n-\mu]^2+s_n^2+\beta_0\right\}-\frac{\gamma}{2}[\mu-\mu_0]^2+\gamma b_0\right).$$

As a function of τ this is the same as in Exercise 12. As a function of μ, it is also the same as Exercise 12 if we replace γ_0 by γ. As a function of γ, it looks like the p.d.f. of the gamma distribution with parameters $a_0+1/2$ and $b_0+(\mu-\mu_0)/2$.

(d) This time, I ran 10 chains of length 10000 each for three different simulations. The three intervals are found by sorting the μ values and using the 2500th and 97500th values. The interval are $(154.4, 216.3)$, $(154.6, 215.8)$, and $(154.4, 215.9)$.

14. The exercise should have included that the prior hyperparameters are $\alpha_0 = 0.5$, $\mu_0 = 0$, $\lambda_0 = 1$, and $\beta_0 = 0.5$.

(a) I used 10 chains of length 10000 each.

(b) The histogram of predicted values is in Fig. S.12.6. There are two main differences between this

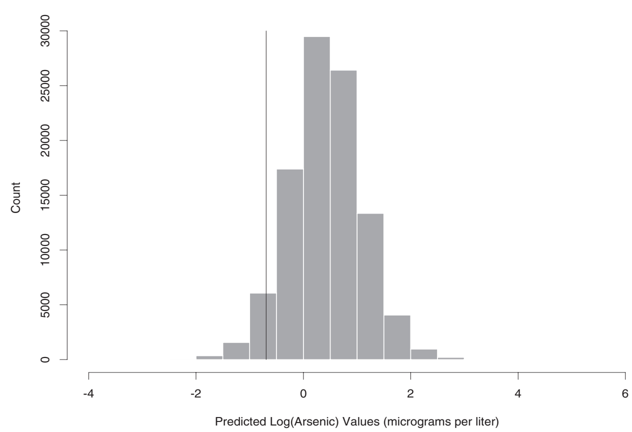

Figure S.12.6: Histogram of Log-arsenic predictions for Exercise 14b in Sec. 12.5.

histogram and the one in Fig. 12.10 in the text. First, the distribution of log-arsenic is centered at slightly higher values in this histogram. Second, the distribution is much less spread out in this histogram. (Notice the difference in horizontal scales between the two figures.)

(c) The median of predicted arsenic concentration is 1.525 in my simulation, compared to the smaller value 1.231 in Example 12.5.8, about 24% higher.

15. (a) For each censored observation X_{n+i}, we observe only that $X_{n+i} \le c$. The probability of $X_{n+i} \le c$ given θ is $1-\exp(-c\theta)$. The likelihood times prior is a constant times

$$\theta^{n+\alpha-1}[1-\exp(-c\theta)]^m\exp\left(-\theta\sum_{i=1}^n x_i\right). \qquad \text{(S.12.5)}$$

We can treat the unobserved values $X_{n+1},\dots,X_{n+m}$ as parameters. The conditional distribution of X_{n+i} given θ and given that $X_{n+i}\le c$ has the p.d.f.

$$g(x|\theta)=\frac{\theta\exp(-\theta x)}{1-\exp(-\theta c)},\ \text{for } 0<x<c. \tag{S.12.6}$$

If we multiply the conditional p.d.f. of $(X_{n+1},\dots,X_{n+m})$ given θ times Eq. (S.12.5), we get

$$\theta^{n+m+\alpha-1}\exp\left(-\theta\sum_{i=1}^{n+m}x_i\right),$$

for $\theta>0$ and $0<x_i<c$ for $i=n+1,\dots,n+m$. As a function of θ, this looks like the p.d.f. of the gamma distribution with parameters $n+m+\alpha$ and $\sum_{i=1}^{n+m}x_i$. As a function of X_{n+i}, it looks like the p.d.f. in Eq. (S.12.6). So, Gibbs sampling can work as follows. Pick a starting value for θ, such as one over the average of the uncensored values. Then simulate the censored observations with p.d.f. (S.12.6). This can be done using the quantile function

$$G^{-1}(p)=-\frac{\log(1-p[1-\exp(-c\theta)])}{\theta}.$$

Then, simulate a new θ from the gamma distribution mentioned above to complete one iteration.

(b) For each censored observation X_{n+i}, we observe only that $X_{n+i}\ge c$. The probability of $X_{n+i}\ge c$ given θ is $\exp(-c\theta)$. The likelihood times prior is a constant times

$$\theta^{n+\alpha-1}\exp\left(-\theta\left[mc+\sum_{i=1}^{n}x_i\right]\right). \tag{S.12.7}$$

We could treat the unobserved values $X_{n+1},\dots,X_{n+m}$ as parameters. The conditional distribution of X_{n+i} given θ and given that $X_{n+i}\ge c$ has the p.d.f.

$$g(x|\theta)=\theta\exp(-\theta[x-c]),\ \text{for } x>c. \tag{S.12.8}$$

If we multiply the conditional p.d.f. of $(X_{n+1},\dots,X_{n+m})$ given θ times Eq. (S.12.7), we get

$$\theta^{n+m+\alpha-1}\exp\left(-\theta\sum_{i=1}^{n+m}x_i\right),$$

for $\theta>0$ and $x_i>c$ for $i=n+1,\dots,n+m$. As a function of θ, this looks like the p.d.f. of the gamma distribution with parameters $n+m+\alpha$ and $\sum_{i=1}^{n+m}x_i$. As a function of X_{n+i}, it looks like the p.d.f. in Eq. (S.12.8). So, Gibbs sampling can work as follows. Pick a starting value for θ, such as the M.L.E., $\dfrac{n+m}{mc+\sum_{i=1}^{n}x_i}$. Then simulate the censored observations with p.d.f. (S.12.8). This can be done using the quantile function

$$G^{-1}(p)=c-\frac{\log(1-p)}{\theta}.$$

Then, simulate a new θ from the gamma distribution mentioned above to complete one interaction. In this part of the exercise, Gibbs sampling is not really needed because the posterior distribution of θ is available in closed form. Notice that (S.12.7) is a constant times the p.d.f. of the gamma distribution with parameters $n+\alpha$ and $mc+\sum_{i=1}^{n}x_i$, which is then the posterior distribution of θ.

16. (a) The joint p.d.f. of (X_i, Z_i) can be found from the joint p.d.f. of (X_i, Y_i) and the transformation $h(x, y) = (x, x + y)$. The joint p.d.f. of (X_i, Y_i) is $f(x, y) = \lambda\mu \exp(-x\lambda - y\mu)$ for $x, y > 0$. The inverse transformation is $h^{-1}(x, z) = (x, z - x)$, with Jacobian equal to 1. So, the joint p.d.f. of $(X_i Z_i)$ is

$$g(x, z) = f(x, z - x) = \lambda\mu \exp(-x[\lambda - \mu] - z\mu), \text{ for } 0 < x < z,\ z > 0.$$

The marginal p.d.f. of Z_i is the integral of this over x, namely

$$g_2(z) = \frac{\lambda\mu}{\lambda - \mu} [1 - \exp(-z[\lambda - \mu])] \exp(-z\mu),$$

for $z > 0$. The conditional p.d.f. of X_i given $Z_i = z$ is the ratio

$$\frac{g(x, z)}{g_2(z)} = \frac{\lambda - \mu}{1 - \exp(-z[\lambda - \mu])} \exp(-x[\lambda - \mu]), \text{ for } 0 < x < z. \quad \text{(S.12.9)}$$

The conditional c.d.f. of X_i given $Z_i = z$ is the integral of this, which is the formula in the text.

