BETA

Chapter 1

Probability Basics

1.1 From Percentages to Probabilities

Basic Exercises

1.1 Let G stand for Graham, B for Baucus, C for Conrad, M for Murkowski, and K for Kyl.
a) {G, B, C, M, K}.
b) 3/5.

1.2 a) {{G,B},{G,C},{G,M}, {G,K},{B,C},{B,M},{B,K}, {C,M},{C,K},{M,K}}.
b) 3/10, as there are exactly three possible subcommittees consisting of two Democrats: {G,B},{G,C},{B,C}, out of a total of ten possible subcommittees of two senators.
c) 1/10, as there is exactly one possible subcommittee consisting of two Republicans, namely {M,K}, out of a total of ten possible choices for the subcommittee.
d) $(10-4)/10 = 6/10$, as out of ten possible committees, exactly four committees do not consist of one Republican and one Democrat.

1.3 a) {(G,B),(B,G),(G,C),(C,G),(G,M),(M,G), (G,K),(K,G),(B,C),(C,B),
(B,M),(M,B),(B,K),(K,B), (C,M),(M,C),(C,K),(K,C),(M,K),(K,M)},
where first letter in each pair $(\cdot,\cdot)$ corresponds to the name of the chair, while the second one corresponds to the name of the vice-chair.
b) Equally probable. In the first scenario the probability is 1/5, in the second scenario the probability is 4/20=1/5, since four outcomes (G,B),(G,C), (G,M),(G,K) out of twenty correspond to an event that Graham is the chair of the committee.
c) 12/20=3/5, since 12 ordered pairs out of 20 start with G,B or C.

1.4 Note that the total number of units (in thousands) is equal to
471+1,470+11,715+23,468+24,476+21,327+13,782+15,647=112,356.
Therefore,
a) $\dfrac{23,468}{112,356} \approx 0.2089$;
b) $\dfrac{24,476 + 21,327 + 13,782 + 15,647}{112,356} = \dfrac{75,232}{112,356} \approx 0.6696$;
c) $\dfrac{471 + 1,470}{112,356} = \dfrac{1,941}{112,356} \approx 0.0173$;
d) 0;
e) 1;

f) All U.S. housing units.

1.5 The total number of murder victims between 20 and 59 years old is equal to $2916 + 2175 + 1842 + 1581 + 1213 + 888 + 540 + 372 = 11,527$. Then

a) $\frac{1,213}{11,527} \approx 0.1052$;

b) $\frac{11,527 - 2,916}{11,527} = \frac{8,611}{11,527} \approx 0.7470$;

c) $\frac{888 + 540 + 372}{11,527} = \frac{1,800}{11,527} \approx 0.1562$;

d) $\frac{2916 + 2175 + 372}{11,527} = \frac{5,463}{11,527} \approx 0.4739$;

e) All murder victims in the U.S. who are between 20 and 59 years old.

1.6 The answers will vary. In general, suppose the number of left-handed female students in the class is ℓ_1 and the number of left-handed male students in the class is ℓ_2. Moreover, suppose the class consists of n_1 girls and n_2 boys. Then

a) $n_1/(n_1 + n_2)$;

b) $(\ell_1 + \ell_2)/(n_1 + n_2)$;

c) $\ell_1/(n_1 + n_2)$;

d) $(n_2 - \ell_2)/(n_1 + n_2)$.

1.7 a) $0.38 + (0.62)(0.2) = 0.504$, since 38% of students can speak only English, plus another 20% of the bilingual students do not speak Spanish.

b) $(0.62)(0.8)(0.1) = 0.0496$

1.8 a) The answers will vary since the experiment is random.

b) $n_5(E)/5$, where $n_5(E)$ is the number of heads in 5 tosses. The answer is based on the frequentist interpretation of probability.

c) $n_{10}(E)/10$, where $n_{10}(E)$ equals to the number of heads in 10 tosses. The answer is based on the frequentist interpretation of probability.

d) $n_{20}(E)/20$, where $n_{20}(E)$ equals to the number of heads in 20 tosses. The answer is based on the frequentist interpretation of probability.

e) In general, the sequence $n(E)/n$ is not monotone in n and may fail to have a limit.

1.9 On each day of the year, the probability that a Republican governor is chosen equals to $28/50 = 0.56$. Thus, a Republican governor is chosen 56% of time. Since 2004 is a leap year (with 366 days), the number of days when a Republican governor is selected to read the invocation is approximately equal to 205, since $366(0.56) = 204.96$

1.10 a) About 31.4 percent of human gestation periods exceed 9 months.

b) The favorite in a horse race finishes in the money in about two thirds of the races.

c) When a large number of traffic fatalities is considered, about 40% of the cases involve an intoxicated or alcohol-impaired driver or nonoccupant.

1.11 a) $4000 \times 0.314 = 1256$.

b) $500 \times (2/3) \approx 333$.

c) $389 \times 0.4 \approx 156$.

Advanced Exercises

1.12 The odds that the ball lands on red should be 18 to 20 for the bet to be fair.

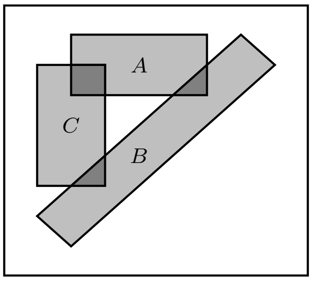
1.13 a) If p_1 is the probability that Fusaichi Pegasus wins the race, then $(1 - p_1)/p_1 = 3/5$. Thus, $p_1 = 5/8$.
b) If p_2 is the probability that Red Bullet wins the race, then $(1 - p_2)/p_2 = 9/2$. Thus, $p_2 = 2/11$. (Note: Posted odds at horse races are always "odds against").

1.2 Set Theory

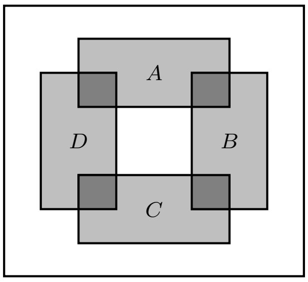
Basic Exercises

1.14 Yes. If the sets are pairwise disjoint, then the intersection of *all* the sets must be empty, since that intersection must be contained in *pairwise* intersections of the original sets.

1.15 The answers will vary. One example is given below:



1.16 The answers will vary. One example is given below:



1.17 The answers will vary. Say, take {{a,b,e},{b,c,f},{c,d,e},{d,a,f}}, where a,b,c,d,e,f are distinct points, then every pairwise intersection is not empty, but every three sets have an empty intersection.

1.18 a) $\bigcap_{n=1}^{4}[0, 1/n] = [0, 1/4]$ and $\bigcup_{n=1}^{4}[0, 1/n] = [0, 1]$.
b)

$$\bigcap_{n=1}^{\infty}[0, 1/n] = \{0\}, \qquad \bigcup_{n=1}^{\infty}[0, 1/n] = [0, 1],$$

1.19 In part a) assume the convention that $[a, b]$ is empty for $b < a$.
a) $(1, 2)$; **b)** $[1, 2)$; **c)** $[1, 2]$; **d)** $\{3\}$; **e)** $\emptyset$; **f)** $\emptyset$; **g)** $[5, 6)$.

1.20 a) $(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 3), (3, 4), (3, 5)$.

b) $(1,3)$, $(1,4)$, $(1,5)$, $(2,3)$, $(2,4)$, $(2,5)$, $(3,3)$, $(3,4)$, $(3,5)$, $(3,1)$, $(3,2)$, $(4,1)$, $(4,2)$, $(4,3)$, $(5,1)$, $(5,2)$, $(5,3)$.
c) $(1,1)$, $(1,2)$, $(2,1)$, $(2,2)$, $(4,4)$, $(4,5)$, $(5,4)$, $(5,5)$.
d) $A = (\{1,2\} \times \{1,2\}) \cup (\{4,5\} \times \{4,5\})$.

1.21 a) $(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)$.
b) $(0,0,1), (0,0,2), (0,1,1), (0,1,2), (1,0,1), (1,0,2), (1,1,1), (1,1,2)$.
c) (a,f), (a,g), (a,h), (b,f), (b,g), (b,h), (c,f), (c,g), (c,h), (d,f), (d,g), (d,h), (e,f), (e,g), (e,h).
d) The answer is the same as in part c).

1.22 $[1,2] \times [1,2]$.

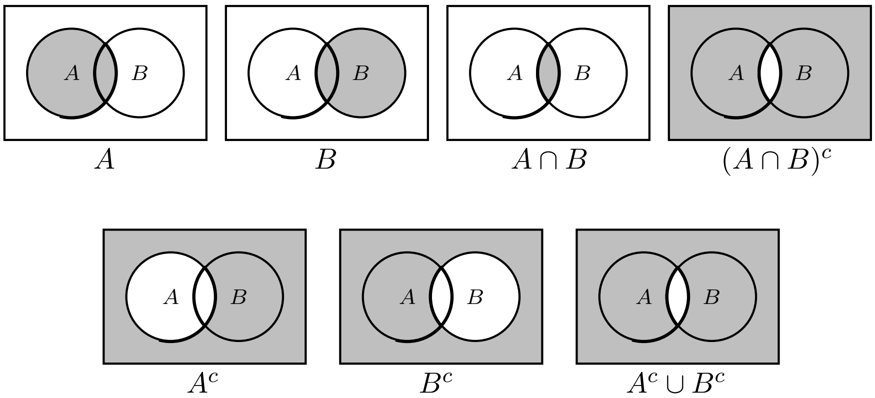
1.23 The answers will vary. For example, take $I_n = (0, 1/n)$, for $n \in \mathcal{N}$. Then $I_j \cap I_k = (0, \frac{1}{\max(j,k)}) \neq \emptyset$ for all j and k, but $\bigcap_{n=1}^{\infty}(0, 1/n) = \emptyset$.

1.24 The answers will vary. For example, take $I_n = [0, 1/n]$ for all $n \in \mathcal{N}$, then $\bigcap_{n=1}^{\infty} I_n = \{0\}$. (Note: the requirement $I_j \cap I_k \neq \emptyset$ could be omitted in the formulation of the problem).

1.25 The answers will vary. One possibility is $\mathcal{R} = \bigcup_{k \in \mathcal{Z}} [k, k+1)$.

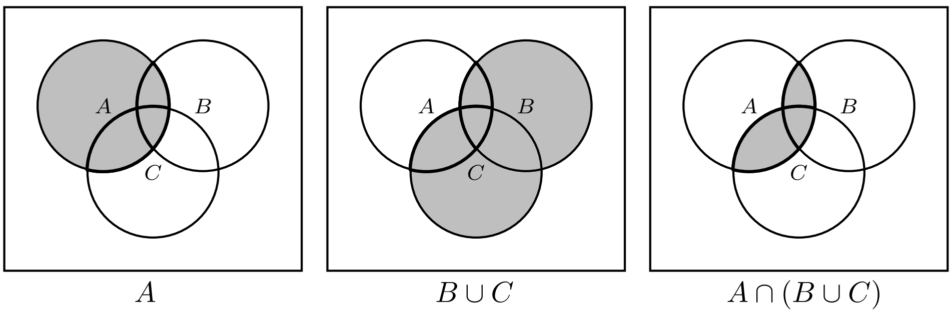
Theory Exercises

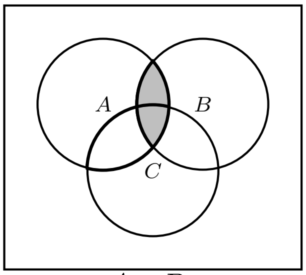
1.26 a) Verification of $(A \cap B)^c = A^c \cup B^c$ using Venn diagrams:



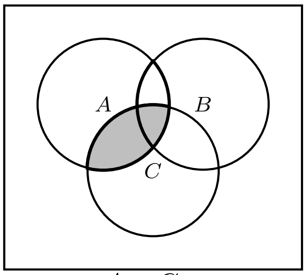
b) $x \in (A \cap B)^c$ if and only if $x \notin A \cap B$, which is true if and only if $x \notin A$ or $x \notin B$. The latter is equivalent to $x \in A^c$ or $x \in B^c$, which is in turn equivalent to $x \in A^c \cup B^c$.
c) By the first de Morgan's law, $(A^c \cup B^c)^c = (A^c)^c \cap (B^c)^c = A \cap B$, thus, $(A \cap B)^c = A^c \cup B^c$.

1.27 a) Verification of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ using Venn diagrams:

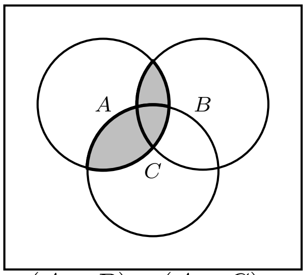




$A \cap B$

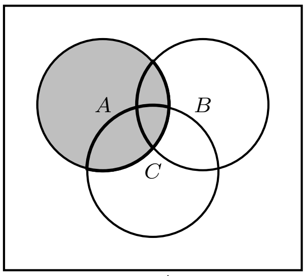


$A \cap C$

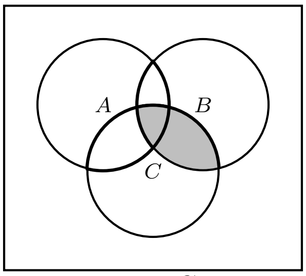


$(A \cap B) \cup (A \cap C)$

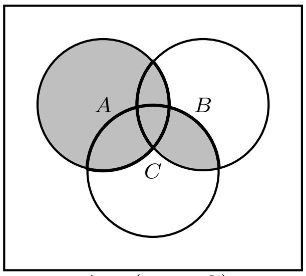
Verification of $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$:



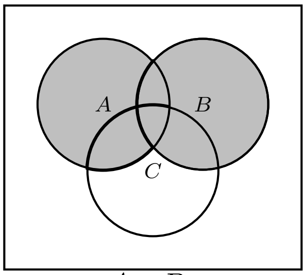
A



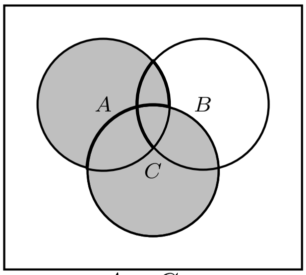
$B \cap C$



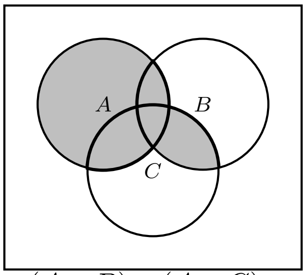
$A \cup (B \cap C)$



$A \cup B$



$A \cup C$



$(A \cup B) \cap (A \cup C)$

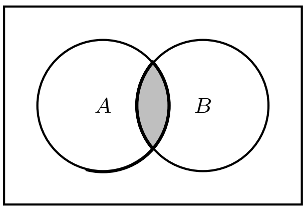
b) Proof of distributivity:

$$\begin{aligned} A \cap (B \cup C) &= \{x : x \in A \text{ and } x \text{ belongs to } (B \text{ or } C)\} \\ &= \{x : x \in A \cap B \text{ or } x \in A \cap C\} = (A \cap B) \cup (A \cap C). \end{aligned}$$

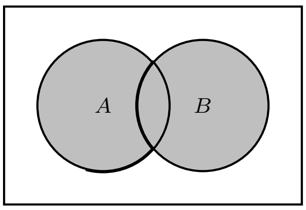
Moreover, using the de Morgan's laws and the above equation, one also obtains that

$$\begin{aligned} A \cup (B \cap C) = (A^c \cap (B \cap C)^c)^c = (A^c \cap (B^c \cup C^c))^c = ((A^c \cap B^c) \cup (A^c \cap C^c))^c \\ = (A^c \cap B^c)^c \cap (A^c \cap C^c)^c = (A \cup B) \cap (A \cup C). \end{aligned}$$

1.28 a) Commutativity is obvious from the Venn diagrams below:

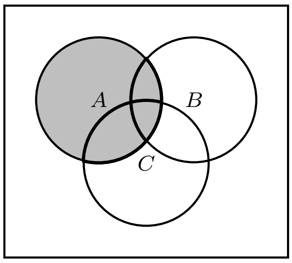


$A \cap B = B \cap A$

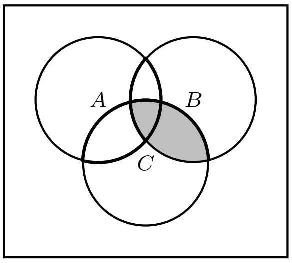


$A \cup B = B \cup A$

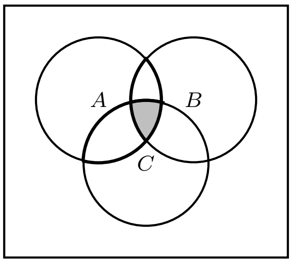
Verification of $A \cap (B \cap C) = (A \cap B) \cap C$ using Venn diagrams:



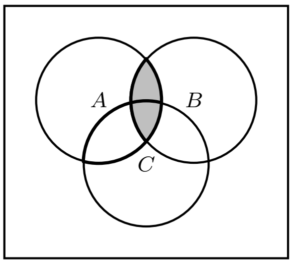
A



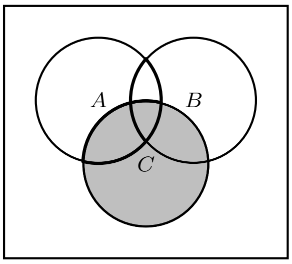
$B \cap C$



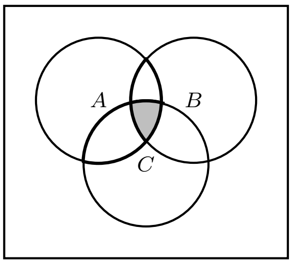
$A \cap (B \cap C)$



$A \cap B$

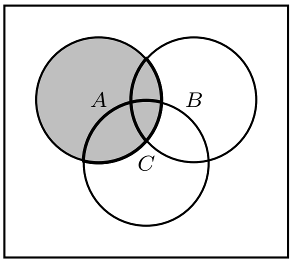


C

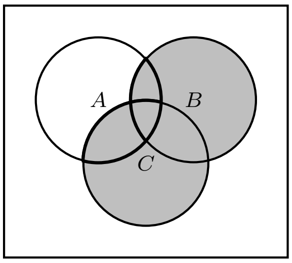


$(A \cap B) \cap C$

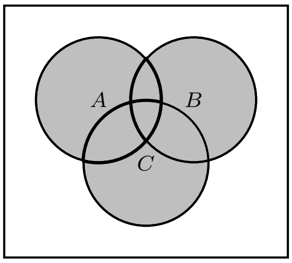
Verification of $A \cup (B \cup C) = (A \cup B) \cup C$ using Venn diagrams:



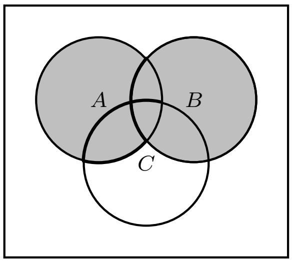
A



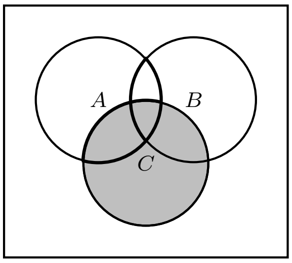
$B \cup C$



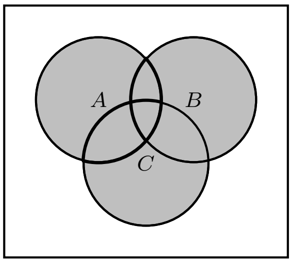
$A \cup (B \cup C)$



$A \cup B$



C



$(A \cup B) \cup C$

b) Proof of commutativity:

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

$$A \cup B = \{x : x \in A \text{ or } x \in B\} = B \cup A.$$

Proof of associativity:

$$\begin{aligned} A \cap (B \cap C) &= \{x : x \in A \text{ and } x \in B \cap C\} \\ &= \{x : x \in A \text{ and } x \in B \text{ and } x \in C\} \\ &= \{x : x \in A \cap B \text{ and } x \in C\} = (A \cap B) \cap C. \end{aligned}$$

$$\begin{aligned} A \cup (B \cup C) &= \{x : x \in A \text{ or } x \in B \cup C\} = \{x : x \in A \text{ or } x \in B \text{ or } x \in C\} \\ &= \{x : x \in A \cup B \text{ or } x \in C\} = (A \cup B) \cup C. \end{aligned}$$

1.29 a) $A \cup \emptyset$ consists of all the elements that are in A or in $\emptyset$, but $\emptyset$ has no elements, thus $A \cup \emptyset = A$.
b) Every element of A must be contained in the union of sets A and B, thus $A \subset A \cup B$.
c) Suppose $B \subset A$, then every element of B is also an element of A, implying that $A \cup B = A$. Conversely, suppose $A \cup B = A$, then $A^c = (A \cup B)^c = B^c \cap A^c$, implying that $B \cap A^c = B \cap (B^c \cap A^c) = (B \cap B^c) \cap A^c = \emptyset \cap A^c = \emptyset$. Thus, there are no elements of B which are not in A. In other words, every element of B must be in A, i.e. $B \subset A$.

1.30 a) $A \cap \emptyset$ consists of all elements that are in both A and $\emptyset$. Since $\emptyset$ has no elements, $A \cap \emptyset$ must be empty.
b) Every element of $A \cap B$ must be an element of A (and B), thus, $A \cap B \subset A$.
c) Suppose $A \subset B$. If $x \in A$, then $x \in B$, implying that $x \in A \cap B$. Therefore, $A \subset (A \cap B) \subset A$, where the last inclusion holds by part (b), thus, $A = A \cap B$. Conversely, suppose $A = A \cap B$, then

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

thus, there are no elements of A which are not in B, implying that every element of A must be in B, i.e. $A \subset B$.

1.31 a) To see that $(A \cap B) \cup (A \cap B^c) = A$, first note that $A \cap B \subset A$ and $A \cap B^c \subset A$, implying that $(A \cap B) \cup (A \cap B^c) \subset A$. On the other hand, every element from A is either in B (in which case it belongs to $A \cap B$) or not in B (in which case it belongs to $A \cap B^c$). Thus, $A \subset (A \cap B) \cup (A \cap B^c)$. Therefore, the required equality follows.
b) If $A \cap B = \emptyset$, then $x \in A$ implies that $x \notin B$ (since otherwise we will find $x \in A \cap B = \emptyset$, which is impossible). In other words, if $x \in A$ then $x \in B^c$. Thus, $A \subset B^c$.
c) If $A \subset B$, then $A = A \cap B$, implying that $A^c = (A \cap B)^c = A^c \cup B^c$. Therefore, $B^c \subset (A^c \cup B^c) = A^c$.

1.32 First let us show that

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

Note that $x \in (\bigcup_{n=1}^{\infty} A_n)^c$ if and only if $x \notin \bigcup_{n=1}^{\infty} A_n$, which is true if and only if $x \notin A_n$ for all $n \in \mathcal{N}$. The latter is equivalent to $x \in A_n^c$ for all $n \in \mathcal{N}$, which, in turn, holds if and only if $x \in \bigcap_{n=1}^{\infty} A_n^c$. To see the second identity, note (from the above with A_n^c in place of A_n) that

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} (A_n^c)^c = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c,$$

and, upon taking complements, we obtain that $(\bigcap_{n=1}^{\infty} A_n)^c = \bigcup_{n=1}^{\infty} A_n^c$.

1.33 First let us show that
$$B \cap (\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} (B \cap A_n).$$
Note that $x \in B \cap (\bigcup_{n=1}^{\infty} A_n)$ if and only if $x \in B$ and $x \in A_n$ for some $n \in \mathcal{N}$, which is equivalent to $x \in B \cap A_n$ for some $n \in \mathcal{N}$. The latter statement holds if and only if $x \in \bigcup_{n=1}^{\infty} (B \cap A_n)$. Thus, the required identity holds. Next note that (from the above with A_n^c in place of A_n, and B^c in place of B)
$$\bigcup_{n=1}^{\infty} (B^c \cap A_n^c) = B^c \cap (\bigcup_{n=1}^{\infty} A_n^c),$$
which, in view of Exercise 1.32, implies that
$$\bigcap_{n=1}^{\infty} (B \cup A_n) = \left(\bigcup_{n=1}^{\infty} (B \cup A_n)^c \right)^c = \left(\bigcup_{n=1}^{\infty} (B^c \cap A_n^c) \right)^c$$
$$= \left(B^c \cap (\bigcup_{n=1}^{\infty} A_n^c) \right)^c = B \cup \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c = B \cup \left(\bigcap_{n=1}^{\infty} A_n \right).$$

1.34 a) If $B \subset \bigcup_{n=1}^{\infty} A_n$, then every given point x of B belongs to A_n for some $n \in \mathcal{N}$, thus, $x \in B \cap A_n$ for some $n \in \mathcal{N}$, implying that $x \in \bigcup_{n=1}^{\infty} (B \cap A_n)$. Thus, $B \subset \bigcup_{n=1}^{\infty} (B \cap A_n)$. Conversely, note that $B \cap A_n \subset B$ for all $n \in \mathcal{N}$, thus, $\bigcup_{n=1}^{\infty} (B \cap A_n) \subset B$. Therefore, $B = \bigcup_{n=1}^{\infty} (B \cap A_n)$.
b) Required equality follows at once from part (a) with $B = E$.
c) If $A_1, A_2, \dots$ are pairwise disjoint, then $A_j \cap A_k = \emptyset$ for all $j \neq k$, then $(A_j \cap E) \bigcap (A_k \cap E) \subset (A_j \bigcap A_k) = \emptyset$ for all $j \neq k$, implying that $(A_1 \cap E), (A_2 \cap E), \dots$ are pairwise disjoint.
d) If $A_1, A_2, \dots$ form a partition of U, then, by part (b), every $E \subset U$ can be written as $E = \bigcup_{n=1}^{\infty} (A_n \cap E)$, where the latter union is the union of pairwise disjoint sets $(A_1 \cap E)$, $(A_2 \cap E), \dots$, due to part (c).

Advanced Exercises

1.35 a) Take an arbitrary point $x \in \liminf_{n\to\infty} A_n$. Then there exists an $n \in \mathcal{N}$ such that $x \in A_k$ for all $k \geq n$. Therefore, $x \in \bigcup_{k=m}^{\infty} A_k$ for every $m \in \mathcal{N}$, implying that $x \in \limsup_{n\to\infty} A_n$. Thus, $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$.
b) The set $\liminf_{n\to\infty} A_n$ consists of all the points that belong to all but finitely many A_n's. The set $\limsup_{n\to\infty} A_n$ consists of all the points that belong to infinitely many A_n's. Thus, clearly, the limit inferior is a subset of the limit superior.
c) On the one hand, $\limsup_{n\to\infty} A_n = [-1, 1]$, since every point in $[-1, 1]$ belongs to infinitely many A_n's and no point outside $[-1, 1]$ belongs to infinitely many A_n's. To see the latter, take an arbitrary $x \notin [-1, 1]$, then there exists $n \in \mathcal{N}$ such that $1 + \frac{1}{n} < |x|$, then $x \notin A_k$ for all $k \geq n$. Thus $x \notin \limsup_{n\to\infty} A_n$.
On the other hand, $\liminf_{n\to\infty} A_n = \{0\}$. Indeed, note that $\{0\} \in A_n$ for all $n \in \mathcal{N}$ and $A_n \cap A_{n+1} = \{0\}$ for all $n \in \mathcal{N}$, implying that $\bigcap_{k=n}^{\infty} A_k = \{0\}$ for all $n \in \mathcal{N}$. Thus, $\{0\} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

1.36 Define a function $f : \mathcal{N} \to \mathcal{Z}$ by $f(2k) = k$ and $f(2k-1) = -k+1$ for all $k \in \mathcal{N}$. Clearly,

f is a one-to-one function from $\mathcal{N}$ onto $\mathcal{Z}$, whose inverse f^{-1}, for each $z \in \mathcal{Z}$, is given by

$$f^{-1}(z) = \begin{cases} 2z, & \text{if } z \in \{1, 2, 3, \dots\}, \\ 1 - 2z, & \text{if } z \in \{0, -1, -2, -3, \dots\}, \end{cases}$$

thus $\mathcal{Z}$ and $\mathcal{N}$ are equivalent sets and $\mathcal{Z}$ is a countably infinite set.

1.37 Define a function $f : \mathcal{N}^2 \to \mathcal{N}$ by $f(m, n) = 2^{m-1}(2n - 1)$. Clearly, for every point $z \in \mathcal{N}$ there exists a unique pair $(m, n) \in \mathcal{N}$ such that $z = 2^{m-1}(2n - 1)$. Thus, f is a one-to-one function from $\mathcal{N}^2$ onto $\mathcal{N}$, implying that $\mathcal{N}^2$ is a countably infinite set.

1.38 Take an arbitrary sequence $(x_k)_{k=1}^{\infty}$, then its range $R(x_1, x_2, \dots) = \bigcup_{k=1}^{\infty}\{x_k\}$ is a countable set. Indeed, either the range $R(x_1, x_2, \dots)$ is finite or $R(x_1, x_2, \dots)$ can be written as a set $\{a_j : j \in \mathcal{N}\}$, where each $a_j = x_m$ for some $m \in \mathcal{N}$ and $a_1, a_2, \dots$ are all distinct. In the latter case, define a function $f : \mathcal{N} \to \{a_j : j \in \mathcal{N}\}$ by $f(n) = a_n$, then f is a one-to-one function from $\mathcal{N}$ onto $\{a_j : j \in \mathcal{N}\} = R(x_1, x_2, \dots)$. It follows therefore that the range of an arbitrary sequence is countable.
Conversely, let A be an arbitrary nonempty countable set. Then either A is finite or infinitely countable. If A is finite, say, $A = \{a_1, \dots, a_m\}$ for some finite $m \in \mathcal{N}$, then, for every $k \in \mathcal{N}$, let $x_k = a_k$ if $1 \le k \le m$, and $x_k = a_m$ for $k > m$. Then $(x_k)_{k=1}^{\infty}$ is an infinite sequence whose range equals A. Finally, if A is infinitely countable, then there is a one-to-one function f that maps $\mathcal{N}$ onto A. Thus, A is equal to the range of an infinite sequence $(f(n))_{n=1}^{\infty}$. Therefore, we showed that every nonempty countable set is the range of some infinite sequence.

1.39 Let A be a countable set and let g be an arbitrary function defined on A. By Exercise 1.38, A is the range of some infinite sequence $(x_k)_{k=1}^{\infty}$, thus,

$$g(A) = \{g(a) : a \in A\} = \{g(a) : a \in \bigcup_{k=1}^{\infty}\{x_k\}\} = \bigcup_{k=1}^{\infty}\{g(x_k)\},$$

implying that $g(A)$ is the range of an infinite sequence $(g(x_k))_{k=1}^{\infty}$. Thus, by Exercise 1.38, $g(A)$ is countable.

1.40 Let A be a countable set and $B \subset A$. If $B = \emptyset$, then it is finite and therefore countable. If $B \neq \emptyset$, then there exists some point $b \in B$, and we can define a function $g : A \to B$ by

$$g(a) = \begin{cases} a, & \text{if } a \in B, \\ b, & \text{if } a \in A \cap B^c. \end{cases}$$

Clearly, $g(A) = B$ thus, by Exercise 1.39 (which is based on Exercise 1.38), we conclude that B is countable.

1.41 Let A and B be countable sets. If at least one of the two sets is empty, then $A \times B = \emptyset$, in which case the cartesian product of A and B is a finite set and, therefore, countable. Now suppose that both A and B are nonempty. Then A is the range of some sequence $(x_k)_{k=1}^{\infty}$ and B is the range of some sequence $(y_j)_{j=1}^{\infty}$, implying that

$$A \times B = \{(a, b) : a \in A,\ b \in B\} = \{(a, b) : a \in \bigcup_{k=1}^{\infty}\{x_k\},\ b \in \bigcup_{j=1}^{\infty}\{y_j\}\}$$

$$= \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{\infty} \{(x_k, y_j)\} = \bigcup_{k \in \mathcal{K}} \bigcup_{j \in \mathcal{J}} \{(x_k, y_j)\},$$

where $\mathcal{K} \subset \mathcal{N}$ and $\mathcal{J} \subset \mathcal{N}$ are such that, for $k \in \mathcal{K}$, x_k's are all distinct and $\{x_k : k \in \mathcal{K}\}$ coincides with A, and, for $j \in \mathcal{J}$, y_j's are all distinct and $\{y_j : j \in \mathcal{J}\}$ coincides with B. Then

$$A \times B = \{(x_k, y_j) : (k, j) \in \mathcal{K} \times \mathcal{J}\} = g(\mathcal{K} \times \mathcal{J}),$$

where we defined $g((k,j)) = (x_k, y_j)$ for all $(k,j) \in \mathcal{K} \times \mathcal{J}$. By Exercise 1.37, $\mathcal{N}^2$ is countable, thus, by Exercise 1.40, $\mathcal{K} \times \mathcal{J}$, as a subset of the countable set $\mathcal{N}^2$, is itself countable. Then it follows that $g(\mathcal{K} \times \mathcal{J})$ is countable by Exercise 1.39. Hence $A \times B$ is countable. Moreover, suppose $A_1, A_2, \ldots, A_n$ is a finite collection of countable sets, then by induction one can obviously show that $A_1 \times A_2 \times \cdots \times A_n = A_1 \times (A_2 \times \cdots \times A_n)$ is a countable set.

1.42 Note that $\mathcal{Q} = \bigcup_{r \in \mathcal{Z}} \bigcup_{q \in \mathcal{N}} \{\frac{r}{q}\} = g(\mathcal{Z} \times \mathcal{N})$, where we define $g((r,q)) = r/q$ for all $(r,q) \in \mathcal{Z} \times \mathcal{N}$. By Exercises 1.36 and 1.41, $\mathcal{Z} \times \mathcal{N}$ is countable. Thus, by Exercise 1.39, $\mathcal{Q} = g(\mathcal{Z} \times \mathcal{N})$ is also countable.

1.43 a) Suppose the converse is true, i.e. that $[0,1)$ is countable. Then let $\{x_n\}_{n=1}^{\infty}$ be the enumeration of elements of $[0,1)$. For each x_n consider its decimal expansion and let k_n equal the nth digit in the decimal expansion of x_n. For each $n \in \mathcal{N}$ define $m_n = (k_n + 5) \bmod 10$. Consider the point

$$y = 0.m_1 m_2 m_3 \cdots = \sum_{n=1}^{\infty} m_n 10^{-n}.$$

Then, for every given $n \in \mathcal{N}$, $y \neq x_n$ since, by construction, $m_n \neq k_n$ (the points y and x_n differ (at least) in the nth digit in their respective decimal expansions). On the other hand, clearly $y \in [0,1)$. Thus we have a contradiction. Therefore, $[0,1)$ is uncountable.
b) If $(0,1)$ were countable, then $(0,1)$ would have been the range of some sequence $\{x_1, x_2, \ldots\}$, in which case $[0,1)$ would have been the range of the sequence $\{0, x_1, x_2, \ldots\}$, implying that $[0,1)$ is a countable set, which contradicts part (a). Thus, $(0,1)$ is uncountable.
c) Fix arbitrary real $a < b$. Define $f : (0,1) \to (a,b)$ by $f(x) = a + (b-a)x$ for all $x \in (0,1)$. Since f is a one-to-one function from $(0,1)$ onto (a,b), then $(0,1)$ and (a,b) are equivalent sets in the sense of definition on page 19. Thus, (a,b) is uncountable (since otherwise $(0,1)$ would have to be countable, which would contradict part (b)).
d) Let us show that any non-degenerate interval in $\mathcal{R}$ is uncountable. For any $a < b$ we have that, since $(a,b) \subset (a,b] \subset [a,b] \subset (-\infty, +\infty)$ and $(a,b) \subset [a,b)$, then each of the intervals $(a,b]$, $[a,b)$, $[a,b]$ and $\mathcal{R} = (-\infty, \infty)$ is uncountable, since otherwise, by Exercise 1.40, we will have a contradiction with the fact that (a,b) is uncountable (which was shown in part (c)).

1.44 Suppose A_n is a countable set for each $n \in \mathcal{N}$. Upon applying Exercise 1.38 for each fixed $n \in \mathcal{N}$, let $\{a_{n,k}\}_{k=1}^{\infty}$ be the enumeration of all elements of A_n. Next define sets $E_k = \{a_{1,k}, a_{2,k-1}, a_{3,k-2}, \cdots, a_{k,1}\}$ for all $k \in \mathcal{N}$. Then each E_k is a finite set and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{k=1}^{\infty} E_k$. Therefore, in view of Exercise 1.39, it suffices to show that the set J, defined by:

$$J = \bigcup_{k=1}^{\infty} \{(1,k), (2,k-1), (3,k-2), \cdots, (k,1)\}$$

is countable. But $J \subset \mathcal{N}^2$, where (by Exercise 1.41) $\mathcal{N}^2$ is countable. Thus, by Exercise 1.40, J is countable and the required conclusion follows.

1.45 a) To show consistency of the two definitions, take $I = \{1, \cdots, N\}$. Then the following set of functions

$$\{x : \{1, \cdots, N\} \to \cup_{j=1}^{N} A_j \text{ such that } x(i) \in A_i \text{ for all } i \in \{1, \cdots, N\}\}$$

can be readily identified with the set of vectors

$$\{(a_1, \cdots, a_N) : a_i \in A_i \text{ for all } i \in \{1, \cdots, N\}\}$$

since, given an arbitrary $(a_1, \cdots, a_N)$ (with $a_i \in A_i$), we can define a function x on I by putting $x(i) = a_i$ for each $i \in I$, and conversely, given a function x from the first set, the vector $(x(1), \cdots, x(N))$ belongs to the second set.
b) False. To see that the countable Cartesian product of countable sets may be uncountable, note first that every point $x \in [0, 1)$ has a decimal expansion $x = 0.x_1x_2x_3 \cdots = \sum_{k=1}^{\infty} x_k 10^{-k}$ for some $x_k \in \{0, 1, \cdots, 9\}$. For any set E, let E^∞ denote the countable Cartesian product of copies of E. Define a function $f : [0, 1) \to \{0, \cdots, 9\}^\infty$ by $f(x) = (x_1, x_2, x_3, \cdots)$. Clearly f is a one-to-one map from $[0, 1)$ onto $\{0, \cdots, 9\}^\infty$, where, by Exercise 1.1.43(a), the set $[0, 1)$ is uncountable. Thus, $\{0, \cdots, 9\}^\infty$ (which is a countable Cartesian product of finite sets) has to be uncountable.

1.3 Review Exercises

Basic Exercises

1.46 a) $11/16 = 0.6875$; **b)** The 16 states in the South.

1.47 a) $\dfrac{4}{190 + 71 + 61 + 25 + 10 + 4 + 87} = \dfrac{4}{448} = \dfrac{1}{112} \approx 0.00893$
b) $(61 + 25)/448 = 86/448 = 43/224 \approx 0.19196$
c) $(448 - 190)/448 = 258/448 = 129/224 \approx 0.57589$

1.48 a) In a large number of rolls, a die will have three dots facing up 1/6 of the time.
b) In a large number of rolls, a die will have three or more dots facing up two thirds of the time.
c) $10000/6 \approx 1667$.
d) $10000(2/3) \approx 6667$.

1.49 $n(Y) \approx 0.6n$, since $n(Y)/n \approx 0.6$ for large n.

1.50 a) $0, 1, 4$; **b)** $[0, 4)$.

1.51 Associate 0 with a tail and 1 with a head. Then, for example, $(0, 1, 0)$ denotes the outcome that the 1st and 3rd tosses are tails and the 2nd toss is a head. Then "Two heads and one tail in three tosses" corresponds to the following three outcomes $(1, 1, 0)$, $(1, 0, 1)$ and $(0, 1, 1)$. There are eight possible outcomes of the experiment: $(0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 0)$, $(1, 0, 0)$, $(0, 1, 1)$, $(1, 0, 1)$, $(1, 1, 0)$ and $(1, 1, 1)$. Thus, the required probability is 3/8.

1.52 $(A \cap B) \cup (A \cap B^c) = \{x \in A : x \in B\} \cup \{x \in A : x \notin B\} = \{x \in A\} = A$.

1.53 a) $\bigcup_{n=1}^{\infty} (\frac{1}{n+1}, \frac{1}{n}] = (0, 1]$, since $\bigcup_{n=1}^{k} (\frac{1}{n+1}, \frac{1}{n}] = (\frac{1}{k+1}, 1]$ for all $k \in \mathcal{N}$ and $\frac{1}{k+1} \to 0$ as $k \to \infty$.

b) $\bigcup_{n=1}^{\infty}[\frac{1}{n+1},\frac{1}{n}]=(0,1]$, since $\bigcup_{n=1}^{k}[\frac{1}{n+1},\frac{1}{n}]=[\frac{1}{k+1},1]$ for all $k\in\mathcal{N}$, and $\frac{1}{k+1}\to 0$ as $k\to\infty$, and $\{0\}\notin[\frac{1}{n+1},\frac{1}{n}]$ for all $n\in\mathcal{N}$.
c) Part (a) involves a union of pairwise disjoint sets $(\frac{1}{2},1]$, $(\frac{1}{3},\frac{1}{2}]$, $(\frac{1}{4},\frac{1}{3}]$, $(\frac{1}{5},\frac{1}{4}]$, $(\frac{1}{6},\frac{1}{5}]\dots$ Part (b) involves a union of sets that are not pairwise disjoint, since $[\frac{1}{n+2},\frac{1}{n+1}]\cap[\frac{1}{n+1},\frac{1}{n}]=\{\frac{1}{n+1}\}$, but whose intersection is empty since $[\frac{1}{n+3},\frac{1}{n+2}]\cap[\frac{1}{n+1},\frac{1}{n}]=\emptyset$.

1.54 a) Finite, since the required set equals to

$$\{(1,2),(1,3),(1,4),(2,2),(2,3),(2,4),(3,2),(3,3),(3,4)\}.$$

b) Uncountable, since the required set is equal to

$$\{(1,y):2\le y\le 4\}\cup\{(2,y):2\le y\le 4\}\cup\{(3,y):2\le y\le 4\},$$

which is the union of three parallel (nonempty) segments.
c) Uncountable, since the required set is equal to

$$\{(x,1):2\le x\le 4\}\cup\{(x,2):2\le x\le 4\}\cup\{(x,3):2\le x\le 4\},$$

which is the union of three parallel (nonempty) segments.
d) Countably infinite, since the set equals to $\bigcup_{n=1}^{\infty}\{(1,n),(2,n),(3,n)\}$.
e) Uncountable, since the set equals to $\{(x,y):1\le x\le 3,\ 2\le y\le 4\}$.
f) Countably infinite.
g) Finite, since $\bigcap_{n=1}^{\infty}\{n,n+1,n+2\}\subset(\{1,2,3\}\cap\{4,5,6\})=\emptyset$.

1.55 a) $(\{3,4\}^c\cap\{4,5\}^c)^c=(\{3,4\}^c)^c\cup(\{4,5\}^c)^c=\{3,4\}\cup\{4,5\}=\{3,4,5\}$.
b) Note that $\{3,4\}^c=\{1,2,5,6,7,8,\dots\}$ and $\{4,5\}^c=\{1,2,3,6,7,8,\dots\}$, thus, $\{3,4\}^c\cap\{4,5\}^c=\{1,2,6,7,8,\dots\}$, which implies the required equality: $(\{3,4\}^c\cap\{4,5\}^c)^c=\{1,2,6,7,8,\dots\}^c=\{3,4,5\}$.

Theory Exercises

1.56 a) Put $B_1=A_1$ and, for all $n\in\mathcal{N}$, let

$$B_n=A_n\bigcap\left(\bigcup_{k=1}^{n-1}A_k\right)^c=A_1^c\cap\dots\cap A_{n-1}^c\cap A_n.$$

Then $B_n\cap A_m=\emptyset$ for all $m<n$. Since $B_m\subset A_m$ for all m, it follows that $B_n\cap B_m=\emptyset$ for all $m<n$. Thus, the sets $B_1,B_2,\dots$ are pairwise disjoint. Now let us show that for every $n\in\mathcal{N}$, the equality $\bigcup_{j=1}^{n}B_j=\bigcup_{j=1}^{n}A_j$ holds. Note that $B_1\cup B_2=A_1\cup(A_2\cap A_1^c)=A_1\cup A_2$. Moreover,

$$\bigcup_{j=1}^{n}B_j=B_n\cup(\bigcup_{j=1}^{n-1}B_j)=\left(A_n\cap\left(\bigcup_{k=1}^{n-1}A_k\right)^c\right)\cup(\bigcup_{j=1}^{n-1}A_j)=\bigcup_{j=1}^{n}A_j,$$

where the 2nd equality has to be valid due to the induction assumption that $\bigcup_{j=1}^{n-1}B_j=\bigcup_{j=1}^{n-1}A_j$, whereas the last equality holds due to the identity $(E\cap F^c)\cup F=E\cup F$ for arbitrary E,F.
b) Select (pairwise disjoint) sets $B_1,B_2,\dots$ as in part (a). Upon noting that $x\in\bigcup_{j=1}^{\infty}A_j$ if

and only if there exists $n \in \mathcal{N}$ such that $x \in \bigcup_{j=1}^{n} A_j = \bigcup_{j=1}^{n} B_j$, and the latter is true if and only if $x \in \bigcup_{j=1}^{\infty} B_j$, we conclude that $\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} B_j$.

1.57 a) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \subset A \cup (B \cap C)$.
b) The answers will vary. Any choice of A, B, C such that $A \cap C \neq A$ will give a valid example. Say, take $A = \{a\}$ and $B = C = \emptyset$, then

$$(A \cup B) \cap C = \emptyset \neq \{a\} = A \cup (B \cap C).$$

c) By Proposition 1.2, we know that $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. If $A \subset C$, then $A \cap C = A$, implying that $(A \cup B) \cap C = A \cup (B \cap C)$.
d) Assume that $(A \cup B) \cap C = A \cup (B \cap C)$. Suppose that $A \not\subset C$, then there exists $x \in A$ such that $x \notin C$. Then, $x \in A \cup (B \cap C)$ but $x \notin (A \cup B) \cap C$, which is impossible since $(A \cup B) \cap C = A \cup (B \cap C)$. Therefore, $A \subset C$.

1.58 a) Let $\mathcal{C}$ be the class of all sets of the form $\{C \subset U : C \supset A \text{ and } C \supset B\}$. Clearly, since $A \subset A \cup B$ and $B \subset A \cup B$ (and $A \cup B \subset U$), then $(A \cup B) \in \mathcal{C}$, which implies that $(A \cup B) \supset \bigcap\limits_{C \in \mathcal{C}} C$. On the other hand, for every $C \in \mathcal{C}$, we must have that $C \supset (A \cup B)$. Therefore, $(\bigcap\limits_{C \in \mathcal{C}} C) \supset (A \cup B)$, and the required conclusion follows.
b) If $A \subset D$ and $B \subset D$, then $(A \cup B) \subset D$, since every element of $A \cup B$ is in A or in B, and thus belongs to D. In view of part (a), the smallest $D \in \mathcal{C}$ is equal to $A \cup B$. (Here "smallest" set means that every other set in the class contains it).
c) Let $\mathcal{E}$ be the class of all sets of the form $\{E \subset U : E \subset A \text{ and } E \subset B\}$. Clearly, $A \cap B \subset A$ and $A \cap B \subset B$, thus, $A \cap B \in \mathcal{E}$, implying that $A \cap B \subset \bigcup_{E \in \mathcal{E}} E$. On the other hand, for every $E \in \mathcal{E}$, we have that $E \subset (A \cap B)$, since every point of E has to belong to both A and B. Therefore, $(\bigcup_{E \in \mathcal{E}} E) \subset (A \cap B)$ and the required conclusion follows.
d) If $A \supset D$ and $B \supset D$, then $A \cap B \supset D$ and $A \cap B$ is the largest set in $\mathcal{E}$, by part (c). (Here "largest" set is in the sense that every other set in the class is contained in it).
e) Let $A_1, A_2, \ldots$ be a sequence of subsets of U. Let $\mathcal{C}^*$ be the class of all sets of the form $\{C \subset U : C \supset A_n \text{ for all } n \in \mathcal{N}\}$ and let $\mathcal{E}^*$ be the class of all sets of the form $\{E \subset U : C \subset A_n \text{ for all } n \in \mathcal{N}\}$. Then if $D \in \mathcal{C}^*$, then $D \supset (\bigcup_{n=1}^{\infty} A_n)$ and $\bigcap_{C \in \mathcal{C}^*} C = \bigcup_{n=1}^{\infty} A_n$. On the other hand, if $D \in \mathcal{E}^*$, then $D \subset \bigcap_{n=1}^{\infty} A_n$ and $\bigcup_{E \in \mathcal{E}^*} E = \bigcap_{n=1}^{\infty} A_n$.

1.59 a) True, since $A \times (B \cup C) = \{(x, y) : x \in A \text{ and } y \in (B \cup C)\}$
$= \{(x, y) : x \in A \text{ and } y \in B, \text{ or } x \in A \text{ and } y \in C\} = (A \times B) \cup (A \times C)$.
b) True, since $A \times (B \cap C) = \{(x, y) : x \in A \text{ and } y \in (B \cap C)\}$
$= \{(x, y) : x \in A \text{ and } y \in B \text{ and } y \in C\} = (A \times B) \cap (A \times C)$.
c) False. Any non-empty A, B such that $A \neq B$ will give required counterexample. For example, take $A = \{0\}$ and $B = \{1\}$, then $A \times B = \{(0, 1)\} \neq \{(1, 0)\} = B \times A$.

Advanced Exercises

1.60 a) $(71 + 92)/525 \approx 0.31$; **b)** $(71 + 61)/525 \approx 0.251$;
c) $71/(71 + 92) \approx 0.436$; **d)** $71/(71 + 61 + 37) = 71/169 \approx 0.42$;
e) $24/525 \approx 0.046$

1.61 The odds against randomly selecting an adult female Internet user, who believes that having a "cyber affair" is cheating, are 1:3, since three out of every four women believe that having a "cyber affair" is cheating.

1.62 a) $B^c = U \setminus B$;
b) $A \setminus B = \{x : x \in A \text{ and } x \notin B\} = \{x : x \in A \text{ and } x \in B^c\} = A \cap B^c$.
c) $(A \setminus B)^c = (A \cap B^c)^c = A^c \cup B$, by part (b) and the de Morgan's law.

1.63 $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

1.64 a) By Exercises 1.62 and 1.63 and properties of operations of complement, union and intersection, note first that

$$(A \triangle B) \triangle C = ((A \triangle B) \setminus C) \cup (C \setminus (A \triangle B))$$

$$= (((A \setminus B) \cup (B \setminus A)) \setminus C) \cup (C \setminus (A \triangle B))$$

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

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

where the right-hand side of the last equation is symmetric in A, B, C. Therefore, upon interchanging the roles of A and C, one immediately obtains

$$(A \triangle B) \triangle C = (C \triangle B) \triangle A = A \triangle (B \triangle C),$$

with the last equality being obviously true due to symmetry of operation $\triangle$ (since $E \triangle F = F \triangle E$ for arbitrary E, F).
b) $A \triangle U = (A \setminus U) \cup (U \setminus A) = \emptyset \cup A^c = A^c$.
c) $A \triangle \emptyset = (A \setminus \emptyset) \cup (\emptyset \setminus A) = A \cup \emptyset = A$.
d) $A \triangle A = (A \setminus A) \cup (A \setminus A) = \emptyset \cup \emptyset = \emptyset$.

1.65 a) For arbitrary E, F, note that

$$E \triangle F = (E \cup F) \setminus (E \cap F) = (E \cup F) \cap (E \cap F)^c,$$

therefore,

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

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

b) Note that

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

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

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

Since $(A^c \cap (B \triangle C)) \subset (B \triangle C) \subset (A \cup (B \triangle C))$, it follows that

$$(A \cup (B \triangle C)) \supset (A \cup B) \triangle (A \cup C).$$

c) The equality holds if and only $A^c \cap (B \triangle C) = A \cup (B \triangle C) = B \triangle C$, which is true if and only if $A = \emptyset$, since one must have that $A \subset (B \triangle C) \subset A^c$, implying that $A = A \cap A^c = \emptyset$.

1.66 If $B = \emptyset$, then $A \triangle B = A \triangle \emptyset = A$, by Exercise 1.64(c). Conversely, suppose $A = A \triangle B$. Then

$$B = \emptyset \triangle B = (A \triangle A) \triangle B = A \triangle (A \triangle B) = A \triangle A = \emptyset,$$

where the 1st equality holds by Exercise 1.64(c), the 2nd equality holds by Exercise 1.64(d), the 3rd equality is valid by Exercise 1.64(a), the 4th equality is true due to the assumption that $A = A \triangle B$, and the last equality holds by Exercise 1.64(d).

Chapter 2

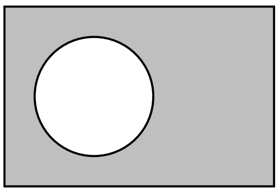
Mathematical Probability

2.1 Sample Space and Events

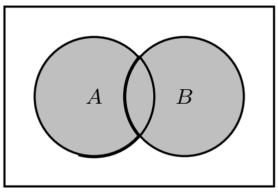
Basic Exercises

2.1 The required Venn diagrams are given by:

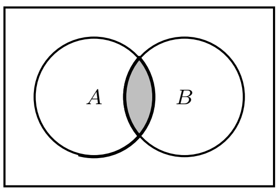
a) E^c= "E does not occur":



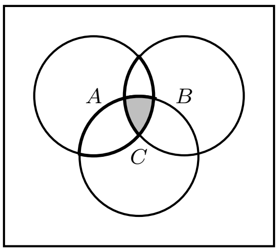
b) $A \cup B$= "Either A or B or both occur":



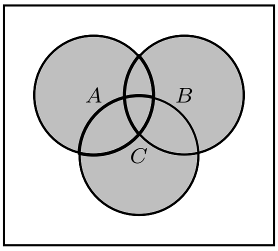
c) $A \cap B$= "Both A and B occur":



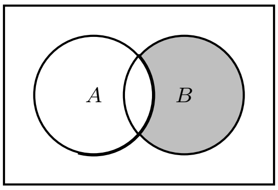
d) $A \cap B \cap C$= "All three events A, B, C occur at once":



e) $A \cup B \cup C$="At least one of the events A, B, C occurs":



f) $A^c \cap B$="B occurs but A does not occur":



2.2 a) False, unless both A and B are empty. A simple counterexample is obtained by taking $C = A \neq \emptyset$ and $A \cap B = \emptyset$, then A and B are mutually exclusive, but A and C are not mutually exclusive since $A \cap C = A \neq \emptyset$.
b) True, since the condition $A \cap B \neq \emptyset$ violates definition 2.3 of mutual exclusiveness of the events A, B, C regardless of the choice of C.

2.3 a) The sample space is $\Omega = [0, 1)$, since the distance between the center of the petri dish and the center of the first colony of bacteria can equal any number between 0 and 1. (Note that $\{0\}$ is a possible outcome since the center of the colony can be located exactly at the center of the dish, whereas $\{1\}$ is excluded from the sample space under the assumption that bacteria can only develop in the interior of the unit circle formed by the petri dish.)
b) $[1/4, 1/2]$= spot is at least 1/4 and no more than 1/2 unit from the center of the dish;
c) Distance between the center of the first colony of bacteria and the center of the petri dish does not exceed 1/3 unit.

2.4 a) $\Omega = \{1, 2, 3, 4, 5, 6\}$.
b) $A = \{2, 4, 6\}$, $B = \{4, 5, 6\}$, $C = \{1, 2\}$, $D = \{3\}$.
c) A^c = die comes up odd=$\{1, 3, 5\}$; $A \cap B$= die comes up either 4 or 6=$\{4, 6\}$; $B \cup C$ = die does not come up 3= $\{1, 2, 4, 5, 6\}$.
d) $A \cap B = \{4, 6\}$, thus A and B are not mutually exclusive; $B \cap C = \emptyset$ implies that B and C are mutually exclusive; A, C, D are not mutually exclusive since $A \cap C = \{2\} \neq \emptyset$.

e) The three events B, C, D are mutually exclusive, whereas the four events A, B, C, D are not mutually exclusive.
f) $\{5\}$ = die comes up 5; $\{1,3,5\}$= die comes up odd; $\{1,2,3,4\}$ = die comes up 4 or less.

2.5 a) Since one of the 39 trees is selected at random, the sample space is $\Omega = \{T_1, \cdots, T_{39}\}$, where T_n denotes the percentage of seed damage for the nth tree.

For any given event E, let $N(E)$ denote the number of outcomes in E.
b) B^c= selected tree has less than 20% seed damage, $N(B^c) = 19 + 2 = 21$;
c) $C \cap D$= selected tree has at least 50% but less than 60% seed damage, $N(C \cap D) = 2$;
d) $A \cup D$ = selected tree has either less than 40% or at least 50% seed damage, $N(A \cup D) = 39 - 6 = 33$;
e) C^c = selected tree has either less than 30% or at least 60% seed damage, $N(C^c) = 19 + 2 + 5 + 2 = 28$;
f) $A \cap D$ = selected tree has seed damage which, at the same time, is under 40% and at least 50% $= \emptyset$, thus, $N(A \cap D) = 0$.
g) A and D are mutually exclusive; this is the only pair of mutually exclusive events among A, B, C and D, therefore, it is the only collection of mutually exclusive events among A, B, C, D.

2.6 a) The sample space is given by:
$\Omega = \{(a,b,c) : a,b,c \in \{0,1,\cdots,9\}$ such that $a \neq b$ and $b \neq c$ and $a \neq c\}$.
(The three balls are removed from the urn one at a time, which suggests that the order, in which the balls are removed, is observed. However, for the purposes of part (b) of the problem, order is irrelevant and it suffices to consider a sample space given by $\tilde{\Omega} = \{\{a,b,c\} : a,b,c \in \{0,1,\cdots,9\}$ such that $a \neq b,\ b \neq c,\ c \neq a\}$, i.e. each outcome is an unordered set of three distinct integers in $[0,9]$.)
b) Let E denote the event that an even number of odd-numbered balls is removed. Then, on the sample space Ω,

$$E = \{(2k, 2\ell, 2m), (2r+1, 2n+1, 2k), (2k, 2r+1, 2n+1), (2r+1, 2k, 2n+1) :$$
$$k, \ell, m, r, n \in \{0,1,2,3,4\}, \text{ where } r \neq n \text{ and } k \neq \ell, \ell \neq m \text{ and } k \neq m\}.$$

(If one instead considers the sample space $\tilde{\Omega}$, then
$E = \{\{2k, 2\ell, 2m\}, \{2j, 2r+1, 2n+1\} : k, \ell, m, r, n, j \in \{0,1,2,3,4\},$ where
$k < \ell < m$ and $r < n\}$.)

2.7 $A_2 = \{(1,1)\},\ A_3 = \{(1,2),(2,1)\},\ A_4 = \{(1,3),(2,2),(3,1)\}$,
$A_5 = \{(1,4),(2,3),(3,2),(4,1)\},\ A_6 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$,
$A_7 = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$,
$A_8 = \{(2,6),(3,5),(4,4),(5,3),(6,2)\},\ A_9 = \{(3,6),(4,5),(5,4),(6,3)\}$,
$A_{10} = \{(4,6),(5,5),(6,4)\},\ A_{11} = \{(5,6),(6,5)\},\ A_{12} = \{(6,6)\}$.

2.8 a) The sample space is equal to $\Omega = \{(1,0),(1,1),(2,0),(2,1),(2,2),$
$(3,0),(3,1),(3,2),(3,3),(4,0),(4,1),\ (4,2),(4,3),(4,4),(5,0),(5,1),(5,2),$
$(5,3),(5,4),(5,5),(6,0),(6,1),(6,2),\ (6,3),(6,4),(6,5),(6,6)\}$.
In other words,

$$\Omega = \{(x,y) : x \in \{1,2,3,4,5,6\} \text{ and } y \text{ is integer such that } 0 \leq y \leq x\},$$

since whenever the die turns up x ($x \in \{1,\cdots,6\}$), the total number of heads that follow is anywhere between 0 and x, inclusive.

b) Total # of heads is even $= \{(x, 2k) : x \in \{1, 2, 3, 4, 5, 6\}$ and $k \in \{0, 1, 2, 3\}$, where $2k \le x\}$. More explicitly, total number of heads is even$=\{(1,0),(2,0),(2,2),(3,0),(3,2),(4,0),(4,2),$ $(4,4),(5,0),(5,2),(5,4),(6,0),(6,2),(6,4),(6,6)\}$.

2.9 a) $\Omega = \{\text{T, HT, HHT, HHHT, HHHHT}, \dots\}$.
b) Laura wins $= \{\text{HT, HHHT, HHHHHT}, \dots\}$. In other words,
Laura wins $= \{\underbrace{\text{H...H}}_{2k+1}\text{T} : k$ is nonnegative integer$\}$.

2.10 a) $\Omega = \{\text{H, TH, TTH, TTTH}, \dots\}$.
b) $\Omega = \{\text{HH, THH, HTH, TTHH, THTH, HTTH, TTTHH, TTHTH, THTTH, HTTTH}, \dots\}$. In other words, $\Omega = \{\underbrace{\text{T...T}}_{k}\text{H}\underbrace{\text{T...T}}_{\ell}\text{H} : k$ and ℓ are nonnegative integers$\}$.
c) $\{\text{TTTTTH}\}$.
d) $\{\text{TTTTHH, TTTHTH, TTHTTH, THTTTH, HTTTTH}\}$.

2.11 a) $\Omega = \{4, 5, 6, 7, 8, 9, 10\}$.
b) Note that $A = \{6, 7, 8, 9, 10\}$ and $B = \{4, 5, 6, 7, 8\}$, since "at least 4 women are on the jury" means that "8 or less men are on the jury." Then $A \cup B = \Omega = \{4, 5, 6, 7, 8, 9, 10\}$, $A \cap B = \{6, 7, 8\}$, $A \cap B^c = \{9, 10\}$.
c) A and B are not mutually exclusive since $A \cap B \ne \emptyset$; A and B^c are not mutually exclusive since $A \cap B^c \ne \emptyset$. Since $A^c \cap B^c = \{4, 5\} \cap \{9, 10\} = \emptyset$, thus A^c and B^c are mutually exclusive.

2.12 a) If A and B^c are mutually exclusive, then every element of A is not in B^c, thus, every element of A must belong to B, therefore, $A \subset B$ and B occurs whenever A occurs.
b) "B occurs whenever A occurs" implies that every element of A also belongs to B, i.e. $A \subset B$, which implies that $A \cap B = A$ and $A \cap B^c = \emptyset$.

2.13 a) $A \cap B^c$. **b)** $(A \cap B^c) \cup (A^c \cap B)$.
c) $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$, which is essentially saying that either A occurs but B and C do not occur, or B occurs but A and C do not occur, or C occurs but A and B do not occur.
d) $(A \cap B \cap C)^c$, since at most two of A, B, C occur if and only if all three events do not occur at the same time.

2.14 a) $\Omega = \{(n_x, m_y) : x, y \in \{hearts, diamonds, clubs, spades\}$,
$n, m \in \{ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, king\}$, where $n_x \ne m_y\}$.
(n_x stands for a card of denomination n of suit x.)
b) Since $(queen_{hearts}, ace_{hearts}) \in (A \cap B)$, thus, $A \cap B \ne \emptyset$, therefore, A and B are not mutually exclusive.

2.15 a) $\Omega = (0, \infty)$, where point 0 is excluded under the assumption that several patients cannot arrive at an exactly same time (therefore, only the very first patient can arrive at exactly 6:00 P.M.).
b) $\Omega = \{0, 1, 2, \dots\}$, since the number of people arriving to the emergency room during the first half-hour could be any nonnegative integer.
c) $A \subset B$, since $A=$ "the number of patients who arrive before 6:30 P.M. is at least five," and $B=$ "the number of patients who arrive before 6:30 P.M. is at least four."
d) $\Omega = \{(t_1, t_2, t_3, \dots) : 0 < t_1 < t_2 < t_3 < \dots\}$, where, for patients arriving after 6:00 P.M., t_i denotes the elapsed time (in hours) from 6:00 P.M. until the arrival of the ith patient.
e) False. If event A occurs, then Smith has to wait until the admission of the tenth patient or

longer. Since Jones has to stay only until the tenth patient is admitted, occurrence of A implies that Smith cannot leave the emergency room before Jones. Thus, $A \cap B=$ "Smith and Jones leave at the same time (when the tenth patient is admitted)." In other words, $A \cap B =$ "tenth patient is the first patient with a sprained ankle." Clearly, $A \cap B \neq \emptyset$, thus, the statement given in the problem is false.
f) False. Note that $C=$"first patient with a sprained ankle is either the tenth patient admitted to the emergency room or he arrives sometime after the tenth patient," thus, $A = C$, implying that $A \cap C = A \neq \emptyset$, thus A and C are not mutually exclusive.
g) False. Since, from answers to (e),(f), $A \cap B \neq \emptyset$ and $A = C$, then $B \cap C \neq \emptyset$ and the statement in the problem is false.
h) False. Indeed, suppose first patient with a sprained ankle is the 2nd patient admitted to the emergency room, then Smith leaves right after the 2nd patient, whereas Jones has to wait until the 10th patient. Thus, B occurs but A does not occur.
i) False. Suppose the first patient with a sprained ankle is the 12th patient admitted. Then A occurs but B does not occur.
j) True, since $A = C$ (see (f)).
k) True, since $A = C$.
l) False. The answer follows at once from $A = C$ and part (i).
m) False. The answer follows at once from $A = C$ and part (h).

2.16 Note that for each $n \geq 1$, $(A_{n+1} \cap A_n^c) \cap A_n = \emptyset$. Then, since for every $m \leq n-1$, $(A_{m+1} \cap A_m^c) \subset A_{m+1} \subset A_n$, it follows that

$$((A_{m+1} \cap A_m^c) \cap (A_{n+1} \cap A_n^c)) \subset (A_n \cap (A_{n+1} \cap A_n^c)) = \emptyset.$$

Therefore for every $n \geq 1$ and every $m < n$, the events $(A_{m+1} \cap A_m^c)$ and $(A_{n+1} \cap A_n^c)$ are mutually exclusive. Thus, $A_2 \cap A_1^c$, $A_3 \cap A_2^c$, $A_4 \cap A_3^c, \ldots$ are pairwise mutually exclusive.

2.17 a) The set of all events $=\{\emptyset, \{a\}, \{b\}, \{a,b\}\}$.
b) The set of all events $=\{\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}$.
c) The set of all events $=\{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\}, \{a,b,c\}, \{a,b,d\}, \{a,c,d\}, \{b,c,d\}, \{a,b,c,d\}\}$.
d) If Ω contains n outcomes, then there is 1 event that has 0 outcomes (empty set), n events that contain only 1 outcome, $\binom{n}{2}$ events that contain exactly 2 outcomes, and, more generally, there are $\binom{n}{k}$ events that contain exactly k outcomes from Ω, where $k \in \{1, \ldots, n\}$. Therefore, the total number of events is equal to

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

2.18 a) Take an arbitrary outcome ω from the intersection of all events $A_2, A_4, A_6, \ldots$ and $A_1^c, A_3^c, A_5^c, \ldots$. Then for each $n \geq 1$, we have that $\omega \in (A_n \cup A_{n+1}) \subset (\bigcup_{k=n}^{\infty} A_k)$. Since $\omega \in (\bigcup_{k=n}^{\infty} A_k)$ for all $n \geq 1$, it follows that $\omega \in \bigcap_{n=1}^{\infty}(\bigcup_{k=n}^{\infty} A_k)$.
b) See (a), since the same argument holds.
c) Take an arbitrary outcome ω in the intersection of all the sets A_{10}, A_{100}, $A_{1000}, \cdots$ and A_n^c, for all integers n which are not powers of 10. Then for each $n \geq 1$,

$$\omega \in \bigcup_{k=n}^{10n} A_k \subset (\bigcup_{k=n}^{\infty} A_k),$$

therefore, $\omega \in A^*$ and the required conclusion follows.
d) Take an arbitrary outcome ω which belongs to the intersection of all A_n^c with $n \notin \{86, 2049, 30498541\}$ and which also belongs to the event $A_{86} \cap A_{2049} \cap A_{30498541}$. Then $\omega \notin (\bigcup_{k=30498542}^{\infty} A_k)$, and since we have that $A^* \subset (\bigcup_{k=30498542}^{\infty} A_k)$, it follows that $\omega \notin A^*$.
e) Take an arbitrary outcome $\omega \in A^*$. Then for each $n \geq 1$, $\omega \in (\bigcup_{k=n}^{\infty} A_k)$, which implies that $\omega \in A_k$ for some $k \geq n$. Then ω must belong to A_j for infinitely many j. (Indeed, suppose that the set $J = \{j \geq 1 : \omega \in A_j\}$ is finite, let $j^* = \max\{j : j \in J\}$ and take $n = j^* + 1$. Then there exists $k \geq n$ such that $\omega \in A_k$, therefore, $k \in J$. Then $j^* \geq k \geq n = j^* + 1$, which results in contradiction.)

Conversely, suppose ω belongs to A_j for infinitely many j's. As before, let $J = \{j \geq 1 : \omega \in A_j\}$. Since J is infinite, for every n there exists $j \in J$ such that $j \geq n$. Therefore, for each n, there exists $j \geq n$ such that $\omega \in A_j \subset (\bigcup_{k=n}^{\infty} A_k)$, which implies that $\omega \in \bigcup_{k=n}^{\infty} A_k$ for all n, thus, $\omega \in A^*$. Thus, we showed that $A^* = \{\omega \in \Omega : \omega \in A_j \text{ for infinitely many } j\}$.

2.19 a) By condition (1) in the definition of the σ-algebra, if $A_n \in \mathcal{A}$ for $n = 1, 2, \ldots$, then $A_n^c \in \mathcal{A}$ for $n = 1, 2, \ldots$. Then, by (2), it follows that $(\bigcap_{n=1}^{\infty} A_n^c) \in \mathcal{A}$, which, in turn, implies that $(\bigcap_{n=1}^{\infty} A_n^c)^c \in \mathcal{A}$, by (1). Thus, by the De Morgan's law, $\bigcup_{n=1}^{\infty} A_n = (\bigcap_{n=1}^{\infty} A_n^c)^c \in \mathcal{A}$.
b) Let 2^Ω denote the collection of all possible subsets of Ω. Then:
(1) If $A \in 2^\Omega$, then $A \subset \Omega$, then $A^c = \{\omega \in \Omega : \omega \notin A\} \subset \Omega$, thus $A^c \in 2^\Omega$;
(2) If $A_n \in 2^\Omega$ for all $n \geq 1$, then $A_n \subset \Omega$ for all $n \geq 1$, then $\bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for all } n \geq 1\} \subset \Omega$, therefore, $(\bigcap_{n=1}^{\infty} A_n) \in 2^\Omega$.
(3) If $A_n \in 2^\Omega$ for all $n \geq 1$, then $A_n \subset \Omega$ for all $n \geq 1$, then $\bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for some } n \geq 1\} \subset \Omega$, therefore, $(\bigcup_{n=1}^{\infty} A_n) \in 2^\Omega$.
Thus, 2^Ω is a σ-algebra.
c) Let $\mathcal{A} = \{\emptyset, \Omega\}$, then for each $A \in \mathcal{A}$, either $A = \emptyset$ or $A = \Omega$. Since $\emptyset^c = \Omega$ and $\Omega^c = \emptyset$, we have that if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$. If $A_n \in \mathcal{A}$ for all $n \geq 1$, then either $A_n = \Omega$ for all $n \geq 1$, in which case $\bigcap_{n=1}^{\infty} A_n = \Omega \in \mathcal{A}$, or $A_n = \emptyset$ for some $n \geq 1$, in which case $\bigcap_{n=1}^{\infty} A_n = \emptyset \in \mathcal{A}$. By (a), it follows also that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, thus, $\mathcal{A} = \{\emptyset, \Omega\}$ is a σ-algebra.
d) Let $\mathcal{A} = \{\Omega, E, E^c, \emptyset\}$. Assuming that $E \neq \Omega$ and $E \neq \emptyset$ (otherwise, the problem reduces to (c)), note that since $(E^c)^c = E$ and $\emptyset^c = \Omega$, condition (1) in the definition of σ-algebra is satisfied by $\mathcal{A}$ at once. Next take an arbitrary sequence of events $A_n \in \mathcal{A}$ for all $n \geq 1$, then either it contains at least one $\emptyset$ or both E and E^c, in which case $\bigcap_{i=1}^{\infty} A_i = \emptyset$, or not. In the latter case there are three possibilities: either the sequence $\{A_i : i \geq 1\}$ contains only Ω and (at least one) E as its elements, or only Ω and (at least one) E^c, or only Ω's. Then $\bigcap_{i=1}^{\infty} A_i$ is equal to E or E^c or Ω respectively. In each case we have that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$, i.e. condition (2) of the definition is satisfied. By (a), condition (3) is then satisfied as well. Thus, $\mathcal{A}$ is a σ-algebra.
e) If $\mathcal{D} = \{$collection of all unions (including empty) of $E_1, E_2, \cdots\}$, where $E_1, E_2, \ldots$ are mutually exclusive events such that $\bigcup_{i=1}^{\infty} E_i = \Omega$, then every set $A \in \mathcal{D}$ has the form $A = \bigcup_{i \in I} E_i$ for some subset I of $\{1, 2, 3, \ldots\}$. Then $A^c = \bigcup_{i \in I^c} E_i$, implying that $A^c \in \mathcal{D}$ and condition (1) of the definition of σ-algebra holds. Next, for an arbitrary sequence $\{A_1, A_2, \cdots\}$ of sets in $\mathcal{D}$, there exists a sequence $\{I_1, I_2, \cdots\}$ of subsets of $\{1, 2, \cdots\}$ such that $A_j = \bigcup_{i \in I_j} E_i$ for each j. Then $\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} (\bigcup_{i \in I_j} E_i) = \bigcup_{i \in U} E_i$, where $U = \bigcup_{j=1}^{\infty} I_j$. Thus, $(\bigcup_{j=1}^{\infty} A_j) \in \mathcal{D}$, i.e. condition (3) in the definition of σ-algebra is satisfied. Condition (2) then follows at once from conditions (1),(3) and the De Morgan's law: $\bigcap_{j=1}^{\infty} A_j = \left(\bigcup_{j=1}^{\infty} A_j^c\right)^c$.
f) If Ω is an infinite set, then the class $\mathcal{D}$ of subsets of Ω, given by

$$\mathcal{D} = \{E \subset \Omega : \text{either } E \text{ or } E^c \text{ is finite}\},$$

is not a σ-algebra. Indeed, since Ω is infinite, one can choose in Ω an infinite sequence $\{x_1, x_2, x_3, \cdots\}$ of distinct points. Let $E_i = \{x_i\}$ for all $i = 1, 2, \ldots$. Then $E_i \in \mathcal{D}$ for all $i = 1, 2, \ldots$ Consider a set $G = \{x_2, x_4, x_6, x_8, \cdots\} = \{x_{2k} : k = 1, 2, \ldots\} \subset \Omega$. Then G is infinite and $G^c \supset \{x_1, x_3, x_5, \cdots\}$, therefore, G^c is also infinite. Thus, $G \notin \mathcal{D}$. On the other hand, $G = \bigcup_{k=1}^{\infty} E_{2k}$, where $E_{2k} \in \mathcal{D}$ for all $k = 1, 2, \cdots$. Therefore, condition (3) of the definition of σ-algebra is violated.

2.2 Axioms of Probability

Basic Exercises

2.20 The sample space is $\Omega = \{\omega_1, \cdots, \omega_9\}$, where $\omega_1 =$"Atlantic Ocean," $\omega_2 =$"Pacific Ocean," $\omega_3 =$"Gulf of Mexico," $\omega_4 =$"Great Lakes," $\omega_5 =$"Other lakes," $\omega_6 =$"Rivers and canals," $\omega_7 =$"Bays and sounds," $\omega_8 =$"Harbors," $\omega_9 =$"Other."
a) The probability assignment on the finite space Ω is legitimate since each $P(\{\omega_i\}) \geq 0$ for all $i \in \{1, \cdots, 9\}$ and

$$\sum_{i=1}^{9} P(\{\omega_i\}) = 0.011 + 0.059 + 0.271 + 0.018 + 0.003 + 0.211 + 0.094 + 0.099 + 0.234 = 1.$$

b) $P(\{\omega_1, \omega_2\}) = P(\{\omega_1\}) + P(\{\omega_2\}) = 0.011 + 0.059 = 0.07$
c) $P(\{\omega_4, \omega_5, \omega_8\}) = 0.018 + 0.003 + 0.099 = 0.12$
d) $P(\{\omega_3, \omega_7, \omega_8, \omega_9\}) = 0.271 + 0.094 + 0.099 + 0.234 = 0.698$

2.21 For $\Omega = \{\text{H}, \text{T}\}$, the set of all events of Ω is given by $\{\emptyset, \{\text{H}\}, \{\text{T}\}, \Omega\}$.
a) The probability assignment $P(\{\text{H}\}) = p$ and $P(\{\text{T}\}) = 1 - p$, where $0 \leq p \leq 1$, uniquely determines a probability measure on the events of Ω by Proposition 2.3, since $p \geq 0$, $1 - p \geq 0$ and $p + (1 - p) = 1$.
b) $P(\{\text{H}\}) = p$, $P(\{\text{T}\}) = 1 - p$, $P(\Omega) = 1$ and $P(\emptyset) = 0$.

2.22 The sample space is given by $\Omega = \{1, 2, 3, 4, 5, 6\}$.
a) Assignment #3 is not legitimate as $\sum_{k=1}^{6} P(\{k\}) = 6 \times 0.2 = 1.2 > 1$. Assignment #4 is not legitimate as it assigns negative values to outcomes $\{3\}$ and $\{5\}$. Assignments #1,2,5 are legitimate as they satisfy conditions of Proposition 2.3.
b) First note that $A = \{2, 4, 6\}$, $B = \{4, 5, 6\}$, $C = \{1, 2\}$ and $D = \{3\}$.
For assignment #1, $P(A) = P(B) = 3/6 = 1/2$, $P(C) = 2/6 = 1/3$ and $P(D) = 1/6$.
For assignment #2, $P(A) = 0.15 + 0.05 + 0.2 = 0.4$, $P(B) = 0.05 + 0.1 + 0.2 = 0.35$, $P(C) = 0.1 + 0.15 = 0.25$, $P(D) = 0.4$.
For assignment #5, $P(A) = 1/8 + 0 + 1/8 = 1/4$, $P(B) = 0 + 7/16 + 1/8 = 9/16$, $P(C) = 1/16 + 1/8 = 3/16$ and $P(D) = 1/4$.
c) The die comes up odd with probability 1/2 for assignment #1;
with probability $0.1 + 0.4 + 0.1$=0.6 for assignment #2, and
with probability $1/16 + 1/4 + 7/16 = 3/4$ for assignment #5.
d) If the die is balanced, then, by symmetry, all outcomes are equally likely to occur, thus assignment #1 should be used.
e) Based on the frequentist interpretation of probability, assignment #2 should be used.

2.23 a) Assignments #1,2,4,5 are legitimate. Indeed, each probability is non-negative and the

sum of probabilities of all distinct outcomes of the experiment equals one. (For #5, note that $\sum_{k=0}^{\infty} p(1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1$, and #1 is a special case of #5 with $p = 1/2$).
Assignment #3 is not legitimate since the sum of probabilities of all possible outcomes is equal to $\sum_{k=2}^{\infty} \frac{1}{k} = \infty$.
b) Let A_n denote the event that success occurs by the nth attempt. Then $P(A_2) = P(\{s\}) + P(\{f, s\})$, implying that:
#1: $P(A_2) = 3/4$. #2: $P(A_2) = 2/3$. #4: $P(A_2) = 0$.
#5: $P(A_2) = 2p - p^2 = 1 - (1-p)^2$.
Note also that $P(A_4) = P(A_2) + P(\{f, f, s\}) + P(\{f, f, f, s\})$. Then
#1: $P(A_4) = 15/16$. #2: $P(A_4) = 1$. #4: $P(A_4) = 1$.

#5: $P(A_4) = p\sum_{k=0}^{3}(1-p)^k = p \cdot \dfrac{1-(1-p)^4}{1-(1-p)} = 1-(1-p)^4$.

Similarly, $P(A_6) = P(A_4) + P(\{f, f, f, f, s\}) + P(\{f, f, f, f, f, s\})$. Thus,
#1: $P(A_6) = 63/64$. #2: $P(A_6) = 1$. #4: $P(A_6) = 1$.

#5: $P(A_6) = p\sum_{k=0}^{5}(1-p)^k = p \cdot \dfrac{1-(1-p)^6}{1-(1-p)} = 1-(1-p)^6$.

c) If $p = 0$ assignment #5 is not legitimate, since every possible outcome has probability zero, implying that $P(\Omega) = 0$, which is impossible. Note that $p = 0$ means that success never occurs. Let us add to Ω an outcome $\{f, f, f, \cdots\}$ and assign it probability one. Then the probability assignment, extended in that fashion, becomes legitimate.

2.24 a) Note that $B = (B \cap A) \cup (B \cap A^c)$, where $B \cap A$ and $B \cap A^c$ are mutually exclusive. Then, by the additivity axiom of probability,

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

b) Note that $A \cup B = A \cup (B \cap A^c)$, where the events A and $B \cap A^c$ are mutually exclusive, then, by the additivity axiom of probability,

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

c) From part (a), $P(A \cap B) = P(B) - P(B \cap A^c)$, where $P(B \cap A^c) = P(A \cup B) - P(A)$ by part (b). Thus, if $A \cup B = \Omega$, then

$$P(A \cap B) = P(B) - (P(A \cup B) - P(A)) = P(B) - 1 + P(A).$$

2.25 a) The probability assignment is legitimate because probability of each outcome is nonnegative and probabilities of all outcomes sum to 1, since there are 36 possible outcomes and each has probability 1/36.
b) $P(A_2) = P(A_{12}) = 1/36$, $P(A_3) = P(A_{11}) = 2/36$,
$P(A_4) = P(A_{10}) = 3/36$, $P(A_5) = P(A_9) = 4/36$,
$P(A_6) = P(A_8) = 5/36$, $P(A_7) = 6/36$.
c) The answers will vary. For example, let us take $P((1, 1)) = 1$, while the remaining 35 outcomes are assigned probability zero. Then $P(A_2) = 1$, whereas $P(A_n) = 0$ for all $n = 3, \cdots, 12$.
d) If the die is balanced, then the probability assignment from part (c) is not reasonable, whereas the assignment from part (a) is reasonable since, due to obvious symmetry, each face

of the die is equally likely to occur.

2.26 a) Since $0.01 \geq 0$ and $\sum_{k=1}^{100} 0.01 = 1$, the assignment is legitimate.
b) $P(A) = 50(0.01) = 0.5$; $P(B) = \sum_{k=1}^{31} 0.01 = 0.31$, since 31 is the largest integer that is less than or equal to 10π. Note also that the number of primes between 1 and 100 is equal to 25. Hence $P(C) = 25(0.01) = 0.25$.

2.27 a) $\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4\}\}$. **b)** 1/16.
c) $1 - \sum_{k=1}^{4} P((k, k)) = 1 - (4/16) = 3/4$.

2.28 a) $1/N$. **b)** m/N.

Theory Exercises

2.29 a) If $P(A_i) = 1/6$ for each i and $n = 10$, then $(*)$ tells us that

$$P(A_1 \cup \cdots \cup A_{10}) \leq \frac{10}{6},$$

which is trivial, since probability of any event must always be less than or equal to 1. On the other hand, if $P(A_i) = 1/60$ for each i and $n = 10$, then inequality $(*)$ states that $P(\bigcup_{i=1}^{10} A_i) \leq 10/60 = 1/6$, suggesting that the union of the ten events occurs no more frequently than 1/6 of the time, thus $(*)$ indeed provides some useful information. Similarly, if $P(A_i) = 1/6$ and $n = 4$, then inequality $(*)$ states that $P(\bigcup_{i=1}^{4} A_i) \leq 4/6 = 2/3$, which provides useful upper bound on the probability of occurrence of the union of the four events.
b) Upon using Exercise 2.24(b) and then Exercise 2.24(a), one obtains that

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2 \cap A_1^c) = P(A_1) + (P(A_2) - P(A_2 \cap A_1)) \\ &\leq P(A_1) + P(A_2), \end{aligned}$$

where the last inequality follows from the nonnegativity axiom, i.e. since $P(A_2 \cap A_1) \geq 0$. Let us proceed further by induction. Suppose $(*)$ holds for all $2 \leq n \leq k$. Let $B = \bigcup_{n=1}^{k} A_n$. Then

$$P(\bigcup_{n=1}^{k+1} A_n) = P(B \cup A_{k+1}) \leq P(B) + P(A_{k+1}) \leq \sum_{n=1}^{k} P(A_n) + P(A_{k+1}).$$

Advanced Exercises

2.30 a) Since $P((i, j)) \geq 0$ for all $i, j \in \mathcal{Z}_+$ and

$$P(\Omega) = \sum_{i,j=0}^{\infty} P((i, j)) = \sum_{i,j=0}^{\infty} e^{-5} \frac{2^i 3^j}{i!j!} = e^{-5} \left[\sum_{i=0}^{\infty} \frac{2^i}{i!}\right] \left[\sum_{j=0}^{\infty} \frac{3^j}{j!}\right] = e^{-5} e^2 e^3 = 1,$$

the probability assignment is legitimate.
b)

$$\begin{aligned} P(\text{point is on the } x\text{–axis}) &= P\left(\{(i, 0) : i = 0, 1, 2, \dots\}\right) \\ &= \sum_{j=0}^{\infty} e^{-5} \frac{2^i 3^0}{i!0!} = e^{-5} e^2 = e^{-3}. \end{aligned}$$

c)

$$P(\{(i,j) : i,j \in \{0,1,2,3\}) = \sum_{i,j=0}^{3} e^{-5}\frac{2^i 3^j}{i!j!}$$

$$= e^{-5}(\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!})(\frac{3^0}{0!}+\frac{3^1}{1!}+\frac{3^2}{2!}+\frac{3^3}{3!}) = e^{-5}\,\frac{19\times 13}{3} = \frac{247}{3e^5} \approx 0.5548.$$

d) For each $i \in \mathcal{Z}_+$,

$$P(\{(i,j) : j \in \{0,1,2,\dots\}\}) = \sum_{j=0}^{\infty} e^{-5}\,\frac{2^i 3^j}{i!j!} = \frac{2^i}{i!}e^{-5}e^3 = e^{-2}\,\frac{2^i}{i!}.$$

e) For each $j \in \mathcal{Z}_+$,

$$P(\{(i,j) : i \in \{0,1,2,\dots\}\}) = \sum_{i=0}^{\infty} e^{-5}\,\frac{2^i 3^j}{i!j!} = \frac{3^j}{j!}e^{-5}e^2 = e^{-3}\,\frac{3^j}{j!}.$$

2.31 a) For each $m, n \in \mathcal{N}$, we have that $P(\{(\omega_{1m}, \omega_{2n})\}) = p_{1m}p_{2n} \geq 0$, since both $p_{1m} \geq 0$ and $p_{2n} \geq 0$. Also $\Omega = \Omega_1 \times \Omega_2$ is a countable space, since both Ω_1 and Ω_2 are countable, and

$$\sum_{m,n=1}^{\infty} P(\{(\omega_{1m}, \omega_{2n})\}) = \sum_{m,n=1}^{\infty} p_{1m}p_{2n} = (\sum_{m=1}^{\infty} p_{1m})(\sum_{n=1}^{\infty} p_{2n}) = 1 \cdot 1 = 1.$$

Therefore, by Proposition 2.3, the probability assignment is legitimate.

b)

$$P(\{(\omega_{11}, \omega_{21}), (\omega_{12}, \omega_{24})\}) = p_{11}p_{21} + p_{12}p_{24}.$$

c)

$$P(\{(\omega_{1m}, \omega_{2n}) : n \in \mathcal{N}\}) = \sum_{n=1}^{\infty} P(\{(\omega_{1m}, \omega_{2n})\}) = p_{1m}(\sum_{n=1}^{\infty} p_{2n}) = p_{1m}.$$

d)

$$P(\{(\omega_{1m}, \omega_{2n}) : m \in \mathcal{N}\}) = \sum_{m=1}^{\infty} P(\{(\omega_{1m}, \omega_{2n})\}) = p_{2n}(\sum_{m=1}^{\infty} p_{1m}) = p_{2n}.$$

2.32 The answers will vary. As an example, suppose $\mathcal{Q} = \{x_1, x_2, x_3, \cdots\}$ is some enumeration of rational numbers. Let $P(\{x_k\}) = e^{-1}/(k-1)!$, then $P(\{\omega\}) > 0$ for all $\omega \in \mathcal{Q}$ and $\sum_{k=1}^{\infty} P(\{x_k\}) = e^{-1}\sum_{k=1}^{\infty}(1/(k-1)!) = e^{-1}\sum_{k=0}^{\infty}(1/k!) = e^{-1}e = 1$.

2.33 Suppose that $\Omega = [0,1] \times [0,1]$ is a countable set. Then, by Corollary 2.1,

$$\sum_{\omega \in \Omega} P(\{\omega\}) = 1.$$

On the other hand, since $P(\{\omega\}) = 0$ for every $\omega \in \Omega$, it follows that

$$\sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{\omega \in \Omega} 0 = 0,$$

which contradicts the previous condition. Since legitimacy of the probability assignment in the problem is taken for a fact, one has to conclude that the set $[0, 1] \times [0, 1]$ is uncountable.

2.34 Bruno de Finetti's thought experiment: Suppose you're willing to pay $\$x$ to be promised that, if the Red Sox win next year's World Series, you will receive \$1.
a) If $x > 1$, then your opponent should sell the promise to you, because if the Red Sox win next year's World Series, then your opponent's net profit is $(\$x - \$1) > \$0$, whereas if the Red Sox lose next year's World Series, then your opponent's net profit is $\$x > \$1$. Thus, in both cases your opponent makes a positive profit.
b) If $x < 0$, your opponent's optimal strategy is to buy the promise from you. Indeed, if the Red Sox win next year's World Series, then your opponent's net profit is $1 + |x|$ dollars. If the Red Sox lose next year's World Series, then your opponent's net profit is $|x|$ dollars. Thus, regardless of the outcome of the World Series, your opponent is making a positive profit.
c) If $x < 1$ but you're certain that the Red Sox will win, then you fear your opponent's decision to buy the promise from you, because in the latter case you have to sell the promise, which is worth \$1, for a lower price. In other words, you'll suffer a loss of $(1 - x) > 0$ dollars. However, if $x = 1$, you are not losing money on the deal.

On the other hand, if $x > 0$ but you're certain that the Red Sox will lose, you fear your opponents decision to sell you the promise, as you will have to pay a positive amount of money $\$x$ for the promise that's worth nothing. On the other hand, if $x = 0$ then you pay nothing for the promise, thus have nothing to lose when the Red Sox lose the World Series.
d) Let x denote the price of the promise that corresponds to the Red Sox, let y denote the price of the promise that corresponds to the Diamondbacks, and let z denote the price of the promise that corresponds to the win by either the Red Sox or the Diamondbacks. Then, if $x + y > z$, your opponent's optimal strategy is to sell to you both the promise for the Red Sox and the promise for the Diamondbacks, but to buy from you the promise concerning a victory by either the Red Sox or the Diamondbacks.

Indeed, if the Red Sox win the World Series, your opponent's profit will be $(\$x - \$1) + \$y + (\$1 - \$z) > \$0$, since he sells the promise worth \$1 for $\$x$, plus he sells the promise that's worth nothing for $\$y$, but he pays $\$z$ for the promise worth \$1. Similarly, one shows that if the Diamondbacks win, your opponent's profit is $\$x + (\$y - \$1) + (\$1 - \$z) > \$0$. Finally, if the winner is neither the Red Sox nor the Diamondbacks, then your opponent essentially sells you two promises that are worth nothing for a total of $(\$x + \$y)$ but buys a worthless promise for $\$z$, thus, his net profit is $\$x + \$y - \$z > \$0$. Thus, irrespective of the outcome of the World Series, your opponent ends up with a positive profit of $\$x + \$y - \$z$.
e) Let x, y, z be as in part (d), but now assume that $x + y < z$. Your opponent's optimal strategy is to sell the promise (priced at $\$z$) concerning a victory by either the Red Sox or the Diamondbacks but to buy the promise for the Red Sox (priced at $\$x$) and the promise for the Diamondbacks (priced at $\$y$).

Indeed, if the Red Sox win the World Series, your opponent's net profit is $(\$z - \$1) - \$y + (\$1 - \$x) > \$0$, since he essentially sold the promise, which is worth \$1, for $\$z$, and he paid $\$y$ for the worthless promise, plus he paid $\$x$ for the promise that's worth \$1. Similarly, one shows that if the Diamondbacks win, then your opponent's profit is $(\$z - \$1) + (\$1 - \$y) - \$x > \$0$. Finally, if both the Red Sox and the Diamondbacks lose, then your opponent's profit under this strategy is $\$z - \$x - \$y > \$0$. Thus, irrespective of the outcome of the World Series, your opponent has a positive profit of $\$z - \$x - \$y$.

2.3 Specifying Probabilities

Basic Exercises

2.35 a) $18/38 = 9/19 \approx 0.4737$; **b)** $9/19$; **c)** $2/38 = 1/19 \approx 0.0526$;
d) $20/38 = 10/19 \approx 0.5263$; **e)** $10/19$; **f)** $36/38 \approx 0.9474$

2.36 a) All components are identical and each one can fail with probability 0.5;
b) Note that $\Omega = \{(x_1, \cdots, x_5) : x_j \in \{s, f\}, 1 \le j \le 5\}$, where

$$P((x_1, \cdots, x_5)) = 2^{-5} = 1/32,$$

then for each event E,

$$P(E) = \frac{\text{\# of outcomes in } E}{32}.$$

c) $6/32 = 0.1875$; **d)** $1 - (1/32) = 31/32 = 0.96875$; **e)** $1/32 = 0.03125$

2.37 a) Equal-likelihood model is appropriate if the three blood-type alleles a, b and o occur with the same frequency (i.e. 1/3 of the time) and are equally likely to be inherited by an offspring.
b) $2/6 = 1/3 \approx 0.3333$;
c) No, because different genotypes do not appear in the population with the same frequency.
d) $P(\{bb, bo\}) = 0.007 + 0.116 = 0.123$, which is smaller than in (b).

2.38 Joint events in a contingency table are mutually exclusive since for all $(i, j) \neq (k, \ell)$,

$$(A_i \cap R_j) \cap (A_k \cap R_\ell) = (A_i \cap A_k) \cap (R_j \cap R_\ell) = \emptyset,$$

which is due to $A_i \cap A_k = \emptyset$ when $i \neq k$ and $R_j \cap R_\ell = \emptyset$ when $j \neq \ell$.

2.39 a) 12; **b)** 94; **c)** 42; **d)** 54; **e)** 22;
f) Y_3="The player has 6–10 years of experience"; W_2="The player weighs between 200 and 300 lb"; $W_1 \cap Y_2$="The player weighs under 200 lb and has 1–5 years of experience."
g) $P(Y_3) = 8/94 \approx 0.0851$, or approximately 8.51% of the players on the New England Patriots roster have 6–10 years of experience.
$P(W_2) = 54/94 \approx 0.5745$, or approximately 57.45% of the players on the New England Patriots roster weigh between 200 and 300 lb.
$P(W_1 \cap Y_2) = 9/94 \approx 0.0957$, or approximately 9.57% of the players on the New England Patriots roster weigh under 200 lb and have 1–5 years of experience.
h),i) The joint probability distribution table is given by:

	Y_1	Y_2	Y_3	Y_4	$P(W_i)$
W_1	$\frac{9}{94} \approx .096$	$\frac{9}{94} \approx .096$	$\frac{2}{94} \approx .021$	$\frac{1}{94} \approx .011$	$\frac{21}{94} \approx .223$
W_2	$\frac{22}{94} \approx .234$	$\frac{25}{94} \approx .266$	$\frac{5}{94} \approx .053$	$\frac{2}{94} \approx .021$	$\frac{54}{94} \approx .574$
W_3	$\frac{11}{94} \approx .117$	$\frac{7}{94} \approx .074$	$\frac{1}{94} \approx .011$	0	$\frac{19}{94} \approx .202$
$P(Y_j)$	$\frac{42}{94} \approx .447$	$\frac{41}{94} \approx .436$	$\frac{8}{94} \approx .085$	$\frac{3}{94} \approx .032$	1

2.40 a) Yes. The classical probability model is appropriate since dice are balanced, thus every pair of faces is equally likely to occur (i.e. each pair occurs with probability 1/36).

b) $3/36 \approx 0.0833$; **c)** $4/36 \approx 0.1111$; **d)** $5/36 \approx 0.1389$; **e)** $7/36 \approx 0.1944$

2.41 a) $1/6 \approx 0.1667$; **b)** $4/6 \approx 0.6667$; **c)** $1/6 \approx 0.1667$

2.42 a) $\dfrac{\text{length }[10,15]}{\text{length }[0,30]} = \dfrac{5}{30} \approx 0.1667$; **b)** $\dfrac{\text{length }[10,30]}{\text{length }[0,30]} = \dfrac{20}{30} \approx 0.6667$

2.43 a) $P(\{(x,y) : \sqrt{x^2+y^2} < 1/2\}) = \dfrac{\pi(1/2)^2}{\pi \cdot 1^2} = \dfrac{1}{4} = 0.25$, since area of a circle of radius R is equal to πR^2.
b) $1 - 0.25 = 0.75$;
c) The square in question is given by

$$C = \{(x,y) : -1/\sqrt{2} \le x \le 1/\sqrt{2},\ -1/\sqrt{2} \le y \le 1/\sqrt{2}\},$$

where C is inscribed into a ball B, given by

$$B = \{(x,y) : x^2 + y^2 \le 1\}.$$

Then, since $C \subset B$, it follows that

$$P(B \cap C^c) = \frac{\text{Area}(B) - \text{Area}(C)}{\text{Area}(B)} = \frac{\pi - (\sqrt{2})^2}{\pi} = 1 - \frac{2}{\pi} \approx 0.3634$$

d)

$$\frac{\text{Area}\left(\{(x,y) : -1/4 < x < 1/4, x^2+y^2 \le 1\}\right)}{\text{Area}\left(\{(x,y) : x^2+y^2 \le 1\}\right)} = \frac{2}{\pi}\int_{-1/4}^{1/4} \sqrt{1-x^2}\,dx$$

$$= \frac{4}{\pi}\int_0^{1/4} \sqrt{1-x^2}\,dx = \frac{4}{\pi}\left(\frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\arcsin(x)\right)\Bigg|_0^{1/4}$$

$$= \frac{\sqrt{15}}{8\pi} + \frac{2}{\pi}\arcsin(1/4) \approx 0.315$$

2.44 a) $A =$ "The x coordinate of the selected point in the square exceeds $1/3$," and $P(A) = 2/3$.
b) $B =$ "The y coordinate of the selected point in the square does not exceed 0.7," and $P(B) = 0.7$;
c) $C =$ "The sum of the two coordinates of the point, selected in the square, exceeds 1.2" and, since area of Ω is 1, it follows that

$$P(C) = \text{Area of triangle with vertices at } (0.2,1), (1,1), (1,0.2) = \frac{0.8^2}{2} = 0.32;$$

d) $D =$ "The difference between the two coordinates of the point, selected in Ω, is less than $1/10$," and

$$P(D) = P(\{(x,y) \in \Omega : x - 0.1 < y < x + 0.1\})$$

$$= \frac{\text{Area}(\Omega) - \text{Area}(T_1) - \text{Area}(T_2)}{\text{Area}(\Omega)} = \frac{1 - (0.9^2/2) - (0.9^2/2)}{1} = 0.19,$$

where T_1 denoted the triangle with vertices at $(0,0.1), (0,1), (0.9,1)$ and T_2 denoted the triangle with vertices at $(0.1,0), (1,0), (1,0.9)$.
e) $E =$ "The x and y coordinates of the selected point in Ω are equal" and $P(E) = 0$.

2.45 a) $P(\{(x,y) : 0 \le x \le 1,\ x < y \le 1\} = \frac{1}{2}P(\Omega) = \frac{1}{2}$.
b) The desired probability is given by:

$$P(\{(x,y) \in \Omega : \min(x,y) < 1/2\})$$

$$= \frac{\text{Area}(\Omega) - \text{Area}(\{(x,y) : 1/2 \le x < 1,\ 1/2 \le y < 1\}}{\text{Area}(\Omega)} = 1 - \frac{1}{2^2} = \frac{3}{4}.$$

c) $\frac{1}{4}$, since the required event is the complement of the event in (b).
d) $P(\{(x,y) : 0 \le x < 1/2,\ 0 \le y < 1/2\} = \frac{1}{4}$.
e) $\frac{3}{4}$, since the required event is the complement of the event in (d).
f) $P(\{(x,y) \in \Omega : 1 \le x+y \le 1.5\} = \dfrac{1}{\text{Area}(\Omega)}\,(\text{Area}(E_1) - \text{Area}(E_2))$
$= \frac{1}{2} - \frac{1}{2^3} = \frac{3}{8}$, where E_1 denoted the triangle with vertices at $(0,1),(1,1),(1,0)$ and E_2 denoted the triangle with vertices at $(0.5,1),(1,1),(1,0.5)$.
g) $P\{(x,y) \in \Omega : 1.5 \le x+y \le 2\} = \text{Area}(E_2) = 2^{-3} = 1/8$.

2.46 Note that $\Omega = (0,1)$. Then
a) $P([0.7, 0.8)) = 0.1$;
b) $P\left(\bigcup_{k=0}^{9}[0.1k+0.07, 0.1k+0.08)\right) = 10(0.01) = 0.1$;
c) 0.105; First note that the set $\left\{x \in (0,1) : \sqrt{x} \in \bigcup_{k=0}^{9}[0.1k+0.07, 0.1k+0.08)\right\} =$
$= \bigcup_{k=0}^{9}\left[(0.1k+0.07)^2, (0.1k+0.08)^2\right)$, implying that the probability, that the square root of the selected point has second digit of its decimal expansion equal to 7, is equal to

$$\sum_{k=0}^{9}\left((0.1k+0.08)^2 - (0.1k+0.07)^2\right) = (0.01)\sum_{k=0}^{9}(0.2k+0.15)$$

$$= 0.002\,(\sum_{k=0}^{9} k) + 0.015 = 0.002 \times 45 + 0.015 = 0.105$$

2.47 a) First note that Volume of $\{(x,y,z) : 1/4 < \sqrt{x^2+y^2+z^2} \le 1\}$
$=$ Volume of unit sphere – Volume of sphere $\{(x,y,z) : x^2+y^2+z^2 \le 1/16\}$
$= (4/3)\pi - (4/3)\pi(1/4)^3$, since the volume of a 3-dimensional sphere of radius R equals to $(4/3)\pi R^3$. Therefore, the required probability equals to

$$\frac{(4/3)\pi - (4/3)\pi(1/4)^3}{(4/3)\pi} = 1 - \frac{1}{64} = \frac{63}{64} = 0.984375$$

b) The cube of side length 1 centered at the origin is a subset of the unit sphere, thus, the required probability is given by

$$\frac{\text{Volume (Cube)}}{\text{Volume (Unit sphere)}} = \frac{1^3}{(4/3)\pi} = \frac{3}{4\pi} \approx 0.2387$$

c) 0, since volume of the surface of the 3-dimensional sphere is 0.

2.48 a) $\Omega = \{(x_1, x_2) : 0 \le x_1 \le b,\ 0 \le x_2 \le h - hx_1/b\}$ is the triangle of base b and height h with vertices at $(0,0), (b,0), (0,h)$. Consider triangle $T_x = \{(x_1, x_2) \in \Omega : x_2 > x\}$. Then T_x has vertices at $(0,x), (b(1-x/h), 0), (0,h)$ and represents the complement of the event of interest. Therefore, the required probability is equal to

$$\frac{\text{Area}(\Omega) - \text{Area}(T_x)}{\text{Area}(\Omega)} = 1 - \frac{\text{Area}(T_x)}{\text{Area}(\Omega)} = 1 - \frac{(1/2)(h-x)b(1-x/h)}{(1/2)hb}$$
$$= 1 - \left(1 - \frac{x}{h}\right)^2 = \frac{x}{h}\left(2 - \frac{x}{h}\right).$$

b) $\dfrac{\text{Area}(T_x) - \text{Area}(T_y)}{\text{Area}(\Omega)} = \dfrac{(1/2)(h-x)b(1-x/h) - (1/2)(h-y)b(1-y/h)}{(1/2)hb}$

$= \left(1 - \dfrac{x}{h}\right)^2 - \left(1 - \dfrac{y}{h}\right)^2 = \dfrac{y-x}{h}\left(2 - \dfrac{x+y}{h}\right)$, for all $h \ge y \ge x$.

2.49 Let X be the distance from a point, randomly selected on a given side of an equilateral triangle (with side length ℓ), to the opposite vertex. Then
a)

$$P(X \le x) = \begin{cases} 0, & \text{if } x < \frac{\sqrt{3}}{2}\ell, \\ \sqrt{\frac{4x^2}{\ell^2} - 3}, & \text{if } \frac{\sqrt{3}}{2}\ell \le x \le \ell, \\ 1, & \text{if } x > \ell. \end{cases}$$

Indeed, note that the range of possible values of X is $\left[\frac{\sqrt{3}}{2}\ell, \ell\right]$, since the longest distance X is achieved when the point is chosen at a vertex, whereas the smallest value of X is achieved when the point is chosen exactly in the middle of the side of the triangle (in which case X equals the height h of the triangle, which is $\ell \sin(\pi/3) = \ell\sqrt{3}/2$). Next, for $x \in [\ell\sqrt{3}/2, \ell]$, note that X is at most x if and only if the point is chosen on the side of the triangle within distance $d = \sqrt{x^2 - h^2}$ of the middle point of that side. Therefore, the required probability is given by:

$$\frac{\left|[\frac{\ell}{2} - d, \frac{\ell}{2} + d]\right|}{|[0,\ell]|} = \frac{2d}{\ell} = \frac{2\sqrt{x^2 - h^2}}{\ell} = \frac{2\sqrt{x^2 - (\ell\sqrt{3}/2)^2}}{\ell} = \sqrt{\frac{4x^2}{\ell^2} - 3},$$

where $|\cdot|$ denotes the 1-dimensional volume (i.e. length) of the intervals of interest.
b) For all $x < y$,

$$P(x \le X \le y) = \begin{cases} 0, & \text{if } y < \ell\sqrt{3}/2 \text{ or } x > \ell, \\ \sqrt{\frac{4y^2}{\ell^2} - 3}, & \text{if } x < \ell\sqrt{3}/2 \text{ and } y \in [\frac{\sqrt{3}}{2}\ell, \ell], \\ \sqrt{\frac{4y^2}{\ell^2} - 3} - \sqrt{\frac{4x^2}{\ell^2} - 3}, & \text{if } (x,y) \in [\frac{\sqrt{3}}{2}\ell, \ell]^2, \\ 1 - \sqrt{\frac{4x^2}{\ell^2} - 3}, & \text{if } x \in [\frac{\sqrt{3}}{2}\ell, \ell] \text{ and } y > \ell, \\ 1, & \text{if } x < \ell\sqrt{3}/2 \text{ and } y > \ell. \end{cases}$$

Clearly $P(X = x) = 0$ for every point x, since the 1-dimensional volume of any single point is 0. Thus, by the additivity axiom, it follows that $P(X \le x) = P(X < x) + P(X = x)$ $= P(X < x) + 0 = P(X < x)$ for all x. By the additivity axiom again,

$$P(X \le y) = P(\{X < x\} \cup \{x \le X \le y\}) = P(X < x) + P(x \le X \le y),$$

implying that

$$P(x \le X \le y) = P(X \le y) - P(X < x) = P(X \le y) - P(X \le x),$$

and the required result follows by plugging-in the answer to part (a).

2.50 a) $P(\text{white}) \approx 3194005/4058814 \approx 0.7869$, $P(\text{black}) \approx 622598/4058814 \approx 0.1534$, and $P(\text{other}) \approx (4058814 - 3194005 - 622598)/4058814 \approx 0.0597$;
b) The probabilities are empirical, since we use proportions of occurrence of events for assignment of probability to those events.

2.51 The probability specified by the engineer is subjective, as it relies on an incomplete information about the company, incomplete knowledge of the job market and limited interviewing experience.

2.52 The type of probability specified by the realtor is subjective (but with a hint of empirical evidence behind it if the realtor has data on how long comparable houses stay on the market).

Advanced Exercises

2.53 0. Let $\{r_1, r_2, \cdots\}$ be the enumeration of all rational points in $(0,1)$. Then, by Kolmogorov additivity axiom and Proposition 2.5 in the textbook,

$$P(\mathcal{Q} \cap (0,1)) = P(\bigcup_{i=1}^{\infty}\{r_i\}) = \sum_{i=1}^{\infty} P(\{r_i\}) = \sum_{i=1}^{\infty} \frac{|\{r_i\}|}{|(0,1)|} = \sum_{i=1}^{\infty} \frac{0}{1} = \sum_{i=1}^{\infty} 0 = 0,$$

where $|\cdot|$ stands for the length of the interval in question (i.e. it denotes the one-dimensional volume of the subset).

2.54 Suppose the converse, i.e. that there exists a countably infinite sample space $\Omega = \{\omega_1, \omega_2, \cdots\}$ which permits an equal-likelihood probability model. Then, for some $p \ge 0$, $P(\{\omega_i\}) = p$ for all $i \in \mathcal{N}$. Then, by additivity axiom,

$$P(\Omega) = \sum_{i=1}^{\infty} P(\{\omega_i\}) = \sum_{i=1}^{\infty} p = \begin{cases} \infty, & \text{if } p > 0, \\ 0, & \text{if } p = 0. \end{cases}$$

But this contradicts the Kolmogorov certainty axiom that $P(\Omega) = 1$. Therefore, equal-likelihood probability models are not possible in the case of countably infinite sample spaces.

2.55 1/2. First note that the three line segments can form a triangle whenever the length of each of them is less than the sum of lengths of the other two. Therefore, the required event of interest is the set $T = \{(x,y,z) \in (0,\ell)^3 : x < y+z,\ y < x+z,\ z < x+y\}$ and the required probability is given by

$$\begin{aligned} \frac{|T|}{|(0,\ell)^3|} &= \frac{\left|\{(x,y,z) \in (0,\ell)^3 : x < y+z,\ y < x+z,\ z < x+y\}\right|}{|(0,\ell)^3|} \\ &= \frac{1}{\ell^3}\left(\left|(0,\ell)^3\right| - 3\left|\{(x,y,z) \in (0,\ell)^3 : z > x+y\}\right|\right), \end{aligned}$$

where $|\cdot|$ denotes the three-dimensional volume of a set. Therefore,

$$P(T) = 1 - \frac{3}{\ell^3} \iiint\limits_{\substack{0<x<\ell \\ 0<y<\ell \\ x+y<z<\ell}} dxdydz = 1 - \frac{3}{\ell^3} \iint\limits_{\substack{x>0,y>0 \\ x+y<\ell}} (\ell-(x+y))dxdy$$

$$= 1 - \frac{3}{\ell^3}\int_0^\ell (\int_0^{\ell-x} (\ell-x-y)dy)dx = 1 - \frac{3}{\ell^3}\int_0^\ell \frac{1}{2}(\ell-x)^2dx$$

$$= 1 + \left(\left.\frac{(\ell-x)^3}{2\ell^3}\right|_0^\ell \right) = 1 - \frac{1}{2} = \frac{1}{2}.$$

2.56 1/2. Whenever two numbers, say, a and b, are selected from the interval $(0,\ell)$, the resulting three segments of lengths $\min\{a,b\}$, $|b-a|$ and $\ell - \max\{a,b\}$ can form a triangle whenever the length of each of the three segments is less than the sum of the lengths of the other two. Thus, the problem reduces to Exercise 2.55.

2.57 Buffon's needle problem: $2\ell/(\pi d)$. Let us determine position of the needle by specifying the distance x from the center of the needle to the closest parallel line and the angle y (in radians) that the needle forms with that line. Thus, the experiment is equivalent to selecting at random a point in the set $[0,d/2]\times[0,\pi]$. The needle will intersect a line if and only if $\frac{x}{\sin(y)} < \frac{\ell}{2}$. Thus, the required probability is given by:

$$\frac{|\{(x,y)\in[0,d/2]\times[0,\pi] : x < \ell\sin(y)/2\}|}{|[0,d/2]\times[0,\pi]|} = \frac{1}{\pi(d/2)}\int_0^\pi (\int_0^{\ell\sin(y)/2} dx)dy$$

$$= \frac{2}{\pi d}\int_0^\pi \frac{\ell}{2}\sin(y)dy = \frac{\ell}{\pi d}(-\cos(y))\Big|_0^\pi = \frac{2\ell}{\pi d},$$

where in the above $|\cdot|$ denotes the two-dimensional volume of a set.

2.4 Basic Properties of Probability

2.58 The answers will vary. One such example is obtained if we let $\Omega = [0,1]$ and suppose that $P(E) = |E|$ for any event $E \subset \Omega$, where $|\cdot|$ is the one-dimensional volume of E. (Assume that any interval in $[0,1]$ is an event). Take $A = (1/3, 2/3)$ and $B = [1/3, 2/3]$. Then A is a proper subset of B, but $P(A) = 1/3 = P(B)$.

2.59 The answers will vary. One needs to find sets A and B such that $P(A) \leq P(B)$ but $A \not\subset B$. Consider, for example, $\Omega = \{1,2,3\}$ with $P(\{1\}) = P(\{2\}) = P(\{3\}) = 1/3$. Then taking $A = \{1\}$ and $B = \{2,3\}$ produces the desired result.

2.60 Let L be the event that the selected person is a lawyer, and let R be the event that the selected person is a Republican. By the domination principle, L is more probable than $R \cap L$, since $(R\cap L) \subset L$.

2.61 If X denotes the number of Republicans on the subcommittee, then, by the complementation rule,

$$P(X \geq 1) = 1 - P(X=0) = 1 - 1/6 = 5/6.$$

If the complementation rule is not used, one needs to determine and sum the probabilities of the events $\{X = 1\}$ and $\{X = 2\}$, which takes longer.

2.62 a) $1 - 0.271 = 0.729$;
b) $0.011 + 0.059 + 0.018 + 0.003 + 0.211 + 0.094 + 0.099 + 0.234 = 0.729$;
c) The method used in part (a) requires less work than the one in (b).

2.63 a) $1 - (42/94) = 52/94 \approx 0.553$;
b) $1 - (19/94) = 75/94 \approx 0.798$;
c) Using the general addition rule:

$$P(Y_1 \cup W_3) = P(Y_1) + P(W_3) - P(Y_1 \cap W_3) = \frac{42}{94} + \frac{19}{94} - \frac{11}{94} = \frac{50}{94} \approx 0.532,$$

whereas computation without the general addition rule gives:

$$P(Y_1 \cup W_3) = \frac{9 + 22 + 11 + 7 + 1 + 0}{94} = \frac{50}{94} \approx 0.532;$$

clearly, the second computation involves more terms.

2.64 Using the complementation rule:

$$P\left(\{(x,y) \in [0,1]^2 : |x-y| \le \frac{1}{4}\}\right) = 1 - P\left(\{(x,y) \in [0,1]^2 : |x-y| > \frac{1}{4}\}\right)$$
$$= 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16} = 0.4375,$$

since the set $\{(x,y) \in [0,1]^2 : |x-y| > 1/4\}$ is the union of two disjoint (equal) right triangles Δ_1 and Δ_2 (where Δ_1 has vertices $(0, \frac{1}{4}), (0,1), (\frac{3}{4}, 1)$, while Δ_2 has vertices $(1, \frac{3}{4}), (1,0), (\frac{1}{4}, 0)$). If the complementation rule is not used, then the required set is the union of two disjoint equal trapezoids T_1 and T_2, where, say, T_1 has vertices at the points $(0,0), (0, \frac{1}{4}), (\frac{3}{4}, 1), (1,1)$ and its area equals to

$$|T_1| = \frac{1}{2}\left(\frac{3\sqrt{2}}{4} + \sqrt{2}\right)\frac{\sqrt{2}}{8} = \frac{7}{32},$$

since T_1 has bases of lengths $3\sqrt{2}/4$ and $\sqrt{2}$, and height of size $\sqrt{2}/8$. Thus,

$$P\left(\{(x,y) \in [0,1]^2 : |x-y| \le \frac{1}{4}\}\right) = 2P(T_1) = 2 \times \frac{7}{32} = \frac{7}{16} = 0.4375;$$

the first method (using the complementation rule) is simpler.

2.65 a) $0.51 + 0.071 - 0.041 = 0.54$;
b) $1 - 0.51 = 0.49$;
c) $0.51 - 0.041 = 0.469$;
d) $0.071 - 0.041 = 0.03$

2.66 a) No. A and B are not mutually exclusive since

$$P(A) + P(B) = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} \ne \frac{1}{2} = P(A \cup B).$$

b) $P(A \cap B) = P(A) + P(B) - P(A \cup B) = \frac{7}{12} - \frac{1}{2} = \frac{1}{12} \approx 0.0833$

2.67 a) $P(B) = P(A \cup B) - P(A) + P(A \cap B) = \frac{5}{8} - \frac{1}{3} + \frac{1}{10} = \frac{47}{120} \approx 0.3917$;
b) $P(A \cap B^c) = P(A) - P(A \cap B) = \frac{1}{3} - \frac{1}{10} = \frac{7}{30} \approx 0.2333$;
c) $P(A \cup B^c) = P(A) + P(B^c) - P(A \cap B^c) = \frac{1}{3} + (1 - \frac{47}{120}) - \frac{7}{30} \approx 0.7083$;
d) $P(A^c \cup B^c) = P((A \cap B)^c) = 1 - P(A \cap B) = 1 - \frac{1}{10} = 0.9$

2.68 Let L be the event that the selected person enjoys his/her personal life, and J be the event that the selected person enjoys his/her job. Then
a) 85%, since $P(J \cup L) = 1 - P(J^c \cap L^c) = 1 - 0.15 = 0.85$;
b) 84%, since $P(J) = P(J \cap L) + P(J \cap L^c) = 0.04 + 0.8 = 0.84$;
c) 1%, since $P(L \cap J^c) = P(J^c) - P(J^c \cap L^c) = (1 - 0.84) - 0.15 = 0.01$;
d) 5%, since $P(L) = P(L \cap J) + P(L \cap J^c) = 0.04 + 0.01 = 0.05$

2.69 a) $P(T \cap H^c) + P(H \cap T^c) = (P(T) - P(T \cap H)) + (P(H) - P(H \cap T))$
$= 0.47 - 0.119 + 0.334 - 0.119 = 0.566$;
b) Note first that

$$\begin{aligned} P(T \cap H^c \cap E^c) &= P(T \cap H^c) - P(T \cap H^c \cap E) \\ &= (P(T) - P(T \cap H)) - (P(T \cap E) - P(T \cap E \cap H)) \\ &= P(T) - P(T \cap H) - P(T \cap E) + P(T \cap E \cap H) \\ &= 0.47 - 0.119 - 0.151 + 0.048 = 0.248; \end{aligned}$$

similarly, one establishes that

$$P(T^c \cap H \cap E^c) = 0.334 - 0.119 - 0.104 + 0.048 = 0.159;$$

$$P(T^c \cap H^c \cap E) = 0.346 - 0.151 - 0.104 + 0.048 = 0.139;$$

thus,

$$\begin{aligned} &P(T \cap H^c \cap E^c) + P(T^c \cap H \cap E^c) + P(T^c \cap H^c \cap E) \\ &= 0.248 + 0.159 + 0.139 = 0.546; \end{aligned}$$

c) $P(T^c \cap H^c \cap E^c) = P((T \cup H \cup E)^c) = 1 - P(T \cup H \cup E) = 1 - .824 = .176$, where we used the result in Example 2.31 in the textbook.
d) $P(T \cap H \cap E^c) = P(T \cap H) - P(T \cap H \cap E) = 0.119 - 0.048 = 0.071$;
e) $1 - P(\text{none}) - P(\text{exactly one}) - P(T \cap H \cap E) = 1 - 0.176 - 0.546 - 0.048$
$= 0.23$

2.70 a) Taking $N = 2$ in Proposition 2.10 immediately gives the required result.
b) For mutually exclusive events A and B, $P(A \cap B) = P(\emptyset) = 0$. Then, by the general addition rule,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - 0 = P(A) + P(B),$$

which is the additivity axiom for A and B.
c) If $A_1, \cdots, A_N$ are mutually exclusive, then $A_{k_1} \cap \cdots \cap A_{k_n} = \emptyset$ for all integer $n \geq 2$ and $1 \leq k_1 < \cdots < k_n \leq N$, implying that

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{k=1}^{N} P(A_k) - \sum_{k_1<k_2} 0 + \cdots + (-1)^{n+1} \sum_{k_1<k_2<\cdots<k_n} 0 + \cdots + (-1)^{N+1} 0,$$

and the required conclusion follows.

2.71 The required expression is given by:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup A_4) &= \sum_{i=1}^{4} P(A_i) - P(A_1 \cap A_2) - P(A_1 \cap A_3) \\ &- P(A_1 \cap A_4) - P(A_2 \cap A_3) - P(A_2 \cap A_4) \\ &- P(A_3 \cap A_4) + P(A_1 \cap A_2 \cap A_3) \\ &+ P(A_1 \cap A_2 \cap A_4) + P(A_1 \cap A_3 \cap A_4) \\ &+ P(A_2 \cap A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_3 \cap A_4). \end{aligned}$$

Theory Exercises

2.72 For $N = 2$ the inclusion-exclusion principle holds due to Proposition 2.9 and Exercise 2.70(a). Assume that the principle is valid for all $N \leq m-1$. Then

$$P(\bigcup_{i=1}^{m} A_i) = P((\bigcup_{i=1}^{m-1} A_i) \cup A_m) = P(\bigcup_{i=1}^{m-1} A_i) + P(A_m) - P(A_m \cap (\bigcup_{i=1}^{m-1} A_i)), \tag{2.1}$$

where

$$\begin{aligned} P(\bigcup_{i=1}^{m-1} A_i) &= \sum_{i=1}^{m-1} P(A_i) + \sum_{1 \leq k_1 < k_2 \leq m-1} P(A_{k_1} \cap A_{k_2}) + \cdots + \\ &+ (-1)^{n+1} \sum_{1 \leq k_1 < k_2 < \cdots < k_n \leq m-1} P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_n}) \\ &+ \cdots + (-1)^m P(A_1 \cap A_2 \cap \cdots \cap A_{m-1}), \end{aligned}$$

and

$$\begin{aligned} P(A_m \cap (\bigcup_{i=1}^{m-1} A_i)) &= P(\bigcup_{i=1}^{m-1} (A_i \cap A_m)) \\ &= \sum_{i=1}^{m-1} P(A_i \cap A_m) + \sum_{1 \leq k_1 < k_2 \leq m-1} P(A_{k_1} \cap A_{k_2} \cap A_m) + \cdots + \\ &+ (-1)^{n+1} \sum_{1 \leq k_1 < k_2 < \cdots < k_n \leq m-1} P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_n} \cap A_m) \\ &+ \cdots + (-1)^m P(A_1 \cap A_2 \cap \cdots \cap A_{m-1} \cap A_m). \end{aligned}$$

Upon plugging-in the last two equations into equation (2.1), one obtains the desired inclusion-exclusion principle for $N = m$.

2.73 a) Note that $B_n = A_n \cap A_{n-1}^c = A_n \cap (A_{n-1}^c \cap \cdots \cap A_1^c)$, since $A_1 \subset A_2 \subset \cdots$ implies that $A_{n-1}^c = (\cup_{i=1}^{n-1} A_i)^c = \cap_{i=1}^{n-1} A_i^c$. Therefore, B_n is the event that A_n occurs and none of the events $A_1, \cdots, A_{n-1}$ occur.

b) By part (a) and the solution to Exercise 1.56, the required result follows at once.

c) By the additivity axiom and since $P(A_{n-1} \cap A_n) = P(A_{n-1})$ (which holds by $A_{n-1} \subset A_n$ and Exercise 1.30(c)), it follows that for all $n \geq 2$,

$$P(B_n) = P(A_n \cap A_{n-1}^c) = P(A_n) - P(A_n \cap A_{n-1}) = P(A_n) - P(A_{n-1}).$$

d) Since $B_1, B_2, \cdots$ are mutually exclusive, then, by the (countable) additivity axiom and part (c),

$$\begin{aligned} P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) &= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) \\ &= \lim_{n\to\infty} \left(P(A_1) + \sum_{i=2}^{n} (P(A_i) - P(A_{i-1})) \right) \\ &= \lim_{n\to\infty} (P(A_1) + P(A_n) - P(A_1)) = \lim_{n\to\infty} P(A_n). \end{aligned}$$

e) By parts (b) and (d), it follows that

$$P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \lim_{n\to\infty} P(A_n).$$

f) $A_1 \supset A_2 \supset \cdots$ implies that $A_1^c \subset A_2^c \subset \cdots$, then by the de Morgan's law and Proposition 2.11(a),

$$\begin{aligned} P(\bigcap_{n=1}^{\infty} A_n) = P\left(\left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \right) &= 1 - P(\bigcup_{n=1}^{\infty} A_n^c) = 1 - \lim_{n\to\infty} P(A_n^c) \\ &= \lim_{n\to\infty} (1 - P(A_n^c)) = \lim_{n\to\infty} P(A_n). \end{aligned}$$

2.74 a) $P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) \geq P(A_1) + P(A_2) - 1$, by the general addition rule and since $P(A_1 \cup A_2) \leq 1$.
b) For $N = 1$, the inequality $P(A_1) \geq P(A_1) - (1-1)$ holds. By (a), the Bonferroni's inequality also holds for $N = 2$. Next assume that the inequality is valid for all $N \leq m-1$. Then

$$\begin{aligned} P(\bigcap_{n=1}^{m} A_n) = P(A_m \cap (\bigcap_{n=1}^{m-1} A_n)) &\geq P(A_m) + P(\bigcap_{n=1}^{m-1} A_n) - 1 \\ &\geq P(A_m) + \left(\sum_{n=1}^{m-1} P(A_n) - ((m-1) - 1) \right) - 1 \\ &= \sum_{n=1}^{m} P(A_n) - (m-2) - 1 = \sum_{n=1}^{m} P(A_n) - (m-1), \end{aligned}$$

and the desired result follows by induction.

2.75 a) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$, by the general addition rule and since $P(A_1 \cap A_2) \geq 0$.
b) The result follows by induction, since

$$P(\bigcup_{i=1}^{N} A_i) = P((\bigcup_{i=1}^{N-1} A_i) \cup A_N) \leq P(\bigcup_{i=1}^{N-1} A_i) + P(A_N) \leq (\sum_{i=1}^{N-1} P(A_i)) + P(A_N).$$

c) By Proposition 2.11 and Boole's inequality for a finite number of events,

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{n=1}^{\infty} (\cup_{i=1}^{n} A_i)) = \lim_{n\to\infty} P(\cup_{i=1}^{n} A_i) \le \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

d) By the de Morgan's law and Boole's inequality,

$$P(\bigcap_{n=1}^{N} A_n) = P\left(\left(\bigcup_{n=1}^{N} A_n^c\right)^c\right) = 1 - P(\bigcup_{n=1}^{N} A_n^c)$$

$$\ge 1 - \sum_{n=1}^{N} P(A_n^c) = 1 - \sum_{n=1}^{N}(1 - P(A_n)) = 1 - N + \sum_{n=1}^{N} P(A_n).$$

Advanced Exercises

2.76 a) Let A_n be the event that the first head occurs on the nth toss. Then $A = \bigcup_{n=1}^{\infty} A_n$. By Proposition 2.11,

$$P(A) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} A_i) = \lim_{n\to\infty} (1 - P(\bigcap_{i=1}^{n} A_i^c))$$

$$= 1 - \lim_{n\to\infty} P(\underbrace{T\cdots T}_{n}) = 1 - \lim_{n\to\infty} \frac{1}{2^n} = 1 - 0 = 1.$$

b) $P(\bigcup_{k=0}^{\infty} A_{2k+1}) = \sum_{k=0}^{\infty} P(A_{2k+1}) = \sum_{k=0}^{\infty} \frac{1}{2^{2k+1}} = \frac{1}{2}\sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{1}{2} \times \frac{1}{1-\frac{1}{4}} = \frac{2}{3}.$

c) By parts (a) and (b), the required probability equals to $P(A) - \frac{2}{3} = 1 - \frac{2}{3} = \frac{1}{3}$.

2.77 First Borel-Cantelli lemma: If A^* is the event that infinitely many of the A_ns occur, then, by Exercise 2.18(e),

$$P(A^*) = P(\bigcap_{n=1}^{\infty}(\bigcup_{i=n}^{\infty} A_i)).$$

Let $C_n = \bigcup_{i=n}^{\infty} A_i$, then $C_n \supset C_{n+1}$ for all n, and by Proposition 2.11(b),

$$P(\bigcap_{n=1}^{\infty} C_n) = \lim_{n\to\infty} P(C_n).$$

Since $\sum_{n=1}^{\infty} P(A_n) < \infty$ and by the Boole's inequality, it follows that

$$0 \le P(C_n) = P(\bigcup_{i=n}^{\infty} A_i) \le \sum_{i=n}^{\infty} P(A_i) \longrightarrow 0, \quad \text{as } n \to \infty.$$

Therefore,

$$P(A^*) = P(\bigcap_{n=1}^{\infty} C_n) = \lim_{n\to\infty} P(C_n) = 0.$$

2.78 a) For $N = 2$, the equality holds by the axiom of additivity for two events. Assume that the equality holds for all $N \leq m - 1$. Then

$$P(\bigcup_{n=1}^{m} A_n) = P((\bigcup_{n=1}^{m-1} A_n) \cup A_m) = P(\bigcup_{n=1}^{m-1} A_n) + P(A_m)$$

$$= \sum_{n=1}^{m-1} P(A_n) + P(A_m) = \sum_{n=1}^{m} P(A_n),$$

where we used the fact that if $A_1, A_2, \cdots, A_m$ are mutually exclusive, then the events $(\cup_{i=1}^{m-1} A_i)$ and A_m are mutually exclusive, and the induction assumption.
b) For mutually exclusive $A_1, A_2, \cdots$,

$$P(\bigcup_{n=1}^{\infty} A_n) = P\left(\bigcup_{N=1}^{\infty} (\bigcup_{n=1}^{N} A_n)\right) = \lim_{N\to\infty} P(\bigcup_{n=1}^{N} A_n)$$

$$= \lim_{N\to\infty} \sum_{n=1}^{N} P(A_n) = \sum_{n=1}^{\infty} P(A_n),$$

where the 2nd equality holds by Proposition 2.11(a), while the 3rd equality holds by the additivity axiom for a finite number of events.
c) Clearly, axiom of countable additivity implies the axiom of finite additivity, since $A_1 \cup A_2 = A_1 \cup A_2 \cup \emptyset \cup \emptyset \cup \cdots$, thus,

$$P(A_1 \cup A_2) = P(A_1 \cup A_2 \cup \emptyset \cup \emptyset \cup \cdots) = P(A_1) + P(A_2) + 0 + 0 + \cdots = P(A_1) + P(A_2)$$

for arbitrary mutually exclusive events A_1 and A_2. Taking into account that the axiom of countable additivity implies the continuity property of probability measures (by Proposition 2.11, which was proved in Exercise 2.73), then the axiom of countable additivity implies both the axiom of finite additivity and the continuity. Conversely, the axioms of finite additivity and continuity together imply the countable additivity of probability measures by part (b). Therefore, the axioms of finite additivity and continuity together are equivalent to the axiom of countable additivity.

2.5 Review Exercises

2.79 $\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\}$.

2.80 a)

$$\Omega = \{\{a,b\},\{a,c\},\{a,d\},\{a,e\},\{a,f\},\{b,c\},\{b,d\},$$
$$\{b,e\},\{b,f\},\{c,d\},\{c,e\},\{c,f\},\{d,e\},\{d,f\},\{e,f\}\}.$$

b) $\{\{a,b\},\{a,c\},\{a,d\},\{a,e\},\{a,f\},\{b,f\},\{c,f\},\{d,f\},\{e,f\}\}$.
c) $\Omega = \{\{a,d\},\{a,e\},\{a,f\},\{b,d\},\{b,e\},\{b.f\},\{c,d\},\{c,e\},\{c,f\}\}$.
The event of interest is $\{\{a,d\},\{a,e\},\{a,f\},\{b,f\},\{c,f\}\}$.
d) The sample space is given by:

$$\Omega = \{(a,b),(a,c),(a,d),(a,e),(a,f),\\ (b,a),(b,c),(b,d),(b,e),(b,f),\\ (c,a),(c,b),(c,d),(c,e),(c,f),\\ (d,a),(d,b),(d,c),(d,e),(d,f),\\ (e,a),(e,b),(e,c),(e,d),(e,f)\}.$$

The event of interest is equal to:
$\{(a,d),(d,a),(a,e),(e,a),(a,f),(f,a),(b,f),(f,b),(c,f),(f,c)\}$.

2.81 The cars are classified into 6 categories. Let x_i represent the number of cars (in a given set of 100 cars) that are in category i, where $i = 1, \cdots, 6$.

$$\Omega = \{(x_1, x_2, x_3, x_4, x_5, x_6) : x_i \in \mathcal{N} \cup \{0\} \text{ for } i \in \{1, \cdots, 6\} \text{ and } \sum_{i=1}^{6} x_i = 100\}.$$

2.82 a) Not mutually exclusive, since the intersection of the two given events is not empty (in fact, the intersection contains the following outcomes: 6 men and 6 women; 5 men and 7 women; 7 men and 5 women).
b) Yes, the events are mutually exclusive, since a jury of 12 people cannot have at least 5 men and at least 8 women as members, thus, the intersection of the two events is empty.
c) The three events are mutually exclusive, since the number of men (and women) selected in each case is different, thus, every pairwise intersection is empty.
d) The events are not mutually exclusive, since the intersection of the three events is not empty.

2.83 a) $\Omega = \{(\mathrm{A,B,C}), (\mathrm{A,C,B}), (\mathrm{B,A,C}), (\mathrm{B,C,A}), (\mathrm{C,A,B}), (\mathrm{C,B,A})\}$.
b) $\{(\mathrm{A,B,C}), (\mathrm{A,C,B})\}$.
c) $A^c \cap B^c \cap C^c =$"A is not the first alternate, B is not the second and C is not the third" $=\{(\mathrm{C,A,B}), (\mathrm{B,C,A})\}$.
d) $A \cap B \cap C =$"A is the first alternate, B is the second and C is the third"$=(\mathrm{A,B,C})\}$.
e) $A \cap C=$"A is the first alternate and C is the third"$= \{(\mathrm{A,B,C})\}$.
f) The equal-likelihood model is appropriate, since who is the first alternate, who is the second and who is the third is determined by chance (and the sample space is finite). Since the sample space contains six outcomes, each outcome is assigned probability 1/6.
g) $P(A^c \cap B^c \cap C^c) = 2/6 = 1/3$ and $P(A \cap B \cap C) = P(A \cap C) = 1/6$.

2.84 Answers will vary.

2.85 $A \cup B^c =$"A occurs or B does not occur";
$A^c \cap B^c =$"Neither A nor B occurs";
$(A \cap B^c) \cup (A^c \cap B) =$"Exactly one of the events A,B occurs";
$A \cap B \cap C =$"Events A,B, and C occur";
$A \cap (B \cup C) =$"A occurs and either B or C occurs";
$A \cup (B \cap C) =$"Either A occurs or both B and C occur";
$A \cup B \cup C =$"At least one of the events A, B, C occurs";
$A^c \cup B^c \cup C^c =$"At least one of the events A, B and C does not occur";
$\cap_{n=1}^{\infty} A_n =$"All events $A_1, A_2, \cdots$ occur";
$\cup_{n=1}^{\infty} A_n =$"At least one of $A_1, A_2, \cdots$ occurs."

2.86 $(A \cap B^c) \cup (A^c \cap B)$. (Other equivalent descriptions are also possible, say, $(A \cup B) \cap (A \cap B)^c$).

2.87 a) True, by definition.
b) Not true. The following provides a simple counter-example: Take $\Omega = \{1, 2, 3\}$, $A = \{1, 3\}$, $B = \{2\}$ and $C = \{2, 3\}$. Clearly, A and B are mutually exclusive, but events A, B, C are not mutually exclusive since $A \cap C \neq \emptyset$.

