Introduction to Probability
2nd Edition
Problem Solutions

(last updated: 9/18/15)

© Dimitri P. Bertsekas and John N. Tsitsiklis

Massachusetts Institute of Technology

WWW site for book information and orders

http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts

CHAPTER 1

Solution to Problem 1.1. We have

$$A = \{2, 4, 6\}, \qquad B = \{4, 5, 6\},$$

so $A \cup B = \{2, 4, 5, 6\}$, and

$$(A \cup B)^c = \{1, 3\}.$$

On the other hand,

$$A^c \cap B^c = \{1, 3, 5\} \cap \{1, 2, 3\} = \{1, 3\}.$$

Similarly, we have $A \cap B = \{4, 6\}$, and

$$(A \cap B)^c = \{1, 2, 3, 5\}.$$

On the other hand,

$$A^c \cup B^c = \{1, 3, 5\} \cup \{1, 2, 3\} = \{1, 2, 3, 5\}.$$

Solution to Problem 1.2. (a) By using a Venn diagram it can be seen that for any sets S and T, we have

$$S = (S \cap T) \cup (S \cap T^c).$$

(Alternatively, argue that any x must belong to either T or to T^c, so x belongs to S if and only if it belongs to $S \cap T$ or to $S \cap T^c$.) Apply this equality with $S = A^c$ and $T = B$, to obtain the first relation

$$A^c = (A^c \cap B) \cup (A^c \cap B^c).$$

Interchange the roles of A and B to obtain the second relation.

(b) By De Morgan's law, we have

$$(A \cap B)^c = A^c \cup B^c,$$

and by using the equalities of part (a), we obtain

$$(A \cap B)^c = \big((A^c \cap B) \cup (A^c \cap B^c)\big) \cup \big((A \cap B^c) \cup (A^c \cap B^c)\big) = (A^c \cap B) \cup (A^c \cap B^c) \cup (A \cap B^c).$$

(c) We have $A = \{1, 3, 5\}$ and $B = \{1, 2, 3\}$, so $A \cap B = \{1, 3\}$. Therefore,

$$(A \cap B)^c = \{2, 4, 5, 6\},$$

and

$$A^c \cap B = \{2\}, \qquad A^c \cap B^c = \{4, 6\}, \qquad A \cap B^c = \{5\}.$$

Thus, the equality of part (b) is verified.

Solution to Problem 1.5. Let G and C be the events that the chosen student is a genius and a chocolate lover, respectively. We have $\mathbf{P}(G) = 0.6$, $\mathbf{P}(C) = 0.7$, and $\mathbf{P}(G \cap C) = 0.4$. We are interested in $\mathbf{P}(G^c \cap C^c)$, which is obtained with the following calculation:

$$\mathbf{P}(G^c \cap C^c) = 1 - \mathbf{P}(G \cup C) = 1 - \big(\mathbf{P}(G) + \mathbf{P}(C) - \mathbf{P}(G \cap C)\big) = 1 - (0.6 + 0.7 - 0.4) = 0.1.$$

Solution to Problem 1.6. We first determine the probabilities of the six possible outcomes. Let $a = \mathbf{P}(\{1\}) = \mathbf{P}(\{3\}) = \mathbf{P}(\{5\})$ and $b = \mathbf{P}(\{2\}) = \mathbf{P}(\{4\}) = \mathbf{P}(\{6\})$. We are given that $b = 2a$. By the additivity and normalization axioms, $1 = 3a + 3b = 3a + 6a = 9a$. Thus, $a = 1/9$, $b = 2/9$, and $\mathbf{P}(\{1, 2, 3\}) = 4/9$.

Solution to Problem 1.7. The outcome of this experiment can be any finite sequence of the form $(a_1, a_2, \ldots, a_n)$, where n is an arbitrary positive integer, $a_1, a_2, \ldots, a_{n-1}$ belong to $\{1, 3\}$, and a_n belongs to $\{2, 4\}$. In addition, there are possible outcomes in which an even number is never obtained. Such outcomes are infinite sequences $(a_1, a_2, \ldots)$, with each element in the sequence belonging to $\{1, 3\}$. The sample space consists of all possible outcomes of the above two types.

Solution to Problem 1.8. Let p_i be the probability of winning against the opponent played in the ith turn. Then, you will win the tournament if you win against the 2nd player (probability p_2) and also you win against at least one of the two other players [probability $p_1 + (1 - p_1)p_3 = p_1 + p_3 - p_1p_3$]. Thus, the probability of winning the tournament is

$$p_2(p_1 + p_3 - p_1p_3).$$

The order $(1, 2, 3)$ is optimal if and only if the above probability is no less than the probabilities corresponding to the two alternative orders, i.e.,

$$p_2(p_1 + p_3 - p_1p_3) \geq p_1(p_2 + p_3 - p_2p_3),$$

$$p_2(p_1 + p_3 - p_1p_3) \geq p_3(p_2 + p_1 - p_2p_1).$$

It can be seen that the first inequality above is equivalent to $p_2 \geq p_1$, while the second inequality above is equivalent to $p_2 \geq p_3$.

Solution to Problem 1.9. (a) Since $\Omega = \cup_{i=1}^n S_i$, we have

$$A = \bigcup_{i=1}^n (A \cap S_i),$$

while the sets $A \cap S_i$ are disjoint. The result follows by using the additivity axiom.

(b) The events $B \cap C^c$, $B^c \cap C$, $B \cap C$, and $B^c \cap C^c$ form a partition of Ω, so by part (a), we have

$$\mathbf{P}(A) = \mathbf{P}(A \cap B \cap C^c) + \mathbf{P}(A \cap B^c \cap C) + \mathbf{P}(A \cap B \cap C) + \mathbf{P}(A \cap B^c \cap C^c). \quad (1)$$

The event $A \cap B$ can be written as the union of two disjoint events as follows:

$$A \cap B = (A \cap B \cap C) \cup (A \cap B \cap C^c),$$

so that

$$\mathbf{P}(A \cap B) = \mathbf{P}(A \cap B \cap C) + \mathbf{P}(A \cap B \cap C^c). \tag{2}$$

Similarly,

$$\mathbf{P}(A \cap C) = \mathbf{P}(A \cap B \cap C) + \mathbf{P}(A \cap B^c \cap C). \tag{3}$$

Combining Eqs. (1)-(3), we obtain the desired result.

Solution to Problem 1.10. Since the events $A \cap B^c$ and $A^c \cap B$ are disjoint, we have using the additivity axiom repeatedly,

$$\mathbf{P}\big((A \cap B^c) \cup (A^c \cap B)\big) = \mathbf{P}(A \cap B^c) + \mathbf{P}(A^c \cap B) = \mathbf{P}(A) - \mathbf{P}(A \cap B) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

Solution to Problem 1.14. (a) Each possible outcome has probability 1/36. There are 6 possible outcomes that are doubles, so the probability of doubles is $6/36 = 1/6$.

(b) The conditioning event (sum is 4 or less) consists of the 6 outcomes

$$\big\{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\big\},$$

2 of which are doubles, so the conditional probability of doubles is $2/6 = 1/3$.

(c) There are 11 possible outcomes with at least one 6, namely, $(6,6)$, $(6,i)$, and $(i,6)$, for $i = 1, 2, \ldots, 5$. Thus, the probability that at least one die is a 6 is 11/36.

(d) There are 30 possible outcomes where the dice land on different numbers. Out of these, there are 10 outcomes in which at least one of the rolls is a 6. Thus, the desired conditional probability is $10/30 = 1/3$.

Solution to Problem 1.15. Let A be the event that the first toss is a head and let B be the event that the second toss is a head. We must compare the conditional probabilities $\mathbf{P}(A \cap B \,|\, A)$ and $\mathbf{P}(A \cap B \,|\, A \cup B)$. We have

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

and

$$\mathbf{P}(A \cap B \,|\, A \cup B) = \frac{\mathbf{P}\big((A \cap B) \cap (A \cup B)\big)}{\mathbf{P}(A \cup B)} = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A \cup B)}.$$

Since $\mathbf{P}(A \cup B) \geq \mathbf{P}(A)$, the first conditional probability above is at least as large, so Alice is right, regardless of whether the coin is fair or not. In the case where the coin is fair, that is, if all four outcomes HH, HT, TH, TT are equally likely, we have

$$\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{1/4}{1/2} = \frac{1}{2}, \qquad \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A \cup B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

A generalization of Alice's reasoning is that if A, B, and C are events such that $B \subset C$ and $A \cap B = A \cap C$ (for example, if $A \subset B \subset C$), then the event A is at least as likely if we know that B has occurred than if we know that C has occurred. Alice's reasoning corresponds to the special case where $C = A \cup B$.

Solution to Problem 1.16. In this problem, there is a tendency to reason that since the opposite face is either heads or tails, the desired probability is 1/2. This is, however, wrong, because given that heads came up, it is more likely that the two-headed coin was chosen. The correct reasoning is to calculate the conditional probability

$$\begin{aligned} p &= \mathbf{P}(\text{two-headed coin was chosen} \,|\, \text{heads came up}) \\ &= \frac{\mathbf{P}(\text{two-headed coin was chosen and heads came up})}{\mathbf{P}(\text{heads came up})}. \end{aligned}$$

We have

$$\mathbf{P}(\text{two-headed coin was chosen and heads came up}) = \frac{1}{3},$$

$$\mathbf{P}(\text{heads came up}) = \frac{1}{2},$$

so by taking the ratio of the above two probabilities, we obtain $p = 2/3$. Thus, the probability that the opposite face is tails is $1 - p = 1/3$.

Solution to Problem 1.17. Let A be the event that the batch will be accepted. Then $A = A_1 \cap A_2 \cap A_3 \cap A_4$, where A_i, $i = 1, \ldots, 4$, is the event that the ith item is not defective. Using the multiplication rule, we have

$$\mathbf{P}(A) = \mathbf{P}(A_1)\mathbf{P}(A_2 \,|\, A_1)\mathbf{P}(A_3 \,|\, A_1 \cap A_2)\mathbf{P}(A_4 \,|\, A_1 \cap A_2 \cap A_3) = \frac{95}{100} \cdot \frac{94}{99} \cdot \frac{93}{98} \cdot \frac{92}{97} = 0.812.$$

Solution to Problem 1.18. Using the definition of conditional probabilities, we have

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

Solution to Problem 1.19. Let A be the event that Alice does not find her paper in drawer i. Since the paper is in drawer i with probability p_i, and her search is successful with probability d_i, the multiplication rule yields $\mathbf{P}(A^c) = p_i d_i$, so that $\mathbf{P}(A) = 1 - p_i d_i$. Let B be the event that the paper is in drawer j. If $j \neq i$, then $A \cap B = B$, $\mathbf{P}(A \cap B) = \mathbf{P}(B)$, and we have

$$\mathbf{P}(B \,|\, A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B)}{\mathbf{P}(A)} = \frac{p_j}{1 - p_i d_i}.$$

Similarly, if $i = j$, we have

$$\mathbf{P}(B \,|\, A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B)\mathbf{P}(A \,|\, B)}{\mathbf{P}(A)} = \frac{p_i(1 - d_i)}{1 - p_i d_i}.$$

Solution to Problem 1.20. (a) Figure 1.1 provides a sequential description for the three different strategies. Here we assume 1 point for a win, 0 for a loss, and 1/2 point

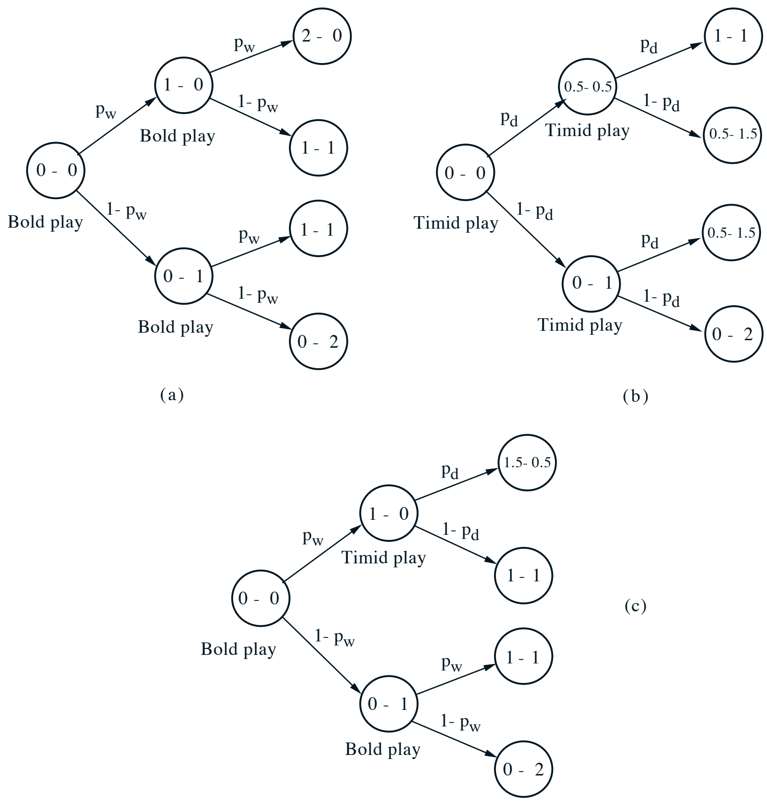

Figure 1.1: Sequential descriptions of the chess match histories under strategies (i), (ii), and (iii).

for a draw. In the case of a tied 1-1 score, we go to sudden death in the next game, and Boris wins the match (probability p_w), or loses the match (probability $1 - p_w$).

(i) Using the total probability theorem and the sequential description of Fig. 1.1(a), we have

$$\mathbf{P}(\text{Boris wins}) = p_w^2 + 2p_w(1 - p_w)p_w.$$

The term p_w^2 corresponds to the win-win outcome, and the term $2p_w(1 - p_w)p_w$ corresponds to the win-lose-win and the lose-win-win outcomes.

(ii) Using Fig. 1.1(b), we have

$$\mathbf{P}(\text{Boris wins}) = p_d^2 p_w,$$

corresponding to the draw-draw-win outcome.

(iii) Using Fig. 1.1(c), we have

$$\mathbf{P}(\text{Boris wins}) = p_w p_d + p_w(1 - p_d)p_w + (1 - p_w)p_w^2.$$

The term $p_w p_d$ corresponds to the win-draw outcome, the term $p_w(1-p_d)p_w$ corresponds to the win-lose-win outcome, and the term $(1-p_w)p_w^2$ corresponds to lose-win-win outcome.

(b) If $p_w < 1/2$, Boris has a greater probability of losing rather than winning any one game, regardless of the type of play he uses. Despite this, the probability of winning the match with strategy (iii) can be greater than 1/2, provided that p_w is close enough to 1/2 and p_d is close enough to 1. As an example, if $p_w = 0.45$ and $p_d = 0.9$, with strategy (iii) we have

$$\mathbf{P}(\text{Boris wins}) = 0.45 \cdot 0.9 + 0.45^2 \cdot (1-0.9) + (1-0.45) \cdot 0.45^2 \approx 0.54.$$

With strategies (i) and (ii), the corresponding probabilities of a win can be calculated to be approximately 0.43 and 0.36, respectively. What is happening here is that with strategy (iii), Boris is allowed to select a playing style *after* seeing the result of the first game, while his opponent is not. Thus, by being able to dictate the playing style in each game after receiving partial information about the match's outcome, Boris gains an advantage.

Solution to Problem 1.21. Let $p(m,k)$ be the probability that the starting player wins when the jar initially contains m white and k black balls. We have, using the total probability theorem,

$$p(m,k) = \frac{m}{m+k} + \frac{k}{m+k}\big(1 - p(m,k-1)\big) = 1 - \frac{k}{m+k} p(m,k-1).$$

The probabilities $p(m,1), p(m,2), \ldots, p(m,n)$ can be calculated sequentially using this formula, starting with the initial condition $p(m,0) = 1$.

Solution to Problem 1.22. We derive a recursion for the probability p_i that a white ball is chosen from the ith jar. We have, using the total probability theorem,

$$p_{i+1} = \frac{m+1}{m+n+1} p_i + \frac{m}{m+n+1}(1-p_i) = \frac{1}{m+n+1} p_i + \frac{m}{m+n+1},$$

starting with the initial condition $p_1 = m/(m+n)$. Thus, we have

$$p_2 = \frac{1}{m+n+1} \cdot \frac{m}{m+n} + \frac{m}{m+n+1} = \frac{m}{m+n}.$$

More generally, this calculation shows that if $p_{i-1} = m/(m+n)$, then $p_i = m/(m+n)$. Thus, we obtain $p_i = m/(m+n)$ for all i.

Solution to Problem 1.23. Let $p_{i,n-i}(k)$ denote the probability that after k exchanges, a jar will contain i balls that started in that jar and $n-i$ balls that started in the other jar. We want to find $p_{n,0}(4)$. We argue recursively, using the total probability theorem. We have

$$\begin{aligned}
p_{n,0}(4) &= \frac{1}{n}\cdot\frac{1}{n}\cdot p_{n-1,1}(3),\\
p_{n-1,1}(3) &= p_{n,0}(2) + 2\cdot\frac{n-1}{n}\cdot\frac{1}{n}\cdot p_{n-1,1}(2) + \frac{2}{n}\cdot\frac{2}{n}\cdot p_{n-2,2}(2),\\
p_{n,0}(2) &= \frac{1}{n}\cdot\frac{1}{n}\cdot p_{n-1,1}(1),\\
p_{n-1,1}(2) &= 2\cdot\frac{n-1}{n}\cdot\frac{1}{n}\cdot p_{n-1,1}(1),\\
p_{n-2,2}(2) &= \frac{n-1}{n}\cdot\frac{n-1}{n}\cdot p_{n-1,1}(1),\\
p_{n-1,1}(1) &= 1.
\end{aligned}$$

Combining these equations, we obtain

$$p_{n,0}(4) = \frac{1}{n^2}\left(\frac{1}{n^2} + \frac{4(n-1)^2}{n^4} + \frac{4(n-1)^2}{n^4}\right) = \frac{1}{n^2}\left(\frac{1}{n^2} + \frac{8(n-1)^2}{n^4}\right).$$

Solution to Problem 1.24. Intuitively, there is something wrong with this rationale. The reason is that it is not based on a correctly specified probabilistic model. In particular, the event where both of the other prisoners are to be released is not properly accounted in the calculation of the posterior probability of release.

To be precise, let A, B, and C be the prisoners, and let A be the one who considers asking the guard. Suppose that all prisoners are a priori equally likely to be released. Suppose also that if B and C are to be released, then the guard chooses B or C with equal probability to reveal to A. Then, there are four possible outcomes:

(1) A and B are to be released, and the guard says B (probability 1/3).

(2) A and C are to be released, and the guard says C (probability 1/3).

(3) B and C are to be released, and the guard says B (probability 1/6).

(4) B and C are to be released, and the guard says C (probability 1/6).

Thus,

$$\begin{aligned}
\mathbf{P}(\text{A is to be released} \mid \text{guard says B}) &= \frac{\mathbf{P}(\text{A is to be released and guard says B})}{\mathbf{P}(\text{guard says B})}\\
&= \frac{1/3}{1/3 + 1/6} = \frac{2}{3}.
\end{aligned}$$

Similarly,

$$\mathbf{P}(\text{A is to be released} \mid \text{guard says C}) = \frac{2}{3}.$$

Thus, regardless of the identity revealed by the guard, the probability that A is released is equal to 2/3, the a priori probability of being released.

Solution to Problem 1.25. Let $\overline{m}$ and $\underline{m}$ be the larger and the smaller of the two amounts, respectively. Consider the three events

$$A = \{X < \underline{m}), \qquad B = \{\underline{m} < X < \overline{m}), \qquad C = \{\overline{m} < X).$$

Let $\overline{A}$ (or $\overline{B}$ or $\overline{C}$) be the event that A (or B or C, respectively) occurs *and* you first select the envelope containing the larger amount $\overline{m}$. Let $\underline{A}$ (or $\underline{B}$ or $\underline{C}$) be the event that A (or B or C, respectively) occurs *and* you first select the envelope containing the smaller amount $\underline{m}$. Finally, consider the event

$$W = \{\text{you end up with the envelope containing } \overline{m}\}.$$

We want to determine $\mathbf{P}(W)$ and check whether it is larger than $1/2$ or not.

By the total probability theorem, we have

$$\mathbf{P}(W \,|\, A) = \frac{1}{2}\big(\mathbf{P}(W \,|\, \overline{A}) + \mathbf{P}(W \,|\, \underline{A})\big) = \frac{1}{2}(1+0) = \frac{1}{2},$$

$$\mathbf{P}(W \,|\, B) = \frac{1}{2}\big(\mathbf{P}(W \,|\, \overline{B}) + \mathbf{P}(W \,|\, \underline{B})\big) = \frac{1}{2}(1+1) = 1,$$

$$\mathbf{P}(W \,|\, C) = \frac{1}{2}\big(\mathbf{P}(W \,|\, \overline{C}) + \mathbf{P}(W \,|\, \underline{C})\big) = \frac{1}{2}(0+1) = \frac{1}{2}.$$

Using these relations together with the total probability theorem, we obtain

$$\begin{aligned}
\mathbf{P}(W) &= \mathbf{P}(A)\mathbf{P}(W \,|\, A) + \mathbf{P}(B)\mathbf{P}(W \,|\, B) + \mathbf{P}(C)\mathbf{P}(W \,|\, C) \\
&= \frac{1}{2}\big(\mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C)\big) + \frac{1}{2}\mathbf{P}(B) \\
&= \frac{1}{2} + \frac{1}{2}\mathbf{P}(B).
\end{aligned}$$

Since $\mathbf{P}(B) > 0$ by assumption, it follows that $\mathbf{P}(W) > 1/2$, so your friend is correct.

Solution to Problem 1.26. (a) We use the formula

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

Since all crows are black, we have $\mathbf{P}(B) = 1 - q$. Furthermore, $\mathbf{P}(A) = p$. Finally, $\mathbf{P}(B \,|\, A) = 1 - q = \mathbf{P}(B)$, since the probability of observing a (black) crow is not affected by the truth of our hypothesis. We conclude that $\mathbf{P}(A \,|\, B) = \mathbf{P}(A) = p$. Thus, the new evidence, while compatible with the hypothesis "all cows are white," does not change our beliefs about its truth.

(b) Once more,

$$\mathbf{P}(A \,|\, C) = \frac{\mathbf{P}(A \cap C)}{\mathbf{P}(C)} = \frac{\mathbf{P}(A)\mathbf{P}(C \,|\, A)}{\mathbf{P}(C)}.$$

Given the event A, a cow is observed with probability q, and it must be white. Thus, $\mathbf{P}(C \,|\, A) = q$. Given the event A^c, a cow is observed with probability q, and it is white with probability $1/2$. Thus, $\mathbf{P}(C \,|\, A^c) = q/2$. Using the total probability theorem,

$$\mathbf{P}(C) = \mathbf{P}(A)\mathbf{P}(C \,|\, A) + \mathbf{P}(A^c)\mathbf{P}(C \,|\, A^c) = pq + (1-p)\frac{q}{2}.$$

Hence,

$$\mathbf{P}(A \,|\, C) = \frac{pq}{pq + (1-p)\dfrac{q}{2}} = \frac{2p}{1+p} > p.$$

Thus, the observation of a white cow makes the hypothesis "all cows are white" more likely to be true.

Solution to Problem 1.27. Since Bob tosses one more coin that Alice, it is impossible that they toss both the same number of heads and the same number of tails. So Bob tosses either more heads than Alice or more tails than Alice (but not both). Since the coins are fair, these events are equally likely by symmetry, so both events have probability 1/2.

An alternative solution is to argue that if Alice and Bob are tied after $2n$ tosses, they are equally likely to win. If they are not tied, then their scores differ by at least 2, and toss $2n+1$ will not change the final outcome. This argument may also be expressed algebraically by using the total probability theorem. Let B be the event that Bob tosses more heads. Let X be the event that after each has tossed n of their coins, Bob has more heads than Alice, let Y be the event that under the same conditions, Alice has more heads than Bob, and let Z be the event that they have the same number of heads. Since the coins are fair, we have $\mathbf{P}(X) = \mathbf{P}(Y)$, and also $\mathbf{P}(Z) = 1 - \mathbf{P}(X) - \mathbf{P}(Y)$. Furthermore, we see that

$$\mathbf{P}(B \mid X) = 1, \qquad \mathbf{P}(B \mid Y) = 0, \qquad \mathbf{P}(B \mid Z) = \frac{1}{2}.$$

Now we have, using the total probability theorem,

$$\begin{aligned}
\mathbf{P}(B) &= \mathbf{P}(X) \cdot \mathbf{P}(B \mid X) + \mathbf{P}(Y) \cdot \mathbf{P}(B \mid Y) + \mathbf{P}(Z) \cdot \mathbf{P}(B \mid Z) \\
&= \mathbf{P}(X) + \frac{1}{2} \cdot \mathbf{P}(Z) \\
&= \frac{1}{2} \cdot \big(\mathbf{P}(X) + \mathbf{P}(Y) + \mathbf{P}(Z)\big) \\
&= \frac{1}{2}.
\end{aligned}$$

as required.

Solution to Problem 1.30. Consider the sample space for the hunter's strategy. The events that lead to the correct path are:

(1) Both dogs agree on the correct path (probability p^2, by independence).

(2) The dogs disagree, dog 1 chooses the correct path, and hunter follows dog 1 [probability $p(1-p)/2$].

(3) The dogs disagree, dog 2 chooses the correct path, and hunter follows dog 2 [probability $p(1-p)/2$].

The above events are disjoint, so we can add the probabilities to find that under the hunter's strategy, the probability that he chooses the correct path is

$$p^2 + \frac{1}{2}p(1-p) + \frac{1}{2}p(1-p) = p.$$

On the other hand, if the hunter lets one dog choose the path, this dog will also choose the correct path with probability p. Thus, the two strategies are equally effective.

Solution to Problem 1.31. (a) Let A be the event that a 0 is transmitted. Using the total probability theorem, the desired probability is

$$\mathbf{P}(A)(1-\epsilon_0)+\big(1-\mathbf{P}(A)\big)(1-\epsilon_1)=p(1-\epsilon_0)+(1-p)(1-\epsilon_1).$$

(b) By independence, the probability that the string 1011 is received correctly is

$$(1-\epsilon_0)(1-\epsilon_1)^3.$$

(c) In order for a 0 to be decoded correctly, the received string must be 000, 001, 010, or 100. Given that the string transmitted was 000, the probability of receiving 000 is $(1-\epsilon_0)^3$, and the probability of each of the strings 001, 010, and 100 is $\epsilon_0(1-\epsilon_0)^2$. Thus, the probability of correct decoding is

$$3\epsilon_0(1-\epsilon_0)^2+(1-\epsilon_0)^3.$$

(d) When the symbol is 0, the probabilities of correct decoding with and without the scheme of part (c) are $3\epsilon_0(1-\epsilon_0)^2+(1-\epsilon_0)^3$ and $1-\epsilon_0$, respectively. Thus, the probability is improved with the scheme of part (c) if

$$3\epsilon_0(1-\epsilon_0)^2+(1-\epsilon_0)^3>(1-\epsilon_0),$$

or

$$(1-\epsilon_0)(1+2\epsilon_0)>1,$$

which is equivalent to $0<\epsilon_0<1/2$.

(e) Using Bayes' rule, we have

$$\mathbf{P}(0\,|\,101)=\frac{\mathbf{P}(0)\mathbf{P}(101\,|\,0)}{\mathbf{P}(0)\mathbf{P}(101\,|\,0)+\mathbf{P}(1)\mathbf{P}(101\,|\,1)}.$$

The probabilities needed in the above formula are

$$\mathbf{P}(0)=p,\quad \mathbf{P}(1)=1-p,\quad \mathbf{P}(101\,|\,0)=\epsilon_0^2(1-\epsilon_0),\quad \mathbf{P}(101\,|\,1)=\epsilon_1(1-\epsilon_1)^2.$$

Solution to Problem 1.32. The answer to this problem is not unique and depends on the assumptions we make on the reproductive strategy of the king's parents.

Suppose that the king's parents had decided to have exactly two children and then stopped. There are four possible and equally likely outcomes, namely BB, GG, BG, and GB (B stands for "boy" and G stands for "girl"). Given that at least one child was a boy (the king), the outcome GG is eliminated and we are left with three equally likely outcomes (BB, BG, and GB). The probability that the sibling is male (the conditional probability of BB) is 1/3 .

Suppose on the other hand that the king's parents had decided to have children until they would have a male child. In that case, the king is the second child, and the sibling is female, with certainty.

Solution to Problem 1.33. Flip the coin twice. If the outcome is heads-tails, choose the opera. if the outcome is tails-heads, choose the movies. Otherwise, repeat the process, until a decision can be made. Let A_k be the event that a decision was made at the kth round. Conditional on the event A_k, the two choices are equally likely, and we have

$$\mathbf{P}(\text{opera}) = \sum_{k=1}^{\infty} \mathbf{P}(\text{opera} \,|\, A_k)\mathbf{P}(A_k) = \sum_{k=1}^{\infty} \frac{1}{2}\mathbf{P}(A_k) = \frac{1}{2}.$$

We have used here the property $\sum_{k=0}^{\infty} \mathbf{P}(A_k) = 1$, which is true as long as $\mathbf{P}(\text{heads}) > 0$ and $\mathbf{P}(\text{tails}) > 0$.

Solution to Problem 1.34. The system may be viewed as a series connection of three subsystems, denoted 1, 2, and 3 in Fig. 1.19 in the text. The probability that the entire system is operational is $p_1 p_2 p_3$, where p_i is the probability that subsystem i is operational. Using the formulas for the probability of success of a series or a parallel system given in Example 1.24, we have

$$p_1 = p, \qquad p_3 = 1 - (1-p)^2,$$

and

$$p_2 = 1 - (1-p)\big(1 - p\big(1 - (1-p)^3\big)\big).$$

Solution to Problem 1.35. Let A_i be the event that exactly i components are operational. The probability that the system is operational is the probability of the union $\cup_{i=k}^{n} A_i$, and since the A_i are disjoint, it is equal to

$$\sum_{i=k}^{n} \mathbf{P}(A_i) = \sum_{i=k}^{n} p(i),$$

where $p(i)$ are the binomial probabilities. Thus, the probability of an operational system is

$$\sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

Solution to Problem 1.36. (a) Let A denote the event that the city experiences a black-out. Since the power plants fail independent of each other, we have

$$\mathbf{P}(A) = \prod_{i=1}^{n} p_i.$$

(b) There will be a black-out if either all n or any $n-1$ power plants fail. These two events are disjoint, so we can calculate the probability $\mathbf{P}(A)$ of a black-out by adding their probabilities:

$$\mathbf{P}(A) = \prod_{i=1}^{n} p_i + \sum_{i=1}^{n} \left((1 - p_i) \prod_{j \neq i} p_j \right).$$

Here, $(1-p_i)\prod_{j\neq i} p_j$ is the probability that $n-1$ plants have failed and plant i is the one that has not failed.

Solution to Problem 1.37. The probability that k_1 voice users and k_2 data users simultaneously need to be connected is $p_1(k_1)p_2(k_2)$, where $p_1(k_1)$ and $p_2(k_2)$ are the corresponding binomial probabilities, given by

$$p_i(k_i) = \binom{n_i}{k_i} p_i^{k_i}(1-p_i)^{n_i-k_i}, \qquad i = 1, 2.$$

The probability that more users want to use the system than the system can accommodate is the sum of all products $p_1(k_1)p_2(k_2)$ as k_1 and k_2 range over all possible values whose total bit rate requirement $k_1r_1+k_2r_2$ exceeds the capacity c of the system. Thus, the desired probability is

$$\sum_{\{(k_1,k_2)\,|\,k_1r_1+k_2r_2>c,\; k_1\leq n_1,\; k_2\leq n_2\}} p_1(k_1)p_2(k_2).$$

Solution to Problem 1.38. We have

$$p_T = \mathbf{P}(\text{at least 6 out of the 8 remaining holes are won by Telis}),$$

$$p_W = \mathbf{P}(\text{at least 4 out of the 8 remaining holes are won by Wendy}).$$

Using the binomial formulas,

$$p_T = \sum_{k=6}^{8} \binom{8}{k} p^k (1-p)^{8-k}, \qquad p_W = \sum_{k=4}^{8} \binom{8}{k} (1-p)^k p^{8-k}.$$

The amount of money that Telis should get is $10 \cdot p_T/(p_T + p_W)$ dollars.

Solution to Problem 1.39. Let the event A be the event that the professor teaches her class, and let B be the event that the weather is bad. We have

$$\mathbf{P}(A) = \mathbf{P}(B)\mathbf{P}(A\,|\,B) + \mathbf{P}(B^c)\mathbf{P}(A\,|\,B^c),$$

and

$$\mathbf{P}(A\,|\,B) = \sum_{i=k}^{n} \binom{n}{i} p_b^i (1-p_b)^{n-i},$$

$$\mathbf{P}(A\,|\,B^c) = \sum_{i=k}^{n} \binom{n}{i} p_g^i (1-p_g)^{n-i}.$$

Therefore,

$$\mathbf{P}(A) = \mathbf{P}(B)\sum_{i=k}^{n} \binom{n}{i} p_b^i (1-p_b)^{n-i} + \big(1-\mathbf{P}(B)\big)\sum_{i=k}^{n} \binom{n}{i} p_g^i (1-p_g)^{n-i}.$$

Solution to Problem 1.40. Let A be the event that the first $n-1$ tosses produce an even number of heads, and let E be the event that the nth toss is a head. We can obtain an even number of heads in n tosses in two distinct ways: 1) there is an even number of heads in the first $n-1$ tosses, and the nth toss results in tails: this is the event $A \cap E^c$; 2) there is an odd number of heads in the first $n-1$ tosses, and the nth toss results in heads: this is the event $A^c \cap E$. Using also the independence of A and E,

$$\begin{aligned} q_n &= \mathbf{P}\big((A \cap E^c) \cup (A^c \cap E)\big) \\ &= \mathbf{P}(A \cap E^c) + \mathbf{P}(A^c \cap E) \\ &= \mathbf{P}(A)\mathbf{P}(E^c) + \mathbf{P}(A^c)\mathbf{P}(E) \\ &= (1-p)q_{n-1} + p(1-q_{n-1}). \end{aligned}$$

We now use induction. For $n = 0$, we have $q_0 = 1$, which agrees with the given formula for q_n. Assume, that the formula holds with n replaced by $n-1$, i.e.,

$$q_{n-1} = \frac{1 + (1-2p)^{n-1}}{2}.$$

Using this equation, we have

$$\begin{aligned} q_n &= p(1-q_{n-1}) + (1-p)q_{n-1} \\ &= p + (1-2p)q_{n-1} \\ &= p + (1-2p)\frac{1 + (1-2p)^{n-1}}{2} \\ &= \frac{1 + (1-2p)^n}{2}, \end{aligned}$$

so the given formula holds for all n.

Solution to Problem 1.41. We have

$$\mathbf{P}(N = n) = \mathbf{P}(A_{1,n-1} \cap A_{n,n}) = \mathbf{P}(A_{1,n-1})\mathbf{P}(A_{n,n} \mid A_{1,n-1}),$$

where for $i \le j$, $A_{i,j}$ is the event that contestant i's number is the smallest of the numbers of contestants $1, \ldots, j$. We also have

$$\mathbf{P}(A_{1,n-1}) = \frac{1}{n-1}.$$

We claim that

$$\mathbf{P}(A_{n,n} \mid A_{1,n-1}) = \mathbf{P}(A_{n,n}) = \frac{1}{n}.$$

The reason is that by symmetry, we have

$$\mathbf{P}(A_{n,n} \mid A_{i,n-1}) = \mathbf{P}(A_{n,n} \mid A_{1,n-1}), \qquad i = 1, \ldots, n-1,$$

while by the total probability theorem,

$$\begin{aligned} \mathbf{P}(A_{n,n}) &= \sum_{i=1}^{n-1} \mathbf{P}(A_{i,n-1})\mathbf{P}(A_{n,n} \mid A_{i,n-1}) \\ &= \mathbf{P}(A_{n,n} \mid A_{1,n-1}) \sum_{i=1}^{n-1} \mathbf{P}(A_{i,n-1}) \\ &= \mathbf{P}(A_{n,n} \mid A_{1,n-1}). \end{aligned}$$

Hence

$$\mathbf{P}(N = n) = \frac{1}{n-1} \cdot \frac{1}{n}.$$

An alternative solution is also possible, using the counting methods developed in Section 1.6. Let us fix a particular choice of n. Think of an outcome of the experiment as an ordering of the values of the n contestants, so that there are $n!$ equally likely outcomes. The event $\{N = n\}$ occurs if and only if the first contestant's number is smallest among the first $n-1$ contestants, and contestant n's number is the smallest among the first n contestants. This event can occur in $(n-2)!$ different ways, namely, all the possible ways of ordering contestants $2, \ldots, n-1$. Thus, the probability of this event is $(n-2)!/n! = 1/(n(n-1))$, in agreement with the previous solution.

Solution to Problem 1.49. A sum of 11 is obtained with the following 6 combinations:

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

A sum of 12 is obtained with the following 6 combinations:

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

Each combination of 3 distinct numbers corresponds to 6 permutations, while each combination of 3 numbers, two of which are equal, corresponds to 3 permutations. Counting the number of permutations in the 6 combinations corresponding to a sum of 11, we obtain $6+6+3+6+3+3 = 27$ permutations. Counting the number of permutations in the 6 combinations corresponding to a sum of 12, we obtain $6+6+3+3+6+1 = 25$ permutations. Since all permutations are equally likely, a sum of 11 is more likely than a sum of 12.

Note also that the sample space has $6^3 = 216$ elements, so we have $\mathbf{P}(11) = 27/216$, $\mathbf{P}(12) = 25/216$.

Solution to Problem 1.50. The sample space consists of all possible choices for the birthday of each person. Since there are n persons, and each has 365 choices for their birthday, the sample space has 365^n elements. Let us now consider those choices of birthdays for which no two persons have the same birthday. Assuming that $n \leq 365$, there are 365 choices for the first person, 364 for the second, etc., for a total of $365 \cdot 364 \cdots (365 - n + 1)$. Thus,

$$\mathbf{P}(\text{no two birthdays coincide}) = \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

It is interesting to note that for n as small as 23, the probability that there are two persons with the same birthday is larger than 1/2.

Solution to Problem 1.51. (a) We number the red balls from 1 to m, and the white balls from $m+1$ to $m+n$. One possible sample space consists of all pairs of integers (i,j) with $1 \leq i,j \leq m+n$ and $i \neq j$. The total number of possible outcomes is $(m+n)(m+n-1)$. The number of outcomes corresponding to red-white selection, (i.e., $i \in \{1, \ldots, m\}$ and $j \in \{m+1, \ldots, m+n\}$) is mn. The number of outcomes corresponding to white-red selection, (i.e., $i \in \{m+1, \ldots, m+n\}$ and $j \in \{1, \ldots, m\}$) is also mn. Thus, the desired probability that the balls are of different color is

$$\frac{2mn}{(m+n)(m+n-1)}.$$

Another possible sample space consists of all the possible ordered color pairs, i.e., $\{RR, RW, WR, WW\}$. We then have to calculate the probability of the event $\{RW, WR\}$. We consider a sequential description of the experiment, i.e., we first select the first ball and then the second. In the first stage, the probability of a red ball is $m/(m+n)$. In the second stage, the probability of a red ball is either $m/(m+n-1)$ or $(m-1)/(m+n-1)$ depending on whether the first ball was white or red, respectively. Therefore, using the multiplication rule, we have

$$\mathbf{P}(RR) = \frac{m}{m+n} \cdot \frac{m-1}{m-1+n}, \qquad \mathbf{P}(RW) = \frac{m}{m+n} \cdot \frac{n}{m-1+n},$$

$$\mathbf{P}(WR) = \frac{n}{m+n} \cdot \frac{m}{m+n-1}, \qquad \mathbf{P}(WW) = \frac{n}{m+n} \cdot \frac{n-1}{m+n-1}.$$

The desired probability is

$$\begin{aligned}\mathbf{P}\big(\{RW, WR\}\big) &= \mathbf{P}(RW) + \mathbf{P}(WR) \\ &= \frac{m}{m+n} \cdot \frac{n}{m-1+n} + \frac{n}{m+n} \cdot \frac{m}{m+n-1} \\ &= \frac{2mn}{(m+n)(m+n-1)}.\end{aligned}$$

(b) We calculate the conditional probability of all balls being red, given any of the possible values of k. We have $\mathbf{P}(R \,|\, k=1) = m/(m+n)$ and, as found in part (a), $\mathbf{P}(RR \,|\, k=2) = m(m-1)/(m+n)(m-1+n)$. Arguing sequentially as in part (a), we also have $\mathbf{P}(RRR \,|\, k=3) = m(m-1)(m-2)/(m+n)(m-1+n)(m-2+n)$. According to the total probability theorem, the desired answer is

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

Solution to Problem 1.52. The probability that the 13th card is the first king to be dealt is the probability that out of the first 13 cards to be dealt, exactly one was a king, and that the king was dealt last. Now, given that exactly one king was dealt in the first 13 cards, the probability that the king was dealt last is just 1/13, since each "position" is equally likely. Thus, it remains to calculate the probability that there was exactly one king in the first 13 cards dealt. To calculate this probability we count the "favorable" outcomes and divide by the total number of possible outcomes. We first count the favorable outcomes, namely those with exactly one king in the first 13 cards dealt. We can choose a particular king in 4 ways, and we can choose the other 12 cards in $\binom{48}{12}$ ways, therefore there are $4 \cdot \binom{48}{12}$ favorable outcomes. There are $\binom{52}{13}$ total outcomes, so the desired probability is

$$\frac{1}{13} \cdot \frac{4 \cdot \dbinom{48}{12}}{\dbinom{52}{13}}.$$

For an alternative solution, we argue as in Example 1.10. The probability that the first card is not a king is 48/52. Given that, the probability that the second is not a king is 47/51. We continue similarly until the 12th card. The probability that the 12th card is not a king, given that none of the preceding 11 was a king, is 37/41. (There are $52-11=41$ cards left, and $48-11=37$ of them are not kings.) Finally, the conditional probability that the 13th card is a king is 4/40. The desired probability is

$$\frac{48 \cdot 47 \cdots 37 \cdot 4}{52 \cdot 51 \cdots 41 \cdot 40}.$$

Solution to Problem 1.53. Suppose we label the classes A, B, and C. The probability that Joe and Jane will both be in class A is the number of possible combinations for class A that involve both Joe and Jane, divided by the total number of combinations for class A. Therefore, this probability is

$$\frac{\binom{88}{28}}{\binom{90}{30}}.$$

Since there are three classes, the probability that Joe and Jane end up in the same class is

$$3 \cdot \frac{\binom{88}{28}}{\binom{90}{30}}.$$

A much simpler solution is as follows. We place Joe in one class. Regarding Jane, there are 89 possible "slots", and only 29 of them place her in the same class as Joe. Thus, the answer is 29/89, which turns out to agree with the answer obtained earlier.