(b) The likelihood times prior is

$$\frac{\lambda^{n+a-1}\mu^{n+b-1}}{(\lambda - \mu)^{n-k}} \exp\left(-\lambda \sum_{i=1}^{k} x_i - \mu \sum_{i=1}^{k} y_i\right) \prod_{i=k+1}^{n} [1 - \exp(-[\lambda - \mu] z_i)]. \quad \text{(S.12.10)}$$

We can treat the unobserved pairs (X_i, Y_i) for $i = k + 1, \ldots, n$ as parameters. Since we observe $X_i + Y_i = Z_i$, we shall just treat X_i as a parameter. The conditional p.d.f. of X_i given the other parameters and Z_i is in (S.12.9). Multiplying the product of those p.d.f.'s for $i = k + 1, \ldots, n$ times (S.12.10) gives

$$\lambda^{n+a-1}\mu^{n+b-1} \exp\left(-\lambda \sum_{i=1}^{n} x_i - \mu \left[\sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} (z_i - x_i)\right]\right), \quad \text{(S.12.11)}$$

where $0 < x_i < z_i$ for $i = k + 1, \ldots, n$. As a function of λ, (S.12.11) looks like the p.d.f. of the gamma distribution with parameters $n + a$ and $\sum_{i=1}^{n} x_i$. As a function of μ it looks like the p.d.f. of the gamma distribution with parameters $n + b$ and $\sum_{i=1}^{n} y_i$. As a function of x_i $(i = k + 1, \ldots, n)$, it looks like the p.d.f. in (S.12.9). So, Gibbs sampling can work as follows. Pick starting values for μ and λ, such as one over the averages of the observed values of the X_i's and Y_i's. Then simulate the unobserved X_i values for $i = k + 1, \ldots, n$ using the probability integral transform. Then simulate new λ and μ values using the gamma distributions mentioned above to complete one iteration.

12.6 The Bootstrap

Commentary

The bootstrap has become a very popular technique for solving non-Bayesian problems that are not amenable to analysis. The nonparametric bootstrap can be implemented without much of the earlier material in this chapter. Indeed, one need only know how to simulate from a discrete uniform distribution (Example 12.3.11) and compute simulation standard errors (Sec. 12.2).

The software R has a function `boot` that is available after issuing the command `library(boot)`. The first three arguments to `boot` are a vector `data` containing the original sample, a function `f` to compute the statistic whose distribution is being bootstrapped, and the number of bootstrap samples to create. For the nonparametric bootstrap, the function `f` must have at least two arguments. The first will always be `data`, and the second will be a vector `inds` of integers of the same dimension as `data`. This vector `inds` will choose the bootstrap sample. The function should return the desired statistic computed from the sample `data[inds]`. Any additional arguments to `f` can be passed to `boot` by setting them explicitly at the end of the argument list. For the parametric bootstrap, `boot` needs the optional arguments `sim="parametric"` and `ran.gen`. The function `ran.gen` tells how to generate the bootstrap samples, and it takes two arguments. The first argument will be `data`. The second argument is anything else that you need to generate the samples, for example, estimates of parameters based on the orginal data. Also, `f` needs at least one argument which will be a simulated data set. Any additional arguments can be passed explicitly to `boot`.

Solutions to Exercises

1. We could start by estimating θ by the M.L.E., $1/\overline{X}$. Then we would use the exponential distribution with parameter $1/\overline{X}$ for the distribution $\hat{F}$ in the bootstrap. The bootstrap estimate of the variance of $\overline{X}$ is the variance of a sample average $\overline{X}^*$ of a sample of size n from the distribution $\hat{F}$, i.e., the exponential distribution with parameter $1/\overline{X}$. The variance of $\overline{X}^*$ is $1/n$ times the variance of a single observation from $\hat{F}$, which equals $\overline{X}^2$. So, the bootstrap estimate is $\overline{X}^2/n$.

2. The numbers $x_1, \ldots, x_n$ are known when we sample from F_n. Let $i_1, \ldots, i_n \in \{1, \ldots, n\}$. Since $X_j = x_{i_j}$ if and only if $J_j = i_j$, we can compute

$$\Pr(X_1^* = x_{i_1}, \ldots, X_n^* = x_{i_n}) = \Pr(J_1 = i_1, \ldots, J_n = i_n) = \prod_{j=1}^n \Pr(J_j = i_j) = \prod_{j=1}^n \Pr(X_j^* = x_{i_j}).$$

   The second equality follows from the fact that $J_1, \ldots, J_n$ are a random sample with replacement from the set $\{1, \ldots, n\}$.

3. Let $n = 2k + 1$. The sample median of a nonparametric bootstrap sample is the $k + 1$st smallest observation in the bootstrap sample. Let x denote the smallest observation in the original sample. Assume that there are ℓ observations from the original sample that equal x. (Usually $\ell = 1$, but it is not necessary.) The sample median from the bootstrap sample equals x from the original data set if and only if at least $k + 1$ observations in the bootstrap sample equal x. Since each observation in the bootstrap equals x with probability ℓ/n and the bootstrap observations are independent, the probability that at least $k + 1$ of them equal x is

$$\sum_{i=k+1}^{n} \binom{n}{i} \left(\frac{\ell}{n}\right)^i \left(1 - \frac{\ell}{n}\right)^{n-i}.$$

4. For each bootstrap sample, compute the sample median. The bias estimate is the average of all of these sample medians minus the original sample median, 201.3. I started with a pilot sample of size 2000 and estimated the bias as 0.545. The sample variance of the 2000 sample medians was 3.435. This led me to estimate the necessary simulation size as

$$\left[\Phi^{-1}\left(\frac{1 + 0.9}{2}\right) \frac{3.435^{1/2}}{0.02}\right]^2 = 23234.$$

   So, I did 30000 bootstrap samples. The new estimate of bias was 0.5564, with a simulation standard error of 0.011.

5. This exercise is performed in a manner similar to Exercise 4.

(a) In this case, I did three simulations of size 50000 each. The three estimates of bias were -1.684, -1.688, and -1.608.

(b) Each time, the estimated sample size needed to achieve the desired accuracy was between 48000 and 49000.

6. (a) For each bootstrap sample, compute the sample median. The estimate we want is the sample variance of these values. I did a pilot simulation of size 2000 and got a sample variance of 18.15. I did another simulation of size 10000 and got a sample variance of 18.87.

(b) To achieve the desired accuracy, we would need a simulation of size

$$\left[\Phi^{-1}\left(\frac{1+0.95}{2}\right)\frac{18.87^{1/2}}{0.005}\right]^2 = 2899533.$$

That is, we would need about three million bootstrap samples.

7. (a) Each bootstrap sample consists of $\overline{X}^{*(i)}$ having a normal distribution with mean 0 and variance $31.65/11$, $\overline{Y}^{*(i)}$ having the normal distribution with mean 0 and variance $68.8/10$, $S_X^{2*(i)}$ being 31.65 times a χ^2 random variable with 10 degrees of freedom, and $S_Y^{2*(i)}$ being 68.8 times a χ^2 random variable with 9 degrees of freedom. For each sample, we compute the statistic $U^{(i)}$ displayed in Example 12.6.10 in the text. We then compute what proportion of the absolute values of the 10000 statistics exceed the 0.95 quantile of the t distribution with 19 degrees of freedom, 1.729. In three separate simulations, I got proportions of 0.1101, 0.1078, and 0.1115.

(b) To correct the level of the test, we need the 0.9 quantile of the distribution of $|U|$. For each simulation, we sort the 10000 $|U^{(i)}|$ values and select the 9000th value. In my three simulations, this value was 1.773, 1.777, and 1.788.

(c) To compute the simulation standard error of the sample quantile, I chose to split the 10000 samples into eight sets of size 1250. For each set, I sort the $|U^{(i)}|$ values and choose the 1125th one. The simulation standard error is then the the square-root of one-eighth of the the sample variance of these eight values. In my three simulations, I got the values 0.0112, 0.0136, and 0.0147.

8. The correlation is the ratio of the covariance to the square-root of the product of the variances. The mean of X^* is $E(X^*) = \overline{X}$, and the mean of Y^* is $E(Y^*) = \overline{Y}$. The variance of X^* is $\sum_{i=1}^{n}(X_i - \overline{X})^2/n$, and the variance of Y^* is $\sum_{i=1}^{n}(Y_i - \overline{Y})^2/n$. The covariance is

$$E[(X^* - \overline{X})(Y^* - \overline{Y})] = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Dividing this by the square-root of the product of the variances yields (12.6.2).

9. (a) For each bootstrap sample, compute the sample correlation $R^{(i)}$. Then compute the sample variance of $R^{(1)}, \ldots, R^{(1000)}$. This is the approximation to the bootstrap estimate of the variance of the sample correlation. I did three separate simulations and got sample variances of 4.781×10^{-4}, 4.741×10^{-4}, and 4.986×10^{-4}.