2.88 a) The assignment determines a unique probability measure on Ω by Proposition 2.3, since each outcome is assigned a non-negative probability and the sum of probabilities of all individual outcomes in Ω sum to one.

b) The probabilities of the 16 events of the random experiment are give by:

Event	Probability	Event	Probability
$\emptyset$	0	$\{\mathrm{HT}, \mathrm{TH}\}$	$2p(1-p)$
$\{\mathrm{HH}\}$	p^2	$\{\mathrm{HT}, \mathrm{TT}\}$	$p(1-p) + (1-p)^2 = 1-p$
$\{\mathrm{HT}\}$	$p(1-p)$	$\{\mathrm{TH}, \mathrm{TT}\}$	$p(1-p) + (1-p)^2 = 1-p$
$\{\mathrm{TH}\}$	$p(1-p)$	$\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}\}$	$p + p(1-p) = 2p - p^2$
$\{\mathrm{TT}\}$	$(1-p)^2$	$\{\mathrm{HH}, \mathrm{HT}, \mathrm{TT}\}$	$p + (1-p)^2 = p^2 - p + 1$
$\{\mathrm{HH}, \mathrm{HT}\}$	$p^2 + p(1-p) = p$	$\{\mathrm{HH}, \mathrm{TH}, \mathrm{TT}\}$	$p + (1-p)^2 = p^2 - p + 1$
$\{\mathrm{HH}, \mathrm{TH}\}$	$p^2 + p(1-p) = p$	$\{\mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$	$2p(1-p) + (1-p)^2 = 1 - p^2$
$\{\mathrm{HH}, \mathrm{TT}\}$	$p^2 + (1-p)^2 = 2p^2 - 2p + 1$	Ω	1

2.89 a) The sample space is given by:
$\Omega = \{(w, w, w), (w, w, b), (w, b, w), (b, w, w), (w, b, b), (b, w, b), (b, b, w), (b, b, b)\}$, and each outcome in Ω is assigned probability $1/8$.
b) $P(\{(b, b, b)\}) = 1/8$;
c) $P(\{(w, w, w), (w, w, b), (w, b, w), (b, w, w)\}) = 4/8 = 0.5$;

2.90 a) For $\Omega = \{(x, y) : x^2 + y^2 < 1\}$ (with $|\Omega| = \pi$, where $|\cdot|$ denotes the two-dimensional volume)

$$P(\{(x, y) \in \Omega : \sqrt{x^2 + y^2} > \frac{1}{2}\}) = 1 - P\left(\left\{(x, y) : x^2 + y^2 \le \frac{1}{2^2}\right\}\right)$$

$$= 1 - \frac{\pi(1/2)^2}{\pi} = \frac{3}{4} = 0.75;$$

b) Let C be the cube with vertices at $(0, 1/2), (1/2, 0), (0, -1/2), (-1/2, 0)$. Then $P(\{(x, y) \in \Omega : |x| + |y| > \frac{1}{2}\}) = 1 - \frac{|C|}{\pi} = 1 - \frac{1}{2\pi} \approx 0.8408$
c) $P\left(\{(x, y) \in \Omega : \max\{|x|, |y|\} > \frac{1}{2}\right) = 1 - P(\max\{|x|, |y|\} \le \frac{1}{2})$
$= 1 - P(\{(x, y) : |x| \le \frac{1}{2}, |y| \le \frac{1}{2}\}) = 1 - 1/\pi \approx 0.6817$.

2.91 a) Each $p_n = P(\{n\}) = 1/2^n$ is non-negative and

$$\sum_{n=1}^{\infty} p_n = \sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{1 - (1/2)} - 1 = 2 - 1 = 1,$$

thus the probability assignment is legitimate.
b) $\sum_{k=0}^{\infty} p_{2k+1} = \sum_{k=0}^{\infty} \frac{1}{2^{2k+1}} = \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{1}{2} \times \frac{1}{1 - (1/4)} = \frac{2}{3}$.

2.92 a) Note that $B_{n+1} \subset B_n$ for all $n = 3, 4, \cdots$, implying that $P(B_{n+1}) \leq P(B_n)$ by the domination principle. In particular, it follows that $P(B_6) \leq P(B_5)$.
b) $P(\bigcap_{n=3}^{\infty} B_n) = \lim_{n\to\infty} P(B_n)$, by Proposition 2.11(b) (in the textbook).
c) Let B denote the event that you will never reach your goal. The probability of reaching the goal after just three tosses is equal to $(1/2)^3 = 1/8$; the probability of first reaching the goal after exactly six tosses is equal to $P(\text{THHHHH})+P(\text{HTHHHH})+ P(\text{HHTHHH})=3(1/2)^6 > 0$. Clearly these two events are contained in B^c, thus $P(B^c) \geq (1/8) + 3(1/2)^6 > 1/8$, implying that $P(B) = 1 - P(B^c) < 1 - (1/8) = 7/8$.

2.93 Note that $\Omega = \{(x, y) : 0 \leq x \leq 120, 0 \leq y \leq 120\}$, where x and y denote the arrival times of woman #1 and woman #2 respectively, measured in minutes after 3 P.M.. Then

$$P(\{(x, y) \in [0, 120]^2 : |x - y| \leq 40\}) = 1 - P(\{(x, y) \in [0, 120]^2 : |x - y| > 40\})$$

$$= 1 - \frac{(120 - 40)^2}{120^2} = 1 - \frac{2^2}{3^2} = \frac{5}{9},$$

where we used the fact that $\{(x, y) \in [0, 120]^2 : |x - y| > 40\}$ is the union of two (mutually exclusive) equal right triangles (with legs equal to $120 - 40 = 80$ each).

2.94 Subjective. South Carolina is a large state, with different types of microclimate in different parts of the state, thus, no single mathematical probability model can be realistically constructed for the entire state. The empirical method does not apply here because a particular day is chosen.

2.95 a) Empirical method is most likely here. Manufacturing processes typically involve large numbers of almost identical experiments, and a relative frequency of defective bolts can be used to assign probability of occurrence of a defective bolt.
b) Subjective. Since the event involves a particular race (and a particular horse), it cannot be placed in the context of a repeatable experiment, thus the empirical method cannot be used. Clearly, mathematical probability modeling is also not possible here.
c) Empirical. One can collect information on a large number of races, look at the proportion of races won by the favorite, and use it to assign the probability.
d) The answers will vary.

2.96 a) $P(A \cap H^c) + P(H \cap A^c) = (0.8 - 0.35) + (0.45 - 0.35) = 0.55$
b) $P(A^c \cap H^c) = 1 - P(A \cup H) = 1 - (0.8 + 0.45 - 0.35) = 0.1$

2.97 Total frequencies (in thousands) are given by:
V_1: $151,473$; V_2: $4,461$; V_3: $87,160$; C_1: $209,539$; C_2: $20,391$; C_3: $13,164$; Total: $243,094$.
a) $243,094 - 151,473 = 91,621$;
b) $20,391$;
c) $4,461$;
d) 320;
e) $4,461 + 20,391 - 320 = 24,532$;
f) C_1="A randomly selected vehicle is from the U.S."; V_3="A randomly selected vehicle is a truck"; $C_1 \cap V_3$="A randomly selected vehicle is a truck from the U.S.";
g) $P(C_1) = 209,539/243,094 \approx 0.862$; $P(V_3) = 87,160/243,094 \approx 0.359$;
$P(C_1 \cap V_3) = 75,940/243,094 \approx 0.312$;
h) $P(C_1 \cup V_3) = (129,728 + 3,871 + 75,940 + 6,933 + 4,287)/243,094 \approx 0.908$;

i) $P(C_1 \cup V_3) \approx 0.862 + 0.359 - 0.312 = 0.909$;
j) The joint probability distribution table is given by:

	C_1	C_2	C_3	$P(V_i)$
V_1	0.534	0.054	0.035	0.623
V_2	0.016	0.001	0.001	0.018
V_3	0.312	0.029	0.018	0.359
$P(C_j)$	0.862	0.084	0.054	1

2.98 a) Note that $\Omega = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$, where, say, outcome $(2,3,1)$ denotes that ball #2 was drawn first, ball #3 was drawn second and ball #1 was drawn third. Then

$$P(A_1) = P(\{(1,2,3),(1,3,2)\}) = 2/6 = 1/3,$$

$$P(A_2) = P(\{(1,2,3),(3,2,1)\}) = 2/6 = 1/3,$$

$$P(A_3) = P(\{(1,2,3),(2,1,3)\}) = 2/6 = 1/3.$$

b) A_1, A_2 and A_3 are not mutually exclusive, since $(1,2,3) \in A_1 \cap A_2 \cap A_3$, for example.

2.99 a) $P(\{\text{THHH}, \text{THHT}, \text{HTHH}, \text{HHHH}\}) = 4/(2^4) = 4/16 = 0.25$, where the event in question should be understood as: "The first tail, if there is one at all in four tosses, is followed by (at least) two consecutive heads."
b) $P(\{\text{HHHH},\text{HHHT},\text{THHH}\}) = 3/(2^4) = 3/16 = 0.1875$

2.100 The sample space can be taken to be

$$\Omega = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\},$$

where, for example, $(2,3,1)$ denotes the event that husband #2 (i.e. from the 2nd couple) dances with wife #1 and husband #3 dances with wife #2 and husband #1 dances with wife #3. Then
a) $P(\{(1,2,3)\}) = 1/6$;
b) $P(\{(2,3,1),(3,1,2)\}) = 2/6 = 1/3$;
c) $1 - (1/3) = 2/3$.

2.101 a) $P(W^c \cap D^c \cap I^c) = 1 - P(W \cup D \cup I) = 1 - [0.4 + 0.3 + 0.23 + (-0.15 - 0.13 - 0.09) + 0.05] = 0.39$, where we applied the complementation rule, de Morgan's law and the inclusion-exclusion principle.
b) By the inclusion-exclusion principle,

$$P((W \cap D) \cup (W \cap I) \cup (D \cap I)) = P(W \cap D) + P(W \cap I) + P(D \cap I)$$

$$-3P(W \cap D \cap I) + P(W \cap D \cap I)$$

$$= P(W \cap D) + P(W \cap I) + P(D \cap I) - 2P(W \cap D \cap I)$$

$$= 0.15 + 0.13 + 0.09 - 2(0.05) = 0.27$$

c) $1 - 0.39 - 0.27 = 0.34$, where we used the answers to parts (a),(b).

2.102 Let N be the event that some non-legitimate email arrives during a given hour and let L be the event that some legitimate email arrives during the hour. Then

$$P(N^c \cap L^c) = 1 - P(N \cup L) = 1 - (P(N) + P(L) - P(N \cap L))$$
$$= 1 - (0.5 + 0.7 - 0.4) = 0.2,$$

where we applied the complementation rule, de Morgan's law and the inclusion-exclusion principle.

Theory Exercises

2.103 Finite additivity follows from equation (2.4) (in the textbook) by induction, since for mutually exclusive events $A_1, \cdots, A_n$, if finite additivity holds for unions of $(n-1)$ (or less) events, then

$$P(\bigcup_{i=1}^{n} A_i) = P((\bigcup_{i=1}^{n-1} A_i) \cup A_n) = P(\bigcup_{i=1}^{n-1} A_i) + P(A_n) = \sum_{i=1}^{n-1} P(A_i) + P(A_n).$$

On the other hand, the countable additivity property cannot be obtained from finite additivity.

2.104 a) If $P(A) = P(B) = 0$, then $P(A \cup B) = 0$ by Boole's inequality, since

$$0 \le P(A \cup B) \le P(A) + P(B) = 0 + 0 = 0.$$

b) If $P(A) = P(B) = 1$, then $P(A^c) = P(B^c) = 0$, which, by part (a), implies that

$$P(A \cap B) = 1 - P((A \cap B)^c) = 1 - P(A^c \cup B^c) = 1 - 0 = 1.$$

c) Yes, parts (a) and (b) remain valid for countably many events. To see that (a) holds for countably many events, apply Exercise 2.75(c) (i.e. Boole's inequality for a countably infinite number of events) and proceed as in part (a). To prove part (b) for countably many events, use part (a) for countably many events and the de Morgan's law (Proposition 1.4).

d) For uncountably many events, the results are not valid. Indeed, take $\Omega = (0,1)$, where for each event $E \subset \Omega$, $P(E) = |E|$ (where $|\cdot|$ denotes the one-dimensional volume). Then, if (a) were true for uncountably many events, then one would have that

$$P(\Omega) = P\left(\bigcup_{x \in (0,1)} \{x\}\right) = 0,$$

since $|\{x\}| = 0$ for each point $x \in (0,1)$. But this contradicts $P(\Omega) = 1$. On the other hand, if (b) were true for an uncountable number of events, one would have that

$$P(\emptyset) = P\left(\bigcap_{x \in (0,1)} ((0,1) \setminus \{x\})\right) = 1,$$

since $|(0,1) \setminus \{x\}| = |(0,x)| + |(x,1)| = x + 1 - x = 1$. But then there is a contradiction with $P(\emptyset) = 1 - P(\Omega) = 0$.

Advanced Exercises

2.105 a) 4/52. Note that $\Omega = \{((x,y) : x \in \{1,2,\cdots,52\}, y \in \{1,2,\cdots,52\}\}$, where we enumerated all the cards in the deck starting from aces. Then

$$P(\{(x,y) : x \in \{1,2,\cdots,52\}, y \in \{1,2,3,4\}) = \frac{52 \times 4}{52 \times 52} = \frac{4}{52}.$$

b) 4/52. Note that $\Omega = \{(x,y) : x \in \{1,2,\cdots,52\}, y \in \{1,2,\cdots,52\}, x \neq y\}$, where the cards are enumerated starting from aces. Then

$$P(\{(x,y) : x \in \{1,2,\cdots,52\}, y \in \{1,2,3,4\}, x \neq y\}) = \frac{52 \times 4 - 4}{52 \times 52 - 52} = \frac{4 \times 51}{52 \times 51} = \frac{4}{52}.$$

2.106 a) $A \setminus B$="A occurs but B does not."
b) $P(A \setminus B) = P(A \cap B^c) = P(A) - P(A \cap B)$, by Exercise 1.62(b) and the law of partitions.
c) By Exercise 1.30(c), if $B \subset A$, then $A \cap B = B$, which by (b) implies that

$$P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(B).$$

2.107 a) $A \triangle B$="Exactly one of the events A and B occurs."
b) Using the answer to Exercise 1.63 and the law of partitions, one obtains that

$$P(A \triangle B) = P((A \setminus B) \cup (B \setminus A)) = P(A \setminus B) + P(B \setminus A) = P(A \cap B^c) + P(B \cap A^c)$$
$$= (P(A) - P(A \cap B)) + (P(B) - P(B \cap A)) = P(A) + P(B) - 2P(A \cap B).$$

2.108 a) "All but finitely many of $A_1, A_2, \cdots$ occur."
b) "Infinitely many of $A_1, A_2, \cdots$ occur."
c) The event in part (a) is contained in the event given in (b) by Exercise 1.35(a).

2.109 Let A_i be the event: "5" on ith throw, where $i = 1, 2, \cdots$. Since $P(A_i \cap A_j) \neq 0$ for $i \neq j$, then $P(\cup_{i=1}^{8} A_i) \neq \sum_{i=1}^{8} P(A_i)$. (The result is impossible since probability cannot exceed 1).

2.110 In the notation of Exercise 2.109, $P(A_i) = 1/6$, $P(A_i \cap A_j) = 1/(6^2)$ for $i \neq j$, and $P(A_i \cap A_j \cap A_k) = 1/(6^3)$ for distinct i, j, k. Then

$$P(A_1 \cup A_2 \cup A_3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{6^2} - \frac{1}{6^2} - \frac{1}{6^2} + \frac{1}{6^3} = \frac{91}{216} \approx 0.421$$

2.111 a) $\Omega = \{(x_1, \cdots, x_n) : n \in \mathcal{N},\ x_i \in \{1,2,3,4,5\} \text{ for all } i \neq n,\ x_n = 6\} \bigcup \{(x_1, x_2, \cdots) : x_i \in \{1,2,3,4,5\} \text{ for all } i \in \mathcal{N}\}$.
b) For each $n \in \mathcal{N}$, assign

$$P(\{(x_1, \cdots, x_n)\}) = \frac{5^{n-1}}{6^n}.$$

Note that $\sum_{n=1}^{\infty} P(\{(x_1, \cdots, x_n)\}) = \sum_{n=1}^{\infty} \frac{5^{n-1}}{6^n} = \frac{1}{6} \times \frac{1}{1-(5/6)} = 1$; thus, for all $x_j \in \{1,2,3,4,5\}$, $j \in \mathcal{N}$, assign $P(\{(x_1, x_2, \cdots)\}) = 0$.
c) 1, by solution to (b).
d) $P(\text{Tom}) = \displaystyle\sum_{k=0}^{\infty} \frac{5^{(3k+1)-1}}{6^{3k+1}} = \frac{1}{6} \times \frac{1}{1-(5/6)^3} = \frac{6^2}{6^3 - 5^3} = \frac{36}{91} \approx 0.3956;$

$$P(\text{Dick}) = \sum_{k=0}^{\infty} \frac{5^{(3k+2)-1}}{6^{3k+2}} = \frac{5}{6^2} \times \frac{1}{1-(5/6)^3} = \frac{30}{91} \approx 0.3297;$$

$$P(\text{Harry}) = 1 - \frac{36}{91} - \frac{30}{91} = \frac{25}{91} \approx 0.2747.$$

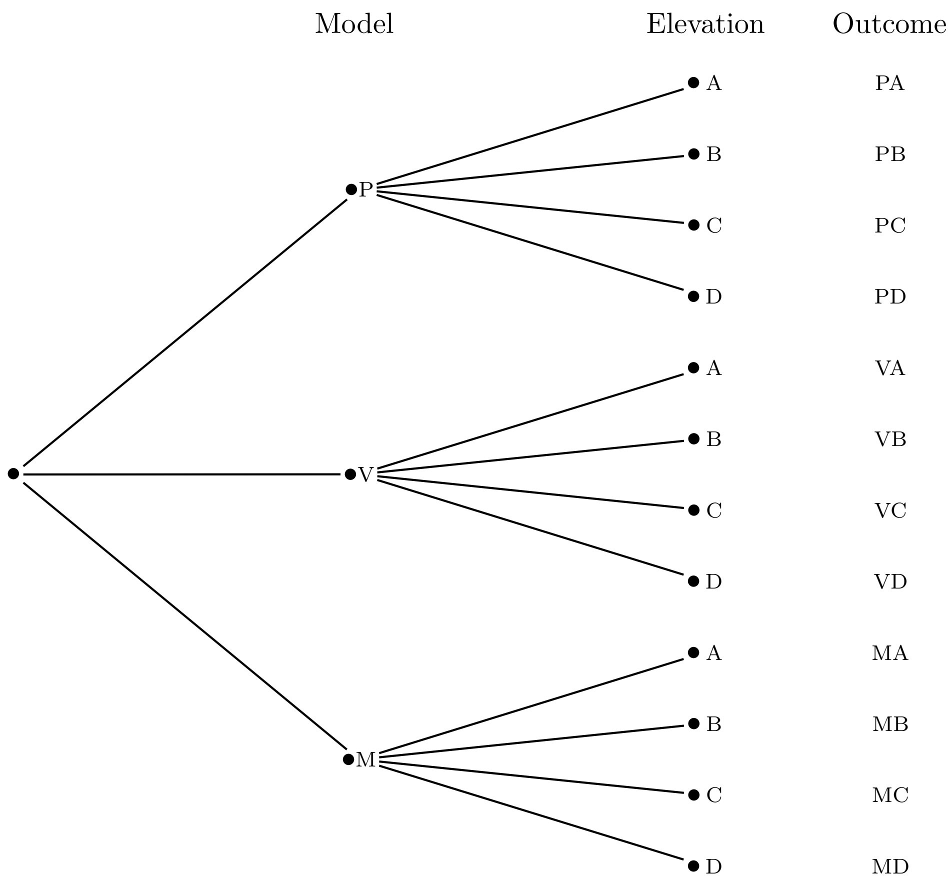
Chapter 3

Combinatorial Probability

3.1 The Basic Counting Rule

Basic Exercises

3.1 a) The tree diagram is given by:



b) 12.
c) $3 \times 4 = 12$.

3.2 $2 \times 4 + 2 \times 3 = 14$.

3.3 $5 \times 3 \times 6 \times 4 = 360$, by the BCR.

3.4 $39 \times 19 \times 8 \times 6 \times 24 \times 5 \times 5 \times 6 \times 5 = 640,224,000$.

3.5 a) 10^8, since there are 10 choices (i.e. $0, 1, \cdots, 9$) for each of the eight digits.
b) $10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 = 1,814,400$.
c) $10 \times 9^7 = 47,829,690$, since there are 10 choices for the 1st digit and 9 choices for every digit starting from the 2nd upto the 8th.
d) $10 \times 9 \times 8^6 = 23,592,960$, since there are 10 choices for the 1st digit, 9 choices for the 2nd and 8 choices for each of the remaining six digits.

3.6 $9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 = 362,880$.

3.7 a) $10^4 \times 8 = 80,000$. **b)** $10^3 \times 9 = 9,000$. **c)** $10^7 \times 8 \times 9 = 80,000 \times 9,000 = 720,000,000$.

3.8 a) $9 \times 10^2 = 900$. **b)** $9 \times 10^6 = 9,000,000$. **c)** $(9 \times 10^2) \times (9 \times 10^6) = 81 \times 10^8 = 8,100,000,000$.

3.9 a) $6^2 = 36$.
b) $4 \times 1 = 4$, since, in order for the sum to be 5, the first die can have 1, 2, 3 or 4 dots facing up (i.e. a total of four choices) whereas the 2nd die must have exactly $(5 - x_1)$ dots facing up, where x_1 is the number rolled on the 1st die.
c) 6.
d) $6 \times 3 = 18$. Note that the sum of the two dice is even if and only if either both dice show even numbers or both dice show odd numbers. Thus, although there are six possibilities for the 1st die, there are only three possibilities for the 2nd die (i.e. the 2nd die must show 1,3 or 5 if the 1st die shows an odd number, and the 2nd die must show 2,4 or 6 if the 1st die shows an even number).

3.10 a) $2^5 = 32$. **b)** 5. **c)** $2^5 - 1 = 31$. **d)** 5+1=6.

3.11 a) $6^4 = 1296$. **b)** $6 \times 5 \times 4 \times 3 = 360$. **c)** $2 \times 6^3 = 432$.
d) $6^4 - 5^4 = 671$, since there are 6^4 possible words when all six letters are used and there are 5^4 possible words when only five letters are used (i.e. when letter c is excluded).

3.12 a) n^n, since each of the n elements in A can be mapped to n possible elements in B.
b) $n \times (n-1) \times \cdots \times 2 \times 1$. Since no two elements from A can be mapped to the same element in B, then while there are n possible choices for the value of the function at the 1st element in A, there are just $(n-1)$ possible choices for the value of the function at the 2nd element in A, thus there are only $(n-2)$ possible choices for the value of the function at the 3rd element in A, etc.

3.13 a) $(5 \times 4 \times 3 \times 2 \times 1)/2 = 60$. Note that the number of arrangements where c is before d is the same as the number of arrangements where c is behind d (due to symmetry). On the other hand, the total number of arrangements of five people in a line is $5 \times 4 \times 3 \times 2 \times 1$.
b) $4 \times 4 \times 3 \times 2 \times 1 = 96$.

3.14 The number of arrangements is equal to $(n-k-1) \times 2 \times ((n-2)!)$ for $k \in \{0, 1, \cdots, n-2\}$, where $(n-2)! = (n-2) \times (n-3) \times \cdots \times 2 \times 1$. Indeed, among the two positions occupied by you and your best friend in the line, let x denote the position closest to the beginning of the line. Since there should be k people standing between you and your best friend, the possible values of x are $1, 2, \cdots, (n-k-1)$. Note that there are two choices regarding who (you or your friend) is standing in position x. Finally, keeping the positions occupied by you and your friend fixed (i.e. positions numbered x, $x+k+1$), there are $(n-2)!$ possibilities for arranging the remaining $(n-2)$ people in the $(n-2)$ available positions.

3.15 a) $4\times 3\times 2\times 1\times 2^4 = 384$, since there are $4\times 3\times 2\times 1$ ways to arrange the order of sitting for four married couples (where each couple is viewed as a single unit), and there are two sitting choices for each of the four couples (i.e. whether wife sits first and her husband sits next or the other way around).
b) $(3\times 2\times 1)\times 2^4 = 96$. First let us view each married couple as a single unit. With respect to couple #1 sitting at the round table, there are 3 choices as to which couple (#2, #3 or #4) is sitting to its right and then there are 2 choices as to which couple is sitting to its left (the remaining fourth couple has to sit in the position at the table opposite to couple #1). In addition, for every given arrangement of couples at the table, there are two choices per each couple as to whether a husband sits to the right from his wife or vice versa. The answer then follows by the BCR.

3.16 $n+\frac{n(n-1)}{2}=\frac{n(n+1)}{2}$, since there are n dominos which have equal numbers on both subrectangles, and there are $n(n-1)$ possible ordered pairs of distinct numbers selected from $\{1,\cdots,n\}$, which results in $n(n-1)/2$ possible dominos with unequal numbers on the two rectangles.

3.17 $n(n-1)/2$. The number of possible handshakes is equal to the number of possible unordered pairs of people chosen from the n people at the party.

3.18 a) $52\times 51\times 50\times 49\times 48 = 311,875,200$.

b) $(52\times 51\times 50\times 49\times 48)/(5\times 4\times 3\times 2\times 1) = 2,598,960$. The answer is obtained upon noting that (# of five-card draws) $\times$ (# of possible arrangements of five cards)= # of five-card studs.

3.19 a) $n\times(n-1)\times\cdots\times 1$. Since none of the n boxes are left empty, then each box contains exactly one ball. Since balls and boxes are all distinguishable, then the problem is equivalent to finding the number of possible (ordered) arrangements of n (different) objects in a line.
b) $n(n-1)\dfrac{n(n-1)}{2}((n-2)!) = (n!)\dfrac{n(n-1)}{2}$. Since exactly one box out of n is to be left empty, then one box contains 0 balls, one box contains two balls and $(n-2)$ boxes contain one ball each. There are $n(n-1)$ ways to choose the pair of boxes, where the first box is designated to contain 0 and the second box is designated to contain 2 balls, respectively. Next there are $n(n-1)/2$ ways to choose the (unordered) set of two balls (from n available balls) that go into the box designated to contain two balls. Finally there are $(n-2)! = (n-2)(n-3)\times\cdots\times 1$ ways to arrange the remaining $(n-2)$ balls between the $n-2$ boxes that contain one ball each. The required answer then follows by the BCR.
c) $n^n-(n\times(n-1)\times\cdots\times 1) = n^n-(n!)$, since, in general, there are n^n ways to assign n different balls to n different boxes (by Exercise 3.12(a)), while there are $n!$ ways to arrange the balls so that no box is empty (by part (a) of the current problem).

Theory Exercises

3.20 To establish the BCR by method of mathematical induction, first note that the BCR was proved for $r=2$ in Proposition 3.1. Next assume that the BCR holds for $r=n$, i.e. if there are m_1 possibilities for the 1st action, and for each possibility of the 1st action there are

m_2 possibilities for the 2nd action, and so on, then there are $m = m_1 \cdot m_2 \cdots m_n$ possibilities altogether for the n actions. Let us establish the BCR for $r = n+1$. Let $f_1, \cdots, f_m$ (where $m = m_1 \cdot m_2 \cdots m_n$) denote all possible outcomes occurring as a result of the first n actions. Let $s_1, \cdots, s_{m_{n+1}}$ be the outcomes of the $(n+1)$st action (taken by itself). Then we can enumerate the possibilities for all $(n+1)$ actions as follows:

$$\begin{array}{cccc} f_1s_1 & f_1s_2 & \ldots & f_1s_{m_{n+1}} \\ f_2s_1 & f_2s_2 & \ldots & f_2s_{m_{n+1}} \\ \vdots & \vdots & \ddots & \vdots \\ f_ms_1 & f_ms_2 & \ldots & f_ms_{m_{n+1}}, \end{array}$$

where f_ks_ℓ represents selecting possibility f_k, occurring as a result of the first n actions, and possibility s_ℓ for the $(n+1)$st action. Each row in the display contains m_{n+1} possibilities and there are m rows, therefore, the total number of possibilities for the $(n+1)$ actions is equal to $m \cdot m_{n+1} = m_1 \ldots m_n \cdot m_{n+1}$. The desired conclusion therefore follows.

Advanced Exercises

3.21 a) $\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}$, thus, the total number of subsets of $\Omega = \{a,b,c\}$ equals 8.
b) $2^3 = 8$. Each subset $A \subset \Omega$ is fully (and uniquely) determined by whether $a \in A$ or $a \notin A$, and by whether $b \in A$ or $b \notin A$, and by whether $c \in A$ or $c \notin A$. Since there are two possibilities for each of the three elements, then the required answer is $2 \times 2 \times 2$.
c) 2^n. Each subset $A \subset \Omega$ is fully (and uniquely) determined by whether each of the n elements of Ω belongs to A or not. Since there are two possibilities for each of the n elements, then the required answer is $\underbrace{2 \times \cdots \times 2}_{n}$.

3.22 a) $2 \times 8 = 2^4 = 16$. For every subset $A \subset \{a,b,c\}$, one can construct two different subsets of $\{a,b,c,d\}$: A and $A \cup \{d\}$. On the other hand, every subset of $\{a,b,c,d\}$ can be constructed in that fashion from the subsets of $\{a,b,c\}$. Therefore, the number of possible subsets of $\{a,b,c,d\}$ is equal to twice the number of possible subsets of $\{a,b,c\}$.
b) Yes. Given an arbitrary subset A of $E = \{\omega_1, \cdots, \omega_n\}$, one can construct two different subsets of $\{\omega_1, \cdots, \omega_n, \omega_{n+1}\}$: A and $A \cup \{\omega_{n+1}\}$. Moreover every subset of $\{\omega_1, \cdots, \omega_n, \omega_{n+1}\}$ can be constructed from subsets of E in that fashion. Therefore, the number of subsets of $\{\omega_1, \cdots, \omega_{n+1}\}$ is equal to twice the number of subsets of $\{\omega_1, \cdots, \omega_n\}$.
c) For any set $\{\omega_1\}$, there are two possible subsets: $\emptyset$ and $\{\omega_1\}$. Thus, the number of subsets of a set, consisting of one element, is equal to $2^1 = 2$. Now suppose that any set of n elements has 2^n subsets. Take an arbitrary set $\{\omega_1, \cdots, \omega_n, \omega_{n+1}\}$ of $(n+1)$ elements. Then the number of possible subsets of $\{\omega_1, \cdots, \omega_n, \omega_{n+1}\}$ is equal to $2 \times (\#\text{ of subsets of } \{\omega_1, \cdots, \omega_n\}) = 2 \times 2^n$, by part (b). Thus, every set of $n+1$ elements has 2^{n+1} subsets. The required conclusion then follows by induction.

3.2 Permutations and Combinations

3.23 For all $k, j \in \mathcal{N}$, with $j < k$,

$$k! = k(k-1)\cdots(k-j+1)\underbrace{(k-j)(k-j-1)\cdots 2 \cdot 1}_{(k-j)!} = k(k-1)\cdots(k-j+1)((k-j)!).$$

For $k = j \in \mathcal{N}$, the following identity obviously holds:

$$k! = k(k-1)\cdots 1 \cdot \underbrace{(0!)}_{=1} = k(k-1)\cdots(k-j+1)((k-j)!).$$

Thus, the required equality holds for all $j \leq k$ $(k, j \in \mathcal{N})$.

3.24 a) $\frac{7!}{(7-3)!} = 210$. **b)** $\frac{5!}{(5-2)!} = 20$. **c)** $\frac{8!}{(8-4)!} = 1680$. **d)** 1. **e)** $9! = 362880$.

3.25 $(18)_3 = \frac{18!}{(18-3)!} = 18 \cdot 17 \cdot 16 = 4896$.

3.26 $(30)_4 = \frac{30!}{(30-4)!} = 30 \cdot 29 \cdot 28 \cdot 27 = 657720$.

3.27 $7! = 5040$.

3.28 a) $\frac{7!}{3!4!} = 35$. **b)** $\frac{5!}{2!3!} = 10$. **c)** $\frac{8!}{4!4!} = 70$. **d)** 1. **e)** 1.

3.29 $\binom{18}{3} = \frac{18!}{3!15!} = 816$.

3.30 $\binom{100}{3} = 161,700$. **b)** $(100)_3 = 970,200$.

3.31 $\binom{42}{5} = 850,668$.

3.32 $\binom{53}{5} \times 42 = 120,526,770$.

3.33 a) $\binom{80}{10} = 1,646,492,110,120$. **b)** $\binom{20}{10} = 184,756$.

3.34 a) $\binom{14}{3} = 364$. **b)** $(14)_3 = 2184$.
c) $(14)_3 - 12 \times (3!) = 2184 - 12 \times 6 = 2112$, since we want to exclude $12 \times (3!)$ ordered selections of the committee where the two specified people will have to serve together (there are 12 possibilities for selection of the third person in addition to the two specified people). Another way of solving the problem is given by:

$$\left(\binom{2}{0}\binom{12}{3} + \binom{2}{1}\binom{12}{2}\right)(3!) = 2112,$$

since $\binom{2}{0}\binom{12}{3}$ represents the number of ways to select an unordered set of 3 people different from the two specified, $\binom{2}{1}\binom{12}{2}$ represents the number of ways to select an unordered set of 3 people of whom exactly one is one of the two specified. The sum has to be multiplied by $3!$ to obtain the number of ordered selections from the number of unordered selections.

3.35 $\binom{6}{3}10^3 \cdot 26^3 = 351,520,000$. Note that there are $\binom{6}{3}$ ways to choose a set of three positions (out of six) that are occupied by letters. Once positions for letters (and, thus, for numbers) are fixed, there are 26^3 arrangements of letters in the three positions occupied by letters and there are 10^3 possible arrangements of numbers in the three positions occupied by numbers.

3.36 a) $10 \times 4 = 40$, since straight flush corresponds to poker hands of the following form: $\{A, 2, 3, 4, 5\}, \{2, 3, 4, 5, 6\}, \cdots, \{10, J, Q, K, A\}$ (with five cards of the same suit), where A stands for ace, J for jack, Q for queen and K for king, thus one has ten different possibilities for each suit and there are four suits.
b) $(13)_2\binom{4}{4}\binom{4}{1} = 624$, since there are $(13)_2 = 13 \cdot 12$ ways to choose an ordered pair (w, x) of denominations from the 13 available denominations and, for that pair of denominations, there are $\binom{4}{4}\binom{4}{1}$ ways to pick the suits of the cards.

c) $(13)_2\binom{4}{3}\binom{4}{2} = 3744$.

d) $4\left(\binom{13}{5} - 10\right) = 5108$. (For each suit, take the number of possible hands of that suit and subtract the number of straight flushes of that suit).

e) $10(4^5 - 4) = 10200$.

f) $\binom{13}{3} \cdot 3 \cdot \binom{4}{3}\binom{4}{1}\binom{4}{1} = 54912$.

g) $\binom{13}{3} \cdot 3 \cdot \binom{4}{2}\binom{4}{2}\binom{4}{1} = 123552$.

h) $\binom{13}{4} \cdot 4 \cdot \binom{4}{2}\binom{4}{1}\binom{4}{1}\binom{4}{1} = 1,098,240$.

3.37 Take the answers obtained in Exercise 3.36 and multiply by 5! to obtain the number of possible hands when the order in which the cards are received matters. Namely, one obtains:
Straight flushes (when order matters): $40(5!) = 4800$;
Four of a kind (when order matters): $624(5!) = 74880$;
Full house (when order matters): $3744(5!) = 449,280$;
Flush (when order matters): $5108(5!) = 612,960$;
Straight (when order matters): $10200(5!) = 1,224,000$;
Three of a kind (when order matters): $54912(5!) = 6,589,440$;
Two pair (when order matters): $123552(5!) = 14,826,240$;
One pair (when order matters): $1098240(5!) = 131,788,800$.

3.38 In Proposition 2.10, each sum of the form $\sum_{1\leq k_1<k_2<\cdots<k_n\leq N} P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_n})$ contains $\binom{N}{n}$ terms.

3.39 a) $\binom{100}{5} = \frac{100\times99\times98\times97\times96}{5!} = 75,287,520$.

b) $\frac{100\times98\times96\times94\times92}{5!} = 67,800,320$. (Equivalently, one could compute $\binom{50}{5}2^5$, since one could pick the five states (out of 50), whose senators serve on the committee, and then pick one senator out of two for each of the five states selected).

3.40 $\binom{3}{2}\binom{2}{0}\binom{3}{1} = 9$.

3.41 a), b) Selection of k objects out of n without replacement is equivalent to dividing n objects into a "desired" and an "undesired" subcategories of sizes k and $(n-k)$, respectively, which is in turn equivalent to selecting $(n-k)$ objects (as "undesired") out of n without replacement. Therefore, $\binom{n}{k} = \binom{n}{k,\, n-k} = \binom{n}{n-k}$.

3.42 a) $\binom{50+80}{25} = \binom{130}{25} \approx 3.855 \times 10^{26}$. **b)** $\binom{50}{13}\binom{80}{12} \approx 2.138 \times 10^{25}$. **c)** $\binom{50}{25} \approx 1.264 \times 10^{14}$.

3.43 $\frac{9!}{4!2!} = 7560$.

3.44 a) $\binom{30}{8,7,4,11} = \frac{30!}{8!7!4!11!} \approx 1.363 \times 10^{15}$.

b) $\binom{15}{8}\binom{9}{7}\binom{6}{4} = 3,474,900$.

3.45 a) $\binom{20}{8}(8)_3 = 125,970 \times 336 = 42,325,920$, by the BCR and since for the 1st step we have: $\binom{20}{8} = 125,970$, while for the 2nd step we have: $(8)_3 = 336$.

b) $\binom{20}{5}(15)_3 = 42,325,920$.

c) $(20)_3\binom{17}{5} = 42,325,920$.

d) All three quantities in the equation provide expressions for computing the number of ways to choose j committee members who serve as (different) officers plus $k-j$ committee members who are not officers, from a list of n candidates. Thus, the three expressions should be equal. Note that the first one corresponds to first choosing all k committee members and then choosing j officers from among them. The second expression corresponds to first selecting $k-j$ committee members who are not officers, and then selecting j people who will serve as committee officers from the remaining pool of $n-(k-j)$ candidates. Finally, the third expression corresponds to first choosing j people who will serve as committee officers, and then choosing the remaining $k-j$ committee members from the $n-j$ candidates left.

e) Note that

$$\binom{n}{k}(k)_j = \frac{n!}{k!(n-k)!}\cdot\frac{k!}{(k-j)!} = \frac{n!}{(k-j)!}\cdot\frac{1}{(n-k)!} \tag{3.1}$$

$$= \frac{n!}{(k-j)!(n-(k-j))!}\cdot\frac{(n-(k-j))!}{(n-k)!} = \binom{n}{k-j}(n-(k-j))_j.$$

Also

$$(n)_j\binom{n-j}{k-j} = \frac{n!}{(n-j)!}\cdot\frac{(n-j)!}{(k-j)!(n-k)!} = \frac{n!}{(k-j)!}\cdot\frac{1}{(n-k)!} = \binom{n}{k}(k)_j,$$

where the last equality follows from (3.1).

Theory Exercises

3.46 For arbitrary $a_1,\cdots,a_n,b_1,\cdots,b_n$, we have that

$$\begin{aligned}(a_1+b_1)(a_2+b_2)\cdots(a_n+b_n) &= \sum_{k_1,\cdots,k_n\in\{0,1\}} a_1^{k_1}\dots a_n^{k_n}b_1^{1-k_1}\dots b_n^{1-k_n}\\ &= \sum_{\ell=0}^{n}\ \sum_{\substack{k_1,\cdots,k_n\in\{0,1\}:\\ k_1+\cdots+k_n=\ell}} a_1^{k_1}\dots a_n^{k_n}b_1^{1-k_1}\dots b_n^{1-k_n}.\end{aligned}$$

Note that there are $\binom{n}{\ell}$ different ways to pick ℓ indices out of n and put corresponding k_j's equal to 1 (while the rest are zeros). Therefore, upon taking $a_1=\cdots=a_n=a$ and $b_1=\cdots=b_n=b$, we obtain that

$$(a+b)^n = \sum_{\ell=0}^{n}\ \sum_{\substack{k_1,\cdots,k_n\in\{0,1\}:\\ k_1+\cdots+k_n=\ell}} a^\ell b^{n-\ell} = \sum_{\ell=0}^{n}\binom{n}{\ell}a^\ell b^{n-\ell}.$$

3.47 a) Note that $(a+b)^0=1$, $(a+b)^1=a+b$, $(a+b)^2=a^2+2ab+b^2$, $(a+b)^3=a^3+3a^2b+3ab^2+b^3$, $(a+b)^4=a^4+4a^3b+6a^2b^2+4ab^3+b^4$, $(a+b)^5=a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$, $(a+b)^6=a^6+6a^5b+15a^4b^2+20a^3b^3+15a^2b^4+6ab^5+b^6$.

b) The Pascal's triangle is given by:

$$\begin{array}{ccccccccccccc} &&&&&&1&&&&&&\\ &&&&&1&&1&&&&&\\ &&&&1&&2&&1&&&&\\ &&&1&&3&&3&&1&&&\\ &&1&&4&&6&&4&&1&&\\ &1&&5&&10&&10&&5&&1&\\ 1&&6&&15&&20&&15&&6&&1 \end{array}$$

c) For all $n, k \in \mathcal{N}$, with $k \le n$,

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

3.48 a) For $n = 1$,

$$\sum_{j=0}^{k}\binom{1}{j}\binom{m}{k-j}=\binom{1}{0}\binom{m}{k}+\binom{1}{1}\binom{m}{k-1}=\binom{m}{k}+\binom{m}{k-1}=\binom{m+1}{k},$$

where the first equality holds since $\binom{1}{j} = 0$ for all $j > 1$, while the last equality holds by Exercise 3.47(c). Thus, the Vandermonde's identity is verified for $n = 1$. Next assume that the Vandermonde's identity is valid for $n = N$. Let us show that the identity is valid for $n = N+1$.

$$\begin{aligned}
\sum_{j=0}^{k}\binom{N+1}{j}\binom{m}{k-j} &= \sum_{j=1}^{k}\Bigg(\underbrace{\binom{N}{j-1}+\binom{N}{j}}_{=\binom{N+1}{j}\text{ by 3.47(c)}}\Bigg)\binom{m}{k-j}+\underbrace{\binom{N+1}{0}}_{=1}\binom{m}{k}\\
&=\sum_{j=1}^{k}\binom{N}{j-1}\binom{m}{k-j}+\sum_{j=1}^{k}\binom{N}{j}\binom{m}{k-j}+\binom{m}{k}\\
&=\sum_{\ell=0}^{k-1}\binom{N}{\ell}\binom{m}{k-1-\ell}+\sum_{j=0}^{k}\binom{N}{j}\binom{m}{k-j}\quad(\text{where } \ell=j-1)\\
&=\binom{N+m}{k-1}+\binom{N+m}{k}=\binom{N+1+m}{k},
\end{aligned}$$

where the last line follows from the Vandermonde's identity for $n = N$ and Exercise 3.47(c).

b) Fix an arbitrary $k \in \{0, \cdots, n+m\}$. Clearly, there are $\binom{n}{j}$ subsets of $\Omega_1 = \{a_1, \cdots, a_n\}$ consisting of j elements, and there are $\binom{m}{k-j}$ subsets of $\Omega_2 = \{b_1, \cdots, b_m\}$ consisting of $k-j$ elements. Since $\Omega_1 \cap \Omega_2 = \emptyset$, there are $\binom{n}{j}\binom{m}{k-j}$ sets of the form $E \cup G$, where E is a subset of Ω_1 of size j and G is a subset of Ω_2 of size $k-j$ (clearly, then $E \cup G$ is a subset of size k of $\Omega_1 \cup \Omega_2$). Since each subset of $\Omega_1 \cup \Omega_2$ consisting of k elements is, for some j ($0 \le j \le k$), the union of a subset of Ω_1 of size j and a subset of Ω_2 of size $k-j$, then the total number of possible subsets of size k of $\Omega_1 \cup \Omega_2$ is equal to $\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j}$. On the other hand, $\Omega_1 \cup \Omega_2$ is a set of $n+m$ elements, thus the number of possible subsets of size k of $\Omega_1 \cup \Omega_2$ is equal to $\binom{n+m}{k}$. Therefore, $\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$.

3.49 First note that

$$(a_1+a_2+\cdots+a_m)^n=\sum_{j_1,\cdots,j_n=1}^{m} a_{j_1}a_{j_2}\ldots a_{j_n}.$$

For all $n_1, \cdots, n_m \in \{0, \cdots, n\}$ such that $n_1 + \cdots + n_m = n$, define

$$J_{n_1,\cdots,n_m} = \{(j_1, \cdots, j_n) : n_k \text{ of the indices } j_1, \cdots, j_n \text{ are equal to } k, \text{ for all } k = 1, \cdots, m\}.$$

Note that the total number of elements in $J_{n_1,\cdots,n_m}$ is equal to $\binom{n}{n_1,\cdots,n_m}$, since it's the total number of ways to distribute n indices into m categories so that n_1 of them are equal to 1, n_2 of them are equal to 2, etc., n_m of them are equal to m. Moreover,

$$\{(j_1, \cdots, j_n) : j_k \in \{1, \cdots, m\} \text{ for all } k = 1, \cdots, n\} = \bigcup_{\substack{n_1,\cdots,n_m \in \mathcal{N}\cup\{0\}: \\ n_1+\cdots+n_m=n}} J_{n_1,\cdots,n_m}.$$

Therefore,

$$\begin{aligned}
\sum_{j_1,\cdots,j_n=1}^{m} a_{j_1} a_{j_2} \dots a_{j_n} &= \sum_{\substack{n_1,\cdots,n_m \in \mathcal{N}\cup\{0\}: \\ n_1+\cdots+n_m=n}} \left(\sum_{(j_1,\cdots,j_n)\in J_{n_1,\cdots,n_m}} a_{j_1} a_{j_2} \dots a_{j_n} \right) \\
&= \sum_{\substack{n_1,\cdots,n_m \in \mathcal{N}\cup\{0\}: \\ n_1+\cdots+n_m=n}} \left(\sum_{(j_1,\cdots,j_n)\in J_{n_1,\cdots,n_m}} a_1^{n_1} a_2^{n_2} \dots a_m^{n_m} \right) \\
&= \sum_{\substack{n_1,\cdots,n_m \in \mathcal{N}\cup\{0\}: \\ n_1+\cdots+n_m=n}} \binom{n}{n_1, \cdots, n_m} a_1^{n_1} a_2^{n_2} \dots a_m^{n_m}.
\end{aligned}$$

Therefore,

$$(a_1 + a_2 + \cdots + a_m)^n = \sum_{\substack{n_1,\cdots,n_m \in \mathcal{N}\cup\{0\}: \\ n_1+\cdots+n_m=n}} \binom{n}{n_1, \cdots, n_m} a_1^{n_1} a_2^{n_2} \dots a_m^{n_m}.$$

Advanced Exercises

3.50 For $k = 0, \cdots, n$, the number of subsets of size k of a set, containing n elements, is equal to $\binom{n}{k}$. Thus, the total number of subsets of a set of n elements is equal to:

$$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k \cdot 1^{n-k} = (1+1)^n = 2^n,$$

by the Binomial theorem.

3.51 To compute the number of distinct positive integer-valued solutions to $x_1 + \cdots + x_r = m$, where $r, m \in \mathcal{N}, r \leq m$, imagine that there are m indistinguishable objects placed in a line and we want to subdivide them into r nonempty groups. Then to every positive integer-valued solution $(x_1, \cdots, x_r)$ of the equation $x_1 + \cdots + x_r = m$, we can associate the distribution of objects of the form:

$$\underbrace{\mathrm{o} \cdots \mathrm{o}}_{x_1} | \underbrace{\mathrm{o} \cdots \mathrm{o}}_{x_2} | \dots | \underbrace{\mathrm{o} \cdots \mathrm{o}}_{x_r},$$

where "o" stands for the object in question, while "|" denotes a line separating adjacent groups. Note that since there are r different groups, then there must be $r-1$ separating lines, and those have to be placed in distinct spaces between the m objects. Note that any such distribution of objects is uniquely determined by the positions of separating lines and vice versa. Since there are $m-1$ spots available (between the m objects) where the $r-1$ separating lines can be placed (with no more than one line put between every pair of objects), then there are $\binom{m-1}{r-1}$ ways to subdivide m indistinguishable objects into r groups of positive integer-valued sizes $(x_1, x_1, \cdots, x_r)$. Moreover, there is a one-to-one correspondence between such subdivisions and the positive integer-valued solutions to the equation $x_1 + \cdots + x_r = m$. Therefore, the total number of positive integer-valued solutions to $x_1 + \cdots + x_r = m$ is equal to $\binom{m-1}{r-1}$.

3.52 For $r, m \in \mathcal{N}$, let $n_0(r, m)$ denote the number of *non-negative* integer-valued solutions to the equation $x_1+\cdots+x_r = m$ and let $n(r, m)$ be the number of *positive* integer-valued solutions to the equation $x_1 + \cdots + x_r = m$. Then

$$n_0(r, m) = \sum_{k=0}^{r-1} \binom{r}{k} n(r-k, m),$$

since, if $(x_1, \cdots, x_{r-k})$ is a positive integer-valued solution to $x_1 + \cdots + x_{r-k} = m$, then $(x_1, \cdots, x_{r-k}, \underbrace{0, \cdots, 0}_{k})$ is a non-negative integer-valued solution to $x_1+\cdots+x_r = m$, and there are $\binom{r}{k}$ different ways to pick which k of the r possible x_i's are zeros, where $k = 0, 1, \cdots, r-1$. By Exercise 3.51, $n(r-k, m) = \binom{m-1}{r-k-1}$. Therefore,

$$n_0(r, m) = \sum_{k=0}^{r-1} \binom{r}{k} \binom{m-1}{r-1-k} = \binom{r+m-1}{r-1},$$

where the last equality holds by the Vandermonde's identity (Exercise 3.48).

An even simpler solution is obtained by noting that the number of non-negative integer-valued solutions of $x_1 + \cdots + x_r = m$ is the same as the number of positive integer-valued solutions of $y_1 + \cdots + y_r = m + r$ (seen by letting $y_i = x_i + 1,\ i = 1, \cdots, r$). Thus, the answer is $\binom{m+r-1}{r-1}$, by Exercise 3.51.

3.53 a) $\binom{10+4-1}{4-1} = \binom{13}{3} = 286$, by Exercise 3.52. **b)** $\binom{10-1}{4-1} = \binom{9}{3} = 84$, by Exercise 3.51.

3.54 $\binom{100+6-1}{6-1} = \binom{105}{5} = 96,560,646$. Refer to the solution to Exercise 2.81 and note that the number of outcomes in Ω is equal to the number of non-negative integer-valued solutions to the equation: $x_1+x_2+x_3+x_4+x_5+x_6 = 100$. The required answer then follows by Exercise 3.52.

3.55 a) If the balls are all distinguishable and some urns are allowed to stay empty, then the number of possible outcomes is equal to N^n, since each of the n balls can be placed in any of the N (distinguishable) urns.
b) If the balls are indistinguishable and some urns are allowed to stay empty, then the number of possible outcomes coincides with the number of non-negative integer-valued solutions to the equation $x_1 + \cdots + x_N = n$ (where x_k corresponds to the number of balls in urn #k, $k = 1, \cdots, N$), and is equal to $\binom{n+N-1}{N-1}$ by Exercise 3.52.

3.3 Applications of Counting Rules to Probability

3.56 Let X be the number of people (out of 20) who are cured. Note that the outcomes in the sample space Ω are equally likely. Then, according to the 1st method, (if $N(E)$ denotes the number of outcomes in E)

$$P(X \geq 15) = \frac{N(\{X \geq 15\})}{N(\Omega)} = \frac{\binom{20}{15} + \binom{20}{16} + \binom{20}{17} + \binom{20}{18} + \binom{20}{19} + \binom{20}{20}}{2^{20}} = \frac{21,700}{1,048,576} \approx 0.021,$$

where the numerator is obtained by computing the number of ways to pick a set of k people, who are cured, from among 20 (which is $\binom{20}{k}$) and then summing the results over $k = 15, \cdots, 20$. The 2nd method gives:

$$P(X \geq 15) = \sum_{k=15}^{20} P(X = k) = \sum_{k=15}^{20} \binom{20}{k} \frac{1}{2^{20}} \approx 0.021$$

3.57 a) $\dfrac{13}{\binom{52}{4}} = \dfrac{1}{20,825} \approx 0.000048$, since there are 13 possibilities for the choice of the denomination and then there is just one way to pick 4 cards of that denomination, while $\binom{52}{4}$ is the number of ways to pick a set of four cards out of 52 and represents the total number of possible outcomes in the sample space.

b) $\dfrac{\binom{13}{4}4^4}{\binom{52}{4}} \approx 0.6761$, since there are $\binom{13}{4}$ ways to pick four different denominations (out of a total of 13), and then there are 4 ways in which one can pick specific suits for each selected denomination. On the other hand, $\binom{52}{4}$ represents the total number of ways to select a set of four different cards from the deck of 52.

3.58 $1 - \dfrac{\binom{8}{2}}{\binom{10}{2}} = 1 - \dfrac{28}{45} = \dfrac{17}{45} \approx 0.378$, where we used the complementation rule and the fact that there are $\binom{8}{2}$ ways to pick two winning tickets from among the eight that are not yours.

3.59 a) $1/13 \approx 0.0769$, since, by symmetry of the problem, the 7th card is equally likely to be of any one of the 13 denominations.

b) $\dfrac{(48)_6 \times 4}{(52)_7} \approx 0.0524$, since there are $(48)_6$ ways to select the first six cards from the 48 "non-aces" (taking into account the order), and for each choice of those six cards there are 4 possibilities for the ace chosen next; on the other hand, $(52)_7$ represents the total number of possible ordered selections of seven cards from the deck of 52, and is the number of outcomes in the sample space.

3.60 $\dfrac{\binom{N-M}{j}\binom{M}{n-j}}{\binom{N}{n}}.$

3.61 a) $1 - P(\text{no multiple birthdays}) = 1 - \dfrac{(365)_{38}}{365^{38}} \approx 0.8641;$

b) $1 - P(\text{no multiple birthdays}) = 1 - \dfrac{(365)_N}{365^N}.$

c) The required probabilities for $N = 1, \cdots, 70$ are given in the following table:

N	1	2	3	4	5	6	7	8	9	10
Prob.	0	0.003	0.008	0.016	0.027	0.040	0.056	0.074	0.095	0.117

N	11	12	13	14	15	16	17	18	19	20
Prob.	0.141	0.167	0.194	0.223	0.253	0.284	0.315	0.347	0.379	0.411

N	21	22	23	24	25	26	27	28	29	30
Prob.	0.444	0.476	0.507	0.538	0.569	0.598	0.627	0.654	0.681	0.706

N	31	32	33	34	35	36	37	38	39	40
Prob.	0.730	0.753	0.775	0.795	0.814	0.832	0.849	0.864	0.878	0.891

N	41	42	43	44	45	46	47	48	49	50
Prob.	0.903	0.914	0.924	0.933	0.941	0.948	0.955	0.961	0.966	0.970

N	51	52	53	54	55	56	57	58	59	60
Prob.	0.974	0.978	0.981	0.984	0.986	0.988	0.990	0.992	0.993	0.994

N	61	62	63	64	65	66	67	68	69	70
Prob.	0.995	0.996	0.997	0.997	0.998	0.998	0.998	0.999	0.999	0.999

d) $N = 23$, by using the table in (c).

3.62 a) $\dfrac{\binom{6}{2}(6)_4(4)_3}{(10)_7} = \dfrac{3}{14} \approx 0.2143$; Note that a total of seven balls are drawn if and only if there are two red balls among the first six balls selected and the seventh ball is red. Let Ω consist of all ordered selections of seven balls chosen without replacement from a set of 10 (distinguishable) balls (where, say, ball #1 through ball #4 are red, while the rest are black). Then $N(\Omega) = (10)_7$. On the other hand, there are $\binom{6}{2}$ ways to decide which two of the first six balls selected are red. Moreover, for a given order in which red and black balls appear, there are $(4)_3$ ways to pick the specific balls for the red ones, and there are $(6)_4$ ways to pick the specific balls for the black balls.

b) $\dfrac{\binom{6}{2}6^4 4^3}{10^7} \approx 0.1244$; Here Ω is defined as in (a) but the selection is with replacement, resulting in $N(\Omega) = 10^7$. As in (a), there are $\binom{6}{2}$ ways to decide which two of the first six balls selected are red, but now, for a given order in which red and black balls are chosen, there are 4^3 ways to pick the specific red balls and 6^4 ways to pick the specific black ones.

3.63 $\dfrac{3!4!3!5!}{12!} = \dfrac{1}{4620} \approx 0.000216$, since there are 3! ways of permuting the disciplines, 4! ways of permuting mathematicians among themselves, 3! ways of permuting chemists among themselves and 5! ways of permuting physicists among themselves.

3.64 $\dfrac{\binom{k}{k}\binom{N-k}{n-k}}{\binom{N}{n}} = \dfrac{\binom{N-k}{n-k}}{\binom{N}{n}}.$

3.65 a) $\dfrac{(N)_n}{N^n} = \dfrac{N(N-1)\cdots(N-(n-1))}{N^n}.$

b) For each $n \in \mathcal{N}$,

$$\frac{(N)_n}{N^n} = \frac{N(N-1)\cdots(N-(n-1))}{N^n} = \frac{N}{N} \times \frac{N-1}{N} \times \cdots \times \frac{N-(n-1)}{N}$$
$$= 1 \times \left(1 - \frac{1}{N}\right) \times \cdots \times \left(1 - \frac{n-1}{N}\right) \longrightarrow 1, \text{ as } N \to \infty.$$

Thus, when a random sample is selected (with replacement) and the population size is large relative to the sample size, any member of the population is unlikely to be chosen more than once.

3.66 a) $\dfrac{\binom{6}{1}\binom{94}{4}}{\binom{100}{5}} \approx 0.243$;

b) $\dfrac{\binom{6}{0}\binom{94}{5}}{\binom{100}{5}} + \dfrac{\binom{6}{1}\binom{94}{4}}{\binom{100}{5}} \approx 0.972$;

c) $1 - \dfrac{\binom{6}{0}\binom{94}{5}}{\binom{100}{5}} \approx 0.271$

3.67 a) $P(E) = \dfrac{N(E)}{N(\Omega)} = \dfrac{\binom{1}{1}\binom{N-1}{n-1}n!}{(N)_n} = \dfrac{n}{N}$, since there are $\binom{1}{1}\binom{N-1}{n-1}$ ways to select an unordered sample of size n (without replacement) containing the specified member, implying that there are $n!\binom{1}{1}\binom{N-1}{n-1}$ such ordered samples; on the other hand, the sample space consists of all ordered samples of size n taken (without replacement) from a population of size N, implying that $N(\Omega) = (N)_n$.
b) By symmetry of the problem, since all the members of the population are equally likely to be selected, the kth member selected is equally likely to be any one of the N possible members of the population. Thus, $P(A_k) = 1/N$. Since $A_1, A_2, \cdots, A_n$ are mutually exclusive and $E = \cup_{k=1}^{n} A_k$, then

$$P(E) = \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n} \frac{1}{N} = \frac{n}{N}.$$

3.68 a) By Example 3.21(c), for $E =$"At least one woman (among N) gets her own key," the required probability is given by:

$$P(E) = \sum_{n=1}^{N} \frac{(-1)^{n+1}}{n!}.$$

Thus, for $N = 1, \cdots, 6$, the desired probabilities are given by:

N	1	2	3	4	5	6
$P(E)$	1	0.5	$2/3 \approx 0.66667$	$5/8 = 0.625$	$19/30 \approx 0.63333$	$91/144 \approx 0.63194$

b) Since $1 - e^{-1} \approx 0.63212$, it is easy to see that $P(E)$ is larger than $1 - e^{-1}$ for odd values of N and $P(E)$ is smaller than $1 - e^{-1}$ for even values but $|1 - e^{-1} - P(E)|$ decreases monotonously as N increases.
c) The approximation error at $N = 6$ does not exceed 0.0002, and one can show that it decreases further quite rapidly as N grows larger. Thus,the approximation is extremely good for $N \geq 6$.

3.69 a) 10!=3,628,800. **b)** $1/(10!) \approx 2.756 \times 10^{-7}$.

3.70 a) $\dfrac{(N-1)(N-2)\cdots(N-(n-1))\cdot 1}{N(N-1)\cdots(N-(n-1))} = \dfrac{(N-1)_{n-1}}{(N)_n} = \dfrac{1}{N}$.
b) Arrange the keys in a line in the order in which they will be tried. Then the correct key is equally likely to be in any one of the N positions, thus, the required probability is $1/N$.
c) (i) $1 - P(\text{first } n \text{ keys are incorrect}) = 1 - \dfrac{(N-1)_n}{(N)_n} = 1 - \dfrac{N-n}{N} = \dfrac{n}{N}$.
(ii) Let A_k be the event that the correct key is found on the kth try, $k = 1, \cdots, n$. Then $A_1, \cdots, A_n$ are mutually exclusive and $P(A_k) = 1/N$, $k = 1, \cdots, n$. Therefore, the required probability is given by:

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n} \frac{1}{N} = \frac{n}{N}.$$

3.71 a) $\dfrac{(N-1)^{n-1}\cdot 1}{N^n} = \dfrac{1}{N}\left(1-\dfrac{1}{N}\right)^{n-1}$, since $(N-1)^{n-1}$ is the number of ways to pick the wrong keys in the first $(n-1)$ attempts, while N^n is the total number of ways to select n keys out of N with replacement, when the order of selection matters.
b) (i) $1 - P(\text{first } n \text{ keys are incorrect}) = 1 - \dfrac{(N-1)^n}{N^n} = 1 - \left(1-\dfrac{1}{N}\right)^n$.
(ii) Let A_k be the event that the correct key is found for the first time on the kth try, $k = 1, \cdots, n$. Then $A_1, \cdots, A_n$ are mutually exclusive and $P(A_k) = \frac{1}{N}(1-\frac{1}{N})^{k-1}$, $k = 1, \cdots, n$. Then the required probability is given by:

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n} \frac{1}{N}\left(1-\frac{1}{N}\right)^{k-1} = \frac{1}{N}\sum_{\ell=0}^{n-1}\left(1-\frac{1}{N}\right)^{\ell}$$

$$= \frac{1}{N}\cdot\frac{1-(1-\frac{1}{N})^n}{1-(1-\frac{1}{N})} = 1 - \left(1-\frac{1}{N}\right)^n.$$

3.72 a) $1 - \dfrac{1}{2^{15}} \approx 0.99997$;
b) Let X be the number of questions that the student answers correctly. Then

$$P(X \geq 9) = \sum_{k=9}^{15} P(X=k) = \sum_{k=9}^{15} \frac{\binom{15}{k}}{2^{15}} = \frac{9949}{2^{15}} \approx 0.3036$$

3.73 a) Take the answers to Exercise 3.36 and divide by $\binom{52}{5} = 2598960$ to obtain the required probabilities. Namely,

$$P(\text{straight flush}) = \frac{40}{\binom{52}{5}} \approx 0.0000154, \quad P(\text{four of a kind}) = \frac{624}{\binom{52}{5}} \approx 0.00024,$$

$$P(\text{full house}) = \frac{3744}{\binom{52}{5}} \approx 0.00144, \quad P(\text{flush}) = \frac{5108}{\binom{52}{5}} \approx 0.001965,$$

$$P(\text{straight}) = \frac{10200}{\binom{52}{5}} \approx 0.00392, \quad P(\text{three of a kind}) = \frac{54912}{\binom{52}{5}} \approx 0.02113,$$

$$P(\text{two pair}) = \frac{123552}{\binom{52}{5}} \approx 0.04754, \quad P(\text{one pair}) = \frac{1098240}{\binom{52}{5}} \approx 0.422557$$

b) No, the answers will be the same. The number of five-card poker studs, counted in each event, is equal to the number of respective five-card draw poker hands times 5!. On the other hand, the total number of possible five-card poker studs is equal to the 5! times the total number of possible five-card draw poker hands. The same factor 5!, that enters both the numerator and denominator (when we use equation (3.4) in the text), will cancel out.

3.74 $\dfrac{\binom{14}{10}\binom{6}{0} + \binom{14}{4}\binom{6}{6}}{\binom{20}{10}} = \dfrac{2002}{184756} \approx 0.0108$, since one of the daughter cells consists of all normal subunits if and only if either daughter cell #1 consists of 10 normal subunits (and 0 mutant subunits) or it consists of 4 normal subunits and 6 mutant subunits.

3.75 $\dfrac{n(n-1)\binom{n}{2}((n-2)!)}{n^n} = \dfrac{n!\binom{n}{2}}{n^n}$. Exactly one box stays empty if and only if exactly one of the n boxes contains two balls, $n-2$ boxes contain one ball each and one of the boxes contains 0 balls. Note that there are n ways to choose which box stays empty, then there are $n-1$ choices of the box that contains two balls and there are $\binom{n}{2}$ ways to fill it. Next there are $(n-2)!$ ways to distribute the remaining $(n-2)$ balls between the remaining $n-2$ boxes that must contain one ball each. On the other hand, the total number of outcomes in the sample space is equal to the number of ways to put n balls into n boxes (where some boxes could stay empty). Since each of the n balls can be put into any of the n boxes, then $N(\Omega) = n^n$ and the required conclusion follows. (Note that throughout this solution all the boxes and balls are taken to be distinguishable).

3.76 $\dfrac{(n-k-1) \times 2 \times ((n-2)!)}{n!} = \dfrac{2(n-k-1)}{n(n-1)}$. Take the answer to Exercise 3.14 and divide it by $n!$.

Theory Exercises

3.77 Note that the number of members (in the entire population), who have the specified attribute, is equal to Np. Therefore, we have that:

a) $\dfrac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$, since there are $\binom{Np}{k}$ ways to pick k members of the sample from among the Np members of the population who have the specified attribute, and there are $\binom{N-Np}{n-k} = \binom{N(1-p)}{n-k}$ ways to pick the remaining $n-k$ members of the sample from the $N - Np$ members of the population who lack the specified attribute, while $\binom{N}{n}$ represents the total number of ways to pick an (unordered) sample of size n without replacement from the population of size N.

b) $\dfrac{\binom{n}{k}(Np)^k(N(1-p))^{n-k}}{N^n} = \dbinom{n}{k} p^k (1-p)^{n-k}$. If we consider ordered selection with replacement, then there are N^n different samples possible. On the other hand, there are $\binom{n}{k}$ possible ways to pick locations of the k members (who have the specified attribute) in the sample of size n, and, given any one such choice, there are $(Np)^k(N-Np)^{n-k}$ ways to pick the specific members.

c) The proof is the same as in the solution to Exercise 5.70.

d) If the population size N is large relative to the sample size n, then, when sampling is done *without replacement*, the probability of the event, that exactly k members with the specified attribute are selected, can be well approximated by the corresponding probability for sampling *with replacement.*

Advanced Exercises

3.78 $\frac{1}{n!}\left(1-\sum_{k=1}^{N-n}\frac{(-1)^{k+1}}{k!}\right)$. Note that exactly n of the N women get their own keys if and only if n women get their own keys while the other $N-n$ women get the wrong keys. There are $\binom{N}{n}$ ways to fix the set of those n women who get their own keys. Once that set is fixed, let us consider the remaining $N-n$ women and their $N-n$ keys. Applying the result of Example 3.21(c), we obtain that

$$P\left(\text{none of the remaining } (N-n) \text{ women gets her own key}\right) = 1-\sum_{k=1}^{N-n}\frac{(-1)^{k+1}}{k!}.$$

Note also that the above probability, that none of the remaining $(N-n)$ women gets her own key, is also equal to

$$\frac{\#\text{ of ways to arrange } (N-n) \text{ keys among } (N-n) \text{ women so that none has her own}}{(N-n)!},$$

thus, the number of ways to arrange $(N-n)$ remaining keys among the $(N-n)$ remaining women in such a way that none has her own key is equal to

$$(N-n)!\left(1-\sum_{k=1}^{N-n}\frac{(-1)^{k+1}}{k!}\right).$$

Therefore, the total number of ways, in which exactly n of the N women get their own keys, is equal to

$$\binom{N}{n}(N-n)!\left(1-\sum_{k=1}^{N-n}\frac{(-1)^{k+1}}{k!}\right).$$

Finally, note that there are $N!$ possible ways to arrange all N keys among the N women. Therefore, the required probability is equal to

$$\frac{\binom{N}{n}(N-n)!\left(1-\sum_{k=1}^{N-n}\frac{(-1)^{k+1}}{k!}\right)}{N!} = \frac{1-\sum_{k=1}^{N-n}\frac{(-1)^{k+1}}{k!}}{n!}.$$

3.79 Since the number of the U.S. adults is very large relative to the sample size $n = 10$, then, by Exercise 3.65, when sampling is done with replacement, any member of the population is unlikely to be selected more than once. Therefore, in the current problem we can approximate the desired answers by considering sampling with replacement instead of the sampling without replacement. Therefore,

a) $\binom{10}{5}\frac{1}{2^{10}} = \frac{252}{2^{10}} \approx 0.2461$

b) $\binom{10}{8}\frac{1}{2^{10}} + \binom{10}{9}\frac{1}{2^{10}} + \binom{10}{10}\frac{1}{2^{10}} = \frac{56}{2^{10}} \approx 0.0547$

c) The probabilities obtained in (a) and (b) are only approximately correct since we replaced sampling without replacement with sampling with replacement.

d) For even N,

$$P(\text{exactly five are Democrats}) = \frac{\binom{0.5N}{5}\binom{0.5N}{5}}{\binom{N}{10}},$$

$$P(\text{8 or more are Democrats}) = \frac{\binom{0.5N}{8}\binom{0.5N}{2} + \binom{0.5N}{9}\binom{0.5N}{1} + \binom{0.5N}{10}}{\binom{N}{10}}.$$

(If N is not even, then $0.5N$ has to be replaced with the appropriate closest larger and smaller integers).

3.80 a) 1/52. We need to calculate how many of the 52! possible orderings of the cards have the ace of spades immediately following the first ace. Note that each ordering of the 52 cards can be obtained by first ordering the 51 cards different from the ace of spades and then inserting the ace of space into that ordering. There are 51! possible orderings of the 51 cards different from the ace of spades. Moreover, for each such ordering, there is only one way in which the ace of spades can be inserted immediately after the first ace. Therefore,

$$P(\text{the ace of spades follows the first ace}) = \frac{51!}{52!} = \frac{1}{52}.$$

b) 1/52, by the same argument as in (a).

c) 4/52. Let A_1 be the event that the ace of spades follows right after the first ace, A_2 be the event that the ace of diamonds follows right after the first ace, A_3 be the event that the ace of hearts follows right after the first ace, and A_4 be the event that the ace of clubs follows right after the first ace. Then, clearly, events A_1, A_2, A_3, A_4 are mutually exclusive and

$$P(\bigcup_{i=1}^{4} A_i) = \sum_{i=1}^{4} P(A_i) = \sum_{i=1}^{4} \frac{1}{52} = \frac{4}{52}.$$

d) 4/52, by an argument similar to the one given in (c).

3.81 a) Infinitely many, since at noon the box contains all the $\mathcal{N}$-numbered balls except for those numbered $10k$, $k = 1, 2, \ldots$.

b) 0, since for every $k \in \mathcal{N}$, the ball $\#k$ is removed from the box at $\frac{1}{2^{k-1}}$ minute to noon.

c) The number of basketballs left in the box at noon is equal to 0 with probability 1. Indeed, let E_n be the event that ball number 1 is still in the box after the first n withdrawals. Then

$$P(E_n) = \frac{9 \times 18 \times 27 \times \cdots \times (9n)}{10 \times 19 \times 28 \times \cdots \times (9n+1)}.$$

The event that ball number 1 is still in the box at noon is equal to $\bigcap_{n=1}^{\infty} E_n$. Note that events E_n are decreasing (i.e. $E_n \supset E_{n+1}$ for each $n \in \mathcal{N}$), thus, by the continuity property of probability measure,

$$P(\text{ball \#1 is still in the box at noon}) = P(\bigcap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} P(E_n) = \prod_{n=1}^{\infty} \frac{9n}{9n+1}.$$

Let us show that $\displaystyle\prod_{n=1}^{\infty} \frac{9n}{9n+1} = 0$ or, equivalently, that $\displaystyle\left[\prod_{n=1}^{\infty} \frac{9n}{9n+1}\right]^{-1} = \prod_{n=1}^{\infty} \frac{9n+1}{9n} = \infty.$ Note that for all $m \geq 1$,

$$\prod_{n=1}^{\infty} \frac{9n+1}{9n} = \prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) \geq \prod_{n=1}^{m}\left(1+\frac{1}{9n}\right) > \frac{1}{9}+\frac{1}{18}+\frac{1}{27}+\cdots+\frac{1}{9m} = \frac{1}{9}\sum_{i=1}^{m}\frac{1}{i}.$$

Taking $m \to \infty$, and, in view of $\sum_{i=1}^{\infty} \frac{1}{i} = \infty$, it follows that $\prod_{n=1}^{\infty} \frac{9n+1}{9n} = \infty$. Therefore,

$$P(\text{ball \#1 is still in the box at noon}) = 0.$$

Similarly, one shows that for arbitrary $k \in \mathcal{N}$,

$$P(\text{ball \#}k \text{ is still in the box at noon}) = 0.$$

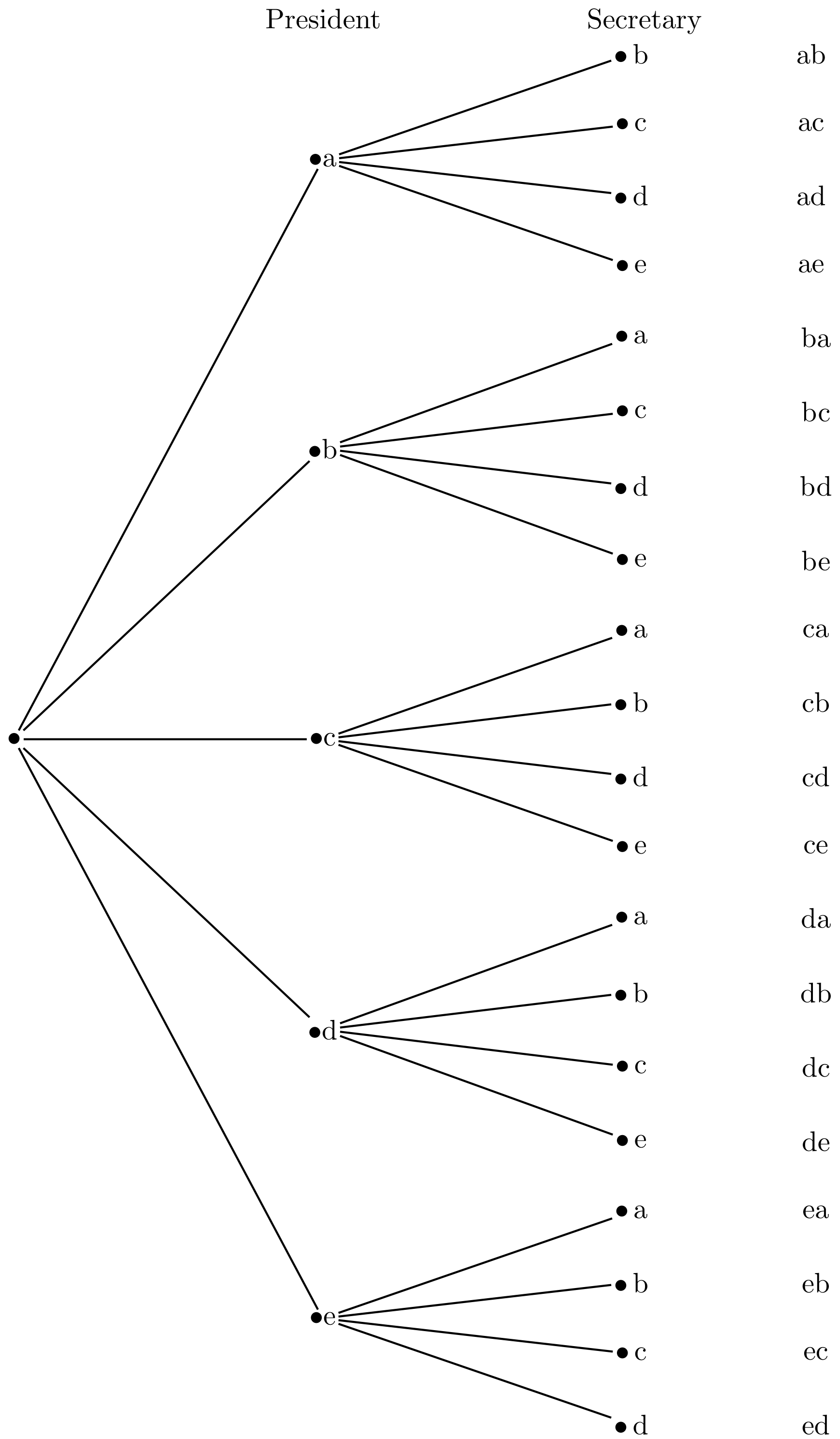
Therefore, by the Boole's inequality (Exercise 2.75),

$$\begin{aligned} P(\text{box is not empty at noon}) &= P\left(\bigcup_{k=1}^{\infty}\{\text{ball \#}k \text{ is in the box at noon}\}\right) \\ &\leq \sum_{k=1}^{\infty} P(\text{ball \#}k \text{ is in the box at noon}) = \sum_{k=1}^{\infty} 0 = 0. \end{aligned}$$

3.4 Review Exercises

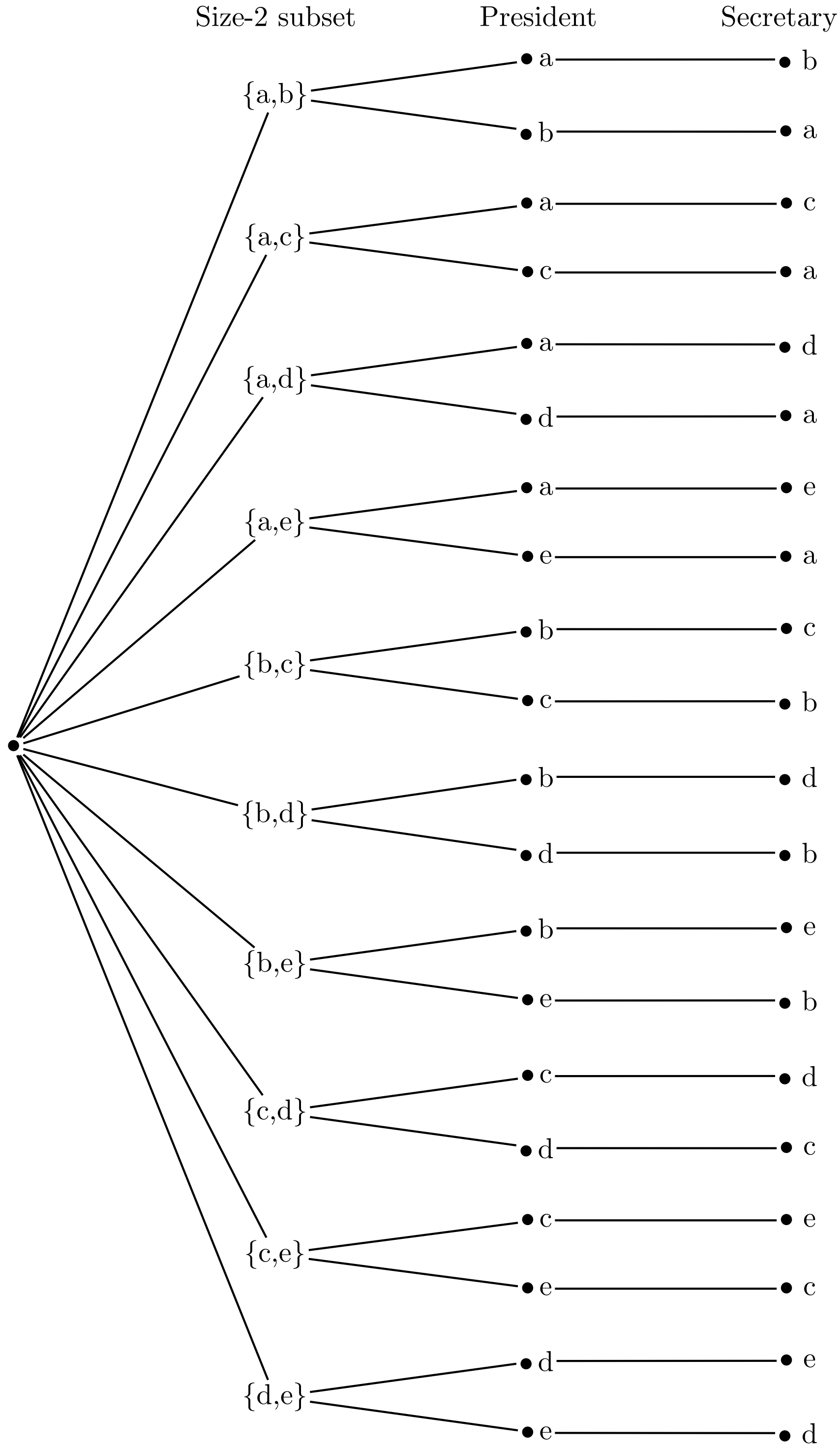
Basic Exercises

3.82 a) The required tree diagram is of the form:

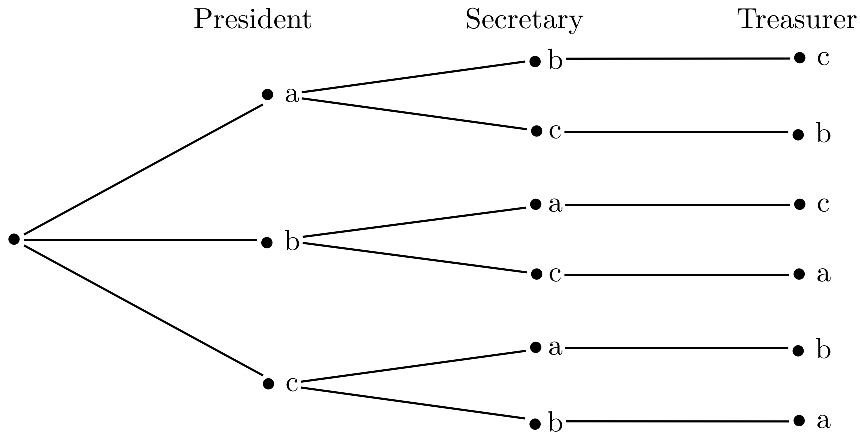


b) The required tree diagram is the same as in (a), but one should switch the words "President" and "Secretary".

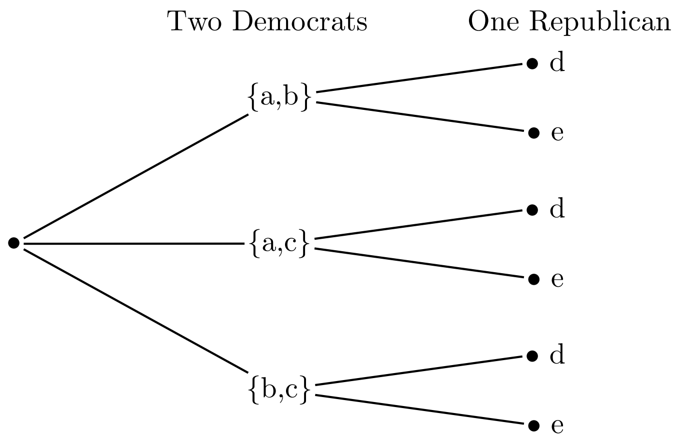
c) The required tree diagram is given by:



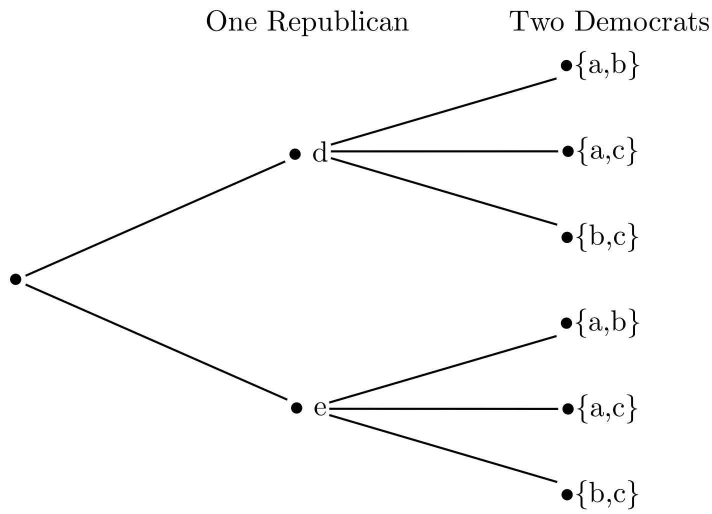
d) The required tree diagram is given by:



e) The first required diagram has the form:



The second required diagram is given by:



3.83 a) $(10)_6 = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151,200.$
b) $10^6 = 1,000,000.$
c) $\frac{(10)_6}{10^6} = 0.1512$

3.84 $6 \times 8 \times 5 \times 19 \times 16 \times 14 = 1,021,440.$

3.85 $(15)_3 = 2730.$

3.86 $\binom{45}{6} = 8,145,060.$

3.87 a) $8 \times 3 \times (10)_3 = 17280.$
b) $8 \times 3 \times \binom{10}{3} = 2880.$

3.88 $\binom{50}{5,10,10,25} \approx 1.241 \times 10^{24}$

3.89 $\binom{100}{10,7,15,8,60} \approx 1.163 \times 10^{49}$

3.90 a) $\binom{12}{2} = 66.$
b) $(12)_3 = 12 \times 11 \times 10 = 1,320.$

3.91 $\dfrac{8!}{2!3!2!} = 1680.$

3.92 a) $\binom{10}{3,3,4} = 4,200.$

b) $3!\binom{10}{3,3,4} = 25,200.$

c) $3!3!\binom{10}{3,3,4} = 151,200.$

3.93 a) $\dfrac{11!}{4!4!2!} = 34,650.$
b) $\dfrac{13!}{2!4!2!} = 64,864,800.$

3.94 $1 - \dfrac{\binom{4}{0}\binom{48}{5}}{\binom{52}{5}} \approx 0.341$

3.95 $\dfrac{\binom{4}{2}\binom{48}{11}}{\binom{52}{13}} \approx 0.2135$

b) $\dfrac{(4)_3\binom{13}{8}\binom{13}{4}\binom{13}{1}}{\binom{52}{13}} \approx 0.0004521$

c) $\dfrac{\frac{4!}{2!}\binom{13}{5}\binom{13}{5}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}} \approx 0.03174$

d) $\dfrac{\binom{39}{13}}{\binom{52}{13}} = 0.0128$

e) 0.0511, since let A_1 denote the event that a bridge hand is void in hearts, A_2 be the event that a bridge hand is void in diamonds, A_3 be the event that a bridge hand is void in spades, and let A_4 be the event that a bridge hand is void in clubs. Then, by the inclusion-exclusion formula,

$$P\Big(\bigcup_{i=1}^4 A_i\Big) = \sum_{i=1}^4 P(A_i) - \sum_{1\le i<j\le 4} P(A_i\cap A_j) + \sum_{1\le i<j<k\le 4} P(A_i\cap A_j\cap A_k) - P\Big(\bigcap_{i=1}^4 A_i\Big),$$

where $P(A_i) = \frac{\binom{39}{13}}{\binom{52}{13}}$, $P(A_i\cap A_j) = \frac{\binom{26}{13}}{\binom{52}{13}}$, $P(A_i\cap A_j\cap A_k) = \frac{\binom{13}{13}}{\binom{52}{13}}$ for all $i,j,k \in \{1,\cdots,4\}$, where i,j,k are all distinct. Also, $P\Big(\bigcap_{i=1}^4 A_i\Big) = 0$. Therefore, upon applying the result of Exercise 3.38, we obtain that

$$P\Big(\bigcup_{i=1}^4 A_i\Big) = \binom{4}{1}\cdot\frac{\dbinom{39}{13}}{\dbinom{52}{13}} - \binom{4}{2}\cdot\frac{\dbinom{26}{13}}{\dbinom{52}{13}} + \binom{4}{3}\cdot\frac{\dbinom{13}{13}}{\dbinom{52}{13}} - \binom{4}{4}\cdot 0 \approx 0.0511$$

3.96 a) $\dfrac{\dbinom{4}{1,1,1,1}\dbinom{48}{12,12,12,12}}{\dbinom{52}{13,13,13,13}} \approx 0.1055$

b) $\dfrac{\binom{4}{3}\binom{48}{23}}{\binom{52}{26}} \approx 0.2497$

3.97 $1-P(\text{first ball is red})-P(\text{1st ball is black but 2nd ball is red}) = 1-\dfrac{10}{26}-\dfrac{16\times 10}{26\times 25}\approx 0.3692$

3.98 a) $\left(1-\dfrac{6}{36}\right)/2 = 5/12 \approx 0.4167$, since the events "green die shows a larger number than the red die" and "red die shows a larger number than the green die" have the same probability (by symmetry), while the complement of their union, i.e. the event "both dice show the same number" has probability 6/36.

b) $\dfrac{\binom{3}{2}6\times 5+\binom{3}{3}6}{6^3} = 4/9.$

3.99 Let x_1 and x_2 be the values on the first and second selected chips, respectively. Then
a) 4/45. For sampling without replacement, the event of interest is given by

$$\{(x_1,x_2)\in\{1,\cdots,10\}^2 : x_1+x_2=10, x_1\neq x_2\} = \{(x_1,x_2)\in\{1,\cdots,10\}^2 : x_1+x_2=10\}\setminus\{(5,5)\}.$$

Then the number of elements in the event of interest is equal to the number of positive integer-valued solutions of the equation $x_1+x_2=10$ minus one. Thus, upon using Exercise 3.51, we obtain that the number of elements in the event of interest equals to $\binom{10-1}{2-1}-1=8$. Therefore, since $10\times 9=90$ is the total number of elements in the sample space, the required probability

equals to $8/90 = 4/45 \approx 0.0889$
b) 0.09, since there are 10^2 elements in the sample space, while the required event,

$$\{(x_1, x_2) \in \{1, \cdots, 10\}^2 : x_1 + x_2 = 10\},$$

by Exercise 3.51, contains 9 elements.

3.100 a) $\dfrac{\binom{10}{1}\binom{20}{2}\binom{30}{3}}{\binom{60}{6}} = \dfrac{1}{6.49} \approx 0.1541$

b) $\dfrac{10 \times 20^2 \times 30^3 \times \frac{6!}{2!3!}}{60^6} \approx 0.1389$, since there are $6!/(2!3!)$ different color arrangements possible for the ordered selection of six balls, of which 2 are white and 3 are blue (and 1 is red); moreover, for every such arrangement of colors, there are $10 \times 20^2 \times 30^3$ possible choices of balls.

3.101 $\dfrac{\binom{k-1}{M-1}}{\binom{N}{M}}$, since if the kth item inspected is the last defective, then there are $M-1$ defective items among the first $k-1$ inspected, and, thus, there are $\binom{k-1}{M-1}$ ways to pick which of the first $k-1$ items are defective. On the other hand, the total number of ways to pick which of the N items are defective is equal to $\binom{N}{M}$, thus, the required answer follows.

3.102 a) $\dfrac{\binom{6}{6}}{\binom{42}{6}} \approx 0.00000019$

b) $\dfrac{\binom{6}{4}\binom{36}{2}}{\binom{42}{6}} \approx 0.0018$

c) $\dfrac{\binom{6}{0}\binom{36}{6} + \binom{6}{1}\binom{36}{5} + \binom{6}{2}\binom{36}{4}}{\binom{42}{6}} \approx 0.9709$

d) $0.00000019N$, by (a) and assuming that the N tickets are all different.

3.103 $1 - 5/8 = 3/8 = 0.375$, since, by the matching problem, i.e. Example 3.21(c), with $N = 4$, we have that

$$P(\text{at least one man sits across from his wife}) = \sum_{n=1}^{4} \frac{(-1)^{n+1}}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} = \frac{5}{8}.$$

3.104 $\dfrac{\binom{8}{6}6!}{8^6} = \dfrac{(8)_6}{8^6} \approx 0.0769$, since there are $\binom{8}{6}$ ways to select the six floors where someone gets off, and, for every such choice, there are 6! ways to pick the order in which the six people leave the elevator.

3.105 $\dfrac{\binom{k}{k}\binom{N-k}{n-k}}{\binom{N}{n}} = \dfrac{\binom{N-k}{n-k}}{\binom{N}{n}}$. (Same as Exercise 3.64).

3.106 a) $\dfrac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$. Note that a total of n members of the population are marked, and the two samples have exactly k members in common if in the sample of size m exactly k members are marked. Thus, one has to count the number of ways to choose k members out of n marked members times the number of ways to choose $m-k$ remaining members of the second sample out of $N-n$ "unmarked" members, and then divide by the total number of ways to pick a sample of size m from the population of size N.

b) $\dfrac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$. Note that now a total of m members of the population will be marked, and the two samples have k members in common if k members of the first sample are selected from those m that are later marked, while the remaining $n-k$ members of the first sample are selected from those $N-m$ members that are left "unmarked".

c) The equality holds since, by (a),(b), both expressions represent the probability that exactly k members of the population are selected both times.

d)

$$\frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}} = \frac{\frac{n!}{k!(n-k)!} \times \frac{(N-n)!}{(m-k)!(N-n-(m-k))!}}{\frac{N!}{m!(N-m)!}} = \frac{m!n!(N-n)!(N-m)!}{N!k!(n-k)!(m-k)!(N-n-(m-k))!}$$

$$= \frac{m!}{k!(m-k)!} \times \frac{(N-m)!}{(n-k)!(N-m-(n-k))!} \times \frac{n!(N-n)!}{N!} = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

3.107 a) 6^n. **b)** 5^{n-1}. **c)** $\dfrac{5^{n-1}}{6^n}$.

Advanced Exercises

3.108 Let E_k denote the event that the first 6 occurs on the kth toss, $k \in \mathcal{N}$.
Then $P(E_k) = \frac{5^{k-1}}{6^k}$, since the first $k-1$ tosses must result in one of the five outcomes $1, \cdots, 5$, while the kth toss has to be a 6. Then

a)

$$P\Bigl(\bigcup_{k=1}^{\infty} E_k\Bigr) = \sum_{k=1}^{\infty} \frac{5^{k-1}}{6^k} = \frac{1}{6}\sum_{\ell=0}^{\infty}\left(\frac{5}{6}\right)^{\ell} = \frac{1}{6} \times \frac{1}{1-\frac{5}{6}} = 1.$$

b)

$$P(\text{first 6 occurs on an odd–numbered toss}) = P(\bigcup_{k=0}^{\infty} E_{2k+1}) = \sum_{k=0}^{\infty} \frac{5^{2k}}{6^{2k+1}} = \frac{1}{6} \times \frac{1}{1-(\frac{5}{6})^2} = \frac{6}{11},$$

$$P(\text{first 6 occurs on an even–numbered toss}) = P(\bigcup_{k=1}^{\infty} E_{2k}) = 1 - \frac{6}{11} = \frac{5}{11}.$$

3.109 a) $\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$. The answer follows by the same argument as in Exercise 3.106.

b) Assuming that the proportion of tagged deer in the sample of n recaptured animals is the same as in the entire population of N deer, one arrives at the equation:

$$\frac{M}{N} = \frac{k}{n},$$

thus, $N = \frac{Mn}{k} = \frac{10\times 8}{3} \approx 27$.

3.110 a) Suppose $0! = x$, then $1 = 1! = (0+1)! = (0+1)\cdot(0!) = 0! = x$, unless $(n+1)! = (n+1)\cdot n!$ is allowed to be violated for $n = 0$.
b) Suppose that when you press the CLEAR key, the display reads x. Then, when you press "14, ENTER", the display should show the product of x and 14, then, when you subsequently press "3, ENTER", the display should show the product of $14x$ and 3. Finally, when we subsequently press "6, ENTER", the display should show the product of $14x \times 3$ and 6, i.e. the final result should equal to $x \times 14 \times 3 \times 6 = 252x$. Since the display shows 252, then $252x = 252$, implying that $x = 1$. Thus, multiplication of no numbers gives 1. Since 9^0 represents the multiplication of 9 zero times, then $9^0 = 1$.
c) Since the number of ways to choose three men and no women must be the same as the number of ways to choose three men, then $\binom{5}{3}\binom{2}{0} = \binom{5}{3}$, implying that $10\binom{2}{0} = 10$, thus $\binom{2}{0} = 1$. Therefore, by the combinations rule, $\binom{2}{0} = \frac{2!}{0!2!} = \frac{1}{0!} = 1$, which implies that $0! = 1$.

3.111 a) $\binom{2N}{N}$, since every polygonal line is uniquely determined by the locations of N people who have \$10 bills in a line of length $2N$.
b) Each polygonal line should terminate at the point $(2N, 0)$, since, overall, the line should travel from the origin N steps up and N steps down (regardless of the order in which those steps are made).
c) Event E^c ="Someone has to wait for change" occurs if and only if there are more people with \$20 than those with \$10 among the first k people in line for some $k = 1, \cdots, 2N$. The latter is equivalent to the fact that the corresponding polygonal line travels more steps up than down at some point k in time, which is equivalent to the line going above the x-axis. Therefore, event E occurs if and only if the polygonal line does not go above the x-axis.
d) $\binom{2N}{N+1} = \frac{(2N)!}{(N+1)!(N-1)!}$. Let us call our polygonal line *good* if it never goes above the x-axis, otherwise let us call it *bad*. In order to count the number of bad polygonal lines, note that for each bad polygonal line one can construct a polygonal path that starts at $(0, 0)$ and ends at $(2N, 2)$, by reflecting the portion of the original polygonal line to the right of the first bad point with respect to the horizontal line $y = 1$. Conversely, to any polygonal path that starts at $(0, 0)$ and ends at $(2N, 2)$ there corresponds a bad polygonal line of the original form. Thus, the total number of bad polygonal lines equals to the number of possible polygonal paths that start at $(0, 0)$ and end at $(2N, 2)$, and that number is equal to $\binom{2N}{N+1}$, since the path should climb $N+1$ steps up (out of $2N$) while taking the remaining $N-1$ steps down, in an arbitrary order.
e) By (a),(c),(d),

$$P(E^c) = P(\text{random polygonal line is bad}) = \frac{\binom{2N}{N+1}}{\binom{2N}{N}} = \frac{(2N)!}{(N+1)!(N-1)!} \cdot \frac{N!N!}{(2N)!} = \frac{N}{N+1}.$$

Thus,

$$P(E) = 1 - P(E^c) = 1 - \frac{N}{N+1} = \frac{1}{N+1}.$$

BETA

Chapter 4

Conditional Probability and Independence

4.1 Conditional Probability

Basic Exercises

4.1 The answers will vary. For example, consider an experiment where a balanced die is rolled twice. Let A be the event that 6 occurs on the 1st roll, and let B be the event that 5 occurs on the 2nd roll. Then $P(A) = \frac{6}{36} = \frac{1}{6}$ and $P(A|B) = \frac{1}{6}$. Thus, $P(A|B) = P(A)$.

4.2 a) Fix an arbitrary event $E \subset \Omega$. Since $P(E|\Omega)$ and $P(E)$ both compute the probability of event E given the same sample space Ω, then one must have $P(E \,|\, \Omega) = P(E)$.
b) For all events $E \subset \Omega$, $P(E \,|\, \Omega) = \frac{P(E\cap\Omega)}{P(\Omega)} = \frac{P(E)}{1} = P(E)$.
c) By (a), one can view unconditional probabilities as conditional, where the event on which one conditions is the sample space Ω.

4.3 a) $4/52 = 1/13$, i.e. a king is selected with probability 1/13.
b) $4/12 = 1/3$, i.e. given that a face card is selected, it is a king with probability 1/3.
c) 1/13, i.e. given that a heart is selected, it is a king with probability 1/13.
d) 0, i.e. given that a non-face card is selected, the probability that it is a king is 0.
e) 12/52=3/13, i.e. the probability of selecting a face card is 3/13.
f) 1, i.e. given that a king is selected, the probability that a face card is selected equals 1.
g) 3/13, i.e. given that a heart is selected, the probability that it is a face card equals 3/13.
h) 8/48=1/6, i.e. given that the card chosen is not a king, the probability that a face card is selected equals to 1/6.

4.4 a) $$\frac{23,468}{471+1,470+11,715+23,468+24,476+21,327+13,782+15,647} = \frac{23,468}{112,356} \approx 0.2089$$
b) $$\frac{23,468}{112,356-471} = \frac{23,468}{111,885} \approx 0.20975$$
c) $$\frac{1,470+11,715+23,468}{111,885} = \frac{36,653}{111,885} \approx 0.3276$$
d) Approximately 20.89% of all U.S. housing units have exactly four rooms. Among the U.S. housing units consisting of at least two rooms, 20.975% have exactly four rooms. Among the

U.S. housing units with at least two rooms, 32.76% have two, three or four rooms.

4.5 a) $\dfrac{320}{1164} \approx 0.2749$

b) $\dfrac{36}{253} \approx 0.1423$

c) Part (a): Approximately 27.49% of all faculty members are Assistant Professors. Part (b): Approximately 14.23% of the faculty members in their 50s are Assistant Professors.

d) The table describing the conditional distribution of rank by age and the marginal probability distribution of rank is given by:

		Age					
		< 30	$30-39$	$40-49$	$50-59$	$60+$	$P(R_i)$
Rank	Full professor (R_1)	$\frac{2}{68}$	$\frac{52}{402}$	$\frac{156}{348}$	$\frac{145}{253}$	$\frac{75}{93}$	$\frac{430}{1164}$
	Associate professor (R_2)	$\frac{3}{68}$	$\frac{170}{402}$	$\frac{125}{348}$	$\frac{68}{253}$	$\frac{15}{93}$	$\frac{381}{1164}$
	Assistant professor (R_3)	$\frac{57}{68}$	$\frac{163}{402}$	$\frac{61}{348}$	$\frac{36}{253}$	$\frac{3}{93}$	$\frac{320}{1164}$
	Instructor (R_4)	$\frac{6}{68}$	$\frac{17}{402}$	$\frac{6}{348}$	$\frac{4}{253}$	0	$\frac{33}{1164}$
	Total	1	1	1	1	1	1

4.6 a) $42/94 \approx 0.4468$; **b)** $19/94 \approx 0.2021$; **c)** $11/19 \approx 0.5789$; **d)** $11/42 \approx 0.2619$;

e) Approximately 44.68% of players on the New England Patriots roster are rookies. Approximately 20.21% of players on the New England Patriots roster weigh more than 300 pounds. 57.89% of rookies on the team weigh more than 300 pounds. Approximately 26.19% of the players, who weigh over 300 pounds, are rookies.

4.7 a) Probability distribution of weight for rookies is given by:

Weight	$P(W_i)$
Under 200 (W_1)	$9/42 \approx 0.2143$
$200-300$ (W_2)	$22/42 \approx 0.5238$
Over 300 (W_3)	$11/42 \approx 0.2619$

b) Probability distribution of years of experience for players whose weight is over 300 pounds is given by:

Years of Experience	$P(Y_i)$
Rookie (Y_1)	$11/19 \approx 0.579$
$1-5$ (Y_2)	$7/19 \approx 0.368$
$6-10$ (Y_3)	$1/19 \approx 0.053$
$10+$ (Y_4)	0

c) Conditional distribution of weight by years of experience and the marginal distribution of weight are given in the following table:

		Years of Experience				
		Rookie	$1-5$	$6-10$	$10+$	$P(W_i)$
Weight	Under 200 (W_1)	9/42	9/41	2/8	1/3	21/94
	$200-300$ (W_2)	22/42	25/41	5/8	2/3	54/94
	Over 300 (W_3)	11/42	7/41	1/8	0	19/94
	Total	1	1	1	1	1

d) Conditional probability distribution of years of experience by weight and the marginal probability distribution of years of experience are given by:

		Weight			
		Under 200	$200-300$	Over 300	$P(Y_i)$
Years of Experience	Rookie (Y_1)	9/21	22/54	11/19	42/94
	$1-5$ (Y_2)	9/21	25/54	7/19	41/94
	$6-10$ (Y_3)	2/21	5/54	1/19	8/94
	$10+$ (Y_4)	1/21	2/54	0	3/94
	Total	1	1	1	1

4.8 a) 0.529; **b)** 0.153; **c)** 0.084; **d)** $0.084/0.153 \approx 0.549$; **e)** $0.084/0.529 \approx 0.1588$;
f) 52.9% of all U.S. citizens 15 years of age or older live with spouse. 15.3% of all U.S. citizens 15 years of age or older are over 64. 8.4% of all U.S. citizens 15 years of age or older live with spouse and are older than 64. Among the U.S. citizens who are over 64, 54.9% live with spouse. Among the U.S. citizens who are 15 years of age or older and live with spouse, 15.88% are over 64 years old.
g) Conditional probability distributions of age by living arrangement and the marginal probability distribution of age are given by:

		Living Arrangement			
		Alone	With spouse	With others	$P(A_i)$
Age	$15-24$ (A_1)	6/126	17/529	154/345	0.177
	$25-44$ (A_2)	38/126	241/529	122/345	0.401
	$45-64$ (A_3)	35/126	187/529	47/345	0.269
	Over 64 (A_4)	47/126	84/529	22/345	0.153
	Total	1	1	1	1

h) Conditional probability distributions of living arrangement by age and the marginal probability distribution of living arrangement are given by:

		Age				
		$15-24$	$25-44$	$45-64$	Over 64	$P(L_i)$
Living Arrangement	Alone (L_1)	6/177	38/401	35/269	47/153	0.126
	With spouse (L_2)	17/177	241/401	187/269	84/153	0.529
	With others (L_3)	154/177	122/401	47/269	22/153	0.345
	Total	1	1	1	1	1

4.9 a) $\dfrac{13 \times 48 \times 44 \times 40}{(13)_4 \times 4^4} = \dfrac{1}{4}$, since if we consider the ordered selection of the four cards, then there are 13 spades to choose from for the 1st card, after that there are 48 cards (whose denomination is different from the denomination of the 1st selected card) to choose from for the 2nd card, then (similarly) there are 44 cards to choose from for the 3rd card and finally there are 40 cards to choose from for the 4th card. On the other hand, the event that four cards selected (taking into account the order) have different denominations consists of $(13)_4 \times 4^4$ outcomes, since there are $(13)_4$ ways to fix the four different denominations for the 1st, 2nd, 3rd and 4th cards selected and for every such choice, there are 4^4 ways to pick the specific cards.

b) $\dfrac{12 \times 39 \times 26 \times 13}{52 \times 39 \times 26 \times 13} = \dfrac{12}{52} = \dfrac{3}{13}$, since there are 12 face cards in the deck, and with every subsequent selection of the cards, there are 13 cards less to choose from, since the suits must be all different.

4.10 0.141; Let X be the number of times that 3 occurs (in 12 tosses). Then

$$P(X \geq 4 \,|\, X \geq 1) = \frac{P((X \geq 4) \cap (X \geq 1))}{P(X \geq 1)} = \frac{P(X \geq 4)}{1 - P(X = 0)},$$

where

$$P(X \geq 4) = 1 - \sum_{k=0}^{3} P(X = k) = 1 - \left(\frac{5^{12}}{6^{12}} + \frac{\binom{12}{1} 5^{11}}{6^{12}} + \frac{\binom{12}{2} 5^{10}}{6^{12}} + \frac{\binom{12}{3} 5^{9}}{6^{12}} \right) \approx 0.1252,$$

and $1 - P(X = 0) = 1 - \frac{5^{12}}{6^{12}} \approx 0.8878$, implying that $P(X \geq 4 \,|\, X \geq 1) = \frac{0.1252}{0.8878} \approx 0.141$

4.11 Assume that boys and girls are equally likely to be born. Then
a) 1/2. **b)** 1/3.

4.12 $\dfrac{\binom{7}{2} \times 4^5}{5^7} \approx 0.27525$; Let us fix the attention on the first 8 tosses. Then there are 5^7 outcomes where all first seven tosses are different from 6. On the other hand, there are $\binom{7}{2}$ ways to pick which of the first seven tosses result in 4, and there are 4^5 ways to pick the values for the five tosses that are different from 4 and 6.

4.13 38.776%, since $0.019/0.049 = 19/49 \approx 0.38776$, by the conditional probability rule.

4.14 a) $\dfrac{1^4 \times 2^4}{2^8} = \dfrac{1}{16}$.
b) $\dfrac{4}{\binom{8}{3}} = \dfrac{1}{14}$.
c) $1 - \dfrac{1}{14} = \dfrac{13}{14}$, using (b), since conditional probability is a probability measure and, thus, the complementation rule is applicable.

4.15 a) 1/2. **b)** 1/2.
c) 3/4, since the required conditional probability is equal to the ratio of the area of a trapezoid with vertices at the points $(1/2, 0)$, $(1, 0)$, $(1, 1)$ and $(1/2, 1/2)$ to the area of the rectangle with vertices at the points $(1/2, 0)$, $(1, 0)$, $(1, 1)$ and $(1/2, 1)$.

4.16 For any set $A \subset \mathcal{R}^2$, let $|A|$ denote the area of A. Then
a)

$$\frac{\big|\{(x, y) : x > 1/2,\ x^2 + y^2 < 1\}\big|}{\pi} = \frac{\frac{\pi}{3} - \frac{\sqrt{3}}{4}}{\pi} = \frac{1}{3} - \frac{\sqrt{3}}{4\pi} \approx 0.1955$$

b) 1/2.
c)

$$\frac{\big|\{(x, y) : x > 1/2,\ x > y,\ x^2 + y^2 < 1\}\big|}{|\{(x, y) : x > 1/2,\ x^2 + y^2 < 1\}|} = \frac{\frac{\pi}{24} - \frac{\sqrt{3}-1}{8}}{\frac{1}{3} - \frac{\sqrt{3}}{4\pi}} \approx 0.2015,$$

where we used (a) and the fact that

$$|\{(x,y) : x > 1/2,\ x > y,\ x^2+y^2 < 1\}| = \int_{1/2}^{1/\sqrt{2}} (\int_x^{\sqrt{1-x^2}} dy)dx$$

$$= \int_{\frac{1}{2}}^{\frac{1}{\sqrt{2}}} \sqrt{1-x^2}\,dx - \left(\frac{x^2}{2}\bigg|_{1/2}^{1/\sqrt{2}}\right) = \int_{\frac{\pi}{4}}^{\frac{\pi}{3}} \sin^2(\varphi)d\varphi - \frac{3}{8}$$

$$= \int_{\frac{\pi}{4}}^{\frac{\pi}{3}} \frac{1}{2}(1-\cos(2\varphi))d\varphi - \frac{3}{8} = \frac{\pi}{24} - \frac{1}{4}\sin(2\varphi)\bigg|_{\frac{\pi}{4}}^{\frac{\pi}{3}} - \frac{3}{8} = \frac{\pi}{24} - \frac{\sqrt{3}-1}{8}.$$

Theory Exercises

4.17 Consider events A and B such that $P(A) > 0$ and $P(B) > 0$. Then

a) $P(B \mid A) > P(B)$ if and only if $\dfrac{P(B \cap A)}{P(A)} > P(B)$, which is true if and only if $P(A \cap B) > P(A)P(B)$ (since $P(A) > 0$). The latter is true if and only if $\dfrac{P(A \cap B)}{P(B)} > P(A)$ (since $P(B) > 0$), but that is equivalent to $P(A \mid B) > P(A)$.

b) $P(B \mid A) < P(B)$ if and only if $\dfrac{P(B \cap A)}{P(A)} < P(B)$, which is true if and only if $P(A \cap B) < P(A)P(B)$ (since $P(A) > 0$). The latter is true if and only if $\dfrac{P(A \cap B)}{P(B)} < P(A)$ (since $P(B) > 0$), but that is equivalent to $P(A \mid B) < P(A)$.

c) $P(B \mid A) = P(B)$ if and only if $\dfrac{P(B \cap A)}{P(A)} = P(B)$, which is true if and only if $P(A \cap B) = P(A)P(B)$ (since $P(A) > 0$). The latter is true if and only if $\dfrac{P(A \cap B)}{P(B)} = P(A)$ (since $P(B) > 0$), but that is equivalent to $P(A \mid B) = P(A)$.

d) If the events are positively correlated, then, given the occurrence of one of them, the chances that the other event occurs increase. If the events are negatively correlated, then, given the occurrence of one of them, the probability of the other event decreases. If the events are independent, then their probabilities remain unchanged irrespective of the occurrence of one of the events.

4.18 Note that

i) $P_A(B) = P(B \mid A) = \dfrac{P(B \cap A)}{P(A)} \geq 0$ for any event B.

ii) $P_A(\Omega) = P(\Omega \mid A) = \dfrac{P(\Omega \cap A)}{P(A)} = \dfrac{P(A)}{P(A)} = 1$, since $\Omega \cap A = A$.

iii) For arbitrary mutually exclusive events $(B_n, n \in \mathcal{N})$,

$$P_A\left(\bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n \mid A\right) = \frac{P\left((\bigcup_{n=1}^{\infty} B_n) \cap A\right)}{P(A)} = \frac{P\left(\bigcup_{n=1}^{\infty}(B_n \cap A)\right)}{P(A)}$$

$$= \frac{\sum_{n=1}^{\infty} P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} \frac{P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} P(B_n \mid A) = \sum_{n=1}^{\infty} P_A(B_n),$$

where we used that (unconditional) probability is countably additive and events $B_n \cap A$ (where $n \in \mathcal{N}$) are mutually exclusive. Thus, all three Kolmogorov axioms hold for $P_A(\cdot)$, implying that $P_A(\cdot) = P(\cdot \,|\, A)$ is a probability measure.

4.19 Note that $P(A \cap B) > 0$ implies that $P(A) > 0$ and $P(B \,|\, A) > 0$. Then

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

thus,

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

4.20 Note that

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

where the last equality holds by Exercise 4.19. Therefore, conditioning first on the occurrence of A and then on the occurrence of B is the same as conditioning on the occurrence of both A and B at once (i.e. conditioning on $A \cap B$).

Advanced Exercises

4.21 a) Since for any two events A and B, $P(A \,|\, B) = \dfrac{P(A \cap B)}{P(B)}$, then, upon multiplying both sides by $P(B)$, one obtains that $P(A \cap B) = P(A \,|\, B)P(B)$. Therefore,

$$P(W \cap R_1) = P(W \,|\, R_1)P(R_1) = 0.12 \times 0.35 = 0.042,$$

$$P(W \cap R_2) = P(W \,|\, R_2)P(R_2) = 0.18 \times 0.60 = 0.108,$$

$$P(W \cap R_3) = P(W \,|\, R_3)P(R_3) = 0.13 \times 0.05 = 0.0065,$$

$$P(B \cap R_1) = P(B \,|\, R_1)P(R_1) = 0.81 \times 0.35 = 0.2835,$$

$$P(B \cap R_2) = P(B \,|\, R_2)P(R_2) = 0.72 \times 0.60 = 0.432,$$

$$P(B \cap R_3) = P(B \,|\, R_3)P(R_3) = 0.75 \times 0.05 = 0.0375,$$

$$P(O \cap R_1) = P(O \,|\, R_1)P(R_1) = 0.07 \times 0.35 = 0.0245,$$

$$P(O \cap R_2) = P(O \,|\, R_2)P(R_2) = 0.10 \times 0.60 = 0.06,$$

$$P(O \cap R_3) = P(O \,|\, R_3)P(R_3) = 0.12 \times 0.05 = 0.006$$

b) Using (a), we obtain that

$$P(W) = P(W \cap R_1) + P(W \cap R_2) + P(W \cap R_3) = 0.1565,$$

$$P(B) = P(B \cap R_1) + P(B \cap R_2) + P(B \cap R_3) = 0.753,$$

$$P(O) = P(O \cap R_1) + P(O \cap R_2) + P(O \cap R_3) = 0.0905$$

c) Conditional probability distributions of religion by occupation is given by:

		Occupation		
		W	B	O
Religion	R_1	$\frac{0.042}{0.1565} \approx 0.268$	$\frac{0.2835}{0.753} \approx 0.376$	$\frac{0.0245}{0.0905} \approx 0.271$
	R_2	$\frac{0.108}{0.1565} \approx 0.690$	$\frac{0.432}{0.753} \approx 0.574$	$\frac{0.06}{0.0905} \approx 0.663$
	R_3	$\frac{0.0065}{0.1565} \approx 0.042$	$\frac{0.0375}{0.753} \approx 0.050$	$\frac{0.006}{0.0905} \approx 0.066$
	Total	1	1	1

d) Since $P(R_1|W) = 0.268 < 0.35 = P(R_1)$, the randomly chosen member of the community is less likely to be of Religion 1 if you are told that this person is a white-collar worker than if you are not given that information.

4.22 Let S_i $(i = 1, 2)$ denote the event that a randomly selected student passes the exam on the ith try. Then
a) $P(\text{pass}) = P(S_1) + P(S_1^c \cap S_2) = 0.6 + P(S_2 \,|\, S_1^c)P(S_1^c) = 0.6 + 0.8(1 - 0.6) = 0.92$
b) $P(S_1 \,|\, \text{pass}) = \dfrac{P(S_1)}{P(\text{pass})} = \dfrac{0.6}{0.92} \approx 0.652$

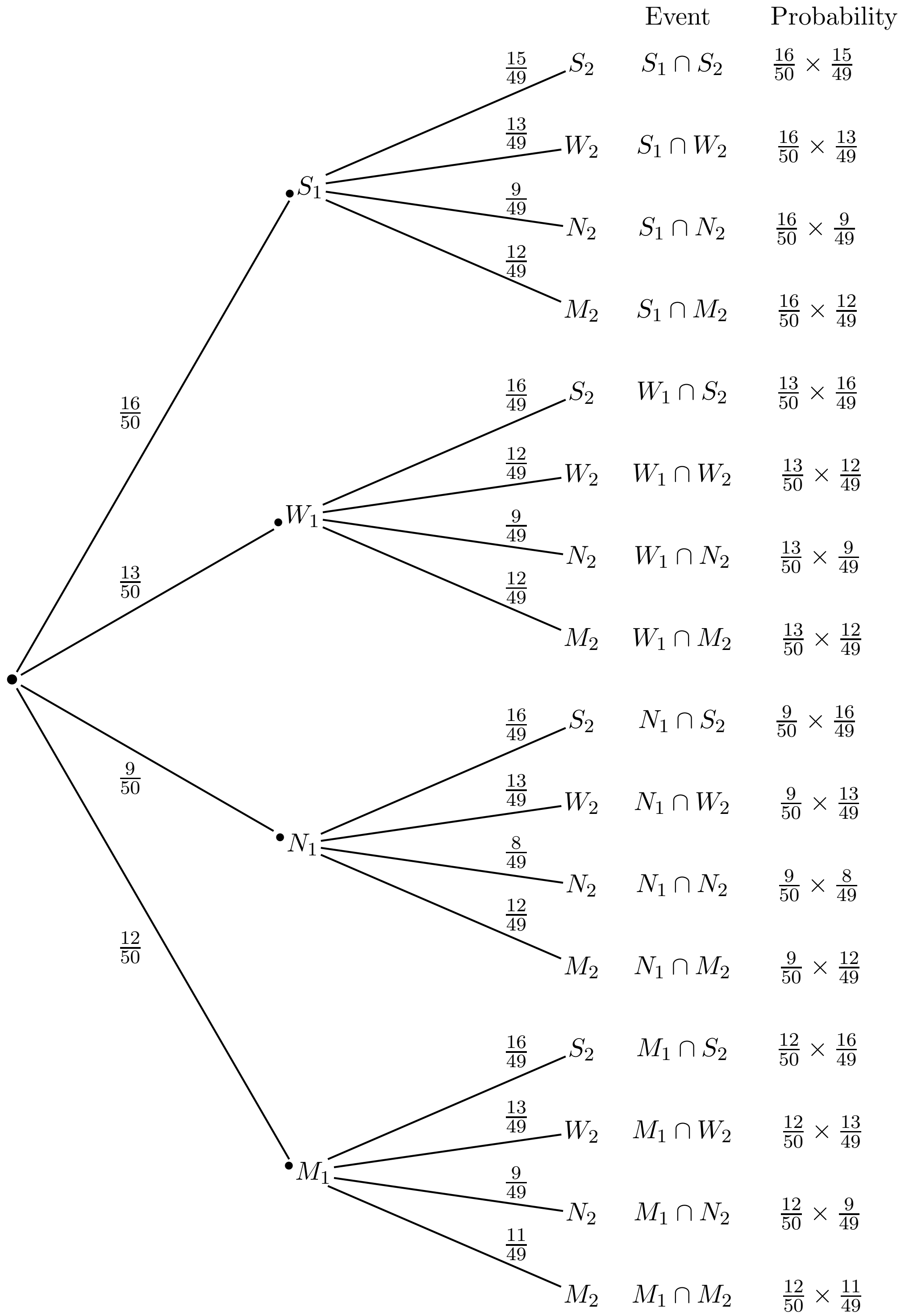
4.2 The General Multiplication Rule

Basic Exercises

4.23 a) $P(U \cap R) = P(R \,|\, U)P(U) = 0.25 \times 0.36 = 0.09$, where U denotes the event that a randomly selected American with Internet access is a regular user, and R denotes the event that a randomly selected American with Internet access finds that the Web has reduced his/her social contact.
b) 9% of Americans with Internet access are regular Internet users who feel that the Web has reduced their social contact.

4.24 a) $\dfrac{1}{6} \times \dfrac{2}{5} = \dfrac{1}{15}$.
b) $P(\text{both are less than 3}) = \dfrac{2}{6} \times \dfrac{1}{5} = \dfrac{1}{15}$. $P(\text{both are greater than 3}) = \dfrac{3}{6} \times \dfrac{2}{5} = \dfrac{1}{5}$.

4.25 a) $\dfrac{9}{50} \times \dfrac{13}{49} = \dfrac{117}{2450} \approx 0.04776$
b) $\dfrac{16}{50} \times \dfrac{15}{49} = \dfrac{240}{2450} \approx 0.098$
c) For $i = 1, 2$, let S_i denote the event that the ith selected state is from the South, W_i denote the event that the ith selected state is from the West, N_i denote the event that the ith selected state is from the Northeast, and M_i denote the event that the ith selected state is from the Midwest. Then the required tree diagram is of the form:



d) $P(\text{both in South}) + P(\text{both in West}) + P(\text{both in Northeast}) + P(\text{both in Midwest})$
$= \frac{16}{50} \times \frac{15}{49} + \frac{13}{50} \times \frac{12}{49} + \frac{9}{50} \times \frac{8}{49} + \frac{12}{50} \times \frac{11}{49} = \frac{600}{2450} \approx 0.2449$

e) $2 \times \frac{12}{50} \times \frac{13}{49} = \frac{312}{2450} \approx 0.1273$

4.26 a) $\frac{9}{50} \times \frac{16}{49} \times \frac{15}{48} = \frac{2160}{117600} \approx 0.0184$

b) $\frac{13}{50} \times \frac{12}{49} \times \frac{11}{48} = \frac{1716}{117600} \approx 0.0146$
c) $4 \times 4 \times 4 = 64$.
d) $3!\,(P(S_1 \cap W_2 \cap N_3) + P(S_1 \cap W_2 \cap M_3) + P(W_1 \cap N_2 \cap M_3) + P(S_1 \cap N_2 \cap M_3))$
$= 3!\left(\frac{16}{50} \times \frac{13}{49} \times \frac{9}{48} + \frac{16}{50} \times \frac{13}{49} \times \frac{12}{48} + \frac{13}{50} \times \frac{9}{49} \times \frac{12}{48} + \frac{16}{50} \times \frac{9}{49} \times \frac{12}{48}\right) = \frac{45000}{117600} \approx 0.38265$, where the factor 3! appears since the three states can be selected in any order.
e) $\dfrac{\binom{16}{2}\binom{12}{1}}{\binom{50}{3}} = \dfrac{8640}{117600} \approx 0.0735$

4.27 Let S_i $(i = 1, 2, 3)$ denote the event that the ith try is a success, i.e. that on the ith try the student passes the exam. Then
a) $P(\text{student passes on the 2nd try}) = P(S_2 \cap S_1^c) = p(S_2 \,|\, S_1^c)P(S_1^c) = 0.54 \times 0.4 = 0.216$
b) $P(\text{student passes on the 3rd try}) = P(S_3 \cap S_2^c \cap S_1^c) = P(S_3 \,|\, S_1^c \cap S_2^c)P(S_2^c \,|\, S_1^c)P(S_1^c)$
$= 0.48 \times (1 - 0.54) \times 0.4 = 0.08832$
c) 90.432%, since $P(S_1) + P(S_2 \cap S_1^c) + P(S_3 \cap S_2^c \cap S_1^c) = 0.6 + 0.216 + 0.08832 = 0.90432$

4.28 a) $\frac{(N-1)(N-2)\ldots(N-(n-2))(N-(n-1))\cdot 1}{N(N-1)\ldots(N-n+1)} = \frac{1}{N}$.
b) $\frac{N-1}{N} \times \frac{N-2}{N-1} \times \cdots \times \frac{N-(n-1)}{N-(n-2)} \times \frac{1}{N-(n-1)} = \frac{1}{N}$.
c) $\frac{1}{N}$, since, by symmetry, the required probability should equal to the probability that the very first key on the key ring is the required house key.

4.29 $P(G) = P(G \,|\, M)P(M) + P(G \,|\, M^c)P(M^c) = 0.174 \times 0.486 + 0.045 \times 0.514 = 0.107694$, where G is the event that a randomly selected American over the age of seven is a golfer, while M is the event that the selected person is male. I.e. approximately 10.8% of Americans, 7 years old or older, play golf.

4.30 $0.127 \times 0.83 + 0.207 \times 0.54 + 0.22 \times 0.43 + 0.165 \times 0.37 + 0.109 \times 0.27 + 0.172 \times 0.2 = 0.43667$

4.31 $\displaystyle\sum_{i=1}^{6} P(\text{two heads} \,|\, \{i\})P(\{i\}) = \frac{1}{6}\left(0 + \frac{1}{2^2} + \frac{\binom{3}{2}}{2^3} + \frac{\binom{4}{2}}{2^4} + \frac{\binom{5}{2}}{2^5} + \frac{\binom{6}{2}}{2^6}\right) \approx 0.2578$, where $\{i\}$ denotes the event that the die shows i, $i = 1, \cdots, 6$.

4.32 a) $P(R_2) = P(R_2 \,|\, R_1)P(R_1) + P(R_2 \,|\, B_1)P(B_1) = \dfrac{r+c}{r+b+c} \cdot \dfrac{r}{r+b} + \dfrac{r}{r+b+c} \cdot \dfrac{b}{r+b}$
$= \dfrac{r}{r+b}$.
b) $r/(r+b)$, since

$$\begin{aligned}
P(R_3) &= P(R_3 \,|\, R_1 \cap R_2)P(R_2 \,|\, R_1)P(R_1) + P(R_3 \,|\, R_1 \cap B_2)P(B_2 \,|\, R_1)P(R_1) \\
&+ P(R_3 \,|\, B_1 \cap R_2)P(R_2 \,|\, B_1)P(B_1) + P(R_3 \,|\, B_1 \cap B_2)P(B_2 \,|\, B_1)P(B_1) \\
&= \frac{r+2c}{r+b+2c} \cdot \frac{r+c}{r+b+c} \cdot \frac{r}{r+b} + \frac{r+c}{r+b+2c} \cdot \frac{b}{r+b+c} \cdot \frac{r}{r+b} \\
&+ \frac{r+c}{r+b+2c} \cdot \frac{r}{r+b+c} \cdot \frac{b}{r+b} + \frac{r}{r+b+2c} \cdot \frac{b+c}{r+b+c} \cdot \frac{b}{r+b} \\
&= \frac{r}{r+b}.
\end{aligned}$$

c) $P(R_1 \,|\, R_2) = \dfrac{P(R_2|R_1)P(R_1)}{P(R_2)} = \dfrac{\frac{r+c}{r+b+c} \cdot \frac{r}{r+b}}{\frac{r}{r+b}} = \dfrac{r+c}{r+b+c}$.

d) $P(B_2 \cap R_3) = P(R_1 \cap B_2 \cap R_3) + P(B_1 \cap B_2 \cap R_3) = \dfrac{r}{r+b} \cdot \dfrac{b}{r+b+c} \cdot \dfrac{r+c}{r+b+2c} + \dfrac{b}{r+b}$
$\times \dfrac{b+c}{r+b+c} \cdot \dfrac{r}{r+b+2c} = \dfrac{rb}{(r+b)(r+b+c)}$.

4.33 $\frac{5}{7} \times \frac{2}{3} + \frac{6}{7} \times \frac{1}{3} = \frac{16}{21} \approx 0.7619$

4.34 Let $\{k\}$ be the event that the second number selected is k ($k = 0, \cdots, 9$). Then, by the law of total probability,

$$P(\{k\}) = \sum_{j=0}^{k} \frac{1}{10-j} \cdot \frac{1}{10}, \quad k \in \{0, \cdots, 9\}.$$

Namely, $P(\{0\}) = 0.01$, $P(\{1\}) \approx 0.02111$, $P(\{2\}) \approx 0.03361$, $P(\{3\} \approx 0.04790$,
$P(\{4\}) \approx 0.06456$, $P(\{5\}) \approx 0.08456$, $P(\{6\}) \approx 0.10956$, $P(\{7\}) \approx 0.14290$, $P(\{8\}) \approx 0.19290$,
$P(\{9\}) \approx 0.29290$

4.35 $P(F_1 \,|\, M_2) = \dfrac{P(F_1 \cap M_2)}{P(M_2)} = \dfrac{P(M_2 \,|\, F_1)P(F_1)}{P(M_2 \,|\, F_1)P(F_1) + P(M_2 \,|\, M_1)P(M_1)}$
$= \dfrac{\frac{17}{39} \times \frac{23}{40}}{\frac{17}{39} \times \frac{23}{40} + \frac{16}{39} \times \frac{17}{40}} = \dfrac{391}{663} \approx 0.5897$

Theory Exercises

4.36 Since $P(A_2 \,|\, A_1) = P(A_2 \cap A_1)/P(A_1)$, then $P(A_1 \cap A_2) = P(A_2 \,|\, A_1)P(A_1)$, thus, Equation (4.5) holds for $n = 2$. Now suppose that Equation (4.5) in the textbook is true for all $n \leq N$. Then

$$P\Big(\bigcap_{n=1}^{N+1} A_n\Big) = P\Big((\bigcap_{n=1}^{N} A_n) \cap A_{N+1}\Big) = P\Big(\bigcap_{n=1}^{N} A_n\Big)P\Big(A_{N+1} \,\Big|\, \bigcap_{n=1}^{N} A_n\Big)$$
$$= P(A_1)P(A_2 \,|\, A_1) \ldots P(A_N \,|\, A_1 \cap \cdots \cap A_{N-1})P(A_{N+1} \,|\, A_1 \cap \cdots \cap A_N),$$

thus, Equation (4.5) holds by mathematical induction principle.

Advanced Exercises

4.37 a) 47.37%. Let D denote the event that a randomly selected female in the study takes a different mate, and let O denote the event that offspring of a randomly selected female dies. Then

$$0.63 = P(D) = P(D \,|\, O)P(O) + P(D \,|\, O^c)(1 - P(O))$$
$$= 0.43P(O) + 0.81(1 - P(O)) = 0.81 - 0.38P(O),$$

implying that $P(O) = \frac{18}{38} = \frac{9}{19} \approx 0.4737$
b) 20.37%. Using (a), one obtains that

$$P(D \cap O) = P(D \,|\, O)P(O) = 0.43 \times \frac{9}{19} \approx 0.2037$$

c) 32.33%, since

$$P(O \mid D) = \frac{P(D \cap O)}{P(D)} = \frac{0.43 \times \frac{9}{19}}{0.63} \approx 0.3233$$

4.38 41/75. Note $P(R_3) = P(R_3 \mid R_1 \cap R_2)P(R_2 \mid R_1)P(R_1) + P(R_3 \mid G_1 \cap R_2)P(R_2 \mid G_1)P(G_1)$ $+P(R_3 \mid R_1 \cap G_2)P(G_2 \mid R_1)P(R_1) + P(R_3 \mid G_1 \cap G_2)P(G_2 \mid G_1)P(G_1)$, where R_i denotes the event that a red ball is taken from Urn $\#i$ (and is placed in Urn $\#(i+1)$), $i = 1, 2, 3$; G_i denotes a similar event but for a green ball. Thus,

$$P(R_3) = \frac{3}{5} \times \frac{5}{6} \times \frac{2}{5} + \frac{3}{5} \times \frac{4}{6} \times \frac{3}{5} + \frac{2}{5} \times \frac{1}{6} \times \frac{2}{5} + \frac{2}{5} \times \frac{2}{6} \times \frac{3}{5} = \frac{41}{75} \approx 0.5467$$

4.39 Let $F(r, b, n) = P(n\text{th ball selected is red if initially there are } r \text{ red and } b \text{ black balls})$, $r, b, n \in \mathcal{N}$. By Exercise 4.32(a), $P(F(r, b, 2)) = \frac{r}{r+b}$. Next suppose that $P(F(r, b, n-1)) = \frac{r}{r+b}$. Let us show that $P(F(r, b, n)) = \frac{r}{r+b}$. Let R_i be the event that the ith ball is red, and B_i be the event that the ith ball is black, when starting with r red and b black balls in the urn. Then

$$\begin{aligned} P(F(r, b, n)) &= P(R_n \mid R_1)P(R_1) + P(R_n \mid B_1)P(B_1) \\ &= P\left(F(r+c, b, n-1)\right) \cdot \frac{r}{r+b} + P\left(F(r, b+c, n-1)\right) \cdot \frac{b}{r+b} \\ &= \frac{r+c}{r+c+b} \cdot \frac{r}{r+b} + \frac{r}{r+b+c} \cdot \frac{b}{r+b} = \frac{r}{r+b}. \end{aligned}$$

Thus, by induction in n, the required conclusion follows for all $n \in \mathcal{N}$.

4.40 1/3 (provided, of course, that $n \geq 2$). Let B_k denote the event that the bowl containing k red marbles and $(n-k)$ green marbles is selected, $k = 0, \cdots, n$. Fix arbitrary $n \geq 2$. Then

$$P(\text{two red marbles are selected}) = \sum_{k=0}^{n} P(\text{two red marbles are selected} \mid B_k)P(B_k)$$

$$= \sum_{k=0}^{n} \frac{\binom{k}{2}}{\binom{n}{2}} \cdot \frac{1}{n+1} = \frac{1}{n+1} \cdot \sum_{k=0}^{n} \frac{k(k-1)}{n(n-1)} = \frac{1}{(n+1)n(n-1)} \sum_{k=1}^{n} k(k-1)$$

$$= \frac{1}{(n+1)n(n-1)} \left(\sum_{k=1}^{n} k^2 - \sum_{k=1}^{n} k \right) = \frac{1}{(n+1)n(n-1)} \left(\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} \right) = \frac{1}{3}.$$

4.41 a) $p = 4/7$, $q = 2/7$, $r = 1/7$. Let us call a person, who comes as an incumbent into round n after winning the $(n-1)$st round, the nth incumbent, $n \geq 2$. Let I_n be the event that the nth incumbent wins the game, let C_n be the event that the nth challenger wins the game, and B_n be the event that the nth benchwarmer wins the game, $n \geq 2$. Also let S_n denote the event that the nth incumbent wins the nth round. Then

$$p = P(I_n) = P(I_n \mid S_n)P(S_n) + P(I_n \mid S_n^c)P(S_n^c) = 1 \cdot \frac{1}{2} + P(B_{n+1}) \cdot \frac{1}{2} = \frac{1}{2} + \frac{r}{2},$$

since the nth incumbent, who loses the nth round, is the benchwarmer in the $(n+1)$st round. Similarly,

$$q = P(C_n) = P(C_n \mid S_n)P(S_n) + P(C_n \mid S_n^c)P(S_n^c) = 0 \cdot \frac{1}{2} + P(I_{n+1}) \cdot \frac{1}{2} = \frac{p}{2},$$

since the challenger, who wins the round, becomes an incumbent in the next round. Also,

$$r = P(B_n) = P(B_n \,|\, S_n)P(S_n) + P(B_n \,|\, S_n^c)P(S_n^c) = 0 \cdot \frac{1}{2} + P(C_{n+1}) \cdot \frac{1}{2} = \frac{q}{2}.$$

Upon solving the system: $p = \frac{1}{2} + \frac{r}{2}$, $q = \frac{p}{2}$, $r = \frac{q}{2}$, we obtain that $p = \frac{4}{7}$, $q = \frac{2}{7}$ and $r = \frac{1}{7}$.

b) $P(\text{Tom wins game}) = P(\text{Tom wins game} \,|\, \text{Tom wins 1st round})P(\text{Tom wins 1st round})$ $+P(\text{Tom wins game} \,|\, \text{Tom loses 1st round})P(\text{Tom loses 1st round}) = p \cdot \frac{1}{2} + r \cdot \frac{1}{2} = \frac{5}{14}$. By a similar argument (or, simply, by symmetry), $P(\text{Dick wins game}) = \frac{5}{14}$.
Note also that $P(\text{Harry wins game}) = q = \frac{2}{7}$.
c) (i) By symmetry, $P(\text{Tom wins game}) = P(\text{Dick wins game}) = P(\text{Harry wins game}) = 1/3$.
(ii) Conditioning on whether Tom is chosen as a benchwarmer in the first round and applying results of (b), we obtain that

$$P(\text{Tom wins game}) = \frac{2}{7} \times \frac{1}{3} + \frac{5}{14} \times \frac{2}{3} = \frac{1}{3}.$$

Likewise, $P(\text{Dick wins game}) = 1/3$ and $P(\text{Harry wins game}) = 1/3$.