Solution to Problem 1.54. (a) Since the cars are all distinct, there are 20! ways to line them up.

(b) To find the probability that the cars will be parked so that they alternate, we count the number of "favorable" outcomes, and divide by the total number of possible outcomes found in part (a). We count in the following manner. We first arrange the US cars in an ordered sequence (permutation). We can do this in 10! ways, since there are 10 distinct cars. Similarly, arrange the foreign cars in an ordered sequence, which can also be done in 10! ways. Finally, interleave the two sequences. This can be done in two different ways, since we can let the first car be either US-made or foreign. Thus, we have a total of $2 \cdot 10! \cdot 10!$ possibilities, and the desired probability is

$$\frac{2 \cdot 10! \cdot 10!}{20!}.$$

Note that we could have solved the second part of the problem by neglecting the fact that the cars are distinct. Suppose the foreign cars are indistinguishable, and also that the US cars are indistinguishable. Out of the 20 available spaces, we need to choose 10 spaces in which to place the US cars, and thus there are $\binom{20}{10}$ possible outcomes. Out of these outcomes, there are only two in which the cars alternate, depending on whether we start with a US or a foreign car. Thus, the desired probability is $2/\binom{20}{10}$, which coincides with our earlier answer.

Solution to Problem 1.55. We count the number of ways in which we can safely place 8 distinguishable rooks, and then divide this by the total number of possibilities. First we count the number of favorable positions for the rooks. We will place the rooks one by one on the 8×8 chessboard. For the first rook, there are no constraints, so we have 64 choices. Placing this rook, however, eliminates one row and one column. Thus, for the second rook, we can imagine that the illegal column and row have been removed, thus leaving us with a 7×7 chessboard, and with 49 choices. Similarly, for the third rook we have 36 choices, for the fourth 25, etc. In the absence of any restrictions, there are $64 \cdot 63 \cdots 57 = 64!/56!$ ways we can place 8 rooks, so the desired probability is

$$\frac{64 \cdot 49 \cdot 36 \cdot 25 \cdot 16 \cdot 9 \cdot 4}{\frac{64!}{56!}}.$$

Solution to Problem 1.56. (a) There are $\binom{8}{4}$ ways to pick 4 lower level classes, and $\binom{10}{3}$ ways to choose 3 higher level classes, so there are

$$\binom{8}{4}\binom{10}{3}$$

valid curricula.

(b) This part is more involved. We need to consider several different cases:

(i) Suppose we do not choose L_1. Then both L_2 and L_3 must be chosen; otherwise no higher level courses would be allowed. Thus, we need to choose 2 more lower level classes out of the remaining 5, and 3 higher level classes from the available 5. We then obtain $\binom{5}{2}\binom{5}{3}$ valid curricula.

(ii) If we choose L_1 but choose neither L_2 nor L_3, we have $\binom{5}{3}\binom{5}{3}$ choices.

(iii) If we choose L_1 and choose one of L_2 or L_3, we have $2 \cdot \binom{5}{2}\binom{5}{3}$ choices. This is because there are two ways of choosing between L_2 and L_3, $\binom{5}{2}$ ways of choosing 2 lower level classes from $L_4, \ldots, L_8$, and $\binom{5}{3}$ ways of choosing 3 higher level classes from $H_1, \ldots, H_5$.

(iv) Finally, if we choose L_1, L_2, and L_3, we have $\binom{5}{1}\binom{10}{3}$ choices.

Note that we are not double counting, because there is no overlap in the cases we are considering, and furthermore we have considered every possible choice. The total is obtained by adding the counts for the above four cases.

Solution to Problem 1.57. Let us fix the order in which letters appear in the sentence. There are 26! choices, corresponding to the possible permutations of the 26-letter alphabet. Having fixed the order of the letters, we need to separate them into words. To obtain 6 words, we need to place 5 separators ("blanks") between the letters. With 26 letters, there are 25 possible positions for these blanks, and the number of choices is $\binom{25}{5}$. Thus, the desired number of sentences is $25!\binom{25}{5}$. Generalizing, the number of sentences consisting of w nonempty words using exactly once each letter from a l-letter alphabet is equal to

$$l!\binom{l-1}{w-1}.$$

Solution to Problem 1.58. (a) The sample space consists of all ways of drawing 7 elements out of a 52-element set, so it contains $\binom{52}{7}$ possible outcomes. Let us count those outcomes that involve exactly 3 aces. We are free to select any 3 out of the 4 aces, and any 4 out of the 48 remaining cards, for a total of $\binom{4}{3}\binom{48}{4}$ choices. Thus,

$$\mathbf{P}(\text{7 cards include exactly 3 aces}) = \frac{\dbinom{4}{3}\dbinom{48}{4}}{\dbinom{52}{7}}.$$

(b) Proceeding similar to part (a), we obtain

$$\mathbf{P}(\text{7 cards include exactly 2 kings}) = \frac{\dbinom{4}{2}\dbinom{48}{5}}{\dbinom{52}{7}}.$$

(c) If A and B stand for the events in parts (a) and (b), respectively, we are looking for $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$. The event $A \cap B$ (having exactly 3 aces and exactly 2 kings) can occur by choosing 3 out of the 4 available aces, 2 out of the 4 available kings, and 2 more cards out of the remaining 44. Thus, this event consists of $\binom{4}{3}\binom{4}{2}\binom{44}{2}$ distinct outcomes. Hence,

$$\mathbf{P}(\text{7 cards include 3 aces and/or 2 kings}) = \frac{\dbinom{4}{3}\dbinom{48}{4} + \dbinom{4}{2}\dbinom{48}{5} - \dbinom{4}{3}\dbinom{4}{2}\dbinom{44}{2}}{\dbinom{52}{7}}.$$

Solution to Problem 1.59. Clearly if $n > m$, or $n > k$, or $m - n > 100 - k$, the probability must be zero. If $n \le m$, $n \le k$, and $m - n \le 100 - k$, then we can find the probability that the test drive found n of the 100 cars defective by counting the total number of size m subsets, and then the number of size m subsets that contain n lemons. Clearly, there are $\binom{100}{m}$ different subsets of size m. To count the number of size m subsets with n lemons, we first choose n lemons from the k available lemons, and then choose $m - n$ good cars from the $100 - k$ available good cars. Thus, the number of ways to choose a subset of size m from 100 cars, and get n lemons, is

$$\binom{k}{n}\binom{100-k}{m-n},$$

and the desired probability is

$$\frac{\binom{k}{n}\binom{100-k}{m-n}}{\binom{100}{m}}.$$

Solution to Problem 1.60. The size of the sample space is the number of different ways that 52 objects can be divided in 4 groups of 13, and is given by the multinomial formula

$$\frac{52!}{13!\,13!\,13!\,13!}.$$

There are 4! different ways of distributing the 4 aces to the 4 players, and there are

$$\frac{48!}{12!\,12!\,12!\,12!}$$

different ways of dividing the remaining 48 cards into 4 groups of 12. Thus, the desired probability is

$$\frac{4!\dfrac{48!}{12!\,12!\,12!\,12!}}{\dfrac{52!}{13!\,13!\,13!\,13!}}.$$

An alternative solution can be obtained by considering a different, but probabilistically equivalent method of dealing the cards. Each player has 13 slots, each one of which is to receive one card. Instead of shuffling the deck, we place the 4 aces at the top, and start dealing the cards one at a time, with each free slot being equally likely to receive the next card. For the event of interest to occur, the first ace can go anywhere; the second can go to any one of the 39 slots (out of the 51 available) that correspond to players that do not yet have an ace; the third can go to any one of the 26 slots (out of the 50 available) that correspond to the two players that do not yet have an ace; and finally, the fourth, can go to any one of the 13 slots (out of the 49 available) that correspond to the only player who does not yet have an ace. Thus, the desired probability is

$$\frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49}.$$

By simplifying our previous answer, it can be checked that it is the same as the one obtained here, thus corroborating the intuitive fact that the two different ways of dealing the cards are probabilistically equivalent.

CHAPTER 2

Solution to Problem 2.1. Let X be the number of points the MIT team earns over the weekend. We have

$$
\begin{aligned}
\mathbf{P}(X=0) &= 0.6\cdot 0.3 = 0.18,\\
\mathbf{P}(X=1) &= 0.4\cdot 0.5\cdot 0.3 + 0.6\cdot 0.5\cdot 0.7 = 0.27,\\
\mathbf{P}(X=2) &= 0.4\cdot 0.5\cdot 0.3 + 0.6\cdot 0.5\cdot 0.7 + 0.4\cdot 0.5\cdot 0.7\cdot 0.5 = 0.34,\\
\mathbf{P}(X=3) &= 0.4\cdot 0.5\cdot 0.7\cdot 0.5 + 0.4\cdot 0.5\cdot 0.7\cdot 0.5 = 0.14,\\
\mathbf{P}(X=4) &= 0.4\cdot 0.5\cdot 0.7\cdot 0.5 = 0.07,\\
\mathbf{P}(X>4) &= 0.
\end{aligned}
$$

Solution to Problem 2.2. The number of guests that have the same birthday as you is binomial with $p = 1/365$ and $n = 499$. Thus the probability that exactly one other guest has the same birthday is

$$
\binom{499}{1}\frac{1}{365}\left(\frac{364}{365}\right)^{498} \approx 0.3486.
$$

Let $\lambda = np = 499/365 \approx 1.367$. The Poisson approximation is $e^{-\lambda}\lambda = e^{-1.367}\cdot 1.367 \approx 0.3483$, which closely agrees with the correct probability based on the binomial.

Solution to Problem 2.3. (a) Let L be the duration of the match. If Fischer wins a match consisting of L games, then $L-1$ draws must first occur before he wins. Summing over all possible lengths, we obtain

$$
\mathbf{P}(\text{Fischer wins}) = \sum_{l=1}^{10}(0.3)^{l-1}(0.4) = 0.571425.
$$

(b) The match has length L with $L < 10$, if and only if $(L-1)$ draws occur, followed by a win by either player. The match has length $L = 10$ if and only if 9 draws occur. The probability of a win by either player is 0.7. Thus

$$
p_L(l) = \mathbf{P}(L=l) = \begin{cases} (0.3)^{l-1}(0.7), & l = 1,\ldots,9,\\ (0.3)^9, & l = 10,\\ 0, & \text{otherwise.}\end{cases}
$$

Solution to Problem 2.4. (a) Let X be the number of modems in use. For $k < 50$, the probability that $X = k$ is the same as the probability that k out of 1000 customers need a connection:

$$
p_X(k) = \binom{1000}{k}(0.01)^k(0.99)^{1000-k}, \qquad k = 0, 1, \ldots, 49.
$$

The probability that $X = 50$, is the same as the probability that 50 or more out of 1000 customers need a connection:

$$p_X(50) = \sum_{k=50}^{1000} \binom{1000}{k} (0.01)^k (0.99)^{1000-k}.$$

(b) By approximating the binomial with a Poisson with parameter $\lambda = 1000 \cdot 0.01 = 10$, we have

$$p_X(k) = e^{-10} \frac{10^k}{k!}, \qquad k = 0, 1, \ldots, 49,$$

$$p_X(50) = \sum_{k=50}^{1000} e^{-10} \frac{10^k}{k!}.$$

(c) Let A be the event that there are more customers needing a connection than there are modems. Then,

$$\mathbf{P}(A) = \sum_{k=51}^{1000} \binom{1000}{k} (0.01)^k (0.99)^{1000-k}.$$

With the Poisson approximation, $\mathbf{P}(A)$ is estimated by

$$\sum_{k=51}^{1000} e^{-10} \frac{10^k}{k!}.$$

Solution to Problem 2.5. (a) Let X be the number of packets stored at the end of the first slot. For $k < b$, the probability that $X = k$ is the same as the probability that k packets are generated by the source:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots, b-1,$$

while

$$p_X(b) = \sum_{k=b}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - \sum_{k=0}^{b-1} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Let Y be the number of number of packets stored at the end of the second slot. Since $\min\{X, c\}$ is the number of packets transmitted in the second slot, we have $Y = X - \min\{X, c\}$. Thus,

$$p_Y(0) = \sum_{k=0}^{c} p_X(k) = \sum_{k=0}^{c} e^{-\lambda} \frac{\lambda^k}{k!},$$

$$p_Y(k) = p_X(k+c) = e^{-\lambda} \frac{\lambda^{k+c}}{(k+c)!}, \qquad k = 1, \ldots, b-c-1,$$

$$p_Y(b-c) = p_X(b) = 1 - \sum_{k=0}^{b-1} e^{-\lambda} \frac{\lambda^k}{k!}.$$

(b) The probability that some packets get discarded during the first slot is the same as the probability that more than b packets are generated by the source, so it is equal to

$$\sum_{k=b+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!},$$

or

$$1 - \sum_{k=0}^{b} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Solution to Problem 2.6. We consider the general case of part (b), and we show that $p > 1/2$ is a necessary and sufficient condition for $n = 2k+1$ games to be better than $n = 2k-1$ games. To prove this, let N be the number of Celtics' wins in the first $2k-1$ games. If A denotes the event that the Celtics win with $n = 2k+1$, and B denotes the event that the Celtics win with $n = 2k-1$, then

$$\mathbf{P}(A) = \mathbf{P}(N \geq k+1) + \mathbf{P}(N = k) \cdot \big(1 - (1-p)^2\big) + \mathbf{P}(N = k-1) \cdot p^2,$$

$$\mathbf{P}(B) = \mathbf{P}(N \geq k) = \mathbf{P}(N = k) + \mathbf{P}(N \geq k+1),$$

and therefore

$$\begin{aligned}
\mathbf{P}(A) - \mathbf{P}(B) &= \mathbf{P}(N = k-1) \cdot p^2 - \mathbf{P}(N = k) \cdot (1-p)^2 \\
&= \binom{2k-1}{k-1} p^{k-1}(1-p)^k p^2 - \binom{2k-1}{k} (1-p)^2 p^k (1-p)^{k-1} \\
&= \frac{(2k-1)!}{(k-1)!\,k!} p^k (1-p)^k (2p-1).
\end{aligned}$$

It follows that $\mathbf{P}(A) > \mathbf{P}(B)$ if and only if $p > \frac{1}{2}$. Thus, a longer series is better for the better team.

Solution to Problem 2.7. Let random variable X be the number of trials you need to open the door, and let K_i be the event that the ith key selected opens the door.

(a) In case (1), we have

$$\begin{aligned}
p_X(1) &= \mathbf{P}(K_1) = \frac{1}{5}, \\
p_X(2) &= \mathbf{P}(K_1^c)\mathbf{P}(K_2 \,|\, K_1^c) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5}, \\
p_X(3) &= \mathbf{P}(K_1^c)\mathbf{P}(K_2^c \,|\, K_1^c)\mathbf{P}(K_3 \,|\, K_1^c \cap K_2^c) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}.
\end{aligned}$$

Proceeding similarly, we see that the PMF of X is

$$p_X(x) = \frac{1}{5}, \qquad x = 1, 2, 3, 4, 5.$$

We can also view the problem as ordering the keys in advance and then trying them in succession, in which case the probability of any of the five keys being correct is 1/5.

In case (2), X is a geometric random variable with $p = 1/5$, and its PMF is

$$p_X(k) = \frac{1}{5} \cdot \left(\frac{4}{5}\right)^{k-1}, \qquad k \geq 1.$$

(b) In case (1), we have

$$p_X(1) = \mathbf{P}(K_1) = \frac{2}{10},$$
$$p_X(2) = \mathbf{P}(K_1^c)\mathbf{P}(K_2 \,|\, K_1^c) = \frac{8}{10} \cdot \frac{2}{9},$$
$$p_X(3) = \mathbf{P}(K_1^c)\mathbf{P}(K_2^c \,|\, K_1^c)\mathbf{P}(K_3 \,|\, K_1^c \cap K_2^c) = \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{2}{8} = \frac{7}{10} \cdot \frac{2}{9}.$$

Proceeding similarly, we see that the PMF of X is

$$p_X(x) = \frac{2 \cdot (10 - x)}{90}, \qquad x = 1, 2, \ldots, 10.$$

Consider now an alternative line of reasoning to derive the PMF of X. If we view the problem as ordering the keys in advance and then trying them in succession, the probability that the number of trials required is x is the probability that the first $x - 1$ keys do not contain either of the two correct keys and the xth key is one of the correct keys. We can count the number of ways for this to happen and divide by the total number of ways to order the keys to determine $p_X(x)$. The total number of ways to order the keys is 10! For the xth key to be the first correct key, the other key must be among the last $10 - x$ keys, so there are $10 - x$ spots in which it can be located. There are 8! ways in which the other 8 keys can be in the other 8 locations. We must then multiply by two since either of the two correct keys could be in the xth position. We therefore have $2 \cdot 10 - x \cdot 8!$ ways for the xth key to be the first correct one and

$$p_X(x) = \frac{2 \cdot (10 - x)8!}{10!} = \frac{2 \cdot (10 - x)}{90}, \qquad x = 1, 2, \ldots, 10,$$

as before.

In case (2), X is again a geometric random variable with $p = 1/5$.

Solution to Problem 2.8. For $k = 0, 1, \ldots, n - 1$, we have

$$\frac{p_X(k+1)}{p_X(k)} = \frac{\binom{n}{k+1} p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{p}{1-p} \cdot \frac{n-k}{k+1}.$$

Solution to Problem 2.9. For $k = 1, \ldots, n$, we have

$$\frac{p_X(k)}{p_X(k-1)} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1}(1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)} = \frac{(n+1)p - kp}{k - kp}.$$

If $k \leq k^*$, then $k \leq (n+1)p$, or equivalently $k - kp \leq (n+1)p - kp$, so that the above ratio is greater than or equal to 1. It follows that $p_X(k)$ is monotonically nondecreasing. If $k > k^*$, the ratio is less than one, and $p_X(k)$ is monotonically decreasing, as required.

Solution to Problem 2.10. Using the expression for the Poisson PMF, we have, for $k \geq 1$,

$$\frac{p_X(k)}{p_X(k-1)} = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \cdot \frac{(k-1)!}{\lambda^{k-1} \cdot e^{-\lambda}} = \frac{\lambda}{k}.$$

Thus if $k \leq \lambda$ the ratio is greater or equal to 1, and it follows that $p_X(k)$ is monotonically increasing. Otherwise, the ratio is less than one, and $p_X(k)$ is monotonically decreasing, as required.

Solution to Problem 2.13. We will use the PMF for the number of girls among the natural children together with the formula for the PMF of a function of a random variable. Let N be the number of natural children that are girls. Then N has a binomial PMF

$$p_N(k) = \begin{cases} \binom{5}{k} \cdot \left(\frac{1}{2}\right)^5, & \text{if } 0 \leq k \leq 5, \\ 0, & \text{otherwise.} \end{cases}$$

Let G be the number of girls out of the 7 children, so that $G = N + 2$. By applying the formula for the PMF of a function of a random variable, we have

$$p_G(g) = \sum_{\{n \mid n+2=g\}} p_N(n) = p_N(g-2).$$

Thus

$$p_G(g) = \begin{cases} \binom{5}{g-2} \cdot \left(\frac{1}{2}\right)^5, & \text{if } 2 \leq g \leq 7, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 2.14. (a) Using the formula $p_Y(y) = \sum_{\{x \mid x \bmod(3)=y\}} p_X(x)$, we obtain

$$\begin{aligned} p_Y(0) &= p_X(0) + p_X(3) + p_X(6) + p_X(9) = 4/10, \\ p_Y(1) &= p_X(1) + p_X(4) + p_X(7) = 3/10, \\ p_Y(2) &= p_X(2) + p_X(5) + p_X(8) = 3/10, \\ p_Y(y) &= 0, \qquad \text{if } y \notin \{0, 1, 2\}. \end{aligned}$$

(b) Similarly, using the formula $p_Y(y) = \sum_{\{x \mid 5 \bmod(x+1)=y\}} p_X(x)$, we obtain

$$p_Y(y) = \begin{cases} 2/10, & \text{if } y = 0, \\ 2/10, & \text{if } y = 1, \\ 1/10, & \text{if } y = 2, \\ 5/10, & \text{if } y = 5, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 2.15. The random variable Y takes the values $k \ln a$, where $k = 1, \ldots, n$, if and only if $X = a^k$ or $X = a^{-k}$. Furthermore, Y takes the value 0, if and only if $X = 1$. Thus, we have

$$p_Y(y) = \begin{cases} \dfrac{2}{2n+1}, & \text{if } y = \ln a,\ 2\ln a, \ldots, k \ln a, \\ \dfrac{1}{2n+1}, & \text{if } y = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 2.16. (a) The scalar a must satisfy

$$1 = \sum_x p_X(x) = \frac{1}{a} \sum_{x=-3}^{3} x^2,$$

so

$$a = \sum_{x=-3}^{3} x^2 = (-3)^2 + (-2)^2 + (-1)^2 + 1^2 + 2^2 + 3^2 = 28.$$

We also have $\mathbf{E}[X] = 0$ because the PMF is symmetric around 0.

(b) If $z \in \{1, 4, 9\}$, then

$$p_Z(z) = p_X(\sqrt{z}) + p_X(-\sqrt{z}) = \frac{z}{28} + \frac{z}{28} = \frac{z}{14}.$$

Otherwise $p_Z(z) = 0$.

(c) $\text{var}(X) = \mathbf{E}[Z] = \sum_z z p_Z(z) = \sum_{z \in \{1,4,9\}} \frac{z^2}{14} = 7.$

(d) We have

$$\begin{aligned} \text{var}(X) &= \sum_x (x - \mathbf{E}[X])^2 p_X(x) \\ &= 1^2 \cdot \big(p_X(-1) + p_X(1)\big) + 2^2 \cdot \big(p_X(-2) + p_X(2)\big) + 3^2 \cdot \big(p_X(-3) + p_X(3)\big) \\ &= 2 \cdot \frac{1}{28} + 8 \cdot \frac{4}{28} + 18 \cdot \frac{9}{28} \\ &= 7. \end{aligned}$$

Solution to Problem 2.17. If X is the temperature in Celsius, the temperature in Fahrenheit is $Y = 32 + 9X/5$. Therefore,

$$\mathbf{E}[Y] = 32 + 9\mathbf{E}[X]/5 = 32 + 18 = 50.$$

Also

$$\text{var}(Y) = (9/5)^2 \text{var}(X),$$

where var(X), the square of the given standard deviation of X, is equal to 100. Thus, the standard deviation of Y is $(9/5) \cdot 10 = 18$. Hence a normal day in Fahrenheit is one for which the temperature is in the range $[32, 68]$.

Solution to Problem 2.18. We have

$$p_X(x) = \begin{cases} 1/(b-a+1), & \text{if } x = 2^k, \text{ where } a \le k \le b,\ k \text{ integer}, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\mathbf{E}[X] = \sum_{k=a}^{b} \frac{1}{b-a+1} 2^k = \frac{2^a}{b-a+1}(1 + 2 + \cdots + 2^{b-a}) = \frac{2^{b+1} - 2^a}{b-a+1}.$$

Similarly,

$$\mathbf{E}[X^2] = \sum_{k=a}^{b} \frac{1}{b-a+1} (2^k)^2 = \frac{4^{b+1} - 4^a}{3(b-a+1)},$$

and finally

$$\text{var}(X) = \frac{4^{b+1} - 4^a}{3(b-a+1)} - \left(\frac{2^{b+1} - 2^a}{b-a+1}\right)^2.$$

Solution to Problem 2.19. We will find the expected gain for each strategy, by computing the expected number of questions until we find the prize.

(a) With this strategy, the probability of finding the location of the prize with i questions, where $i = 1, \ldots, 8$, is $1/10$. The probability of finding the location with 9 questions is $2/10$. Therefore, the expected number of questions is

$$\frac{2}{10} \cdot 9 + \frac{1}{10} \sum_{i=1}^{8} i = 5.4.$$

(b) It can be checked that for 4 of the 10 possible box numbers, exactly 4 questions will be needed, whereas for 6 of the 10 numbers, 3 questions will be needed. Therefore, with this strategy, the expected number of questions is

$$\frac{4}{10} \cdot 4 + \frac{6}{10} \cdot 3 = 3.4.$$

Solution to Problem 2.20. The number C of candy bars you need to eat is a geometric random variable with parameter p. Thus the mean is $\mathbf{E}[C] = 1/p$, and the variance is $\text{var}(C) = (1-p)/p^2$.

Solution to Problem 2.21. The expected value of the gain for a single game is infinite since if X is your gain, then

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k} = \sum_{k=1}^{\infty} 1 = \infty.$$

Thus if you are faced with the choice of playing for given fee f or not playing at all, and your objective is to make the choice that maximizes your expected net gain, you would be willing to pay any value of f. However, this is in strong disagreement with the behavior of individuals. In fact experiments have shown that most people are willing to pay only about \$20 to \$30 to play the game. The discrepancy is due to a presumption that the amount one is willing to pay is determined by the expected gain. However, expected gain does not take into account a person's attitude towards risk taking.

Solution to Problem 2.22. (a) Let X be the number of tosses until the game is over. Noting that X is geometric with probability of success

$$\mathbf{P}\big(\{HT, TH\}\big) = p(1-q) + q(1-p),$$

we obtain

$$p_X(k) = \big(1 - p(1-q) - q(1-p)\big)^{k-1}\big(p(1-q) + q(1-p)\big), \qquad k = 1, 2, \ldots$$

Therefore

$$\mathbf{E}[X] = \frac{1}{p(1-q) + q(1-p)}$$

and

$$\text{var}(X) = \frac{pq + (1-p)(1-q)}{\big(p(1-q) + q(1-p)\big)^2}.$$

(b) The probability that the last toss of the first coin is a head is

$$\mathbf{P}\big(HT \,|\, \{HT, TH\}\big) = \frac{p(1-q)}{p(1-q) + (1-q)p}.$$

Solution to Problem 2.23. Let X be the total number of tosses.

(a) For each toss after the first one, there is probability 1/2 that the result is the same as in the preceding toss. Thus, the random variable X is of the form $X = Y + 1$, where Y is a geometric random variable with parameter $p = 1/2$. It follows that

$$p_X(k) = \begin{cases} (1/2)^{k-1}, & \text{if } k \geq 2, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\mathbf{E}[X] = \mathbf{E}[Y] + 1 = \frac{1}{p} + 1 = 3.$$

We also have

$$\text{var}(X) = \text{var}(Y) = \frac{1-p}{p^2} = 2.$$

(b) If $k > 2$, there are $k - 1$ sequences that lead to the event $\{X = k\}$. One such sequence is $H \cdots HT$, where $k-1$ heads are followed by a tail. The other $k-2$ possible sequences are of the form $T \cdots TH \cdots HT$, for various lengths of the initial $T \cdots T$

segment. For the case where $k = 2$, there is only one (hence $k - 1$) possible sequence that leads to the event $\{X = k\}$, namely the sequence HT. Therefore, for any $k \geq 2$,

$$\mathbf{P}(X = k) = (k-1)(1/2)^k.$$

It follows that

$$p_X(k) = \begin{cases} (k-1)(1/2)^k, & \text{if } k \geq 2, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\mathbf{E}[X] = \sum_{k=2}^{\infty} k(k-1)(1/2)^k = \sum_{k=1}^{\infty} k(k-1)(1/2)^k = \sum_{k=1}^{\infty} k^2(1/2)^k - \sum_{k=1}^{\infty} k(1/2)^k = 6-2 = 4.$$

We have used here the equalities

$$\sum_{k=1}^{\infty} k(1/2)^k = \mathbf{E}[Y] = 2,$$

and

$$\sum_{k=1}^{\infty} k^2(1/2)^k = \mathbf{E}[Y^2] = \text{var}(Y) + \big(\mathbf{E}[Y]\big)^2 = 2 + 2^2 = 6,$$

where Y is a geometric random variable with parameter $p = 1/2$.

Solution to Problem 2.24. (a) There are 21 integer pairs (x, y) in the region

$$R = \big\{(x, y) \mid -2 \leq x \leq 4,\ -1 \leq y - x \leq 1\big\},$$

so that the joint PMF of X and Y is

$$p_{X,Y}(x, y) = \begin{cases} 1/21, & \text{if } (x, y) \text{ is in } R, \\ 0, & \text{otherwise.} \end{cases}$$

For each x in the range $[-2, 4]$, there are three possible values of Y. Thus, we have

$$p_X(x) = \begin{cases} 3/21, & \text{if } x = -2, -1, 0, 1, 2, 3, 4, \\ 0, & \text{otherwise.} \end{cases}$$

The mean of X is the midpoint of the range $[-2, 4]$:

$$\mathbf{E}[X] = 1.$$

The marginal PMF of Y is obtained by using the tabular method. We have

$$p_Y(y) = \begin{cases} 1/21, & \text{if } y = -3, \\ 2/21, & \text{if } y = -2, \\ 3/21, & \text{if } y = -1, 0, 1, 2, 3, \\ 2/21, & \text{if } y = 4, \\ 1/21, & \text{if } y = 5, \\ 0, & \text{otherwise.} \end{cases}$$

The mean of Y is

$$\mathbf{E}[Y] = \frac{1}{21}\cdot(-3+5) + \frac{2}{21}\cdot(-2+4) + \frac{3}{21}\cdot(-1+1+2+3) = 1.$$

(b) The profit is given by

$$P = 100X + 200Y,$$

so that

$$\mathbf{E}[P] = 100\cdot\mathbf{E}[X] + 200\cdot\mathbf{E}[Y] = 100\cdot 1 + 200\cdot 1 = 300.$$

Solution to Problem 2.25. (a) Since all possible values of (I, J) are equally likely, we have

$$p_{I,J}(i,j) = \begin{cases} \frac{1}{\sum_{k=1}^n m_k}, & \text{if } j \le m_i,\\ 0, & \text{otherwise.} \end{cases}$$

The marginal PMFs are given by

$$p_I(i) = \sum_{j=1}^m p_{I,J}(i,j) = \frac{m_i}{\sum_{k=1}^n m_k}, \qquad i = 1,\ldots,n,$$

$$p_J(j) = \sum_{i=1}^n p_{I,J}(i,j) = \frac{l_j}{\sum_{k=1}^n m_k}, \qquad j = 1,\ldots,m,$$

where l_j is the number of students that have answered question j, i.e., students i with $j \le m_i$.

(b) The expected value of the score of student i is the sum of the expected values $p_{ij}a + (1-p_{ij})b$ of the scores on questions j with $j = 1,\ldots,m_i$, i.e.,

$$\sum_{j=1}^{m_i}\big(p_{ij}a + (1-p_{ij})b\big).$$

Solution to Problem 2.26. (a) The possible values of the random variable X are the ten numbers $101,\ldots,110$, and the PMF is given by

$$p_X(k) = \begin{cases} \mathbf{P}(X > k-1) - \mathbf{P}(X > k), & \text{if } k = 101,\ldots 110,\\ 0, & \text{otherwise.} \end{cases}$$

We have $\mathbf{P}(X > 100) = 1$ and for $k = 101,\ldots 110$,

$$\begin{aligned} \mathbf{P}(X > k) &= \mathbf{P}(X_1 > k, X_2 > k, X_3 > k)\\ &= \mathbf{P}(X_1 > k)\,\mathbf{P}(X_2 > k)\,\mathbf{P}(X_3 > k)\\ &= \frac{(110-k)^3}{10^3}. \end{aligned}$$

It follows that

$$p_X(k) = \begin{cases} \dfrac{(111-k)^3 - (110-k)^3}{10^3}, & \text{if } k = 101, \ldots 110, \\ 0, & \text{otherwise.} \end{cases}$$

(An alternative solution is based on the notion of a CDF, which will be introduced in Chapter 3.)

(b) Since X_i is uniformly distributed over the integers in the range $[101, 110]$, we have $\mathbf{E}[X_i] = (101 + 110)/2 = 105.5$. The expected value of X is

$$\mathbf{E}[X] = \sum_{k=-\infty}^{\infty} k \cdot p_X(k) = \sum_{k=101}^{110} k \cdot p_x(k) = \sum_{k=101}^{110} k \cdot \frac{(111-k)^3 - (110-k)^3}{10^3}.$$

The above expression can be evaluated to be equal to 103.025. The expected improvement is therefore 105.5 - 103.025 = 2.475.

Solution to Problem 2.31. The marginal PMF p_Y is given by the binomial formula

$$p_Y(y) = \binom{4}{y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{4-y}, \qquad y = 0, 1, \ldots, 4.$$

To compute the conditional PMF $p_{X|Y}$, note that given that $Y = y$, X is the number of 1's in the remaining $4 - y$ rolls, each of which can take the 5 values $1, 3, 4, 5, 6$ with equal probability $1/5$. Thus, the conditional PMF $p_{X|Y}$ is binomial with parameters $4 - y$ and $p = 1/5$:

$$p_{X|Y}(x \,|\, y) = \binom{4-y}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{4-y-x},$$

for all nonnegative integers x and y such that $0 \leq x + y \leq 4$. The joint PMF is now given by

$$\begin{aligned} p_{X,Y}(x, y) &= p_Y(y) p_{X|Y}(x \,|\, y) \\ &= \binom{4}{y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{4-y} \binom{4-y}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{4-y-x}, \end{aligned}$$

for all nonnegative integers x and y such that $0 \leq x + y \leq 4$. For other values of x and y, we have $p_{X,Y}(x, y) = 0$.

Solution to Problem 2.32. Let X_i be the random variable taking the value 1 or 0 depending on whether the first partner of the ith couple has survived or not. Let Y_i be the corresponding random variable for the second partner of the ith couple. Then, we have $S = \sum_{i=1}^m X_i Y_i$, and by using the total expectation theorem,

$$\begin{aligned} \mathbf{E}[S \,|\, A = a] &= \sum_{i=1}^{m} \mathbf{E}[X_i Y_i \,|\, A = a] \\ &= m\mathbf{E}[X_1 Y_1 \,|\, A = a] \\ &= m\mathbf{E}[Y_1 \,|\, X_1 = 1,\ A = a]\mathbf{P}(X_1 = 1 \,|\, A = a) \\ &= m\mathbf{P}(Y_1 = 1 \,|\, X_1 = 1,\ A = a)\mathbf{P}(X_1 = 1 \,|\, A = a). \end{aligned}$$

We have

$$\mathbf{P}(Y_1 = 1 \,|\, X_1 = 1,\, A = a) = \frac{a-1}{2m-1}, \qquad \mathbf{P}(X_1 = 1 \,|\, A = a) = \frac{a}{2m}.$$

Thus

$$\mathbf{E}[S \,|\, A = a] = m\, \frac{a-1}{2m-1} \cdot \frac{a}{2m} = \frac{a(a-1)}{2(2m-1)}.$$

Note that $\mathbf{E}[S \,|\, A = a]$ does not depend on p.

Solution to Problem 2.38. (a) Let X be the number of red lights that Alice encounters. The PMF of X is binomial with $n = 4$ and $p = 1/2$. The mean and the variance of X are $\mathbf{E}[X] = np = 2$ and $\mathrm{var}(X) = np(1-p) = 4 \cdot (1/2) \cdot (1/2) = 1$.

(b) The variance of Alice's commuting time is the same as the variance of the time by which Alice is delayed by the red lights. This is equal to the variance of $2X$, which is $4\mathrm{var}(X) = 4$.

Solution to Problem 2.39. Let X_i be the number of eggs Harry eats on day i. Then, the X_i are independent random variables, uniformly distributed over the set $\{1, \ldots, 6\}$. We have $X = \sum_{i=1}^{10} X_i$, and

$$\mathbf{E}[X] = \mathbf{E}\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} \mathbf{E}[X_i] = 35.$$

Similarly, we have

$$\mathrm{var}(X) = \mathrm{var}\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} \mathrm{var}(X_i),$$

since the X_i are independent. Using the formula of Example 2.6, we have

$$\mathrm{var}(X_i) = \frac{(6-1)(6-1+2)}{12} \approx 2.9167,$$

so that $\mathrm{var}(X) \approx 29.167$.

Solution to Problem 2.40. Associate a success with a paper that receives a grade that has not been received before. Let X_i be the number of papers between the ith success and the $(i+1)$st success. Then we have $X = 1 + \sum_{i=1}^{5} X_i$ and hence

$$\mathbf{E}[X] = 1 + \sum_{i=1}^{5} \mathbf{E}[X_i].$$

After receiving $i-1$ different grades so far ($i-1$ successes), each subsequent paper has probability $(6-i)/6$ of receiving a grade that has not been received before. Therefore, the random variable X_i is geometric with parameter $p_i = (6-i)/6$, so $\mathbf{E}[X_i] = 6/(6-i)$. It follows that

$$\mathbf{E}[X] = 1 + \sum_{i=1}^{5} \frac{6}{6-i} = 1 + 6 \sum_{i=1}^{5} \frac{1}{i} = 14.7.$$

Solution to Problem 2.41. (a) The PMF of X is the binomial PMF with parameters $p = 0.02$ and $n = 250$. The mean is $\mathbf{E}[X] = np = 250 \cdot 0.02 = 5$. The desired probability is

$$\mathbf{P}(X = 5) = \binom{250}{5}(0.02)^5(0.98)^{245} = 0.1773.$$

(b) The Poisson approximation has parameter $\lambda = np = 5$, so the probability in (a) is approximated by

$$e^{-\lambda}\frac{\lambda^5}{5!} = 0.1755.$$

(c) Let Y be the amount of money you pay in traffic tickets during the year. Then

$$\mathbf{E}[Y] = \sum_{i=1}^{5} 50 \cdot \mathbf{E}[Y_i],$$

where Y_i is the amount of money you pay on the ith day. The PMF of Y_i is

$$\mathbf{P}(Y_i = y) = \begin{cases} 0.98, & \text{if } y = 0, \\ 0.01, & \text{if } y = 10, \\ 0.006, & \text{if } y = 20, \\ 0.004, & \text{if } y = 50. \end{cases}$$

The mean is

$$\mathbf{E}[Y_i] = 0.01 \cdot 10 + 0.006 \cdot 20 + 0.004 \cdot 50 = 0.42.$$

The variance is

$$\text{var}(Y_i) = \mathbf{E}[Y_i^2] - \big(\mathbf{E}[Y_i]\big)^2 = 0.01 \cdot (10)^2 + 0.006 \cdot (20)^2 + 0.004 \cdot (50)^2 - (0.42)^2 = 13.22.$$

The mean of Y is

$$\mathbf{E}[Y] = 250 \cdot \mathbf{E}[Y_i] = 105,$$

and using the independence of the random variables Y_i, the variance of Y is

$$\text{var}(Y) = 250 \cdot \text{var}(Y_i) = 3,305.$$

(d) The variance of the sample mean is

$$\frac{p(1-p)}{250}$$

so assuming that $|p - \hat{p}|$ is within 5 times the standard deviation, the possible values of p are those that satisfy $p \in [0, 1]$ and

$$(p - 0.02)^2 \leq \frac{25p(1-p)}{250}.$$

This is a quadratic inequality that can be solved for the interval of values of p. After some calculation, the inequality can be written as $275p^2 - 35p + 0.1 \leq 0$, which holds if and only if $p \in [0.0025, 0.1245]$.

Solution to Problem 2.42. (a) Noting that

$$\mathbf{P}(X_i = 1) = \frac{\text{Area}(S)}{\text{Area}\big([0,1] \times [0,1]\big)} = \text{Area}(S),$$

we obtain

$$\mathbf{E}[S_n] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] = \mathbf{E}[X_i] = \text{Area}(S),$$

and

$$\text{var}(S_n) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(X_i) = \frac{1}{n}\text{var}(X_i) = \frac{1}{n}\big(1 - \text{Area}(S)\big)\text{Area}(S),$$

which tends to zero as n tends to infinity.

(b) We have

$$S_n = \frac{n-1}{n}S_{n-1} + \frac{1}{n}X_n.$$

(c) We can generate S_{10000} (up to a certain precision) as follows :

1. Initialize S to zero.
2. For $i = 1$ to 10000
3. Randomly select two real numbers a and b (up to a certain precision) independently and uniformly from the interval $[0, 1]$.
4. If $(a - 0.5)^2 + (b - 0.5)^2 < 0.25$, set x to 1 else set x to 0.
5. Set $S := (i - 1)S/i + x/i$.
6. Return S.

By running the above algorithm, a value of S_{10000} equal to 0.7783 was obtained (the exact number depends on the random number generator). We know from part (a) that the variance of S_n tends to zero as n tends to infinity, so the obtained value of S_{10000} is an approximation of $\mathbf{E}[S_{10000}]$. But $\mathbf{E}[S_{10000}] = \text{Area}(S) = \pi/4$, this leads us to the following approximation of π:

$$4 \cdot 0.7783 = 3.1132.$$

(d) We only need to modify the test done at step 4. We have to test whether or not $0 \leq \cos \pi a + \sin \pi b \leq 1$. The obtained approximation of the area was 0.3755.

CHAPTER 3

Solution to Problem 3.1. The random variable $Y = g(X)$ is discrete and its PMF is given by

$$p_Y(1) = \mathbf{P}(X \le 1/3) = 1/3, \qquad p_Y(2) = 1 - p_Y(1) = 2/3.$$

Thus,

$$\mathbf{E}[Y] = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 2 = \frac{5}{3}.$$

The same result is obtained using the expected value rule:

$$\mathbf{E}[Y] = \int_0^1 g(x) f_X(x)\, dx = \int_0^{1/3} dx + \int_{1/3}^1 2\, dx = \frac{5}{3}.$$

Solution to Problem 3.2. We have

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^{\infty} \frac{\lambda}{2} e^{-\lambda |x|}\, dx = 2 \cdot \frac{1}{2} \int_0^{\infty} \lambda e^{-\lambda x}\, dx = 2 \cdot \frac{1}{2} = 1,$$

where we have used the fact $\int_0^\infty \lambda e^{-\lambda x} dx = 1$, i.e., the normalization property of the exponential PDF. By symmetry of the PDF, we have $\mathbf{E}[X] = 0$. We also have

$$\mathbf{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{\lambda}{2} e^{-\lambda |x|} dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2},$$

where we have used the fact that the second moment of the exponential PDF is $2/\lambda^2$. Thus

$$\mathrm{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 = 2/\lambda^2.$$

Solution to Problem 3.5. Let $A = bh/2$ be the area of the given triangle, where b is the length of the base, and h is the height of the triangle. From the randomly chosen point, draw a line parallel to the base, and let A_x be the area of the triangle thus formed. The height of this triangle is $h - x$ and its base has length $b(h - x)/h$. Thus $A_x = b(h - x)^2/(2h)$. For $x \in [0, h]$, we have

$$F_X(x) = 1 - \mathbf{P}(X > x) = 1 - \frac{A_x}{A} = 1 - \frac{b(h - x)^2/(2h)}{bh/2} = 1 - \left(\frac{h - x}{h}\right)^2,$$

while $F_X(x) = 0$ for $x < 0$ and $F_X(x) = 1$ for $x > h$.