(b) The approximation to the bootstrap bias estimate is the sample average of $R^{(1)}, \ldots, R^{(1000)}$ minus the original sample correlation, 0.9670. In my three simulations, I got the values -0.0030, -0.0022, and -0.0026. It looks like 1000 is not enough bootstrap samples to get a good estimate of this bias.

(c) For the simulation standard error of the variance estimate, we use the square-root of Eq. (12.2.3) where each $Y^{(i)}$ in (12.2.3) is $R^{(i)}$ in this exercise. In my three simulations, I got the values 2.231×10^{-5}, 2.734×10^{-5}, and 3.228×10^{-5}. For the simulation standard error of the bias estimate, we just note that the bias estimate is an average, so we need only calculate the square-root of 1/1000 times the sample variance of $R^{(1)}, \dots, R^{(1000)}$. In my simulations, I got 6.915×10^{-4}, 6.886×10^{-4}, and 7.061×10^{-4}.

10. For both parts (a) and (b), we need 10000 bootstrap samples. From each bootstrap sample, we compute the sample median. Call these values $M^{*(i)}$, for $i = 1, \dots, 10000$. The median of the original data is $M = 152.5$.

   (a) Sort the $M^{*(i)}$ values from smallest to largest. The percentile interval just runs from the 500th sorted value to the 9500th sorted value. I ran three simulations and got the following three intervals: $[148, 175]$, $[148, 175]$, and $[146.5, 175]$.

   (b) Choose a measure of spread and compute it from the original sample. Call the value Y. For each bootstrap sample, compute the same measure of spread $Y^{*(i)}$. I choose the median absolute deviation, which is $Y = 19$ for this data set. Then sort the values $(M^{*(i)} - M)/Y^{*(i)}$. Find the 500th and 9500th sorted values Z_{500} and Z_{9500}. The percentile-t confidence interval runs from $M - ZY$ to $M + ZY$. In my three simulations, I got the intervals $[143, 181]$, $[142.6, 181]$, and $[141.9, 181]$.

   (c) The sample average of the beef hot dog values is 156.9, and the value of σ' is 22.64. The confidence interval based on the normal distribution use the t distribution quantile $T_{19}^{-1}(0.95) = 1.729$ and equals $156.9 \pm 1.729 \times 22.64/20^{1/2}$, or $[148.1, 165.6]$. This interval is considerably shorter than either of the bootstrap intervals.

11. (a) If X^* has the distribution F_n, then $\mu = E(X^*) = \overline{X}$,

$$\begin{aligned} \sigma^2 &= \text{Var}(X^*) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2, \text{ and} \\ E([X-\mu]^3) &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^3. \end{aligned}$$

   Plugging these values into the formula for skewness (see Definition 4.4.1) yields the formula for M_3 given in this exercise.

   (b) The summary statistics of the 1970 fish price data are $\overline{X} = 41.1$, $\sum_{i=1}^{n}(X_i - \overline{X})^2/n = 1316.5$, and $\sum_{i=1}^{n}(X_i - \overline{X})^3/n = 58176$, so the sample skewness is $M_3 = 1.218$. For each bootstrap sample, we also compute the sample skewness $M_3^{*(i)}$ for $i = 1, \dots, 1000$. The bias of M_3 is estimated by the sample average of the $M_3^{*(i)}$'s minus M_3. I did three simulations and got the values -0.2537, -0.2936, and -0.2888. To estimate the standard deviation of M_3, compute the sample standard deviation of the $M_3^{*(i)}$'s. In my three simulations, I got 0.5480, 0.5590, and 0.5411.

12. We want to show that the distribution of R is the same for all parameter vectors $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ that share the same value of ρ. Let $\boldsymbol{\theta}_1 = (\mu_{x1}, \mu_{y1}, \sigma_{x1}^2, \sigma_{y1}^2, \rho)$ and $\boldsymbol{\theta}_2 = (\mu_{x2}, \mu_{y2}, \sigma_{x2}^2, \sigma_{y2}^2, \rho)$ be two parameter vectors that share the same value of ρ. Let $a_x = \sigma_{x2}/\sigma_{x1}$, $a_y = \sigma_{y2}/\sigma_{y1}$, $b_x = \mu_{x2} - \mu_{x1}$, and $b_y - \mu_{y2} - \mu_{y1}$. For $i = 1, 2$, let $\boldsymbol{W}_i$ be a sample of size n from a bivariate normal distribution with parameter vector $\boldsymbol{\theta}_i$, and let R_i be the sample correlation. We want to show that R_1 and R_2 have

the same distribution. Write $\boldsymbol{W}_i = [(X_{i1}, Y_{i1}), \ldots, (X_{in}, Y_{in})]$ for $i = 1, 2$. Define $X'_{ij} = a_x(X_{ij} + b_x)$, $Y'_{ij} = a_y(Y_{ij} + b_y)$ for $j = 1, \ldots, n$. Then it is trivial to see that $\boldsymbol{W}'_2 = [(X'_{21}, Y'_{21}), \ldots, (X'_{2n}, Y'_{2n})]$ has the same distribution as $\boldsymbol{W}_2$. Let R'_2 be the sample correlation computed from $\boldsymbol{W}'_2$. Then R'_2 and R_2 have the same distribution. We complete the proof by showing that $R'_2 = R_1$. Hence R'_2 and R_1 and R_2 all have the same distribution. To see that $R'_2 = R_1$, let $\overline{X}_1 = \sum_{j=1}^{n} X_{1j}$ and similarly $\overline{Y}_1$, $\overline{X}'_2$ and $\overline{Y}'_2$. Then $\overline{X}'_2 = a_x(\overline{X}_1 + b_x)$ and $\overline{Y}'_2 = a_y(\overline{Y}_1 + b_y)$. So, for each j, $X'_{2j} - \overline{X}'_2 = a_x(X_{1j} - \overline{X}_1)$ and $Y'_{2j} - \overline{Y}'_2 = a_y(Y_{1j} - \overline{Y}_1)$. Since $a_x, a_y > 0$, it follows that

$$
\begin{aligned}
R'_2 &= \frac{\sum_{j=1}^{n}(X'_{2j} - \overline{X}'_2)(Y'_{2j} - \overline{Y}'_2)}{\left(\left[\sum_{j1}^{n}(X'_{2j} - \overline{X}'_2)^2\right]\left[\sum_{j1}^{n}(Y'_{2j} - \overline{Y}'_2)^2\right]\right)^{1/2}} \\
&= \frac{a_x a_y \sum_{j=1}^{n}(X_{1j} - \overline{X}_1)(Y1_{1j} - \overline{Y}1_1)}{\left(\left[a_x^2 \sum_{j1}^{n}(X_{1j} - \overline{X}_1)^2\right]\left[a_y^2 \sum_{j1}^{n}(Y_{1j} - \overline{Y}_1)^2\right]\right)^{1/2}} \\
&= R_1.
\end{aligned}
$$

12.7 Supplementary Exercises

Solutions to Exercises

1. For the random number generator that I have been using for these solutions, Fig. S.12.7 contains one such normal quantile plot. It looks fairly straight. On the horizontal axis I plotted the sorted

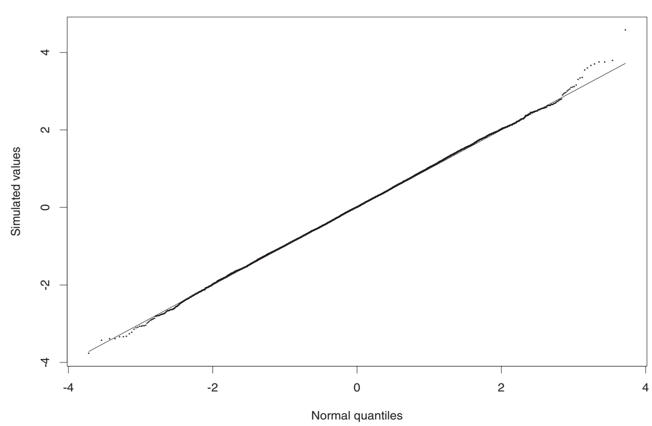

Figure S.12.7: Normal quantile plot for Exercise 1 in Sec. 12.7. A straight line has been added for reference.

pseudo-normal values and on the vertical axis, I plotted the values $\Phi^{-1}(i/10001)$ for $i = 1, \ldots, 10000$.