4.42 43/459. Let X be the number of queens, excluding the queen of spades, in the second stack. Let Q denote the event that the card drawn from the second stack is a queen. Then

$$P(Q) = \sum_{n=0}^{3} P(Q \,|\, X = n)P(X = n) = \sum_{n=0}^{3} \frac{n+1}{27} \times \frac{\binom{3}{n}\binom{48}{26-n}}{\binom{51}{26}} = \frac{43}{459} \approx 0.09368$$

4.3 Independent Events

4.43 a) By Definitions 4.2 and 4.3, if A and B are independent (with $P(A) > 0$), then

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

implying that $P(B \cap A) = P(A)P(B)$.
b) If $P(A) > 0$ and $P(A \cap B) = P(A)P(B)$, then

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

4.44 a) No. $P(C_1) = \frac{9.3}{61.4} \neq \frac{1.3}{25.8} = P(C_1 \,|\, S_2)$, thus, C_1 and S_2 are not independent.
b) No. $P(S_1) = \frac{35.6}{61.4} \neq \frac{9.8}{21.4} = P(S_1 \,|\, C_2)$, thus, S_1 and C_2 are not independent.

4.45 Since $P(A_2) = 0.401$, $P(L_1) = 0.126$ and $P(A_2 \cap L_1) = 0.038 \neq 0.401 \times 0.126$, then $P(A_2 \cap L_1) \neq P(A_2)P(L_1)$, implying that A_2 and L_1 are not independent. In other words, for a randomly selected U.S. citizen, 15 years of age or older, the event that the selected person is between 25 and 44 years old and the event that the person lives alone are not independent (thus, if one of the two events does in fact occur, then the probability of the other event changes).

4.46 a) $\frac{1}{6^5} \approx 0.0001286$

b) $\frac{6}{6^5} = \frac{1}{6^4} \approx 0.0007716$

c) $\dfrac{\binom{6}{2} \times 2 \times \frac{5!}{2!3!}}{6^5} = \dfrac{300}{6^5} \approx 0.03858$, since there are $\binom{6}{2}$ ways to pick which two of the six numbers are rolled, then there are two ways to decide which one of the two selected numbers will be doubled (and which one will be tripled). Next fix one such selection, say, three 1's and two 2's. Then there are $\frac{5!}{3!2!}$ possible ordered arrangements of three 1's and two 2's in 5 positions. Similar argument holds for any other such selection, and the required result follows.

4.47 a) Suppose first that $P(A) = 0$, then, for an arbitrary event B, we have $A \cap B \subset A$, implying that $0 \leq P(A \cap B) \leq P(A) = 0$, thus $P(A \cap B) = 0 = P(A)P(B)$, which by Definition 4.4 implies independence of A and B.

On the other hand, if $P(A) = 1$, then $P(A^c) = 0$, and, by the just proved statement, we have that A^c is independent of any other event. Then, by Proposition 4.4, A is independent of any other event.

b) If A and A are independent, then $P(A) = P(A \cap A) = P(A)P(A)$, implying that either $P(A) = 0$ or $P(A) = 1$.

4.48 Yes, A and B are independent, since

$$P(A) = \frac{|[0.6, 0.7)|}{|(0,1)|} = 0.1,$$

$$P(B) = \frac{\sum_{k=0}^{9} \left|\left[0.04 + \frac{k}{10}, 0.05 + \frac{k}{10}\right)\right|}{|(0,1)|} = \frac{10 \times 0.01}{1} = 0.1,$$

and

$$P(A \cap B) = \frac{|[0.64, 0.65)|}{|(0,1)|} = 0.01 = 0.1 \times 0.1 = P(A)P(B).$$

4.49 False. The counter-example will vary. For instance, take an arbitrary event A with $0 < P(A) < 1$ and let $B = \Omega$ and $C = A$. Then, by Exercise 4.47, A and B are independent, also B and C are independent, but A and C are not independent.

4.50 a) $0.255 \times 0.238 \times 0.393 \approx 0.02385$

b) $3 \times 0.255^2 \times 0.114 \approx 0.0222$

4.51 $2^n - 1 - n$. For each $m \in \{2, \cdots, n\}$, one needs to check that $P(A_{k_1} \cap \cdots \cap A_{k_m}) = \prod_{i=1}^{m} P(A_{k_i})$ for every (unordered) set $\{k_1, \cdots, k_m\}$ of m distinct integers $k_i \in \{1, \cdots, n\}$, $i = 1, \cdots, m$. There are $\binom{n}{m}$ such sets (for each m), therefore, one needs to check a total of $\sum_{m=2}^{n} \binom{n}{m}$ equations to check for mutual independence of n events. By Binomial theorem,

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

4.52 a) By Example 4.15(d), assignments #1, #2 and #3 correspond to independent tosses of the coin. On the other hand, assignment #4 does not correspond to independent tosses since $P(\text{HTH}) = 0.050 \neq 0.220 = P(\text{HHT})$, whereas, as established in Example 4.15(d), individual outcomes with the same number of heads must have the same probability.

b) 0.5, 1 and 0.2 for assignments #1, #2 and #3, respectively. If p is the probability of head on a given toss, then $P(\text{HHH}) = p^3$, implying that $p = (P(\text{HHH}))^{1/3}$, and the required answer easily follows in each case.

4.53 a) By the inclusion-exclusion principle (and after some elementary algebra),

$$\begin{aligned} P\Big(\bigcup_{j=1}^{5} A_j\Big) &= p_1+p_2+p_3+p_4+p_5-p_1p_2-p_1p_3-p_1p_4-p_1p_5-p_2p_3-p_2p_4-p_2p_5 \\ &- p_3p_4-p_3p_5-p_4p_5+p_1p_2p_3+p_1p_2p_4+p_1p_2p_5+p_1p_3p_4+p_1p_3p_5+p_1p_4p_5 \\ &+ p_2p_3p_4+p_2p_3p_5+p_2p_4p_5+p_3p_4p_5-p_1p_2p_3p_4-p_1p_2p_3p_5-p_1p_2p_4p_5 \\ &- p_1p_3p_4p_5-p_2p_3p_4p_5+p_1p_2p_3p_4p_5 \\ &= 1-(1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5). \end{aligned}$$

b) For a series system,

$$P(\text{system is functioning}) = P\Big(\bigcap_{j=1}^{n} A_j\Big) = \prod_{j=1}^{n} p_j.$$

For a parallel system,

$$P(\text{system is functioning}) = P\Big(\bigcup_{j=1}^{n} A_j\Big) = 1-P\Big(\bigcap_{j=1}^{n} A_j^c\Big) = 1-\prod_{j=1}^{n} P(A_j^c) = 1-\prod_{j=1}^{n}(1-p_j),$$

where we applied the complementation rule, De Morgan's law and Proposition 4.5.

4.54 $244/495 \approx 0.4929$; Let X be the sum of scores on the two dice in the first roll. Then

$$P(\text{win}) = P(X=7)+P(X=11)+\sum_{k\in\{4,5,6,8,9,10\}} P(\text{win}\,|\,X=k)P(X=k)$$

$$= \frac{6}{36}+\frac{2}{36}+\sum_{k\in\{4,5,6,8,9,10\}} \frac{P(X=k)}{P(X=k)+P(X=7)} \times P(X=k)$$

$$= \frac{8}{36}+\frac{\frac{3}{36}\times\frac{3}{36}}{\frac{3}{36}+\frac{6}{36}}+\frac{\frac{4}{36}\times\frac{4}{36}}{\frac{4}{36}+\frac{6}{36}}+\frac{\frac{5}{36}\times\frac{5}{36}}{\frac{5}{36}+\frac{6}{36}}+\frac{\frac{5}{36}\times\frac{5}{36}}{\frac{5}{36}+\frac{6}{36}}+\frac{\frac{4}{36}\times\frac{4}{36}}{\frac{4}{36}+\frac{6}{36}}+\frac{\frac{3}{36}\times\frac{3}{36}}{\frac{3}{36}+\frac{6}{36}} = \frac{244}{495} \approx 0.4929$$

4.55 a) If A and B are mutually exclusive, then $P(A\cap B) = P(\emptyset) = 0$.
b) If A and B are independent with $P(A) > 0$ and $P(B) > 0$, then $P(A\cap B) = P(A)P(B) > 0$, which by (a) implies that A and B cannot be mutually exclusive.
c) The answers will vary.

4.56 Let H_i denote the event that the ith toss results in a head, $i = 1, 2$.
a) $P(H_1) = P(H_1\,|\,\text{balanced})P(\text{balanced})+P(H_1\,|\,\text{biased})P(\text{biased}) = 0.5\cdot\frac{1}{2}+p\cdot\frac{1}{2} = (p+0.5)/2$.
b) $(p+0.5)/2$. Note that

$$\begin{aligned} P(H_2) &= P(H_1\cap H_2\,|\,\text{balanced})P(\text{balanced})+P(H_1\cap H_2\,|\,\text{biased})P(\text{biased}) \\ &+ P(H_1^c\cap H_2\,|\,\text{balanced})P(\text{balanced})+P(H_1^c\cap H_2\,|\,\text{biased})P(\text{biased}) \\ &= \frac{1}{4}\cdot\frac{1}{2}+p^2\cdot\frac{1}{2}+\frac{1}{4}\cdot\frac{1}{2}+(1-p)p\cdot\frac{1}{2} = \frac{1}{4}+\frac{p}{2}. \end{aligned}$$

c) $P(H_1 \cap H_2) = \frac{1}{4} \cdot \frac{1}{2} + p^2 \cdot \frac{1}{2} = 0.5(p^2 + 0.25)$.
d) Events H_1 and H_2 are independent if and only if $P(H_1 \cap H_2) = P(H_1)P(H_2)$, which, by (a)–(c), holds if and only if

$$0.5(p^2 + 0.25) = [(p + 0.5)/2]^2,$$

which is equivalent to $p^2 - p + 0.25 = 0$, or $(p - 0.5)^2 = 0$. Therefore, H_1 and H_2 are independent if and only if $p = 0.5$, which means that the second coin has to be balanced in order for H_1 and H_2 to be independent.

4.57 a) Exactly two of the four children are boys means that either first two children born are boys and the 3rd and 4th are girls, or the 1st and 3rd are boys and the 2nd and 4th are girls, or the 1st and 4th are boys and the 2nd and 3rd are girls, or the 2nd and 3rd are boys and the 1st and 4th are girls, or the 2nd and 4th are boys and the 1st and 3rd are girls, or the first two are girls and the last two are boys.
b) $P(S_2) = 6p^2(1 - p)^2$.
c) $P(S_k) = \binom{4}{k} p^k (1 - p)^{4-k}$, $k = 0, 1, 2, 3, 4$. If the nth child born is a boy, let us say that the nth position is occupied by a boy. Then, if exactly k of the four children born are boys, then there are $\binom{4}{k}$ ways to pick a set of positions occupied by boys. Fix one such set of positions for boys, say, the first k children are boys and the last $4 - k$ children born are girls. Probability of the latter event happening is $p^k(1-p)^{4-k}$. The same holds for every other fixed set of positions for boys. Thus, the required result follows.

Theory Exercises

4.58 Suppose A and B are independent, then

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c),$$

implying that A and B^c are independent. Interchanging the roles of A and B in the above argument leads to the conclusion that B and A^c are independent. Finally note that

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) \\ &= 1 - (P(A) + P(B) - P(A)P(B)) = (1 - P(A))(1 - P(B)) = P(A^c)P(B^c), \end{aligned}$$

implying that A^c and B^c are independent.

4.59 Suppose $P(E) = p > 0$. Let A_n be the event that E occurs for the first time on the nth try, $n = 1, 2, \ldots$. Then

$$P(E \text{ eventually occurs}) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} (1 - p)^{n-1} p$$

$$= p \sum_{k=0}^{\infty} (1 - p)^k = p \cdot \frac{1}{1 - (1 - p)} = 1,$$

since $0 \leq 1 - p < 1$.

4.60 Let E_n be the event that A occurs for the first time on the nth try and B does not occur at all in the first n repetitions of the experiment. Since A and B are mutually exclusive, then the probability that neither A nor B occurs on a given try is equal to $1-P(A\cup B) = 1-P(A)-P(B)$, implying that $P(E_n) = (1-P(A)-P(B))^{n-1}P(A)$ for each $n = 1, 2, \ldots$. Then

$$P(\{A \text{ occurs before } B\}) = P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty}(1-P(A)-P(B))^{n-1}P(A)$$

$$= \frac{P(A)}{1-(1-P(A)-P(B))} = \frac{P(A)}{P(A)+P(B)},$$

since $1 \geq P(A)+P(B) > 0$.

Advanced Exercises

4.61 a) True, provided that $P(A\cap B) > 0$, since

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

On the other hand, if $P(A\cap B) = 0$, then $P(C\,|\,A\cap B)$ may not be defined, even though $P(C\,|\,A)$ is defined.

b) False. Counterexamples may vary. For instance, consider rolling a pair of balanced dice, and let B be the event that the first die shows an odd number, and C be the event that the second die shows an odd number. Let A be the event that the sum of scores on the two dice is even. Then it's easy to see that B and C are independent, but $P(B\,|\,A) = P(C\,|\,A) = P(B\cap C\,|\,A) = 0.5$, implying that $P(B\cap C\,|\,A) \neq P(B\,|\,A)P(C\,|\,A)$.

c) False. Counterexamples may vary. For instance, consider events A, B, C such that $B = A$, $C = A^c$ and $0 < P(A) < 1$. Then

$$P(B\cap C\,|A) = \frac{P(B\cap C\cap A)}{P(A)} = \frac{P(\emptyset)}{P(A)} = 0 = 1\times 0 = P(B\,|\,A)P(C\,|\,A),$$

thus, B and C are conditionally independent given A, but

$$P(B\cap C) = P(\emptyset) = 0 \neq \underbrace{P(A)}_{>0}\underbrace{(1-P(A))}_{>0} = P(B)P(C),$$

i.e. B and C are not independent.

4.62 a) Since $1-x \leq e^{-x}$, by independence of $A_1, A_2\ldots$, for all $n \in \mathcal{N}$ we have that

$$P\left(\bigcap_{k=n}^{n+j} A_k^c\right) = \prod_{k=n}^{n+j} P(A_k^c) = \prod_{k=n}^{n+j}(1-P(A_k)) \leq \prod_{k=n}^{n+j} e^{-P(A_k)} = \exp\left[-\sum_{k=n}^{n+j} P(A_k)\right] \longrightarrow 0,$$

as $j\to\infty$, since the series $\sum_k P(A_k)$ diverges. Hence $P(\bigcap_{k=n}^{\infty} A_k^c) = \lim_{j\to\infty} P(\bigcap_{k=n}^{n+j} A_k^c) = 0$. Thus, $0 \leq P(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c) \leq \sum_{n=1}^{\infty} P(\bigcap_{k=n}^{\infty} A_k^c) = \sum_{n=1}^{\infty} 0 = 0$. Therefore, by the De Morgan's law,

$$P(A^*) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = 1 - P\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\right) = 1-0 = 1.$$

b) Fix an event E such that $0 < P(E) < 1$ and let $A_n = E$ for all $n \in \mathcal{N}$. Then, clearly, $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P(E) = \infty$, but $P(A^*) = P(E) < 1$. Thus, the second Borel-Cantelli lemma fails without the independence assumption.

4.63 a) If $r \leq 1/2$, then the amoeba population eventually becomes extinct with probability 1; if $r > 1/2$, then the extinction probability equals to $(1-r)/r$. Let E_n be the event that the amoeba population is eventually extinct if the initial size of the population is n, and let $p_n = P(E_n)$. Then,

$$p_1 = 1 \cdot P(\text{1st amoeba dies}) + P(E_1 \,|\, \text{1st amoeba splits}) P(\text{1st amoeba splits})$$

$$= (1-r) + p_2\, r = (1-r) + p_1^2 r.$$

Note that the above quadratic equation $rp_1^2 - p_1 + (1-r) = 0$ has the following roots: 1 and $(1-r)/r$. Thus, when $r \leq 1/2$, $(1-r)/r \geq 1$, implying that the required extinction probability p_1 equals to 1. On the other hand, if $r > 1/2$, then the required extinction probability equals to $(1-r)/r$.
b) For $r = 1/3$ and $r = 1/2$, the extinction probability is 1. For $r = 3/4$, the extinction probability equals to $1/3$. For $r = 9/10$, the extinction probability equals to $1/9$.
c) We use contextual independence to conclude that $P(E_1 \,|\, \text{1st amoeba splits}) = p_2 = p_1^2$.

4.64 Note that in every repetition of the experiment, either event E or E^c occurs. Then $\{n(E) = k\} = \{E \text{ occurs in } k \text{ repetitions and } E^c \text{ occurs in } n-k \text{ repetitions}\}$. There are $\binom{n}{k}$ ways to pick a set of those repetitions where E occurs. Therefore, using independence of the repetitions, we obtain that

$$P(n(E) = k) = \binom{n}{k} (P(E))^k (P(E^c))^{n-k} = \binom{n}{k} p^k (1-p)^{n-k},$$

where $k = 0, 1, \cdots, n$.

4.65 $\binom{2n}{n}/4^n$. Let X be the number of heads obtained by Jan in n independent tosses, and let Y be the number of heads obtained by Jean in n independent tosses. Then

$$P(X = Y) = \sum_{j=0}^{n} P(\{X = j\} \cap \{Y = j\}) = \sum_{j=0}^{n} P(X = j) P(Y = j),$$

where, by Exercise 4.64, $P(X = j) = P(Y = j) = \binom{n}{j} \frac{1}{2^n}$. Therefore,

$$P(X = Y) = \sum_{j=0}^{n} \binom{n}{j} \binom{n}{j} \frac{1}{2^n} \cdot \frac{1}{2^n} = \frac{1}{4^n} \sum_{j=0}^{n} \binom{n}{j} \binom{n}{n-j} = \frac{1}{4^n} \binom{2n}{n},$$

since $\binom{n}{j} = \binom{n}{n-j}$ and the last equality holds by the Vandermonde's identity (Exercise 3.48).

4.4 Bayes's Rule

Basic Exercises

4.66 a) 0.045

b) $\dfrac{0.045 \times 0.514}{0.045 \times 0.514 + 0.174 \times 0.486} \approx 0.2148$, by the Bayes's rule.

c) 4.5% of American females 7 years old or older play golf. Approximately 21.48% of Americans, who are 7 years old or older and play golf, are female.

4.67 $\dfrac{.54 \times .207}{.83 \times .127 + .54 \times .207 + .43 \times .22 + .37 \times .165 + .27 \times .109 + .2 \times .172} \approx 0.256$, by the Bayes's rule. Thus, approximately 25.6% of adult moviegoers are between 25 and 34 years old.

4.68 $\dfrac{0.6 \times 0.4}{0.6 \times 0.4 + 0.8 \times 0.32 + 0.3 \times 0.28} \approx 0.4138$

4.69 0.0989; Let N be the event that a selected policyholder is *normal*, and let A be the event that a randomly selected policyholder has an accident in a given year. Then, by the Bayes's rule,

$$P(N^c \,|\, A^c) = \frac{P(A^c \,|\, N^c)P(N^c)}{P(A^c \,|\, N)P(N) + P(A^c \,|\, N^c)P(N^c)} = \frac{0.4 \times 0.18}{0.8 \times 0.82 + 0.4 \times 0.18} \approx 0.0989$$

4.70 $495/8784 \approx 0.05635$; Upon using the notation and result of Exercise 4.54, we obtain that

$$P(X = 4 \,|\, \text{win}) = \frac{P(\text{win} \,|\, X = 4)P(X = 4)}{P(\text{win})} = \frac{\frac{3/36}{3/36+6/36} \times \frac{3}{36}}{\frac{244}{495}} = \frac{495}{244 \times 36} = \frac{495}{8784} \approx 0.05635$$

4.71 a) 1/6, 1/3, 1/2. Let I denote the event that the first urn is selected, II be the event that the second urn is selected, III be the event that the third urn is selected and R be the event that the randomly chosen marble is red. Then, by the Bayes's rule,

$$P(I \,|\, R) = \frac{\frac{1}{10} \times \frac{1}{3}}{\frac{1}{10} \times \frac{1}{3} + \frac{2}{10} \times \frac{1}{3} + \frac{3}{10} \times \frac{1}{3}} = \frac{1}{6}, \quad P(II \,|\, R) = \frac{\frac{2}{10} \times \frac{1}{3}}{\frac{1}{10} \times \frac{1}{3} + \frac{2}{10} \times \frac{1}{3} + \frac{3}{10} \times \frac{1}{3}} = \frac{2}{6} = \frac{1}{3},$$

$$P(III \,|\, R) = \frac{\frac{3}{10} \times \frac{1}{3}}{\frac{1}{10} \times \frac{1}{3} + \frac{2}{10} \times \frac{1}{3} + \frac{3}{10} \times \frac{1}{3}} = \frac{3}{6} = \frac{1}{2}.$$

b) 1/6, 1/3, 1/2; the result is not affected by the addition of one million green marbles in each urn. Indeed,

$$P(I \,|\, R) = \frac{\frac{1}{1{,}000{,}010} \times \frac{1}{3}}{\frac{1}{1{,}000{,}010} \times \frac{1}{3} + \frac{2}{1{,}000{,}010} \times \frac{1}{3} + \frac{3}{1{,}000{,}010} \times \frac{1}{3}} = \frac{1}{6}.$$

Similarly, one shows that $P(II \,|\, R)$ and $P(III \,|\, R)$ remain unchanged.

c) 3/8, 1/3, 7/24. The result will be affected by the addition of extra 1,000,000 green marbles, with the new probabilities given by: 0.3333337, 1/3, 0.333333; in general, as the number of extra green marbles added to urns increases, each required probability converges to 1/3.

Let G denote the event that the selected marble is green. Then

$$P(I \,|\, G) = \frac{\frac{9}{10} \times \frac{1}{3}}{\frac{9}{10} \times \frac{1}{3} + \frac{8}{10} \times \frac{1}{3} + \frac{7}{10} \times \frac{1}{3}} = \frac{9}{24} = \frac{3}{8}, \quad P(II \,|\, G) = \frac{\frac{8}{10} \times \frac{1}{3}}{\frac{9}{10} \times \frac{1}{3} + \frac{8}{10} \times \frac{1}{3} + \frac{7}{10} \times \frac{1}{3}} = \frac{1}{3},$$

$$P(III\,|\,G) = \frac{\frac{7}{10}\times\frac{1}{3}}{\frac{9}{10}\times\frac{1}{3}+\frac{8}{10}\times\frac{1}{3}+\frac{7}{10}\times\frac{1}{3}} = \frac{7}{24}.$$

If extra 1,000,000 green marbles are added, then

$$P(I\,|\,G) = \frac{\frac{1{,}000{,}009}{1{,}000{,}010}\times\frac{1}{3}}{\frac{1{,}000{,}009}{1{,}000{,}010}\times\frac{1}{3}+\frac{1{,}000{,}008}{1{,}000{,}010}\times\frac{1}{3}+\frac{1{,}000{,}007}{1{,}000{,}010}\times\frac{1}{3}} = \frac{1,000,009}{3,000,024} \approx 0.3333337,$$

Similarly,

$$P(II\,|\,G) = \frac{1,000,008}{3,000,024} = \frac{1}{3},\quad P(III\,|\,G) = \frac{1,000,007}{3,000,024} = 0.333333$$

In general, note that if each urn contains n marbles, where the first urn contains exactly one red marble, the second urn contains exactly two red marbles and the third urn contains exactly three red marbles (and the rest of the marbles are green), then

$$P(I\,|\,G) = \frac{\frac{n-1}{n}\times\frac{1}{3}}{\frac{n-1}{n}\times\frac{1}{3}+\frac{n-2}{n}\times\frac{1}{3}+\frac{n-3}{n}\times\frac{1}{3}} = \frac{n-1}{3n-6} = \frac{1-\frac{1}{n}}{3-\frac{6}{n}} \longrightarrow \frac{1}{3},\quad \text{as } n\to\infty.$$

Similarly,

$$P(II\,|\,G) = \frac{n-2}{3n-6} = \frac{1}{3},\quad P(III\,|\,G) = \frac{n-3}{3n-6} = \frac{1-\frac{3}{n}}{3-\frac{6}{n}} \longrightarrow \frac{1}{3},\quad \text{as } n\to\infty.$$

4.72 1/2. Let S be the event that the first chosen coin is silver, let A denote the event that chest A is selected. Similarly define events B, C and D (corresponding to chests B, C and D). Then

$$P(\text{the other drawer has silver coin}\,|\,S) = P(C\,|\,S) = \frac{1\times\frac{1}{4}}{\frac{1}{2}\times\frac{1}{4}+\frac{1}{2}\times\frac{1}{4}+1\times\frac{1}{4}+0\times\frac{1}{4}} = \frac{1}{2},$$

by the Bayes' rule (where events A, B, C and D form the needed partition of the sample space).

4.73 0.8033; Let A_n be the event that n eggs are broken in a randomly selected carton (containing 12 eggs), $n = 0, 1, 2, 3$. Let B be the event that a randomly selected egg from a carton is broken. Then, by the Bayes's rule and upon noting that $P(B\,|\,A_n) = \frac{n}{12}$, we obtain that

$$P(A_1\,|\,B) = \frac{\frac{1}{12}\times 0.192}{\frac{1}{12}\times 0.192+\frac{2}{12}\times 0.022+\frac{3}{12}\times 0.001} \approx 0.8033$$

4.74 a) 9/17. Let C_i be the event that coin #i is tossed, $i = 1, 2$. Then

$$P(C_1\,|\,\{\text{HT}\}) = \frac{P(\{\text{HT}\}\,|\,C_1)P(C_1)}{P(\{\text{HT}\}\,|\,C_1)P(C_1)+P(\{\text{HT}\}\,|\,C_2)P(C_2)} = \frac{\frac{1}{4}\times\frac{1}{2}}{\frac{1}{4}\times\frac{1}{2}+\frac{2}{3}\times\frac{1}{3}\times\frac{1}{2}} = \frac{9}{17}.$$

b) 9/17, same as in (a).
c) 9/17, since

$$P(C_1\,|\,\{\text{HT,TH}\}) = \frac{P(\{\text{HT,TH}\}\,|\,C_1)P(C_1)}{P(\{\text{HT,TH}\}\,|\,C_1)P(C_1)+P(\{\text{HT,TH}\}\,|\,C_2)P(C_2)} = \frac{\frac{1}{2}\times\frac{1}{2}}{\frac{1}{2}\times\frac{1}{2}+(2\times\frac{2}{3}\times\frac{1}{3})\times\frac{1}{2}}$$

4.75 25/42. Let H_i the event that the ith toss is a head, $i = 1, 2$. Then

$$P(H_2 \mid H_1) = \frac{P(H_1 \cap H_2)}{P(H_1)} = \frac{P(H_1 \cap H_2 \mid C_1)P(C_1) + P(H_1 \cap H_2 \mid C_2)P(C_2)}{P(H_1 \mid C_1)P(C_1) + P(H_1 \mid C_2)P(C_2)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} + \frac{2}{3} \times \frac{2}{3} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2}} = \frac{25}{42} \approx 0.5952$$

Advanced Exercises

4.76 a) 96% of people who have the disease will test positive. 98% of people who do not have the disease will test negative.

b) $\dfrac{0.96 \times 0.001}{0.96 \times 0.001 + 0.02 \times 0.999} = \dfrac{48}{1047} \approx 0.046$

c) Approximately 4.6% of people testing positive actually have the disease.

d) The prior probability of illness is 0.001, whereas the posterior probability of illness, given a positive test, is 0.046, which is 46 times higher than the prior probability.

e) Let D denote the event that the selected person has the disease, and let $+/-$ denote the corresponding positive/negative result of the test. Then the first step of the algorithm maps $(P(D), P(D^c)) \cdot (P(+ \mid D), P(+ \mid D^c))$ to $(10^3 P(D), 10^3 P(D^c)) \cdot (10^2 P(+ \mid D), 10^2 P(+ \mid D^c))$ (to work with integers), and the latter is further mapped to $\big(10^5 P(D)P(+|D), 10^5 P(D^c)P(+ \mid D^c)\big)$. Finally that vector is mapped to

$$\left(\frac{10^5 P(+ \mid D)P(D)}{10^5(P(+ \mid D)P(D) + P(+|D^c)P(D^c))}, \frac{10^5 P(+ \mid D^c)P(D^c)}{10^5(P(+ \mid D)P(D) + P(+|D^c)P(D^c))}\right),$$

which is nothing but the Bayes's rule to compute $(P(D \mid +), P(D^c \mid +))$, but with both the numerator and denominator multiplied by 10^5 in order to work with integers.

4.77 $\frac{100 p_j N_j}{\sum_{k=1}^m p_k N_k}$%. Let S_j be the event that a randomly selected member of the population belongs to stratum j $(j = 1, \cdots, m)$. Let A be the event that a randomly selected member of the population has the specified attribute. Let $N = \sum_{k=1}^m N_k$ denote the population size. Then

$$P(S_j \mid A) = \frac{P(A \mid S_j)P(S_j)}{\sum_{k=1}^m P(A \mid S_k)P(S_k)} = \frac{p_j \times \frac{N_j}{N}}{\sum_{k=1}^m p_k \times \frac{N_k}{N}} = \frac{p_j N_j}{\sum_{k=1}^m p_k N_k},$$

i.e. $\frac{100 p_j N_j}{\sum_{k=1}^m p_k N_k}$% of members having the specified attribute belong to stratum j $(1 \le j \le m)$.

4.78 a) $P(p = 0.4 \mid 6 \text{ heads in 10 tosses}) = \frac{P(6 \text{ heads in 10 tosses} \mid p=0.4)\frac{1}{2}}{P(6 \text{ heads in 10 tosses} \mid p=0.4)\frac{1}{2} + P(6 \text{ heads in 10 tosses} \mid p=0.55)\frac{1}{2}}$

$= \dfrac{\binom{10}{6}0.4^6 \times 0.6^4}{\binom{10}{6}0.4^6 \times 0.6^4 + \binom{10}{6}0.55^6 \times 0.45^4} \approx 0.3186$; $P(p = 0.55 \mid 6 \text{ heads in 10 tosses}) \approx 1 - 0.3186$ $= 0.6814$;

b) 0.0611, 0.9389; $P(p = 0.4 \mid 6 \text{ heads in first 10 tosses and 8 heads in the next 10 tosses})$

$$= \frac{\binom{10}{6}0.4^6 \times 0.6^4 \times \binom{10}{8}0.4^8 \times 0.6^2 \times \frac{1}{2}}{\binom{10}{6}0.4^6 \times 0.6^4 \times \binom{10}{8}0.4^8 \times 0.6^2 \times \frac{1}{2} + \binom{10}{6}0.55^6 \times 0.45^4 \times \binom{10}{8}0.55^8 \times 0.45^2 \times \frac{1}{2}}$$

$$= \frac{0.4^{14} \times 0.6^6}{0.4^{14} \times 0.6^6 + 0.55^{14} \times 0.45^6} \approx 0.0611;$$

$P(p = 0.55 \mid 6 \text{ heads in first 10 tosses and 8 heads in the next 10 tosses}) \approx 1 - 0.0611 = 0.9389$

4.5 Review Exercises

Basic Exercises

4.79 a) 3/8, since there are 3 red marbles among the marbles numbered #8 through #15.
b) 3/10, since exactly three of the ten red marbles are numbered in the #8–#15 range.
c) $P(\text{2nd is red} \mid \text{1st is red}) = 9/14$; $P(\text{2nd is red} \mid \text{1st is green}) = 10/14 = 5/7$.
d) $\dfrac{9}{14} \times \dfrac{10}{15} = \dfrac{3}{7}$.
e) $\dfrac{9}{14} \times \dfrac{10}{15} + \dfrac{10}{14} \times \dfrac{5}{15} = \dfrac{2}{3}$.
f) $\dfrac{\frac{3}{7}}{\frac{2}{3}} = \dfrac{9}{14}$, where we used (d) and (e).
g) $1 - \dfrac{9}{14} = \dfrac{5}{14}$, where we used (f) and the complementation rule for conditional probability.

4.80 $26/57 \approx 0.456$, since

$$P(\text{N has two aces and S has one ace}) = \frac{\binom{4}{2,1,1}\binom{48}{11,12,25}}{\binom{52}{13,13,26}}, \quad P(\text{N has two aces}) = \frac{\binom{4}{2}\binom{48}{11}}{\binom{52}{13}},$$

implying that

$$P(\text{S has one ace} \mid \text{N has two aces}) = \frac{\dfrac{\dbinom{4}{2,1,1}\dbinom{48}{11,12,25}}{\dbinom{52}{13,13,26}}}{\dfrac{\dbinom{4}{2}\dbinom{48}{11}}{\dbinom{52}{13}}} = \frac{26}{57}.$$

4.81 a) $\dfrac{3263}{9269} \approx 0.352$;
b) Not independent, since $P(\text{college} \mid \text{private}) = \dfrac{3263}{9269} \neq \dfrac{14889}{68335} = P(\text{college})$.
c) Conditional distributions of type by level and the marginal distribution of type are given by:

		Level			
		Elementary	High school	College	$P(T_i)$
Type	Public (T_1)	$\frac{33903}{38543} \approx 0.8796$	$\frac{13537}{14903} \approx 0.9083$	$\frac{11626}{14889} \approx 0.7808$	$\frac{59066}{68335} \approx 0.8644$
	Private (T_2)	$\frac{4640}{38543} \approx 0.1204$	$\frac{1366}{14903} \approx 0.0917$	$\frac{3263}{14889} \approx 0.2192$	$\frac{9269}{68335} \approx 0.1356$
	Total	1	1	1	1

d) The conditional distributions of level by type and the marginal distribution of level are given by:

		Type		
		Public	Private	$P(L_i)$
Level	Elementary (L_1)	$\frac{33903}{59066} \approx 0.574$	$\frac{4640}{9269} \approx 0.5006$	$\frac{38543}{68335} \approx 0.564$
	High school (L_2)	$\frac{13537}{59066} \approx 0.2292$	$\frac{1366}{9269} \approx 0.1474$	$\frac{14903}{68335} \approx 0.2181$
	College (L_3)	$\frac{11626}{59066} \approx 0.1968$	$\frac{3263}{9269} \approx 0.352$	$\frac{14889}{68335} \approx 0.2179$
	Total	1	1	1

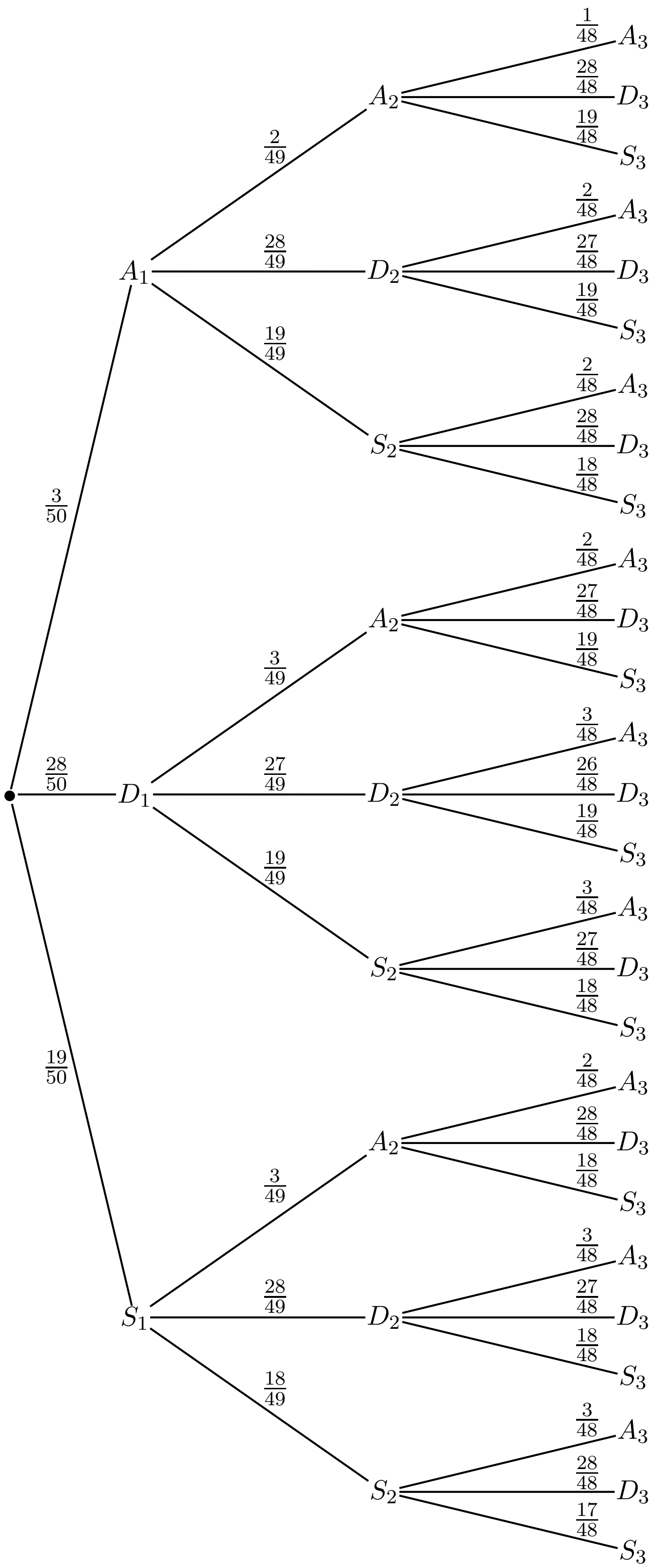
4.82 a) $\dfrac{3}{50} \times \dfrac{19}{49} \times \dfrac{18}{48} = \dfrac{171}{19600} \approx 0.0087$

b) $\dfrac{\binom{3}{1}\binom{19}{2}}{\binom{50}{3}} \approx 0.0262$

c) $\dfrac{28 \times 27 \times 26}{50 \times 49 \times 48} \approx 0.1671$

d) $\dfrac{28 \times 27 \times 26 + 3 \times 2 \times 1 + 19 \times 18 \times 17}{50 \times 49 \times 48} \approx 0.2166$

e) Let A_i denote the event that the ith student selected received a master of arts, let D_i denote the event that the ith student selected received a master of public administration, and let S_i denote the event that the ith student selected received a master of science, $i = 1, 2, 3$. Then the required tree diagram is of the form:



Event	Probability
$A_1 \cap A_2 \cap A_3$	0.00005
$A_1 \cap A_2 \cap D_3$	0.00143
$A_1 \cap A_2 \cap S_3$	0.00097
$A_1 \cap D_2 \cap A_3$	0.00143
$A_1 \cap D_2 \cap D_3$	0.01929
$A_1 \cap D_2 \cap S_3$	0.01357
$A_1 \cap S_2 \cap A_3$	0.00097
$A_1 \cap S_2 \cap D_3$	0.01357
$A_1 \cap S_2 \cap S_3$	0.00872
$D_1 \cap A_2 \cap A_3$	0.00143
$D_1 \cap A_2 \cap D_3$	0.01929
$D_1 \cap A_2 \cap S_3$	0.01357
$D_1 \cap D_2 \cap A_3$	0.01929
$D_1 \cap D_2 \cap D_3$	0.16714
$D_1 \cap D_2 \cap S_3$	0.12214
$D_1 \cap S_2 \cap A_3$	0.01357
$D_1 \cap S_2 \cap D_3$	0.12214
$D_1 \cap S_2 \cap S_3$	0.08143
$S_1 \cap A_2 \cap A_3$	0.00097
$S_1 \cap A_2 \cap D_3$	0.01357
$S_1 \cap A_2 \cap S_3$	0.00872
$S_1 \cap D_2 \cap A_3$	0.01357
$S_1 \cap D_2 \cap D_3$	0.12214
$S_1 \cap D_2 \cap S_3$	0.08143
$S_1 \cap S_2 \cap A_3$	0.00872
$S_1 \cap S_2 \cap D_3$	0.08143
$S_1 \cap S_2 \cap S_3$	0.04944

4.83 Special multiplication rule. Sampling with replacement makes selection of any given member of the sample independent from selection of all other members of the sample, thus the conditional probabilities, present in the general multiplication rule, are equal to the corresponding unconditional probabilities in the case of sampling with replacement.

4.84 80%. Let J be the event that a person questioned by Kushel enjoys his/her job, and L be the event that a person enjoys his/her personal life. Then

$$P(J \mid L) = \frac{P(J \cap L)}{1 - (P(J^c \cap L^c) + P(J \cap L^c))} = \frac{0.04}{1 - (0.15 + 0.8)} = \frac{4}{5} = 0.8$$

4.85 0; 1/3. Clearly, if the sum of the two dice is 9, the 1st die cannot come up 1, thus, the corresponding probability is 0. If the sum of the dice is 4, then the 1st die comes up 1 if and only if the outcome $(1, 3)$ occurs. Since the sum of the dice is 4 whenever one of the three outcomes (1,3),(2,2),(3,1) occurs, the required probability that 1st die is 1, given the sum of the dice is 4, is equal to 1/3.

4.86 The answers will vary.

4.87 $\frac{\binom{k-1}{M-1}}{\binom{N}{M}}$. Let D be the event that the kth item selected is defective, and A be the event that there are $M-1$ defective items among the first $k-1$ items selected. Then

$$P(k\text{th item selected is the last defective}) = P(D \cap A) = P(D \mid A)P(A)$$

$$= \frac{1}{N-(k-1)} \times \frac{\binom{M}{M-1}\binom{N-M}{k-1-(M-1)}}{\binom{N}{k-1}} = \frac{1}{N-k+1} \times \frac{\frac{M!}{(M-1)!} \cdot \frac{(N-M)!}{(k-M)!(N-k)!}}{\frac{N!}{(k-1)!(N-k+1)!}} = \frac{\binom{k-1}{M-1}}{\binom{N}{M}}.$$

4.88 a) 4/10. Let A_k be the event that there are k women among the first three students selected, $k = 0, 1, 2, 3$. Then

$$P(\text{4th selected student is female}) = \sum_{k=0}^{3} P(\text{4th selected student is female} \mid A_k)P(A_k)$$

$$= \sum_{k=0}^{3} \frac{4-k}{7} \cdot \frac{\binom{4}{k}\binom{6}{3-k}}{\binom{10}{3}} = 0.4$$

b) 4/10. By symmetry, the fourth student selected is as likely to be a woman as the first student selected, and the probability that the first student selected is female is 4/10.
c) 4/10. Let f_1, f_2, f_3, f_4 denote the four female students in a class of n students (with the remaining $n-4$ students being male), $n \geq 4$. The students are selected from the class one after another without replacement. Let $F_{i,n}$ be the event that f_i is the fourth student selected from the class (of n), $i = 1, \ldots, 4$. Clearly, $P(F_{i,n}) = P(F_{1,n})$ for all $i = 2, 3, 4$. Let us show that $P(F_{1,n}) = 1/n$ for all $n \geq 4$. Clearly the desired claim is true for $n = 4$, since $4P(F_{1,4}) = P(F_{1,4}) + \cdots + P(F_{4,4}) = 1$, implying that $P(F_{1,4}) = 1/4$. Next suppose that $P(F_{1,n}) = 1/n$ for $n = k$. Then, for $n = k+1$, we have that

$$\begin{aligned} P(F_{1,k+1}) &= P(F_{1,k+1} | \text{“}f_1 \text{ is among first } k \text{ selected”})P(\text{“}f_1 \text{ is among first } k \text{ selected”}) \\ &= P(F_{1,k})P(\text{“}f_1 \text{ is among first } k \text{ selected”}) = \frac{1}{k} \times \frac{k}{k+1} = \frac{1}{k+1}, \end{aligned}$$

where event "f_1 is among first k selected" refers to a class of size $k+1$, thus it is a complement of the event "f_1 is the last student selected from the class of size $(k+1)$", implying that

$$P(\text{“}f_1 \text{ is among first } k \text{ selected”}) = 1 - \frac{k!}{(k+1)!} = \frac{k}{k+1}.$$

By mathematical induction principle, it follows that $P(F_{1,n}) = 1/n$ for all $n \geq 4$. Therefore, for all $n \in \{4, 5 \ldots\}$,

$$P\left(\bigcup_{i=1}^{4} F_{i,n}\right) = \sum_{i=1}^{4} P(F_{i,n}) = 4P(F_{1,n}) = \frac{4}{n}.$$

Upon putting $n = 10$, we conclude that the required probability is equal to $4/10$.

4.89 a) 0.5177; By the inclusion-exclusion formula,

$$P(\text{at least one 5 in four rolls of a die}) = \binom{4}{1}\frac{1}{6} - \binom{4}{2}\frac{1}{6^2} + \binom{4}{3}\frac{1}{6^3} - \binom{4}{4}\frac{1}{6^4} = \frac{671}{1296} \approx 0.5177$$

b) 0.5177, since

$$P(\text{at least one 5 in four rolls of a die}) = 1 - P(\text{no 5s in four rolls}) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177$$

c) In the context of the problem it is natural to assume that tosses are independent of one another. This contextual independence is used in our solutions to (a) and (b).

4.90 a) $1 - (3/4)^8 \approx 0.8999$
b) $1 - (3/4)^8 - 8 \times (3/4)^7 \times (1/4) \approx 0.6329$
c) $0.6329/0.8999 \approx 0.7033$, where we used the answers to (a) and (b).
d) $\dfrac{\frac{3}{4} \cdot \left(1 - \left(\frac{3}{4}\right)^7 - 7 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^6\right)}{0.6329} \approx 0.658$

4.91 a) $0.91^4 \approx 0.6857$
b) $0.91^3 \times 0.09 \approx 0.0678$
c) $4 \times 0.91^3 \times 0.09 \approx 0.2713$
d) $\binom{4}{2} 0.91^2 0.09^2 \approx 0.0402$

4.92 Not independent. If the gender and activity limitation were independent, one should have had that $P(\text{limitation} \,|\, \text{female}) = P(\text{limitation} \,|\, \text{male})$, which is not the case here.

4.93 a) Yes, independent, since for all $a, b \in [0, 1]$,

$$P(\{(x, y) \in \Omega : x < a,\ y < b\}) = ab = P(\{(x, y) \in \Omega : x < a\})P(\{(x, y) \in \Omega : y < b\}),$$

where $\Omega = \{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$.
b) Yes, independent, since for all $a, b \in [0, 1]$ (with Ω as in (a)),

$$P(\{(x, y) \in \Omega : x < a,\ y > b\}) = a(1 - b) = P(\{(x, y) \in \Omega : x < a\})P(\{(x, y) \in \Omega : y > b\}).$$

4.94 a) No, not independent when $a \in (0,1)$ and $b \in (0,1)$. For $\Omega = \{(x,y) : 0 \le y \le x \le 1\}$, we have $P((x,y) \in \Omega : x < a) = \frac{a^2/2}{1/2} = a^2$, $P(\{(x,y) \in \Omega : y < b\}) = 1 - \frac{(1-b)^2/2}{1/2} = b(2-b)$, and

$$P(\{(x,y) \in \Omega : x{<}a,\ y{<}b\}) = \begin{cases} a^2, & \text{if } a \le b, \\ a^2 - (a-b)^2, & \text{if } a > b. \end{cases}$$

Therefore, it's easy to see that

$$P(\{(x,y) \in \Omega : x < a,\ y < b\}) \ne P(\{(x,y) \in \Omega : x < a\})P(\{(x,y) \in \Omega : y < b\}).$$

b) Yes, the events are independent, since

$$P(\{(x,y) \in \Omega\colon x < a\})P(\{(x,y) \in \Omega\colon y < bx\}) = a^2 b = \frac{\frac{a(ba)}{2}}{\frac{1}{2}} = P(\{(x,y) \in \Omega\colon x{<}a,\ y{<}bx\}).$$

4.95 $1 - 0.9999^{748} \approx 0.0721$, i.e. there is about 7.21% chance that at least one "critically 1" item fails.

4.96 a) $0.208 \times 0.471 + 0.276 \times 0.529 \approx 0.244$;

b) 0.471;

c) $\dfrac{0.208 \times 0.471}{0.244} \approx 0.4015$

d) Approximately 24.4% of U.S. engineers and scientists have masters degree as their highest degree obtained. 47.1% of all U.S. engineers and scientists are engineers. Among the U.S. engineers and scientists with masters degree as their highest degree obtained approximately 40.15% are engineers.

4.97 a) 2/3, since the required conditional probability equals to

$$\frac{P(\text{both sides are red})}{P(\text{first side is red})} = \frac{\frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{2}{3}.$$

b) The red face showing is equally likely to be any one of the three red sides, and the other side can be red in two of those three possibilities.

4.98 a) 2.3%, since

$$P(\text{a randomly selected item is defective}) = 0.03 \times 0.5 + 0.02 \times 0.3 + 0.01 \times 0.2 = 0.023$$

b) 26.1%, since

$$P(\text{item is from Factory II} \,|\, \text{item is defective}) = \frac{0.02 \times 0.3}{0.023} = \frac{6}{23} \approx 0.261$$

4.99

$$P(\text{2nd side is head} \,|\, \text{1st side is head}) = \frac{P(\text{both sides are heads})}{P(\text{1st side is head})} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

(In the above, $P(\text{1st side is head}) = \frac{1}{2}$ can be established either directly by symmetry, or by the law of total probability, since $P(\text{1st side is head}) = \frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 = \frac{1}{2}$).

4.100 Let A be the event that the student knows the answer, and C be the event that the student gives the correct answer to the question.

a) $\dfrac{500p}{(4p+1)}\%$, since

$$P(A \mid C) = \frac{p}{p + \frac{1}{5} \cdot (1-p)} = \frac{5p}{4p+1}.$$

b) If $p = 10^{-6}$, then

$$P(A \mid C) = \frac{5p}{4p+1} = \frac{5}{4+\frac{1}{p}} = \frac{5}{4+10^6} \approx \frac{5}{10^6},$$

i.e. the probability is approximately five in one million that a student who answers correctly knows the answer.

c) If $p = 1 - 10^{-6}$, then $1 - p = 10^{-6}$ and $p \approx 1$, implying that

$$P(A^c \mid C) = 1 - \frac{5p}{4p+1} = \frac{1-p}{4p+1} \approx \frac{10^{-6}}{5} = \frac{1}{5 \times 10^6},$$

i.e. the probability is about one in five million that a student who answers correctly is ignorant of the answer.

Theory Exercises

4.101 a) For each $N \in \mathcal{N}$, by independence of $A_1, A_2, \ldots,$

$$P\Big(\bigcap_{n=1}^{N} A_n\Big) = \prod_{n=1}^{N} P(A_n).$$

Let $B_N = \bigcap_{n=1}^{N} A_n$. Then, in view of $B_N \supset B_{N+1}$ for all $N \in \mathcal{N}$ and continuity properties of the probability measure (Proposition 2.11), one must have that

$$P\Big(\bigcap_{N=1}^{\infty} B_N\Big) = \lim_{N\to\infty} P(B_N) = \lim_{N\to\infty} \prod_{n=1}^{N} P(A_n) = \prod_{n=1}^{\infty} P(A_n).$$

The required conclusion then easily follows upon noting that $\bigcap_{N=1}^{\infty} B_N = \bigcap_{N=1}^{\infty} (\bigcap_{n=1}^{N} A_n) = \bigcap_{n=1}^{\infty} A_n$.

b) Using the complementation rule, De Morgan's law, independence of $A_1^c, A_2^c, \ldots$ (which holds by Proposition 4.5) and the result of (a), one obtains that

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = 1 - P\Big(\bigcap_{n=1}^{\infty} A_n^c\Big) = 1 - \prod_{n=1}^{\infty} P(A_n^c) = 1 - \prod_{n=1}^{\infty} (1 - P(A_n)).$$

4.102 Using definition of conditional probability and the multiplication rule (Proposition 4.2), one obtains that

$$\frac{P(A \mid B)}{P(A^c \mid B)} = \frac{P(A \cap B)/P(B)}{P(A^c \cap B)/P(B)} = \frac{P(A \cap B)}{P(A^c \cap B)} = \frac{P(B \mid A)P(A)}{P(B \mid A^c)P(A^c)} = \frac{P(A)}{P(A^c)} \cdot \frac{P(B \mid A)}{P(B \mid A^c)}.$$

Advanced Exercises

4.103 a) For $i = 0, 1, \cdots, \min(k, d)$,

$$p_i = \sum_{\ell=\max(i,\, n+d-N)}^{\min(n+i-k,\, d)} \frac{\binom{\ell}{i}\binom{n-\ell}{k-i}}{\binom{n}{k}} \cdot \frac{\binom{d}{\ell}\binom{N-d}{n-\ell}}{\binom{N}{n}};$$

and $p_i = 0$ for all $i \notin \{0, 1, \cdots, \min(k, d)\}$.
To obtain the answer, note first that $P(\text{there are } \ell \text{ defectives in the 1st sample}) = 0$ unless $\ell \in \{0, \cdots, \min(d, n)\}$ and $n - \ell \le N - d$.
On the other hand, for all integers $\ell \in \{\max(0, n + d - N), \cdots, \min(n, d)\}$,

$$P(\text{there are } \ell \text{ defectives in the 1st sample}) = \frac{\binom{d}{\ell}\binom{N-d}{n-\ell}}{\binom{N}{n}}.$$

Similarly, for $i \in \{0, \cdots, \ell\}$ such that $n - \ell \ge k - i$, we have that

$$P(\text{there are } i \text{ defectives in the 2nd sample} \,|\, \text{there are } \ell \text{ defectives in the 1st sample}) = \frac{\binom{\ell}{i}\binom{n-\ell}{k-i}}{\binom{n}{k}},$$

and the above conditional probability is zero otherwise. Thus, by the law of total probability,

$$p_i = \sum_{\ell=\max(0,n+d-N)}^{\min(n,d)} P(i \text{ defectives in the 2nd sample} \,|\, \ell \text{ defectives in the 1st sample}) \cdot \frac{\binom{d}{\ell}\binom{N-d}{n-\ell}}{\binom{N}{n}}$$

$$= \sum_{\ell=\max(i,\, n+d-N)}^{\min(n+i-k,\, d)} \frac{\binom{\ell}{i}\binom{n-\ell}{k-i}}{\binom{n}{k}} \cdot \frac{\binom{d}{\ell}\binom{N-d}{n-\ell}}{\binom{N}{n}}.$$

b) $p_0 \approx 0.9035497$, $p_1 \approx 0.09295779$, $p_2 \approx 0.003435811$, $p_3 \approx 5.63 \times 10^{-5}$, $p_4 \approx 4.03 \times 10^{-7}$, $p_5 \approx 9.87 \times 10^{-10}$.
c)

$$\frac{\binom{N-d}{n} / \binom{N}{n}}{p_0} = \frac{\binom{N-d}{n}\binom{n}{k}}{\sum_{\ell=\max(0,n+d-N)}^{\min(n-k,d)} \binom{d}{\ell}\binom{N-d}{n-\ell}\binom{n-\ell}{k}}.$$

4.104 a) When sampling with replacement, for $i = 0, \cdots, k$,

$$p_i = \sum_{\ell=0}^{n} \binom{k}{i} (\ell/n)^i ((n-\ell)/n)^{k-i} \binom{n}{\ell} (d/N)^\ell ((N-d)/N)^{n-\ell},$$

since, for $\ell = 0, \cdots, n$,

$$P(\ell \text{ defectives in the 1st sample}) = \binom{n}{\ell} (d/N)^\ell ((N-d)/N)^{n-\ell},$$

$$P(i \text{ defectives in the 2nd sample} \,|\, \ell \text{ defectives in the 1st sample}) = \binom{k}{i}(\ell/n)^i((n-\ell)/n)^{k-i}.$$

b) $p_0 \approx 0.9122056$, $p_1 \approx 0.0767048$, $p_2 \approx 0.01005675$, $p_3 \approx 0.0009548676$, $p_4 \approx 7.32 \times 10^{-5}$, $p_5 \approx 4.6 \times 10^{-6}$, $p_6 \approx 2.33 \times 10^{-7}$, $p_7 \approx 9.15 \times 10^{-9}$, $p_8 \approx 2.63 \times 10^{-10}$, $p_9 \approx 4.96 \times 10^{-12}$, $p_{10} \approx 4.59 \times 10^{-14}$.
c)

$$\frac{((N-d)/N)^n}{p_0} = \frac{(N-d)^n}{\sum_{\ell=0}^{n}((n-\ell)/n)^k \binom{n}{\ell} d^\ell (N-d)^{n-\ell}}.$$

4.105 275/1296. Let X be the number of tosses until first 6 occurs. Then

$$P(X \text{ is even}) = \sum_{k=1}^{\infty} P(X=2k) = \sum_{k=1}^{\infty}(5/6)^{2k-1}(1/6) = \frac{5}{36}\sum_{m=0}^{\infty}(25/36)^m = \frac{5}{36}\cdot\frac{1}{1-\frac{25}{36}} = \frac{5}{11},$$

implying that

$$P(X=4 \,|\, X \text{ is even}) = \frac{(5/6)^3(1/6)}{5/11} = \frac{275}{1296} \approx 0.2122$$

4.106 Prisoner's paradox. No, the probability remains equal to 1/3. Let A be the event that the jailer tells that a will be freed, and let B be the event that the jailer tells that b will be freed. Also, let E_b be the event that b is to be executed, and E_c be the event that c is to be executed. Then

$$P(E_c \,|\, A) = \frac{P(A \,|\, E_c)P(E_c)}{P(A \,|\, E_c)P(E_c) + P(A \,|\, E_b)P(E_b)} = \frac{\frac{1}{2}\times\frac{1}{3}}{\frac{1}{2}\times\frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{3}.$$

Similarly, $P(E_c \,|\, B) = 1/3$.

4.107 p. By symmetry, the kth member selected is as likely to have the specified attribute as the 1st member selected.

4.108 $r/(r+g)$. Let R_k denote the event that a ball chosen from urn #k is red. Let $p_k = P(R_k)$, $k \in \{1, \cdots, n\}$. Then for all $k \in \{2, \cdots, n\}$,

$$\begin{aligned} p_k = P(R_k) &= P(R_k \,|\, R_{k-1})P(R_{k-1}) + P(R_k \,|\, R_{k-1}^c)P(R_{k-1}^c) \\ &= \frac{r+1}{r+g+1}\cdot p_{k-1} + \frac{r}{r+g+1}\cdot(1-p_{k-1}) = \frac{r+p_{k-1}}{r+g+1} \end{aligned}$$

and $p_1 = \dfrac{r}{r+g}$. The required answer then follows by induction.

4.109 a) The required conditional probability equals to

$$\frac{P(\text{all } n+1 \text{ children selected are boys})}{P(\text{first } n \text{ children selected are boys})} = \frac{\sum_{k=0}^{N}(k/N)^{n+1}\cdot\frac{1}{N+1}}{\sum_{k=0}^{N}(k/N)^n\cdot\frac{1}{N+1}} = \frac{\sum_{k=0}^{N}(k/N)^{n+1}}{\sum_{k=0}^{N}(k/N)^n},$$

by the law of total probability and since $P(\text{all } n \text{ children selected are boys} \mid k\text{th family is chosen}) = (k/N)^n$.
b) If N is large, we can use Riemann integral approximations

$$\frac{1}{N}\sum_{k=0}^{N}(k/N)^{n+1} \approx \int_0^1 x^{n+1}dx = \frac{1}{n+2}, \qquad \frac{1}{N}\sum_{k=0}^{N}(k/N)^{n} \approx \int_0^1 x^{n}dx = \frac{1}{n+1},$$

and conclude that the answer to (a) is given by:

$$\frac{\sum_{k=0}^{N}(k/N)^{n+1}}{\sum_{k=0}^{N}(k/N)^{n}} \approx \frac{\frac{1}{n+2}}{\frac{1}{n+1}} = \frac{n+1}{n+2}.$$

4.110 The required result follows upon noting that any sequence of outcomes for the n repetitions of the experiment that leads to outcome E_i occurring k_i times for $i = 1, \cdots, m$, will have probability $p_1^{k_1}p_2^{k_2}\dots p_m^{k_m}$, by independence of repetitions. As there are $\binom{n}{k_1,\dots,k_m} = \frac{n!}{k_1!k_2!\dots k_m!}$ such sequences of outcomes (since there are $n!/(k_1!k_2!\dots k_m!)$ different permutations of n elements of which k_1 are alike, k_2 are alike,..., k_m are alike), it follows that

$$P(n(E_1) = k_1, \dots, n(E_m) = k_m) = \binom{n}{k_1, \dots, k_m} p_1^{k_1}\dots p_m^{k_m}.$$

4.111 $i/10$. Let E_i denote the event that you ruin your opponent, given that your starting capital is equal to \$$i$ and your opponent's starting capital is equal to \$$(10-i)$, and let $p_i = P(E_i)$, where $i = 0, \dots, 10$. Then

$$p_i = P(E_i \mid \text{H on 1st toss})P(\text{H}) + P(E_i \mid \text{T on 1st toss})P(\text{T}) = p_{i+1}\cdot\frac{1}{2} + p_{i-1}\cdot\frac{1}{2},$$

for all $i = 1, \dots, 9$, and obviously $p_0 = 0$ and $p_{10} = 1$. Therefore,

$$p_{i+1} = 2p_i - p_{i-1} = 2(2p_{i-1} - p_{i-2}) - p_{i-1} = 3p_{i-1} - 2p_{i-2} = \cdots = (2+k)p_{i-k} - (1+k)p_{i-k-1},$$

for all $k = 0, \cdots, i-1$. Taking $k = i-1$ in the above equation yields that

$$p_{i+1} = (i+1)p_1 - ip_0 = (i+1)p_1, \quad i \in \{0, \cdots, 9\},$$

since $p_0 = 0$. Therefore, $p_{10} = 10p_1 = 1$, implying that $p_1 = 1/10$ and, thus, $p_i = ip_1 = i/10$ for all $i \in \{0, \cdots, 10\}$.

4.112 Randomized Response: Let S denote the event that a sensitive question is chosen, I denote the event that an innocuous question is chosen and Y be the event that the answer is positive. Suppose $P(Y \mid I)$ is known. Let X be the number of positive responses in the sample, and n be the (known) size of the sample. Then, assuming that the coin is balanced (thus, $P(I) = P(S) = 1/2$), by the law of total probability,

$$P(Y) = P(Y \mid S)\frac{1}{2} + P(Y \mid I)\frac{1}{2}, \text{ thus, } P(Y \mid S) = 2P(Y) - P(Y \mid I).$$

On the other hand, $P(Y)$ is naturally estimated by X/n, which is a proportion of positive responses in the sample. Thus, a natural estimate of $P(Y \mid S)$ is given by $\dfrac{2X}{n} - P(Y \mid I)$.

4.113 Assume that $0 < P(\text{head}) = p < 1$ for the coin in question. Consider successive pairs of tosses of the coin. If both of a pair's tosses show the same side, disregard and toss another pair. When pair's tosses differ, treat outcome $\{\text{HT}\}$ as a "head" of a balanced coin, and treat $\{\text{TH}\}$ as a "tail" of a balanced coin (since $P(\{\text{HT}\}) = P(\{\text{TH}\})$). This is a well-defined procedure for simulating a balanced coin provided that only finitely many pairs of tosses of the original coin are needed to simulate one toss of a balanced coin. The latter indeed is true (with probability one), since

$$\sum_{k=1}^{\infty} P(\text{1st pair showing either HT or TH occurs on the } k\text{th pair's toss})$$

$$= \sum_{k=1}^{\infty} \left[(1-p)^2 + p^2\right]^{k-1} 2p(1-p) = 1.$$

CHAPTER FIVE

Instructor's Solutions Manual

Anna Amirdjanova

University of Michigan, Ann Arbor

Neil A. Weiss

Arizona State University

FOR

A Course In

Probability

Neil A. Weiss

Arizona State University

PEARSON Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

Publisher: Greg Tobin
Editor-in-Chief: Deirdre Lynch
Associate Editor: Sara Oliver Gordus
Editorial Assistant: Christina Lepre
Production Coordinator: Kayla Smith-Tarbox
Senior Author Support/Technology Specialist: Joe Vetere
Compositor: Anna Amirdjanova and Neil A. Weiss
Accuracy Checker: Delray Schultz
Proofreader: Carol A. Weiss

Chapter 5

Discrete Random Variables and Their Distributions

5.1 From Variables to Random Variables

Basic Exercises

5.1 Answers will vary.

5.2

a) Yes, X is a random variable because it is a real-valued function whose domain is the sample space of a random experiment, namely, the random experiment of tossing a balanced die twice and observing the two faces showing. Yes, X is a discrete random variable because it is a random variable with a finite (and, hence, countable) range, namely, the set $\{2, 3, \ldots, 12\}$.

b)–f) We have the following table:

Part	Event	The sum of the two faces showing ...	Subset of the sample space
b)	$\{X = 7\}$	is seven.	$\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$
c)	$\{X > 10\}$	exceeds 10 (i.e., is either 11 or 12).	$\{(5, 6), (6, 5), (6, 6)\}$
d)	$\{X = 2 \text{ or } 12\}$	is either 2 or 12.	$\{(1, 1), (6, 6)\}$
e)	$\{4 \le X \le 6\}$	is four, five, or six.	$\{(1, 3), (2, 2), (3, 1), (1, 4), (2, 3), (3, 2), (4, 1), (1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$
f)	$\{X \in A\}$	is an even integer.	$\{(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), (4, 2), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (6, 2), (6, 4), (6, 6)\}$

5.3

a) Yes, X is a random variable because it is a real-valued function whose domain is the sample space of a random experiment, namely, the random experiment of observing which of five components of a unit are working and which have failed. Yes, X is a discrete random variable because it is a random variable with a finite (and, hence, countable) range, namely, the set $\{0, 1, 2, 3, 4, 5\}$.

b)–e) We have the following table:

Part	Event	In words ...	Subset of the sample space
b)	$\{X \ge 4\}$	At least four (i.e., four or five) of the components are working.	$\{(s, s, s, s, f), (s, s, s, f, s), (s, s, f, s, s), (s, f, s, s, s), (f, s, s, s, s), (s, s, s, s, s)\}$
c)	$\{X = 0\}$	None of the components are working.	$\{(f, f, f, f, f)\}$
d)	$\{X = 5\}$	All of the components are working.	$\{(s, s, s, s, s)\}$
e)	$\{X \ge 1\}$	At least one of the components is working.	$\{(f, f, f, f, f)\}^c$

f) $\{X = 2\}$ **g)** $\{X \geq 2\}$ **h)** $\{X \leq 2\}$ **i)** $\{2 \leq X \leq 4\}$

5.4
a) Let m denote male and f denote female. Then we can represent each possible outcome as an ordered triple, where each entry of the triple is either m or f. For instance, (m, f, f) represents the outcome that the first born is male, the second born is female, and the third born is female. Therefore, a sample space for this random experiment is $\Omega = \{\,(x_1, x_2, x_3) : x_j \in \{m, f\},\ 1 \leq j \leq 3\,\}$.
b) We have the following table:

ω	(m,m,m)	(m,m,f)	(m,f,m)	(m,f,f)	(f,m,m)	(f,m,f)	(f,f,m)	(f,f,f)
$Y(\omega)$	0	1	1	2	1	2	2	3

c) Yes, Y is a random variable because it is a real-valued function whose domain is the sample space of a random experiment. Yes, Y is a discrete random variable because it is a random variable with a finite (and, hence, countable) range, namely, the set $\{0, 1, 2, 3\}$.

5.5 Take $\Omega = \{\,(x, y) : -3 \leq x \leq 3, -3 \leq y \leq 3\,\}$ and, for brevity, set $d(x, y) = \sqrt{x^2 + y^2}$.
a) We have

$$S(x, y) = \begin{cases} 10, & \text{if } d(x, y) \leq 1; \\ 5, & \text{if } 1 < d(x, y) \leq 2; \\ 0, & \text{otherwise.} \end{cases}$$

b) Yes, S is a discrete random variable because it is a random variable with a finite (and, hence, countable) range, namely, the set $\{0, 5, 10\}$.
c)–h) We have the following table:

Part	Event	The archer's score is ...	Subset of the sample space
c)	$\{S = 5\}$	5 points; that is, she hits the ring with inner radius 1 foot and outer radius 2 feet centered at the origin.	$\{\,(x, y) \in \Omega : 1 < d(x, y) \leq 2\,\}$
d)	$\{S > 0\}$	positive; that is, she hits the disk of radius 2 feet centered at the origin.	$\{\,(x, y) \in \Omega : d(x, y) \leq 2\,\}$
e)	$\{S \leq 7\}$	at most 7 points; that is, she does not hit the bull's eye.	$\{\,(x, y) \in \Omega : d(x, y) > 1\,\}$
f)	$\{5 < S \leq 15\}$	more than 5 points and at most 15 points; that is, she hits the bull's eye.	$\{\,(x, y) \in \Omega : d(x, y) \leq 1\,\}$
g)	$\{S < 15\}$	less than 15 points, which is certain.	Ω
h)	$\{S < 0\}$	negative, which is impossible.	$\emptyset$

5.6
a) Let n denote nondefective and d denote defective. Then we can represent each possible outcome as an ordered five-tuple, where each entry of the five-tuple is either n or d. For instance, (n, d, n, n, d) represents the outcome that the first, third, and fourth parts selected are nondefective and the second and fifth parts selected are defective. Thus, a sample space for this random experiment is

$$\Omega = \left\{(x_1, x_2, x_3, x_4, x_5) : x_j \in \{n, d\},\ 1 \leq j \leq 5\right\}.$$

b) We have $Y = 1$ if $X \leq 1$, and $Y = 0$ if $X > 1$.

5.7

a) Let r denote red and w denote white. Then we can represent each possible outcome as an ordered triple, where the first entry of the triple is the number of the urn chosen, and the second and third entries are the colors of the first and second balls selected, respectively. For instance (II, r, w) represents the outcome that Urn II is chosen, the first ball selected is red, and the second ball selected is white. Hence, a possible sample space is $\Omega = \{(\text{I}, r, r), (\text{I}, r, w), (\text{I}, w, r), (\text{II}, r, w), (\text{II}, w, r), (\text{II}, w, w)\}$.

b) We have the following table:

ω	(I, r, r)	(I, r, w)	(I, w, r)	(II, r, w)	(II, w, r)	(II, w, w)
$X(\omega)$	2	1	1	1	1	0

5.8

a) As the interview pool consists of three people, the number of women in the interview pool must be between zero and three, inclusive. Hence, the possible values of the random variable X are 0, 1, 2, and 3.

b) The number of ways that the interview pool can contain exactly x women ($x = 0, 1, 2, 3$) is the number of ways that x women can be chosen from the five women and $3 - x$ men can be chosen from the six men. The number of ways that x women can be chosen from the five women is $\binom{5}{x}$, and the number of ways that $3 - x$ men can be chosen from the six men is $\binom{6}{3-x}$. Consequently, by the BCR, the number of ways that the interview pool can contain exactly x women equals $\binom{5}{x}\binom{6}{3-x}$. We, therefore, have the following table:

x	Number of ways $\{X = x\}$ can occur
0	$\binom{5}{0}\binom{6}{3} = 1 \cdot 20 = 20$
1	$\binom{5}{1}\binom{6}{2} = 5 \cdot 15 = 75$
2	$\binom{5}{2}\binom{6}{1} = 10 \cdot 6 = 60$
3	$\binom{5}{3}\binom{6}{0} = 10 \cdot 1 = 10$

c) $\{X \leq 1\}$ is the event that at most one (i.e., 0 or 1) woman is in the interview pool. From the table in part (b), we see that the number of ways that event can occur is $20 + 75 = 95$.

5.9

a) We have the following table:

ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$Y(\omega)$	3	1	1	-1	1	-1	-1	-3

b) The event $\{Y = 0\}$ is, in words, that the number of heads equals the number of tails, which is impossible. Hence, as a subset of the sample space, we have $\{Y = 0\} = \emptyset$.

5.10 Denote the two salads by s_1 and s_2, the three entrees by e_1, e_2, and e_3, and the two desserts by d_1 and d_2. Also, let s_0, e_0, and d_0 represent a choice of no salad, entree, and dessert, respectively. Thus, each possible choice of a meal is of the form $s_i e_j d_k$, where $i, k \in \{0, 1, 2\}$ and $j \in \{0, 1, 2, 3\}$. We note that the sample space consists of $3 \cdot 4 \cdot 3 = 36$ possible outcomes.

a) We have the following table:

ω	$X(\omega)$	ω	$X(\omega)$	ω	$X(\omega)$	ω	$X(\omega)$
$s_0e_0d_0$	0.00	$s_0e_3d_0$	3.50	$s_1e_2d_0$	4.00	$s_2e_1d_0$	4.00
$s_0e_0d_1$	1.50	$s_0e_3d_1$	5.00	$s_1e_2d_1$	5.50	$s_2e_1d_1$	5.50
$s_0e_0d_2$	1.50	$s_0e_3d_2$	5.00	$s_1e_2d_2$	5.50	$s_2e_1d_2$	5.50
$s_0e_1d_0$	2.50	$s_1e_0d_0$	1.00	$s_1e_3d_0$	4.50	$s_2e_2d_0$	4.50
$s_0e_1d_1$	4.00	$s_1e_0d_1$	2.50	$s_1e_3d_1$	6.00	$s_2e_2d_1$	6.00
$s_0e_1d_2$	4.00	$s_1e_0d_2$	2.50	$s_1e_3d_2$	6.00	$s_2e_2d_2$	6.00
$s_0e_2d_0$	3.00	$s_1e_1d_0$	3.50	$s_2e_0d_0$	1.50	$s_2e_3d_0$	5.00
$s_0e_2d_1$	4.50	$s_1e_1d_1$	5.00	$s_2e_0d_1$	3.00	$s_2e_3d_1$	6.50
$s_0e_2d_2$	4.50	$s_1e_1d_2$	5.00	$s_2e_0d_2$	3.00	$s_2e_3d_2$	6.50

b) The event $\{X \le 3\}$ is, in words, that the meal costs \$3.00 or less. As as subset of the sample space, we have that

$$\{X \le 3\} = \{s_0e_0d_0, s_0e_0d_1, s_0e_0d_2, s_0e_1d_0, s_0e_2d_0, s_1e_0d_0, s_1e_0d_1, s_1e_0d_2, s_2e_0d_0, s_2e_0d_1, s_2e_0d_2\}.$$

5.11 We can take the sample space to be $\Omega = \{w, \ell w, \ell\ell w, \dots\}$, where

$$\underbrace{\ell \dots \ell}_{n-1 \text{ times}} w$$

represents the outcome that you don't win a prize the first $n-1$ weeks and you do win a prize the nth week; that is, it takes n weeks before you win a prize.

a) Yes, W is a random variable because it is a real-valued function whose domain is the sample space Ω. We have

$$W(\underbrace{\ell \dots \ell}_{n-1 \text{ times}} w) = n, \qquad n \in \mathcal{N}.$$

b) Yes, W is a discrete random variable because, as we see from part (a), the range of W is $\mathcal{N}$.

c) $\{W > 1\}$ is the event that it takes you more than 1 week to win a prize or, equivalently, you don't win a prize the first week.

d) $\{W \le 10\}$ is the event that you win a prize by the 10th week.

e) $\{15 \le W < 20\}$ is the event that you first win a prize between weeks 15 and 19, inclusive; that is, it takes you at least 15 weeks but less than 20 weeks to win a prize.

5.12

a) Define $g(x) = \lfloor m + (n-m+1)x \rfloor$. For $0 < x < 1$, we have

$$m - 1 < m + (n-m+1)x - 1 < g(x) \le m + (n-m+1)x < n + 1.$$

Because g is integer valued, it follows that $g\big((0,1)\big) \subset \{m, m+1, \dots, n\}$. However, if k is an integer with $m+1 \le k \le n$, then

$$0 < \frac{k-m}{n-m+1} < 1 \qquad \text{and} \qquad g\left(\frac{k-m}{n-m+1}\right) = k.$$

Furthermore,

$$g\left(\frac{1}{2(n-m+1)}\right) = \lfloor m + 1/2 \rfloor = m.$$

Thus, $g\big((0,1)\big) = \{m, m+1, \dots, n\}$ and, therefore, the possible values of Y are $m, m+1, \dots, n$.

b) Yes, Y is a discrete random variable because, from part (a), its range is finite and, hence, countable.

c) For $u \in \mathcal{R}$, we have $\lfloor u \rfloor \leq u < \lfloor u \rfloor + 1$. Hence,

$$Y = y \quad \text{if and only if} \quad y \leq m + (n - m + 1)X < y + 1.$$

Therefore,

$$\{Y = y\} = \left\{ \frac{y - m}{n - m + 1} \leq X < \frac{y + 1 - m}{n - m + 1} \right\}.$$

Theory Exercises

5.13 Let X be a random variable with countable range, say, R. Then we have $\{X \in R\} = \Omega$. Hence,

$$P(X \in R) = P(\Omega) = 1,$$

so that Definition 5.2 holds with $K = R$.

5.14
a) Suppose to the contrary that there is an $n \in \mathcal{N}$ such that $\{ x \in \mathcal{R} : P(X = x) > 1/n \}$ contains more than $n - 1$ elements. Then there is an $m \in \mathcal{N}$ with $m \geq n$ and real numbers $x_1, \ldots, x_m$ such that $P(X = x_j) > 1/n$ for $j = 1, \ldots, m$. Therefore,

$$P\big(X \in \{x_1, \ldots, x_m\}\big) = \sum_{j=1}^{m} P(X = x_j) > m/n \geq 1.$$

Thus, $P\big(X \in \{x_1, \ldots, x_m\}\big) > 1$, which is impossible.
b) We have

$$\{ x \in \mathcal{R} : P(X = x) \neq 0 \} = \bigcup_{n=1}^{\infty} \{ x \in \mathcal{R} : P(X = x) > 1/n \}.$$

From part (b), each set in the union is finite and, hence, countable. Therefore, the set on the left of the preceding display is countable, being a countable union of countable sets.
c) For any random variable X, we have $P(X = x) = 0$ except for countably many real numbers, x.

Advanced Exercises

5.15 We provide two examples of which the first uses the hint. Let $\Omega = [0, 1]$ and let P be length in the extended sense. Note that P is a probability measure on Ω. Define $X(\omega) = \omega$ if $\omega \in C$, and $X(\omega) = 2$ otherwise, where C is a set as described in the hint. The range of X is $C \cup \{2\}$, which is uncountable because C is uncountable. Let $K = \{2\}$. Then K is countable and

$$P(X \in K) = P(X = 2) = P\big([0, 1] \setminus C\big) = P\big([0, 1]\big) - P(C) = |[0, 1]| - |C| = 1 - 0 = 1.$$

Hence, X is a discrete random variable with an uncountable range.

An a second example, let $\Omega = [0, 1]$ and define P on the events of Ω by

$$P(E) = \begin{cases} 1, & \text{if } 1 \in E; \\ 0, & \text{if } 1 \notin E. \end{cases}$$

It is easy to see that P is a probability measure on Ω. Now, define $X(\omega) = \omega$ for all $\omega \in \Omega$. The range of X is $[0, 1]$, which is uncountable. Let $K = \{1\}$. Then K is countable and

$$P(X \in K) = P(X = 1) = P(\{1\}) = 1.$$

Hence, X is a discrete random variable with an uncountable range.

5.16 From Exercise 5.14, we know that $A = \{ x \in \mathcal{R} : P(X = x) \neq 0 \}$ is countable. Let $x_1, x_2, \ldots$ be an enumeration of A. As probabilities are always between 0 and 1, inclusive, we have $0 \leq P(X \in A) \leq 1$, However, by the additivity axiom,

$$\begin{aligned}\sum_x P(X = x) &= \sum_{x \in A} P(X = x) + \sum_{x \notin A} P(X = x) = \sum_{n=1}^{\infty} P(X = x_n) + 0 \\ &= \sum_{n=1}^{\infty} P\big(X \in \{x_n\}\big) = P\left(X \in \bigcup_{n=1}^{\infty} \{x_n\}\right) = P(X \in A).\end{aligned}$$

Hence,

$$0 \leq \sum_x P(X = x) \leq 1.$$

5.17 Suppose X is a discrete random variable. Then there is a countable set K such that $P(X \in K) = 1$. Applying the additivity axiom, we get

$$1 = P(X \in K) = \sum_{x \in K} P(X = x) \leq \sum_x P(X = x) \leq 1,$$

where the last inequality follows from Exercise 5.16. Hence, $\sum_x P(X = x) = 1$.

Conversely, suppose $\sum_x P(X = x) = 1$. From Exercise 5.14, the set $A = \{x \in \mathcal{R} : P(X = x) \neq 0\}$ is countable. Applying the additivity axiom, we get

$$1 = \sum_x P(X = x) = \sum_{x \in A} P(X = x) = P(X \in A).$$

Hence, there is a countable set A such that $P(X \in A) = 1$; that is, X is a discrete random variable.

5.18 Suppose that X is a discrete random variable. Then, from Definition 5.2, there is a countable set $K \subset \mathcal{R}$ such that $P(X \in K) = 1$. From the complementation rule,

$$P(X \in K^c) = 1 - P(X \in K) = 1 - 1 = 0,$$

and, in particular then, $P(X = x) = 0$ if $x \notin K$. Let $x_0 \in K^c$ and define

$$X_0(\omega) = \begin{cases} X(\omega), & \text{if } X(\omega) \in K; \\ x_0, & \text{if } X(\omega) \notin K. \end{cases}$$

The range of X_0 is a subset of (perhaps equal to) $K \cup \{x_0\}$, which is countable, being the union of two countable sets. Therefore, the range of X_0 is countable, being a subset of a countable set. We note that

$$\{X_0 = x\} = \begin{cases} \{X = x\}, & \text{if } x \in K; \\ \{X \in K^c\}, & \text{if } x = x_0; \\ \emptyset, & \text{otherwise.} \end{cases}$$

Hence,

$$\begin{aligned} P(X_0 = x) &= \begin{cases} P(X = x), & \text{if } x \in K; \\ P(X \in K^c), & \text{if } x = x_0; \\ 0, & \text{otherwise.} \end{cases} \\ &= \begin{cases} P(X = x), & \text{if } x \in K; \\ 0, & \text{if } x \notin K. \end{cases} \\ &= P(X = x). \end{aligned}$$

Thus, X_0 is a random variable with a countable range such that $P(X_0 = x) = P(X = x)$ for all $x \in \mathcal{R}$.

Conversely, suppose that there is a random variable X_0 with a countable range such that

$$P(X_0 = x) = P(X = x), \qquad x \in \mathcal{R}.$$

Let R denote the range of X_0. Applying the additivity axiom twice, we conclude that

$$P(X \in R) = \sum_{x \in R} P(X = x) = \sum_{x \in R} P(X_0 = x) = P(X_0 \in R) = P(\Omega) = 1.$$

Consequently, there is a countable set $R \in \mathcal{R}$ such that $P(X \in R) = 1$, meaning that X is a discrete random variable.

5.19 The sample space for this random experiment consists of all real numbers between 0 and 1, exclusive: $\Omega = (0, 1) = \{ x : 0 < x < 1 \}$. Because a number is being selected at random, a geometric probability model is appropriate. Thus, for each event E,

$$P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{1} = |E|, \tag{$*$}$$

where $|E|$ denotes the length (in the extended sense) of the set E.
a) Yes, X is a random variable because it is a real-valued function whose domain is Ω. In fact, X is defined on Ω by $X(x) = x$; that is, X is the identity function on Ω.
b) No, X is not a discrete random variable. Indeed, as we will presently show, $P(X = x) = 0$ for all $x \in \mathcal{R}$. This result, in turn, implies that $\sum_x P(X = x) = 0 \neq 1$. Therefore, from Exercise 5.17, X is not a discrete random variable.

To show that $P(X = x) = 0$ for all $x \in \mathcal{R}$, we can proceed as follows. Clearly, the range of X is the interval $(0, 1)$ and, consequently, we have $P(X = x) = 0$ if $x \notin (0, 1)$. So, from now on, we assume that $x \in (0, 1)$. From Equation $(*)$,

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

Alternatively, for $n \in \mathcal{N}$ sufficiently large, we have $[x, x + 1/n] \subset (0, 1) = \Omega$. Let $A_n = [x, x + 1/n]$. Then $\{X = x\} \subset \{X \in A_n\}$ for all $n \in \mathcal{N}$ and, hence, from Equation $(*)$,

$$P(X = x) \leq P(X \in A_n) = |X \in A_n| = |A_n| = \frac{1}{n},$$

for n sufficiently large. Thus, $P(X = x) = 0$.

5.20 A sample space for this random experiment is the unit disk: $\Omega = \{ (x, y) : x^2 + y^2 < 1 \}$. Because the dish is smeared with a uniform suspension of bacteria, use of a geometric probability model is reasonable. Specifically, we can think of the location of the center of the first spot (visible bacteria colony) as a point selected at random from the unit disk. Consequently, for each event E,

$$P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{\pi}, \tag{$**$}$$

where $|E|$ denotes the area (in the extended sense) of the set E.
a) Yes, Z is a random variable because it is a real-valued function whose domain is Ω. In fact, Z is defined on Ω by $Z(x, y) = \sqrt{x^2 + y^2}$.
b) No, Z is not a discrete random variable. Indeed, as we will presently show, $P(Z = z) = 0$ for all $z \in \mathcal{R}$. This result, in turn, implies that $\sum_z P(Z = z) = 0 \neq 1$. Therefore, from Exercise 5.17, Z is not a discrete random variable.

To show that $P(Z = z) = 0$ for all $z \in \mathcal{R}$, we can proceed as follows. Clearly, the range of Z is the interval $[0, 1)$ and, hence, $P(Z = z) = 0$ if $z \notin [0, 1)$. So, from now on, we assume that $z \in [0, 1)$. The event $\{Z = z\}$ is that the center of the first spot to appear is z units from the center of the petri dish or, equivalently, that the center of the first spot to appear lies on the circle $C = \{ (x, y) : x^2 + y^2 = z^2 \}$. Hence, from Equation (∗∗),

$$P(Z = z) = P(C) = \frac{|C|}{\pi} = 0,$$

where the last equality follows from the fact that the area of a circle (boundary of a disk) is 0. Alternatively, let

$$A_n = \left\{ (x, y) : z \le \sqrt{x^2 + y^2} \le z + 1/n \right\}, \qquad n \in \mathcal{N},$$

and note that $A_n \subset \Omega$ for n sufficiently large. As $\{Z = z\} \subset \{z \le Z \le z + 1/n\} = A_n$ for all $n \in \mathcal{N}$, we have, in view of Equation (∗∗), that

$$P(Z = z) \le P(A_n) = \frac{|A_n|}{\pi} = \frac{\pi(z + 1/n)^2 - \pi z^2}{\pi} = \frac{2z}{n} + \frac{1}{n^2},$$

for sufficiently large n. Hence, $P(Z = z) = 0$.

5.21 We can take the sample space for this random experiment to be all positive real numbers; that is, $\Omega = (0, \infty) = \{ x \in \mathcal{R} : x > 0 \}$. The outcome of the random experiment is x means that the duration of the total solar eclipse is x minutes.
a) Note that X is a random variable because it is a real-valued function whose domain is Ω. In fact, X is defined on Ω by $X(x) = x$; that is, X is the identity function on Ω. However, we claim that X is not a discrete random variable. Indeed, as we will presently show, $P(X = x) = 0$ for all $x \in \mathcal{R}$. This result, in turn, implies that $\sum_x P(X = x) = 0 \ne 1$. Hence, from Exercise 5.17, X is not a discrete random variable.

To show that $P(X = x) = 0$ for all $x \in \mathcal{R}$, we can proceed as follows. Clearly, the range of X is (a subset of) the interval $(0, \infty)$ and, thus, we have $P(X = x) = 0$ if $x \notin (0, \infty)$. So, from now on, we assume that $x \in (0, \infty)$. Intuitively speaking, it is impossible that we would ever observe any specific duration, x, more than once. Hence, in n independent repetitions of the random experiment (observing the duration of a solar eclipse), we would have $n(\{X = x\}) \le 1$. Consequently, from the frequentist interpretation of probability, for large n,

$$0 \le P(X = x) \approx \frac{n(\{X = x\})}{n} \le \frac{1}{n}.$$

Because $1/n \to 0$ as $n \to \infty$, we conclude that $P(X = x) = 0$.
b) Note that Y is a random variable because it is a real-valued function whose domain is Ω. In fact, Y is defined on Ω by $Y(x) = \lfloor x \rfloor$. Clearly, the range of Y is (a subset of) the nonnegative integers and, hence, is countable. Therefore, Y is a discrete random variable.

5.22 *Note:* Referring to the argument presented in Exercise 5.21(a), we see that, generally speaking, a random variable, X, denoting the measurement of a quantity such as time, weight, height, or temperature, has the property that $P(X = x) = 0$ for all $x \in \mathcal{R}$. Therefore, in view of Exercise 5.17, such a random variable is not discrete.
a) From the note, as T_6 denotes a (random) time, it is not a discrete random variable.
b) The number, X_3, of patients who arrive during the first 3 hours is a discrete random variable because its range is (a subset of) the nonnegative integers, which is countable.
c) We have $S = \lfloor T_6 \rfloor$. Hence, S is a discrete random variable because its range is (a subset of) the nonnegative integers, which is countable.
d) From the note, as the elapsed time between the arrivals of the sixth and eighth patients denotes a (random) time, it is not a discrete random variable.

e) From the note, as the elapsed time between the arrival of the last patient to arrive before 3:00 A.M. and the arrival of the first patient to arrive after 3:00 A.M. denotes a (random) time, it is not a discrete random variable.
f) The event $\{X_3 \geq 6\}$ occurs if and only if at least six patients arrive during the first 3 hours, which happens if and only if the sixth patient arrives within the first 3 hours, that is, if and only if the event $\{T_6 \leq 3\}$ occurs. Hence, $\{X_3 \geq 6\} = \{T_6 \leq 3\}$.

5.23
a) We note that a value of X is always a ratio of two integers. Hence, the range of X is countable, being a subset of the set of rational numbers (which is a countable set).
b) First we show that the range of X, which we denote R, consists precisely of the rational numbers in the interval $(0, 1]$, which we denote K. As a value of X is always a ratio of two positive integers of which the numerator does not exceed the denominator, we have $R \subset K$. Conversely, suppose that $x \in K$. Then we can write $x = m/n$ where m and n are positive integers with $m \leq n$. Now, from the domination principle and the general multiplication rule,

$$\begin{aligned} P(X = x) &= P(X = m/n) \geq P(N = n,\ M = m) = P(N = n)P(M = m \mid N = n) \\ &= p(1-p)^{n-1} \cdot \frac{1}{n} > 0. \end{aligned}$$

Thus, $x \in R$. We have now shown that $R = K$; that is, the range of X consists of all rational numbers in $(0, 1]$. The required result now follows from the fact that between any two rational numbers is another rational number.

5.2 Probability Mass Functions

Basic Exercises

5.24 Because the dice are balanced, each of the 36 possible outcomes, as shown in Fig. 2.1 on page 27, is equally likely. In other words, a classical probability model is appropriate here. So, for each event, E,

$$P(E) = \frac{N(E)}{36}.$$

a) To illustrate the determination of the PMF of X, let's obtain $p_X(5)$. We have

$$p_X(5) = P(X = 5) = P\big(\{\,(x, y) \in \Omega : x + y = 5\,\}\big) = P\big(\{(1, 4), (2, 3), (3, 2), (4, 1)\}\big) = \frac{4}{36}.$$

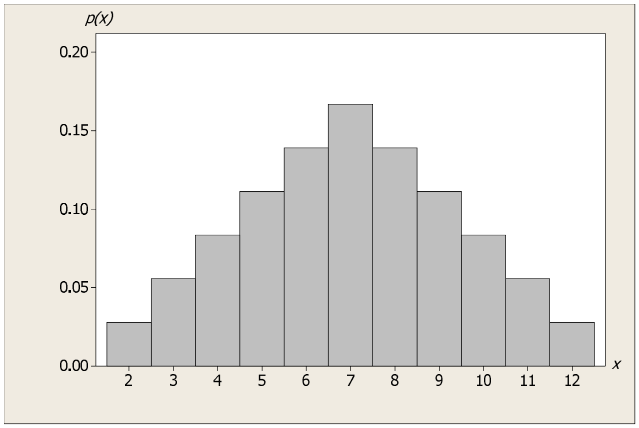
Proceeding similarly, we obtain the following table for the PMF of X:

x	2	3	4	5	6	7	8	9	10	11	12
$p_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Alternatively, we have

$$p_X(x) = \begin{cases} (x-1)/36, & \text{if } x = 2, 3, 4, 5, 6, 7; \\ (13-x)/36, & \text{if } x = 8, 9, 10, 11, 12; \\ 0, & \text{otherwise.} \end{cases}$$

b) Following is a probability histogram of X. In the graph, we use $p(x)$ instead of $p_X(x)$.



c) Referring to part (a) and applying the FPF, we get

$$P(\text{win on first roll}) = P(X = 7 \text{ or } 11) = p_X(7) + p_X(11) = \frac{6}{36} + \frac{2}{36} = \frac{2}{9}.$$

d) Referring to part (a) and applying the FPF, we get

$$P(\text{loss on first roll}) = P(X = 2, 3, \text{ or } 12) = p_X(2) + p_X(3) + p_X(12) = \frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{1}{9}.$$

e) In part (c), we have $A = \{7, 11\}$ and, in part (d), we have $A = \{2, 3, 12\}$.

5.25 Recall that a sample space for this random experiment is

$$\Omega = \left\{ (x_1, x_2, x_3, x_4, x_5) : x_j \in \{s, f\},\ 1 \le j \le 5 \right\},$$

where s and f denote a working and failed component, respectively.

a) As the components act independently and the probability of a working component is 0.8, we deduce that the probability that any individual outcome consists of exactly k working components is $(0.8)^k(0.2)^{5-k}$. Because there are $\binom{5}{k}$ outcomes in which exactly k of the components are working, we conclude that

$$p_X(x) = \begin{cases} \dbinom{5}{x}(0.8)^x(0.2)^{5-x}, & \text{if } x = 0, 1, \ldots, 5; \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively, by evaluating the preceding formula, we get the following table:

x	0	1	2	3	4	5
$p_X(x)$	0.00032	0.00640	0.05120	0.20480	0.40960	0.32768

b) Here $A = \{ x \in \mathcal{R} : 1 \le x \le 3 \}$. Applying the FPF and referring to part (a), we get

$$\begin{aligned} P(1 \le X \le 3) &= \sum_{1 \le x \le 3} p_X(x) = p_X(1) + p_X(2) + p_X(3) \\ &= 0.00640 + 0.05120 + 0.20480 = 0.2624. \end{aligned}$$

c) The required event is $\{X = 5\}$. From part (a), we have $P(X = 5) = p_X(5) = 0.32768$.

d) The required event is $\{X \geq 1\}$. From part (a) and the complementation rule,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - p_X(0) = 1 - 0.00032 = 0.99968.$$

e) Arguing as in part (a), we get that

$$p_X(x) = \begin{cases} \binom{5}{x} p^x (1-p)^{5-x}, & \text{if } x = 0, 1, \ldots, 5; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\begin{aligned} P(1 \leq X \leq 3) &= \sum_{1 \leq x \leq 3} p_X(x) = p_X(1) + p_X(2) + p_X(3) \\ &= \binom{5}{1} p^1 (1-p)^{5-1} + \binom{5}{2} p^2 (1-p)^{5-2} + \binom{5}{3} p^3 (1-p)^{5-3} \\ &= 5p(1-p)^4 + 10p^2(1-p)^3 + 10p^3(1-p)^2 \\ &= 5p(1-p)^2 \left((1-p)^2 + 2p(1-p) + 2p^2\right) \\ &= 5p(1-p)^2(1+p^2). \end{aligned}$$

Also,

$$P(X = 5) = p_X(5) = \binom{5}{5} p^5 (1-p)^{5-5} = 1 \cdot p^5 (1-p)^0 = p^5$$

and

$$P(X \geq 1) = 1 - P(X = 0) = 1 - p_X(0) = 1 - \binom{5}{0} p^0 (1-p)^{5-0} = 1 - (1-p)^5.$$

5.26 As genders in successive births are independent and the probability of a female child is p, the probability that any individual outcome consists of exactly k females is $p^k(1-p)^{3-k}$. Referring now to the table in the solution to Exercise 5.4(b), we obtain the following table:

ω	(m, m, m)	(m, m, f)	(m, f, m)	(m, f, f)	(f, m, m)	(f, m, f)	(f, f, m)	(f, f, f)
$P(\{\omega\})$	$(1-p)^3$	$p(1-p)^2$	$p(1-p)^2$	$p^2(1-p)$	$p(1-p)^2$	$p^2(1-p)$	$p^2(1-p)$	p^3
$Y(\omega)$	0	1	1	2	1	2	2	3

Hence, the PMF of Y is given by the following table:

y	0	1	2	3
$p_Y(y)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Alternatively,

$$p_Y(y) = \begin{cases} \binom{3}{y} p^y (1-p)^{3-y}, & \text{if } y = 0, 1, 2, 3; \\ 0, & \text{otherwise.} \end{cases}$$

b) From part (a), we have $P(Y = 1) = p_Y(1) = 3p(1-p)^2$.

c) From the FPF and part (a), we get

$$P(Y \le 1) = \sum_{y \le 1} p_Y(y) = p_Y(0) + p_Y(1) = (1-p)^3 + 3p(1-p)^2$$
$$= (1-p)^2\big((1-p) + 3p\big) = (1-p)^2(1+2p).$$

d) From the complementation rule and part (a),

$$P(Y \ge 1) = 1 - P(Y = 0) = 1 - p_Y(0) = 1 - (1-p)^3 = \big(3 - 3p + p^2\big)p.$$

5.27 As in Exercise 5.5, we take $\Omega = \{\,(x, y) : -3 \le x \le 3, -3 \le y \le 3\,\}$ and set $d(x, y) = \sqrt{x^2 + y^2}$. Because the archer will actually hit the target and is equally likely to hit any portion of the target, a geometric probability model is appropriate here. Hence, for each event E,

$$P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{36},$$

where $|E|$ denotes the area of the event E.
a) Referring to the solution of Exercise 5.5(a), we see that

$$\big|\{S = s\}\big| = \begin{cases} \big|\{\,(x, y) \in \Omega : d(x, y) > 2\}\big|, & \text{if } s = 0; \\ \big|\{\,(x, y) \in \Omega : 1 < d(x, y) \le 2\}\big|, & \text{if } s = 5; \\ \big|\{\,(x, y) \in \Omega : d(x, y) \le 1\}\big|, & \text{if } s = 10; \\ \big|\emptyset\big|, & \text{otherwise.} \end{cases} = \begin{cases} 36 - 4\pi, & \text{if } s = 0; \\ 3\pi, & \text{if } s = 5; \\ \pi, & \text{if } s = 10; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$p_S(s) = P(S = s) = \begin{cases} 1 - \pi/9, & \text{if } s = 0; \\ \pi/12, & \text{if } s = 5; \\ \pi/36, & \text{if } s = 10; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, to four decimal places, the PMF of S is given by the following table:

s	0	5	10
$p_S(s)$	0.6509	0.2618	0.0873

Note: In parts (b)–(g), we refer both to the table in part (a) and to the table presented in the solution to parts (c)–(h) of Exercise 5.5.
b) We have $P(S = 5) = p_S(5) = 0.2618$. The probability is 0.2618 that the archer scores exactly 5 points; that is, that she hits the ring with inner radius 1 foot and outer radius 2 feet centered at the origin.
c) From the FPF,

$$P(S > 0) = \sum_{s > 0} p_S(s) = p_S(5) + p_S(10) = 0.2618 + 0.0873 = 0.3491.$$

The probability is 0.3491 that the archer's score is positive; that is, that she hits the disk of radius 2 feet centered at the origin.
d) From the FPF,

$$P(S \le 7) = \sum_{s \le 7} p_S(s) = p_S(0) + p_S(5) = 0.6509 + 0.2618 = 0.9127.$$

The probability is 0.9127 that the archer's score is at most 7 points; that is, that she does not hit the bull's eye.

e) From the FPF,

$$P(5 < S \leq 15) = \sum_{5<s\leq 15} p_S(s) = p_S(10) = 0.0873.$$

The probability is 0.0873 that the archer's score is more than 5 points and at most 15 points; that is, that she hits the bull's eye.

f) From the FPF,

$$P(S < 15) = \sum_{s<15} p_S(s) = p_S(0) + p_S(5) + p_S(10) = 0.6509 + 0.2618 + 0.0873 = 1.$$

The probability is 1 that the archer's score is less than 15 points, which is of course the case because the most that the archer can score is 10 points.

g) We have $P(S < 0) = P(\emptyset) = 0$. The probability is 0 that the archer's score is negative, which is of course the case because the least the archer can score is 0 points.

5.28 We first determine the PMF of X for general p. There are several ways to accomplish this task. Here we refer to the table in the solution to Exercise 5.7(b) and compute the probability of each possible outcome, ω. To illustrate the computations, let us find $P\{(\text{I}, r, r)\}$. Set

$$\begin{aligned} A &= \text{event Urn I is chosen,} \\ B &= \text{event the first ball selected is red, and} \\ C &= \text{event the second ball selected is red.} \end{aligned}$$

Note that $\{(\text{I}, r, r)\} = A \cap B \cap C$. Applying the general multiplication rule for three events, we get

$$P\{(\text{I}, r, r)\} = P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = p \cdot \frac{4}{5} \cdot \frac{3}{4} = \frac{3}{5}p.$$

Proceeding similarly, we get the following table:

ω	(I, r, r)	(I, r, w)	(I, w, r)	(II, r, w)	(II, w, r)	(II, w, w)
$P(\{\omega\})$	$\frac{3}{5}p$	$\frac{1}{5}p$	$\frac{1}{5}p$	$\frac{1}{5}(1-p)$	$\frac{1}{5}(1-p)$	$\frac{3}{5}(1-p)$
$X(\omega)$	2	1	1	1	1	0

Noting that $\frac{1}{5}p + \frac{1}{5}p + \frac{1}{5}(1-p) + \frac{1}{5}(1-p) = \frac{2}{5}$, we conclude that

$$p_X(x) = \begin{cases} \frac{3}{5}(1-p), & \text{if } x = 0; \\ \frac{2}{5}, & \text{if } x = 1; \\ \frac{3}{5}p, & \text{if } x = 2; \\ 0, & \text{otherwise.} \end{cases}$$

a) When $p = 0.5$, we have

$$p_X(x) = \begin{cases} 0.3, & \text{if } x = 0, 2; \\ 0.4, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) When $p = 0.4$, we have

$$p_X(x) = \begin{cases} 0.36, & \text{if } x = 0; \\ 0.4, & \text{if } x = 1; \\ 0.24, & \text{if } x = 2; \\ 0, & \text{otherwise.} \end{cases}$$

5.29
a) The number of possible interview pools is the number of ways that three applicants can be chosen from 11 applicants, which is $\binom{11}{3} = 165$. Because the selection is done randomly, a classical probability model is appropriate here. Thus, for each event E,

$$P(E) = \frac{N(E)}{N(\Omega)} = \frac{N(E)}{165},$$

where $N(E)$ denotes the number of elements (outcomes) of E. In particular,

$$p_X(x) = P(X = x) = \frac{N(\{X = x\})}{165}$$

for each $x \in \mathcal{R}$. Referring now to the table in the solution to Exercise 5.8(b), we conclude that

$$p_X(x) = \begin{cases} \frac{20}{165}, & \text{if } x = 0; \\ \frac{75}{165}, & \text{if } x = 1; \\ \frac{60}{165}, & \text{if } x = 2; \\ \frac{10}{165}, & \text{if } x = 3; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} \frac{4}{33}, & \text{if } x = 0; \\ \frac{5}{11}, & \text{if } x = 1; \\ \frac{4}{11}, & \text{if } x = 2; \\ \frac{2}{33}, & \text{if } x = 3; \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively, we have the following table, where the probabilities are given to four decimal places:

x	0	1	2	3
$p_X(x)$	0.1212	0.4545	0.3636	0.0606

b) From the FEF and part (a),

$$P(X = 1 \text{ or } 2) = \sum_{x=1 \text{ or } 2} p_X(x) = p_X(1) + p_X(2) = 0.4545 + 0.3636 = 0.8181,$$

or, more precisely, $5/11 + 4/11 = 9/11$.

5.30 We assume, as is reasonable, that successive tosses of the coin are independent.
a) Referring to the table in the solution to Exercise 5.9(a), we obtain the following table:

ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$P(\{\omega\})$	$(0.5)^3$	$(0.5)^3$	$(0.5)^3$	$(0.5)^3$	$(0.5)^3$	$(0.5)^3$	$(0.5)^3$	$(0.5)^3$
$Y(\omega)$	3	1	1	-1	1	-1	-1	-3

Therefore,

$$p_Y(y) = \begin{cases} (0.5)^3, & \text{if } y = -3 \text{ or } 3; \\ 3 \cdot (0.5)^3, & \text{if } y = -1 \text{ or } 1; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 0.125, & \text{if } y = -3 \text{ or } 3; \\ 0.375, & \text{if } y = -1 \text{ or } 1; \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively, we have the following table:

y	-3	-1	1	3
$p_Y(y)$	0.125	0.375	0.375	0.125

b) Referring to the table in the solution to Exercise 5.9(a), we obtain the following table:

ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$P(\{\omega\})$	$(0.2)^3$	$(0.2)^2(0.8)$	$(0.2)^2(0.8)$	$(0.2)(0.8)^2$	$(0.2)^2(0.8)$	$(0.2)(0.8)^2$	$(0.2)(0.8)^2$	$(0.8)^3$
$Y(\omega)$	3	1	1	-1	1	-1	-1	-3

Therefore,

$$p_Y(y) = \begin{cases} (0.8)^3, & \text{if } y = -3; \\ 3 \cdot (0.2)(0.8)^2, & \text{if } y = -1; \\ 3 \cdot (0.2)^2(0.8), & \text{if } y = 1; \\ (0.2)^3, & \text{if } y = 3; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 0.512, & \text{if } y = -3; \\ 0.384, & \text{if } y = -1; \\ 0.096, & \text{if } y = 1; \\ 0.008, & \text{if } y = 3; \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively, we have the following table:

y	-3	-1	1	3
$p_Y(y)$	0.512	0.384	0.096	0.008

c) Referring to the table in the solution to Exercise 5.9(a), we obtain the following table:

ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$P(\{\omega\})$	p^3	$p^2(1-p)$	$p^2(1-p)$	$p(1-p)^2$	$p^2(1-p)$	$p(1-p)^2$	$p(1-p)^2$	$(1-p)^3$
$Y(\omega)$	3	1	1	-1	1	-1	-1	-3

Therefore,

$$p_Y(y) = \begin{cases} (1-p)^3, & \text{if } y = -3; \\ 3p(1-p)^2, & \text{if } y = -1; \\ 3p^2(1-p), & \text{if } y = 1; \\ p^3, & \text{if } y = 3; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} (1-p)^3, & \text{if } y = -3; \\ 3p(1-p)^2, & \text{if } y = -1; \\ 3p^2(1-p), & \text{if } y = 1; \\ p^3, & \text{if } y = 3; \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively, we have the following table:

y	-3	-1	1	3
$p_Y(y)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

d) The result of part (a) follows immediately by substituting $p = 0.5$ in the result of part (c), and the result of part (b) follows immediately by substituting $p = 0.2$ in the result of part (c).

5.31 We assume, as is reasonable, that the results from one week to the next are independent.
a) Referring to the solution to Exercise 5.11(a), we get, for $n \in \mathcal{N}$,

$$P(W = n) = P(\{\underbrace{\ell \ldots \ell}_{n-1 \text{ times}} w\}) = \underbrace{(1-p) \cdots (1-p)}_{n-1 \text{ times}}\, p = p(1-p)^{n-1}.$$

Therefore,

$$p_W(w) = \begin{cases} p(1-p)^{w-1}, & \text{if } w \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases}$$

b) From the FPF and part (a),

$$P(W > n) = \sum_{w>n} p_W(w) = \sum_{k=n+1}^{\infty} p_W(k) = \sum_{k=n+1}^{\infty} p(1-p)^{k-1}$$

$$= p(1-p)^n \sum_{j=0}^{\infty}(1-p)^j = p(1-p)^n \frac{1}{1-(1-p)} = (1-p)^n.$$

c) For the event $\{W > n\}$ to occur, you must not win a prize during any of the first n weeks.
d) The probability that you don't win a prize during any of the first n weeks is

$$\underbrace{(1-p)\cdots(1-p)}_{n \text{ times}} = (1-p)^n.$$

The result found here, although the same as that in part (b), is much simpler to obtain.
e) We have $P(W > 1) = (1-p)^1 = 1-p$.
f) From the complementation rule,

$$P(W \le 10) = 1 - P(W > 10) = 1 - (1-p)^{10}.$$

g) We note that $\{W > 15\} = \{15 < W \le 20\} \cup \{W > 20\}$ and that the events in the union are mutually exclusive. Hence,

$$P(W > 15) = P(15 < W \le 20) + P(W > 20)$$

and, consequently,

$$P(15 < W \le 20) = P(W > 15) - P(W > 20) = (1-p)^{15} - (1-p)^{20}.$$

h) We note that $\{W > 14\} = \{14 < W \le 19\} \cup \{W > 19\}$ and that the events in the union are mutually exclusive. Hence,

$$P(W > 14) = P(14 < W \le 19) + P(W > 19)$$

and, consequently, because W is integer valued,

$$P(15 \le W < 20) = P(14 < W \le 19) = P(W > 14) - P(W > 19) = (1-p)^{14} - (1-p)^{19}.$$

5.32
a) Referring to Exercise 5.12, we see that the range of Y is $R = \{m, m+1, \dots, n\}$. If $k \in R$, then

$$P(Y = k) = P\big(\lfloor m + (n-m+1)X \rfloor = k\big) = P\big(k \le m + (n-m+1)X < k+1\big)$$

$$= P\left(\frac{k-m}{n-m+1} \le X < \frac{k+1-m}{n-m+1}\right) = P\left(X \in \left[\frac{k-m}{n-m+1}, \frac{k+1-m}{n-m+1}\right)\right)$$

$$= \left|\left[\frac{k-m}{n-m+1}, \frac{k+1-m}{n-m+1}\right)\right| = \frac{k+1-m}{n-m+1} - \frac{k-m}{n-m+1} = \frac{1}{n-m+1}.$$

Hence,

$$p_Y(y) = \begin{cases} 1/(n-m+1), & \text{if } y \in \{m, m+1, \dots, n\}; \\ 0, & \text{otherwise.} \end{cases}$$

b) The result of part (a) shows that we can use a basic random number generator to obtain a random integer within any specified range. Specifically, to obtain a random integer between m and n, inclusive, we first use a basic random number generator to get a real number, x, between 0 and 1, and then we compute $\lfloor m + (n-m+1)x \rfloor$.

5.33

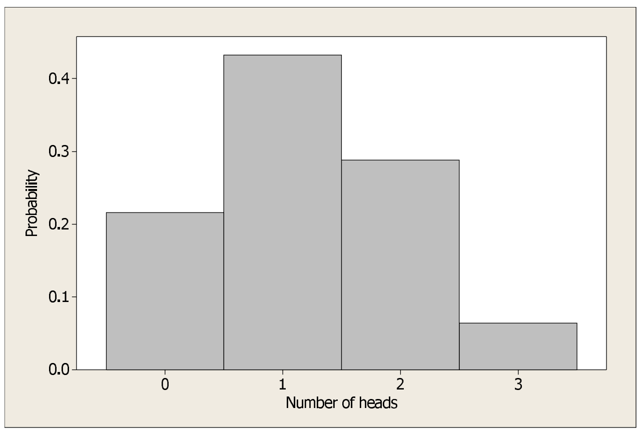
a) Answers will vary for the simulation, but the probability distribution is as in the first and fourth columns of Table 5.10 on page 192, and the probability histogram is as in Fig. 5.2(b) on page 193.

b) Answers will vary for the simulation, but the probability distribution is as in the first and fourth columns of Table 5.10, and the probability histogram is as in Fig. 5.2(b).

c) Answers will vary for the simulation. Referring to Example 5.5, specifically, to Table 5.9 on page 189, we find that the probability distribution is as given in the following table:

Heads	0	1	2	3
Probability	0.216	0.432	0.288	0.064

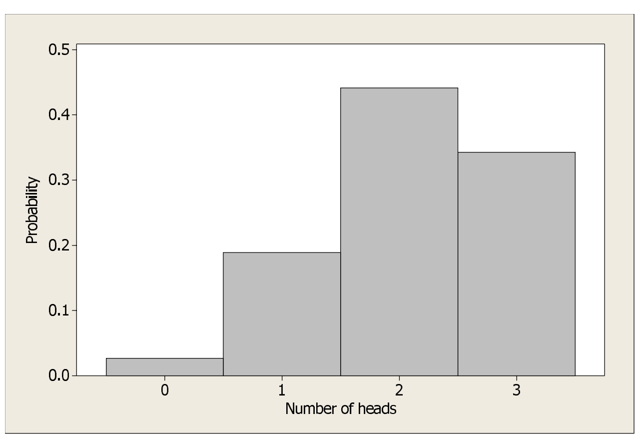
A probability histogram is as follows:



d) Answers will vary for the simulation. Referring to Example 5.5, specifically, to Table 5.9 on page 189, we find that the probability distribution is as given in the following table:

Heads	0	1	2	3
Probability	0.027	0.189	0.441	0.343

A probability histogram is as follows:



5.34 We assume, as is reasonable, that the tosses of the coins are independent. To simplify the discussion, we first determine a formula for the PMF of the total number of heads, W, obtained in n tosses of a balanced coin. On each toss, there are two possible outcomes, a head or a tail; hence, in n tosses of the coin, there are 2^n possible outcomes. And, because the coin is balanced, these outcomes are equally likely. Hence, a classical probability model is appropriate here. We note that, for $k = 0, 1, \ldots, n$, the event $\{W = k\}$ can occur in $\binom{n}{k}$ ways, namely, the number of ways that we can select k positions for heads from the n positions. Thus,

$$P(W = k) = \frac{N(\{W = k\})}{N(\Omega)} = \frac{\binom{n}{k}}{2^n} = \frac{1}{2^n}\binom{n}{k}.$$

Consequently,

$$p_W(w) = \frac{1}{2^n}\binom{n}{w}, \qquad w = 0, 1, \ldots, n,$$

and $p_W(w) = 0$ otherwise.

a) We have $n = 2$, so that

$$p_X(x) = \frac{1}{2^2}\binom{2}{x} = \frac{1}{4}\binom{2}{x}, \qquad x = 0, 1, 2.$$

b) We have $n = 2$, so that

$$p_Y(y) = \frac{1}{2^2}\binom{2}{y} = \frac{1}{4}\binom{2}{y}, \qquad y = 0, 1, 2.$$

c) We have $n = 2$, so that

$$p_X(y) = \frac{1}{2^2}\binom{2}{y} = \frac{1}{4}\binom{2}{y}, \qquad y = 0, 1, 2.$$

d) The possible values of $X + Y$ are 0, 1, 2, 3, and 4. We note that $X + Y$ gives the total number of heads in four tosses of a balanced coin. Therefore, we have $n = 4$, so that

$$p_{X+Y}(z) = \frac{1}{2^4}\binom{4}{z} = \frac{1}{16}\binom{4}{z}, \qquad z = 0, 1, 2, 3, 4.$$

e) From parts (a) and (b),

$$p_X(z) + p_Y(z) = \frac{1}{4}\binom{2}{z} + \frac{1}{4}\binom{2}{z} = \frac{1}{2}\binom{2}{z}, \qquad z = 0, 1, 2.$$

f) The possible values of XY are 0, 1, 2, and 4. Using the independence of X and Y and parts (a) and (b), we get

$$P(XY = 1) = P(X = 1,\ Y = 1) = P(X = 1)P(Y = 1) = \left(\frac{1}{4}\cdot 2\right)\left(\frac{1}{4}\cdot 2\right) = \frac{1}{4},$$

$$\begin{aligned} P(XY = 2) &= P(X = 1,\ Y = 2) + P(X = 2,\ Y = 1) = 2P(X = 1)P(Y = 2) \\ &= 2\left(\frac{1}{4}\cdot 2\right)\left(\frac{1}{4}\right) = \frac{1}{4}, \end{aligned}$$

$$P(XY = 4) = P(X = 2,\ Y = 2) = P(X = 2)P(Y = 2) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16},$$

and, from the complementation rule,

$$P(XY = 0) = 1 - P(XY = 1) - P(XY = 2) - P(XY = 4) = 1 - \frac{1}{4} - \frac{1}{4} - \frac{1}{16} = \frac{7}{16}.$$

Therefore,

$$p_{XY}(z) = \begin{cases} 7/16, & \text{if } z = 0; \\ 1/4, & \text{if } z = 1, 2; \\ 1/16, & \text{if } z = 4. \end{cases}$$

g) The possible values of X^2 are 0, 1, and 4. From part (a),

$$p_{X^2}(z) = P(X^2 = z) = P\left(X = \sqrt{z}\right) = p_X\left(\sqrt{z}\right) = \frac{1}{4}\binom{2}{\sqrt{z}}, \qquad z = 0, 1, 4.$$

5.35 A sample space for the random experiment is

$$\Omega = \left\{ (x, y, z) : x, y, z \in \{1, 2, 3, 4, 5, 6\} \right\},$$

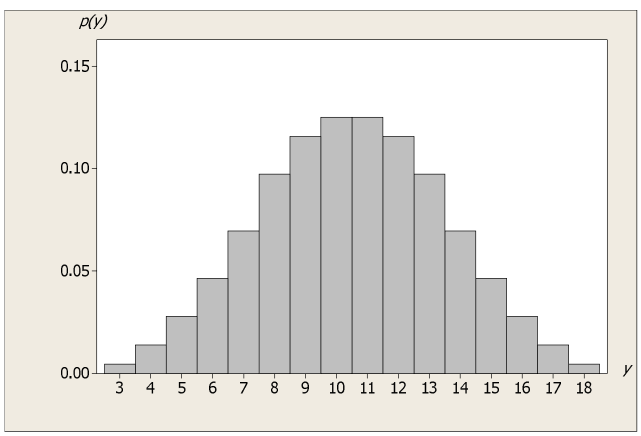
where x, y, and z denote the faces showing on the first, second, and third dice, respectively. As there are six possibilities for each die, the total number of possible outcomes is $6^3 = 216$. And, because the dice are balanced, each possible outcome is equally likely. In other words, a classical probability model is appropriate here. So, for each event E,

$$P(E) = \frac{N(E)}{N(\Omega)} = \frac{N(E)}{216}.$$

a) The possible sums of the three faces showing are between 3 and 18, inclusive, and we find that the number of ways these sums can occur are 1, 3, 6, 10, 15, 21, 25, 27, 27, 25, 21, 15, 10, 6, 3, 1, respectively. Hence, the PMF of Y is given by the following table:

y	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$p_Y(y)$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

b) Following is a probability histogram of Y. In the graph, we use $p(y)$ instead of $p_Y(y)$.



c) Both probability histograms are symmetric. The general shapes of the two histograms are similar, although the one in Exercise 5.24 (sum of two dice) has a more triangular shape, whereas the one here (sum of three dice) is closer to being bell shaped.

Theory Exercises

5.36 Let $\Omega = \{ x \in \mathcal{R} : p(x) \neq 0 \}$. By assumption, Ω is a countable set. From properties (a) and (c) of Proposition 5.1, we see that the sequence $\{ p(\omega) : \omega \in \Omega \}$ satisfies the conditions of Proposition 2.3.

Hence, there is a unique probability measure P on the events of Ω such that $P(\{\omega\}) = p(\omega)$ for all $\omega \in \Omega$. Define $X\colon \Omega \to \mathcal{R}$ by $X(\omega) = \omega$. Because Ω is countable, so is the range of X. Consequently, X is a discrete random variable. Now let $x \in \mathcal{R}$. If $x \notin \Omega$, then

$$p_X(x) = P(X = x) = P(\emptyset) = 0 = p(x),$$

whereas, if $x \in \Omega$, then

$$p_X(x) = P(X = x) = P(\{x\}) = p(x).$$

Thus, for all $x \in \mathcal{R}$, we have $p_X(x) = p(x)$; that is, p is the PMF of the random variable X.

Advanced Exercises

5.37 We can take the sample space to be $\Omega = \{s, fs, ffs, \ldots\}$, where

$$\underbrace{f \ldots f}_{n-1 \text{ times}} s$$

represents the outcome that the goal is realized on the nth attempt. Note that

$$X(\underbrace{f \ldots f}_{n-1 \text{ times}} s) = n, \qquad n \in \mathcal{N}.$$

a) Using the independence of the results of the attempts, we have, for $n \in \mathcal{N}$,

$$P(X = n) = P(\{\underbrace{f \ldots f}_{n-1 \text{ times}} s\}) = \underbrace{(1-p) \cdots (1-p)}_{n-1 \text{ times}}\, p = p(1-p)^{n-1}.$$

Therefore,

$$p_X(x) = \begin{cases} p(1-p)^{x-1}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases}$$

b) Referring to the result of part (a), we get the following results.

$p = 0.25$:

$$p_X(x) = \begin{cases} 0.25 \cdot (1-0.25)^{x-1}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 0.25 \cdot (0.75)^{x-1}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases}$$

$p = 0.5$:

$$p_X(x) = \begin{cases} 0.5 \cdot (1-0.5)^{x-1}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} (0.5)^{x}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases}$$

$p = 1$:

$$p_X(x) = \begin{cases} 1 \cdot (1-1)^{x-1}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 1, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases}$$

In this last case ($p = 1$), the goal will be realized with certainty on the first attempt and, hence, $p_X(1) = 1$.

c) As $0 < p \le 1$, we have, from the FPF, part (a), and the formula for a geometric series, that

$$P(X < \infty) = P(X \in \mathcal{R}) = \sum_{x \in \mathcal{R}} p_X(x) = \sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{k=0}^{\infty} (1-p)^k = p\,\frac{1}{1-(1-p)} = 1.$$

Thus, the goal will eventually be realized (with probability 1).

d) From the FPF and part (a),

$$\begin{aligned}
P(X \text{ is even}) &= \sum_{x \text{ is even}} p_X(x) = \sum_{n=1}^{\infty} p_X(2n) = \sum_{n=1}^{\infty} p(1-p)^{2n-1} = \sum_{k=0}^{\infty} p(1-p)^{2k+1} \\
&= p(1-p)\sum_{k=0}^{\infty}(1-p)^{2k} = p(1-p)\sum_{k=0}^{\infty}\big((1-p)^2\big)^k = p(1-p)\frac{1}{1-(1-p)^2} \\
&= \frac{1-p}{2-p}.
\end{aligned}$$

e) Let $E = \{X \text{ is even}\}$ and let A denote the event that the goal is realized on the first attempt. We note that $P(E \mid A) = 0$ and that, because the results of the attempts are independent, $P(E \mid A^c) = P(E^c)$. Hence, by the law of total probability and the complementation rule,

$$\begin{aligned}
P(E) &= P(A)P(E \mid A) + P(A^c)P(E \mid A^c) = p \cdot 0 + (1-p)P(E^c) \\
&= (1-p)\big(1 - P(E)\big) = 1 - p - (1-p)P(E).
\end{aligned}$$

Solving for $P(E)$, we find that $P(E) = (1-p)/(2-p)$.

f) When $p = 0$, we have $P(X = \infty) = 1$; that is, X always equals ∞. In particular, then, X is not a real-valued function and, hence, is not a random variable.

5.38 We first recall from calculus that

$$\ln(1-t) = -\sum_{k=1}^{\infty}\frac{t^k}{k}, \qquad -1 \le t < 1.$$

Now, from the solution to Exercise 5.23(b), we know that the range of X consists of all rational numbers in the interval (0, 1], which we denote K. Let $x \in K$. Then x can be expressed uniquely in the form m/n, where m and n are relatively prime positive integers with $m \le n$. Noting that

$$\{X = x\} = \{M/N = m/n\} = \bigcup_{k=1}^{\infty}\{N = kn,\ M = km\}$$

and that the events in the union are mutually exclusive, we get

$$\begin{aligned}
p_X(x) = P(X = x) &= \sum_{k=1}^{\infty} P(N = kn,\ M = km) \\
&= \sum_{k=1}^{\infty} P(N = kn)P(M = km \mid N = kn) = \sum_{k=1}^{\infty} p(1-p)^{kn-1} \cdot \frac{1}{kn} \\
&= \frac{p}{n(1-p)}\sum_{k=1}^{\infty}\frac{(1-p)^{kn}}{k} = \frac{p}{n(1-p)}\sum_{k=1}^{\infty}\frac{\big((1-p)^n\big)^k}{k} \\
&= -\frac{p}{n(1-p)}\ln\big(1-(1-p)^n\big) = -\frac{p\ln\big(1-(1-p)^n\big)}{n(1-p)}.
\end{aligned}$$

Hence,

$$p_X(x) = -\frac{p\ln\big(1-(1-p)^n\big)}{n(1-p)}$$

if $x \in K$ and its denominator in lowest terms is n, and $p_X(x) = 0$ otherwise.

5.3 Binomial Random Variables

Basic Exercises

5.39

a) Here the random experiment consists of administering pyrantel pamoate to a child with pinworm infestation; each administration (to a different child) constitutes one trial. The specified event is that a child is cured; thus, a success is a cure and a failure is a noncure. The success probability, p, is the probability that, for any particular treated child, the child is cured; the failure probability, $q = 1 - p$, is the probability that, for any particular treated child, the child is not cured. By assumption, $p = 0.9$, so $q = 0.1$. Finally, it can be reasonably assumed that the results of administering pyrantel pamoate are independent from one child to another.

b) With independence, an outcome consisting of exactly k cures has probability $(0.9)^k(0.1)^{3-k}$. Thus, we have the following table:

Outcome	Probability	Outcome	Probability
(s, s, s)	$(0.9)^3(0.1)^0 = 0.729$	(f, s, s)	$(0.9)^2(0.1)^1 = 0.081$
(s, s, f)	$(0.9)^2(0.1)^1 = 0.081$	(f, s, f)	$(0.9)^1(0.1)^2 = 0.009$
(s, f, s)	$(0.9)^2(0.1)^1 = 0.081$	(f, f, s)	$(0.9)^1(0.1)^2 = 0.009$
(s, f, f)	$(0.9)^1(0.1)^2 = 0.009$	(f, f, f)	$(0.9)^0(0.1)^3 = 0.001$

c) From the table in part (b), we see that the outcomes in which exactly two of the three children are cured are (s, s, f), (s, f, s), and (f, s, s).

d) From the table in part (b), we see that each of the three outcomes in part (c) has probability 0.081. The three probabilities are the same because each probability is obtained by multiplying two success probabilities of 0.9 and one failure probability of 0.1.

e) From parts (c) and (d), we see that the probability that exactly two of the three children will be cured is equal to $3 \cdot 0.081 = 0.243$.

f) Referring to the table in part (b), we get

$$\begin{aligned}
P(X = 0) &= P\big(\{(f, f, f)\}\big) = 0.001, \\
P(X = 1) &= P\big(\{(s, f, f), (f, s, f), (f, f, s)\}\big) = P\big(\{(s, f, f)\}\big) + P\big(\{(f, s, f)\}\big) + P\big(\{(f, f, s)\}\big) \\
&= 0.009 + 0.009 + 0.009 = 0.027, \\
P(X = 2) &= P\big(\{(s, s, f), (s, f, s), (f, s, s)\}\big) = P\big(\{(s, s, f)\}\big) + P\big(\{(s, f, s)\}\big) + P\big(\{(f, s, s)\}\big) \\
&= 0.081 + 0.081 + 0.081 = 0.243, \\
P(X = 3) &= P\big(\{(s, s, s)\}\big) = 0.729.
\end{aligned}$$

Thus, we have the following table for the PMF of X:

x	0	1	2	3
$p_X(x)$	0.001	0.027	0.243	0.729

g) Applying Procedure 5.1 with $n = 3$ and $p = 0.9$, we get

$$p_X(x) = \binom{3}{x}(0.9)^x(0.1)^{3-x}, \qquad x = 0, 1, 2, 3,$$

and $p_X(x) = 0$ otherwise. Substituting successively $x = 0$, 1, 2, and 3 into the preceding display and doing the necessary algebra gives the same values as shown in the table in part (f).

5.40
a) In this case, there are N^n possible outcomes of which MN^{n-1} have the property that the kth member sampled has the specified attribute. Consequently, the probability is $MN^{n-1}/N^n = M/N$ that the kth member sampled has the specified attribute.
b) In this case, there are $(N)_n$ possible outcomes of which $M(N-1)_{n-1}$ have the property that the kth member sampled has the specified attribute. Thus, the probability is $M(N-1)_{n-1}/(N)_n = M/N$ that the kth member sampled has the specified attribute.

5.41 For random sampling with replacement, a member selected is returned to the population for possible reselection; hence, the result of one trial has no effect on the results of other trials (i.e., the trials are independent). However, for random sampling without replacement, a member selected isn't returned to the population for possible reselection; hence, the result of one trial affects the results of the other trials (i.e., the trials aren't independent).

5.42 Let X denote the number of times the favorite finishes in the money in the next five races. We note that $X \sim \mathcal{B}(5, 0.67)$. Therefore,

$$p_X(x) = \binom{5}{x}(0.67)^x(0.33)^{5-x}, \qquad x = 0, 1, 2, 3, 4, 5,$$

and $p_X(x) = 0$ otherwise.
a) We have

$$P(X = 2) = \binom{5}{2}(0.67)^2(0.33)^{5-2} = 10 \cdot (0.67)^2(0.33)^3 = 0.161.$$

b) We have

$$P(X = 4) = \binom{5}{4}(0.67)^4(0.33)^{5-4} = 5 \cdot (0.67)^4(0.33)^1 = 0.332.$$

c) From the FPF,

$$\begin{aligned} P(X \geq 4) &= \sum_{x \geq 4} p_X(x) = p_X(4) + p_X(5) = \binom{5}{4}(0.67)^4(0.33)^{5-4} + \binom{5}{5}(0.67)^5(0.33)^{5-5} \\ &= 5 \cdot (0.67)^4(0.33) + 1 \cdot (0.67)^5 = 0.468. \end{aligned}$$

d) From the FPF,

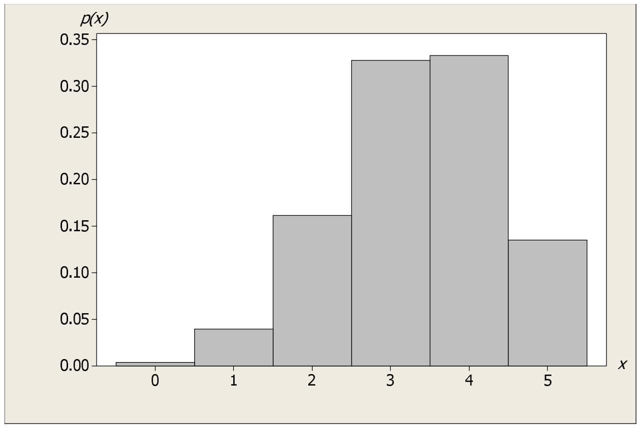
$$\begin{aligned} P(2 \leq X \leq 4) &= \sum_{2 \leq x \leq 4} p_X(x) = p_X(2) + p_X(3) + p_X(4) \\ &= \binom{5}{2}(0.67)^2(0.33)^{5-2} + \binom{5}{3}(0.67)^3(0.33)^{5-3} + \binom{5}{4}(0.67)^4(0.33)^{5-4} \\ &= 10 \cdot (0.67)^2(0.33)^3 + 10 \cdot (0.67)^3(0.33)^2 + 5 \cdot (0.67)^4(0.33) = 0.821. \end{aligned}$$

e) Proceeding as above, we obtain the following table for the PMF of the random variable X:

x	0	1	2	3	4	5
$p_X(x)$	0.00391	0.03973	0.16132	0.32753	0.33249	0.13501

f) The probability distribution of X is left skewed because $p = 0.67 > 0.5$.

g) A probability histogram for X is as follows. Note that, in the graph, we use $p(x)$ in place of $p_X(x)$.



5.43
a) Let

$$\begin{aligned} E &= \text{event that the child has sickle cell anemia,} \\ A &= \text{event that the father passes hemoglobin S to the child, and} \\ B &= \text{event that the mother passes hemoglobin S to the child.} \end{aligned}$$

Then $E = A \cap B$ and, by independence,

$$P(E) = P(A \cap B) = P(A)P(B) = 0.5 \cdot 0.5 = 0.25.$$

b) Let Y denote the number of children of the five who have sickle cell anemia. Assuming independence from one birth to the next, we know, in view of part (a), that $Y \sim \mathcal{B}(5, 0.25)$. Hence, from the complementation rule,

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) = 1 - p_Y(0) \\ &= 1 - \binom{5}{0}(0.25)^0(0.75)^5 = 1 - (0.75)^5 \\ &= 0.763. \end{aligned}$$

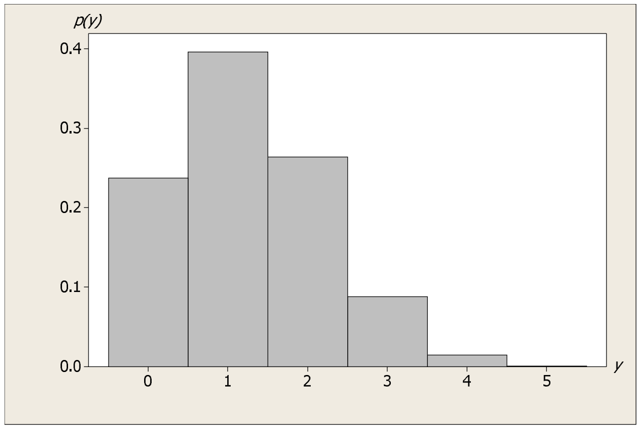
c) As we noted in part (b), $Y \sim \mathcal{B}(5, 0.25)$. Hence, the PMF of Y is given by

$$p_Y(y) = \binom{5}{y}(0.25)^y(0.75)^{5-y}, \qquad y = 0, 1, 2, 3, 4, 5,$$

and $p_Y(y) = 0$ otherwise. Applying the preceding formula, we obtain the following table:

y	0	1	2	3	4	5
$p_Y(y)$	0.237305	0.395508	0.263672	0.087891	0.014648	0.000977

d) A probability histogram for Y is as follows. Note that, in the graph, we use $p(y)$ in place of $p_Y(y)$.



5.44 Let X denote the number of girls in a family with five children. Assuming independence from one birth to the next, then, because all sex distributions are equally likely, we have $X \sim \mathcal{B}(5, 0.5)$. Thus,

$$P(X = 3) = p_X(3) = \binom{5}{3}(0.5)^3(1 - 0.5)^{5-3} = 10 \cdot (0.5)^5 = 0.3125.$$

Hence, 31.25% of families with five children have three girls and two boys.

5.45
a) The assumption of Bernoulli trials is appropriate if, for this baseball player, successive trials at bat are independent, and the probability of a hit on any given at bat coincides with the player's batting average.
b) Let X denote the number of hits by this player in his next four times at bat. Under the assumptions of part (a), we have $X \sim \mathcal{B}(4, 0.26)$. Hence, by the FPF,

$$\begin{aligned} P(X \geq 2) &= \sum_{x \geq 2} p_X(x) = p_X(2) + p_X(3) + p_X(4) \\ &= \binom{4}{2}(0.26)^2(0.74)^{4-2} + \binom{4}{3}(0.26)^3(0.74)^{4-3} + \binom{4}{4}(0.26)^4(0.74)^{4-4} \\ &= 6 \cdot (0.26)^2(0.74)^2 + 4 \cdot (0.26)^3(0.74) + 1 \cdot (0.26)^4 = 0.279. \end{aligned}$$

5.46 Let X denote the number of voters in the sample who intend to vote "yes." Then $X \sim \mathcal{B}(20, 0.6)$. More voters in the sample intend to vote "yes" than don't intend to vote "yes" if and only if $X \geq 11$. Hence, by the FPF, the required probability is

$$P(X \geq 11) = \sum_{x \geq 11} p_X(x) = \sum_{k=11}^{20} \binom{20}{k}(0.6)^k(0.4)^{20-k} = 0.755.$$

5.47 Let X denote the number of successes in 15 Bernoulli trials.
a) Here $X \sim \mathcal{B}(15, 0.5)$. Therefore, from the FPF,

$$P(6 \leq X \leq 9) = \sum_{6 \leq x \leq 9} p_X(x) = \sum_{k=6}^{9} \binom{15}{k}(0.5)^k(1 - 0.5)^{15-k} = (0.5)^{15} \sum_{k=6}^{9} \binom{15}{k} = 0.698.$$

We have $(n+1)p = 16 \cdot 0.5 = 8$, which is an integer. Hence, from Proposition 5.4(b) on page 204, the most likely number of successes are 7 and 8.
b) Here $X \sim \mathcal{B}(15, 0.4)$. Therefore, from the FPF,

$$P(6 \le X \le 9) = \sum_{6 \le x \le 9} p_X(x) = \sum_{k=6}^{9} \binom{15}{k} (0.4)^k (0.6)^{15-k} = 0.563.$$

We have $(n+1)p = 16 \cdot 0.4 = 6.4$, which is not an integer. Hence, from Proposition 5.4(a), the most likely number of successes is $\lfloor 6.4 \rfloor = 6$.