The PDF is obtained by differentiating the CDF. We have

$$f_X(x) = \frac{dF_X}{dx}(x) = \begin{cases} \dfrac{2(h - x)}{h^2}, & \text{if } 0 \le x \le h, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 3.6. Let X be the waiting time and Y be the number of customers found. For $x < 0$, we have $F_X(x) = 0$, while for $x \geq 0$,

$$F_X(x) = \mathbf{P}(X \leq x) = \frac{1}{2}\mathbf{P}(X \leq x \,|\, Y = 0) + \frac{1}{2}\mathbf{P}(X \leq x \,|\, Y = 1).$$

Since

$$\mathbf{P}(X \leq x \,|\, Y = 0) = 1,$$

$$\mathbf{P}(X \leq x \,|\, Y = 1) = 1 - e^{-\lambda x},$$

we obtain

$$F_X(x) = \begin{cases} \dfrac{1}{2}(2 - e^{-\lambda x}), & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the CDF has a discontinuity at $x = 0$. The random variable X is neither discrete nor continuous.

Solution to Problem 3.7. (a) We first calculate the CDF of X. For $x \in [0, r]$, we have

$$F_X(x) = \mathbf{P}(X \leq x) = \frac{\pi x^2}{\pi r^2} = \left(\frac{x}{r}\right)^2.$$

For $x < 0$, we have $F_X(x) = 0$, and for $x > r$, we have $F_X(x) = 1$. By differentiating, we obtain the PDF

$$f_X(x) = \begin{cases} \dfrac{2x}{r^2}, & \text{if } 0 \leq x \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$\mathbf{E}[X] = \int_0^r \frac{2x^2}{r^2}\,dx = \frac{2r}{3}.$$

Also

$$\mathbf{E}[X^2] = \int_0^r \frac{2x^3}{r^2}\,dx = \frac{r^2}{2},$$

so

$$\text{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 = \frac{r^2}{2} - \frac{4r^2}{9} = \frac{r^2}{18}.$$

(b) Alvin gets a positive score in the range $[1/t, \infty)$ if and only if $X \leq t$, and otherwise he gets a score of 0. Thus, for $s < 0$, the CDF of S is $F_S(s) = 0$. For $0 \leq s < 1/t$, we have

$$F_S(s) = \mathbf{P}(S \leq s) = \mathbf{P}(\text{Alvin's hit is outside the inner circle}) = 1 - \mathbf{P}(X \leq t) = 1 - \frac{t^2}{r^2}.$$

For $1/t < s$, the CDF of S is given by

$$F_S(s) = \mathbf{P}(S \leq s) = \mathbf{P}(X \leq t)\mathbf{P}(S \leq s \,|\, X \leq t) + \mathbf{P}(X > t)\mathbf{P}(S \leq s \,|\, X > t).$$

We have

$$\mathbf{P}(X \le t) = \frac{t^2}{r^2}, \qquad \mathbf{P}(X > t) = 1 - \frac{t^2}{r^2},$$

and since $S = 0$ when $X > t$,

$$\mathbf{P}(S \le s \,|\, X > t) = 1.$$

Furthermore,

$$\mathbf{P}(S \le s \,|\, X \le t) = \mathbf{P}(1/X \le s \,|\, X \le t) = \frac{\mathbf{P}(1/s \le X \le t)}{\mathbf{P}(X \le t)} = \frac{\dfrac{\pi t^2 - \pi(1/s)^2}{\pi r^2}}{\dfrac{\pi t^2}{\pi r^2}} = 1 - \frac{1}{s^2 t^2}.$$

Combining the above equations, we obtain

$$\mathbf{P}(S \le s) = \frac{t^2}{r^2}\left(1 - \frac{1}{s^2 t^2}\right) + 1 - \frac{t^2}{r^2} = 1 - \frac{1}{s^2 r^2}.$$

Collecting the results of the preceding calculations, the CDF of S is

$$F_S(s) = \begin{cases} 0, & \text{if } s < 0, \\ 1 - \dfrac{t^2}{r^2}, & \text{if } 0 \le s < 1/t, \\ 1 - \dfrac{1}{s^2 r^2}, & \text{if } 1/t \le s. \end{cases}$$

Because F_S has a discontinuity at $s = 0$, the random variable S is not continuous.

Solution to Problem 3.8. (a) By the total probability theorem, we have

$$F_X(x) = \mathbf{P}(X \le x) = p\mathbf{P}(Y \le x) + (1-p)\mathbf{P}(Z \le x) = pF_Y(x) + (1-p)F_Z(x).$$

By differentiating, we obtain

$$f_X(x) = pf_Y(x) + (1-p)f_Z(x).$$

(b) Consider the random variable Y that has PDF

$$f_Y(y) = \begin{cases} \lambda e^{\lambda y}, & \text{if } y < 0 \\ 0, & \text{otherwise,} \end{cases}$$

and the random variable Z that has PDF

$$f_Z(z) = \begin{cases} \lambda e^{-\lambda z}, & \text{if } y \ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

We note that the random variables $-Y$ and Z are exponential. Using the CDF of the exponential random variable, we see that the CDFs of Y and Z are given by

$$F_Y(y) = \begin{cases} e^{\lambda y}, & \text{if } y < 0, \\ 1, & \text{if } y \ge 0, \end{cases}$$

$$F_Z(z) = \begin{cases} 0, & \text{if } z < 0, \\ 1 - e^{-\lambda z}, & \text{if } z \geq 0. \end{cases}$$

We have $f_X(x) = pf_Y(x) + (1-p)f_Z(x)$, and consequently $F_X(x) = pF_Y(x) + (1-p)F_Z(x)$. It follows that

$$\begin{aligned} F_X(x) &= \begin{cases} pe^{\lambda x}, & \text{if } x < 0, \\ p + (1-p)(1 - e^{-\lambda x}), & \text{if } x \geq 0, \end{cases} \\ &= \begin{cases} pe^{\lambda x}, & \text{if } x < 0, \\ 1 - (1-p)e^{-\lambda x}, & \text{if } x \geq 0. \end{cases} \end{aligned}$$

Solution to Problem 3.11. (a) X is a standard normal, so by using the normal table, we have $\mathbf{P}(X \leq 1.5) = \Phi(1.5) = 0.9332$. Also $\mathbf{P}(X \leq -1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$.

(b) The random variable $(Y-1)/2$ is obtained by subtracting from Y its mean (which is 1) and dividing by the standard deviation (which is 2), so the PDF of $(Y-1)/2$ is the standard normal.

(c) We have, using the normal table,

$$\begin{aligned} \mathbf{P}(-1 \leq Y \leq 1) &= \mathbf{P}\big(-1 \leq (Y-1)/2 \leq 0\big) \\ &= \mathbf{P}(-1 \leq Z \leq 0) \\ &= \mathbf{P}(0 \leq Z \leq 1) \\ &= \Phi(1) - \Phi(0) \\ &= 0.8413 - 0.5 \\ &= 0.3413, \end{aligned}$$

where Z is a standard normal random variable.

Solution to Problem 3.12. The random variable $Z = X/\sigma$ is a standard normal, so

$$\mathbf{P}(X \geq k\sigma) = \mathbf{P}(Z \geq k) = 1 - \Phi(k).$$

From the normal tables we have

$$\Phi(1) = 0.8413, \qquad \Phi(2) = 0.9772, \qquad \Phi(3) = 0.9986.$$

Thus $\mathbf{P}(X \geq \sigma) = 0.1587$, $\mathbf{P}(X \geq 2\sigma) = 0.0228$, $\mathbf{P}(X \geq 3\sigma) = 0.0014$.

We also have

$$\mathbf{P}\big(|X| \leq k\sigma\big) = \mathbf{P}\big(|Z| \leq k\big) = \Phi(k) - \mathbf{P}(Z \leq -k) = \Phi(k) - \big(1 - \Phi(k)\big) = 2\Phi(k) - 1.$$

Using the normal table values above, we obtain

$$\mathbf{P}(|X| \leq \sigma) = 0.6826, \qquad \mathbf{P}(|X| \leq 2\sigma) = 0.9544, \qquad \mathbf{P}(|X| \leq 3\sigma) = 0.9972,$$

where t is a standard normal random variable.

Solution to Problem 3.13. Let X and Y be the temperature in Celsius and Fahrenheit, respectively, which are related by $X = 5(Y - 32)/9$. Therefore, 59 degrees Fahrenheit correspond to 15 degrees Celsius. So, if Z is a standard normal random variable, we have using $\mathbf{E}[X] = \sigma_X = 10$,

$$\mathbf{P}(Y \le 59) = \mathbf{P}(X \le 15) = \mathbf{P}\left(Z \le \frac{15 - \mathbf{E}[X]}{\sigma_X}\right) = \mathbf{P}(Z \le 0.5) = \Phi(0.5).$$

From the normal tables we have $\Phi(0.5) = 0.6915$, so $\mathbf{P}(Y \le 59) = 0.6915$.

Solution to Problem 3.15. (a) Since the area of the semicircle is $\pi r^2/2$, the joint PDF of X and Y is $f_{X,Y}(x, y) = 2/\pi r^2$, for (x, y) in the semicircle, and $f_{X,Y}(x, y) = 0$, otherwise.

(b) To find the marginal PDF of Y, we integrate the joint PDF over the range of X. For any possible value y of Y, the range of possible values of X is the interval $[-\sqrt{r^2 - y^2}, \sqrt{r^2 - y^2}]$, and we have

$$f_Y(y) = \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{2}{\pi r^2}\,dx = \begin{cases} \dfrac{4\sqrt{r^2 - y^2}}{\pi r^2}, & \text{if } 0 \le y \le r, \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$\mathbf{E}[Y] = \frac{4}{\pi r^2}\int_0^r y\sqrt{r^2 - y^2}\,dy = \frac{4r}{3\pi},$$

where the integration is performed using the substitution $z = r^2 - y^2$.

(c) There is no need to find the marginal PDF f_Y in order to find $\mathbf{E}[Y]$. Let D denote the semicircle. We have, using polar coordinates

$$\mathbf{E}[Y] = \int\int_{(x,y)\in D} y f_{X,Y}(x, y)\,dx\,dy = \int_0^\pi \int_0^r \frac{2}{\pi r^2}\,s(\sin\theta)s\,ds\,d\theta = \frac{4r}{3\pi}.$$

Solution to Problem 3.16. Let A be the event that the needle will cross a horizontal line, and let B be the probability that it will cross a vertical line. From the analysis of Example 3.11, we have that

$$\mathbf{P}(A) = \frac{2l}{\pi a}, \qquad \mathbf{P}(B) = \frac{2l}{\pi b}.$$

Since at most one horizontal (or vertical) line can be crossed, the expected number of horizontal lines crossed is $\mathbf{P}(A)$ [or $\mathbf{P}(B)$, respectively]. Thus the expected number of crossed lines is

$$\mathbf{P}(A) + \mathbf{P}(B) = \frac{2l}{\pi a} + \frac{2l}{\pi b} = \frac{2l(a + b)}{\pi ab}.$$

The probability that at least one line will be crossed is

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

Let X (or Y) be the distance from the needle's center to the nearest horizontal (or vertical) line. Let Θ be the angle formed by the needle's axis and the horizontal lines as in Example 3.11. We have

$$\mathbf{P}(A \cap B) = \mathbf{P}\left(X \leq \frac{l \sin\Theta}{2},\ Y \leq \frac{l\cos\Theta}{2}\right).$$

We model the triple (X, Y, Θ) as uniformly distributed over the set of all (x, y, θ) that satisfy $0 \leq x \leq a/2$, $0 \leq y \leq b/2$, and $0 \leq \theta \leq \pi/2$. Hence, within this set, we have

$$f_{X,Y,\Theta}(x, y, \theta) = \frac{8}{\pi ab}.$$

The probability $\mathbf{P}(A \cap B)$ is

$$\begin{aligned}
\mathbf{P}\big(X \leq (l/2)\sin\Theta,\ Y \leq (l/2)\cos\Theta\big) &= \iint_{\substack{x \leq (l/2)\sin\theta \\ y \leq (l/2)\cos\theta}} f_{X,Y,\Theta}(x, y, \theta)\, dx\, dy\, d\theta \\
&= \frac{8}{\pi ab} \int_0^{\pi/2} \int_0^{(l/2)\cos\theta} \int_0^{(l/2)\sin\theta} dx\, dy\, d\theta \\
&= \frac{2l^2}{\pi ab} \int_0^{\pi/2} \cos\theta\, \sin\theta\, d\theta \\
&= \frac{l^2}{\pi ab}.
\end{aligned}$$

Thus we have

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) = \frac{2l}{\pi a} + \frac{2l}{\pi b} - \frac{l^2}{\pi ab} = \frac{l}{\pi ab}\big(2(a + b) - l\big).$$

Solution to Problem 3.18. (a) We have

$$\mathbf{E}[X] = \int_1^3 \frac{x^2}{4}\, dx = \frac{x^3}{12}\Big|_1^3 = \frac{27}{12} - \frac{1}{12} = \frac{26}{12} = \frac{13}{6},$$

$$\mathbf{P}(A) = \int_2^3 \frac{x}{4}\, dx = \frac{x^2}{8}\Big|_2^3 = \frac{9}{8} - \frac{4}{8} = \frac{5}{8}.$$

We also have

$$\begin{aligned}
f_{X|A}(x) &= \begin{cases} \dfrac{f_X(x)}{\mathbf{P}(A)}, & \text{if } x \in A, \\ 0, & \text{otherwise,} \end{cases} \\
&= \begin{cases} \dfrac{2x}{5}, & \text{if } 2 \leq x \leq 3, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

from which we obtain

$$\mathbf{E}[X \mid A] = \int_2^3 x \cdot \frac{2x}{5}\,dx = \frac{2x^3}{15}\bigg|_2^3 = \frac{54}{15} - \frac{16}{15} = \frac{38}{15}.$$

(b) We have

$$\mathbf{E}[Y] = \mathbf{E}[X^2] = \int_1^3 \frac{x^3}{4}\,dx = 5,$$

and

$$\mathbf{E}[Y^2] = \mathbf{E}[X^4] = \int_1^3 \frac{x^5}{4}\,dx = \frac{91}{3}.$$

Thus,

$$\mathrm{var}(Y) = \mathbf{E}[Y^2] - \big(\mathbf{E}[Y]\big)^2 = \frac{91}{3} - 5^2 = \frac{16}{3}.$$

Solution to Problem 3.19. (a) We have, using the normalization property,

$$\int_1^2 cx^{-2}\,dx = 1,$$

or

$$c = \frac{1}{\displaystyle\int_1^2 x^{-2}\,dx} = 2.$$

(b) We have

$$\mathbf{P}(A) = \int_{1.5}^2 2x^{-2}\,dx = \frac{1}{3},$$

and

$$f_{X|A}(x \mid A) = \begin{cases} 6x^{-2}, & \text{if } 1.5 < x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

(c) We have

$$\mathbf{E}[Y \mid A] = \mathbf{E}[X^2 \mid A] = \int_{1.5}^2 6x^{-2}x^2\,dx = 3,$$

$$\mathbf{E}[Y^2 \mid A] = \mathbf{E}[X^4 \mid A] = \int_{1.5}^2 6x^{-2}x^4\,dx = \frac{37}{4},$$

and

$$\mathrm{var}(Y \mid A) = \frac{37}{4} - 3^2 = \frac{1}{4}.$$

Solution to Problem 3.20. The expected value in question is

$$\begin{aligned}\mathbf{E}[\text{Time}] = &\big(5 + \mathbf{E}[\text{stay of 2nd student}]\big) \cdot \mathbf{P}(\text{1st stays no more than 5 minutes}) \\ &+ \big(\mathbf{E}[\text{stay of 1st} \mid \text{stay of 1st } \ge 5] + \mathbf{E}[\text{stay of 2nd}]\big) \\ &\cdot \mathbf{P}(\text{1st stays more than 5 minutes}).\end{aligned}$$

We have $\mathbf{E}[\text{stay of 2nd student}] = 30$, and, using the memorylessness property of the exponential distribution,

$$\mathbf{E}[\text{stay of 1st} \,|\, \text{stay of 1st } \geq 5] = 5 + \mathbf{E}[\text{stay of 1st}] = 35.$$

Also

$$\mathbf{P}(\text{1st student stays no more than 5 minutes}) = 1 - e^{-5/30},$$

$$\mathbf{P}(\text{1st student stays more than 5 minutes}) = e^{-5/30}.$$

By substitution we obtain

$$\mathbf{E}[\text{Time}] = (5 + 30) \cdot (1 - e^{-5/30}) + (35 + 30) \cdot e^{-5/30} = 35 + 30 \cdot e^{-5/30} = 60.394.$$

Solution to Problem 3.21. (a) We have $f_Y(y) = 1/l$, for $0 \leq y \leq l$. Furthermore, given the value y of Y, the random variable X is uniform in the interval $[0, y]$. Therefore, $f_{X|Y}(x \,|\, y) = 1/y$, for $0 \leq x \leq y$. We conclude that

$$f_{X,Y}(x, y) = f_Y(y) f_{X|Y}(x \,|\, y) = \begin{cases} \dfrac{1}{l} \cdot \dfrac{1}{y}, & 0 \leq x \leq y \leq l, \\ 0, & \text{otherwise.} \end{cases}$$

(b) We have

$$f_X(x) = \int f_{X,Y}(x, y)\, dy = \int_x^l \frac{1}{ly}\, dy = \frac{1}{l} \ln(l/x), \qquad 0 \leq x \leq l.$$

(c) We have

$$\mathbf{E}[X] = \int_0^l x f_X(x)\, dx = \int_0^l \frac{x}{l} \ln(l/x)\, dx = \frac{l}{4}.$$

(d) The fraction Y/l of the stick that is left after the first break, and the further fraction X/Y of the stick that is left after the second break are independent. Furthermore, the random variables Y and X/Y are uniformly distributed over the sets $[0, l]$ and $[0, 1]$, respectively, so that $\mathbf{E}[Y] = l/2$ and $\mathbf{E}[X/Y] = 1/2$. Thus,

$$\mathbf{E}[X] = \mathbf{E}[Y]\, \mathbf{E}\left[\frac{X}{Y}\right] = \frac{l}{2} \cdot \frac{1}{2} = \frac{l}{4}.$$

Solution to Problem 3.22. Define coordinates such that the stick extends from position 0 (the left end) to position 1 (the right end). Denote the position of the first break by X and the position of the second break by Y. With method (ii), we have $X < Y$. With methods (i) and (iii), we assume that $X < Y$ and we later account for the case $Y < X$ by using symmetry.

Under the assumption $X < Y$, the three pieces have lengths X, $Y - X$, and $1 - Y$. In order that they form a triangle, the sum of the lengths of any two pieces must exceed the length of the third piece. Thus they form a triangle if

$$X < (Y - X) + (1 - Y), \qquad (Y - X) < X + (1 - Y), \qquad (1 - Y) < X + (Y - X).$$

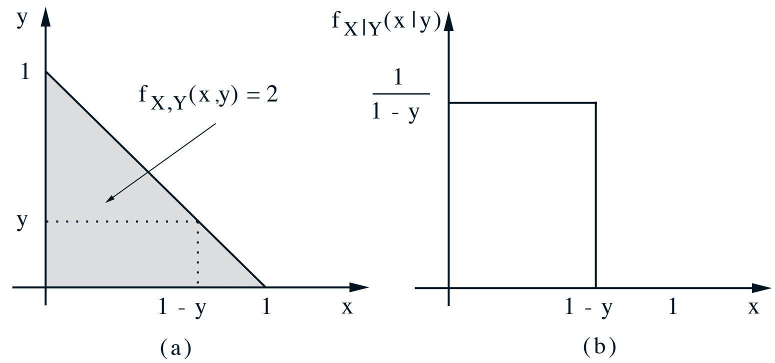

Figure 3.1: (a) The joint PDF. (b) The conditional density of X.

These conditions simplify to

$$X < 0.5, \qquad Y > 0.5, \qquad Y - X < 0.5.$$

Consider first method (i). For X and Y to satisfy these conditions, the pair (X, Y) must lie within the triangle with vertices $(0, 0.5)$, $(0.5, 0.5)$, and $(0.5, 1)$. This triangle has area $1/8$. Thus the probability of the event that the three pieces form a triangle *and* $X < Y$ is $1/8$. By symmetry, the probability of the event that the three pieces form a triangle *and* $X > Y$ is $1/8$. Since there two events are disjoint and form a partition of the event that the three pieces form a triangle, the desired probability is $1/8 + 1/8 = 1/4$.

Consider next method (ii). Since X is uniformly distributed on $[0, 1]$ and Y is uniformly distributed on $[X, 1]$, we have for $0 \le x \le y \le 1$,

$$f_{X,Y}(x, y) = f_X(x)\, f_{Y\,|\,X}(y\,|\,x) = 1 \cdot \frac{1}{1-x}.$$

The desired probability is the probability of the triangle with vertices $(0, 0.5)$, $(0.5, 0.5)$, and $(0.5, 1)$:

$$\int_0^{1/2}\int_{1/2}^{x+1/2} f_{X,Y}(x,y)dydx = \int_0^{1/2}\int_{1/2}^{x+1/2} \frac{1}{1-x}dydx = \int_0^{1/2} \frac{x}{1-x}dydx = -\frac{1}{2}+\ln 2.$$

Consider finally method (iii). Consider first the case $X < 0.5$. Then the larger piece after the first break is the piece on the right. Thus, as in method (ii), Y is uniformly distributed on $[X, 1]$ and the integral above gives the probability of a triangle being formed and $X < 0.5$. Considering also the case $X > 0.5$ doubles the probability, giving a final answer of $-1 + 2\ln 2$.

Solution to Problem 3.23. (a) The area of the triangle is $1/2$, so that $f_{X,Y}(x, y) = 2$, on the triangle indicated in Fig. 3.1(a), and zero everywhere else.

(b) We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{1-y} 2\,dx = 2(1-y), \qquad 0 \le y \le 1.$$

(c) We have

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{1-y}, \qquad 0 \le x \le 1-y.$$

The conditional density is shown in the figure.

Intuitively, since the joint PDF is constant, the conditional PDF (which is a "slice" of the joint, at some fixed y) is also constant. Therefore, the conditional PDF must be a uniform distribution. Given that $Y = y$, X ranges from 0 to $1-y$. Therefore, for the PDF to integrate to 1, its height must be equal to $1/(1-y)$, in agreement with the figure.

(d) For $y > 1$ or $y < 0$, the conditional PDF is undefined, since these values of y are impossible. For $0 \le y < 1$, the conditional mean $\mathbf{E}[X\,|\,Y = y]$ is obtained using the uniform PDF in Fig. 3.1(b), and we have

$$\mathbf{E}[X\,|\,Y = y] = \frac{1-y}{2}, \qquad 0 \le y < 1.$$

For $y = 1$, X must be equal to 0, with certainty, so $\mathbf{E}[X\,|\,Y = 1] = 0$. Thus, the above formula is also valid when $y = 1$. The conditional expectation is undefined when y is outside $[0, 1]$.

The total expectation theorem yields

$$\mathbf{E}[X] = \int_0^1 \frac{1-y}{2} f_Y(y)\,dy = \frac{1}{2} - \frac{1}{2}\int_0^1 y f_Y(y)\,dy = \frac{1-\mathbf{E}[Y]}{2}.$$

(e) Because of symmetry, we must have $\mathbf{E}[X] = \mathbf{E}[Y]$. Therefore, $\mathbf{E}[X] = \big(1-\mathbf{E}[X]\big)/2$, which yields $\mathbf{E}[X] = 1/3$.

Solution to Problem 3.24. The conditional density of X given that $Y = y$ is uniform over the interval $[0, (2-y)/2]$, and we have

$$\mathbf{E}[X\,|\,Y = y] = \frac{2-y}{4}, \qquad 0 \le y \le 2.$$

Therefore, using the total expectation theorem,

$$\mathbf{E}[X] = \int_0^2 \frac{2-y}{4} f_Y(y)\,dy = \frac{2}{4} - \frac{1}{4}\int_0^2 y f_Y(y)\,dy = \frac{2-\mathbf{E}[Y]}{4}.$$

Similarly, the conditional density of Y given that $X = x$ is uniform over the interval $[0, 2(1-x)]$, and we have

$$\mathbf{E}[Y\,|\,X = x] = 1 - x, \qquad 0 \le x \le 1.$$

Therefore,

$$\mathbf{E}[Y] = \int_0^1 (1-x) f_X(x)\, dx = 1 - \mathbf{E}[X].$$

By solving the two equations above for $\mathbf{E}[X]$ and $\mathbf{E}[Y]$, we obtain

$$\mathbf{E}[X] = \frac{1}{3}, \qquad \mathbf{E}[Y] = \frac{2}{3}.$$

Solution to Problem 3.25. Let C denote the event that $X^2 + Y^2 \geq c^2$. The probability $\mathbf{P}(C)$ can be calculated using polar coordinates, as follows:

$$\begin{aligned}
\mathbf{P}(C) &= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} \int_c^{\infty} r e^{-r^2/2\sigma^2}\, dr\, d\theta \\
&= \frac{1}{\sigma^2} \int_c^{\infty} r e^{-r^2/2\sigma^2}\, dr \\
&= e^{-c^2/2\sigma^2}.
\end{aligned}$$

Thus, for $(x, y) \in C$,

$$f_{X,Y|C}(x,y) = \frac{f_{X,Y}(x,y)}{\mathbf{P}(C)} = \frac{1}{2\pi\sigma^2} e^{-\dfrac{1}{2\sigma^2}(x^2+y^2-c^2)}.$$

Solution to Problem 3.34. (a) Let A be the event that the first coin toss resulted in heads. To calculate the probability $\mathbf{P}(A)$, we use the continuous version of the total probability theorem:

$$\mathbf{P}(A) = \int_0^1 \mathbf{P}(A \,|\, P = p) f_P(p)\, dp = \int_0^1 p^2 e^p\, dp,$$

which after some calculation yields

$$\mathbf{P}(A) = e - 2.$$

(b) Using Bayes' rule,

$$\begin{aligned}
f_{P|A}(p) &= \frac{\mathbf{P}(A|P=p) f_P(p)}{\mathbf{P}(A)} \\
&= \begin{cases} \dfrac{p^2 e^p}{e-2}, & 0 \leq p \leq 1, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

(c) Let B be the event that the second toss resulted in heads. We have

$$\begin{aligned}
\mathbf{P}(B \,|\, A) &= \int_0^1 \mathbf{P}(B \,|\, P = p, A) f_{P|A}(p)\, dp \\
&= \int_0^1 \mathbf{P}(B \,|\, P = p) f_{P|A}(p)\, dp \\
&= \frac{1}{e-2} \int_0^1 p^3 e^p\, dp.
\end{aligned}$$

After some calculation, this yields

$$\mathbf{P}(B \mid A) = \frac{1}{e-2} \cdot (6-2e) = \frac{0.564}{0.718} \approx 0.786.$$

CHAPTER 4

Solution to Problem 4.1. Let $Y = \sqrt{|X|}$. We have, for $0 \le y \le 1$,

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(\sqrt{|X|} \le y) = \mathbf{P}(-y^2 \le X \le y^2) = y^2,$$

and therefore by differentiation,

$$f_Y(y) = 2y, \qquad \text{for } 0 \le y \le 1.$$

Let $Y = -\ln|X|$. We have, for $y \ge 0$,

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(\ln|X| \ge -y) = \mathbf{P}(X \ge e^{-y}) + \mathbf{P}(X \le -e^{-y}) = 1 - e^{-y},$$

and therefore by differentiation

$$f_Y(y) = e^{-y}, \qquad \text{for } y \ge 0,$$

so Y is an exponential random variable with parameter 1. This exercise provides a method for simulating an exponential random variable using a sample of a uniform random variable.

Solution to Problem 4.2. Let $Y = e^X$. We first find the CDF of Y, and then take the derivative to find its PDF. We have

$$\mathbf{P}(Y \le y) = \mathbf{P}(e^X \le y) = \begin{cases} \mathbf{P}(X \le \ln y), & \text{if } y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} f_Y(y) &= \begin{cases} \dfrac{d}{dx} F_X(\ln y), & \text{if } y > 0, \\ 0, & \text{otherwise,} \end{cases} \\ &= \begin{cases} \dfrac{1}{y} f_X(\ln y), & \text{if } y > 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

When X is uniform on $[0, 1]$, the answer simplifies to

$$f_Y(y) = \begin{cases} \dfrac{1}{y}, & \text{if } 0 < y \le e, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 4.3. Let $Y = |X|^{1/3}$. We have

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}\big(|X|^{1/3} \le y\big) = \mathbf{P}\big(-y^3 \le X \le y^3\big) = F_X(y^3) - F_X(-y^3),$$

and therefore, by differentiating,

$$f_Y(y) = 3y^2 f_X(y^3) + 3y^2 f_X(-y^3), \qquad \text{for } y > 0.$$

Let $Y = |X|^{1/4}$. We have

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}\big(|X|^{1/4} \le y\big) = \mathbf{P}(-y^4 \le X \le y^4) = F_X(y^4) - F_X(-y^4),$$

and therefore, by differentiating,

$$f_Y(y) = 4y^3 f_X(y^4) + 4y^3 f_X(-y^4), \qquad \text{for } y > 0.$$

Solution to Problem 4.4. We have

$$F_Y(y) = \begin{cases} 0, & \text{if } y \le 0, \\ \mathbf{P}(5 - y \le X \le 5) + \mathbf{P}(20 - y \le X \le 20), & \text{if } 0 \le y \le 5, \\ \mathbf{P}(20 - y \le X \le 20), & \text{if } 5 < y \le 15, \\ 1, & \text{if } y > 15. \end{cases}$$

Using the CDF of X, we have

$$\mathbf{P}(5 - y \le X \le 5) = F_X(5) - F_X(5 - y),$$

$$\mathbf{P}(20 - y \le X \le 20) = F_X(20) - F_X(20 - y).$$

Thus,

$$F_Y(y) = \begin{cases} 0, & \text{if } y \le 0, \\ F_X(5) - F_X(5 - y) + F_X(20) - F_X(20 - y), & \text{if } 0 \le y \le 5, \\ F_X(20) - F_X(20 - y), & \text{if } 5 < y \le 15, \\ 1, & \text{if } y > 15. \end{cases}$$

Differentiating, we obtain

$$f_Y(y) = \begin{cases} f_X(5 - y) + f_X(20 - y), & \text{if } 0 \le y \le 5, \\ f_X(20 - y), & \text{if } 5 < y \le 15, \\ 0, & \text{otherwise,} \end{cases}$$

consistent with the result of Example 3.14.

Solution to Problem 4.5. Let $Z = |X - Y|$. We have

$$F_Z(z) = P\big(|X - Y| \le z\big) = 1 - (1 - z)^2.$$

(To see this, draw the event of interest as a subset of the unit square and calculate its area.) Taking derivatives, the desired PDF is

$$f_Z(z) = \begin{cases} 2(1 - z), & \text{if } 0 \le z \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 4.6. Let $Z = |X - Y|$. To find the CDF, we integrate the joint PDF of X and Y over the region where $|X - Y| \leq z$ for a given z. In the case where $z \leq 0$ or $z \geq 1$, the CDF is 0 and 1, respectively. In the case where $0 < z < 1$, we have

$$F_Z(z) = \mathbf{P}(X - Y \leq z,\, X \geq Y) + \mathbf{P}(Y - X \leq z,\, X < Y).$$

The events $\{X - Y \leq z,\, X \geq Y\}$ and $\{Y - X \leq z,\, X < Y\}$ can be identified with subsets of the given triangle. After some calculation using triangle geometry, the areas of these subsets can be verified to be $z/2 + z^2/4$ and $1/4 - (1-z)^2/4$, respectively. Therefore, since $f_{X,Y}(x, y) = 1$ for all (x, y) in the given triangle,

$$F_Z(z) = \left(\frac{z}{2} + \frac{z^2}{4}\right) + \left(\frac{1}{4} - \frac{(1-z)^2}{4}\right) = z.$$

Thus,

$$F_Z(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ z, & \text{if } 0 < z < 1, \\ 1, & \text{if } z \geq 1. \end{cases}$$

By taking the derivative with respect to z, we obtain

$$f_Z(z) = \begin{cases} 1, & \text{if } 0 \leq z \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 4.7. Let X and Y be the two points, and let $Z = \max\{X, Y\}$. For any $t \in [0, 1]$, we have

$$\mathbf{P}(Z \leq t) = \mathbf{P}(X \leq t)\mathbf{P}(Y \leq t) = t^2,$$

and by differentiating, the corresponding PDF is

$$f_Z(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ 2z, & \text{if } 0 \leq z \leq 1, \\ 0, & \text{if } z \geq 1. \end{cases}$$

Thus, we have

$$\mathbf{E}[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz = \int_0^1 2z^2 dz = \frac{2}{3}.$$

The distance of the largest of the two points to the right endpoint is $1 - Z$, and its expected value is $1 - \mathbf{E}[Z] = 1/3$. A symmetric argument shows that the distance of the smallest of the two points to the left endpoint is also $1/3$. Therefore, the expected distance between the two points must also be $1/3$.

Solution to Problem 4.8. Note that $f_X(x)$ and $f_Y(z - x)$ are nonzero only when $x \geq 0$ and $x \leq z$, respectively. Thus, in the convolution formula, we only need to integrate for x ranging from 0 to z:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx = \lambda^2 e^{-z} \int_0^z dx = \lambda^2 z e^{-\lambda z}.$$

Solution to Problem 4.9. Let $Z = X - Y$. We will first calculate the CDF $F_Z(z)$ by considering separately the cases $z \geq 0$ and $z < 0$. For $z \geq 0$, we have (see the left side of Fig. 4.6)

$$\begin{aligned}
F_Z(z) &= \mathbf{P}(X - Y \leq z) \\
&= 1 - \mathbf{P}(X - Y > z) \\
&= 1 - \int_0^\infty \left(\int_{z+y}^\infty f_{X,Y}(x,y)\,dx \right) dy \\
&= 1 - \int_0^\infty \mu e^{-\mu y} \left(\int_{z+y}^\infty \lambda e^{-\lambda x}\,dx \right) dy \\
&= 1 - \int_0^\infty \mu e^{-\mu y} e^{-\lambda(z+y)}\,dy \\
&= 1 - e^{-\lambda z} \int_0^\infty \mu e^{-(\lambda+\mu)y}\,dy \\
&= 1 - \frac{\mu}{\lambda+\mu} e^{-\lambda z}.
\end{aligned}$$

For the case $z < 0$, we have using the preceding calculation

$$F_Z(z) = 1 - F_Z(-z) = 1 - \left(1 - \frac{\lambda}{\lambda+\mu} e^{-\mu(-z)} \right) = \frac{\lambda}{\lambda+\mu} e^{\mu z}.$$

Combining the two cases $z \geq 0$ and $z < 0$, we obtain

$$F_Z(z) = \begin{cases} 1 - \dfrac{\mu}{\lambda+\mu} e^{-\lambda z}, & \text{if } z \geq 0, \\ \dfrac{\lambda}{\lambda+\mu} e^{\mu z}, & \text{if } z < 0. \end{cases}$$

The PDF of Z is obtained by differentiating its CDF. We have

$$f_Z(z) = \begin{cases} \dfrac{\lambda\mu}{\lambda+\mu} e^{-\lambda z}, & \text{if } z \geq 0, \\ \dfrac{\lambda\mu}{\lambda+\mu} e^{\mu z}, & \text{if } z < 0. \end{cases}$$

For an alternative solution, fix some $z \geq 0$ and note that $f_Y(x - z)$ is nonzero only when $x \geq z$. Thus,

$$\begin{aligned}
f_{X-Y}(z) &= \int_{-\infty}^\infty f_X(x) f_Y(x - z)\,dx \\
&= \int_z^\infty \lambda e^{-\lambda x} \mu e^{-\mu(x-z)}\,dx \\
&= \lambda\mu e^{\lambda z} \int_z^\infty e^{-(\lambda+\mu)x}\,dx \\
&= \lambda\mu e^{\lambda z} \frac{1}{\lambda+\mu} e^{-(\lambda+\mu)z} \\
&= \frac{\lambda\mu}{\lambda+\mu} e^{-\mu z},
\end{aligned}$$

in agreement with the earlier answer. The solution for the case $z < 0$ is obtained with a similar calculation.

Solution to Problem 4.10. We first note that the range of possible values of Z are the integers from the range $[1, 5]$. Thus we have

$$p_Z(z) = 0, \quad \text{if } z \neq 1, 2, 3, 4, 5.$$

We calculate $p_Z(z)$ for each of the values $z = 1, 2, 3, 4, 5$, using the convolution formula. We have

$$p_Z(1) = \sum_x p_X(x)p_Y(1-x) = p_X(1)p_Y(0) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6},$$

where the second equality above is based on the fact that for $x \neq 1$ either $p_X(x)$ or $p_Y(1-x)$ (or both) is zero. Similarly, we obtain

$$p_Z(2) = p_X(1)p_Y(1) + p_X(2)p_Y(0) = \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{18},$$

$$p_Z(3) = p_X(1)p_Y(2) + p_X(2)p_Y(1) + p_X(3)p_Y(0) = \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{3},$$

$$p_Z(4) = p_X(2)p_Y(2) + p_X(3)p_Y(1) = \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{6},$$

$$p_Z(5) = p_X(3)p_Y(2) = \frac{1}{3} \cdot \frac{1}{6} = \frac{1}{18}.$$

Solution to Problem 4.11. The convolution of two Poisson PMFs is of the form

$$\sum_{i=0}^{k} \frac{\lambda^i e^{-\lambda}}{i!} \cdot \frac{\mu^{k-i} e^{-\mu}}{(k-i)!} = e^{-(\lambda+\mu)} \sum_{i=0}^{k} \frac{\lambda^i \mu^{k-i}}{i!\,(k-i)!}.$$

We have

$$(\lambda+\mu)^k = \sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i} = \sum_{i=0}^{k} \frac{k!}{i!\,(k-i)!} \lambda^i \mu^{k-i}.$$

Thus, the desired PMF is

$$\frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^{k} \frac{k!\,\lambda^i \mu^{k-i}}{i!\,(k-i)!} = \frac{e^{-(\lambda+\mu)}}{k!} (\lambda+\mu)^k,$$

which is a Poisson PMF with mean $\lambda + \mu$.

Solution to Problem 4.12. Let $V = X + Y$. As in Example 4.10, the PDF of V is

$$f_V(v) = \begin{cases} v, & 0 \leq v \leq 1, \\ 2 - v, & 1 \leq v \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Let $W = X + Y + Z = V + Z$. We convolve the PDFs f_V and f_Z, to obtain

$$f_W(w) = \int f_V(v) f_Z(w-v)\, dv.$$

We first need to determine the limits of the integration. Since $f_V(v) = 0$ outside the range $0 \le v \le 2$, and $f_W(w - v) = 0$ outside the range $0 \le w - v \le 1$, we see that the integrand can be nonzero only if

$$0 \le v \le 2, \qquad \text{and} \qquad w - 1 \le v \le w.$$

We consider three separate cases. If $w \le 1$, we have

$$f_W(w) = \int_0^w f_V(v) f_Z(w - v)\, dv = \int_0^w v\, dv = \frac{w^2}{2}.$$

If $1 \le w \le 2$, we have

$$\begin{aligned} f_W(w) &= \int_{w-1}^{w} f_V(v) f_Z(w - v)\, dv \\ &= \int_{w-1}^{1} v\, dv + \int_1^w (2 - v)\, dv \\ &= \frac{1}{2} - \frac{(w-1)^2}{2} - \frac{(w-2)^2}{2} + \frac{1}{2}. \end{aligned}$$

Finally, if $2 \le w \le 3$, we have

$$f_W(w) = \int_{w-1}^{2} f_V(v) f_Z(w - v)\, dv = \int_{w-1}^{2} (2 - v)\, dv = \frac{(3-w)^2}{2}.$$

To summarize,

$$f_W(w) = \begin{cases} w^2/2, & 0 \le w \le 1, \\ 1 - (w-1)^2/2 - (2-w)^2/2, & 1 \le w \le 2, \\ (3-w)^2/2, & 2 \le w \le 3, \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Problem 4.13. We have $X - Y = X + Z - (a + b)$, where $Z = a + b - Y$ is distributed identically with X and Y. Thus, the PDF of $X + Z$ is the same as the PDF of $X + Y$, and the PDF of $X - Y$ is obtained by shifting the PDF of $X + Y$ to the left by $a + b$.

Solution to Problem 4.14. For all $z \ge 0$, we have, using the independence of X and Y, and the form of the exponential CDF,

$$\begin{aligned} F_Z(z) &= \mathbf{P}\big(\min\{X, Y\} \le z\big) \\ &= 1 - \mathbf{P}\big(\min\{X, Y\} > z\big) \\ &= 1 - \mathbf{P}(X > z,\, Y > z) \\ &= 1 - \mathbf{P}(X > z)\mathbf{P}(Y > z) \\ &= 1 - e^{-\lambda z} e^{-\mu z} \\ &= 1 - e^{-(\lambda + \mu) z}. \end{aligned}$$

This is recognized as the exponential CDF with parameter $\lambda + \mu$. Thus, the minimum of two independent exponentials with parameters λ and μ is an exponential with parameter $\lambda + \mu$.

Solution to Problem 4.17. Because the covariance remains unchanged when we add a constant to a random variable, we can assume without loss of generality that X and Y have zero mean. We then have

$$\text{cov}(X-Y, X+Y) = \mathbf{E}\big[(X-Y)(X+Y)\big] = \mathbf{E}[X^2] - \mathbf{E}[Y^2] = \text{var}(X) - \text{var}(Y) = 0,$$

since X and Y were assumed to have the same variance.