2. The plots for this exercise are formed the same way as that in Exercise 1 except we replace the normal pseudo-random values by the appropriate gamma pseudo-random values and we replace Φ^{-1} by the

quantile function of the appropriate gamma distribution. Two of the plots are in Fig. S.12.8. The plots are pretty straight except in the extreme upper tail, where things are expected to be highly variable.

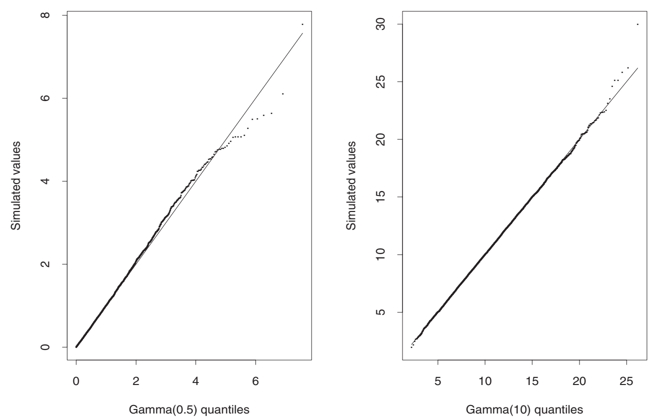

Figure S.12.8: Gamma quantile plots for Exercise 2 in Sec. 12.7. The left plot has parameters 0.5 and 1 and the right plot has parameters 10 and 1. Straight lines have been added for reference.

3. Once again, the plots are drawn in a fashion similar to Exercise 1. This time, we notice that the plot with one degree of freedom has some really serious non-linearity. This is the Cauchy distribution which has very long tails. The extreme observations from a Cauchy sample are very variable. Two of the plots are in Fig. S.12.9.

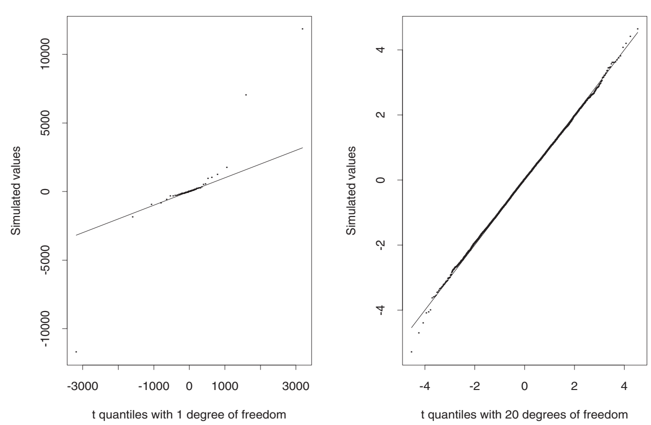

Figure S.12.9: Two t quantile plots for Exercise 3 in Sec. 12.7. The left plot has 1 degree of freedom, and the right plot has 20 degrees of freedom. Straight lines have been added for reference.

4. (a) I simulated 1000 pairs three times and got the following average values: 1.478, 1.462, 1.608. It looks like 1000 is not enough to be very confident of getting the average within 0.01.
   (b) Using the same three sets of 1000, I computed the sample variance each time and got 1.8521, 1.6857, and 2.5373.
   (c) Using (12.2.5), it appears that we need from 120000 to 170000 simulations.

5. (a) To simulate a noncentral t random variable, we can simulate independent Z and W with Z having the normal distribution with mean 1.936 and variance 1, and W having the χ^2 distribution with 14 degrees of freedom. Then set $T = Z/(W/14)^{1/2}$.

   (b) I did three separate simulations of size 1000 each and got the following three proportions with $T > 1.761$: 0.571, 0.608, 0.577. The simulation standard errors were 0.01565, 0.01544, and 0.01562.

   (c) Using (12.2.5), we find that we need a bit more than 16000 simulated values.

6. (a) For each sample, we compute the numbers of observations in each of the four intervals $(-\infty, 3.575)$, $[3.575, 3.912)$, $[3.912, 4.249)$, and $[4.249, \infty)$. Then we compute the Q statistic as we did in Example 10.1.6. We then compare each Q statistic to the three critical values 7.779, 9.488, and 13.277. We compute what proportion of the 10000 Q's is above each of these three critical values. I did three separate simulations of size 10000 each and got the proportions: 0.0495, 0.0536, and 0.0514 for the 0.9 critical value (7.779). I got 0.0222, 0.0247, and 0.0242 for the 0.95 critical value (9.488). I got 0.0025, 0.0021, and 0.0029 for the 0.99 critical value (13.277). It looks like the test whose nominal level is 0.1 has size closer to 0.05, while the test whose nominal level is 0.05 has level closer to 0.025.

   (b) For the power calculation, we perform exactly the same calculations with samples from the different normal distribution. I performed three simulations of size 1000 each for this exercise also. I got the proportions: 0.5653, 0.5767, and 0.5796 for the 0.9 critical value (7.779). I got 0.4560, 0.4667, and 0.4675 for the 0.95 critical value (9.488). I got 0.2224, 0.2280, and 0.2333 for the 0.99 critical value (13.277).

7. (a) We need to compute the same Q statistics as in Exercise 6(b) using samples from ten different normal distributions. For each of the ten distributions, we also compute the 0.9, 0.95 and 0.99 sample quantiles of the 10000 Q statistics. Here is a table of the simulated quantiles:

		Quantile		
μ	σ^2	0.9	0.95	0.99
3.8	0.25	3.891	4.976	7.405
3.8	0.80	4.295	5.333	8.788
3.9	0.25	3.653	4.764	6.405
3.9	0.80	4.142	5.133	7.149
4.0	0.25	3.825	5.104	7.405
4.0	0.80	4.554	5.541	8.635
4.1	0.25	3.861	5.255	8.305
4.1	0.80	4.505	5.658	8.637
4.2	0.25	4.193	5.352	8.260
4.2	0.80	4.087	4.981	7.677

   (b) The quantiles change a bit as the distributions change, but they are remarkably stable.

   (c) Instead of starting with normal samples, we start with samples having a t distribution as described in the exercise. We compute the Q statistic for each sample and see what proportion of our 10000 Q statistics is greater than 5.2. In three simulations of this sort I got proportions of 0.12 0.118, and 01.24.

8. (a) The product of likelihood times prior is

$$\exp\left(-\frac{u_0(\psi-\psi_0)^2}{2} - \sum_{i=1}^{p} \tau_i \left[\beta_0 + \frac{n_i(\mu_i - \overline{y}_i)^2 + w_i + \lambda(\mu_i - \psi)^2}{2}\right]\right)$$

$$\times \quad \lambda^{p/2+\gamma_0-1}\exp(-\lambda\delta_0)\prod_{i=1}^{p}\tau_i^{\alpha_0+[n_i+1]/2-1},$$

where $w_i = \sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)^2$ for $i = 1, \ldots, p$.

(b) As a function of μ_i, τ_i, or ψ this looks the same as it did in Example 12.5.6 except that λ_0 needs to be replaced by λ wherever it occurs. As a function of λ, it looks like the p.d.f. of the gamma distribution with parameters $p/2+\gamma_0$ and $\delta_0 + \sum_{i=1}^{p}\tau_i(\mu_i-\psi)^2/2$.

(c) I ran six Markov chains for 10000 iterations each, producing 60000 parameter vectors. The requested posterior means and simulation standard errors were

Parameter	μ_1	μ_2	μ_3	μ_4	$1/\tau_1$	$1/\tau_2$	$1/\tau_3$	$1/\tau_4$
Posterior mean	156.9	158.7	118.8	160.6	486.7	598.8	479.2	548.4
Sim. std. err.	0.009583	0.01969	0.02096	0.01322	0.8332	0.8286	0.5481	0.9372

The code at the end of this manual was modified following the suggestions in the exercise in order to produce the above output. The same was done in Exercise 9.