5.48 Let X denote the number of 1s in 100 throws of a balanced die. Then $X \sim \mathcal{B}(100, 1/6)$.
a) From the FPF,

$$P(X \le 4) = \sum_{x \le 4} p_X(x) = \sum_{k=0}^{4} \binom{100}{k} (1/6)^k (5/6)^{100-k} = 0.0000940.$$

We would be skeptical that the die is really balanced because the occurrence of at most four 1s in 100 throws of a balanced die is an extremely unlikely event.
b) From the FPF,

$$P(X > 32) = \sum_{x > 32} p_X(x) = \sum_{k=33}^{100} \binom{100}{k} (1/6)^k (5/6)^{100-k} = 0.0000496.$$

As in part (a), we would be skeptical that the die is really balanced because the chance of getting a 1 more than 32 times in 100 throws of a balanced die is very small.

5.49 The event of an "odd man" means that either $n-1$ people get heads and one person gets a tail or $n-1$ people get tails and one person gets a head. Letting X denote the number of heads in n tosses of a coin with probability p of heads, we note that $X \sim \mathcal{B}(n, p)$. Hence, the probability of the required event is

$$\begin{aligned} P(X = 1 \text{ or } n-1) &= p_X(1) + p_X(n-1) = \binom{n}{1} p^1 (1-p)^{n-1} + \binom{n}{n-1} p^{n-1} (1-p)^1 \\ &= np(1-p)^{n-1} + np^{n-1}(1-p) = npq\big(q^{n-2} + p^{n-2}\big), \end{aligned}$$

where $q = 1 - p$.

5.50 The probability that a decimal digit is either 6 or 7 equals $2/10 = 0.2$. In a sequence of n decimal digits, let X_n denote the number of digits that are either 6 or 7. Then $X_n \sim \mathcal{B}(n, 0.2)$. We want to determine the condition on n so that $P(X_n \ge 1) > 0.95$ or, equivalently, $P(X_n = 0) < 0.05$. However,

$$P(X_n = 0) = p_{X_n}(0) = \binom{n}{0} (0.2)^0 (0.8)^n = (0.8)^n.$$

Hence, we need $(0.8)^n < 0.05$. Taking logarithms, this inequality becomes

$$n \ln(0.8) < \ln(0.05) \qquad \text{or} \qquad n > \frac{\ln(0.05)}{\ln(0.8)} \qquad \text{or} \qquad n > 13.4.$$

Therefore, a sequence of decimal digits has probability exceeding 0.95 of containing at least one 6 or 7 if and only if it is of length 14 or more.

5.51
a) Consider the experiment of randomly selecting a number from the interval (0, 1), and let the specified event be that the number obtained exceeds 0.7. In three independent repetitions of this random experiment, the median of the three numbers obtained exceeds 0.7 if and only if two or more of them exceed 0.7—that is, if and only if two or more of the three trials result in success.
b) In part (a), the specified event is that a number selected at random from the interval (0, 1) exceeds 0.7, which has probability $1 - 0.7 = 0.3$.

5.52 Let X denote the total number of red marbles selected. Also, let

$$\begin{aligned} I &= \text{event Urn I is chosen, and} \\ II &= \text{event Urn II is chosen.} \end{aligned}$$

The probabilities that a red marble is selected from Urns I and II are 0.5 and 0.55, respectively. Because the marbles are selected with replacement, we see that, if Urn I is chosen, then $X \sim \mathcal{B}(6, 0.5)$, whereas, if Urn II is chosen, then $X \sim \mathcal{B}(6, 0.55)$. Hence, for $x = 0, 1, \dots, 6$,

$$P(X = x \mid I) = \binom{6}{x}(0.5)^x(0.5)^{6-x} \qquad \text{and} \qquad P(X = x \mid II) = \binom{6}{x}(0.55)^x(0.45)^{6-x}.$$

As each urn is equally likely to be the one chosen, we have $P(I) = P(II) = 0.5$. Applying Bayes's rule, we now conclude that, for $x = 0, 1, \dots, 6$,

$$\begin{aligned} P(II \mid X = x) &= \frac{P(II)P(X = x \mid II)}{P(I)P(X = x \mid I) + P(II)P(X = x \mid II)} \\ &= \frac{0.5 \cdot \binom{6}{x}(0.55)^x(0.45)^{6-x}}{0.5 \cdot \binom{6}{x}(0.5)^x(0.5)^{6-x} + 0.5 \cdot \binom{6}{x}(0.55)^x(0.45)^{6-x}} \\ &= \frac{(0.55)^x(0.45)^{6-x}}{(0.5)^x(0.5)^{6-x} + (0.55)^x(0.45)^{6-x}} = \left(1 + \frac{1}{(1.1)^x(0.9)^{6-x}}\right)^{-1}. \end{aligned}$$

Evaluating this last expression, we get the following table, where the probabilities are displayed to four decimal places:

x	0	1	2	3	4	5	6
$P(II \mid X = x)$	0.3470	0.3938	0.4425	0.4925	0.5425	0.5917	0.6392

Theory Exercises

5.53
a) Applying the binomial theorem, we get

$$1 = 1^n = \big(p + (1 - p)\big)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k}.$$

b) Let $X \sim \mathcal{B}(n, p)$. Then, as a PMF must sum to 1, we get

$$1 = \sum_x p_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k}.$$

5.54 Suppose that X is symmetric. Then, in particular, we have

$$p^n = \binom{n}{n} p^n (1-p)^{n-n} = p_X(n) = p_X(n-n) = p_X(0) = \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n.$$

Therefore, $p = 1 - p$, which implies that $p = 1/2$.

Conversely, suppose that $p = 1/2$. Let $R = \{0, 1, \dots, n\}$. We note that $n - x \in R$ if and only if $x \in R$. Hence, for $x \notin R$, we have $p_X(x) = 0 = p_X(n-x)$. Moreover, for $x \in R$,

$$p_X(x) = \binom{n}{x} 2^{-n} = \binom{n}{n-x} 2^{-n} = p_X(n-x).$$

Thus, we have shown that $p_X(x) = p_X(n-x)$ for all x; that is, X is symmetric.

5.55

a) We have

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

and

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

b) Using integration by parts with $u = (1-t)^k$ and $dv = t^{n-k-1}\,dt$, we get

$$\begin{aligned}\int_0^q t^{n-k-1}(1-t)^k\,dt &= \left[(1-t)^k \cdot \frac{t^{n-k}}{n-k}\right]_0^q - \int_0^q \frac{t^{n-k}}{n-k} \cdot k(1-t)^{k-1}(-1)\,dt \\ &= \frac{1}{n-k}\,p^k q^{n-k} + \frac{k}{n-k}\int_0^q t^{n-k}(1-t)^{k-1}\,dt.\end{aligned}$$

c) Multiplying both sides of the equation in part (b) by $(n-k)\binom{n}{k}$, we get

$$(n-k)\binom{n}{k}\int_0^q t^{n-k-1}(1-t)^k\,dt = p_X(k) + k\binom{n}{k}\int_0^q t^{n-k}(1-t)^{k-1}\,dt.$$

Applying now the identities in part (a) gives

$$n\binom{n-1}{k}\int_0^q t^{n-k-1}(1-t)^k\,dt = p_X(k) + n\binom{n-1}{k-1}\int_0^q t^{n-k}(1-t)^{k-1}\,dt.$$

Now let

$$A_k = n\binom{n-1}{k}\int_0^q t^{n-k-1}(1-t)^k\,dt.$$

Then we have $A_k = p_X(k) + A_{k-1}$, or $p_X(k) = A_k - A_{k-1}$. Therefore,

$$P(X \le k) = p_X(0) + \sum_{j=1}^{k} p_X(j) = p_X(0) + \sum_{j=1}^{k}(A_j - A_{j-1}) = p_X(0) + A_k - A_0.$$

However,

$$A_0 = n\int_0^q t^{n-1}\,dt = q^n = p_X(0).$$

Consequently,

$$P(X \le k) = A_k = n\binom{n-1}{k}\int_0^q t^{n-k-1}(1-t)^k\,dt = n\binom{n-1}{k}B_q(n-k, k+1).$$

d) Let $Y = n - X$. Then Y is the number of failures in n Bernoulli trials with success probability p. Interchanging the roles of success and failure, we see that $Y \sim \mathcal{B}(n, q)$. Applying part (c) to Y with k replaced by $n-k-1$, we get

$$\begin{aligned} P(X > k) &= P(n - Y > k) = P(Y < n-k) = P(Y \le n-k-1) \\ &= n\binom{n-1}{n-k-1}\int_0^p t^{n-(n-k-1)-1}(1-t)^{n-k-1}\,dt \\ &= n\binom{n-1}{k}\int_0^p t^k(1-t)^{n-k-1}\,dt = n\binom{n-1}{k}B_p(k+1, n-k). \end{aligned}$$

Advanced Exercises

5.56

a) Let X denote the total number of red marbles obtained in the five draws. We note that, because the sampling is with replacement, $X \sim \mathcal{B}(5, k/12)$. Therefore,

$$P(X = 2) = \binom{5}{2}\left(\frac{k}{12}\right)^2\left(1 - \frac{k}{12}\right)^3 = \frac{10}{(12)^5}\,k^2(12-k)^3.$$

Consider the function $g(x) = x^2(12-x)^3$ on the interval $[0, 12]$. Now,

$$g'(x) = 2x(12-x)^3 - 3x^2(12-x)^2 = x(12-x)^2\,(24 - 2x - 3x) = x(12-x)^2(24-5x).$$

Thus, $g'(x) > 0$ for $x \in (0, 4.8)$ and $g'(x) < 0$ for $x \in (4.8, 12)$. Consequently, g is strictly increasing between 0 and 4.8 and strictly decreasing between 4.8 and 12. As $g(4) = 8192$ and $g(5) = 8575$, it follows that $P(X = 2)$ is maximized when $k = 5$.

b) Referring to part (a), we see that the maximum probability is

$$\frac{10}{(12)^5}g(5) = \frac{10}{(12)^5}\cdot 8575 = 0.345.$$

5.57 Let p denote the probability that any particular engine works. Also, let X and Y denote the number of working engines for the first and second rockets, respectively. We know that $X \sim \mathcal{B}(2, p)$ and $Y \sim \mathcal{B}(4, p)$. Rocket 1 achieves its mission if and only if $X \ge 1$, and rocket 2 achieves its mission if and only if $Y \ge 2$. By assumption, then, $P(X \ge 1) = P(Y \ge 2)$. Now,

$$P(X \ge 1) = 1 - P(X = 0) = 1 - \binom{2}{0}p^0(1-p)^2 = 1 - (1-p)^2 = 2p - p^2$$

and

$$\begin{aligned} P(Y \ge 2) &= 1 - P(Y = 0) - P(Y = 1) = 1 - \binom{4}{0}p^0(1-p)^4 - \binom{4}{1}p^1(1-p)^3 \\ &= 1 - (1-p)^4 - 4p(1-p)^3 = 6p^2 - 8p^3 + 3p^4. \end{aligned}$$

Therefore, we want to find p so that $2p - p^2 = 6p^2 - 8p^3 + 3p^4$, or $3p^4 - 8p^3 + 7p^2 - 2p = 0$. However,

$$3p^4 - 8p^3 + 7p^2 - 2p = p(p-1)^2(3p-2).$$

As the common probability of achieving the mission is nonzero, we know that $p \neq 0$. Consequently, either $p = 2/3$ or $p = 1$.

5.4 Hypergeometric Random Variables

Basic Exercises

5.58 When we use the word "definition" in this exercise, we are referring to Definition 5.5. If $r = 0$, then, by definition, $\binom{m}{r} = \binom{m}{0} = 1$; similarly, if $r < 0$, then, by definition, $\binom{m}{r} = 0$. If $1 \le r \le m$, then, by definition,

$$\binom{m}{r} = \frac{m(m-1)\cdots(m-r+1)}{r!} = \frac{m(m-1)\cdots(m-r+1)}{r!} \cdot \frac{(m-r)!}{(m-r)!}$$
$$= \frac{m(m-1)\cdots(m-r+1)(m-r)!}{r!\,(m-r)!} = \frac{m!}{r!\,(m-r)!}.$$

If $r > m$, then, by definition,

$$\binom{m}{r} = \frac{m(m-1)\cdots(m-r+1)}{r!} = \frac{m(m-1)\cdots 1 \cdot 0 \cdot (-1)\cdots(m-r+1)}{r!} = \frac{0}{r!} = 0.$$

Hence,

$$\binom{m}{r} = \begin{cases} \dfrac{m!}{r!\,(m-r)!}, & \text{if } 1 \le r \le m; \\ 1, & \text{if } r = 0; \\ 0, & \text{if } r > m \quad \text{or} \quad r < 0. \end{cases}$$

5.59 Referring to Definition 5.5 on page 211 yields the following results.

a) Because $3 > 0$,

$$\binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{3!} = \frac{7 \cdot 6 \cdot 5}{6} = 35.$$

b) Because $-3 < 0$,

$$\binom{7}{-3} = 0.$$

c) Because $3 > 0$, we have

$$\binom{-7}{3} = \frac{(-7)\cdot(-8)\cdot(-9)}{3!} = -\frac{7 \cdot 8 \cdot 9}{6} = -84.$$

d) Because $-3 < 0$,

$$\binom{-7}{-3} = 0.$$

e) Because $7 > 0$,

$$\binom{3}{7} = \frac{3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1) \cdot (-2) \cdot (-3)}{7!} = \frac{0}{7!} = 0.$$

f) Because $-7 < 0$,

$$\binom{3}{-7} = 0.$$

g) Because $7 > 0$,

$$\binom{-3}{7} = \frac{(-3)\cdot(-4)\cdot(-5)\cdot(-6)\cdot(-7)\cdot(-8)\cdot(-9)}{7!} = -\frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 8\cdot 9}{7!} = -\frac{8\cdot 9}{2\cdot 1} = -36.$$

h) Because $-7 < 0$,

$$\binom{-3}{-7} = 0.$$

i) Because $3 > 0$, we have

$$\binom{1/4}{3} = \frac{(1/4)\cdot(1/4-1)\cdot(1/4-2)}{3!} = \frac{(1/4)\cdot(-3/4)\cdot(-7/4)}{3!} = \frac{1\cdot 3\cdot 7}{4^3\cdot 6} = \frac{7}{128}.$$

j) Because $-3 < 0$,

$$\binom{1/4}{-3} = 0.$$

k) Because $3 > 0$, we have

$$\binom{-1/4}{3} = \frac{(-1/4)\cdot(-1/4-1)\cdot(-1/4-2)}{3!} = \frac{(-1/4)\cdot(-5/4)\cdot(-9/4)}{3!} = -\frac{1\cdot 5\cdot 9}{4^3\cdot 6} = -\frac{15}{128}.$$

l) Because $-3 < 0$,

$$\binom{-1/4}{-3} = 0.$$

m) No, because 1/4 is not an integer.

5.60 We refer to Defintion 5.5 on page 211.
a) If $r = 0$, then

$$\binom{-1}{r} = \binom{-1}{0} = 1 = (-1)^0 = (-1)^r.$$

If $r > 0$, then

$$\binom{-1}{r} = \frac{(-1)\cdot(-2)\cdots(-1-r+1)}{r!} = \frac{(-1)^r 1\cdot 2\cdots r}{r!} = (-1)^r\frac{r!}{r!} = (-1)^r.$$

b) If $r = 0$, then

$$\binom{-m}{r} = \binom{-m}{0} = 1 = \binom{m-1}{0} = (-1)^0\binom{m+0-1}{0} = (-1)^r\binom{m+r-1}{0}.$$

If $r > 0$, then

$$\binom{-m}{r} = \frac{(-m)\cdot(-m-1)\cdots(-m-r+1)}{r!} = \frac{(-1)^r m\cdot(m+1)\cdots(m+r-1)}{r!}$$
$$= (-1)^r\frac{(m+r-1)\cdot(m+r-2)\cdots m}{r!} = (-1)^r\binom{m+r-1}{r}.$$

5.61

a) We first note that there are $\binom{52}{5}$ possible equally likely outcomes of the random experiment. Hence, for each event E,

$$P(E) = \frac{N(E)}{\binom{52}{5}}.$$

Let A be the event that exactly three face cards are obtained. Three face cards can be selected from the 12 face cards in $\binom{12}{3}$ ways, and two non-face cards can be selected from the 40 non-face cards in $\binom{40}{2}$ ways. Hence, by the BCR,

$$P(A) = \frac{N(A)}{\binom{52}{5}} = \frac{\binom{12}{3}\binom{40}{2}}{\binom{52}{5}} = 0.0660.$$

b) Let the specified attribute be "face card." Here $N = 52$, $M = 12$, $N - M = 40$, and $n = 5$. Noting that $p = 12/52 = 3/13$, we see that the number, X, of face cards obtained has the $\mathcal{H}(52, 5, 3/13)$ distribution. Consequently, in view of Equation (5.13) on page 210,

$$p_X(x) = \frac{\binom{12}{x}\binom{40}{5-x}}{\binom{52}{5}}, \qquad x = 0, 1, 2, 3, 4, 5,$$

and $p_X(x) = 0$ otherwise.

5.62 We first note that $X \sim \mathcal{H}(N, n, p)$. It follows that $\hat{p}$ is a discrete random variable whose possible values are (contained among) $0,\ 1/n,\ 2/n,\ \ldots,\ (n-1)/n,\ 1$. Therefore, from Equation (5.12) on page 210, the PMF of $\hat{p}$ is

$$p_{\hat{p}}(x) = P(\hat{p} = x) = P(X/n = x) = P(X = nx) = p_X(nx) = \frac{\binom{Np}{nx}\binom{N(1-p)}{n(1-x)}}{\binom{N}{n}},$$

if $x = 0, 1/n, 2/n, \ldots, (n-1)/n, 1$; and $p_{\hat{p}}(x) = 0$ otherwise.

5.63

a) Of the five pills chosen, let X denote the number of placebos selected. Here $N = 10$, $M = 5$, and $N - M = 5$. Noting that $p = 5/10 = 0.5$, we see that $X \sim \mathcal{H}(10, 5, 0.5)$. Hence, from the complementation rule and Equation (5.13) on page 210,

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) = 1 - p_X(0) - p_X(1) \\ &= 1 - \frac{\binom{5}{0}\binom{5}{5}}{\binom{10}{5}} - \frac{\binom{5}{1}\binom{5}{4}}{\binom{10}{5}} = 1 - \frac{6}{252} = 0.897. \end{aligned}$$

b) We can solve this problem in several ways. One way is to use the general multiplication rule. Let E be the event that the first three pills you select are placebos and, for $1 \leq j \leq 3$, let A_j be the event that the jth pill you select is a placebo. Then $E = A_1 \cap A_2 \cap A_3$ and, by the general multiplication rule for three events,

$$P(E) = P(A_1 \cap A_1 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) = \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8} = \frac{1}{12} = 0.0833.$$

Another method is to use the BCR as follows:

$$P(E) = \frac{5 \cdot 4 \cdot 3}{10 \cdot 9 \cdot 8} = \frac{60}{720} = \frac{1}{12} = 0.0833.$$

Yet another method is to use the $\mathcal{H}(10, 3, 0.5)$ distribution:

$$P(E) = \frac{\binom{5}{3}\binom{5}{0}}{\binom{10}{3}} = \frac{10 \cdot 1}{120} = \frac{1}{12} = 0.0833.$$

5.64 Let Y denote the number of households of the six sampled that have a VCR. Because the sampling is without replacement, $Y \sim \mathcal{H}(N, 6, 0.842)$, where N is the number of U.S. households. Because the sample size is small relative to the population size, we can use the binomial distribution with parameters $n = 6$ and $p = 0.842$ to approximate probabilities for Y.

a) We have

$$P(Y = 4) \approx \binom{6}{4}(0.842)^4(0.158)^2 = 0.188.$$

b) We have

$$\begin{aligned} P(Y \geq 4) &= P(Y = 4) + P(Y = 5) + P(Y = 6) \\ &\approx \binom{6}{4}(0.842)^4(0.158)^2 + \binom{6}{5}(0.842)^5(0.158)^1 + \binom{6}{6}(0.842)^6(0.158)^0 \\ &= 0.946. \end{aligned}$$

c) From the complementation rule,

$$\begin{aligned} P(Y \leq 4) = 1 - P(Y > 4) &= 1 - P(Y = 5) - P(Y = 6) \\ &\approx 1 - \binom{6}{5}(0.842)^5(0.158)^1 - \binom{6}{6}(0.842)^6(0.158)^0 \\ &= 0.242. \end{aligned}$$

d) We have

$$\begin{aligned} P(2 \leq Y \leq 5) &= P(Y = 2) + P(Y = 3) + P(Y = 4) + P(Y = 5) \\ &\approx \binom{6}{2}(0.842)^2(0.158)^4 + \binom{6}{3}(0.842)^3(0.158)^3 + \binom{6}{4}(0.842)^4(0.158)^2 \\ &\quad + \binom{6}{5}(0.842)^5(0.158)^1 \\ &= 0.643. \end{aligned}$$

e) As we noted, because the sample size is small relative to the population size, Y has approximately the binomial distribution with parameters $n = 6$ and $p = 0.842$. Hence,

$$p_Y(y) \approx \binom{6}{y}(0.842)^y(0.158)^{6-y}, \qquad y = 0, 1, \ldots, 6,$$

and $p_Y(y) = 0$ otherwise.
f) The PMF obtained in part (e) is only approximately correct because the sampling is without replacement, implying that, strictly speaking, the trials are not independent.
g) As we noted, $Y \sim \mathcal{H}(N, 6, 0.842)$. Hence, from Equation (5.12) on page 210,

$$p_Y(y) = \frac{\binom{0.842N}{y}\binom{0.158N}{6-y}}{\binom{N}{6}}, \qquad y = 0, 1, \ldots, 6,$$

and $p_Y(y) = 0$ otherwise.

5.65 Let X denote the number of defective parts of the three selected. Also, let

$$\begin{aligned} I &= \text{event that Bin I is chosen, and} \\ II &= \text{event that Bin II is chosen.} \end{aligned}$$

Given that Bin I is chosen, $X \sim \mathcal{H}(20, 3, 1/4)$, whereas, given that Bin II is chosen, $X \sim \mathcal{H}(15, 3, 4/15)$. Hence,

$$P(X = 2 \mid I) = \frac{\binom{5}{2}\binom{15}{1}}{\binom{20}{3}} = \frac{5}{38} \qquad \text{and} \qquad P(X = 2 \mid II) = \frac{\binom{4}{2}\binom{11}{1}}{\binom{15}{3}} = \frac{66}{455}.$$

Moreover, because the bin chosen is selected randomly, we have $P(I) = P(II) = 1/2$. Hence, by Bayes's rule,

$$\begin{aligned} P(I \mid X = 2) &= \frac{P(I)P(X = 2 \mid I)}{P(I)P(X = 2 \mid I) + P(II)P(X = 2 \mid II)} = \frac{\frac{1}{2}\cdot\frac{5}{38}}{\frac{1}{2}\cdot\frac{5}{38} + \frac{1}{2}\cdot\frac{66}{455}} \\ &= \frac{\frac{5}{38}}{\frac{5}{38} + \frac{66}{455}} = 0.476. \end{aligned}$$

5.66 We have $N = 10$, $n = 3$, and $p = 0.6$.
a) If the sampling is without replacement, then $X \sim \mathcal{H}(10, 3, 0.6)$. Thus,

$$p_X(x) = \frac{\binom{6}{x}\binom{4}{3-x}}{\binom{10}{3}}, \qquad x = 0, 1, 2, 3,$$

and $p_X(x) = 0$ otherwise. Evaluating the nonzero probabilities, we get the following table:

x	0	1	2	3
$p_X(x)$	0.0333	0.3000	0.5000	0.1667

b) If the sampling is with replacement, then $X \sim \mathcal{B}(3, 0.6)$. Thus,

$$p_X(x) = \binom{3}{x}(0.6)^x(0.4)^{3-x}, \qquad x = 0, 1, 2, 3,$$

and $p_X(x) = 0$ otherwise. Evaluating the nonzero probabilities, we get the following table:

x	0	1	2	3
$p_X(x)$	0.064	0.288	0.432	0.216

c) The results of parts (a) and (b) differ substantially, which is not surprising because the sample size is not small relative to the population size.

5.67 We have $N = 100$, $n = 5$, and $p = 0.06$.
a) We know that $X \sim \mathcal{H}(100, 5, 0.06)$. Thus,

$$p_X(x) = \frac{\binom{6}{x}\binom{94}{5-x}}{\binom{100}{5}}, \qquad x = 0, 1, \ldots, 5,$$

and $p_X(x) = 0$ otherwise. Evaluating the nonzero probabilities, we get the following table:

x	0	1	2	3	4	5
$p_X(x)$	0.72908522	0.24302841	0.02670642	0.00116115	0.00001873	0.00000008

b) The appropriate approximating binomial distribution is the one with parameters $n = 5$ and $p = 0.06$. Thus,

$$p_X(x) = \binom{5}{x}(0.06)^x(0.94)^{5-x}, \qquad x = 0, 1, \ldots, 5,$$

and $p_X(x) = 0$ otherwise. Evaluating the nonzero probabilities and juxtaposing the results with those of part (a), we get the following table:

Defectives x	Hypergeometric probability	Binomial probability
0	0.72908522	0.73390402
1	0.24302841	0.23422469
2	0.02670642	0.02990102
3	0.00116115	0.00190858
4	0.00001873	0.00006091
5	0.00000008	0.00000078

c) Referring to part (b), we see that, here, the binomial approximation to the hypergeometric distribution is quite good for the two smallest values of x, moderately good for the next value of x, and then degrades rapidly as x increases. This fact is not surprising because the rule of thumb for the approximation is barely met—the sample size is exactly 5% of the population size.

5.68 In this case, we have $N = 80{,}000$, $M = 30{,}000$, $N - M = 50{,}000$, and $n = 100$. Noting that $p = 30{,}000/80{,}000 = 0.375$, we conclude that the number, X, of words that the student knows has the $\mathcal{H}(80{,}000, 100, 0.375)$ distribution. Hence,

$$p_X(x) = \frac{\binom{30{,}000}{x}\binom{50{,}000}{100-x}}{\binom{80{,}000}{100}}, \qquad x = 0, 1, \ldots, 100,$$

and $p_X(x) = 0$ otherwise.

a) We have

$$P(X = 40) = p_X(40) = \frac{\binom{30{,}000}{40}\binom{50{,}000}{60}}{\binom{80{,}000}{100}} = 0.0712.$$

b) From the FPF,

$$P(35 \le X \le 40) = \sum_{35 \le x \le 40} p_X(x) = \sum_{k=35}^{40} \frac{\binom{30{,}000}{k}\binom{50{,}000}{100-k}}{\binom{80{,}000}{100}} = 0.465.$$

c) Yes, a binomial approximation to the hypergeometric distribution would be justified in this case because the sample size is very small relative to the population size; in fact, the sample size is only 0.125% of the population size.

d) As this other student knows 38 of the 100 words sampled, her sample proportion of words known in this dictionary is 0.38. So, we would estimate that she knows about $0.38 \cdot 80{,}000 = 30{,}400$ words in this dictionary. Assuming that this student knows very few words that aren't in this dictionary, we would estimate the size of her vocabulary at 30,400 words.

Theory Exercises

5.69 The number of possible samples of size n from the population of size N is $\binom{N}{n}$. As the selection is done randomly, a classical probability model is appropriate. Hence, for each event E,

$$P(E) = \frac{N(E)}{N(\Omega)} = \frac{N(E)}{\binom{N}{n}}.$$

Now, let $k \in \{0, 1, \ldots, n\}$. We want to determine the number of ways that the event $\{X = k\}$ can occur. There are Np members of the population with the specified attribute, of which k are to be selected; that can be done in $\binom{Np}{k}$ ways. And there are $N(1-p)$ members of the population without the specified attribute, of which $n - k$ are to be selected; that can be done in $\binom{N(1-p)}{n-k}$ ways. Hence, by the BCR,

$$P(\{X = k\}) = \frac{N(\{X = k\})}{\binom{N}{n}} = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}.$$

Consequently,

$$p_X(x) = \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, \ldots, n,$$

and $p_X(x) = 0$ otherwise.

5.70 We know that X has the hypergeometric distribution with parameters N, n, and p. Recalling that $q = 1 - p$, we have, for $x = 0, 1, \ldots, n$,

$$p_X(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} = \frac{\dfrac{(Np)_x}{x!} \cdot \dfrac{(Nq)_{n-x}}{(n-x)!}}{\dfrac{(N)_n}{n!}} = \binom{n}{x}\frac{(Np)_x \cdot (Nq)_{n-x}}{(N)_n}$$

$$= \binom{n}{x}\frac{(Np)(Np-1)\cdots(Np-x+1)\cdot(Nq)(Nq-1)\cdots\big(Nq-(n-x)+1\big)}{N(N-1)\cdots(N-n+1)}$$

$$= \binom{n}{x}\frac{p\left(p-\frac{1}{N}\right)\cdots\left(p-\frac{x-1}{N}\right)\cdot q\left(q-\frac{1}{N}\right)\cdots\left(q-\frac{n-x-1}{N}\right)}{1\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}.$$

Hence,

$$\lim_{N\to\infty} p_X(x) = \binom{n}{x}\frac{p^x \cdot q^{n-x}}{1} = \binom{n}{x}p^x(1-p)^{n-x}.$$

5.71

a) Here we consider ordered samples with replacement. From Example 3.4(c) on page 90, there are N^n such samples. The problem is to find the probability that exactly k of the n members sampled have the specified attribute. Because the sampling is random, each possible sample is equally likely to be the one selected. We note that there are $\binom{n}{k}$ ways to choose the k positions in the ordered sample for the members that have the specified attribute. Once that option has been determined, there are $(Np)^k$ choices for which members with the specified attribute occupy the k positions and $(Nq)^{n-k}$ choices for which members without the specified attribute occupy the remaining $n-k$ positions. Hence, by the BCR,

$$P(\{X = k\}) = \frac{N(\{X = k\})}{N(\Omega)} = \frac{\binom{n}{k} \cdot (Np)^k \cdot (Nq)^{n-k}}{N^n} = \binom{n}{k}\frac{(Np)^k(Nq)^{n-k}}{N^n}.$$

Consequently,

$$p_X(x) = \binom{n}{x}\frac{(Np)^x(Nq)^{n-x}}{(N)^n}, \qquad x = 0, 1, \ldots, n,$$

and $p_X(x) = 0$ otherwise.

b) For $x = 0, 1, \ldots, n$, we have

$$\binom{n}{x}\frac{(Np)^x(Nq)^{n-x}}{(N)^n} = \binom{n}{x}\frac{(Np)^x(Nq)^{n-x}}{(N)^x(N)^{n-x}} = \binom{n}{x}p^x q^{n-x},$$

which is the PMF of a binomial distribution with parameters n and p.

c) Here we consider ordered samples without replacement. From Example 3.8(a) on page 96, there are $(N)_n$ such samples. The problem is to find the probability that exactly k of the n members sampled have the specified attribute. Because the sampling is random, each possible sample is equally likely to be the one selected. We note that there are $\binom{n}{k}$ ways to choose the k positions in the ordered sample for the members that have the specified attribute. Once that option has been determined, there are $(Np)_k$ choices for which members with the specified attribute occupy the k positions and $(Nq)_{n-k}$ choices for which members without the specified attribute occupy the remaining $n-k$ positions. Hence, by the BCR,

$$P(\{X=k\}) = \frac{N(\{X=k\})}{N(\Omega)} = \frac{\binom{n}{k}\cdot(Np)_k\cdot(Nq)_{n-k}}{(N)_n} = \binom{n}{k}\frac{(Np)_k(Nq)_{n-k}}{(N)_n}.$$

Consequently,

$$p_X(x) = \binom{n}{x}\frac{(Np)_x(Nq)_{n-x}}{(N)_n}, \qquad x = 0, 1, \ldots, n,$$

and $p_X(x) = 0$ otherwise.

d) For $x = 0, 1, \ldots, n$, we have, in view of the permutations rule (Proposition 3.2 on page 95) and the combinations rule (Proposition 3.4 on page 99),

$$\binom{n}{x}\frac{(Np)_x(Nq)_{n-x}}{(N)_n} = \frac{n!}{x!\,(n-x)!}\,\frac{\dfrac{(Np)!}{(Np-x)!}\,\dfrac{(Nq)!}{(Nq-(n-x))!}}{\dfrac{N!}{(N-n)!}}$$

$$= \frac{\dfrac{(Np)!}{x!\,(Np-x)!}\,\dfrac{(Nq)!}{(n-x)!\,(Nq-(n-x))!}}{\dfrac{N!}{n!\,(N-n)!}} = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}},$$

which is the PMF of a hypergeometric distribution with parameters N, n, and p.

5.72

a) Applying Vandermonde's identity with $k=n$, $n=M$, and $m=N-M$ yields

$$\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k} = \binom{M+(N-M)}{n} = \binom{N}{n}.$$

The required result now follows upon dividing both sides of the preceding display by $\binom{N}{n}$.

b) Let $X \sim \mathcal{H}(N, n, M/N)$. Then, as a PMF must sum to 1, we get

$$1 = \sum_x p_X(x) = \sum_{k=0}^{n}\frac{\binom{N\cdot M/N}{k}\binom{N\cdot(1-M/N)}{n-k}}{\binom{N}{n}} = \sum_{k=0}^{n}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

c) Applying part (b) with $n=k$, $M=n$, and $N=n+m$ yields

$$1 = \sum_{j=0}^{k}\frac{\binom{n}{j}\binom{(n+m)-n}{k-j}}{\binom{n+m}{k}} = \sum_{j=0}^{k}\frac{\binom{n}{j}\binom{m}{k-j}}{\binom{n+m}{k}} = \frac{1}{\binom{n+m}{k}}\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j}.$$

The required result now follows upon multiplying both sides of the preceding display by $\binom{n+m}{k}$.

Advanced Exercises

5.73 Suppose that the committee were randomly selected. Let X denote the number of men chosen. Noting that $12/20 = 0.6$, we have that $X \sim \mathcal{H}(20, 5, 0.6)$. The probability that one or fewer men would be chosen is

$$P(X \le 1) = \sum_{x \le 1} p_X(x) = p_X(0) + p_X(1) = \frac{\binom{12}{0}\binom{8}{5}}{\binom{20}{5}} + \frac{\binom{12}{1}\binom{8}{4}}{\binom{20}{5}} = 0.0578.$$

In other words, if the committee were randomly selected, there is less than a 6% chance that at most one man would be chosen. Consequently, the men do have reason to complain that the committee was not randomly selected.

5.74

a) Let X denote the number of red balls of the six sampled. Then $X \sim \mathcal{H}(12, 6, 2/3)$. Noting that at least two of the balls sampled must be red, we have

$$P(X = x) = \frac{\binom{8}{x}\binom{4}{6-x}}{\binom{12}{6}}, \qquad x = 2, 3, \dots, 6,$$

and $P(X = x) = 0$ otherwise.

b) Let Y denote the number of red balls of the eight sampled. Then $Y \sim \mathcal{H}(12, 8, 1/2)$. Noting that at least two of the balls sampled must be red and that at most six of the balls sampled can be red, we have

$$P(Y = x) = \frac{\binom{6}{x}\binom{6}{8-x}}{\binom{12}{8}}, \qquad x = 2, 3, \dots, 6,$$

and $P(Y = x) = 0$ otherwise.

c) We claim that the quantities on the right of the two preceding displays are equal for all $x = 2, 3, \dots, 6$. But that result is just the identity presented in Exercise 3.106(c) with $N = 12$, $n = 8$, and $m = 6$. Hence, there are no values of x for which the answer to part (a) is greater than the answer to part (b), and there are no values of x for which the answer to part (a) is smaller than the answer to part (b).

5.75

a) As the first class contains only four upper-division students, at least two of the six students sampled must be lower-division students. This type of restriction does not arise in the second class because there are 40 upper-division students. Consequently, as both classes contain at least six lower-division students, the set of possible values of the number of lower-division students sampled is $\{2, 3, \dots, 6\}$ for the first class and $\{0, 1, \dots, 6\}$ for the second class.

b) A greater difference exists for the first class between the probability distributions of the number of lower-division students obtained for sampling with replacement (binomial distribution) and for sampling without replacement (hypergeometric distribution). Indeed, the sample size is smaller relative to the population size in the second class than in the first class, namely, 5% and 50%, respectively.

5.76 Let A denote the event that a shipment is good—it contains three or fewer defective parts. Set X equal to the number of defective parts in a random sample of size 10 from a shipment. We want to find the smallest integer k so that $P(X < k \mid A) \geq 0.95$. To that end, let A_j denote the event that the shipment contains exactly j defective parts. We note that the A_js are mutually exclusive and that $A = \bigcup_{j=0}^{3} A_j$. Suppose that $P(X < k \mid A_j) \geq 0.95$ for $0 \leq j \leq 3$. Then,

$$P(X < k \mid A) = \frac{P(\{X < k\} \cap A)}{P(A)} = \frac{\sum_{j=0}^{3} P(\{X < k\} \cap A_j)}{\sum_{j=0}^{3} P(A_j)} = \frac{\sum_{j=0}^{3} P(A_j) P(X < k \mid A_j)}{\sum_{j=0}^{3} P(A_j)}$$

$$\geq \frac{\sum_{j=0}^{3} P(A_j) \cdot 0.95}{\sum_{j=0}^{3} P(A_j)} = 0.95 \cdot \frac{\sum_{j=0}^{3} P(A_j)}{\sum_{j=0}^{3} P(A_j)} = 0.95.$$

Thus, we want the smallest k such that $P(X < k \mid A_j) \geq 0.95$ for $0 \leq j \leq 3$. Now, given that event A_j occurs, we know that $X \sim \mathcal{H}(100, 10, 0.0j)$. Thus,

$$P(X < k \mid A_j) = \sum_{i=0}^{k-1} \frac{\binom{j}{i}\binom{100-j}{10-i}}{\binom{100}{10}} = \frac{\sum_{i=0}^{\min\{k-1,j\}} \binom{j}{i}\binom{100-j}{10-i}}{\binom{100}{10}}.$$

We therefore obtain the following table:

k	$P(X < k \mid A_0)$	$P(X < k \mid A_1)$	$P(X < k \mid A_2)$	$P(X < k \mid A_3)$
1	1	0.9	0.8091	0.7265
2	1	1	0.9909	0.9742

From the table, we see that the smallest k such that $P(X < k \mid A_j) \geq 0.95$, for $0 \leq j \leq 3$, is $k = 2$.

5.5 Poisson Random Variables

Basic Exercises

5.77 The PMF of X is given by

$$p_X(x) = e^{-3}\frac{3^x}{x!}, \qquad x = 0, 1, 2, \ldots,$$

and $p_X(x) = 0$ otherwise.

a) We have

$$P(X = 3) = p_X(3) = e^{-3}\frac{3^3}{3!} = 4.5e^{-3} = 0.224.$$

b) From the FPF,

$$P(X < 3) = \sum_{x<3} p_X(x) = p_X(0) + p_X(1) + p_X(2) = e^{-3}\frac{3^0}{0!} + e^{-3}\frac{3^1}{1!} + e^{-3}\frac{3^2}{2!}$$

$$= e^{-3}\left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right) = 8.5e^{-3} = 0.423.$$

c) From the complementation rule and the FPF,

$$P(X > 3) = 1 - P(X \le 3) = 1 - \sum_{x \le 3} p_X(x) = 1 - \big(p_X(0) + p_X(1) + p_X(2) + p_X(3)\big)$$

$$= 1 - \left(e^{-3}\frac{3^0}{0!} + e^{-3}\frac{3^1}{1!} + e^{-3}\frac{3^2}{2!} + e^{-3}\frac{3^3}{3!}\right) = 1 - 13e^{-3} = 0.353.$$

d) From the complementation rule and part (c),

$$P(X \le 3) = 1 - P(X > 3) = 1 - 0.353 = 0.647.$$

e) From the complementation rule and part (b),

$$P(X \ge 3) = 1 - P(X < 3) = 1 - 0.423 = 0.577.$$

5.78 Let X denote the number of wars that began during the selected calendar year. The PMF of X is given by

$$p_X(x) = e^{-0.7}\frac{(0.7)^x}{x!}, \qquad x = 0, 1, 2, \dots,$$

and $p_X(x) = 0$ otherwise.
a) We have

$$P(X = 0) = p_X(0) = e^{-0.7}\frac{(0.7)^0}{0!} = e^{-0.7} = 0.497.$$

b) From the FPF,

$$P(X \le 2) = \sum_{x \le 2} p_X(x) = p_X(0) + p_X(1) + p_X(2)$$

$$= e^{-0.7}\frac{(0.7)^0}{0!} + e^{-0.7}\frac{(0.7)^1}{1!} + e^{-0.7}\frac{(0.7)^2}{2!}$$

$$= e^{-0.7}\left(\frac{(0.7)^0}{0!} + \frac{(0.7)^1}{1!} + \frac{(0.7)^2}{2!}\right) = 1.945e^{-0.7} = 0.966.$$

c) From the FPF,

$$P(1 \le X \le 3) = \sum_{1 \le x \le 3} p_X(x) = p_X(1) + p_X(2) + p_X(3)$$

$$= e^{-0.7}\frac{(0.7)^1}{1!} + e^{-0.7}\frac{(0.7)^2}{2!} + e^{-0.7}\frac{(0.7)^3}{3!}$$

$$= e^{-0.7}\left(\frac{(0.7)^1}{1!} + \frac{(0.7)^2}{2!} + \frac{(0.7)^3}{3!}\right) = 0.498.$$

5.79 Let X denote the number of Type I calls made from the motel during a 1-hour period. The PMF of X is given by

$$p_X(x) = e^{-1.7}\frac{(1.7)^x}{x!}, \qquad x = 0, 1, 2, \dots,$$

and $p_X(x) = 0$ otherwise.
a) We have

$$P(X = 1) = p_X(1) = e^{-1.7}\frac{(1.7)^1}{1!} = 1.7e^{-1.7} = 0.311.$$

b) From the FPF,

$$
\begin{aligned}
P(X \leq 2) &= \sum_{x \leq 2} p_X(x) = p_X(0) + p_X(1) + p_X(2) \\
&= e^{-1.7}\,\frac{(1.7)^0}{0!} + e^{-1.7}\,\frac{(1.7)^1}{1!} + e^{-1.7}\,\frac{(1.7)^2}{2!} \\
&= e^{-1.7}\left(\frac{(1.7)^0}{0!} + \frac{(1.7)^1}{1!} + \frac{(1.7)^2}{2!}\right) = 4.145e^{-1.7} = 0.757.
\end{aligned}
$$

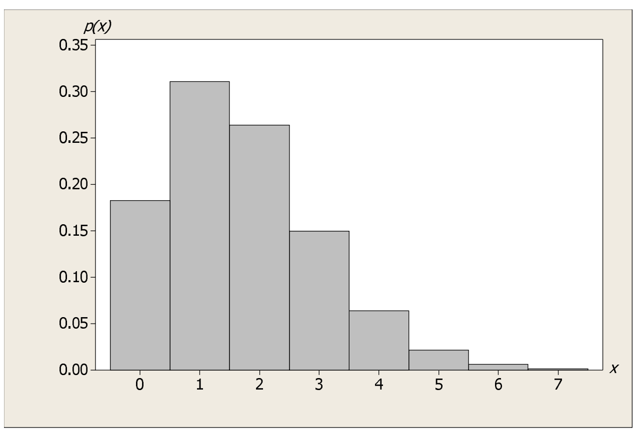
c) From the complementation rule and FPF,

$$
\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) = 1 - \sum_{x<2} p_X(x) = 1 - \big(p_X(0) + p_X(1)\big) \\
&= 1 - \left(e^{-1.7}\,\frac{(1.7)^0}{0!} + e^{-1.7}\,\frac{(1.7)^1}{1!}\right) = 1 - 2.7e^{-1.7} = 0.507.
\end{aligned}
$$

d) Using the formula for the PMF of X, we obtain the following table:

x	0	1	2	3	4	5	6	7
$p_X(x)$	0.183	0.311	0.264	0.150	0.064	0.022	0.006	0.001

e) In the following probability histogram, we use $p(x)$ instead of $p_X(x)$.



5.80 Let X denote the number of males of the 10,000 sampled who have Fragile X Syndrome. We have $n = 10{,}000$ and $p = 1/1500$. Referring to Proposition 5.7 on page 220 and noting that $np = 20/3$, we get

$$
p_X(x) \approx e^{-20/3}\,\frac{(20/3)^x}{x!}, \qquad x = 0, 1, \ldots, 10{,}000.
$$

a) From the complementation rule and FPF,

$$
\begin{aligned}
P(X > 7) &= 1 - P(X \leq 7) = 1 - \sum_{x \leq 7} p_X(x) = 1 - \sum_{k=0}^{7} p_X(k) \\
&\approx 1 - e^{-20/3} \sum_{k=0}^{7} \frac{(20/3)^k}{k!} = 0.352.
\end{aligned}
$$

b) From the FPF,

$$P(X \le 10) = \sum_{x \le 10} p_X(x) = \sum_{k=0}^{10} p_X(k) \approx e^{-20/3} \sum_{k=0}^{10} \frac{(20/3)^k}{k!} = 0.923.$$

c) From the note following Proposition 5.7, the maximum error is

$$np^2 = 10{,}000 \cdot (1/1500)^2 = \frac{1}{225} \approx 0.004.$$

5.81 Let X denote the number of golfers who would get a hole in one on the sixth hole. We have $n = 155$ and $p = 1/3709$.
a) From the complementation rule and FPF,

$$P(X \ge 4) = 1 - P(X < 4) = 1 - \sum_{x<4} p_X(x) = 1 - \sum_{k=0}^{3} \binom{155}{k} (1/3709)^k (3708/3709)^{155-k}$$
$$= 1 - 0.999999882 = 0.000000118.$$

b) Referring to Proposition 5.7 on page 220 and noting that $np = 155/3709$, we get

$$p_X(x) \approx e^{-155/3709} \frac{(155/3709)^x}{x!}, \qquad x = 0, 1, \ldots, 155.$$

Therefore, from the complementation rule and FPF,

$$P(X \ge 4) = 1 - P(X < 4) = 1 - \sum_{x<4} p_X(x) = 1 - \sum_{k=0}^{3} p_X(k)$$
$$\approx 1 - e^{-155/3709} \sum_{k=0}^{3} \frac{(155/3709)^k}{k!} = 1 - 0.999999877 = 0.000000123.$$

c) We are assuming that the results of attempts at holes in one by PGA golfers are independent and that the chance of making a hole in one is the same for all such golfers and for every such attempt. These assumptions appear reasonable because the occurrence or non-occurrence of a hole in one by one PGA golfer probably does not change the odds of making a hole in one for other PGA golfers. Morever, because all PGA golfers are quite skillful and because the chance of making a hole in one is so small, it appears reasonable to presume that the probability of making a hole in one is (approximately) $1/3709$ for all PGA golfers.

5.82 We have

$$p_X(x) = e^{-20} \frac{(20)^x}{x!}, \qquad x = 0, 1, 2, \ldots,$$

and $p_X(x) = 0$ otherwise.
a) From the complementation rule,

$$P(X \ge 20) = 1 - P(X < 20) = 1 - P(X \le 20) + P(X = 20)$$
$$= 1 - 0.559 + e^{-20} \frac{(20)^{20}}{20!} = 1 - 0.559 + 0.0888 = 0.530.$$

b) From the conditional probability rule and part (a),

$$P(X = 25 \mid X \ge 20) = \frac{P(X = 25,\ X \ge 20)}{P(X \ge 20)} = \frac{P(X = 25)}{P(X \ge 20)} = \frac{e^{-20}(20)^{25}/25!}{0.530} = 0.0841.$$

5.83 Let X denote the number of colds that a person has had during the year. Also, let E denote the event that the drug was effective for the person. As the drug is effective for 8 of 10 people, we have $P(E) = 0.8$. Moreover, given E, $X \sim \mathcal{P}(0.75)$, whereas, given E^c, $X \sim \mathcal{P}(3)$. Therefore, by Bayes's rule,

$$\begin{aligned} P(E \mid X = 1) &= \frac{P(E)P(X = 1 \mid E)}{P(E)P(X = 1 \mid E) + P(E^c)P(X = 1 \mid E^c)} \\ &= \frac{0.8 \cdot e^{-0.75}(0.75)^1/1!}{0.8 \cdot e^{-0.75}(0.75)^1/1! + 0.2 \cdot e^{-3}\, 3^1/1!} \\ &= \frac{1}{1 + e^{-2.25}} = 0.905. \end{aligned}$$

5.84 Let X denote the number of people in a random sample of size n who have characteristic α. Because the sampling is without replacement, $X \sim \mathcal{H}(N, n, 0.005)$, where N is the population size. Because N is not specified, we must use the corresponding binomial distribution, $\mathcal{B}(n, 0.005)$, as the "exact" distribution of X. We want $P(X \geq 1) > 0.9$ or, equivalently, $P(X = 0) < 0.1$. Using the binomial distribution, this condition becomes

$$(0.995)^n < 0.1 \qquad \text{or} \qquad n > \frac{\ln 0.1}{\ln 0.995} \approx 459.4.$$

Hence, using the binomial distribution, the sample size must be at least 460. Applying the Poisson approximation, the condition becomes

$$e^{-0.005n} < 0.1 \qquad \text{or} \qquad n > -\frac{\ln 0.1}{0.005} \approx 460.5.$$

Thus, using the Poisson approximation, the sample size must be at least 461. The answers obtained by using the binomial distribution and Poisson approximation agree almost perfectly (i.e., they differ by 1).

5.85 Let X denote the number of eggs in a nest, and let Y denote the number of eggs observed. We know that $X \sim \mathcal{P}(4)$. Moreover, because an empty nest cannot be recognized, we have, for $k \in \mathcal{N}$,

$$\begin{aligned} P(Y = k) = P(X = k \mid X > 0) &= \frac{P(X = k,\ X > 0)}{P(X > 0)} = \frac{P(X = k)}{P(X > 0)} \\ &= \frac{P(X = k)}{1 - P(X = 0)} = \frac{e^{-4}\, 4^k/k!}{1 - e^{-4}} = \left(\frac{e^{-4}}{1 - e^{-4}}\right)\frac{4^k}{k!}. \end{aligned}$$

Hence,

$$p_Y(y) = \left(\frac{e^{-4}}{1 - e^{-4}}\right)\frac{4^y}{y!}, \qquad y = 1, 2, 3, \ldots,$$

and $p_Y(y) = 0$ otherwise.

Theory Exercises

5.86

a) We first note that, because $np_n \to \lambda$ as $n \to \infty$, we have

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} \frac{np_n}{n} = \frac{\lambda}{\lim_{n\to\infty} n} = 0.$$

Now, let x be a nonnegative integer. Then, for n sufficiently large,

$$
\begin{aligned}
p_{X_n}(x) &= \binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{(n)_x}{x!} p_n^x (1-p_n)^{-x} (1-p_n)^n \\
&= \frac{(n)_x}{x!} \frac{n^x}{n^x} p_n^x (1-p_n)^{-x} \left(1 - \frac{np_n}{n}\right)^n = \frac{(n)_x}{n^x} \cdot \frac{(np_n)^x}{x!} \cdot (1-p_n)^{-x} \cdot \left(1 - \frac{np_n}{n}\right)^n .
\end{aligned}
$$

However,

$$
\lim_{n\to\infty} \frac{(n)_x}{n^x} = \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} = \lim_{n\to\infty} 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) = 1
$$

and, by the hint,

$$
\lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^n = e^{-\lambda}.
$$

Therefore,

$$
\lim_{n\to\infty} p_{X_n}(x) = 1 \cdot \frac{\lambda^x}{x!} \cdot (1-0)^{-x} \cdot e^{-\lambda} = e^{-\lambda} \frac{\lambda^x}{x!}.
$$

b) Let $X \sim \mathcal{B}(n, p)$, where n is large and p is small. Set $X_n = X$, $p_n = p$, and $\lambda = np$. Note that $np_n = \lambda$. By part (a), we have, for $x = 0, 1, \ldots, n$,

$$
p_X(x) = p_{X_n}(x) \approx e^{-\lambda} \frac{\lambda^x}{x!} = e^{-np} \frac{(np)^x}{x!},
$$

which is Equation (5.21) on page 220. Thus, part (a) provides the theoretical basis for Proposition 5.7.

5.87 For $k \in \mathcal{N}$, we have

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

Therefore, It follows that

$$
\begin{aligned}
p_X(k) &> p_X(k-1), \quad \text{if } k < \lambda; \\
p_X(k) &= p_X(k-1), \quad \text{if } k = \lambda; \\
p_X(k) &< p_X(k-1), \quad \text{if } k > \lambda.
\end{aligned}
\tag{$*$}
$$

Let's now consider two cases. Recall that $m = \lfloor \lambda \rfloor$.

Case 1: λ isn't an integer.

We have $k < \lambda$ if and only if $k \leq m$ and we have $k > \lambda$ if and only if $k \geq m + 1$. Thus, from Relations ($*$), we see that $p_X(k)$ is strictly increasing as k goes from 0 to m and is strictly decreasing as k goes from m to ∞.

Case 2: λ is an integer.

We have $k < \lambda$ if and only if $k \leq m - 1$ and we have $k > \lambda$ if and only if $k \geq m + 1$. Thus, from Relations ($*$), we see that $p_X(k)$ is strictly increasing as k goes from 0 to $m - 1$, that $p_X(m-1) = p_X(m)$, and that $p_X(k)$ is strictly decreasing as k goes from m to ∞.

5.88

a) Using integration by parts with $u = t^{k+1}$ and $dv = e^{-t}\,dt$, we get

$$\frac{1}{(k+1)!}\int_a^\infty t^{k+1}e^{-t}\,dt = -\frac{1}{(k+1)!}\Big[t^{k+1}e^{-t}\Big]_a^\infty + \frac{1}{(k+1)!}\int_a^\infty (k+1)t^k e^{-t}\,dt$$
$$= e^{-a}\frac{a^{k+1}}{(k+1)!} + \frac{1}{k!}\int_a^\infty t^k e^{-t}\,dt.$$

b) We want to prove that

$$\int_0^\infty t^k e^{-t}\,dt = k!, \qquad k = 0, 1, 2, \ldots. \tag{$*$}$$

Now, because $\int_0^\infty e^{-t}\,dt = 1 = 0!$, Equation $(*)$ holds for $k = 0$. Assuming that Equation $(*)$ holds for k, we get, in view of part (a), that

$$\int_0^\infty t^{k+1}e^{-t}\,dt = e^{-0}\,0^{k+1} + \frac{(k+1)!}{k!}\int_0^\infty t^k e^{-t}\,dt = 0 + (k+1)k! = (k+1)!,$$

as required.

c) Let

$$A_k = \frac{1}{k!}\int_\lambda^\infty t^k e^{-t}\,dt, \qquad k = 0, 1, 2, \ldots.$$

Noting that $A_0 = e^{-\lambda}$, we get, from the FPF and part (a), that

$$P(X \le k) = \sum_{x \le k} p_X(x) = \sum_{j=0}^{k} e^{-\lambda}\frac{\lambda^j}{j!} = e^{-\lambda} + \sum_{j=1}^{k} e^{-\lambda}\frac{\lambda^j}{j!} = A_0 + \sum_{j=1}^{k}(A_j - A_{j-1})$$
$$= A_k = \frac{1}{k!}\int_\lambda^\infty t^k e^{-t}\,dt = \frac{1}{k!}\Gamma(k+1, \lambda).$$

d) Applying, in turn, the complementation rule, part (c), and part (b), we get

$$P(X > k) = 1 - P(X \le k) = 1 - \frac{1}{k!}\int_\lambda^\infty t^k e^{-t}\,dt = 1 - \frac{1}{k!}\left(\int_0^\infty t^k e^{-t}\,dt - \int_0^\lambda t^k e^{-t}\,dt\right)$$
$$= 1 - \frac{1}{k!}\left(k! - \int_0^\lambda t^k e^{-t}\,dt\right) = \frac{1}{k!}\int_0^\lambda t^k e^{-t}\,dt = \frac{1}{k!}\gamma(k+1, \lambda).$$

Advanced Exercises

5.89 Let X denote the number of defective bolts in a box of 100. Then $X \sim \mathcal{B}(100, 0.015)$. We can compute probabilites for X exactly by using the binomial PDF

$$p_X(x) = \binom{100}{x}(0.015)^x(0.985)^{100-x}, \qquad x = 0, 1, \ldots, 100,$$

and $p_X(x) = 0$ otherwise. However, it is computationally simpler to use the Poisson approximation:

$$p_X(x) \approx e^{-1.5}\frac{(1.5)^x}{x!}, \qquad x = 0, 1, \ldots, 100.$$

a) We have

$$P(X = 0) = p_X(0) \approx e^{-1.5}\,\frac{(1.5)^0}{0!} = e^{-1.5} = 0.223.$$

b) From the FPF,

$$\begin{aligned} P(X \le 2) &= \sum_{x \le 2} p_X(x) = \sum_{k=0}^{2} p_X(k) \\ &\approx e^{-1.5}\,\frac{(1.5)^0}{0!} + e^{-1.5}\,\frac{(1.5)^1}{1!} + e^{-1.5}\,\frac{(1.5)^2}{2!} \\ &= 3.625e^{-1.5} = 0.809. \end{aligned}$$

c) We again use the Poisson approximation. Let X denote the number of defective bolts in a box that contains $100 + m$ bolts. We know that

$$X \sim \mathcal{B}(100 + m, 0.015) \approx \mathcal{P}(1.5 + 0.015m).$$

We want to choose the smallest m so that $P(X \le m) \ge 0.95$. Now,

$$P(X \le m) = \sum_{x \le m} p_X(x) = \sum_{k=0}^{m} p_X(x) \approx e^{-(1.5+0.015m)} \sum_{x=0}^{m} \frac{(1.5 + 0.015m)^x}{x!}.$$

Using the preceding result, we obtain the following table:

m	0	1	2	3	4
$P(X \le m)$	0.223	0.553	0.801	0.929	0.978

From the table, we see that the required m is 4; that is, the minimum number of bolts that must be placed in a box to ensure that 100 or more of the bolts will be nondefective at least 95% of the time is 104.

5.90 Let X denote the number of customers per day who arrive at Grandma's Fudge Shoppe, and let Y denote the number who make a purchase. We know that $X \sim \mathcal{P}(\lambda)$. Furthermore, given that $X = n$, we have $Y \sim \mathcal{B}(n, p)$. Applying the law of total probability and Equation (5.26) on page 222, we get, for each nonnegative integer k, that

$$\begin{aligned} P(Y = k) &= \sum_{n=k}^{\infty} P(X = n)P(Y = k \mid X = n) = \sum_{n=k}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\binom{n}{k} p^k (1-p)^{n-k} \\ &= e^{-\lambda}\frac{\lambda^k p^k}{k!} \sum_{n=k}^{\infty} \frac{\lambda^{n-k}}{(n-k)!}(1-p)^{n-k} = e^{-\lambda}\frac{(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{\big(\lambda(1-p)\big)^{n-k}}{(n-k)!} \\ &= e^{-\lambda}\frac{(\lambda p)^k}{k!} \sum_{j=0}^{\infty} \frac{\big(\lambda(1-p)\big)^{j}}{j!} = e^{-\lambda}\frac{(\lambda p)^k}{k!} e^{\lambda(1-p)} = e^{-p\lambda}\frac{(p\lambda)^k}{k!}. \end{aligned}$$

Hence, $Y \sim \mathcal{P}(p\lambda)$, that is, Y has the Poisson distribution with parameter $p\lambda$.

5.91 For nonnegative integers x and y,

$$p_X(x) = e^{-\lambda}\,\frac{\lambda^x}{x!} \qquad \text{and} \qquad p_Y(y) = e^{-\mu}\,\frac{\mu^y}{y!},$$

and $p_X(x) = p_Y(y) = 0$ otherwise.

a) Let z be a nonnegative integer. Then, by the law of total probability, the assumed independence condition, and the binomial theorem,

$$\begin{aligned}
p_{X+Y}(z) &= P(X+Y=z) = \sum_{x=0}^{\infty} P(X=x)P(X+Y=z \mid X=x) \\
&= \sum_{x=0}^{\infty} P(X=x)P(Y=z-x \mid X=x) = \sum_{x=0}^{\infty} P(X=x)P(Y=z-x) \\
&= \sum_{x=0}^{z} e^{-\lambda}\,\frac{\lambda^x}{x!}\cdot e^{-\mu}\,\frac{\mu^{z-x}}{(z-x)!} = e^{-(\lambda+\mu)}\,\frac{1}{z!}\sum_{x=0}^{z}\binom{z}{x}\lambda^x\mu^{z-x} = e^{-(\lambda+\mu)}\,\frac{1}{z!}(\lambda+\mu)^z. \\
&= e^{-(\lambda+\mu)}\,\frac{(\lambda+\mu)^z}{z!}.
\end{aligned}$$

Therefore, $X+Y \sim \mathcal{P}(\lambda+\mu)$.

b) From the assumed independence condition, we have

$$P(X+Y=3 \mid X=1) = P(Y=2 \mid X=1) = P(Y=2) = p_Y(2) = e^{-\mu}\,\frac{\mu^2}{2!} = \frac{\mu^2}{2}e^{-\mu}.$$

5.92 From the exponential series,

$$e^t = \sum_{k=0}^{\infty}\frac{t^k}{k!} = \sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!} + \sum_{k=0}^{\infty}\frac{t^{2k+1}}{(2k+1)!}.$$

It follows that

$$e^{-t} = \sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!} - \sum_{k=0}^{\infty}\frac{t^{2k+1}}{(2k+1)!}$$

and, therefore, that

$$e^t + e^{-t} = 2\sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!}.$$

Applying the FPF, we get that

$$\begin{aligned}
P(X \text{ is even}) &= \sum_{x \text{ is even}} p_X(x) = \sum_{k=0}^{\infty} p_X(2k) = \sum_{k=0}^{\infty} e^{-\lambda}\,\frac{\lambda^{2k}}{(2k)!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} \\
&= e^{-\lambda}\,\frac{e^{\lambda}+e^{-\lambda}}{2} = \frac{1+e^{-2\lambda}}{2}.
\end{aligned}$$

5.6 Geometric Random Variables

Basic Exercises

5.93
a) From Proposition 5.9 on page 229, we see that $X \sim \mathcal{G}(0.67)$. Hence, $p_X(x) = (0.67)(0.33)^{x-1}$ if $x \in \mathcal{N}$, and $p_X(x) = 0$ otherwise.
b) From part (a),

$$P(X = 3) = p_X(3) = (0.67)(0.33)^{3-1} = (0.67)(0.33)^2 = 0.0730.$$

From Proposition 5.10 on page 231,

$$P(X \geq 3) = P(X > 2) = (1 - 0.67)^2 = (0.33)^2 = 0.109.$$

From the complementation rule and Proposition 5.10,

$$P(X \leq 3) = 1 - P(X > 3) = 1 - (1 - 0.67)^3 = 1 - (0.33)^3 = 0.964.$$

c) We want to determine n so that $P(X \leq n) \geq 0.99$ or, equivalently, $P(X > n) \leq 0.01$. From Proposition 5.10, this last relation is equivalent to $(1 - 0.67)^n \leq 0.01$. Thus, we want

$$(0.33)^n \leq 0.01 \qquad \text{or} \qquad n \geq \frac{\ln 0.01}{\ln 0.33} \approx 4.2.$$

Hence, you must bet on five (or more) races to be at least 99% sure of winning something.

5.94 Let X denote the number of at-bats up to and including the first hit. Then $X \sim \mathcal{G}(0.26)$.
a) We have

$$P(X = 5) = p_X(5) = (0.26)(1 - 0.26)^{5-1} = (0.26)(0.74)^4 = 0.0780.$$

b) From Proposition 5.10 on page 231,

$$P(X > 5) = (1 - 0.26)^5 = (0.74)^5 = 0.222.$$

c) From Equation (5.31) on page 230 and Proposition 5.10 on page 231,

$$\begin{aligned} P(3 \leq X \leq 10) &= P(2 < X \leq 10) = P(X > 2) - P(X > 10) \\ &= (1 - 0.26)^2 - (1 - 0.26)^{10} = (0.74)^2 - (0.74)^{10} \\ &= 0.498. \end{aligned}$$

5.95
a) From the FPF,

$$\begin{aligned} P(X \text{ is even}) &= \sum_{x \text{ is even}} p_X(x) = \sum_{k=1}^{\infty} p_X(2k) = \sum_{k=1}^{\infty} p(1-p)^{2k-1} \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} \left((1-p)^2\right)^k = \frac{p}{1-p} \cdot \frac{(1-p)^2}{1-(1-p)^2} = \frac{1-p}{2-p}. \end{aligned}$$

b) From part (a) and the complementation rule,

$$P(X \text{ is odd}) = 1 - P(X \text{ is even}) = 1 - \frac{1-p}{2-p} = \frac{1}{2-p}.$$

c) From the lack-of-memory property, the complementation rule, and Proposition 5.10 on page 231,

$$\begin{aligned} P(2 \le X \le 9 \mid X \ge 4) &= P(4 \le X \le 9 \mid X \ge 4) = P(4 \le X \le 9 \mid X > 3) \\ &= \sum_{k=4}^{9} P(X = k \mid X > 3) = \sum_{j=1}^{6} P(X = 3 + j \mid X > 3) = \sum_{j=1}^{6} P(X = j) \\ &= P(X \le 6) = 1 - P(X > 6) = 1 - (1-p)^6. \end{aligned}$$

d) From the conditional probability rule, the complementation rule, and Proposition 5.10, we have, for $k = 1, 2, \ldots, n$, that

$$P(X = k \mid X \le n) = \frac{P(X = k,\ X \le n)}{P(X \le n)} = \frac{P(X = k)}{1 - P(X > n)} = \frac{p(1-p)^{k-1}}{1 - (1-p)^n}.$$

e) Applying part (d) with $n - 1$ replacing n and with $n - k$ replacing k, we have, for $k = 1, 2, \ldots, n - 1$, that

$$P(X = n - k \mid X < n) = P(X = n - k \mid X \le n - 1) = \frac{p(1-p)^{(n-k)-1}}{1 - (1-p)^{n-1}} = \frac{p(1-p)^{n-k-1}}{1 - (1-p)^{n-1}}.$$

5.96 Applying Equation (5.35) on page 233, we see that

$$P(X = 10 \mid X > 6) = P(X = 6 + 4 \mid X > 6) = P(X = 4).$$

Hence, it is the first equation that is implied by the lack-of-memory property.

5.97

a) From the conditional probability rule and Proposition 5.10 on page 231,

$$P(X = k \mid X > n) = \frac{P(X = k,\ X > n)}{P(X > n)} = \frac{P(X = k)}{P(X > n)} = \frac{p(1-p)^{k-1}}{(1-p)^n} = p(1-p)^{k-n-1}.$$

b) From the lack-of-memory property, Equation (5.35) on page 233,

$$P(X = k \mid X > n) = P\big(X = n + (k - n) \mid X > n\big) = P(X = k - n) = p(1-p)^{k-n-1}.$$

5.98 Let

$$\begin{aligned} I &= \text{event that Coin I is selected, and} \\ II &= \text{event that Coin II is selected.} \end{aligned}$$

Because the coin is selected at random, $P(I) = P(II) = 0.5$. Also, given I, we have $X \sim \mathcal{G}(1 - p_1)$, whereas, given II, we have $X \sim \mathcal{G}(1 - p_2)$.

a) Let $k \in \mathcal{N}$. Then, by the law of total probability,

$$\begin{aligned} P(X = k) &= P(I)P(X = k \mid I) + P(II)P(X = k \mid II) = 0.5(1-p_1)p_1^{k-1} + 0.5(1-p_2)p_2^{k-1} \\ &= 0.5\left((1-p_1)p_1^{k-1} + (1-p_2)p_2^{k-1}\right). \end{aligned}$$

Hence,

$$p_X(x) = 0.5\left((1-p_1)p_1^{x-1} + (1-p_2)p_2^{x-1}\right), \qquad x \in \mathcal{N},$$

and $p_X(x) = 0$ otherwise.

b) If $p_1 = p_2$, say, both equal to p, then, by part (a),

$$p_X(x) = 0.5\left((1-p)p^{x-1} + (1-p)p^{x-1}\right) = (1-p)p^{x-1} = (1-p)\big(1 - (1-p)\big)^{x-1}$$

if $x \in \mathcal{N}$, and $p_X(x) = 0$ otherwise. Thus, in this case, $X \sim \mathcal{G}(1 - p)$.

5.99

a) When sampling is with replacement, the sampling process constitutes Bernoulli trials with success probability p. Consequently, from Proposition 5.9 on page 229, X has the geometric distribution with parameter p. Thus,

$$p_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \dots,$$

and $p_X(x) = 0$ otherwise.

b) When sampling is without replacement, the sampling process doesn't constitute Bernoulli trials. The event $\{X = x\}$ occurs if and only if none of the first $x - 1$ members sampled have the specified attribute (event A) and the xth member sampled has the specified attribute (event B). Several approaches are possible for obtaining $P(X = x)$. Here, we proceed as follows. Applying the hypergeometric distribution with parameters N, $x - 1$, and p, we get

$$P(A) = \frac{\binom{Np}{0}\binom{N(1-p)}{x-1}}{\binom{N}{x-1}} = \frac{\binom{N(1-p)}{x-1}}{\binom{N}{x-1}}.$$

Clearly, $P(B \mid A) = Np/(N - x + 1)$. Hence, by the general multiplication rule,

$$\begin{aligned} P(X = x) = P(A \cap B) = P(A)P(B \mid A) &= \frac{\binom{N(1-p)}{x-1}}{\binom{N}{x-1}} \frac{Np}{N - x + 1} \\ &= \frac{\big(N(1-p)\big)_{x-1}(Np)}{(N)_x} = p\frac{\big(N(1-p)\big)_{x-1}}{(N-1)_{x-1}}. \end{aligned}$$

Thus,

$$p_X(x) = p\frac{\big(N(1-p)\big)_{x-1}}{(N-1)_{x-1}}, \qquad x = 1, 2, \dots, N(1-p) + 1,$$

and $p_X(x) = 0$ otherwise.

c) Here we consider the specified attributed to be "house key." We observe that, in the notation of part (b), we have $p = 1/N$. Let C_n denote the event that you get the house key on the nth try, where $1 \le n \le N$. Applying the result of part (b) with $x = n$, we get

$$P(C_n) = p_X(n) = (1/N)\frac{\big(N(1-1/N)\big)_{n-1}}{(N-1)_{n-1}} = (1/N)\frac{(N-1)_{n-1}}{(N-1)_{n-1}} = 1/N.$$

5.100

a) Let A_n denote the event that Tom wins on toss $2n + 1$. Also, let X_1 and X_2 denote the number of tosses required for Tom and Dick to get a head, respectively. Observe that $X_j \sim \mathcal{G}(p_j)$, $j = 1, 2$. We have

$$\begin{aligned} P(A_n) &= P(X_1 = n+1,\ X_2 > n) = P(X_1 = n+1)P(X_2 > n) \\ &= p_1(1-p_1)^{(n+1)-1}(1-p_2)^n = p_1\big((1-p_1)(1-p_2)\big)^n. \end{aligned}$$

Letting T denote the event that Tom wins and noting that $A_1, A_2, \ldots$ are mutually exclusive events, we get

$$P(T) = P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n) = p_1 \sum_{n=0}^{\infty}\big((1-p_1)(1-p_2)\big)^n$$
$$= p_1 \cdot \frac{1}{1-(1-p_1)(1-p_2)} = \frac{p_1}{p_1+p_2-p_1p_2}.$$

b) If both coins are balanced, then $p_1 = p_2 = 0.5$ and, hence,

$$P(T) = \frac{p_1}{p_1+p_2-p_1p_2} = \frac{0.5}{0.5+0.5-0.5\cdot 0.5} = \frac{1/2}{3/4} = \frac{2}{3}.$$

c) For the game to be fair, we must have $P(T) = 0.5$. In view of part (a), this happens if and only if

$$\frac{p_1}{p_1+p_2-p_1p_2} = \frac{1}{2} \qquad \text{or} \qquad p_1 = \frac{p_2}{1+p_2}.$$

d) If Dick's coin is balanced, $p_2 = 0.5$. Then, in view of part (c), the game will be fair if and only if

$$p_1 = \frac{0.5}{1+0.5} \qquad \text{or} \qquad p_1 = \frac{1}{3}.$$

5.101

a) Let T denote the event that Tom wins and let E denote the event that Tom goes first. By assumption, we have $P(E) = 0.5$. Also, from Exercise 5.100(a), $P(T \mid E) = p_1/(p_1+p_2-p_1p_2)$. Interchanging the roles of Tom and Dick, it follows as well from Exercise 5.100(a) that

$$P(T \mid E^c) = 1 - P(T^c \mid E^c) = 1 - \frac{p_2}{p_1+p_2-p_1p_2} = \frac{p_1(1-p_2)}{p_1+p_2-p_1p_2}.$$

Hence, by the law of total probability,

$$P(T) = P(E)P(T \mid E) + P(E^c)P(T \mid E^c) = 0.5 \cdot \frac{p_1}{p_1+p_2-p_1p_2} + 0.5 \cdot \frac{p_1(1-p_2)}{p_1+p_2-p_1p_2}$$
$$= \frac{p_1 + p_1(1-p_2)}{2(p_1+p_2-p_1p_2)} = \frac{p_1(2-p_2)}{2(p_1+p_2-p_1p_2)}.$$

b) For the game to be fair, we must have $P(T) = 0.5$. In view of part (a), this happens if and only if

$$\frac{p_1(2-p_2)}{2(p_1+p_2-p_1p_2)} = \frac{1}{2} \qquad \text{or} \qquad p_1 = p_2,$$

a result that we could also obtain by using a symmetry argument.

5.102 From the solution to Exercise 5.85, the probability of observing a nest with six eggs is

$$\left(\frac{e^{-4}}{1-e^{-4}}\right)\frac{4^6}{6!} \approx 0.106.$$

Hence, the number of nests examined until a nest containing six eggs is observed, Y, has the geometric distribution with parameter (approximately) 0.106. That is,

$$p_Y(y) = (0.106)(1-0.106)^{y-1} = (0.106)(0.894)^{y-1}, \qquad y \in \mathcal{N},$$

and $p_Y(y) = 0$ otherwise.

Theory Exercises

5.103 We must verify that p_X satistfies the three properties stated in Proposition 5.1 on page 187. Clearly, we have $p_X(x) \geq 0$ for all x. Moreover, $\{ x \in \mathcal{R} : p_X(x) \neq 0 \} = \mathcal{N}$, which is countable. Finally, applying Equation (5.30) on page 230, we get

$$\sum_x p_X(x) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p \sum_{j=0}^{\infty} (1-p)^j = p \cdot \frac{1}{1-(1-p)} = 1.$$

5.104

a) In Bernoulli trials, the trials are independent and, furthermore, the success probability is positive. Therefore, from Proposition 4.6, in repeated Bernoulli trials, the probability is 1 that a success will eventually occur.

b) If we let X denote the number of trials up to and including the first success, then $X \sim \mathcal{G}(p)$. The probability that a success will eventually occur is

$$P(X < \infty) = \sum_{x<\infty} p_X(x) = \sum_x p_X(x) = 1.$$

5.105 We first observe that

$$\{X > a\} = \{a < X \leq b\} \cup \{X > b\}$$

and that the two events on the right of this equation are mutually exclusive. Hence, by the additivity property of a probability measure,

$$P(X > a) = P(a < X \leq b) + P(X > b) \qquad \text{or} \qquad P(a < X \leq b) = P(X > a) - P(X > b).$$

5.106 We first recall from Proposition 4.1 on page 134 that $P(\cdot \mid X > n)$ is a probability measure. If Equation (5.35) holds, then

$$P(X > n+k \mid X > n) = P\left(\bigcup_{j=k+1}^{\infty} \{X = n+j\} \mid X > n \right) = \sum_{j=k+1}^{\infty} P(X = n+j \mid X > n)$$

$$= \sum_{j=k+1}^{\infty} P(X = j) = \sum_{j=k+1}^{\infty} P(X = j) = P(X > k).$$

Conversely, suppose that $P(X > n+k \mid X > n) = P(X > k)$ for all $n, k \in \mathcal{N}$. Then,

$$\begin{aligned} P(X = n+k \mid X > n) &= P(X > n+k-1 \mid X > n) - P(X > n+k \mid X > n) \\ &= P(X > k-1) - P(X > k) = P(X = k), \end{aligned}$$

so that Equation (5.35) holds.

Advanced Exercises

5.107

a) If the lack-of-memory property holds for X, then

$$P(X > m \mid X > n) = P\big(X > n + (m-n) \mid X > n\big) = P(X > m-n),$$

where the last equality follows from Exercise 5.106. Hence, Equation $(*)$ is assured to hold by the lack-of-memory property.

b) The question is whether it is possible for X to satisfy Equation $(**)$. Suppose that it does. Then, we are assuming tacitly that $P(X > n) \neq 0$ for all $n \in \mathcal{N}$. By the conditional probability rule, we have

$$P(X > m) = P(X > m \mid X > n) = \frac{P(X > m,\ X > n)}{P(X > n)} = \frac{P(X > m)}{P(X > n)},$$

which implies that $P(X > n) = 1$ for all $n \in \mathcal{N}$. Note that this equation holds as well for $n = 0$ because X is positive-integer valued. Consequently, by Equation (5.32) on page 230,

$$p_X(n) = P(X > n - 1) - P(X > n) = 1 - 1 = 0, \qquad n \in \mathcal{N},$$

which, in turn, implies that

$$1 = \sum_x p_X(x) = \sum_{n=1}^{\infty} p_X(n) = 0,$$

which is a contradiction. Hence, it is not possible for X to satisfy Equation $(**)$.

5.108

a) For $k = 2, 3, \dots,$ we have, by the law of total probability and the independence condition, that

$$\begin{aligned} P(X + Y = k) &= \sum_n P(X = n)P(X + Y = k \mid X = n) \\ &= \sum_n P(X = n)P(Y = k - n \mid X = n) = \sum_n P(X = n)P(Y = k - n) \\ &= \sum_{n=1}^{k-1} p(1-p)^{n-1} p(1-p)^{k-n-1} = \sum_{n=1}^{k-1} p^2(1-p)^{k-2} = (k-1)p^2(1-p)^{k-2}. \end{aligned}$$

Hence,

$$p_{X+Y}(z) = (z-1)p^2(1-p)^{z-2}, \qquad z = 2, 3, \dots,$$

and $p_{X+Y}(z) = 0$ otherwise.

b) No, because, for instance, $p_{X+Y}(1) = 0$.

c) For convenience, set $Z = \min\{X, Y\}$. Then, by the independence condition and Proposition 5.10 on page 231, we have, for $n \in \mathcal{N}$, that

$$\begin{aligned} P(Z > n) &= P(\min\{X, Y\} > n) = P(X > n,\ Y > n) = P(X > n)P(Y > n) = (1-p)^n(1-p)^n \\ &= (1-p)^{2n} = \big((1-p)^2\big)^n = \Big(1 - \big(1 - (1-p)^2\big)\Big)^n = \Big(1 - \big(2p - p^2\big)\Big)^n. \end{aligned}$$

Applying again Proposition 5.10, we see that Z has the same tail probabilities as a geometric random variable with parameter $2p - p^2$. As tail probabilities determine the distribution, we conclude that $Z \sim \mathcal{G}\big(2p - p^2\big)$.

5.109

a) We observe that the event $\{X = x,\ Y = y\}$ occurs if and only if the first $x - 1$ trials are failures, the xth trial is success, the next $y - 1$ trials are failures, and the following trial is success. Hence,

$$P(X = x,\ Y = y) = (1-p)^{x-1} p (1-p)^{y-1} p = p^2(1-p)^{x+y-2}.$$

Applying the conditional probability rule, we find that

$$P(Y = y \mid X = x) = \frac{P(X = x,\ Y = y)}{P(X = x)} = \frac{p^2(1-p)^{x+y-2}}{p(1-p)^{x-1}} = p(1-p)^{y-1}.$$

Now applying the law of total probability, we get

$$\begin{aligned} P(Y = y) &= \sum_x P(X = x)P(Y = y \mid X = x) = \sum_x p_X(x)p(1-p)^{y-1} \\ &= p(1-p)^{y-1}\sum_x p_X(x) = p(1-p)^{y-1} \cdot 1 = p(1-p)^{y-1}. \end{aligned}$$

Thus, $P(Y = y \mid X = x) = P(Y = y)$, which means that events $\{X = x\}$ and $\{Y = y\}$ are independent.
b) We know that $X \sim \mathcal{G}(p)$ and, from the solution to part (a), we see that $Y \sim \mathcal{G}(p)$. Furthermore, from part (a), the independence condition of Exercise 5.108 is satisfied. Hence, from Exercise 5.108(a),

$$p_{X+Y}(z) = (z-1)p^2(1-p)^{z-2}, \qquad z = 2, 3, \ldots,$$

and $p_{X+Y}(z) = 0$ otherwise.
c) The random variable $X + Y$ gives the number of trials until the second success.
d) Event $\{X + Y = z\}$ occurs if and only if there is exactly one success in the first $z - 1$ trials and the zth trial is a success. Let A denote the former event and B the latter. Because Bernoulli trials are independent, events A and B are independent. Applying the binomial distribution, we have

$$P(A) = \binom{z-1}{1} p^1(1-p)^{(z-1)-1} = (z-1)p(1-p)^{z-2}.$$

Hence, for $z = 2, 3, \ldots,$

$$\begin{aligned} p_{X+Y}(z) &= P(X + Y = z) = P(A \cap B) = P(A)P(B) \\ &= (z-1)p(1-p)^{z-2} \cdot p = (z-1)p^2(1-p)^{z-2}. \end{aligned}$$

This result agrees with the one found in part (b).

5.7 Other Important Discrete Random Variables

Basic Exercises

5.110 We note that the only possible values of $I_{\{X=1\}}$ and X are 0 and 1. Hence, to prove that those two random variables are equal, it suffices to show that $X = 1$ if and only if $I_{\{X=1\}} = 1$. However,

$$X(\omega) = 1 \Leftrightarrow \omega \in \{X = 1\} \Leftrightarrow I_{\{X=1\}}(\omega) = 1.$$

5.111
a) We note that both $I_{E\cap F}$ and $I_E \cdot I_F$ are 1 if $E \cap F$ occurs and both are 0 otherwise. Consequently, we have $I_{E\cap F} = I_E \cdot I_F$.
b) Suppose that $E \cap F = \emptyset$. The only possible values of $I_{E\cup F}$ are 0 and 1 and, because $E \cap F = \emptyset$, the same is true for $I_E + I_F$. However,

$$I_{E\cup F}(\omega) = 0 \Leftrightarrow \omega \notin E \cup F \Leftrightarrow \omega \notin E \text{ and } \omega \notin F \Leftrightarrow I_E(\omega) = I_F(\omega) = 0 \Leftrightarrow (I_E + I_F)(\omega) = 0.$$

Consequently, $I_{E\cup F} = I_E + I_F$. Conversely, suppose that $E \cap F \neq \emptyset$. Let $\omega_0 \in E \cap F$. Then

$$I_{E\cup F}(\omega_0) = 1 \neq 2 = 1 + 1 = I_E(\omega_0) + I_F(\omega_0) = (I_E + I_F)(\omega_0).$$

Consequently, $I_{E\cup F} \neq I_E + I_F$. We have now established that a necessary and sufficient condition for $I_{E\cup F} = I_E + I_F$ is that $E \cap F = \emptyset$, that is, that E and F are mutually exclusive.

c) We claim that $I_{E \cup F} = I_E + I_F - I_{E \cap F}$, that is, that

$$I_{E \cup F}(\omega) = I_E(\omega) + I_F(\omega) - I_{E \cap F}(\omega), \qquad \omega \in \Omega.$$

To establish the preceding equation, we consider four cases.

Case 1: $\omega \in (E \cup F)^c$

Then $\omega \notin E \cup F$, $\omega \notin E$, $\omega \notin F$, and $\omega \notin E \cap F$. Hence,

$$I_{E \cup F}(\omega) = 0 = 0 + 0 - 0 = I_E(\omega) + I_F(\omega) - I_{E \cap F}(\omega).$$

Case 2: $\omega \in E \cap F$

Then $\omega \in E \cup F$, $\omega \in E$, and $\omega \in F$. Hence,

$$I_{E \cup F}(\omega) = 1 = 1 + 1 - 1 = I_E(\omega) + I_F(\omega) - I_{E \cap F}(\omega).$$

Case 3: $\omega \in E \cap F^c$

Then $\omega \in E \cup F$, $\omega \in E$, $\omega \notin F$, and $\omega \notin E \cap F$. Hence,

$$I_{E \cup F}(\omega) = 1 = 1 + 0 - 0 = I_E(\omega) + I_F(\omega) - I_{E \cap F}(\omega).$$

Case 4: $\omega \in E^c \cap F$

Then $\omega \in E \cup F$, $\omega \notin E$, $\omega \in F$, and $\omega \notin E \cap F$. Hence,

$$I_{E \cup F}(\omega) = 1 = 0 + 1 - 0 = I_E(\omega) + I_F(\omega) - I_{E \cap F}(\omega).$$

5.112 The only possible values of $I_{\bigcup_n A_n}$ are 0 and 1 and, because $A_1, A_2, \ldots$ are mutually exclusive, the same is true for $\sum_n I_{A_n}$. Hence, to prove that those two random variables are equal, it suffices to show that $I_{\bigcup_n A_n} = 0$ if and only if $\sum_n I_{A_n} = 0$. However,

$$I_{\bigcup_n A_n}(\omega) = 0 \Leftrightarrow \omega \notin \bigcup_n A_n \Leftrightarrow \omega \notin A_n \text{ for all } n \Leftrightarrow I_{A_n}(\omega) = 0 \text{ for all } n \Leftrightarrow \left(\sum_n I_{A_n}\right)(\omega) = 0.$$

5.113

a) For $1 \le k \le n$, let E_k denote the event that the kth trial results in success. Then $I_{E_k} = 0$ if the kth trial results in failure and $I_{E_k} = 1$ if the kth trial results in success. Hence, $\sum_{k=1}^n I_{E_k}$ gives the total number of successes in the n trials; that is, $X = \sum_{k=1}^n I_{E_k}$.

b) For $1 \le k \le n$, let X_k be 1 or 0 depending on whether the kth trial results in success or failure, respectively. Then each X_k is a Bernoulli random variable with parameter p and, moreover, $\sum_{k=1}^n X_k$ gives the total number of successes in the n trials; that is, $X = \sum_{k=1}^n X_k$.

5.114 Because we have a classical probability model, we know that $P(\{\omega\}) = 1/N$ for all $\omega \in \Omega$. Thus, for $k = 1, 2, \ldots, N$, we have

$$P(X = k) = P(\{\,\omega : X(\omega) = k\,\}) = P(\{\omega_k\}) = 1/N.$$

Hence, $p_X(x) = 1/N$ if $x \in \{1, 2, \ldots, N\}$, and $p_X(x) = 0$ otherwise. In other words, X has the discrete uniform distribution on the set $\{1, 2, \ldots, N\}$.

5.115 We have

$$p_X(x) = \begin{cases} 1/10, & \text{if } x \in S; \\ 0, & \text{otherwise.} \end{cases}$$

a) For $x \in S$,

$$P(Y = y \mid X = x) = \begin{cases} 1/9, & \text{if } y \in S \setminus \{x\}; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, by the law of total probability, for $y \in S$,

$$\begin{aligned} P(Y = y) &= \sum_x P(X = x)P(Y = y \mid X = x) = \sum_{x \in S} \frac{1}{10} P(Y = y \mid X = x) \\ &= \frac{1}{10} \sum_{x \in S} P(Y = y \mid X = x) = \frac{1}{10} \sum_{x \in S \setminus \{y\}} \frac{1}{9} = \frac{1}{10}\left(9 \cdot \frac{1}{9}\right) = \frac{1}{10}. \end{aligned}$$

Therefore,

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

Thus, Y has the discrete uniform distribution on S.
b) By symmetry, the value of Y is equally likely to be any of the 10 digits. Therefore,

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

Thus, Y has the discrete uniform distribution on S.

5.116
a) From Proposition 5.13 on page 241, we see that $X \sim \mathcal{NB}(3, 0.67)$. Hence,

$$p_X(x) = \binom{x-1}{2}(0.67)^3(0.33)^{x-3}, \qquad x = 3, 4, \ldots,$$

and $p_X(x) = 0$ otherwise.
b) From part (a),

$$P(X = 4) = p_X(4) = \binom{4-1}{2}(0.67)^3(0.33)^{4-3} = 3(0.67)^3(0.33) = 0.298.$$

From the complementation rule, the FPF, and part (a),

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) = 1 - \sum_{x<4} p_X(x) = 1 - \sum_{k=3}^{3} p_X(k) = 1 - p_X(3) \\ &= 1 - \binom{3-1}{2}(0.67)^3(0.33)^{3-3} = 1 - (0.67)^3 = 0.699. \end{aligned}$$

From the FPF and part (a),

$$\begin{aligned} P(X \leq 4) &= \sum_{x \leq 4} p_X(x) = \sum_{k=3}^{4} p_X(k) = \binom{3-1}{2}(0.67)^3(0.33)^{3-3} + \binom{4-1}{2}(0.67)^3(0.33)^{4-3} \\ &= (0.67)^3 + 3(0.67)^3(0.33) = 0.599. \end{aligned}$$

5.117 Let X denote the number of at-bats up to and including the second hit. Then $X \sim \mathcal{NB}(2, 0.26)$, so that

$$p_X(x) = \binom{x-1}{1}(0.26)^2(0.74)^{x-2} = (x-1)(0.26)^2(0.74)^{x-2}, \qquad x = 2, 3, \dots,$$

and $p_X(x) = 0$ otherwise.
a) We have

$$P(X = 5) = p_X(5) = (5-1)(0.26)^2(0.74)^{5-2} = 0.110.$$

b) From the complementation rule and the FPF,

$$\begin{aligned} P(X > 5) &= 1 - P(X \le 5) = 1 - \sum_{x \le 5} p_X(x) = 1 - \sum_{k=2}^{5} p_X(k) \\ &= 1 - \sum_{k=2}^{5}(k-1)(0.26)^2(0.74)^{k-2} = 1 - (0.26)^2 \sum_{k=2}^{5}(k-1)(0.74)^{k-2} \\ &= 0.612. \end{aligned}$$

Alternatively, because $X > 5$ if and only if the number of hits in the first five at-bats is at most 1, we have

$$P(X > 5) = \sum_{k=0}^{1}\binom{5}{k}(0.26)^k(0.74)^{5-k} = (0.74)^5 + 5(0.26)(0.74)^4 = 0.612.$$

c) From the FPF,

$$P(3 \le X \le 10) = \sum_{3 \le x \le 10} p_X(x) = \sum_{k=3}^{10} p_X(k) = (0.26)^2 \sum_{k=3}^{10}(k-1)(0.74)^{k-2} = 0.710.$$

5.118 For each $r \in \mathcal{N}$, we know that X is a positive-integer valued random variable. From Proposition 5.11 on page 234, the only positive-integer valued random variables that have the lack-of-memory property are geometric random variables. A negative binomial random variable is a geometric random variable if and only if $r = 1$. Hence, X has the lack-of-memory property if and only if $r = 1$.

5.119 Let E denote the event of a 7 or 11 when two balanced dice are rolled. Then

$$P(E) = \frac{6+2}{36} = \frac{2}{9}.$$

Let X denote the number of rolls required for the fourth occurrence of E. Then $X \sim \mathcal{NB}(4, 2/9)$. Hence, by the complementation rule and the FPF,

$$\begin{aligned} P(X > 6) &= 1 - P(X \le 6) = 1 - \sum_{x \le 6} p_X(x) = 1 - \sum_{k=4}^{6} p_X(k) \\ &= 1 - \sum_{k=4}^{6}\binom{k-1}{4-1}(2/9)^4(7/9)^{k-4} = 1 - (2/9)^4 \sum_{k=4}^{6}\binom{k-1}{3}(7/9)^{k-4} \\ &= 0.975. \end{aligned}$$

5.120

a) Consider Bernoulli trials with success probability p. Let X denote the number of trials until the rth success and let Y denote the number of successes in the first n trials. Then $X \sim \mathcal{NB}(r, p)$ and $Y \sim \mathcal{B}(n, p)$. Event $\{X > n\}$ occurs if and only if the rth success occurs after the nth trial, which happens if and only if the number of successes in the first n trials is less than r, that is, event $\{Y < r\}$ occurs. Therefore, $\{X > n\} = \{Y < r\}$ and, in particular, $P(X > n) = P(Y < r)$.

b) We have

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \dots$$

and

$$p_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0, 1, \dots, n.$$

Applying part (a) and the FPF, we get

$$\sum_{x=n+1}^{\infty} \binom{x-1}{r-1} p^r (1-p)^{x-r} = \sum_{y=0}^{r-1} \binom{n}{y} p^y (1-p)^{n-y}.$$

c) From the complementation rule and the FPF,

$$P(X > n) = 1 - P(X \le n) = 1 - \sum_{x \le n} p_X(x) = 1 - \sum_{k=r}^{n} p_X(k).$$

Hence, with this computation, we must evaluate $n - r + 1$ terms of the PMF of X.

d) From the FPF,

$$P(Y < r) = \sum_{y<r} p_Y(y) = \sum_{k=0}^{r-1} p_Y(k).$$

Hence, with this computation, we must evaluate r terms of the PMF of Y.

e) From parts (c) and (d), in computing $P(X > n)$, the difference between the numbers of terms that must be evaluated using the PMF of X and the PMF of Y is

$$(n - r + 1) - r = n - 2r + 1 = n\left(1 - \frac{2r-1}{n}\right).$$

Thus, when n is large relative to r, it is necessary to evaluate roughly n more terms using the PMF of X than using the PMF of Y. In other words, when n is large relative to r, there is considerable computational savings in using the PMF of Y rather than the PMF of X to evaluate $P(X > n)$.

5.121 For $x = r, r+1, \dots,$

$$\binom{x-1}{r-1} = \binom{x-1}{x-r} = \binom{r + (x-r) - 1}{x-r} = (-1)^{x-r}\binom{-r}{x-r},$$

where the last equality follows from Exercise 5.60(b). Hence, for $x = r, r+1, \dots,$

$$\binom{x-1}{r-1} p^r (1-p)^{x-r} = (-1)^{x-r}\binom{-r}{x-r} p^r (1-p)^{x-r} = \binom{-r}{x-r} p^r (p-1)^{x-r}.$$

5.122 Let X and Z denote the calls on which the second and fifth sales are made, respectively. Also, let $Y \sim \mathcal{NB}(3, 0.2)$. Then

$$P(X = j) = \binom{j-1}{2-1}(0.2)^2(0.8)^{j-2} = (j-1)(0.2)^2(0.8)^{j-2}, \qquad j = 2, 3, \ldots,$$

and, arguing as in Example 5.26(c), we find that

$$\begin{aligned} P(Z = k \mid X = j) = P(Y = k - j) &= \binom{k-j-1}{3-1}(0.2)^3(0.8)^{k-j-3} \\ &= \binom{k-j-1}{2}(0.2)^3(0.8)^{k-j-3}, \qquad k = j+3, j+4, \ldots; \ j = 2, 3, \ldots. \end{aligned}$$

From the FPF and the preceding display, it follows immediately that

$$P(Z \le 15 \mid X = j) = \sum_{k=j+3}^{15} \binom{k-j-1}{2}(0.2)^3(0.8)^{k-j-3}, \qquad j = 2, 3, \ldots, 12.$$

a) Applying the law of total probability, we now get that

$$\begin{aligned} P(X \le 5,\ Z = 15) &= \sum_{j=2}^{5} P(X = j)P(Z = 15 \mid X = j) \\ &= \sum_{j=2}^{5} (j-1)(0.2)^2(0.8)^{j-2}\binom{15-j-1}{2}(0.2)^3(0.8)^{15-j-3} \\ &= (0.2)^5(0.8)^{10}\sum_{j=2}^{5}(j-1)\binom{14-j}{2} = 0.0156. \end{aligned}$$

b) Again applying the law of total probability, we get

$$\begin{aligned} P(X \le 5,\ Z \le 15) &= \sum_{j=2}^{5} P(X = j)P(Z \le 15 \mid X = j) \\ &= \sum_{j=2}^{5}(j-1)(0.2)^2(0.8)^{j-2}\left(\sum_{k=j+3}^{15}\binom{k-j-1}{2}(0.2)^3(0.8)^{k-j-3}\right) \\ &= (0.2)^5\sum_{j=2}^{5}(j-1)\left(\sum_{k=j+3}^{15}\binom{k-j-1}{2}(0.8)^{k-5}\right) = 0.104. \end{aligned}$$

Theory Exercises

5.123 We note that $I_E(\omega) = 1$ if and only if $\omega \in E$. Hence,

$$p_{I_E}(1) = P(I_E = 1) = P\big(\{\, \omega \in \Omega : I_E(\omega) = 1 \,\}\big) = P(\{\, \omega \in \Omega : \omega \in E) = P(E).$$

Moreover, because the only two possible values of I_E are 0 and 1, we have, from the complementation rule, that

$$p_{I_E}(0) = P(I_E = 0) = 1 - P(I_E = 1) = 1 - p_{I_E}(1) = 1 - P(E).$$

Hence,

$$p_{I_E}(x) = \begin{cases} 1 - P(E), & \text{if } x = 0; \\ P(E), & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases}$$

5.124 The possible values of X are $r, r+1, \ldots$. For such a positive integer, the event $\{X = k\}$ occurs if and only if the rth success is obtained on the kth trial. That happens if and only if (1) among the first $k-1$ trials, exactly $r-1$ are successes, and (2) the kth trial is a success. Denote the former event by A and the latter event by B. We have $\{X = k\} = A \cap B$ and, because the trials are independent, event A and event B are independent. Hence,

$$P(X = k) = P(A \cap B) = P(A)P(B).$$

The number of successes in the first $k-1$ trials has the binomial distribution with parameters $k-1$ and p. Thus, by Proposition 5.3 on page 201,

$$P(A) = \binom{k-1}{r-1} p^{r-1}(1-p)^{(k-1)-(r-1)} = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}.$$

Also, the probability that the kth trial is a success is p, so $P(B) = p$. Therefore,

$$P(X = k) = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \cdot p = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

Consequently, the PMF of the random variable X is

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \ldots,$$

and $p_X(x) = 0$ otherwise.

5.125 We must verify that p_X satisfies the three properties stated in Proposition 5.1 on page 187. Clearly, we have $p_X(x) \geq 0$ for all x. Moreover, $\{\, x \in \mathcal{R} : p_X(x) \neq 0 \,\} = \{r, r+1, \ldots\}$, which is countable, being a subset of the countable set $\mathcal{N}$. Applying Exercise 5.121 and the binomial series [Equation (5.45) on page 241], we get

$$\begin{aligned} \sum_x p_X(x) &= \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = \sum_{k=r}^{\infty} \binom{-r}{k-r} p^r (p-1)^{k-r} \\ &= p^r \sum_{j=0}^{\infty} \binom{-r}{j} (p-1)^j = p^r \big(1 + (p-1)\big)^{-r} = p^r p^{-r} = 1. \end{aligned}$$

5.126

a) In Bernoulli trials, the trials are independent and, furthermore, the success probability is positive. Therefore, from Proposition 4.6, in repeated Bernoulli trials, the probability is 1 that a success will eventually occur. Because of independence in Bernoulli trials, once a success occurs, the subsequent process of Bernoulli trials is independent of the prior trials and behaves the same probabilistically as the process of Bernoulli trials starting from the beginning. Hence, we conclude that the probability is 1 that a second success will eventually occur. Continuing in this way, we deduce that, for each positive integer r, the probability is 1 that an rth success will eventually occur.

b) If we let X denote the number of trials up to and including the rth success, then $X \sim \mathcal{NB}(r, p)$. The probability that an rth success will eventually occur is

$$P(X < \infty) = \sum_{x<\infty} p_X(x) = \sum_x p_X(x) = 1,$$

where the last equality is a consequence of Exercise 5.125.

5.127 We must verify that the function defined in Equation (5.50) satisfies the three properties stated in Proposition 5.1 on page 187. We note that $p(0) = \binom{-\alpha}{0} p^\alpha (p-1)^0 = p^\alpha > 0$. Also, from Definition 5.5 on page 211, we have, for each $k \in \mathcal{N}$,

$$\begin{aligned}\binom{-\alpha}{k} p^\alpha (p-1)^k &= \frac{(-\alpha)(-a-1)\cdots(-\alpha-k+1)}{k!} p^\alpha (-1)^k (1-p)^k \\ &= \frac{\alpha(a+1)\cdots(\alpha+k-1)}{k!} p^\alpha (1-p)^k > 0.\end{aligned}$$

Hence, $p(y) \geq 0$ for all y. Moreover, $\{\, y \in \mathcal{R} : p(y) \neq 0 \,\} = \{0, 1, \ldots\}$, which is countable, being the union of the countable sets $\{0\}$ and $\mathcal{N}$. Applying now the binomial series, Equation (5.45) on page 241, we get

$$\sum_y p(y) = \sum_{k=0}^{\infty} \binom{-\alpha}{k} p^\alpha (p-1)^k = p^\alpha \sum_{k=0}^{\infty} \binom{-\alpha}{k} (p-1)^k = p^\alpha \big(1 + (p-1)\big)^{-\alpha} = 1.$$

Advanced Exercises

5.128

a) Let k be an integer between 0 and N, inclusive. Let R_k denote the event that the smoker finds the matchbox in his right pocket empty at the moment when there are k matches in the matchbox in his left pocket; and let L_k denote the event that the smoker finds the matchbox in his left pocket empty at the moment when there are k matches in the matchbox in his right pocket. Events R_k and L_k are mutually exclusive, have the same probability (by symmetry), and have union $\{Y = k\}$. To find $P(R_k)$, we consider a success to be that the smoker goes to his right pocket for a match. Event R_k occurs if and only if the $(N+1)$st success occurs on trial $(N+1) + (N-k) = 2N+1-k$. Applying the negative binomial PMF with parameters $N+1$ and 1/2, we see that the probability of that happening is

$$P(R_k) = \binom{(2N+1-k)-1}{(N+1)-1} \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{(2N+1-k)-(N+1)} = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N+1-k}.$$

Therefore,

$$P(Y = k) = P(R_k) + P(L_k) = 2P(R_k) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k}.$$

Consequently,

$$p_Y(y) = \binom{2N-y}{N} \left(\frac{1}{2}\right)^{2N-y}, \qquad y = 0, 1, \ldots, N,$$

and $p_Y(y) = 0$ otherwise.

b) Referring to part (a), we obtain the following table for $N = 5$:

y	0	1	2	3	4	5
$p_Y(y)$	0.246094	0.246094	0.218750	0.164063	0.093750	0.031250

And, for $N = 10$, we have

y	$p_Y(y)$	y	$p_Y(y)$
0	0.176197	6	0.061096
1	0.176197	7	0.034912
2	0.166924	8	0.016113
3	0.148376	9	0.005371
4	0.122192	10	0.000977
5	0.091644		

c) We proceed as in part (a) until we apply the negative binomial PDF. At that point, we use the one with parameters $N + 1$ and p. Hence, we find that

$$P(R_k) = \binom{2N-k}{N} p^{N+1}(1-p)^{N-k} \quad \text{and} \quad P(L_k) = \binom{2N-k}{N}(1-p)^{N+1}p^{N-k}.$$

Therefore,

$$P(Y = k) = P(R_k) + P(L_k) = \binom{2N-k}{N}\left(p^{N+1}(1-p)^{N-k} + (1-p)^{N+1}p^{N-k}\right).$$

Consequently,

$$p_Y(y) = \binom{2N-y}{N}\left(p^{N+1}(1-p)^{N-y} + (1-p)^{N+1}p^{N-y}\right), \qquad y = 0, 1, \ldots N,$$

and $p_Y(y) = 0$ otherwise.

d) Referring to part (c), we obtain the following table for $N = 5$:

y	0	1	2	3	4	5
$p_Y(y)$	0.200658	0.217380	0.216760	0.187730	0.126720	0.050752

And, for $N = 10$, we have

y	$p_Y(y)$	y	$p_Y(y)$
0	0.117142	6	0.098410
1	0.126903	7	0.068997
2	0.134867	8	0.039308
3	0.138435	9	0.016240
4	0.134671	10	0.003670
5	0.121357		

e) The answers here are the same as those in part (d) because, if we replace p by $1 - p$, the probabilities for Y do not change. We can see this fact by interchanging the roles of the left and right pockets or, more directly, by replacing p by $1 - p$ (and, therefore, $1 - p$ by p) in the formula for the PMF of Y determined in part (c).

5.129

a) Consider a success a win of one point by Jane on any particular trial and let X denote the number of trials until the 21st success. Then $X \sim \mathcal{NB}(21, p)$. Jane wins if and only if she gets 21 points before Joan does, that is, if and only if the 21st success occurs on or before trial 41. Hence, by the FPF, the required probability is

$$P(X \leq 41) = \sum_{x \leq 41} p_X(x) = \sum_{k=21}^{41} p_X(k) = \sum_{k=21}^{41} \binom{k-1}{20} p^{21}(1-p)^{k-21}.$$

b) Suppose that $p = 1/2$, that is, that Jane and Joan are evenly matched. Then, by symmetry, the probability is $1/2$ that Jane wins the game. Hence, in view of part (a), we have

$$\frac{1}{2} = \sum_{k=21}^{41} \binom{k-1}{20} \left(\frac{1}{2}\right)^{21} \left(1-\frac{1}{2}\right)^{k-21} = \sum_{k=21}^{41} \binom{k-1}{20} 2^{-k}.$$

5.130

a) For each nonnegative integer y,

$$\begin{aligned}
p_{Y_r}(y) &= \binom{-r}{y} p_r^r (p_r - 1)^y = \frac{(-r)(-r-1)\cdots(-r-y+1)}{y!} \left(1 - \frac{r(1-p_r)}{r}\right)^r \left(\frac{r(p_r-1)}{r}\right)^y \\
&= \frac{r(r+1)\cdots(r+y-1)}{r^y} \left(1 - \frac{r(1-p_r)}{r}\right)^r \frac{\big(r(1-p_r)\big)^y}{y!} \\
&= \left(1+\frac{1}{r}\right)\left(1+\frac{2}{r}\right)\cdots\left(1+\frac{y-1}{r}\right) \cdot \left(1 - \frac{r(1-p_r)}{r}\right)^r \cdot \frac{\big(r(1-p_r)\big)^y}{y!}.
\end{aligned}$$

Now,

$$\lim_{r\to\infty} \left(1+\frac{1}{r}\right)\left(1+\frac{2}{r}\right)\cdots\left(1+\frac{y-1}{r}\right) = 1$$

and, because $r(1-p_r) \to \lambda$ as $r \to \infty$,

$$\lim_{r\to\infty} \frac{\big(r(1-p_r)\big)^y}{y!} = \frac{\lambda^y}{y!}.$$

Moreover, from the hint given in Exercise 5.86(a),

$$\lim_{r\to\infty} \left(1 - \frac{r(1-p_r)}{r}\right)^r = e^{-\lambda}.$$

Hence,

$$\lim_{r\to\infty} p_{Y_r}(y) = 1 \cdot e^{-\lambda} \cdot \frac{\lambda^y}{y!} = e^{-\lambda}\frac{\lambda^y}{y!}.$$

b) Because $r(1-p_r) \to \lambda$ as $r \to \infty$, we have

$$\lim_{r\to\infty} p_r = \lim_{r\to\infty}\left(1 - \frac{r(1-p_r)}{r}\right) = 1 - \lim_{r\to\infty}\frac{r(1-p_r)}{r} = 1 - \frac{\lambda}{\lim_{r\to\infty} r} = 1 - 0 = 1.$$

c) Suppose that $X \sim \mathcal{NB}(r, p)$, where r is large and p is close to 1. Let $Y = X - r$. Then the PDF of Y is given by Equation (5.49). Now,

$$p_X(x) = P(X = x) = P(Y + r = x) = P(Y = x - r) = p_Y(x - r).$$

Hence, in view of parts (a) and (b),

$$p_X(x) \approx e^{-r(1-p)} \frac{\big(r(1-p)\big)^{x-r}}{(x-r)!}, \qquad x = r, r+1, \ldots.$$

In other words, for large r and p close to 1, the PMF of a $\mathcal{NB}(r, p)$ random variable can be approximated by the right side of the previous display.

5.8 Functions of a Discrete Random Variable

Basic Exercises

5.131

a) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are -6, -3, 0, 3, 6, and 9. Here $g(x) = 3x$. We note that g is one-to-one and, solving the equation $y = 3x$ for x, we find that $g^{-1}(y) = y/3$. Hence, from Equation (5.54) on page 248, we have

$$p_Y(y) = p_{3X}(y) = p_X(y/3).$$

Then, for instance,

$$p_Y(6) = p_X(6/3) = p_X(2) = 0.15.$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	-6	-3	0	3	6	9
$p_Y(y)$	0.10	0.20	0.20	0.15	0.15	0.20

b) The transformation results in a change of scale. That is, if the units in which X is measured are changed by a multiple of b, then bX will have the same relative value (in new units) as X does in old units.
c) Yes, Corollary 5.1 applies to a change of scale (i.e., a function of the form $g(x) = bx$) because such a function is one-to-one on all of $\mathcal{R}$.

5.132

a) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are -4, -3, -2, -1, 0, and 1. Here $g(x) = x - 2$. We note that g is one-to-one and, solving the equation $y = x - 2$ for x, we find that $g^{-1}(y) = y + 2$. Hence, from Equation (5.54) on page 248, we have

$$p_Y(y) = p_{X-2}(y) = p_X(y + 2).$$

Then, for instance,

$$p_Y(-1) = p_X(-1 + 2) = p_X(1) = 0.15.$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	-4	-3	-2	-1	0	1
$p_Y(y)$	0.10	0.20	0.20	0.15	0.15	0.20

b) The transformation results in a change of location. That is, if the units in which X is measured are changed by an addition of a, then $a + X$ will have the same relative value (in the new units) as X does in the old units.
c) Yes, Corollary 5.1 applies to a change of location $\big($i.e., a function of the form $g(x) = a + x\big)$ because such a function is one-to-one on all of $\mathcal{R}$.

5.133
a) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are -8, -5, -2, 1, 4, and 7. Here $g(x) = 3x - 2$. We note that g is one-to-one and, solving the equation $y = 3x - 2$ for x, we find that $g^{-1}(y) = (y + 2)/3$. Hence, from Equation (5.54) on page 248, we have

$$p_Y(y) = p_{3X-2}(y) = p_X\big((y + 2)/3\big).$$

Then, for instance,

$$p_Y(4) = p_X\big((4 + 2)/3\big) = p_X(2) = 0.15.$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	-8	-5	-2	1	4	7
$p_Y(y)$	0.10	0.20	0.20	0.15	0.15	0.20

b) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are $a - 2b$, $a - b$, a, $a + b$, $a + 2b$, and $a + 3b$. Here $g(x) = a + bx$. We note that g is one-to-one and, solving the equation $y = a + bx$ for x, we find that $g^{-1}(y) = (y - a)/b$. Hence, from Equation (5.54) on page 248, we have

$$p_Y(y) = p_{a+bX}(y) = p_X\big((y - a)/b\big).$$

Then, for instance,

$$p_Y(a + 2b) = p_X\Big(\big((a + 2b) - a\big)/b\Big) = p_X(2) = 0.15.$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	$a - 2b$	$a - b$	a	$a + b$	$a + 2b$	$a + 3b$
$p_Y(y)$	0.10	0.20	0.20	0.15	0.15	0.20

c) Yes, Corollary 5.1 applies to a simultaneous location and scale change $\big($i.e., a function of the form $g(x) = a + bx\big)$ because such a function is one-to-one on all of $\mathcal{R}$.

5.134
a) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are -33, -5, -1, 3, 31, and 107. Here $g(x) = 4x^3 - 1$. We note that g is one-to-one and, solving the equation $y = 4x^3 - 1$ for x, we find that $g^{-1}(y) = \big((y + 1)/4\big)^{1/3}$. Hence, from Equation (5.54) on page 248, we have

$$p_Y(y) = p_{4X^3-1}(y) = p_X\Big(\big((y + 1)/4\big)^{1/3}\Big).$$

Then, for instance,

$$p_Y(31) = p_X\Big(\big((31 + 1)/4\big)^{1/3}\Big) = p_X(2) = 0.15.$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	-33	-5	-1	3	31	107
$p_Y(y)$	0.10	0.20	0.20	0.15	0.15	0.20

b) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are 0, 1, 16, and 81. Here $g(x) = x^4$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14 on page 247. For instance,

$$\begin{aligned} p_Y(16) &= \sum_{x \in g^{-1}(\{16\})} p_X(x) = \sum_{\{x : x^4 = 16\}} p_X(x) = \sum_{x \in \{-2,2\}} p_X(x) \\ &= p_X(-2) + p_X(2) = 0.10 + 0.15 = 0.25. \end{aligned}$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	0	1	16	81
$p_Y(y)$	0.20	0.35	0.25	0.20

c) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are 0, 1, 2, and 3. Here $g(x) = |x|$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14. For instance,

$$\begin{aligned} p_Y(2) &= \sum_{x \in g^{-1}(\{2\})} p_X(x) = \sum_{\{x : |x| = 2\}} p_X(x) = \sum_{x \in \{-2,2\}} p_X(x) \\ &= p_X(-2) + p_X(2) = 0.10 + 0.15 = 0.25. \end{aligned}$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	0	1	2	3
$p_Y(y)$	0.20	0.35	0.25	0.20

d) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are 4, 6, 8, and 10. Here $g(x) = 2|x| + 4$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14. For instance,

$$\begin{aligned} p_Y(8) &= \sum_{x \in g^{-1}(\{8\})} p_X(x) = \sum_{\{x : 2|x|+4 = 8\}} p_X(x) = \sum_{x \in \{-2,2\}} p_X(x) \\ &= p_X(-2) + p_X(2) = 0.10 + 0.15 = 0.25. \end{aligned}$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	4	6	8	10
$p_Y(y)$	0.20	0.35	0.25	0.20

e) As the possible values of X are -2, -1, 0, 1, 2, and 3, the possible values of Y are $\sqrt{5}$, $\sqrt{8}$, $\sqrt{17}$, and $\sqrt{32}$. Here $g(x) = \sqrt{3x^2 + 5}$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14. For instance,

$$p_Y\big(\sqrt{17}\big) = \sum_{x \in g^{-1}(\{\sqrt{17}\})} p_X(x) = \sum_{\{x:\sqrt{3x^2+5}=\sqrt{17}\}} p_X(x) = \sum_{x\in\{-2,2\}} p_X(x)$$
$$= p_X(-2) + p_X(2) = 0.10 + 0.15 = 0.25.$$

Proceeding similarly, we obtain the following table for the PMF of Y.

y	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{17}$	$\sqrt{32}$
$p_Y(y)$	0.20	0.35	0.25	0.20

5.135
a) Because the range of X is $\mathcal{N}$, the range of Y is $R_Y = \{n/(n+1) : n \in \mathcal{N}\}$.
b) Here $g(x) = x/(x+1)$, which is one-to-one on the range of X. Solving the equation $y = x/(x+1)$ for x, we find that $g^{-1}(y) = y/(1-y)$. Hence, from Equation (5.54) on page 248, we have

$$p_Y(y) = p_{X/(X+1)}(y) = p_X\big(y/(1-y)\big)$$

if $y \in R_Y$, and $p_Y(y) = 0$ otherwise.
c) Because the range of X is $\mathcal{N}$, the range of Z is $R_Z = \{(n+1)/n : n \in \mathcal{N}\}$. Here $g(x) = (x+1)/x$, which is one-to-one on the range of X. Solving the equation $z = (x+1)/x$ for x, we determine that $g^{-1}(z) = 1/(z-1)$. Hence, from Equation (5.54), we have

$$p_Z(z) = p_{(X+1)/X}(z) = p_X\big(1/(z-1)\big)$$

if $z \in R_Z$, and $p_Z(z) = 0$ otherwise.

5.136 We know that $p_X(x) = p(1-p)^{x-1}$ if $x \in \mathcal{N}$, and $p_X(x) = 0$ otherwise.
a) Let $Y = \min\{X, m\}$. We note that, because the range of X is $\mathcal{N}$, the range of Y is $\{1, 2, \ldots, m\}$. Here $g(x) = \min\{x, m\}$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14 on page 247. For $y = 1, 2, \ldots, m-1$,

$$p_Y(y) = \sum_{x\in g^{-1}(\{y\})} p_X(x) = \sum_{\{x:\min\{x,m\}=y\}} p_X(x) = \sum_{x\in\{y\}} p_X(x) = p_X(y) = p(1-p)^{y-1}.$$

Also,

$$p_Y(m) = \sum_{x\in g^{-1}(\{m\})} p_X(x) = \sum_{\{x:\min\{x,m\}=m\}} p_X(x) = \sum_{x\ge m} p_X(x)$$
$$= \sum_{k=m}^{\infty} p(1-p)^{k-1} = (1-p)^{m-1}.$$

Hence,

$$p_Y(y) = \begin{cases} p(1-p)^{y-1}, & \text{if } y = 1, 2, \ldots, m-1; \\ (1-p)^{m-1}, & \text{if } y = m; \\ 0, & \text{otherwise.} \end{cases}$$

b) Let $Z = \max\{X, m\}$. We note that, because the range of X is $\mathcal{N}$, the range of Z is $\{m, m+1, \ldots\}$. Here $g(x) = \max\{x, m\}$, which is not one-to-one on the range of X. Hence, to find the PDF of Z, we apply Proposition 5.14. For $z = m+1, m+2, \ldots,$

$$p_Z(z) = \sum_{x \in g^{-1}(\{z\})} p_X(x) = \sum_{\{x : \max\{x,m\}=z\}} p_X(x) = \sum_{x \in \{z\}} p_X(x) = p_X(z) = p(1-p)^{z-1}.$$

Also,

$$\begin{aligned} p_Z(m) &= \sum_{x \in g^{-1}(\{m\})} p_X(x) = \sum_{\{x : \max\{x,m\}=m\}} p_X(x) = \sum_{x \le m} p_X(x) \\ &= \sum_{k=1}^{m} p(1-p)^{k-1} = 1 - (1-p)^m. \end{aligned}$$

Hence,

$$p_Z(z) = \begin{cases} 1-(1-p)^m, & \text{if } z = m; \\ p(1-p)^{z-1}, & \text{if } z = m+1, m+2, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

5.137 We know that $p_X(x) = e^{-3}\, 3^x/x!$ if $x = 0, 1, \ldots,$ and $p_X(x) = 0$ otherwise. Because the range of X is $\{0, 1, \ldots\}$, the range of Y is also $\{0, 1, \ldots\}$. Here $g(x) = |x-3|$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14 on page 247. For $y = 1, 2,$ and 3,

$$\begin{aligned} p_Y(y) &= \sum_{x \in g^{-1}(\{y\})} p_X(x) = \sum_{\{x : |x-3|=y\}} p_X(x) = \sum_{x \in \{3-y, 3+y\}} p_X(x) \\ &= p_X(3-y) + p_X(3+y) = e^{-3}\,\frac{3^{3-y}}{(3-y)!} + e^{-3}\,\frac{3^{3+y}}{(3+y)!} \\ &= 27e^{-3}\left(\frac{3^{-y}}{(3-y)!} + \frac{3^y}{(3+y)!}\right). \end{aligned}$$

Also, for $y = 0, 4, 5, \ldots,$

$$\begin{aligned} p_Y(y) &= \sum_{x \in g^{-1}(\{y\})} p_X(x) = \sum_{\{x : |x-3|=y\}} p_X(x) = \sum_{x \in \{3+y\}} p_X(x) \\ &= p_X(3+y) = e^{-3}\,\frac{3^{3+y}}{(3+y)!} = 27e^{-3}\,\frac{3^y}{(3+y)!}. \end{aligned}$$

Hence,

$$p_Y(y) = \begin{cases} 27e^{-3}\,\dfrac{3^y}{(3+y)!}, & \text{if } y = 0, 4, 5, \ldots; \\[2ex] 27e^{-3}\left(\dfrac{3^{-y}}{(3-y)!} + \dfrac{3^y}{(3+y)!}\right), & \text{if } y = 1, 2, 3; \\[2ex] 0, & \text{otherwise.} \end{cases}$$

5.138 We know that $p_X(x) = 1/9$ if $x = 0, \pi/4, 2\pi/4, \dots, 2\pi$, and $p_X(x) = 0$ otherwise.
a) Let $Y = \sin X$. Here $g(x) = \sin x$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14 on page 247. To that end, we first construct the following table:

x	0	$\pi/4$	$2\pi/4$	$3\pi/4$	$4\pi/4$	$5\pi/4$	$6\pi/4$	$7\pi/4$	$8\pi/4$
$\sin x$	0	$1/\sqrt{2}$	1	$1/\sqrt{2}$	0	$-1/\sqrt{2}$	-1	$-1/\sqrt{2}$	0

From the table, we see that the PDF of Y is as follows:

y	-1	$-1/\sqrt{2}$	0	$1/\sqrt{2}$	1
$p_Y(y)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

b) Let $Z = \cos X$. Here $g(x) = \cos x$, which is not one-to-one on the range of X. Hence, to find the PDF of Z, we apply Proposition 5.14. To that end, we first construct the following table:

x	0	$\pi/4$	$2\pi/4$	$3\pi/4$	$4\pi/4$	$5\pi/4$	$6\pi/4$	$7\pi/4$	$8\pi/4$
$\cos x$	1	$1/\sqrt{2}$	0	$-1/\sqrt{2}$	-1	$-1/\sqrt{2}$	0	$1/\sqrt{2}$	1

From the table, we see that the PDF of Z is as follows:

z	-1	$-1/\sqrt{2}$	0	$1/\sqrt{2}$	1
$p_Z(z)$	$1/9$	$2/9$	$2/9$	$2/9$	$2/9$

c) No, $\tan X$ is not a random variable, because it is not a real-valued function on the sample space. For instance, we know that $P(X = \pi/2) = 1/9 > 0$, which implies that $\{X = \pi/2\} \neq \emptyset$. Now, let ω be such that $X(\omega) = \pi/2$. Then $\tan X(\omega) = \tan(\pi/2) = \infty$.

5.139 We know that $X \sim \mathcal{NB}(3, p)$ and, hence, that

$$p_X(x) = \binom{x-1}{2} p^3 (1-p)^{x-3}, \qquad x = 3, 4, \dots,$$

and $p_X(x) = 0$ otherwise.
a) As X denotes the number of trials until the third success, the number of failures by the third success is $X - 3$. Thus, the proportion of failures by the third success is $(X-3)/X = 1 - 3/X$.
b) Because the range of X is $\{3, 4, \dots\}$, the range of Y is $R_Y = \{\, 1 - 3/n : n = 3, 4, \dots \,\}$. Here we have $g(x) = 1 - 3/x$, which is one-to-one on the range of X. Solving the equation $y = 1 - 3/x$ for x, we find that $g^{-1}(y) = 3/(1-y)$. Hence, from Equation (5.54) on page 248,

$$\begin{aligned} p_Y(y) = p_{1-3/X}(y) = p_X\big(3/(1-y)\big) &= \binom{3/(1-y) - 1}{2} p^3 (1-p)^{3/(1-y)-3} \\ &= \binom{(2+y)/(1-y)}{2} p^3 (1-p)^{3y/(1-y)} \end{aligned}$$

if $y \in R_Y$, and $p_Y(y) = 0$ otherwise.

5.140 Let X denote the number of trials until the fifth success and let $Y = \min\{X, 17\}$. We want to determine the PDF of Y. Now, we know that $X \sim \mathcal{NB}(5, p)$. Hence,

$$p_X(x) = \binom{x-1}{4} p^5 (1-p)^{x-5}, \qquad x = 5, 6, \ldots,$$

and $p_X(x) = 0$ otherwise. Here $g(x) = \min\{x, 17\}$, which is not one-to-one on the range of X. Hence, to find the PDF of Y, we apply Proposition 5.14 on page 247. For $y = 5, 6, \ldots, 16$,

$$\begin{aligned} p_Y(y) &= \sum_{x \in g^{-1}(\{y\})} p_X(x) = \sum_{\{x : \min\{x,17\} = y\}} p_X(x) = \sum_{x \in \{y\}} p_X(x) \\ &= p_X(y) = \binom{y-1}{4} p^5 (1-p)^{y-5}. \end{aligned}$$

Also,

$$\begin{aligned} p_Y(17) &= \sum_{x \in g^{-1}(\{17\})} p_X(x) = \sum_{\{x : \min\{x,17\} = 17\}} p_X(x) = \sum_{x \geq 17} p_X(x) \\ &= \sum_{k=17}^{\infty} p_X(k) = 1 - \sum_{k=5}^{16} p_X(k) = 1 - \sum_{k=5}^{16} \binom{k-1}{4} p^5 (1-p)^{k-5}. \end{aligned}$$

Hence,

$$p_Y(y) = \begin{cases} \binom{y-1}{4} p^5 (1-p)^{y-5}, & \text{if } y = 5, 6, \ldots, 16; \\ 1 - \sum_{k=5}^{16} \binom{k-1}{4} p^5 (1-p)^{k-5}, & \text{if } y = 17; \\ 0, & \text{otherwise.} \end{cases}$$

Advanced Exercises

5.141

a) The range of Y is $\{0, 1, 2\}$. Let $g(x) = x \pmod 3$. Applying Proposition 5.14 on page 247, we get

$$\begin{aligned} p_Y(0) &= \sum_{g(x)=0} p_X(x) = \sum_{x \in \{3,6,9,\ldots\}} p_X(x) = \sum_{k=1}^{\infty} p_X(3k) = \sum_{k=1}^{\infty} p(1-p)^{3k-1} \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} \big((1-p)^3\big)^k = \frac{p(1-p)^2}{1-(1-p)^3} = \frac{(1-p)^2}{3-3p+p^2} \end{aligned}$$

and

$$\begin{aligned} p_Y(1) &= \sum_{g(x)=1} p_X(x) = \sum_{x \in \{1,4,7,\ldots\}} p_X(x) = \sum_{k=0}^{\infty} p_X(3k+1) = \sum_{k=0}^{\infty} p(1-p)^{(3k+1)-1} \\ &= p \sum_{k=0}^{\infty} \big((1-p)^3\big)^k = \frac{p}{1-(1-p)^3} = \frac{1}{3-3p+p^2} \end{aligned}$$

and

$$p_Y(2) = \sum_{g(x)=2} p_X(x) = \sum_{x\in\{2,5,8,\ldots\}} p_X(x) = \sum_{k=0}^{\infty} p_X(3k+2) = \sum_{k=0}^{\infty} p(1-p)^{(3k+2)-1}$$

$$= p(1-p)\sum_{k=0}^{\infty}\big((1-p)^3\big)^k = \frac{p(1-p)}{1-(1-p)^3} = \frac{1-p}{3-3p+p^2}.$$

Hence,

$$p_Y(y) = \begin{cases} (1-p)^2/(3-3p+p^2), & \text{if } y=0; \\ 1/(3-3p+p^2), & \text{if } y=1; \\ (1-p)/(3-3p+p^2), & \text{if } y=2; \\ 0, & \text{otherwise.} \end{cases}$$

b) The range of Y is $\{0, 1, 2\}$. Let $g(x) = x \pmod 3$. We first note that, for $n \in \mathcal{N}$,

$$p_X(n+1) = P(X=n+1) = p(1-p)^{(n+1)-1} = (1-p)p(1-p)^{n-1} = (1-p)p_X(n)$$

and

$$p_X(n+2) = P(X=n+2) = p(1-p)^{(n+2)-1} = (1-p)^2 p(1-p)^{n-1} = (1-p)^2 p_X(n).$$

Applying Proposition 5.14, we get

$$p_Y(0) = \sum_{g(x)=0} p_X(x) = \sum_{x\in\{3,6,9,\ldots\}} p_X(x) = \sum_{k=1}^{\infty} p_X(3k).$$

Also,

$$p_Y(1) = \sum_{g(x)=1} p_X(x) = \sum_{x\in\{1,4,7,\ldots\}} p_X(x) = \sum_{k=0}^{\infty} p_X(3k+1) = p_X(1) + \sum_{k=1}^{\infty} p_X(3k+1)$$

$$= p + \sum_{k=1}^{\infty}(1-p)p_X(3k) = p + (1-p)p_Y(0)$$

and

$$p_Y(2) = \sum_{g(x)=2} p_X(x) = \sum_{x\in\{2,5,8,\ldots\}} p_X(x) = \sum_{k=0}^{\infty} p_X(3k+2) = p_X(2) + \sum_{k=1}^{\infty} p_X(3k+2)$$

$$= p(1-p) + \sum_{k=1}^{\infty}(1-p)^2 p_X(3k) = p(1-p) + (1-p)^2 p_Y(0)$$

Therefore,

$$1 = p_Y(0) + p_Y(1) + p_Y(2) = p_Y(0) + \big(p + (1-p)p_Y(0)\big) + \big(p(1-p) + (1-p)^2 p_Y(0)\big)$$

$$= p + p(1-p) + \big(1 + (1-p) + (1-p)^2\big)p_Y(0) = 2p - p^2 + \left(3 - 3p + p^2\right) p_Y(0).$$

Solving for $p_Y(0)$ yields

$$p_Y(0) = \frac{1-2p+p^2}{3-3p+p^2} = \frac{(1-p)^2}{3-3p+p^2}.$$

Consequently, we also have

$$
\begin{aligned}
p_Y(1) = p + (1-p)p_Y(0) &= p + (1-p)\frac{(1-p)^2}{3-3p+p^2} \\
&= \frac{\left(3p - 3p^2 + p^3\right) + (1-p)^3}{3-3p+p^2} = \frac{1}{3-3p+p^2}
\end{aligned}
$$

and

$$
\begin{aligned}
p_Y(2) &= p(1-p) + (1-p)^2 p_Y(0) = p(1-p) + (1-p)^2 \frac{(1-p)^2}{3-3p+p^2} \\
&= \frac{p(1-p)\left(3-3p+p^2\right) + (1-p)^4}{3-3p+p^2} = \frac{(1-p)\left(\left(3p-3p^2+p^3\right)+(1-p)^3\right)}{3-3p+p^2} \\
&= \frac{1-p}{3-3p+p^2}.
\end{aligned}
$$

Hence,

$$
p_Y(y) = \begin{cases} (1-p)^2/(3-3p+p^2), & \text{if } y = 0; \\ 1/(3-3p+p^2), & \text{if } y = 1; \\ (1-p)/(3-3p+p^2), & \text{if } y = 2; \\ 0, & \text{otherwise.} \end{cases}
$$

Review Exercises for Chapter 5

Basic Exercises

5.142

a) The possible values of X are 1, 2, 3, and 4.
b) The event that the student selected is a junior can be represented as $\{X = 3\}$.
c) From the table, the total number of undergraduate students is

$$6{,}159 + 6{,}790 + 8{,}141 + 11{,}220 = 32{,}310.$$

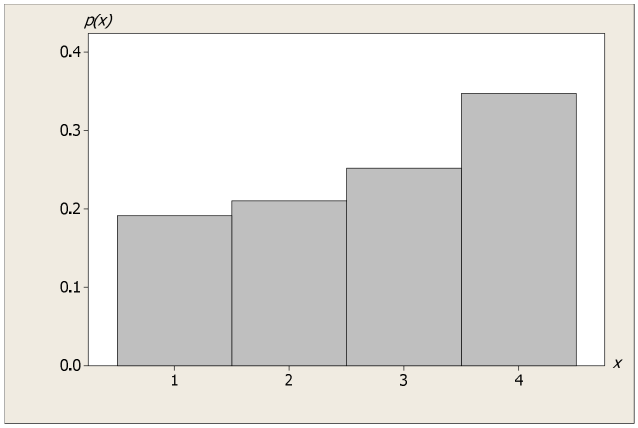
Hence, because the selection is done randomly, we have

$$P(X = 3) = \frac{N\{X = 3\}}{N(\Omega)} = \frac{8{,}141}{32{,}310} = 0.252.$$

Thus, 25.2% of the undergraduate students are juniors.
d) Proceeding as in part (c), we obtain the following table for the PMF of X:

x	1	2	3	4
$p_X(x)$	0.191	0.210	0.252	0.347

e) Following is a probability histogram for X. In the graph, we use $p(x)$ instead of $p_X(x)$.



5.143
a) The event that the number of busy lines is exactly four can be expressed as $\{Y = 4\}$.
b) The event that the number of busy lines is at least four can be expressed as $\{Y \geq 4\}$.
c) The event that the number of busy lines is between two and four, inclusive, can be expressed as $\{2 \leq Y \leq 4\}$.
d) We have

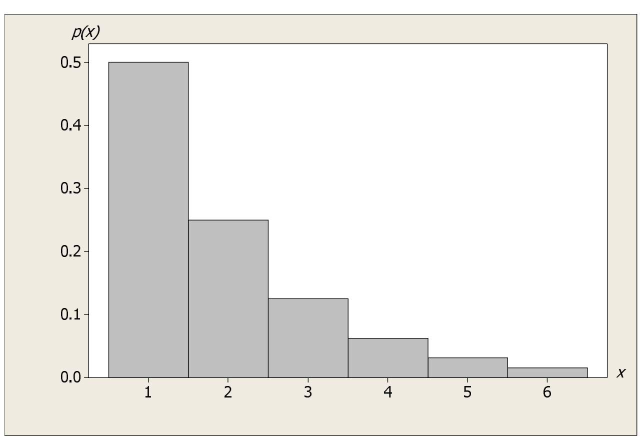
$$P(Y = 4) = p_Y(4) = 0.174,$$

$$P(Y \geq 4) = \sum_{y \geq 4} p_Y(y) = \sum_{k=4}^{6} p_Y(k) = 0.174 + 0.105 + 0.043 = 0.322,$$

and

$$P(2 \leq Y \leq 4) = \sum_{2 \leq y \leq 4} p_Y(y) = \sum_{k=2}^{4} p_Y(k) = 0.232 + 0.240 + 0.174 = 0.646.$$

5.144 We note that $X \sim \mathcal{G}(0.5)$. Hence, $p_X(x) = 2^{-x}$ if $x \in \mathcal{N}$, and $p_X(x) = 0$ otherwise.
a) Following is a partial probability histogram for X. In the graph, we use $p(x)$ instead of $p_X(x)$.



b) We can't draw the entire probability histogram because the range of X is infinite, namely, $\mathcal{N}$.

c) By the FPF,

$$P(X > 4) = \sum_{x>4} p_X(x) = \sum_{k=5}^{\infty} (1/2)^k = \frac{(1/2)^5}{1-(1/2)} = (1/2)^4 = \frac{1}{16}.$$

d) From the complementation rule and the FPF,

$$P(X > 4) = 1 - P(X \le 4) = 1 - \sum_{x \le 4} p_X(x) = 1 - \sum_{k=1}^{4} (1/2)^k$$

$$= 1 - \frac{(1/2)-(1/2)^5}{1-(1/2)} = 1 - \left(1-(1/2)^4\right) = (1/2)^4 = \frac{1}{16}.$$

e) Proceeding as in parts (c) and (d), respectively, we get

$$P(X > x) = \sum_{y>x} p_X(y) = \sum_{k=x+1}^{\infty} (1/2)^k = \frac{(1/2)^{x+1}}{1-(1/2)} = (1/2)^x = 2^{-x}$$

and

$$P(X > x) = 1 - P(X \le x) = 1 - \sum_{y \le x} p_X(y) = 1 - \sum_{k=1}^{x} (1/2)^k$$

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

f) Let N denote the set of prime numbers. Then

$$P(X \in N) = \sum_{x \in N} p_X(x) \ge p_X(2) + p_X(3) + p_X(5) + p_X(7) + p_X(11)$$

$$= \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^7} + \frac{1}{2^{11}} = 0.41455,$$

to five decimal places. Also, from part (e),

$$P(X \in N) = \sum_{x \in N} p_X(x) \le p_X(2) + p_X(3) + p_X(5) + p_X(7) + p_X(11) + P(X > 12)$$

$$= \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^7} + \frac{1}{2^{11}} + \frac{1}{2^{12}} = 0.41479,$$

to five decimal places. Thus, $0.41455 \le P(X \in N) \le 0.41479$ and, hence, $P(X \in N) = 0.415$ to three significant digits.