Solution to Problem 4.18. We have

$$\text{cov}(R, S) = \mathbf{E}[RS] - \mathbf{E}[R]\mathbf{E}[S] = \mathbf{E}[WX + WY + X^2 + XY] = \mathbf{E}[X^2] = 1,$$

and

$$\text{var}(R) = \text{var}(S) = 2,$$

so

$$\rho(R, S) = \frac{\text{cov}(R, S)}{\sqrt{\text{var}(R)\text{var}(S)}} = \frac{1}{2}.$$

We also have

$$\text{cov}(R, T) = \mathbf{E}[RT] - \mathbf{E}[R]\mathbf{E}[T] = \mathbf{E}[WY + WZ + XY + XZ] = 0,$$

so that

$$\rho(R, T) = 0.$$

Solution to Problem 4.19. To compute the correlation coefficient

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y},$$

we first compute the covariance:

$$\begin{aligned}
\text{cov}(X, Y) &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] \\
&= \mathbf{E}[aX + bX^2 + cX^3] - \mathbf{E}[X]\mathbf{E}[Y] \\
&= a\mathbf{E}[X] + b\mathbf{E}[X^2] + c\mathbf{E}[X^3] \\
&= b.
\end{aligned}$$

We also have

$$\begin{aligned}
\text{var}(Y) &= \text{var}(a + bX + cX^2) \\
&= \mathbf{E}\big[(a + bX + cX^2)^2\big] - \big(\mathbf{E}[a + bX + cX^2]\big)^2 \\
&= (a^2 + 2ac + b^2 + 3c^2) - (a^2 + c^2 + 2ac) \\
&= b^2 + 2c^2,
\end{aligned}$$

and therefore, using the fact $\mathrm{var}(X) = 1$,

$$\rho(X, Y) = \frac{b}{\sqrt{b^2 + 2c^2}}.$$

Solution to Problem 4.22. If the gambler's fortune at the beginning of a round is a, the gambler bets $a(2p-1)$. He therefore gains $a(2p-1)$ with probability p, and loses $a(2p-1)$ with probability $1-p$. Thus, his expected fortune at the end of a round is

$$a\Big(1 + p(2p-1) - (1-p)(2p-1)\Big) = a\Big(1 + (2p-1)^2\Big).$$

Let X_k be the fortune after the kth round. Using the preceding calculation, we have

$$\mathbf{E}[X_{k+1} \,|\, X_k] = \Big(1 + (2p-1)^2\Big) X_k.$$

Using the law of iterated expectations, we obtain

$$\mathbf{E}[X_{k+1}] = \Big(1 + (2p-1)^2\Big) \mathbf{E}[X_k],$$

and

$$\mathbf{E}[X_1] = \Big(1 + (2p-1)^2\Big) x.$$

We conclude that

$$\mathbf{E}[X_n] = \Big(1 + (2p-1)^2\Big)^n x.$$

Solution to Problem 4.23. (a) Let W be the number of hours that Nat waits. We have

$$\mathbf{E}[X] = \mathbf{P}(0 \le X \le 1)\mathbf{E}[W \,|\, 0 \le X \le 1] + \mathbf{P}(X > 1)\mathbf{E}[W \,|\, X > 1].$$

Since $W > 0$ only if $X > 1$, we have

$$\mathbf{E}[W] = \mathbf{P}(X > 1)\mathbf{E}[W \,|\, X > 1] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

(b) Let D be the duration of a date. We have $\mathbf{E}[D \,|\, 0 \le X \le 1] = 3$. Furthermore, when $X > 1$, the conditional expectation of D given X is $(3 - X)/2$. Hence, using the law of iterated expectations,

$$\mathbf{E}[D \,|\, X > 1] = \mathbf{E}\Big[\mathbf{E}[D \,|\, X] \ \ |\, X > 1\Big] = \mathbf{E}\left[\frac{3-X}{2} \ \Big|\ X > 1\right].$$

Therefore,

$$\begin{aligned}
\mathbf{E}[D] &= \mathbf{P}(0 \le X \le 1)\mathbf{E}[D \,|\, 0 \le X \le 1] + \mathbf{P}(X > 1)\mathbf{E}[D \,|\, X > 1] \\
&= \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot \mathbf{E}\left[\frac{3-X}{2} \ \Big|\ X > 1\right] \\
&= \frac{3}{2} + \frac{1}{2}\left(\frac{3}{2} - \frac{\mathbf{E}[X \,|\, X > 1]}{2}\right) \\
&= \frac{3}{2} + \frac{1}{2}\left(\frac{3}{2} - \frac{3/2}{2}\right) \\
&= \frac{15}{8}.
\end{aligned}$$

(c) The probability that Pat will be late by more than 45 minutes is 1/8. The number of dates before breaking up is the sum of two geometrically distributed random variables with parameter 1/8, and its expected value is $2 \cdot 8 = 16$.

Solution to Problem 4.24. (a) Consider the following two random variables:

$X =$ amount of time the professor devotes to his task [exponentially distributed with parameter $\lambda(y) = 1/(5 - y)$];

$Y =$ length of time between 9 a.m. and his arrival (uniformly distributed between 0 and 4).

Note that $\mathbf{E}[Y] = 2$. We have

$$\mathbf{E}[X \mid Y = y] = \frac{1}{\lambda(y)} = 5 - y,$$

which implies that

$$\mathbf{E}[X \mid Y] = 5 - Y,$$

and

$$\mathbf{E}[X] = \mathbf{E}\big[\mathbf{E}[X \mid Y]\big] = \mathbf{E}[5 - Y] = 5 - \mathbf{E}[Y] = 5 - 2 = 3.$$

(b) Let Z be the length of time from 9 a.m. until the professor completes the task. Then,

$$Z = X + Y.$$

We already know from part (a) that $\mathbf{E}[X] = 3$ and $\mathbf{E}[Y] = 2$, so that

$$\mathbf{E}[Z] = \mathbf{E}[X] + \mathbf{E}[Y] = 3 + 2 = 5.$$

Thus the expected time that the professor leaves his office is 5 hours after 9 a.m.

(c) We define the following random variables:

$W =$ length of time between 9 a.m. and arrival of the Ph.D. student (uniformly distributed between 9 a.m. and 5 p.m.).

$R =$ amount of time the student will spend with the professor, if he finds the professor (uniformly distributed between 0 and 1 hour).

$T =$ amount of time the professor will spend with the student.

Let also F be the event that the student finds the professor.

To find $\mathbf{E}[T]$, we write

$$\mathbf{E}[T] = \mathbf{P}(F)\mathbf{E}[T \mid F] + \mathbf{P}(F^c)\mathbf{E}[T \mid F^c]$$

Using the problem data,

$$\mathbf{E}[T \mid F] = \mathbf{E}[R] = \frac{1}{2}$$

(this is the expected value of a uniformly distribution ranging from 0 to 1),

$$\mathbf{E}[T \mid F^c] = 0$$

(since the student leaves if he does not find the professor). We have

$$\mathbf{E}[T] = \mathbf{E}[T \mid F]\mathbf{P}(F) = \frac{1}{2}\mathbf{P}(F),$$

so we need to find $\mathbf{P}(F)$.

In order for the student to find the professor, his arrival should be between the arrival and the departure of the professor. Thus

$$\mathbf{P}(F) = \mathbf{P}(Y \le W \le X + Y).$$

We have that W can be between 0 (9 a.m.) and 8 (5 p.m.), but $X + Y$ can be any value greater than 0. In particular, it may happen that the sum is greater than the upper bound for W. We write

$$\mathbf{P}(F) = \mathbf{P}(Y \le W \le X + Y) = 1 - \big(\mathbf{P}(W < Y) + \mathbf{P}(W > X + Y)\big)$$

We have

$$\mathbf{P}(W < Y) = \int_0^4 \frac{1}{4} \int_0^y \frac{1}{8}\, dw\, dy = \frac{1}{4}$$

and

$$\begin{aligned}
\mathbf{P}(W > X + Y) &= \int_0^4 \mathbf{P}(W > X + Y \mid Y = y) f_Y(y)\, dy \\
&= \int_0^4 \mathbf{P}(X < W - Y \mid Y = y) f_Y(y)\, dy \\
&= \int_0^4 \int_y^8 F_{X|Y}(w - y) f_W(w) f_Y(y)\, dw\, dy \\
&= \int_0^4 \frac{1}{4} \int_y^8 \frac{1}{8} \int_0^{w-y} \frac{1}{5 - y} e^{-\frac{x}{5-y}}\, dx\, dw\, dy \\
&= \frac{12}{32} + \frac{1}{32} \int_0^4 (5 - y) e^{-\frac{8-y}{5-y}}\, dy.
\end{aligned}$$

Integrating numerically, we have

$$\int_0^4 (5 - y) e^{-\frac{8-y}{5-y}}\, dy = 1.7584.$$

Thus,

$$\mathbf{P}(Y \le W \le X + Y) = 1 - \big(\mathbf{P}(W < Y) + \mathbf{P}(W > X + Y)\big) = 1 - 0.68 = 0.32.$$

The expected amount of time the professor will spend with the student is then

$$\mathbf{E}[T] = \frac{1}{2}\mathbf{P}(F) = \frac{1}{2}\, 0.32 = 0.16 = 9.6 \text{ mins}.$$

Next, we want to find the expected time the professor will leave his office. Let Z be the length of time measured from 9 a.m. until he leaves his office. If the professor doesn't spend any time with the student, then Z will be equal to $X+Y$. On the other hand, if the professor is interrupted by the student, then the length of time will be equal to $X+Y+R$. This is because the professor will spend the same amount of total time on the task regardless of whether he is interrupted by the student. Therefore,

$$\mathbf{E}[Z] = \mathbf{P}(F)\mathbf{E}[Z \mid F] + \mathbf{P}(F^c)\mathbf{E}[Z \mid F^c] = \mathbf{P}(F)\mathbf{E}[X+Y+R] + \mathbf{P}(F^c)\mathbf{E}[X+Y].$$

Using the results of the earlier calculations,

$$\mathbf{E}[X+Y] = 5,$$

$$\mathbf{E}[X+Y+R] = \mathbf{E}[X+Y] + \mathbf{E}[R] = 5 + \frac{1}{2} = \frac{11}{2}.$$

Therefore,

$$\mathbf{E}[Z] = 0.68 \cdot 5 + 0.32 \cdot \frac{11}{2} = 5.16.$$

Thus the expected time the professor will leave his office is 5.16 hours after 9 a.m.

Solution to Problem 4.29. The transform is given by

$$M(s) = \mathbf{E}[e^{sX}] = \frac{1}{2}e^{s} + \frac{1}{4}e^{2s} + \frac{1}{4}e^{3s}.$$

We have

$$\mathbf{E}[X] = \frac{d}{ds}M(s)\Big|_{s=0} = \frac{1}{2} + \frac{2}{4} + \frac{3}{4} = \frac{7}{4},$$

$$\mathbf{E}[X^2] = \frac{d^2}{ds^2}M(s)\Big|_{s=0} = \frac{1}{2} + \frac{4}{4} + \frac{9}{4} = \frac{15}{4},$$

$$\mathbf{E}[X^3] = \frac{d^3}{ds^3}M(s)\Big|_{s=0} = \frac{1}{2} + \frac{8}{4} + \frac{27}{4} = \frac{37}{4}.$$

Solution to Problem 4.30. The transform associated with X is

$$M_X(s) = e^{s^2/2}.$$

By taking derivatives with respect to s, we find that

$$\mathbf{E}[X] = 0, \quad \mathbf{E}[X^2] = 1, \quad \mathbf{E}[X^3] = 0, \quad \mathbf{E}[X^4] = 3.$$

Solution to Problem 4.31. The transform is

$$M(s) = \frac{\lambda}{\lambda - s}.$$

Thus,

$$\frac{d}{ds}M(s) = \frac{\lambda}{(\lambda-s)^2}, \qquad \frac{d^2}{ds^2}M(s) = \frac{2\lambda}{(\lambda-s)^3}, \qquad \frac{d^3}{ds^3}M(s) = \frac{6\lambda}{(\lambda-s)^4},$$

$$\frac{d^4}{ds^4}M(s) = \frac{24\lambda}{(\lambda - s)^5}, \qquad \frac{d^5}{ds^5}M(s) = \frac{120\lambda}{(\lambda - s)^6}.$$

By setting $s = 0$, we obtain

$$\mathbf{E}[X^3] = \frac{6}{\lambda^3}, \qquad \mathbf{E}[X^4] = \frac{24}{\lambda^4}, \qquad \mathbf{E}[X^5] = \frac{120}{\lambda^5}.$$

Solution to Problem 4.32. (a) We must have $M(0) = 1$. Only the first option satisfies this requirement.

(b) We have

$$\mathbf{P}(X = 0) = \lim_{s \to -\infty} M(s) = e^{2(e^{-1}-1)} \approx 0.2825.$$

Solution to Problem 4.33. We recognize this transform as corresponding to the following mixture of exponential PDFs:

$$f_X(x) = \begin{cases} \frac{1}{3} \cdot 2e^{-2x} + \frac{2}{3} \cdot 3e^{-3x}, & \text{for } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

By the inversion theorem, this must be the desired PDF.

Solution to Problem 4.34. For $i = 1, 2, 3$, let X_i, $i = 1, 2, 3$, be a Bernoulli random variable that takes the value 1 if the ith player is successful. We have $X = X_1 + X_2 + X_3$. Let $q_i = 1 - p_i$. Convolution of the PMFs of X_1 and X_2 yields the PMF of $Z = X_1 + X_2$:

$$p_Z(z) = \begin{cases} q_1 q_2, & \text{if } z = 0, \\ q_1 p_2 + p_1 q_2, & \text{if } z = 1, \\ p_1 p_2, & \text{if } z = 2, \\ 0, & \text{otherwise.} \end{cases}$$

Convolution of the PMFs of Z and X_3 yields the PMF of $X = X_1 + X_2 + X_3$:

$$p_X(x) = \begin{cases} q_1 q_2 q_3, & \text{if } x = 0, \\ p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3, & \text{if } x = 1, \\ q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3, & \text{if } x = 2, \\ p_1 p_2 p_3, & \text{if } x = 3, \\ 0, & \text{otherwise.} \end{cases}$$

The transform associated with X is the product of the transforms associated with X_i, $i = 1, 2, 3$. We have

$$M_X(s) = (q_1 + p_1 e^s)(q_2 + p_2 e^s)(q_3 + p_3 e^s).$$

By carrying out the multiplications above, and by examining the coefficients of the terms e^{ks}, we obtain the probabilities $\mathbf{P}(X = k)$. These probabilities are seen to coincide with the ones computed by convolution.

Solution to Problem 4.35. We first find c by using the equation

$$1 = M_X(0) = c \cdot \frac{3+4+2}{3-1},$$

so that $c = 2/9$. We then obtain

$$\mathbf{E}[X] = \frac{dM_X}{ds}(s)\Big|_{s=0} = \frac{2}{9} \cdot \frac{(3-e^s)(8e^{2s}+6e^{3s}) + e^s(3+4e^{2s}+2e^{3s})}{(3-e^s)^2}\Big|_{s=0} = \frac{37}{18}.$$

We now use the identity

$$\frac{1}{3-e^s} = \frac{1}{3} \cdot \frac{1}{1-e^s/3} = \frac{1}{3}\left(1 + \frac{e^s}{3} + \frac{e^{2s}}{9} + \cdots\right),$$

which is valid as long as s is small enough so that $e^s < 3$. It follows that

$$M_X(s) = \frac{2}{9} \cdot \frac{1}{3} \cdot (3 + 4e^{2s} + 2e^{3s}) \cdot \left(1 + \frac{e^s}{3} + \frac{e^{2s}}{9} + \cdots\right).$$

By identifying the coefficients of e^{0s} and e^s, we obtain

$$p_X(0) = \frac{2}{9}, \qquad p_X(1) = \frac{2}{27}.$$

Let $A = \{X \neq 0\}$. We have

$$p_{X|\{X\in A\}}(k) = \begin{cases} \dfrac{p_X(k)}{\mathbf{P}(A)}, & \text{if } k \neq 0, \\ 0, & \text{otherwise,} \end{cases}$$

so that

$$\begin{aligned} \mathbf{E}[X \mid X \neq 0] &= \sum_{k=1}^{\infty} k p_{X|A}(k) \\ &= \sum_{k=1}^{\infty} \frac{k p_X(k)}{\mathbf{P}(A)} \\ &= \frac{\mathbf{E}[X]}{1 - p_X(0)} \\ &= \frac{37/18}{7/9} \\ &= \frac{37}{14}. \end{aligned}$$

Solution to Problem 4.36. (a) We have $U = X$ if $X = 1$, which happens with probability $1/3$, and $U = Z$ if $X = 0$, which happens with probability $2/3$. Therefore, U is a mixture of random variables and the associated transform is

$$M_U(s) = \mathbf{P}(X=1)M_Y(s) + \mathbf{P}(X=0)M_Z(s) = \frac{1}{3} \cdot \frac{2}{2-s} + \frac{2}{3}e^{3(e^s-1)}.$$

(b) Let $V = 2Z + 3$. We have

$$M_V(s) = e^{3s} M_Z(2s) = e^{3s} e^{3(e^{2s}-1)} = e^{3(s-1+e^{2s})}.$$

(c) Let $W = Y + Z$. We have

$$M_W(s) = M_Y(s) M_Z(s) = \frac{2}{2-s} e^{3(e^s-1)}.$$

Solution to Problem 4.37. Let X be the number of different types of pizza ordered. Let X_i be the random variable defined by

$$X_i = \begin{cases} 1, & \text{if a type } i \text{ pizza is ordered by at least one customer,} \\ 0, & \text{otherwise.} \end{cases}$$

We have $X = X_1 + \cdots + X_n$, and by the law of iterated expectations,

$$\mathbf{E}[X] = \mathbf{E}\big[\mathbf{E}[X \mid K]\big] = \mathbf{E}\big[\mathbf{E}[X_1 + \cdots + X_n \mid K]\big] = n\, \mathbf{E}\big[\mathbf{E}[X_1 \mid K]\big].$$

Furthermore, since the probability that a customer does not order a pizza of type 1 is $(n-1)/n$, we have

$$\mathbf{E}[X_1 \mid K = k] = 1 - \left(\frac{n-1}{n}\right)^k,$$

so that

$$\mathbf{E}[X_1 \mid K] = 1 - \left(\frac{n-1}{n}\right)^K.$$

Thus, denoting

$$p = \frac{n-1}{n},$$

we have

$$\mathbf{E}[X] = n\, \mathbf{E}\big[1 - p^K\big] = n - n\, \mathbf{E}\big[p^K\big] = n - n\, \mathbf{E}\big[e^{K \log p}\big] = n - n M_K(\log p).$$

Solution to Problem 4.41. (a) Let N be the number of people that enter the elevator. The corresponding transform is $M_N(s) = e^{\lambda(e^s-1)}$. Let $M_X(s)$ be the common transform associated with the random variables X_i. Since X_i is uniformly distributed within $[0, 1]$, we have

$$M_X(s) = \frac{e^s - 1}{s}.$$

The transform $M_Y(s)$ is found by starting with the transform $M_N(s)$ and replacing each occurrence of e^s with $M_X(s)$. Thus,

$$M_Y(s) = e^{\lambda(M_X(s)-1)} = e^{\lambda\left(\frac{e^s-1}{s}-1\right)}.$$

(b) We have using the chain rule

$$\mathbf{E}[Y] = \frac{d}{ds} M_Y(s)\bigg|_{s=0} = \frac{d}{ds} M_X(s)\bigg|_{s=0} \cdot \lambda e^{\lambda(M_X(s)-1)}\bigg|_{s=0} = \frac{1}{2}\cdot\lambda = \frac{\lambda}{2},$$

where we have used the fact that $M_X(0) = 1$.

(c) From the law of iterated expectations we obtain

$$\mathbf{E}[Y] = \mathbf{E}\big[\mathbf{E}[Y \,|\, N]\big] = \mathbf{E}\big[N\mathbf{E}[X]\big] = \mathbf{E}[N]\mathbf{E}[X] = \frac{\lambda}{2}.$$

Solution to Problem 4.42. Take X and Y to be normal with means 1 and 2, respectively, and very small variances. Consider the random variable that takes the value of X with some probability p and the value of Y with probability $1-p$. This random variable takes values near 1 and 2 with relatively high probability, but takes values near its mean (which is $3-2p$) with relatively low probability. Thus, this random variable is not normal.

Now let N be a random variable taking only the values 1 and 2 with probabilities p and $1-p$, respectively. The sum of a number N of independent normal random variables with mean equal to 1 and very small variance is a mixture of the type discussed above, which is not normal.

Solution to Problem 4.43. (a) Using the total probability theorem, we have

$$\mathbf{P}(X > 4) = \sum_{k=0}^{4} \mathbf{P}(k \text{ lights are red})\mathbf{P}(X > 4 \,|\, k \text{ lights are red}).$$

We have

$$\mathbf{P}(k \text{ lights are red}) = \binom{4}{k}\left(\frac{1}{2}\right)^4.$$

The conditional PDF of X given that k lights are red, is normal with mean k minutes and standard deviation $(1/2)\sqrt{k}$. Thus, X is a mixture of normal random variables and the transform associated with its (unconditional) PDF is the corresponding mixture of the transforms associated with the (conditional) normal PDFs. However, X is not normal, because a mixture of normal PDFs need not be normal. The probability $\mathbf{P}(X > 4 \,|\, k \text{ lights are red})$ can be computed from the normal tables for each k, and $\mathbf{P}(X > 4)$ is obtained by substituting the results in the total probability formula above.

(b) Let K be the number of traffic lights that are found to be red. We can view X as the sum of K independent normal random variables. Thus the transform associated with X can be found by replacing in the binomial transform $M_K(s) = (1/2 + (1/2)e^s)^4$ the occurrence of e^s by the normal transform corresponding to $\mu = 1$ and $\sigma = 1/2$. Thus

$$M_X(s) = \left(\frac{1}{2} + \frac{1}{2}\left(e^{\frac{(1/2)^2 s^2}{2} + s}\right)\right)^4.$$

Note that by using the formula for the transform, we cannot easily obtain the probability $\mathbf{P}(X > 4)$.

Solution to Problem 4.44. (a) Using the random sum formulas, we have

$$\mathbf{E}[N] = \mathbf{E}[M]\,\mathbf{E}[K],$$

$$\mathrm{var}(N) = \mathbf{E}[M]\,\mathrm{var}(K) + \big(\mathbf{E}[K]\big)^2 \mathrm{var}(M).$$

(b) Using the random sum formulas and the results of part (a), we have

$$\mathbf{E}[Y] = \mathbf{E}[N]\,\mathbf{E}[X] = \mathbf{E}[M]\,\mathbf{E}[K]\,\mathbf{E}[X],$$

$$\begin{aligned}\mathrm{var}(Y) &= \mathbf{E}[N]\,\mathrm{var}(X) + \big(\mathbf{E}[X]\big)^2 \mathrm{var}(N)\\ &= \mathbf{E}[M]\,\mathbf{E}[K]\,\mathrm{var}(X) + \big(\mathbf{E}[X]\big)^2 \Big(\mathbf{E}[M]\,\mathrm{var}(K) + \big(\mathbf{E}[K]\big)^2 \mathrm{var}(M)\Big).\end{aligned}$$

(c) Let N denote the total number of widgets in the crate, and let X_i denote the weight of the ith widget. The total weight of the crate is

$$Y = X_1 + \cdots + X_N,$$

with

$$N = K_1 + \cdots + K_M,$$

so the framework of part (b) applies. We have

$$\mathbf{E}[M] = \frac{1}{p}, \qquad \mathrm{var}(M) = \frac{1-p}{p^2}, \qquad \text{(geometric formulas)},$$

$$\mathbf{E}[K] = \mu, \qquad \mathrm{var}(M) = \mu, \qquad \text{(Poisson formulas)},$$

$$\mathbf{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{var}(M) = \frac{1}{\lambda^2}, \qquad \text{(exponential formulas)}.$$

Using these expressions into the formulas of part (b), we obtain $\mathbf{E}[Y]$ and $\mathrm{var}(Y)$, the mean and variance of the total weight of a crate.

C H A P T E R 5

Solution to Problem 5.1. (a) We have $\sigma_{M_n} = 1/\sqrt{n}$, so in order that $\sigma_{M_n} \leq 0.01$, we must have $n \geq 10,000$.

(b) We want to have

$$\mathbf{P}\Big(|M_n - h| \leq 0.05\Big) \geq 0.99.$$

Using the facts $h = \mathbf{E}[M_n]$, $\sigma^2_{M_n} = 1/n$, and the Chebyshev inequality, we have

$$\begin{aligned}
\mathbf{P}\Big(|M_n - h| \leq 0.05\Big) &= \mathbf{P}\Big(|M_n - \mathbf{E}[M_n]| \leq 0.05\Big) \\
&= 1 - \mathbf{P}\Big(|M_n - \mathbf{E}[M_n]| \geq 0.05\Big) \\
&\geq 1 - \frac{1/n}{(0.05)^2}.
\end{aligned}$$

Thus, we must have

$$1 - \frac{1/n}{(0.05)^2} \geq 0.99,$$

which yields $n \geq 40,000$.

(c) Based on Example 5.3, $\sigma^2_{X_i} \leq (0.6)^2/4$, so he should use 0.3 meters in place of 1.0 meters as the estimate of the standard deviation of the samples X_i in the calculations of parts (a) and (b). In the case of part (a), we have $\sigma_{M_n} = 0.3/\sqrt{n}$, so in order that $\sigma_{M_n} \leq 0.01$, we must have $n \geq 900$. In the case of part (b), we have $\sigma_{M_n} = 0.3/\sqrt{n}$, so in order that $\sigma_{M_n} \leq 0.01$, we must have $n \geq 900$. In the case of part (a), we must have

$$1 - \frac{0.09/n}{(0.05)^2} \geq 0.99,$$

which yields $n \geq 3,600$.

Solution to Problem 5.4. Proceeding as in Example 5.5, the best guarantee that can be obtained from the Chebyshev inequality is

$$\mathbf{P}\Big(|M_n - f| \geq \epsilon\Big) \leq \frac{1}{4n\epsilon^2}.$$

(a) If ϵ is reduced to half its original value, and in order to keep the bound $1/(4n\epsilon^2)$ constant, the sample size n must be made four times larger.

(b) If the error probability δ is to be reduced to $\delta/2$, while keeping ϵ the same, the sample size has to be doubled.

Solution to Problem 5.5. In cases (a), (b), and (c), we show that Y_n converges to 0 in probability. In case (d), we show that Y_n converges to 1 in probability.

(a) For any $\epsilon > 0$, we have

$$\mathbf{P}\Big(|Y_n| \geq \epsilon\Big) = 0,$$

for all n with $1/n < \epsilon$, so $\mathbf{P}\big(|Y_n| \geq \epsilon\big) \to 0$.

(b) For all $\epsilon \in (0,1)$, we have

$$\mathbf{P}\big(|Y_n| \geq \epsilon\big) = \mathbf{P}\big(|X_n|^n \geq \epsilon\big) = \mathbf{P}\big(X_n \geq \epsilon^{1/n}\big) + \mathbf{P}\big(X_n \leq -\epsilon^{1/n}\big) = 1 - \epsilon^{1/n},$$

and the two terms in the right-hand side converge to 0, since $\epsilon^{1/n} \to 1$.

(c) Since $X_1, X_2, \ldots$ are independent random variables, we have

$$\mathbf{E}[Y_n] = \mathbf{E}[X_1] \cdots \mathbf{E}[X_n] = 0.$$

Also

$$\mathrm{var}(Y_n) = \mathbf{E}[Y_n^2] = \mathbf{E}[X_1^2] \cdots \mathbf{E}[X_n^2] = \mathrm{var}(X_1)^n = \left(\frac{4}{12}\right)^n,$$

so $\mathrm{var}(Y_n) \to 0$. Since all Y_n have 0 as a common mean, from Chebyshev's inequality it follows that Y_n converges to 0 in probability.

(d) We have for all $\epsilon \in (0,1)$, using the independence of $X_1, X_2, \ldots$,

$$\begin{aligned}
\mathbf{P}\big(|Y_n - 1| \geq \epsilon\big) &= \mathbf{P}\big(\max\{X_1, \ldots, X_n\} \geq 1 + \epsilon\big) + \mathbf{P}\big(\max\{X_1, \ldots, X_n\} \leq 1 - \epsilon\big) \\
&= \mathbf{P}(X_1 \leq 1 - \epsilon, \ldots, X_n \leq 1 - \epsilon) \\
&= \big(\mathbf{P}(X_1 \leq 1 - \epsilon)\big)^n \\
&= \left(1 - \frac{\epsilon}{2}\right)^n.
\end{aligned}$$

Hence $\mathbf{P}\big(|Y_n - 1| \geq \epsilon\big) \to 0$.

Solution to Problem 5.8. Let S be the number of times that the result was odd, which is a binomial random variable, with parameters $n = 100$ and $p = 0.5$, so that $\mathbf{E}[X] = 100 \cdot 0.5 = 50$ and $\sigma_S = \sqrt{100 \cdot 0.5 \cdot 0.5} = \sqrt{25} = 5$. Using the normal approximation to the binomial, we find

$$\mathbf{P}(S > 55) = \mathbf{P}\left(\frac{S - 50}{5} > \frac{55 - 50}{5}\right) \approx 1 - \Phi(1) = 1 - 0.8413 = 0.1587.$$

A better approximation can be obtained by using the de Moivre-Laplace approximation, which yields

$$\begin{aligned}
\mathbf{P}(S > 55) = \mathbf{P}(S \geq 55.5) &= \mathbf{P}\left(\frac{S - 50}{5} > \frac{55.5 - 50}{5}\right) \\
&\approx 1 - \Phi(1.1) = 1 - 0.8643 = 0.1357.
\end{aligned}$$

Solution to Problem 5.9. (a) Let S be the number of crash-free days, which is a binomial random variable with parameters $n = 50$ and $p = 0.95$, so that $\mathbf{E}[X] = 50 \cdot 0.95 = 47.5$ and $\sigma_S = \sqrt{50 \cdot 0.95 \cdot 0.05} = 1.54$. Using the normal approximation to the binomial, we find

$$\mathbf{P}(S \geq 45) = \mathbf{P}\left(\frac{S - 47.5}{1.54} \geq \frac{45 - 47.5}{1.54}\right) \approx 1 - \Phi(-1.62) = \Phi(1.62) = 0.9474.$$

A better approximation can be obtained by using the de Moivre-Laplace approximation, which yields

$$\mathbf{P}(S \geq 45) = \mathbf{P}(S > 44.5) = \mathbf{P}\left(\frac{S - 47.5}{1.54} \geq \frac{44.5 - 47.5}{1.54}\right)$$
$$\approx 1 - \Phi(-1.95) = \Phi(1.95) = 0.9744.$$

(b) The random variable S is binomial with parameter $p = 0.95$. However, the random variable $50 - S$ (the number of crashes) is also binomial with parameter $p = 0.05$. Since the Poisson approximation is exact in the limit of small p and large n, it will give more accurate results if applied to $50 - S$. We will therefore approximate $50 - S$ by a Poisson random variable with parameter $\lambda = 50 \cdot 0.05 = 2.5$. Thus,

$$\begin{aligned}
\mathbf{P}(S \geq 45) &= \mathbf{P}(50 - S \leq 5) \\
&= \sum_{k=0}^{5} \mathbf{P}(n - S = k) \\
&= \sum_{k=0}^{5} e^{-\lambda} \frac{\lambda^k}{k!} \\
&= 0.958.
\end{aligned}$$

It is instructive to compare with the exact probability which is

$$\sum_{k=0}^{5} \binom{50}{k} 0.05^k \cdot 0.95^{50-k} = 0.962.$$

Thus, the Poisson approximation is closer. This is consistent with the intuition that the normal approximation to the binomial works well when p is close to 0.5 or n is very large, which is not the case here. On the other hand, the calculations based on the normal approximation are generally less tedious.

Solution to Problem 5.10. (a) Let $S_n = X_1 + \cdots + X_n$ be the total number of gadgets produced in n days. Note that the mean, variance, and standard deviation of S_n is $5n$, $9n$, and $3\sqrt{n}$, respectively. Thus,

$$\begin{aligned}
\mathbf{P}(S_{100} < 440) &= \mathbf{P}(S_{100} \leq 439.5) \\
&= \mathbf{P}\left(\frac{S_{100} - 500}{30} < \frac{439.5 - 500}{30}\right) \\
&\approx \Phi(\frac{439.5 - 500}{30}) \\
&= \Phi(-2.02) \\
&= 1 - \Phi(2.02) \\
&= 1 - 0.9783 \\
&= 0.0217.
\end{aligned}$$

(b) The requirement $\mathbf{P}(S_n \geq 200 + 5n) \leq 0.05$ translates to

$$\mathbf{P}\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}\right) \leq 0.05,$$

or, using a normal approximation,

$$1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) \leq 0.05,$$

and

$$\Phi\left(\frac{200}{3\sqrt{n}}\right) \geq 0.95.$$

From the normal tables, we obtain $\Phi(1.65) \approx 0.95$, and therefore,

$$\frac{200}{3\sqrt{n}} \geq 1.65,$$

which finally yields $n \leq 1632$.

(c) The event $N \geq 220$ (it takes at least 220 days to exceed 1000 gadgets) is the same as the event $S_{219} \leq 1000$ (no more than 1000 gadgets produced in the first 219 days). Thus,

$$\begin{aligned}
\mathbf{P}(N \geq 220) &= \mathbf{P}(S_{219} \leq 1000) \\
&= \mathbf{P}\left(\frac{S_{219} - 5 \cdot 219}{3\sqrt{219}} \leq \frac{1000 - 5 \cdot 219}{3\sqrt{219}}\right) \\
&= 1 - \Phi(2.14) \\
&= 1 - 0.9838 \\
&= 0.0162.
\end{aligned}$$

Solution to Problem 5.11. Note that W is the sample mean of 16 independent identically distributed random variables of the form $X_i - Y_i$, and a normal approximation is appropriate. The random variables $X_i - Y_i$ have zero mean, and variance equal to $2/12$. Therefore, the mean of W is zero, and its variance is $(2/12)/16 = 1/96$. Thus,

$$\mathbf{P}\big(|W| < 0.001\big) = \mathbf{P}\left(\frac{|W|}{\sqrt{1/96}} < \frac{0.001}{\sqrt{1/96}}\right) \approx \Phi\big(0.001\sqrt{96}\big) - \Phi\big(-0.001\sqrt{96}\big)$$

$$= 2\Phi\big(0.001\sqrt{96}\big) - 1 = 2\Phi(0.0098) - 1 \approx 2 \cdot 0.504 - 1 = 0.008.$$

Let us also point out a somewhat different approach that bypasses the need for the normal table. Let Z be a normal random variable with zero mean and standard deviation equal to $1/\sqrt{96}$. The standard deviation of Z, which is about 0.1, is much larger than 0.001. Thus, within the interval $[-0.001, 0.001]$, the PDF of Z is approximately constant. Using the formula $\mathbf{P}(z - \delta \leq Z \leq z + \delta) \approx f_Z(z) \cdot 2\delta$, with $z = 0$ and $\delta = 0.001$, we obtain

$$\mathbf{P}\big(|W| < 0.001\big) \approx \mathbf{P}(-0.001 \leq Z \leq 0.001) \approx f_Z(0) \cdot 0.002 = \frac{0.002}{\sqrt{2\pi}(1/\sqrt{96})} = 0.0078.$$

C H A P T E R 6

Solution to Problem 6.1. (a) The random variable R is binomial with parameters p and n. Hence,

$$p_R(r) = \binom{n}{r}(1-p)^{n-r}p^r, \qquad \text{for } r = 0, 1, 2, \ldots, n,$$

$\mathbf{E}[R] = np$, and $\text{var}(R) = np(1-p)$.

(b) Let A be the event that the first item to be loaded ends up being the only one on its truck. This event is the union of two disjoint events:

(i) the first item is placed on the red truck and the remaining $n-1$ are placed on the green truck, and,

(ii) the first item is placed on the green truck and the remaining $n-1$ are placed on the red truck.

Thus, $\mathbf{P}(A) = p(1-p)^{n-1} + (1-p)p^{n-1}$.

(c) Let B be the event that at least one truck ends up with a total of exactly one package. The event B occurs if exactly one or both of the trucks end up with exactly 1 package, so

$$\mathbf{P}(B) = \begin{cases} 1, & \text{if } n = 1, \\ 2p(1-p), & \text{if } n = 2, \\ \binom{n}{1}(1-p)^{n-1}p + \binom{n}{n-1}p^{n-1}(1-p), & \text{if } n = 3, 4, 5, \ldots \end{cases}$$

(d) Let $D = R - G = R - (n - R) = 2R - n$. We have $\mathbf{E}[D] = 2\mathbf{E}[R] - n = 2np - n$. Since $D = 2R - n$, where n is a constant,

$$\text{var}(D) = 4\text{var}(R) = 4np(1-p).$$

(e) Let C be the event that each of the first 2 packages is loaded onto the red truck. Given that C occurred, the random variable R becomes

$$2 + X_3 + X_4 + \cdots + X_n.$$

Hence,

$$\mathbf{E}[R \,|\, C] = \mathbf{E}[2 + X_3 + X_4 + \cdots + X_n] = 2 + (n-2)\mathbf{E}[X_i] = 2 + (n-2)p.$$

Similarly, the conditional variance of R is

$$\text{var}(R \,|\, C) = \text{var}(2 + X_3 + X_4 + \cdots + X_n) = (n-2)\text{var}(X_i) = (n-2)p(1-p).$$

Finally, given that the first two packages are loaded onto the red truck, the probability that a total of r packages are loaded onto the red truck is equal to the probability that $r-2$ of the remaining $n-2$ packages go to the red truck:

$$p_{R|C}(r) = \binom{n-2}{r-2}(1-p)^{n-r}p^{r-2}, \qquad \text{for } r = 2, \ldots, n.$$

Solution to Problem 6.2. (a) Failed quizzes are a Bernoulli process with parameter $p = 1/4$. The desired probability is given by the binomial formula:

$$\binom{6}{2}p^2(1-p)^4 = \frac{6!}{4!\,2!}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^4.$$

(b) The expected number of quizzes up to the third failure is the expected value of a Pascal random variable of order three, with parameter $1/4$, which is $3 \cdot 4 = 12$. Subtracting the number of failures, we have that the expected number of quizzes that Dave will pass is $12 - 3 = 9$.

(c) The event of interest is the intersection of the following three independent events:

A: there is exactly one failure in the first seven quizzes.

B: quiz eight is a failure.

C: quiz nine is a failure.

We have

$$\mathbf{P}(A) = \binom{7}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^6, \qquad \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{4},$$

so the desired probability is

$$\mathbf{P}(A \cap B \cap C) = 7\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^6.$$

(d) Let B be the event that Dave fails two quizzes in a row before he passes two quizzes in a row. Let us use F and S to indicate quizzes that he has failed or passed, respectively. We then have

$$\begin{aligned}
\mathbf{P}(B) &= \mathbf{P}(FF \cup SFF \cup FSFF \cup SFSFF \cup FSFSFF \cup SFSFSFF \cup \cdots) \\
&= \mathbf{P}(FF) + \mathbf{P}(SFF) + \mathbf{P}(FSFF) + \mathbf{P}(SFSFF) + \mathbf{P}(FSFSFF) \\
&\quad + \mathbf{P}(SFSFSFF) + \cdots \\
&= \left(\frac{1}{4}\right)^2 + \frac{3}{4}\left(\frac{1}{4}\right)^2 + \frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 \\
&\quad + \frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \cdots \\
&= \left[\left(\frac{1}{4}\right)^2 + \frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \cdots\right] \\
&\quad + \left[\frac{3}{4}\left(\frac{1}{4}\right)^2 + \frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\left(\frac{1}{4}\right)^2 + \cdots\right].
\end{aligned}$$

Therefore, $\mathbf{P}(B)$ is the sum of two infinite geometric series, and

$$\mathbf{P}(B) = \frac{\left(\frac{1}{4}\right)^2}{1 - \frac{1}{4}\cdot\frac{3}{4}} + \frac{\frac{3}{4}\left(\frac{1}{4}\right)^2}{1 - \frac{3}{4}\cdot\frac{1}{4}} = \frac{7}{52}.$$

Solution to Problem 6.3. The answers to these questions are found by considering suitable Bernoulli processes and using the formulas of Section 6.1. Depending on the specific question, however, a different Bernoulli process may be appropriate. In some cases, we associate trials with slots. In other cases, it is convenient to associate trials with busy slots.

(a) During each slot, the probability of a task from user 1 is given by $p_1 = p_{1|B}p_B = (5/6)\cdot(2/5) = 1/3$. Tasks from user 1 form a Bernoulli process and

$$\mathbf{P}(\text{first user 1 task occurs in slot 4}) = p_1(1-p_1)^3 = \frac{1}{3}\cdot\left(\frac{2}{3}\right)^3.$$

(b) This is the probability that slot 11 was busy and slot 12 was idle, given that 5 out of the 10 first slots were idle. Because of the fresh-start property, the conditioning information is immaterial, and the desired probability is

$$p_B \cdot p_I = \frac{5}{6}\cdot\frac{1}{6}.$$

(c) Each slot contains a task from user 1 with probability $p_1 = 1/3$, independent of other slots. The time of the 5th task from user 1 is a Pascal random variable of order 5, with parameter $p_1 = 1/3$. Its mean is given by

$$\frac{5}{p_1} = \frac{5}{1/3} = 15.$$

(d) Each busy slot contains a task from user 1 with probability $p_{1|B} = 2/5$, independent of other slots. The random variable of interest is a Pascal random variable of order 5, with parameter $p_{1|B} = 2/5$. Its mean is

$$\frac{5}{p_{1|B}} = \frac{5}{2/5} = \frac{25}{2}.$$

(e) The number T of tasks from user 2 until the 5th task from user 1 is the same as the number B of busy slots until the 5th task from user 1, minus 5. The number of busy slots ("trials") until the 5th task from user 1 ("success") is a Pascal random variable of order 5, with parameter $p_{1|B} = 2/5$. Thus,

$$p_B(t) = \binom{t-1}{4}\left(\frac{2}{5}\right)^5\left(1-\frac{2}{5}\right)^{t-5}, \qquad t = 5, 6, \ldots.$$

Since $T = B - 5$, we have $p_T(t) = p_B(t+5)$, and we obtain

$$p_T(t) = \binom{t+4}{4}\left(\frac{2}{5}\right)^5\left(1-\frac{2}{5}\right)^t, \qquad t = 0, 1, \ldots.$$

Using the formulas for the mean and the variance of the Pascal random variable B, we obtain

$$\mathbf{E}[T] = \mathbf{E}[B] - 5 = \frac{25}{2} - 5 = 7.5,$$

and

$$\mathrm{var}(T) = \mathrm{var}(B) = \frac{5\big(1-(2/5)\big)}{(2/5)^2}.$$

Solution to Problem 6.8. The total number of accidents between 8 am and 11 am is the sum of two independent Poisson random variables with parameters 5 and $3 \cdot 2 = 6$, respectively. Since the sum of independent Poisson random variables is also Poisson, the total number of accidents has a Poisson PMF with parameter 5+6=11.

Solution to Problem 6.9. As long as the pair of players is waiting, all five courts are occupied by other players. When all five courts are occupied, the time until a court is freed up is exponentially distributed with mean $40/5 = 8$ minutes. For our pair of players to get a court, a court must be freed up $k+1$ times. Thus, the expected waiting time is $8(k+1)$.

Solution to Problem 6.10. (a) This is the probability of no arrivals in 2 hours. It is given by

$$P(0,2) = e^{-0.6 \cdot 2} = 0.301.$$

For an alternative solution, this is the probability that the first arrival comes after 2 hours:

$$\mathbf{P}(T_1 > 2) = \int_2^\infty f_{T_1}(t)\,dt = \int_2^\infty 0.6e^{-0.6t}\,dt = e^{-0.6 \cdot 2} = 0.301.$$

(b) This is the probability of zero arrivals between time 0 and 2, and of at least one arrival between time 2 and 5. Since these two intervals are disjoint, the desired probability is the product of the probabilities of these two events, which is given by

$$P(0,2)\big(1 - P(0,3)\big) = e^{-0.6 \cdot 2}(1 - e^{-0.6 \cdot 3}) = 0.251.$$

For an alternative solution, the event of interest can be written as $\{2 \le T_1 \le 5\}$, and its probability is

$$\int_2^5 f_{T_1}(t)\,dt = \int_2^5 0.6e^{-0.6t}\,dt = e^{-0.6 \cdot 2} - e^{-0.6 \cdot 5} = 0.251.$$

(c) If he catches at least two fish, he must have fished for exactly two hours. Hence, the desired probability is equal to the probability that the number of fish caught in the first two hours is at least two, i.e.,

$$\sum_{k=2}^\infty P(k,2) = 1 - P(0,2) - P(1,2) = 1 - e^{-0.6 \cdot 2} - (0.6 \cdot 2)e^{-0.6 \cdot 2} = 0.337.$$

For an alternative approach, note that the event of interest occurs if and only if the time Y_2 of the second arrival is less than or equal to 2. Hence, the desired probability is

$$\mathbf{P}(Y_2 \le 2) = \int_0^2 f_{Y_2}(y)\, dy = \int_0^2 (0.6)^2 y e^{-0.6y}\, dy.$$

This integral can be evaluated by integrating by parts, but this is more tedious than the first approach.