9. (a) The product of likelihood times prior is

$$\exp\left(-\frac{u_0(\psi-\psi_0)^2}{2} - \sum_{i=1}^{p}\tau_i\left[\beta + \frac{n_i(\mu_i-\overline{y}_i)^2 + w_i + \lambda_0(\mu_i-\psi)^2}{2}\right]\right)$$
$$\times \quad \beta^{p\alpha_0+\epsilon_0-1}\exp(-\beta\phi_0)\prod_{i=1}^{p}\tau_i^{\alpha_0+[n_i+1]/2-1},$$

where $w_i = \sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)^2$ for $i = 1, \ldots, p$.

(b) As a function of μ_i, τ_i, or ψ this looks the same as it did in Example 12.5.6 except that β_0 needs to be replaced by β wherever it occurs. As a function of β, it looks like the p.d.f. of the gamma distribution with parameters $p\alpha_0+\epsilon_0$ and $\phi_0 + \sum_{i=1}^{p}\tau_i$.

(c) I ran six Markov chains for 10000 iterations each, producing 60000 parameter vectors. The requested posterior means and simulation standard errors were

Parameter	μ_1	μ_2	μ_3	μ_4	$1/\tau_1$	$1/\tau_2$	$1/\tau_3$	$1/\tau_4$
Posterior mean	156.6	158.3	120.6	159.7	495.1	609.2	545.3	570.4
Sim. std. err.	0.01576	0.01836	0.02140	0.03844	0.4176	1.194	0.8968	0.7629

10. (a) The numerator of the likelihood ratio statistic is the maximum of the likelihood function over all parameter values in the alternative hypothesis, while the denominator is the maximum of the likelihood over all values in the null hypothesis. Both the numerator and denominator have a factor of $\prod_{i=1}^{k}\binom{n_i}{X_i}$ that will divide out in the ratio, so we shall ignore these factors. In this example, the maximum over the alternative hypothesis will be the maximum over all parameter values, so we

would set $p_i = X_i/n_i$ in the likelihood to get

$$\prod_{i=1}^{k} \left(\frac{X_i}{n_i}\right)^{X_i} \left(1 - \frac{X_i}{n_i}\right)^{n_i - X_i} = \frac{\prod_{i=1}^{k} X_i^{X_i}(n_i - X_i)^{n_i - X_i}}{\prod_{i=1}^{k} n_i^{n_i}}.$$

For the denominator, all of the p_i are equal, hence the likelihood to be maximized is $p^{X_1+\cdots+X_k}(1-p)^{n_1+\cdots+n_k-X_1-\cdots-X_k}$. This is maximized at $p = \sum_{j=1}^{k} X_j / \sum_{j=1}^{k} n_j$, to yield

$$\left(\frac{\sum_{j=1}^{k} X_j}{\sum_{j=1}^{k} n_j}\right)^{\sum_{j=1}^{k} X_j} \left(1 - \frac{\sum_{j=1}^{k} X_j}{\sum_{j=1}^{k} n_j}\right)^{\sum_{j=1}^{k}(n_j - X_j)} = \frac{\left(\sum_{j=1}^{k} X_j\right)^{\sum_{j=1}^{k} X_j} \left(\sum_{j=1}^{k}(n_j - X_j)\right)^{\sum_{j=1}^{k}(n_j - X_j)}}{\left(\sum_{j=1}^{k} n_j\right)^{\sum_{j=1}^{k} n_j}}.$$

The ratio of these two maxima is a positive constant times the statistic stated in the exercise. The likelihood ratio test rejects the null hypothesis when the statistic is greater than a constant.

(b) Call the likelihood ratio test statistic T. The distribution of T, under the assumption that H_0 is true, that is $p_1 = \cdots = p_k$ still depends on the common value of the p_i's, call it p. If the sample sizes are large, the distribution should not depend very much on p, but it will still depend on p. Let $F_p(\cdot)$ denote the c.d.f. of T when p is the common value of all p_i's. If we reject the null hypothesis when $T \geq c$, the test will be of level α_0 so long as

$$1 - F_p(c) \leq \alpha_0, \text{ for all } p. \tag{S.12.12}$$

If c satisfies (S.12.12) then all larger c satisfy (S.12.12), so we want the smallest c that satisfies (S.12.12). Eq. (S.12.12) is equivalent to $F_p(c) \geq 1 - \alpha_0$ for all p, which, in turn, is equivalent to $c \geq F_p^{-1}(1-\alpha_0)$ for all p. The smallest c that satisfies this last inequality is $c = \sup_p F_p^{-1}(1-\alpha_0)$. To approximate c by simulation, proceed as follows. Pick a collection of reasonable values of p and a large number v of simulations to perform. For each value of p, perform v simulations as follows. Simulate k independent binomial random variables with parameters n_i and p, and compute the value of T. Sort the v values of T and approximate $F_p^{-1}(1-\alpha_0)$ by the $(1-\alpha_0)v$th sorted value. Let c be the largest of these values over the different chosen values of p. It should be clear that the distribution of T is the same for p as it is for $1-p$, so one need only check values of p between 0 and 1/2.

(c) To compute the p-value, we first find the observed value t of T, and then find $\sup_p \Pr(T \geq t)$ under the assumption that the each $p_i = p$ for $i = 1, \ldots, k$. In Table 2.1, the X_i values are $X_1 = 22$, $X_2 = 25$, $X_3 = 16$, $X_4 = 10$, while the sample sizes are $n_1 = 40$, $n_2 = 38$, $n_3 = 38$, $n_4 = 34$. The observed value of T is

$$t = \frac{22^{22}18^{18}25^{25}13^{13}16^{16}22^{22}10^{10}24^{24}}{73^{73}77^{77}} = \exp(-202.17).$$

A pilot simulation showed that the maximum over p of $1 - F_p(t)$ occurs at $p = 0.5$, so a larger simulation was performed with $p = 0.5$. The estimated p-value is 0.01255 with a simulation standard error of 0.0039.

11. (a) We shall use the same approach as in Exercise 12 of Sec. 12.6. Let the parameter be $\theta = (\mu, \sigma_1, \sigma_2)$ (where μ is the common value of $\mu_1 = \mu_2$). Each pair of parameter values θ and θ' that have the

same value of σ_2/σ_1 can be obtained from each other by multiplying μ, σ_1 and σ_2 by the same positive constant and adding some other constant to the resulting μ. That is, there exist $a > 0$ and b such that $\theta' = (a\mu + b, a\sigma_1, a\sigma_2)$. If $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ have the distribution determined by θ, then $X_i' = aX_i + b$ for $i = 1, \ldots, m$ and $Y_j' = aY_j + b$ for $j = 1, \ldots, n$ have the distribution determined by θ'. We need only show that the statistic V in (9.6.13) has the same value when it is computed using the X_i's and Y_j's as when it is computed using the X_i''s and Y_j''s. It is easy to see that the numerator of V computed with the X_i''s and Y_j' equals a times the numerator of V computed using the X_i's and Y_j's. The same is true of the denominator, hence V has the same value either way and it must have the same distribution when the parameter is θ as when the parameter is θ'.

(b) By the same reasoning as in part (a), the value of ν is the same whether it is calculated with the X_i's and Y_j's or with the X_i''s and Y_j''s. Hence the distribution of ν (thought of as a random variable before observing the data) depends on the parameter only through σ_2/σ_1.

(c) For each simulation with ratio r, we can simulate $\overline{X}_m$ having the standard normal distribution and S_X^2 having the χ^2 distribution with 9 degrees of freedom. Then simulate $\overline{Y}_n$ having the normal distribution with mean 0 and variance r^2 and S_Y^2 equal to r^2 times a χ^2 random variable with 10 degrees of freedom. Make the four random variables independent when simulating. Then compute V and ν. Compute the three quantiles $T_\nu^{-1}(0.9)$, $T_\nu^{-1}(0.95)$ and $T_\nu^{-1}(0.99)$ and check whether V is greater than each quantile. Our estimates are the proportions of the 10000 simulations in which the value of V are greater than each quantile. Here are the results from one of my simulations:

	Probability		
r	0.9	0.95	0.99
1.0	0.1013	0.0474	0.0079
1.5	0.0976	0.0472	0.0088
2.0	0.0979	0.0506	0.0093
3.0	0.0973	0.0463	0.0110
5.0	0.0962	0.0476	0.0117
10.0	0.1007	0.0504	0.0113

The upper tail probabilities are very close to their nominal values.