5.145

a) We have $p = 0.4$ because there is a 40% chance that a drinker is involved.

b) With independence, an outcome consisting of exactly k successes has probability $(0.4)^k(0.6)^{3-k}$. Thus, we have the following table:

Outcome	Probability	Outcome	Probability
(s, s, s)	$(0.4)^3(0.6)^0 = 0.064$	(f, s, s)	$(0.4)^2(0.6)^1 = 0.096$
(s, s, f)	$(0.4)^2(0.6)^1 = 0.096$	(f, s, f)	$(0.4)^1(0.6)^2 = 0.144$
(s, f, s)	$(0.4)^2(0.6)^1 = 0.096$	(f, f, s)	$(0.4)^1(0.6)^2 = 0.144$
(s, f, f)	$(0.4)^1(0.6)^2 = 0.144$	(f, f, f)	$(0.4)^0(0.6)^3 = 0.216$

c) From the table in part (b), we see that the outcomes in which exactly two of three traffic fatalities involve a drinker are (s, s, f), (s, f, s), and (f, s, s).
d) From the table in part (b), we see that each of the three outcomes in part (c) has probability 0.096. The three probabilities are the same because each probability is obtained by multiplying two success probabilities of 0.4 and one failure probability of 0.6.
e) From parts (c) and (d), we see that the probability that exactly two of three traffic fatalities involve a drinker is equal to $3 \cdot 0.096 = 0.288$.
f) $X \sim \mathcal{B}(3, 0.4)$; that is, X has the binomial distribution with parameters 3 and 0.4.
g) Referring to the table in part (b), we get

$$\begin{aligned}
P(X = 0) &= P\big(\{(f, f, f)\}\big) = 0.216, \\
P(X = 1) &= P\big(\{(s, f, f), (f, s, f), (f, f, s)\}\big) = P\big(\{(s, f, f)\}\big) + P\big(\{(f, s, f)\}\big) + P\big(\{(f, f, s)\}\big) \\
&= 0.144 + 0.144 + 0.144 = 0.432, \\
P(X = 2) &= P\big(\{(s, s, f), (s, f, s), (f, s, s)\}\big) = P\big(\{(s, s, f)\}\big) + P\big(\{(s, f, s)\}\big) + P\big(\{(f, s, s)\}\big) \\
&= 0.096 + 0.096 + 0.096 = 0.288, \\
P(X = 3) &= P\big(\{(s, s, s)\}\big) = 0.064.
\end{aligned}$$

Thus, we have the following table for the PMF of X:

x	0	1	2	3
$p_X(x)$	0.216	0.432	0.288	0.064

h) Applying Procedure 5.1 with $n = 3$ and $p = 0.4$, we get

$$p_X(x) = \binom{3}{x}(0.4)^x(0.6)^{3-x}, \qquad x = 0, 1, 2, 3,$$

and $p_X(x) = 0$ otherwise. Substituting successively $x = 0$, 1, 2, and 3 into the preceding display and doing the necessary algebra gives the same values as shown in the table in part (g).

5.146 Let Y denote the number of traffic fatalities examined up to and including the first one that involves a drinker. Then $Y \sim \mathcal{G}(0.4)$. From Proposition 5.9 on page 229,

$$p_Y(y) = (0.4)(0.6)^{y-1}, \qquad y = 1, 2, \ldots,$$

and $p_Y(y) = 0$ otherwise. Also, from Proposition 5.10 on page 231,

$$P(Y > n) = (0.6)^n, \qquad n \in \mathcal{N}.$$

a) We have

$$P(Y = 4) = p_Y(4) = (0.4)(0.6)^{4-1} = (0.4)(0.6)^3 = 0.0864.$$

b) From the complementation rule,

$$P(Y \le 4) = 1 - P(Y > 4) = 1 - (0.6)^4 = 0.870.$$

c) We have

$$P(Y \ge 4) = P(Y > 3) = (0.6)^3 = 0.216.$$

d) We want to find the largest n such that $P(Y > n) > 0.2$. Thus,

$$(0.6)^n > 0.2 \qquad \text{or} \qquad n < \frac{\ln 0.2}{\ln 0.6} \approx 3.15.$$

Hence, $n = 3$.
e) See the beginning of the solution to this exercise.

5.147 Let X denote the number of the eight tires tested that get at least 35,000 miles. We know that $X \sim \mathcal{B}(8, 0.9)$. Therefore, by the FPF,

$$P(X \geq 0.75 \cdot 8) = P(X \geq 6) = \sum_{x \geq 6} p_X(x) = \sum_{k=6}^{8} \binom{8}{k} (0.9)^k (0.1)^{8-k} = 0.962.$$

5.148 Let X denote the number of "no-shows" when n reservations are made. Then $X \sim \mathcal{B}(n, 0.04)$. We want to find the largest n so that $P(X = 0) \geq 0.8$. Noting that

$$P(X = 0) = \binom{n}{0} (0.04)^0 (0.96)^n = (0.96)^n,$$

we have the requirement that

$$(0.96)^n \geq 0.8 \qquad \text{or} \qquad n \leq \frac{\ln 0.8}{\ln 0.96} \approx 5.5.$$

Hence, $n = 5$.

5.149 Let X denote the number of moderate-income families of five sampled that spend more than half their income on housing. Then $X \sim \mathcal{H}(N, 5, 0.76)$, where N is the number of moderate-income families in the United States. Because the sample size (5) is small relative to the population size (N), we know, from Proposition 5.6 on page 215, that $X \approx \mathcal{B}(5, 0.76)$. Thus,

$$p_X(x) \approx \binom{5}{x} (0.76)^x (0.24)^{5-x}, \qquad x = 0, 1, \ldots, 5.$$

a) We have

$$P(X = 3) = p_X(3) \approx \binom{5}{3} (0.76)^3 (0.24)^2 = 0.253.$$

b) From the FPF,

$$P(X \leq 3) = \sum_{x \leq 3} p_X(x) = \sum_{k=0}^{3} p_X(k) \approx \sum_{k=0}^{3} \binom{5}{k} (0.76)^k (0.24)^{5-k} = 0.346.$$

c) From the FPF,

$$P(X \geq 4) = \sum_{x \geq 4} p_X(x) = \sum_{k=4}^{5} p_X(k) \approx \sum_{k=4}^{5} \binom{5}{k} (0.76)^k (0.24)^{5-k} = 0.654.$$

d) From the complementation rule and the FPF,

$$\begin{aligned} P(1 \leq X \leq 4) &= 1 - P(X = 0 \text{ or } 5) = 1 - \big(p_X(0) + p_X(5)\big) \\ &\approx 1 - \left(\binom{5}{0} (0.76)^0 (0.24)^5 + \binom{5}{5} (0.76)^5 (0.24)^0 \right) \\ &= 1 - (0.24)^5 - (0.76)^5 = 0.746. \end{aligned}$$

e) Applying the formula presented at the beginning of this solution for the PDF of a $\mathcal{B}(5, 0.76)$ random variable, we obtain the following table:

x	0	1	2	3	4	5
$p_X(x)$	0.000796	0.012607	0.079847	0.252850	0.400346	0.253553

f) The PMF obtained in part (e) assumes sampling with replacement, whereas the sampling is actually without replacement. Thus, the binomial approximation to the hypergeometric distribution (Proposition 5.6) has been used.
g) As we noted, $X \sim \mathcal{H}(N, 5, 0.76)$. Hence,

$$p_X(x) = \frac{\binom{0.76N}{x}\binom{0.24N}{5-x}}{\binom{N}{5}}, \qquad x = 0, 1, \ldots, 5.$$

5.150
a) Let X denote the number of successes in n Bernoulli trials with success probability p. We know that $X \sim \mathcal{B}(n, p)$. Let X_j be 0 or 1, respectively, depending on whether a failure or success occurs on trial j. We want to find $P(X_j = 1 \mid X = n)$. The event $\{X_j = 1,\ X = n\}$ occurs if and only if trial j results in a success and among the other $n-1$ trials exactly $k-1$ successes occur. The former event has probability p and the latter event has probability

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

Moreover, by the independence of Bernoulli trials, the two events are independent. Applying the conditional probability rule, the special multiplication rule, and the binomial PMF, we get

$$P(X_j = 1 \mid X = n) = \frac{P(X_j = 1,\ X = n)}{P(X = n)} = \frac{p \cdot \binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}.$$

b) Given that exactly k successes occur in n Bernoulli trials, we can think of the positions (trial numbers) of the k successes and $n-k$ failures as obtained by randomly placing k "s"s and $n-k$ "f"s in n boxes, one in each box. By symmetry, the letter ("s" or "f") placed in box j is equally likely to be any of the n letters. As k of the n letters are "s"s, the probability is k/n that an "s" goes into box j—that is, the probability is k/n that trial j results in success.

5.151
a) According to the frequentist interpretation of probability (see page 5), we have $P(E) \approx n(E)/n$. Intuitively, then, we estimate $P(E)$ to be $n(E)/n$.
b) We note that $n(E) \sim \mathcal{B}(n, p)$, where $p = P(E)$. Suppose that E occurs k times in the n trials, that is, that event $\{n(E) = k\}$ occurs. The probability of that happening is $\binom{n}{k}p^k(1-p)^{n-k}$. We want to choose p to maximize that quantity, which we denote $f(p)$. We have

$$\begin{aligned} f'(p) &= \binom{n}{k}\left(kp^{k-1} \cdot (1-p)^{n-k} - p^k \cdot (n-k)(1-p)^{n-k-1}\right) \\ &= \binom{n}{k}p^{k-1}(1-p)^{n-k-1}\big(k(1-p) - (n-k)p\big) = \binom{n}{k}p^{k-1}(1-p)^{n-k-1}(k - np). \end{aligned}$$

From the previous display, we see that $f'(p) = 0$ if and only if $p = k/n$ and, moreover, that $f'(p) > 0$ if $p < k/n$ and $f'(p) < 0$ if $p > k/n$. Therefore, f takes its maximum value at $p = k/n$. In other words, the maximum likelihood estimate of $P(E)$ is the random variable $\hat{p} = n(E)/n$.
c) The maximum likelihood estimate of $P(E)$ coincides with the intuitive estimate based on the frequentist interpretation of probability.

5.152 Referring to Definition 5.5 on page 211, we obtain the following results.

a) Because $4 > 0$,
$$\binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4!} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{24} = 126.$$
b) Because $-4 < 0$,
$$\binom{9}{-4} = 0.$$
c) Because $4 > 0$,
$$\binom{-9}{4} = \frac{(-9) \cdot (-10) \cdot (-11) \cdot (-12)}{4!} = \frac{9 \cdot 10 \cdot 11 \cdot 12}{24} = 495.$$
d) Because $-4 < 0$,
$$\binom{-9}{-4} = 0.$$
e) Because $9 > 0$,
$$\binom{4}{9} = \frac{4 \cdot 3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1) \cdot (-2) \cdot (-3) \cdot (-4)}{9!} = \frac{0}{9!} = 0.$$
f) Because $-9 < 0$,
$$\binom{4}{-9} = 0.$$
g) Because $9 > 0$,
$$\begin{aligned}\binom{-4}{9} &= \frac{(-4) \cdot (-5) \cdot (-6) \cdot (-7) \cdot (-8) \cdot (-9) \cdot (-10) \cdot (-11) \cdot (-12)}{9!} \\ &= -\frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 10 \cdot 11 \cdot 12}{9!} = -\frac{10 \cdot 11 \cdot 12}{3 \cdot 2 \cdot 1} = -220.\end{aligned}$$
h) Because $-9 < 0$,
$$\binom{-4}{-9} = 0.$$
i) Because $4 > 0$,
$$\binom{1/2}{4} = \frac{(1/2) \cdot (-1/2) \cdot (-3/2) \cdot (-5/2)}{4!} = -\frac{15}{16 \cdot 24} = -\frac{5}{128}.$$
j) Because $-4 < 0$,
$$\binom{1/2}{-4} = 0.$$
k) Because $4 > 0$,
$$\binom{-1/2}{4} = \frac{(-1/2) \cdot (-3/2) \cdot (-5/2) \cdot (-7/2)}{4!} = \frac{1 \cdot 3 \cdot 5 \cdot 7}{16 \cdot 24} = \frac{35}{128}.$$
l) Because $-4 < 0$,
$$\binom{-1/2}{-4} = 0.$$

5.153 Let Y denote the number of tests required under the alternative scheme. Also, let X denote the number of people in the sample who have the disease. We know that $X \sim \mathcal{H}(N, n, p)$, where N is the number of people in the population. As N is very large, we can use the binomial approximation to the hypergeometric distribution. Hence, $X \approx \mathcal{B}(n, p)$. We have

$$Y = \begin{cases} 1, & \text{if } X = 0; \\ n+1, & \text{if } X \geq 1. \end{cases}$$

Now,

$$p_Y(1) = P(Y = 1) = P(X = 0) \approx \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n$$

and, by the complementation rule,

$$p_Y(n+1) = 1 - p_Y(0) = 1 - (1-p)^n.$$

Hence,

$$p_Y(y) = \begin{cases} (1-p)^n, & \text{if } y = 1; \\ 1-(1-p)^n, & \text{if } y = n+1; \\ 0, & \text{otherwise.} \end{cases}$$

5.154

a) The number of patients, X, that arrive within the first hour of the doctor's arrival has, by assumption, the Poisson distribution with parameter $\lambda = 6.9$. The first patient arrives more than 1 hour after the doctor arrives if and only if no patients arrive within the first hour of the doctor's arrival, which means that $X = 0$. Hence, the required probability is

$$P(X = 0) = e^{-6.9} \frac{(6.9)^0}{0!} = e^{-6.9} = 0.00101.$$

b) According to Proposition 5.8 on page 225, the most probable number of patients that arrive within the first hour of the doctor's arrival is $\lfloor 6.9 \rfloor = 6$. From the complementation rule and the FPF,

$$P(|X-6| > 1) = 1 - P(|X-6| \leq 1) = 1 - \sum_{|x-6|\leq 1} p_X(x) = 1 - \sum_{k=5}^{7} e^{-6.9} \frac{(6.9)^k}{k!} = 0.569.$$

c) From the complementation rule and the FPF,

$$P(X \geq 5) = 1 - P(X < 5) = 1 - \sum_{x<5} p_X(x) = 1 - \sum_{k=0}^{4} e^{-6.9} \frac{(6.9)^k}{k!} = 0.818.$$

d) Let Y denote the number of patients observed by the inspector during a 1-hour period. Also, let

L = event the inspector inspects the large hospital, and

S = event the inspector inspects the small hospital.

As the inspector tosses a balanced coin to decide which hospital to inspect, we have $P(L) = P(S) = 1/2$. Furthermore, given that event L occurs, $Y \sim \mathcal{P}(6.9)$, whereas, given that event S occurs, $Y \sim \mathcal{P}(2.6)$. Therefore, by the law of total probability,

$$\begin{aligned} P(Y = 3) &= P(L)P(Y = 3 \mid L) + P(S)P(Y = 3 \mid S) = \frac{1}{2} \cdot e^{-6.9} \frac{(6.9)^3}{3!} + \frac{1}{2} \cdot e^{-2.6} \frac{(2.6)^3}{3!} \\ &= 0.136. \end{aligned}$$

e) From Bayes's rule,

$$P(S \mid Y = 3) = \frac{P(S)P(Y = 3 \mid S)}{P(L)P(Y = 3 \mid L) + P(S)P(Y = 3 \mid S)} = \frac{\frac{1}{2} \cdot e^{-2.6} \frac{(2.6)^3}{3!}}{\frac{1}{2} \cdot e^{-6.9} \frac{(6.9)^3}{3!} + \frac{1}{2} \cdot e^{-2.6} \frac{(2.6)^3}{3!}}$$

$$= \frac{1}{e^{-4.3}(6.9/2.6)^3 + 1} = 0.798.$$

5.155

a) By the assumptions, we see that $X \sim \mathcal{B}(2500, 0.001)$. Hence,

$$p_X(x) = \binom{2500}{x}(0.001)^x(0.999)^{2500-x}, \qquad x = 0, 1, \ldots, 2500,$$

and $p_X(x) = 0$ otherwise.

b) Noting that $2500 \cdot 0.001 = 2.5$, we see, from Proposition 5.7 on page 220, that

$$p_X(x) \approx e^{-2.5}\frac{(2.5)^x}{x!}, \qquad x = 0, 1, \ldots, 2500.$$

c) From the complementation rule, the FPF, and part (a),

$$P(X \geq 5) = 1 - P(X < 5) = 1 - \sum_{x<5} p_X(x) = 1 - \sum_{k=0}^{4}\binom{2500}{k}(0.001)^k(0.999)^{2500-k} = 0.1087.$$

And, from the complementation rule, the FPF, and part (b),

$$P(X \geq 5) = 1 - P(X < 5) = 1 - \sum_{x<5} p_X(x) \approx 1 - \sum_{k=0}^{4} e^{-2.5}\frac{(2.5)^k}{k!} = 0.1088.$$

The Poisson approximation to the binomial probability is excellent here.

d) Answers will vary. We might, however, be justified in claiming that typographical errors are not independent. For instance, if a typographical error is the result of the typist misplacing his fingers on the keyboard, then subsequent characters are likely to be in error.

5.156 We have

$$p_X(x) = e^{-1.75}\frac{(1.75)^x}{x!}, \qquad x = 0, 1, \ldots,$$

and $p_X(x) = 0$ otherwise.

a) We have

$$P(X = 2) = p_X(2) = e^{-1.75}\frac{(1.75)^2}{2!} = 0.266.$$

b) From the FPF,

$$P(4 \leq X \leq 6) = \sum_{4 \leq x \leq 6} p_X(x) = \sum_{k=4}^{6} e^{-1.75}\frac{(1.75)^k}{k!} = e^{-1.75}\sum_{k=4}^{6}\frac{(1.75)^k}{k!} = 0.0986.$$

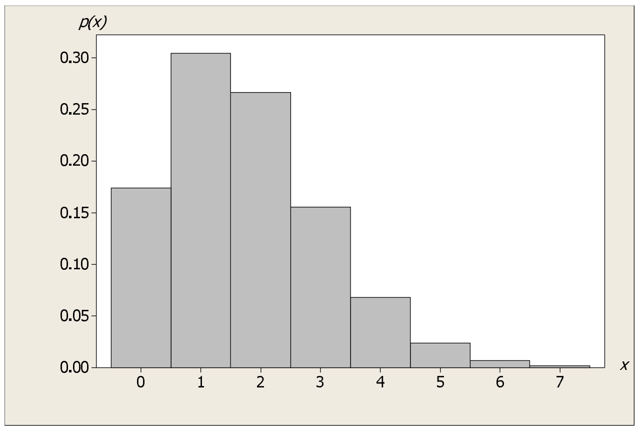
c) From the complementation rule,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1.75}\frac{(1.75)^0}{0!} = 0.826.$$

d) Applying the formula for the PMF of the $\mathcal{P}(1.75)$ distribution, presented at the beginning of the solution to this exercise, we get the following table:

x	0	1	2	3	4	5	6	7
$p_X(x)$	0.174	0.304	0.266	0.155	0.068	0.024	0.007	0.002

e) A partial probability histogram for the random variable X is as follows. Note that, in the graph, we have used $p(x)$ in place of $p_X(x)$.



f) From part (e), the probability histogram of X is right skewed. This property holds for all Poisson random variables, as we learned in Proposition 5.8 on page 225.

5.157

a) We first recall that the number of possible five-card draw poker hands is $\binom{52}{5}$. A hand of four of a kind is of the form $\{x, x, x, x, y\}$, where x and y are distinct denominations. There are $\binom{13}{1}$ choices for x and then $\binom{12}{1}$ choices for y. Then there are $\binom{4}{4}$ ways to choose the four xs from the four xs, and there are $\binom{4}{1}$ ways to choose the one y from the four ys. Hence, the probability of getting four of a kind is

$$\frac{\binom{13}{1}\binom{12}{1}\binom{4}{4}\binom{4}{1}}{\binom{52}{5}} = \frac{624}{2{,}598{,}960} = 0.00024.$$

b) Let X denote the number of times that we get four of a kind in 10,000 hands of five-card draw. Then $X \sim \mathcal{B}(10{,}000, 0.00024)$. Noting that $10{,}000 \cdot 0.00024 = 2.4$, the Poisson approximation to this binomial distribution yields

$$p_X(x) \approx e^{-2.4}\frac{(2.4)^x}{x!}, \qquad x = 0, 1, \ldots, 10{,}000.$$

Hence,

$$P(X = 2) = p_X(2) \approx e^{-2.4}\frac{(2.4)^2}{2!} = 0.261.$$

Also, from the complementation rule and the FPF,

$$P(X \geq 2) = 1 - P(X < 2) = 1 - \sum_{x<2} p_X(x) = 1 - p_X(0) - p_X(1)$$
$$\approx 1 - e^{-2.4}\,\frac{(2.4)^0}{0!} - e^{-2.4}\,\frac{(2.4)^1}{1!} = 1 - 3.4e^{-2.4} = 0.692.$$

5.158 Suppose that $X \sim \mathcal{B}(n, p)$, where n is large and p is large (i.e., close to 1). The random variable $Y = n - X$ gives the number of failures in the n Bernoulli trials and, hence, by interchanging the roles of success and failure, we see that $Y \sim \mathcal{B}(n, 1-p)$. As n is large and $1-p$ is small, we can apply Proposition 5.7 on page 220 to conclude that

$$p_Y(y) \approx e^{-n(1-p)}\,\frac{\big(n(1-p)\big)^y}{y!}, \qquad y = 0, 1, \dots, n.$$

Consequently, from Equation (5.54) on page 248,

$$p_X(x) = p_{n-Y}(x) = p_Y(n-x) \approx e^{-n(1-p)}\,\frac{\big(n(1-p)\big)^{n-x}}{(n-x)!}, \qquad x = 0, 1, \dots, n.$$

Thus, in the sense of the preceding display, we can use the Poisson distribution with parameter $n(1-p)$ to approximate probabilities for X.

5.159

a) Let X denote weekly demand to the nearest gallon. Then $X \sim \mathcal{P}(\lambda)$. Letting Y denote weekly gas sales, we see that $Y = \min\{X, m\}$. The range of Y is $\{0, 1, \dots, m\}$. For $y = 0, 1, \dots, m-1$,

$$p_Y(y) = P(Y = y) = P\big(\min\{X, m\} = y\big) = P(X = y) = p_X(y) = e^{-\lambda}\,\frac{\lambda^y}{y!}.$$

Also, from the complementation rule and the FPF,

$$p_Y(m) = P(Y = m) = P\big(\min\{X, m\} = m\big) = P(X \geq m) = 1 - P(X < m)$$
$$= 1 - \sum_{x<m} p_X(x) = 1 - \sum_{k=0}^{m-1} e^{-\lambda}\,\frac{\lambda^k}{k!} = 1 - e^{-\lambda}\sum_{k=0}^{m-1}\frac{\lambda^k}{k!}.$$

Hence,

$$p_Y(y) = \begin{cases} e^{-\lambda}\,\dfrac{\lambda^y}{y!}, & \text{if } y = 0, 1, \dots, m-1; \\[2ex] 1 - e^{-\lambda}\displaystyle\sum_{k=0}^{m-1}\frac{\lambda^k}{k!}, & \text{if } y = m; \\[2ex] 0, & \text{otherwise.} \end{cases}$$

Note: We could also apply Proposition 5.14 on page 247 to solve this problem.

b) The gas station runs out of gas before the end of the week if and only if the demand equals or exceeds m, which is

$$P(X \geq m) = \sum_{x \geq m} p_X(x) = \sum_{k=m}^{\infty} e^{-\lambda}\,\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=m}^{\infty}\frac{\lambda^k}{k!}.$$

5.160 For convenience, set $p = e^{-6.9}(6.9)^7/7!$. For each $n \in \mathcal{N}$, let E_n denote the event that exactly seven patients arrive during the nth hour after 12:00 P.M.. By assumption, $E_1, E_2, \ldots$ are independent events and, moreover, $P(E_n) = p$ for all $n \in \mathcal{N}$. Referring to Proposition 4.5 on page 152 and applying the complementation rule, we conclude that

$$P(X = k) = P\big(E_1^c \cap \cdots \cap E_{k-1}^c \cap E_k\big) = P\big(E_1^c\big) \cdots P\big(E_{k-1}^c\big) P(E_k) = p(1-p)^{k-1}, \qquad k \in \mathcal{N}.$$

Therefore, $X \sim \mathcal{G}(p)$.

a) We have

$$P(X = 4) = p(1-p)^3 = 0.0918.$$

b) From Proposition 5.10 on page 231,

$$P(X > 6) = (1-p)^6 = 0.380.$$

c) From the lack-of-memory property of geometric random variables, Exercise 5.106, and part (b), we deduce that

$$P(X > 15 \mid X > 9) = P(X > 9 + 6 \mid X > 9) = P(X > 6) = 0.380.$$

5.161

a) Referring to Proposition 5.10 on page 231, we get

$$P(X > 6) = (1-p)^6 = \binom{6}{0} p^0 (1-p)^6 = P(Y = 0).$$

b) Consider Bernoulli trials with success probability p. Let X denote the number of trials up to and including the first success, and let Y denote the number of successes in the first six trials. Then $X \sim \mathcal{G}(p)$ and $Y \sim \mathcal{B}(6, p)$. Now, $X > 6$ if and only if the first success occurs after the sixth trial, which happens if and only if no successes occur during the first six trials, which means that $Y = 0$. Consequently, we have $\{X > 6\} = \{Y = 0\}$ and, hence, in particular, $P(X > 6) = P(Y = 0)$.

5.162

a) From Proposition 5.13 on page 241, we know that $X \sim \mathcal{NB}(3, p)$ and that

$$p_X(x) = \binom{x-1}{2} p^3 (1-p)^{x-3}, \qquad x = 3, 4, \ldots,$$

and $p_X(x) = 0$ otherwise.

b) From the FPF,

$$P(X \le 40) = \sum_{x \le 40} p_X(x) = \sum_{k=3}^{40} p_X(k).$$

Hence, we see that 38 terms of the PMF of X must be added to obtain $P(X \le 40)$. We have

$$P(X \le 40) = \sum_{k=3}^{40} \binom{k-1}{2} p^3 (1-p)^{k-3}.$$

c) Let Y denote the number of successes in the first 40 trials. Then, we have $Y \sim \mathcal{B}(40, p)$. We note that $\{X > 40\} = \{Y < 3\}$. Hence, from the complementation rule and the FPF,

$$P(X \le 40) = 1 - P(X > 40) = 1 - P(Y < 3) = 1 - \sum_{y < 3} p_Y(y) = 1 - \sum_{k=0}^{2} p_Y(k).$$

Using the preceding display and the PMF of a $\mathcal{B}(40, p)$ random variable, we get

$$P(X \leq 40) = 1 - \sum_{k=0}^{2} \binom{40}{k} p^k (1-p)^{40-k}.$$

d) To evaluate $P(X \leq 40)$, it is far easier to use the method of part (c) than the direct method of part (b).

5.163 Consider a "success" a year in which at least one serious accident occurs. Let X denote the number of years until the third success (i.e., until the third year in which at least one serious accident occurs). From the assumptions, we know that $X \sim \mathcal{NB}(3, 0.4)$. The problem is to determine the probability that there will be at least 3 years in which no serious accidents occur before the third year in which at least one serious accident occurs, that is, the probability of at least three failures before the third success. Equivalently, we want $P(X \geq 6)$. From the complementation rule and the FPF,

$$P(X \geq 6) = 1 - P(X < 6) = 1 - \sum_{x<6} p_X(x) = 1 - \sum_{k=3}^{5} \binom{k-1}{2} (0.4)^3 (0.6)^{k-3} = 0.683.$$

Advanced Exercises

5.164 Let X_j, $1 \leq j \leq n$, denote the number of the jth member sampled. Because the sampling is with replacement, $X_1, \ldots, X_n$ are independent random variables. And, because the sampling is random, each X_j has the discrete uniform distribution on the set $S = \{1, 2, \ldots, N\}$.

a) Let $X = \max\{X_j : 1 \leq j \leq n\}$. From the FPF, for $k \in S$,

$$P(X_j \leq k) = \sum_{x \leq k} p_{X_j}(x) = \sum_{i=1}^{k} \frac{1}{N} = \frac{k}{N}.$$

Hence,

$$\begin{aligned} P(X \leq k) &= P\big(\max\{X_j : 1 \leq j \leq n\} \leq k\big) = P(X_1 \leq k, \ldots, X_n \leq k) \\ &= P(X_1 \leq k) \cdots P(X_n \leq k) = \left(\frac{k}{N}\right)^n. \end{aligned}$$

Consequently,

$$P(X = k) = P(X \leq k) - P(X \leq k-1) = \left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n = \frac{k^n - (k-1)^n}{N^n}.$$

Thus,

$$p_X(x) = \frac{x^n - (x-1)^n}{N^n}, \qquad x = 1, 2, \ldots, N,$$

and $p_X(x) = 0$ otherwise.

b) Let $Y = \min\{X_j : 1 \leq j \leq n\}$. From the FPF, for $k \in S$,

$$P(X_j > k) = \sum_{x>k} p_{X_j}(x) = \sum_{i=k+1}^{N} \frac{1}{N} = \frac{N-k}{N}.$$

Hence,

$$\begin{aligned} P(Y > k) &= P\big(\min\{X_j : 1 \leq j \leq n\} > k\big) = P(X_1 > k, \ldots, X_n > k) \\ &= P(X_1 > k) \cdots P(X_n > k) = \left(\frac{N-k}{N}\right)^n. \end{aligned}$$

Consequently,

$$P(Y = k) = P(Y > k-1) - P(Y > k) = \left(\frac{N-(k-1)}{N}\right)^n - \left(\frac{N-k}{N}\right)^n$$
$$= \frac{(N-k+1)^n - (N-k)^n}{N^n}.$$

Thus,

$$p_Y(y) = \frac{(N-y+1)^n - (N-y)^n}{N^n}, \qquad y = 1, 2, \ldots, N,$$

and $p_Y(y) = 0$ otherwise.

5.165

a) Let X denote the largest numbered member sampled. Note that event $\{X = k\}$ occurs if and only if exactly $n-1$ of the members sampled have numbers at most $k-1$ and one member sampled has number k. Hence,

$$P(X = k) = \frac{N(\{X = k\})}{\binom{N}{n}} = \frac{\binom{k-1}{n-1}\binom{1}{1}\binom{N-k}{0}}{\binom{N}{n}} = \frac{\binom{k-1}{n-1}}{\binom{N}{n}}, \qquad k = n, n+1, \ldots, N.$$

Thus,

$$p_X(x) = \frac{\binom{x-1}{n-1}}{\binom{N}{n}}, \qquad x = n, n+1, \ldots, N,$$

and $p_X(x) = 0$ otherwise.

b) Let Y denote the smallest numbered member sampled. Note that event $\{Y = k\}$ occurs if and only if exactly $n-1$ of the members sampled have numbers at least $k+1$ and one member sampled has number k. Hence,

$$P(Y = k) = \frac{N(\{Y = k\})}{\binom{N}{n}} = \frac{\binom{k-1}{0}\binom{1}{1}\binom{N-k}{n-1}}{\binom{N}{n}} = \frac{\binom{N-k}{n-1}}{\binom{N}{n}}, \qquad k = 1, 2, \ldots, N-n+1.$$

Thus,

$$p_Y(y) = \frac{\binom{N-y}{n-1}}{\binom{N}{n}}, \qquad y = 1, 2, \ldots, N-n+1,$$

and $p_Y(y) = 0$ otherwise.

5.166 We have

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

and

$$p_Y(y) = \binom{m}{y} p^y (1-p)^{m-y}, \quad y = 0, 1, \ldots, m.$$

a) For $z = 0, 1, \dots, n+m$, we have

$$\begin{aligned}
P(X+Y=z) &= \sum_x P(X=x)P(X+Y=z \mid X=x) \\
&= \sum_x P(X=x)P(Y=z-x \mid X=x) = \sum_x P(X=x)P(Y=z-x) \\
&= \sum_{x=0}^{z} \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{z-x} p^{z-x}(1-p)^{m-(z-x)} \\
&= p^z(1-p)^{n+m-z} \sum_{x=0}^{z} \binom{n}{x}\binom{m}{z-x} = \binom{n+m}{z} p^z (1-p)^{n+m-z},
\end{aligned}$$

where the last equality follows from Vandermonde's identity. Thus, $X+Y \sim \mathcal{B}(n+m, p)$.
b) For $x = 0, 1, \dots, z$, we have, in view of the conditional probability rule, the independence condition, and part (a), that

$$\begin{aligned}
P(X=x \mid X+Y=z) &= \frac{P(X=x,\ X+Y=z)}{P(X+Y=z)} \\
&= \frac{P(X=x,\ Y=z-x)}{P(X+Y=z)} = \frac{P(X=x)P(Y=z-x)}{P(X+Y=z)} \\
&= \frac{\binom{n}{x} p^x (1-p)^{n-x} \binom{m}{z-x} p^{z-x}(1-p)^{m-(z-x)}}{\binom{n+m}{z} p^z (1-p)^{n+m-z}} \\
&= \frac{\binom{n}{x}\binom{m}{z-x}}{\binom{n+m}{z}}.
\end{aligned}$$

Thus, given that $X+Y=z$, we have $X \sim \mathcal{H}\big(n+m, z, n/(n+m)\big)$.

5.167 If the sample is drawn with replacement, then the number of nondefective items of the six sampled has the $\mathcal{B}(6, 2/3)$ distribution. Hence, its PMF is

$$\binom{6}{k}(2/3)^k(1/3)^{6-k}, \qquad k = 0, 1, \dots 6.$$

However, if the sample is drawn without replacement., then the number of nondefective items of the six sampled has the $\mathcal{H}(12, 6, 2/3)$ distribution. Hence, its PMF is

$$\frac{\binom{8}{k}\binom{4}{6-k}}{\binom{12}{6}}, \qquad k = 0, 1, \dots 6.$$

The following table provides both PMFs.

k	Binomial	Hypergeometric
0	0.00137	0.00000
1	0.01646	0.00000
2	0.08230	0.03030
3	0.21948	0.24242
4	0.32922	0.45455
5	0.26337	0.24242
6	0.08779	0.03030

The engineer will probably estimate the number of nondefectives in the lot of 12 to be roughly twice the number of nondefectives in the sample of six:

$$M \approx 12 \cdot \frac{k}{6} = 2k.$$

From the table, we see that it's more likely for the number of nondefectives in the sample to be close to four—that is, for the engineer's estimate to be close to eight—if the sampling is without replacement (hypergeometric). For instance, the probability that the number of nondefectives in the sample will be within one of four is 0.812 for the binomial and 0.939 for the hypergeometric. Hence, a better estimate of the number of nondefectives in the lot is obtained by sampling without replacement.

5.168

a) We would expect the binomial approximation to be worse when $n = 50$ than when $n = 8$ because, in the former case, the sample size is larger relative to the size of the population than in the latter case. In fact, here, when $n = 50$, the rule of thumb for use of a binomial distribution to approximate a hypergeometric distribution is just barely met—the sample size of 50 is exactly 5% of the population size of 1000.

b) We have the following table, where the probabilities are rounded to eight decimal places and where we displayed probabilities until they are zero to that number of decimal places:

Successes x	Hypergeometric probability	Binomial probability	Successes x	Hypergeometric probability	Binomial probability
0	0.00001038	0.00001427	15	0.02854165	0.02991866
1	0.00013826	0.00017841	16	0.01507892	0.01636177
2	0.00089640	0.00109274	17	0.00723473	0.00818088
3	0.00377132	0.00437095	18	0.00316049	0.00374957
4	0.01157781	0.01283965	19	0.00125978	0.00157877
5	0.02765180	0.02953120	20	0.00045900	0.00061177
6	0.05349306	0.05537101	21	0.00015309	0.00021849
7	0.08617037	0.08701158	22	0.00004679	0.00007200
8	0.11793007	0.11692182	23	0.00001312	0.00002191
9	0.13921652	0.13640879	24	0.00000337	0.00000616
10	0.14344797	0.13981901	25	0.00000080	0.00000160
11	0.13023588	0.12710819	26	0.00000017	0.00000039
12	0.10498345	0.10327540	27	0.00000003	0.00000009
13	0.07561264	0.07547049	28	0.00000001	0.00000002
14	0.04891210	0.04986443			

c) As we see by comparing the table in part (b) to Table 5.15, the accuracy of the binomial approximation to the hypergeometric is superior when $n = 8$ than when $n = 50$—just as expected.

5.169

a) The likelihood ratio for M based on an observed value k of X is

$$\frac{L_k(M)}{L_k(M-1)} = \frac{\binom{M}{k}\binom{N-M}{n-k}\Big/\binom{N}{n}}{\binom{M-1}{k}\binom{N-(M-1)}{n-k}\Big/\binom{N}{n}}$$
$$= \frac{M(N-M+1-n+k)}{(M-k)(N-M+1)} = 1 + \frac{(N+1)k - nM}{(M-k)(N-M+1)},$$

where the last equality is obtained by subtracting and adding $(M-k)(N-M+1)$ from the numerator of the penultimate expression, and then simplifying algebraically. It follow that $L_k(M)$ takes its maximum value at $M = \lfloor (N+1)k/n \rfloor$. Hence, the maximum likelihood estimate of M is given by

$$\widehat{M} = \left\lfloor (N+1)\frac{X}{n} \right\rfloor = \lfloor (N+1)\hat{p} \rfloor.$$

b) In this case, $N = 80{,}000$, $n = 100$, and $X = 38$. Hence, from part (a), the maximum likelihood estimate of the student's vocabulary is

$$\widehat{M} = \left\lfloor (80{,}000+1)\cdot\frac{38}{100} \right\rfloor = \lfloor 30{,}400.38 \rfloor = 30{,}400.$$

5.170

a) The random variable X has PMF given by

$$p_X(x) = p(1-p)^x, \qquad x = 0, 1, 2, \ldots,$$

and $p_X(x) = 0$ otherwise. Given that $X = n$, the random variable Y has the binomial distribution with parameters n and p. Let k be a nonnegative integer. Applying the law of total probability yields

$$P(Y=k) = \sum_x P(X=x)P(Y=k \mid X=x) = \sum_{n=0}^{\infty} P(X=n)P(Y=k \mid X=n)$$
$$= \sum_{n=k}^{\infty} p(1-p)^n \binom{n}{k} p^k (1-p)^{n-k} = \frac{p^{k+1}}{(1-p)^k}\sum_{n=k}^{\infty}\binom{n}{k}(1-p)^{2n}.$$

Now we let $m = n+1$, $r = k+1$, and $p_0 = 1-(1-p)^2 = p(2-p)$. Then

$$P(Y=k) = \frac{p^r}{(1-p)^{r-1}}\sum_{m=r}^{\infty}\binom{m-1}{r-1}(1-p_0)^{m-1}$$
$$= \left(\frac{p^r}{(1-p)^{r-1}}\right)\left(\frac{(1-p_0)^{r-1}}{p_0^r}\right)\sum_{m=r}^{\infty}\binom{m-1}{r-1}p_0^r(1-p_0)^{m-r}.$$

The terms in the last sum are from the PDF of a negative binomial random variable with parameters r and p_0, summed over all possible values of such a random variable. Hence, the sum is 1. Thus,

$$P(Y = k) = \left(\frac{p^r}{(1-p)^{r-1}}\right)\left(\frac{(1-p_0)^{r-1}}{p_0^r}\right)$$
$$= \left(\frac{p}{p_0}\right)\left(\frac{p(1-p_0)}{p_0(1-p)}\right)^{r-1} = \left(\frac{p}{p_0}\right)\left(\frac{p(1-p_0)}{p_0(1-p)}\right)^{k}.$$

Simple algebra shows that

$$\frac{p}{p_0} = \frac{1}{2-p} \qquad \text{and} \qquad \frac{p(1-p_0)}{p_0(1-p)} = 1 - \frac{1}{2-p}.$$

Consequently,

$$p_Y(y) = \left(\frac{1}{2-p}\right)\left(1 - \frac{1}{2-p}\right)^y, \qquad y = 0, 1, 2, \ldots,$$

and $p_Y(y) = 0$ otherwise.

b) Let $Z = Y + 1$. We note that the function $g(y) = y + 1$ is one-to-one. Solving for y in the equation $z = y + 1$, we see that $g^{-1}(z) = z - 1$. Applying Equation (5.54) on page 248 and the result of part (a) yields, for $z \in \mathcal{N}$,

$$p_Z(z) = p_{g(Y)}(z) = p_Y\big(g^{-1}(z)\big) = p_Y(z-1) = \left(\frac{1}{2-p}\right)\left(1 - \frac{1}{2-p}\right)^{z-1}.$$

Thus, $Y + 1 \sim \mathcal{G}\big(1/(2-p)\big)$.

5.171

a) We consider a "success" to be a defective item. From independence, we know that repeated observations of success (defective) or failure (nondefective) constitute Bernoulli trials with success probability p. Hence, the time (item number) of the kth success (defective item) has the negative binomial distribution with parameters k and p. Consequently, the probability that the nth item is the kth defective one is

$$\binom{n-1}{k-1} p^k (1-p)^{n-k}, \qquad n = k, k+1, \ldots.$$

b) Here we consider a "success" to be a rejected item (i.e., defective and inspected). From independence, we know that repeated observations of success or failure constitute Bernoulli trials with success probability pq. Hence, the time (item number) of the kth success (rejected item) has the negative binomial distribution with parameters k and pq. Consequently, the probability that the nth item is the kth rejected one is

$$\binom{n-1}{k-1} (pq)^k (1-pq)^{n-k}, \qquad n = k, k+1, \ldots.$$

c) We give two solutions.

Method 1: Let Z denote the number of items until the first rejection. From the law of partitions, we have, for each nonnegative integer k, that

$$P(W = k) = \sum_n P(W = k, Z = n) = \sum_{n=k+1}^{\infty} P(W = k, Z = n).$$

Event $\{W = k,\ Z = n\}$ occurs if and only if among the first $n-1$ items exactly k are defective, none of these k defective items are inspected, and the nth item is inspected and defective. The probability of this happening is

$$P(W = k, Z = n) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \cdot (1-q)^k \cdot pq$$
$$= q(1-q)^k \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k}.$$

Therefore, letting $r = k+1$, we get

$$P(W = k) = q(1-q)^k \sum_{n=k+1}^{\infty} \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k}$$
$$= q(1-q)^k \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = q(1-q)^k,$$

where, in the last equation, we used the fact that the values of the PMF of a negative binomial random variable with parameters r and p must sum to 1. Consequently,

$$p_W(w) = q(1-q)^w, \qquad w = 0, 1, 2, \ldots,$$

and $p_W(w) = 0$ otherwise.

Method 2: Let k be a nonnegative integer. The number of defective items until the first rejection equals k if and only if the first k defective items are not inspected and the next defective item is inspected, which has probability $(1-q)^k q$. Consequently,

$$p_W(w) = q(1-q)^w, \qquad w = 0, 1, 2, \ldots,$$

and $p_W(w) = 0$ otherwise.
d) Let $Z = W + 1$. We note that the function $g(w) = w + 1$ is one-to-one. Solving for w in the equation $z = w + 1$, we see that $g^{-1}(z) = z - 1$. Applying Equation (5.54) on page 248 and the result of part (c) yields, for $z \in \mathcal{N}$,

$$p_Z(z) = p_{g(W)}(z) = p_W\big(g^{-1}(z)\big) = p_W(z-1) = q(1-q)^{z-1}.$$

Thus, $W + 1 \sim \mathcal{G}(q)$ and represent the number of defective items up to and including the first rejection.
e) In the preceding (first) scenario, a defective item is rejected if and only if it is inspected, which has probability q. Here, in the second scenario, a defective item is rejected if and only if it is inspected and detected as defective, which has probability qr. Consequently, in the second scenario, the answers to parts (b)–(d) are obtained from those in the first scenario by replacing q by qr.

5.172 Let X denote the number of 6s in four throws of one balanced die and let Y denote the number of pairs of 6s in 24 throws of two balanced dice. Then $X \sim \mathcal{B}(4, 1/6)$ and $Y \sim \mathcal{B}(24, 1/36)$.
a) From the complementation rule,

$$P(X \ge 1) = 1 - P(X = 0) = 1 - \binom{4}{0}(1/6)^0(5/6)^{4-0} = 1 - (5/6)^4 = 0.518$$

and

$$P(Y \ge 1) = 1 - P(Y = 0) = 1 - \binom{24}{0}(1/36)^0(35/36)^{24-0} = 1 - (35/36)^{24} = 0.491.$$

Hence, the gambler is not right. The chance of getting at least one 6 when one balanced die is thrown four times exceeds that of getting at least one pair of 6s when two balanced dice are thrown 24 times.

b) From the complementation rule,

$$P(X \geq 2) = 1 - P(X < 2) = 1 - \sum_{k=0}^{1} \binom{4}{k} (1/6)^k (5/6)^{4-k} = 0.132$$

and

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - \sum_{k=0}^{1} \binom{24}{k} (1/36)^k (35/36)^{24-k} = 0.143.$$

5.173 Let C and D denote the events that the coin and die experiments are chosen, respectively. Given that D occurs, $X \sim \mathcal{G}(1/6)$, whereas, given that C occurs, $X \sim \mathcal{G}(1/2)$. Also, $P(C) = P(D) = 1/2$.
a) From the lack-of-memory property of the geometric distribution,

$$P(X = 8 \,|\, \{X > 5\} \cap D) = (1/6)(5/6)^{3-1} = \frac{25}{216}.$$

b) From the lack-of-memory property of the geometric distribution,

$$P(X = 8 \,|\, \{X > 5\} \cap C) = (1/2)(1/2)^{3-1} = \frac{1}{8}.$$

c) From Bayes's rule

$$\begin{aligned} P(D \,|\, X > 5) &= \frac{P(D)P(X > 5 \,|\, D)}{P(C)P(X > 5 \,|\, C) + P(D)P(X > 5 \,|\, D)} \\ &= \frac{(1/2)P(X > 5 \,|\, D)}{(1/2)P(X > 5 \,|\, C) + (1/2)P(X > 5 \,|\, D)} = \frac{1}{1 + \dfrac{P(X > 5 \,|\, C)}{P(X > 5 \,|\, D)}}. \end{aligned}$$

Now, by Proposition 5.10 on page 231,

$$P(X > 5 \,|\, D) = (5/6)^5 \qquad \text{and} \qquad P(X > 5 \,|\, C) = (1/2)^5.$$

Hence,

$$P(D \,|\, X > 5) = \frac{1}{1 + \dfrac{(1/2)^5}{(5/6)^5}} = \frac{3125}{3368}.$$

d) From the conditional form of the complementation rule and part (c),

$$P(C \,|\, X > 5) = 1 - P(D \,|\, X > 5) = 1 - \frac{3125}{3368} = \frac{243}{3368}.$$

e) Applying the conditional form of the law of total probability, Exercise 4.20, and the results of parts (a)–(d), we get

$$\begin{aligned} &P(X = 8 \,|\, X > 5) \\ &\quad = P_{\{X>5\}}(X = 8) = P_{\{X>5\}}(C)P_{\{X>5\}}(X = 8 \,|\, C) + P_{\{X>5\}}(D)P_{\{X>5\}}(X = 8 \,|\, D) \\ &\quad = P(C \,|\, X > 5)P(X = 8 \,|\, \{X > 5\} \cap C) + P(D \,|\, X > 5)P(X = 8 \,|\, \{X > 5\} \cap D) \\ &\quad = \frac{243}{3368} \cdot \frac{1}{8} + \frac{3125}{3368} \cdot \frac{25}{216} = 0.116. \end{aligned}$$

f) From the law of total probability,

$$\begin{aligned} P(X=3) &= P(C)P(X=3 \mid C) + P(D)P(X=3 \mid D) \\ &= \frac{1}{2} \cdot \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 + \frac{1}{2} \cdot \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = 0.120. \end{aligned}$$

g) From parts (e) and (f),

$$P(X = 5 + 3 \mid X > 5) = P(X = 8 \mid X > 5) = 0.116 \neq 0.120 = P(X = 3).$$

Therefore, X does not have the lack-of-memory property.
h) If five failures have occurred (i.e., the first five trials resulted in failure), then it is more likely that the die experiment was performed than the coin experiment, because the success probability in the former is much smaller than that in the latter ($1/6$ vs. $1/2$). Therefore, you should feel less sure of a success on the sixth trial, given these five failures, than you were of a success on the first trial, when both experiments were equally likely to be performed.
i) Knowing that the die experiment has been chosen, $X \sim \mathcal{G}(1/6)$. In particular, then, X has the lack-of-memory property. Therefore, the probability of a success on the sixth trial, given these five failures, is equal to the probability of a success on the first trial.

5.174 Let F and B denote the events that the balanced and biased coin is in your hand, respectively. We use P_1 and P_2 for probabilities involving the first and second experiments, respectively. For a coin with probability p of a head, let X denote the number of heads in six tosses and let Y denote the number of tosses until the third head. Then $X \sim \mathcal{B}(6, p)$ and $Y \sim \mathcal{NB}(3, p)$.
a) We note that, for the first experiment, event E occurs if and only if exactly three heads occur in the six tosses. Hence,

$$P_1(E \mid F) = \binom{6}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3 = \frac{5}{16} \qquad \text{and} \qquad P_1(E \mid B) = \binom{6}{3}\left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)^3 = \frac{256}{3125}.$$

b) We note that, for the second experiment, event E occurs if and only if the third head occurs on the sixth toss. Hence,

$$P_2(E \mid F) = \binom{6-1}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3 = \frac{5}{32} \qquad \text{and} \qquad P_2(E \mid B) = \binom{6-1}{2}\left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)^3 = \frac{128}{3125}.$$

c) From Bayes's rule and part (a),

$$P_1(F \mid E) = \frac{P_1(F)P_1(E \mid F)}{P_1(F)P_1(E \mid F) + P_1(B)P_1(E \mid B)} = \frac{(0.4) \cdot \frac{5}{16}}{(0.4) \cdot \frac{5}{16} + (0.6) \cdot \frac{256}{3125}} = \frac{15{,}625}{21{,}769}.$$

d) From Bayes's rule and part (b),

$$P_2(F \mid E) = \frac{P_2(F)P_2(E \mid F)}{P_2(F)P_2(E \mid F) + P_2(B)P_2(E \mid B)} = \frac{(0.4) \cdot \frac{5}{32}}{(0.4) \cdot \frac{5}{32} + (0.6) \cdot \frac{128}{3125}} = \frac{15{,}625}{21{,}769}.$$

e) As we see, the answers to parts (c) and (d) are identical. This equality is a result of the fact that $P_2(E \mid F) = P_1(E \mid F)/2$ and $P_2(E \mid B) = P_1(E \mid B)/2$. Consequently, when we apply Bayes's rule in part (d), we get the same answer as in part (c).

BETA

Chapter 6

Jointly Discrete Random Variables

6.1 Joint and Marginal Probability Mass Functions: Bivariate Case

Basic Exercises

6.1 a) $P(X = Y) = \sum_x P(X = Y = x) = \sum_{x=1}^{\infty} p^2(1-p)^{2x-2} = p^2 \sum_{x=0}^{\infty}((1-p)^2)^x = p^2 \times \frac{1}{1-(1-p)^2} = \frac{p}{2-p}$, which represents the probability that both electrical components have the same lifetimes.
b) $(1-p)/(2-p)$, since

$$P(X > Y) = \sum_x P(X = x, Y < x) = \sum_x P(X = x, Y \leq x-1) = \sum_{x=2}^{\infty}\sum_{y=1}^{x-1} P(X = x, Y = y)$$

$$= \sum_{x=2}^{\infty}\sum_{y=1}^{x-1} p^2(1-p)^{x+y-2} = p^2 \sum_{x=2}^{\infty}(1-p)^{x-1}\left(\sum_{y=0}^{x-2}(1-p)^y\right)$$

$$= p^2 \sum_{x=2}^{\infty}(1-p)^{x-1} \cdot \frac{1-(1-p)^{x-1}}{1-(1-p)} = \frac{1-p}{2-p}.$$

This represents the probability that the first component outlasts the second.
c) Because the joint PMF of X and Y is symmetric in x and y (since $P(X = x, Y = y) = P(X = y, Y = x)$ for all x, y) we must have $P(Y > X) = P(X > Y)$.
d) Since $1 = P(\Omega) = P(X > Y) + P(X = Y) + P(X < Y) = 2P(X > Y) + \frac{p}{2-p}$, then

$$1 - \frac{p}{2-p} = 2P(X > Y), \text{ implying that } \frac{1-p}{2-p} = P(X > Y).$$

6.2 a) $p_{X,Y}(2,1) = P(X = 2, Y = 1) = \dfrac{N(\{X = 2, Y = 1\})}{N(\Omega)} = \dfrac{8}{40} = 0.2$. The probability that a randomly selected student has exactly 2 siblings, where one of the siblings is a sister and the other is a brother.
b) The joint PMF of X and Y is given by: $p_{X,Y}(0,0) = p_{X,Y}(1,0) = p_{X,Y}(2,1) = 0.2,$

$p_{X,Y}(1,1) = 0.225$, $p_{X,Y}(2,0) = p_{X,Y}(3,1) = p_{X,Y}(4,3) = 0.025$, $p_{X,Y}(2,2) = p_{X,Y}(3,2) = 0.05$, and $p_{X,Y}(x,y) = 0$ for all other x, y.

c)

Siblings, X \ Sisters, Y	0	1	2	3	$p_X(x)$
0	0.2	0	0	0	0.2
1	0.2	0.225	0	0	0.425
2	0.025	0.2	0.05	0	0.275
3	0.0	0.025	0.05	0	0.075
4	0	0	0	0.025	0.025
$p_Y(y)$	0.425	0.45	0.1	0.025	1

d) Using the law of partitions (Proposition 2.8), we have for each x:

$$p_X(x) = P(X = x) = \sum_{k=0}^{3} P(X = x, Y = k) = p_{X,Y}(x,0) + p_{X,Y}(x,1) + p_{X,Y}(x,2) + p_{X,Y}(x,3).$$

e) If the student has no brothers, then the number of siblings, that the student has, should equal to the number of sisters the student has:

$$P(X = Y) = p_{X,Y}(0,0) + p_{X,Y}(1,1) + p_{X,Y}(2,2) + p_{X,Y}(3,3) = 0.475$$

f) The probability that the student has no sisters is : $P(Y = 0) = p_Y(0) = 0.425$

6.3 a) Since X is chosen uniformly from the 10 decimal digits, $P(X = x) = \frac{1}{10}$ for all $x \in \{0, 1, \cdots, 9\}$ and $P(X = x) = 0$ otherwise. Once X has been chosen, Y is uniformly chosen from the remaining 9 decimal digits, thus $P(Y = y \,|\, X = x) = \frac{1}{9}$ for $y \in \{0, 1, \cdots, 9\} \setminus \{x\}$ and $P(Y = y \,|\, X = x) = 0$ otherwise. Thus, upon using the multiplication rule, one obtains that $P(X = x, Y = y) = P(Y = y|X = x)P(X = x) = \frac{1}{90}$ for all $x, y \in \{0, 1, \cdots, 9\}$ such that $x \neq y$, and $P(X = x, Y = y) = 0$ otherwise.
b) Using the FPF, $P(X = Y) = \sum_{i=0}^{9} P(X = Y = i) = 0$. However, the result is easily obtained upon realizing that by definition of Y, Y can never equal X and thus $P(X = Y) = 0$.
c) Using the FPF, $P(X > Y) = \sum_{i=1}^{9} \sum_{j=0}^{i-1} P(X = i, Y = j) = \sum_{i=1}^{9} \sum_{j=0}^{i-1} (1/90) = \sum_{i=1}^{9} (i/90) = \frac{1}{2}$. The result is easy to obtain directly from $P(X = Y) = 0$ and symmetry of the distribution of X, Y:

$$1 = P(\Omega) = P(X > Y) + P(X < Y) = 2P(X > Y), \text{ thus, } \; P(X > Y) = \frac{1}{2}.$$

d) Using Proposition 6.2, $P(Y = y) = \sum_{x \in \{0,1,\cdots,9\} \setminus \{y\}} P(X = x, Y = y) = \frac{9}{90} = \frac{1}{10}$.
e) Using symmetry, each y is equally likely to be chosen, and thus $P(Y = y) = \frac{1}{10}$.

6.4 a)

I_F \ I_E	0	1	$p_{I_F}(x)$
0	$P(E^c \cap F^c)$	$P(E \cap F^c)$	$P(F^c)$
1	$P(E^c \cap F)$	$P(E \cap F)$	$P(F)$
p_{I_E}	$P(E^c)$	$P(E)$	1

b) Since E and E^c partition Ω (as do F and F^c), the law of partitions (Proposition 2.8) and definition of indicator random variables yield the desired result.
c) An indicator random variable indicates whether a specified event occurs. Knowledge of the probabilities of two different events occurring does not in general provide any knowledge of the probability of both events occurring.
d) The joint PMF of two indicator random variables can be determined by the marginal PMFs only when the two events in question are either mutually exclusive or independent.

6.5 By partitioning, for arbitrary $x_0 \in \mathcal{R}$ for discrete random variables X and Y we have that

$$p_X(x_0) = P(X = x_0) = P(X = x_0, Y \in \mathcal{R}) = \sum_{x=x_0} \sum_{y \in \mathcal{R}} P(X = x, Y = y) = \sum_{y \in \mathcal{R}} p_{X,Y}(x_0, y),$$

which is Equation (6.3). Similarly, for arbitrary $y_0 \in \mathcal{R}$,

$$p_Y(y_0) = P(Y = y_0) = P(X \in \mathcal{R}, Y = y_0) = \sum_{x \in \mathcal{R}} \sum_{y=y_0} P(X = x, Y = y) = \sum_{x \in \mathcal{R}} p_{X,Y}(x, y_0)$$

which is Equation (6.4).

6.6 a) The joint PMF of X and Y is given by: $p_{X,Y}(x, y) = \frac{1}{36}$ for $x = y \in \{1, \cdots, 6\}$, and $p_{X,Y}(x, y) = \frac{2}{36}$ for all $x < y$, where $x, y \in \{1, \cdots, 6\}$, and $p_{X,Y}(x, y) = 0$ otherwise.
b) The required table is given by:

		Larger value, Y						
		1	2	3	4	5	6	$p_X(x)$
	1	0.0278	0.0556	0.0556	0.0556	0.0556	0.0556	0.3058
Smaller value, X	2	0	0.0278	0.0556	0.0556	0.0556	0.0556	0.2502
	3	0	0	0.0278	0.0556	0.0556	0.0556	0.1946
	4	0	0	0	0.0278	0.0556	0.0556	0.1390
	5	0	0	0	0	0.0278	0.0556	0.0834
	6	0	0	0	0	0	0.0278	0.0278
	$p_Y(y)$	0.0278	0.0834	0.1390	0.1946	0.2502	0.3058	1

c) $P(X = Y) = p_{X,Y}(1, 1) + p_{X,Y}(2, 2) + \ldots + p_{X,Y}(6, 6) = 0.1667$, which is the probability that both times the die comes up the same number.
d) $P(X < Y) = 1 - P(X = Y) = 0.8333$, which is the probability that the value of the larger of the two faces showing is strictly larger than the value of the smaller of the two faces showing. The solution is easily computed since $P(X \leq Y) = 1$.
e) $P(X > Y) = 0$. By definition of the variables X and Y, X can not be strictly larger than Y.
f) $P(Y = 2X) = p_{X,Y}(1, 2) + p_{X,Y}(2, 4) + p_{X,Y}(3, 6) = 0.1667$, which is the probability that the larger value of the two faces showing is exactly twice the value of the smaller of the two faces.
g) $P(Y \geq 2X) = p_{X,Y}(1, 2) + p_{X,Y}(1, 3) + p_{X,Y}(1, 4) + p_{X,Y}(1, 5) + p_{X,Y}(1, 6) + p_{X,Y}(2, 4) + p_{X,Y}(2, 5) + p_{X,Y}(2, 6) + p_{X,Y}(3, 6) = 0.5$, which represents the probability that the larger of the two faces is at least twice the value of the smaller of the two faces showing.
h) $P(Y - X \geq 3) = p_{X,Y}(1, 4) + p_{X,Y}(1, 5) + p_{X,Y}(1, 6) + p_{X,Y}(2, 5) + p_{X,Y}(2, 6) + p_{X,Y}(3, 6) = 0.3333$, which represents the probability that the difference between the larger of the two faces and the smaller of the two faces is at least 3.

6.7 a)

$$p_{X,Y}(x,y) = \binom{3}{x,\, y,\, 3-x-y}\left(\frac{1}{4}\right)^x\left(\frac{1}{4}\right)^y\left(\frac{1}{2}\right)^{3-x-y}$$

for integer $0 \le x, y \le 3$ where $x + y \le 3$, and $p_{X,Y}(x,y) = 0$ otherwise. The result is obtained upon noting that there are $\binom{3}{x,y,3-x-y}$ ways to select the order (of political affiliations) in which x Greens, y Democrats and $(3-x-y)$ Republicans are chosen into the sample of size 3 without replacement. Moreover, the probability of every such ordering is the same and equals to $(\frac{2}{8})^x(\frac{2}{8})^y(\frac{4}{8})^{3-x-y}$.

b)

Green, X \ Democrats, Y	0	1	2	3	$p_X(x)$
0	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{3}{32}$	$\frac{1}{64}$	$\frac{27}{64}$
1	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{3}{64}$	0	$\frac{27}{64}$
2	$\frac{3}{32}$	$\frac{3}{64}$	0	0	$\frac{9}{64}$
3	$\frac{1}{64}$	0	0	0	$\frac{1}{64}$
$p_Y(y)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$	1

c) Using the FPF, $P(X > Y) = p_{X,Y}(1,0) + p_{X,Y}(2,0) + p_{X,Y}(3,0) + p_{X,Y}(2,1) = \frac{3}{16} + \frac{3}{32} + \frac{1}{64} + \frac{3}{64} = \frac{11}{32}$, i.e. the probability, that there are more Greens than Democrats in the sample, is equal to $11/32$.

d) If no Republicans are chosen, then every person in the sample must either be a Green or a Democrat, thus the event that no Republicans are chosen is represented by $\{X + Y = 3\}$.

e) Using the FPF and (a), $P(X+Y=3) = p_{X,Y}(3,0) + p_{X,Y}(2,1) + p_{X,Y}(1,2) + p_{X,Y}(0,3) = \frac{1}{64} + \frac{3}{64} + \frac{3}{64} + \frac{1}{64} = \frac{1}{8}$.

f) Let Z be the number of Republicans in the sample. Then $Z \sim \mathcal{B}\left(3, \frac{1}{2}\right)$, and

$$P(Z=0) = \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

g) $X \sim \mathcal{B}\left(3, \frac{1}{4}\right)$, i.e. $p_X(x) = \binom{3}{x}(0.25)^x(0.75)^{3-x}$ for $x \in \{0,1,2,3\}$, and $p_X(x) = 0$ for all $x \notin \{0,1,2,3\}$.

h) $X + Y \sim \mathcal{B}\left(3, \frac{1}{2}\right)$, i.e. $p_{X+Y}(u) = \binom{3}{u}/8$ for all $u \in \{0,1,2,3\}$, and $p_{X+Y}(u) = 0$ otherwise.

i) Using the FPF and (a),

$$\begin{aligned} P(1 \le X+Y \le 2) &= p_{X,Y}(1,0) + p_{X,Y}(2,0) + p_{X,Y}(0,1) + p_{X,Y}(1,1) + p_{X,Y}(0,2) \\ &= \frac{3}{16} + \frac{3}{32} + \frac{3}{16} + \frac{3}{16} + \frac{3}{32} = \frac{3}{4} \end{aligned}$$

j) Using (h),

$$P(1 \le X+Y \le 2) = P(X+Y=1) + P(X+Y=2) = \binom{3}{1}\frac{1}{8} + \binom{3}{2}\frac{1}{8} = \frac{3}{4}.$$

6.8 a) The joint PMF of X and Y is given by:

$$p_{X,Y}(x,y) = \frac{\binom{4}{3-(x+y)}\binom{2}{x}\binom{2}{y}}{\binom{8}{3}}$$

for $(x,y) \in \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2),(2,1)\}$, and $p_{X,Y}(x,y) = 0$ otherwise.

b)

		Democrats, Y			
		0	1	2	$p_X(x)$
Green, X	0	0.0714	0.2143	0.0714	0.3571
	1	0.2143	0.2857	0.0357	0.5357
	2	0.0714	0.0357	0	0.1071
	$p_Y(y)$	0.3571	0.5357	0.1071	1

c) $P(X > Y) = p_{X,Y}(1,0) + p_{X,Y}(2,0) + p_{X,Y}(2,1) = 0.3214$, which represents the probability that more Greens are chosen than Democrats.
d) $\{X + Y = 3\}$.
e) Let E represent the event that no Republicans are chosen. Then, by (d), $E = \{(1,2),(2,1)\}$. Using the FPF and (a), $P(E) = p_{X,Y}(1,2) + p_{X,Y}(2,1) = 0.0714$.
f) Let R represent the number of Republicans in the sample.
Then R is $\mathcal{H}(N = 8, n = 3, p = 0.5)$, i.e. the probability distribution of R is given by:

$$p_R(r) = \frac{\binom{4}{r}\binom{4}{3-r}}{\binom{8}{3}}, \quad r \in \{0,1,2,3\}.$$

Thus, $p_R(0) = \frac{4}{56} \approx 0.0714$
g) X is $\mathcal{H}(N = 8, n = 3, p = 1/4)$ and the PMF of X is given by:

$$p_X(x) = \frac{\binom{2}{x}\binom{6}{3-x}}{\binom{8}{3}}$$

for $x = 0,1,2$, $p_X(x) = 0$ for all other values.
h) $X + Y$ is $\mathcal{H}(N = 8, n = 3, p = 0.5)$. The PMF for $X + Y$ is given as:

$$p_{X+Y}(z) = \frac{\binom{4}{z}\binom{4}{3-z}}{\binom{8}{3}}$$

for $z = 0,1,2,3$, $p_{X+Y}(z) = 0$ for all other values.
i) Let F represent the event that the sum of the total number of Green and Democrats chosen are either 1 or 2. Then, $F = \{(0,1),(0,2),(1,0),(1,1),(2,0)\}$. Using the FPF and part (a), $P(1 \le X + Y \le 2) = p_{X,Y}(0,1) + p_{X,Y}(0,2) + p_{X,Y}(1,0) + p_{X,Y}(1,1) + p_{X,Y}(2,0) = 0.8571$

j) Using (h), $P(1 \le X + Y \le 2) = p_{X+Y}(1) + p_{X+Y}(2) = \frac{\binom{4}{1}\binom{4}{2}}{\binom{8}{3}} + \frac{\binom{4}{2}\binom{4}{1}}{\binom{8}{3}} \approx 0.8571$

6.9 a) Using Proposition 5.9, $X \sim \mathcal{G}(p)$.
b) Using Proposition 5.13, $Y \sim \mathcal{NB}(2,p)$.

c) $p_{X,Y}(x,y) = p^2(1-p)^{y-2}$ for $x, y \in \mathcal{N}$, $x < y$ and $p_{X,Y}(x,y) = 0$ otherwise. The answer follows from

$$P(X=x, Y=y) = P(\underbrace{F...F}_{x-1} S \underbrace{F...F}_{y-x-1} S) = p^2(1-p)^{x-1+(y-x-1)} = p^2(1-p)^{y-2}, \quad x < y; x, y \in \mathcal{N}.$$

d) For X, using Proposition 6.2, for $x \in \mathcal{N}$,

$$P(X=x) = \sum_y P(X=x, Y=y) = \sum_{y=x+1}^{\infty} p^2(1-p)^{y-2}$$

$$= p^2(1-p)^{x-1} \sum_{y=0}^{\infty} (1-p)^y = \frac{p^2(1-p)^{x-1}}{p} = p(1-p)^{x-1}$$

and $p_X(x) = 0$ otherwise, which agrees with (a).
For Y, using Proposition 6.2, for $y = 2, 3 \ldots$,

$$P(Y=y) = \sum_x P(X=x, Y=y) = \sum_{x=1}^{y-1} p^2(1-p)^{y-2} = p^2(1-p)^{y-2} \sum_{x=1}^{y-1} 1$$

$$= (y-1)p^2(1-p)^{y-2} = \binom{y-1}{2-1} p^2(1-p)^{y-2}$$

and $p_Y(y) = 0$ otherwise, which agrees with (b).
e) $Y - X$ is the number of trials between the first and the second success.
f) Once the first success has occurred, since each individual trial is independent, the number of additional trials needed before the second success will have the same distribution as the number of trials needed before the first success. thus $Y - X \sim \mathcal{G}(p)$
g) For $z \in \mathcal{N}$

$$P(Y-X=z) = \sum_x P(X=x, Y=x+z) = \sum_{x=1}^{\infty} p^2(1-p)^{x+z-2} = p^2(1-p)^{z-2} \sum_{x=1}^{\infty} (1-p)^x$$

$$= p^2(1-p)^{z-1} \sum_{x=0}^{\infty} (1-p)^x = \frac{p^2(1-p)^{z-1}}{1-(1-p)} = \frac{p^2(1-p)^{z-1}}{p} = p(1-p)^{z-1}$$

and $P(Y - X = z) = 0$ otherwise, which agrees with (f).

6.10 a) The sample space for the head-tail outcomes in three tosses of the coin is:
$\{(H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)\}$
b) The joint PMF of X and Y is given by:

Total winnings after 3^{rd} toss, Y \ Total winnings after 2^{nd} toss, X	-2	0	2
-3	0.125	0	0
-1	0.125	0.25	0
1	0	0.25	0.125
3	0	0	0.125

c) Let E be the event $X > Y$. Then $E = \{(-2,-3),(0,-1),(2,1)\}$. Using **a)** and the FPF, $P(E) = p_{X,Y}(-2,-3) + p_{X,Y}(0,-1) + p_{X,Y}(2,1) = 0.5$
d) First, we note that X can never equal Y. Thus $X > Y$ or $X < Y$. Now, once we have performed the first two tosses, the value of X is fixed, and Y has a 50-50 chance of being \$1 larger or \$1 smaller than X depending on the result of the third coin. Thus $P(X > Y) = P(X < Y) = 0.5$.
e) The PMF of X is given as follows:

x	-2	0	2
$p_X(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

f) $X =\#$ of heads $-\#$ of tails$= Z - (2 - Z) = -2 + 2Z$, where Z is the number of heads in the first two tosses. Since $Z \sim \mathcal{B}(2, 0.5)$, it follows that $p_X(x) = \binom{2}{1+x/2}\frac{1}{4}$ if $x = -2, 0, 2$, and $p_X(x) = 0$ otherwise.
g) The PMF of Y is given as follows:

y	-3	-1	1	3
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

h) $Y = W-(3-W) = -3+2W$, where W is the number of heads in the first three tosses. Since $W \sim \mathcal{B}(3, 0.5)$, it follows that $p_Y(y) = \binom{3}{(3+y)/2}\frac{1}{8}$ if $y = -3, -1, 1, 3$, and $p_Y(y) = 0$ otherwise.
i) $p_{Y-X}(-1) = p_{Y-X}(1) = 0.5$
j) $p_{Y-X}(-1) = p_{X,Y}(-2,-3) + p_{X,Y}(0,-1) + p_{X,Y}(2,1) = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = 0.5$
$p_{Y-X}(1) = p_{X,Y}(-2,-1) + p_{X,Y}(0,1) + p_{X,Y}(2,3) = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = 0.5$

6.11a) The sample space for the head-tail outcomes in three tosses of the coin is:
$\{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}$
b) The joint PMF of X and Y is given by:

		Total \$ after 2^{nd} toss, X		
		-2	0	2
Total \$ after 3^{rd} toss, Y	-3	q^3	0	0
	-1	pq^2	$2pq^2$	0
	1	0	$2p^2q$	p^2q
	3	0	0	p^3

c) Let E be the event $X > Y$. Then $E = \{(-2,-3),(0,-1),(2,1)\}$. Using **a)** and the FPF, $P(E) = p_{X,Y}(-2,-3) + p_{X,Y}(0,-1) + p_{X,Y}(2,1) = q^3 + 2pq^2 + p^2q = q(q^2 + 2pq + p^2) = q(q+p)^2 = q(1)^2 = q$
d) Event $\{X > Y\}$ means that the third toss is a tail. Thus $P(X > Y) = q$.
e) The PMF of X is given as follows:

x	-2	0	2
$p_X(x)$	q^2	$2pq$	p^2

f) $X = -2 + 2Z$ where Z is the number of heads in the first two tosses. Since $Z \sim \mathcal{B}(2, p)$, it follows that $p_X(x) = \binom{2}{1+x/2}p^{1+x/2}q^{1-x/2}$ if $x = -2, 0, 2$, and $p_X(x) = 0$ otherwise.
g) The PMF of Y is given as follows:

y	-3	-1	1	3
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

h) $Y = -3 + 2W$ where W is the number of heads in the first three tosses. Since $W \sim \mathcal{B}(3, p)$, it follows that $p_Y(y) = \binom{3}{(3+y)/2} p^{(3+y)/2} q^{(3-y)/2}$ if $y = -3, -1, 1, 3$, and $p_Y(y) = 0$ otherwise.
i) $p_{Y-X}(-1) = q$, $p_{Y-X}(1) = p$
j) $p_{Y-X}(-1) = p_{X,Y}(-2,-3) + p_{X,Y}(0,-1) + p_{X,Y}(2,1) = q^3 + 2pq^2 + p^2 q = q(q^2 + 2pq + p^2) = q(q+p)^2 = q(1)^2 = q$
$p_{Y-X}(1) = p_{X,Y}(-2,-1) + p_{X,Y}(0,1) + p_{X,Y}(2,3) = pq^2 + 2p^2 q + p^3 = p(q^2 + 2pq + p^2) = p(q+p)^2 = p$

6.12 a) The sample space for the head-tail outcomes in four tosses of the coin is:
$\{(H,H,H,H), (H,H,H,T), (H,H,T,H), (H,H,T,T), (H,T,H,H), (H,T,H,T), (H,T,T,H), (H,T,T,T), (T,H,H,H), (T,H,H,T), (T,H,T,H), (T,H,T,T), (T,T,H,H), (T,T,H,T), (T,T,T,H), (T,T,T,T)\}$
b) The joint PMF of X and Y is given by:

		\$ after 3^{rd} toss, X			
		-3	-1	1	3
\$ after 4^{th} toss, Y	-4	$\frac{1}{16}$	0	0	0
	-2	$\frac{1}{16}$	$\frac{3}{16}$	0	0
	0	$\frac{3}{16}$	$\frac{3}{16}$	0	0
	2	0	0	$\frac{3}{16}$	$\frac{1}{16}$
	4	0	0	0	$\frac{1}{16}$

c) Let E be the event $X > Y$. Then $E = \{(-3,-4), (-1,-2), (1,0), (3,2)\}$. Using **a)** and the FPF, $P(E) = p_{X,Y}(-3,-4) + p_{X,Y}(-1,-2) + p_{X,Y}(1,0) + p_{X,Y}(3,2) = 0.5$
d) Again, first, we note that X can never equal Y. Thus $X > Y$ or $X < Y$. Now, once we have performed the first three tosses, the value of X is fixed, and Y has a 50-50 chance of being \$1 larger or \$1 smaller than X depending on the result of the fourth coin. Thus $P(X > Y) = P(X < Y) = 0.5$.
e) The PMF of X is given as follows:

x	-3	-1	1	3
$p_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

f) $X = -3 + 2Z$ where Z is the number of heads in the first three tosses. Since $Z \sim \mathcal{B}(3, 0.5)$, it follows that $p_X(x) = \binom{3}{(3+x)/2} \frac{1}{8}$ if $y = -3, -1, 1, 3$, and $p_Y(y) = 0$ otherwise.
g) The PMF of Y is given as follows:

y	-4	-2	0	2	4
$p_Y(y)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

h) $Y = -4 + 2W$ where W is the number of heads in the first four tosses. Since $W \sim \mathcal{B}(4, 0.5)$, it follows that $p_Y(y) = \binom{4}{2+y/2} \frac{1}{16}$ if $y = -4, -2, 0, 2, 4$, and $p_Y(y) = 0$ otherwise.
i) $p_{Y-X}(-1) = p_{Y-X}(1) = 0.5$
j) $p_{Y-X}(-1) = p_{X,Y}(-3,-4) + p_{X,Y}(-1,-2) + p_{X,Y}(1,0) + p_{X,Y}(3,2) = \frac{1}{16} + \frac{3}{16} + \frac{3}{16} + \frac{1}{16} = 0.5$
$p_{Y-X}(-1) = p_{X,Y}(-3,-2) + p_{X,Y}(-1,0) + p_{X,Y}(1,2) + p_{X,Y}(3,4) = \frac{1}{16} + \frac{3}{16} + \frac{3}{16} + \frac{1}{16} = 0.5$

6.13 a) Using Proposition 6.2, the marginal PMF of X is:

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_{y=0}^{\infty} e^{-(\lambda+\mu)} \frac{\lambda^x \mu^y}{x! y!}$$

$$= \frac{e^{-\lambda}\lambda^x}{x!} \underbrace{\sum_{y=0}^{\infty} \frac{e^{-\mu}\mu^y}{y!}}_{=1} = \frac{e^{-\lambda}\lambda^x}{x!}$$

for $x = 0, 1, \ldots$ and $P(X = x) = 0$ otherwise. Thus $X \sim \mathcal{P}(\lambda)$.
Using Proposition 6.2 and the same argument, the marginal PMF of Y is:

$$P(Y = y) = \sum_x P(X = x, Y = y) = \sum_{x=0}^{\infty} e^{-(\lambda+\mu)} \frac{\lambda^x \mu^y}{x!y!}$$

$$= \frac{e^{-\mu}\mu^y}{y!} \underbrace{\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}}_{=1} = \frac{e^{-\mu}\mu^y}{y!}$$

for $y = 0, 1, \ldots$ and $P(Y = y) = 0$ otherwise. Thus $Y \sim \mathcal{P}(\mu)$.
b) The joint PMF is the product of the two marginal PMF's.

6.14) Let $p : \mathcal{R}^2 \to \mathcal{R}$ be defined by $p(x, y) = p_X(x)p_Y(y)$. Since $p_X(x)$ and $p_Y(y)$ are both PMF's, $p_X(x) \geq 0$ for all $x \in \mathcal{R}$ and $p_Y(y) \geq 0$ for all $y \in \mathcal{R}$, thus $p(x, y) \geq 0$ for all $(x, y) \in \mathcal{R}^2$. For the second property, we note that for any $y \in \mathcal{R}$, if $p_Y(y) = 0$, then $p(x, y) = 0$. Therefore, $\{(x, y) \in \mathcal{R}^2 : p(x, y) \neq 0\} \subset \{y \in \mathcal{R} : p_Y(y) \neq 0\}$ and since the subset of any countable set is countable and $\{y \in \mathcal{R} : p_Y(y) \neq 0\}$ is countable since $p_Y(y)$ is a PMF, we must have that $\{(x, y) \in \mathcal{R}^2 : p(x, y) \neq 0\}$ is countable. Finally, since $p_X(x)$ and $p_Y(y)$ are both PMF's,

$$\sum\sum\nolimits_{(x,y)} p(x, y) = \sum\sum\nolimits_{(x,y)} p_X(x)p_Y(y) = \sum_x p_X(x) \sum_y p_Y(y) = 1$$

Thus, $p(x, y)$ satisfies the three properties and is a joint PMF.

6.15 a) Using Proposition 6.2, the PMF for X is:

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y q(x)r(y) = q(x) \sum_y r(y) \qquad \text{for } x, y \in \mathcal{R}$$

and, the PMF for Y is:

$$P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x q(x)r(y) = r(y) \sum_x q(x) \qquad \text{for } x, y \in \mathcal{R}$$

b) From Proposition 6.1 c), we have that $\sum\sum_{(x,y)} p_{X,Y}(x, y) = 1$ which implies that

$$\sum_x \sum_y p_{X,Y}(x, y) = \sum_x \sum_y q(x)r(y) = \sum_x q(x) \sum_y r(y) = 1$$

Thus we have

$$p_{X,Y}(x, y) = q(x)r(y) = q(x)r(y) \sum_x q(x) \sum_y r(y)$$

$$= \left(q(x) \sum_y r(y)\right) \left(r(y) \sum_x q(x)\right) = p_X(x)p_Y(y)$$

c) If $\sum_x q(x) = \sum_y r(y) = 1$ then q and r would be the marginal PMFs of X and Y respectively.

Theory Exercises

6.16 Clearly, $p_{X,Y}(x,y) = P(X = x, Y = y) \geq 0$ for all $x, y \in \mathcal{R}$. Next note that $p_X(x) = 0$ implies that $0 \leq p_{X,Y}(x,y) = P(\{X = x\} \cap \{Y = y\}) \leq P(X = x) = p_X(x) = 0$, thus, $\{(x,y) : p_X(x) = 0\} \subset \{(x,y) : p_{X,Y}(x,y) = 0\}$. Therefore, $\{(x,y) : p_{X,Y}(x,y) \neq 0\} \subset (\{(x,y) : p_X(x) \neq 0\})$. Similarly one shows: $\{(x,y) : p_{X,Y}(x,y) \neq 0\} \subset (\{(x,y) : p_Y(y) \neq 0\})$. Thus,

$$\{(x,y) : p_{X,Y}(x,y) \neq 0\} \subset \{(x,y) : p_X(x) \neq 0,\, p_Y(y) \neq 0\} = \{x : p_X(x) \neq 0\} \times \{y : p_Y(y) \neq 0\}.$$

Since the cartesian product of two countable sets is a countable set, and the subset of a countable set is a countable set, then $\{(x,y) : p_{X,Y}(x,y) \neq 0\}$ is countable. Finally note that, since X, Y are discrete, we have that

$$1 = P(\Omega) = P\left((X,Y) \in \{(x,y) : p_{X,Y}(x,y) \neq 0\}\right) = \sum_{(x,y):\, p_{X,Y}(x,y) \neq 0} p_{X,Y}(x,y) = \sum_{(x,y)} p_{X,Y}(x,y).$$

6.17 Suppose that (a) $p(x,y) \geq 0$ for all $x, y \in \mathcal{R}^2$; (b) $\{(x,y) \in \mathcal{R}^2 : p(x,y) \neq 0\}$ is countable; (c) $\sum_{(x,y)} p(x,y) = 1$. Let $\Omega = \{(x,y) \in \mathcal{R}^2 : p(x,y) \neq 0\}$. Then Ω is countable and, in view of conditions (a),(c) and Proposition 2.3, there exists a unique probability measure P such that $P(\{(\omega_1, \omega_2)\}) = p(\omega_1, \omega_2)$ for all $(\omega_1, \omega_2) \in \Omega$. Define $X((\omega_1, \omega_2)) = \omega_1$ and $Y((\omega_1, \omega_2)) = \omega_2$ for all $(\omega_1, \omega_2) \in \Omega$. Then X and Y are discrete random variables and

$$p_{X,Y}(x,y) = P(\{(\omega_1, \omega_2) \in \Omega : X(\omega_1, \omega_2) = x \text{ and } Y(\omega_1, \omega_2) = y\})$$

$$= P(\{(x,y)\}) = \begin{cases} p(x,y), & \text{if } (x,y) \in \Omega, \\ 0, & \text{if } (x,y) \notin \Omega. \end{cases}$$

On the other hand, note that $p(x,y) = 0$ for all $(x,y) \notin \Omega$, by definition of Ω. Therefore, $p_{X,Y}(x,y) = p(x,y)$ for all $(x,y) \in \mathcal{R}^2$.

6.18 By the law of partitioning, $\{X = x\} = \cup_y \{X = x, Y = y\}$ and hence, $P(X = x) = P(\cup_y (X = x, Y = y)) = \sum_y P(X = x, Y = y)$. Using the same logic, $\{Y = y\} = \cup_x \{X = x, Y = y\}$ and hence, $P(Y = y) = P(\cup_x \{X = x, Y = y\}) = \sum_x P(X = x, Y = y)$.

6.19 Note first that

$$P\left((X,Y) \in \{(x,y) : p_{X,Y}(x,y) > 0\}\right) = \sum_{(x,y):\, p_{X,Y}(x,y) > 0} p_{X,Y}(x,y) = \sum_{(x,y)} p_{X,Y}(x,y) = 1,$$

by Proposition 6.1(b),(c), and the law of partitioning. Therefore, for every set $A \subset \mathcal{R}^2$,

$$0 \leq P((X,Y) \in A \cap \{(x,y) : p_{X,Y}(x,y) = 0\}) \leq P((X,Y) \in \{(x,y) : p_{X,Y}(x,y) = 0\}) = 0.$$

Therefore, for arbitrary $A \subset \mathcal{R}^2$,

$$\begin{aligned} P((X,Y) \in A) &= P((X,Y) \in A \cap \{(x,y) : p_{X,Y}(x,y) = 0\}) \\ &+ P((X,Y) \in A \cap \{(x,y) : p_{X,Y}(x,y) > 0\}) \\ &= 0 + \sum_{(x,y) \in A:\, p_{X,Y}(x,y) > 0} p_{X,Y}(x,y) = \sum_{(x,y) \in A} p_{X,Y}(x,y). \end{aligned}$$

Advanced Exercises

6.20 a) Since the first number chosen is selected at random from the first N positive integers, $p(X = x) = \frac{1}{N}$ for $1 \leq x \leq N$ and $p(X = x) = 0$ otherwise. Given the first number $X = x$ has already been selected, the second number is selected at random from the first x positive integers, and thus $P(Y = y|X = x) = \frac{1}{x}$ for integer $1 \leq y \leq x$ and $P(Y = y|X = x) = 0$ otherwise. Using the general multiplication rule (Proposition 4.2),

$$P(X = x, Y = y) = P(Y = y|X = x)P(X = x) = \frac{1}{xN} \quad \text{for integer } 1 \leq y \leq x \leq N$$

and $P(X = x, Y = y) = 0$ otherwise.
b) Using the law of partitions (Proposition 2.8),

$$P(Y = y) = \sum_x P(X = x, Y = y) = \sum_{x=y}^{N} P(X = x, Y = y) = \frac{1}{N}\sum_{x=y}^{N}\frac{1}{x}$$

Thus the marginal PMF of Y is $P(Y = y) = \frac{1}{N}\sum_{x=y}^{N}\frac{1}{x}$ for $1 \leq y \leq N$ and $P(Y = y) = 0$ otherwise. Now to prove $\sum_y P(Y = y) = 1$ we rewrite the sum as:

$$\sum_y P(Y = y) = \sum_{y=1}^{N} P(Y = y) = \sum_{y=1}^{N}\left(\frac{1}{N}\sum_{x=y}^{N}\frac{1}{x}\right) = \frac{1}{N}\sum_{y=1}^{N}\sum_{x=y}^{N}\frac{1}{x}$$

Now we note that we can rewrite the summands as follows:

$$\begin{array}{rccccccccccc}
\displaystyle\sum_{y=1}^{N}\sum_{x=y}^{N}\frac{1}{x} = & \frac{1}{1} & + & \frac{1}{2} & + & \frac{1}{3} & + & \frac{1}{4} & + & \dots & + & \frac{1}{N} \\
+ & & & \frac{1}{2} & + & \frac{1}{3} & + & \frac{1}{4} & + & \dots & + & \frac{1}{N} \\
+ & & & & & \frac{1}{3} & + & \frac{1}{4} & + & \dots & + & \frac{1}{N} \\
+ & & & & & & & \frac{1}{4} & + & \dots & + & \frac{1}{N} \\
+ & & & & & & & & & \vdots & & \vdots \\
+ & & & & & & & & & & & \frac{1}{N} \\
\hline
= & 1 & + & 1 & + & 1 & + & 1 & + & \dots & + & 1 \quad = \mathrm{N}
\end{array}$$

Thus we have

$$\sum_y P(Y = y) = \sum_{y=1}^{N} P(Y = y) = \sum_{y=1}^{N}\left(\frac{1}{N}\sum_{x=y}^{N}\frac{1}{x}\right) = \frac{1}{N}\sum_{y=1}^{N}\sum_{x=y}^{N}\frac{1}{x} = \frac{1}{N}N = 1$$

c) $P(X = Y) = \sum_{x=1}^{N} P(X = x, Y = x) = \sum_{x=1}^{N}\frac{1}{Nx} = \frac{1}{N}\sum_{x=1}^{N}\frac{1}{x}$
d) Since $\sum_{x=1}^{N}\frac{1}{x} \sim \ln N$ as $N \to \infty$, $\frac{1}{N}\sum_{x=1}^{N}\frac{1}{x} \sim \frac{\ln N}{N}$ as $N \to \infty$.

6.2 Joint and Marginal Probability Mass Functions: Multivariate Case

Basic Exercises

6.21 $2^m - 2$. There are a total of 2^m subsets of $X_1, \ldots, X_m$ including the set of all of them, and the empty set. Excluding those two subsets, there are $2^m - 2$ subsets of random variables for which there are valid marginal PMFs.

6.22 First, we determine the univariate marginal PMFs. To obtain the marginal PMF of X, we have:

$$p_X(x) = \sum\sum_{(y,z)} p_{X,Y,Z}(x,y,z) = \sum_{y=0}^{2}\sum_{z=0}^{2} \frac{x+2y+z}{63}$$

$$= \frac{x}{63} + \frac{x+2}{63} + \frac{x+4}{63} + \frac{x+1}{63} + \frac{x+2+1}{63} + \frac{x+4+1}{63} + \frac{x+2}{63} + \frac{x+2+2}{63} + \frac{x+4+2}{63}$$

$$= \frac{x+3}{7} \qquad \text{for } x \in \{0,1\}$$

and $p_X(x) = 0$ otherwise.
Now, the marginal PMF of Y is given as:

$$p_Y(y) = \sum\sum_{(x,z)} p_{X,Y,Z}(x,y,z) = \sum_{x=0}^{1}\sum_{z=0}^{2} \frac{x+2y+z}{63}$$

$$= \frac{2y}{63} + \frac{2y+1}{63} + \frac{2y+1}{63} + \frac{2y+1+1}{63} + \frac{2y+2}{63} + \frac{2y+1+2}{63}$$

$$= \frac{4y+3}{21} \qquad \text{for } y \in \{0,1,2\}$$

and $p_Y(y) = 0$ otherwise.
and finaly, the marginal PMF of Z is given as:

$$p_Z(z) = \sum\sum_{(x,y)} p_{X,Y,Z}(x,y,z) = \sum_{x=0}^{1}\sum_{y=0}^{2} \frac{x+2y+z}{63}$$

$$= \frac{z}{63} + \frac{z+1}{63} + \frac{z+2}{63} + \frac{z+1+2}{63} + \frac{z+4}{63} + \frac{z+1+4}{63}$$

$$= \frac{2z+5}{21} \qquad \text{for } z \in \{0,1,2\}$$

and $p_Z(z) = 0$ otherwise.
Next we determine the bivariate marginal PMFs. To obtain the marginal PMF of X and Y, we have:

$$p_{X,Y}(x,y) = \sum_z p_{X,Y,Z}(x,y,z) = \sum_{z=0}^{2} \frac{x+2y+z}{63} = \frac{x+2y}{63} + \frac{x+2y+1}{63} + \frac{x+2y+2}{63}$$

$$= \frac{x+2y+1}{21} \qquad \text{for } x \in \{0,1\},\ y \in \{0,1,2\}$$

and $p_{X,Y}(x,y) = 0$ otherwise.
Next, the marginal PMF of X and Z, is given by:

$$p_{X,Z}(x,z) = \sum_{y} p_{X,Y,Z}(x,y,z) = \sum_{y=0}^{2} \frac{x+2y+z}{63} = \frac{x+z}{63} + \frac{x+z+2}{63} + \frac{x+z+4}{63}$$

$$= \frac{x+z+2}{21} \qquad \text{for } x \in \{0,1\},\ z \in \{0,1,2\}$$

and $p_{X,Z}(x,z) = 0$ otherwise.
Finaly, the marginal PMF of Y and Z is given as:

$$p_{Y,Z}(y,z) = \sum_{x} p_{X,Y,Z}(x,y,z) = \sum_{x=0}^{1} \frac{x+2y+z}{63} = \frac{2y+z}{63} + \frac{1+2y+z}{63}$$

$$= \frac{4y+2z+1}{63} \qquad \text{for } y,z \in \{0,1,2\}$$

and $p_{Y,Z}(y,z) = 0$ otherwise.

6.23 a) We have that since X is chosen uniformly from the 10 decimal digits, $P(X=x) = \frac{1}{10}$ for $x = 0,1,\ldots,9$. Once X is chosen, Y is chosen uniformly from the remaining 9 decimal digits, thus, $P(Y=y|X=x) = \frac{1}{9}$ for $y = 0,1,\ldots,9$ where $y \neq x$. Once Y is also chosen, then Z is chosen uniformly from the remaining 8 decimal digits, implying that $P(Z=z \,|\, X=x, Y=y) = \frac{1}{8}$ for $z = 0,1,\ldots,9$, such that z, y and x are all distinct. Thus, using the general multiplication rule, $P(X=x,Y=y,Z=z) = P(Z=z|X=x,Y=y)P(Y=y|X=x)P(X=x) = \frac{1}{720}$ for $x,y,z \in \{0,1,\ldots,9\}$ such that x, y and z are all distinct, and $P(X=x,Y=y,Z=z) = 0$ otherwise.
b) Since X, Y and Z must all be different, by the law of partitioning, $1 = P(\Omega) = P(X > Y > Z) + P(X > Z > Y) + P(Y > X > Z) + P(Y > Z > X) + P(Z > X > Y) + P(Z > Y > X)$. By symmetry, $P(X > Y > Z) = P(X > Z > Y) = P(Y > X > Z) = P(Y > Z > X) = P(Z > X > Y) = P(Z > Y > X)$ and thus $1 = 6P(X > Y > Z)$ so $P(X > Y > Z) = \frac{1}{6}$.
c) $P(X > Y > Z) = \sum_{x=2}^{9}\sum_{y=1}^{x-1}\sum_{z=0}^{y-1} P(X=x,Y=y,Z=z) = \sum_{x=2}^{9}\sum_{y=1}^{x-1}\sum_{z=0}^{y-1} \frac{1}{720} = \frac{1}{6}$.
d) First, we determine the univariate marginal PMFs. To obtain the marginal PMF of X, we have:

$$p_X(x) = \sum_{y,z} P(X=x,Y=y,Z=z) = \sum_{\substack{y=0 \\ y\neq x}}^{9} \sum_{\substack{z=0 \\ z\notin\{y,x\}}}^{9} \frac{1}{720} = \frac{1}{10} \qquad \text{for } x = 0,1,\ldots,9$$

and $p_X(x) = 0$ otherwise.
Now, the univariate marginal PMFs of Y and Z are computed using an identical argument and have the same form as X.

Next we determine the bivariate marginal PMFs. The marginal PMF of X and Y is given by:

$$p_{X,Y}(x,y) = \sum_{z} P(X=x,Y=y,Z=z) = \sum_{\substack{z=0 \\ z\notin\{x,y\}}}^{9} \frac{1}{720} = \frac{1}{90} \qquad \text{for } x,y = 0,1,\ldots,9,\ x \neq y$$

and $p_{X,Y}(x,y) = 0$ otherwise.
Now, the marginal PMF of X and Z and the marginal PMF of Y and Z are computed using an identical argument and have the same distribution.
e) For each of the univariate marginal PMFs, since each x, y or z is equally likely by symmetry, each random variable has a uniform distribution over the 10 decimal digits. For the bivariate marginal PMFs, each pair of X, Y, or X, Z, or Y, Z is equally likely to occur, and thus they are each uniformly distributed over the 90 total possibilities.

6.24 a) Using Proposition 5.13, and realizing that all three random variables are negative binomial random variables, the marginal PMF of X is:

$$p_X(x) = \binom{x-1}{0} p\,(1-p)^{x-1} = p\,(1-p)^{x-1} \qquad \text{for } x = 1, 2, \ldots$$

and $p_X(x) = 0$ otherwise.
The marginal PMF of Y is:

$$p_Y(y) = \binom{y-1}{1} p^2(1-p)^{y-2} = (y-1)p^2(1-p)^{y-2} \qquad \text{for } y = 2, 3, \ldots$$

and $p_Y(y) = 0$ otherwise.
The marginal PMF of Z is:

$$p_Z(z) = \binom{z-1}{2} p^3(1-p)^{z-3} \qquad \text{for } z = 3, 4, \ldots$$

and $p_Z(z) = 0$ otherwise.
b) First, since the time needed to wait between succesive trials is geometric, and $\{X = x, Y = y, Z = z\}$ is the same as $\{X = x, Y - X = y - x, Z - Y = z - y\}$, the joint PMF of X, Y and Z is:

$$p_{X,Y,Z}(x,y,z) = p(1-p)^{z-y-1}p(1-p)^{y-x-1}p(1-p)^{x-1} = p^3(1-p)^{z-3}$$

for integer $1 \leq x < y < z$.
c) For the bivariate marginal PMFs, first we consider X, Y. Again, since $\{X = x, Y = y\} = \{X = x, Y - X = y - x\}$, the marginal PMF of X and Y is:

$$p_{X,Y}(x,y) = p^2(1-p)^{y-2}$$

for integer $1 \leq x < y$. Similarily, since the amount of time needed to wait for the second success has negative binomial distribution with $r = 2$, the marginal PMF for Y and Z is:

$$p_{Y,Z}(y,z) = (y-1)p^2(1-p)^{y-2}p(1-p)^{z-y-1} = (y-1)p^3(1-p)^{z-3}$$

for integer $2 \leq y < z$.
Now, considering X and Z, the time needed to wait from the first success to the third is negative binomially distributed with $r = 2$ and since $\{X = x, Z = z\} = \{X = x, Z - X = z - x\}$ we have that the marginal PMF of X and Z is:

$$p_{X,Z}(x,z) = p(1-p)^{x-1}(z-x-1)p^2(1-p)^{z-x-2} = (z-x-1)p^3(1-p)^{z-3}$$

for integer $3 \leq x+2 \leq z$.
d) First, for X and Y,

$$P(X=x, Y=y) = \sum_{z} P(X=x, Y=y, Z=z) = \sum_{z=y+1}^{\infty} p^3(1-p)^{z-3}$$

$$= p^3(1-p)^{y-2} \sum_{z=0}^{\infty}(1-p)^z = p^2(1-p)^{y-2} = p_{X,Y}(x,y),$$

for $x \in \mathcal{N}$, $y = x+1, \ldots$.
Next, for Y and Z,

$$P(Y=y, Z=z) = \sum_{x} P(X=x, Y=y, Z=z) = \sum_{x=1}^{y-1} p^3(1-p)^{z-3}$$

$$= (y-1)p^3(1-p)^{z-3} = p_{Y,Z}(y,z),$$

for $y = 2, 3, \ldots$, $z = y+1, \ldots$.
Finally, for X and Z,

$$P(X=x, Z=z) = \sum_{y} P(X=x, Y=y, Z=z) = \sum_{y=x+1}^{z-1} p^3(1-p)^{z-3}$$

$$= (z-x-1)p^3(1-p)^{z-3} = p_{X,Z}(x,z),$$

for $x \in \mathcal{N}$, $z = x+2, \ldots$.
e) Once we've correctly identified the limits of the summands, using Proposition 6.5, we have

$$P(X=x) = \sum\sum_{(y,z)} P(X=x, Y=y, Z=z) = \sum_{y=x+1}^{\infty} \sum_{z=y+1}^{\infty} p^3(1-p)^{z-3} = p(1-p)^{x-1} = p_X(x)$$

for all $x \in \mathcal{N}$,

$$P(Y=y) = \sum\sum_{(x,z)} P(X=x, Y=y, Z=z) = \sum_{z=y+1}^{\infty} \sum_{x=1}^{y-1} p^3(1-p)^{z-3}$$

$$= (y-1)p^2(1-p)^{y-2} = p_Y(y), \quad y = 2, 3, \ldots,$$

$$P(Z=z) = \sum\sum_{(x,y)} P(X=x, Y=y, Z=z) = \sum_{x=1}^{z-2} \sum_{y=x+1}^{z-1} p^3(1-p)^{z-3}$$

$$= \binom{z-1}{2} p^3(1-p)^{z-3} = p_Z(z), \quad z = 3, 4, \ldots.$$

6.25 a) The random variables X, Y, Z have the multinomial distribution with parameters $n=3$ and $p_1 = \frac{1}{4}, p_2 = \frac{1}{4}$, and $p_3 = \frac{1}{2}$. Thus the joint PMF of X, Y, Z is:

$$p_{X,Y,Z}(x,y,z) = \binom{3}{x, y, z} \left(\frac{1}{4}\right)^x \left(\frac{1}{4}\right)^y \left(\frac{1}{2}\right)^z$$

if x, y, and z, are nonegative integers whose sum is 3 and otherwise $p_{X,Y,Z}(x,y,z) = 0$.
b)

$$P(X > Y > Z) = p_{X,Y,Z}(2,1,0) = \frac{3}{16},$$

which is the probability that more Greens are chosen than Democrats and more Democrats are chosen than Republicans.

$$P(X \leq Y < Z) = p_{X,Y,Z}(0,0,3) + p_{X,Y,Z}(0,1,2) = \frac{5}{16},$$

which represents the probability that at least as many Democrats are chosen as Greens, and that more Republicans are chosen than either of the other groups.

$$P(X \leq Z \leq Y) = p_{X,Y,Z}(0,3,0) + p_{X,Y,Z}(0,2,1) + p_{X,Y,Z}(1,1,1) = \frac{19}{64},$$

which represents the probability that there are at least as many Republicans as Greens chosen, and at least as many Democrats as Republicans chosen.
c) Since each selection is independent from other selections, $X \sim \mathcal{B}\left(3, \frac{1}{4}\right)$, $Y \sim \mathcal{B}\left(3, \frac{1}{4}\right)$, and $Z \sim \mathcal{B}\left(3, \frac{1}{2}\right)$.
d) If two of the three parties are grouped and considered as one new group, we have $X + Y \sim \mathcal{B}\left(3, \frac{1}{2}\right)$, $X + Z \sim \mathcal{B}\left(3, \frac{3}{4}\right)$ and $Y + Z \sim \mathcal{B}\left(3, \frac{3}{4}\right)$
e) $X + Y$ and Z have the multinomial distribution with $n = 3$, $p_1 = \frac{1}{2}$ and $p_2 = \frac{1}{2}$. Thus the joint PMF of $X + Y$, Z is

$$p_{X+Y,\,Z}(u,v) = \binom{3}{u}\frac{1}{8}$$

if $u = 0,1,2,3$ and $v = 3 - u$, and otherwise $p_{X+Y,Z}(u,v) = 0$.
f) Using the FPF,

$$P(1 \leq X + Z \leq 2) = p_{X,Y,Z}(2,1,0) + p_{X,Y,Z}(0,2,1) + p_{X,Y,Z}(1,2,0)$$

$$+p_{X,Y,Z}(0,1,2) + p_{X,Y,Z}(1,1,1) = \frac{9}{16}.$$

g) Using (d), $P(1 \leq X + Z \leq 2) = P(X + Z = 1) + P(X + Z = 2) = \frac{9}{16}$.
h) Since three subjects must be chosen, $p_{X+Y+Z}(3) = 1$ and $p_{X+Y+Z}(u) = 0$ for all $u \neq 3$.

6.26 a) The random variables X_1, X_2, X_3, X_4 have the multinomial distribution with parameters $n = 5$ and $p_1 = \frac{6}{40}, p_2 = \frac{15}{40}, p_3 = \frac{12}{40}$ and $p_4 = \frac{7}{40}$. Thus the joint PMF of X_1, X_2, X_3, X_4 is:

$$p_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4) = \binom{5}{x_1, x_2, x_3, x_4}\left(\frac{6}{40}\right)^{x_1}\left(\frac{15}{40}\right)^{x_2}\left(\frac{12}{40}\right)^{x_3}\left(\frac{7}{40}\right)^{x_4}$$

if x_1, x_2, x_3, and x_4 are nonegative integers whose sum is 5 and otherwise

$$p_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4) = 0.$$

b) 0.0554, since $P(X_1 = 0, X_2 = 3, X_3 = 1, X_4 = 1) = p_{X_1,X_2,X_3,X_4}(0,3,1,1)$

$= \binom{5}{0,3,1,1}\left(\frac{6}{40}\right)^0\left(\frac{15}{40}\right)^3\left(\frac{12}{40}\right)^1\left(\frac{7}{40}\right)^1 \approx 0.0554$
c) $P(X_2 \geq 3) = p_{X_1,X_2,X_3,X_4}(2,3,0,0) + p_{X_1,X_2,X_3,X_4}(1,3,1,0) + p_{X_1,X_2,X_3,X_4}(1,3,0,1)$ $+ p_{X_1,X_2,X_3,X_4}(0,3,2,0) + p_{X_1,X_2,X_3,X_4}(0,3,1,1) + p_{X_1,X_2,X_3,X_4}(0,3,0,2)$ $+ p_{X_1,X_2,X_3,X_4}(1,4,0,0) + p_{X_1,X_2,X_3,X_4}(0,4,1,0) + p_{X_1,X_2,X_3,X_4}(0,4,0,1)$ $+p_{X_1,X_2,X_3,X_4}(0,5,0,0) = 0.0119+0.0475+0.0277+0.0475+0.0554+0.0161+0.0148+0.0297+0.0173 + 0.0074 \approx 0.2752$
d) $P(X_2 \geq 3, X_3 \geq 1) = p_{X_1,X_2,X_3,X_4}(1,3,1,0)+p_{X_1,X_2,X_3,X_4}(0,3,2,0)+p_{X_1,X_2,X_3,X_4}(0,3,1,1)+p_{X_1,X_2,X_3,X_4}(0,4,1,0) = 0.0475 + 0.0475 + 0.0554 + 0.0297 \approx 0.1325$

6.27 a) Let X_i be the number of times the i^{th} face appears in 18 tosses. Then $X_1, X_2, X_3, X_4, X_5, X_6$ have the multinomial distribution with $n = 18$ and $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$. Thus $P(X_1 = X_2 = X_3 = X_4 = X_5 = X_6 = 3) = \binom{18}{3,3,3,3,3,3}\frac{1}{6^{18}} \approx 0.00135$
b) 0.00135, since $\{X_i \leq 3 \text{ for all } i = 1,\ldots,6,\ X_1+\ldots+X_6 = 18\} = \{X_i = 3 \text{ for all } i = 1,\ldots,6\}$.
c) 0.00135, since $\{X_i \geq 3 \text{ for all } i = 1,\ldots,6,\ X_1+\ldots+X_6 = 18\} = \{X_i = 3 \text{ for all } i = 1,\ldots,6\}$.

6.28 0.1136; Let X_1 be the number of good items, X_2 be the number of salvageable items, and X_3 be the number of scrapped items. Then the random variables X_1, X_2, X_3 have the multinomial distribution with parameters $n = 50$ and $p_1 = 0.97$, $p_2 = 0.03 \times \frac{2}{3} = 0.02$, and $p_3 = 0.03 \times \frac{1}{3} = 0.01$. Thus

$$p_{X_1,X_2,X_3}(48,1,1) = \binom{50}{48,1,1}(0.97)^{48}(0.02)^1(0.01)^1 \approx 0.1136$$

6.29 a) $X_2 \sim \mathcal{B}(n, p_2)$ because when concerned with only the value of X_2, the other 4 variables can be grouped into 1 new larger variable, and thus we have a multinomial distribution with only two variables.
b) $X_3 + X_5 \sim \mathcal{B}(n, p_3 + p_5)$ because after grouping X_3 and X_5, the other 3 variables can be grouped as well, and again result in a Binomial distribution.
c) $X_1 + X_4 + X_5 \sim \mathcal{B}(n, p_1 + p_4 + p_5)$ using the same reasoning as above.
d) X_1+X_2 and $X_3+X_4+X_5$ have the multinomial distribtion with n and p_1+p_2, and $p_3+p_4+p_5$ using the same regrouping argument as above.
e) X_1, $X_2 + X_3$, and $X_4 + X_5$ have the multinomial distribution with n and p_1, $p_2 + p_3$ and $p_4 + p_5$ using the same regrouping argument as above.
f) X_1, $X_2 + X_3 + X_4$, and X_5 have the multinomial distribution with n and p_1, $p_2 + p_3 + p_4$ and p_5 using the same regrouping argument as above.

6.30 a) By the binomial theorem,

$$\frac{1}{(q+r)^{n-x}}\sum_{y=0}^{n-x}\binom{n}{x,y,n-x-y}q^y r^{n-x-y} = \frac{1}{(q+r)^{n-x}}\sum_{y=0}^{n-x}\frac{n!}{x!y!(n-x-y)!}q^y r^{n-x-y}$$

$$= \frac{1}{(q+r)^{n-x}}\sum_{y=0}^{n-x}\frac{n!}{x!(n-x)!}\cdot\frac{(n-x)!}{y!(n-x-y)!}q^y r^{n-x-y}$$

$$= \frac{1}{(q+r)^{n-x}}\binom{n}{x}\underbrace{\sum_{y=0}^{n-x}\binom{n-x}{y}q^y r^{n-x-y}}_{=(q+r)^{n-x}} = \binom{n}{x}.$$

b) Proving the required equality is equivalent to showing that the following holds:

$$\binom{n}{x}(1-(q+r))^x(q+r)^{n-x} = \sum_{y=0}^{n-x}\binom{n}{x, y, n-x-y}(1-(q+r))^x q^y r^{n-x-y}$$

which clearly is the equation used to obtain the marginal PMF of X from a multinomial distribution by using Proposition 6.5.

6.31 If p_i is the proportion of people in the population of size N with attribute i, then Np_i is the total number of people with attribute i. Then $\binom{Np_i}{x_i}$ is the number of ways that x_i people can be chosen with attribte i from the population. Thus, $\binom{Np_1}{x_1}\times\cdots\times\binom{Np_m}{x_m}$ is the total number of ways to choose $x_1, \ldots x_m$ people with attributes $a_1, \ldots, a_m$ respectively. If $n = x_1 + \ldots + x_n$, then since $\binom{N}{n}$ is the total number of ways to pick n people from a population of N people, we have:

$$p_{X_1,X_2,\ldots,X_m}(x_1, x_2, \ldots, x_m) = \frac{\binom{Np_1}{x_1}\cdots\binom{Np_m}{x_m}}{\binom{N}{n}}$$

6.32 a) X, Y, Z have a multiple hypergeometric distribution with $N = 8$, $n = 3$, $p_x = \frac{1}{4}$, $p_y = \frac{1}{4}$ and $p_z = \frac{1}{2}$.
b)

$$P(X > Y > Z) = p_{X,Y,Z}(2, 1, 0) = \frac{1}{28},$$

which is the probability that more Greens are chosen than either Democrats or Republicans, and more Democrats are chosen than Republicans.

$$P(X \leq Y < Z) = p_{X,Y,Z}(0, 0, 3) + p_{X,Y,Z}(0, 1, 2) = \frac{2}{7},$$

which is the probability that at least as many Democrats are chosen as Greens, and that more Republicans are chosen than either other group.

$$P(X \leq Z \leq Y) = p_{X,Y,Z}(0, 2, 1) + p_{X,Y,Z}(1, 1, 1) = \frac{5}{14},$$

which is the probability that there are at least as many Republicans as Greens chosen, and at least as many Democrats as Republicans chosen.
c)

$$P(X = x) = \frac{\binom{2}{x}\binom{6}{3-x}}{\binom{8}{3}}$$

for $x = 0, 1, 2$ and $P(X = x) = 0$ otherwise.

$$P(Y = y) = \frac{\binom{2}{y}\binom{6}{3-y}}{\binom{8}{3}}$$

for $y = 0, 1, 2$ and $P(Y = y) = 0$ otherwise.

$$P(Z = z) = \frac{\binom{4}{z}\binom{4}{3-z}}{\binom{8}{3}}$$

for $z = 0, 1, 2, 3$ and $P(Z = z) = 0$ otherwise.
d)

$$P(X+Y=u) = \frac{\binom{4}{u}\binom{4}{3-u}}{\binom{8}{3}}$$

for $u = 0, 1, 2, 3$ and $P(X+Y=u) = 0$ otherwise.

$$P(X+Z=v) = \frac{\binom{6}{v}\binom{2}{3-v}}{\binom{8}{3}}$$

for $v = 0, 1, 2, 3$ and $P(X+Z=v) = 0$ otherwise.

$$P(Y+Z=w) = \frac{\binom{6}{w}\binom{2}{3-w}}{\binom{8}{3}}$$

for $w = 0, 1, 2, 3$ and $P(Y+Z=w) = 0$ otherwise.
e) $X+Y, Z$ have a multiple hypergeometric distribution with $N = 8$, $n = 3$, $p_1 = \frac{1}{2}$, and $p_2 = \frac{1}{2}$.
f) Using the FPF,

$$P(1 \le X+Z \le 2) = p_{X,Y,Z}(2,1,0) + p_{X,Y,Z}(0,2,1) + p_{X,Y,Z}(1,2,0)$$
$$+p_{X,Y,Z}(0,1,2) + p_{X,Y,Z}(1,1,1) = \frac{9}{14}$$

g) Using **d)**, $P(1 \le X+Z \le 2) = P(X+Z=1) + P(X+Z=2) = \frac{9}{14}$
h) Since three subjects must be chosen, $P_{X+Y+Z}(3) = 1$ and $P_{X+Y+Z}(u) = 0$ otherwise.

6.33 a) X_1, X_2, X_3, X_4 have a multiple hypergeometric distribution with $N = 40$, $n = 5$, $p_1 = \frac{6}{40}$, $p_2 = \frac{15}{40}$, $p_3 = \frac{12}{40}$, and $p_4 = \frac{7}{40}$.
b) $P(X_1 = 0, X_2 = 3, X_3 = 1, X_4 = 1) = p_{X_1,X_2,X_3,X_4}(0,3,1,1) \approx 0.0581$
c) $P(X_2 \ge 3) = p_{X_1,X_2,X_3,X_4}(2,3,0,0) + p_{X_1,X_2,X_3,X_4}(1,3,1,0) + p_{X_1,X_2,X_3,X_4}(1,3,0,1)$
$+ p_{X_1,X_2,X_3,X_4}(0,3,2,0) + p_{X_1,X_2,X_3,X_4}(0,3,1,1) + p_{X_1,X_2,X_3,X_4}(0,3,0,2)$
$+ p_{X_1,X_2,X_3,X_4}(1,4,0,0) + p_{X_1,X_2,X_3,X_4}(0,4,1,0) + p_{X_1,X_2,X_3,X_4}(0,4,0,1)$
$+ p_{X_1,X_2,X_3,X_4}(0,5,0,0) \approx 0.264$
d) $P(X_2 \ge 3, X_3 \ge 1) = p_{X_1,X_2,X_3,X_4}(1,3,1,0) + p_{X_1,X_2,X_3,X_4}(0,3,2,0) + p_{X_1,X_2,X_3,X_4}(0,3,1,1) + p_{X_1,X_2,X_3,X_4}(0,4,1,0) \approx 0.178$

6.34 a) When sampling with replacement, the distribution is multinomial and thus:

$$p_{X_1,X_2,X_3}(x_1,x_2,x_3) = 0.4\left(\binom{6}{x_1,x_2,x_3}\left(\frac{5}{14}\right)^{x_1}\left(\frac{11}{28}\right)^{x_2}\left(\frac{1}{4}\right)^{x_3}\right)$$
$$+0.6\left(\binom{6}{x_1,x_2,x_3}\left(\frac{1}{4}\right)^{x_1}\left(\frac{5}{12}\right)^{x_2}\left(\frac{1}{3}\right)^{x_3}\right),$$

for all non-negative integers x_1, x_2, x_3 whose sum is 6.
b) When sampling without replacement, the distribution is multiple hypergeometric and thus:

$$p_{X_1,X_2,X_3}(x_1,x_2,x_3) = 0.4\left(\frac{\binom{10}{x_1}\binom{11}{x_2}\binom{7}{x_3}}{\binom{28}{6}}\right) + 0.6\left(\frac{\binom{9}{x_1}\binom{15}{x_2}\binom{12}{x_3}}{\binom{36}{6}}\right),$$

for all non-negative integers x_1, x_2, x_3 such that $x_1 + x_2 + x_3 = 6$.

Theory Exercises

6.35 Clearly, $p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = P(X_1 = x_1,\cdots,X_m = x_m) \geq 0$ for all $x_1,\ldots,x_m \in \mathcal{R}$. Next, similarly to Exercise 6.16, note that $\{(x_1,\cdots,x_m) : p_{X_i}(x_i) = 0\} \subset \{(x_1,\cdots,x_m) : p_{X_1,\ldots,X_m}(x_1,\cdots,x_m) = 0\}$ for all $i = 1,\ldots,m$, implying that

$$\{(x_1,\cdots,x_m) : p_{X_i}(x_i) \neq 0\} \supset \{(x_1,\cdots,x_m) : p_{X_1,\ldots,X_m}(x_1,\cdots,x_m) \neq 0\} \text{ for all } i = 1,\ldots,m.$$

Therefore,

$$\{(x_1,\cdots,x_m) : p_{X_1,\ldots,X_m}(x_1,\cdots,x_m) \neq 0\} \subset \bigcap_{i=1}^{m}\{(x_1,\cdots,x_m) : p_{X_i}(x_i) \neq 0\}.$$

Note that $\bigcap_{i=1}^{m}\{(x_1,\cdots,x_m) : p_{X_i}(x_i) \neq 0\} = \{x_1 : p_{X_1}(x_1) \neq 0\} \times \ldots \times \{x_m : p_{X_m}(x_m) \neq 0\}$, which is a cartesian product of a finite collection of countable sets, thus, is itself countable. Therefore, $\{(x_1,\cdots,x_m) : p_{X_1,\ldots,X_m}(x_1,\cdots,x_m) \neq 0\}$, as a subset of a countable set, is countable. Finally note that, since the random variables are discrete, we have that

$$1 = P(\Omega) = P\left((X_1,\cdots,X_m) \in \{(x_1,\ldots,x_m) : p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) \neq 0\}\right)$$

$$= \sum_{(x_1,\ldots,x_m):\, p_{X_1,\ldots,X_m}(x_1,\ldots,x_m)\neq 0} p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = \sum_{(x_1,\ldots,x_m)} p_{X_1,\ldots,X_m}(x_1,\ldots,x_m).$$

6.36 Suppose that (a) $p(x_1,\ldots,x_m) \geq 0$ for all $x_1,\ldots,x_m \in \mathcal{R}^m$; (b) $\{(x_1,\ldots,x_m) \in \mathcal{R}^m : p(x_1,\ldots,x_m) \neq 0\}$ is countable; (c) $\sum_{(x_1,\ldots,x_m)} p(x_1,\ldots,x_m) = 1$. Let $\Omega = \{(x_1,\ldots,x_m) \in \mathcal{R}^m : p(x_1,\ldots,x_m) \neq 0\}$. Then Ω is countable and, in view of conditions (a),(c) and Proposition 2.3, there exists a unique probability measure P such that $P(\{(\omega_1,\ldots,\omega_m)\}) = p(\omega_1,\ldots,\omega_m)$ for all $(\omega_1,\ldots,\omega_m) \in \Omega$. Define $X_i((\omega_1,\ldots,\omega_m)) = \omega_i$ for all i and for all $(\omega_1,\ldots,\omega_m) \in \Omega$. Then $X_1,\ldots,X_m$ are discrete random variables and

$$p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = P(\{(\omega_1,\ldots,\omega_m) \in \Omega : X_i((\omega_1,\ldots,\omega_m)) = x_i \text{ for all } i\})$$

$$= P(\{(x_1,\ldots,x_m)\}) = \begin{cases} p(x_1,\ldots,x_m), & \text{if } (x_1,\ldots,x_m) \in \Omega, \\ 0, & \text{if } (x_1,\ldots,x_m) \notin \Omega. \end{cases}$$

On the other hand, note that $p(x_1,\ldots,x_m) = 0$ for all $(x_1,\ldots,x_m) \notin \Omega$, by definition of Ω. Therefore, $p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = p(x_1,\ldots,x_m)$ for all $(x_1,\ldots,x_m) \in \mathcal{R}^m$.

6.37 Similarly to Exercise 6.18, the required result follows at once by the law of partitioning and since for every integer-valued $1 \leq k_1 < \cdots < k_j \leq m$ (with $j \in \{1,\cdots,m\}$),

$$\left\{X_{k_1} = x_{k_1},\cdots,X_{k_j} = x_{k_j}\right\} = \bigcup_{(y_1,\ldots,y_m):\, y_{k_i}=x_{k_i} \text{ for all } i=1,\ldots,j} \{X_1 = y_1,\ldots,X_m = y_m\}.$$

6.38 Since the set where $p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) > 0$ must be countable, let us consider A in Proposition 6.6 to be countable. If it wasn't, we could partition it into a set where for every element in the set, $p_{X_1,\ldots,X_m}(x,\ldots,x_m) = 0$ (which we could then disregard) and a countable set where for every element in that set, $p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) > 0$. Now, using the law of partitions, we can partition A into countably many subsets, one subset for every element of the set, and then the desired result is an immediate consequence.

6.39 For $k = 1, \cdots, m$ and $i \in \{1, \cdots, n\}$, let $N_i = k$ whenever E_k is the ith event to occur. Moreover, for arbitrary given non-negative integers $x_1, \cdots, x_m$ with $x_1 + \cdots + x_m = n$, define $\mathcal{K} = \{(k_1, ..., k_n) \in \{1, ..., m\}^n : x_1$ of all the k_j's are equal to 1, etc., x_m of all the k_j's are equal to $m\}$. Then the cardinality of $\mathcal{K}$ is equal to $\binom{n}{x_1, \cdots, x_m}$ and

$$P(X_1 = x_1, \cdots, X_m = x_m) = \sum_{(k_1, \cdots, k_n) \in \mathcal{K}} P(N_1 = k_1, \cdots, N_n = k_n)$$

$$= \sum_{(k_1, \cdots, k_n) \in \mathcal{K}} p_{k_1} \ldots p_{k_n} = \sum_{(k_1, \cdots, k_n) \in \mathcal{K}} p_1^{x_1} \ldots p_m^{x_m} = \binom{n}{x_1, \cdots, x_m} p_1^{x_1} \ldots p_m^{x_m}.$$

Advanced Exercises

6.40 a) If we have that there are n items to be distributed to m places, then knowing $x_1 + \ldots + x_{m-1}$ of them have been distributed to $X_1, \ldots, X_{m-1}$ is the same as knowing that $x_1 + \ldots + x_{m-1}$ of them have been distributed to $X_1, \ldots, X_{m-1}$ and that $n - (x_1 + \ldots + x_{m-1}$ of them go to X_m since logically we must have a place for every item. Therefore, $\{X_1 = x_1, \ldots, X_{m-1} = x_{m-1}\} = \{X_1 = x_1, \ldots, X_{m-1} = x_{m-1}, X_m = x_m\}$
b) Using Proposition 3.5, the result follows immediately.
c) When $m = 2$, we have $p_{X_1}(x_1) = \binom{n}{x_1, n-x_1} p_1^{x_1}(1 - p_1)^{n-x_1} = \binom{n}{x_1} p_1^{x_1}(1 - p_1)^{n-x_1}$ which is clearly the form for the binomial distribution.
d) Consider the marginal PMF for $X_{j_1}, \ldots, X_{j_k}$ for any $1 \leq k \leq m$. We would have by grouping, $\{X_{j_1} = x_{j_1}, \ldots, X_{j_k} = x_{j_k}\} = \{X_{j_1} = x_{j_1}, \ldots, X_{j_k} = x_{j_k}, X_{j_{k+1}} + \ldots + X_{j_m} = n - (x_{j_1} + \ldots + x_{j_k})\}$ Thus, using Proposition 3.5, we have

$$p_{X_{j_1}, \ldots, X_{j_k}}(x_{j_1}, \ldots, x_{j_k}) = \binom{n}{x_{j_1}, \ldots, x_{j_k}, n - (x_{j_1} + \ldots + x_{j_k})}$$

$$\times p_{j_1}^{x_{j_1}} \cdots p_{j_k}^{x_{j_k}} (1 - (p_{j_1} + \ldots + p_{j_k}))^{n-(x_{j_1} + \ldots + x_{j_k})}$$

which is clearly in the same form as the PMF given in **b)**.

6.41 a) For non-negative $x_1, \ldots, x_n$ such that $x_1 + \ldots + x_n = n$

$$\frac{\binom{Np_1}{x_1} \cdots \binom{Np_m}{x_m}}{\binom{N}{n}} = \binom{n}{x_1, \ldots, x_m} (Np_1)_{x_1} \cdots (Np_m)_{x_m} \cdot \frac{1}{(N)_n}$$

Note that for all $1 \leq i \leq n$,
$(Np_i)_{x_i} = (Np_i - x_i + 1)(Np_i - x_i + 2) \ldots (Np_i) = N^{x_i}(p_i - \dfrac{x_i - 1}{N})(p_i - \dfrac{x_i - 2}{N}) \ldots p_i$. Thus, $\dfrac{1}{N^n}(Np_1)_{x_1} \cdots (Np_m)_{x_m} \longrightarrow p_1^{x_1} \cdots p_m^{x_m}$, as $N \to \infty$. Note also that
$(N)_n = N^n(1 - \dfrac{n-1}{N})(1 - \dfrac{n-2}{N}) \ldots 1$, which implies that $\frac{(N)_n}{N^n} \to 1$ as $N \to \infty$. Therefore,

$$\frac{(Np_1)_{x_1}(Np_m)_{x_m}}{(N)_n} \longrightarrow p_1^{x_1} \cdots p_m^{x_m},$$

as $N \to \infty$, and the required result follows.
b) When N is large relative to n, the change in the probabilities from successive removals changes so minutely as to not affect the overall distribution.

6.42 a) Let N be the total population of U.S. residents. Then the exact distribution of X_1, X_2, X_3, X_4 is multiple hypergeometric and thus:

$$P_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4) = \frac{\binom{N*0.19}{x_1}\binom{N*0.231}{x_2}\binom{N*0.355}{x_3}\binom{N*0.224}{x_4}}{\binom{N}{n}}$$

b) When N is large relative to n, we can approximate the multiple hypergeometric with the multinomial since the difference in corresponding probabilities when sampling is done with and without replacement are negligible. Thus the approximate distribution is:

$$P_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4) \approx \binom{n}{x_1,x_2,x_3,x_4}(0.19)^{x_1}(0.231)^{x_2}(0.355)^{x_3}(0.244)^{x_4}.$$

c) When $n = 5$, $P_{X_1,X_2,X_3,X_4}(1,1,2,1) \approx 0.08098$

6.43 Assuming that the politician is running for an office important enough to have her have so many more constituents than 40, we can use the multinomial approximation to the multiple hypergeometric. This being done, were her claim to be true, we would have expected to see 26 people who favor her issue and 12 who don't. Even if we assume that every one of the "undecided" people sampled eventually sides with her, it would still seem that she is overestimating her popularity on the issue, and it would seem prudent for her opponents to use this to their advantage. Some may note however, that either way, she still has a majority of the people on her side, and thus this may not be the issue of choice for her opponents.

6.3 Conditional Probability Mass Functions

Basic Exercises

6.44 a)

		Bedrooms, X			
		2	3	4	Total
Bathrooms, Y	2	$\frac{3}{19}$	$\frac{14}{19}$	$\frac{2}{19}$	1.00
	3	0	$\frac{12}{23}$	$\frac{11}{23}$	1.00
	4	0	$\frac{2}{7}$	$\frac{5}{7}$	1.00
	5	0	0	1.00	1.00
	$p_X(x)$	$\frac{3}{50}$	$\frac{14}{25}$	$\frac{19}{50}$	1.00

b) $P(X = 3 \mid Y = 2) + P(X = 4 \mid Y = 2) = \frac{14}{19} + \frac{2}{19} = \frac{16}{19} \approx 0.8421$
c) About 84.21% of all 2 bathroom houses have at least 3 bedrooms.

6.45 a) For each number of siblings X, the conditional PMF of the number of sisters Y is given in the table in (b) and is computed as $P(Y = y|X = x) = \frac{P(X=x,Y=y)}{P(X=x)}$.

b)

		Sisters, Y				
		0	1	2	3	Total
Siblings, X	0	1	0	0	0	1.00
	1	0.471	0.529	0	0	1.00
	2	0.091	0.727	0.182	0	1.00
	3	0.0	0.333	0.667	0	1.00
	4	0	0	0	1	1.00
	$p_Y(y)$	0.425	0.45	0.1	0.025	1.00

c) $P(Y \geq 1|X = 2) = \frac{10}{11} \approx 0.909$
d) For each number of sisters Y, the conditional PMF of the number of siblings X is given in the table in (e) and is computed as $P(X = x|Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)}$.
e)

		Siblings, X					
		0	1	2	3	4	Total
Sisters, Y	0	$\frac{8}{17} \approx 0.471$	$\frac{8}{17} \approx 0.471$	$\frac{1}{17} \approx 0.059$	0	0	1.0
	1	0	0.5	$\frac{8}{18} \approx 0.444$	$\frac{1}{18} \approx 0.056$	0	1.0
	2	0	0	0.5	0.5	0	1.0
	3	0	0	0	0	1	1.0
	$p_X(x)$	$\frac{8}{40} = 0.2$	$\frac{17}{40} = 0.425$	$\frac{11}{40} = 0.275$	$\frac{3}{40} = 0.075$	0.025	1.0

f) $P(X \geq 2 \,|\, Y = 1) = 0.5$

6.46 a) Given arbitrary $x \in \{0, 1, \cdots, 9\}$,

$$P(Y = y \,|\, X = x) = \frac{1}{9}$$

for $y \in \{0, 1, \cdots, 9\} \setminus \{x\}$, which shows that $Y|X$ is uniformly distributed on the remaining 9 decimal digits.
b) $P(X = x \,|\, Y = y) = \dfrac{P(X = x, Y = y)}{P(Y = y)} = \dfrac{\frac{1}{90}}{\sum_{k \in \{0,\cdots,9\} \setminus \{y\}} P(X = k, Y = y)}$
$= \dfrac{\frac{1}{90}}{\sum_{k \in \{0,\cdots,9\} \setminus \{y\}} \frac{1}{90}} = \dfrac{1}{9}$ for $y \in \{0, 1, \cdots, 9\} \setminus \{x\}$.
c) $P(3 \leq X \leq 4 \,|\, Y = 2) = P(X = 3 \,|\, Y = 2) + P(X = 4 \,|\, Y = 2) = \frac{2}{9}$

6.47 a)

		I_F		
		0	1	p_{I_E}
I_E	0	$P(F^c \,|\, E^c)$	$P(F \,|\, E^c)$	1
	1	$P(F^c \,|\, E)$	$P(F \,|\, E)$	1
	$p_{I_F}(x)$	$P(F^c)$	$P(F)$	1

b) The result, that guarantees that the values in each row in the above table sum to 1, is Proposition 4.1.

6.48 a) The required table for the conditional PMFs of Y given $X = x$ and the marginal PMF of Y is given by:

Larger value, Y

Smaller value, X	1	2	3	4	5	6	Total
1	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	1
2	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	1
3	0	0	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	1
4	0	0	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	1
5	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$	1
6	0	0	0	0	0	1	1
$p_Y(y)$	0.0278	0.0834	0.1390	0.1946	0.2502	0.3058	1

b) The required table for the conditional PMFs of X given $Y = y$ and the marginal PMF of X is given by:

Smaller value, X

Larger value, Y	1	2	3	4	5	6	Total
1	1	0	0	0	0	0	1
2	$\frac{2}{3}$	$\frac{1}{3}$	0	0	0	0	0
3	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	0	0	0	1
4	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	0	0	1
5	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	0	1
6	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	1
$p_X(x)$	0.3058	0.2502	0.1946	0.1390	0.0834	0.0278	1

6.49 a) The conditional PMF of Y given $X = x$ ($x \in \mathcal{N}$) is equal to $P(Y = y|X = x) = p(1-p)^{y-x-1}$ for $y = x+1, x+2, \ldots$ and $P(Y = y|X = x) = 0$ otherwise, since by Exercise 6.9 the number of trias needed after the first success to get the second success has geometric distribution with parameter p.
b) For each $x \in \mathcal{N}$, $P(Y = y|X = x) = \frac{P(X=x,Y=y)}{P(X=x)} = p(1-p)^{y-x-1}$ for $y = x+1, \cdots$.
c) The conditional PMF of X given $Y = y$ ($y \in \{2, 3, \cdots\}$) is equal to $P(X = x|Y = y) = \frac{1}{y-1}$ for $x = 1, \ldots, y-1$ and $P(X = x|Y = y) = 0$ otherwise, since once we know when the second success must occur, any of the previous $(y-1)$ trials could have been the one where the first success occured, and thus is uniformly distributed.
d) $P(X = x|Y = y) = \dfrac{(1-p)^{y-2}p^2}{\binom{y-1}{1}p^2(1-p)^{y-2}} = \dfrac{1}{y-1}$ for $y \in \{2, 3, \cdots\}$ and $x \in \{1, \cdots, y-1\}$.
e) The conditional PMF of $Y - X$ given $X = x$ equals $P(Y - X = z|X = x) = p(1-p)^{z-1}$ for $z = 1, \ldots$ and $P(Y - X = x|X = x) = 0$ otherwise, since the occurence of the first success is independent of the number of trials between the first and second success, and thus as shown before, the number of trials needed between successess is geometric with parameter p.
f) $P(Y - X = z|X = x) = P(Y = x + z|X = x) = p(1-p)^{z-1}$, by (b).

6.50 a) As shown in Exercise 6.13, $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x \in \mathcal{Z}_+$, thus

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{e^{-(\lambda+\mu)}\lambda^x\mu^y/(x!y!)}{e^{-\lambda}\lambda^x/x!} = \frac{e^{-\mu}\mu^y}{y!},\ y = 0, 1, \ldots,$$

thus, $Y|X=x$ has $\mathcal{P}(\mu)$ distribution.
b) As shown in Exercise 6.13, $P(Y=y)=\frac{e^{-\mu}\mu^y}{y!}$, $y\in\mathcal{Z}_+$, thus

$$P(X=x|Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}=\frac{e^{-(\lambda+\mu)}\lambda^x\mu^y/(x!y!)}{e^{-\mu}\mu^y/y!}=\frac{e^{-\lambda}\lambda^x}{x!},\ x=0,1,\ldots,$$

thus, $X|Y=y$ has $\mathcal{P}(\lambda)$ distribution.
c) The conditional distribution of Y given X is the same as the marginal distribution of Y.
d) The conditional distribution of X given Y is the same as the marginal distribution of X.

6.51 a) $p_{Y|X}(y|x)=\frac{p_{X,Y}(x,y)}{p_X(x)}=\frac{p_X(x)p_Y(y)}{p_X(x)}=p_Y(y)$
b) $p_{X|Y}(x|y)=\frac{p_{X,Y}(x,y)}{p_Y(y)}=\frac{p_X(x)p_Y(y)}{p_Y(y)}=p_X(x)$

6.52 a) From Exercise 6.22, we have that $p_Y(y)=\frac{4y+3}{21}$, and thus

$$p_{X,Z|Y}(x,z|y)=\frac{p_{X,Y,Z}(x,y,z)}{p_Y(y)}=\frac{x+2y+z}{12y+9}$$

Therefore we have

$$p_{X,Z|Y}(x,z|1)=\frac{x+z+2}{21}$$

b) Given that $Y=1$, the probability that all the family's automobile-accident losses will be reimbursed is:

$$P(X+Y+Z\le 2|Y=1)=p_{X,Z|Y}(0,0|1)+p_{X,Z|Y}(1,0|1)+p_{X,Z|Y}(0,1|1)=\frac{2}{21}+\frac{3}{21}+\frac{3}{21}=\frac{8}{21}.$$

c) From Exercise 6.22, we have that $p_{X,Y}(x,y)=\frac{x+2y+1}{21}$, and thus

$$p_{Z|X,Y}(z|x,y)=\frac{p_{X,Y,Z}(x,y,z)}{p_{X,Y}(x,y)}=\frac{x+2y+z}{3x+6y+3}$$

Therefore we have

$$p_{Z|X,Y}(z|0,1)=\frac{z+2}{9}$$

d) Given that $X=0$ and $Y=1$, the probability that all the family's automobile-accident losses will be reimbursed is:

$$P(X+Y+Z\le 2|X=0,Y=1)=p_{Z|X,Y}(0,1,0)+p_{Z|X,Y}(0,1,1)=\frac{5}{9}$$

6.53 a) For $x\in\{0,\cdots,n\}$,

$$p_{Y,Z|X}(y,z|x)=\frac{p_{X,Y,Z}(x,y,z)}{p_X(x)}=\frac{\binom{n}{x,y,z}p^xq^yr^z}{\binom{n}{x}p^x(q+r)^{y+z}}=\frac{(n-x)!}{y!z!}\left(\frac{q}{1-p}\right)^y\left(\frac{r}{1-p}\right)^z$$
$$=\binom{n-x}{y,z}\left(\frac{q}{1-p}\right)^y\left(\frac{r}{1-p}\right)^z,$$

where y, z are nonnegative integers whose sum is $n - x$. Thus $Y, Z|X = x$ has a multinomial distribution with parameters $n - x$ and $q/(1-p)$, $r/(1-p)$.
b) Since we must have that $X + Y + Z = n$, then if $X = x$ and $Y = y$, then we must have $Z = n - x - y$. Hence

$$p_{Z|X,Y}(z|x,y) = \begin{cases} 1 & \text{if } z = n - x - y \\ 0 & \text{otherwise} \end{cases}$$

6.54 a)

$$p_{X_1,\ldots,X_{m-1}|X_m}(x_1,\ldots,x_{m-1}|x_m) = \frac{p_{X_1,\ldots,X_m}(x_1,\ldots,x_m)}{p_{X_m}(x_m)} = \frac{\binom{n}{x_1,\ldots,x_m} {p_1}^{x_1} \times \cdots \times {p_m}^{x_m}}{\binom{n}{x_m} {p_m}^{x_m}(1-p_m)^{n-x_m}}$$

$$= \frac{(n-x_m)!}{x_1! \cdots x_{m-1}!} \left(\frac{p_1}{1-p_m}\right)^{x_1} \times \cdots \times \left(\frac{p_{m-1}}{1-p_m}\right)^{x_{m-1}}$$

$$= \binom{n-x_m}{x_1,\ldots,x_{m-1}} \left(\frac{p_1}{1-p_m}\right)^{x_1} \times \cdots \times \left(\frac{p_{m-1}}{1-p_m}\right)^{x_{m-1}},$$

which is a multinomial distribution with parameters $n - x_m$ and $p_1/(1-p_m), \cdots, p_{m-1}/(1-p_m)$.
b) Here $m = 3$, $n = 10$, $p_1 = \frac{18}{38}$, $p_2 = \frac{18}{38}$, and $p_3 = \frac{2}{38}$. Thus,

$$p_{X_1,X_2|X_3}(x_1,x_2,1) = \binom{n-1}{x_1,x_2} \left(\frac{1}{2}\right)^{x_1+x_2}.$$

6.55 a) If the first $m - 1$ random variables are grouped together, we have:

$$p_{X_m}(x_m) = \frac{\binom{N(1-p_m)}{n-x_m}\binom{Np_m}{x_m}}{\binom{N}{n}}, \quad x_m = 0, 1, \cdots, Np_m,$$

thus the conditional distribution of $X_1, \ldots, X_{m-1}$ given $X_m = x_m$ is:

$$p_{X_1,\ldots,X_{m-1}|X_m}(x_1,\ldots,x_{m-1}|x_m) = \frac{p_{X_1,\ldots,X_m}(x_1,\ldots,x_m)}{p_{X_m}(x_m)}$$

$$= \frac{\binom{Np_1}{x_1} \cdots \binom{Np_{m-1}}{x_{m-1}}}{\binom{N(1-p_m)}{n-x_m}},$$

which represents a multiple hypergeometric distribution with parameters $N(1-p_m)$, $n - x_m$ and $p_1/(1-p_m), \cdots, p_{m-1}/(1-p_m)$.
b) Here $m = 4$, $N = 40$, $n = 5$, $Np_1 = 6$, $Np_2 = 15$, $Np_3 = 12$, $Np_4 = 7$, thus

$$p_{X_1,X_2,X_3|X_4}(x_1,x_2,x_3,1) = \frac{\binom{6}{x_1}\binom{15}{x_2}\binom{12}{x_3}}{\binom{33}{4}},$$

where x_1, x_2, x_3 are non-negative integers whose sum is 4.