(d) The expected number of fish caught is equal to the expected number of fish caught during the first two hours (which is $2\lambda = 2 \cdot 0.6 = 1.2$), plus the expectation of the number N of fish caught after the first two hours. We have $N = 0$ if he stops fishing at two hours, and $N = 1$, if he continues beyond the two hours. The event $\{N = 1\}$ occurs if and only if no fish are caught in the first two hours, so that $\mathbf{E}[N] = \mathbf{P}(N = 1) = P(0, 2) = 0.301$. Thus, the expected number of fish caught is $1.2 + 0.301 = 1.501$.

(e) Given that he has been fishing for 4 hours, the future fishing time is the time until the first fish is caught. By the memorylessness property of the Poisson process, the future time is exponential, with mean $1/\lambda$. Hence, the expected total fishing time is $4 + (1/0.6) = 5.667$.

Solution to Problem 6.11. We note that the process of departures of customers who have bought a book is obtained by splitting the Poisson process of customer departures, and is itself a Poisson process, with rate $p\lambda$.

(a) This is the time until the first customer departure in the split Poisson process. It is therefore exponentially distributed with parameter $p\lambda$.

(b) This is the probability of no customers in the split Poisson process during an hour, and using the result of part (a), equals $e^{-p\lambda}$.

(c) This is the expected number of customers in the split Poisson process during an hour, and is equal to $p\lambda$.

Solution to Problem 6.12. Let X be the number of different types of pizza ordered. Let X_i be the random variable defined by

$$X_i = \begin{cases} 1, & \text{if a type } i \text{ pizza is ordered by at least one customer,} \\ 0, & \text{otherwise.} \end{cases}$$

We have $X = X_1 + \cdots + X_n$, and $\mathbf{E}[X] = n\mathbf{E}[X_1]$.

We can think of the customers arriving as a Poisson process, and with each customer independently choosing whether to order a type 1 pizza (this happens with probability $1/n$) or not. This is the situation encountered in splitting of Poisson processes, and the number of type 1 pizza orders, denoted Y_1, is a Poisson random variable with parameter λ/n. We have

$$\mathbf{E}[X_1] = \mathbf{P}(Y_1 > 0) = 1 - \mathbf{P}(Y_1 = 0) = 1 - e^{-\lambda/n},$$

so that

$$\mathbf{E}[X] = n\mathbf{E}[X_1] = n\big(1 - e^{-\lambda/n}\big).$$

Solution to Problem 6.13. (a) Let R be the total number of messages received during an interval of duration t. Note that R is a Poisson random variable with arrival rate $\lambda_A + \lambda_B$. Therefore, the probability that exactly nine messages are received is

$$\mathbf{P}(R = 9) = \frac{\big((\lambda_A + \lambda_B)T\big)^9 e^{-(\lambda_A+\lambda_B)t}}{9!}.$$

(b) Let R be defined as in part (a), and let W_i be the number of words in the ith message. Then,

$$N = W_1 + W_2 + \cdots + W_R,$$

which is a sum of a random number of random variables. Thus,

$$\begin{aligned}
\mathbf{E}[N] &= \mathbf{E}[W]\mathbf{E}[R] \\
&= \left(1 \cdot \frac{2}{6} + 2 \cdot \frac{3}{6} + 3 \cdot \frac{1}{6}\right)(\lambda_A + \lambda_B)t \\
&= \frac{11}{6}(\lambda_A + \lambda_B)t.
\end{aligned}$$

(c) Three-word messages arrive from transmitter A in a Poisson manner, with average rate $\lambda_A p_W(3) = \lambda_A/6$. Therefore, the random variable of interest is Erlang of order 8, and its PDF is given by

$$f(x) = \frac{(\lambda_A/6)^8 x^7 e^{-\lambda_A x/6}}{7!}.$$

(d) Every message originates from either transmitter A or B, and can be viewed as an independent Bernoulli trial. Each message has probability $\lambda_A/(\lambda_A+\lambda_B)$ of originating from transmitter A (view this as a "success"). Thus, the number of messages from transmitter A (out of the next twelve) is a binomial random variable, and the desired probability is equal to

$$\binom{12}{8}\left(\frac{\lambda_A}{\lambda_A+\lambda_B}\right)^8\left(\frac{\lambda_B}{\lambda_A+\lambda_B}\right)^4.$$

Solution to Problem 6.14. (a) Let X be the time until the first bulb failure. Let A (respectively, B) be the event that the first bulb is of type A (respectively, B). Since the two bulb types are equally likely, the total expectation theorem yields

$$\mathbf{E}[X] = \mathbf{E}[X \,|\, A]\mathbf{P}(A) + \mathbf{E}[X \,|\, B]\mathbf{P}(B) = 1 \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{2}{3}.$$

(b) Let D be the event of no bulb failures before time t. Using the total probability theorem, and the exponential distributions for bulbs of the two types, we obtain

$$\mathbf{P}(D) = \mathbf{P}(D \,|\, A)\mathbf{P}(A) + \mathbf{P}(D \,|\, B)\mathbf{P}(B) = \frac{1}{2}e^{-t} + \frac{1}{2}e^{-3t}.$$

(c) We have

$$\mathbf{P}(A \,|\, D) = \frac{\mathbf{P}(A \cap D)}{\mathbf{P}(D)} = \frac{\frac{1}{2}e^{-t}}{\frac{1}{2}e^{-t} + \frac{1}{2}e^{-3t}} = \frac{1}{1 + e^{-2t}}.$$

(d) We first find $\mathbf{E}[X^2]$. We use the fact that the second moment of an exponential random variable T with parameter λ is equal to $\mathbf{E}[T^2] = \mathbf{E}[T]^2 + \mathrm{var}(T) = 1/\lambda^2 + 1/\lambda^2 = 2/\lambda^2$. Conditioning on the two possible types of the first bulb, we obtain

$$\mathbf{E}[X^2] = \mathbf{E}[X^2 \,|\, A]\mathbf{P}(A) + \mathbf{E}[X^2 \,|\, B]\mathbf{P}(B) = 2 \cdot \frac{1}{2} + \frac{2}{9} \cdot \frac{1}{2} = \frac{10}{9}.$$

Finally, using the fact $\mathbf{E}[X] = 2/3$ from part (a),

$$\mathrm{var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{10}{9} - \frac{2^2}{3^2} = \frac{2}{3}.$$

(e) This is the probability that out of the first 11 bulbs, exactly 3 were of type A and that the 12th bulb was of type A. It is equal to

$$\binom{11}{3}\left(\frac{1}{2}\right)^{12}.$$

(f) This is the probability that out of the first 12 bulbs, exactly 4 were of type A, and is equal to

$$\binom{12}{4}\left(\frac{1}{2}\right)^{12}.$$

(g) The PDF of the time between failures is $(e^{-x} + 3e^{-3x})/2$, for $x \geq 0$, and the associated transform is

$$\frac{1}{2}\left(\frac{1}{1-s} + \frac{3}{3-s}\right).$$

Since the times between successive failures are independent, the transform associated with the time until the 12th failure is given by

$$\left[\frac{1}{2}\left(\frac{1}{1-s} + \frac{3}{3-s}\right)\right]^{12}.$$

(h) Let Y be the total period of illumination provided by the first two type-B bulbs. This has an Erlang distribution of order 2, and its PDF is

$$f_Y(y) = 9ye^{-3y}, \qquad y \geq 0.$$

Let T be the period of illumination provided by the first type-A bulb. Its PDF is

$$f_T(t) = e^{-t}, \qquad t \geq 0.$$

We are interested in the event $T < Y$. We have

$$\mathbf{P}(T < Y \,|\, Y = y) = 1 - e^{-y}, \qquad y \geq 0.$$

Thus,

$$\mathbf{P}(T < Y) = \int_0^\infty f_Y(y)\mathbf{P}(T < Y \,|\, Y = y)\,dy = \int_0^\infty 9ye^{-3y}\big(1 - e^{-y}\big)\,dy = \frac{7}{16},$$

as can be verified by carrying out the integration.

We now describe an alternative method for obtaining the answer. Let T_1^A be the period of illumination of the first type-A bulb. Let T_1^B and T_2^B be the period of illumination provided by the first and second type-B bulb, respectively. We are interested in the event $\{T_1^A < T_1^B + T_2^B\}$. We have

$$\begin{aligned}
\mathbf{P}(T_1^A < T_1^B + T_2^B) &= \mathbf{P}(T_1^A < T_1^B) + \mathbf{P}(T_1^A \geq T_1^B)\,\mathbf{P}(T_1^A < T_1^B + T_2^B \,|\, T_1^A \geq T_1^B) \\
&= \frac{1}{1+3} + \mathbf{P}(T_1^A \geq T_1^B)\,\mathbf{P}(T_1^A - T_1^B < T_2^B \,|\, T_1^A \geq T_1^B) \\
&= \frac{1}{4} + \frac{3}{4}\mathbf{P}(T_1^A - T_1^B < T_2^B \,|\, T_1^A \geq T_1^B).
\end{aligned}$$

Given the event $T_1^A \geq T_1^B$, and using the memorylessness property of the exponential random variable T_1^A, the remaining time $T_1^A - T_1^B$ until the failure of the type-A bulb is exponentially distributed, so that

$$\mathbf{P}(T_1^A - T_1^B < T_2^B \,|\, T_1^A \geq T_1^B) = \mathbf{P}(T_1^A < T_2^B) = \mathbf{P}(T_1^A < T_1^B) = \frac{1}{4}.$$

Therefore,

$$\mathbf{P}(T_1^A < T_1^B + T_2^B) = \frac{1}{4} + \frac{3}{4}\cdot\frac{1}{4} = \frac{7}{16}.$$

(i) Let V be the total period of illumination provided by type-B bulbs while the process is in operation. Let N be the number of light bulbs, out of the first 12, that are of type B. Let X_i be the period of illumination from the ith type-B bulb. We then have $V = Y_1 + \cdots + Y_N$. Note that N is a binomial random variable, with parameters $n = 12$ and $p = 1/2$, so that

$$\mathbf{E}[N] = 6, \qquad \mathrm{var}(N) = 12\cdot\frac{1}{2}\cdot\frac{1}{2} = 3.$$

Furthermore, $\mathbf{E}[X_i] = 1/3$ and $\mathrm{var}(X_i) = 1/9$. Using the formulas for the mean and variance of the sum of a random number of random variables, we obtain

$$\mathbf{E}[V] = \mathbf{E}[N]\mathbf{E}[X_i] = 2,$$

and

$$\mathrm{var}(V) = \mathrm{var}(X_i)\mathbf{E}[N] + \mathbf{E}[X_i]^2\mathrm{var}(N) = \frac{1}{9}\cdot 6 + \frac{1}{9}\cdot 3 = 1.$$

(j) Using the notation in parts (a)-(c), and the result of part (c), we have

$$\begin{aligned}\mathbf{E}[T \mid D] &= t + \mathbf{E}[T - t \mid D \cap A]\mathbf{P}(A \mid D) + \mathbf{E}[T - t \mid D \cap B]\mathbf{P}(B \mid D) \\ &= t + 1 \cdot \frac{1}{1 + e^{-2t}} + \frac{1}{3}\left(1 - \frac{1}{1 + e^{-2t}}\right) \\ &= t + \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{1 + e^{-2t}}.\end{aligned}$$

Solution to Problem 6.15. (a) The total arrival process corresponds to the merging of two independent Poisson processes, and is therefore Poisson with rate $\lambda = \lambda_A + \lambda_B = 7$. Thus, the number N of jobs that arrive in a given three-minute interval is a Poisson random variable, with $\mathbf{E}[N] = 3\lambda = 21$, $\mathrm{var}(N) = 21$, and PMF

$$p_N(n) = \frac{(21)^n e^{-21}}{n!}, \qquad n = 0, 1, 2, \ldots.$$

(b) Each of these 10 jobs has probability $\lambda_A/(\lambda_A + \lambda_B) = 3/7$ of being of type A, independently of the others. Thus, the binomial PMF applies and the desired probability is equal to

$$\binom{10}{3}\left(\frac{3}{7}\right)^3\left(\frac{4}{7}\right)^7.$$

(c) Each future arrival is of type A with probability $\lambda_A/(\lambda_A + \lambda_B) = 3/7$, independently of other arrivals. Thus, the number K of arrivals until the first type A arrival is geometric with parameter 3/7. The number of type B arrivals before the first type A arrival is equal to $K - 1$, and its PMF is similar to a geometric, except that it is shifted by one unit to the left. In particular,

$$p_K(k) = \left(\frac{3}{7}\right)\left(\frac{4}{7}\right)^k, \qquad k = 0, 1, 2, \ldots.$$

(d) The fact that at time 0 there were two type A jobs in the system simply states that there were exactly two type A arrivals between time -1 and time 0. Let X and Y be the arrival times of these two jobs. Consider splitting the interval $[-1, 0]$ into many time slots of length δ. Since each time instant is equally likely to contain an arrival and since the arrival times are independent, it follows that X and Y are independent uniform random variables. We are interested in the PDF of $Z = \max\{X, Y\}$. We first find the CDF of Z. We have, for $z \in [-1, 0]$,

$$\mathbf{P}(Z \le z) = \mathbf{P}(X \le z \text{ and } Y \le z) = (1 + z)^2.$$

By differentiating, we obtain

$$f_Z(z) = 2(1 + z), \qquad -1 \le z \le 0.$$

(e) Let T be the arrival time of this type B job. We can express T in the form $T = -K + X$, where K is a nonnegative integer and X lies in [0,1]. We claim that X

is independent from K and that X is uniformly distributed. Indeed, conditioned on the event $K = k$, we know that there was a single arrival in the interval $[-k, -k+1]$. Conditioned on the latter information, the arrival time is uniformly distributed in the interval $[-k, k+1]$ (cf. Problem 6.18), which implies that X isuniformly distributed in $[0, 1]$. Since this conditional distribution of X is the same for every k, it follows that X is independent of $-K$.

Let D be the departure time of the job of interest. Since the job stays in the system for an integer amount of time, we have that D is of the form $D = L + X$, where L is a nonnegative integer. Since the job stays in the system for a geometrically distributed amount of time, and the geometric distribution has the memorylessness property, it follows that L is also memoryless. In particular, L is similar to a geometric random variable, except that its PMF starts at zero. Furthermore, L is independent of X, since X is determined by the arrival process, whereas the amount of time a job stays in the system is independent of the arrival process. Thus, D is the sum of two independent random variables, one uniform and one geometric. Therefore, D has "geometric staircase" PDF, given by

$$f_D(d) = \left(\frac{1}{2}\right)^{\lfloor d \rfloor}, \qquad d \geq 0,$$

and where $\lfloor d \rfloor$ stands for the largest integer below d.

Solution to Problem 6.16. (a) The random variable N is equal to the number of successive interarrival intervals that are smaller than τ. Interarrival intervals are independent and each one is smaller than τ with probability $1 - e^{-\lambda\tau}$. Therefore,

$$\mathbf{P}(N = 0) = e^{-\lambda\tau}, \qquad \mathbf{P}(N = 1) = e^{-\lambda\tau}\left(1-e^{-\lambda\tau}\right), \qquad \mathbf{P}(N = k) = e^{-\lambda\tau}\left(1-e^{-\lambda\tau}\right)^k,$$

so that N has a distribution similar to a geometric one, with parameter $p = e^{-\lambda\tau}$, except that it shifted one place to the left, so that it starts out at 0. Hence,

$$\mathbf{E}[N] = \frac{1}{p} - 1 = e^{\lambda\tau} - 1.$$

(b) Let T_n be the nth interarrival time. The event $\{N \geq n\}$ indicates that the time between cars $n-1$ and n is less than or equal to τ, and therefore $\mathbf{E}[T_n \,|\, N \geq n] = \mathbf{E}[T_n \,|\, T_n \leq \tau]$. Note that the conditional PDF of T_n is the same as the unconditional one, except that it is now restricted to the interval $[0, \tau]$, and that it has to be suitably renormalized so that it integrates to 1. Therefore, the desired conditional expectation is

$$\mathbf{E}[T_n \,|\, T_n \leq \tau] = \frac{\displaystyle\int_0^\tau s\lambda e^{-\lambda s}\, ds}{\displaystyle\int_0^\tau \lambda e^{-\lambda s}\, ds}.$$

This integral can be evaluated by parts. We will provide, however, an alternative approach that avoids integration.

We use the total expectation formula

$$\mathbf{E}[T_n] = \mathbf{E}[T_n \,|\, T_n \leq \tau]\mathbf{P}(T_n \leq \tau) + \mathbf{E}[T_n \,|\, T_n > \tau]\mathbf{P}(T_n > \tau).$$

We have $\mathbf{E}[T_n] = 1/\lambda$, $\mathbf{P}(T_n \le \tau) = 1 - e^{-\lambda\tau}$, $\mathbf{P}(T_n > \tau) = e^{-\lambda\tau}$, and $\mathbf{E}[T_n \,|\, T_n > \tau] = \tau + (1/\lambda)$. (The last equality follows from the memorylessness of the exponential PDF.) Using these equalities, we obtain

$$\frac{1}{\lambda} = \mathbf{E}[T_n \,|\, T_n \le \tau]\big(1 - e^{-\lambda\tau}\big) + \left(\tau + \frac{1}{\lambda}\right) e^{-\lambda\tau},$$

which yields

$$\mathbf{E}[T_n \,|\, T_n \le \tau] = \frac{\dfrac{1}{\lambda} - \left(\tau + \dfrac{1}{\lambda}\right) e^{-\lambda\tau}}{1 - e^{-\lambda\tau}}.$$

(c) Let T be the time until the U-turn. Note that $T = T_1 + \cdots + T_N + \tau$. Let v denote the value of $\mathbf{E}[T_n \,|\, T_n \le \tau]$. We find $\mathbf{E}[T]$ using the total expectation theorem:

$$\begin{aligned}
\mathbf{E}[T] &= \tau + \sum_{n=0}^{\infty} \mathbf{P}(N = n)\mathbf{E}[T_1 + \cdots + T_N \,|\, N = n] \\
&= \tau + \sum_{n=0}^{\infty} \mathbf{P}(N = n) \sum_{i=1}^{n} \mathbf{E}[T_i \,|\, T_1 \le \tau, \ldots, T_n \le \tau, T_{n+1} > \tau] \\
&= \tau + \sum_{n=0}^{\infty} \mathbf{P}(N = n) \sum_{i=1}^{n} \mathbf{E}[T_i \,|\, T_i \le \tau] \\
&= \tau + \sum_{n=0}^{\infty} \mathbf{P}(N = n) n v \\
&= \tau + v\mathbf{E}[N],
\end{aligned}$$

where $\mathbf{E}[N]$ was found in part (a) and v was found in part (b). The second equality used the fact that the event $\{N = n\}$ is the same as the event $\{T_1 \le \tau, \ldots, T_n \le \tau, T_{n+1} > \tau\}$. The third equality used the independence of the interarrival times T_i.

Solution to Problem 6.17. We will calculate the expected length of the photographer's waiting time T conditioned on each of the two events: A, which is that the photographer arrives while the wombat is resting or eating, and A^c, which is that the photographer arrives while the wombat is walking. We will then use the total expectation theorem as follows:

$$\mathbf{E}[T] = \mathbf{P}(A)\mathbf{E}[T \,|\, A] + \mathbf{P}(A^c)\mathbf{E}[T \,|\, A^c].$$

The conditional expectation $\mathbf{E}[T \,|\, A]$ can be broken down in three components:

(i) The expected remaining time up to when the wombat starts its next walk; by the memorylessness property, this time is exponentially distributed and its expected value is 30 secs.

(ii) A number of walking and resting/eating intervals (each of expected length 50 secs) during which the wombat does not stop; if N is the number of these intervals, then $N + 1$ is geometrically distributed with parameter $1/3$. Thus the expected length of these intervals is $(3 - 1) \cdot 50 = 100$ secs.

(iii) The expected waiting time during the walking interval in which the wombat stands still. This time is uniformly distributed between 0 and 20, so its expected value is 10 secs.

Collecting the above terms, we see that

$$\mathbf{E}[T \mid A] = 30 + 100 + 10 = 140.$$

The conditional expectation $\mathbf{E}[T \mid A^c]$ can be calculated using the total expectation theorem, by conditioning on three events: B_1, which is that the wombat does not stop during the photographer's arrival interval (probability 2/3); B_2, which is that the wombat stops during the photographer's arrival interval after the photographer arrives (probability 1/6); B_3, which is that the wombat stops during the photographer's arrival interval before the photographer arrives (probability 1/6). We have

$$\begin{aligned}\mathbf{E}[T \mid A^c, B_1] &= \mathbf{E}[\text{photographer's wait up to the end of the interval}] + \mathbf{E}[T \mid A] \\ &= 10 + 140 = 150.\end{aligned}$$

Also, it can be shown that if two points are randomly chosen in an interval of length l, the expected distance between the two points is $l/3$ (an end-of-chapter problem in Chapter 3), and using this fact, we have

$$\mathbf{E}[T \mid A^c, B_2] = \mathbf{E}[\text{photographer's wait up to the time when the wombat stops}] = \frac{20}{3}.$$

Similarly, it can be shown that if two points are randomly chosen in an interval of length l, the expected distance between each point and the nearest endpoint of the interval is $l/3$. Using this fact, we have

$$\begin{aligned}\mathbf{E}[T \mid A^c, B_3] &= \mathbf{E}[\text{photographer's wait up to the end of the interval}] + \mathbf{E}[T \mid A] \\ &= \frac{20}{3} + 140.\end{aligned}$$

Applying the total expectation theorem, we see that

$$\mathbf{E}[T \mid A^c] = \frac{2}{3} \cdot 150 + \frac{1}{6} \cdot \frac{20}{3} + \frac{1}{6}\left(\frac{20}{3} + 140\right) = 125.55.$$

To apply the total expectation theorem and obtain $\mathbf{E}[T]$, we need the probability $\mathbf{P}(A)$ that the photographer arrives during a resting/eating interval. Since the expected length of such an interval is 30 seconds and the length of the complementary walking interval is 20 seconds, we see that $\mathbf{P}(A) = 30/50 = 0.6$. Substituting in the equation

$$\mathbf{E}[T] = \mathbf{P}(A)\mathbf{E}[T \mid A] + \big(1 - \mathbf{P}(A)\big)\mathbf{E}[T \mid A^c],$$

we obtain

$$\mathbf{E}[T] = 0.6 \cdot 140 + 0.4 \cdot 125.55 = 134.22.$$

CHAPTER 7

Solution to Problem 7.1. We construct a Markov chain with state space $S = \{0, 1, 2, 3\}$. We let $X_n = 0$ if an arrival occurs at time n. Also, we let $X_n = i$ if the last arrival up to time n occurred at time $n - i$, for $i = 1, 2, 3$. Given that $X_n = 0$, there is probability 0.2 that the next arrival occurs at time $n + 1$, so that $p_{00} = 0.2$, and $p_{01} = 0.8$. Given that $X_n = 1$, the last arrival occurred at time $n - 1$, and there is zero probability of an arrival at time $n + 1$, so that $p_{12} = 1$. Given that $X_n = 2$, the last arrival occurred at time $n - 2$. We then have

$$\begin{aligned} p_{20} &= \mathbf{P}(X_{n+1} = 0 \,|\, X_n = 2) \\ &= \mathbf{P}(T = 3 \,|\, T \geq 3) \\ &= \frac{\mathbf{P}(T = 3)}{\mathbf{P}(T \geq 3)} \\ &= \frac{3}{8}, \end{aligned}$$

and $p_{23} = 5/8$. Finally, given that $X_n = 3$, an arrival is guaranteed at time $n + 1$, so that $p_{30} = 1$.

Solution to Problem 7.2. It cannot be described as a Markov chain with states L and R, because $\mathbf{P}(X_{n+1} = L \,|\, X_n = R, X_{n-1} = L) = 1/2$, while $\mathbf{P}(X_{n+1} = L \,|\, X_n = R, X_{n-1} = R, X_{n-1} = L) = 0$.

Solution to Problem 7.3. The answer is no. To establish this, we need to show that the Markov property fails to hold, that is we need to find scenarios that lead to the same state and such that the probability law for the next state is different for different scenarios.

Let X_n be the 4-state Markov chain corresponding to the original example. Let us compare the two scenarios $(Y_0, Y_1) = (1, 2)$ and $(Y_0, Y_1) = (2, 2)$. For the first scenario, the information $(Y_0, Y_1) = (1, 2)$ implies that $X_0 = 2$ and $X_1 = 3$, so that

$$\mathbf{P}(Y_2 = 2 \,|\, Y_0 = 1, Y_1 = 2) = \mathbf{P}\big(X_2 \in \{3, 4\} \,|\, X_1 = 3\big) = 0.7.$$

For the second scenario, the information $(Y_0, Y_1) = (2, 2)$ is not enough to determine X_1, but we can nevertheless assert that $\mathbf{P}(X_1 = 4 \,|\, Y_0 = Y_1 = 2) > 0$. (This is because the conditioning information $Y_0 = 2$ implies that $X_0 \in \{3, 4\}$, and for either choice of X_0, there is positive probability that $X_1 = 4$.)

We then have

$$\begin{aligned} &\mathbf{P}(Y_2 = 2 \,|\, Y_0 = Y_1 = 2) \\ &\qquad = \mathbf{P}(Y_2 = 2 \,|\, X_1 = 4,\ Y_0 = Y_1 = 2)\mathbf{P}(X_1 = 4 \,|\, Y_0 = Y_1 = 2) \\ &\qquad\quad + \mathbf{P}(Y_2 = 2 \,|\, X_1 = 3,\ Y_0 = Y_1 = 2)\big(1 - \mathbf{P}(X_1 = 4 \,|\, Y_0 = Y_1 = 2)\big) \\ &\qquad = 1 \cdot \mathbf{P}(X_1 = 4 \,|\, Y_0 = Y_1 = 2) + 0.7\big(1 - \mathbf{P}(X_1 = 4 \,|\, Y_0 = Y_1 = 2)\big) \\ &\qquad = 0.7 + 0.3 \cdot \mathbf{P}(X_1 = 4 \,|\, Y_0 = Y_1 = 2) \\ &\qquad > 0.7. \end{aligned}$$

Thus, $\mathbf{P}(Y_2 = 2 \,|\, Y_0 = 1, Y_1 = 2) \neq \mathbf{P}(Y_2 = 2 \,|\, Y_0 = Y_1 = 2)$, which implies that Y_n does not have the Markov property.

Solution to Problem 7.4. (a) We introduce a Markov chain with state equal to the distance between spider and fly. Let n be the initial distance. Then, the states are $0, 1, \ldots, n$, and we have

$$p_{00} = 1, \qquad p_{0i} = 0, \quad \text{for } i \neq 0,$$

$$p_{10} = 0.4, \qquad p_{11} = 0.6, \qquad p_{1i} = 0, \quad \text{for } i \neq 0, 1,$$

and for all $i \neq 0, 1$,

$$p_{i(i-2)} = 0.3, \qquad p_{i(i-1)} = 0.4, \qquad p_{ii} = 0.3, \qquad p_{ij} = 0, \quad \text{for } j \neq i-2, i-1, i.$$

(b) All states are transient except for state 0 which forms a recurrent class.

Solution to Problem 7.5. It is periodic with period 2. The two corresponding subsets are $\{2, 4, 6, 7, 9\}$ and $\{1, 3, 5, 8\}$.

Solution to Problem 7.10. For the first model, the transition probability matrix is

$$\begin{bmatrix} 1-b & b \\ r & 1-r \end{bmatrix}.$$

We need to exclude the cases $b = r = 0$ in which case we obtain a periodic class, and the case $b = r = 1$ in which case there are two recurrent classes. The balance equations are of the form

$$\pi_1 = (1-b)\pi_1 + r\pi_2, \qquad \pi_2 = b\pi_1 + (1-r)\pi_2,$$

or

$$b\pi_1 = r\pi_2.$$

This equation, together with the normalization equation $\pi_1 + \pi_2 = 1$, yields the steady-state probabilities

$$\pi_1 = \frac{r}{b+r}, \qquad \pi_2 = \frac{b}{b+r}.$$

For the second model, we need to exclude the case $b = r = 1$ that makes the chain periodic with period 2, and the case $b = 1$, $r = 0$, which makes the chain periodic with period $\ell + 1$. The balance equations are of the form

$$\begin{aligned}
\pi_1 &= (1-b)\pi_1 + r\big(\pi_{(2,1)} + \cdots + \pi_{(2,\ell-1)}\big) + \pi_{(2,\ell)}, \\
\pi_{(2,1)} &= b\pi_1, \\
\pi_{(2,i)} &= (1-r)\pi_{(2,i-1)}, \qquad i = 2, \ldots, \ell.
\end{aligned}$$

The last two equations can be used to express $\pi_{(2,i)}$ in terms of π_1,

$$\pi_{(2,i)} = (1-r)^{i-1} b\pi_1, \qquad i = 1, \ldots, \ell.$$

Substituting into the normalization equation $\pi_1 + \sum_{i=1}^{\ell} \pi_{(2,i)} = 1$, we obtain

$$1 = \left(1 + b\sum_{i=1}^{\ell}(1-r)^{i-1}\right)\pi_1 = \left(1 + \frac{b\big(1-(1-r)^{\ell}\big)}{r}\right)\pi_1,$$

or

$$\pi_1 = \frac{r}{r + b\big(1-(1-r)^{\ell}\big)}.$$

Using the equation $\pi_{(2,i)} = (1-r)^{i-1}b\pi_1$, we can also obtain explicit formulas for the $\pi_{(2,i)}$.

Solution to Problem 7.11. We use a Markov chain model with 3 states, H, M, and E, where the state reflects the difficulty of the most recent exam. We are given the transition probabilities

$$\begin{bmatrix} r_{HH} & r_{HM} & r_{HE} \\ r_{MH} & r_{MM} & r_{ME} \\ r_{EH} & r_{EM} & r_{EE} \end{bmatrix} = \begin{bmatrix} 0 & .5 & .5 \\ .25 & .5 & .25 \\ .25 & .25 & .5 \end{bmatrix}.$$

It is easy to see that our Markov chain has a single recurrent class, which is aperiodic. The balance equations take the form

$$\pi_1 = \frac{1}{4}(\pi_2 + \pi_3),$$
$$\pi_2 = \frac{1}{2}(\pi_1 + \pi_2) + \frac{1}{4}\pi_3,$$
$$\pi_3 = \frac{1}{2}(\pi_1 + \pi_3) + \frac{1}{4}\pi_2,$$

and solving these with the constraint $\sum_i \pi_i = 1$ gives

$$\pi_1 = \frac{1}{5}, \qquad \pi_2 = \pi_3 = \frac{2}{5}.$$

Solution to Problem 7.12. (a) This is a generalization of Example 7.6. We may proceed as in that example and introduce a Markov chain with states $0, 1, \ldots, n$, where state i indicates that there i available rods at Alvin's present location. However, that Markov chain has a somewhat complex structure, and for this reason, we will proceed differently.

We consider a Markov chain with states $0, 1, \ldots, n$, where state i indicates that Alvin is off the island and has i rods available. Thus, a transition in this Markov chain reflects two trips (going to the island and returning). It is seen that this is a birth-death process. This is because if there are i rods off the island, then at the end of the round trip, the number of rods can only be $i-1$, i or $i+1$.

We now determine the transition probabilities. When $i > 0$, the transition probability $p_{i,i-1}$ is the probability that the weather is good on the way to the island, but is bad on the way back, so that $p_{i,i-1} = p(1-p)$. When $0 < i < n$, the transition probability $p_{i,i+1}$ is the probability that the weather is bad on the way to the island, but is good on the way back, so that $p_{i,i+1} = p(1-p)$. For $i = 0$, the transition probability $p_{i,i+1} = p_{0,1}$ is just the probability that the weather is good on the way back, so that $p_{0,1} = p$. The transition probabilities p_{ii} are then easily determined because the sum of the transition probabilities out of state i must be equal to 1. To summarize, we have

$$p_{ii} = \begin{cases} (1-p)^2 + p^2, & \text{for } i > 0, \\ 1-p, & \text{for } i = 0, \\ 1-p+p^2, & \text{for } i = n, \end{cases}$$

$$p_{i,i+1}, = \begin{cases} (1-p)p, & \text{for } 0 < i < n, \\ p, & \text{for } i = 0, \end{cases}$$

$$p_{i,i-1} = \begin{cases} (1-p)p, & \text{for } i > 0, \\ 0, & \text{for } i = 0. \end{cases}$$

Since this is a birth-death process, we can use the local balance equations. We have

$$\pi_0 p_{01} = \pi_1 p_{10},$$

implying that

$$\pi_1 = \frac{\pi_0}{1-p},$$

and similarly,

$$\pi_n = \cdots = \pi_2 = \pi_1 = \frac{\pi_0}{1-p}.$$

Therefore,

$$1 = \sum_{i=0}^{n} \pi_i = \pi_0 \left(1 + \frac{n}{1-p}\right),$$

which yields

$$\pi_0 = \frac{1-p}{n+1-p}, \qquad \pi_i = \frac{1}{n+1-p}, \quad \text{for all } i > 0.$$

(b) Assume that Alvin is off the island. Let A denote the event that the weather is nice but Alvin has no fishing rods with him. Then,

$$\mathbf{P}(A) = \pi_0 p = \frac{p-p^2}{n+1-p}.$$

Suppose now that Alvin is on the island. The probability that he has no fishing rods with him is again π_0, by the symmetry of the problem. Therefore, $\mathbf{P}(A)$ is the same. Thus, irrespective of his location, the probability that the weather is nice but Alvin cannot fish is $(p-p^2)/(n+1-p)$.

Solution to Problem 7.13. (a) The local balance equations take the form

$$0.6\pi_1 = 0.3\pi_2, \qquad 0.2\pi_2 = 0.2\pi_3.$$

They can be solved, together with the normalization equation, to yield

$$\pi_1 = \frac{1}{5}, \qquad \pi_2 = \pi_3 = \frac{2}{5}.$$

(b) The probability that the first transition is a birth is

$$0.6\pi_1 + 0.2\pi_2 = \frac{0.6}{5} + \frac{0.2 \cdot 2}{5} = \frac{1}{5}.$$

(c) If the state is 1, which happens with probability 1/5, the first change of state is certain to be a birth. If the state is 2, which happens with probability 2/5, the probability that the first change of state is a birth is equal to $0.2/(0.3 + 0.2) = 2/5$. Finally, if the state is 3, the probability that the first change of state is a birth is equal to 0. Thus, the probability that the first change of state that we observe is a birth is equal to

$$1 \cdot \frac{1}{5} + \frac{2}{5} \cdot \frac{2}{5} = \frac{9}{25}.$$

(d) We have

$$\begin{aligned}\mathbf{P}(\text{state was 2} \,|\, \text{first transition is a birth}) &= \frac{\mathbf{P}(\text{state was 2 and first transition is a birth})}{\mathbf{P}(\text{first transition is a birth})} \\ &= \frac{\pi_2 \cdot 0.2}{1/5} = \frac{2}{5}.\end{aligned}$$

(e) As shown in part (c), the probability that the first change of state is a birth is 9/25. Furthermore, the probability that the state is 2 and the first change of state is a birth is $2\pi_2/5 = 4/25$. Therefore, the desired probability is

$$\frac{4/25}{9/25} = \frac{4}{9}.$$

(f) In a birth-death process, there must be as many births as there are deaths, plus or minus 1. Thus, the steady-state probability of births must be equal to the steady-state probability of deaths. Hence, in steady-state, half of the state changes are expected to be births. Therefore, the conditional probability that the first observed transition is a birth, given that it resulted in a change of state, is equal to 1/2. This answer can also be obtained algebraically:

$$\mathbf{P}(\text{birth} \,|\, \text{change of state}) = \frac{\mathbf{P}(\text{birth})}{\mathbf{P}(\text{change of state})} = \frac{1/5}{\frac{1}{5} \cdot 0.6 + \frac{2}{5} \cdot 0.5 + \frac{2}{5} \cdot 0.2} = \frac{1/5}{2/5} = \frac{1}{2}.$$

(g) We have

$$\mathbf{P}(\text{leads to state 2} \,|\, \text{change}) = \frac{\mathbf{P}(\text{change that leads to state 2})}{\mathbf{P}(\text{change})} = \frac{\pi_1 \cdot 0.6 + \pi_3 \cdot 0.2}{2/5} = \frac{1}{2}.$$

This is intuitive because for every change of state that leads into state 2, there must be a subsequent change of state that leads away from state 2.

Solution to Problem 7.14. (a) Let p_{ij} be the transition probabilities and let π_i be the steady-state probabilities. We then have

$$\mathbf{P}(X_{1000} = j, X_{1001} = k, X_{2000} = l \,|\, X_0 = i) = r_{ij}(1000) p_{jk} r_{kl}(999) \approx \pi_j p_{jk} \pi_l.$$

(b) Using Bayes' rule, we have

$$\mathbf{P}(X_{1000} = i \mid X_{1001} = j) = \frac{\mathbf{P}(X_{1000} = i, X_{1001} = j)}{\mathbf{P}(X_{1001} = j)} = \frac{\pi_i p_{ij}}{\pi_j}.$$

Solution to Problem 7.15. Let $i = 0, 1 \ldots, n$ be the states, with state i indicating that there are exactly i white balls. The nonzero transition probabilities are

$$p_{00} = \epsilon, \qquad p_{01} = 1 - \epsilon, \qquad p_{nn} = \epsilon, \qquad p_{n,n-1} = 1 - \epsilon,$$

$$p_{i,i-1} = (1 - \epsilon)\frac{i}{n}, \qquad p_{i,i+1} = (1 - \epsilon)\frac{n - i}{n}, \qquad i = 1, \ldots, n - 1.$$

The chain has a single recurrent class, which is aperiodic. In addition, it is a birth-death process. The local balance equations take the form

$$\pi_i (1 - \epsilon)\frac{n - i}{n} = \pi_{i+1}(1 - \epsilon)\frac{i + 1}{n}, \qquad i = 0, 1, \ldots n - 1,$$

which leads to

$$\pi_i = \frac{n(n-1)\ldots(n-i+1)}{1 \cdot 2 \cdots i}\pi_0 = \frac{n!}{i!\,(n-i)!}\pi_0 = \binom{n}{i}\pi_0.$$

We recognize that this has the form of a binomial distribution, so that for the probabilities to add to 1, we must have $\pi_0 = 1/2^n$. Therefore, the steady-state probabilities are given by

$$\pi_j = \binom{n}{j}\left(\frac{1}{2}\right)^n, \qquad j = 0, \ldots, n.$$

Solution to Problem 7.16. Let $j = 0, 1 \ldots, m$ be the states, with state j corresponding to the first urn containing j white balls. The nonzero transition probabilities are

$$p_{j,j-1} = \left(\frac{j}{m}\right)^2, \qquad p_{j,j+1} = \left(\frac{m - j}{m}\right)^2, \qquad p_{jj} = \frac{2j(m - j)}{m^2}.$$

The chain has a single recurrent class that is aperiodic. This chain is a birth-death process and the steady-state probabilities can be found by solving the local balance equations:

$$\pi_j \left(\frac{m - j}{m}\right)^2 = \pi_{j+1}\left(\frac{j + 1}{m}\right)^2, \qquad j = 0, 1, \ldots, m - 1.$$

The solution is of the form

$$\pi_j = \pi_0 \left(\frac{m(m-1)\cdots(m-j+1)}{1 \cdot 2 \cdots j}\right)^2 = \pi_0 \left(\frac{m!}{j!\,(m-j)!}\right)^2 = \pi_0 \binom{m}{j}^2.$$

We recognize this as having the form of the hypergeometric distribution (Problem 61 of Chapter 1, with $n = 2m$ and $k = m$), which implies that $\pi_0 = \binom{2m}{m}$, and

$$\pi_j = \frac{\binom{m}{j}^2}{\binom{2m}{m}}, \qquad j = 0, 1, \ldots, m.$$

Solution to Problem 7.17. (a) The states form a recurrent class, which is aperiodic since all possible transitions have positive probability.

(b) The Chapman-Kolmogorov equations are

$$r_{ij}(n) = \sum_{k=1}^{2} r_{ik}(n-1)p_{kj}, \qquad \text{for } n > 1, \quad \text{and } i,\, j = 1, 2,$$

starting with $r_{ij}(1) = p_{ij}$, so they have the form

$$r_{11}(n) = r_{11}(n-1)(1-\alpha) + r_{12}(n-1)\beta, \qquad r_{12}(n) = r_{11}(n-1)\alpha + r_{12}(n-1)(1-\beta),$$

$$r_{21}(n) = r_{21}(n-1)(1-\alpha) + r_{22}(n-1)\beta, \qquad r_{22}(n) = r_{21}(n-1)\alpha + r_{22}(n-1)(1-\beta).$$

If the $r_{ij}(n-1)$ have the given form, it is easily verified by substitution in the Chapman-Kolmogorov equations that the $r_{ij}(n)$ also have the given form.

(c) The steady-state probabilities π_1 and π_2 are obtained by taking the limit of $r_{i1}(n)$ and $r_{i2}(n)$, respectively, as $n \to \infty$. Thus, we have

$$\pi_1 = \frac{\beta}{\alpha + \beta}, \qquad \pi_2 = \frac{\alpha}{\alpha + \beta}.$$

Solution to Problem 7.18. Let the state be the number of days that the gate has survived. The balance equations are

$$\begin{aligned}
\pi_0 &= \pi_0 p + \pi_1 p + \cdots + \pi_{m-1} p + \pi_m, \\
\pi_1 &= \pi_0 (1-p), \\
\pi_2 &= \pi_1 (1-p) = \pi_0 (1-p)^2,
\end{aligned}$$

and similarly

$$\pi_i = \pi_0 (1-p)^i, \qquad i = 1, \ldots, m.$$

We have using the normalization equation

$$1 = \pi_0 + \sum_{i=1}^{m} \pi_i = \pi_0 \left(1 + \sum_{i=1}^{m} (1-p)^i \right),$$

so

$$\pi_0 = \frac{p}{1 - (1-p)^{m+1}}.$$

The long-term expected frequency of gate replacements is equal to the long-term expected frequency of visits to state 0, which is π_0. Note that if the natural lifetime m of a gate is very large, then π_0 is approximately equal to p.