12. I used the same simulations as in Exercise 11 but computed the statistic U from (9.6.3) instead of V and compared U to the quantiles of the t distribution with 19 degrees of freedom. The proportions are below:

	Probability		
r	0.9	0.95	0.99
1.0	0.1016	0.0478	0.0086
1.5	0.0946	0.0461	0.0090
2.0	0.0957	0.0483	0.0089
3.0	0.0929	0.0447	0.0112
5.0	0.0926	0.0463	0.0124
10.0	0.0964	0.0496	0.0121

These values are also very close to the nominal values.

13. (a) The fact that $E(\hat{\beta}_1) = \beta_1$ depends only on the fact that each Y_i has mean $\beta_0 + x_i\beta_1$. It does not depend on the distribution of Y_i (as long as the distribution has finite mean). Since $\hat{\beta}_1$ is a linear function of $Y_1, \ldots, Y_n$, its variance depends only on the variances of the Y_i's (and the fact that

they are independent). It doesn't depend on any other feature of the distribution. Indeed, we can write

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)Y_i}{\sum_{j=1}^{n}(x_j - \overline{x}_n)^2} = \sum_{i=1}^{n} a_i Y_i,$$

where $a_i = (x_i - \overline{x}_n)/\sum_{j=1}^{n}(x_j - \overline{x}_n)^2$. Then $\text{Var}(\hat{\beta}_1) = \sum_{i=1}^{n} a_i^2 \text{Var}(Y_i)$. This depends only on the variances of the Y_i's, which do not depend on β_0 or β_1.

(b) Let T have the t distribution with k degrees of freedom. Then Y_i has the same distribution as $\beta_0 + \beta_1 x_i + \sigma T$, whose variance is $\sigma^2 \text{Var}(T)$. Hence, $\text{Var}(Y_i) = \sigma^2 \text{Var}(T)$. It follows that

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \text{Var}(T) \sum_{i=1}^{n} a_i^2.$$

Let $v = \text{Var}(T) \sum_{i=1}^{n} a_i^2$.

(c) There are several possible simulation schemes to estimate v. The simplest might be to notice that

$$\sum_{i=1}^{n} a_i^2 = \frac{1}{\sum_{j=1}^{n}(x_i - \overline{x}_n)^2},$$

so that we only need to estimate $\text{Var}(T)$. This could be done by simulating lots of t random variables with k degrees of freedom and computing the sample variance. In fact, we can actually calculate v in closed form if we wish. According to Exercise 1 in Sec. 8.4, $\text{Var}(T) = k/(k-2)$.

14. As we noted in Exercise 13(c), the value of v is

$$\frac{5/3}{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2} = 3.14 \times 10^{-3}.$$

15. (a) We are trying to approximate the value a that makes $\ell(a) = E[L(\theta, a)|\boldsymbol{x}]$ the smallest. We have a sample $\theta^{(1)}, \ldots, \theta^{(v)}$ from the posterior distribution of θ, so we can approximate $\ell(a)$ by $\hat{\ell}(a) = \sum_{i=1}^{v} L(\theta^{(i)}, a)/v$. We could then do a search through many values of a to find the value that minimizes $\hat{\ell}(a)$. We could use either brute force or mathematical software for minimization. Of course, we would only have the value of a that minimizes $\hat{\ell}(a)$ rather than $\ell(a)$.

(b) To compute a simulation standard error, we could draw several (say k) samples from the posterior (or split one large sample into k smaller ones) and let Z_i be the value of a that minimizes the ith version of $\hat{\ell}$. Then compute S in Eq. (12.2.2) and let the simulation standard error be $S/k^{1/2}$.

16. (In the displayed formula, on the right side of the = sign, all θ's should have been μ's.) The posterior hyperparameters are all given in Example 12.5.2, so we can simulate as many μ values as we want to estimate the posterior mean of $L(\theta, a)$. We simulated 100000 t random variables with 22 degrees of freedom and multiplied each one by 15.214 and added 183.95 to get a sample of μ values. For each value of a near 183.95, we computed $\hat{\ell}(a)$ and found that $a = 182.644$ gave the smallest value. We then repeated the entire exercise for a total of five times. The other four a values were 182.641, 182.548, 182.57 and 182.645. The simulation standard error is then 0.0187

R Code For Two Text Examples

If you are not using *R* or if you are an expert, you should not bother reading this section.

The code below (with comments that start `#`) is used to perform the calculations in Examples 12.5.6 and 12.5.7 in the text. The reason that the code appears to be so elaborate is that I realized that Exercises 8 and 9 in Sec. 12.7 asked to perform essentially the same analysis, each with one additional parameter. Modifying the code below to handle those exercises is relatively straightforward. Significantly less coding would be needed if one were going to perform the analysis only once. For example, one would need the three functions that simulate each parameter given the others, plus the function called `hierchain` that could be used both for burn-in and the later runs. The remaining calculations could be done by typing some additional commands at the *R* prompt or in a text file to be `source`d.

In the first printing, there was an error in these examples. For some reason (my mistake, obviously) the w_i values were recorded in reverse order when the simulations were performed. That is, w_4 was used as if it were w_1, w_3 was used as if it were w_2, etc. The $\overline{y}_i$ and n_i values were in the correct order, otherwise the error could have been fixed by reordering the hot dog type names, but no such luck. Because the w_i were such different numbers, the effect on the numerical output was substantial. Most notably, the means of the $1/\tau_i$ are not nearly so different as stated in the first printing.

The data file `hotdogs.csv` contains four columns separated by commas with the data in Table 11.15 along with a `header` row:

```
Beef,Meat,Poultry,Specialty
 186,173,129,155
 181,191,132,170
 176,182,102,114
 149,190,106,191
 184,172,94,162
 190,147,102,146
 158,146,87,140
 139,139,99,187
 175,175,107,180
 148,136,113,,
 152,179,135,,
 111,153,142,,
 141,107,86,,
 153,195,143,,
 190,135,152,,
 157,140,146,,
 131,138,144,,
 149,,,
 135,,,
 132,,,
```

The commas with nothing after them indicate that the data in the next column has run out already, and `NA` (not available) values will be produced in *R*. Most *R* functions have sensible ways to deal with `NA` values, generally by including the optional argument `na.rm=T` or something similar. By default, *R* uses all values (including `NA`'s) to compute things like `mean` or `var`. Hence, the the result will be `NA` if one does not change the default.

First, we list some code that sets up the data, summary statistics, and prior hyperparameters. The lines that appear below were part of a file `hotdogmcmc-example.r`. They were executed by typing

```
source("hotdogmcmc-example.r")
```

at the *R* prompt in a command window. The `source` function reads a file of text and treats each line as if it had been typed at the *R* prompt.

```
# Read the data from a comma-separated file with a header row.
hotdogs=read.table("hotdogs.csv",header=T,sep=",")
# Compute the summary statistics
# First, the sample sizes: how many are not NA?
n=apply(hotdogs,2,function(x){sum(!is.na(x))})
# Next, the sample means: remember to remove the NA values
ybar=apply(hotdogs,2,mean,na.rm=T)
# Next, the wi values (sum of squared deviations from sample mean)
w=apply(hotdogs,2,var,na.rm=T)*(n-1)
# Set the prior hyperparameters:
hyp=list(lambda0=1, alpha0=1, beta0=0.1, u0=0.001, psi0=170)
# Set the initial values of parameters.  These will be perturbed to be
#   used as starting values for independent Markov chains.
tau=(n-1)/w
psi=(hyp$psi0*hyp$u0+hyp$lambda0*sum(tau*ybar))/(hyp$u0+hyp$lambda0*sum(tau))
mu=(n*ybar+hyp$lambda0*psi)/(n+hyp$lambda0)
```

Next, we list a series of functions that perform major parts of the calculation. The programs are written specifically for these examples, using variable names like `ybar`, `n`, `w`, `mu`, `tau`, and `psi` so that the reader can easily match what the programs are doing to the example. If one had wished to have a general hierarchical model program, one could have made the programs more generic at the cost of needing special routines to deal with the particular structure of the examples. Each of these functions is stored in a text file, and the `source` function is used to read the lines which in turn define the function for use by *R*. That is, after each file has been "`source`ed," the function whose name appears to the left of the = sign becomes available for use. It's arguments appear in parentheses after the word `function` on the first line.