Theory Exercises

6.56 Let $P(Y = y \,|\, X = x) = g(y)$ and $P(X = x) = h(x)$. Then, $P(X = x, Y = y) = h(x)g(y)$ and $P(Y = y) = \sum_x h(x)g(y) = g(y)\sum_x h(x)$. Since $h(x)$ is the PMF of X, $\sum_x h(x) = 1$, and thus, $P(Y = y) = g(y) = P(Y = y \,|\, X = x)$.

6.57 Clearly the general multiplication rule holds for all x such that $p_X(x) > 0$. Additionally, if $p_X(x) = 0$, then we must have that $p_{X,Y}(x, y) = 0$ as well. Thus Equation 6.15 holds for all real x.

Advanced Exercises

6.58 a) Let $y = n - (x_{k+1} + \ldots + x_m)$ and $p_y = 1 - (p_{k+1} + \ldots + p_m)$. From Exercise 6.40, we know

$$P(X_{k+1} = x_{k+1}, \ldots, X_m = x_m) = \binom{n}{y, x_{k+1}, \ldots, x_m} p_y^y p_{k+1}^{x_{k+1}} \cdots p_x^{x_m}$$

Thus the conditional distribution of $X_1, \ldots, X_k$ given $X_{k+1} = x_{k+1}, \ldots, X_m = x_m$ is

$$p_{X_1,\ldots,X_k|X_{k+1},\ldots,X_m}(x_1, \ldots, x_k|x_{k+1} \ldots, x_m) = \frac{P(X_1 = x_1, \ldots, X_m = x_m)}{P(X_{k+1} = x_{k+1}, \ldots, X_m = x_m)}$$

$$= \frac{\binom{n}{x_1,\ldots,x_m} p_1^{x_1} \cdots p_m^{x_m}}{\binom{n}{y,x_{k+1},\ldots,x_m} p_y^y p_{k+1}^{x_{k+1}} \cdots p_x^{x_m}} = \frac{\frac{n!}{x_1!\cdots x_m!} p_1^{x_1} \cdots p_k^{x_k}}{\frac{n!}{y! x_{k+1}! \cdots x_m!} p_y^y} = \frac{y!}{x_1! \cdots x_k!} \frac{p_1^{x_1} \cdots p_k^{x_l}}{p_y^{x_1+\ldots+x_k}}$$

$$= \binom{y}{x_1, \ldots, x_k} \left(\frac{p_1}{p_y}\right)^{x_1} \cdots \left(\frac{p_k}{p_y}\right)^{x_k}$$

Hence, $X_1, \ldots, X_k | X_{k+1} = x_{k+1}, \ldots, X_m = x_m$ is multinomial with parameters y and $\frac{p_1}{p_y}, \ldots, \frac{p_k}{p_y}$.
b) The result of **a)** says that once we know that out of the n total trials, $y - n$ of them are accounted for by $X_{k+1}, \ldots, X_m$, the remaining y are distributed multinomially among the first k X, with relative prbabilities instead of the original ones.

6.59 a) $P(Y = y|X = x) = \frac{1}{x}$ for $y = 1, \ldots, x$ and $P(Y = y|X = x) = 0$ otherwise. In other words, $Y|X = x$ has uniform distribution on $\{1, \cdots, x\}$ (where $x \in \{1, \cdots, N\}$).
b) Using the result from Exercise 6.20, the conditional distribution of X given that $Y = y$ is:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{\frac{1}{Nx}}{\frac{1}{N}\sum_{j=y}^{N} \frac{1}{j}} = \frac{1}{\sum_{j=y}^{N} \frac{x}{j}},$$

c) First, as $N \to \infty$, for a fixed y, $\sum_{n=y}^{N} n^{-1} \sim \sum_{n=1}^{N} n^{-1}$. Second, using the fact that $\sum_{n=1}^{N} n^{-1} \sim \ln N$ as $N \to \infty$, we have that $\frac{1}{\sum_{n=y}^{N} n^{-1}} \sim \frac{1}{\ln N}$ as $N \to \infty$. Therefore, finally, we have that $\frac{1}{x\sum_{n=y}^{N} n^{-1}} \sim \frac{1}{x \ln N}$ as $N \to \infty$.

6.4 Independent Random Variables

Basic Exercises

6.60 a) First, we have that $p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y q(x)r(y) = q(x)\sum_y r(y)$ and similarly, $p_Y(y) = r(y)\sum_x q(x)$. Also we note that $1 = \sum_{(x,y)} p_{X,Y}(x,y) = \sum_{(x,y)} q(x)r(y) = \left(\sum_x q(r)\right)\left(\sum_y r(y)\right)$. Thus we have:

$$p_{X,Y}(x,y) = q(x)r(y) = q(x)r(y)\underbrace{\left(\sum_x q(r)\right)\left(\sum_y r(y)\right)}_{=1}$$

$$= \left(q(x)\sum_y r(y)\right)\left(r(y)\sum_x q(x)\right) = p_X(x)p_Y(y)$$

b) q and r are the marginal PMFs of X and Y respectively only when $\sum_x q(x) = 1 = \sum_y r(y)$.
c) Generalizing to the multivariate case of m discrete random variables, we assume

$$p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = q_1(x_1)\times\cdots\times q_m(x_m) \text{ for all } x_1,\cdots,x_m \in \mathcal{R}.$$

Then we have that

$$p_{X_i}(x_i) = \sum_{\{x_1,\ldots,x_i,\ldots,x_m\}\in\mathcal{R}^{i-1}\times\{x_i\}\times\mathcal{R}^{m-i}} p_{X_1,\ldots,X_m}(x_1,\ldots,x_m)$$

$$= \sum_{\{x_1,\ldots,x_i,\ldots,x_m\}\in\mathcal{R}^{i-1}\times\{x_i\}\times\mathcal{R}^{m-i}} q_1(x_1)\times\cdots\times q_m(x_m)$$

$$= q_i(x_i)\sum_{\{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m\}\in\mathcal{R}^{m-1}} q_1(x_1)\times\cdots\times q_{i-1}(x_{i-1})\times q_{i+1}(x_{i+1})\times\cdots\times q_m(x_m)$$

$$= q_i(x_i)\left(\sum_{x_1} q_1(x_1)\right)\cdots\left(\sum_{x_{i-1}} q_{i-1}(x_{i-1})\right)\left(\sum_{x_{i+1}} q_{i+1}(x_{i+1})\right)\cdots\left(\sum_{x_m} q_m(x_m)\right)$$

We also have that

$$1 = \sum_{\mathcal{R}^m} p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = \sum_{\mathcal{R}^m} q_1(x_1)\times\cdots\times q_m(x_m) = \left(\sum_{x_1} q_1(x_1)\right)\cdots\left(\sum_{x_m} q_m(x_m)\right)$$

thus

$$p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) = q_1(x_1)\times\cdots\times q_m(x_m)$$

$$= q_1(x_1)\times\cdots\times q_m(x_m)\left(\left(\sum_{x_1} q_1(x_1)\right)\cdots\left(\sum_{x_m} q_m(x_m)\right)\right)^{m-1}$$

$$= \prod_{i=1}^{m}\left(q_i(x_i)\left(\sum_{x_1} q_1(x_1)\right)\cdots\left(\sum_{x_{i-1}} q_{i-1}(x_{i-1})\right)\left(\sum_{x_{i+1}} q_{i+1}(x_{i+1})\right)\cdots\left(\sum_{x_m} q_m(x_m)\right)\right)$$

$$= \prod_{i=1}^{m} p_{X_1}(x_i),$$

thus, by Proposition 6.14, $X_1, \cdots, X_m$ are independent.

6.61 X and Y are not independent random variables. For example, $p_{X,Y}(0,1) = 0$ but $p_X(0)p_Y(1) = \frac{8}{40} \times \frac{18}{40} = \frac{9}{100}$.

6.62 X and Y are not independent random variables since $p_{X,Y}(1,1) = 0$ but $p_X(1)p_Y(1) = \frac{1}{100}$

6.63 X and Y are not independent random variables since $p_{X,Y}(2,1) = 0$ but $p_X(2)p_Y(1) = 0.007$

6.64 a) X and Y are not independent since Y must be greater than X and therefore depends on X.
b) Using Proposition 6.10, we know that while $p_X(2) > 0$ and $p_Y(2) > 0$, $p_{X,Y}(2,2) = 0$.
c) Using Proposition 6.11, we know that while $p_Y(2) > 0$, $p_{Y|X}(2|2) = 0$.
d) Once the first success occurs, its occurence has no effect on how much longer one must wait till the second success.
e) Using Proposition 6.10 and Exercise 6.9, for all $x, z \in \mathcal{R}$,

$$p_{X,Y-X}(x,z) = P(X = x, Y - X = z) = P(X = x, Y = z + x) = p^2(1-p)^{x+z-2}$$
$$= \left(p(1-p)^{x-1}\right)\left(p(1-p)^{z-1}\right) = p_X(x)p_{Y-X}(z).$$

Therefore, X and $Y - X$ are independent.
f) Using Proposition 6.11,

$$P(Y - X = z|X = x) = \frac{P(Y - X = z, X = x)}{P(X = x)} = \frac{p^2(1-p)^{x+z-2}}{p(1-p)^{x-1}} = p(1-p)^{z-1} = p_{Y-X}(z)$$

Therefore X and $Y - X$ are independent.

6.65 If some two of the random variables $X_1, \cdots, X_m$ are not independent, then the m random variables $X_1, \cdots, X_m$ are not independent. To see that, suppose the converse, i.e. suppose for some $i, j \in \{1, \cdots, m\}$, X_i and X_j are not independent, but $X_1, \cdots, X_m$ are independent. The latter independence, by Definition 6.5, implies the equality that $P(X_1 \in A_1, \ldots, X_m \in A_m) = P(X_1 \in A_1) \ldots P(X_m \in A_m)$ for all subsets $A_1, \cdots, A_m$ of real numbers. Thus, taking $A_k = \Omega$ for all $k \notin \{i, j\}$, we obtain that $P(X_i \in A_i, X_j \in A_j) = P(X_i \in A_i)P(X_j \in A_j)$ for all subsets A_i, A_j of $\mathcal{R}$, which contradicts the fact that X_i and X_j are not independent.

6.66 a) Since there are only 5 people in the sample, we know that $X_1 + X_2 \leq 5$ and thus if $X_1 = 5$, we must have $X_2 = 0$
b) If X_1, X_2, X_3, X_4 were independent than any subset of them would be as well (see the argument in Exercise 6.65), and (a) gives an example of a subset that is not independent, and thus, X_1, X_2, X_3, X_4 are not independent.

6.67 a) Since there are only 5 people in the sample, we know that $X_1 + X_2 \leq 5$ and thus if $X_1 = 5$, we must have $X_2 = 0$
b) If X_1, X_2, X_3, X_4 were independent than any subset of them would be as well, and **a)** gives an example of a subset that is not independent, and thus, X_1, X_2, X_3, X_4 are not independent.

6.68 Using the same logic as in Exercise 6.66 and 6.67, we know that $X_1 + X_2 \leq n$ and thus if

$X_1 = n$, we must have $X_2 = 0$. Therefore X_1 and X_2 are not independent. Thus $X_1, \ldots, X_m$ can't be independent by Exercise 6.65.

6.69 Yes, X and Y are independent, since $p_{X,Y}(x,y) = \frac{1}{100} = \frac{1}{10} \cdot \frac{1}{10} = p_X(x)p_Y(y)$.

6.70 a) Since X must be at least as large as Y, then Y depends on X.
b) Note that $p_X(0) = \frac{1}{55}, p_Y(1) = \frac{9}{55}$, $p_{X,Y}(0,1) = 0 \neq p_X(0)p_Y(1)$, thus X and Y are not independent by Proposition 6.10.
c) Since $p_Y(1) = \frac{9}{55}$, $p_{Y|X}(1|0) = 0$, thus X and Y are not independent by Proposition 6.11.

6.71 Let X denote the time (to the nearest minute) it takes for a student to complete the mathematics exam, and let Y denote the corresponding time for the verbal exam.
a) $P(X > Y) = \sum_{x=1}^{\infty} \sum_{y=1}^{x-1} (1-0.02)^{x-1} 0.02 (1-0.025)^{y-1} 0.025 = \frac{49}{89} \approx 0.5506$
b) $P(X \leq 2Y) = \sum_{y=1}^{\infty} \sum_{x=1}^{2y} 0.98^{x-1} 0.975^{y-1} 0.02 \cdot 0.025 \approx 0.6225$
c) $P(X+Y=z) = \sum_{x=1}^{z-1} P(X=x, Y=z-x) = \sum_{x=1}^{z-1} 0.98^{x-1}(0.02)0.975^{z-x-1}0.025 = 0.1\left((0.98)^{z-1} - (0.975)^{z-1}\right)$. If Z denotes the total time that a student takes to complete both exams, then $p_Z(z) = 0.1\left((0.98)^{z-1} - (0.975)^{z-1}\right)$ for $z = 2, 3, 4, \ldots$ and $p_Z(z) = 0$ otherwise.
d) Since our variables are in minutes, we are interested in the probability that the total time is less than or equal to 120 minutes. Then

$$P(Z \leq 120) = \sum_{z=2}^{120} p_Z(z) = 0.1(\sum_{z=2}^{120} 0.98^{z-1} - \sum_{z=2}^{120} 0.975^{z-1}) = 0.749$$

6.72 a) By Proposition 6.13, $\sqrt{X_1^2 + X_2^2 + X_3^2}$ and $\sqrt{X_4^2 + X_5^2 + X_6^2}$ are independent.
b) $\sin(X_1X_2X_3)$ and $\cos(X_3X_5)$ are not independent since if they were, $\frac{\sin^{-1}(\sin(X_1X_2X_3))}{X_1X_2} = X_3$ would have to be independent of $\frac{\cos^{-1}(\cos(X_3X_5))}{X_5} = X_3$.
c) By Proposition 6.13, $\sin(X_1X_2X_5)$ and $\cos(X_3X_4)$ are independent.
d) $X_1 - X_2$ and $X_1 + X_2$ are not necessarily independent. For example, if $X_2 = 0$ with probability one and X_1 takes at least two possible values with positive probabilities, then clearly $X_1 - X_2$ and $X_1 + X_2$ are not independent, although X_1, X_2 are independent.
e) By Proposition 6.13, $\min\{X_1, X_2, X_7\}$ and $\max\{X_3, X_4\}$ are independent.
f) By Proposition 6.13, $\sum_{k=1}^{5} X_k \sin(kt)$ and $\sum_{k=6}^{7} X_k \sin(kt)$ are independent.

6.73 a) Let Z denote the total number of patients that arrive each hour at both emergency rooms combined. Letting X and Y being the number of patients that arrive at the big and small emergency rooms respectively, we have: for all $z \in \mathcal{Z}_+$,

$$P(Z=z) = P(X+Y=z) = \sum_{x=0}^{z} P(X=x, Y=z-x) = \sum_{x=0}^{z} \left(\frac{e^{-6.9}6.9^x}{x!}\right)\left(\frac{e^{-2.6}2.6^{z-x}}{(z-x)!}\right)$$

$$= e^{-(6.9+2.6)} \sum_{x=0}^{z} \frac{6.9^x 2.6^{z-x}}{x!(z-x)!} = e^{-9.5} \sum_{x=0}^{z} \frac{z!}{z!} \cdot \frac{6.9^x 2.6^{z-x}}{x!(z-x)!} = \frac{e^{-9.5}}{z!} \sum_{x=0}^{z} \binom{z}{x} 6.9^x 2.6^{z-x}$$

and using the binomial theorem, $\sum_{x=0}^{z} \binom{z}{x} 6.9^x 2.6^y = (6.9+2.6)^z = 9.5^z$. Thus we have that $P(Z=z) = \frac{e^{-9.5}9.5^z}{z!}$ for $z \in \mathcal{Z}_+$. Hence, $Z \sim \mathcal{P}(9.5)$.
b) By definition of the Poisson distribution, we expect 9.5 visits per hour in the two hospitals combined.

c) $P(Z \le 8) = \sum_{z=0}^{8} P(Z = z) = e^{-9.5}\left(1 + \frac{9.5}{1!} + \frac{9.5^2}{2!} + \cdots + \frac{9.5^8}{8!}\right) \approx 0.3918$

6.74 Let Y be the number of people making a depost during the time interval. Let X be the number of people entering the bank in the time interval. Then

$$\begin{aligned}
P(Y = y) &= \sum_x P(X = x, Y = y) = \sum_x P(Y = y | X = x) P(X = x) \\
&= \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} \cdot \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-\lambda} p^y}{y!} \sum_{x=y}^{\infty} \frac{(1-p)^{x-y}\lambda^x}{(x-y)!} \\
&= \frac{e^{-\lambda} p^y}{y!} \sum_{x=0}^{\infty} \frac{(1-p)^x \lambda^{x+y}}{x!} = \frac{e^{-\lambda} p^y \lambda^y}{y!} \sum_{x=0}^{\infty} \frac{(1-p)^x \lambda^x}{x!} \\
&= \frac{e^{-\lambda}(p\lambda)^y}{y!} \underbrace{\sum_{x=0}^{\infty} \frac{((1-p)\lambda)^x}{x!}}_{=e^{(1-p)\lambda}} = \frac{e^{-\lambda+\lambda-p\lambda}(p\lambda)^y}{y!} = \frac{e^{-p\lambda}(p\lambda)^y}{y!}
\end{aligned}$$

Thus $Y \sim \mathcal{P}(p\lambda)$.

6.75 Each person entering the bank either is type 1 (with probability p_1) or not, independently of the types of other people entering and of the number of people entering the bank, thus, by Exercise 6.74, X_1 has $\mathcal{P}(\lambda p_1)$ distribution. A similar argument shows that X_2 has $\mathcal{P}(\lambda p_2)$ distribution and X_3 has $\mathcal{P}(\lambda p_3)$ distribution. On the other hand, let $N = X_1 + X_2 + X_3$ denote the number of people entering the bank, then N has $\mathcal{P}(\lambda)$ distribution. Therefore, for all $x_1, x_2, x_3 \in \mathcal{Z}_+$,

$$\begin{aligned}
p_{X_1,X_2,X_3}(x_1, x_2, x_3) &= \sum_{k=0}^{\infty} P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \,|\, N = k) P(N = k) \\
&= \binom{x_1 + x_2 + x_3}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3} \cdot \frac{e^{-\lambda}\lambda^{x_1+x_2+x_3}}{(x_1 + x_2 + x_3)!} \\
&= e^{-\lambda} \frac{(\lambda p_1)^{x_1}}{x_1!} \cdot \frac{(\lambda p_2)^{x_2}}{x_2!} \cdot \frac{(\lambda p_3)^{x_3}}{x_3!} = p_{X_1}(x_1) p_{X_2}(x_2) p_{X_3}(x_3),
\end{aligned}$$

where in the second equality we used the fact that $P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \,|\, N = k) = 0$ unless $k = x_1 + x_2 + x_3$, and that, given $N = x_1 + x_2 + x_3$, the distribution of X_1, X_2, X_3 is multinomial with parameters $x_1 + x_2 + x_3$ and p_1, p_2, p_3. The last equality easily follows upon noting that $e^{-\lambda} = e^{-\lambda p_1} e^{-\lambda p_2} e^{-\lambda p_3}$. Finally, if at least one of x_1, x_2, x_3 is not a non-negative integer, then $p_{X_1,X_2,X_3}(x_1, x_2, x_3) = 0 = p_{X_1}(x_1) p_{X_2}(x_2) p_{X_3}(x_3)$. Thus, X_1, X_2, X_3 are independent Poisson random variables with parameters λp_1, λp_2 and λp_3 respectively.

Theory Exercises

6.76 By Definition 6.5, if $X_1, \cdots, X_m$ are independent, then $P(X_1 \in A_1, \cdots, X_m \in A_m) = P(X_1 \in A_1) \ldots P(X_m \in A_m)$ for all subsets $A_1, \ldots, A_m$ of real numbers, therefore, for every subset $\mathcal{J} = \{i_1, \cdots, i_k\} \subset \{1, \cdots, m\}$, upon taking $A_i = \Omega$ for all $i \in \{1, \cdots, m\} \setminus \mathcal{J}$, we obtain

$$P(X_{i_1} \in A_{i_1}, \cdots, X_{i_k} \in A_{i_k}) = P(X_{i_1} \in A_1) \ldots P(X_{i_k} \in A_k) \cdot \underbrace{\prod_{i \notin \mathcal{J}} P(X_i \in \Omega)}_{=1} = \prod_{i \in \mathcal{J}} P(X_i \in A_i),$$

which, by Definition 4.5, implies that $\{X_1 \in A_1\}, \ldots, \{X_m \in A_m\}$ are mutually independent.

Conversely, if $\{X_1 \in A_1\}, \ldots, \{X_m \in A_m\}$ are mutually independent for all subsets of $\mathcal{R}$, then $P(X_1 \in A_1, \cdots, X_m \in A_m) = \prod_{i=1}^{m} P(X_i \in A_i)$ for all subsets of $\mathcal{R}$, thus, by Definition 6.5, $X_1, \ldots, X_m$ are independent.

6.77 Suppose $X_1, \ldots, X_m$ are independent. Let $x_1, \ldots, x_m$ be any real numbers, and set $A_1 = \{x_1\}, \ldots, A_m = \{x_m\}$. Then

$$p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m) = P(X_1 \in A_1, \ldots, X_m \in A_m)$$
$$= P(X_1 \in A_1) \cdots P(X_m \in A_m) = P(X_1 = x_1) \cdots P(X_m = x_m) = p_{X_1}(x_1) \cdots p_{X_m}(x_m)$$

Conversely, suppose

$$p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = p_{X_1}(x_1) \cdots p_{X_m}(x_m) \text{ for all } x_1, \cdots, x_m \in \mathcal{R}.$$

Then, let $A_1, \ldots, A_m$ be any m subsets of $\mathcal{R}$. Applying the above equation and the FPF gives

$$P(X_1 \in A_1, \ldots, X_m \in A_m) = P((X_1, \ldots, X_m) \in A_1 \times \cdots \times A_m)$$
$$= \sum_{(x_1,\ldots,x_m) \in A_1 \times \cdots \times A_m} \cdots \sum p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \sum_{(x_1,\ldots,x_m) \in A_1 \times \cdots \times A_m} \cdots \sum p_{X_1}(x_1) \cdots p_{X_m}(x_m)$$
$$= \left(\sum_{x_1 \in A_1} p_{X_1}(x_1) \right) \cdots \left(\sum_{x_m \in A_m} p_{X_m}(x_m) \right) = P(X_1 \in A_1) \cdots P(X_m \in A_m)$$

as required.

6.78 Note that for any event E, we have $\{\omega : I_E(\omega) = 1\} = E$ and $\{\omega : I_E(\omega) = 0\} = E^c$. Suppose first that the random variables $I_{E_1}, \ldots, I_{E_m}$ are independent, then by Exercise 6.76, the events $\{I_{E_1} = 1\}, \ldots, \{I_{E_m} = 1\}$ are independent, implying that the events $E_1, \ldots, E_m$ are independent.

Conversely, suppose that $E_1, \ldots, E_m$ are independent, then, by Proposition 4.5, the events $F_1, \ldots, F_m$ are also independent, where each F_i is either E_i or E_i^c, $i = 1, \cdots, m$. The latter, in view of the earlier note, implies that the events $\{I_{E_1} = x_1\}, \ldots, \{I_{E_m} = x_m\}$ are independent for all $x_1, \cdots, x_m \in \{0, 1\}$. Thus, $p_{I_{E_1},\ldots,I_{E_m}}(x_1, \cdots, x_m) = p_{I_{E_1}}(x_1) \ldots p_{I_{E_m}}(x_m)$ for all $x_1, \ldots, x_m \in \{0, 1\}$. On the other hand, if at least one of $x_1, \ldots, x_m$ is not in $\{0, 1\}$, then we have that $p_{I_{E_1},\ldots,I_{E_m}}(x_1, \cdots, x_m) = 0 = p_{I_{E_1}}(x_1) \ldots p_{I_{E_m}}(x_m)$. Therefore, $p_{I_{E_1},\ldots,I_{E_m}}(x_1, \cdots, x_m) = p_{I_{E_1}}(x_1) \ldots p_{I_{E_m}}(x_m)$ for all real $x_1, \ldots, x_m$, which by Exercise 6.77, implies independence of $I_{E_1}, \ldots, I_{E_m}$.

Advanced Exercises

6.79 a) The minimum of two random variables is greater than a value x if and only if both of the random variables are greater than x. Thus, for all $x \in \mathcal{N}$, by independence of X, Y,

$$P(\min\{X, Y\} > x) = P(X > x, Y > x) = P(X > x)P(Y > x)$$
$$= (1 - p)^x (1 - q)^x = (1 - (p + q - pq))^x,$$

implying that for all $x \in \mathcal{N}$,

$$p_{\min\{X,Y\}}(x) = P(\min\{X,Y\} > x-1) - P(\min\{X,Y\} > x)$$
$$= (1-(p+q-pq))^{x-1} - (1-(p+q-pq))^{x} = (1-(p+q-pq))^{x-1}(p+q-pq),$$

and we conclude that $\min\{X,Y\} \sim \mathcal{G}(p+q-pq)$.
b) Since both X and Y must be positive integers, then if $X+Y=z$, there are $z-1$ ways that could happen, i.e. $(X=1, Y=z-1), \ldots, (X=z-1, Y=1)$. We can compute $P(X+Y=z)$ as follows:

$$P(X+Y=z) = \sum_{x=1}^{z-1} P(X=x, Y=z-x) = \sum_{x=1}^{z-1} p(1-p)^{x-1}q(1-q)^{z-x-1}$$

$$= pq(1-q)^{z-2}\sum_{x=0}^{z-2}\left(\frac{1-p}{1-q}\right)^{x} = pq(1-q)^{z-2}\frac{1-\left(\frac{1-p}{1-q}\right)^{z-1}}{1-\frac{1-p}{1-q}}$$
$$= \frac{pq(1-q)^{z-1}\left(1-\left(\frac{1-p}{1-q}\right)^{z-1}\right)}{p-q} = \frac{pq((1-q)^{z-1}-(1-p)^{z-1})}{p-q}.$$

c) By (b) we have that

$$P(X+Y=z) = \frac{pq((1-q)^{z-1}-(1-p)^{z-1})}{p-q}$$
$$= \frac{pq((1-q)-(1-p))\left[(1-q)^{z-2}+(1-q)^{z-3}(1-p)+(1-q)^{z-4}(1-p)^2+\cdots+(1-p)^{z-2}\right]}{p-q}$$
$$= pq\sum_{k=0}^{z-2}(1-q)^k(1-p)^{z-2-k}, \quad z \in \{2,3,\cdots\}.$$

Thus, when $p=q$, we have that for $z \in \{2,3,\cdots\}$,

$$P(X+Y=z) = p^2\sum_{k=0}^{z-2}(1-p)^{z-2} = (z-1)p^2(1-p)^{z-2} = \binom{z-1}{2-1}p^2(1-p)^{z-2}.$$

Therefore, $X+Y \sim \mathcal{NB}(2,p)$.

6.80 a) First we note,

$$p_X(x) = \begin{cases} \frac{1}{4}, & x=-1 \\ \frac{1}{2}, & x=0 \\ \frac{1}{4}, & x=1 \end{cases}$$

and

$$p_Y(y) = \begin{cases} \frac{1}{4}, & y=-1 \\ \frac{1}{2}, & y=0 \\ \frac{1}{4}, & y=1 \end{cases}$$

X and Y are not independent, since $p_{X,Y}(1,1) = 0 \neq \frac{1}{16} = \frac{1}{4} \cdot \frac{1}{4} = p_X(1)p_Y(1)$
b) First we note

$$p_{X+Y}(z) = \begin{cases} \frac{1}{2}, & z = -1 \\ \frac{1}{2}, & z = 1 \end{cases}$$

and

$$p_{X-Y}(z) = \begin{cases} \frac{1}{2}, & z = -1 \\ \frac{1}{2}, & z = 1 \end{cases}$$

Now,

$$\begin{aligned} P(X+Y=-1, X-Y=-1) &= P(X=-1, Y=0) \\ = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} &= P(X+Y=-1)P(X-Y=-1) \\ P(X+Y=-1, X-Y=1) &= P(X=0, Y=-1) \\ = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} &= P(X+Y=-1)P(X-Y=1) \\ P(X+Y=1, X-Y=-1) &= P(X=0, Y=1) \\ = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} &= P(X+Y=1)P(X-Y=-1) \\ P(X+Y=1, X-Y=1) &= P(X=1, Y=0) \\ = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} &= P(X+Y=1)P(X-Y=1) \end{aligned}$$

Thus $X+Y$ and $X-Y$ are independent.

6.81 First note that $P(X_1 = x_1, \ldots, X_m = x_m \,|\, X = x) = 0$ if $x_1 + \ldots + x_m \neq x$. Then

$$P(X_1 = x_1) = \sum_{x_2=0}^{\infty} \cdots \sum_{x_m=0}^{\infty} P(X_1 = x_1, \cdots, X_m = x_m \,|\, X = x_1 + \cdots + x_m) P(X = x_1 + \cdots + x_m)$$

$$= \sum_{x_2=0}^{\infty} \cdots \sum_{x_m=0}^{\infty} \binom{x_1 + \ldots + x_m}{x_1, \ldots, x_m} p_1^{x_1} \cdots p_m^{x_m} \frac{e^{-\lambda} \lambda^{x_1 + \ldots + x_m}}{(x_1 + \ldots + x_m)!}$$

$$= \sum_{x_2=0}^{\infty} \cdots \sum_{x_m=0}^{\infty} \frac{e^{-(p_1 + \ldots + p_m)\lambda} \lambda^{x_1} \cdots \lambda^{x_m}}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m}$$

$$= \left(\frac{e^{-p_1 \lambda} (p_1 \lambda)^{x_1}}{x_1!} \right) \underbrace{\left(\sum_{x_2=0}^{\infty} \frac{e^{-p_2 \lambda} (p_2 \lambda)^{x_2}}{x_2!} \right)}_{=1} \cdots \underbrace{\left(\sum_{x_m=0}^{\infty} \frac{e^{-p_m \lambda} (p_m \lambda)^{x_m}}{x_m!} \right)}_{=1} = \frac{e^{-p_1 \lambda} (p_1 \lambda)^{x_1}}{x_1!}$$

for $x_1 \in \mathcal{Z}_+$. Thus, $X_1 \sim \mathcal{P}(p_1 \lambda)$. Moreover, a similar argument implies that $X_i \sim \mathcal{P}(p_i \lambda)$ for each $i = 1, \cdots, m$. Next note that

$$p_{X_1, \ldots, X_m}(x_1, \ldots, x_m) = \sum_{x=0}^{\infty} P(X_1 = x_1, \cdots, X_m = x_m \,|\, X = x) P(X = x)$$

$$= P(X_1 = x_1, \cdots, X_m = x_m \,|\, X = x_1 + \cdots + x_m) P(X = x_1 + \cdots + x_m)$$

$$= \binom{x_1 + \ldots + x_m}{x_1, \ldots, x_m} p_1^{x_1} \cdots p_m^{x_m} \frac{e^{-\lambda} \lambda^{x_1 + \ldots + x_m}}{(x_1 + \ldots + x_m)!} = \frac{e^{-(p_1 + \ldots + p_m)\lambda} \lambda^{x_1} \cdots \lambda^{x_m}}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m}$$

$$= \left(\frac{e^{-p_1 \lambda} (p_1 \lambda)^{x_1}}{x_1!} \right) \cdots \left(\frac{e^{-p_m \lambda} (p_m \lambda)^{x_m}}{x_m!} \right) = p_{X_1}(x_1) \cdots p_{X_m}(x_m),$$

thus, $X_i \sim \mathcal{P}(p_i \lambda)$ and $X_1, \ldots, X_m$ are independent.

6.82 a) Since $X_1, \cdots, X_m$ are independent and identically distributed, then

$$P(\min\{X_1, ..., X_m\} > x) = P(X_1 > x, ..., X_m > x) = P(X_1 > x) \cdots P(X_m > x) = (P(X_1 > x))^m.$$

Thus,

$$P(\min\{X_1, \ldots, X_m\} = x) = P(\min\{X_1, \ldots, X_m\} > x - 1) - P(\min\{X_1, \ldots, X_m\} > x)$$
$$= (P(X_1 > x - 1))^m - (P(X_1 > x))^m.$$

b) Using Proposition 5.10 and (a),

$$P(\min\{X_1, \ldots, X_m\} > x) = (1 - p)^{mx}.$$

Therefore, for all $x \in \mathcal{N}$,

$$P(\min\{X_1, \cdots, X_m\} = x) = ((1 - p)^m)^{x-1} - ((1 - p)^m)^x = ((1 - p)^m)^{x-1}(1 - (1 - p)^m),$$

implying that $\min\{X_1, \cdots, X_m\} \sim \mathcal{G}(1 - (1 - p)^m)$.

c) The maximum of several random variables is less than a value x if and only if all of the random variables are less than x. Thus, since $X_1, \cdots, X_m$ are independent and identically distributed,

$$P(\max\{X_1, ..., X_m\} < x) = P(X_1 < x, ..., X_m < x) = P(X_1 < x) \cdots P(X_m < x) = (P(X_1 < x))^m.$$

Therefore,

$$P(\max\{X_1, \cdots, X_m\} = x) = P(\max\{X_1, \cdots, X_m\} < x + 1) - P(\max\{X_1, \cdots, X_m\} < x)$$
$$= (P(X_1 < x + 1))^m - (P(X_1 < x))^m = (P(X_1 \leq x))^m - (P(X_1 \leq x - 1))^m$$
$$= (1 - P(X_1 > x))^m - (1 - P(X_1 > x - 1))^m.$$

d) In the case when each X_i is geometrically distributed with a common parameter $p \in (0, 1)$, we have for $x \in \mathcal{N}$,

$$P(\max\{X_1, \ldots, X_m\} = x) = (1 - (1 - p)^x)^m - (1 - (1 - p)^{x-1})^m.$$

Let us show that $\max\{X_1, \cdots, X_m\}$ is not a geometric random variable. Suppose the converse, i.e. $\max\{X_1, \cdots, X_m\} \sim \mathcal{G}(q)$. Then we must have that

$$P(\max\{X_1, \cdots, X_m\} = 1) = q = p^m,$$

and, thus, in view of (a),

$$(1 - (1 - p)^2)^m - p^m = P(\max\{X_1, \cdots, X_m\} = 2) = q(1 - q) = p^m(1 - p^m).$$

In other words, we must have that

$$p^m(2-p)^m - p^m = p^m(1-p^m), \text{ i.e. } (2-p)^m = 2-p^m,$$

but the latter equality holds for $m \geq 2$ only if $p = 1$. Thus, for $m = 2, 3, \ldots$, $\max\{X_1, \cdots, X_m\}$ is not a geometric random variable.

6.83 a) The PMF of a constant random variable X is

$$p_X(x) = \begin{cases} 1, & x = c \\ 0, & x \neq c \end{cases}$$

b) First suppose that X is independent of itself. Let us show that it must be a constant random variable. If X is independent of itself,

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

for all x, y. Specifically, if $x = y$, then $P(X = x) = (P(X = x))^2$ for all x. Hence

$$(P(X = x))^2 - P(X = x) = 0 \qquad \Rightarrow \qquad P(X = x)(P(X = x) - 1) = 0$$

And thus $P(X = x)$ must equal either 0 or 1 for all x. And therefore, it must equal 1 for exactly one value of x, and 0 for all other values of x, and thus it is constant.

Next, assume that X is a constant random variable equal to c. For any x, y, we have

$$P(X = x, X = y) = \begin{cases} 1, & \text{if } x = y = c \\ 0, & \text{if otherwise} \end{cases}$$

Additionally,

$$P(X = x)P(X = y) = 0 \qquad \text{unless } x = y = c$$

and $P(X = x)P(X = y) = 1 \times 1 = 1$ when $x = y = c$. Thus,

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

for all x and y, and therefore, the constant random variable X is independent of itself.

c) First, assume that X is independent of all other random variables defined on the same sample space. Then X and $-X$ are independent, implying that

$$p_X(x)p_{-X}(-x) = p_{X,-X}(x, -x) = P(X = x,\, -X = -x) = P(X = x) = p_X(x),$$

thus, for each $x \in \mathcal{R}$, either $p_X(x) = 0$ or $p_{-X}(-x) = p_X(x) = 1$. Therefore, X equals to a constant with probability one.

Next assume that $P(X = c) = 1$ for some $c \in \mathcal{R}$ ($P(X = x) = 0$ for all $x \neq c$). Then for any given random variable Y defined on the same probability space, for all $y \in \mathcal{R}$,

$$p_{X,Y}(c, y) = P(X = c, Y = y) = P(Y = y) = 1 \cdot p_Y(y) = p_X(c)p_Y(y),$$

and $p_{X,Y}(x, y) = P(X = x, Y = y) = 0 = 0 \cdot P(Y = y) = p_X(x)p_Y(y)$ for all $x \neq c$, for all $y \in \mathcal{R}$. Thus, a constant random variable is independent of any other random variable.

6.84 a) $P(X = x, Y = y \,|\, Z = z) = P(X = x \,|\, Z = z)P(Y = y \,|\, Z = z)$ for all $z \in \mathcal{R}$ such that $P(Z = z) > 0$.
b) X_1 and X_3 are not independent, since $P(X_1 = 1, X_3 = -3) = 0 \neq p(1-p)^3 = p_{X_1}(1)p_{X_3}(-3)$.
c) X_1 and X_3 are conditionally independent given X_2.
For $X_2 = -2$,

$$P(X_1 = -1, X_3 = -3|X_2 = -2) = 1-p = 1\cdot(1-p) = P(X_1 = -1|X_2 = -2)P(X_3 = -3|X_2 = -2)$$

$$P(X_1 = -1, X_3 = -1|X_2 = -2) = p = 1 \cdot p = P(X_1 = -1|X_2 = -2)P(X_3 = -1|X_2 = -2)$$

and $P(X_1 = x_1, X_3 = x_3|X_2 = -2) = 0 = P(X_1 = x_1|X_2 = -2)P(X_3 = x_3|X_2 = -2)$ for all other (x_1, x_3).
For $X_2 = 0$,

$$P(X_1 = -1, X_3 = -1|X_2 = 0) = \frac{1-p}{2} = \frac{1}{2}\cdot(1-p) = P(X_1 = -1|X_2 = 0)P(X_3 = -1|X_2 = 0)$$

$$P(X_1 = -1, X_3 = 1|X_2 = 0) = \frac{p}{2} = \frac{1}{2}\cdot p = P(X_1 = -1|X_2 = 0)P(X_3 = 1|X_2 = 0)$$

$$P(X_1 = 1, X_3 = -1|X_2 = 0) = \frac{1-p}{2} = \frac{1}{2}\cdot(1-p) = P(X_1 = 1|X_2 = 0)P(X_3 = -1|X_2 = 0)$$

$$P(X_1 = 1, X_3 = 1|X_2 = 0) = \frac{p}{2} = \frac{1}{2}\cdot p = P(X_1 = 1|X_2 = 0)P(X_3 = 1|X_2 = 0)$$

and $P(X_1 = x_1, X_3 = x_3|X_2 = 0) = 0 = P(X_1 = x_1|X_2 = 0)P(X_3 = x_3|X_2 = 0)$ for all other (x_1, x_3).
For $X_2 = 2$,

$$P(X_1 = 1, X_3 = 1|X_2 = 2) = 1 - p = 1 \cdot (1 - p) = P(X_1 = 1|X_2 = 2)P(X_3 = 1|X_2 = 2)$$

$$P(X_1 = 1, X_3 = 3|X_2 = 2) = p = 1 \cdot p = P(X_1 = 1|X_2 = 2)P(X_3 = 3|X_2 = 2)$$

and $P(X_1 = x_1, X_3 = x_3|X_2 = 2) = 0 = P(X_1 = x_1|X_2 = 2)P(X_3 = x_3|X_2 = 2)$ for all other (x_1, x_3).
d) X_1 and X_2 are not independent. We have $P(X_1 = 1, X_2 = -2) = 0 \neq p(1 - p)^2 = p_{X_1}(1)p_{X_2}(-2)$.
e) X_1 and X_2 are not conditionally independent given X_3 since

$$P(X_1 = -1, X_2 = 0 \,|\, X_3 = -1) = \frac{1}{3}$$

but $P(X_1 = -1 \,|\, X_3 = -1)P(X_2 = 0 \,|\, X_3 = -1) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$.

6.5 Functions of Two or More Discrete Random Variables

Basic Exercises

6.85 a)

z	4	5	6	7	8	9
$p_{X+Y}(z)$	0.06	0.28	0.28	0.26	0.10	0.02

b)

z	-1	0	1	2
$p_{X-Y}(z)$	0.06	0.4	0.5	0.04

c)

z	2	3	4	5
$p_{\max\{X,Y\}}(z)$	0.06	0.52	0.4	0.02

d)

z	2	3	4
$p_{\min\{X,Y\}}(z)$	0.38	0.5	0.12

6.86 a)

z	0	1	2
$p_{X-Y}(z)$	$\frac{19}{40}$	$\frac{19}{40}$	$\frac{1}{20}$

b) $X - Y$ represents tha number of brothers a student has.

6.87 a)

z	0	1	2	3	4	5
$p_{Y-X}(z)$	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$

b) $Y - X$ is the magnitude of the difference between the two values of the die.

6.88 a) Let X, Y be the arrival times of the two people, and let Z be the number of minutes that the first person to arrive waits for the second person to arrive, i.e. $Z = \max\{X, Y\} - \min\{X, Y\}$. Clearly, the possible values of Z are $0, 1, \cdots, 60$, and

$$p_Z(0) = \sum_{x=0}^{60} p_{X,Y}(x, x) = \sum_{x=0}^{60} \frac{1}{61^2} = \frac{1}{61},$$

and for $z \in \{1, \cdots, 60\}$, using the symmetry $(p_{X,Y}(x, y) = p_{X,Y}(y, x))$,

$$p_Z(z) = 2\sum_{x=0}^{60-z} p_{X,Y}(x, x + z) = 2\sum_{x=0}^{60-z} \frac{1}{61^2} = \frac{2(61 - z)}{61^2} = \frac{2(61 - z)}{3721},$$

and $p_Z(z) = 0$ otherwise.

b) $P(Z \le 10) = \sum_{z=0}^{10} p_Z(z) = \frac{1171}{3721} \approx 0.3147$

c) $p_{\min\{X,Y\}}(z) = \frac{121-2z}{3721}$ for $z \in \{0, \cdots, 60\}$ and $p_{\min\{X,Y\}}(z) = 0$ for all $z \notin \{0, \cdots, 60\}$, since

$$P(\min\{X, Y\} = z) = p_{X,Y}(z, z) + 2P(X = z, Y > z) = \frac{1}{61^2} + 2 \cdot \frac{1}{61} \cdot \frac{60 - z}{61} = \frac{121 - 2z}{3721}$$

for all $z \in \{0, \cdots, 60\}$.

d) $p_{\max\{X,Y\}}(z) = \frac{2z+1}{3721}$ for $z \in \{0, \cdots, 60\}$ and $p_{\min\{X,Y\}}(z) = 0$ for all $z \notin \{0, \cdots, 60\}$, since

$$\begin{aligned} P(\max\{X, Y\} = z) &= \sum_{k=0}^{z} P(\max\{X, Y\} = z, \min\{X, Y\} = k) \\ &= \frac{1}{61^2} + \sum_{k=0}^{z-1} \frac{2}{61^2} = \frac{2z + 1}{3721}, \quad z = 0, \cdots, 60. \end{aligned}$$

6.89 a) Let $Z = |X - Y|$. Then there are $N+1$ different ways that $Z = 0$. Thus $P(Z=0) = \frac{N+1}{(N+1)^2} = \frac{1}{N+1}$. Assuming that $X > Y$, if $Z = z$ for $z = 1, \ldots, N$, then $X - Y = z$. There are $N+1-z$ ways for this to happen. Thus there are $2(N+1-z)$ ways for $Z = z$ for $z = 1, \ldots, N$, hence $P(Z = z) = \frac{2(N+1-z)}{(N+1)^2}$ for $z = 1, \ldots, N$ and $P(Z = z) = 0$ otherwise.
b) For $z = 1, \cdots, N$,

$$p_Z(z) = \sum_{x=0}^{N-z} p_{X,Y}(x, x+z) + \sum_{y=0}^{N-z} p_{X,Y}(y+z, y) = \frac{2(N+1-z)}{(N+1)^2};$$

$$p_Z(0) = \sum_{x=0}^{N} p_{X,Y}(x, x) = \sum_{x=0}^{N} \frac{1}{(N+1)^2} = \frac{1}{N+1},$$

and $p_Z(z) = 0$ for all $z \notin \{0, \cdots, N\}$.

6.90 a) Let $Z = g(X, Y) = X + Y$. Then for $z = 0, \ldots, N$,

$$p_Z(z) = \sum_{x=0}^{z} p_{X,Y}(x, z-x) = \sum_{x=0}^{z} \frac{1}{(N+1)^2} = \frac{z+1}{(N+1)^2};$$

for $z = N+1, \cdots, 2N$,

$$p_Z(z) = \sum_{x=z-N}^{N} p_{X,Y}(x, z-x) = \sum_{x=z-N}^{N} \frac{1}{(N+1)^2} = \frac{2N-z+1}{(N+1)^2}.$$

Thus, $p_Z(z) = \frac{N+1-|N-z|}{(N+1)^2}$ for $z = 0, \ldots, 2N$, and $p_Z(z) = 0$ otherwise.
b) Note that for $z \in \{0, \cdots, N\}$,

$$\begin{aligned} P(\min\{X,Y\} = z) &= \sum_{k=z}^{N} P(\min\{X,Y\} = z, \max\{X,Y\} = k) = p_{X,Y}(z,z) + \sum_{k=z+1}^{N} \frac{2}{(N+1)^2} \\ &= \frac{1}{(N+1)^2} + \frac{2(N-z)}{(N+1)^2} = \frac{2(N-z)+1}{(N+1)^2}; \end{aligned}$$

and $p_{\min\{X,Y\}}(z) = 0$ for all $z \notin \{0, \cdots, N\}$.
c) For $z \in \{0, \cdots, N\}$,

$$\begin{aligned} P(\max\{X,Y\} = z) &= \sum_{k=0}^{z} P(\min\{X,Y\} = k, \max\{X,Y\} = z) = p_{X,Y}(z,z) + \sum_{k=0}^{z-1} \frac{2}{(N+1)^2} \\ &= \frac{1}{(N+1)^2} + \frac{2z}{(N+1)^2} = \frac{2z+1}{(N+1)^2}; \end{aligned}$$

and $p_{\max\{X,Y\}}(z) = 0$ for all $z \notin \{0, \cdots, N\}$.

6.91 a) Following Example 6.22(b), we have

$$p_{\min\{X,Y\}}(z) = \sum\sum\nolimits_{\min\{x,y\}=z} p_{X,Y}(x,y) = p_{X,Y}(z,z) + \sum_{y=z+1}^{\infty} p_{X,Y}(z,y) + \sum_{x=z+1}^{\infty} p_{X,Y}(x,z)$$

$$= pq((1-p)(1-q))^{z-1} + \sum_{y=z+1}^{\infty} pq(1-p)^{z-1}(1-q)^{y-1} + \sum_{x=z+1}^{\infty} pq(1-p)^{x-1}(1-q)^{z-1}$$

$$= pq((1-p)(1-q))^{z-1} + pq(1-p)^{z-1}(1-q)^{z}\sum_{y=0}^{\infty}(1-q)^{y} + pq(1-p)^{z}(1-q)^{z-1}\sum_{x=0}^{\infty}(1-p)^{x}$$

$$= pq((1-p)(1-q))^{z-1} + p(1-p)^{z-1}(1-q)^{z} + q(1-p)^{z}(1-q)^{z-1}$$

$$= pq((1-p)(1-q))^{z-1} + p(1-q)((1-p)(1-q))^{z} + q(1-p)((1-p)(1-q))^{z-1}$$

$$= (pq + p(1-q) + q(1-p))(1-(p+q-pq))^{z-1}$$

$$= (p+q-pq)(1-(p+q-pq))^{z-1}$$

for $z \in \mathcal{N}$ and $p_{\min\{X,Y\}}(x,y) = 0$ otherwise.

b) Let $Z = \min\{X, Y\}$. Then using tail probabilities, $P(Z = z) = P(Z \geq z) - P(Z \geq z+1)$. Now

$$P(Z \geq z) = P(X \geq z, Y \geq z) = P(X \geq z)P(Y \geq z) = \left(\sum_{x=z}^{\infty} P(X = x)\right)\left(\sum_{y=z}^{\infty} P(Y = y)\right)$$

$$= \left(\sum_{x=z}^{\infty} p(1-p)^{x-1}\right)\left(\sum_{y=z}^{\infty} q(1-q)^{y-1}\right)$$

$$= p(1-p)^{z-1}\left(\sum_{x=0}^{\infty}(1-p)^{x}\right) q(1-q)^{z-1}\left(\sum_{y=0}^{\infty}(1-q)^{y}\right)$$

$$= ((1-p)(1-q))^{z-1}$$

Similarily, $P(Z \geq z+1) = ((1-p)(1-q))^{z}$. Thus $P(Z = z) = ((1-p)(1-q))^{z-1} - ((1-p)(1-q))^{z} = (p+q-pq)(1-(p+q-pq))^{z-1}$, Hence, $Z \sim \mathcal{G}(p+q-pq)$.

c) If $p = q$, then $P(Z = z) = (2p - p^2)(1 - (2p - p^2)^{z-1}$ which agrees with Example 6.22 (b).

Theory Exercises

6.92 Note that

$$P(g(X_1, \cdots, X_m) = z) = P\left(\bigcup_{(x_1,\cdots,x_m)\in g^{-1}(\{z\})} \{(X_1, \cdots, X_m) = (x_1, \cdots, x_m)\}\right)$$

$$= P\left(\bigcup_{(x_1,\cdots,x_m)\in g^{-1}(\{z\}):\ p_{X_1,\cdots,X_m}(x_1,\cdots,x_m)>0} \{(X_1, \cdots, X_m) = (x_1, \cdots, x_m)\}\right)$$

$$= \sum_{(x_1,\cdots,x_m)\in g^{-1}(\{z\}):\ p_{X_1,\cdots,X_m}(x_1,\cdots,x_m)>0} p_{X_1,\cdots,X_m}(x_1, \cdots, x_m)$$

$$= \sum_{(x_1,\cdots,x_m)\in g^{-1}(\{z\})} p_{X_1,\cdots,X_m}(x_1, \cdots, x_m),$$

where we used the fact that the set of $(x_1, \cdots, x_m)$, such that $p_{X_1,\cdots,X_m}(x_1, \cdots, x_m) > 0$, is countable, and then applied the countable additivity of the probability measure. (The last equality holds since arbitrary sums of zeros are always equal to zero).

6.93 a) Let $g(x, y) = x + y$ and $g(X, Y) = Z$. Then $g^{-1}(\{z\}) = \{(x, y) : x + y = z\} = \{(x, y) : y = z - x\}$. Thus $g^{-1}(\{z\})$ is the set where x can vary freely, and $y = z - x$. Therefore,

$$p_Z(z) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x, y) = \sum_x p_{X,Y}(x, z - x).$$

b) Let $g(x, y) = y - x$ and $g(X, Y) = Z$. Then $g^{-1}(\{z\}) = \{(x, y) : y - x = z\} = \{(x, y) : y = x + z\}$. Thus $g^{-1}(\{z\})$ is the set where x can vary freely, and $y = x + z$. Therefore,

$$p_Z(z) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x, y) = \sum_x p_{X,Y}(x, x + z).$$

c) If X and Y are independent, then

$$p_{X+Y}(z) = \sum_x p_X(x)p_Y(z - x)$$

and

$$p_{Y-X}(z) = \sum_x p_X(x)p_Y(x + z).$$

6.94 First, since

$$\{X + Y = z\} = \bigcup_x \{X + Y = z, X = x\} = \bigcup_x \{X = x, Y = z - x\}$$

we have $p_{X+Y}(z) = \sum_x p_{X,Y}(x, z - x)$. Second, since

$$\{Y - X = z\} = \bigcup_x \{Y - X = z, X = x\} = \bigcup_x \{X = x, Y = x + z\}$$

we have $p_{Y-X}(z) = \sum_x p_{X,Y}(x, x + z)$.

6.95 a) Let $g(x, y) = xy$ and $g(X, Y) = Z$. Then $g^{-1}(\{z\}) = \{(x, y) : xy = z\} = \{(x, y) : y = z/x\}$. Thus $g^{-1}(\{z\})$ is the set where x can vary freely, and $y = z/x$. Therefore,

$$p_Z(z) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x, y) = \sum_x p_{X,Y}(x, z/x).$$

b) Let $g(x, y) = y/x$ and $g(X, Y) = Z$. Then $g^{-1}(\{z\}) = \{(x, y) : y/x = z\} = \{(x, y) : y = xz\}$. Thus $g^{-1}(\{z\})$ is the set where x can vary freely, and $y = xz$. Therefore,

$$p_Z(z) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x, y) = \sum_x p_{X,Y}(x, xz).$$

c) If X and Y are independent, then

$$p_{XY}(z) = \sum_x p_X(x) p_Y(z/x)$$

and

$$p_{Y/X}(z) = \sum_x p_X(x) p_Y(xz).$$

6.96 First, since

$$\{XY = z\} = \bigcup_x \{XY = z, X = x\} = \bigcup_x \{X = x, Y = z/x\}$$

we have $p_{XY}(z) = \sum_x p_{X,Y}(x, z/x)$. Second, since

$$\{Y/X = z\} = \bigcup_x \{Y/X = z, X = x\} = \bigcup_x \{X = x, Y = xz\}$$

we have $p_{Y/X}(z) = \sum_x p_{X,Y}(x, xz)$.

Advanced Exercises

6.97 a) $X + Y \sim \mathcal{B}(m + n, p)$. Observing m independent trial each with success probability p and then independently observing n independent trails each with success probability p is the same as observing $m + n$ independent trials each with success probability p. Moreover, if X is the number of successes in m independent trials and Y is the number of successes in the next n independent trials, then $X + Y$ is the number of successes in $m + n$ independent trials (with probability of success p in each trial).
b) Using Proposition 6.16 and the convention that $\binom{b}{a} = 0$ if either $a > b$ or $a < 0$, we have that for $z \in \{0, \cdots, m + n\}$,

$$P(X + Y = z) = \sum\sum_{x+y=z} p_{X,Y}(x, y) = \sum_{x=0}^{z} P(X = x, Y = z - x) = \sum_{x=0}^{z} P(X = x)P(Y = z - x)$$

$$= \sum_{x=0}^{z} \binom{m}{x} p^x (1-p)^{n-x} \binom{n}{z-x} p^{z-x} (1-p)^{n-z-x} = p^z (1-p)^{m+n-z} \underbrace{\sum_{x=0}^{z} \binom{m}{x}\binom{n}{z-x}}_{=\binom{n+m}{k} \text{ by Vandermonde's identity}}$$

$$= \binom{m+n}{z} p^z (1-p)^{m+n-z},$$

and, therefore, $X + Y \sim \mathcal{B}(m + n, p)$
c) Claim : $X_1 + \ldots + X_m \sim \mathcal{B}(n_1 + \ldots + n_m, p)$. Proof is done by induction.

By (b), the claim is true for $m = 2$. Next assume that the claim is true for all $m \leq k$, and let $Y = X_1 + \ldots + X_k$. Since X_{k+1} is independent of $X_1, \ldots, X_k$, then X_{k+1} is independent of a function of $X_1, \ldots, X_k$, namely their sum, Y. Thus, by the induction assumption, X_{k+1} and

Y are two independent Binomial random variables with identical success probability p and the number of trials equal to n_{k+1} and $n_1 + \ldots + n_k$ respectively. Then using (b), we have that $X_1 + \ldots + X_k + X_{k+1} = X_{k+1} + Y \sim \mathcal{B}(n_1 + \ldots + n_{k+1}, p)$ and therefore the claim is true for $m = k+1$. Thus by the principle of mathematical induction, the claim is true for all $m \in \mathcal{N}$.

6.98 a) Using the binomial theorem, we have for $z \in \mathcal{Z}_+$,

$$P(X+Y=z) = \sum_{x=0}^{z} P(X=x, Y=z-x) = \sum_{x=0}^{z} P(X=x)P(Y=z-x)$$

$$= \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\mu}\mu^{z-x}}{(z-x)!} = \frac{e^{-(\lambda+\mu)}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!}\lambda^x \mu^{z-x}$$

$$= \frac{e^{-(\lambda+\mu)}}{z!} \underbrace{\sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x}}_{=(\lambda+\mu)^z} = \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}.$$

Hence, $X + Y \sim \mathcal{P}(\lambda + \mu)$.
b) Claim : $X_1 + \ldots + X_m \sim \mathcal{P}(\lambda_1 + \ldots + \lambda_m)$. Proof by induction.

By (a), the claim is true for $m = 2$. Assume that the claim is true for all $m \leq k$. Let $Y = X_1 + \ldots + X_k$, then $Y \sim \mathcal{P}(\lambda_1 + \cdots + \lambda_k)$. Since X_{k+1} is independent of $X_1, \ldots, X_k$, it is independent of a function of $X_1, \ldots, X_k$, namely their sum, Y. Thus X_{k+1} and Y are two independent Poisson random variables with rates, λ_{k+1} and $\lambda_1 + \ldots + \lambda_k$ respectively. Then using (a), we have that $X_1 + \ldots + X_k + X_{k+1} = X_{k+1} + Y \sim \mathcal{P}(\lambda_1 + \ldots + \lambda_{k+1})$ and therefore the claim is valid for $m = k+1$. Thus by the induction principle, the claim is true for all $m \in \mathcal{N}$.

6.99 Let K_X be the countable subset of $\mathcal{R}$ such that $P(X \in K_X) = 1$ and K_Y be the countable subset of $\mathcal{R}$ such that $P(Y \in K_Y) = 1$. Then $K_X \times K_Y$ is a countable subset of $\mathcal{R}^2$ and, if we let $G = \{g(x,y) : (x,y) \in K_X \times K_Y\}$, then G is countable by Exercise 1.39. Moreover, $P(g(X,Y) \in G) = P((X,Y) \in K_X \times K_Y) = 1$. Thus, $g(X,Y)$ is a discrete random variable.

6.6 Sums of Discrete Random Variables

Basic Exercises

6.100 Using Proposition 6.18, we have:

$$p_{X+Y}(4) = \sum_x p_{X,Y}(x, 4-x) = p_{X,Y}(2,2) = 0.06$$

$$p_{X+Y}(5) = \sum_x p_{X,Y}(x, 5-x) = p_{X,Y}(3,2) = 0.28$$

$$p_{X+Y}(6) = \sum_x p_{X,Y}(x, 6-x) = p_{X,Y}(4,2) + p_{X,Y}(3,3) = 0.04 + 0.24 = 0.28$$

$$p_{X+Y}(7) = \sum_x p_{X,Y}(x, 7-x) = p_{X,Y}(4,3) + p_{X,Y}(3,4) = 0.22 + 0.04 = 0.26$$

$$p_{X+Y}(8) = \sum_x p_{X,Y}(x, 8-x) = p_{X,Y}(4,4) = 0.1$$

$$p_{X+Y}(9) = \sum_x p_{X,Y}(x, 9-x) = p_{X,Y}(4,5) = 0.02$$

6.101 a) The random variable $M = \dfrac{X+Y}{2}$ is the mean value of two faces obtained in two tosses of the die.
b) The PMF of M is given by:

$$p_M(z) = \begin{cases} \dfrac{2z-1}{36}, & \text{if } z \in \{1, 1.5, 2, 2.5, 3, 3.5\}, \\ \dfrac{13-2z}{36}, & \text{if } z \in \{4, 4.5, 5, 5.5, 6\} \\ 0, & \text{otherwise.} \end{cases}$$

Indeed, by Proposition 6.18,

$$p_{X+Y}(z) = \sum_{x=1}^{z-1} p_{X,Y}(x, z-x) = \sum_{x=1}^{z-1} \frac{1}{36} = \frac{z-1}{36}, \quad \text{for } z \in \{2, \cdots, 7\},$$

$$p_{X+Y}(z) = \sum_{x=z-6}^{6} p_{X,Y}(x, z-x) = \frac{6-(z-6-1)}{36} = \frac{13-z}{36}, \quad \text{for } z \in \{8, \cdots, 12\},$$

and $p_{X+Y}(z) = 0$ for all $z \notin \{2, \cdots, 12\}$. Then by Proposition 5.14,

$$p_M(z) = p_{X+Y}(2z) = \begin{cases} \frac{2z-1}{36}, & \text{if } 2z \in \{2, \cdots, 7\}, \\ \frac{13-2z}{36}, & \text{if } 2z \in \{8, \cdots, 12\}, \\ 0, & \text{otherwise.} \end{cases}$$

The required conclusion for PMF of $M = (X+Y)/2$ then follows.

6.102 Let $Z = X + Y$. Then using Proposition 6.19, for $z = 0, \ldots, N$

$$p_Z(z) = p_{X+Y}(z) = \sum_{x=0}^{z} p_{X,Y}(x, z-x) = \sum_{x=0}^{z} \frac{1}{(N+1)^2} = \frac{z+1}{(N+1)^2},$$

and for $z = N+1, \ldots, 2N$

$$p_Z(z) = p_{X+Y}(z) = \sum_{x=z-N}^{2N} p_{X,Y}(x, z-x) = \sum_{x=z-N}^{2N} \frac{1}{(N+1)^2} = \frac{2N-z+1}{(N+1)^2},$$

which is the same answer as obtained earlier in Exercise 6.90(a).

6.103 No, the sum of two independent geometric random variables with the same success probability parameter p is not a geometric random variable. It is instead a negative binomial random variables with parameters 2 and the common success probability p.

6.104 Using Proposition 6.19, if X is the number of claims in the first week, Y is the number of claims in the second week, and $Z = X + Y$ is the total number of claims received during the two weeks, then,

$$p_Z(z) = \sum_{x=0}^{z} P(X = x)P(Y = z - x) = \sum_{x=0}^{z} 2^{-(x+1)}2^{-(z-x+1)} = \sum_{x=0}^{z} 2^{-(z+2)} = \frac{z+1}{2^{z+2}}$$

for $z = 0, 1, \ldots$ and $p_Z(z) = 0$ otherwise.

6.105 a) For $u \in \mathcal{N}$,

$$p_{X+1}(u) = p_X(u-1) = \frac{1}{2^u} = \frac{1}{2}(1 - \frac{1}{2})^{u-1}, \text{ thus, } X + 1 \sim \mathcal{G}(0.5).$$

Similarly, $Y + 1$ is $\mathcal{G}(0.5)$.
b) By Exercise 6.103, $X + Y + 2 \sim \mathcal{NB}(2, 0.5)$, since independence of X and Y implies independence of $X + 1$ and $Y + 1$.
c) Using (b) and Proposition 6.17,

$$p_{X+Y}(z) = p_{X+Y+2}(z+2) = \binom{z+1}{1}\frac{1}{2^{z+2}} = \frac{z+1}{2^{z+2}}$$

for $z = 0, 1, 2 \cdots$, and $p_{X+Y}(z) = 0$ for all $z \notin \mathcal{Z}_+$. The result agrees with the answer obtained in Exercise 6.104.

6.106 a) Part (a) of Proposition 6.20 says that the sum of independent Binomial random variables with varying n and identical probabilities of success will also be a Binomial random variable with the number of trials being the sum of all the number of trials from each of the random variables, and the success probability will remain the same. Part (b) of Proposition 6.20 says that the sum of independent Poisson random variables with varying parameters will be a Poisson random variable with its parameter the sum of each of the random variables parameters. Part (c) of Proposition 6.20 says that the sum of independent Negative Binomial random varaibles with varying r and identical probabilities of success will also be a Negative Binomial random variable with r being the sum of all the r's from each of the random variables, and the success probability will remain the same.

6.107 From Exercise 6.106, $X_1 + \ldots + X_r \sim \mathcal{NB}(r, p)$.

6.108 Since $X + Y \sim \mathcal{P}(\lambda + \mu)$, then for $z = 0, 1, \ldots,$

$$p_{X|X+Y}(x|z) = \frac{P(X = x, X + Y = z)}{P(X + Y = z)} = \frac{P(X = x)P(Y = z - x)}{P(X + Y = z)} = \frac{\frac{e^{-(\lambda+\mu)}\lambda^x\mu^{z-x}}{x!(z-x)!}}{\frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}}$$

$$= \frac{\lambda^x\mu^{z-x}z!}{(\lambda+\mu)^z x!(z-x)!} = \binom{z}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^x\left(\frac{\mu}{\lambda+\mu}\right)^{z-x}, \quad x = 0, \cdots, z.$$

Thus $(X \mid X + Y = z) \sim \mathcal{B}\left(z, \frac{\lambda}{\lambda+\mu}\right)$.

6.109 a) Since $X + Y \sim \mathcal{NB}(r + s, p)$, then for $z = r + s, r + s + 1, \cdots,$

$$p_{X|X+Y}(x|z) = \frac{P(X = x, X + Y = z)}{P(X + Y = z)} = \frac{P(X = x)P(Y = z - x)}{P(X + Y = z)}$$

$$= \frac{\left(\binom{x-1}{r-1}p^r(1-p)^{x-r}\right)\left(\binom{z-x-1}{s-1}p^s(1-p)^{z-x-s}\right)}{\binom{z-1}{(r+s)-1}p^{r+s}(1-p)^{z-(r+s)}} = \frac{\binom{x-1}{r-1}\binom{z-x-1}{s-1}}{\binom{z-1}{(r+s)-1}}, \quad x = r, r+1, \cdots, z-s.$$

b) Since a conditional PMF is a PMF, then $\sum_{x=r}^{z-s} p_{X|X+Y}(x|z) = 1$ and the desired result follows.
c) Using (b) and Proposition 6.19, for $z = r+s, r+s+1, \ldots$

$$p_{X+Y}(z) = \sum_x p_X(x)p_Y(z-x)$$

$$= p^{r+s}(1-p)^{z-(r+s)} \sum_{x=r}^{z-s} \binom{x-1}{r-1}\binom{z-x-1}{s-1}$$

$$= \binom{z-1}{r+s-1} p^{r+s}(1-p)^{z-(r+s)}.$$

Hence, $X + Y \sim \mathcal{NB}(r+s, p)$

6.110 Let X denote the number of people waiting when the Ice Cream Shoppe opens, Y denote the number of customers that arrive during the time that the shop is open, and let $Z = X + Y$ be the total number of customers. Then for $z \leq N$

$$p_Z(z) = \sum_{x=0}^{z} P(X=x)P(Y=z-x) = \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^{z-x}}{(N+1)(z-x)!} = \frac{e^{-\lambda}}{N+1}\sum_{x=0}^{z}\frac{\lambda^{z-x}}{(z-x)!}$$

and for $z > N$

$$p_Z(z) = \sum_{x=0}^{N} P(X=x)P(Y=z-x) = \sum_{x=0}^{N} \frac{e^{-\lambda}\lambda^{z-x}}{(N+1)(z-x)!} = \frac{e^{-\lambda}}{N+1}\sum_{x=0}^{N}\frac{\lambda^{z-x}}{(z-x)!}$$

and $p_Z(z) = 0$ otherwise.

6.111 a) For $z = 0, \ldots, m+n$, using the convention that $\binom{b}{a} = 0$ if $a < 0$ or $a > b$, we have that

$$p_{X+Y}(z) = \sum_{x=0}^{z} P(X=x)P(Y=z-x) = \sum_{x=0}^{z}\binom{m}{x}p^x(1-p)^{m-x}\binom{n}{z-x}q^{z-x}(1-q)^{n-z+x}$$

$$= (1-p)^m q^z (1-q)^{n-z} \sum_{x=0}^{z}\binom{m}{x}\binom{n}{z-x}\left(\frac{p(1-q)}{q(1-p)}\right)^x$$

and $p_{X+Y}(z) = 0$ otherwise.
b) If $p = q$, then from Proposition 6.20, $X + Y \sim \mathcal{B}(m+n, p)$.
c) Letting $p = q$ in (a), we have

$$p_{X+Y}(z) = (1-p)^m p^z (1-p)^{n-z} \sum_{x=0}^{z}\binom{m}{x}\binom{n}{z-x}\left(\frac{p(1-p)}{p(1-p)}\right)^x$$

$$= p^z(1-p)^{m+n-z} \underbrace{\sum_{x=0}^{z} \binom{m}{x}\binom{n}{z-x}}_{=\binom{m+n}{z}} = \binom{m+n}{z} p^z(1-p)^{m+n-z},$$

by the Vandermonde's identity, and thus $X+Y \sim \mathcal{B}(m+n,p)$.

6.112 a) The total claim amount distribution is compound Poisson.
b) The answers will vary.

Theory Exercises

6.113 Applying Proposition 6.16 to function $g(x+y) = y+x$ gives

$$p_{X+Y}(z) = \sum_{(x,y):\, x+y=z} p_{X,Y}(z) = \sum_{y} \sum_{x=z-y}^{z-y} p_{X,Y}(x,y) = \sum_{y} p_{X,Y}(z-y,y).$$

6.114 a) If X and Y are independent Poisson random variables with parameters λ and μ, then so is their sum with parameter $\lambda+\mu$. In otherwords, if X is counting the number of arrivals when the arrival rate is λ, and Y is counting the number of arrivals when the arrival rate is μ, then the total number of arrivals will have an arrival rate of $\lambda+\mu$.
b) Using the binomial theorem, we have

$$P(X+Y=z) = \sum_{x=0}^{z} P(X=x, Y=z-x) = \sum_{x=0}^{z} P(X=x)P(Y=z-x)$$

$$= \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\mu}\mu^{z-x}}{(z-x)!} = \frac{e^{-(\lambda+\mu)}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda^x \mu^{z-x}$$

$$= \frac{e^{-(\lambda+\mu)}}{z!} \underbrace{\sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x}}_{=(\lambda+\mu)^z} = \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}, \quad z = 0, 1, \ldots.$$

Hence, $X+Y \sim \mathcal{P}(\lambda+\mu)$.

6.115 Using the binomial series formula (5.45) we have that for $|t|<1$,

$$\begin{aligned}(1+t)^a(1+t)^b &= \left(\sum_{j=0}^{\infty}\binom{a}{j}t^j\right)\left(\sum_{k=0}^{\infty}\binom{b}{k}t^k\right) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{a}{j}\binom{b}{k}t^{j+k} \\ &= \sum_{j=0}^{\infty}\sum_{\ell=j}^{\infty}\binom{a}{j}\binom{b}{\ell-j}t^{\ell} = \sum_{\ell=0}^{\infty}\left\{\sum_{j=0}^{\ell}\binom{a}{j}\binom{b}{\ell-j}\right\}t^{\ell};\end{aligned}$$

on the other hand,

$$(1+t)^a(1+t)^b = (1+t)^{a+b} = \sum_{\ell=0}^{\infty}\binom{a+b}{\ell}t^{\ell}.$$

Thus, we have the following equality for corresponding polynomials (in t):

$$\sum_{\ell=0}^{\infty}\left\{\sum_{j=0}^{\ell}\binom{a}{j}\binom{b}{\ell-j}\right\}t^{\ell}=\sum_{\ell=0}^{\infty}\binom{a+b}{\ell}t^{\ell},$$

implying that the coefficients in front of t^{ℓ} on both sides of the equality are the same, namely, for $\ell=0,1,\cdots,$

$$\sum_{j=0}^{\ell}\binom{a}{j}\binom{b}{\ell-j}=\binom{a+b}{\ell}.$$

6.116 a) The sum of two independent negative binomial random variables with parameters (r,p) and (s,p) is again a negative binomial random variable with parameters $(r+s,p)$.
b) If X and Y are independent negative binomial random variables where X counts the number of trials until the r^{th} success and Y counts the number of trials until the s^{th} success, then their sum will count the number of trials until the $r+s^{th}$ success.
c) Let $Z=X+Y$. Then using the representation of the negative binomial PMF given by equation (5.44),

$$p_Z(z)=P(X+Y=z)=\sum_{x=r}^{z-s}P(X=x,Y=z-x)$$

$$=\sum_{x=r}^{z-s}\binom{-r}{x-r}p^r(1-p)^{x-r}\binom{-s}{z-x-s}p^s(1-p)^{z-x-s}$$

$$=p^{r+s}(1-p)^{z-(r+s)}\sum_{x=0}^{z-(r+s)}\binom{-r}{x}\binom{-s}{z-(r+s)-x}=p^{r+s}(1-p)^{z-(r+s)}\binom{-(r+s)}{z-(r+s)}$$

where the last step follows from Exercise 6.115. Therefore, $X+Y\sim\mathcal{NB}(r+s,p)$.

6.117 Proposition 6.20 (a) was proved in Exercise 6.97(c). Proposition 6.20 (c) was proved in Exercise 6.98(b). For Proposition 6.20 (b) we have the following:
Claim : $X_1+\ldots+X_m\sim\mathcal{NB}(r_1+\ldots+r_m,p)$. Proof is done by induction. By Exercise 6.116, the claim is true for $m=2$. Assume that the claim is true for all $m\leq k$. Let us show that the claim also holds for $m=k+1$. Let $Y=X_1+\ldots+X_k$. Since X_{k+1} is independent of $X_1,\ldots,X_k$, it is independent of a function of $X_1,\ldots,X_k$, namely their sum, Y. Thus X_{k+1} and Y are two independent Negative Binomial with identical probabilities of success and numbers of successes r_{k+1} and $r_1+\ldots+r_k$ respectively. Then $X_1+\ldots+X_k+X_{k+1}=X_{k+1}+Y\sim\mathcal{NB}(r_1+\ldots+r_{k+1},p)$ and, therefore, the claim is true for $m=k+1$. Thus, by induction, the claim is true for all $m\in\mathcal{N}$.

Advanced Exercises

6.118 We have

$$p_{X+Y+Z}(w)=\sum\sum\sum_{x+y+z=w}p_{X,Y,Z}(x,y,z)=\sum\sum\sum_{x+y+z=w}p_X(x)p_Y(y)p_Z(z)$$

$$=\sum_x p_X(x)\left(\sum_y p_Y(y)p_Z(w-x-y)\right)=\sum_x p_X(x)p_{Y+Z}(w-x),$$

by Proposition 6.18.

6.119 a) Note that for all $\omega \in \mathcal{R}$,

$$\{X_1 + \ldots + X_m = w\} = \bigcup_{(x_1,\ldots,x_{m-1})} \cdots \bigcup \{X_1 = x_1, \ldots, X_{m-1} = x_{m-1}, X_m = w - (x_1 + \ldots + x_{m-1})\}$$

and the desired result follows from countable additivity of the probability measure and the fact that the set of $(x_1, \cdots, x_{m-1})$, where $p_{X_1,\cdots,X_{m-1}}(x_1, \cdots, x_{m-1}) > 0$, is countable.
b) In addition to (a), we have for $i = 1, \cdots, m-1$,

$$p_{X_1+\cdots+X_m}(w) = \sum_{(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_m)} \sum p_{X_1,\cdots,X_m}\left(x_1, \cdots, x_{i-1}, w - \sum_{k\in\{1,\cdots,m\}\setminus\{i\}} x_k, x_{i+1}, \cdots, x_m\right).$$

c) If $X_1, \cdots, X_m$ are independent, then for $i = 1, \cdots, m$,

$$p_{X_1+\cdots+X_m}(w) = \sum_{(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_m)} \sum \left(\prod_{k\in\{1,\cdots,m\}\setminus\{i\}} p_{X_k}(x_k)\right) p_{X_i}\left(w - \sum_{k\in\{1,\cdots,m\}\setminus\{i\}} x_k\right).$$

d) For three random variables X, Y, Z, we have

$$p_{X+Y+Z}(w) = \sum_{(x,y)} \sum p_{X,Y,Z}(x, y, w - (x+y))$$

or

$$p_{X+Y+Z}(w) = \sum_{(x,z)} \sum p_{X,Y,Z}(x, w - (x+z), z)$$

or

$$p_{X+Y+Z}(w) = \sum_{(y,z)} \sum p_{X,Y,Z}(w - (y+z), y, z).$$

If X, Y, Z are independent, then

$$p_{X+Y+Z}(w) = \sum_{(x,y)} \sum p_X(x) p_Y(y) p_Z(w - x - y)$$

or

$$p_{X+Y+Z}(w) = \sum_{(x,z)} \sum p_X(x) p_Y(w - x - z) p_Z(z)$$

or

$$p_{X+Y+Z}(w) = \sum_{(y,z)} \sum p_X(w - y - z) p_Y(y) p_Z(z).$$

6.120 For $m = 2, 3, \ldots$, by independence of $X_1, \cdots, X_m$ and Exercise 6.19(c), we have

$$\begin{aligned} p_{X_1+\cdots+X_m}(w) &= \sum_{(x_1,\cdots,x_{m-1})} \cdots \sum \left(\prod_{k=1}^{m-1} p_{X_k}(x_k)\right) p_{X_m}(w - (x_1 + \cdots + x_{m-1})) \\ &= \sum_{(x_1,\cdots,x_{m-1})} \cdots \sum p_{X_1,\cdots,X_{m-1}}(x_1, \cdots, x_{m-1}) p_{X_m}(w - (x_1 + \cdots + x_{m-1})) \end{aligned}$$

$$= \sum_x \left(\sum_{x_1+\ldots+x_{m-1}=x} \cdots \sum p_{X_1,\cdots,X_{m-1}}(x_1,\cdots,x_{m-1}) \right) p_{X_m}(w-x)$$

$$= \sum_x \underbrace{\left(\sum_{(x_1,\ldots,x_{m-2})} \cdots \sum p_{X_1,\cdots,X_{m-1}}\left(x_1,\cdots,x_{m-2},x-(x_1+\cdots+x_{m-2})\right) \right)}_{=p_{X_1+\cdots+X_{m-1}}(x),\ \text{by Exercise 6.19(a)}} p_{X_m}(w-x)$$

$$= \sum_x p_{X_1+\cdots+X_{m-1}}(x) p_{X_m}(w-x), \quad \text{for each } w \in \mathcal{R}.$$

6.121 The PMF of $X+Y+Z$ is given by:

w	1	2	3	4	5
$p_{X+Y+Z}(w)$	$\frac{4}{63}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{17}{63}$	$\frac{1}{9}$

6.7 Review Exercises

Basic Exercises

6.122 a)

		College, Y		
		1	2	$p_X(x)$
Gender, X	1	0.171	0.257	0.429
	2	0.229	0.343	0.571
	$p_Y(y)$	0.4	0.6	1.0

b) Because the first column must sum to 14, and there are 15 different combinations of non-negative integers that sum to 14, there are 15 different ways of filling in the blanks.
c) Since there are 15 different ways of filling in the table, there are 15 different joint PMF's that have the same pair of marginal PMFs.

6.123 a) $p_{X,Y}(3,6) = \frac{35}{250} = 0.14$, which is the probability that a family has 3 televisions and 6 radios.
b)

		Radios, Y						
		3	4	5	6	7	8	$p_X(x)$
Televisions, X	0	0.000	0.004	0.004	0.004	0.004	0.000	0.016
	1	0.000	0.024	0.060	0.072	0.008	0.000	0.164
	2	0.004	0.032	0.128	0.176	0.056	0.004	0.400
	3	0.000	0.024	0.148	0.140	0.036	0.004	0.352
	4	0.000	0.004	0.028	0.024	0.012	0.000	0.068
	$p_Y(y)$	0.004	0.088	0.368	0.416	0.116	0.008	1.000

c) $\{2X = Y\}$ and $P(2X = Y) = p_{X,Y}(2,4) + p_{X,Y}(3,6) + p_{X,Y}(4,8) = 0.172$

d) $\{Y \geq X+5\}$ and $P(X+5 \leq Y) = p_{X,Y}(0,5) + p_{X,Y}(0,6) + p_{X,Y}(0,7) + p_{X,Y}(0,8) + p_{X,Y}(1,6) + p_{X,Y}(1,7) + p_{X,Y}(1,8) + p_{X,Y}(2,7) + p_{X,Y}(2,8) + p_{X,Y}(3,8) = 0.156$

6.124 a) See cells in the table in (b) below, where $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$ and represents the probability of a family owning y radios given we know they own x televisions.
b)

		Radios, Y						
		3	4	5	6	7	8	Total
Televisions, X	0	0.000	0.250	0.250	0.250	0.250	0.000	1.000
	1	0.000	0.146	0.366	0.439	0.049	0.000	1.000
	2	0.010	0.080	0.320	0.440	0.140	0.010	1.000
	3	0.000	0.068	0.420	0.398	0.102	0.011	1.000
	4	0.000	0.059	0.412	0.353	0.176	0.000	1.000
	$p_Y(y)$	0.004	0.088	0.368	0.416	0.116	0.008	1.000

c) $P(Y \leq 5|X=2) = p_{Y|X}(3|2) + p_{Y|X}(4|2) + p_{Y|X}(5|2) = 0.41$
d) See cells in the table in (e) below, where $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$ and represents the probability of a family owning x televisions given we know they own y radios.
e)

		Radios, Y						
		3	4	5	6	7	8	$p_X(x)$
Televisions, X	0	0.000	0.045	0.011	0.010	0.034	0.000	0.016
	1	0.000	0.273	0.163	0.173	0.069	0.000	0.164
	2	1.000	0.364	0.348	0.423	0.483	0.500	0.400
	3	0.000	0.273	0.402	0.337	0.310	0.500	0.352
	4	0.000	0.045	0.076	0.058	0.103	0.000	0.068
	Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000

f) $P(X \geq 3|Y=5) = p_{X|Y}(3|5) + p_{X|Y}(4|5) = 0.478$, i.e. approximately 47.8% of the town's families with five radios have at least three televisions.

6.125 a) Clearly, $p_{Y|X}(0\,|\,0) = 1$ and for $x = 1, \cdots, n$, $(Y\,|\,X=x) \sim \mathcal{B}(x, \alpha)$.
b) $p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x) = \binom{x}{y}\alpha^y(1-\alpha)^{x-y}\binom{n}{x}p^x(1-p)^{n-x}$ for all $y = 0, \ldots, x$ and $x = 0, \ldots, n$ and $p_{X,Y}(x,y) = 0$ otherwise.
c) The marginal PMF of Y is given as:

$$p_Y(y) = \sum_x p_{X,Y}(x,y) = \sum_{x=y}^{n} \binom{x}{y}\binom{n}{x}\alpha^y p^x (1-\alpha)^{x-y}(1-p)^{n-x}$$

$$= \sum_{x=y}^{n} \frac{x!}{y!(x-y)!}\frac{n!}{x!(n-x)!}\alpha^y p^x (1-\alpha)^{x-y}(1-p)^{n-x}$$

$$= \frac{n!}{y!}\alpha^y p^y \sum_{x=0}^{n-y} \frac{(n-y)!}{(n-y)!} \cdot \frac{1}{x!(n-y-x)!} p^x (1-\alpha)^x (1-p)^{n-y-x}$$

$$= \binom{n}{y}(\alpha p)^y \underbrace{\sum_{x=0}^{n-y}\binom{n-y}{x}(p(1-\alpha))^x(1-p)^{n-y-x}}_{=(p(1-\alpha)+(1-p))^{n-y}}$$

$$= \binom{n}{y}(\alpha p)^y(1-\alpha p)^{n-y}, \quad y = 0, \cdots, n.$$

Thus $Y \sim \mathcal{B}(n, \alpha p)$.
d) The required conditional PMF is given by: for $x = y, \cdots, n$,

$$p_{X\,|\,Y}(x\,|\,y) = \frac{\binom{x}{y}\binom{n}{x}\alpha^y(1-\alpha)^{x-y}p^x(1-p)^{n-x}}{\binom{n}{y}(\alpha p)^y(1-\alpha p)^{n-y}} = \binom{n-y}{x-y} r^{x-y}(1-r)^{n-x},$$

where $r = \frac{(1-\alpha)p}{1-\alpha p}$.
e) Using (d), we obtain that

$$p_{X-Y\,|\,Y}(z\,|\,y) = P(X-Y=z\,|\,Y=y) = P(X=z+y\,|\,Y=y)$$

$$= p_{X\,|\,Y}(z+y\,|\,y) = \binom{n-y}{z} r^z(1-r)^{n-y-z},$$

where $r = (1-\alpha)p/(1-\alpha p)$. Thus $(X-Y\,|\,Y=y) \sim \mathcal{B}(n-y, r)$.

6.126 Let X, Y, Z be the number of holes that Jan wins, that Jean wins and that are tied, respectively. Then X, Y, Z have a multinomial distribution with parameters $n = 18$ and p, q, r. Thus for nonnegative integers x, y, z, where $x + y + z = 18$,

$$p_{X,Y,Z}(x,y,z) = \binom{18}{x, y, z} p^x q^y r^z$$

and $p_{X,Y,Z}(x,y,z) = 0$ otherwise.

6.127 Because the distribution is symmetric,

$$P(|X-Y| \geq 3) = P(X \geq 3+Y) + P(Y \geq 3+X) = 2P(X \geq 3+Y)$$

$$= 2\sum_{y=1}^{\infty}\sum_{x=3+y}^{\infty} p^2(1-p)^{x+y-2} = \frac{2(1-p)^3}{2-p}.$$

6.128 a) For $z = 0$, $P(|X-Y| = 0) = \sum_{i=1}^{\infty} P(X = Y = i) = \sum_{i=1}^{\infty} p^2(1-p)^{2i-2} = p^2\sum_{i=0}^{\infty}((1-p)^2)^i = \frac{p^2}{2p-p^2} = \frac{p}{2-p}$. Again by symmetry, for $z = 1, 2\ldots$, $P(|X-Y| = z) = P(X = Y+z) + P(Y = X+z) = 2P(X = Y+z)$. Now

$$P(X = Y+z) = \sum_{y=1}^{\infty} P(X = y+z, Y = y) = \sum_{y=1}^{\infty} p^2(1-p)^{2y+z-2} = \frac{p(1-p)^z}{2-p}.$$

Therefore, for $z \in \mathcal{N}$, $P(|X-Y| = z) = \frac{2p(1-p)^z}{2-p}$.
b) $P(|X-Y| \geq 3) = 1 - \{P(|X-Y| = 0) + P(|X-Y| = 1) + P(|X-Y| = 2)\} = \dfrac{2(1-p)^3}{2-p}$.

6.129 a) Since the total area of the target is 36 square feet, and the area of the bulls-eye is π square feet, the probability of our archer hitting the bulls-eye is $p = \frac{\pi}{36}$. The area of the 5 point ring is 3π, thus the probability of our archer hitting the 5 point ring is $q = \frac{\pi}{12}$. The probability of hitting the remaining region is $r = 1 - p - q$. Thus the joint PMF of the random variables X, Y, and Z is multinomial with parameters 4, p, q, and r.
b) Since each shot is independent, $X \sim \mathcal{B}(4, p)$, $Y \sim \mathcal{B}(4, q)$ and $Z \sim \mathcal{B}(4, 1 - p - q)$.
c) The probability that the archer hits either the bulls-eye or the 5 point ring is $p + q$. Thus, $X + Y \sim \mathcal{B}(4, p + q)$. $X + Y$ represents the number of shots fired that score some positive amount of points.
d) $T = 10X + 5Y + 0 \cdot Z = 10X + 5Y$. The PMF of T is given by:

t	$p_T(t)$	t	$p_T(t)$
0	0.1795346	25	0.0218368
5	0.2888283	30	0.0048621
10	0.2705220	35	0.0006959
15	0.1628840	40	0.0000580
20	0.0707783		

6.130 a) Using the multivariate form of FPF and the binomial formaula,

$$p_{X_1,X_2,X_3+X_4}(x_1, x_2, z) = \sum_{y=0}^{z} P(X_1 = x_1, X_2 = x_2, X_3 = y, X_4 = z - y)$$

$$= \sum_{y=0}^{z} \binom{n}{x_1, x_2, y, z-y} p_1^{x_1} p_2^{x_2} p_3^{y} p_4^{z-y} = \frac{n!}{x_1!x_2!z!} p_1^{x_1} p_2^{x_2} \underbrace{\sum_{y=0}^{z} \frac{z!}{y!(z-y)!} p_3^{y} p_4^{z-y}}_{=(p_3+p_4)^z}$$

$$= \binom{n}{x_1, x_2, z} p_1^{x_1} p_2^{x_2} (p_3 + p_4)^z.$$

Thus $X_1, X_2, X_3 + X_4$ have a multinomial distribution with parameters n, p_1, p_2, and $p_3 + p_4$
b) Logically, we can group X_3 and X_4 and consider the sum as a new variable, which represents the number of selected members of the population that have 3rd or 4th attribute. Then $p_3 + p_4$ represents the proportion of the population that has either 3rd or 4th attributes. Hence we have 3 variables $X_1, X_2, X_3 + X_4$ and $X_1, X_2, X_3 + X_4$ has a multinomial distribution with parameters n and $p_1, p_2, p_3 + p_4$.

6.131 a) For notational purposes, let $U = U_{0.3}$ and $V = U_{0.6}$.

		V: 0	1	2	$p_U(u)$
U	0	0.16	0.24	0.09	0.49
	1	0.00	0.24	0.18	0.42
	2	0.00	0.00	0.09	0.09
	$p_V(v)$	0.16	0.48	0.36	1.00

For the PMF of V, refer to the last row of the above table.
b) For the PMF of U, refer to the last column of the above table.

c) For the joint PMF of U, V refer to the cells of the above table.
d) The conditional PMFs of U for each possible value of V are provided in the cell columns of the following table.

		V 0	1	2	$p_U(u)$
U	0	1.00	0.50	0.25	0.49
	1	0.00	0.50	0.50	0.42
	2	0.00	0.00	0.25	0.09
	Total	1.00	1.00	1.00	1.00

e) The conditional PMFs of V for each possible value of U are provided in the cell columns of the following table.

		V 0	1	2	Total
U	0	0.327	0.490	0.184	1.000
	1	0.000	0.571	0.429	1.000
	2	0.000	0.000	1.000	1.000
	$p_V(v)$	0.160	0.480	0.360	1.000

6.132 a) Let X, Y and Z represent the number of Republicans, Democrats and Independents selected to serve on the committee, respectively. Then X, Y, Z has a multiple hypergeometric distribution with parameters $N = 100$, $n = 10$ $p_X = 0.5$, $p_Y = 0.3$ and $p_Z = 0.2$. The probability that the committee will be perfectly representitive is

$$P(X = 5, Y = 3, Z = 2) = \frac{\binom{50}{5}\binom{30}{3}\binom{20}{2}}{\binom{100}{10}} \approx 0.0944$$

b) The probability that the committee will be close to being perfectly representitive is given by $P(4 \leq X \leq 6, 2 \leq Y \leq 4, 1 \leq Z \leq 3) = p_{X,Y,Z}(4,3,3) + p_{X,Y,Z}(4,4,2) + p_{X,Y,Z}(5,2,3) + p_{X,Y,Z}(5,3,2) + p_{X,Y,Z}(5,4,1) + p_{X,Y,Z}(6,2,2) + p_{X,Y,Z}(6,3,1) = 0.503$

6.133 a) Let $X_1, \ldots, X_7$ represent the number of red, orange, yellow, green, blue, violet, and brown N&Ns selected from a packet respectively. Then $X_1, \ldots, X_7$ has a multiple hypergeometric distribution with parameters $N = 70$, $n = 7$, $p_1 = \frac{1}{7}$, $p_2 = \frac{1}{10}$ $p_3 = \frac{6}{35}$ $p_4 = \frac{3}{14}$ $p_5 = \frac{1}{7}$ $p_6 = \frac{4}{35}$ and $p_7 = \frac{4}{35}$. The probability that all seven colors are chosen is: $P(X_1 = X_2 = X_3 = X_4 = X_5 = X_6 = X_7 = 1) = 0.00673$.
b) The probability that only one color is chosen is: $P(X_1 = 7) + P(X_2 = 7) + P(X_3 = 7) + P(X_4 = 7) + P(X_5 = 7) + P(X_6 = 7) + P(X_7 = 7) = 0.00000624$.
c) $P(X_1 = 2, X_3 = 2, X_4 = 2, X_5 = 1) = 0.00260$

6.134 From example 6.25 we know that the distribution of the number of serious accidents during a weekday is Poisson with parameter $2.4 \times 0.6 = 1.44$ and the distribution of the number of serious accidents during a weekend day is Poisson with parameter $1.3 \times 0.4 = 0.52$. Thus, using Proposition 6.20(b), the distribution of serious accidents during a week is Poisson with parameter

$$\underbrace{1.44 + 1.44 + 1.44 + 1.44 + 1.44}_{\text{5 weekdays}} + \underbrace{0.52 + 0.52}_{\text{saturday and sunday}} = 8.24$$

6.135 a) The marginal PMF of X is

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_{y=0}^{\infty} \frac{x^y e^{-x}}{y!M} = \frac{e^{-x}}{M} \underbrace{\sum_{y=0}^{\infty} \frac{x^y}{y!}}_{=e^x} = \frac{1}{M}$$

for $x = 1, \ldots, M$ and $p_X(x) = 0$ otherwise. Thus, X is uniformly distributed on $\{1, \cdots, M\}$.
b) The marginal PMF of Y is

$$p_Y(y) = \sum_x p_{X,Y}(x,y) = \sum_{x=1}^{M} \frac{x^y e^{-x}}{y!M} = \frac{1}{y!M} \sum_{x=1}^{M} x^y e^{-x}$$

for $y = 0, \ldots$ and $p_Y(y) = 0$ otherwise.
c) $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{e^{-x}x^y}{y!}$. Thus $(Y \mid X = x) \sim \mathcal{P}(x)$.
d) X and Y are not independent random variables since the conditional distribution of Y given $X = x$ is not the same as the marginal distribution of Y.

6.136 Let X_1, X_2, X_3 and X_4 denote, respectively, the numbers of the n items sampled that are in the four specified categories (nondefective-uninspected, nondefective-inspected, defective-uninspected, and defective-inspected). Then the random variables X_1, X_2, X_3, X_4 have the multinomial distribution with parameters n and $(1-p)(1-q), (1-p)q, p(1-q), pq$.

6.137 a) If the sample is taken without replacement, then X_1, X_2, X_3, X_4 have the multiple hypergeometric distribution with parameters N, n and $(1-p)(1-q), (1-p)q, p(1-q), pq$.
b) It is appropriate to use the multinomial distribution to approximate the multiple hypergeometric distribution when the population (lot) size N is large relative to the sample size n.

6.138 The conditional distribution of the number of undetected defectives, given that the number of detected defectives equals k, is binomial with parameters $n-k$ and $\frac{p(1-q)}{1-pq}$, since

$$P(X_3 = x | X_4 = k) = \frac{\binom{n}{x,k,n-x-k}(p(1-q))^x (pq)^k (1-p)^{n-x-k}}{\binom{n}{k}(pq)^k(1-pq)^{n-k}}$$

$$= \frac{(n-k)!}{x!(n-x-k)!} \left(\frac{p(1-q)}{1-pq}\right)^x \left(\frac{1-p}{1-pq}\right)^{n-k-x}, \quad x = 0, \cdots, n-k.$$

6.139 a) The conditional PMF of the number of people, N, entering the bank during the specified time interval, given the number of people, X, making a deposit is

$$P(N = n \mid X = x) = \frac{P(X = x, N = n)}{P(X = x)} = \frac{\frac{e^{-\lambda}\lambda^n}{n!}\binom{n}{x}p^x(1-p)^{n-x}}{\frac{e^{-p\lambda}(p\lambda)^x}{x!}}$$

$$= \frac{e^{-\lambda}\lambda^n n! p^x (1-p)^{n-x} x!}{n! x! (n-x)! e^{-p\lambda}(p\lambda)^x} = \frac{e^{-(1-p)\lambda}((1-p)\lambda)^{n-x}}{(n-x)!}, \quad n = x, x+1, \cdots.$$

b) From above, we see $(N - X \mid X = x) \sim \mathcal{P}((1-p)\lambda)$.
c) X and $N - X$ are independent since the answer in (b) does not depend on x.
d) The number of people who make a deposit, X, and the number of people who don't make a deposit, $N - X$, are independent.

6.140 The distribution of the total number of people at the car wash is Poisson with parameter $\lambda + \mu + \nu$ (Proposition 6.20). Thus the conditional joint distribution of the hourly numbers of customers requesting the three options, given the total number n of customers during the hour, is multinomial with parameters n and $\frac{\lambda}{\lambda+\mu+\nu}, \frac{\mu}{\lambda+\mu+\nu}, \frac{\nu}{\lambda+\mu+\nu}$

6.141 Let X be the number of successes in n trials, and let Y be the number of the trial on which the first success occurs. Note that the probability of $k-1$ successes in $n-y$ trials is $\binom{n-y}{k-1}p^{k-1}(1-p)^{n-y-k+1} = P(X = k|Y = y)$. Thus, for $k = 0, \cdots, n$,

$$P(Y = y|X = k) = \frac{P(X = k, Y = y)}{P(X = k)} = \frac{P(X = k|Y = y)P(Y = y)}{P(X = k)}$$

$$= \frac{\left(\binom{n-y}{k-1}p^{k-1}(1-p)^{n-y-k+1}\right)\left(p(1-p)^{y-1}\right)}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{\binom{n-y}{k-1}}{\binom{n}{k}}, \quad y = 1, \cdots, n-k+1.$$

6.142 a) For $y = 0, 1, \cdots, 9$, we have $p_{X|Y}(x|y) = \dfrac{p_{Y\,|\,X}(y\,|\,x)p_X(x)}{\sum_{k=0}^{y} p_{Y\,|\,X}(y\,|\,k)p_X(k)} = \dfrac{\frac{1}{10-x}}{\sum_{k=0}^{y}\frac{1}{10-k}}$,
where $x = 0, \cdots, y$, since $p_X(x) = 1/10$ for all $x = 0, \cdots, 9$ and $p_{Y\,|\,X}(y\,|\,x) = 1/(10-x)$ for $y = x, x+1, \cdots, 9$ and $x \in \{0, \cdots, 9\}$.
b) The PMF of $(X \mid Y = 6)$ is given by:

x	$p_{X\|Y}(x\|6)$	x	$p_{X\|Y}(x\|6)$
0	0.0913	4	0.1521
1	0.1014	5	0.1825
2	0.1141	6	0.2282
3	0.1304		

6.143 a) If $T = n$, then X_1 through X_{n-1} must not exceed m. Additionally, we must also have that $X_n > m$. Thus, $X_n \in \{m+1, \ldots, N\}$. Thus $\{T = n\}$ is the union of the events $\{X_1 \le m, \ldots, X_{n-1} \le m, X_n = k\}$ for $k = m+1, \ldots, N$.
b) The probability of the event $\{X_1 \le m, \ldots, X_{n-1} \le m, X_n = k\}$ is $\left(\frac{m}{N}\right)^{n-1} \cdot \frac{1}{N}$. Thus

$$P(T = n) = \sum_{k=m+1}^{N} P(X_1 \le m, \ldots, X_{n-1} \le m, X_n = k) = \sum_{k=m+1}^{N} \left(\frac{m}{N}\right)^{n-1} \cdot \frac{1}{N}$$

$$= \left(\frac{m}{N}\right)^{n-1}\left(\frac{N-m}{N}\right) = \left(1 - \frac{m}{N}\right)\left(\frac{m}{N}\right)^{n-1}$$

Therefore, $T \sim \mathcal{G}\left(1 - \frac{m}{N}\right)$
c) Consider the random experiment of selecting a number at random from the first N positive

integers. Let the specified event (i.e., a success) be that the number selected exceeds m. Independent repititions of this experiment constitute Bernoulli trials with success probability $1-\frac{m}{N}$. The random variable T is the time of the first success, which has the geometric distribution with parameter $1-\frac{m}{N}$.

6.144 Using the law of total probability,

$$P(U=u)=\sum_{k=1}^{N} P(U=u, X_1=k)=\sum_{k=1}^{N} P(X_1=k)P(U=u \mid X_1=k).$$

Thus, for $u=2$,

$$P(U=2)=\sum_{k=1}^{N} P(X_1=k)P(U=2 \mid X_1=k)=\sum_{k=1}^{N} \frac{1}{N}\cdot P(X_2 \geq k)$$

$$=\sum_{k=1}^{N} \frac{1}{N}\cdot\frac{N-k+1}{N}=\frac{1}{N^2}(N+(N-1)+\cdots+1)=\frac{1}{N^2}\cdot\frac{N(N+1)}{2}=\frac{N+1}{2N}.$$

For $u=3,\ldots$, we have that $P(U=u, X_1=1)=0$ and

$$P(U=u)=\sum_{k=1}^{N} P(U=u, X_1=k)=\sum_{k=2}^{N} P(U=u, X_1=k)=\sum_{k=2}^{N} P(X_1=k)P(U=u|X_1=k)$$

$$=\frac{1}{N}\sum_{k=2}^{N} P(X_2<k,\ldots,X_{u-1}<k, X_u \geq k)=\frac{1}{N}\sum_{k=2}^{N}\left(1-\frac{k-1}{N}\right)\left(\frac{k-1}{N}\right)^{u-2}$$

and $P(U=u)=0$ otherwise.

6.145 Folowing exercise 6.140, and letting $p_j=\frac{\lambda_j}{\lambda_1+\ldots+\lambda_m}$, we have that the condition joint PMF of $X_1,\ldots,X_m$ given that $X_1+\ldots+X_m=z$ is multinomial with parameters z and $p_1,\ldots,p_m$.

6.146 a) The marginal PMF of X_m is Binomial with parameters n and p_m.
b) The conditional PMF of $X_1,\ldots,X_{m-1}$ given $X_m=x_m$ is multinomial with parameters $n-x_m$ and $\frac{p_1}{1-p_m},\ldots,\frac{p_{m-1}}{1-p_m}$

6.147 a) Assuming the two factories are independent, by Proposition 6.20, the total number of defective components per day is Binomial with parameters $m+n$ and p.
b) Let X represent the number of defective components from Factory A and Z represent the total number of defective components from both factories combined. Then the conditional probability distribution of the number of defective components per day from Factory A, given the combined total of defective components is

$$p_{X|Z}(x|z)=P(X=x|Z=z)=\frac{P(X=x, Y=z-x)}{P(Z=z)}$$

$$=\frac{\binom{m}{x}p^x(1-p)^{m-x}\binom{n}{z-x}p^{z-x}(1-p)^{n-z+x}}{\binom{m+n}{z}p^z(1-p)^{m+n-z}}=\frac{\binom{m}{x}\binom{n}{z-x}}{\binom{m+n}{z}}$$

Thus $X|Z=z$ has a hypergeometric distribution with parameters $m+n$, z and $\frac{m}{m+n}$.

6.148 a) If l represents the daily output from Factory C, then by Proposition 6.20, the total number of defective components per day is Binomial with parameters $m+n+l$ and p.
b) $X|Z=z$ has a hypergeometric distribution with parameters $m+n+l$, z and $\frac{m}{m+n+l}$.

6.149 a) Using Exercise 6.31, the conditional joint PMF of $X_1,\ldots,X_m$ given $X_1+\ldots+X_m=z$ is multiple hypergeometric with parameters N,z and $\frac{n_1}{N},\ldots,\frac{n_m}{N}$, where $N=n_1+\cdots+n_m$. One can also obtain the required answer by noting that $X_1+\cdots+X_m$ has $\mathcal{B}(n_1+\cdots+n_m,\ p)$ distribution, thus, for all $x_i\in\{1,\cdots,n_i\}$, $i=1,\ldots,m$,

$$P(X_1=x_1,\cdots,X_m=x_m|X_1+\cdots+X_m=z)=\frac{P(X_1=x_1,\ldots,X_m=x_m,X_1+\cdots+X_m=z)}{P(X_1+\cdots+X_m=z)}$$

$$=\begin{cases}0, & \text{if } z\neq x_1+\cdots+x_m,\\ \dfrac{\binom{n_1}{x_1}p^{x_1}(1-p)^{n_1-x_1}\times\cdots\times\binom{n_m}{x_m}p^{x_m}(1-p)^{n_m-x_m}}{\binom{n_1+\cdots+n_m}{z}p^z(1-p)^{n_1+\cdots+n_m-z}}, & \text{if } z=x_1+\cdots+x_m;\end{cases}$$

$$=\begin{cases}0, & \text{if } z\neq x_1+\cdots+x_m,\\ \dfrac{\binom{n_1}{x_1}\times\ldots\times\binom{n_m}{x_m}}{\binom{n_1+\cdots+n_m}{z}}, & \text{if } z=x_1+\cdots+x_m.\end{cases}$$

b) Using the fact that the j^{th} univariate marginal PMF of a multiple hypergeometric distribution with parameters N,z and $\frac{n_1}{N},\ldots,\frac{n_m}{N}$ has the hypergeometric distribution with parameters N,z and $\frac{n_j}{N}$, the required results immediately follow.

6.150 a) By definition, X_j is Bernoulli with parameter p.
b) If $X=x$, then we need to have had x successes in n trials, thus X_j is Bernoulli with parameter $\frac{x}{n}$.
c) Using the definition of conditional PMF, we have

$$P(X_j=k|X=x)=\frac{P(X_j=k,X=x)}{P(X=x)}=\frac{P(X=x|X_j=k)P(X_j=k)}{P(X=x)}.$$

For $k=0$, we have

$$P(X_j=0|X=x)=\frac{\left(\binom{n-1}{x}p^x(1-p)^n-x-1\right)(1-p)}{\binom{n}{x}p^x(1-p)^{n-x}}=1-\frac{x}{n}.$$

For $k=1$, we have

$$P(X_j=1|X=x)=\frac{\left(\binom{n-1}{x-1}p^{x-1}(1-p)^n-x\right)p}{\binom{n}{x}p^x(1-p)^{n-x}}=\frac{x}{n}.$$

Therefore, the desired result is proved.

Theory Exercises

6.151 a) Note that for $x_1,\cdots,x_m\in\mathcal{N}$,

$$P(X_1=x_1,\ldots,X_m=x_m)=P(\underbrace{F\ldots F}_{x_1-1}S\underbrace{F\ldots F}_{x_2-1}SF\ldots S\underbrace{F\ldots F}_{x_m-1}S)$$

$$= \left((1-p)^{x_1-1}p\right)\left((1-p)^{x_2-1}p\right)\ldots\left((1-p)^{x_m-1}p\right),$$

and $p_{X_1,\cdots,X_m}(x_1,\cdots,x_m) = 0$ otherwise. Thus, by Exercise 6.60(c), $X_1,\ldots,X_m$ are independent random variables, where $p_{X_i}(x_i) = (1-p)^{x_i-1}p$ for $x_i \in \mathcal{N}$ and $p_{X_i}(x_i) = 0$ for $x_i \notin \mathcal{N}$, $i = 1,\ldots,m$. Therefore, $X_1,\ldots X_m$ are independent identically distributed geometric random variables with parameter p.
b) Clearly, $X = X_1 + \cdots + X_r$, where, by (a), $X_1,\ldots,X_r$ are independent geometric random variables with parameter p.
c) Using (a) and (b), if $X \sim \mathcal{NB}(r,p)$ and $Y \sim \mathcal{NB}(s,p)$ and are independent, then $X = X_1 + \cdots + X_r$ and $Y = Y_1 + \cdots + Y_s$ for some independent geometric random variables $X_1,\ldots,X_r,Y_1,\ldots,Y_s$ with parameter p. Then their sum $X+Y$ can be written as the sum of $r+s$ independent geometric random variables with success parameter p and represents the number of Bernoulli trials up to and including the $(r+s)$th success. Thus, $X+Y \sim \mathcal{NB}(r+s,p)$.

Advanced Exercises

6.152 a) Since S_N is the sum of independent identically distributed random variables, S_N has a compound distribution.
b) Yes. Since N has a Poisson distribution, S_N is has a compound Poisson distribution.
c) S_N is the total number of trials needed to achieve the N^{th} success.
d) Using Exercise 6.151 (a), we have that

$$p_{S_N}(x) = \sum_{n=0}^{\infty} p_N(n)p_{S_n}(x) = \sum_{n=1}^{x} \frac{e^{-\lambda}\lambda^n}{n!}\binom{x-1}{n-1}p^n(1-p)^{x-n},\ x \in \mathcal{N},$$

and $p_{S_N}(x) = 0$ for all $x \notin \mathcal{N}$.
e) $P(S_N = 3) = e^{-\lambda}\left[\lambda p(1-p)^2 + \lambda^2p^2(1-p) + \frac{\lambda^3}{6}p^3\right]$.

6.153 a) Considering the joint PMF of the random variables $X_1,\ldots X_k$ for $x_1,\ldots,x_k$, let $y = n - (x_1 + \ldots + x_k)$ and $q = 1 - (p_1 + \ldots + p_k)$. Then,

$$P(X_1 = x_1,\ldots,X_k = x_k) = \binom{n}{x_1,\ldots,x_k,y}p_1^{x_1}\cdots p_k^{x_k}q^y$$

for $x_1,\ldots,x_k \in \mathcal{Z}_+$ such that $x_1+\cdots+x_k \in \{0,\ldots,n\}$, and $p_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = 0$ otherwise.
b) The conditional distribution of $X_1,\ldots,X_{m-2}$ given $X_{m-1} = x_{m-1}$ and $X_m = x_m$ is multinomial with parameters $n - (x_{m-1} + x_m)$ and $\frac{p_1}{1-(p_{m-1}+p_m)},\ldots,\frac{p_{m-2}}{1-(p_{m-1}+p_m)}$.

6.154 a) $T \sim \mathcal{G}(1 - p_X(0))$, since T counts the number of Bernoulli trials until the 1st success, where "success"$=\{X \neq 0\}$. In particular, this implies that T is finite with probability one.
b) X_T represents the first strictly positive value obtained in the sequence $X_1, X_2,\ldots$
c) For $x \in \mathcal{N}$, $t \in \mathcal{N}$, we have

$$p_{X_T,T}(x,t) = P(X_1 = 0,\ldots,X_{t-1} = 0, X_t = x) = (p_X(0))^{t-1}p_X(x),$$

and $p_{X_T,T}(x,t) = 0$ otherwise.
d) The PMF of X_T is:

$$p_{X_T}(x) = \sum_{t=1}^{\infty} P(X_T = x, T = t) = \sum_{t=1}^{\infty}(p_X(0))^{t-1}p_X(x) = \frac{p_X(x)}{1-p_X(0)}$$

for $x = 1, 2, \ldots$ and $p_{X_T}(x) = 0$ otherwise.
e) X_T and T are independent since $p_{X_T,T}(x,t) = p_{X_t}(x)p_T(t)$ for all x and t.

6.155 Since X, Y are both integer-valued, using partitioning of events, we have that

$$\begin{aligned} P(X \le x, Y \le y) &= P(X = x, Y = y) + P(X = x, Y \le y-1) \\ &+ P(X \le x-1, Y = y) + P(X \le x-1, Y \le y-1). \end{aligned}$$

On the other hand,

$$P(X \le x, Y \le y-1) = P(X = x, Y \le y-1) + P(X \le x-1, Y \le y-1),$$

$$P(X \le x-1, Y \le y) = P(X \le x-1, Y = y) + P(X \le x-1, Y \le y-1).$$

Therefore,

$$P(X \le x, Y \le y) - P(X \le x, Y \le y-1) - P(X \le x-1, Y \le y) + P(X \le x-1, Y \le y-1) = p_{X,Y}(x,y).$$

6.156 $P(U = u, V = v) = P(\min\{X_1, X_2\} = u, \max\{X_3, X_4\} = v)$
$= P(X_1 = u, X_2 = u, X_3 = v, X_4 = v) + P(X_1 = u, X_2 = u, X_3 = v, X_4 < v)$
$+ P(X_1 = u, X_2 = u, X_3 < v, X_4 = v) + P(X_1 = u, X_2 > u, X_3 = v, X_4 = v)$
$+ P(X_1 = u, X_2 > u, X_3 = v, X_4 < v) + P(X_1 = u, X_2 > u, X_3 < v, X_4 = v)$
$+ P(X_1 > u, X_2 = u, X_3 = v, X_4 = v) + P(X_1 > u, X_2 = u, X_3 = v, X_4 < v)$
$+ P(X_1 > u, X_2 = u, X_3 < v, X_4 = v)$
$= \dfrac{1}{10^4}\left(1 + v + v + (9-u) + v(9-u) + v(9-u) + (9-u) + v(9-u) + v(9-u)\right)$
$= \dfrac{19 + 38v - 4vu - 2u}{10^4} = \dfrac{1+2v}{100} \cdot \dfrac{19-2u}{100} = p_V(v)p_U(u).$

6.157 Note that, when $u, v \in \{0, 1, \ldots, N\}$ and $u < v$, then

$$P(U = u, V = v) = P(X = u, Y = v) + P(X = v, Y = u) = \frac{2}{(N+1)^2};$$

if $u = v \in \{0, 1, \ldots, N\}$, then

$$P(U = u, V = u) = \frac{1}{(N+1)^2},$$

and $P(U = u, V = v) = 0$ for all other values of (u, v). Thus, the joint PMF of U and V is:

$$p_{U,V}(u,v) = \begin{cases} \dfrac{1}{(N+1)^2}, & \text{if } u = v \in \{0, 1, \ldots, N\}, \\ \dfrac{2}{(N+1)^2}, & \text{if } u < v \text{ and } u, v \in \{0, 1, \ldots, N\}, \\ 0, & \text{otherwise.} \end{cases}$$

6.158 a) For $x_1, \ldots, x_{m-1} \in \{0, \cdots, n\}$, where $0 < x_1 + \ldots + x_{m-1} \le n$, let $x = x_1 + \ldots + x_{m-1}$

and $p_n = p_{1n} + \ldots + p_{(m-1)n}$. Also, note that $np_n \to \lambda_1 + \ldots + \lambda_{m-1}$ as $n \to \infty$. Let $\lambda = \lambda_1 + \ldots + \lambda_{m-1}$. Then

$$p_{X_{1n},\ldots,X_{(m-1),n}}(x_1,\ldots,x_{m-1}) = \binom{n}{x_1,\ldots,x_{m-1},n-x} p_{1n}^{x_1}\cdots p_{(m-1)n}^{x_{m-1}}(1-p_n)^{n-x}$$

$$= \frac{1}{x_1!\cdots x_{m-1}!}\cdot\frac{n(n-1)\cdots(n-x+1)}{n^x}(np_{1n})^{x_1}\cdots(np_{(m-1)n})^{x_{m-1}}\Big[1-\frac{np_n}{n}\Big]^{n-x},$$

where

$$\frac{n(n-1)\cdots(n-x+1)}{n^x} = \frac{n}{n}\times\frac{n-1}{n}\times\cdots\times\frac{n-x+1}{n}$$

$$= 1\times(1-\frac{1}{n})\times(1-\frac{2}{n})\times\cdots\times(1-\frac{x-1}{n}) \longrightarrow 1, \text{ as } n\to\infty.$$

Also, by assumption, for $1 \le i \le m-1$, $(np_{in})^{x_i} \to \lambda_i^{x_i}$ as $n \to \infty$. Moreover,

$$\Big[1-\frac{np_n}{n}\Big]^{n-x} = \Big[1-\frac{np_n}{n}\Big]^{-x}\Big[1-\frac{np_n}{n}\Big]^{n} \longrightarrow 1\cdot e^{-\lambda}, \text{ as } n\to\infty.$$

Thus, for all $x_1,\ldots,x_{m-1} \in \{0,\cdots,n\}$, where $0 < x_1 + \ldots + x_{m-1} \le n$,

$$p_{X_{1n},\ldots,X_{(m-1)n}}(x_1,\ldots,x_{m-1}) \longrightarrow \frac{1}{x_1!\cdots x_{m-1}!}\lambda_1^{x_1}\cdots\lambda_{m-1}^{x_{m-1}}e^{-\lambda}, \text{ as } n\to\infty.$$

Also $p_{X_{1n},\ldots,X_{(m-1)n}}(0,\ldots,0) = (1-p_n)^n = (1-\frac{np_n}{n})^n \longrightarrow e^{-\lambda}$, thus, the required conclusion follows for all $x_1,\ldots,x_{m-1} \in \{0,\cdots,n\}$ such that $0 \le x_1 + \ldots + x_{m-1} \le n$.

b) For a fixed m, as n grows large with respect to m, the multinomial distribution with parameters n and $p_{1n},\ldots,p_{mn}$ is well approximated by a multivariate Poisson distribution with parameters $np_{1n},\ldots,np_{(m-1)n}$.

c) When $m = 2$, the multinomial distribution in question is just a binomial distribution, whereas the multivariate Poisson distribution in (a) reduces to a univariate Poisson, thus, the required result follows at once from (a).

BETA

Chapter 7

Expected Value of Discrete Random Variables

7.1 From Averages to Expected Values

Basic Exercises

7.1 Using Exercise 5.24, we have

$$\mathcal{E}(X) = \sum_x xP(X = x) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36}$$

$$+8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7.$$

7.2 Let X be the score of the archer. Then using Exercise 5.27, we have

$$\mathcal{E}(X) = \sum_x xP(X = x) = 5 \times \frac{\pi}{12} + 10 \times \frac{\pi}{36} = \frac{25\pi}{36}.$$

7.3 a) Using Exercise 5.28, when $p = 0.5$ we have

$$\mathcal{E}(X) = \sum_x xP(X = x) = 1 \times 0.4 + 2 \times 0.3 = 1.$$

b) Again using Exercise 5.28, when $p = 0.4$ we have

$$\mathcal{E}(X) = \sum_x xP(X = x) = 1 \times 0.4 + 2 \times 0.24 = 0.88$$

7.4 a) Let X be the number of women in the interview pool. Then, using Exercise 5.29 (a), we have

$$\mathcal{E}(X) = \sum_x xP(X = x) = 1 \times \frac{5}{11} + 2 \times \frac{12}{33} + 3 \times \frac{2}{33} = \frac{15}{11}.$$

b) Note that X has a hypergeometric distribution with parameters $N = 11$, $n = 3$ and $p = \frac{5}{11}$. Then, using Table 7.4, we have that $\mathcal{E}(X) = np = 3 \cdot \frac{5}{11} = \frac{15}{11}$.

7.5 Since the total torque acting on the seesaw must be 0, we have the following equation:

$$\sum_{k=1}^{m}(x_k - \bar{x})p_k = 0.$$

Then $\sum_{k=1}^{m} x_k p_k - \sum_{k=1}^{m} \bar{x} p_k = 0$, implying that

$$\sum_{k=1}^{m} x_k P(X = x_k) - \bar{x} \underbrace{\sum_{k=1}^{m} P(X = x_k)}_{=1} = 0.$$

Therefore, $\mathcal{E}(X) = \bar{x}$.

7.6 a) Note first that Ω contains a finite number of elements, say, $\Omega = \{\omega_1, \cdots, \omega_N\}$ for some finite $N \in \mathcal{N}$. For each $x \in \mathcal{R}$, let K_x be the set of all $\omega \in \Omega$ such that $X(\omega) = x$; also let $\{x_1, \cdots, x_M\}$ (with $M \leq N$ and all distinct $x_1, \cdots, x_M$) denote the range of X. Then $\Omega = \bigcup_{i=1}^{M} K_{x_i}$ and

$$\sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) = \sum_{i=1}^{M}\Big(\sum_{\omega \in K_{x_i}} X(\omega)P(\{\omega\})\Big) = \sum_{i=1}^{M} x_i \Big(\sum_{\omega \in K_{x_i}} P(\{\omega\})\Big)$$

$$= \sum_{i=1}^{M} x_i P(K_{x_i}) = \sum_{i=1}^{M} x_i P(X = x_i) = \mathcal{E}(X).$$

b) If there are N equally likely outcomes, then $P(\{\omega_i\}) = \frac{1}{N}$ for each possible outcome $\omega_i \in \Omega$ and $\mathcal{E}(X) = \frac{1}{N}\sum_{i=1}^{N} X(\omega_i)$.

7.7 a) 82.4, since consider a finite population $\{\omega_1, \cdots, \omega_N\}$ of (arbitrary but fixed) size $N \in \mathcal{N}$. Let y_i be the value of the variable in question for the ith member of the population (i.e. for ω_i), $i = 1, \cdots, N$. If X represents the value of the variable for a *randomly* selected member of the population, then, by Exercise 7.6(b),

$$\mathcal{E}(X) = \frac{1}{N}\sum_{i=1}^{N} X(\omega_i) = \frac{1}{N}\sum_{i=1}^{N} y_i = \bar{y} = 82.4$$

b) $82.4n$. For large n, the average value of X in n independent observations is approximately equal to $\mathcal{E}(X)$, implying that the sum of values of X in n independent observations is approximately equal to $n\mathcal{E}(X)$.

7.8 a) 0.217242. Let X be the total number of people tested until the first positive test occurs. Then X has a geometric distribution (with some parameter p). Since, on average, the first positive test occurs when the 25^{th} person is tested, then $\mathcal{E}(X) = 25 = 1/p$, implying that $p = 1/25$. Thus,

$$P(X \leq 6) = \sum_{x=1}^{6}\left(\frac{1}{25}\right)\left(1 - \frac{1}{25}\right)^{x-1} = 0.217242$$

b) 125. Let Y be the total number of people tested until the fifth positive test occurs. Then,

Y has $\mathcal{NB}(5, 1/25)$ distribution. From Table 7.4, and since the average number of people that the physician must test until obtaining the fifth positive test is the expectation of Y, it follows that

$$\mathcal{E}(Y) = \frac{r}{p} = \frac{5}{\frac{1}{25}} = 125.$$

7.9 18. Since the number of children cured after treatment with the drug has a binomial distribution with parameters $n = 20$ and $p = 0.9$, the expected number of children cured equals to $np = 18$.

7.10 15. Distribution of the number of times that the favorite will finish in the money is binomial with parameters n and 0.67. Additionally, we want the expectation to be at least 10. If X is the number of times that the favorite will finish in the money, then we have $np = n0.67 = \mathcal{E}(X) \geq 10$, implying that $n \geq 14.9254$, and, since n must be an integer, it follows that n is at least 15.

7.11 1.25; If both parents have the sickle cell trait, then the probability that they will pass it on is 0.25, and the number of children out of 5 who will get it has a binomial distribution with parameters $n = 5$ and $p = 0.25$. Thus the expected number of children who will inherit the sickle cell trait is equal to $np = 1.25$.

7.12 a) Let X represent the number of undergraduate students selected. Then when there is no replacement, $X \sim \mathcal{H}(10, 3, 0.6)$ and the expected number of underclassmen is $\mathcal{E}(X) = np = 1.8$.
b) When there is replacement, $X \sim \mathcal{B}(3, 0.6)$ and the expected number of underclassmen is $\mathcal{E}(X) = np = 1.8$.
c) The results are the same since a binomial random variable with parameters n and p has the same expectation as a hypergeometric random variable with parameters N, n and p.

7.13 From example 5.14, we have that X is hypergeometric with parameters 100, 5 and 0.06. Thus using Table 7.4, $\mathcal{E}(X) = np = 0.3$. This means that on average, when 5 TVs are randomly selected without replacement, we expect to see 0.3 defective TVs.

7.14 a) When sampling without replacement, the distribution of people with the specific attribute is hypergeometric with parameters N, n and p, and the expectation is np.
b) When sampling with replacement, the distribution of people with the specific attribute is binomial with parameters n and p, and the expectation is np.

7.15 a) 15. Since X, the number of cars that use the drive-up window (at a given fast-food restaurant) between 5:00 P.M. and 6:00 P.M., is Poisson with some parameter λ, and, on average, 15 cars use it during that hour, then $\lambda = E(X) = 15$.
b) 63.472%, since

$$P(|X - 15| \leq 3) = \sum_{x=12}^{18} \frac{e^{-15}\, 15^x}{x!} \approx 0.63472$$

7.16 The means are the same.

7.17 Assuming that all the customers at Downtown Coffee Shop make their selections independently and have the same chance of buying a café mocha with no whipped cream, then the expected number of customers (per hour) who buy a café mocha with no whipped cream is $0.3 \times 31 = 9.3$

7.18 $4/(1-e^{-4})$. From Exercise 5.85, letting Y be the number of eggs observed in a non-empty nest, we have for $y \in \mathcal{N}$,

$$p_Y(y) = \frac{e^{-4} \cdot \frac{4^y}{y!}}{1-e^{-4}},$$

thus,

$$\mathcal{E}(Y) = \sum_y yP(Y=y) = \sum_{y=1}^{\infty} y \frac{e^{-4} \cdot \frac{4^y}{y!}}{1-e^{-4}} = \frac{4}{1-e^{-4}} \approx 4.07463$$

7.19 If E is an event, then $I_E \sim \mathcal{B}(1, P(E))$. Therefore, $\mathcal{E}(I_E) = 1 \cdot P(E) = P(E)$.

7.20 a) Let X be the number of at bats until the first hit. Then $X \sim \mathcal{G}(0.260)$, and $\mathcal{E}(X) = 1/0.260 \approx 3.84615$;
b) Let Y be the number of at bats until the second hit. Then $Y \sim \mathcal{NB}(2, 0.260)$, and $\mathcal{E}(Y) = 2/0.260 \approx 7.69231$;
c) Let Z be the number of at bats until the tenth hit. Then $Z \sim \mathcal{NB}(10, 0.260)$, and $\mathcal{E}(Z) = 10/0.260 \approx 38.4615$;

7.21 a) The answers will vary.
b) By definition of expectation,

$$\mathcal{E}(X) = \sum_x xP(X=x) = \sum_{x=0}^{9} \frac{x}{10} = 4.5;$$

The sample average, for large sample size n, should be close to $\mathcal{E}(X) = 4.5$

7.22 $(N+1)/2$. By definition of expectation,

$$\mathcal{E}(X) = \sum_x xP(X=x) = \sum_{x=1}^{N} \frac{x}{N} = \frac{1}{N}\sum_{x=1}^{N} x = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}.$$

Theory Exercises

7.23 Note that $k\binom{Np}{k} = Np\binom{Np-1}{k-1}$ for $k \geq 1$, and $\binom{N}{n} = \frac{N}{n} \times \binom{N-1}{n-1}$ for $n \geq 1$, thus,

$$\mathcal{E}(X) = \sum_{k=0}^{N} k \cdot \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}} = \sum_{k=1}^{N} \frac{Np\binom{Np-1}{k-1}\binom{N-Np}{(n-1)-(k-1)}}{\frac{N}{n} \times \binom{N-1}{n-1}}$$

$$= np\sum_{\ell=0}^{N-1} \frac{\binom{Np-1}{\ell}\binom{N-Np}{(n-1)-\ell}}{\binom{N-1}{n-1}} = np\sum_{\ell=0}^{N-1} \frac{\binom{(N-1)\widetilde{p}}{\ell}\binom{(N-1)(1-\widetilde{p})}{(n-1)-\ell}}{\binom{N-1}{n-1}} = np,$$

since the last sum is the sum of probabilities of all possible values of a hypergeometric random variable with parameters $N-1$, $n-1$ and $\widetilde{p}$, where we put $\widetilde{p} = (Np-1)/(N-1)$.

7.24 The required conclusion follows from:

$$\mathcal{E}(X) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = -p\sum_{x=0}^{\infty} \frac{d}{dp}(1-p)^x = -p\frac{d}{dp}\sum_{x=0}^{\infty}(1-p)^x = -p\frac{d}{dp}\left(\frac{1}{p}\right) = \frac{1}{p}.$$

7.25 a) Since a negative binomial random variable with parameters r and p can be represented as a sum of r geometric random variables, each of which has an expectation of $1/p$, then the expected value of a negative binomial random variable is r/p.
b) Note that

$$\mathcal{E}(X) = \sum_{x=r}^{\infty}(r-(x-r))p_X(x)$$

$$= r\sum_{x=r}^{\infty}\binom{-r}{x-r}p^r(p-1)^{x-r} + \sum_{x=r}^{\infty}\binom{-r}{x-r}(x-r)p^r(p-1)^{x-r}$$

$$= r + p^r(p-1)\sum_{j=0}^{\infty}\binom{-r}{j}\frac{d}{dp}(p-1)^j.$$

c) By (b) and the binomial series theorem (see equation (5.45) on page 241 in the text),

$$r + p^r(p-1)\sum_{j=0}^{\infty}\binom{-r}{j}\frac{d}{dp}(p-1)^j = r + p^r(p-1)\frac{d}{dp}\sum_{j=0}^{\infty}\binom{-r}{j}(p-1)^j$$

$$= r + p^r(p-1)\frac{d}{dp}(1+(p-1))^{-r} = r + p^r(p-1)\frac{d}{dp}p^{-r} = r + p^r(p-1)\frac{(-r)}{p^{r+1}} = \frac{r}{p}.$$

d) Since a negative binomial random variable, with parameters $r = 1$ and p, is a geometric random variable with parameter p, the desired result immediately follows from (c).

Advanced Exercises

7.26 α. Indeed, note that

$$\mathcal{E}(X) = \sum_{x=0}^{\infty} x\cdot\frac{\alpha^x}{(1+\alpha)^{x+1}} = \sum_{x=1}^{\infty} x\cdot\left(1-\frac{1}{1+\alpha}\right)^x\left(\frac{1}{1+\alpha}\right) = \left(1-\frac{1}{1+\alpha}\right)\mathcal{E}(Y),$$

where Y is $\mathcal{G}(\frac{1}{1+\alpha})$. Therefore,

$$\mathcal{E}(X) = \left(1-\frac{1}{1+\alpha}\right)(1+\alpha) = \alpha.$$

7.27 a) Let $A^+ = \{x_1, x_2, \ldots\}$ be the enumeration of all positive x such that $p_X(x) > 0$. For each $N \in \mathcal{N}$, consider $S_N^+ = \sum_{i=1}^{N} x_i p_X(x_i)$. Since $x_i > 0$ and $p_X(x_i) > 0$ for all i, then S_N^+ is a monotone increasing sequence and must have a limit as $N \to \infty$, where the limit is either a finite positive constant or $+\infty$. Since $\mathcal{E}(X^+) = \lim_{N\to\infty} S_N^+$, the required conclusion then follows.
b) Let $A^- = \{y_1, y_2, \ldots\}$ be the enumeration of all negative x such that $p_X(x) > 0$. For each $N \in \mathcal{N}$, define $S_N^- = \sum_{i=1}^{N}(-y_i)p_X(y_i)$. Since $-y_i > 0$ and $p_X(y_i) > 0$ for all i, then S_N^- is a monotone increasing sequence and must have a limit as $N \to \infty$, where the limit is either a finite positive constant or $+\infty$. Since $\sum_{x<0} x p_X(x) = -\mathcal{E}(X^-) = -\lim_{N\to\infty} S_N^-$, the required

conclusion follows.
c) Note that

$$\mathcal{E}|X| = \sum_x |x| p_X(x) = \sum_{x>0} x p_X(x) + \sum_{x<0} (-x) p_X(x) = \mathcal{E}(X^+) + \mathcal{E}(X^-).$$

Since $\mathcal{E}(X^+)$ and $\mathcal{E}(X^-)$ are both positive, then $\mathcal{E}|X|$ is finite if and only if $\mathcal{E}(X^+)$ and $\mathcal{E}(X^-)$ are both finite. Therefore, X has finite expectation if and only if both $\mathcal{E}(X^+)$ and $\mathcal{E}(X^-)$ are real numbers, and in the latter case we have that

$$\mathcal{E}(X) = \sum_x x p_X(x) = \sum_{x>0} x p_X(x) + \sum_{x<0} x p_X(x) = \mathcal{E}(X^+) - \mathcal{E}(X^-).$$

d) The answers will vary. For example, the random variable X with the following PMF:

$$p_X(x) = \frac{6}{\pi^2 x^2}$$

for $x \in \mathcal{N}$, and $p_X(x) = 0$ otherwise, has an infinite expectation since

$$\mathcal{E}(X) = \sum_{x=1}^{\infty} x P(X = x) = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x} = \infty.$$

7.28 a) The probability that the first head appears on the n^{th} toss is $\left(\frac{1}{2}\right)^n$. Thus

$$\mathcal{E}(X) = \sum_{x=1}^{\infty} 2^x \left(\frac{1}{2}\right)^x = \sum_{x=1}^{\infty} 1,$$

which clearly diverges.
b) Since $\mathcal{E}(X) = \mathcal{E}(X^+) = \infty$ and $\mathcal{E}(X^-) = 0$, we have that $\mathcal{E}(X) = \infty$.
c) The result of (b) is a paradox since this implies that there is no value that the house could charge to make this game fair. No matter what finite value the house charges, everybody should play since in the long run the house will pay out an infinite amount of money. The only way this game could be "fair" would be if the house charged an infinite amount of money to play!

7.2 Basic Properties of Expected Value

Basic Exercises

7.29 a) Since each of the six sides are equally likely, the expectation of Y is $\sum_{y=1}^{6} \frac{y}{6} = 3.5$. The expectation of Z is also 3.5;
b) Since $X = Y + Z$, by linearity of expectation, $\mathcal{E}(X) = \mathcal{E}(Y + Z) = \mathcal{E}(Y) + \mathcal{E}(Z) = 7$.
c) In this case, since the PMFs of Y and Z are easier to compute than the PMF of X, the method used in (b) is easier than the work done in Exercise 7.1.