Solution to Problem 7.28. (a) For $j < i$, we have $p_{ij} = 0$. Since the professor will continue to remember the highest ranking, even if he gets a lower ranking in a subsequent year, we have $p_{ii} = i/m$. Finally, for $j > i$, we have $p_{ij} = 1/m$, since the class is equally likely to receive any given rating.

(b) There is a positive probability that on any given year, the professor will receive the highest ranking, namely $1/m$. Therefore, state m is accessible from every other state. The only state accessible from state m is state m itself. Therefore, m is the only recurrent state, and all other states are transient.

(c) This question can be answered by finding the mean first passage time to the absorbing state m starting from i. It is simpler though to argue as follows: since the probability of achieving the highest ranking in a given year is $1/m$, independent of the current state, the required expected number of years is the expected number of trials to the first success in a Bernoulli process with success probability $1/m$. Thus, the expected number of years is m.

Solution to Problem 7.29. (a) There are 3 different paths that lead back to state 1 after 6 transitions. One path makes two self-transitions at state 2, one path makes two self-transitions at state 4, one path makes one self-transition at state 2 and one self-transition at state 4. By adding the probabilities of these three paths, we obtain

$$r_{11}(6) = \frac{2}{3} \cdot \frac{3}{5} \cdot \left(\frac{1}{3} \cdot \frac{2}{5} + \frac{1}{9} + \frac{4}{25}. \right) = \frac{182}{1125}.$$

(b) The time T until the process returns to state 1 is equal to 2 (the time it takes for the transitions from 1 to 2 and from 3 to 4), plus the time it takes for the state to move from state 2 to state 3 (this is geometrically distributed with parameter $p = 2/3$), plus the time it takes for the state to move from state 4 to state 1 (this is geometrically distributed with parameter $p = 3/5$). Using the formulas $\mathbf{E}[X] = 1/p$ and $\mathrm{var}(X) = (1-p)/p^2$ for the mean and variance of a geometric random variable, we find that

$$\mathbf{E}[T] = 2 + \frac{3}{2} + \frac{5}{3} = \frac{31}{6},$$

and

$$\mathrm{var}(T) = \left(1 - \frac{2}{3}\right) \cdot \frac{3^2}{2^2} + \left(1 - \frac{3}{5}\right) \cdot \frac{5^2}{3^2} = \frac{67}{36}.$$

(c) Let A be the event that X_{999}, X_{1000}, and X_{1001} are all different. Note that

$$\mathbf{P}(A \,|\, X_{999} = i) = \begin{cases} 2/3, & \text{for } i = 1, 2, \\ 3/5, & \text{for } i = 3, 4. \end{cases}$$

Thus, using the total probability theorem, and assuming that the process is in steady-state at time 999, we obtain

$$\mathbf{P}(A) = \frac{2}{3}(\pi_1 + \pi_2) + \frac{3}{5}(\pi_3 + \pi_4) = \frac{2}{3} \cdot \frac{15}{31} + \frac{3}{5} \cdot \frac{16}{31} = \frac{98}{155}.$$

Solution to Problem 7.30. (a) States 4 and 5 are transient, and all other states are recurrent. There are two recurrent classes. The class $\{1, 2, 3\}$ is aperiodic, and the class $\{6, 7\}$ is periodic.

(b) If the process starts at state 1, it stays within the aperiodic recurrent class $\{1, 2, 3\}$, and the n-step transition probabilities converge to steady-state probabilities π_i. We have $\pi_i = 0$ for $i \notin \{1, 2, 3\}$. The local balance equations take the form

$$\pi_1 = \pi_2, \qquad \pi_2 = 6\pi_3.$$

Using also the normalization equation, we obtain

$$\pi_1 = \pi_2 = \frac{6}{13}, \qquad \pi_3 = \frac{1}{13}.$$

(c) Because the class $\{6, 7\}$ is periodic, there are no steady-state probabilities. In particular, the sequence $r_{66}(n)$ alternates between 0 and 1, and does not converge.

(d) (i) The probability that the state increases by one during the first transition is equal to

$$0.5\pi_1 + 0.1\pi_2 = \frac{18}{65}.$$

(d) (ii) The probability that the process is in state 2 and that the state increases is

$$0.1\pi_2 = \frac{0.6}{13}.$$

Thus, the desired conditional probability is equal to

$$\frac{0.6/13}{18/65} = \frac{1}{6}.$$

(d) (iii) If the state is 1 (probability 6/13), it is certain to increase at the first change of state. if the state is 2 (probability 6/13), it has probability 1/6 of increasing at the first change of state. Finally, if the state is 3, it cannot increase at the first change of state. Therefore, the probability that the state increases at the first change of state is equal to

$$\frac{6}{13} + \frac{1}{6} \cdot \frac{6}{13} = \frac{7}{13}.$$

(e) (i) Let a_4 and a_5 be the probability that the class $\{1, 2, 3\}$ is eventually reached, starting from state 4 and 5, respectively. We have

$$a_4 = 0.2 + 0.4a_4 + 0.2a_5,$$
$$a_5 = 0.7a_4,$$

which yields

$$a_4 = 0.2 + 0.4a_4 + 0.14a_4,$$

and $a_4 = 10/23$. Also, the probability that the class $\{6, 7\}$ is reached, starting from state 4, is $1 - (10/23) = 13/23$.

(e) (ii) Let μ_4 and μ_5 be the expected times until a recurrent state is reached, starting from state 4 and 5, respectively. We have

$$\mu_4 = 1 + 0.4\mu_4 + 0.2\mu_5,$$
$$\mu_5 = 1 + 0.7\mu_4.$$

Substituting the second equation into the first, and solving for μ_4, we obtain

$$\mu_4 = \frac{60}{23}.$$

Solution to Problem 7.36. Define the state to be the number of operational machines. The corresponding continuous-time Markov chain is the same as a queue with arrival rate λ and service rate μ (the one of Example 7.15). The required probability is equal to the steady-state probability π_0 for this queue.

Solution to Problem 7.37. We consider a continuous-time Markov chain with state $n = 0, 1, \ldots, 4$, where

$$n = \text{number of people waiting.}$$

For $n = 0, 1, 2, 3$, the transitions from n to $n+1$ have rate 1, and the transitions from $n+1$ to n have rate 2. The balance equations are

$$\pi_n = \frac{\pi_{n-1}}{2}, \qquad n = 1, \ldots, 4,$$

so that $\pi_n = \pi_0/2^n$, $n = 1, \ldots, 4$. Using the normalization equation $\sum_{i=0}^{4} \pi_i = 1$, we obtain

$$\pi_0 = \frac{1}{1 + 2^{-1} + 2^{-2} + 2^{-3} + 2^{-4}} = \frac{16}{31}.$$

A passenger who joins the queue (in steady-state) will find n other passengers with probability $\pi_n/(\pi_0 + \pi_1 + \pi_2 + \pi_3)$, for $n = 0, 1, 2, 3$. The expected number of passengers found by Penelope is

$$\mathbf{E}[N] = \frac{\pi_1 + 2\pi_2 + 3\pi_3}{\pi_0 + \pi_1 + \pi_2 + \pi_3} = \frac{(8 + 2 \cdot 4 + 3 \cdot 2)/31}{(16 + 8 + 4 + 2)/31} = \frac{22}{30} = \frac{11}{15}.$$

Since the expected waiting time for a new taxi is 1/2 minute, the expected waiting time (by the law of iterated expectations) is

$$\mathbf{E}[T] = \mathbf{E}[N] \cdot \frac{1}{2} = \frac{11}{30}.$$

Solution to Problem 7.38. Define the state to be the number of pending requests. Thus there are $m + 1$ states, numbered $0, 1, \ldots, m$. At state i, with $1 \leq i \leq m$, the transition rate to $i - 1$ is

$$q_{i,i-1} = \mu.$$

At state i, with $0 \leq i \leq m - 1$, the transition rate to $i + 1$ is

$$q_{i,i+1} = (m - i)\lambda.$$

This is a birth-death process, for which the steady-state probabilities satisfy

$$(m - i)\lambda\pi_i = \mu\pi_{i+1}, \qquad i = 0, 1, \ldots, m - 1,$$

together with the normalization equation

$$\pi_1 + \cdots + \pi_m = 1.$$

The solution to these equations yields the steady-state probabilities.

CHAPTER 8

Solution to Problem 8.1. There are two hypotheses:

$$H_0 : \text{the phone number is 2537267,}$$
$$H_1 : \text{the phone number is not 2537267,}$$

and their prior probabilities are

$$\mathbf{P}(H_0) = \mathbf{P}(H_1) = 0.5.$$

Let B be the event that Artemisia obtains a busy signal when dialing this number. Under H_0, we expect a busy signal with certainty:

$$\mathbf{P}(B \mid H_0) = 1.$$

Under H_1, the conditional probability of B is

$$\mathbf{P}(B \mid H_1) = 0.01.$$

Using Bayes' rule we obtain the posterior probability

$$\mathbf{P}(H_0 \mid B) = \frac{\mathbf{P}(B \mid H_0)\mathbf{P}(H_0)}{\mathbf{P}(B \mid H_0)\mathbf{P}(H_0) + \mathbf{P}(B \mid H_1)\mathbf{P}(H_1)} = \frac{0.5}{0.5 + 0.005} \approx 0.99.$$

Solution to Problem 8.2. (a) Let K (or $\bar{K}$) be the event that Nefeli knew (or did not know, respectively) the answer to the first question, and let C be the event that she answered the question correctly. Using Bayes' rule, we have

$$\mathbf{P}(K \mid C) = \frac{\mathbf{P}(K)\,\mathbf{P}(C \mid K)}{\mathbf{P}(K)\,\mathbf{P}(C \mid K) + \mathbf{P}(\bar{K})\,\mathbf{P}(C \mid \bar{K})} = \frac{0.5 \cdot 1}{0.5 \cdot 1 + 0.5 \cdot \frac{1}{3}} = \frac{3}{4}.$$

(b) The probability that Nefeli knows the answer to a question that she answered correctly is $3/4$ by part (a), so the posterior PMF is binomial with $n = 6$ and $p = 3/4$.

Solution to Problem 8.3. (a) Let X denote the random wait time. We have the observation $X = 30$. Using Bayes' rule, the posterior PDF is

$$f_{\Theta|X}(\theta \mid 30) = \frac{f_\Theta(\theta) f_{X|\Theta}(30 \mid \theta)}{\int f_\Theta(\theta') f_{X|\Theta}(30 \mid \theta')\, d\theta'}.$$

Using the given prior, $f_\Theta(\theta) = 10\theta$ for $\theta \in [0, 1/5]$, we obtain

$$f_{\Theta|X}(\theta \mid 30) = \begin{cases} \dfrac{10\,\theta f_{X|\Theta}(30 \mid \theta)}{\int_0^{1/5} 10\,\theta' f_{X|\Theta}(30 \mid \theta')\, d\theta'}, & \text{if } \theta \in [0, 1/5], \\ 0, & \text{otherwise.} \end{cases}$$

We also have $f_{X|\Theta}(30\,|\,\theta) = \theta e^{-30\theta}$, so the posterior is

$$f_{\Theta|X}(\theta\,|\,30) = \begin{cases} \dfrac{\theta^2 e^{-30\theta}}{\int_0^{1/5} (\theta')^2 e^{-30\theta'}\,d\theta'}, & \text{if } \theta \in [0, 1/5], \\ 0, & \text{otherwise.} \end{cases}$$

The MAP rule selects $\hat{\theta}$ that maximizes the posterior (or equivalently its numerator, since the denominator is a positive constant). By setting the derivative of the numerator to 0, we obtain

$$\frac{d}{d\theta}\left(\theta^2 e^{-30\theta}\right) = 2\,\theta e^{-30\theta} - 30\,\theta^2 e^{-30\theta} = (2 - 30\,\theta)\theta e^{-30\theta} = 0.$$

Therefore, $\hat{\theta} = 2/30$.

The conditional expectation estimator is

$$\mathbf{E}[\Theta\,|\,X = 30] = \frac{\int_0^{1/5} \theta^3 e^{-30\theta}\,d\theta}{\int_0^{1/5} (\theta')^2 e^{-30\theta'}\,d\theta'}.$$

(b) Let X_i denote the random wait time for the ith day, $i = 1, \ldots, 5$. We have the observation vector $X = x$, where $x = (30, 25, 15, 40, 20)$. Using Bayes' rule, the posterior PDF is

$$f_{\Theta|X}(\theta\,|\,x) = \frac{f_\Theta(\theta) f_{X|\Theta}(x\,|\,\theta)}{\int f_\Theta(\theta) f_{X|\Theta}(x\,|\,\theta')\,d\theta'}.$$

In view of the independence of the X_i, we have for $\theta \in [0, 1/5]$,

$$\begin{aligned} f_{X|\Theta}(x\,|\,\theta) &= f_{X_1|\Theta}(x_1\,|\,\theta) \cdots f_{X_5|\Theta}(x_5\,|\,\theta) \\ &= \theta e^{-x_1\theta} \cdots \theta e^{-x_5\theta} \\ &= \theta^5 e^{-(x_1 + \cdots + x_5)\theta} \\ &= \theta^5 e^{-(30+25+15+40+20)\theta} \\ &= \theta^5 e^{-130\theta}. \end{aligned}$$

Using the given prior, $f_\Theta(\theta) = 10\theta$ for $\theta \in [0, 1/5]$, we obtain the posterior

$$f_{\Theta|X}(\theta\,|\,x) = \begin{cases} \dfrac{\theta^6 e^{-130\theta}}{\int_0^{1/5} (\theta')^6 e^{-130\theta'}\,d\theta'}, & \text{if } \theta \in [0, 1/5], \\ 0, & \text{otherwise.} \end{cases}$$

To derive the MAP rule, we set the derivative of the numerator to 0, obtaining

$$\frac{d}{d\theta}\left(\theta^6 e^{-130\theta}\right) = 6\theta^5 e^{-130\theta} - 130\,\theta^6 e^{-130\theta} = (6 - 130\,\theta)\theta^5 e^{-30\theta} = 0.$$

Therefore,

$$\hat{\theta} = \frac{6}{130}.$$

The conditional expectation estimator is

$$\mathbf{E}\big[\Theta \mid X = (30, 25, 15, 40, 20)\big] = \frac{\int_0^{1/5} \theta^7 e^{-130\theta}\, d\theta}{\int_0^{1/5} (\theta')^6 e^{-130\theta'}\, d\theta'}.$$

Solution to Problem 8.4. (a) Let X denote the random variable representing the number of questions answered correctly. For each value $\theta \in \{\theta_1, \theta_2, \theta_3\}$, we have using Bayes' rule,

$$p_{\Theta \mid X}(\theta \mid k) = \frac{p_\Theta(\theta)\, p_{X \mid \Theta}(k \mid \theta)}{\sum_{i=1}^{3} p_\Theta(\theta_i)\, p_{X \mid \Theta}(k \mid \theta_i)}.$$

The conditional PMF $p_{X \mid \Theta}$ is binomial with $n = 10$ and probability of success p_i equal to the probability of answer correctly a question, given that the student is of category i, i.e.,

$$p_i = \theta_i + (1 - \theta_i) \cdot \frac{1}{3} = \frac{2\theta_i + 1}{3}.$$

Thus we have

$$p_1 = \frac{1.6}{3}, \qquad p_2 = \frac{2.4}{3}, \qquad p_3 = \frac{2.9}{3}.$$

For a given number of correct answers k, the MAP rule selects the category i for which the corresponding binomial probability $\binom{10}{k} p_i^k (1 - p_i)^{10-k}$ is maximized.

(b) The posterior PMF of M is given by

$$p_{M|X}(m \mid X = k) = \sum_{i=1}^{3} p_{\Theta \mid X}(\theta_i \mid X = k)\, \mathbf{P}(M = m \mid X = k, \Theta = \theta_i).$$

The probabilities $p_{\Theta \mid X}(\theta_i \mid X = k)$ were calculated in part (a), and the probabilities $\mathbf{P}(M = m \mid X = k, \Theta = \theta_i)$ are binomial and can be calculated in the manner described in Problem 2(b). For $k = 5$, the posterior PMF can be explicitly calculated for $m = 0, \ldots, 5$. The MAP and LMS estimates can be obtained from the posterior PMF.

The probabilities $p_{\Theta \mid X}(\theta_i \mid X = k)$ were calculated in part (a),

$$p_{\Theta \mid X}(\theta_1 \mid X = 5) \approx 0.9010, \quad p_{\Theta \mid X}(\theta_2 \mid X = 5) \approx 0.0989, \quad p_{\Theta \mid X}(\theta_3 \mid X = 5) \approx 0.0001.$$

The probability that the student knows the answer to a question that she answered correctly is

$$q_i = \frac{\theta_i}{\theta_i + (1 - \theta_i)/3}$$

for $i = 1, 2, 3$. The probabilities $\mathbf{P}(M = m \mid X = k, \Theta = \theta_i)$ are binomial and are given by

$$\mathbf{P}(M = m \mid X = k, \Theta = \theta_i) = \binom{k}{m} q_i^m (1 - q_i)^{k-m}$$

For $k = 5$, the posterior PMF can be explicitly calculated for $m = 0, \ldots, 5$

$$p_{M|X}(0 \mid X = 5) \approx 0.0145,$$

$$p_{M|X}(1 \,|\, X = 5) \approx 0.0929,$$
$$p_{M|X}(2 \,|\, X = 5) \approx 0.2402,$$
$$p_{M|X}(3 \,|\, X = 5) \approx 0.3173,$$
$$p_{M|X}(4 \,|\, X = 5) \approx 0.2335,$$
$$p_{M|X}(5 \,|\, X = 5) \approx 0.1015,$$

It follows that the MAP estimate is

$$\hat{m} = 3.$$

The conditional expectation estimate is

$$\mathbf{E}[M|X = 5] = \sum_{m=1}^{5} m p_{M|X}(m \,|\, X = 5) \approx 2.9668 \approx 3.$$

Solution to Problem 8.5. According to the MAP rule, we need to maximize over $\theta \in [0, 1]$ the posterior PDF

$$f_{\Theta|X}(\theta \,|\, k) = \frac{f_\Theta(\theta) p_{X|\Theta}(k \,|\, \theta)}{\displaystyle\int f_\Theta(\theta') p_{X|\Theta}(k \,|\, \theta')\, d\theta'},$$

where X is the number of heads observed. Since the denominator is a positive constant, we only need to maximize

$$f_\Theta(\theta)\, p_{X|\Theta}(k \,|\, \theta) = \binom{n}{k} \left(2 - 4\left|\frac{1}{2} - \theta\right|\right) \theta^k \,(1 - \theta)^{n-k}.$$

The function to be minimized is differentiable except at $\theta = 1/2$. This leads to two different possibilities: (a) the maximum is attained at $\theta = 1/2$; (b) the maximum is attained at some $\theta < 1/2$, at which the derivative is equal to zero; (c) the maximum is attained at some $\theta > 1/2$, at which the derivative is equal to zero.

Let us consider the second possibility. For $\theta < 1/2$, we have $f_\Theta(\theta) = 4\theta$. The function to be maximized, ignoring the constant term $4\binom{n}{k}$, is

$$\theta^{k+1}(1 - \theta)^{n-k}.$$

By setting the derivative to zero, we find $\hat{\theta} = (k + 1)/(n + 1)$, provided that $(k + 1)/(n + 1) < 1/2$. Let us now consider the third possibility. For $\theta > 1/2$, we have $f_\Theta(\theta) = 4(1 - \theta)$. The function to be maximized, ignoring the constant term $4\binom{n}{k}$, is

$$\theta^k(1 - \theta)^{n-k+1}.$$

By setting the derivative to zero, we find $\hat{\theta} = k/(n+1)$, provided that $k/(n+1) > 1/2$. If neither condition $(k + 1)/(n + 1) < 1/2$ and $k/(n + 1) > 1/2$ holds, we must have

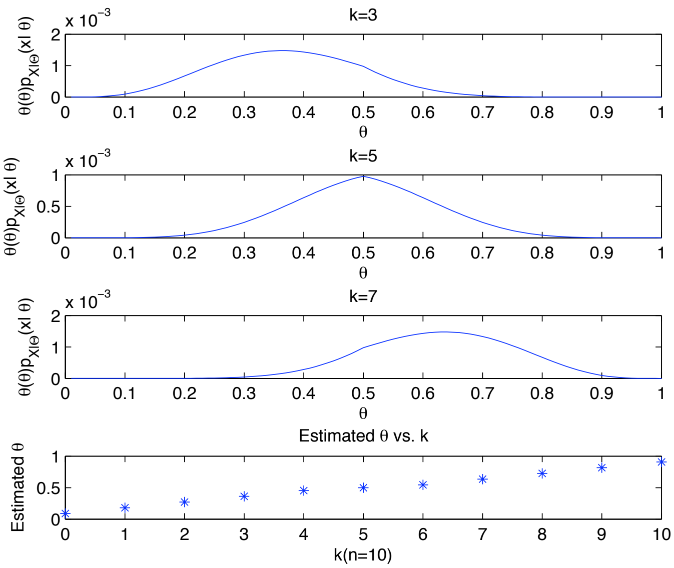

Figure 8.1: (a)-(c) Plots of the function $f_\theta(\theta)\theta^k(1-\theta)^{n-k}$ in Problem 8.5, when $n = 10$, and for $k = 3, 5, 7$, respectively. (d) The MAP estimate $\hat{\theta}$ as a function of k, when $n = 10$.

the first possibility, with the maximum attained at $\hat{\theta} = 1/2$. To summarize, the MAP estimate is given by

$$\hat{\theta} = \begin{cases} \dfrac{k+1}{n+1}, & \text{if } \dfrac{k+1}{n+1} < \dfrac{1}{2}, \\ \dfrac{1}{2}, & \text{if } \dfrac{k}{n+1} \le \dfrac{1}{2} \le \dfrac{k+1}{n+1}, \\ \dfrac{k}{n+1}, & \text{if } \dfrac{1}{2} < \dfrac{k}{n+1}. \end{cases}$$

Figure 8.1 shows a plot of the function $f_\theta(\theta)\,\theta^k(1-\theta)^{n-k}$, for three different values of k, as well as a plot of $\hat{\theta}$ as function of k, all for the case where $n = 10$.

Solution to Problem 8.6. (a) First we calculate the values of c_1 and c_2. We have

$$c_1 = \frac{1}{\int_5^{60} e^{-0.04x}\,dx} \approx 0.0549,$$

$$c_2 = \frac{1}{\int_5^{60} e^{-0.16x}\,dx} \approx 0.3561.$$

Next we derive the posterior probability of each hypothesis,

$$\begin{aligned}
p_{\Theta|T}(1\,|\,20) &= \frac{0.3 f_{T\,|\,\Theta}(x\,|\,\Theta=1)}{0.3 f_{T\,|\,\Theta}(x\,|\,\Theta=1) + 0.7 f_{T\,|\,\Theta}(x\,|\,\Theta=2)} \\
&= \frac{0.3\cdot 0.0549\, e^{-0.04\cdot 20}}{0.3\cdot 0.0549\, e^{-0.04\cdot 20} + 0.7\cdot 0.3561\, e^{-0.16\cdot 20}} \\
&= 0.4214,
\end{aligned}$$

and

$$\begin{aligned}
p_{\Theta|T}(2\,|\,20) &= \frac{0.7 f_{T\,|\,\Theta}(x\,|\,\Theta=2)}{0.3 f_{T\,|\,\Theta}(x\,|\,\Theta=1) + 0.7 f_{T\,|\,\Theta}(x\,|\,\Theta=2)} \\
&= \frac{0.7\cdot 0.3561\, e^{-0.16\cdot 20}}{0.3\cdot 0.0549\, e^{-0.04\cdot 20} + 0.7\cdot 0.3561\, e^{-0.16\cdot 20}} \\
&= 0.5786.
\end{aligned}$$

Therefore she would accept the hypothesis that the problem is not difficult, and the probability of error is

$$p_e = p_{\Theta|T}(1\,|\,20) = 0.4214.$$

(b) We write the posterior probability of each hypothesis,

$$\begin{aligned}
&p_{\Theta|T_1,T_2,T_3,T_4,T_5}(1\,|\,20,10,25,15,35) \\
&= \frac{0.3 f_{T_1,T_2,T_3,T_4,T_5\,|\,\Theta}(20,10,25,15,35\,|\,\Theta=1)}{0.3 f_{T_1,T_2,T_3,T_4,T_5\,|\,\Theta}(20,10,25,15,35\,|\,\Theta=1) + 0.7 f_{T_1,T_2,T_3,T_4,T_5\,|\,\Theta}(20,10,25,15,35\,|\,\Theta=2)} \\
&= \frac{0.3\cdot 0.0549^5 \exp(-0.04\cdot(20+10+25+15+35))}{0.3\cdot 0.0549^5 \exp(-0.04\cdot(20+10+25+15+35)) + 0.7\cdot 0.3561^5 \exp(-0.16\cdot(20+10+25+15+35))} \\
&= 0.9171,
\end{aligned}$$

and similarly

$$\begin{aligned}
&p_{\Theta|T_1,T_2,T_3,T_4,T_5}(2\,|\,20,10,25,15,35) \\
&= \frac{0.3 f_{T_1,T_2,T_3,T_4,T_5\,|\,\Theta}(20,10,25,15,35\,|\,\Theta=2)}{0.3 f_{T_1,T_2,T_3,T_4,T_5\,|\,\Theta}(20,10,25,15,35\,|\,\Theta=1) + 0.7 f_{T_1,T_2,T_3,T_4,T_5\,|\,\Theta}(20,10,25,15,35\,|\,\Theta=2)} \\
&= 0.0829.
\end{aligned}$$

So this time the professor would accept the hypothesis that the problem is difficult. The probability of error is 0.0829, much lower than the case of a single observation.

Solution to Problem 8.7. (a) Let H_1 and H_2 be the hypotheses that box 1 or 2, respectively, was chosen. Let $X = 1$ if the drawn ball is white, and $X = 2$ if it is black. We introduce a parameter/random variable Θ, taking values θ_1 and θ_2, corresponding to H_1 and H_2, respectively. We have the following prior distribution for Θ:

$$p_\Theta(\theta_1) = p, \qquad p_\Theta(\theta_2) = 1 - p,$$

where p is given. Using Bayes' rule, we have

$$\begin{aligned} p_{\Theta|X}(\theta_1 \,|\, 1) &= \frac{p_\Theta(\theta_1)p_{X|\Theta}(1 \,|\, \theta_1)}{p_\Theta(\theta_1)p_{X|\Theta}(1 \,|\, \theta_1) + p_\Theta(\theta_2)p_{X|\Theta}(1 \,|\, \theta_2)} \\ &= \frac{2p/3}{2p/3 + (1-p)/3} \\ &= \frac{2p}{1+p}. \end{aligned}$$

Similarly we calculate the other conditional probabilities of interest:

$$p_{\Theta|X}(\theta_2 \,|\, 1) = \frac{1-p}{1+p}, \qquad p_{\Theta|X}(\theta_1 \,|\, 2) = \frac{p}{2-p}, \qquad p_{\Theta|X}(\theta_2 \,|\, 2) = \frac{2-2p}{2-p}.$$

If a white ball is drawn ($X = 1$), the MAP rule selects box 1 if

$$p_{\Theta|X}(\theta_1 \,|\, 1) > p_{\Theta|X}(\theta_2 \,|\, 1),$$

that is, if

$$\frac{2p}{1+p} > \frac{1-p}{1+p},$$

or $p > 1/3$, and selects box 2 otherwise. If a black ball is drawn ($X = 2$), the MAP rule is selects box 1 if

$$p_{\Theta|X}(\theta_1 \,|\, 2) > p_{\Theta|X}(\theta_2 \,|\, 2),$$

that is, if

$$\frac{p}{2-p} > \frac{2-2p}{2-p},$$

or $p > 2/3$, and selects box 2 otherwise.

Suppose now that the two boxes have equal prior probabilities ($p = 1/2$). Then, the MAP rule decides on box 1 (or box 2) if $X = 1$ (or $X = 2$, respectively). Given an initial choice of box 1 ($\Theta = \theta_1$), the probability of error is

$$e_1 = \mathbf{P}(X = 2 \,|\, \theta_1) = \frac{1}{3}.$$

Similarly, for an initial choice of box 2 ($\Theta = \theta_2$), the probability of error is

$$e_2 = \mathbf{P}(X = 1 \,|\, \theta_2) = \frac{1}{3}.$$

The overall probability of error of the MAP decision rule is obtained using the total probability theorem:

$$\mathbf{P}(\text{error}) = p_\Theta(\theta_1)e_1 + p_\Theta(\theta_2)e_2 = \frac{1}{2}\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{3}.$$

Thus, whereas prior to knowing the data (the value of X), the probability of error for either decision was 1/2, after knowing the data and using the MAP rule, the probability of error is reduced to 1/3. This is in fact a general property of the MAP rule: with more data, the probability of error cannot increase, regardless of the observed value of X (see Problem 8.9).

Solution to Problem 8.8. (a) Let K be the number of heads observed before the first tail, and let $p_{K|H_i}(k)$ be the PMF of K when hypothesis H_i is true. Note that event $K = k$ corresponds to a sequence of k heads followed by a tail, so that

$$p_{K|H_i}(k) = (1 - q_i)q_i^k, \qquad k = 0, 1, \ldots, \quad i = 1, 2.$$

Using Bayes' rule, we obtain

$$\begin{aligned} \mathbf{P}(H_1 \,|\, K = k) &= \frac{p_{K|H_1}(k)\mathbf{P}(H_1)}{p_K(k)} \\ &= \frac{\frac{1}{2}(1 - q_1)q_1^k}{\frac{1}{2}(1 - q_1)q_1^k + \frac{1}{2}(1 - q_0)q_0^k} \\ &= \frac{(1 - q_1)q_1^k}{(1 - q_1)q_1^k + (1 - q_0)q_0^k}. \end{aligned}$$

(b) An error occurs in two cases: if H_0 is true and $K \geq k^*$, or if H_1 is true and $K < k^*$. So, the probability of error, denoted by p_e, is

$$\begin{aligned} p_e &= \mathbf{P}(K \geq k^* \,|\, H_0)\mathbf{P}(H_0) + \mathbf{P}(K < k^* \,|\, H_1)\mathbf{P}(H_1) \\ &= \sum_{k=k^*}^{\infty} p_{K|H_0}(k)\mathbf{P}(H_0) + \sum_{k=0}^{k^*-1} p_{K|H_1}(k)\mathbf{P}(H_1) \\ &= \mathbf{P}(H_0) \sum_{k=k^*}^{\infty} (1 - q_0)q_0^k + \mathbf{P}(H_1) \sum_{k=0}^{k^*-1} (1 - q_1)q_1^k \\ &= \mathbf{P}(H_0)(1 - q_0)\frac{q_0^{k^*}}{1 - q_0} + \mathbf{P}(H_1)(1 - q_1)\frac{1 - q_1^{k^*}}{1 - q_1} \\ &= \mathbf{P}(H_1) + \mathbf{P}(H_0)q_0^{k^*} - \mathbf{P}(H_1)q_1^{k^*} \\ &= \frac{1}{2}(1 + q_0^{k^*} - q_1^{k^*}). \end{aligned}$$

To find the value of k^* that minimizes p_e, we temporarily treat k^* as a continuous variable and differentiate p_e with respect to k^*. Setting this derivative to zero, we obtain

$$\frac{dp_e}{dk^*} = \frac{1}{2}\big((\log q_0)q_0^{k^*} - (\log q_1)q_1^{k^*}\big) = 0.$$

The solution to this equation is

$$\overline{k} = \frac{\log(|\log q_0|) - \log(|\log q_1|)}{|\log q_0| - |\log q_1|}.$$

As k^* ranges from 0 to $\overline{k}$, the derivative of p_e is nonzero, so that p_e is monotonic. Since $q_1 > q_0$, the derivative is negative at $k^* = 0$. This implies that p_e is monotonically decreasing as k^* ranges from 0 to $\overline{k}$. Similarly, the derivative of p_e is positive for very large values of k^*, which implies that p_e is monotonically increasing as k^* ranges from $\overline{k}$ to infinity. It follows that $\overline{k}$ minimizes p_e. However, k^* can only take integer values, so the integer k^* that minimizes p_e is either $\lfloor \overline{k} \rfloor$ or $\lceil \overline{k} \rceil$, whichever gives the lower value of P_e.

We now derive the form of the MAP decision rule, which minimizes the probability of error, and show that it is of the same type as the decision rules we just studied. With the MAP decision rule, for any given k, we accept H_1 if

$$\mathbf{P}(K = k \,|\, H_1)\mathbf{P}(H_1) > \mathbf{P}(K = k \,|\, H_0)\mathbf{P}(H_0),$$

and accept H_0 otherwise. Note that if

$$(1 - q_1)q_1^k\mathbf{P}(H_1) > (1 - q_0)q_0^k\mathbf{P}(H_0),$$

then

$$(1 - q_1)q_1^{k+1}\mathbf{P}(H_1) > (1 - q_0)q_0^{k+1}\mathbf{P}(H_0),$$

since $q_1 > q_0$. Similarly, if

$$(1 - q_1)q_1^k\mathbf{P}(H_1) < (1 - q_0)q_0^k\mathbf{P}(H_0),$$

then

$$(1 - q_1)q_1^{k-1}\mathbf{P}(H_1) < (1 - q_0)q_0^{k-1}\mathbf{P}(H_0).$$

This implies that if we decide in favor of H_1 when a value k is observed, then we also decide in favor of H_1 when a larger value is observed. Similarly, if we decide in favor H_0 when a value k is observed, then we also decide in favor of H_0 when a smaller value k is observed. Therefore, the MAP rule is of the type considered and optimized earlier, and thus will not result in a lower value of p_e.

(c) As in part (b) we have

$$p_e = \mathbf{P}(H_1) + \mathbf{P}(H_0)q_0^{k^*} - \mathbf{P}(H_1)q_1^{k^*}.$$

Consider the case where $\mathbf{P}(H_1) = 0.7$, $q_0 = 0.3$ and $q_1 = 0.7$. Using the calculations in part (b), we have

$$\overline{k} = \frac{\log\left(\dfrac{\mathbf{P}(H_0)\log(v_0)}{\mathbf{P}(H_1)\log(v_1)}\right)}{\log\left(\dfrac{v_1}{v_0}\right)} \approx 0.43$$

Thus, the optimal value of k^* is either $\lfloor \overline{k} \rfloor = 0$ or $\lceil \overline{k} \rceil = 1$. We find that with either choice the probability of error p_e is the same and equal to 0.3. Thus, either choice minimizes the probability of error.

Note that $\overline{k}$ decreases as $\mathbf{P}(H_1)$ increases from 0.7 to 1.0. So the choice $k^* = 0$ remains optimal in this range. As a result, we always decide in favor of H_1, and the probability of error is $p_e = \mathbf{P}(H_0) = 1 - \mathbf{P}(H_1)$.

Solution to Problem 8.10. Let Θ be the car speed and let X be the radar's measurement. Similar to Example 8.11, the joint PDF of Θ and X is uniform over the set of pairs (θ, x) that satisfy $55 \leq \theta \leq 75$ and $\theta \leq x \leq \theta + 5$. As in Example 8.11, for any given x, the value of Θ is constrained to lie on a particular interval, the posterior PDF of Θ is uniform over that interval, and the conditional mean is the midpoint of that interval. In particular,

$$\mathbf{E}[\Theta \mid X = x] = \begin{cases} \dfrac{x}{2} + 27.5, & \text{if } 55 \le x \le 60, \\ x - 2.5, & \text{if } 60 \le x \le 75, \\ \dfrac{x}{2} + 35, & \text{if } 75 \le x \le 80. \end{cases}$$

Solution to Problem 8.11. From Bayes' rule,

$$\begin{aligned} p_{\Theta|X}(\theta \mid x) &= \frac{p_{X|\Theta}(x \mid \theta) p_\Theta(\theta)}{p_X(x)} \\ &= \frac{p_{X|\Theta}(x \mid \theta) p_\Theta(\theta)}{\displaystyle\sum_{i=1}^{100} p_{X|\Theta}(x \mid i) p_\Theta(i)} \\ &= \frac{\dfrac{1}{\theta} \cdot \dfrac{1}{100}}{\displaystyle\sum_{i=x}^{100} \dfrac{1}{i} \cdot \dfrac{1}{100}} \\ &= \begin{cases} \dfrac{\dfrac{1}{\theta}}{\displaystyle\sum_{i=x}^{100} \dfrac{1}{i}}, & \text{for } \theta = x, x+1, \ldots, 100, \\ 0, & \text{for } \theta = 1, 2, \ldots, x-1. \end{cases} \end{aligned}$$

Given that $X = x$, the posterior probability is maximized at $\hat{\theta} = x$, and this is the MAP estimate of Θ given x. The LMS estimate is

$$\hat{\theta} = \mathbf{E}[\Theta \mid X = x] = \sum_{\theta=1}^{100} \theta p_{\Theta|X}(\theta \mid x) = \frac{101 - x}{\displaystyle\sum_{i=x}^{100} \frac{1}{i}}.$$

Figure 8.2 plots the MAP and LMS estimates of Θ as a function of X.

Solution to Problem 8.12. (a) The posterior PDF is

$$\begin{aligned} f_{\Theta|X_1,\ldots,X_n}(\theta \mid x_1, \ldots, x_n)\, d\theta &= \frac{f_\Theta(\theta) f_{X|\Theta}(x \mid \theta)}{\displaystyle\int_0^1 f_\Theta(\theta') f_{X|\Theta}(x \mid \theta')\, d\theta'} \\ &= \begin{cases} \dfrac{\dfrac{1}{\theta^n}}{\dfrac{1}{n-1}\overline{x}^{1-n} - \dfrac{1}{n-1}}, & \text{if } \overline{x} \le \theta, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

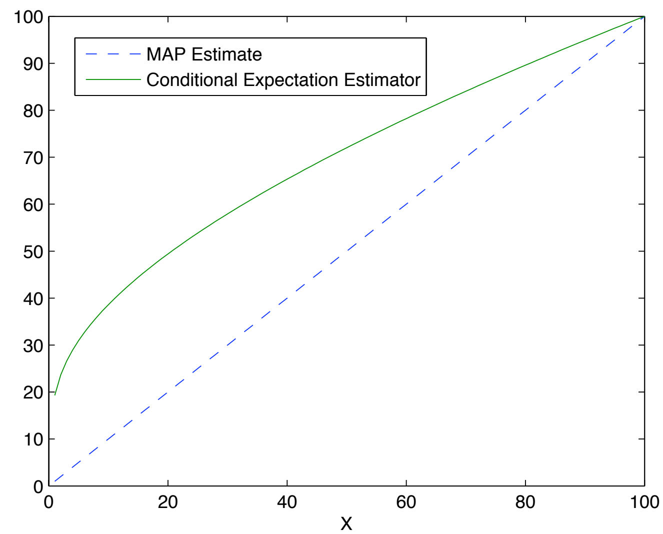

Figure 8.2: MAP and LMS estimates of Θ as a function of X in Problem 8.11.

Using the definition of conditional expectation we obtain

$$\begin{aligned}
\mathbf{E}[\Theta \mid X_1 = x_1, \ldots, X_n = x_n] &= \int_0^1 \theta \cdot f_{\Theta|X_1,\ldots,X_n}(\theta \mid x_1, \ldots, x_n)\, d\theta \\
&= \int_{\overline{x}}^1 \theta \cdot f_{\Theta|X_1,\ldots,X_n}(\theta \mid x_1, \ldots, x_n)\, d\theta \\
&= \frac{\dfrac{1}{n-2}\overline{x}^{2-n} - \dfrac{1}{n-2}}{\dfrac{1}{n-1}\overline{x}^{1-n} - \dfrac{1}{n-1}} \\
&= \frac{n-1}{n-2} \cdot \frac{\overline{x}^{2-n} - 1}{\overline{x}^{1-n} - 1} \\
&= \frac{n-1}{n-2} \cdot \frac{\overline{x}^{2-n} - 1}{\overline{x}^{1-n} - 1} \\
&= \frac{n-1}{n-2} \cdot \frac{x(1 - x^{n-2})}{1 - x^{n-1}}.
\end{aligned}$$

(b) The conditional mean squared error of the MAP estimator is

$$\begin{aligned}
\mathbf{E}\big[(\hat{\Theta} - \Theta)^2 \mid X_1 = x_1, \ldots, X_n = x_n\big] &= \mathbf{E}\left[(\overline{x} - \Theta)^2 \mid X_1 = x_1, \ldots, X_n = x_n\right] \\
&= \overline{x}^2 - 2\overline{x}\frac{n-1}{n-2} \cdot \frac{\overline{x}^{2-n} - 1}{\overline{x}^{1-n} - 1} + \left(\frac{n-1}{n-3} \cdot \frac{\overline{x}^{3-n} - 1}{\overline{x}^{1-n} - 1}\right),
\end{aligned}$$

and the conditional mean squared error for the LMS estimator is

$$\mathbf{E}\big[(\hat{\Theta}-\Theta)^2 \,|\, X_1 = x_1, \ldots, X_n = x_n\big] = \mathbf{E}\left[\left(\frac{\overline{x}^{2-n}-1}{\overline{x}^{1-n}-1} - \Theta\right)^2 \;\middle|\; X_1 = x_1, \ldots, X_n = x_n\right]$$
$$= -\left(\frac{n-1}{n-2}\cdot\frac{\overline{x}^{2-n}-1}{\overline{x}^{1-n}-1}\right)^2 + \left(\frac{n-1}{n-3}\cdot\frac{\overline{x}^{3-n}-1}{\overline{x}^{1-n}-1}\right).$$

We plot in Fig. 8.3 the estimators and the corresponding conditional mean squared errors as a function of $\overline{x}$, for the case where $n=5$.

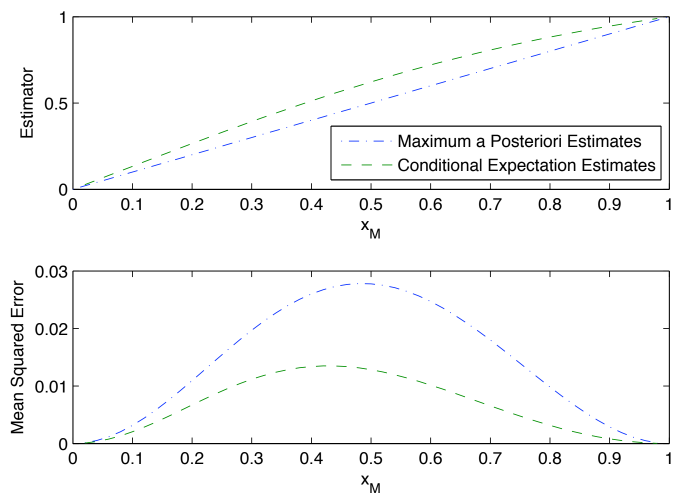

Figure 8.3: The MAP and LMS estimates, and their conditional mean squared errors, as functions of $\overline{x}$ in Problem 8.12.