First, we have the functions that simulate the next values of the parameters in each Markov chain:

```
mugen=function(i,tau,psi,n,ybar,w,hyp){
#
# Simulate a new mu[i] value
#
  (n[i]*ybar[i]+hyp$lambda0*psi)/(n[i]+hyp$lambda0)+rnorm(1,0,1)/sqrt(tau[i]*
    (n[i]+hyp$lambda0))
}
taugen=function(i,mu,psi,n,ybar,w,hyp){
#
# Simulate a new tau[i] value
#
  rgamma(1,hyp$alpha0+0.5*(n[i]+1))/(hyp$beta0+0.5*(w[i]+n[i]*(mu[i]-ybar[i])^2+
   hyp$lambda0*(mu[i]-psi)^2))
}
psigen=function(mu,tau,n,ybar,w,hyp){
#
# Simulate a new psi value
#
  (hyp$psi0*hyp$u0+hyp$lambda0*sum(tau*mu))/(hyp$u0+hyp$lambda0*sum(tau))+
```

```
    rnorm(1,0,1)/sqrt(hyp$u0+hyp$lambda0*sum(tau))
}
```

Next is the function that does burn-in and computes the F statistics described in the text. If the F statistics are too large, this function would have to be run again from the start with more burn-in. One could rewrite the function to allow it to start over from the end of the previous burn-in if one wished. (One would have to preserve the accumulated means and sums of squared deviations.)

```
burnchain=function(nburn,start,nchain,mu,tau,psi,n,ybar,w,hyp,stand){
#
# Perform "nburn" burn-in for "nchain" Markov chains and check the F statistics
#   starting after "start". The initial values are "mu", "tau", "psi"
#   and are perturbed by "stand" times random variables. The data are
#   "n", "ybar", "w". The prior  hyperparameters are "hyp".
#
# ngroup is the number of groups
ngroup=length(ybar)
# Set up the perturbed starting values for the different chains.
#   First, store 0 in all values
muval=matrix(0,nchain,ngroup)
tauval=muval
psival=rep(0,nchain)
# Next, for each chain, perturb the starting values using random
#   normals or lognormals
for(l in 1:nchain){
      muval[l,]=mu+stand*rnorm(ngroup)/sqrt(tau)
      tauval[l,]=tau*exp(rnorm(ngroup)*stand)
      psival[l]=psi+stand*rnorm(1)/sqrt(hyp$u0)
# Save the starting vectors for all chains just so we can see what
#   they were.
      startvec=cbind(muval,tauval,psival)
}
# The next matrices/vectors will store the accumulated means "...a" and sums
#   of squared deviations "...v" so that we don't need to store all of the
#   burn-in simulations when computing the F statistics.
#   See Exercise 23(b) in Sec. 7.10 of the text.
muacca=matrix(0,nchain,ngroup)
tauacca=muacca
psiacca=rep(0,nchain)
muaccv=muacca
tauaccv=muacca
psiaccv=psiacca
# The next matrix will store the burn-in F statistics so that we can
#   see if we need more burn-in.
fs=matrix(0,nburn-start+1,2*ngroup+1)
# Loop through the burn-in
for(i in 1:nburn){
# Loop throught the chains
      for(l in 1:nchain){
```

```
# Loop through the coordinates
   for(j in 1:ngroup){
# Generate the next mu
    muval[l,j]=mugen(j,tauval[l,],psival[l],n,ybar,w,hyp)
# Accumulate the average mu (muacca) and the sum of squared deviations (muaccv)
    muaccv[l,j]=muaccv[l,j]+(i-1)*(muval[l,j]-muacca[l,j])^2/i
    muacca[l,j]=muacca[l,j]+(muval[l,j]-muacca[l,j])/i
# Do the same for tau
    tauval[l,j]=taugen(j,muval[l,],psival[l],n,ybar,w,hyp)
    tauaccv[l,j]=tauaccv[l,j]+(i-1)*(tauval[l,j]-tauacca[l,j])^2/i
    tauacca[l,j]=tauacca[l,j]+(tauval[l,j]-tauacca[l,j])/i
        }
# Do the same for psi
psival[l]=psigen(muval[l,],tauval[l,],n,ybar,w,hyp)
        psiaccv[l]=psiaccv[l]+(i-1)*(psival[l]-psiacca[l])^2/i
        psiacca[l]=psiacca[l]+(psival[l]-psiacca[l])/i
  }
# Once we have enough burn-in, start computing the F statistics (see
#   p. 826 in the text)
  if(i>=start){
mub=i*apply(muacca,2,var)
muw=apply(muaccv,2,mean)/(i-1)
taub=i*apply(tauacca,2,var)
tauw=apply(tauaccv,2,mean)/(i-1)
psib=i*var(psiacca)
psiw=mean(psiaccv)/(i-1)
fs[i-start+1,]=c(mub/muw,taub/tauw,psib/psiw)
  }
}
# Return a list with useful information: the last value of each
#   parameter for all chains, the F statistics, the input information,
#   and the starting vectors.  The return value will contain enough
#   information to allow us to start all the Markov chains and
#   simulate them as long as we wish.
list(mu=muval,tau=tauval,psi=psival,fstat=fs,nburn=nburn,start=start,
n=n,ybar=ybar,w=w,hyp=hyp,nchain=nchain,startvec=startvec)
}
```

A similar, but simpler, function will simulate a single chain after we have finished burn-in:

```
hierchain=function(nsim,mu,tau,psi,n,ybar,w,hyp){
#
# Run a Markov chain for "nsim" simulations from initial values "mu",
#   "tau", "psi"; the data are "n", "ybar", "w"; the prior
#   hyperparameters are "hyp".
#
# ngroup is the number of groups
ngroup=length(ybar)
# Set up matrices to hold the simulated parameter values
```

```
psiex=rep(0,nsim)
muex=matrix(0,nsim,ngroup)
tauex=muex
# Loop through the simulations
for(i in 1:nsim){
# Loop through the coordinates
   for(j in 1:ngroup){
# Generate the next value of mu
    temp=mugen(j,tau,psi,n,ybar,w,hyp)
    mu[j]=temp
# Store the value of mu
    muex[i,j]=temp
# Do the same for tau
    temp=taugen(j,mu,psi,n,ybar,w,hyp)
    tau[j]=temp
    tauex[i,j]=temp
        }
# Do the same for psi
temp=psigen(mu,tau,n,ybar,w,hyp)
psi=temp
psiex[i]=temp
}
# Return a list with useful information: The simulated values
list(mu=muex,tau=tauex,psi=psiex)
}
```

Next, we have a function that will run several independent chains and put the results together. It calls the previous function once for each chain.

```
stackchains=function(burn,nsim){
#
# Starting from the information in "burn", obtained from "burnchain",
#   run "nsim" additional simulations for each chain and stack the
#   results on top of each other.  The
#   results from chain i can be extracted by using rows
#   (i-1)*nsim to i*nsim of each parameter matrix
#
# Set up storage for parameter values
  muex=NULL
  tauex=NULL
  psiex=NULL
# Loop through the chains
for(l in 1:burn$nchain){
# Extract the last burn-in parameter value for chain l
      mu=burn$mu[l,]
      tau=burn$tau[l,]
      psi=burn$psi[l]
# Run the chain nsim times
      temp=hierchain(nsim,mu,tau,psi,burn$n,burn$ybar,burn$w,burn$hyp)
```

```
# Extract the simulated values from each chain and stack them.
          muex=rbind(muex,temp$mu)
      tauex=rbind(tauex,temp$tau)
      psiex=c(psiex,temp$psi)
}
# Return a list with useful information: the simulated values, the
#   number of simulations per chain, and the number of chains.
list(mu=muex,tau=tauex,psi=psiex,nsim=nsim,nchain=burn$nchain)
}
```