7.30 a) The PMF of $W = YZ$ is given by:

w	1	2	3	4	5	6	8	9	10	12	15	16	18	20	24	25	30	36
$p_W(w)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{36}$

Thus the expectation is

$$\mathcal{E}(W) = \sum_w wP(W = w) = \frac{49}{4} = 12.25$$

b) Since $W = YZ$, let $g(y, z) = yz$. Then $g(Y, Z) = W$ and by the FEF,

$$\mathcal{E}(W) = \mathcal{E}(g(Y, Z)) = \sum_{(y,z)}\sum g(y, z)p_{Y,Z}(y, z) = \sum_{y=1}^{6}\sum_{z=1}^{6} yz\,\frac{1}{36} = \frac{(1 + \cdots + 6)^2}{36} = \frac{21^2}{36} = 12.25$$

c) Since Y and Z are independent, then $\mathcal{E}(W) = \mathcal{E}(YZ) = \mathcal{E}(Y)\mathcal{E}(Z) = \frac{7}{2} \times \frac{7}{2} = \frac{49}{4} = 12.25$

7.31 a) The PMF of $U = \min\{Y, Z\}$ is $p_U(u) = (13-2u)/36$ for $u = 1, 2, 3, 4, 5, 6$, and $p_U(u) = 0$ otherwise. Thus the expectation of U is

$$\mathcal{E}(U) = \sum_u uP(U = u) = \sum_{u=1}^{6} \frac{13u - 2u^2}{36} = \frac{91}{36} \approx 2.52778$$

b) By the FEF,

$$\mathcal{E}(U) = \mathcal{E}(\min(Y, Z)) = \sum_{y=1}^{6}\sum_{z=1}^{6} \min(y, z)\frac{1}{36} = \frac{1}{36}(\sum_{y=1}^{6}\sum_{z=1}^{y} z) + \frac{1}{36}(\sum_{y=1}^{5}\sum_{z=y+1}^{6} y)$$

$$= \frac{1}{36}\sum_{y=1}^{6} \frac{y(y+1)}{2} + \frac{1}{36}\sum_{y=1}^{5} y(6 - y) = \frac{56}{36} + \frac{35}{36} = \frac{91}{36} \approx 2.52778$$

7.32 a) The PMF of $V = \max(Y, Z)$ is equal to $p_V(v) = (2v - 1)/36$ for $v = 1, 2, 3, 4, 5, 6$, and $p_V(v) = 0$ otherwise. Thus the expectation of V is

$$\mathcal{E}(V) = \sum_v vP(V = v) = \sum_{v=1}^{6} \frac{2v^2 - v}{36} = \frac{161}{36} \approx 4.47222$$

b) By the FEF,

$$\mathcal{E}(V) = \mathcal{E}(\max(Y, Z)) = \sum_{y=1}^{6}\sum_{z=1}^{6} \max(y, z)\frac{1}{36} = \frac{1}{36}(\sum_{y=1}^{6}\sum_{z=1}^{y} y) + \frac{1}{36}(\sum_{y=1}^{5}\sum_{z=y+1}^{6} z)$$

$$= \frac{1}{36}\sum_{y=1}^{6} y^2 + \frac{1}{36}\sum_{y=1}^{5}\left(21 - \frac{y(y+1)}{2}\right) = \frac{91}{36} + \frac{70}{36} = \frac{161}{36} \approx 4.47222$$

7.33 In this problem it is useful to note first that, by the binomial series theorem, we have the equality

$$(1 + t)^a - (1 - t)^a = \sum_{j=0}^{\infty} \binom{a}{j} \underbrace{(t^j - (-t)^j)}_{=0 \text{ if } j \text{ is even}} = 2 \sum_{j>0:\, j \text{ is odd}} \binom{a}{j} t^j. \tag{7.1}$$

a) Let X_i equal the number of successes in the ith Bernoulli trial, $i = 1, \cdots, n$. Then $Y = 1_{\{X_1+\cdots+X_n \text{ is odd}\}}$ and

$$\begin{aligned} p_{X_1,\cdots,X_n}(x_1,\cdots,x_n) &= p_{X_1}(x_1)\dots p_{X_n}(x_n) = p^{x_1}(1-p)^{1-x_1}\dots p^{x_n}(1-p)^{1-x_n} \\ &= p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)} \end{aligned}$$

for all $(x_1,\dots,x_n) \in \{0,1\}^n$ (and $p_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = 0$ otherwise). Then

$$\mathcal{E}(Y) = \sum_{x_1=0}^{1}\cdots\sum_{x_n=0}^{1} 1_{\{x_1+\cdots+x_n \text{ is odd}\}}\, p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$$

$$= \sum_{j>0:\ j \text{ is odd}} \left(\sum_{x_1=0}^{1}\cdots\sum_{x_n=0}^{1} 1_{\{x_1+\cdots+x_n=j\}}\right) p^j(1-p)^{n-j} = \sum_{j>0:\ j \text{ is odd}} \binom{n}{j} p^j(1-p)^{n-j}$$

$$= (1-p)^n\,\frac{1}{2}\left(\left(1+\frac{p}{1-p}\right)^n - \left(1-\frac{p}{1-p}\right)^n\right) = \frac{1}{2}(1-(1-2p)^n),$$

where we used equality (7.1) (which was proved earlier).

b) Let S_n denote the number of successes in n Bernoulli trials. Then

$$\mathcal{E}(Y) = \mathcal{E}(1_{\{S_n \text{ is odd}\}}) = P(S_n \text{ is odd}) = \sum_{j>0:\ j \text{ is odd}} P(S_n = j)$$

$$= \sum_{j>0:\ j \text{ is odd}} \binom{n}{j} p^j(1-p)^{n-j} = (1-p)^n \sum_{j>0:\ j \text{ is odd}} \binom{n}{j}\left(\frac{p}{1-p}\right)^j$$

$$= (1-p)^n\,\frac{1}{2}\left(\left(1+\frac{p}{1-p}\right)^n - \left(1-\frac{p}{1-p}\right)^n\right) = \frac{1}{2}(1-(1-2p)^n),$$

where we used equality (7.1) (which was proved earlier).

7.34 a) X, Y and Z have a multinomial distribution with parameters $n = 4$, and $\frac{\pi}{36}, \frac{\pi}{12}$ and $1 - \frac{\pi}{9}$. Thus, the joint PMF of X, Y, Z is given by

$$p_{X,Y,Z}(x,y,z) = \binom{4}{x,y,z}\left(\frac{\pi}{36}\right)^x\left(\frac{\pi}{12}\right)^y\left(1-\frac{\pi}{9}\right)^z$$

for all $x, y, z \in \mathcal{Z}_+$ such that $x + y + z = 4$; and $p_{X,Y,Z}(x,y,z) = 0$ otherwise.

b) Since $T = 10X + 5Y + 0Z$, let $g(x,y,z) = 10x + 5y$. Then $g(X,Y,Z) = T$ and by the FEF,

$$\mathcal{E}(T) = \sum\sum\sum_{(x,y,z)} g(x,y,z)p_{X,Y,Z}(x,y,z) \approx 8.72665$$

c) We have that $X \sim \mathcal{B}\left(4, \frac{\pi}{36}\right)$, $Y \sim \mathcal{B}\left(4, \frac{\pi}{12}\right)$ and $Z \sim \mathcal{B}\left(4, 1-\frac{\pi}{9}\right)$. Thus $\mathcal{E}(X) = \frac{\pi}{9}$, $\mathcal{E}(Y) = \frac{\pi}{3}$ and $\mathcal{E}(Z) = 4\left(1-\frac{\pi}{9}\right)$.

d) Since $T = 10X + 5Y + 0Z$, then, by the linearity property of expected value, $\mathcal{E}(T) = 10\mathcal{E}(X) + 5\mathcal{E}(Y) = \frac{25\pi}{9} \approx 8.72665$

e) Let W_i be the score obtained on the ith shot. Then, from Exercise 7.2, we have that $\mathcal{E}(W_i) = \frac{25\pi}{36}$ for each i. Since $T = W_1 + W_2 + W_3 + W_4$, then $\mathcal{E}(T) = 4\mathcal{E}(W_1) = \frac{25\pi}{9}$.

f) We can compute the PMF of T, and then use the definition of expectation directly.

7.35 1. Let E_i be the event that the ith man sits across from his wife, $i = 1, \ldots, N$. Then the total number of men who sit across from their wives is equal to $\sum_{i=1}^{N} 1_{E_i}$, thus, the corresponding expectation is equal to

$$\mathcal{E}(\sum_{i=1}^{N} 1_{E_i}) = \sum_{i=1}^{N} \mathcal{E}(1_{E_i}) = \sum_{i=1}^{N} P(E_i) = \sum_{i=1}^{N} \frac{1}{N} = 1.$$

7.36 Since we expect to have np_1 items with 1 defect and np_2 items with 2 or more defects, the expected cost of repairing the defective items in the sample will be $\$1 \times np_1 + \$3 \times np_2$.

7.37 14.7; Let X_k be the number of throws from the appearance of the $(k-1)^{th}$ distinct number until the appearance of the k^{th} distinct number, where $X_1 = 1$. Then $X_i \sim \mathcal{G}\left(\frac{7-i}{6}\right)$ for $i = 1, 2, 3, 4, 5, 6$. Thus $\mathcal{E}(X_i) = \frac{6}{7-i}$, and the expected number of throws of a balanced die to get all six possible numbers is equal to $\sum_{i=1}^{6} \mathcal{E}(X_i) = \sum_{i=1}^{6} \frac{6}{7-i} = 14.7$

7.38 a) For investment A,

$$\mathcal{E}(Y_A) = 1 \times 0.5 + 4 \times 0.4 = 2.1$$

For investment B,

$$\mathcal{E}(Y_B) = 1 \times 0.3 + 16 \times 0.2 = 3.5$$

Investment A has a lower expected yield than investment B.
b) For investment A,

$$\mathcal{E}(\sqrt{Y_A}) = 1 \times 0.5 + 2 \times 0.4 = 1.3$$

For investment B,

$$\mathcal{E}(\sqrt{Y_B}) = 1 \times 0.3 + 4 \times 0.2 = 1.1$$

Using the square root utility function, which decreases the effect of large earnings, investment A has a higher expected utility than investment B.
c) For investment A,

$$\mathcal{E}(Y_A^{3/2}) = 1 \times 0.5 + 8 \times 0.4 = 3.7$$

For investment B,

$$\mathcal{E}(Y_B^{3/2}) = 1 \times 0.3 + 64 \times 0.2 = 13.1$$

Using the above utility function, which increases the effect of large earnings, investment B has a higher expected utility than investment A.

7.39 a) $\frac{16}{17}$. Let X_2 be the number of items (in a given lot of 17 items) that are inspected by both engineers. Then X_2 has $\mathcal{B}(17, (4/17)^2)$ distribution. Therefore, $\mathcal{E}(X_2) = 17(\frac{4}{17})^2 = \frac{16}{17}$.
b) $\frac{169}{17}$. Let X_0 be the number of items (in a given lot of 17 items) that are inspected by neither engineer. Then X_0 is $\mathcal{B}(17, (13/17)^2)$ and, thus, $\mathcal{E}(X_0) = 17(\frac{13}{17})^2 = \frac{169}{17}$.
c) $\frac{104}{17}$. Let X_1 be the number of items (in a lot of 17) inspected by just one of the two engineers. Then X_1 is $\mathcal{B}(n = 17, p = 2 \times \frac{4}{17} \times \frac{13}{17})$, thus, $\mathcal{E}(X_1) = 17 \times 2 \times \frac{4}{17} \times \frac{13}{17} = \frac{104}{17}$.
d) 17, since $X_0 + X_1 + X_2 = 17$.

7.40 a) $\frac{1}{p(2-p)}$. Let A and B represent the time it takes for each of the two components to fail.

Let X represent the time it takes for the first component to fail (i.e. $X = \min(A, B)$). Then, upon using tail probabilities, one obtains that

$$\mathcal{E}(X) = \sum_{x=0}^{\infty} P(X > x) = \sum_{x=0}^{\infty} P(A > x, B > x) = \sum_{x=0}^{\infty} P(A > x)P(B > x)$$

$$= \sum_{x=0}^{\infty} (1-p)^{2x} = \frac{1}{1-(1-p)^2} = \frac{1}{(2-p)p}.$$

b) $\frac{3-2p}{p(2-p)}$. Let Y represent the time that it takes for the second component to fail (i.e. $Y = \max(A, B)$). Then, using tail probabilities,

$$\mathcal{E}(Y) = \sum_{y=0}^{\infty} P(Y > y) = \sum_{y=0}^{\infty} (1 - P(Y < y+1)) = \sum_{y=0}^{\infty} (1 - P(A < y+1, B < y+1))$$

$$= \sum_{y=0}^{\infty} (1 - P(A < y+1)P(B < y+1)) = \sum_{y=0}^{\infty} [1 - (1 - P(A > y))(1 - P(B > y))]$$

$$= \sum_{y=0}^{\infty} \left[1 - (1-(1-p)^y)^2\right] = \sum_{y=0}^{\infty} (1-p)^y(2-(1-p)^y) = 2\sum_{y=0}^{\infty}(1-p)^y - \sum_{y=0}^{\infty}(1-p)^{2y}$$

$$= \frac{2}{p} - \frac{1}{(2-p)p} = \frac{3-2p}{(2-p)p}.$$

7.41 a) $\frac{1210}{61}$. Let A and B represent the arrival times of the two people. Let X represent the time it takes for the first person to arrive (i.e. $X = \min(A, B)$). Then, using tail probabilities,

$$\mathcal{E}(X) = \sum_{x=0}^{\infty} P(X > x) = \sum_{x=0}^{59} P(A > x, B > x) = \sum_{x=0}^{59} P(A > x)P(B > x)$$

$$= \sum_{x=0}^{59} \left(\frac{60-x}{61}\right)^2 = \frac{1210}{61} \approx 19.8361$$

b) $\frac{2450}{61}$. Let Y represent the time that it takes for the second person to arrive (i.e. $Y = \max(A, B)$). Then, using tail probabilities,

$$\mathcal{E}(Y) = \sum_{y=0}^{\infty} P(Y > y) = \sum_{y=0}^{59} (1 - P(Y < y+1)) = \sum_{y=0}^{59} (1 - P(A < y+1, B < y+1))$$

$$= \sum_{y=0}^{59} (1 - P(A < y+1)P(B < y+1)) = \sum_{y=0}^{59} \left[1 - \left(\frac{y+1}{61}\right)^2\right] = \frac{2450}{61} \approx 40.1639$$

c) $\frac{1240}{61}$. Let Z represent the time that the first person has to wait for the second. Then $Z = |A - B| = Y - X$, thus, using answers to (a),(b), one deduces that

$$\mathcal{E}(Z) = \mathcal{E}(Y) - \mathcal{E}(X) = \frac{2450}{61} - \frac{1210}{61} = \frac{1240}{61} \approx 20.3279$$

7.42 For any random variable X with finite expectation, since the random variable $(X - \mathcal{E}(X))^2$ is always nonnegative, so is its expectation, and thus

$$\mathcal{E}(X - \mathcal{E}(X))^2 \geq 0, \text{ i.e. } \mathcal{E}(X^2 - 2\mathcal{E}(X)X + (\mathcal{E}(X))^2) \geq 0,$$

which (by linearity of expectation) implies that $\mathcal{E}(X^2) - (\mathcal{E}(X))^2 \geq 0$.

7.43 a) X and Y are linearly uncorrelated if and only if

$$\mathcal{E}\left((X - \mathcal{E}(X))\left(Y - \mathcal{E}(Y)\right)\right) = 0,$$

which is equivalent to

$$\mathcal{E}\left(XY - X\mathcal{E}(Y) - Y\mathcal{E}(X) + \mathcal{E}(X)\mathcal{E}(Y)\right) = 0$$

or, equivalently,

$$\mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) = 0,$$

thus, the desired result follows.
b) Yes. If X and Y are independent, then $\mathcal{E}(XY) = \mathcal{E}(X)\mathcal{E}(Y)$, which, by (a), implies that X and Y are linearly uncorrelated.
c) No. For example, the random variables with joint PDF shown in Table 7.8 are linearly uncorrelated but not independent.

Theory Exercises

7.44 a) Since $X = Y_1 + \cdots + Y_r$ for some (independent) $\mathcal{G}(p)$ random variables $Y_1, \ldots, Y_r$, then

$$\mathcal{E}(X) = \sum_{i=1}^{r} \mathcal{E}(Y_i) = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}.$$

b) No, independence of the geometric random variables is not required for the result in (a), since the proof in (a) relies only on the linearity property of expectation (which holds regardless of the dependence structure between the random variables).

7.45 To see that for arbitrary discrete random variables defined on the same probability space with finite expectations and arbitrary $c_1, \cdots, c_m \in \mathbb{R}$, $m \in \mathcal{N}$, the equality

$$\mathcal{E}\left(\sum_{j=1}^{m} c_j X_j\right) = \sum_{j=1}^{m} c_j \mathcal{E}(X_j) \tag{7.2}$$

holds, note first that (7.2) holds for $m = 2$ by Proposition 7.3. Next assume that (7.2) holds for all $m \leq n$. Then

$$\mathcal{E}\left(\sum_{j=1}^{n+1} c_j X_j\right) = \mathcal{E}\left((\sum_{j=1}^{n} c_j X_j) + c_{n+1}X_{n+1}\right) = \mathcal{E}\left(\sum_{j=1}^{n} c_j X_j\right) + c_{n+1}\mathcal{E}(X_{n+1})$$

$$= \sum_{j=1}^{n} c_j \mathcal{E}(X_j) + c_{n+1}\mathcal{E}(X_{n+1}) = \sum_{j=1}^{n+1} c_j \mathcal{E}(X_j).$$

Thus, (7.2) holds for $m = n+1$, which by mathematical induction implies that (7.2) holds for arbitrary $m \in \mathcal{N}$.

7.46 a) Let $X_1, \ldots, X_m$ be independent discrete random variables defined on the same sample space and having finite expectation, and let $c_1, \ldots, c_m \in \mathcal{R}$. Claim: $\prod_{j=1}^m X_j$ has finite expectation and $\mathcal{E}(\prod_{j=1}^m X_j) = \prod_{j=1}^m \mathcal{E}(X_j)$. Note that the claim holds true for $m = 2$ in view of Proposition 7.5. Next assume that the claim is true for all $m \leq n$. Then

$$\mathcal{E}\left(\prod_{j=1}^{n+1} X_j\right) = \mathcal{E}\left((\prod_{j=1}^{n} X_j)X_{n+1}\right) = \mathcal{E}\left(\prod_{j=1}^{n} X_j\right)\mathcal{E}(X_{n+1}) = \left(\prod_{j=1}^{n} \mathcal{E}(X_j)\right)\mathcal{E}(X_{n+1}) = \prod_{j=1}^{n+1} \mathcal{E}(X_j),$$

and the required conclusion follows by mathematical induction principle.

b) Here we have

$$\begin{aligned}
\mathcal{E}\left(\prod_{j=1}^{m} X_j\right) &= \sum_{(x_1,\ldots,x_m)} \!\!\cdots \sum x_1 \cdots x_m p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) \\
&= \sum_{(x_1,\ldots,x_m)} \!\!\cdots \sum x_1 \cdots x_m p_{X_1}(x_1) \cdots p_{X_m}(x_m) \\
&= \sum_{x_1} x_1 p_{X_1}(x_1) \cdots \sum_{x_m} x_m p_{X_m}(x_m) = \prod_{j=1}^{m} \mathcal{E}(X_j),
\end{aligned}$$

where all the series involved are absolutely convergent.

7.47 By Exercise 7.27, it follows that

$$\begin{aligned}
|\mathcal{E}(X)| = |\mathcal{E}(X^+) - \mathcal{E}(X^-)| \leq |\mathcal{E}(X^+)| + |\mathcal{E}(X^-)| &= \sum_{x>0} x p_X(x) + \sum_{x<0} (-x) p_X(x) \\
&= \sum_{x} |x| p_X(x) = \mathcal{E}|X|.
\end{aligned}$$

7.48 First, since $P(|X| \leq M) = 1$, then $P(X = x) = 0$ for all x such that $|x| > M$, therefore,

$$\begin{aligned}
\mathcal{E}(|X|) = \sum_{x} |x| P(X = x) = \sum_{|x| \leq M} |x| P(X = x) &\leq \sum_{|x| \leq M} M P(X = x) \\
&= M \sum_{|x| \leq M} P(X = x) = M.
\end{aligned}$$

Then, in light of Exercise 7.47, the desired result follows.

7.49 a) For any real $t > 0$ and any nonnegative random variable X,

$$\begin{aligned}
tP(X \geq t) = t\sum_{x=t}^{\infty} P(X = x) = \sum_{x=t}^{\infty} tP(X = x) &\leq \sum_{x=t}^{\infty} xP(X = x) \\
&\leq \sum_{x=0}^{\infty} xP(X = x) = \mathcal{E}(X).
\end{aligned}$$

b) Consider an arbitrary nonnegative random variable X. First, if $X \geq t$ then $tI_{\{X\geq t\}} = t \leq X$. Second, if $0 \leq X < t$ then $tI_{\{X\geq t\}} = 0 \leq X$. Thus, $tI_{\{X\geq t\}} \leq X$ for all $t > 0$.
c) The desired result follows upon taking expectation in the inequality in (b), due to the monotonicity property of expectation.

Advanced Exercises

7.50 a) 4.5; Let Y be the first digit of the decimal expansion of X. Then the PMF of Y is given by:

$$p_Y(y) = P\left(\frac{y}{10} \leq X < \frac{y+1}{10}\right) = 0.1 \;\text{ for } y \in \{0, 1, \ldots, 9\},$$

and $p_Y(y) = 0$ otherwise, implying that $\mathcal{E}(Y) = \sum_{j=0}^{9} 0.1j = 4.5$
b) 6.15; Let U be the first digit of the decimal expansion of $\sqrt{X}$. Then the PMF of U is given by:

$$p_U(u) = P\left(\frac{u}{10} \leq \sqrt{X} < \frac{u+1}{10}\right) = P\left(\frac{u^2}{100} \leq X < \frac{(u+1)^2}{100}\right) = \frac{(u+1)^2 - u^2}{100} = \frac{2u+1}{100}$$

for $u \in \{0, 1, \ldots, 9\}$ (and $p_U(u) = 0$ otherwise). Therefore, $\mathcal{E}(U) = \sum_{u=0}^{9} \frac{u(2u+1)}{100} = \frac{123}{20} = 6.15$
c) 4.665; Let V be the second digit of the decimal expansion of $\sqrt{X}$. Then the PMF of V is given by:

$$p_V(v) = \sum_{k=0}^{9} P\left(\frac{10k+v}{100} \leq \sqrt{X} < \frac{10k+v+1}{100}\right) = \sum_{k=0}^{9} P\left(\frac{(10k+v)^2}{10^4} \leq X < \frac{(10k+v+1)^2}{10^4}\right)$$

$$= \sum_{k=0}^{9} \frac{(10k+v+1)^2 - (10k+v)^2}{10^4} = \sum_{k=0}^{9} \frac{20k+2v+1}{10^4} = \frac{91+2v}{10^3}, \quad v \in \{0, 1, \ldots, 9\};$$

and $p_V(v) = 0$ otherwise. Then $\mathcal{E}(V) = \sum_{v=0}^{9} \frac{91v+2v^2}{10^3} = 4.665$

7.51 Let A_i be the event that the ith customer is a woman and the $(i+1)$st customer is a man, $i = 1, \ldots, 49$. Then the total number of times that a woman is followed by a man is equal to $\sum_{i=1}^{49} 1_{A_i}$, thus the expected number of times that a woman is followed by a man is equal to

$$\mathcal{E}\left(\sum_{i=1}^{49} 1_{A_i}\right) = \sum_{i=1}^{49} P(A_i) = \sum_{i=1}^{49} 0.6 \times 0.4 = 49 \times 0.24 = 11.76,$$

where above we assume that there is always a 40% chance that a randomly selected customer is a man and that customers are independent of each other.

On the other hand, if we are told that, in a given sample of fifty customers who entered the store, exactly twenty customers (40%) were men (and customers were sampled without replacement), then the expected number of times that a woman was followed by a man is equal to

$$\mathcal{E}\left(\sum_{i=1}^{49} 1_{A_i}\right) = \sum_{i=1}^{49} P(A_i) = \sum_{i=1}^{49} \frac{\binom{48}{19}}{\binom{50}{20}} = 12.$$

7.52 a) Since $X \geq 0$, then $|t^X| \leq 1$ for all $t \in [-1, 1]$. Thus, the desired result follows from Exercise 7.48.
b) Clearly, for a nonnegative integer-valued random variable X,

$$P_X(t) = \mathcal{E}(t^X) = \sum_{n=0}^{\infty} t^n p_X(n).$$

c) Note that

$$P'_X(t) = \frac{d}{dt}\mathcal{E}(t^X) = \frac{d}{dt}\sum_{x=0}^{\infty} t^x P(X = x) = \sum_{x=1}^{\infty} x t^{x-1} P(X = x),$$

thus,

$$P'_X(1) = \sum_{x=1}^{\infty} x P(X = x) = \mathcal{E}(X).$$

7.53 a) The PGF for a $\mathcal{B}(n, p)$ random variable X is given by:

$$P_X(t) = \sum_{k=0}^{n} t^k \binom{n}{k} p^k (1-p)^{n-k} = (pt + 1 - p)^n,$$

thus, $P'_X(t) = n(pt + 1 - p)^{n-1} p$, implying that $\mathcal{E}(X) = P'_X(1) = np$.
b) The PGF for a $\mathcal{P}(\lambda)$ random variable X is given by:

$$P_X(t) = \sum_{k=0}^{\infty} t^k \cdot \frac{e^{-\lambda}\lambda^k}{k!} = e^{\lambda(t-1)},$$

thus, $P'_X(t) = \lambda e^{\lambda(t-1)}$ and $\mathcal{E}(X) = P'_X(1) = \lambda$.
c) The PGF for a $\mathcal{G}(p)$ random variable X is given by:

$$P_X(t) = \sum_{k=1}^{\infty} t^k p (1-p)^{k-1} = \frac{pt}{1 - (1-p)t} = \frac{pt}{1 - t + pt},$$

thus, $P'_X(t) = \dfrac{p}{(1 - t + pt)^2}$, implying that $\mathcal{E}(X) = P'_X(1) = \dfrac{1}{p}$.
d) The PGF for a $\mathcal{NB}(r, p)$ random variable X is given by:

$$\begin{aligned}
P_X(t) &= \sum_{k=r}^{\infty} t^k \binom{k-1}{r-1} p^r (1-p)^{k-r} \\
&= t^r p^r (1 - t(1-p))^{-r} \underbrace{\sum_{k=r}^{\infty} \binom{k-1}{r-1} (1 - t(1-p))^r (t(1-p))^{k-r}}_{=1} \\
&= \left(\frac{tp}{1 - t + tp}\right)^r.
\end{aligned}$$

Thus, $P'_X(t) = \dfrac{rp^r t^{r-1}}{(1-t+tp)^{r+1}}$, which implies that $\mathcal{E}(X) = P'_X(1) = \dfrac{r}{p}$.
e) The PGF for the Uniform distribution on $\{1, \ldots, N\}$ is

$$P_X(t) = \sum_{k=1}^{N} t^k \frac{1}{N} = \frac{t(1-t^N)}{N(1-t)},$$

then $P'_X(t) = \dfrac{1 - t^N + Nt^N(t-1)}{N(1-t)^2} = \dfrac{1 + t + \cdots + t^{N-1} - Nt^N}{N(1-t)} = \displaystyle\sum_{k=0}^{N-1} \frac{t^k - t^N}{N(1-t)}$
$= \displaystyle\sum_{k=0}^{N-1} \frac{t^k(1-t^{N-k})}{N(1-t)} = \sum_{k=0}^{N-1} \frac{t^k}{N}(1 + t + \cdots + t^{N-k-1})$. Therefore,

$$\mathcal{E}(X) = P'_X(1) = \sum_{k=0}^{N-1} \frac{N-k}{N} = \frac{1}{N}\sum_{\ell=1}^{N} \ell = \frac{N(N+1)}{2N} = \frac{N+1}{2}.$$

7.3 Variance of Discrete Random Variables

Basic Exercises

7.54 a) 5.833; Using Equation 7.30 and Exercise 7.1, we have

$$\mathrm{Var}(X) = \mathcal{E}\left((X - \mu_X)^2\right) = \sum_{x=2}^{12} (x-7)^2 p_X(x) = 5.833$$

b) 5.833; Since $\mathcal{E}(X^2) = 54.833$, $\mathrm{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = 5.833$

7.55 10.512; From Exercise 5.27, we have that $\mathcal{E}(X^2) = \frac{175\pi}{36}$. Thus, using the computing formula for variance, $\mathrm{Var}(X) = \frac{175\pi}{36} - (\frac{25\pi}{36})^2 \approx 10.512$

7.56 a) 0.6; Using Exercise 5.28, we have that when $p = 0.5$, $\mathcal{E}(X^2) = 1.6$ and $\mathcal{E}(X) = 1$, thus, using the computing formula for variance, $\mathrm{Var}(X) = 1.6 - 1^2 = 0.6$.
b) 0.5856; Using Exercise 5.28, we have that when $p = 0.4$, $\mathcal{E}(X^2) = 1.36$ and $\mathcal{E}(X) = 0.88$, thus, using the computing formula for variance, $\mathrm{Var}(X) = 0.5856$.

7.57 a) 72/121. Let X be the number of women chosen. Using Exercise 5.29 (a), we have that $\mathcal{E}(X^2) = \frac{27}{11}$, and using the computing formula for variance, $\mathrm{Var}(X) = \frac{27}{11} - (\frac{15}{11})^2 = \frac{72}{121}$.
b) 72/121. Note that X has a hypergeometric distribution with parameters $N = 11$, $n = 3$ and $p = \frac{5}{11}$. Thus, using Table 7.10, we have that $\mathrm{Var}(X) = \frac{3\times(5/11)\times(6/11)\times(11-3)}{11-1} = \frac{72}{121}$, which is the same result as in (a).

7.58 10.24; Let $\{\omega_1, \ldots, \omega_N\}$ be the finite population in question and y_i $(i = 1, \cdots, N)$ be the value of the variable of interest for the ith member of the population (i.e. for ω_i). If X is the value of the variable of interest for a randomly selected member of the population, then, by an argument similar to the one given in Exercise 7.7, we have

$$\mathrm{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = \mathcal{E}(X^2) - \bar{y}^2 = \frac{1}{N}\sum_{i=1}^{N} X^2(\omega_i) - \bar{y}^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = 3.2^2 = 10.24$$

7.59 a) 600. Let X be the number of people tested until the first positive test occurs. Then X has a $\mathcal{G}(p)$ distribution (with some $p \in (0,1)$). Since, on average, the first positive test occurs with the 25^{th} person, then $\mathcal{E}(X) = 25 = 1/p$, implying that $p = 1/25 = 0.04$, thus, using Table 7.10, one obtains that $\text{Var}(X) = (1-p)/p^2 = 0.96/0.0016 = 600$.
b) 3000. Let Y be the number of tests performed until the fifth person who tests positive is found. Then Y has a negative binomial distribution with parameters 5 and $\frac{1}{25}$. Thus using Table 7.10, $\text{Var}(Y) = 3000$.

7.60 1.8; Let X be the number of children cured. Then X has $\mathcal{B}(20, 0.9)$ distribution and $\text{Var}(X) = 20 \times 0.9 \times 0.1 = 1.8$.

7.61 $\frac{\sqrt{15}}{4}$. If both parents have the sickle cell trait, then the probability that they will pass it on to their child is 0.25, and the number of children (out of 5) who will have the sickle cell anemia has a binomial distribution with parameters 5 and 0.25. Thus the variance is $\frac{15}{16} = 0.9375$ and the standard deviation is $\frac{\sqrt{15}}{4} \approx 0.968$

7.62 a) $\frac{14}{25}$. When sampling without replacement, the number of selected undergraduates has a hypergeometric distribution with parameters $N = 10$, $n = 3$ and $p = 0.6$, thus the variance is $\frac{14}{25} = 0.56$
b) When sampling with replacement, the number of selected undergraduates has a binomial distribution with $n = 3$ and $p = 0.6$, thus the variance is $\frac{18}{25} = 0.72$
c) While the expectations for the hypergeometric and binomial distributions (given above) are the same, the corresponding variances are different.

7.63 $\text{Var}(X) \approx 0.271$, since X has a hypergeometric distribution with parameters $N = 100$, $n = 5$ and $p = 0.06$.

7.64 a) $\left(\frac{N-n}{N-1}\right) np(1-p)$. Without replacement, if X represents the number of members sampled that have the specified attribute, then X has a hypergeometric distribution with parameters N, n and p.
b) $np(1-p)$. With replacement, if X represents the number of members sampled that have the specified attribute, then X has a $\mathcal{B}(n,p)$ distribution and the required answer follows.

7.65 15. Since the arrival of cars has a Poisson distribution with parameter $\lambda = 15$, then the variance is 15.

7.66 9.3; If the purchases are independent, then the distribution of the number of customers per hour who buy a café mocha with no whipped cream is Poisson with $\lambda = 31 \times 0.3 = 9.3$; since the variance of $\mathcal{P}(\lambda)$ is equal to λ, the desired answer then follows.

7.67 3.77; From Exercise 5.85, letting X be the number of eggs observed in a nonempty nest, we find that $\mathcal{E}(X^2) = \frac{20}{1-e^{-4}}$, since

$$\mathcal{E}(X^2) = \sum_{k=1}^{\infty} k^2 \frac{e^{-4}4^k}{k!(1-e^{-4})} = \frac{4}{1-e^{-4}} \sum_{k=1}^{\infty} k \cdot \frac{e^{-4}4^{k-1}}{(k-1)!} = \frac{4}{1-e^{-4}} \underbrace{\sum_{\ell=0}^{\infty} (\ell+1) \frac{e^{-4}4^{\ell}}{\ell!}}_{=\mathcal{E}(Y+1)} = \frac{20}{1-e^{-4}},$$

where Y is $\mathcal{P}(4)$, thus $\mathcal{E}(Y+1) = \mathcal{E}(Y) + 1 = 4 + 1 = 5$. Therefore, by the computing formula for expectation and Exercise 7.18,

$$\mathrm{Var}(X) = \frac{20}{1-e^{-4}} - \left(\frac{4}{1-e^{-4}}\right)^2 \approx 3.77054$$

7.68 a) Since the distribution of the occurrence of the first hit has a geometric distribution with parameter 0.26, the variance is $\frac{1-0.26}{0.26^2} \approx 10.94675$
b) Since the distribution of the occurrence of the second hit has a negative binomial distribution with parameters 2, and 0.26, the variance is 21.8935
c) Since the distribution of the occurrence of the tenth hit has a negative binomial distribution with parameters 10 and 0.26, the variance is 109.4675

7.69 Noting that $I_E^2 = I_E$, we have that $\mathcal{E}(I_E^2) = \mathcal{E}(I_E) = P(E)$, thus, by the computing formula for expectation, $\mathrm{Var}(I_E) = \mathcal{E}(I_E^2) - (\mathcal{E}(I_E))^2 = P(E) - (P(E))^2 = P(E)(1-P(E))$.

7.70 Since I_E has a binomial distribution with parameters $n = 1$ and $p = P(E)$, then

$$\mathrm{Var}(I_E) = np(1-p) = 1 \cdot P(E)(1-P(E)) = P(E)(1-P(E)).$$

7.71 By linearity of expectation,

$$\mathcal{E}(X^*) = \mathcal{E}\left(\frac{X-\mu_X}{\sigma_X}\right) = \frac{\mathcal{E}(X)-\mu_X}{\sigma_X} = 0.$$

By Proposition 7.10,

$$\mathrm{Var}(X^*) = \mathrm{Var}\left(\frac{X-\mu_X}{\sigma_X}\right) = \frac{1}{\sigma_X^2}\mathrm{Var}(X-\mu_X) = \frac{1}{\sigma_X^2}\mathrm{Var}(X) = 1.$$

7.72 a) First note that X_n is symmetric about 0, thus $\mathcal{E}(X_n) = 0$. Then

$$\mathrm{Var}(X_n) = \mathcal{E}(X_n^2) = 0^2(1-\frac{1}{n^2}) + n^2\frac{2}{2n^2} = 1.$$

Then, by Chebyshev's inequality,

$$P(|X_n| \geq 3) \leq \frac{1}{9}.$$

b) By direct calculation, for all $n \geq 3$, $P(|X_n| \geq 3) = \frac{1}{n^2}$.
c) The upper bound, provided by the Chebyshev's inequality, is exact for $n = 3$, but gets worse as n increases, since the exact probability goes to 0 as n goes to infinity.

7.73 a) The probability that a random variable is less than k standard deviations away from its mean is at least $1 - \frac{1}{k^2}$.
b) The probability that a random variable is less than 2 standard deviations away from its mean is at least $\frac{3}{4}$. The probability that a random variable is less than 3 standard deviations away from its mean is at least $\frac{8}{9}$.
c) Proposition 7.11 is equivalent to

$$P(|X-\mu_X| < t) \geq 1 - \frac{\mathrm{Var}(X)}{t^2}$$

for all positive real t. Then the desired relation (*) follows upon taking t equal to $k\sigma_X$.

7.74 $(N^2-1)/12$. Here, since $p_X(x) = \frac{1}{N}$ for all $x \in \{1, 2, \ldots, N\}$ and $p_X(x) = 0$ otherwise, then

$$\mathcal{E}(X) = \sum_{x=1}^{N} \frac{x}{N} = \frac{1}{N}\sum_{x=1}^{N} x = \frac{N+1}{2}$$

and

$$\mathcal{E}(X^2) = \sum_{x=1}^{N} \frac{x^2}{N} = \frac{1}{N}\sum_{x=1}^{N} x^2 = \frac{(N+1)(2N+1)}{6},$$

implying that

$$\text{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = \frac{N^2-1}{12}.$$

7.75 a) Let $f(c) = \mathcal{E}((Y-c)^2)$. Then $f'(c) = \mathcal{E}(2(c-Y)) = 2(c - \mathcal{E}(Y))$. Therefore $f'(c) > 0$ for all $c > \mathcal{E}(Y)$, and $f'(\mathcal{E}(Y)) = 0$, and $f'(c) < 0$ for all $c < \mathcal{E}(Y)$. Thus, $f(c)$ achieves its minimum at the point $c = \mathcal{E}(Y)$.
b) By (a), $\min_c \mathcal{E}((Y-c)^2) = f(\mathcal{E}(Y)) = \mathcal{E}(Y - \mathcal{E}(Y))^2 = \text{Var}(Y)$.

Theory Exercises

7.76 First note that if $|x| < 1$ then $|x|^{r-1} \le 1$ for all $r \in \mathcal{N}$, thus, $|x|^{r-1} \le |x|^r + 1$. On the other hand, if $|x| \ge 1$, then $|x|^{r-1} \le |x|^r$ for all $r \in \mathcal{N}$, thus, $|x|^{r-1} \le |x|^r + 1$. Therefore, for all $x \in \mathcal{R}$ and $r \in \mathcal{N}$, the inequality $|x|^{r-1} \le |x|^r + 1$ holds. Thus, for a discrete random variable X with finite rth moment (for some arbitrary fixed $r \in \mathcal{N}$), we have that

$$\mathcal{E}|X|^{r-1} = \sum_x |x|^{r-1} p_X(x) \le \sum_x (|x|^r + 1) p_X(x) = \sum_x |x|^r p_X(x) + \sum_x p_X(x) = \mathcal{E}|X|^r + 1 < \infty.$$

Thus X has a moment of order $r-1$. Therefore, inductively, X has all lower order moments.

7.77 First assume that $X - \mu_X$ has a finite r^{th} moment. Then, by Exercise 7.76, it follows that $\mathcal{E}|X - \mu_X|^k < \infty$ for all $k \in \{0, 1, \ldots, r\}$, which implies that

$$\mathcal{E}|X|^r = \mathcal{E}|X - \mu_X + \mu_X|^r \le \mathcal{E}(|X - \mu_X| + |\mu_X|)^r = \mathcal{E}\left(\sum_{k=0}^{r} \binom{r}{k} |X - \mu_X|^k |\mu_X|^{r-k}\right)$$

$$= \sum_{k=0}^{r} \binom{r}{k} |\mu_X|^{r-k} \mathcal{E}|X - \mu_X|^k < \infty,$$

in view of the binomial theorem and linearity property of expected value. Thus, X has a finite rth moment.

Next assume that X has a finite r^{th} moment. Then from Exercise 7.76, X has a finite k^{th} moment for all $0 \le k \le r$. Hence

$$\mathcal{E}|X - \mu_X|^r = \sum_x |x - \mu_X|^r p_X(x) \le \sum_x (|x| + |\mu_X|)^r p_X(x) = \sum_x \left(\sum_{k=0}^{r} \binom{r}{k} |x|^k |\mu_X|^{r-k}\right) p_X(x)$$

$$= \sum_{k=0}^{r} \binom{r}{k} |\mu_X|^{r-k} \left(\sum_x |x|^k p_X(x) \right) < \infty,$$

hence, $X - \mu_X$ has a finite r^{th} moment.

7.78 Since $P(|X| > M) = 0$, then $p_X(x) = 0$ for all real x such that $|x| > M$. Then for all $r \in \mathcal{N}$,

$$\sum_x |x|^r p_X(x) = \sum_{x:\, |x| \le M} |x|^r p_X(x) \le \sum_{x:\, |x| \le M} M^r p_X(x) = M^r P(|X| \le M) = M^r,$$

implying that X has finite moments of all orders and

$$|\mathcal{E}(X^r)| = \left| \sum_x x^r p_X(x) \right| \le \sum_x |x^r| p_X(x) \le M^r.$$

7.79 a) Consider a finite population $\{\omega_1, \ldots, \omega_N\}$ and a variable of interest y defined on it. Let y_i be the value of y on the ith member of the population (i.e. $y_i = y(\omega_i)$) and let $\mu = (\sum_{i=1}^N y_i)/N$ be the population mean. Then the population variance σ^2 is given by:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \mu)^2.$$

Suppose a member is selected at random from the population and X is equal to the corresponding value of y for that member. Then, whenever member ω_i is chosen, X takes value y_i (for each $i = 1, \ldots, N$). Therefore,

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X(\omega_i) - \mu)^2.$$

b) By (a) and since $P(\{\omega_i\}) = \frac{1}{N}$ for each $i = 1, \ldots, N$, then

$$\sigma^2 = \sum_{i=1}^{N} (X(\omega_i) - \mu)^2 P(\{\omega_i\}),$$

which, by Exercise 7.6, implies that $\sigma^2 = \mathcal{E}((X - \mu)^2)$. By Proposition 7.1, $\mathcal{E}(X) = \mu$, thus, $\sigma^2 = \mathcal{E}((X - \mathcal{E}(X))^2) = \mathrm{Var}(X)$.

7.80 Note that $\{(X - \mu_X)^2 \ge t^2\} = \{|X - \mu_X| \ge t\}$. Thus, by Markov's inequality,

$$P(|X - \mu_X| \ge t) = P((X - \mu_X)^2 \ge t^2) \le \frac{\mathcal{E}((X - \mu_X)^2)}{t^2} = \frac{\mathrm{Var}(X)}{t^2}.$$

7.81 First note that if $|X - \mu_X| > 0$, then there exists $n \in \mathcal{N}$ such that $|X - \mu_X| \ge 1/n$. On the other hand, $\{|X - \mu_X| \ge 1/n\} \subset \{|X - \mu_X| > 0\}$ for each $n \in \mathcal{N}$. Hence, $\{|X - \mu_X| > 0\} = \bigcup_{n=1}^{\infty} \{|X - \mu_X| \ge 1/n\}$.

Now, assume that a random variable X has zero variance. Then

$$P(|X-\mu_X|>0)=P\left(\bigcup_{n=1}^{\infty}\{|X-\mu_X|\geq\frac{1}{n}\}\right)\leq\sum_{n=1}^{\infty}P\left(|X-\mu_X|\geq\frac{1}{n}\right)$$
$$\leq\sum_{n=1}^{\infty}\frac{\mathrm{Var}(X)}{\frac{1}{n^2}}=\sum_{n=1}^{\infty}\frac{0}{\frac{1}{n^2}}=0.$$

Therefore, $P(X=\mu_X)=1$ and X is a constant random variable.

On the other hand, if a random variable X is constant, then, for some constant $c\in\mathcal{R}$, $P(X=c)=1$ and $P(X^2=c^2)=1$, implying that $\mathrm{Var}(X)=\mathcal{E}(X^2)-(\mathcal{E}(X))^2=c^2-c^2=0$.

7.82 Assume that properties (a) and (b) in Proposition 7.10 are satisfied. Then

$$\mathrm{Var}(a+bX)=\mathrm{Var}(bX)=b^2\mathrm{Var}(X)$$

holds for all constants $a,b\in\mathcal{R}$. Conversely, suppose equation (7.33) is satisfied. Then, upon taking $a=0$ and $b=c$ in (7.33), one arrives at property (a), whereas taking $a=c$ and $b=1$ reduces (7.33) to property (b). Thus, equation (7.33) is equivalent to properties (a) and (b) together.

Advanced Exercises

7.83 First, we recall that $\mathcal{E}(X)=\alpha$. Next note that

$$\mathcal{E}(X^2)=\sum_{x=0}^{\infty}x^2\frac{\alpha^x}{(1+\alpha)^{x+1}}=\sum_{x=1}^{\infty}x^2\left(\frac{\alpha}{1+\alpha}\right)^x\left(1-\frac{\alpha}{1+\alpha}\right)=\frac{\alpha}{1+\alpha}\mathcal{E}(Y^2),$$

where Y is a geometric random variable with parameter $p=1-\frac{\alpha}{1+\alpha}=\frac{1}{1+\alpha}$. Since

$$\mathcal{E}(Y^2)=\mathrm{Var}(Y)+(\mathcal{E}(Y))^2=\frac{1-p}{p^2}+\frac{1}{p^2}=\frac{2-p}{p^2}=2(1+\alpha)^2-(1+\alpha)=(1+\alpha)(1+2\alpha),$$

then

$$\mathcal{E}(X^2)=\alpha+2\alpha^2,\ \text{ and }\ \mathrm{Var}(X)=\alpha+2\alpha^2-\alpha^2=\alpha+\alpha^2.$$

7.84 By definition of PGF, for a nonnegative-integer valued random variable X with finite variance, we have that

$$P_X'(t)=\sum_{n=0}^{\infty}nt^{n-1}p_X(n)$$

so $P_X'(1)=\mathcal{E}(X)$ and

$$P_X''(t)=\sum_{n=0}^{\infty}(n^2-n)t^{n-2}p_X(n)$$

so $P_X''(1)=\mathcal{E}(X^2-X)=\mathcal{E}(X^2)-\mathcal{E}(X)$, implying that

$$\mathrm{Var}(X)=P_X''(1)+P_X'(1)-(P_X'(1))^2.$$

7.85 a) $np(1-p)$. From Exercise 7.53 (a), for a $\mathcal{B}(n,p)$ random variable X, we have that

$$P_X''(t) = np(n-1)(pt+(1-p))^{n-2}p, \text{ thus } P_X''(1) = np^2(n-1).$$

Thus, by Exercise 7.84, $\text{Var}(X) = np^2(n-1) + np - n^2p^2 = np(1-p)$.
b) λ. Using Exercise 7.53 (b), for a $\mathcal{P}(\lambda)$ random variable X,

$$P_X''(t) = \lambda^2 e^{\lambda(t-1)}, \text{ and } P_X''(1) = \lambda^2.$$

Thus, by Exercise 7.84, $\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.
c) $(1-p)/p^2$. Using Exercise 7.53 (c), for a $\mathcal{G}(p)$ random variable X,

$$P_X''(t) = \frac{2p(1-p)}{(1-t+pt)^3}, \text{ thus } P_X''(1) = \frac{2(1-p)}{p^2}.$$

Thus, by Exercise 7.84, $\text{Var}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.
d) $r(1-p)/p^2$. Using Exercise 7.53 (d), for a $\mathcal{NB}(r,p)$ random variable X,

$$P_X''(t) = \frac{rp^r t^{r-2}}{(1-t+tp)^{r+2}}\,(r-1+2t-2tp)\,, \text{ thus } P_X''(1) = \frac{r(r+1-2p)}{p^2}.$$

Thus, by Exercise 7.84, $\text{Var}(X) = \frac{r(r+1-2p)}{p^2} + \frac{r}{p} - \frac{r^2}{p^2} = \frac{r(1-p)}{p^2}$.
e) $(N^2-1)/12$. Using Exercise 7.53 (e), for discrete uniform on $\{1,\ldots,N\}$ random variable X,

$$P_X''(1) = \frac{N^2-1}{3},$$

which, by Exercise 7.84, implies that $\text{Var}(X) = \frac{N^2-1}{3} + \frac{N+1}{2} - \frac{(N+1)^2}{4} = \frac{N^2-1}{12}$.

7.4 Variance, Covariance, and Correlation

Basic Exercises

7.86 a) $\text{Var}(Y) = \text{Var}(Z) = 35/12$.
b) $\text{Var}(X) = \text{Var}(Y+Z) = \text{Var}(Y) + \text{Var}(Z) = 35/6 \approx 5.8333$, since Y and Z are independent.
c) Since the PMF of Y is much more easily obtained than the PMF of X, the method used in (b) is easier than the method used to obtain $\text{Var}(X)$ in Exercise 7.54.

7.87 a) See solution to Exercise 7.34 (a).
b) Note that $T = 10X + 5Y$, and let $g(x,y,z) = (10x+5y)^2$. Then $g(X,Y,Z) = T^2$ and using the FEF, we have

$$\mathcal{E}(T^2) = \sum\sum\sum_{(x,y,z)\in\mathcal{Z}_+^3:\, x+y+z=4} (10x+5y)^2 \frac{1}{36^4}\binom{4}{x,y,z}\pi^x(3\pi)^y(36-4\pi)^z = 118.202$$

c) From Exercise 7.34, we have $\mathcal{E}(T) = 8.72665$, so using the computing formula for variance, we have $\text{Var}(T) = 42.0479$.
d) From Exercise 7.55, it follows that the variance of the score from one shot is 10.512. Since

the shots are independent, we can add their variances, and thus $\mathrm{Var}(T) = 4 \times 10.512 = 42.0479$
e) One could also compute the PMF of T and compute the variance directly from definition.

7.88 a) Using Table 6.7, we first find that $\mathcal{E}(X+Y) = 6.12$. Then

$$\mathrm{Var}(X+Y) = \sum_{z=4}^{9} (z - 6.12)^2 p_{X+Y}(z) = 1.3456$$

b) The variance is the same.

7.89 The variance is equal to 1 when $N \geq 2$, and 0 when $N = 1$. Fix $N \geq 2$. Using the same notation as in Exercise 7.35, we have

$$\mathrm{Var}\left(\sum_{i=1}^{N} 1_{E_i}\right) = \sum_{i=1}^{N} \mathrm{Var}(1_{E_i}) + 2\sum\sum_{i<j} \mathrm{Cov}(1_{E_i}, 1_{E_j}).$$

Note that

$$\mathrm{Var}(1_{E_i}) = \frac{1}{N}\left(1 - \frac{1}{N}\right), \quad i = 1, \ldots, N;$$

also, for all $i, j \in \{1, \ldots, N\}$ such that $i \neq j$,

$$\mathrm{Cov}(1_{E_i}, 1_{E_j}) = \mathcal{E}(1_{E_i}1_{E_j}) - \mathcal{E}(1_{E_i})\mathcal{E}(1_{E_j}) = P(E_i \cap E_j) - P(E_i)P(E_j) = \frac{1}{N(N-1)} - \frac{1}{N^2},$$

and thus

$$\mathrm{Var}\left(\sum_{i=1}^{N} 1_{E_i}\right) = N \cdot \frac{1}{N}\left(1 - \frac{1}{N}\right) + 2\binom{N}{2}\left(\frac{1}{N(N-1)} - \frac{1}{N^2}\right) = 1.$$

(Clearly, when $N = 1$, the required variance is equal to 0).

7.90 $X_i = 1_{E_i}$ and $X_j = 1_{E_j}$ are positively correlated since $P(E_i|E_j) = \frac{1}{N-1} > \frac{1}{N} = P(E_i)$ for all $i \neq j$ and the result of Example 7.20 applies.

7.91 38.99; From Exercise 7.37, we have that $X_i \sim \mathcal{G}\left(\frac{7-i}{6}\right)$ for $i = 2, \ldots, 6$, and $X_1 = 1$. Since $X_1, \cdots, X_6$ are independent,

$$\mathrm{Var}\left(\sum_{i=1}^{6} X_i\right) = \sum_{i=1}^{6} \mathrm{Var}(X_i) = 0 + \frac{1-5/6}{(5/6)^2} + \frac{1-4/6}{(4/6)^2} + \frac{1-3/6}{(3/6)^2} + \frac{1-2/6}{(2/6)^2} + \frac{1-1/6}{(1/6)^2} = 38.99$$

7.92 a) Clearly, if one has more siblings, there is a greater chance for more sisters, thus, the correlation coefficient between X and Y must be positive.
b)

$$\rho(X,Y) = \frac{\mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{1.5 - 1.3 \times (29/40)}{\sqrt{0.91}\sqrt{0.549375}} \approx 0.788478$$

7.93 a) Since trials are independent, then X_{ik} and $X_{j\ell}$ are independent for $i \neq j$ (and all k and ℓ), implying that $\mathrm{Cov}(X_{ik}, X_{j\ell}) = 0$ for $i \neq j$ and all k and ℓ.
b) For all $k \neq \ell$, $X_{ik}X_{i\ell} = 0$, since E_k and E_ℓ are mutually exclusive. Thus for $k \neq \ell$,

$$\mathrm{Cov}(X_{ik}, X_{i\ell}) = \mathcal{E}(X_{ik}X_{i\ell}) - \mathcal{E}(X_{ik})\mathcal{E}(X_{i\ell}) = 0 - p_k p_\ell = -p_k p_\ell.$$

c) Since $X_k = \sum_{i=1}^n X_{ik}$, then for all $k \neq \ell$, by (a),(b) we have that

$$\mathrm{Cov}(X_k, X_\ell) = \mathrm{Cov}\left(\sum_{i=1}^n X_{ik}, \sum_{j=1}^n X_{j\ell}\right) = \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_{ik}, X_{j\ell}) = \sum_{i=1}^n \mathrm{Cov}(X_{ik}, X_{i\ell}) = -np_k p_\ell.$$

d) For $k \neq \ell$, since there are only n experiments, the more times E_k occurs, the less times E_ℓ can possibly occur. Thus, X_k and X_ℓ must be negatively correlated.
e) Since $\mathrm{Var}(X_j) = np_j(1-p_j)$, then for all $k \neq \ell$,

$$\rho(X_k, X_\ell) = \frac{\mathrm{Cov}(X_k, X_\ell)}{\sqrt{\mathrm{Var}(X_k)}\sqrt{\mathrm{Var}(X_\ell)}} = \frac{-np_k p_\ell}{\sqrt{np_k(1-p_k)np_\ell(1-p_\ell)}} = -\sqrt{\frac{p_k}{1-p_k} \cdot \frac{p_\ell}{1-p_\ell}};$$

and, clearly, $\rho(X_k, X_\ell) = 1$ when $k = \ell$.

7.94 Since $T = 10X + 5Y$,

$$\begin{aligned}\mathrm{Var}(T) &= \mathrm{Var}(10X + 5Y) = 100\mathrm{Var}(X) + 25\mathrm{Var}(Y) + 100\mathrm{Cov}(X, Y) \\ &= 100\left(4 \cdot \frac{\pi}{36}(1 - \frac{\pi}{36})\right) + 25\left(4 \cdot \frac{3\pi}{36}(1 - \frac{3\pi}{36})\right) - 100 \cdot \frac{4\pi(3\pi)}{36^2} \approx 42.0479\end{aligned}$$

7.95 Let C represent the cost of repairing the defective items in the sample. Also, let X be the number of items with exactly one defect in the sample, and Y be the number of items with at least two defects in the sample. Then $C = X + 3Y$, and, by equation (*) in Exercise 7.93,

$$\mathrm{Var}(C) = \mathrm{Var}(X+3Y) = \mathrm{Var}(X)+9\mathrm{Var}(Y)+6\mathrm{Cov}(X,Y) = np_1(1-p_1)+9np_2(1-p_2)-6np_1p_2.$$

7.96 Using the PMFs given in Example 6.7, the mean of the total number of accidents is $\frac{22}{7}$ and the standard deviation is $\sqrt{\frac{172}{147}} \approx 1.0817$

7.97 Using the joint and marginal PMF given in Exercise 6.6, one obtains that

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{\frac{441}{36} - \frac{91}{36} \cdot \frac{161}{36}}{\sqrt{\frac{301}{36} - \frac{91^2}{36^2}} \cdot \sqrt{\frac{791}{36} - \frac{161^2}{36^2}}} = \frac{35}{73}.$$

Clearly, X and Y are positively correlated, although mildly so. The latter is natural, since if X is large, then Y must be large as well. However, it is not very close to 1 since if X is small, Y can vary greatly.

7.98 $\mathrm{Cov}(X, Y) = 8.8$ and $\rho(X, Y) \approx 0.999175$, thus, X and Y have a strong positive linear correlation. Indeed, let A represent the size of a surgical claim and B represent the size of a hospital claim. Then $X = A + B$ and $Y = A + 1.2B$. Now based on the information provided, $\mathcal{E}(A) = 5$, $\mathcal{E}(A^2) = 27.4$, $\mathcal{E}(B) = 7$, $\mathcal{E}(B^2) = 51.4$ and $\mathrm{Var}(A + B) = 8$. Then $\mathrm{Var}(A) = \mathrm{Var}(B) = 2.4$ and

$$8 = \mathrm{Var}(A + B) = \mathrm{Var}(A) + \mathrm{Var}(B) + 2\mathrm{Cov}(A, B) = 2.4 + 2.4 + 2\mathrm{Cov}(A, B),$$

which implies that $\mathrm{Cov}(A, B) = 1.6$, which in turn implies that $\mathcal{E}(AB) = 36.6$ Thus,

$$\mathrm{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(A^2 + 2.2AB + 1.2B^2) - (12)(13.4)$$

$$= \mathcal{E}(A^2) + 2.2\mathcal{E}(AB) + 1.2\mathcal{E}(B^2) - 160.8 = 8.8$$

Moreover, $\text{Var}(X) = 8$ and $\text{Var}(Y) = 9.696$. Thus $\rho = 0.999175$. Clearly, there is a strong positive linear correlation between X and Y.

7.99 a) Note that $x^2 + y^2 - 2|x||y| = (|x| - |y|)^2 \geq 0$ for all $x, y \in \mathcal{R}$, thus, $\frac{1}{2}(x^2 + y^2) \geq |xy|$. Therefore,

$$\mathcal{E}|XY| \leq \frac{1}{2}\mathcal{E}(X^2 + Y^2) < \infty,$$

hence, XY has finite mean.
b) First we note

$$(x+y)^2 = x^2 + 2xy + y^2 \leq 2(x^2 + y^2)$$

where the inequality follows from the inequality shown in (a). Since X and Y have finite variances, then both X and Y have finite second moments, which by the just proved inequality implies that $X + Y$ has a finite second moment. Thus, variance of $X + Y$ is well-defined and, by the computing formula for the variance, we have that

$$\begin{aligned} 0 \leq \text{Var}(X+Y) &= \mathcal{E}(X+Y)^2 - (\mathcal{E}(X) + \mathcal{E}(Y))^2 \\ &\leq 2\left(\mathcal{E}\left(X^2\right) + \mathcal{E}\left(Y^2\right)\right) - (\mathcal{E}(X) + \mathcal{E}(Y))^2 < \infty. \end{aligned}$$

7.100 By linearity of expectation we have that

$$\text{Cov}(X,Y) = \mathcal{E}((X - \mu_X)(Y - \mu_Y)) = \mathcal{E}(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y)$$

$$= \mathcal{E}(XY) - \mu_X \mathcal{E}(Y) - \mu_Y \mathcal{E}(X) + \mu_X \mu_Y = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y).$$

7.101 Let X denote the value of the variable for a randomly selected member of the population and let $X_1, \ldots, X_n$ denote, respectively, the values of the variable for the n members of the random sample. As the sampling is with replacement, the random variables $X_1, \ldots, X_n$ are independent and constitute a random sample of size n from the distribution of X. Therefore, by Example 7.16 (b), $\text{Var}(\bar{X}_n) = \text{Var}(X)/n = \sigma^2/n$, where the last step follows from Proposition 7.9.

7.102 First we note that $\mathcal{E}(X^*) = 0 = \mathcal{E}(Y^*)$, thus

$$\text{Cov}(X^*, Y^*) = \mathcal{E}\left(\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right) = \frac{1}{\sigma_X \sigma_Y}\mathcal{E}\left(XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y\right)$$

$$= \frac{1}{\sigma_X \sigma_Y}\left(\mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y)\right) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

The quantity is unitless since the units of the numerator and denominator are both the product of the units of X and Y.

7.103 a) Since $\mathcal{E}(X) = 0$, we have that

$$\text{Cov}(X,Y) = \mathcal{E}(XY) = \mathcal{E}(X^3) = (-1)^3 \times \frac{1}{3} + 0^3 \times \frac{1}{3} + 1^3 \times \frac{1}{3} = 0,$$

thus, $\rho(X,Y) = 0$.
b) The correlation coefficient is a measure of linear association, whereas here the functional

relationship is quadratic.

7.104 a) Upon taking derivatives with respect to a and b, we arrive at the following two equations for a and b:

$$\mathcal{E}(Y - a - bX) = 0,$$
$$\mathcal{E}(XY - aX - bX^2) = 0.$$

From the first equation it follows that

$$a = \mathcal{E}(Y) - b\mathcal{E}(X).$$

Let us plug-in the latter expression into the second equation. Then

$$\mathcal{E}(XY - (\mathcal{E}(Y) - b\,\mathcal{E}(X))X - bX^2) = 0,$$

hence

$$\mathcal{E}(XY) - \mathcal{E}(Y)\mathcal{E}(X) - b(\mathcal{E}(X^2) - (\mathcal{E}(X))^2) = 0,$$

and thus

$$b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \frac{\sigma_Y}{\sigma_X}\rho(X,Y).$$

b) For values of a, b from part (a), we have that

$$\mathcal{E}(Y - a - bX)^2 = \mathcal{E}\Big((Y - \mathcal{E}(Y)) - b(X - \mathcal{E}(X))\Big)^2$$

$$= \mathrm{Var}(Y) - 2b\mathrm{Cov}(X,Y) + b^2\mathrm{Var}(X) = \mathrm{Var}(Y) - \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)}.$$

c) From (b), it follows that $\mathrm{Var}(Y) - \dfrac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)} \geq 0$, therefore,

$$\frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} \leq 1, \text{ thus } (\rho(X,Y))^2 \leq 1,$$

which implies part (a) of Proposition 7.16.

d) If $\rho = 1$ or -1, then $(\mathrm{Cov}(X,Y))^2 = \mathrm{Var}(X)\mathrm{Var}(Y)$. Pick a and b as in (a), then we have $\mathcal{E}(Y - (a + bX))^2 = 0$, which implies that $Y = a + bX$ with probability 1. Note also that since $b = \frac{\sigma_Y}{\sigma_X}\rho(X,Y)$, then b has the same sign as $\rho(X,Y)$.

Conversely, suppose $Y = \alpha + \beta X$ for some real constants α and $\beta > 0$. Then the mean square error, $\mathcal{E}\left([Y - (a + bX)]^2\right)$, is zero (and therefore minimized) when $a = \alpha$ and $b = \beta$, which, by (a), implies that $\beta = \frac{\sigma_Y}{\sigma_X}\rho(X,Y)$. Since $\sigma_Y = |\beta|\sigma_X$, it follows that

$$\rho(X,Y) = \frac{\beta}{|\beta|} = \begin{cases} 1, & \text{if } \beta > 0, \\ -1, & \text{if } \beta < 0. \end{cases}$$

e) ρ measures how closely a linear function of one variable fits another variable (in the sense of the mean square error).

Theory Exercises

7.105 a) $\mathrm{Cov}(X,X) = \mathcal{E}((X-\mu_X)(X-\mu_X)) = \mathcal{E}((X-\mu_X)^2) = \mathrm{Var}(X)$.
b) $\mathrm{Cov}(X,Y) = \mathcal{E}\left((X-\mu_X)(Y-\mu_Y)\right) = \mathcal{E}\left((Y-\mu_Y)(X-\mu_X)\right) = \mathrm{Cov}(Y,X)$.
c) $\mathrm{Cov}(cX,Y) = \mathcal{E}((cX-\mu_{cX})(Y-\mu_Y)) = \mathcal{E}((cX-c\mu_X)(Y-\mu_Y))$
$= \mathcal{E}(c(X-\mu_X)(Y-\mu_Y)) = c\,\mathrm{Cov}(X,Y)$.
d) Note that

$$\mathrm{Cov}(X,Y+Z) = \mathcal{E}\left((X-\mu_X)(Y+Z-\mu_{Y+Z})\right) = \mathcal{E}\left((X-\mu_X)((Y-\mu_Y)+(Z-\mu_Z))\right)$$

$$= \mathcal{E}(X-\mu_X)(Y-\mu_Y) + \mathcal{E}(X-\mu_X)(Z-\mu_Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z).$$

e) $\mathrm{Cov}(a+bX,Y) = \mathcal{E}\left(\left(a+bX-\mu_{(a+bX)}\right)(Y-\mu_Y)\right) = \mathcal{E}((a+bX-a-b\mu_X)(Y-\mu_Y))$
$= b\mathrm{Cov}(X,Y)$.

7.106 By induction and Proposition 7.12 (parts (b),(c),(d)), we have that for arbitrary random variables $X, Y_1, \ldots, Y_n$ defined on the same sample space and having finite variances,

$$\mathrm{Cov}\left(X, \sum_{k=1}^{n} b_k Y_k\right) = \sum_{k=1}^{n} b_k \mathrm{Cov}(X, Y_k).$$

Thus,

$$\mathrm{Cov}\left(\sum_{j=1}^{m} a_j X_j, \sum_{k=1}^{n} b_k Y_k\right) = \sum_{k=1}^{n} b_k \mathrm{Cov}\left(\sum_{j=1}^{m} a_j X_j, Y_k\right) = \sum_{k=1}^{n} b_k \mathrm{Cov}\left(Y_k, \sum_{j=1}^{m} a_j X_j\right)$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{n} a_j b_k \mathrm{Cov}(Y_k, X_j) = \sum_{j=1}^{m}\sum_{k=1}^{n} a_j b_k \mathrm{Cov}(X_j, Y_k).$$

7.107 Clearly,

$$\mathrm{Cov}(X,Y) = \mathcal{E}((X-\mu_X)(Y-\mu_Y)) = \mathcal{E}(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y)$$

$$= \mathcal{E}(XY) - \mu_X\mathcal{E}(Y) - \mu_Y\mathcal{E}(X) + \mu_X\mu_Y = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y).$$

7.108 By Proposition 7.12, we have that

$$\mathrm{Var}\left(\sum_{k=1}^{m} X_k\right) = \mathrm{Cov}\left(\sum_{k=1}^{m} X_k, \sum_{\ell=1}^{m} X_\ell\right) = \sum_{k=1}^{m}\sum_{\ell=1}^{m} \mathrm{Cov}(X_k, X_\ell)$$

$$= \sum_{k=1}^{m} \mathrm{Cov}(X_k, X_k) + \sum_{k=1}^{m} \sum_{\substack{\ell=1 \\ \ell\neq k}}^{m} \mathrm{Cov}(X_k, X_\ell) = \sum_{k=1}^{m} \mathrm{Var}(X_k) + 2\sum_{1\le k<\ell\le m} \mathrm{Cov}(X_k, X_\ell).$$

Advanced Exercises

7.109 For all $1 \le j \le n$, we have that

$$\mathcal{E}((X_j - \bar{X}_n)^2) = \mathcal{E}(X_j^2) - 2\mathcal{E}(X_j\bar{X}_n) + \mathcal{E}(\bar{X}_n^2),$$

where $\mathcal{E}(X_j^2) = \mathcal{E}(X^2)$ and, by independence of X_i and X_j for $i \neq j$,

$$\mathcal{E}(X_j \bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}(X_j X_i) = \frac{1}{n}\left(\sum_{i \neq j} \underbrace{\mathcal{E}(X_i X_j)}_{=\mathcal{E}(X_i)\mathcal{E}(X_j)} + \mathcal{E}(X_j^2)\right) = \frac{1}{n}\left(\mathcal{E}(X^2) + (n-1)\left(\mathcal{E}(X)\right)^2\right)$$

and

$$\mathcal{E}(\bar{X}_n^2) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} \mathcal{E}(X_i X_j) = \frac{1}{n}\left(\mathcal{E}(X^2) + (n-1)\left(\mathcal{E}(X)\right)^2\right).$$

Thus, $\mathcal{E}((X_j - \bar{X}_n)^2) = \frac{n-1}{n}\sigma_X^2$ for all $1 \leq j \leq n$ and, therefore,

$$\mathcal{E}(S_n^2) = \frac{1}{n-1}\sum_{j=1}^{n} \mathcal{E}((X_j - \bar{X}_n)^2) = \frac{1}{n-1}\sum_{j=1}^{n} \frac{n-1}{n}\sigma_X^2 = \sigma_X^2.$$

7.110 a) First note that $S_n^2 = T_n^2$ since

$$\begin{aligned} T_n^2 &= \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_j - \bar{Y}_n\right)^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \mu_X - \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)\right)^2 \\ &= \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \mu_X - \bar{X}_n + \mu_X\right)^2 = S_n^2. \end{aligned}$$

Next, since $\bar{X}_n = \bar{Y}_n + \mu_X$, we have that

$$\begin{aligned} \operatorname{Cov}(\bar{X}_n, S_n^2) &= \mathcal{E}(\bar{X}_n S_n^2) - \mathcal{E}(\bar{X}_n)\mathcal{E}(S_n^2) = \mathcal{E}((\bar{Y}_n + \mu_X)T_n^2) - \mathcal{E}(\bar{Y}_n + \mu_X)\mathcal{E}(T_n^2) \\ &= \mathcal{E}(\bar{Y}_n T_n^2) - \mathcal{E}(\bar{Y}_n)\mathcal{E}(T_n^2) = \operatorname{Cov}(\bar{Y}_n, T_n^2). \end{aligned}$$

b) Note that $\mathcal{E}(Y_i) = 0$ and $\mathcal{E}(Y_i^3) = 0$ for all i, thus, $\mathcal{E}(Y_i Y_j Y_k) = 0$ for all i, j, k. Then

$$\operatorname{Cov}(\bar{Y}_n, T_n^2) = \mathcal{E}(\bar{Y}_n T_n^2) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{E}\left(Y_i(Y_j^2 - 2Y_j\bar{Y}_n + \bar{Y}_n^2)\right)$$

$$= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathcal{E}(Y_i Y_j^2) - 2\mathcal{E}(Y_i Y_j \bar{Y}_n) + \mathcal{E}(Y_i \bar{Y}_n^2)\right] = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}[0 - 2\cdot 0 + 0] = 0.$$

c) By (a),(b), $\operatorname{Cov}(\bar{X}_n, S_n^2) = \operatorname{Cov}(\bar{Y}_n, T_n^2) = 0$, thus, $\rho(\bar{X}_n, S_n^2) = 0$, implying that $\bar{X}_n$ and S_n^2 are uncorrelated.

7.111 a) Let n_k denote the number of times that the kth member of the population occurs in the (randomly selected) sample. Also let $x(\omega_k)$ be the value of the variable of interest for the kth member of the population. Then $\sum_{k=1}^{N} x(\omega_k)n_k$ represents the sum of values of the variable of interest in the sample. On the other hand, the sum of values of the variable of interest in the sample must be equal to $n\bar{X}_n$. Therefore,

$$n\bar{X}_n = \sum_{k=1}^{N} x(\omega_k)n_k, \text{ implying that } \bar{X}_n = \frac{1}{n}\sum_{k=1}^{N} x(\omega_k)n_k = \frac{1}{n}\sum_{k=1}^{N} x(\omega_k)I_k,$$

where the latter equality holds since the sample is taken without replacement.
b) $\mathcal{E}(I_k) = P(k\text{th member of the population is in the sample}) = \frac{n}{N}$. Also, since $I_k^2 = I_k$, then

$$\mathrm{Var}(I_k) = \mathcal{E}(I_k^2) - (\mathcal{E}(I_k))^2 = \mathcal{E}(I_k) - (\mathcal{E}(I_k))^2 = \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n(N-n)}{N^2}.$$

Moreover, for all $i \neq j$,

$$\mathrm{Cov}(I_i, I_j) = \mathcal{E}(I_i I_j) - (n/N)^2$$

$$= P(\text{both } i\text{th and } j\text{th members of the population are in the sample}) - (n/N)^2$$

$$= \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} = -\frac{n(N-n)}{N^2(N-1)}.$$

c) From (a),(b), we have that

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{N} x(\omega_i) I_i\right) = \frac{1}{n^2}\sum_{i=1}^{N} x^2(\omega_i)\mathrm{Var}(I_i) + \frac{2}{n^2}\sum_{1\le i<j\le N} x(\omega_i)x(\omega_j)\mathrm{Cov}(I_i, I_j)$$

$$= \frac{N-n}{nN^2}\sum_{i=1}^{N} x^2(\omega_i) - \frac{2(N-n)}{nN^2(N-1)}\sum_{1\le i<j\le N} x(\omega_i)x(\omega_j).$$

d) Let $\bar{x}_N = N^{-1}\sum_{i=1}^{N} x(\omega_i)$, then it is easy to see that

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x(\omega_i) - \bar{x}_N)^2 = \frac{N-1}{N^2}\sum_{i=1}^{N} x^2(\omega_i) - \frac{2}{N^2}\sum_{1\le i<j\le N} x(\omega_i)x(\omega_j),$$

which implies that

$$\mathrm{Var}(\bar{X}_n) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}.$$

e) Sampling without replacement yields a smaller variance of the sample mean than sampling with replacement. Note also that when N is large relative to n, then

$$\frac{N-n}{N-1} = 1 - \frac{n-1}{N-1} \approx 1, \text{ implying that } \mathrm{Var}(\bar{X}_n) \approx \frac{\sigma^2}{n},$$

thus, when the population size is large relative to the sample size, the sample mean, in the case of sampling without replacement, has a variance which is approximately equal to the variance of the sample mean in the case of sampling with replacement. This is not surprising since when a relatively small sample is taken with replacement from a large population, the probability of selecting any given member of the population more than once is very small, which amounts to essentially sampling without replacement.

7.5 Conditional Expectation

Basic Exercises

7.112 a) $\mathcal{E}(Y|X = x) = \mathcal{E}(Y)$, since if Y and X are independent then the mean of Y should not depend on the value that X takes.
b) Since Y is independent of X, $p_{Y|X}(y|x) = p_Y(y)$ for all x, y, implying that

$$\mathcal{E}(Y \mid X = x) = \sum_y y p_{Y|X}(y \mid x) = \sum_y y p_Y(y) = \mathcal{E}(Y).$$

7.113 Note that the joint PMF of X, Y has the form $p_{X,Y}(x, y) = \left(p(1-p)^{x-1}\right)\left(p(1-p)^{y-1}\right)$, i.e. it can be factored into a function of x alone and a function of y alone, thus, by Exercise 6.60, it follows that X and Y are independent geometric random variables with parameter p each. Thus, by Exercise 7.112, $\mathcal{E}(Y|X = x) = \mathcal{E}(Y) = \frac{1}{p}$.

7.114 By Proposition 7.18, $\mathcal{E}(Y) = \mathcal{E}(\mathcal{E}(Y \mid X))$, where the right-hand side represents expectation of the conditional expectation, thus, the formula is often referred to as the double expectation formula, since the right-hand side involves double (iterated) expectation.

7.115 a) The following table displays the conditional expectation of the number of sisters, given the number of siblings.

Siblings x	Conditional expectation $\mathcal{E}(Y\|X = x)$
0	0
1	$9/17 \approx 0.529$
2	$12/11 \approx 1.091$
3	$5/3 \approx 1.667$
4	3

The mean number of sisters for students with 0 siblings is equal to 0. The mean number of sisters for students with 1 sibling is equal to 9/17. The mean number of sisters for students with 2 siblings is equal to 12/11. The mean number of sisters for students with 3 siblings is equal to 5/3. The mean number of sisters for students with 4 siblings is equal to 3.

b) The following table displays the conditional variance of the number of sisters, given the number of siblings.

Siblings x	Conditional variance $\mathrm{Var}(Y\|X = x)$
0	0
1	$72/289 \approx 0.249$
2	$32/121 \approx 0.264$
3	$2/9 \approx 0.222$
4	0

7.116 a) By the law of total expectation and solution to Exercise 7.115(a),

$$\mathcal{E}(Y) = \sum_x \mathcal{E}(Y|X=x)p_X(x) = \frac{9}{17} \times \frac{17}{40} + \frac{12}{11} \times \frac{11}{40} + \frac{5}{3} \times \frac{3}{40} + 3 \times \frac{1}{40} = \frac{29}{40} = 0.725$$

b) Using the definition of expected value we have

$$\mathcal{E}(Y) = \sum_y y p_Y(y) = 0 \times \frac{17}{40} + 1 \times \frac{18}{40} + 2 \times \frac{4}{40} + 3 \times \frac{1}{40} = \frac{29}{40} = 0.725,$$

which is the same as in (a).
c) We have

$$\mathrm{Var}(Y) = \mathcal{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathcal{E}(Y \mid X))$$

$$= \sum_x \mathrm{Var}(Y|X=x)p_X(x) + \sum_x \left(\mathcal{E}(Y \mid X=x)\right)^2 p_X(x) - \left(\sum_x \mathcal{E}(Y|X=x)p_X(x)\right)^2$$

$$\approx 0.1953 + 0.8797 - 0.5256 = 0.5494$$

d) Using the definition of variance,

$$\mathrm{Var}(Y) = \mathcal{E}(Y^2) - (\mathcal{E}(Y))^2 = \frac{43}{40} - \frac{29^2}{40^2} = \frac{879}{1600} \approx 0.5494,$$

which is the same answer as in (c).

7.117 a) Using solution to Exercise 6.46(a), for $x = 0, 1, \ldots, 9$, we have that that

$$\mathcal{E}(Y \mid X=x) = \sum_{y \in \{0,1,\ldots,9\} \setminus \{x\}} y \times \frac{1}{9} = \frac{1}{9}\left((\sum_{y=0}^{9} y) - x \right) = \frac{45 - x}{9}.$$

b) Using solution to (a) and Exercise 6.46(a), for $x = 0, 1, \ldots, 9$, we have that

$$\mathrm{Var}(Y \mid X=x) = \sum_{y \in \{0,\ldots,9\} \setminus \{x\}} \frac{y^2}{9} - \left(\frac{45-x}{9}\right)^2 = \frac{540 + 90x - 10x^2}{81}.$$

7.118 a) By the law of total expectation and Exercise 7.117, we have that

$$\mathcal{E}(Y) = \sum_x \mathcal{E}(Y|X=x)p_X(x) = \sum_{x=0}^{9} \frac{45-x}{9} \cdot \frac{1}{10} = 4.5$$

b) Using the definition of expected value and the result of Exercise 6.3(d), we have that

$$\mathcal{E}(Y) = \sum_y y p_Y(y) = \sum_{y=0}^{9} \frac{y}{10} = 4.5,$$

which is the same answer as in (a).
c) We have

$$\operatorname{Var}(Y) = \mathcal{E}(\operatorname{Var}(Y \mid X)) + \operatorname{Var}(\mathcal{E}(Y \mid X))$$

$$= \sum_x \operatorname{Var}(Y \mid X = x) p_X(x) + \sum_x \left(\mathcal{E}(Y \mid X = x)\right)^2 p_X(x) - \left(\sum_x \mathcal{E}(Y \mid X = x) p_X(x)\right)^2$$

$$= \sum_{x=0}^{9} \frac{540 + 90x - 10x^2}{81} \cdot \frac{1}{10} + \sum_{x=0}^{9} \left(\frac{45 - x}{9}\right)^2 \cdot \frac{1}{10} - 4.5^2 = 8.25$$

d) Using the definition of variance,

$$\operatorname{Var}(Y) = \mathcal{E}(Y^2) - (\mathcal{E}(Y))^2 = \sum_{y=0}^{9} \frac{y^2}{10} - 4.5^2 = 8.25,$$

which is the same answer as in (c).

7.119 ($1457)0.458 + ($2234)0.326 + ($2516)0.216 = $1939.046

7.120 a) $14809.2; Let N equal the hourly number of customers entering the bank to make a deposit. Then N has $\mathcal{P}(\lambda = 25.8)$ distribution. Let X_i be the amount deposited by the ith customer (during a given 1-hour period). Then $\mathcal{E}(X_i) = \$574$ and $\sigma_{X_i} = \$3167$, and $X_1, X_2, \ldots$ are independent. The total amount deposited during that 1-hour period is equal to $X_1 + \cdots + X_N$, and the required expectation is given by:

$$\mathcal{E}(X_1 + \cdots + X_N) = \sum_{n=0}^{\infty} \mathcal{E}(X_1 + \cdots + X_n \mid N = n) p_N(n)$$

$$= \sum_{n=0}^{\infty} (n \cdot \$574) \frac{e^{-25.8} 25.8^n}{n!} = 25.8(\$574) = \$14809.2$$

b) $16348.44; By the law of total variance, we have that

$$\operatorname{Var}(X_1 + \cdots + X_N) = \mathcal{E}(\operatorname{Var}(X_1 + \cdots + X_N \mid N)) + \operatorname{Var}(\mathcal{E}(X_1 + \cdots + X_N \mid N)),$$

where

$$\operatorname{Var}(X_1 + \cdots + X_N \mid N) = N \operatorname{Var}(X_i) = 3167^2 N$$

and

$$\mathcal{E}(X_1 + \cdots + X_N \mid N) = N \mathcal{E}(X_i) = 574N.$$

Therefore,

$$\operatorname{Var}(X_1 + \cdots + X_N) = 3167^2 \mathcal{E}(N) + 574^2 \operatorname{Var}(N) = (3167^2 + 574^2) 25.8 = 267271617,$$

implying that $\sigma_{X_1 + \cdots + X_N} = 16348.44$

7.121 $6.28; Let Y be the amount that the company makes on a randomly selected insurance

policy, a be the size of the net premium and X is the amount of loss.
Then $Y = a - \max\{0, X - \$200\}$, where for $x \in \{1, 2 \cdots, 5\}$,

$$P(X = \$100x) = P(X = \$100x \,|\, X > 0)P(X > 0) = \frac{c}{x}(0.1)$$

and $p_X(0) = 0.9$ (and $p_X(x) = 0$ for all $x \notin \{0, 100, 200, 300, 400, 500\}$). Thus,

$$\mathcal{E}(Y) = a - \mathcal{E}\left(\max\{0, X - 200\}\right) = a - \left(100 \times \frac{0.1c}{3} + 200 \times \frac{0.1c}{4} + 300 \times \frac{0.1c}{5}\right) = a - \frac{43c}{3}.$$

Note that we must have that $\sum_{x=1}^{5} \frac{c}{x} = 1$, which implies that $c = \frac{60}{137}$. In order for the insurance company to break even, one must have that $\mathcal{E}(Y) = 0$, implying that

$$a = \frac{43c}{3} = \frac{43}{3} \cdot \frac{60}{137} = \frac{860}{137} \approx 6.28$$

7.122 The mean and standard deviation of the claim amounts for all policyholders of the company are equal to \$2375 and \$701.338, respectively. Indeed, let Y denote the claim amount of a randomly selected policyholder, and let X be the indicator of the event that the policyholder is a good driver. Then

$$\mathcal{E}(Y) = \mathcal{E}(Y \,|\, X = 1)P(X = 1) + \mathcal{E}(Y \,|\, X = 0)P(X = 0) = \$2000(0.75) + \$3500(0.25) = \$2375.$$

We also have that

$$\mathrm{Var}(Y) = \mathcal{E}(\mathrm{Var}(Y \,|\, X)) + \mathrm{Var}(\mathcal{E}(Y \,|\, X)),$$

where $\mathrm{Var}(Y \,|\, X) = 200^2 X + 400^2(1 - X) = 160000 - 120000X$ and
$\mathcal{E}(Y \,|\, X) = 2000X + 3500(1 - X) = 3500 - 1500X$. Thus,

$$\mathrm{Var}(Y) = 160000 - 120000\mathcal{E}(X) + \mathrm{Var}(3500 - 1500X) = 70000 + 1500^2(0.75 - 0.75^2) = 491875,$$

since $\mathrm{Var}(X) = 0.75 - 0.75^2$. Therefore, $\sigma_Y \approx 701.338$

7.123 a) Let Z be the indicator random variable of the event that the first trial is a success. (X equals the number of trials until the 1st success). By the law of total expectation, it follows that

$$\mathcal{E}(X^2) = \mathcal{E}(X^2 | Z = 0)p_Z(0) + \mathcal{E}(X^2 | Z = 1)p_Z(1).$$

Proceeding as in Example 7.25, since the trials are independent, we have that $X_{|Z=0}$ has the same distribution as $X + 1$, and thus $\mathcal{E}(X^2 | Z = 0) = \mathcal{E}((X + 1)^2)$. Then

$$\mathcal{E}(X^2) = \mathcal{E}(X^2 + 2X + 1)(1 - p) + p = \mathcal{E}(X^2)(1 - p) + \frac{2(1 - p)}{p} + 1 - p + p = \mathcal{E}(X^2)(1 - p) + \frac{2 - p}{p},$$

implying that

$$\mathcal{E}(X^2) = \frac{2 - p}{p^2}.$$

Therefore,

$$\mathrm{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = \frac{2 - p}{p^2} - \frac{1}{p^2} = \frac{1 - p}{p^2}.$$

b) By the law of total variance,

$$\mathrm{Var}(X) = \mathcal{E}(\mathrm{Var}(X \,|\, Z)) + \mathrm{Var}(\mathcal{E}(X \,|\, Z)),$$

where it is easy to see that

$$\mathrm{Var}(X \,|\, Z) = (1 - Z)\mathrm{Var}(X), \text{ which implies that } \mathcal{E}(\mathrm{Var}(X \mid Z)) = \mathrm{Var}(X)(1 - p);$$

Note also that

$$\mathcal{E}(X \,|\, Z) = (1 - Z)\mathcal{E}(X + 1) + Z = (1 - Z)\left(\frac{1}{p} + 1\right) + Z = 1 - \frac{Z}{p} + \frac{1}{p},$$

thus, $\mathrm{Var}(\mathcal{E}(X \,|\, Z)) = \frac{1}{p^2}\mathrm{Var}(Z) = \frac{1}{p^2}(p - p^2) = \frac{1-p}{p}$. Therefore, we have that

$$\mathrm{Var}(X) = (1 - p)\mathrm{Var}(X) + \frac{1 - p}{p},$$

implying that $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

7.124 Assume $X_r \sim \mathcal{NB}(r, p)$. We already know that $\mathcal{E}(X_1) = \frac{1}{p}$. Assume that $\mathcal{E}(X_{k-1}) = \frac{k-1}{p}$ and, let Z be the indicator random variable of the event that the first trial is a success. Then we have

$$\mathcal{E}(X_k) = \mathcal{E}(X_k \,|\, Z = 0)p_Z(0) + \mathcal{E}(X_k \,|\, Z = 1)p_Z(1).$$

Since the trials are independent, $\mathcal{E}(X_k \,|\, Z = 0) = \mathcal{E}(1 + X_k)$ and $\mathcal{E}(X_k \,|\, Z = 1) = \mathcal{E}(1 + X_{k-1})$. Therefore,

$$\mathcal{E}(X_k) = (1 + \mathcal{E}(X_k))(1 - p) + (1 + \mathcal{E}(X_{k-1}))p = (1 - p)\mathcal{E}(X_k) + 1 - p + p + \frac{k - 1}{p} \cdot p,$$

implying that $\mathcal{E}(X_k) = k/p$, thus, the required equation follows by induction.

7.125 a) Assume $X_r \sim \mathcal{NB}(r, p)$. We already know that $\mathcal{E}(X_r) = \frac{r}{p}$ for all $r \in \mathcal{N}$ and $\mathrm{Var}(X_1) = \frac{1-p}{p^2}$. Assume that $\mathrm{Var}(X_{k-1}) = \frac{(k-1)(1-p)}{p^2}$, then

$$\mathcal{E}(X_{k-1}^2) = \frac{(k - 1)(1 - p)}{p^2} + \frac{(k - 1)^2}{p^2}.$$

Let Z be the indicator random variable of the event that the first trial is a success. Then first note that $(X_k \,|\, Z = 0)$ has the same distribution as $1 + X_k$, while $(X_k \,|\, Z = 1)$ has the same distribution as $1 + X_{k-1}$. Then

$$\begin{aligned}
\mathcal{E}(X_k^2) &= \mathcal{E}(X_k^2 \,|\, Z = 0)p_Z(0) + \mathcal{E}(X_k^2 \,|\, Z = 1)p_Z(1) \\
&= \mathcal{E}((1 + X_k)^2)(1 - p) + \mathcal{E}((1 + X_{k-1})^2)p \\
&= (1 - p)\left(1 + 2\mathcal{E}(X_k) + \mathcal{E}(X_k^2)\right) + p\left(1 + 2\mathcal{E}(X_{k-1}) + \mathcal{E}(X_{k-1}^2)\right) \\
&= (1 - p)(1 + \frac{2k}{p} + \mathcal{E}(X_k^2)) + p\left(1 + \frac{2(k - 1)}{p} + \frac{(k - 1)(1 - p)}{p^2} + \frac{(k - 1)^2}{p^2}\right) \\
&= (1 - p)\mathcal{E}(X_k^2) + \frac{k^2}{p} + \frac{k}{p}(1 - p).
\end{aligned}$$

Hence $\mathcal{E}(X_k^2) = \frac{k^2}{p^2} + \frac{k(1-p)}{p^2}$. Therefore,

$$\mathrm{Var}(X_k) = \mathcal{E}(X_k^2) - (\mathcal{E}(X_k))^2 = \frac{k^2}{p^2} + \frac{k(1-p)}{p^2} - \frac{k^2}{p^2} = \frac{k(1-p)}{p^2}.$$

Hence the required result is true by induction.
b) By the law of total variance,

$$\mathrm{Var}(X_k) = \mathcal{E}(\mathrm{Var}(X_k \,|\, Z)) + \mathrm{Var}(\mathcal{E}(X_k \,|\, Z)),$$

where

$$\mathrm{Var}(X_k \,|\, Z) = (1-Z)\mathrm{Var}(X_k) + Z\mathrm{Var}(X_{k-1})$$

and

$$\mathcal{E}(X_k \,|\, Z) = (1-Z)\mathcal{E}(1+X_k) + Z\mathcal{E}(1+X_{k-1}) = 1 + \frac{k}{p} - \frac{Z}{p},$$

implying that

$$\mathcal{E}(\mathrm{Var}(X_k \,|\, Z)) = (1-p)\mathrm{Var}(X_k) + p\mathrm{Var}(X_{k-1})$$

and

$$\mathrm{Var}(\mathcal{E}(X_k \,|\, Z)) = \frac{1}{p^2}\mathrm{Var}(Z) = \frac{1}{p^2}(p-p^2) = \frac{1-p}{p}.$$

Assume that $\mathrm{Var}(X_{k-1}) = \frac{(k-1)(1-p)}{p^2}$. Then

$$\mathrm{Var}(X_k) = (1-p)\mathrm{Var}(X_k) + \frac{(k-1)(1-p)}{p} + \frac{1-p}{p} = (1-p)\mathrm{Var}(X_k) + \frac{k(1-p)}{p},$$

implying that

$$\mathrm{Var}(X_k) = \frac{k(1-p)}{p^2}.$$

Since $\mathrm{Var}(X_1) = \frac{1-p}{p^2}$, the required conclusion then follows by induction.

7.126 a) $\mathcal{E}(X_k \,|\, \bar{X}_n = x) = x$, in view of the long-run average interpretation of expected value.
b) First, we note that $\mathcal{E}(X_i \,|\, \bar{X}_n = x) = \mathcal{E}(X_k \,|\, \bar{X}_n = x)$ for all $i, k \in \{1, \dots, n\}$, by symmetry. Then

$$x = \mathcal{E}(\bar{X}_n | \bar{X}_n = x) = \mathcal{E}\left(\sum_{i=1}^{n} \frac{X_i}{n} \middle| \bar{X}_n = x\right) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}\left(X_i | \bar{X}_n = x\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}\left(X_k | \bar{X}_n = x\right) = \mathcal{E}(X_k | \bar{X}_n = x).$$

Theory Exercises

7.127 a) Suppose that Y has a finite expectation, then, by definition of expectation, $\mathcal{E}|Y| = \sum_y |y| p_Y(y) < \infty$, and the law of total expectation,

$$\mathcal{E}(|Y|) = \sum_x \underbrace{\mathcal{E}(|Y| \,|\, X = x)}_{\geq 0} p_X(x) < \infty,$$

implies that $\mathcal{E}(|Y|\,|X=x)<\infty$ for all $x\in\mathcal{R}$ with $p_X(x)>0$. The required conclusion then follows.
b) If $\mathrm{Var}(Y)<\infty$, then $\mathcal{E}(Y^2)<\infty$, which, by (a), implies that $Y^2|_{X=x}$ has finite expectation for each $x\in\mathcal{R}$ such that $p_X(x)>0$. The latter implies that $Y|_{X=x}$ has finite variance for each $x\in\mathcal{R}$ such that $p_X(x)>0$.

7.128 a) Recall that (see the argument involving equation (7.70)), by the conditional form of the FEF, $g(Z,Y)|_{X=x}$ has finite expectation if and only if $\sum\sum_{(z,y)}|g(z,y)|p_{Z,Y|X}(z,y\,|\,x)<\infty$, in which case

$$\mathcal{E}(g(Z,Y)\,|\,X=x)=\sum\sum_{(z,y)} g(z,y)p_{Z,Y|X}(z,y\,|\,x).$$

Take $Z=X$ in the above statement. Since for all x such that $p_X(x)>0$, we have that

$$p_{X,Y|X}(z,y\,|\,x)=\frac{P(X=z,Y=y,X=x)}{P(X=x)}=\begin{cases} p_{Y|X}(y\,|\,x), & \text{if } z=x,\\ 0, & \text{if } z\neq x,\end{cases}$$

then $g(X,Y)|_{X=x}$ has finite expectation if and only if $\sum_y|g(x,y)|p_{Y|X}(y\,|\,x)<\infty$, in which case

$$\mathcal{E}(g(X,Y)\,|\,X=x)=\sum_y g(x,y)p_{Y|X}(y\,|\,x).$$

b) Let $g(x,y)=k(x)\ell(x,y)$. By (a),

$$\begin{aligned}\xi(x)\equiv\mathcal{E}(g(X,Y)\,|\,X=x) &= \sum_y g(x,y)p_{Y\,|\,X}(y\,|\,x)\\ &= k(x)\sum_y \ell(x,y)p_{Y\,|\,X}(y\,|\,x)=k(x)\mathcal{E}(\ell(X,Y)\,|\,X=x),\end{aligned}$$

then

$$\mathcal{E}(k(X)\ell(X,Y)\,|\,X)=\xi(X)=k(X)\mathcal{E}(\ell(X,Y)\,|\,X).$$

c) We have

$$\mathcal{E}(\mathcal{E}(g(X,Y)|X))=\sum_x\mathcal{E}(g(x,Y)|X=x)p_X(x)=\sum_x\left(\sum_y g(x,y)p_{Y|X}(y|x)\right)p_X(x)$$

$$=\sum_x\sum_y g(x,y)p_X(x)p_{Y|X}(y|x)=\sum\sum_{(x,y)} g(x,y)p_{X,Y}(x,y)=\mathcal{E}(g(X,Y)).$$

Advanced Exercises

7.129 7/9. Let N be the number of unreimbursed accidents. Then, by Example 6.15,

$$P(N=0\,|\,X=0)=\frac{12}{27}=\frac{4}{9},$$

$$P(N=1\,|\,X=0)=p_{Y,Z\,|\,X}(1,2\,|\,0)+p_{Y,Z\,|\,X}(2,1\,|\,0)=\frac{4}{27}+\frac{5}{27}=\frac{9}{27}=\frac{1}{3}.$$

$$P(N=2\,|\,X=0)=p_{Y,Z\,|\,X}(2,2\,|\,0)=\frac{6}{27}=\frac{2}{9}.$$

$$\mathcal{E}(N) = 0 \cdot P(N = 0 \,|\, X = 0) + 1 \cdot P(N = 1 \,|\, X = 0) + 2 \cdot P(N = 2 \,|\, X = 0) = \frac{1}{3} + 2 \times \frac{2}{9} = \frac{7}{9}.$$

7.130 a) First, by Exercise 7.128, we have that

$$\begin{aligned}
\mathcal{E}(X_k X_\ell \,|\, X_k = x_k) &= \sum_{x_\ell} x_k x_\ell \, p_{X_\ell \,|\, X_k}(x_\ell \,|\, x_k) \\
&= \sum_{x_\ell=0}^{n-x_k} x_k x_\ell \binom{n-x_k}{x_\ell} \left(\frac{p_\ell}{1-p_k}\right)^{x_\ell} \left(1 - \frac{p_\ell}{1-p_k}\right)^{n-x_k-x_\ell} = \frac{p_\ell x_k (n-x_k)}{1-p_k},
\end{aligned}$$

implying that

$$\mathcal{E}(X_k X_\ell \,|\, X_k) = \frac{p_\ell X_k (n - X_k)}{1-p_k}.$$

Thus,

$$\mathcal{E}(X_k X_\ell) = \mathcal{E}(\mathcal{E}(X_k X_\ell \,|\, X_k)) = \sum_{x_k=0}^{n} \frac{p_\ell x_k (n-x_k)}{1-p_k} \binom{n}{x_k} p_k^{x_k} (1-p_k)^{n-x_k} = (n-1) n p_k p_\ell.$$

b) Using (a), we have $\mathrm{Cov}(X_k, X_\ell) = \mathcal{E}(X_k X_\ell) - \mathcal{E}(X_k)\mathcal{E}(X_\ell) = n(n-1)p_k p_\ell - np_k \cdot np_\ell = -np_k p_\ell$ which agrees with Exercise 7.93.

7.131 Note that $(Y \,|\, X = x)$ is $\mathcal{B}(x, p)$. Thus, $\mathcal{E}(Y \,|\, X) = pX$ and $\mathrm{Var}(Y \,|\, X) = p(1-p)X$, implying that $\mathrm{Var}(Y) = \mathrm{Var}(\mathcal{E}(Y \,|\, X)) + \mathcal{E}(\mathrm{Var}(Y \,|\, X)) = p^2 \mathrm{Var}(X) + p(1-p)\mathcal{E}(X) = p\lambda$, and

$$\mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(\mathcal{E}(XY \,|\, X)) - \mathcal{E}(X)\mathcal{E}(\mathcal{E}(Y \,|\, X)) = \mathcal{E}(XpX) - p(\mathcal{E}(X))^2 = p\mathrm{Var}(X) = p\lambda.$$

Thus $\rho(X, Y) = \dfrac{p\lambda}{\sqrt{\lambda}\sqrt{p\lambda}} = \sqrt{p}$.

7.132 a) For each $k \in \mathcal{H}$, by Exercise 7.128,

$$\mathcal{E}((Y - \psi(X))k(X)) = \mathcal{E}(\mathcal{E}((Y - \psi(X))k(X)|X)) = \mathcal{E}(k(X)(\mathcal{E}(Y|X) - \mathcal{E}(Y|X))) = 0.$$

b) Note that $(Y - h(X))^2 - (Y - \psi(X))^2 - 2(Y - \psi(X))(\psi(X) - h(X)) = (h(X) - \psi(X))^2 \geq 0$.
c) Taking the expectation of both sides of the inequality in (b) and applying (a), we obtain that for all $h \in \mathcal{H}$,

$$\mathcal{E}\left((Y - h(X))^2\right) \geq \mathcal{E}\left((Y - \psi(X))^2\right) = \mathcal{E}\left((Y - \mathcal{E}(Y \,|\, X))^2\right).$$

Since $\psi \in \mathcal{H}$, then

$$\mathcal{E}\left((Y - \mathcal{E}(Y \,|\, X))^2\right) = \min_{h \in \mathcal{H}} \mathcal{E}\left((Y - h(X))^2\right).$$

7.133 a) If $\mathcal{E}(Y \,|\, X) = a + bX$ then $\mathcal{E}(Y) = \mathcal{E}(\mathcal{E}(Y|X)) = a + b\mathcal{E}(X)$. By Exercise 7.128, we have that $\mathcal{E}(XY|X) = X\mathcal{E}(Y|X)$. Then

$$(\rho(X, Y))^2 = \frac{(\mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y))^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \frac{[\mathcal{E}(\mathcal{E}(XY|X)) - \mathcal{E}(X)\,(a + b\mathcal{E}(X))]^2}{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

$$= \frac{\left(\mathcal{E}(X(a+bX)) - a\mathcal{E}(X) - b\,(\mathcal{E}(X))^2\right)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \frac{b^2\left(\mathcal{E}(X^2) - \left(\mathcal{E}(X^2)\right)\right)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

$$= \frac{b^2\left(\mathcal{E}(X^2) - (\mathcal{E}(X))^2\right)}{\mathrm{Var}(Y)} = \frac{b^2\mathrm{Var}(X)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(\mathcal{E}(Y|X))}{\mathrm{Var}(Y)}.$$

b) If the conditional expectation is linear, the coefficient of determination is the proportion of the total variance of Y that is explained.

7.6 Review Exercises

Basic Exercises

7.134 31.4, due to the long-run average interpretation of the expected value.

7.135 a) \$290
b) \$2900

7.136 a) The mean of the salesman's daily commission is \$420 with a standard deviation of \$76.68. Indeed, let X be the daily number of sales and Y be the salesman's total daily commission. Then X is $\mathcal{B}(n=20, p=0.6)$ and $Y = 35X$, implying that

$$\mathcal{E}(Y) = 35\mathcal{E}(X) = 35 \times 20 \times 0.6 = 420, \quad \sigma_Y = 35\sigma_X = 35\sqrt{20 \times 0.6 \times 0.4} = 35\sqrt{4.8} \approx 76.68$$

b) In (a) it was assumed that the attempts to make a sale were independent, with the same probability of success (sale), and thus represented the Bernoulli trials.

7.137 Let Y denote the amount the insurance company pays. Then the PMF of Y is

y	$p_Y(y)$
0	$e^{-0.6}$
1000000	$0.6e^{-0.6}$
2000000	$1 - 1.6e^{-0.6}$

Then the mean and standard deviation of Y are equal to \$573089.75 and \$698899.60, respectively.

7.138 a) No, because the mean returns are equal.
b) The second investment is more conservative since its standard deviation is smaller.

7.139 a) The nth factorial moment equals λ^n. Since $k(k-1)\dots(k-n+1) = 0$ for all $k = 0, \dots, n-1$, then

$$\mathcal{E}(X(X-1)\dots(X-n+1)) = \sum_{k=0}^{\infty} k(k-1)\dots(k-n+1)\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} \cdot \frac{e^{-\lambda}\lambda^k}{k!}$$

$$= e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{\lambda^{\ell+n}}{\ell!} = \lambda^n,$$

where we put $\ell = k - n$.
b) From (a) with $n = 1$, $\mathcal{E}(X) = \lambda$; when $n = 2$, $\mathcal{E}(X(X-1)) = \mathcal{E}(X^2) - \lambda = \lambda^2$, implying that

$\mathcal{E}(X^2) = \lambda^2 + \lambda$, thus, $\mathrm{Var}(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda$.

7.140 a) Using Exercise 5.136 (a), we have

$$\mathcal{E}(Y) = \sum_{y=1}^{m-1} yp(1-p)^{y-1} + m(1-p)^{m-1} = \frac{1-(1-p)^m}{p}.$$

b) Using the FEF and the PMF of X, we have

$$\mathcal{E}(Y) = \sum_{x=1}^{\infty} \min(x,m)p(1-p)^{x-1} = \sum_{x=1}^{m} xp(1-p)^{x-1} + pm \sum_{x=m+1}^{\infty} (1-p)^{x-1}$$

$$= \sum_{x=1}^{\infty} xp(1-p)^{x-1} - \sum_{x=m+1}^{\infty} xp(1-p)^{x-1} + pm(1-p)^m \cdot \frac{1}{1-(1-p)}$$

$$= \mathcal{E}(X) - (1-p)^m \sum_{j=1}^{\infty} (j+m)p(1-p)^{j-1} + m(1-p)^m$$

$$= \mathcal{E}(X) - (1-p)^m \mathcal{E}(m+X) + m(1-p)^m$$

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

c) Using tail probabilities of Y, for $m \in \mathcal{N}$, we have

$$\mathcal{E}(Y) = \sum_{y=0}^{\infty} P(Y > y) = \sum_{y=0}^{\infty} P(\min(X,m) > y) = \sum_{y=0}^{m-1} P(X > y) = \sum_{y=0}^{m-1} (1-p)^y = \frac{1-(1-p)^m}{p}.$$

7.141 a) Using the expectation of a geometric random variable, the average time, taken to complete the exam, is $\frac{1}{0.020} = 50$ minutes.
b) Using Exercise 7.140, if there is a 1 hour time limit (i.e. m=60), then the average time, that it takes a student to complete the exam, would be equal to $\frac{1-(1-0.02)^{60}}{0.02} \approx 35.122$ minutes.

7.142 a) The following table represents the conditional expectation of the number of radios for each number of televisions:

Televisions x	Conditional expectation $\mathcal{E}(Y \mid X = x)$
0	5.500
1	5.391
2	5.650
3	5.562
4	5.646

b) The following table represents the conditional variance of the number of radios given the number of televisions:

Televisions x	Conditional variance $\mathrm{Var}(Y \mid X = x)$
0	1.250
1	0.628
2	0.788
3	0.654
4	0.699

7.143 a) From Exercise 7.142 (a), we have $\mathcal{E}(Y) = \sum_x \mathcal{E}(Y \mid X = x)p_X(x) = 5.574$.
b) From Exercise 6.123 we have $\mathcal{E}(Y) = 5.576$. The small difference between this answer and the one in (a) is due to rounding error.
c) From Exercise 7.142 (a) and (b), we have $\mathrm{Var}(Y) = 0.722$.
d) Using Exercise 6.123 we have $\mathrm{Var}(Y) = 0.724$, where the small difference between this answer and the one in (c) exists due to rounding error.

7.144 a) The PMF of X is $p_X(x) = \frac{13-2x}{36}$ for $x = 1, 2, 3, 4, 5, 6$ and $p_X(x) = 0$ otherwise. Thus,

$$\mathcal{E}(X) = \sum_{x=1}^{6} x \cdot \frac{13-2x}{36} = \frac{91}{36}.$$

b) Using the joint PMF of Y and Z and the FEF, we have

$$\mathcal{E}(X) = \sum\sum_{(y,z)\in\{1,\ldots,6\}^2} \min(y,z)\frac{1}{36} = \frac{91}{36}.$$

7.145 a) $\mathrm{Cov}(X, Y) = \rho(X, Y)\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)} = -\frac{1}{2}\sqrt{25}\,\sqrt{144} = -30$.
b) $\mathrm{Cov}(X + Y, X - Y) = \mathrm{Var}(X) - \mathrm{Var}(Y) = 25 - 144 = -119$.
c) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) = 25 + 144 + 2(-30) = 109$.
d) $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y) = 25 + 144 - 2(-30) = 229$.
e) $\mathrm{Var}(3X + 4Y) = 9\mathrm{Var}(X) + 16\mathrm{Var}(Y) + 24\mathrm{Cov}(X, Y) = 1809$.
f) $\mathrm{Cov}(3X + 4Y, -2X + 5Y) = -6\mathrm{Var}(X) + 7\mathrm{Cov}(X, Y) + 20\mathrm{Var}(Y) = 2520$.

7.146 Let X be the number of wells drilled until the third producing well is found. Then $X \sim \mathcal{NB}(3, 1/3)$. Let Y be the amount of the company's profit. Then $\mathcal{E}(Y) = \$7500$ and the standard deviation of Y is \$2121.32, since $Y = 12000 - 500X$ and, thus,

$$\mathcal{E}(Y) = 12000 - 500\mathcal{E}(X) = 7500, \quad \sigma_Y = 500\sigma_X = 500\sqrt{18} = 2121.32$$

7.147 The amount (call it, say, c) that should be put aside annually for cash register repair is equal to \$178.67. Indeed, let X be the annual repair cost. Then, on the other one hand, we need to have $P(X > c) \le 0.05$, on the other hand, by Chebyshev's inequality,

$$P(X > c) = P(X - \mu_X > c - 125) \le P(|X - \mu| > c - 125) \le \frac{144}{(c-125)^2}.$$

Thus, it suffices to pick c so that $\frac{144}{(c-125)^2} = 0.05$, and the required result follows.

7.148 Note that $\mathcal{E}(X) = \frac{1}{10^3}(0.000 + 0.001 + 0.002 + \ldots + 0.999) = 0.4995$, and

$$\sigma_X = \sqrt{\frac{1}{10^3}(0.000^2 + 0.001^2 + 0.002^2 + \ldots + 0.999^2) - 0.4995^2} \approx 0.2887$$

7.149 Since the expected amount paid by the policy is

$$\sum_{x=1}^{5}(5000 - 1000(x-1)) \times 0.1 \times 0.9^{x-1} = 1314.41,$$

the company should charge \$1414.41 (per policy), so that on average it nets a \$100 profit.

7.150 a) If X represents the number of red balls obtained assuming that they are drawn with replacement, then $X \sim \mathcal{B}(2, 0.6)$ with the PMF of X given by:

$$p_X(x) = \begin{cases} \binom{2}{x}0.6^x 0.4^{2-x}, & \text{if } x \in \{0,1,2\}, \\ 0, & \text{otherwise.} \end{cases}$$

b) If Y represents the number of red balls obtained assuming that they are drawn without replacement, then $Y \sim \mathcal{H}(10, 2, 0.6)$ and the PMF of Y is given by:

$$p_Y(y) = \begin{cases} \dfrac{\binom{6}{y}\binom{4}{2-y}}{\binom{10}{2}}, & \text{if } y \in \{0,1,2\}, \\ 0, & \text{otherwise.} \end{cases}$$

c) $\mathcal{E}(X) = 1.2$ and $\mathcal{E}(Y) = 1.2$, since, by definition of expectation and PMFs in (a),(b), we have

$$\mathcal{E}(X) = 1 \times 2(0.6)(0.4) + 2 \times 1(0.6)^2 = 1.2,$$

$$\mathcal{E}(Y) = 1 \times \frac{\binom{6}{1}\binom{4}{1}}{\binom{10}{2}} + 2 \times \frac{\binom{6}{2}\binom{4}{0}}{\binom{10}{2}} = \frac{24}{45} + 2 \times \frac{15}{45} = \frac{54}{45} = 1.2$$

d) From Table 7.4 we have that $\mathcal{E}(X) = 2 \times 0.6 = 1.2$ and $\mathcal{E}(Y) = 2 \times 0.6 = 1.2$;
e) $\mathrm{Var}(X) = 0.48$ and $\mathrm{Var}(Y) = 0.4267$, since

$$\mathrm{Var}(X) = 0^2\binom{2}{0}0.6^0 0.4^2 + 1^2\binom{2}{1}0.6 \cdot 0.4 + 2^2\binom{2}{2}0.6^2 0.4^0 - 1.2^2 = 0.48,$$

$$\mathrm{Var}(Y) = 0^2\frac{\binom{6}{0}\binom{4}{2}}{\binom{10}{2}} + 1^2\frac{\binom{6}{1}\binom{4}{1}}{\binom{10}{2}} + 2^2\frac{\binom{6}{2}\binom{4}{0}}{\binom{10}{2}} - 1.2^2 = \frac{84}{45} - 1.2^2 \approx 0.4267$$

f) From Table 7.10, it follows that $\mathrm{Var}(X) = 2 \times 0.6 \times 0.4 = 0.48$ and
$\mathrm{Var}(Y) = \frac{10-2}{10-1} \times 2 \times 0.6 \times 0.4 \approx 0.4267$

7.151 a) $1 + n(1-(1-p)^n)$. Let X be the number of tests required under the alternative scheme for a random sample of n people. Then

$$\mathcal{E}(X) = 1 \times (1-p)^n + (n+1)\left(1-(1-p)^n\right) = 1 + n\left(1-(1-p)^n\right).$$

b) The alternative scheme is superior whenever $1 + n(1 - (1-p)^n) < n$, that is whenever $p < 1 - \left(\frac{1}{n}\right)^{\frac{1}{n}}$.

7.152 a) By Exercise 7.151, for each group of size n/m, the expected number of tests is equal to $1 + \frac{n}{m}(1 - (1-p)^{\frac{n}{m}})$, thus, the total expected number of tests equals $m + n(1 - (1-p)^{\frac{n}{m}})$.
b) The expected number of tests using the first scheme (where people are tested separately) is 100, using the second (pooled) scheme is 64.4, and using the third (grouped) scheme is 24.9.

7.153 $n\left(\frac{n-1}{n}\right)^n$. Let X denote the number of empty bins, and let A_i denote the event that the i^{th} bin is empty, $i = 1, \ldots, n$. Then $X = \sum_{i=1}^n 1_{A_i}$ and

$$\mathcal{E}(X) = \mathcal{E}\left(\sum_{i=1}^n 1_{A_i}\right) = \sum_{i=1}^n \mathcal{E}(1_{A_i}) = \sum_{i=1}^n P(A_i) = \sum_{i=1}^n \left(\frac{n-1}{n}\right)^n = n\left(\frac{n-1}{n}\right)^n.$$

7.154 a) Let $U = \min(X,Y)$. Then

$$\mathcal{E}(U) = \sum_{u=0}^\infty P(U > u) = \sum_{u=0}^\infty P(X > u)P(Y > u) = \sum_{u=0}^\infty (1-p)^u(1-q)^u = \frac{1}{p+q-pq}.$$

b) Let $V = \max(X,Y)$. Then

$$\mathcal{E}(V) = \sum_{v=0}^\infty P(V > v) = \sum_{v=0}^\infty (1 - P(V \le v)) = \sum_{v=0}^\infty [1 - P(X \le v)P(Y \le v)]$$

$$= \sum_{v=0}^\infty [1 - (1-(1-p)^v)(1-(1-q)^v)] = \frac{1}{p} + \frac{1}{q} - \frac{1}{p+q-pq}.$$

c) From (a),(b) for $p = q$ we have that

$$\mathcal{E}(\min(X,Y)) = \frac{1}{(2-p)p} \quad \text{and } \mathcal{E}(\max(X,Y)) = \frac{3-2p}{(2-p)p}.$$

7.155 Note that $\mathrm{Var}(X) + \mathrm{Var}(Y) = \frac{1}{2} + \frac{13}{4} = \frac{15}{4} = \mathrm{Var}(X+Y)$, since, on the one hand,

$$\mathcal{E}(X) = 1 \cdot \frac{4}{16} + 2 \cdot \frac{8}{16} + 3 \cdot \frac{4}{16} = 2, \quad \mathcal{E}(Y) = 1 \cdot \frac{4}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{4}{16} + 6 \cdot \frac{4}{16} = 3.5,$$

implying that

$$\mathrm{Var}(X) = 1^2 \times \frac{4}{16} + 2^2 \times \frac{8}{16} + 3^2 \times \frac{4}{16} - 2^2 = \frac{1}{2},$$
$$\mathrm{Var}(Y) = \left(1^2 + 3^2 + 4^2 + 6^2\right) \times \frac{4}{16} - 3.5^2 = 3.25 = \frac{13}{4};$$

on the other hand,

$$\mathcal{E}(X+Y) = 2 \times \frac{1}{16} + 3 \times \frac{2}{16} + 4 \times \frac{2}{16} + 5 \times \frac{3}{16} + 6 \times \frac{3}{16} + 7 \times \frac{2}{16} + 8 \times \frac{2}{16} + 9 \times \frac{1}{16} = \frac{11}{2},$$

$$\begin{aligned}\mathrm{Var}(X+Y) &= 2^2\times\frac{1}{16}+3^2\times\frac{2}{16}+4^2\times\frac{2}{16}+5^2\times\frac{3}{16}+6^2\times\frac{3}{16}\\ &+ 7^2\times\frac{2}{16}+8^2\times\frac{2}{16}+9^2\times\frac{1}{16}-\left(\frac{11}{2}\right)^2=\frac{544}{16}-\frac{121}{4}=\frac{15}{4}.\end{aligned}$$

7.156 By Exercise 6.138, the conditional distribution of the number of undetected defectives (in a random sample of size n), given that the number of detected defectives equals k, is binomial with parameters $n-k$ and $\frac{p(1-q)}{1-pq}$, implying that the required conditional expectation is equal to $(n-k)\frac{p(1-q)}{1-pq}$ and the conditional variance is $(n-k)\frac{p(1-p)(1-q)}{(1-pq)^2}$.

7.157 a) The mean of X is $77.7'' = 6'5.7''$ and the standard deviation is $2.759''$.
b) The mean of Y is -2.642 cm and the standard deviation is 7.007 cm, since $Y = 2.54X - 200$ (where recall that $1'' = 2.54\,\mathrm{cm}$), resulting in

$$\mathcal{E}(Y) = 2.54\mathcal{E}(X) - 200 = 2.54\times 77.7 - 200 = -2.642,$$

$$\sigma_Y = 2.54\sigma_X \approx 7.007$$

c) Clearly, since there is a linear relationship between X and Y and $2.54 > 0$, the correlation coefficient is 1.

7.158 Note that

$$\mathcal{E}\left(\frac{1}{X}\right)=\sum_{x=1}^{\infty}\frac{p(1-p)^{x-1}}{x}=\frac{p}{1-p}\underbrace{\sum_{x=1}^{\infty}\frac{(1-p)^x}{x}}_{=-\ln(1-(1-p))}=-\frac{p\ln(p)}{1-p},$$

which is certainly different from $\dfrac{1}{\mathcal{E}(X)}=p$.

7.159 a) 1. Since it takes X trials to receive a single success, one expects a single success in the next X trials, thus the expected value of Y should equal 1.
b) 1. Note that $X \sim \mathcal{G}(p)$ and $(Y\,|\,X=x)\sim\mathcal{B}(x,p)$. So

$$\mathcal{E}(Y)=\sum_x \mathcal{E}(Y|X=x)p_X(x)=\sum_{x=1}^{\infty}(xp)\left(p(1-p)^{x-1}\right)=1.$$

c) $2-2p$. Note that $\mathcal{E}(Y\,|\,X)=pX$ and $\mathrm{Var}(Y\,|\,X)=p(1-p)X$, thus,

$$\mathrm{Var}(\mathcal{E}(Y\,|\,X))=\mathrm{Var}(pX)=p^2\mathrm{Var}(X)=1-p$$

and

$$\mathcal{E}(\mathrm{Var}(Y\,|\,X))=\mathcal{E}(p(1-p)X)=p(1-p)\mathcal{E}(X)=1-p,$$

implying that $\mathrm{Var}(Y)=\mathrm{Var}(\mathcal{E}(Y\,|\,X))+\mathcal{E}(\mathrm{Var}(Y\,|\,X))=2(1-p)$.

7.160 If Y has $\mathcal{P}(\lambda)$, then, since distribution of X_n converges to distribution of Y, we must have that

$$\mathrm{Var(Y)}=\lim_{n\to\infty}\mathrm{Var}(X_n)=\lim_{n\to\infty}np_n(1-p_n)=\lim_{n\to\infty}\left[np_n-\frac{(np_n)^2}{n}\right]=\lambda,$$

since $np_n \to \lambda$ as $n \to \infty$.

7.161 a) If N is large relative to n, then sampling without replacement is well approximated by sampling with replacement, implying that distribution of X is approximately binomial with parameters n and p, thus, $\mathrm{Var}(X) \approx np(1-p)$.
b) Since the exact distribution of X is $\mathcal{H}(N, n, p)$, then $\mathrm{Var}(X) = (\frac{N-n}{N-1})np(1-p) \longrightarrow np(1-p)$ as $N \to \infty$.

7.162 a) $\dfrac{zm}{m+n}$, since, for $z = 0, \ldots, m+n$,

$$\mathcal{E}(X|X+Y=z) = \sum_{x=0}^{m} x p_{X|X+Y=z}(x|z) = \sum_{x=0}^{m} x \, \frac{p_{X,X+Y}(x,z)}{p_{X+Y}(z)} = \sum_{x=0}^{m} x \, \frac{p_X(x) p_Y(z-x)}{p_{X+Y}(z)}$$

$$= \sum_{x=0}^{m} x \, \frac{\binom{m}{x} p^x (1-p)^{m-x} \binom{n}{z-x} p^{z-x} (1-p)^{n-z+x}}{\binom{m+n}{z} p^z (1-p)^{m+n-z}} = \underbrace{\sum_{x=0}^{m} x \, \frac{\binom{m}{x}\binom{n}{z-x}}{\binom{m+n}{z}}}_{=\,\text{mean of } \mathcal{H}(m+n, z, \frac{m}{m+n})} = \frac{zm}{m+n}.$$

b) $\dfrac{zmn(m+n-z)}{(m+n-1)(m+n)^2}$. Indeed, since $(X \mid X+Y=z)$ is $\mathcal{H}(m+n, z, \frac{m}{m+n})$, then

$$\mathrm{Var}(X \mid X+Y=z) = \left(\frac{m+n-z}{m+n-1}\right) z \left(\frac{m}{m+n}\right)\left(\frac{n}{m+n}\right).$$

Theory Exercises

7.163 a) For all x,

$$P(X=x) = \sum_y P(X=x, Y=y) = P(X=x, Y=x) + \sum_{y \neq x} P(X=x, Y=y),$$

then, since $P(Y=X)=1$,

$$1 = \sum_x P(X=x) = \underbrace{\sum_x P(X=x, Y=x)}_{=P(Y=X)=1} + \sum_x \left(\sum_{y \neq x} P(X=x, Y=y) \right),$$

thus,

$$1 = 1 + \sum_x \left(\sum_{y \neq x} P(X=x, Y=y) \right),$$

implying that $\sum_{y \neq x} P(X=x, Y=y) = 0$ for each x.
Therefore, $P(X=x) = P(X=x, Y=x)$ for all x. Similarly, one shows that $P(Y=x) = P(Y=x, X=x)$ for all x, thus, for all $x \in \mathcal{R}$,

$$p_X(x) = P(X=x,\, Y=x) = p_Y(x),$$

i.e. X and Y have the same PMF.
b) By (a) and computing formulas for expectation and variance, X and Y must have the same mean and variance, since

$$\mathcal{E}(X) = \sum_x x p_X(x) = \sum_x x p_Y(x) = \mathcal{E}(Y),$$

$$\operatorname{Var}(X) = \sum_x x^2 p_X(x) - (\mathcal{E}(X))^2 = \sum_x x^2 p_Y(x) - (\mathcal{E}(Y))^2 = \operatorname{Var}(Y).$$

7.164 a) Let σ^2 be the (finite) variance of X. Also, let $\bar{X}_n = \frac{X_1+\ldots+X_n}{n}$, then using properties of expectation and variance, $\mathcal{E}(\bar{X}_n) = \mathcal{E}(X)$ and $\operatorname{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$. Then, by Chebyshev's inequality applied to $\bar{X}_n$, for all $\epsilon > 0$,

$$P(|\bar{X}_n - \mathcal{E}(X)| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2},$$

where the right-hand side goes to 0, as $n \to \infty$. Thus, for each $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mathcal{E}(X)\right| < \epsilon\right) = 1 - \lim_{n\to\infty} P(|\bar{X}_n - \mathcal{E}(X)| \geq \epsilon) = 1 - 0 = 1.$$

b) If $X = 1_E$, then $\mathcal{E}(X) = P(E)$, and $\bar{X}_n$ in (a) represents relative frequency of the occurrence of event E in n independent repetitions of the experiment. Then the weak law of large numbers (given in (a)) states that the relative frequency of event E in a large number n of repetitions of the experiment is approximately equal to $P(E)$ (with probability close to 1), which to a certain extent provides confirmation of the frequentist interpretation of probability.

Advanced Exercises

7.165 $\mathcal{E}(X) = \alpha$ and $\operatorname{Var}(X) = \alpha + \alpha^2\beta$. Indeed, note that

$$\mathcal{E}(X) = \sum_{x=1}^{\infty} x \left(\frac{\alpha}{1+\alpha\beta}\right)^x \cdot \frac{1(1+\beta)\cdots(1+(x-1)\beta)}{x!} p_X(0)$$

$$= \frac{\alpha}{1+\alpha\beta}\left(p_X(0) + \sum_{x=2}^{\infty}\left(\frac{\alpha}{1+\alpha\beta}\right)^{x-1} \cdot \frac{1(1+\beta)\cdots(1+(x-1)\beta)}{(x-1)!} p_X(0)\right)$$

$$= \frac{\alpha}{1+\alpha\beta}\left((1+\beta\cdot 0)p_X(0) + \sum_{y=1}^{\infty}\left(\frac{\alpha}{1+\alpha\beta}\right)^y \frac{1(1+\beta)\cdots(1+(y-1)\beta)(1+y\beta)}{y!} p_X(0)\right)$$

$$= \frac{\alpha}{1+\alpha\beta}\mathcal{E}(1+\beta X) = \frac{\alpha}{1+\alpha\beta}\left(1+\beta\mathcal{E}(X)\right), \text{ implying that } \mathcal{E}(X) = \alpha.$$

Similarly,

$$\mathcal{E}(X^2) = \sum_{x=1}^{\infty} x^2 \left(\frac{\alpha}{1+\alpha\beta}\right)^x \cdot \frac{1(1+\beta)\cdots(1+(x-1)\beta)}{x!} p_X(0)$$

$$= \frac{\alpha}{1+\alpha\beta}\left(p_X(0) + \sum_{x=2}^{\infty} x\left(\frac{\alpha}{1+\alpha\beta}\right)^{x-1} \cdot \frac{1(1+\beta)\cdots(1+(x-1)\beta)}{(x-1)!} p_X(0)\right)$$

$$= \frac{\alpha}{1+\alpha\beta}\left(p_X(0) + \sum_{y=1}^{\infty}(y+1)\left(\frac{\alpha}{1+\alpha\beta}\right)^y \cdot \frac{1(1+\beta)\cdots(1+(y-1)\beta)(1+y\beta)}{y!}p_X(0)\right)$$

$$= \frac{\alpha}{1+\alpha\beta}\left(\sum_{y=0}^{\infty}(y+1)(1+y\beta)p_X(y)\right) = \frac{\alpha}{1+\alpha\beta}\,\mathcal{E}\left((X+1)(1+\beta X)\right)$$

$$= \frac{\alpha}{1+\alpha\beta}\left(1+(\beta+1)\mathcal{E}(X)+\beta\mathcal{E}(X^2)\right) = \frac{\alpha(1+(\beta+1)\alpha)}{1+\alpha\beta} + \frac{\alpha\beta}{1+\alpha\beta}\,\mathcal{E}(X^2),$$

implying that $\mathcal{E}(X^2) = \alpha + \alpha^2\beta + \alpha^2$, thus,

$$\mathrm{Var}(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = \alpha + \alpha^2\beta.$$

7.166 a) Let X denote the total number of coupons that need to be collected until a complete set is obtained. Also, let X_i be the number of coupons that are collected to get the $(i+1)^{th}$ new coupon after the i^{th} new coupon is collected, $i = 0, \ldots, M-1$. Then $X_i \sim \mathcal{G}\left(\frac{M-i}{M}\right)$. Also, $X = \sum_{i=0}^{M-1} X_i$, and therefore by linearity

$$\mathcal{E}(X) = \mathcal{E}\left(\sum_{i=0}^{M-1} X_i\right) = \sum_{i=0}^{M-1}\mathcal{E}(X_i) = \sum_{i=0}^{M-1}\frac{M}{M-i} = M\sum_{j=1}^{M}\frac{1}{j}.$$

b) Since the $X_0, \ldots, X_{M-1}$ are independent, we have that

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=0}^{M-1} X_i\right) = \sum_{i=0}^{M-1}\mathrm{Var}(X_i) = \sum_{i=0}^{M-1}\frac{1-(M-i)/M}{((M-i)/M)^2}$$

$$= M\sum_{i=0}^{M-1}\frac{i}{(M-i)^2} = M\sum_{j=1}^{M}\frac{M-j}{j^2} = M^2\sum_{j=1}^{M}\frac{1}{j^2} - M\sum_{j=1}^{M}\frac{1}{j}.$$

c) Since $\ln(M+1) = \int_0^M \frac{1}{x+1}dx \le \sum_{j=1}^{M}\frac{1}{j} \le 1 + \int_1^M \frac{1}{x}dx = 1 + \ln(M)$, then

$$\underbrace{\frac{\ln(M+1)}{\ln(M)}}_{\longrightarrow 1} \le \frac{\sum_{j=1}^{M}\frac{1}{j}}{\ln(M)} \le \underbrace{\frac{1}{\ln(M)}}_{\longrightarrow 0} + 1,$$

implying that

$$\lim_{M\to\infty}\frac{M\sum_{j=1}^{M}\frac{1}{j}}{M\ln(M)} = 1,$$

thus, by (a), $\mathcal{E}(X) = M\sum_{j=1}^{M}\frac{1}{j} \sim M\ln(M)$, as $M \to \infty$.

7.167 0.244; Let X_i denote the number of claims made by a given (randomly selected) policyholder in year i, $i = 1, 2$. Let H be the event that the policy holder is "high risk". Then we know that $P(H) = 0.1$, $(X_i \,|\, 1_H = 1)$ is $\mathcal{P}(0.6)$ and $(X_i \,|\, 1_H = 0)$ is $\mathcal{P}(0.1)$ for $i = 1, 2$. Moreover, $(X_1 \,|\, 1_H = y)$ and $(X_2 \,|\, 1_H = y)$ are independent (for each $y \in \{0, 1\}$). Then

$$\mathcal{E}(X_2 \,|\, X_1 = 1) = \sum_{x_2=0}^{\infty} x_2\, p_{X_2\,|\,X_1}(x_2 \,|\, 1) = \sum_{x_2=0}^{\infty} x_2 \cdot \frac{P(X_2 = x_2, X_1 = 1)}{P(X_1 = 1)}$$

$$= \sum_{x_2=0}^{\infty} x_2 \cdot \frac{P(X_2 = x_2, X_1 = 1 \,|\, 1_H = 1)P(H) + P(X_2 = x_2, X_1 = 1 \,|\, 1_H = 0)P(H^c)}{P(X_1 = 1 \,|\, 1_H = 1)P(H) + P(X_1 = 1 \,|\, 1_H = 0)P(H^c)}$$

$$= \sum_{x_2=0}^{\infty} x_2 \cdot \frac{P(X_2 = x_2 | \, 1_H = 1)P(X_1 = 1 | \, 1_H = 1)0.1 + P(X_2 = x_2 | \, 1_H = 0)P(X_1 = 1 | \, 1_H = 0)0.9}{P(X_1 = 1 \,|\, 1_H = 1)0.1 + P(X_1 = 1 \,|\, 1_H = 0)0.9}$$

$$= \sum_{x_2=0}^{\infty} x_2 \cdot \frac{\dfrac{e^{-0.6}0.6^{x_2}}{x_2!} \cdot \dfrac{e^{-0.6}0.6}{1!} \cdot 0.1 + \dfrac{e^{-0.1}0.1^{x_2}}{x_2!} \cdot \dfrac{e^{-0.1}0.1}{1!} \cdot 0.9}{\dfrac{e^{-0.6}0.6}{1!} \cdot 0.1 + \dfrac{e^{-0.1}0.1}{1!} \cdot 0.9}$$

$$= \frac{e^{-0.6} \times 0.06}{e^{-0.6} \times .06 + e^{-0.1} \times .09} \underbrace{\left(\sum_{x_2=0}^{\infty} x_2 \cdot \frac{e^{-0.6}0.6^{x_2}}{x_2!} \right)}_{=0.6} + \frac{e^{-0.1} \times 0.09}{e^{-0.6} \times .06 + e^{-0.1} \times .09} \underbrace{\left(\sum_{x_2=0}^{\infty} x_2 \cdot \frac{e^{-0.1}0.1^{x_2}}{x_2!} \right)}_{=0.1}$$

$$= \frac{e^{-0.6} \times 0.036 + e^{-0.1} \times 0.009}{e^{-0.6} \times 0.06 + e^{-0.1} \times 0.09} \approx 0.244$$

7.168 a) 8. Let X represent the number of times that the spelunker chooses the route which returns her to the cavern in 3 hours, let Y represent the number of times that she goes down the route which returns her in 4 hours and let T represent her total time spent underground. Then $T = 1 + 3X + 4Y$. If the route is chosen at random each time, then $p_{X,Y}(x, y) = \binom{x+y}{y} \frac{1}{3^{1+x+y}}$ for $x, y \in \mathcal{Z}_+$. Thus

$$\mathcal{E}(T) = \sum_{(x,y)} \sum (1 + 3x + 4y) p_{X,Y}(x, y) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} (1 + 3x + 4y) \binom{x+y}{y} \frac{1}{3^{1+x+y}} = 8.$$

Alternatively, note that, by symmetry, $\mathcal{E}(X) = \mathcal{E}(Y)$ and $(X + Y + 1)$ is $\mathcal{G}(1/3)$, then $3 = \mathcal{E}(X + Y + 1) = 2\mathcal{E}(X) + 1$, implying that $\mathcal{E}(X) = \mathcal{E}(Y) = 1$. Therefore,

$$\mathcal{E}(T) = 1 + 3\mathcal{E}(X) + 4\mathcal{E}(Y) = 1 + 3 + 4 = 8.$$

b) 4.5; If the spelunker chooses a route at random from among the routes not already tried, then the distribution of the total amount of time spent underground by the spelunker is given by:

t	$p_T(t)$
1	$\frac{1}{3}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
8	$\frac{1}{3}$

and $\mathcal{E}(T) = 4.5$

7.169 $(p + 1)/p^2$. Let X be the number of trials until 2 consecutive successes occur. Let H be the event that the first trial is a success. Then

$$\mathcal{E}(X \,|\, 1_H = 1) = 2p + (2 + \mathcal{E}(X))(1 - p) \text{ and } \mathcal{E}(X \,|\, 1_H = 0) = 1 + \mathcal{E}(X).$$

Therefore,

$$\begin{aligned}\mathcal{E}(X) &= \mathcal{E}(\mathcal{E}(X \mid 1_H)) = \mathcal{E}(X \mid 1_H = 1)P(1_H = 1) + \mathcal{E}(X \mid 1_H = 0)P(1_H = 0) \\ &= 2p^2 + (2 + \mathcal{E}(X))p(1-p) + (1 + \mathcal{E}(X))(1-p),\end{aligned}$$

implying that $\mathcal{E}(X) = (p+1)/p^2$.

7.170 a) 0. Note that for each $i = 1, \ldots, m$, X_i has $\mathcal{B}(n, p_i)$. Moreover, by Exercise 7.130(b), $\operatorname{Cov}(X_i, X_j) = -np_i p_j$ for all $i \neq j$, thus,

$$\operatorname{Var}(X_1 + \ldots + X_m) = \sum_{k=1}^{m} \operatorname{Var}(X_k) + 2 \sum\sum_{1 \le i < j \le m} \operatorname{Cov}(X_i, X_j)$$

$$= \sum_{k=1}^{m} np_k(1-p_k) - 2 \sum\sum_{1 \le i < j \le m} np_i p_j = n \left(\sum_{k=1}^{m} p_k - \left(\sum_{k=1}^{m} p_k \right)^2 \right) = n(1 - 1^2) = 0.$$

b) 0, since $X_1 + \cdots + X_m = n$ and $\operatorname{Var}(n) = 0$.

7.171 a) Since $P_X(t) = \mathcal{E}(t^X)$, where X is a nonnegative-integer valued random variable with finite nth moment, then $P_X^{(1)}(t) = \mathcal{E}(Xt^{X-1})$, $P_X^{(2)}(t) = \mathcal{E}\left(X(X-1)\, t^{X-2}\right)$, etc. up to $P_X^{(n)}(t) = \mathcal{E}\left(X(X-1)\ldots(X-n+1)\, t^{X-n}\right)$, implying that

$$P_X^{(n)}(1) = \mathcal{E}\left(X(X-1)\ldots(X-n+1)\right).$$

b) λ^n. Since for $X \sim \mathcal{P}(\lambda)$, by Exercise 7.53(b), we have that $P_X(t) = e^{\lambda(t-1)}$ and

$$P_X^{(n)}(t) = \lambda^n e^{\lambda(t-1)}, \;\; P_X^{(n)}(1) = \lambda^n,$$

then, by (a), $\mathcal{E}(X(X-1)\ldots(X-n+1)) = P_X^{(n)}(1) = \lambda^n$, which is the same answer as in Exercise 7.139(a).

7.172 a) Suppose that X and Y have the same distribution, then $p_X(x) = p_Y(x)$ for all x, then, by Exercise 7.52(b), for all $t \in \mathcal{R}$,

$$P_X(t) = \sum_{n=0}^{\infty} t^n p_X(n) = \sum_{n=0}^{\infty} t^n p_Y(n) = P_Y(t).$$

b) First note that if X is an arbitrary nonnegative-integer valued random variable with PGF $P_X(t) = \mathcal{E}(t^X)$, then, by Exercise 7.52(b), $P_X(t) = \sum_{n=0}^{\infty} p_X(n)\, t^n$, implying that for all $n \in \mathcal{Z}_+$,

$$P_X^{(n)}(t) = \sum_{k=n}^{\infty} p_X(k)k(k-1)\ldots(k-n+1)t^{k-n}, \;\; \text{and} \;\; P_X^{(n)}(0) = n!p_X(n).$$

Thus,

$$p_X(x) = \begin{cases} \dfrac{P_X^{(x)}(0)}{x!}, & \text{if } x \in \{0, 1, 2, \ldots\}, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, if X and Y are nonnegative-integer valued random variables with the same PGF $P(t)$, then $p_X(x) = \dfrac{P^{(x)}(0)}{x!} = p_Y(x)$ for all $x \in \{0, 1, \ldots\}$ and $p_X(x) = 0 = p_Y(x)$ for all $x \notin \{0, 1, \ldots\}$. Thus, X and Y have the same distribution.
c) By (b) and Exercise 7.53(b), if $P_X(t) = e^{4(t-1)}$, then X is a Poisson random variable with parameter $\lambda = 4$.

7.173 a) $P_{X+Y}(t) = \mathcal{E}(t^{X+Y}) = \mathcal{E}(t^X t^Y) = \mathcal{E}(t^X)\mathcal{E}(t^Y) = P_X(t)P_Y(t)$, where we used independence of X and Y and Proposition 6.9. Thus, the PGF of the sum of independent random variables is equal to the product of individual PGFs.
b) $P_{X_1+\ldots+X_m}(t) = \mathcal{E}\left(t^{X_1+\cdots+X_m}\right) = \mathcal{E}(t^{X_1} \ldots t^{X_m}) = \mathcal{E}(t^{X_1}) \ldots \mathcal{E}(t^{X_m}) = P_{X_1}(t) \ldots P_{X_m}(t)$, where we used independence of $X_1, \ldots, X_m$ and Proposition 6.13.
c) From (b), since $P_{X_i}(t) = P_X(t)$ for $i = 1, \ldots, m$, we have that $P_{X_1+\ldots+X_m}(t) = (P_X(t))^m$.