(c) When $\overline{x}$ is held fixed at 0.5, the MAP estimate also remains fixed at 0.5. On the other hand, the LMS estimate given by the expression found in part (b), can be seen to be larger than 0.5 and converge to 0.5 as $n \to \infty$. Furthermore, the conditional mean squared error decreases to zero as n increases to infinity; see Fig. 8.4.

Solution to Problem 8.14. Here Θ is uniformly distributed in the interval $[4,10]$ and

$$X = \Theta + W,$$

where W is uniformly distributed in the interval $[-1,1]$, and is independent of Θ. The linear LMS estimator of Θ given X is

$$\hat{\Theta} = \mathbf{E}[\Theta] + \frac{\mathrm{cov}(\Theta, X)}{\sigma_X^2}\big(X - \mathbf{E}[X]\big).$$

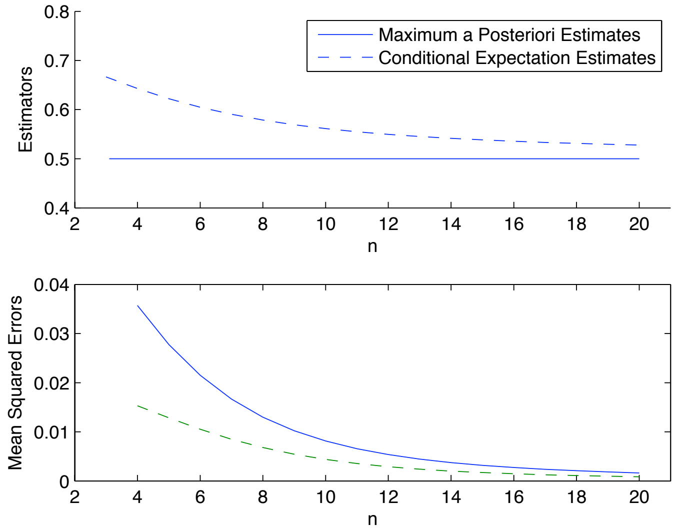

Figure 8.4: Asymptotic behavior of the MAP and LMS estimators, and the corresponding conditional mean squared errors, for fixed $\overline{x} = 0.5$, and $n \to \infty$ in Problem 8.12.

We have

$$\mathbf{E}[X] = \mathbf{E}[\Theta] + \mathbf{E}[W] = \mathbf{E}[\Theta], \qquad \sigma_X^2 = \sigma_\Theta^2 + \sigma_W^2,$$

$$\text{cov}(\Theta, X) = \mathbf{E}\Big[\big(\Theta - \mathbf{E}[\Theta]\big)\big(X - \mathbf{E}[X]\big)\Big] = \mathbf{E}\Big[\big(\Theta - \mathbf{E}[\Theta]\big)^2\Big] = \sigma_\Theta^2,$$

where the last relation follows from the independence of Θ and W. Using the formulas for the mean and variance of the uniform PDF, we have

$$\begin{aligned} \mathbf{E}[\Theta] &= 7, & \sigma_\Theta^2 &= 3, \\ \mathbf{E}[W] &= 0, & \sigma_W^2 &= 1/3. \end{aligned}$$

Thus, the linear LMS estimator is

$$\hat{\Theta} = 7 + \frac{3}{3 + 1/3}\big(X - 7\big),$$

or

$$\hat{\Theta} = 7 + \frac{9}{10}\big(X - 7\big).$$

The mean squared error is $(1 - \rho^2)\sigma_\Theta^2$. We have

$$\rho^2 = \left(\frac{\text{cov}(\Theta, X)}{\sigma_\Theta \sigma_X}\right)^2 = \left(\frac{\sigma_\Theta^2}{\sigma_\Theta \sigma_X}\right)^2 = \frac{\sigma_\Theta^2}{\sigma_X^2} = \frac{3}{3 + 1/3} = \frac{9}{10}.$$

Hence the mean squared error is

$$(1-\rho^2)\sigma_\Theta^2 = \left(1-\frac{9}{10}\right)\cdot 3 = \frac{3}{10}.$$

Solution to Problem 8.15. The conditional mean squared error of the MAP estimator $\hat{\Theta} = X$ is

$$\begin{aligned}
\mathbf{E}\big[(\hat{\Theta}-\Theta)^2 \,|\, X=x\big] &= \mathbf{E}\big[\hat{\Theta}^2 - 2\hat{\Theta}\Theta + \Theta^2 \,|\, X=x\big] \\
&= x^2 - 2x\mathbf{E}[\Theta \,|\, X=x] + \mathbf{E}[\Theta^2 \,|\, X=x] \\
&= x^2 - 2x\frac{101-x}{\displaystyle\sum_{i=x}^{100}\frac{1}{i}} + \frac{\displaystyle\sum_{i=x}^{100} i}{\displaystyle\sum_{i=x}^{100}\frac{1}{i}}.
\end{aligned}$$

The conditional mean squared error of the LMS estimator

$$\hat{\Theta} = \frac{101-X}{\displaystyle\sum_{i=X}^{100}\frac{1}{i}}.$$

is

$$\begin{aligned}
\mathbf{E}[(\hat{\Theta}-\Theta)^2 \,|\, X=x] &= \mathbf{E}[\hat{\Theta}^2 - 2\hat{\Theta}\Theta + \Theta^2 \,|\, X=x] \\
&= \frac{101-x^2}{\displaystyle\sum_{i=x}^{100}\frac{1}{i}} - 2\frac{101-x}{\displaystyle\sum_{i=x}^{100}\frac{1}{i}}\mathbf{E}[\Theta \,|\, X=x] + \mathbf{E}[\Theta^2 \,|\, X=x] \\
&= -\frac{(101-x)^2}{\displaystyle\left(\sum_{i=x}^{100}\frac{1}{i}\right)^2} + \frac{\displaystyle\sum_{i=x}^{100} i}{\displaystyle\sum_{i=x}^{100}\frac{1}{i}}.
\end{aligned}$$

To obtain the linear LMS estimator, we compute the expectation and variance of X. We have

$$\mathbf{E}[X] = \mathbf{E}\big[\mathbf{E}[X \,|\, \Theta]\big] = \mathbf{E}\Big[\frac{\Theta+1}{2}\Big] = \frac{(101/2)+1}{2} = 25.75,$$

and

$$\mathrm{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 = \frac{1}{100}\sum_{x=1}^{100} x^2 \left(\sum_{\theta=x}^{100}\frac{1}{\theta}\right) - (25.75)^2 = 490.19.$$

The covariance of Θ and X is

$$\begin{aligned}\mathrm{cov}(\Theta, X) &= \mathbf{E}\big[(X - \mathbf{E}[X])(\Theta - \mathbf{E}[\Theta])\big] \\ &= \sum_{\theta=1}^{100} \frac{1}{100} \sum_{x=1}^{\theta} \frac{1}{\theta}(x - 25.75)(\theta - 50) \\ &= 416.63.\end{aligned}$$

Applying the linear LMS formula yields

$$\hat{\Theta} = \mathbf{E}[\Theta] + \frac{\mathrm{cov}(\Theta, X)}{\mathrm{var}(X)}\big(X - \mathbf{E}[X]\big) = 50 + \frac{416.63}{490.19}(X - 25.75) = 0.85X + 28.11.$$

The mean squared error of the linear LMS estimator is

$$\begin{aligned}\mathbf{E}\big[(\hat{\Theta} - \Theta)^2 \,|\, X = x\big] &= \mathbf{E}\big[\hat{\Theta}^2 - 2\hat{\Theta}\Theta + \Theta^2 \,|\, X = x\big] \\ &= \hat{\Theta}^2 - 2\hat{\Theta}\mathbf{E}[\Theta \,|\, X = x] + \mathbf{E}[\Theta^2 \,|\, X = x] \\ &= (0.85x + 28.11)^2 - 2\,(0.85x + 28.11)\frac{101 - x}{\sum_{i=x}^{100} \frac{1}{i}} + \frac{\sum_{i=x}^{100} i}{\sum_{i=x}^{100} \frac{1}{i}}.\end{aligned}$$

Figure 8.5 plots the conditional mean squared error of the MAP, LMS, and linear LMS estimators, as a function of x. Note that the conditional mean squared error is lowest for the LMS estimator, but that the linear LMS estimator comes very close.

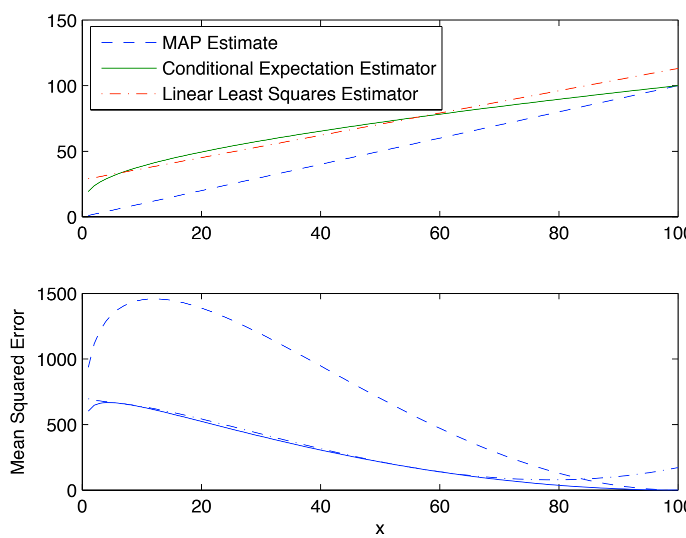

Figure 8.5: Estimators and their conditional mean squared errors in Problem 8.15.

Solution to Problem 8.16. (a) The LMS estimator is

$$g(X) = \mathbf{E}[\Theta \mid X] = \begin{cases} \dfrac{1}{2}X, & \text{if } 0 \le X < 1, \\ X - \dfrac{1}{2}, & \text{if } 1 \le X \le 2. \end{cases}$$

(b) We first derive the conditional variance $\mathbf{E}\big[(\Theta - g(X))^2 \mid X = x\big]$. If $x \in [0,1]$, the conditional PDF of Θ is uniform over the interval $[0, x]$, and

$$\mathbf{E}\Big[\big(\Theta - g(X)\big)^2 \mid X = x\Big] = x^2/12.$$

Similarly, if $x \in [1,2]$, the conditional PDF of Θ is uniform over the interval $[1-x, x]$, and

$$\mathbf{E}\Big[\big(\Theta - g(X)\big)^2 \mid X = x\Big] = 1/12.$$

We now evaluate the expectation and variance of $g(X)$. Note that (Θ, X) is uniform over a region with area $3/2$, so that the constant c must be equal to $2/3$. We have

$$\begin{aligned} \mathbf{E}\big[g(X)\big] &= \mathbf{E}\big[\mathbf{E}[\Theta \mid X]\big] \\ &= \mathbf{E}[\Theta] \\ &= \int\!\!\int \theta f_{X,\Theta}(x,\theta)\, d\theta dx \\ &= \int_0^1 \int_0^x \theta\, \frac{2}{3}\, d\theta\, dx + \int_1^2 \int_{x-1}^x \theta\, \frac{2}{3}\, d\theta\, dx \\ &= \frac{7}{9}. \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathrm{var}\big(g(X)\big) &= \mathrm{var}\big(\mathbf{E}[\Theta \mid X]\big) \\ &= \mathbf{E}\Big[\big(\mathbf{E}[\Theta \mid X]\big)^2\Big] - \Big(\mathbf{E}\big[\mathbf{E}[\Theta \mid X]\big]\Big)^2 \\ &= \int_0^2 \big(\mathbf{E}[\Theta \mid X]\big)^2 f_X(x)\, dx - \big(\mathbf{E}[\Theta]\big)^2 \\ &= \int_0^1 \Big(\frac{1}{2}x\Big)^2 \cdot \frac{2}{3}x\, dx + \int_1^2 \Big(x - \frac{1}{2}\Big)^2 \cdot \frac{2}{3}\, dx - \Big(\frac{7}{9}\Big)^2 \\ &= \frac{103}{648} \\ &= 0.159, \end{aligned}$$

where

$$f_X(x) = \begin{cases} 2x/3, & \text{if } 0 \le x \le 1, \\ 2/3, & \text{if } 1 \le x \le 2. \end{cases}$$

(c) The expectations $\mathbf{E}\Big[\big(\Theta - g(X)\big)^2\Big]$ and $\mathbf{E}\big[\text{var}(\Theta \,|\, X)\big]$ are equal because by the law of iterated expectations,

$$\mathbf{E}\Big[\big(\Theta - g(X)\big)^2\Big] = \mathbf{E}\Big[\mathbf{E}\big[\big(\Theta - g(X)\big)^2 \,|\, X\big]\Big] = \mathbf{E}\big[\text{var}(\Theta \,|\, X)\big].$$

Recall from part (b) that

$$\text{var}(\Theta \,|\, X = x) = \begin{cases} x^2/12, & \text{if } 0 \le x < 1, \\ 1/12, & \text{if } 1 \le x \le 2. \end{cases}$$

It follows that

$$\mathbf{E}\big[\text{var}(\Theta \,|\, X)\big] = \int_x \text{var}(\Theta \,|\, X = x) f_X(x)\, dx = \int_0^1 \frac{x^2}{12} \cdot \frac{2}{3} x\, dx + \int_1^2 \frac{1}{12} \cdot \frac{2}{3}\, dx = \frac{5}{72}.$$

(d) By the law of total variance, we have

$$\text{var}(\Theta) = \mathbf{E}\big[\text{var}(\Theta \,|\, X)\big] + \text{var}\big(\mathbf{E}[\Theta \,|\, X]\big).$$

Using the results from parts (b) and (c), we have

$$\text{var}(\Theta) = \mathbf{E}\big[\text{var}(\Theta \,|\, X)\big] + \text{var}\big(\mathbf{E}[\Theta \,|\, X]\big) = \frac{5}{72} + \frac{103}{648} = \frac{37}{162}.$$

An alternative approach to calculating the variance of Θ is to first find the marginal PDF $f_\Theta(\theta)$ and then apply the definition

$$\text{var}(\Theta) = \int_0^2 \big(\theta - \mathbf{E}[\Theta]\big)^2 f_\Theta(\theta)\, d\theta.$$

(e) The linear LMS estimator is

$$\hat{\Theta} = \mathbf{E}[\Theta] + \frac{\text{cov}(X, \Theta)}{\sigma_X^2}\big(X - \mathbf{E}[X]\big).$$

We have

$$\mathbf{E}[X] = \int_0^1 \int_0^x \frac{2}{3} x\, d\theta\, dx + \int_1^2 \int_{x-1}^x \frac{2}{3} x\, d\theta\, dx = \frac{2}{9} + 1 = \frac{11}{9},$$

$$\mathbf{E}[X^2] = \int_0^1 \int_0^x \frac{2}{3} x^2\, d\theta\, dx + \int_1^2 \int_{x-1}^x \frac{2}{3} x^2\, d\theta\, dx = \frac{1}{6} + \frac{14}{9} = \frac{31}{18},$$

$$\text{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \frac{71}{162},$$

$$\mathbf{E}[\Theta] = \int_0^1 \int_0^x \frac{2}{3} \theta\, d\theta\, dx + \int_1^2 \int_{x-1}^x \frac{2}{3} \theta\, d\theta\, dx = \frac{1}{9} + \frac{2}{3} = \frac{7}{9},$$

$$\mathbf{E}[X\Theta] = \int_0^1 \int_0^x \frac{2}{3} x\theta \, d\theta \, dx + \int_1^2 \int_{x-1}^x \frac{2}{3} x\theta \, d\theta \, dx = \frac{1}{12} + \frac{17}{18} = \frac{37}{36},$$

$$\text{cov}(X, \Theta) = \mathbf{E}[X\Theta] - \mathbf{E}[X]\mathbf{E}[\Theta] = \frac{37}{36} - \frac{11}{9} \cdot \frac{7}{9}.$$

Thus, the linear LMS estimator is

$$\hat{\Theta} = \frac{7}{9} + \frac{\frac{37}{36} - \frac{11}{9} \cdot \frac{7}{9}}{\frac{71}{162}} \left(X - \frac{11}{9} \right) = 0.5626 + 0.1761X.$$

Its mean squared error is

$$\begin{aligned} \mathbf{E}\left[(\Theta - \hat{\Theta})^2\right] &= \mathbf{E}\left[(\Theta - 0.5626 - 0.1761X)^2 \right] \\ &= \mathbf{E}\left[\Theta^2 - 2\Theta\,(0.5626 + 0.1761X) + (0.5626 + 0.1761X)^2 \right]. \end{aligned}$$

After some calculation we obtain the value of the mean squared error, which is approximately 0.2023. Alternatively, we can use the values of $\text{var}(X)$, $\text{var}(\Theta)$, and $\text{cov}(X, \Theta)$ we found earlier, to calculate the correlation coefficient ρ, and then use the fact that the mean squared error is equal to

$$(1 - \rho^2)\text{var}(\Theta),$$

to arrive at the same answer.

Solution to Problem 8.17. We have

$$\text{cov}(\Theta, X) = \mathbf{E}[\Theta^{3/2} W] - \mathbf{E}[\Theta]\,\mathbf{E}[X] = \mathbf{E}[\Theta^{/2}]\,\mathbf{E}[W] - \mathbf{E}[\Theta]\,\mathbf{E}[\Theta] = 0,$$

so the linear LMS estimator of Θ is simply $\hat{\Theta} = \mu$, and does not make use of the available observation.

Let us now consider the transformed observation $Y = X^2 = \Theta W^2$, and linear estimators of the form $\hat{\Theta} = aY + b$. We have

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[\Theta W^2] = \mathbf{E}[\Theta]\,\mathbf{E}[W^2] = \mu, \\ \mathbf{E}[\Theta Y] &= \mathbf{E}[\Theta^2 W^2] = \mathbf{E}[\Theta^2]\,\mathbf{E}[W^2] = \sigma^2 + \mu^2, \\ \text{cov}(\Theta, Y) &= \mathbf{E}[\Theta Y] - \mathbf{E}[\Theta]\,\mathbf{E}[Y] = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2, \\ \text{var}(Y) &= \mathbf{E}[\Theta^2 W^4] - \big(\mathbf{E}[Y]\big)^2 = (\sigma^2 + \mu^2)\mathbf{E}[W^4] - \mu^2. \end{aligned}$$

Thus, the linear LMS estimator of Θ based on Y is of the form

$$\hat{\Theta} = \mu + \frac{\sigma^2}{(\sigma^2 + \mu^2)\mathbf{E}[W^4] - \mu^2} \big(Y - \sigma^2 \big),$$

and makes effective use of the observation: the estimate of Θ, the conditional variance of X becomes large whenever a large value of X^2 is observed.

Solution to Problem 8.18. (a) The conditional CDF of X is given by

$$F_{X|\Theta}(x\,|\,\theta) = \mathbf{P}(X \le x\,|\,\Theta=\theta) = \mathbf{P}(\Theta\cos W \le x\,|\,\Theta=\theta) = \mathbf{P}\left(\cos W \le \frac{x}{\theta}\right).$$

We note that the cosine function is one-to-one and decreasing over the interval $[0, \pi/2]$, so for $0 \le x \le \theta$,

$$F_{X|\Theta}(x\,|\,\theta) = \mathbf{P}\left(W \ge \cos^{-1}\frac{x}{\theta}\right) = 1 - \frac{2}{\pi}\cos^{-1}\frac{x}{\theta}.$$

Differentiation yields

$$f_{X|\Theta}(x\,|\,\theta) = \frac{2}{\pi\sqrt{\theta^2 - x^2}}, \qquad 0 \le x \le \theta.$$

We have

$$f_{\Theta,X}(\theta, x) = f_\Theta(\theta) f_{X|\Theta}(x\,|\,\theta) = \frac{2}{\pi l\sqrt{\theta^2 - x^2}}, \qquad 0 \le \theta \le l, \quad 0 \le x \le \theta.$$

Thus the joint PDF is nonzero over the triangular region $\{(\theta, x)\,|\,0 \le \theta \le l, 0 \le x \le \theta\}$ or equivalently $\{(\theta, x)\,|\,0 \le x \le l, x \le \theta \le l\}$. To obtain $f_X(x)$, we integrate the joint PDF over θ:

$$\begin{aligned}
f_X(x) &= \frac{2}{\pi l}\int_x^l \frac{1}{\sqrt{\theta^2 - x^2}}\,d\theta \\
&= \log\left(\theta + \sqrt{\theta^2 - x^2}\right)\Big|_x^l \\
&= \frac{2}{\pi l}\log\left(\frac{l + \sqrt{l^2 - x^2}}{x}\right), \qquad 0 \le x \le l,
\end{aligned}$$

where we have used the integration formula in the hint.

We have

$$f_{\Theta|X}(\theta\,|\,x) = \frac{f_{\Theta,X}(\theta, x)}{f_X(x)} = \frac{1}{\log\left(\frac{l + \sqrt{l^2 - x^2}}{x}\right)\sqrt{\theta^2 - x^2}}, \qquad x \le \theta \le l.$$

Thus, the LMS estimate is given by

$$\begin{aligned}
\mathbf{E}[\Theta\,|\,X = x] &= \int_{-\infty}^{\infty} \theta f_{\Theta|X}(\theta\,|\,x)\,dx \\
&= \frac{1}{\log\left(\frac{l + \sqrt{l^2 - x^2}}{x}\right)}\int_x^l \frac{\theta}{\sqrt{\theta^2 - x^2}}\,d\theta \\
&= \frac{\sqrt{\theta^2 - x^2}\Big|_x^l}{\log\left(\frac{l + \sqrt{l^2 - x^2}}{x}\right)} \\
&= \frac{\sqrt{l^2 - x^2}}{\log\left(\frac{l + \sqrt{l^2 - x^2}}{x}\right)}, \qquad 0 \le x \le l.
\end{aligned}$$

It is worth noting that $\lim_{x\to 0} \mathbf{E}[\Theta \mid X = x] = 0$ and that $\lim_{x\to l} \mathbf{E}[\Theta \mid X = x] = l$, as one would expect.

(b) The linear LMS estimator is

$$\hat{\Theta} = \mathbf{E}[\Theta] + \frac{\mathrm{cov}(\Theta, X)}{\sigma_X^2}(X - \mathbf{E}[X]).$$

Since Θ is uniformly distributed between 0 and l, it follows that $\mathbf{E}[\Theta] = l/2$. We obtain $\mathbf{E}[X]$ and $\mathbf{E}[X^2]$, using the fact that Θ is independent from W, and therefore also independent from $\cos W$ and $\cos^2 W$. We have

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}[\Theta \cos W] = \mathbf{E}[\Theta]\, \mathbf{E}[\cos W] \\
&= \mathbf{E}[\Theta] \cdot \frac{2}{\pi} \int_0^{\pi/2} \cos w \, dw = \frac{l}{2} \cdot \frac{2}{\pi} \sin w \Big|_0^{\pi/2} = \frac{2}{\pi},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}[X^2] &= \mathbf{E}[\Theta^2 \cos^2 W] = \mathbf{E}\big[\Theta^2 \mathbf{E}[\cos^2 W]\big] = \mathbf{E}[\Theta^2]\, \mathbf{E}[\cos^2 W] \\
&= \frac{1}{l} \int_0^l \theta^2 \, d\theta \cdot \frac{2}{\pi} \int_0^{\pi/2} \cos^2 w \, dw = \frac{l^3}{3l} \cdot \frac{1}{\pi} \int_0^{\pi/2} (1 + \cos 2w) \, dw = \frac{l^2}{3\pi} \cdot \frac{\pi}{2} = \frac{l^2}{6}.
\end{aligned}$$

Thus,

$$\mathrm{var}(X) = \frac{l^2}{6} - \frac{l^2}{\pi^2} = \frac{l^2(\pi^2 - 6)}{6\pi^2}.$$

We also have

$$\mathbf{E}[\Theta X] = \mathbf{E}[\Theta^2 \cos W] = \mathbf{E}[\Theta^2]\, \mathbf{E}[\cos W] = \frac{l^2}{3} \cdot \frac{2}{\pi} = \frac{2l^2}{3\pi}.$$

Hence,

$$\mathrm{cov}(\Theta, X) = \frac{2l^2}{3\pi} - \frac{l}{2} \cdot \frac{l}{\pi} = \frac{l^2}{\pi} \left(\frac{2}{3} - \frac{1}{2} \right) = \frac{l^2}{6\pi}.$$

Therefore,

$$\hat{\Theta} = \frac{l}{2} + \frac{l^2}{6\pi} \cdot \frac{6\pi^2}{l^2(\pi^2 - 6)} \left(X - \frac{l}{\pi} \right) = \frac{l}{2} + \frac{\pi}{\pi^2 - 6} \left(X - \frac{l}{\pi} \right).$$

The mean squared error is

$$\begin{aligned}
(1 - \rho^2)\sigma_\Theta^2 &= \sigma_\Theta^2 - \frac{\mathrm{cov}^2(\Theta, X)}{\sigma_X^2} \\
&= \frac{l^2}{12} - \frac{l^4}{36\pi^2} \cdot \frac{6\pi^2}{l^2(\pi^2 - 6)} \\
&= \frac{l^2}{12} \left(1 - \frac{2}{\pi^2 - 6} \right) \\
&= \frac{l^2}{12} \cdot \frac{\pi^2 - 8}{\pi^2 - 6}.
\end{aligned}$$

Solution to Problem 8.19. (a) Let X be the number of detected photons. From Bayes' rule, we have

$$\begin{aligned}\mathbf{P}(\text{transmitter is on}\,|\,X=k) &= \frac{\mathbf{P}(X=k\,|\,\text{transmitter is on})\cdot\mathbf{P}(\text{transmitter is on})}{\mathbf{P}(X=k)}\\ &= \frac{\mathbf{P}(\Theta+N=k)\cdot p}{\mathbf{P}(N=k)\cdot(1-p)+\mathbf{P}(\Theta+N=k)\cdot p}.\end{aligned}$$

The PMFs of Θ and $\Theta+N$ are

$$p_\Theta(\theta)=\frac{\lambda^\theta e^{-\lambda}}{\theta!},\qquad p_{\Theta+N}(n)=\frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}.$$

Thus, using part (a) we obtain

$$\begin{aligned}\mathbf{P}(\text{transmitter is on}\,|\,X=k) &= \frac{p\cdot\dfrac{(\lambda+\mu)^k e^{-(\lambda+\mu)}}{k!}}{p\cdot\dfrac{(\lambda+\mu)^k e^{-(\lambda+\mu)}}{k!}+(1-p)\cdot\dfrac{\mu^k e^{-\mu}}{k!}}\\ &= \frac{p(\lambda+\mu)^k e^{-\lambda}}{p(\lambda+\mu)^k e^{-\lambda}+(1-p)\mu^k}.\end{aligned}$$

(b) We calculate $\mathbf{P}(\text{transmitter is on}\,|\,X=k)$ and decide that the transmitter is on if and only if this probability is at least 1/2; equivalently, if and only if

$$p(\lambda+\mu)^k e^{-\lambda}\geq(1-p)\mu^k.$$

(c) Let S be the number of transmitted photons, so that S is equal to Θ with probability p, and is equal to 0 with probability $1-p$. The linear LMS estimator is

$$\hat{S}=\mathbf{E}[S]+\frac{\operatorname{cov}(S,X)}{\sigma_X^2}\big(X-\mathbf{E}[X]\big).$$

We calculate all the terms in the preceding expression.

Since Θ and N are independent, S and N are independent as well. We have

$$\begin{aligned}&\mathbf{E}[S]=p\mathbf{E}[\Theta]=p\lambda,\\ &\mathbf{E}[S^2]=p\mathbf{E}[\Theta^2]=p(\lambda^2+\lambda),\\ &\sigma_S^2=\mathbf{E}[S^2]-(\mathbf{E}[S])^2=p(\lambda^2+\lambda)-(p\lambda)^2.\end{aligned}$$

It follows that

$$\mathbf{E}[X]=\mathbf{E}[S]+\mathbf{E}[N]=(p\lambda+\mu),$$

and

$$\begin{aligned}\sigma_X^2 &= \sigma_S^2+\sigma_N^2\\ &= p\big(\lambda^2+\lambda\big)-(p\lambda)^2+\mu\\ &= (p\lambda+\mu)+p(1-p)\lambda^2.\end{aligned}$$

Finally, we calculate $\mathrm{cov}(S,X)$:

$$
\begin{aligned}
\mathrm{cov}(S,X) &= \mathbf{E}\big[(S-\mathbf{E}[S])(X-\mathbf{E}[X])\big] \\
&= \mathbf{E}\big[(S-\mathbf{E}[S])(S-\mathbf{E}[S]+N-\mathbf{E}[N])\big] \\
&= \mathbf{E}\big[(S-\mathbf{E}[S])(S-\mathbf{E}[S])\big] + \mathbf{E}\big[(S-\mathbf{E}[S])(N-\mathbf{E}[N])\big] \\
&= \sigma_S^2 + \mathbf{E}\big[(S-\mathbf{E}[S])(N-\mathbf{E}[N])\big] \\
&= \sigma_S^2 \\
&= p(\lambda^2+\lambda) - (p\lambda)^2,
\end{aligned}
$$

where we have used the fact that $S-\mathbf{E}[S]$ and $N-\mathbf{E}[N]$ are independent, and that $\mathbf{E}\big[S-\mathbf{E}[S]\big] = \mathbf{E}\big[N-\mathbf{E}[N]\big] = 0$.

CHAPTER 9

Solution to Problem 9.1. Let X_i denote the random homework time for the ith week, $i = 1, \ldots, 5$. We have the observation vector $X = x$, where $x = (10, 14, 18, 8, 20)$. In view of the independence of the X_i, for $\theta \in [0, 1]$, the likelihood function is

$$\begin{aligned} f_X(x;\theta) &= f_{X_1}(x_1;\theta)\cdots f_{X_5}(x_5;\theta) \\ &= \theta e^{-x_1\theta}\cdots\theta e^{-x_5\theta} \\ &= \theta^5 e^{-(x_1+\cdots+x_5)\theta} \\ &= \theta^5 e^{-(10+14+18+8+20)\theta} \\ &= \theta^5 e^{-71\theta}. \end{aligned}$$

To derive the ML estimate, we set to 0 the derivative of $f_X(x;\theta)$ with respect to θ, obtaining

$$\frac{d}{d\theta}\big(\theta^5 e^{-71\theta}\big) = 5\theta^4 e^{-71\theta} - 71\,\theta^5 e^{-71\theta} = (5 - 71\,\theta)\theta^4 e^{-71\theta} = 0.$$

Therefore,

$$\hat\theta = \frac{5}{71} = \frac{5}{x_1 + \cdots + x_5}.$$

Solution to Problem 9.2. (a) Let the random variable N be the number of tosses until the kth head. The likelihood function is the Pascal PMF of order k:

$$p_N(n;\theta) = \binom{n-1}{k-1}\theta^k(1-\theta)^{n-k}, \qquad n = k, k+1, \ldots$$

We maximize the likelihood by setting its derivative with respect to θ to zero:

$$0 = k\binom{n-1}{k-1}(1-\theta)^{n-k}\theta^{k-1} - (n-k)\binom{n-1}{k-1}(1-\theta)^{n-k-1}\theta^k,$$

which yields the ML estimator

$$\hat\Theta_1 = \frac{k}{N}.$$

Note that $\hat\Theta_1$ is just the fraction of heads observed in N tosses.

(b) In this case, n is a fixed integer and K is a random variable. The PMF of K is binomial:

$$p_K(k;\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}, \qquad k = 0, 1, 2, \ldots, n.$$

For given n and k, this is a constant multiple of the PMF in part (a), so the same calculation yields the estimator

$$\hat{\Theta}_2 = \frac{K}{n}.$$

We observe that the ML estimator is again the fraction of heads in the observed trials.

Note that although parts (a) and (b) involve different experiments and different random variables, the ML estimates obtained are similar. However, it can be shown that $\hat{\Theta}_2$ is unbiased [since $\mathbf{E}[\hat{\Theta}_2] = \mathbf{E}[K]/n = \theta \cdot n/n = \theta$], whereas $\hat{\Theta}_1$ is not [since $\mathbf{E}[1/N] = 1/\mathbf{E}[N]$].

Solution to Problem 9.3. (a) Let s be the sum of all the ball numbers. Then for all i,

$$\mathbf{E}[X_i] = \frac{s}{k}, \qquad \mathbf{E}[Y_i] = \frac{s}{\overline{k}}.$$

We have

$$\mathbf{E}[\hat{S}] = \mathbf{E}\left[\frac{k}{n}\sum_{i=1}^{n} X_i\right] = \frac{k}{n}\sum_{i=1}^{n}\mathbf{E}\left[X_i\right] = \frac{k}{n}\sum_{i=1}^{n}\frac{s}{k} = s,$$

so $\hat{S}$ is an unbiased estimator of s. Similarly, $\mathbf{E}[\tilde{S}] = s$. Finally, let

$$L = \frac{\overline{S}}{\overline{k}} = \frac{1}{\overline{N}}\sum_{i=1}^{n} X_i = \frac{1}{\overline{N}}\sum_{j=1}^{\overline{N}} Y_j.$$

We have

$$\begin{aligned}
\mathbf{E}[L] &= \sum_{\overline{n}=1}^{n}\mathbf{E}[L \mid \overline{N} = \overline{n}]p_{\overline{N}}(\overline{n}) \\
&= \sum_{\overline{n}=1}^{n}\mathbf{E}\left[\frac{1}{\overline{n}}\sum_{i=1}^{\overline{n}} Y_i \;\middle|\; \overline{N} = \overline{n}\right] p_{\overline{N}}(\overline{n}) \\
&= \sum_{\overline{n}=1}^{n}\mathbf{E}[Y_1]p_{\overline{N}}(\overline{n}) \\
&= \mathbf{E}[Y_1] \\
&= \frac{s}{\overline{k}},
\end{aligned}$$

so that $\overline{S} = \overline{k}\,\mathbf{E}[L] = s$, and $\overline{S}$ is an unbiased estimator of s.

(b) We have

$$\mathrm{var}(\hat{S}) = \frac{k^2}{n}\mathrm{var}(X_1), \qquad \mathrm{var}(\tilde{S}) = \frac{\overline{k}^2}{m}\mathrm{var}(Y_1).$$

Thus,

$$\begin{aligned}
\mathrm{var}(\hat{S}) &= \frac{k^2}{n}\mathrm{var}(X_1) \\
&= \frac{k^2}{n}\left(p\mathbf{E}[Y_1^2] - p^2(\mathbf{E}[Y_1])^2\right) \\
&= \frac{\overline{k}^2}{n}\left(\frac{1}{p}\mathbf{E}[Y_1^2] - \left(\mathbf{E}[Y_1]\right)^2\right) \\
&= \frac{\overline{k}^2}{n}\left(\mathrm{var}(Y_1) + \frac{1-p}{p}\mathbf{E}[Y_1^2]\right) \\
&= \frac{\overline{k}^2}{n}\mathrm{var}(Y_1)\left(1 + \frac{r(1-p)}{p}\right) \\
&= \mathrm{var}(\tilde{S}) \cdot \frac{m}{n} \cdot \frac{p + r(1-p)}{p}.
\end{aligned}$$

It follows that when $m = n$,

$$\frac{\mathrm{var}(\tilde{S})}{\mathrm{var}(\hat{S})} = \frac{p}{p + r(1-p)}.$$

Furthermore, in order for $\mathrm{var}(\hat{S}) \approx \mathrm{var}(\tilde{S})$, we must have

$$m \approx \frac{np}{p + r(1-p)}.$$

(c) We have

$$\begin{aligned}
\mathrm{var}(\overline{S}) &= \mathrm{var}\left(\frac{\overline{k}}{\overline{N}}\sum_{i=1}^{n} X_i\right) \\
&= \overline{k}^2\mathrm{var}\left(\frac{1}{\overline{N}}\sum_{i=1}^{n} X_i\right) \\
&= \overline{k}^2\mathrm{var}\left(\frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}} Y_i\right) \\
&= \overline{k}^2\left(\mathbf{E}[L^2] - \mathbf{E}[L]^2\right),
\end{aligned}$$

where L was defined in part (a):

$$L = \frac{1}{\overline{N}}\sum_{i=1}^{n} X_i = \frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}} Y_i.$$

We showed in part (a) that

$$\mathbf{E}[L] = \mathbf{E}[Y_1],$$

and we will now evaluate $\mathbf{E}[L^2]$. We have

$$\begin{aligned}
\mathbf{E}[L^2] &= \sum_{\overline{n}=1}^{n} \mathbf{E}[L^2 \mid \overline{N} = \overline{n}] p_{\overline{N}}(\overline{n}) \\
&= \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}^2} \mathbf{E}\left[\left(\sum_{i=1}^{\overline{n}} Y_i \right)^2 \;\Big|\; \overline{N} = \overline{n} \right] p_{\overline{N}}(\overline{n}) \\
&= \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}^2} \left(\overline{n}\mathbf{E}[Y_1^2] + \overline{n}(\overline{n}-1)\mathbf{E}[Y_1]^2 \right) p_{\overline{N}}(\overline{n}) \\
&= \mathbf{E}[Y_1^2] \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} p_{\overline{N}}(\overline{n}) + \left(\mathbf{E}[Y_1]\right)^2 \sum_{\overline{n}=1}^{n} \frac{\overline{n}-1}{\overline{n}} p_{\overline{N}}(\overline{n}) \\
&= \left(\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2 \right) \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} p_{\overline{N}}(\overline{n}) + (\mathbf{E}[Y_1])^2,
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{var}(\overline{S}) &= \overline{k}^2 \left(\mathbf{E}[L^2] - \mathbf{E}[L]^2 \right) \\
&= \overline{k}^2 \left(\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2 \right) \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} p_{\overline{N}}(\overline{n}) \\
&= \mathbf{E}\left[\frac{1}{\overline{N}} \right] \overline{k}^2 \left(\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2 \right).
\end{aligned}$$

Thus, we have

$$\frac{\mathrm{var}(\overline{S})}{\mathrm{var}(\hat{S})} = \frac{\mathbf{E}\left[1/\overline{N}\right] \overline{k}^2 (\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2)}{(1/n)\overline{k}^2 \left(\frac{1}{p}\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2 \right)} = \frac{\mathbf{E}\left[n/\overline{N}\right] (\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2)}{\frac{1}{p}\mathbf{E}[Y_1^2] - (\mathbf{E}[Y_1])^2} = \frac{\mathbf{E}\left[n/\overline{N}\right] p}{p + r(1-p)}.$$

We will show that $\mathbf{E}\left[n/\overline{N}\right] p \approx 1$ for large n, so that

$$\frac{\mathrm{var}(\overline{S})}{\mathrm{var}(\hat{S})} \approx \frac{1}{p + r(1-p)}.$$

We show this by proving an upper and a lower bound to

$$\mathbf{E}\left[\frac{1}{\overline{N}} \right] = \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} p_{\overline{N}}(\overline{n}).$$

We have

$$\begin{aligned}
\sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} p_{\overline{N}}(\overline{n}) &= \sum_{\overline{n}=1}^{n} \left(\frac{1}{\overline{n}+1} + \frac{1}{\overline{n}(\overline{n}+1)} \right) p_{\overline{N}}(\overline{n}) \\
&= \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}+1} p_{\overline{N}}(\overline{n}) + \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}(\overline{n}+1)} p_{\overline{N}}(\overline{n}) \\
&= \frac{1}{(n+1)p} + \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}(\overline{n}+1)} p_{\overline{N}}(\overline{n}) \\
&= \frac{1}{(n+1)p} + \frac{1}{(n+1)p} \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} \binom{n+1}{\overline{n}+1} p^{k+1} (1-p)^{n-\overline{n}} \\
&\leq \frac{1}{(n+1)p} + \frac{1}{(n+1)p} \sum_{\overline{n}=2}^{n} \frac{3}{\overline{n}+2} \binom{n+1}{\overline{n}+1} p^{k+1} (1-p)^{n-\overline{n}} \\
&= \frac{1}{(n+1)p} + \frac{3}{(n+1)(n+2)p^2} \sum_{\overline{n}=2}^{n} \binom{n+2}{\overline{n}+2} p^{\overline{n}+2} (1-p)^{n-\overline{n}} \\
&\leq \frac{1}{(n+1)p} + \frac{3}{(n+1)(n+2)p^2}.
\end{aligned}$$

The first inequality comes from bounding $1/\overline{n}$ by $3/(\overline{n}+1)$ when $\overline{n}+1$ is greater than 1, and ignoring the term $\overline{n}=1$. The second inequality holds because

$$\sum_{\overline{n}=2}^{n} \binom{n+2}{\overline{n}+2} p^{\overline{n}+2} (1-p)^{n-\overline{n}} < \sum_{\overline{n}=-2}^{n} \binom{n+2}{\overline{n}+2} p^{\overline{n}+2} (1-p)^{n-\overline{n}} = 1.$$

The preceding calculation also shows that

$$\frac{1}{(n+1)p} \leq \sum_{\overline{n}=1}^{n} \frac{1}{\overline{n}} p_{\overline{N}}(\overline{n}).$$

It follows that for large n, we have $\mathbf{E}\left[1/\overline{N}\right] \approx 1/\big((n+1)p\big)$ or $\mathbf{E}\left[n/\overline{N}\right] p \approx 1$.

Solution to Problem 9.4. (a) Figure 9.1 plots a mixture of two normal distributions. Denoting $\theta = (p_1, \mu_1, \sigma_1, \ldots, p_m, \mu_m, \sigma_m)$, the PDF of each X_i is

$$f_{X_i}(x_i;\theta) = \sum_{j=1}^{m} p_j \cdot \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ \frac{-(x_i-\mu_j)^2}{2\sigma_j^2} \right\}.$$

Using the independence assumption, the likelihood function is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta) = \prod_{i=1}^{n} \left(\sum_{j=1}^{m} p_j \cdot \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ \frac{-(x_i-\mu_j)^2}{2\sigma_j^2} \right\} \right),$$

and the log-likelihood function is

$$\log f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} p_j \cdot \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ \frac{-(x_i-\mu_j)^2}{2\sigma_j^2} \right\} \right).$$

(b) The likelihood function is

$$p_1 \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ \frac{-(x-\mu_1)^2}{2\sigma_1^2} \right\} + (1-p_1) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ \frac{-(x-\mu_2)^2}{2\sigma_2^2} \right\},$$

and is linear in p_1. The ML estimate of p_1 is

$$\hat{p}_1 = \begin{cases} 1, & \text{if } \dfrac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ \dfrac{-(x-\mu_1)^2}{2\sigma_1^2} \right\} > \dfrac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ \dfrac{-(x-\mu_2)^2}{2\sigma_2^2} \right\}, \\ 0, & \text{otherwise,} \end{cases}$$

and the ML estimate of p_2 is $\hat{p}_2 = 1 - \hat{p}_1$.

(c) The likelihood function is the sum of two terms [cf. the solution to part (b)], the first involving μ_1, the second involving μ_2. Thus, we can maximize each term separately and find that the ML estimates are $\hat{\mu}_1 = \hat{\mu}_2 = x$.