The calculations done in Example 12.5.6 begin by applying the above functions and then manipulating the output. The following commands were typed at the *R* command prompt `>`. Notice that some of them produce output that appears in the same window in which the typing is done. The summary statistics appear in Table 12.4 in the text (after correcting the errors).

```
> # Do the burn-in
> hotdog.burn=burnchain(100,100,6,mu,tau,psi,n,ybar,w,hyp,2)
> # Note that the F statistics are all less than 1+0.44m=45.
> hotdog.burn$fstat
         [,1]      [,2]     [,3]     [,4]      [,5]      [,6]     [,7]     [,8]
[1,] 0.799452 0.7756733 1.464278 1.831631 0.9673807 0.4030658 1.161147 2.727503
         [,9]
[1,] 0.548359
> # Now run each chain 10000 more times.
> hotdog.mcmc=stackchains(hotdog.burn,10000)
> # Obtain the data for the summary table in the text
> apply(hotdog.mcmc$mu,2,mean)
[1] 156.5894 158.2559 120.5360 159.5841
> sqrt(apply(hotdog.mcmc$mu,2,var))
[1] 4.893067 5.825234 5.552140 7.615332
> apply(1/hotdog.mcmc$tau,2,mean)
[1] 495.6348 608.4955 542.8819 568.2482
> sqrt(apply(1/hotdog.mcmc$tau,2,var))
[1] 166.0203 221.1775 201.6250 307.3618
> mean(hotdog.mcmc$psi)
[1] 151.0273
> sqrt(var(hotdog.mcmc$psi))
[1] 11.16116
```

Next, we `source` a file that computes values that we can use to assess how similar/different the four groups of hot dogs are.

```
# Compute the six ratios of precisions (or variances)
hotdog.ratio=cbind(hotdog.mcmc$tau[,1]/hotdog.mcmc$tau[,2],
hotdog.mcmc$tau[,1]/hotdog.mcmc$tau[,3],hotdog.mcmc$tau[,1]/hotdog.mcmc$tau[,4],
hotdog.mcmc$tau[,2]/hotdog.mcmc$tau[,3],hotdog.mcmc$tau[,2]/hotdog.mcmc$tau[,4],
hotdog.mcmc$tau[,3]/hotdog.mcmc$tau[,4])
# For each simulation, find the maximum ratio. We need to include one over
#   each ratio also.
hotdog.rmax=apply(cbind(hotdog.ratio,1/hotdog.ratio),1,max)
```

```
# Compute the six differences between means.
hotdog.diff=cbind(hotdog.mcmc$mu[,1]-hotdog.mcmc$mu[,2],
hotdog.mcmc$mu[,2]-hotdog.mcmc$mu[,3],hotdog.mcmc$mu[,3]-hotdog.mcmc$mu[,4],
hotdog.mcmc$mu[,4]-hotdog.mcmc$mu[,1],hotdog.mcmc$mu[,1]-hotdog.mcmc$mu[,3],
hotdog.mcmc$mu[,2]-hotdog.mcmc$mu[,4])
# For each simulation, find the minimum, maximum and average absolute
#   differences.
hotdog.min=apply(abs(hotdog.diff),1,min)
hotdog.max=apply(abs(hotdog.diff),1,max)
hotdog.ave=apply(abs(hotdog.diff),1,mean)
```

Using the results of the above calculations, we now type commands at the prompt that answer various questions. First, what proportion of the time is one of the ratios of standard deviations at least 1.5 (ratio of variances at least 2.25)? In this calculation the vector `hotdog.max>2.25` has coordinates that are either `TRUE` (1) or `FALSE` (0) depending on whether the maximum ratio is greater than 2.25 or not. The `mean` is then the proportion of `TRUE`s.

```
mean(hotdog.rmax>2.25)
[1] 0.3982667
```

Next, compute the 0.01 quantile of the maximum absolute difference between between the means, the median of the minimum difference, and the 0.01 quantile of the average difference. In 99% of the simulations, the difference was greater than the 0.01 quantile.

```
> quantile(hotdog.max,0.01)
     1%
26.3452
> median(hotdog.min)
[1] 2.224152
> quantile(hotdog.ave,0.01)
      1%
13.77761
```

In Example 12.5.7, we needed to simulate a pair of observations (Y_1, Y_3) from each parameter vector and then approximate the 0.05 and 0.95 quantiles of the distribution of $Y_1 - Y_3$ for a prediction interval. The next function allows one to compute a general function of the parameters and find simulation standard errors using Eq. (12.5.1), that is $S/k^{1/2}$.

```
mcmcse=function(simobj,func,entire=FALSE){
#
# Start with the result of a simulation "simobj", compute a vector function
#   "func" from each chain, and then compute formula (12.5.1) for each
#   coordinate as well as the covariance matrix.  If "entire" is TRUE,
#   it also computes the function value on the entire parameter
#   matrix. This may differ from the average over the chains if "func"
#   is not and average and/or if it does additional simulation.
#   Also computes the avearge of the "func"
#   values. The function "func" must take as arguments matrices of mu,
#   tau, and psi values with each row from a single simulation.  It
#   must return a real vector.  For example, if you want two
#   quantiles of the distribution of f(Yi,Yj) where Yi comes from group
```

```
#   i and Yj comes from group j, func should loop through the rows of
#   its input and simulate a pair (Yi,Yj) for each parameter.  Then it
#   should return the appropriate sample quantiles of the simulated
#   values of f(Yj,Yj).
#
# k is the number of chains, nsim the number of simulations
        k=simobj$nchain
nsim=simobj$nsim
# Loop through the chains
for(i in 1:k){
# Extract the parameters for chain i
      mu=simobj$mu[((i-1)*nsim):(i*nsim),]
  tau=simobj$tau[((i-1)*nsim):(i*nsim),]
  psi=simobj$psi[((i-1)*nsim):(i*nsim)]
# Compute the function value based on the parameters of chain i
     if(i==1){
valf=func(mu,tau,psi)
   }else{
valf=rbind(valf,func(mu,tau,psi))
   }
}
# p is how many functions were computed
p=ncol(valf)
# compute the average of each function
ave=apply(valf,2,mean)
# compute formula (12.5.1) for each function
se=sqrt(apply(valf,2,var)*(k-1))/k
#
# Return the average function value, formula (12.5.1), and covariance
#   matrix.  The covariance matrix can be useful if you want to
#   compute a further function of the output and then compute a
#   simulation standard error for that further function. Also computes
#   the function on the entire parameter set if "entire=TRUE".
        if(entire){
list(ave=ave,se=se,covmat=cov(valf)*(k-1)/k^2,
entire=func(simobj$mu,simobj$tau,simobj$psi))
}else{
list(ave=ave,se=se,covmat=cov(valf)*(k-1)/k^2)
}
}
```

The specific function `func` used in Example 12.5.7 is:

```
hotdog13=function(mu,tau,psi){
#
# Compute the 0.05 and 0.95 quantiles of predictive distribution of Y1-Y3
#
  n=nrow(mu)
# Make a place to store the differences.
```

```
  vals=rep(0,n)
# Loop through the parameter vectors.
  for(i in 1:n){
# Simulate a difference.
   vals[i]=rnorm(1)/sqrt(tau[i,1])+mu[i,1]-(rnorm(1)/
sqrt(tau[i,3])+mu[i,3])
  }
# Return the desired quantiles
  quantile(vals,c(0.05,0.95))
}
```

Finally, we use the above functions to compute the prediction interval in Example 12.5.7 along with the simulation standard errors.

```
> hotdog.pred=mcmcse(hotdog.mcmc,hotdog13,T)
> hotdog.pred
$ave
       5%       95%
-18.57540  90.20092

$se
       5%       95%
0.2228034 0.4345629

$covmat
            5%        95%
5%  0.04964136 0.07727458
95% 0.07727458 0.18884493

$entire
       5%       95%
-18.49283  90.62661
```

The final line `hotdog.pred$entire` gives the prediction interval based on the entire collection of 60,000 simulations. The one listed as `hotdog.pred$ave` is the average of the six intervals based on the six independent Markov chains. There is not much difference between them. The simulation standard errors show up as `hotdog.pred$se`. Remember that the numbers in the first printing don't match these because of the error mentioned earlier.