(d) Fix $p_1, \ldots, p_m$ to some positive values. Fix $\mu_2, \ldots, \mu_m$ and $\sigma_2^2, \ldots, \sigma_m^2$ to some arbitrary (respectively, positive) values. If $\mu_1 = x_1$ and σ_1^2 tends to zero, the likelihood $f_{X_1}(x_1;\theta)$ tends to infinity, and the likelihoods $f_{X_i}(x_i;\theta)$ of the remaining points ($i > 1$) are bounded below by a positive number. Therefore, the overall likelihood tends to infinity.

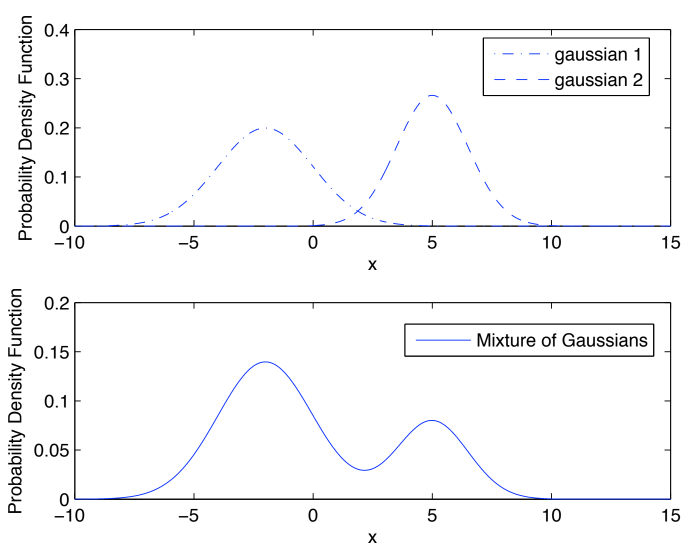

Figure 9.1: The mixture of two normal distributions with $p_1 = 0.7$ and $p_2 = 0.3$ in Problem 9.4.

Solution to Problem 9.5. (a) The PDF of the location X_i of the ith event is

$$f_{X_i}(x_i;\theta) = \begin{cases} c(\theta)\theta e^{-\theta x_i}, & \text{if } m_1 \le x_i \le m_2, \\ 0, & \text{otherwise,} \end{cases}$$

where $c(\theta)$ is a normalization factor,

$$c(\theta) = \frac{1}{\displaystyle\int_{m_1}^{m_2} \theta e^{-\theta x}\,dx} = \frac{1}{e^{-m_1\theta} - e^{-m_2\theta}}.$$

The likelihood function is

$$f_X(x;\theta) = \prod_{i=1}^n f_{X_i}(x_i;\theta) = \prod_{i=1}^n \frac{1}{e^{-m_1\theta} - e^{-m_2\theta}}\theta e^{\theta x_i} = \left(\frac{1}{e^{-m_1\theta} - e^{-m_2\theta}}\right)^n \theta^n \prod_{i=1}^n e^{\theta x_i},$$

and the corresponding log-likelihood function is

$$\log f_X(x;\theta) = \sum_{i=1}^n \log f_{X_i}(x_i;\theta) = -n\log\left(e^{-m_1\theta} - e^{-m_2\theta}\right) + n\log\theta + \theta\sum_{i=1}^n x_i.$$

(b) We plot the likelihood and log-likelihood functions in Fig. 9.2. The ML estimate is approximately 0.26.

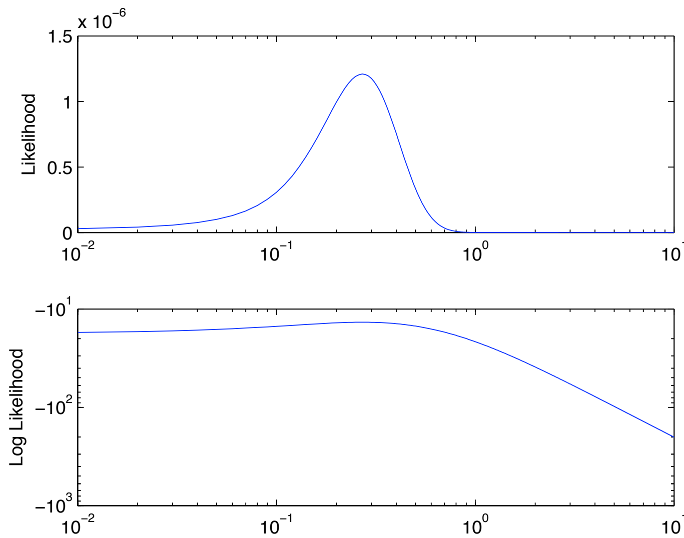

Figure 9.2: Plots of the likelihood and log-likelihood functions in Problem 9.5.

Solution to Problem 9.6. (a) The likelihood function for a single observation x is

$$\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left\{\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right\} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right\},$$

so the likelihood function is

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} \left(\sum_{j=1}^{2} \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left\{\frac{-(x_i-\mu_j)^2}{2\sigma_j^2}\right\} \right).$$

(b) We plot the likelihood as a function of σ_2 and μ_2 in Figure 9.3. The ML estimates are found (by a fine grid/brute force optimization) to be $\hat{\sigma}_2 \approx 7.2$ and $\hat{\mu}_2 \approx 173$.

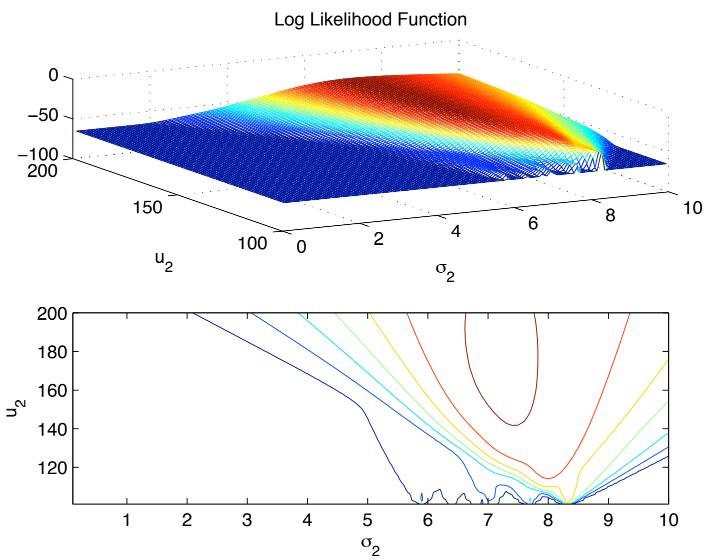

Figure 9.3 Plot of the log-likelihood and its contours as a function of σ_2 and μ_2.

(c) We plot the likelihood as a function of μ_1 and μ_2 in Fig. 9.4. The ML estimates are found (by a fine grid/brute force optimization) to be $\hat{\mu}_1 \approx 174$ and $\hat{\mu}_2 \approx 156$.

(d) Let Θ denote the gender of the student, with $\Theta = 1$ for a female student and $\Theta = 0$ for a male student. Using Bayes' rule, we compare the posterior probabilities,

$$\mathbf{P}(\Theta = 1 \,|\, X = x) = \frac{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right\}}{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right\} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right\}},$$

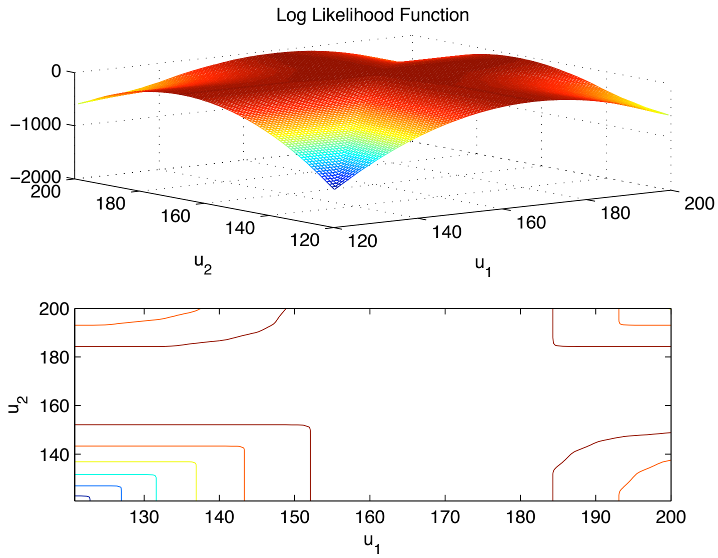

Figure 9.4 Plot of the log-likelihood and its contours as a function of μ_1 and μ_2.

and

$$\mathbf{P}(\Theta = 0 \,|\, X = x) = \frac{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right\}}{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right\} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right\}}.$$

The MAP rule involves a comparison of the two numerators. When $\sigma_1 = \sigma_2$, it reduces to a comparison of $|x - \mu_1|$ to $|x - \mu_2|$. Using the estimates in part (c), we will decide that the student is female if $x < 165$, and male otherwise.

Solution to Problem 9.7. The PMF of X_i is

$$p_{X_i}(x) = e^{-\theta}\frac{\theta^x}{x!}, \qquad x = 0, 1, \ldots.$$

The log-likelihood function is

$$\log p_X(x_1, \ldots, x_n; \theta) = \sum_{i=1}^n \log p_{X_i}(x_i; \theta) = -n\theta + \sum_{i=1}^n x_i \log\theta - \sum_{i=1}^n \log(x_i!),$$

and to maximize it, we set its derivative to 0. We obtain

$$0 = -n + \frac{1}{\theta}\sum_{i=1}^n x_i,$$

which yields the estimator

$$\hat{\Theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

This estimator is unbiased, since $\mathbf{E}[X_i] = \theta$, so that

$$\mathbf{E}[\hat{\Theta}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[X_i] = \theta.$$

It is also consistent, because $\hat{\Theta}_n$ converges to θ in probability, by the weak law of large numbers.

Solution to Problem 9.8. The PDF of X_i is

$$f_{X_i}(x) = \begin{cases} 1/\theta, & \text{if } 0 \le x_i \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is

$$f_X(x_1,\ldots,x_n;\theta) = f_{X_1}(x_1;\theta)\cdots f_{X_n}(x_n;\theta) = \begin{cases} 1/\theta^n, & \text{if } 0 \le \max_{i=1,\ldots,n} x_i \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

We maximize the likelihood function and obtain the ML estimator as

$$\hat{\Theta}_n = \max_{i=1,\ldots,n} X_i.$$

It can be seen that $\hat{\Theta}_n$ converges in probability to θ (the upper endpoint of the interval where X_i takes values); see Example 5.6. Therefore the estimator is consistent.

To check whether $\hat{\Theta}_n$ is unbiased, we calculate its CDF, then its PDF (by differentiation), and then $\mathbf{E}[\hat{\Theta}_n]$. We have, using the independence of the X_i,

$$F_{\hat{\Theta}_n}(x) = \begin{cases} 0, & \text{if } x < 0, \\ x^n/\theta^n, & \text{if } 0 \le x \le \theta, \\ 1, & \text{if } x > \theta, \end{cases}$$

so that

$$f_{\hat{\Theta}_n}(x) = \begin{cases} 0, & \text{if } x < 0, \\ nx^{n-1}/\theta^n, & \text{if } 0 \le x \le \theta, \\ 0, & \text{if } x > \theta. \end{cases}$$

Hence

$$\mathbf{E}[\hat{\Theta}_n] = \frac{n}{\theta^n}\int_0^\theta x\,x^{n-1}\,dx = \frac{n}{\theta^n}\left(\frac{x^{n+1}}{n+1}\right)\Bigg|_0^\theta = \frac{n}{\theta^n}\cdot\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta.$$

Thus $\hat{\Theta}_n$ is not unbiased, but it is asymptotically unbiased.

Some alternative estimators that are unbiased are a scaled version of the ML estimator

$$\hat{\Theta} = \frac{n+1}{n}\max_{i=1,\ldots,n} X_i,$$

or one that relies on the sample mean being an unbiased estimate of $\theta/2$:

$$\hat{\Theta} = \frac{2}{n}\sum_{i=1}^{n} X_i.$$

Solution to Problem 9.9. The PDF of X_i is

$$f_{X_i}(x_i) = \begin{cases} 1, & \text{if } \theta \le x_i \le \theta+1, \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is

$$\begin{aligned} f_X(x_1,\ldots,x_n;\theta) &= f_{X_1}(x_1;\theta)\cdots f_{X_n}(x_n;\theta) \\ &= \begin{cases} 1, & \text{if } \theta \le \min_{i=1,\ldots,n} x_i \le \max_{i=1,\ldots,n} x_i \le \theta+1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Any value in the feasible interval

$$\left[\max_{i=1,\ldots,n} X_i - 1, \min_{i=1,\ldots,n} X_i\right]$$

maximizes the likelihood function and is therefore a ML estimator.

Any choice of estimator within the above interval is consistent. The reason is that $\min_{i=1,\ldots,n} X_i$ converges in probability to θ, while $\max_{i=1,\ldots,n} X_i$ converges in probability to $\theta+1$ (cf. Example 5.6). Thus, both endpoints of the above interval converge to θ.

Let us consider the estimator that chooses the midpoint

$$\hat{\Theta}_n = \frac{1}{2}\left(\max_{i=1,\ldots,n} X_i + \min_{i=1,\ldots,n} X_i - 1\right)$$

of the interval of ML estimates. We claim that it is unbiased. This claim can be verified purely on the basis of symmetry considerations, but nevertheless we provide a detailed calculation. We first find the CDFs of $\max_{i=1,\ldots,n} X_i$ and $\min_{i=1,\ldots,n} X_i$, then their PDFs (by differentiation), and then $\mathbf{E}[\hat{\Theta}_n]$. The details are very similar to the ones for the preceding problem. We have by straightforward calculation,

$$f_{\min_i X_i}(x) = \begin{cases} n(\theta+1-x)^{n-1}, & \text{if } \theta \le x \le \theta+1, \\ 0, & \text{otherwise,} \end{cases}$$

$$f_{\max_i X_i}(x) = \begin{cases} n(x-\theta)^{n-1}, & \text{if } \theta \le x \le \theta+1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} \mathbf{E}\left[\min_{i=1,\ldots,n} X_i\right] &= n\int_{\theta}^{\theta+1} x(\theta+1-x)^{n-1}\,dx \\ &= -n\int_{\theta}^{\theta+1}(\theta+1-x)^n dx + (\theta+1)n\int_{\theta}^{\theta+1}(\theta+1-x)^{n-1}\,dx \\ &= -n\int_0^1 x^n\,dx + (\theta+1)n\int_0^1 x^{n-1}\,dx \\ &= -\frac{n}{n+1} + \theta + 1 \\ &= \theta + \frac{1}{n+1}. \end{aligned}$$

Similarly,

$$\mathbf{E}\left[\max_{i=1,\ldots,n} X_i\right] = \theta + \frac{n}{n+1},$$

and it follows that

$$\mathbf{E}[\hat{\Theta}_n] = \frac{1}{2}\mathbf{E}\left[\max_{i=1,\ldots,n} X_i + \min_{i=1,\ldots,n} X_i - 1\right] = \theta.$$

Solution to Problem 9.10. (a) To compute $c(\theta)$, we write

$$1 = \sum_{k=0}^{\infty} p_K(k;\theta) = \sum_{k=0}^{\infty} c(\theta)e^{-\theta k} = \frac{c(\theta)}{1-e^{-\theta}},$$

which yields $c(\theta) = 1 - e^{-\theta}$.

(b) The PMF of K is a shifted geometric distribution with parameter $p = 1 - e^{-\theta}$ (shifted by 1 to the left, so that it starts at $k = 0$). Therefore,

$$\mathbf{E}[K] = \frac{1}{p} - 1 = \frac{1}{1-e^{-\theta}} - 1 = \frac{e^{-\theta}}{1-e^{-\theta}} = \frac{1}{e^{\theta}-1},$$

and the variance is the same as for the geometric with parameter p,

$$\mathrm{var}(K) = \frac{1-p}{p^2} = \frac{e^{-\theta}}{(1-e^{-\theta})^2}.$$

(c) Let K_i be the number of photons emitted the ith time that the source is triggered. The joint PMF of $K = (K_1, \ldots, K_n)$ is

$$p_K(k_1,\ldots,k_n;\theta) = c(\theta)^n \prod_{i=1}^{n} e^{-\theta k_i} = c(\theta)^n e^{-\theta s_n},$$

where

$$s_n = \sum_{i=1}^{n} k_i.$$

The log-likelihood function is

$$\log p_K(k_1,\ldots,k_n;\theta) = n\log c(\theta) - \theta s_n = n\log\left(1-e^{-\theta}\right) - \theta s_n.$$

We maximize the log-likelihood by setting to 0 the derivative with respect to θ:

$$\frac{d}{d\theta}\log p_K(k_1,\ldots,k_n;\theta) = n\,\frac{e^{-\theta}}{1-e^{-\theta}} - s_n = 0,$$

or

$$e^{-\theta} = \frac{s_n/n}{1+s_n/n}.$$

Taking the logarithm of both sides gives the ML estimate of θ,

$$\hat{\theta}_n = \log\left(1 + \frac{n}{s_n}\right),$$

and the ML estimate of $\psi = 1/\theta$,

$$\hat{\psi}_n = \frac{1}{\hat{\theta}_n} = \frac{1}{\log\left(1 + \dfrac{n}{s_n}\right)}.$$

(d) We verify that $\hat{\Theta}_n$ and $\hat{\Psi}_n$ are consistent estimators of θ and ψ, respectively. Let $S_n = K_1 + \cdots + K_n$. By the strong law of large numbers, we have

$$\frac{S_n}{n} \to \mathbf{E}[K] = \frac{1}{e^\theta - 1},$$

with probability 1. Hence $1 + (n/S_n)$ converges to e^θ, so that

$$\hat{\Theta}_n = \log\left(1 + \frac{n}{S_n}\right) \to \theta,$$

and similarly,

$$\hat{\Psi}_n \to \frac{1}{\theta} = \psi.$$

Since convergence with probability one implies convergence in probability, we conclude that these two estimators are consistent.

Solution to Problem 9.16. (a) We consider a model of the form

$$y = \theta_0 + \theta_1 x,$$

where x is the temperature and y is the electricity consumption. Using the regression formulas, we obtain

$$\hat{\theta}_1 = \frac{\displaystyle\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\displaystyle\sum_{i=1}^n (x_i - \overline{x})^2} = 0.2242, \qquad \hat{\theta}_0 = \overline{y} - \hat{\theta}_1 \overline{x} = 2.1077,$$

where

$$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i = 81.4000, \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^n y_i = 20.3551.$$

The linear regression model is

$$y = 0.2242x + 2.1077.$$

Figure 9.5 plots the data points and the estimated linear relation.

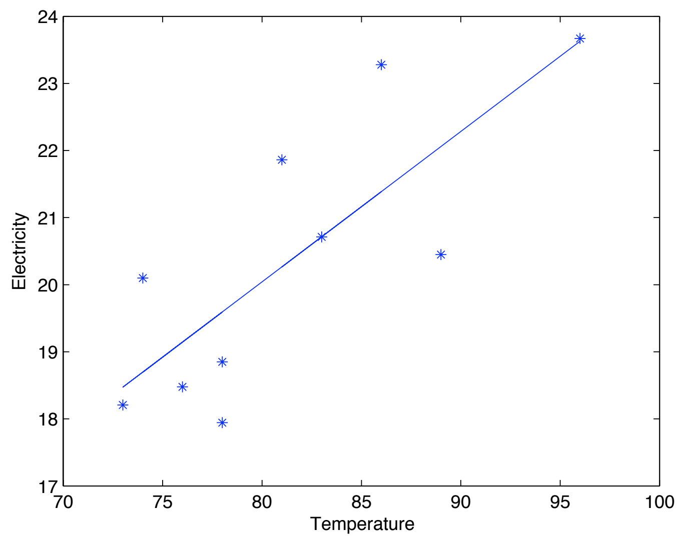

Figure 9.5: Linear regression model of the relationship between temperature and electricity in Problem 9.16.

(b) Using the estimated model with $x = 90$, we obtain

$$y = 0.2242x + 2.1077 = 22.2857.$$

Solution to Problem 9.17. (a) We have

$$\hat{\theta}_1 = \frac{\displaystyle\sum_{i=1}^{5}(x_i - \overline{x})(y_i - \overline{y})}{\displaystyle\sum_{i=1}^{5}(x_i - \overline{x})^2}, \qquad \hat{\theta}_0 = \overline{y} - \hat{\theta}_1\overline{x},$$

where

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = 4.9485, \qquad \overline{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = 134.3527.$$

The resulting ML estimates are

$$\hat{\theta}_1 = 40.6005, \qquad \hat{\theta}_0 = -66.5591.$$

(b) Using the same procedure as in part (a), we obtain

$$\hat{\theta}_1 = \frac{\displaystyle\sum_{i=1}^{5}(x_i^2 - \overline{x})(y_i - \overline{y})}{\displaystyle\sum_{i=1}^{5}(x_i^2 - \overline{x})^2}, \qquad \hat{\theta}_0 = \overline{y} - \hat{\theta}_1\overline{x},$$

where

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i^2 = 33.6560, \qquad \overline{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = 134.3527.$$

which for the given data yields

$$\hat{\theta}_1 = 4.0809, \qquad \hat{\theta}_0 = -2.9948.$$

Figure 9.6 shows the data points (x_i, y_i), $i = 1, \ldots, 5$, the estimated linear model

$$y = 40.6005x - 66.5591,$$

and the estimated quadratic model

$$y = 4.0809x^2 - 2.9948.$$

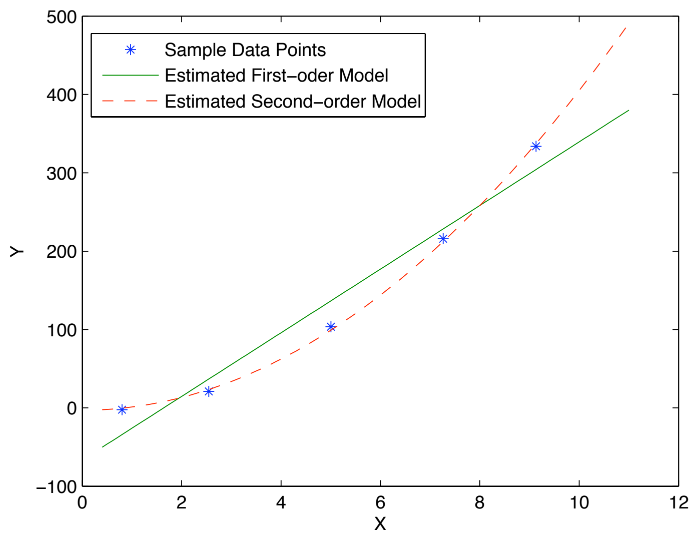

Figure 9.6: Regression plot for Problem 9.17.

(c) This is a Bayesian hypothesis testing problem, where the two hypotheses are:

$$H_1: \quad Y = 40.6005X - 66.5591,$$
$$H_2: \quad Y = 4.0809X^2 - 2.9948.$$

We evaluate the posterior probabilities of H_1 and H_2 given $Y_1, \ldots, Y_5$,

$$\mathbf{P}(H_1 \,|\, Y_1, \ldots, Y_5) = \frac{\mathbf{P}(H_1)\prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_1)}{\mathbf{P}(H_1)\prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_1) + \mathbf{P}(H_2)\prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_2)},$$

and

$$\mathbf{P}(H_2 \,|\, Y_1, \ldots, Y_5) = \frac{\mathbf{P}(H_2) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_2)}{\mathbf{P}(H_1) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_1) + \mathbf{P}(H_2) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_2)}.$$

We compare $\mathbf{P}(H_1) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_1)$ and $\mathbf{P}(H_2) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_2)$, by comparing their logarithms. Using σ^2 to denote the common noise variance in the two models, we have

$$\begin{aligned}
\log\left(\mathbf{P}(H_1) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_1)\right) &= \log\left(\frac{1}{2} \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - \theta_1 x_i - \theta_0)^2}{2\sigma^2}\right\}\right) \\
&= -\sum_{i=1}^{5} \frac{(y_i - \theta_1 x_i - \theta_0)^2}{2\sigma^2} + c \\
&= -\frac{3400.7}{2\sigma^2} + c,
\end{aligned}$$

and

$$\begin{aligned}
\log\left(\mathbf{P}(H_2) \prod_{i=1}^{5} f_{Y_i}(y_i \,|\, H_2)\right) &= \log\left(\frac{1}{2} \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - \theta_1 x_i^2 - \theta_0)^2}{2\sigma^2}\right\}\right) \\
&= -\sum_{i=1}^{5} \frac{(y_i - \theta_1 x_i^2 - \theta_0)^2}{2\sigma^2} + c \\
&= -\frac{52.9912}{2\sigma^2} + c,
\end{aligned}$$

where c is a constant that depends only on σ and n. Using the MAP rule, we select the quadratic model.

Note that when $\sigma_1 = \sigma_2$ and $\mathbf{P}(H_1) = \mathbf{P}(H_2)$, as above, comparing the posterior probabilities is equivalent to comparing the sum of the squared residuals and selecting the model for which the sum is smallest.

Solution to Problem 9.20. We have two hypotheses, the null hypothesis

$$H_0 : \mu_0 = 20, \;\; \sigma_0 = 4,$$

which we want to test against

$$H_1 : \mu_1 = 25, \;\; \sigma_1 = 5.$$

Let $\overline{X}$ be the random variable $\overline{X} = X_1 + X_2 + X_3$. We want the probability of false rejection to be

$$\mathbf{P}\left(\overline{X} > \gamma; H_0\right) = 0.05.$$

Since the mean and variance of $\overline{X}$ under H_0 are $3\mu_0$ and $3\sigma_0^2$, respectively, it follows that

$$\frac{\gamma - 3\mu_0}{\sqrt{3}\,\sigma_0} = \Phi^{-1}(0.95) = 1.644853,$$

and hence

$$\gamma = 1.644853 \cdot \sqrt{3 \cdot 4^2} + 60 = 71.396.$$

The corresponding probability of false acceptance of H_0 is

$$\begin{aligned}
\mathbf{P}\left(\overline{X} \le \gamma; H_1\right) &= \int_{-\infty}^{\gamma} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{3}} e^{\frac{-(x-\mu_1)^2}{2\cdot 3\cdot \sigma_1^2}}\,dx \\
&= \int_{-\infty}^{71.396} \frac{1}{\sqrt{2\pi}5\sqrt{3}} e^{\frac{-(x-75)^2}{2*3*5^2}}\,dx \\
&= \Phi\left(\frac{71.396-75}{5\sqrt{3}}\right) \\
&= \Phi\left(-0.41615\right) = 0.33864.
\end{aligned}$$

Solution to Problem 9.21. We have two hypotheses H_0 and H_1, under which the observation PDFs are

$$f_X(x; H_0) = \frac{1}{5\sqrt{2\pi}} e^{\frac{-(x-60)^2}{2\cdot 25}},$$

and

$$f_X(x; H_1) = \frac{1}{8\sqrt{2\pi}} e^{\frac{-(x-60)^2}{2\cdot 64}}.$$

(a) The probability of false rejection of H_0 is

$$\mathbf{P}(x \in R; H_0) = 2\left(1 - \Phi\left(\frac{\gamma}{5}\right)\right) = 0.1,$$

which yields that $\gamma = 8.25$. The acceptance region of H_0 is $\{x \,|\, 51.75 < x < 68.25\}$, and the probability of false acceptance is

$$\begin{aligned}
\mathbf{P}\left(51.75 < x < 68.25; H_1\right) &= \int_{51.75}^{68.25} \frac{1}{8\sqrt{2\pi}} e^{\frac{-(x-60)^2}{2\cdot 8^2}}\,dx \\
&= 2\Phi\left(\frac{68.25-60}{8}\right) - 1 = 0.697.
\end{aligned}$$

Consider now the LRT. Let $L(x)$ be the likelihood ratio and ξ be the critical value. We have

$$L(x) = \frac{f_X(x; H_1)}{f_X(x; H_0)} = \frac{8}{5} e^{\frac{39}{3200}(x-60)^2},$$

and the rejection region is

$$\left\{ x \,|\, e^{\frac{39}{3200}(x-60)^2} > 5\xi/8 \right\}.$$

This is the same type of rejection region as $R = \left\{ x \,|\, |x-60| > \gamma \right\}$, with ξ and γ being in one-to-one correspondence. Therefore, for the same probability of false rejection, the rejection region of the LRT is also $R = \left\{ x \,|\, |x-60| > \gamma \right\}$.

(b) Let $\overline{X} = \frac{X_1+\cdots+X_n}{n}$. To determine γ, we set

$$\mathbf{P}(\overline{X} \notin R; H_0) = 2\left(1 - \Phi\left(\frac{\gamma\sqrt{n}}{5}\right)\right) = 0.1,$$

which yields

$$\gamma = \frac{\Phi^{-1}(0.95)}{\sqrt{n}}.$$

The acceptance region is

$$R = \left\{x \,|\, 60 - \frac{\Phi^{-1}(0.95)}{\sqrt{n}} < x < 60 + \frac{\Phi^{-1}(0.95)}{\sqrt{n}}\right\},$$

and the probability of false acceptance of H_0 is

$$\mathbf{P}\{\overline{X} \in R; H_1\} = 2\Phi\left(\frac{\Phi^{-1}(0.95)/\sqrt{n}}{8/\sqrt{n}}\right) - 1 = 2\Phi\left(\frac{\Phi^{-1}(0.95)}{8}\right) - 1 = 0.697.$$

We observe that, even if the probability of false rejection is held constant, the probability of false acceptance of H_0 does not decrease with n increasing. This suggests that the form of acceptance region we have chosen is inappropriate for discriminating between these two hypotheses.

(c) Consider now the LRT. Let $L(x)$ be the likelihood ratio and ξ be the critical value. We have

$$L(x) = \frac{f_X(x_1, \ldots, x_n; H_1)}{f_X(x_1, \ldots, x_n; H_0)} = \frac{8}{5} e^{\frac{39}{3200}\sum_{i=1}^n (x_i - 60)^2},$$

and the rejection region is

$$\left\{x \,|\, e^{\frac{39}{3200}\sum_{i=1}^n (x_i - 60)^2} > 5\xi/8\right\}.$$

Solution to Problem 9.22. (a) We want to find k_n satisfying

$$\mathbf{P}\left(X \geq k_n; H_0\right) = \sum_{k=k_n}^{n} \binom{n}{k} \frac{1}{2}^n \leq 0.05.$$

Assuming that n is large enough, we use the normal approximation and obtain

$$\mathbf{P}\left(X \geq k_n; H_0\right) \approx 1 - \Phi\left(\frac{k_n - \frac{1}{2} - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right),$$

so we have

$$\frac{k_{n-}\frac{1}{2} - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \Phi^{-1}\left(0.95\right) = 1.644853,$$

and

$$k_n = \frac{1}{2}n + \frac{1}{2} + 1.644853 \cdot \frac{1}{2}\sqrt{n} = \frac{1}{2}n + 0.822427\sqrt{n} + \frac{1}{2}.$$

(b) The probability of making a correct decision given H_1 should be greater than 0.95, i.e.,

$$\mathbf{P}\left(X \geq k_n; H_1\right) = \sum_{X=k_n}^{n} \binom{n}{k} \frac{3}{5}^k \frac{2}{5}^{n-k} \geq 0.95,$$

which can be approximated by

$$\mathbf{P}\left(X \geq k_n; H_1\right) \approx 1 - \Phi\left(\frac{k_n - \frac{1}{2} - \frac{3}{5}n}{\sqrt{\frac{3}{5}\frac{2}{5}n}}\right) \geq 0.95.$$

Solving the above inequality we obtain

$$n \geq \left(10\left(0.82243 + \frac{\sqrt{6}}{5}1.644853\right)\right)^2 = 265.12$$

Therefore, $\hat{n} = 266$ is the smallest integer that satisfies the requirements on both false rejection and false acceptance probabilities.

(c) The likelihood ratio when $X = k$ is of the form

$$L(k) = \frac{\binom{n}{k}0.6^k(1-0.6)^{n-k}}{\binom{n}{k}0.5^k(1-0.5)^{n-k}} = 0.8^n 1.5^k.$$

Since $L(k)$ monotonically increases, the LRT rule would be to reject H_0 if $X > \gamma$, where γ is a positive integer. We need to guarantee that the false rejection probability is 0.05, i.e.,

$$\mathbf{P}\left(X \geq \gamma; H_0\right) = \sum_{i=\gamma+1}^{n} \binom{n}{i} 0.5^k(1-0.5)^{n-k} \approx 1 - \Phi\left(\frac{\gamma - \frac{1}{2} - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right) = 0.05,$$

which gives $\gamma \approx 147$. Then the false acceptance probability is calculated as

$$\mathbf{P}\left(X < \gamma; H_1\right) = \sum_{i=1}^{\gamma} \binom{n}{i} 0.6^k(1-0.6)^{n-k} \approx \Phi\left(\frac{\gamma - \frac{1}{2} - \frac{3}{5}n}{\sqrt{\frac{3}{5}\frac{2}{5}n}}\right) \approx 0.05.$$

Solution to Problem 9.23. Let H_0 and H_1 be the hypotheses corresponding to λ_0 and λ_1, respectively. Let X be the number of calls received on the given day. We have

$$p_X(k; H_0) = e^{-\lambda_0}\frac{\lambda_0^k}{k!}, \qquad p_X(k; H_1) = e^{-\lambda_1}\frac{\lambda_1^k}{k!}.$$

The likelihood ratio is

$$L(k) = \frac{p_X(k; H_1)}{p_X(k; H_0)} = e^{\lambda_1 - \lambda_0}\left(\frac{\lambda_1}{\lambda_0}\right)^k.$$

The rejection region is of the form

$$R = \Big\{ k \,|\, L(k) > \xi \Big\},$$

or by taking logarithms,

$$R = \Big\{ k \,|\, \log L(k) > \log \xi \Big\} = \Big\{ k \,|\, \lambda_0 - \lambda_1 + k\Big(\log \lambda_1 - \log \lambda_0 \Big) > \log \xi \Big\}.$$

Assuming $\lambda_1 > \lambda_0$, we have

$$R = \{ k \,|\, k > \gamma \},$$

where

$$\gamma = \frac{\lambda_0 - \lambda_1 + \log \xi}{\log \lambda_1 - \log \lambda_0}.$$

To determine the value of γ for a probability of false rejection equal to α, we must have

$$\alpha = \mathbf{P}(k > \gamma; H_0) = 1 - F_X(\gamma; H_0),$$

where $F_X(\,\cdot\,; H_0)$ is the CDF of the Poisson with parameter λ_0.

Solution to Problem 9.24. Let H_0 and H_1 be the hypotheses corresponding to λ_0 and λ_1, respectively. Let $X = (X_1, \ldots, X_n)$ be the observation vector. We have, using the independence of $X_1, \ldots, X_n$,

$$f_X(x_1, \ldots, x_n; H_0) = \lambda_0^n e^{-\lambda_0 (x_1 + \cdots + x_n)}, \qquad f_X(x_1, \ldots, x_n; H_1) = \lambda_1^n e^{-\lambda_1 (x_1 + \cdots + x_n)},$$

for $x_1, \ldots, x_n \geq 0$. The likelihood ratio is

$$L(x) = \frac{f_X(x_1, \ldots, x_n; H_1)}{f_X(x_1, \ldots, x_n; H_0)} = \left(\frac{\lambda_1}{\lambda_0} \right)^n e^{-(\lambda_1 - \lambda_0)(x_1 + \cdots + x_n)}.$$

The rejection region is of the form

$$R = \Big\{ x \,|\, L(x) > \xi \Big\},$$

or by taking logarithms,

$$R = \Big\{ x \,|\, \log L(x) > \log \xi \Big\} = \Big\{ x \,|\, n\Big(\log \lambda_1 - \log \lambda_0 \Big) + (\lambda_0 - \lambda_1)(x_1 + \cdots + x_n) > \log \xi \Big\}.$$

Assuming $\lambda_0 > \lambda_1$, we have

$$R = \{ x \,|\, x_1 + \cdots + x_n > \gamma \},$$

where

$$\gamma = \frac{n(\log \lambda_0 - \log \lambda_1) + \log \xi}{\lambda_0 - \lambda_1}.$$

To determine the value of γ for a probability of false rejection equal to α, we must have

$$\alpha = \mathbf{P}(x_1 + \cdots + x_n > \gamma; H_0) = 1 - F_Y(\gamma; H_0),$$

where $F_Y(\,\cdot\,; H_0)$ is the CDF of $Y = X_1 + \cdots + X_n$, which is an nth order Erlang random variable with parameter λ_0.

Solution to Problem 9.25. (a) Let $\overline{X}$ denote the sample mean for $n = 10$. In order to accept $\mu = 5$, we must have

$$\frac{|\overline{x} - 5|}{1/\sqrt{n}} \le 1.96,$$

or equivalently,

$$\overline{x} \in [5 - 1.96/\sqrt{n}, 5 + 1.96/\sqrt{n}].$$

(b) For $n = 10$, $\overline{X}$ is normal with mean μ and variance $1/10$. The probability of falsely accepting $\mu = 5$ when $\mu = 4$ becomes

$$\mathbf{P}(5-1.96/\sqrt{10} \le \overline{X} \le 5+1.96/\sqrt{10}; \mu = 4) = \Phi(\sqrt{10}+1.96) - \Phi(\sqrt{10}-1.96) \approx 0.114.$$

Solution to Problem 9.26. (a) We estimate the unknown mean and variance as

$$\hat{\mu} = \frac{x_1 + \cdots + x_n}{n} = \frac{8.47 + 10.91 + 10.87 + 9.46 + 10.40}{5} \approx 10.02,$$

and

$$\hat{\sigma}^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \approx 1.09.$$

(b) Using the fact $t(4, 0.05) = 2.132$ (from the t-distribution tables with 4 degrees of freedom), we find that

$$\frac{|\hat{\mu} - \mu|}{\sigma/\sqrt{n}} = \frac{|10.02 - 9|}{1.09/\sqrt{5}} = 2.1859 > 2.132,$$

so we reject the hypothesis $\mu = 9$.

Solution to Problem 9.27. Denoting by $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ the samples of life lengths on the first and second island respectively, we have for each $i = 1, \ldots, n$,

$$x_i \sim N\left(\mu_X, \sigma_X^2\right),$$

$$y_i \sim N\left(\mu_Y, \sigma_Y^2\right).$$

Let $\overline{X}$ and $\overline{Y}$ be the sample means. Using independence between each sample, we have

$$\overline{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right),$$

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right),$$

and using independence between $\overline{X}$ and $\overline{Y}$ we further have

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}\right).$$

To accept the hypothesis $\mu_X = \mu_Y$ at the 95% significance level, we must have

$$\frac{|\overline{x} - \overline{y}|}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}}} < 1.96.$$

Since, using the problem's data $\overline{x} = 181$ and $\overline{y} = 177$, the expression on the left-hand side above can be calculated to be

$$\frac{|\overline{x} - \overline{y}|}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}}} = \frac{|181 - 177|}{\sqrt{\frac{32}{n} + \frac{29}{n}}} \approx 1.62 < 1.92.$$

Therefore we accept the hypothesis.

Solution to Problem 9.28. Let θ be the probability that a single item produced by the machine is defective, and K be the number of defective items out of $n = 600$ samples. Thus K is a binomial random variable and its PMF is

$$p_K(k) = \binom{n}{k} k^{\theta}(n-k)^{1-\theta}, \qquad k = 0, \ldots, n.$$

We have two hypotheses:

$$H_0 : \theta < 0.03, \qquad H_1 : \theta \geq 0.03.$$

We calculate the p-value

$$\alpha^* = \mathbf{P}(K \geq 28; H_0) = \sum_{28}^{600} \binom{600}{k} k^{0.03}(600-k)^{1-0.03},$$

which can be approximated by a normal distribution since $n = 600$ is large:

$$\begin{aligned}
\alpha^* &= \mathbf{P}\left(\frac{K - np}{\sqrt{np(1-p)}} \geq \frac{28 - np}{\sqrt{np(1-p)}}\right) \\
&= \mathbf{P}\left(\frac{K - 600 \cdot 0.03}{\sqrt{600 \cdot 0.03(1-0.03)}} \geq \frac{28 - 600 \cdot 0.03}{\sqrt{600 \cdot 0.03(1-0.03)}}\right) \\
&= \mathbf{P}\left(\frac{K - 18}{\sqrt{17.46}} \geq \frac{28 - 18}{\sqrt{17.46}}\right) \\
&\approx 1 - \Phi(2.39) \\
&\approx 0.84.
\end{aligned}$$

Since α^* is smaller than the 5% level of significance, there is a strong evidence that the null hypothesis should be rejected.

Solution to Problem 9.29. Let X_i be the number of rainy days in the ith year, and let $S = \sum_i^5 X_i$, which is also a Poisson random variable with mean 5μ. We have two hypotheses H_0 ($\mu = 35$) and H_1 ($\mu \geq 35$). Given the level of significance $\alpha = 0.05$ and an observed value $s = 159$, the test would reject H_0 if either

$$\mathbf{P}(S \geq s; H_0) \leq \alpha/2 \qquad \text{or} \qquad \mathbf{P}(S \leq s; H_0) \geq \alpha/2.$$

Therefore the p-value is

$$\begin{aligned}
\alpha^* &= 2 \cdot \min \left\{ \mathbf{P}(S \geq 159; H_0), \mathbf{P}(S \leq 159; H_0) \right\} \\
&= 2 \cdot \mathbf{P}(S \leq 159; H_0) \\
&\approx 2 \cdot \Phi\left(\frac{159 - 5 \cdot 35}{\sqrt{5 \cdot 35}} \right) \approx 0.2262,
\end{aligned}$$

where we use a normal approximation to $\mathbf{P}(S \leq 159; H_0)$. The obtained p-value is above the 5% level, so the test accepts the null hypothesis.

Solution to Problem 9.30. (a) The natural rejection rule is to reject H_0 when the sample mean $\overline{X}$ is greater than some value ξ. So the probability of false rejection is

$$\mathbf{P}(X > \xi; H_0) = 1 - \Phi(\frac{\xi}{\sqrt{\frac{v}{n}}}) = 0.05,$$

which gives

$$\xi = \Phi^{-1}(0.95)\sqrt{\frac{v}{n}} \approx 1.16.$$

Therefore when the observation is $X = 0.96 < 1.16$, we accept the null hypothesis H_0.

(b) With $n = 5$, the critical value is

$$\xi = \Phi^{-1}(0.95)\sqrt{\frac{v}{n}} \approx 0.52.$$

We compute the sample mean $\overline{X} = (0.96 - 0.34 + 0.85 + 0.51 - 0.24)/5 = 0.348$, smaller than ξ. So we still accept the H_0.

(c) Assuming that the variance is unknown, we estimate it by

$$\hat{v} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \approx 0.3680$$

using the t-distribution with $n = 5$, the T value is

$$T = \frac{\overline{X} - 0}{\sqrt{1/n}} \approx 1.2827 < 2.132 = t_{n-1,\alpha}.$$

So we still accept the H_0.