7.174 From Exercise 7.53 and Exercise 7.172(b) we have that the PGF of X satisfies:

$$P_X(t) = (pt + (1-p))^n, \quad \text{if and only if} \quad X \sim \mathcal{B}(n, p);$$

$$P_X(t) = e^{\lambda(t-1)}, \quad \text{if and only if} \quad X \sim \mathcal{P}(\lambda);$$

$$P_X(t) = \left(\frac{pt}{1 - t + pt}\right)^r, \quad \text{if and only if} \quad X \sim \mathcal{NB}(r, p).$$

Let $X_1, \ldots, X_m$ be independent random variables. Then
a) If $X_j \sim \mathcal{B}(n_j, p)$ for $1 \le j \le m$, then, by Exercise 7.173(b),

$$P_{X_1+\cdots+X_m}(t) = \prod_{i=1}^{m} P_{X_i}(t) = \prod_{i=1}^{m} (pt + (1-p))^{n_i} = (pt + (1-p))^{n_1+\cdots+n_m},$$

implying that $X_1 + \cdots + X_m \sim \mathcal{B}(n_1 + \cdots + n_m, p)$.
b) If $X_j \sim \mathcal{P}(\lambda_j)$ for $1 \le j \le m$, then, by Exercise 7.173(b),

$$P_{X_1+\cdots+X_m}(t) = \prod_{i=1}^{m} P_{X_i}(t) = \prod_{i=1}^{m} e^{\lambda_i(t-1)} = e^{(\lambda_1+\cdots+\lambda_m)(t-1)},$$

implying that $X_1 + \cdots + X_m \sim \mathcal{P}(\lambda_1 + \cdots + \lambda_m)$.
c) If $X_j \sim \mathcal{NB}(r_j, p)$ for $1 \le j \le m$, then, by Exercise 7.173(b),

$$P_{X_1+\cdots+X_m}(t) = \prod_{i=1}^{m} P_{X_i}(t) = \prod_{i=1}^{m} \left(\frac{pt}{1-t+pt}\right)^{r_i} = \left(\frac{pt}{1-t+pt}\right)^{r_1+\cdots+r_m},$$

implying that $X_1 + \cdots + X_m \sim \mathcal{NB}(r_1 + \cdots + r_m, p)$.

7.175 a) Since $N, X_1, X_2, \ldots$ are independent, then $S_n = X_1 + \cdots + X_n$ and N are independent for all n, implying that

$$\mathcal{E}\left(t^{S_N} \mid N = n\right) = \mathcal{E}\left(t^{S_n} \mid N = n\right) = \mathcal{E}(t^{S_n}) = (P_X(t))^n,$$

where the last equality holds due to Exercise 7.173(c). Thus, by the law of total expectation, we have that

$$P_{S_N}(t) = \mathcal{E}\left(t^{S_N}\right) = \sum_{n=0}^{\infty} \mathcal{E}\left(t^{S_N} \mid N = n\right) p_N(n) = \sum_{n=0}^{\infty} (P_X(t))^n \, p_N(n) = \sum_{n=0}^{\infty} (P_X(t))^n \cdot \frac{P_N^{(n)}(0)}{n!},$$

where the last equality holds due to the argument given in solution to Exercise 7.172(b).
b) By (a) and Exercise 7.171(a), we have that

$$\mathcal{E}(S_N) = P'_{S_N}(1) = \sum_{n=1}^{\infty} n(P_X(1))^{n-1}P'_X(1)p_N(n) = P'_X(1)\sum_{n=1}^{\infty} np_N(n) = \mathcal{E}(X)\mathcal{E}(N) = \mu\mathcal{E}(N),$$

which is the same result as in equation (7.63) on page 384 in the textbook.
Moreover,

$$\begin{aligned}
\mathcal{E}(S_N(S_N-1)) &= P''_{S_N}(1) = \sum_{n=1}^{\infty}\left[n(n-1)(P'_X(1))^2 + nP''_X(1)\right]p_N(n) \\
&= \sum_{n=1}^{\infty}\left[n(n-1)(\mathcal{E}(X))^2 + n\mathcal{E}(X(X-1))\right]p_N(n) \\
&= \sum_{n=1}^{\infty}\left[n\sigma^2 + n^2\mu^2 - n\mu\right]p_N(n) = (\sigma^2-\mu)\mathcal{E}(N) + \mu^2\mathcal{E}(N^2).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Var}(S_N) &= \mathcal{E}(S_N(S_N-1)) + \mathcal{E}(S_N) - (\mathcal{E}(S_N))^2 \\
&= (\sigma^2-\mu)\mathcal{E}(N) + \mu^2\mathcal{E}(N^2) + \mu\mathcal{E}(N) - \mu^2(\mathcal{E}(N))^2 \\
&= \sigma^2\mathcal{E}(N) + \mu^2\mathrm{Var}(N),
\end{aligned}$$

which is the same result as in equation (7.68) on page 387 in the textbook.
c) Consider Bernoulli trials with success probability p. Let $X_j = 1_{\{j\text{th trial is a success}\}}$ for each $j \in \mathcal{N}$. Also, let N be independent of the X_j s and $N \sim \mathcal{P}(\lambda)$. Then $P_{X_j}(t) = pt + 1 - p$. Thus, by Exercise 7.175(a), we have that

$$P_{S_N}(t) = \sum_{n=0}^{\infty}(pt+1-p)^n \cdot \frac{e^{-\lambda}\lambda^n}{n!} = e^{p\lambda(t-1)},$$

which is the PGF of a Poisson random variable with parameter $p\lambda$. Thus, by Exercise 7.172(b), $S_N \sim \mathcal{P}(p\lambda)$, which is the same conclusion as in Example 6.25.

CHAPTER EIGHT

Instructor's Solutions Manual

Anna Amirdjanova

University of Michigan, Ann Arbor

Neil A. Weiss

Arizona State University

FOR

A Course In

Probability

Neil A. Weiss

Arizona State University

PEARSON Addison Wesley

Boston San Francisco New York

London Toronto Sydney Tokyo Singapore Madrid

Mexico City Munich Paris Cape Town Hong Kong Montreal

Publisher: Greg Tobin
Editor-in-Chief: Deirdre Lynch
Associate Editor: Sara Oliver Gordus
Editorial Assistant: Christina Lepre
Production Coordinator: Kayla Smith-Tarbox
Senior Author Support/Technology Specialist: Joe Vetere
Compositor: Anna Amirdjanova and Neil A. Weiss
Accuracy Checker: Delray Schultz
Proofreader: Carol A. Weiss

Chapter 8

Continuous Random Variables and Their Distributions

8.1 Introducing Continuous Random Variables

Basic Exercises

8.1 Answers will vary, but see the argument in the last paragraph of the solution to Exercise 5.21(a), which is based on the frequentist interpretation of probability.

8.2 The sample space can be taken as $\Omega = \{ (x, y, z) : x^2 + y^2 + z^2 < r^2 \}$. As a point is chosen at random, use of a geometric probability model is reasonable. So, for each event E, $P(E) = |E| \big/ (4\pi r^3/3)$, where $|E|$ denotes the volume (in the extended sense) of the set E. The random variable Z is defined on Ω by $Z(x, y, z) = \sqrt{x^2 + y^2 + z^2}$. Clearly, the range of Z is the interval $[0, r)$ and, hence, $P(Z = w) = 0$ if $w \notin [0, r)$. Now let $w \in [0, r)$. The event $\{Z = w\}$ is that the point chosen lies on the surface $S = \{ (x, y, z) : x^2 + y^2 + z^2 = w^2 \}$. Because the volume of the surface of a sphere (boundary of a sphere) is 0, $P(Z = w) = P(S) = |S| \big/ (4\pi r^3/3) = 0$.

Alternatively, for $w \in [0, r)$, let $A_n = \{w \le Z \le w + 1/n\}$ for each $n \in \mathcal{N}$. We note that A_n is the event that the point chosen is between w units and $w + 1/n$ units from the center of the sphere. Consequently, for sufficiently large n,

$$P(A_n) = \frac{|A_n|}{\frac{4}{3}\pi r^3} = \frac{\frac{4}{3}\pi (w + 1/n)^3 - \frac{4}{3}\pi w^3}{\frac{4}{3}\pi r^3} = \frac{1}{r^3}\left(\frac{3w^2}{n} + \frac{3w}{n^2} + \frac{1}{n^3}\right).$$

Because $\{Z = w\} \subset A_n$, we see that

$$P(Z = w) \le \frac{1}{r^3}\left(\frac{3w^2}{n} + \frac{3w}{n^2} + \frac{1}{n^3}\right)$$

for sufficiently large n. Letting $n \to \infty$ yields $P(Z = w) = 0$.

8.3 The sample space can be taken as $\Omega = \{ \theta : -\pi/2 < \theta < \pi/2 \}$. As an angle is chosen at random, use of a geometric probability model is reasonable. So, for each event E, $P(E) = |E|/\pi$, where $|E|$ denotes the length (in the extended sense) of the set E. Let Θ denote the angle chosen. Then $X = \tan \Theta$, and we have

$$P(X = x) = P(\tan \Theta = x) = P(\Theta = \arctan x)) = P(\{\arctan x\}) = |\{\arctan x\}|/\pi = 0/\pi = 0,$$

for each $x \in \mathcal{R}$. Hence, X is a continuous random variable.

8.4

a) For each $y \in \mathcal{R}$, $P(a + bX = y) = P\big(X = (y - a)/b\big) = 0$, because X is a continuous random variable. Thus, $a + bX$ is a continuous random variable.

b) We have $P\left(X^2 = 0\right) = P(X = 0) = 0$. If $y < 0$, $P\left(X^2 = y\right) = P(\emptyset) = 0$. If $y > 0$,

$$P\big(X^2 = y\big) = P\big(X \in \{-\sqrt{y}, \sqrt{y}\}\big) = P\big(X = -\sqrt{y}\big) + P\big(X = \sqrt{y}\big) = 0 + 0.$$

Thus, X^2 is a continuous random variable.
c) If $|y| > 1$, $P(\sin X = y) = P(\emptyset) = 0$. If $y = 0$,

$$P(\sin X = y) = P(\sin X = 0) = P\big(X \in \{\, n\pi : n \in \mathcal{Z}\,\}\big) = \sum_{n=-\infty}^{\infty} P(X = n\pi) = \sum_{n=-\infty}^{\infty} 0 = 0.$$

Similarly, $P(\sin X = y) = 0$ if $0 < |y| \le 1$. Thus, $\sin X$ is a continuous random variable.

8.5 The sample space can be taken as the interior of the triangle. As a point is chosen at random, use of a geometric probability model is reasonable. Hence, for each event E, $P(E) = |E|\big/(bh/2)$, where $|E|$ denotes the area (in the extended sense) of the set E. The range of Y is the interval $(0, h)$. If $y \notin (0, h)$, then $P(Y = y) = 0$. If $y \in (0, h)$, let L_y denote the intersection of the interior of the triangle with the line y units from the base of the triangle. Then, because the area of a line segment is 0, we have $P(Y = y) = |L_y|\big/(bh/2) = 0$.

8.6 For each $x \in \mathcal{R}$, $P(X = x) = \int_x^x f(t)\,dt = 0$. Thus, X is a continuous random variable.

Theory Exercises

8.7
a) Let $\{x_n\}_{n=1}^{\infty}$ be an enumeration of K. Because the events $\{X = x_1\}$, $\{X = x_2\}$, ... are mutually exclusive, we have

$$P(X \in K) = P\left(X \in \bigcup_{n=1}^{\infty}\{x_n\}\right) = P\left(\bigcup_{n=1}^{\infty}\{X = x_n\}\right) = \sum_{n=1}^{\infty} P(X = x_n) = 0,$$

where the last equation follows from the fact that X is a continuous random variable.
b) Let X be a continuous random variable. If X were also a discrete random variable, then, by Definition 5.2 on page 178, there would be a countable set K such that $P(X \in K) = 1$. But this last equation is impossible in view of part (a).

Advanced Exercises

8.8 From Example 8.1 on page 402, we know X is a continuous random variable. Because the rational numbers in $(0, 1)$, $\mathcal{Q} \cap (0, 1)$, form a countable set, Exercise 8.7(a) shows that $P\big(X \in \mathcal{Q} \cap (0, 1)\big) = 0$; that is, the probability that the number chosen is rational equals 0.

8.9 Set $Y = p(X)$ and let y be in the range of Y. We have

$$P(Y = y) = P\big(p(X) = y\big) = P\big(X \in p^{-1}(\{y\})\big).$$

Now let $A = p^{-1}(\{y\})$ and set $q(x) = p(x) - y$. Then

$$A = p^{-1}(\{y\}) = \{\, x \in \mathcal{R} : p(x) = y \,\} = \{\, x \in \mathcal{R} : q(x) = 0 \,\}.$$

Because p is a polynomial of positive degree, so is q. Hence $A = \{\, x \in \mathcal{R} : q(x) = 0 \,\}$ is finite. Referring now to Exercise 8.7(a), we conclude that $P(Y = y) = P(X \in A) = 0$. Hence Y is a continuous random variable.

8.10 Set $Y = g(X)$ and let y be in the range of Y. Then

$$P(Y = y) = P\big(g(X) = y\big) = P\big(X \in g^{-1}(\{y\})\big).$$

Because g is one-to-one on the range of X, $g^{-1}(\{y\})$ consists of only one number, say, x. Thus, as X is a continuous random variable, $P(Y = y) = P(X = x) = 0$. So, Y is a continuous random variable.

8.11
a) No. For instance, let X be any random variable whose range contains both -1 and 1, and let $g(x) = x^2$. Then g is not one-to-one on the range of X but, by Exercise 8.4, $g(X)$ is a continuous random variable.
b) No. For instance, let X be any continuous random variable and let $g(x) = c$, a constant function. Then $g(X)$ is not a continuous random variable; rather, it is a discrete random variable with PMF given by $p_{g(X)}(y) = 1$ if $y = c$, and $p_{g(X)}(y) = 0$ otherwise.

8.12
a) The range of Y is $(0, 0.75]$, which is a continuum of real numbers.
b) We have $P(Y = 0.75) = P(X \geq 0.75) = 0.25 \neq 0$. Thus, Y is not a continuous random variable.
c) From part (a), $P(Y = y) = 0$ if $y \notin (0, 0.75]$. For $y \in (0, 0.75)$,

$$P(Y = y) = P\big(\min\{X, 0.75\} = y\big) = P(X = y) = 0,$$

because X is a continuous random variable. Thus, $\sum_y P(Y = y) = P(Y = 0.75) = 0.25 < 1$. Hence, from Exercise 5.17 on page 183, Y is not a discrete random variable.

8.13
a) First note that $(-\infty, b] = (-\infty, a] \cup (a, b]$ and that the two sets on the right are disjoint. Therefore,

$$\begin{aligned} F_X(b) = P(X \leq b) &= P\big(X \in (-\infty, b]\big) = P\big(X \in (-\infty, a]\big) + P\big(X \in (a, b]\big) \\ &= P(X \leq a) + P(a < X \leq b) = F_X(a) + P(a < X \leq b). \end{aligned}$$

The required result now follows easily.
b) Let $x \in \mathcal{R}$ and let $\{x_n\}_{n=1}^{\infty}$ be an increasing sequence of real numbers that converges to x. From the domination principle and part (a),

$$0 \leq P(X = x) \leq P(x_n < X \leq x) = F_X(x) - F_X(x_n).$$

Letting $n \to \infty$ and applying the assumed continuity of F_X, we get that $P(X = x) = 0$, as required.

8.14
a) For $k \in \mathcal{N}$, we have $k/n > a$ if and only if $k > \lfloor na \rfloor$ and $k/n \leq b$ if and only if $k \leq \lfloor nb \rfloor$. Thus, for $0 < a < b < 1$,

$$P(a < X_n \leq b) = \frac{N(\{\,k : a < k/n \leq b\,\})}{n} = \frac{N(\{\,k : \lfloor na \rfloor < k \leq \lfloor nb \rfloor\,\})}{n} = \frac{\lfloor nb \rfloor - \lfloor na \rfloor}{n}.$$

b) For $x \in \mathcal{R}$, we have $nx - 1 < \lfloor nx \rfloor \leq nx$ and, hence,

$$x = \lim_{n\to\infty} \frac{nx - 1}{n} \leq \lim_{n\to\infty} \frac{\lfloor nx \rfloor}{n} \leq \lim_{n\to\infty} \frac{nx}{n} = x.$$

This result shows that $\lim_{n\to\infty} \lfloor nx \rfloor / n = x$. Referring back to part (a), we can now conclude that $\lim_{n\to\infty} P(a < X_n \leq b) = b - a$ for $0 < a < b < 1$.

c) Consider the random variable X from Example 8.1. For $0 < x < 1$, $P(X \le x) = |(0, x]| = x$. Consequently, from Exercise 8.13(a), $P(a < X \le b) = b - a$ for $0 < a < b < 1$. Referring now to the result of part (b), we deduce that the limiting distribution of the discrete random variables $X_1, X_2, \ldots$ is that of the continuous random variable X from Example 8.1.

8.2 Cumulative Distribution Functions

Basic Exercises

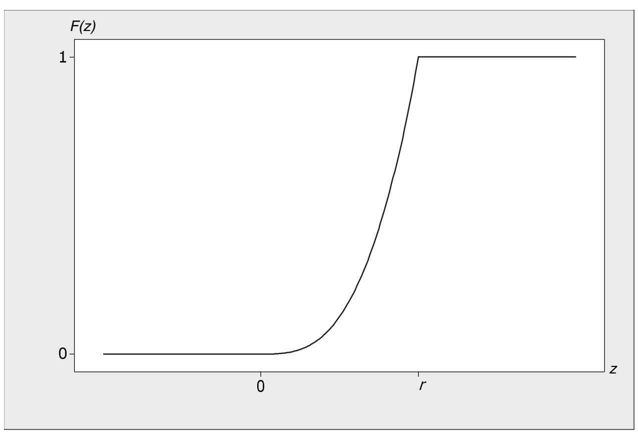
8.15 You may want to refer to the solution to Exercise 8.2.
a) Clearly, $F_Z(z) = 0$ if $z < 0$ and $F_Z(z) = 1$ if $z \ge r$. For $0 \le z < r$,

$$F_Z(z) = \frac{\frac{4}{3}\pi z^3}{\frac{4}{3}\pi r^3} = \frac{z^3}{r^3}.$$

Thus,

$$F_Z(z) = \begin{cases} 0, & \text{if } z < 0; \\ z^3/r^3, & \text{if } 0 \le z < r; \\ 1, & \text{if } z \ge r. \end{cases}$$

Note: In the following graph, we use $F(z)$ instead of $F_Z(z)$.

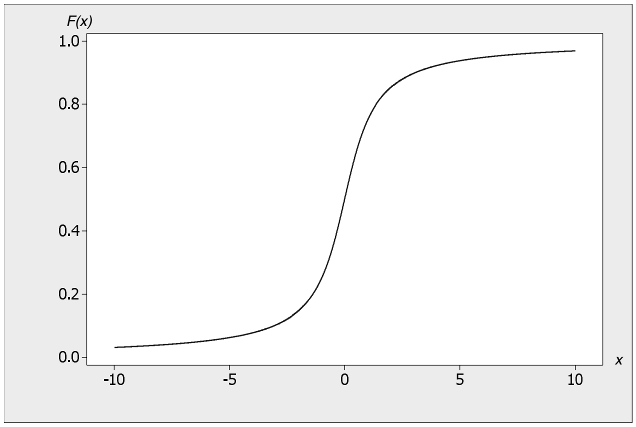


b) Both the graph in part (a) and the display above the graph make clear that the CDF of Z satisfies the four properties in Proposition 8.1.
c) From part (a), we see that the CDF of Z is everywhere continuous. Thus, Z is a continuous random variable.

8.16 You may want to refer to the solution to Exercise 8.3.
a) For each $x \in \mathcal{R}$, event $\{X \le x\}$ occurs if and only if the angle chosen is between $-\pi/2$ and $\arctan x$. Because the angle is chosen at random from the interval $(-\pi/2, \pi/2)$, a classical probability model is appropriate. Thus, for all $x \in \mathcal{R}$,

$$F_X(x) = P(X \le x) = \frac{|\{X \le x\}|}{\pi} = \frac{\arctan x - (-\pi/2)}{\pi} = \frac{1}{2} + \frac{1}{\pi}\arctan x.$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



b) Both the graph in part (a) and the display above the graph make clear that the CDF of X satisfies the four properties in Proposition 8.1.
c) From calculus, we know that the arctan function is everywhere continuous. From that fact, it follows easily that the CDF of X is everywhere continuous. Thus, X is a continuous random variable.

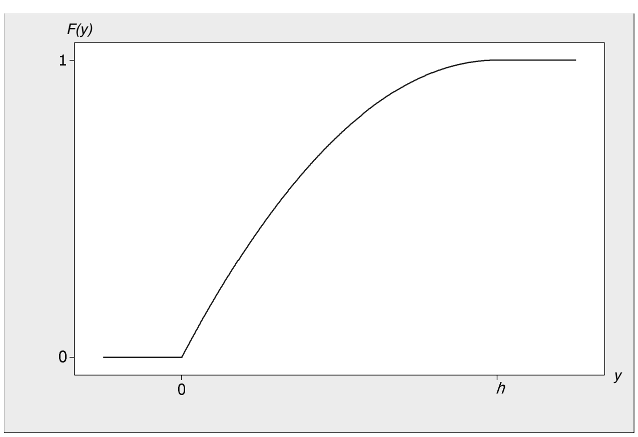
8.17 You may want to refer to the solution to Exercise 8.5.
a) Clearly, $F_Y(y) = 0$ if $y < 0$ and $F_Y(y) = 1$ if $y \geq h$. Now let $0 \leq y < h$. Using similar triangles, we find that

$$F_Y(y) = P(Y \leq y) = 1 - P(Y > y) = 1 - \frac{|\{Y > y\}|}{\frac{1}{2}bh} = 1 - \frac{\frac{1}{2h}b(h-y)^2}{\frac{1}{2}bh} = \frac{2hy - y^2}{h^2}.$$

Thus,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ (2hy - y^2)/h^2, & \text{if } 0 \leq y < h; \\ 1, & \text{if } y \geq h. \end{cases}$$

Note: In the following graph, we use $F(y)$ instead of $F_Y(y)$.

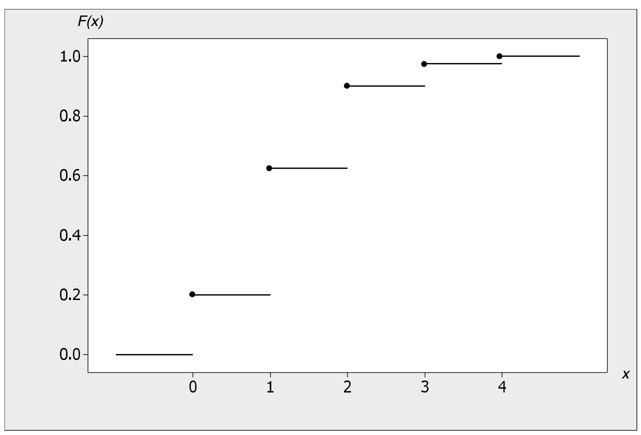


b) Both the graph in part (a) and the display above the graph make clear that the CDF of Y satisfies the four properties in Proposition 8.1.
c) From part (a), we see that the CDF of Y is everywhere continuous. Thus, Y is a continuous random variable.

8.18
a) Referring to Table 5.5 on page 186 and proceeding exactly as in Example 8.3 on page 407, we find that

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 0.2, & \text{if } 0 \le x < 1; \\ 0.625, & \text{if } 1 \le x < 2; \\ 0.9, & \text{if } 2 \le x < 3; \\ 0.975, & \text{if } 3 \le x < 4; \\ 1, & \text{if } x \ge 4. \end{cases}$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.

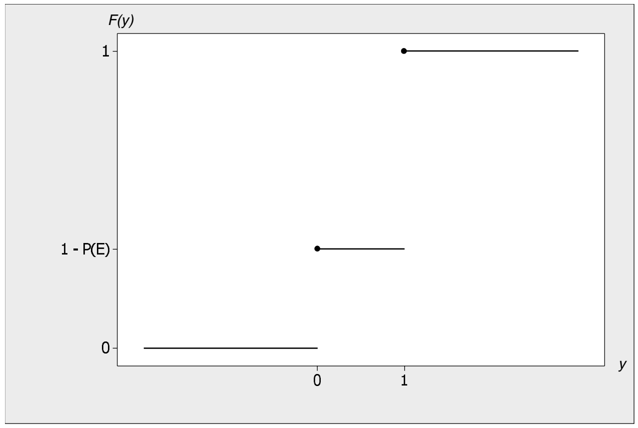


b) Both the graph in part (a) and the display above the graph make clear that the CDF of X satisfies the four properties in Proposition 8.1.
c) From part (a), we see that the CDF of X is not everywhere continuous. Thus, X is not a continuous random variable. It is, in fact, a discrete random variable.

8.19
a) We know that the PDF of Y is given by $p_Y(0) = 1 - P(E)$, $p_Y(1) = P(E)$, and $p_Y(y) = 0$ otherwise. From this we conclude, that the CDF of Y is as follows:

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ 1 - P(E), & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

Note: In the following graph, we use $F(y)$ instead of $F_Y(y)$.

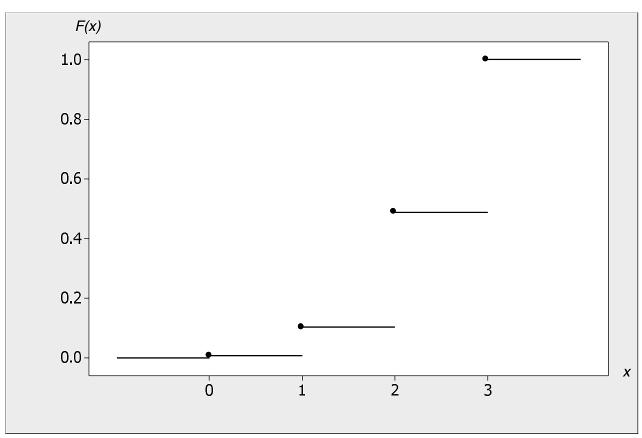


b) Both the graph in part (a) and the display above the graph make clear that the CDF of Y satisfies the four properties in Proposition 8.1.
c) From part (a), we see that the CDF of Y is not everywhere continuous. Thus, Y is not a continuous random variable. It is, in fact, a discrete random variable.

8.20
a) In Example 5.13 on page 202, we found that $X \sim \mathcal{B}(3, 0.8)$. Its PMF is given in Table 5.13 on page 203. From that table we find that

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 0.008, & \text{if } 0 \leq x < 1; \\ 0.104, & \text{if } 1 \leq x < 2; \\ 0.488, & \text{if } 2 \leq x < 3; \\ 1, & \text{if } x \geq 3. \end{cases}$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



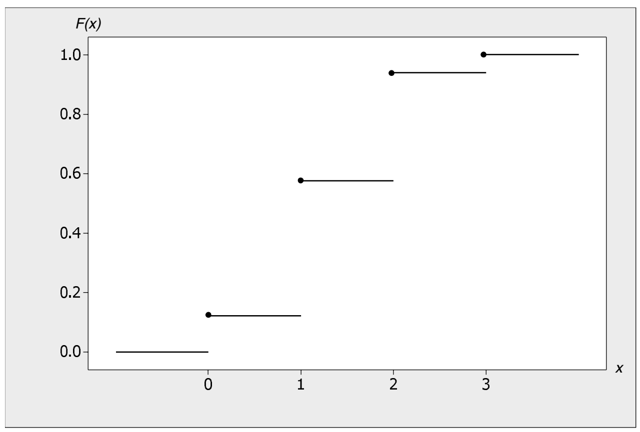
b) Both the graph in part (a) and the display above the graph make clear that the CDF of X satisfies the four properties in Proposition 8.1.
c) From part (a), we see that the CDF of X is not everywhere continuous. Thus, X is not a continuous random variable. It is, in fact, a discrete random variable.

8.21

a) We note that $X \sim \mathcal{H}(11, 3, 5/11)$. Applying the formula for the PDF of a hypergeometric random variable, as given in Equation (5.12) on page 210, we find that $p_X(0) = 4/33$, $p_X(1) = 15/33$, $p_X(2) = 12/33$, $p_X(3) = 2/33$, and $p_X(x) = 0$ otherwise. From this information, we get the CDF of X:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 4/33, & \text{if } 0 \le x < 1; \\ 19/33, & \text{if } 1 \le x < 2; \\ 31/33, & \text{if } 2 \le x < 3; \\ 1, & \text{if } x \ge 3. \end{cases}$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



b) Both the graph in part (a) and the display above the graph make clear that the CDF of X satisfies the four properties in Proposition 8.1.

c) From part (a), we see that the CDF of X is not everywhere continuous. Thus, X is not a continuous random variable. It is, in fact, a discrete random variable.

8.22

a) By the FPF for discrete random variables, $F_X(x) = P(X \le x) = \sum_{t \le x} p_X(t)$.

b) From Equation (8.8) on page 413, we find that $p_X(x) = P(X = x) = F_X(x) - F_X(x-)$.

8.23

a) Clearly, $F_X(x) = 0$ if $x < 1$ and $F_X(x) = 1$ if $x \ge N$. For convenience, set $S = \{1, 2, \dots, N\}$. For $1 \le x < N$,

$$F_X(x) = P(X \le x) = P\big(X \in (-\infty, x]\big) = \frac{N\big((-\infty, x] \cap S\big)}{N(S)} = \frac{\lfloor x \rfloor}{N}.$$

The graph of F_X is a right-continuous step function, which is 0 for $x < 1$ and is 1 for $x \ge N$. It has jumps of magnitude $1/N$ at each of the first N positive integers.

b) Referring to part (a), we have, for $x = 1, 2, \dots, N$,

$$p_X(x) = P(X = x) = F_X(x) - F_X(x-) = \frac{x}{N} - \frac{x-1}{N} = \frac{1}{N}.$$

For all other values of x, we have $p_X(x) = 0$.

8.24
a) Clearly, $F_X(x) = 0$ if $x < 1$. Applying the FPF for discrete random variables and a geometric series formula, we get that, for $x \geq 1$,

$$F_X(x) = P(X \leq x) = \sum_{t \leq x} p_X(t) = \sum_{k=1}^{\lfloor x \rfloor} p(1-p)^{k-1} = p \cdot \frac{1-(1-p)^{\lfloor x \rfloor}}{1-(1-p)} = 1-(1-p)^{\lfloor x \rfloor}.$$

b) Clearly, $F_X(x) = 0$ if $x < 1$. Applying Proposition 5.10 on page 231, we get that, for $x \geq 1$,

$$F_X(x) = P(X \leq x) = 1 - P(X > x) = 1 - P(X > \lfloor x \rfloor) = 1-(1-p)^{\lfloor x \rfloor}.$$

c) The graph of F_X is a right-continuous step function, which is 0 for $x < 1$. At each positive integer n, it has a jump of magnitude $p(1-p)^{n-1}$.
d) Referring to part (a), we have, for $x \in \mathcal{N}$,

$$\begin{aligned} p_X(x) = P(X = x) = F_X(x) - F_X(x-) &= \bigl(1-(1-p)^x\bigr) - \bigl(1-(1-p)^{x-1}\bigr) \\ &= (1-p)^{x-1}\bigl(1-(1-p)\bigr) = p(1-p)^{x-1}. \end{aligned}$$

For all other values of x, we have $p_X(x) = 0$.

8.25 Refer to Proposition 8.2 on page 412 and the formula for the CDF of W at the top of page 408.
a) We have

$$\begin{aligned} P(1 < W \leq 2) &= F_W(2) - F_W(1) = 0.875 - 0.5 = 0.375. \\ P(1 < W < 2) &= F_W(2-) - F_W(1) = 0.5 - 0.5 = 0. \\ P(1 \leq W < 2) &= F_W(2-) - F_W(1-) = 0.5 - 0.125 = 0.375. \\ P(1 \leq W \leq 2) &= F_W(2) - F_W(1-) = 0.875 - 0.125 = 0.75. \end{aligned}$$

b) We have

$$\begin{aligned} P(0.5 < W \leq 2) &= F_W(2) - F_W(0.5) = 0.875 - 0.125 = 0.75. \\ P(0.5 < W < 2) &= F_W(2-) - F_W(0.5) = 0.5 - 0.125 = 0.375. \\ P(0.5 \leq W < 2) &= F_W(2-) - F_W(0.5-) = 0.5 - 0.125 = 0.375. \\ P(0.5 \leq W \leq 2) &= F_W(2) - F_W(0.5-) = 0.875 - 0.125 = 0.75. \end{aligned}$$

c) We have

$$\begin{aligned} P(1 < W \leq 2.75) &= F_W(2.75) - F_W(1) = 0.875 - 0.5 = 0.375. \\ P(1 < W < 2.75) &= F_W(2.75-) - F_W(1) = 0.875 - 0.5 = 0.375. \\ P(1 \leq W < 2.75) &= F_W(2.75-) - F_W(1-) = 0.875 - 0.125 = 0.75. \\ P(1 \leq W \leq 2.75) &= F_W(2.75) - F_W(1-) = 0.875 - 0.125 = 0.75. \end{aligned}$$

d) We have

$$\begin{aligned} P(0.5 < W \leq 2.75) &= F_W(2.75) - F_W(0.5) = 0.875 - 0.125 = 0.75. \\ P(0.5 < W < 2.75) &= F_W(2.75-) - F_W(0.5) = 0.875 - 0.125 = 0.75. \\ P(0.5 \leq W < 2.75) &= F_W(2.75-) - F_W(0.5-) = 0.875 - 0.125 = 0.75. \\ P(0.5 \leq W \leq 2.75) &= F_W(2.75) - F_W(0.5-) = 0.875 - 0.125 = 0.75. \end{aligned}$$

e) For part (a),

$$P(1 < W \leq 2) = p_W(2) = 0.375.$$
$$P(1 < W < 2) = 0.$$
$$P(1 \leq W < 2) = p_W(1) = 0.375.$$
$$P(1 \leq W \leq 2) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$

For part (b),

$$P(0.5 < W \leq 2) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$
$$P(0.5 < W < 2) = p_W(1) = 0.375.$$
$$P(0.5 \leq W < 2) = p_W(1) = 0.375.$$
$$P(0.5 \leq W \leq 2) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$

For part (c),

$$P(1 < W \leq 2.75) = p_W(2) = 0.375.$$
$$P(1 < W < 2.75) = p_W(2) = 0.375.$$
$$P(1 \leq W < 2.75) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$
$$P(1 \leq W \leq 2.75) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$

For part (d),

$$P(0.5 < W \leq 2.75) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$
$$P(0.5 < W < 2.75) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$
$$P(0.5 \leq W < 2.75) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$
$$P(0.5 \leq W \leq 2.75) = p_W(1) + p_W(2) = 0.375 + 0.375 = 0.75.$$

8.26 Note that, for each a and b, all four probabilities are equal to $F_X(b) - F_X(a)$.
a) $F_X(0.8) - F_X(0.2) = 0.8 - 0.2 = 0.6$.
b) $F_X(0.8) - F_X(0) = 0.8 - 0 = 0.8$.
c) $F_X(1.5) - F_X(0.2) = 1 - 0.2 = 0.8$.
d) $F_X(1.5) - F_X(-1) = 1 - 0 = 1$.
e) $F_X(-1) - F_X(-2) = 0 - 0 = 0$.
f) $F_X(2) - F_X(1) = 1 - 1 = 0$.
g) Because X is a continuous random variable.

8.27 Note that, for each a and b, all four probabilities are equal to $F_Z(b) - F_Z(a)$.
a) $F_Z(0.8) - F_Z(0.2) = (0.8)^2 - (0.2)^2 = 0.6$.
b) $F_Z(0.8) - F_Z(0) = (0.8)^2 - 0 = 0.64$.
c) $F_Z(1.5) - F_Z(0.2) = 1 - (0.2)^2 = 0.96$.
d) $F_Z(1.5) - F_Z(-1) = 1 - 0 = 1$.
e) $F_Z(-1) - F_Z(-2) = 0 - 0 = 0$.
f) $F_Z(2) - F_Z(1) = 1 - 1 = 0$.
g) Because Z is a continuous random variable.

8.28 Refer to Proposition 8.2 on page 412 and the formula for the CDF of Y near the bottom of page 410.
a) All four probabilities equal $F_Y(0.8) - F_Y(0.2) = 1 - 0.2 = 0.8$, because F_Y is continuous at $y = 0.8$ and $y = 0.2$.

b) We have

$$
\begin{aligned}
P(0.2 < Y \le 0.75) &= F_Y(0.75) - F_Y(0.2) = 1 - 0.2 = 0.8. \\
P(0.2 < Y < 0.75) &= F_Y(0.75-) - F_Y(0.2) = 0.75 - 0.2 = 0.55. \\
P(0.2 \le Y < 0.75) &= F_Y(0.75-) - F_Y(0.2-) = 0.75 - 0.2 = 0.55. \\
P(0.2 \le Y \le 0.75) &= F_Y(0.75) - F_Y(0.2-) = 1 - 0.2 = 0.8.
\end{aligned}
$$

c) All four probabilities equal $F_Y(1.5) - F_Y(-1) = 1 - 0 = 1$, because F_Y is continuous at $y = 1.5$ and $y = -1$.
d) We have

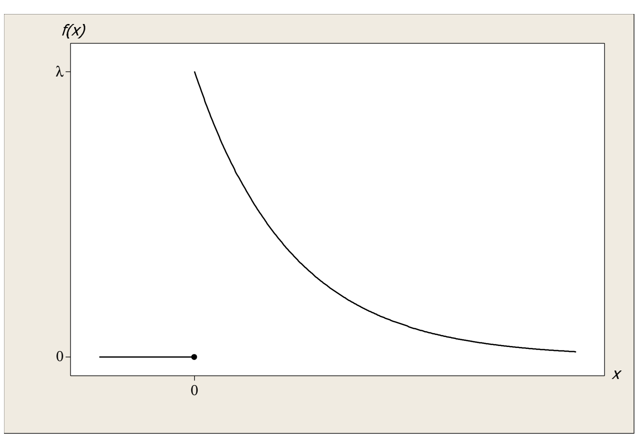
$$
\begin{aligned}
P(-1 < Y \le 0.75) &= F_Y(0.75) - F_Y(-1) = 1 - 0 = 1. \\
P(-1 < Y < 0.75) &= F_Y(0.75-) - F_Y(-1) = 0.75 - 0 = 0.75. \\
P(-1 \le Y < 0.75) &= F_Y(0.75-) - F_Y(-1-) = 0.75 - 0 = 0.75. \\
P(-1 \le Y \le 0.75) &= F_Y(0.75) - F_Y(-1-) = 1 - 0 = 1.
\end{aligned}
$$

e) All four probabilities equal $F_Y(-1) - F_Y(-2) = 0 - 0 = 0$, because F_Y is continuous at $y = -1$ and $y = -2$.
f) We have

$$
\begin{aligned}
P(0.75 < Y \le 2) &= F_Y(2) - F_Y(0.75) = 1 - 1 = 0. \\
P(0.75 < Y < 2) &= F_Y(2-) - F_Y(0.75) = 1 - 1 = 0. \\
P(0.75 \le Y < 2) &= F_Y(2-) - F_Y(0.75-) = 1 - 0.75 = 0.25. \\
P(0.75 \le Y \le 2) &= F_Y(2) - F_Y(0.75-) = 1 - 0.75 = 0.25.
\end{aligned}
$$

8.29
a) We have $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$ for all $x \in \mathcal{R}$.
b) From the first fundamental theorem of calculus, $F_X'(x) = f(x)$ for all x at which f is continuous.
c) All four probabilities equal $\int_a^b f(x)\,dx$.
d) Because, from Exercise 8.6, X is a continuous random variable or, equivalently, because F_X is an everywhere continuous function.

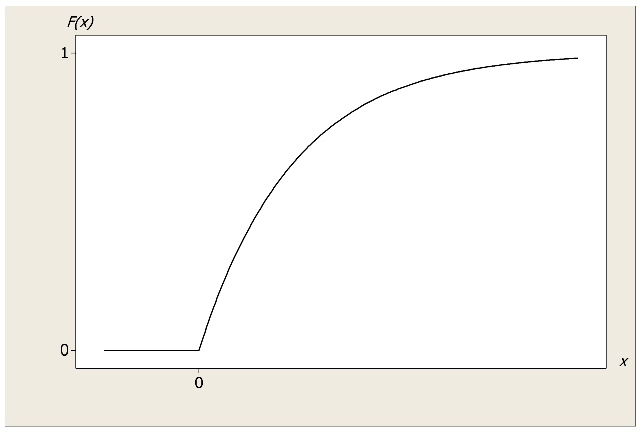
8.30
a) The graph of f is as follows:



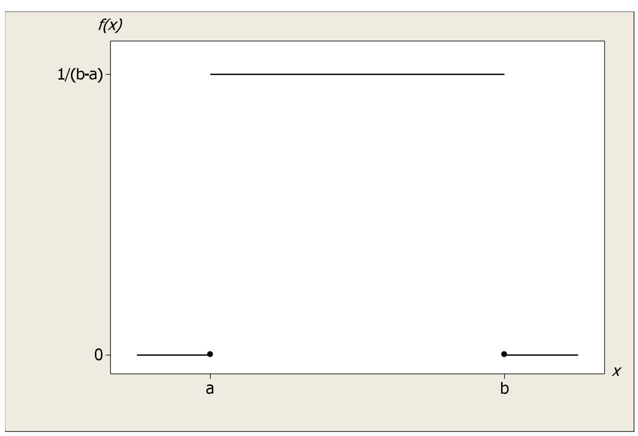
Clearly, $F_X(x) = 0$ if $x < 0$. If $x \geq 0$,

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} \lambda e^{-\lambda t}\,dt = 1 - e^{-\lambda x}.$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



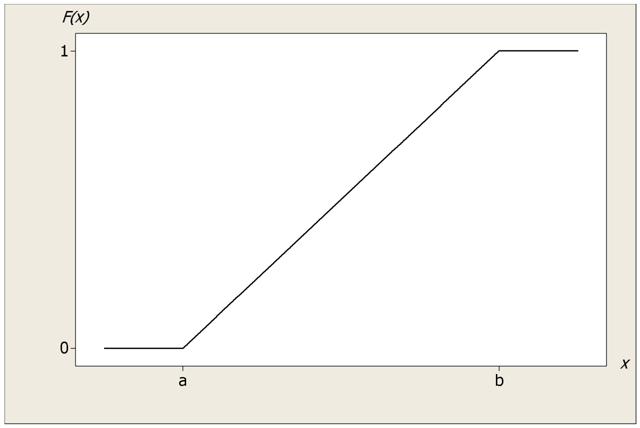
b) The graph of f is as follows:



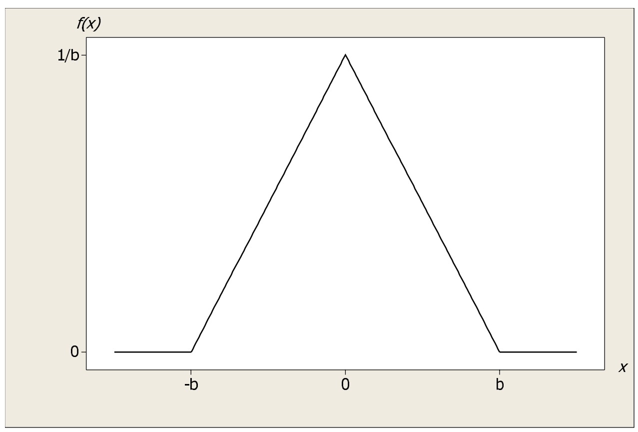
Clearly, $F_X(x) = 0$ if $x < a$, and $F_X(x) = 1$ if $x \geq b$. If $a \leq x < b$,

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{a}^{x} \frac{1}{b-a}\,dt = \frac{x-a}{b-a}.$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



c) The graph of f is as follows:



Clearly, $F_X(x) = 0$ if $x < -b$, and $F_X(x) = 1$ if $x \geq b$. If $-b \leq x < 0$,

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-b}^{x} \frac{1}{b}\left(1 + \frac{t}{b}\right) dt = \frac{1}{2}\left(1 + \frac{x}{b}\right)^2.$$

If $0 \leq x < b$,

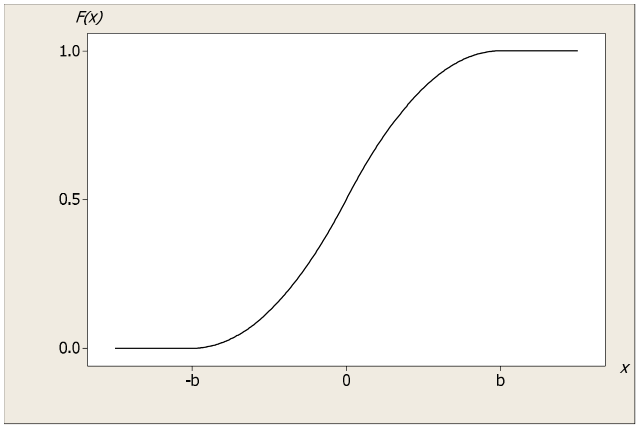
$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-b}^{0} \frac{1}{b}\left(1 + \frac{t}{b}\right) dt + \int_{0}^{x} \frac{1}{b}\left(1 - \frac{t}{b}\right) dt$$
$$= \frac{1}{2} + \frac{1}{2}\left(1 - \left(1 - \frac{x}{b}\right)^2\right) = 1 - \frac{1}{2}\left(1 - \frac{x}{b}\right)^2.$$

Note that, we can also write

$$F_X(x) = \frac{1}{2} + \frac{x}{b} - (\operatorname{sgn} x)\frac{x^2}{2b^2}, \qquad -b \leq x < b,$$

where $\operatorname{sgn} x = -1$, 0, or 1, according as x is negative, zero, or positive, respectively.

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



8.31

a) It is easy to verify that F satisfies properties (a)–(d) of Proposition 8.1 and, hence, is the CDF of some random variable, say, X. We note that

$$P(X = x) = F(x) - F(x-) = \begin{cases} 1, & \text{if } x = 0; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, $\sum_x P(X = x) = 1$ and, hence, by Exercise 5.17 on page 183, X is a discrete random variable. In fact, X is the constant random variable equal to 0.

b) It is easy to verify that F satisfies properties (a)–(d) of Proposition 8.1 and, hence, is the CDF of some random variable, say, X. We note that

$$P(X = x) = F(x) - F(x-) = \begin{cases} 1-p, & \text{if } x = 0; \\ p, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, $\sum_x P(X = x) = 1$ and, hence, by Exercise 5.17 on page 183, X is a discrete random variable. In fact, $X \sim \mathcal{B}(1, p)$ or, equivalently, X has the Bernoulli distribution with parameter p.

c) We note that $\lim_{x\to\infty} F(x) = \infty$ and, hence, property (d) of Proposition 8.1 is violated. Therefore, F is not the CDF of some random variable.

d) It is easy to verify that F satisfies properties (a)–(d) of Proposition 8.1 and, hence, is the CDF of some random variable, say, X. We note that

$$P(X = x) = F(x) - F(x-) = \begin{cases} a_n, & \text{if } x = n \text{ for some nonnegative integer } n; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, $\sum_x P(X = x) = \sum_{n=0}^{\infty} a_n = 1$ and, hence, by Exercise 5.17 on page 183, X is a discrete random variable.

e) It is easy to verify that F satisfies properties (a)–(d) of Proposition 8.1 and, hence, is the CDF of some random variable, say, X. We note that F is everywhere continuous and hence is the CDF of a continuous random variable.

f) We note that $\lim_{x\to\infty} F(x) = \infty$ and, hence, property (d) of Proposition 8.1 is violated. Therefore, F is not the CDF of some random variable.

g) It is easy to verify that F satisfies properties (a)–(d) of Proposition 8.1 and, hence, is the CDF of some random variable, say, X. We note that

$$P(X = x) = F(x) - F(x-) = \begin{cases} 1/8, & \text{if } x = \pm 1; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, X is not a continuous random variable because, say, $P(X = 1) \neq 0$; nor is X a discrete random variable because $\sum_x P(X = x) = 1/4 < 1$. Therefore, X is a random variable of mixed type.

8.32

a) We have

$$F_Y(y) = P(Y \leq y) = P(\max\{X, m\} \leq y) = \begin{cases} 0, & \text{if } y < m; \\ P(X \leq y), & \text{if } y \geq m. \end{cases} = \begin{cases} 0, & \text{if } y < m; \\ F_X(y), & \text{if } y \geq m. \end{cases}$$

b) We have

$$P(Z > z) = P(\min\{X, m\} > z) = \begin{cases} P(X > z), & \text{if } z < m; \\ 0, & \text{if } z \geq m. \end{cases}$$

Therefore,

$$F_Z(z) = P(Z \leq z) = 1 - P(Z > z) = 1 - \begin{cases} P(X > z), & \text{if } z < m; \\ 0, & \text{if } z \geq m. \end{cases} = \begin{cases} F_X(z), & \text{if } z < m; \\ 1, & \text{if } z \geq m. \end{cases}$$

8.33

a) For each $y \in \mathcal{R}$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\max\{X_1, \dots, X_m\} \leq y) = P(X_1 \leq y, \dots, X_m \leq y) \\ &= P(X_1 \leq y) \cdots P(X_m \leq y) = \big(P(X \leq y)\big)^m = \big(F_X(y)\big)^m. \end{aligned}$$

b) For each $z \in \mathcal{R}$,

$$\begin{aligned} P(Z > z) &= P(\min\{X_1, \dots, X_m\} > z) = P(X_1 > z, \dots, X_m > z) = P(X_1 > z) \cdots P(X_m > z) \\ &= \big(P(X > z)\big)^m = \big(1 - P(X \leq z)\big)^m = \big(1 - F_X(z)\big)^m. \end{aligned}$$

Consequently,

$$F_Z(z) = P(Z \leq z) = 1 - P(Z > z) = 1 - \big(1 - F_X(z)\big)^m.$$

8.34 It means that there is no probability in the interval $(a, b]$ or, more precisely, if $A \subset (a, b]$, then $P(X \in A) = 0$. Indeed,

$$P(X \in A) \leq P\big(X \in (a, b]\big) = P(a < X \leq b) = F_X(b) - F_X(a) = c - c = 0.$$

Theory Exercises

8.35 Let $\{x_n\}_{n=1}^{\infty}$ be any increasing sequence of real numbers such that $\lim_{n\to\infty} x_n = x$. For each $n \in \mathcal{N}$, set $A_n = \{X \leq x_n\}$. Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \{X < x\}$. Applying the continuity property of probability gives

$$F_X(x-) = \lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(X \leq x_n) = \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(X < x).$$

8.36

a) First observe that $\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$ and that the two events on the right of the equation are mutually exclusive. Therefore,

$$F_X(b) = P(X \le b) = P(X \le a) + P(a < X \le b) = F_X(a) + P(a < X \le b).$$

Hence, $P(a < X \le b) = F_X(b) - F_X(a)$.

b) First observe that $\{X \le b\} = \{X < a\} \cup \{a \le X \le b\}$ and that the two events on the right of the equation are mutually exclusive. Therefore, in view of Equation (8.7) on page 412,

$$F_X(b) = P(X \le b) = P(X < a) + P(a \le X \le b) = F_X(a-) + P(a \le X \le b).$$

Hence, $P(a \le X \le b) = F_X(b) - F_X(a-)$.

c) First observe that $\{X < b\} = \{X < a\} \cup \{a \le X < b\}$ and that the two events on the right of the equation are mutually exclusive. Therefore, in view of Equation (8.7) on page 412,

$$F_X(b-) = P(X < b) = P(X < a) + P(a \le X < b) = F_X(a-) + P(a \le X < b).$$

Hence, $P(a \le X < b) = F_X(b-) - F_X(a-)$.

Advanced Exercises

8.37 For each $n \in \mathcal{N}$, let $D_n = \{\, x \in \mathcal{R} : F_X(x) - F_X(x-) \ge 1/n \,\}$. We claim that D_n contains at most n numbers. Suppose to the contrary that D_n contains $n+1$ numbers, say, $x_1, \dots, x_{n+1}$. Referring to Equation (8.8) on page 413, we conclude that

$$P(X \in D_n) \ge \sum_{k=1}^{n+1} P(X = x_k) = \sum_{k=1}^{n+1} \big(F_X(x_k) - F_X(x_k-)\big) \ge \sum_{k=1}^{n+1} \frac{1}{n} = \frac{n+1}{n} > 1,$$

which is impossible. Therefore, D_n contains at most n numbers and, in particular, is finite. Now let D denote the set of discontinuites of F_X. We note that $D = \{\, x \in \mathcal{R} : F_X(x) - F_X(x-) > 0 \,\}$ and, hence, that $D = \bigcup_{n=1}^{\infty} D_n$. Consequently, D is a countable union of finite sets and, therefore, D is countable.

8.38

a) We have

$$F_{X_n}(x) = \begin{cases} 0, & \text{if } x < 0; \\ \dfrac{\lfloor nx \rfloor + 1}{n}, & \text{if } 0 \le x < (n-1)/n; \\ 1, & \text{if } x \ge (n-1)/n. \end{cases}$$

b) In the solution to Exercise 8.14(b), we showed that $\lfloor nx \rfloor / n \to x$ as $n \to \infty$ for each $x \in \mathcal{R}$. Therefore, from part (a),

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \frac{\lfloor nx \rfloor + 1}{n} = x,$$

for each $0 \le x < 1$. Consequently, the limiting CDF is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

c) The random variable X from Example 8.4, that is, a randomly chosen number from the interval $(0, 1)$.

8.39

a) We have

$$F_{X_n}(x) = \begin{cases} 0, & \text{if } x < 0; \\ x/n, & \text{if } 0 \le x < n; \\ 1, & \text{if } x \ge n. \end{cases}$$

b) From part (a), if $x < 0$, then $\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} 0 = 0$. Now suppose that $x > 0$. Again from part (a), $F_{X_n}(x) = x/n$ as soon as n exceeds x. Therefore, $\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} x/n = 0$. We have thus shown that $\lim_{n\to\infty} F_{X_n}(x) = 0$ for all $x \in \mathcal{R}$.

c) No, the function that is identically 0 violates property (d) of Proposition 8.1.

d) Answers will vary.

8.40

a) Let $D = \{\, x \in \mathcal{R} : P(X = x) > 0 \,\}$ and set $\alpha = \sum_x P(X = x)$. If $\alpha = 0$, then X is a continuous random variable and we can write

$$F_X = \alpha F_{X_d} + (1-\alpha)F_{X_c},$$

where X_d is any discrete random variable and $X_c = X$. If $\alpha = 1$, then X is a discrete random variable and we can write

$$F_X = \alpha F_{X_d} + (1-\alpha)F_{X_c},$$

where $X_d = X$ and X_c is any continuous random variable.

Now assume that $0 < \alpha < 1$. Define $p\colon \mathcal{R} \to \mathcal{R}$ by $p(x) = \alpha^{-1}P(X = x)$. Clearly, $p(x) \ge 0$ and, by Exercise 8.37, $\{\, x \in \mathcal{R} : p(x) \ne 0 \,\}$ is countable. Moreover,

$$\sum_x p(x) = \sum_x \alpha^{-1}P(X = x) = \alpha^{-1}\sum_x P(X = x) = 1.$$

Thus, p satisfies properties (a), (b), and (c) of Proposition 5.1 on page 187 and, hence, is the PMF of some discrete random variable, say, X_d. Now let

$$F = (1-\alpha)^{-1}(F_X - \alpha F_{X_d}).$$

Note that, for each $a, b \in \mathcal{R}$ with $a < b$,

$$\begin{aligned}(1-\alpha)\big(F(b) - F(a)\big) &= \big(F_X(b) - F_X(a)\big) - \alpha\big(F_{X_d}(b) - F_{X_d}(a)\big) \\ &= P(a < X \le b) - \alpha P(a < X_d \le b).\end{aligned}$$

However,

$$\begin{aligned}\alpha P(a < X_d \le b) &= \alpha \sum_{a<x\le b} p(x) = \sum_{x\in D\cap(a,b]} P(X = x) = P\big(X \in D \cap (a,b]\big) \\ &\le P\big(X \in D \cap (a,b]\big) + P\big(X \in D^c \cap (a,b]\big) = P(a < X \le b).\end{aligned}$$

From the previous two displays, we see that F is nondecreasing and, hence, satisfies property (a) of Proposition 8.1 on page 411. Furthermore, because both F_X and F_{X_d} are CDFs, it now follows easily that F satisfies the other three properties of Proposition 8.1. Thus, F is the CDF of some random variable, say, X_c. Moreover,

$$\begin{aligned}(1-\alpha)P(X_c = x) &= P(X = x) - \alpha P(X_d = x) = P(X = x) - \alpha p(x) \\ &= P(X = x) - P(X = x) = 0,\end{aligned}$$

for all $x \in \mathcal{R}$. Consequently, X_c is a continuous random variable.

b) From Example 8.6, we see that $P(X = 0.75) = 0.25$ and $P(X = x) = 0$ for all $x \neq 0.75$. Therefore, in reference to part (a), we have $\alpha = 0.25$ and

$$p(x) = \alpha^{-1} P(X = x) = 4P(X = x) = \begin{cases} 1, & \text{if } x = 0.75; \\ 0, & \text{otherwise.} \end{cases}$$

Also, $F_{X_d}(x) = 0$ if $x < 0.75$, and $F_{X_d}(x) = 1$ if $x \geq 0.75$. Then

$$F_{X_c}(x) = (1-\alpha)^{-1}\big(F_X(x) - \alpha F_{X_d}(x)\big) = \frac{4}{3}\big(F_X(x) - 0.25 F_{X_d}(x)\big) = \begin{cases} 0, & \text{if } x < 0; \\ \frac{4}{3}x, & \text{if } 0 \leq x < 0.75; \\ 1, & \text{if } x \geq 0.75. \end{cases}$$

c) The discrete random variable is the constant random variable 0.75; the continuous random variable is a number selected at random from the interval (0, 0.75).

8.3 Probability Density Functions

Basic Exercises

8.41
a) From the solution to Exercise 8.15,

$$F_Z(z) = \begin{cases} 0, & \text{if } z < 0; \\ z^3/r^3, & \text{if } 0 \leq z < r; \\ 1, & \text{if } z \geq r. \end{cases}$$

Applying Procedure 8.1 on page 419, we find that

$$f_Z(z) = \begin{cases} 3z^2/r^3, & \text{if } 0 < z < r; \\ 0, & \text{otherwise.} \end{cases}$$

b) That the PDF of Z is quadratic and increasing on the range of Z reflects the fact that we have a geometric probability model and that the rate of change of volume with respect to radius is a quadratic function. In particular, the distance of the point chosen from the center of the sphere is more likely to be large (close to r) than small (close to 0).
c) We have

$$P(Z \leq r/2) = \int_{-\infty}^{r/2} f_Z(z)\,dz = \int_0^{r/2} \frac{3z^2}{r^3}\,dz = \frac{(r/2)^3}{r^3} = \frac{1}{8}.$$

$$P(Z \geq r/4) = \int_{r/4}^{\infty} f_Z(z)\,dz = \int_{r/4}^{r} \frac{3z^2}{r^3}\,dz = \frac{r^3 - (r/4)^3}{r^3} = \frac{63}{64}.$$

$$P(r/4 \leq Z \leq r/2) = \int_{r/4}^{r/2} f_Z(z)\,dz = \int_{r/4}^{r/2} \frac{3z^2}{r^3}\,dz = \frac{(r/2)^3 - (r/4)^3}{r^3} = \frac{7}{64}.$$

d) We have

$$P(Z \le r/2) = F_Z(r/2) = \frac{(r/2)^3}{r^3} = \frac{1}{8}.$$
$$P(Z \ge r/4) = 1 - P(Z < r/4) = 1 - F_Z(r/4) = 1 - \frac{(r/4)^3}{r^3} = 1 - \frac{1}{64} = \frac{63}{64}.$$
$$P(r/4 \le Z \le r/2) = F_Z(r/2) - F_Z(r/4) = \frac{(r/2)^3}{r^3} - \frac{(r/4)^3}{r^3} = \frac{1}{8} - \frac{1}{64} = \frac{7}{64}.$$

8.42 No! Integrating the PDF gets us back to the CDF. We can apply Definition 8.2 on page 406 or Proposition 8.2 on page 412 to the CDF of a random variable to obtain the probability that the random variable falls in any specified interval.

8.43
a) From Exercise 8.16, $F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$ for all $x \in \mathcal{R}$. Applying Procedure 8.1 on page 419, we find that

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathcal{R}.$$

b) The PDF of X takes its maximum at $x = 0$ and decreases monotonically as $|x|$ increases. Thus, the tangent of an angle randomly selected from the interval $(-\pi/2, \pi/2)$ is more likely to be small in magnitude than large in magnitude.
c) We have

$$\begin{aligned} P(X \le 1) &= \int_{-\infty}^{1} f_X(x)\,dx = \int_{-\infty}^{1} \frac{dx}{\pi(1+x^2)} \\ &= \frac{1}{\pi}\big(\arctan 1 - \arctan(-\infty)\big) = \frac{1}{\pi}\left(\frac{\pi}{4} + \frac{\pi}{2}\right) = 0.75. \end{aligned}$$

d) We have

$$P(X \le 1) = F_X(1) = \frac{1}{2} + \frac{1}{\pi}\arctan 1 = \frac{1}{2} + \frac{(\pi/4)}{\pi} = 0.75.$$

8.44
a) From Exercise 8.17,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ (2hy - y^2)/h^2, & \text{if } 0 \le y < h; \\ 1, & \text{if } y \ge h. \end{cases}$$

Applying Procedure 8.1 on page 419, we find that

$$f_Y(y) = \begin{cases} 2(h-y)/h^2, & \text{if } 0 < y < h; \\ 0, & \text{otherwise.} \end{cases}$$

b) The PDF of Y is linear and decreasing on the range of Y. Thus, the distance to the base of a randomly chosen point from the interior of the triangle is more likely to be small (close to 0) than large (close to h).

8.45
a) For $-1 \le x \le 1$,

$$F_X(x) = P(X \le x) = \frac{|\{X \le x\}|}{\pi} = \frac{1}{\pi}\int_{-1}^{x} 2\sqrt{1-t^2}\,dt.$$

Applying Procedure 8.1 on page 419 and using the first fundamental theorem of calculus, we find that

$$f_X(x) = \begin{cases} (2/\pi)\sqrt{1-x^2}, & \text{if } -1 \le x \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) We note that the PDF of X takes its maximum at $x = 0$ and decreases as $|x|$ increases. Roughly, then, for $-1 < x < 1$, the smaller the magnitude of x, the greater is its likelihood. More precisely, the x-coordinate of the center of the first spot to appear is more likely to be small in magnitude than large in magnitude.

c) By symmetry, Y and X have the same distribution. Thus, a PDF of Y is given by

$$f_Y(y) = \begin{cases} (2/\pi)\sqrt{1-y^2}, & \text{if } -1 \le y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

8.46

a) Assume first that $b < 0$. Then

$$F_{a+bX}(y) = P(a + bX \le y) = P\big(X \ge (y-a)/b\big) = 1 - F_X\big((y-a)/b\big).$$

Differentiating now gives

$$f_{a+bX}(y) = -\frac{1}{b} f_X\big((y-a)/b\big).$$

Similarly, if $b > 0$, then

$$f_{a+bX}(y) = \frac{1}{b} f_X\big((y-a)/b\big).$$

Hence, in either case,

$$f_{a+bX}(y) = \frac{1}{|b|} f_X\big((y-a)/b\big).$$

b) Clearly, $F_{X^2}(y) = 0$ for $y < 0$. For $y > 0$,

$$F_{X^2}(y) = P\big(X^2 \le y\big) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Differentiating now gives

$$f_{X^2}(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) = \frac{1}{2\sqrt{y}}\big(f_X(\sqrt{y}) + f_X(-\sqrt{y})\big).$$

Consequently,

$$f_{X^2}(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\big(f_X(\sqrt{y}) + f_X(-\sqrt{y})\big), & \text{if } y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

8.47 We have

$$F_X(x) = P(X \le x) = P\big(X \in (-\infty, x]\big) = \int_{(-\infty,x]} f(t)\,dt = \int_{-\infty}^{x} f(t)\,dt.$$

Hence, by Proposition 8.5 on page 422, f is a PDF of X.

8.48 First note that each of the functions in parts (a)–(e) is nonnegative and, hence, satisfies property (a) of Proposition 8.6 on page 423. To make each one a PDF, we need only divide it by its integral over $\mathcal{R}$. The resulting function will then also satisfy property (b) of Proposition 8.6.

a) From calculus,

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_0^1 x^3(1-x)^2\,dx = \frac{1}{60}.$$

Therefore, the required PDF is

$$f(x) = \begin{cases} 60x^3(1-x)^2, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) From calculus,

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_0^{\infty} e^{-4x}\,dx = \frac{1}{4}.$$

Therefore, the required PDF is

$$f(x) = \begin{cases} 4e^{-4x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

c) From calculus,

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_0^{\infty} x^2 e^{-6x}\,dx = \frac{1}{108}.$$

Therefore, the required PDF is

$$f(x) = \begin{cases} 108x^2 e^{-6x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

d) From calculus,

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \pi.$$

Therefore, the required PDF is

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty.$$

e) From calculus,

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_0^{\pi} \sin x\,dx = 2.$$

Therefore, the required PDF is

$$f(x) = \begin{cases} \frac{1}{2}\sin x, & \text{if } 0 < x < \pi; \\ 0, & \text{otherwise.} \end{cases}$$

8.49 In each part, we check whether f is a nonnegative function (i.e., $f(x) \geq 0$ for all $x \in \mathcal{R}$) and whether its integral equals 1. If that is the case, then f is a PDF; otherwise, it is not.

a) Yes, f is a PDF. We have $F(x) = 0$ if $x < 0$, and, for $x \geq 0$,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^{x} \lambda e^{-\lambda t}\,dt = 1 - e^{-\lambda x}.$$

Therefore,

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

b) Yes, f is a PDF. We have $F(x) = 0$ if $x < a$ and $F(x) = 1$ if $x \geq b$. For $a \leq x < b$,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_a^{x} \frac{1}{b-a}\,dt = \frac{x-a}{b-a}.$$

Therefore,

$$F(x) = \begin{cases} 0, & \text{if } x < a; \\ (x-a)/(b-a), & \text{if } a \le x < b; \\ 1, & \text{if } x \ge b. \end{cases}$$

c) No, f is not a PDF because it is not a nonnegative function.
d) No, f is not a PDF because

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{3} x^2\,dx = \frac{28}{3} \neq 1.$$

e) Yes, f is a PDF. From the solution to Exercise 8.30(c), we know that

$$F(x) = \begin{cases} 0, & \text{if } x < -b; \\ \frac{1}{2}(1 + x/b)^2, & \text{if } -b \le x < 0; \\ 1 - \frac{1}{2}(1 - x/b)^2, & \text{if } 0 \le x < b; \\ 1, & \text{if } x \ge b. \end{cases}$$

8.50 Answers will vary, but one example is the PDF given by

$$f(x) = \begin{cases} 1/\big(2\sqrt{x}\big), & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Its CDF is

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ \sqrt{x}, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

An unbounded CDF is impossible, because all CDFs take values between 0 and 1; that is, if G is a CDF, then $0 \le G(x) \le 1$ for all $x \in \mathcal{R}$, as is seen by referring to Proposition 8.1 on page 411.

8.51 Answers will vary, but one example is the PDF given by

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad -\infty < x < \infty.$$

Its CDF is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x, \qquad -\infty < x < \infty.$$

8.52
a) It is easy to see that F satisfies properties (a)–(d) of Proposition 8.1 on page 411 and, hence, is the CDF of some random variable. Moreover, F is everywhere continuous and, hence, by Proposition 8.3 on page 413, is the CDF of a continuous random variable.
b) Differentiating F shows that a PDF corresponding to F is

$$f(x) = \begin{cases} 1/2, & \text{if } 0 < x < 1; \\ 1/6, & \text{if } 1 < x < 4; \\ 0, & \text{otherwise.} \end{cases}$$

8.53
a) For $2 \le y < 5$,

$$F_Y(y) = P(Y \le y) = \frac{|\{Y \le y\}|}{|(2,5)|} = \frac{|(2,y]|}{|(2,5)|} = \frac{y-2}{3}.$$

Therefore,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 2; \\ (y-2)/3, & \text{if } 2 \le y < 5; \\ 1, & \text{if } y \ge 5. \end{cases}$$

Differentiation now yields

$$f_Y(y) = \begin{cases} 1/3, & \text{if } 2 < y < 5; \\ 0, & \text{otherwise.} \end{cases}$$

b) For $a \le z < b$,

$$F_Z(z) = P(Z \le z) = \frac{|\{Z \le z\}|}{|(a,b)|} = \frac{|(a,z]|}{|(a,b)|} = \frac{z-a}{b-a}.$$

Therefore,

$$F_Z(z) = \begin{cases} 0, & \text{if } z < a; \\ (z-a)/(b-a), & \text{if } a \le z < b; \\ 1, & \text{if } z \ge b. \end{cases}$$

Differentiation now yields

$$f_Z(z) = \begin{cases} 1/(b-a), & \text{if } a < z < b; \\ 0, & \text{otherwise.} \end{cases}$$

8.54 Let X be the loss, in hundreds of thousands of dollars, due to a fire in a commercial building. By assumption, f is a PDF of X. For $0 < x < 20$,

$$P(X > x) = \int_x^\infty f(t)\,dt = \int_x^{20} 0.005(20-t)\,dt = 0.0025(20-x)^2.$$

We want to determine $P(X > 16 \mid X > 8)$. Applying the conditional probability rule yields

$$P(X > 16 \mid X > 8) = \frac{P(X > 8,\ X > 16)}{P(X > 8)} = \frac{P(X > 16)}{P(X > 8)} = \frac{0.0025(20-16)^2}{0.0025(20-8)^2} = \frac{1}{9}.$$

Therefore, of fire losses that exceed \$0.8 million, 11.1% exceed \$1.6 million.

Theory Exercises

8.55 Let a and b be real numbers with $a < b$. By the assumed piecewise continuity of F_X', we can select real numbers $a = c_0 < c_1 < \cdots < c_n = b$ such that F_X' is defined on $[a,b]$, except possibly at the points $c_k,\ 0 \le k \le n$, and that there exists functions $h_k,\ 0 \le k \le n$, continuous on $[c_{k-1}, c_k]$, with $F_X'(x) = h_k(x)$ for all $x \in (c_{k-1}, c_k)$. By definition, then,

$$\int_a^b F_X'(x)\,dx = \sum_{k=1}^n \int_{c_{k-1}}^{c_k} h_k(x)\,dx.$$

As X is a continuous random variable, F_X is an everywhere continuous function. And, because F_X is continuous on $[c_{k-1}, c_k]$ and $F_X'(x) = h_k(x)$ for all $x \in (c_{k-1}, c_k)$, we see that F_X is an antiderivative of h_k on $[c_{k-1}, c_k]$. Therefore, by the second fundamental theorem of calculus,

$$\int_{c_{k-1}}^{c_k} h_k(x)\,dx = F_X(c_k) - F_X(c_{k-1}), \qquad 1 \le k \le n.$$

Consequently,

$$\int_a^b F_X'(x)\,dx = \sum_{k=1}^n \big(F_X(c_k) - F_X(c_{k-1})\big) = F_X(b) - F_X(a).$$

Therefore, by Definition 8.3 on page 417, F_X' is a PDF of X.

Advanced Exercises

8.56 We have

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)^2 = \left(\int_{-\infty}^{\infty} e^{-u^2/2}\,du\right)\left(\int_{-\infty}^{\infty} e^{-v^2/2}\,dv\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(u^2+v^2)/2}\,du\,dv.$$

Changing to polar coordinates and then making the substitution $w = r^2/2$ gives

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)^2 = \int_0^{2\pi}\left(\int_0^{\infty} r\,e^{-r^2/2}\,dr\right)d\theta = \int_0^{2\pi}\left(\int_0^{\infty} e^{-w}\,dw\right)d\theta = 2\pi.$$

Consequently, $\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$, and so the required PDF is given by

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad -\infty < x < \infty.$$

8.57 Making the substitution $y = (x-\mu)/\sigma$ and referring to the preceding exercise, we obtain

$$\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = \sigma\int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \sqrt{2\pi}\,\sigma.$$

Therefore, the required PDF is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

8.58 Assume first that g is strictly decreasing on the range of X. For y in the range of Y,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P\big(g(X) \le y\big) = P\big(X \ge g^{-1}(y)\big) \\ &= 1 - P\big(X \le g^{-1}(y)\big) = 1 - F_X\big(g^{-1}(y)\big). \end{aligned}$$

Applying the chain rule and using the fact from calculus that $(d/dy)g^{-1}(y) = 1/g'\big(g^{-1}(y)\big)$, we get

$$f_Y(y) = F_Y'(y) = -F_X'\big(g^{-1}(y)\big)\frac{d}{dy}g^{-1}(y) = -\frac{1}{g'\big(g^{-1}(y)\big)}\,f_X\big(g^{-1}(y)\big).$$

Similarly, in the case that g is strictly increasing on the range of X, we find that

$$f_Y(y) = \frac{1}{g'\big(g^{-1}(y)\big)}\,f_X\big(g^{-1}(y)\big).$$

Noting that g' is negative in the case of decreasing g and that it is positive in the case of increasing g, we have, in general, that

$$f_Y(y) = \frac{1}{\big|g'\big(g^{-1}(y)\big)\big|}\,f_X\big(g^{-1}(y)\big).$$

8.59 In view of Proposition 8.6 on page 423, we must show that f is a nonnegative function whose integral over $\mathcal{R}$ equals 1. That f is a nonnegative function follows immediately from the facts that each f_k is a nonnegative function and that each α_k is a nonnegative real number. Applying now the facts that each f_k integrates to 1 and that the α_ks sum to 1, we get that

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty}\left(\sum_{k=1}^{n}\alpha_k f_k(x)\right)dx = \sum_{k=1}^{n}\alpha_k\int_{-\infty}^{\infty} f_k(x)\,dx = \sum_{k=1}^{n}\alpha_k = 1.$$

Thus, $f = \sum_{k=1}^{n}\alpha_k f_k$ is a PDF.

8.60

a) From the solution to Exercise 8.33(a), $F_Y(y) = \big(F_X(y)\big)^m$. Differentiation now gives

$$f_Y(y) = m\big(F_X(y)\big)^{m-1} f_X(y).$$

b) From the solution to Exercise 8.33(b), $F_Z(z) = 1 - \big(1 - F_X(z)\big)^m$. Differentiation now gives

$$f_Z(z) = m\big(1 - F_X(z)\big)^{m-1} f_X(z).$$

8.61 Answers will vary, but one possibility is $f = \sum_{k=1}^{\infty} 2^{-k} I_{(k,k+1)}$. Clearly, f is a nonnegative function. Furthermore,

$$\int_{-\infty}^{\infty} f(x)\,dx = \sum_{k=1}^{\infty} \frac{1}{2^k} \int_{-\infty}^{\infty} I_{(k,k+1)}(x)\,dx = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

Consequently, f integrates to 1. Thus, f is a PDF of a continuous random variable. Note that f is discontinuous at each positive integer and, hence, at a countably infinite number of points.

8.62 Answers will vary, but one possibility is $f = \sum_{k=1}^{\infty} I_{(k,k+2^{-k})}$. Clearly, f is a nonnegative function. Furthermore, we have

$$\int_{-\infty}^{\infty} f(x)\,dx = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} I_{(k,k+2^{-k})}(x)\,dx = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

Consequently, f integrates to 1. Thus, f is a PDF of a continuous random variable. Note, however, that for each $x \in \mathcal{R}$, there is a $y > x$ such that $f(y) = 1$. So, $\lim_{x\to\infty} f(x) \neq 0$; in fact, the limit does not exist.

8.63

a) We have

$$G = 1 - 2\int_0^1 F(x)\,dx = 1 - 2\int_0^1 0\,dx = 1 - 0 = 1.$$

Thus, the Gini index equals 1 in the case of completely uneven payroll distribution.

b) We have

$$G = 1 - 2\int_0^1 F(x)\,dx = 1 - 2\int_0^1 x\,dx = 1 - 1 = 0.$$

Thus, the Gini index equals 0 in the case of completely even payroll distribution.

8.64

a) No, because if there were a payroll density function, then F would necessarily be an everywhere continuous function, which it is not.

b) Yes, because we can write $F(x) = \int_0^x 1\,dt$ for $0 \le x \le 1$.

c) First note that $F(x) = \int_0^x 3t^2\,dt = x^3$ for $0 \le x \le 1$. Therefore,

$$G = 1 - 2\int_0^1 F(x)\,dx = 1 - 2\int_0^1 x^3\,dx = 1 - \frac{1}{2} = \frac{1}{2}.$$

Consequently, the Gini index equals 1/2 in the case of a payroll distribution that has a payroll density function of $3x^2$.

8.4 Uniform and Exponential Random Variables

Basic Exercises

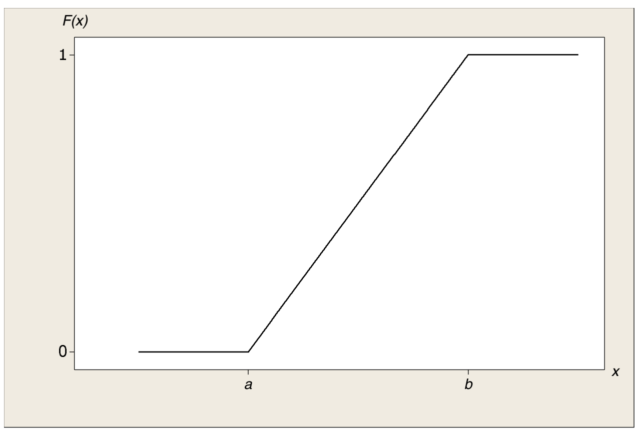
8.65 From Equation (8.20) on page 430, a PDF of X is given by $f_X(x) = 1/(b-a)$ if $a < x < b$, and $f_X(x) = 0$ otherwise. Thus, if $a \le x < b$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_a^x \frac{1}{b-a}\,dt = \frac{x-a}{b-a}.$$

Therefore,

$$F_X(x) = \begin{cases} 0, & \text{if } x < a; \\ (x-a)/(b-a), & \text{if } a \le x < b; \\ 1, & \text{if } x \ge b. \end{cases}$$

Note: In the following graph, we use $F(x)$ instead of $F_X(x)$.



8.66 From Exercise 8.65, we see that $F_X(x) = x/30$ for $0 \le x < 30$.
a) We have

$$P(10 < X < 15) = F_X(15) - F_X(10) = \frac{15}{30} - \frac{10}{30} = \frac{1}{6}$$

and

$$P(X > 10) = 1 - F_X(10) = 1 - \frac{10}{30} = \frac{2}{3}.$$

b) We have

$$P(10 < X < 15) = \frac{|\{10 < X < 15\}|}{30} = \frac{|(10, 15)|}{30} = \frac{5}{30} = \frac{1}{6}$$

and

$$P(X > 10) = \frac{|\{X > 10\}|}{30} = \frac{|(10, 30)|}{30} = \frac{20}{30} = \frac{2}{3}.$$

8.67
a) It seems reasonable that Y should be uniformly distributed. Proceeding purely formally, we note that, when $X = 0$, $Y = c$, and that, when $X = 1$, $Y = c + d$. Thus, we guess that $Y \sim \mathcal{U}(c, c+d)$.

b) We first determine the CDF of Y. If $c \le y < c + d$,

$$F_Y(y) = P(Y \le y) = P(c + dX \le y) = P\big(X \le (y-c)/d\big) = (y-c)/d,$$

where the last equality follows from Example 8.4 on page 408. Differentiation yields

$$f_Y(y) = \begin{cases} 1/d, & \text{if } c < y < c+d; \\ 0, & \text{otherwise.} \end{cases}$$

Referring to Equation (8.20) on page 430 now shows that $Y \sim \mathcal{U}(c, c+d)$.
c) We know that $Y \sim \mathcal{U}(c, c+d)$. Thus we need to choose $c = a$ and $c + d = b$, that is, $c = a$ and $d = b - a$. Hence, if $X \sim \mathcal{U}(0, 1)$, then $a + (b-a)X \sim \mathcal{U}(a, b)$.

8.68
a) Answers will vary. However, the technique is to generate 10,000 numbers with a basic random number generator and then to transform those numbers with $y = -2 + 5x$.
b) The shape of the histogram should be roughly rectangular on the the interval $(-2, 3)$.
c) Answers will vary.

8.69 That F is everywhere continuous follows easily from the facts that $F_X(0) = 1 - e^{-\lambda \cdot 0} = 0$ and that the exponential function, e^x, is everywhere continuous. In particular, then, property (b) of Proposition 8.1 on page 411 is satisfied. Because the exponential function is strictly increasing, $e^{-\lambda x}$ is strictly decreasing. Consequently, the function $1 - e^{-\lambda x}$ is strictly increasing. It now follows that F_X satisfies property (a). Property (c) is obvious from $F_X(x) = 0$ if $x < 0$. Finally,

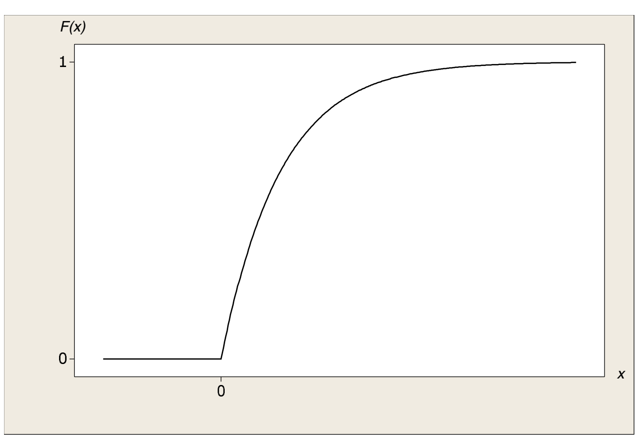
$$F_X(\infty) = \lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} (1 - e^{-\lambda x}) = 1 - 0 = 1,$$

so that property (d) holds.

8.70
a) Clearly, $F_X(x) = 0$ if $x < 0$. For $x \ge 0$,

$$\begin{aligned} F_X(x) = P(X \le x) = P\big(X \in (-\infty, x]\big) &= \int_{-\infty}^{x} f_X(t)\, dt \\ = \int_0^x \lambda e^{-\lambda t}\, dt = -\big(e^{-\lambda x} - 1\big) &= 1 - e^{-\lambda x}. \end{aligned}$$

b) *Note:* In the following graph, we use $F(x)$ instead of $F_X(x)$.



8.71 From Exercise 8.70, we see that $F_X(x) = 1 - e^{-6.9x}$ for $x \geq 0$. Therefore,

$$P(1/4 < X < 1/2) = F_X(1/2) - F_X(1/4) = \left(1 - e^{-6.9 \cdot \frac{1}{2}}\right) - \left(1 - e^{-6.9 \cdot \frac{1}{4}}\right)$$
$$= e^{-1.725} - e^{-3.45} = 0.146$$

and

$$P(X < 1) = F_X(1) = 1 - e^{-6.9 \cdot 1} = 0.999.$$

8.72

a) We have $F_Y(y) = 0$ if $y < 0$. For $y \geq 0$,

$$F_Y(y) = P(Y \leq y) = P(bX \leq y) = P(X \leq y/b) = F_X(y/b) = 1 - e^{-y/b}.$$

Differentiation yields

$$f_Y(y) = \begin{cases} \frac{1}{b}e^{-y/b}, & \text{if } y > 0; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} \frac{1}{b}e^{-\frac{1}{b}y}, & \text{if } y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $Y \sim \mathcal{E}(1/b)$.

b) We know that $bX \sim \mathcal{E}(1/b)$. Thus, we need to choose $1/b = \lambda$, that is, $b = 1/\lambda$. Hence, if $X \sim \mathcal{E}(1)$, then $\lambda^{-1}X \sim \mathcal{E}(\lambda)$.

8.73

a) Let $Y = -\ln X$. We have $F_Y(y) = 0$ if $y < 0$. For $y \geq 0$,

$$F_Y(y) = P(Y \leq y) = P(-\ln X \leq y) = P(\ln X \geq -y)$$
$$= P\left(X \geq e^{-y}\right) = 1 - P\left(X \leq e^{-y}\right) = 1 - e^{-y}.$$

Differentiation yields

$$f_Y(y) = \begin{cases} e^{-y}, & \text{if } y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $Y \sim \mathcal{E}(1)$; that is, $-\ln X \sim \mathcal{E}(1)$.

b) Because $-\ln X \sim \mathcal{E}(1)$, Exercise 8.72(b) shows that $-\lambda^{-1}\ln X = \lambda^{-1}\big(-\ln X\big) \sim \mathcal{E}(\lambda)$.

8.74

a) Answers will vary. However, the technique is to generate 10,000 numbers with a basic random number generator and then to transform those numbers with $y = -(\ln x)/6.9$.

b) The shape of the histogram should be roughly that of the PDF of a random variable having an exponential distribution with parameter 6.9. See Figure 8.12 on page 432.

c) Answers will vary.

8.75 Let X be the duration, in seconds, of a pause during a monologue. By assumption, $X \sim \mathcal{E}(1.4)$. From Exercise 8.70, the CDF of X is given by $F_X(x) = 0$ if $x < 0$, and $F_X(x) = 1 - e^{-1.4x}$ if $x \geq 0$. Also note that, from Proposition 8.8 on page 433, $P(X > x) = e^{-1.4x}$ for $x > 0$.

a) $P(0.5 \leq X \leq 1) = F_X(1) - F_X(0.5) = e^{-1.4 \cdot 0.5} - e^{-1.4 \cdot 1} = 0.250.$

b) $P(X > 1) = e^{-1.4 \cdot 1} = e^{-1.4} = 0.247.$

c) $P(X > 3) = e^{-1.4 \cdot 3} = e^{-4.2} = 0.0150.$

d) We can solve this problem by using the conditional probability rule as follows:

$$P(X > 3 \mid X > 2) = \frac{P(X > 3, X > 2)}{P(X > 2)} = \frac{P(X > 3)}{P(X > 2)} = \frac{e^{-1.4 \cdot 3}}{e^{-1.4 \cdot 2}} = \frac{e^{-4.2}}{e^{-2.8}} = e^{-1.4} = 0.247.$$

Alternatively, we can use the lack-of-memory property of exponential random variables:

$$P(X > 3 \mid X > 2) = P(X > 2 + 1 \mid X > 2) = P(X > 1) = e^{-1.4\cdot 1} = e^{-1.4} = 0.247.$$

8.76 Let X denote the duration, in hours, of a round trip. By assumption, $X \sim \mathcal{E}(1/20)$, or $X \sim \mathcal{E}(0.05)$. From Exercise 8.70, the CDF of X is given by $F_X(x) = 0$ if $x < 0$, and $F_X(x) = 1 - e^{-0.05x}$ if $x \geq 0$. Also note that, from Proposition 8.8 on page 433, $P(X > x) = e^{-0.05x}$ for $x > 0$.

a) $P(X \leq 15) = 1 - e^{-0.05\cdot 15} = 0.528$.

b) $P(15 \leq X \leq 25) = F_X(25) - F_X(15) = e^{-0.05\cdot 15} - e^{-0.05\cdot 25} = 0.186$.

c) $P(X > 25) = e^{-0.05\cdot 25} = e^{-1.25} = 0.287$.

d) Applying the lack-of-memory property of exponential random variables yields

$$\begin{aligned} P(X \leq 40 \mid X > 15) &= 1 - P(X > 40 \mid X > 15) = 1 - P(X > 15 + 25 \mid X > 15) \\ &= 1 - P(X > 25) = 1 - e^{-0.05\cdot 25} = 1 - e^{-1.25} = 0.713. \end{aligned}$$

e) From part (c), the probability is 0.287 that a round trip takes more than 25 hours. So the number, Y, of round trips of five that take more than 25 hours has the binomial distribution with parameters 5 and 0.287; that is, $Y \sim \mathcal{B}(5, 0.287)$. Thus,

$$P(Y = 2) = \binom{5}{2}(0.287)^2(1 - 0.287)^{5-2} = 0.299.$$

Note: If we use the exact value for $P(X > 25)$, namely, $e^{-1.25}$, instead of the rounded value, 0.287, we get an answer of 0.298.

8.77 Let X denote claim size, in thousands of dollars, 10 years ago. Then $X \sim \mathcal{E}(\lambda)$. To determine λ, we note that

$$0.25 = P(X < 1) = 1 - e^{-\lambda\cdot 1},$$

or $e^{-\lambda} = 0.75$. Now let Y denote claim size, in thousands of dollars, today. Then $Y = 2X$. Arguing as in Exercise 8.72(a), we find that $Y \sim \mathcal{E}(\lambda/2)$. Therefore,

$$P(Y < 1) = 1 - e^{-\frac{\lambda}{2}\cdot 1} = 1 - \left(e^{-\lambda}\right)^{1/2} = 1 - \sqrt{0.75} = 0.134.$$

8.78

a) The first relation.

b) Events $\{X > s + t\}$ and $\{X > s\}$ are independent for all $s, t \geq 0$.

c) No. The second relation implies, in particular, that $\{X > s\}$ is independent of itself for all $s \geq 0$. But then

$$\big(P(X > s)\big)^2 = P(X > s)P(X > s) = P(X > s, X > s) = P(X > s), \qquad s \geq 0.$$

From this result, it follows that, for each $s \geq 0$, we have $P(X > s) = 0$ or $P(X > s) = 1$.

We claim that $P(X > s) = 1$ for all $s > 0$. Suppose to the contrary that there is a $t > 0$ such that $P(X > t) < 1$, which, in the current situation, implies that $P(X > t) = 0$. Because X is a positive random variable, we must have $P(X > 0) = 1$. Thus, $F_X(0) = 0$ and $F_X(t) = 1$. As X is a continuous random variable, F_X is an everywhere continuous function. Applying the intermediate value theorem now shows that there is an s $(0 < s < t)$ such that $F_X(s) = 0.5$ or, equivalently, that $P(X > s) = 0.5$, which is a contradiction. Therefore, $P(X > s) = 1$ for all $s \geq 0$.

However, $P(X > s) = 1$ for all $s > 0$ implies that $F_X(s) = 0$ for all $s > 0$. But then we would have $F_X(\infty) = \lim_{x\to\infty} F_X(x) = 0$, which violates property (d) of Proposition 8.1 on page 411. Therefore, it is impossible for the second relation to hold for a positive continuous random variable.

8.79

a) If X had the lack-of-memory property, then $P(X > s + t \mid X > s) = P(X > t) = 1 - t$.

b) Applying the conditional probability rule yields

$$P(X > s + t \mid X > s) = \frac{P(X > s,\ X > s + t)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}$$
$$= \frac{1 - (s + t)}{1 - s} = 1 - \frac{t}{1 - s}.$$

Observe that this probability doesn't equal the one in part (a), showing that X doesn't have the lack-of-memory property, a fact that we already know from Proposition 8.9 on page 434.

8.80 Clearly, $F_X(x) = 0$ if $x < 0$. For $x \geq 0$,

$$P(X > x) = P\big(\min\{X_1, \ldots, X_m\} > x\big) = P(X_1 > x, \ldots, X_m > x)$$
$$= P(X_1 > x) \cdots P(X_m > x) = e^{-\lambda_1 x} \cdots e^{-\lambda_m x} = e^{-(\lambda_1 + \cdots + \lambda_m)x}.$$

Consequently,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-(\lambda_1 + \cdots + \lambda_m)x}, & \text{if } x \geq 0. \end{cases}$$

Differentiation now yields $f_X(x) = (\lambda_1 + \cdots + \lambda_m)e^{-(\lambda_1 + \cdots + \lambda_m)x}$ if $x > 0$, and $f_X(x) = 0$ otherwise. Thus, we see that $\min\{X_1, \ldots, X_m\} \sim \mathcal{E}(\lambda_1 + \cdots + \lambda_m)$.

8.81 Let R denote the average rate, in customers per minute, at which the service representative responds to inquiries, that is, $R = 10/T$. Note that the range of R is the interval $(10/12, 10/8) = (5/6, 5/4)$. For $5/6 \leq r < 5/4$,

$$F_R(r) = P(R \leq r) = P(10/T \leq r) = P(T \geq 10/r) = \frac{12 - 10/r}{12 - 8} = 3 - \frac{5}{2r}.$$

Differentiation now yields

$$f_R(r) = \begin{cases} \dfrac{5}{2r^2}, & \text{if } 5/6 < r < 5/4; \\ 0, & \text{otherwise.} \end{cases}$$

8.82 Let $x > 0$. Noting that $\{X \leq x\} = \{N(x) \geq 1\}$, we get

$$F_X(x) = P(X \leq x) = P\big(N(x) \geq 1\big) = 1 - P\big(N(x) = 0\big) = 1 - e^{-6.9x}\frac{(6.9x)^0}{0!} = 1 - e^{-6.9x}.$$

Therefore, $X \sim \mathcal{E}(6.9)$.

8.83

a) Suppose that M is a median of X. Then

$$F_X(M-) = P(X < M) = 1 - P(X \geq M) \leq 1 - 1/2 = 1/2$$

and $F_X(M) = P(X \leq M) \geq 1/2$. Conversely, suppose that $F_X(M-) \leq 1/2 \leq F_X(M)$. Then we have $P(X \leq M) = F_X(M) \geq 1/2$ and

$$P(X \geq M) = 1 - P(X < M) = 1 - F_X(M-) \geq 1 - 1/2 = 1/2.$$

Thus, M is a median of X.

b) The conditions $P(X < M) \leq 1/2$ and $P(X > M) \leq 1/2$ are precisely the same as the conditions $F_X(M-) \leq 1/2$ and $F_X(M) \geq 1/2$. Therefore, the equivalence follows from part (a).

c) Let $X = I_E$ and let $p = P(E)$. Recall from Exercise 8.19 that

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - p, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

In what follows, you may find it helpful to refer to the graph in the solution to Exercise 8.19. We consider three cases.

$p < 1/2$: Then $F_X(x-) \le 1/2 \le F_X(x)$ if and only if $x = 0$. Thus, the median of X is 0.

$p = 1/2$: Then $F_X(x-) \le 1/2 \le F_X(x)$ for all $0 \le x \le 1$. Thus, all numbers in $[0, 1]$ are medians of X.

$p > 1/2$: Then $F_X(x-) \le 1/2 \le F_X(x)$ if and only if $x = 1$. Thus, the median of X is 1.

d) Let the range of X be the interval (a, b), where $-\infty \le a < b \le \infty$. Note that we have $F_X(a) = 0$ and $F_X(b) = 1$. Because X is a continuous random variable, its CDF is everywhere continuous. Applying the intermediate value theorem, we conclude that there is a number M, with $a < M < b$, such that $F_X(M) = 1/2$ and, because F_X is strictly increasing on the range of X, we know that M is the unique such number.
e) Because X is uniformly distributed on the interval (a, b), we would guess that the median is halfway between the endpoints of the interval, that is, $M = (a + b)/2$.
f) Note that X satisfies the conditions of part (d). From Exercise 8.65, we have $F_X(x) = (x - a)/(b - a)$ for $a \le x < b$. Using algebra to solve the equation $F_X(M) = 1/2$, we find that $M = (a + b)/2$.
g) Note that X satisfies the conditions of part (d). From Exercise 8.70, we have $F_X(x) = 1 - e^{-\lambda x}$ for $x \ge 0$. Using algebra to solve the equation $F_X(M) = 1/2$, we find that $M = \lambda^{-1} \ln 2$.

8.84 First observe that X must have an exponential distribution with parameter 0.004. Therefore, $c = 0.004$. Let Y be the benefit for the policy. Then $Y = \min\{X, 250\}$. Referring now to Exercise 8.32(b), we find that $F_Y(y) = F_X(y)$ if $y < 250$, and $F_Y(y) = 1$ if $y \ge 250$. Because $X \sim \mathcal{E}(0.004)$, it now follows that

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ 1 - e^{-0.004y}, & \text{if } 0 \le y < 250; \\ 1, & \text{if } y \ge 250. \end{cases}$$

Using algebra to solve the equation $F_Y(M) = 1/2$, we find that $M = (\ln 2)/0.004 \approx 173.29$. Thus, the median benefit for this policy is approximately \$173.29.

Theory Exercises

8.85
a) Applying the conditional probability rule to the left side of Equation (8.24) yields

$$P(X > s + t \mid X > s) = \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} = \frac{G(s + t)}{G(s)}.$$

Because the right side of Equation (8.24) is equal to $G(t)$, we see that the equation is equivalent to

$$G(s + t) = G(s)G(t), \qquad s, t \ge 0. \tag{$*$}$$

b) From Equation $(*)$, it follows that, for each $n \in \mathcal{N}$, $G(2/n) = \big(G(1/n)\big)^2$ and therefore, by induction on m, we get from Equation $(*)$ that

$$\begin{aligned} G(m/n) &= G\big((m-1)/n + 1/n)\big) = G\big((m-1)/n\big)\, G(1/n) \\ &= \big(G(1/n)\big)^{m-1} G(1/n) = \big(G(1/n)\big)^m. \end{aligned}$$

Thus,

$$G(m/n) = \big(G(1/n)\big)^m, \qquad m, n \in \mathcal{N}. \tag{$**$}$$

c) Setting $n = 1$ in Equation $(**)$ gives $G(m) = \big(G(1)\big)^m$ for each $m \in \mathcal{N}$. Applying Proposition 8.1(d) on page 411, we conclude that

$$\lim_{m\to\infty} \big(G(1)\big)^m = \lim_{m\to\infty} G(m) = \lim_{m\to\infty} \big(1 - F_X(m)\big) = 0.$$

Consequently, $G(1) < 1$.
d) Assume that $G(1) = 0$. Setting $m = n$ in Equation $(**)$ gives $\big(G(1/n)\big)^n = G(1)$, or

$$G(1/n) = \big(G(1)\big)^{1/n}, \qquad n \in \mathcal{N}. \tag{$***$}$$

We now see that $G(1/n) = 0$ for each $n \in \mathcal{N}$. Because F_X is right-continuous, so is G. Therefore,

$$P(X > 0) = G(0) = \lim_{n\to\infty} G(1/n) = 0$$

and, hence, because X is a positive random variable,

$$P(X = 0) = 1 - P(X < 0) - P(X > 0) = 1 - 0 - 0 = 1.$$

This result contradicts the assumption that X is a continuous random variable. Consequently, $G(1) > 0$.
e) Setting $\lambda = -\ln\big(G(1)\big)$, we note that λ is a positive real number and, in view of Equation $(***)$, we have $G(1/n) = e^{-\lambda/n}$ for each $n \in \mathcal{N}$. Using this result in Equation $(**)$ shows that $G(r) = e^{-\lambda r}$ for all positive rational numbers.
f) Let $x > 0$ and choose a decreasing sequence of rational numbers $\{r_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} r_n = x$. Referring to part (e) and using the right-continuity of G, we conclude that

$$G(x) = \lim_{n\to\infty} G(r_n) = \lim_{n\to\infty} e^{-\lambda r_n} = e^{-\lambda x}.$$

Hence, $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$ and, so, X is an exponential random variable.

Advanced Exercises

8.86 By assumption $G(t) = P(L > t)$ for $t \geq 0$. A thread of length $s + t$ will not break if and only if the individual segments of lengths s and t (say, from 0 to s and from s to $s + t$) do not break. Because there is no breakage interaction between disjoint pieces of thread,

$$G(s + t) = P(L > s + t) = P(L > s)P(L > t) = G(s)G(t).$$

Referring now to Exercise 8.85, we conclude that the random variable L is exponentially distributed with parameter $-\ln\big(G(1)\big)$.

8.87
a) Let X denote the point at which the stick is cut. We can assume that $X \sim \mathcal{U}(0, \ell)$. Also, let Y denote the length of the shorter segment. Then $Y = \min\{X, \ell - X\}$. For $0 \leq y < \ell/2$,

$$P(Y > y) = P(X > y, \ell - X > y) = P(y < X < \ell - y) = \frac{(\ell - y) - y}{\ell} = 1 - \frac{2y}{\ell}.$$

Therefore, $F_Y(y) = 2y/\ell = y/(\ell/2)$ for $0 \leq y < \ell/2$ and, so, $Y \sim \mathcal{U}(0, \ell/2)$. The probability that the shorter segment is less than half as long as the longer segment equals

$$P\big(Y < (\ell - Y)/2\big) = P(Y < \ell/3) = \frac{\ell/3}{\ell/2} = \frac{2}{3}.$$

b) Let R be the ratio of the longer segment to the shorter segment, so that $R = (\ell - Y)/Y$. Referring to part (a), we find that, for $r \geq 1$,

$$F_R(r) = P(R \leq r) = P\left(\frac{\ell - Y}{Y} \leq r\right) = P\big(Y \geq \ell/(r+1)\big) = 1 - \frac{\ell/(r+1)}{\ell/2} = \frac{r-1}{r+1}.$$

Differentiation yields

$$f_R(r) = \begin{cases} 2/(r+1)^2, & \text{if } r > 0; \\ 0, & \text{otherwise.} \end{cases}$$

8.5 Normal Random Variables

Basic Exercises

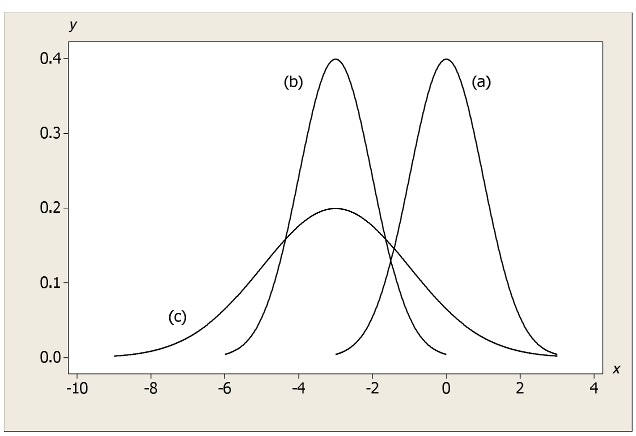
8.88 Because only algebra is involved here, we leave the details to the reader.

8.89
a) Because the exponential function is always positive, f_X is also always positive.
b) As $e^{-y} \downarrow 0$ as $y \to \infty$, it follows easily that $e^{-(x-\mu)^2/2\sigma^2} \downarrow 0$ as $x \to \pm\infty$. This latter relation, in turn, implies that $f_X(x) \downarrow 0$ as $x \to \pm\infty$.
c) Differentiation gives $f_X'(x) = -f_X(x)(x-\mu)/\sigma^2$. This relation shows that $f_X'(x) = 0$ if and only if $x = \mu$. In view of part (b), we conclude that f_X attains its maximum at $x = \mu$.
d) A second differentiation shows that

$$\begin{aligned} f_X''(x) &= -f_X(x)/\sigma^2 - f_X'(x)(x-\mu)/\sigma^2 \\ &= -f_X(x)/\sigma^2 + f_X(x)(x-\mu)^2/\sigma^4 = \sigma^{-4} f_X(x)\big((x-\mu)^2 - \sigma^2\big). \end{aligned}$$

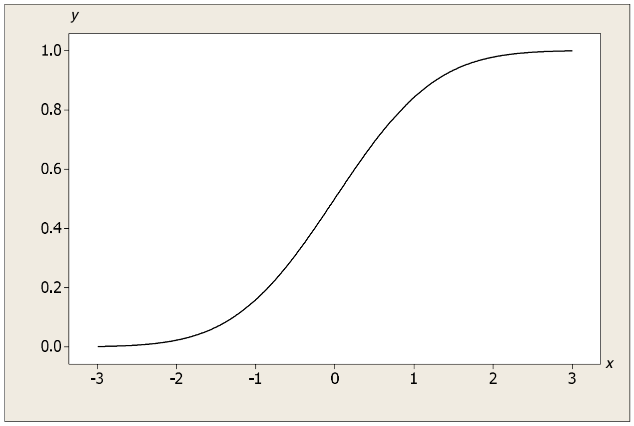
This relation shows that $f_X''(x) = 0$ if and only if $x = \mu \pm \sigma$, which implies that f_X has inflection points at those two values of x.

8.90 The graphs of the three normal PDFs are as follows:



Each PDF is centered at its μ parameter and its spread is determined by its σ parameter—the larger the σ parameter, the flatter and more spread out is the PDF.

8.91 A graph of the PDF is given in the solution to Exercise 8.90, specifically, part (a). A graph of the CDF is as follows:



8.92 The σ parameters for the two distributions are 2 and 1, respectively. Therefore, the former distribution has a wider spread than the latter distribution because its σ parameter is larger.

8.93
a) True, because their σ parameters are identical and the shape of a normal distribution is determined by that parameter.
b) False, because their μ parameters differ and the center of a normal distribution is determined by that parameter.

8.94
a) Referring to Proposition 8.11 on page 443 and Equation (8.37) on page 444, we get that

$$P\left(|X-\mu| \le t\right) = P(\mu - t \le X \le \mu + t) = \Phi\left(\frac{(\mu+t)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-t)-\mu}{\sigma}\right)$$
$$= \Phi(t/\sigma) - \Phi(-t/\sigma) = \Phi(t/\sigma) - \big(1 - \Phi(t/\sigma)\big) = 2\Phi(t/\sigma) - 1.$$

b) Referring to part (a), we find that

$$P\left(|X-\mu| \ge t\right) = 1 - P\left(|X-\mu| \le t\right) = 1 - \big(2\Phi(t/\sigma) - 1\big) = 2\big(1 - \Phi(t/\sigma)\big).$$

8.95 We have

$$\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\cdot\frac{1}{\sqrt{2\pi}}\,e^{-((x-\mu)/\sigma)^2/2} = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-(x-\mu)^2/2\sigma^2}.$$

8.96
a) Because, by Proposition 8.11 on page 443,

$$P(\mu - z\sigma \le X \le \mu + z\sigma) = \Phi\left(\frac{(\mu+z\sigma)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-z\sigma)-\mu}{\sigma}\right) = \Phi(z) - \Phi(-z).$$

b) From part(a) and Equation (8.37) on page 444,

$$P(\mu - z\sigma \le X \le \mu + z\sigma) = 2\Phi(z) - 1.$$

Referring now to Table I on page A-39, we get the following values for $z = 1$, 2, and 3, respectively: 0.6826, 0.9544, and 0.9974.

8.97 They are the same because two normal random variables with the same μ and σ^2 parameters have the same PDF and hence the same probability distribution. In other words, μ and σ^2 together completely determine a normal distribution.

8.98 Here are two reasons why the number of chips in a bag couldn't be exactly normally distributed:

- The number of chips in a bag is a discrete random variable, whereas a normal distribution is the distribution of a continuous random variable.
- A normal random variable has range $\mathcal{R}$, but the number of chips in a bag can't be negative.

8.99
a) Using Equation (8.35) on page 443 and Table I on page A-39, we get

$$P(X > 75) = 1 - P(X \leq 75) = 1 - \Phi\big((75 - 61)/9\big) = 1 - \Phi(1.56) = 1 - 0.9406 = 0.0594.$$

b) Proceeding as in part (a) and referring to Equation (8.37) on page 444 yields

$$\begin{aligned} P(X < 50 \text{ or } X > 70) &= P(X < 50) + P(X > 70) = \Phi\big((50 - 61)/9\big) + 1 - \Phi\big((70 - 61)/9\big) \\ &= \Phi(-1.22) + 1 - \Phi(1) = 1 - \Phi(1.22) + 1 - \Phi(1) = 0.2699. \end{aligned}$$

8.100 Let X denote brain weight of Swedish males. Then $X \sim \mathcal{N}\left(1.40, 0.11^2\right)$.
a) Referring to Proposition 8.11 on page 443 yields

$$\begin{aligned} P(1.50 \leq X \leq 1.70) &= \Phi\big((1.70 - 1.40)/0.11\big) - \Phi\big((1.50 - 1.40)/0.11\big) \\ &= \Phi(2.73) - \Phi(0.91) = 0.9968 - 0.8186 = 0.1782. \end{aligned}$$

b) Using Equation (8.35) on page 443 gives

$$P(X < 1.6) = \Phi\big((1.60 - 1.40)/0.11\big) = \Phi(1.82) = 0.9656.$$

8.101 Let X denote gestation period for women. Then $X \sim \mathcal{N}\left(266, 16^2\right)$.
a) Using Equation (8.35) on page 443 gives

$$P(X < 300) = P(X \leq 300) = F_X(300) = \Phi\big((300 - 266)/16\big) = \Phi(1.25) = 0.9832.$$

Thus, 98.32% of pregnant women give birth before 300 days.
b) Applying the conditional probability rule, referring to part (a), and using the fact that $\Phi(0) = 0.5$, we get

$$\begin{aligned} P(X \leq 300 \mid X > 266) &= \frac{P(266 < x \leq 300)}{P(X > 266)} = \frac{\Phi\big((300 - 266)/16\big) - \Phi\big((266 - 266)/16\big)}{1 - \Phi\big((266 - 266)/16\big)} \\ &= \frac{0.9832 - 0.5}{1 - 0.5} = 0.9664. \end{aligned}$$

Thus, among those women with a longer than average gestation period, 96.64% give birth within 300 days.
c) Yes. For the defendant to have fathered the child, the gestation period would have to be either at least 309 days (August 31, 2001 to July 6, 2002) or at most 93 days (April 4, 2002 to July 6, 2002). The probability of that happening is

$$\begin{aligned} P(X \geq 309) + P(X \leq 93) &= 1 - P(X \leq 309) + P(X \leq 93) \\ &= 1 - \Phi\big((309 - 266)/16\big) + \Phi\big((93 - 266)/16\big) \\ &= 1 - \Phi(2.69) + \Phi(-10.81) \approx 0.004. \end{aligned}$$

Consequently, the probability is about 0.004 (less than a one-half of one percent chance) that the defendant fathered the child. Although possible, it's highly unlikely that the defendant fathered the child.

8.102 Let X be the amount of cola dispensed into the cup. By assumption, $X \sim \mathcal{N}\left(\mu, 0.25^2\right)$. We need to choose μ so that $P(X < 8) = .02$. Thus,

$$0.02 = P(X < 8) = \Phi\big((8-\mu)/0.25\big) = 1 - \Phi\big((\mu - 8)/0.25\big).$$

This means that $(\mu - 8)/0.25 = \Phi^{-1}(0.98) = 2.05$, or $\mu = 8.5125$ oz.

8.103 Let X denote bolt diameter, in millimeters. By assumption, $X \sim \mathcal{N}\left(10, \sigma^2\right)$. The manufacturer's conditions are equivalent to $P(9.7 \le X \le 10.3) = 0.999$. Thus, we need to choose σ so that

$$\begin{aligned}0.999 = P(9.7 \le X \le 10.3) &= \Phi\big((10.3-10)/\sigma\big) - \Phi\big((9.7-10)/\sigma\big)\\ &= \Phi(0.3/\sigma) - \Phi(-0.3/\sigma) = 2\Phi(0.3/\sigma) - 1.\end{aligned}$$

Consequently, $0.3/\sigma = \Phi^{-1}(1.999/2) = \Phi^{-1}(0.9995) = 3.29$, or $\sigma = 0.09$ mm.

8.104 Let X denote measured BAC of a person whose actual BAC equals μ. By assumption, we have $X \sim \mathcal{N}\left(\mu, 0.005^2\right)$.

a) In this case, $\mu = 0.11$. Thus,

$$P(X < 0.10) = \Phi\big((0.10 - 0.11)/0.005\big) = \Phi(-2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228.$$

b) In this case, $\mu = 0.095$. Thus,

$$P(X \ge 0.10) = 1 - \Phi\big((0.10 - 0.095)/0.005\big) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587.$$

8.105 We have $f_X(x) = \left(\alpha/\sqrt{2\pi}\right) e^{-\alpha^2 x^2/2}$. Note that f_X is symmetric about 0. For convenience, let $Y = 1/X^2$. For $y \ge 0$,

$$\begin{aligned}F_Y(y) = P(Y \le y) &= P\left(1/X^2 \le y\right) = P\left(X^2 \ge 1/y\right) = 1 - P\left(X^2 \le 1/y\right)\\ &= 1 - P\left(-y^{-1/2} \le X \le y^{-1/2}\right) = 1 - \left(F_X\left(y^{-1/2}\right) - F_X\left(-y^{-1/2}\right)\right)\\ &= 1 - F_X\left(y^{-1/2}\right) + F_X\left(-y^{-1/2}\right).\end{aligned}$$

Differentiation yields

$$f_Y(y) = \frac{1}{2}y^{-3/2} f_X\left(y^{-1/2}\right) + \frac{1}{2}y^{-3/2} f_X\left(y^{-1/2}\right) = y^{-3/2} f_X\left(y^{-1/2}\right) = \frac{\alpha}{\sqrt{2\pi y^3}}\, e^{-\alpha^2/2y}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

8.106 Let X denote diameter for the produced ball bearings. Then $X \sim \mathcal{N}\left(1.4, 0.025^2\right)$. Let Y denote diameter for the remaining ball bearings. For $1.35 \le y < 1.48$,

$$\begin{aligned}F_Y(y) = P(Y \le y) = P(X \le y \mid 1.35 \le X \le 1.48) &= \frac{P(X \le y,\, 1.35 \le X \le 1.48)}{P(1.35 \le X \le 1.48)}\\ = \frac{P(1.35 \le X \le y)}{P(1.35 \le X \le 1.48)} &= \frac{F_X(y) - F_X(1.35)}{F_X(1.48) - F_X(1.35)}.\end{aligned}$$

Now,

$$\begin{aligned}F_X(1.48) - F_X(1.35) &= \Phi\big((1.48 - 1.4)/0.025\big) - \Phi\big((1.35 - 1.4)/0.025\big)\\ &= \Phi(3.20) - \Phi(-2) = 0.9765.\end{aligned}$$

Noting that $1/0.9765 \approx 1.024$, it now follows that

$$f_Y(y) = 1.024 f_X(y) = 1.024 \cdot \frac{1}{\sqrt{2\pi}\,0.025}\, e^{-(y-1.4)^2/2(0.025)^2} = \frac{40.96}{\sqrt{2\pi}}\, e^{-800(y-1.4)^2}$$

for $1.35 < y < 1.48$, and $f_Y(y) = 0$ otherwise.

8.107 Let X_1 and X_2 denote the amount of soda put in a bottle by Machines I and II, respectively. We know that $X_1 \sim \mathcal{N}\left(16.21, 0.14^2\right)$ and $X_2 \sim \mathcal{N}\left(16.12, 0.07^2\right)$. Also, let A_1 and A_2 denote the events that a randomly selected bottle was filled by Machines I and II, respectively. Because Machine I fills twice as many bottles as Machine II, we have $P(A_1) = 2/3$ and $P(A_2) = 1/3$.

Let B denote the event that a randomly selected bottle contains less than 15.96 oz of soda. We have

$$P(B \mid A_1) = P(X_1 < 15.96) = \Phi\big((15.96 - 16.21)/0.14\big) = \Phi(-1.79) = 0.0367$$

and

$$P(B \mid A_2) = P(X_2 < 15.96) = \Phi\big((15.96 - 16.12)/0.07\big) = \Phi(-2.29) = 0.0110.$$

Applying Bayes's rule now yields

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} = \frac{\frac{2}{3} \cdot 0.0367}{\frac{2}{3} \cdot 0.0367 + \frac{1}{3} \cdot 0.0110} = 0.870.$$

Thus, 87.0% of the bottles that contain less than 15.96 oz are filled by Machine I.

8.108 Let X denote the number of people who don't take the flight. Then $X \sim \mathcal{B}(42, 0.16)$. Thus,

$$p_X(x) = \binom{42}{x}(0.16)^x(0.84)^{42-x}, \qquad x = 0, 1, \ldots, 42. \tag{$*$}$$

Noting that $np = 6.72$ and $np(1-p) = 5.6448$, we can, in view of Proposition 8.12 on page 445, use the normal approximation

$$p_X(x) \approx \frac{1}{\sqrt{11.2896\pi}}\, e^{-(x-6.72)^2/11.2896}, \qquad x = 0, 1, \ldots, 42. \tag{$**$}$$

a) Here we want $P(X = 5) = p_X(5)$. The values obtained by substituting $x = 5$ into Equation $(*)$ and Relation $(**)$ are 0.1408 and 0.1292, respectively.
b) Here we want $P(9 \le X \le 12) = \sum_{x=9}^{12} p_X(x)$. The values obtained by using Equation $(*)$ and Relation $(**)$ are 0.2087 and 0.2181, respectively.
c) Here we want $P(X \ge 1) = 1 - P(X = 0) = 1 - p_X(0)$. The values obtained by substituting $x = 0$ into Equation $(*)$ and Relation $(**)$ are 0.0007 and 0.0031, respectively. Therefore, the values obtained for the required probability are 0.9993 and 0.9969, respectively.
d) Here we want $P(X \le 2) = \sum_{x=0}^{2} p_X(x)$. The values obtained by using Equation $(*)$ and Relation $(**)$ are 0.0266 and 0.0357, respectively.

Note: In this exercise, the normal approximation does not work that well because n is not large and p is relatively small.

8.109 The population under consideration consists of the residents of Anchorage, AK, the number of which we denote by N; the specified attribute is "owns a cell phone"; the proportion of the population that has the specified attribute is 0.56; and the sample size is 500.
a) Because the sampling is without replacement, the exact distribution of X is hypergeometric with parameters N, 500, and 0.56, that is, $X \sim \mathcal{H}(N, 500, 0.56)$.
b) In view of Proposition 5.6 on page 215, the required binomial distribution is the one with parameters 500 and 0.56, that is, $\mathcal{B}(500, 0.56)$.

c) Note that $np = 500 \cdot 0.56 = 280$ and $np(1-p) = 500 \cdot 0.56 \cdot 0.44 = 123.2$. Thus, the required normal distribution is the one with parameters 280 and 123.2, that is, $\mathcal{N}(280, 123.2)$.
d) Referring to part (c), we find that

$$p_X(x) \approx \frac{1}{\sqrt{246.4\pi}} e^{-(x-280)^2/246.4}, \qquad x = 0, 1, \ldots, 500.$$

In this case, we want $P(X = 280)$. Substituting 280 for x in the previous display, we find that

$$P(X = 280) = p_X(280) \approx 0.0359.$$

e) Here we want $P(278 \le X \le 280)$. Referring to part (d), we find that

$$P(278 \le X \le 280) = \sum_{x=278}^{280} p_X(x) \approx 0.0354 + 0.0358 + 0.0359 = 0.1071.$$

Thus, $P(278 \le X \le 280) \approx 0.107$.

8.110 Let X denote the number who have consumed alcohol within the last month. We have $n = 250$ and $p = 0.527$. The problem is to approximate $P(X = 125)$ by using the (local) normal approximation. Note that $np = 250 \cdot 0.527 = 131.75$ and $np(1-p) = 250 \cdot 0.527 \cdot 0.473 = 62.31775$. Thus, the approximating normal distribution is $\mathcal{N}(131.75, 62.31775)$. Referring now to Relation (8.39) on page 445 gives $P(X = 125) = p_X(125) \approx 0.0351$.

8.111 Let X denote the number who have Fragile X Syndrome. We have $n = 10{,}000$ and $p = 1/1500$. Observe that $np = 6.667$ and $np(1-p) = 6.662$. Consequently, the approximating normal distribution is $\mathcal{N}(6.667, 6.662)$.
a) Here we want $P(X > 7)$. Applying the (local) normal approximation gives

$$P(X > 7) = 1 - P(X \le 7) = 1 - \sum_{x=0}^{7} p_X(x) \approx 1 - 0.625 = 0.375.$$

b) Here we want $P(X \le 10)$. Applying the (local) normal approximation gives

$$P(X \le 10) = \sum_{x=0}^{10} p_X(x) \approx 0.930.$$

c) The Poisson approximation is preferable to the normal approximation because the assumptions for the former (large n and small p) are better met than those for the latter (large n and moderate p).

8.112 We first note that X satisfies the conditions of Exercise 8.83(d). Thus, the median of X is the unique number, M, that satisfies the equation $F_X(M) = 1/2$. From Equation (8.35) on page 443, we then have $0.5 = F_X(M) = \Phi\big((M-\mu)/\sigma\big)$. Referring to Table I now shows that $(M-\mu)/\sigma = 0$, or $M = \mu$.

Theory Exercises

8.113 First observe that $\phi(-t) = \phi(t)$ for all $t \in \mathcal{R}$. Therefore, making the substitution $u = -t$, we find that

$$\Phi(-z) = \int_{-\infty}^{-z} \phi(t)\,dt = \int_{z}^{\infty} \phi(-u)\,du = \int_{z}^{\infty} \phi(u)\,du = 1 - \int_{-\infty}^{z} \phi(u)\,du = 1 - \Phi(z)$$

for all $z \in \mathcal{R}$.

Advanced Exercises

8.114
a) First note that, for all $t \in \mathcal{R}$,

$$\phi'(t) = \frac{d}{dt}\left(\frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\right) = \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, \frac{d}{dt}\left(-t^2/2\right) = -t\phi(t).$$

Therefore, for $x > 0$,

$$1 - \Phi(x) = \int_x^\infty \phi(t)\, dt \le \int_x^\infty \frac{t}{x}\phi(t)\, dt = -\frac{1}{x}\int_x^\infty \phi'(t)\, dt = -\frac{1}{x}\big(\phi(\infty) - \phi(x)\big) = \frac{1}{x}\phi(x).$$

For the other inequality, we again use the equality $\phi'(t) = -t\phi(t)$ and integrate by parts twice to obtain

$$\begin{aligned}
1 - \Phi(x) &= \int_x^\infty \phi(t)\, dt = -\int_x^\infty \frac{\phi'(t)}{t}\, dt = \frac{\phi(x)}{x} - \int_x^\infty \frac{\phi(t)}{t^2}\, dt \\
&= \frac{\phi(x)}{x} + \int_x^\infty \frac{\phi'(t)}{t^3}\, dt = \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} + 3\int_x^\infty \frac{\phi(t)}{t^4}\, dt \\
&\ge \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} = \left(\frac{1}{x} - \frac{1}{x^3}\right)\phi(x).
\end{aligned}$$

b) Dividing the inequalities from part (a) by $x^{-1}\phi(x)$ yields the inequalities

$$1 - \frac{1}{x^2} \le \frac{1 - \Phi(x)}{x^{-1}\phi(x)} \le 1.$$

Therefore,

$$1 = \lim_{x\to\infty}\left(1 - \frac{1}{x^2}\right) \le \lim_{x\to\infty} \frac{1 - \Phi(x)}{x^{-1}\phi(x)} \le 1,$$

so that $1 - \Phi(x) \sim x^{-1}\phi(x)$ as $x \to \infty$.

8.6 Other Important Continuous Random Variables

Basic Exercises

8.115 As $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\, dx$, and recalling from calculus that $d\left(x^{t-1}\right)/dt = (\ln x)x^{t-1}$, we have

$$\Gamma'(t) = \int_0^\infty (\ln x)x^{t-1}e^{-x}\, dx$$

and

$$\Gamma''(t) = \int_0^\infty (\ln x)^2 x^{t-1}e^{-x}\, dx.$$

From this latter equation, it follows that $\Gamma''(t) > 0$ for all $t > 0$ and, hence, the graph of the gamma function is always concave up. Noting also that $\Gamma(t) > 0$ for all $t > 0$ and that $\Gamma(1) = \Gamma(2) = 1$, we conclude that the gamma function has a unique minimum, which occurs somewhere between $t = 1$

and $t = 2$. Next we note that

$$\Gamma(t) > \int_0^1 x^{t-1} e^{-x}\,dx > e^{-1} \int_0^1 x^{t-1}\,dx = \frac{1}{et}.$$

Consequently, $\Gamma(t) \to \infty$ as $t \to 0^+$. Hence, the graph of the gamma function is as shown in Figure 8.17 on page 451.

8.116 Using the continuity of the exponential function and the fact that the sum, $\sum_{j=0}^{r-1} (\lambda x)^j / j!$, is a polynomial in x, it follows easily that F is continuous for $x > 0$. Also, it is easy to see that $F(0) = 0$. Therefore, F is an everywhere continuous function. In particular, then, property (b) of Proposition 8.1 is satisfied. Property (c) is obvious from the definition of F. To show that F is nondecreasing, we recall from page 453 that

$$F'(x) = \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, \qquad x > 0.$$

From this equation, it is clear that $F'(x) > 0$ for $x > 0$ and, hence, that F is nondecreasing. Consequently, property (a) is satisfied. Finally, we must verify property (d). To do so, it suffices to show that $\lim_{x\to\infty} e^{-\lambda x} x^j = 0$ for each $0 \le j \le r-1$, which is easy to do by applying L'Hôpital's rule.

8.117

a) The Erlang distribution with parameters r and λ is just the gamma distribution with those two parameters. Thus, a PDF is given by

$$f(x) = \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, \qquad x > 0,$$

and $f(x) = 0$ otherwise.

b) The chi-square distribution with ν degrees of freedom is just the gamma distribution with parameters $\nu/2$ and $1/2$. Thus, a PDF is given by

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \qquad x > 0,$$

and $f(x) = 0$ otherwise.

8.118

a) Let $Y = X^2$. For $y > 0$,

$$F_Y(y) = P(Y \le y) = P\left(X^2 \le y\right) = P\left(-\sqrt{y} \le X \le \sqrt{y}\right) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right).$$

Noting that f_X is symmetric about 0, differentiation now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{2} y^{-1/2} f_X\left(\sqrt{y}\right) + \frac{1}{2} y^{-1/2} f_X\left(-\sqrt{y}\right) = y^{-1/2} f_X\left(\sqrt{y}\right) \\ &= y^{-1/2} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(\sqrt{y})^2/2\sigma^2} = \frac{\left(1/2\sigma^2\right)^{1/2}}{\Gamma(1/2)} y^{1/2-1}\, e^{-(1/2\sigma^2)y}. \end{aligned}$$

Consequently, $X^2 \sim \Gamma\left(1/2, 1/2\sigma^2\right)$.

b) We can proceed as in part (a). Alternatively, we can first apply Proposition 8.10 on page 441 to conclude that $X/\sigma \sim \mathcal{N}(0,1)$ and then apply part (a) to conclude that $X^2/\sigma^2 = (X/\sigma)^2 \sim \Gamma(1/2, 1/2)$. Thus, X^2/σ^2 has the chi-square distribution with 1 degree of freedom.

8.119 Because $Y \sim \mathcal{B}(m+n-1, x)$, we have

$$P(Y=j) = \binom{m+n-1}{j} x^j (1-x)^{m+n-1-j}, \qquad j = 0, 1, \ldots, m+n-1.$$

Referring to Equation (8.55) on page 457 now yields

$$P(Y \geq m) = \sum_{j=m}^{m+n-1} \binom{m+n-1}{j} x^j (1-x)^{m+n-1-j} = F_X(x) = P(X \leq x).$$

8.120 Let $Y = X^{-1} - 1$ and note that the range of Y is $(0, \infty)$ because the range of X is $(0,1)$. For $0 \leq y < \infty$,

$$F_Y(y) = P(Y \leq y) = P\left(X^{-1} - 1 \leq y\right) = P\big(X \geq 1/(y+1)\big) = 1 - F_X\big(1/(y+1)\big).$$

Differentiation now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{(y+1)^2} f_X\big(1/(y+1)\big) = \frac{1}{(y+1)^2} \frac{1}{B(\alpha,\beta)} \left(\frac{1}{y+1}\right)^{\alpha-1} \left(1 - \frac{1}{y+1}\right)^{\beta-1} \\ &= \frac{1}{B(\alpha,\beta)} \frac{1}{(y+1)^{\alpha+1}} \frac{y^{\beta-1}}{(y+1)^{\beta-1}} = \frac{y^{\beta-1}}{B(\alpha,\beta)(y+1)^{\alpha+\beta}} \end{aligned}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

8.121 We used basic calculus to find the derivative of a beta PDF and obtained the following:

$$f'(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-2}(1-x)^{\beta-2}\big((\alpha-1)(1-x) - (\beta-1)x\big), \tag{$*$}$$

which can also be expressed as

$$f'(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-2}(1-x)^{\beta-2}\big((2-\alpha-\beta)x + \alpha - 1\big). \tag{$**$}$$

Now we proceed as follows:

- If $\alpha = \beta = 1$, then $B(\alpha,\beta) = B(1,1) = \Gamma(1)\Gamma(1)/\Gamma(2) = 0!0!/1! = 1$. Therefore,

$$f(x) = \frac{1}{B(1,1)} x^{1-1}(1-x)^{1-1} = 1, \qquad 0 < x < 1,$$

and $f(x) = 0$ otherwise. This function is the PDF of the uniform distribution on $(0,1)$.

- If $\alpha < 1$, then $\alpha - 1 < 0$, so that $x^{\alpha-1} \to \infty$ as $x \to 0^+$. Consequently,

$$\lim_{x \to 0^+} f(x) = \lim_{x \to 0^+} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)} \cdot \infty \cdot 1 = \infty.$$

- If $\beta < 1$, then $\beta - 1 < 0$, so that $(1-x)^{\beta-1} \to \infty$ as $x \to 1^-$. Consequently,

$$\lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)} \cdot 1 \cdot \infty = \infty.$$

- If $\alpha < 1$ and $\beta < 1$, then we know from the previous two bulleted items that $f(x) \to \infty$ as either $x \to 0^+$ or $x \to 1^-$. Also, from Equation $(**)$, $f'(x) = 0$ if and only if $x = (1-\alpha)/(2-\alpha-\beta)$, which lies strictly between 0 and 1. Thus, f takes its unique minimum at this point. We leave it to the reader to show that, on the interval $(0,1)$, the second derivative of f is always positive and, hence, the graph of f is concave up.

- If $\alpha < 1$ and $\beta \geq 1$, then Equation (*) shows that $f'(x) < 0$ for all $x \in (0, 1)$. Thus, f is decreasing thereon.
- If $\alpha \geq 1$ and $\beta < 1$, then Equation (*) shows that $f'(x) > 0$ for all $x \in (0, 1)$. Thus, f is increasing thereon.
- If $\alpha > 1$ and $\beta > 1$, then, from Equation (**), $f'(x) = 0$ if and only if $x = (\alpha - 1)/(\alpha + \beta - 2)$, which lies strictly between 0 and 1. As $f(0) = f(1) = 0$, f takes its unique maximum at this point.
- If $\alpha = \beta$, then

$$\begin{aligned}
f(0.5 - x) &= \frac{1}{B(\alpha,\beta)}(0.5 - x)^{\alpha-1}\big(1 - (0.5 - x)\big)^{\beta-1} = \frac{1}{B(\alpha,\beta)}(0.5 - x)^{\alpha-1}(0.5 + x)^{\beta-1} \\
&= \frac{1}{B(\alpha,\beta)}(0.5 + x)^{\alpha-1}(0.5 - x)^{\beta-1} = \frac{1}{B(\alpha,\beta)}(0.5 + x)^{\alpha-1}\big(1 - (0.5 + x)\big)^{\beta-1} \\
&= f(0.5 + x).
\end{aligned}$$

Consequently, f is symmetric about $x = 0.5$.

8.122 Let $Y = X^{1/\alpha}$. For $0 \leq y < 1$,

$$F_Y(y) = P(Y \leq y) = P\left(X^{1/\alpha} \leq y\right) = P\left(X \leq y^\alpha\right) = F_X\left(y^\alpha\right) = y^\alpha.$$

Differentiation yields $f_Y(y) = \alpha y^{\alpha-1}$ if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Because $B(\alpha, 1) = 1/\alpha$, we see that $X^{1/\alpha}$ has the beta distribution with parameters α and 1.

8.123 Let X denote the proportion of manufactured items that require service within 5 years. From Equation (8.55) on page 457, the CDF of X is given by

$$F_X(x) = \sum_{j=2}^{4} \binom{4}{j} x^j (1 - x)^{4-j} = 6x^2(1 - x)^2 + 4x^3(1 - x) + x^4, \qquad 0 \leq x < 1.$$

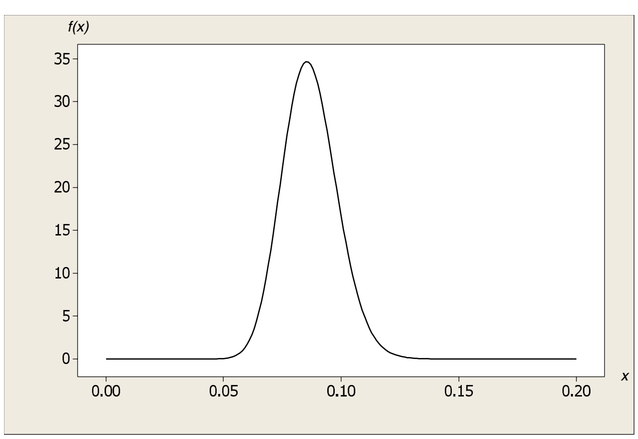
We use this expression to solve parts (a) and (b).

a) $P(X \leq 0.3) = F_X(0.3) = 0.3483$.

b) $P(0.1 < X < 0.20) = F_X(0.2) - F_X(0.1) = 0.1808 - 0.0523 = 0.1285$.

8.124 Let f and F denote the PDF and CDF, respectively, of a beta distribution with parameters 51.3 and 539.75.

a) A graph of the PDF is as follows:



b) $F(0.10) - F(0.07) = 0.8711 - 0.0661 = 0.805$.

8.125 Let $X \sim \mathcal{T}(-1, 1)$. We want to use a linear transformation, $y = c + dx$, that takes the interval $(-1, 1)$ to the interval (a, b). Thus, we need to choose c and d so that $c - d = a$ and $c + d = b$. These relations imply that $c = (b + a)/2$ and $d = (b - a)/2$. Now let $Y = c + dX$. For $a \le y < b$,

$$F_Y(y) = P(Y \le y) = P\big(c + dX \le y\big) = P\big(X \le (y - c)/d\big) = F_X\big((y - c)/d\big).$$

Differentiating yields

$$f_Y(y) = \frac{1}{d} f_X\big((y - c)/d\big) = \frac{2}{b - a} f_X\left(\frac{2}{b - a}\big(y - (a + b)/2\big)\right)$$

for $a < y < b$, and $f_Y(y) = 0$ otherwise. Finally, we show that Y actually has the $\mathcal{T}(a, b)$ distribution. Referring to Equation (8.57) on page 460, we see that, for $a < y < b$,

$$\begin{aligned} f_Y(y) &= \frac{2}{b - a} f_X\left(\frac{2}{b - a}\big(y - (a + b)/2\big)\right) \\ &= \frac{2}{b - a}\left(1 - \left|\frac{2}{b - a}\big(y - (a + b)/2\big)\right|\right) \\ &= \frac{2}{b - a}\left(1 - \frac{|a + b - 2y|}{b - a}\right). \end{aligned}$$

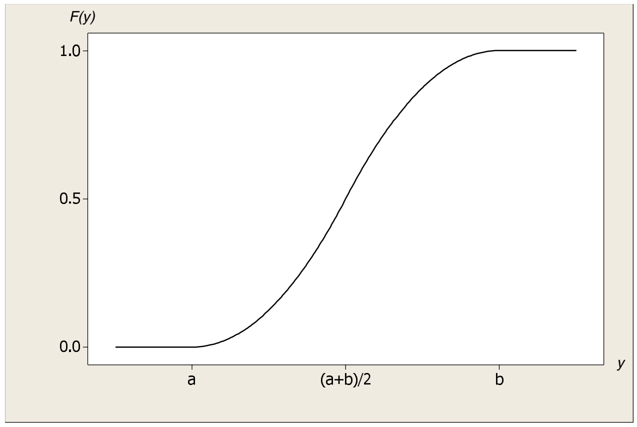
8.126 From Exercise 8.49(e), we find that the CDF of a random variable X with a $\mathcal{T}(-1, 1)$ distribution is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < -1; \\ \frac{1}{2}(1 + x)^2, & \text{if } -1 \le x < 0; \\ 1 - \frac{1}{2}(1 - x)^2, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

Referring now to to the solution of Exercise 8.125, we see that the CDF of a random variable Y with a $\mathcal{T}(a, b)$ distribution is given by

$$\begin{aligned} F_Y(y) &= \begin{cases} 0, & \text{if } y < a; \\ \dfrac{1}{2}\left(1 + \dfrac{2}{b - a}\big(y - (a + b)/2\big)\right)^2, & \text{if } a \le y < (a + b)/2; \\ 1 - \dfrac{1}{2}\left(1 - \dfrac{2}{b - a}\big(y - (a + b)/2\big)\right)^2, & \text{if } (a + b)/2 \le y < b; \\ 1, & \text{if } y \ge b. \end{cases} \\ &= \begin{cases} 0, & \text{if } y < a; \\ 2\left(\dfrac{y - a}{b - a}\right)^2, & \text{if } a \le y < (a + b)/2; \\ 1 - 2\left(\dfrac{b - y}{b - a}\right)^2, & \text{if } (a + b)/2 \le y < b; \\ 1, & \text{if } y \ge b. \end{cases} \end{aligned}$$

Following is a graph of this CDF:



8.127

a) For $x > 0$,

$$f_X(x) = \frac{(0.1)^5}{\Gamma(5)} x^{5-1} e^{-0.1x} = \frac{1}{24 \times 10^5}\, x^4 e^{-x/10},$$

and $f_X(x) = 0$ otherwise. Also, referring to Equation (8.49) on page 453, we see that

$$F_X(x) = 1 - e^{-x/10} \sum_{j=0}^{4} \frac{(x/10)^j}{j!}, \qquad x \geq 0,$$

and $F_X(x) = 0$ otherwise.

b) Referring to part (a), we find that

$$P(X \leq 60) = F_X(60) = 1 - e^{-6} \sum_{j=0}^{4} \frac{6^j}{j!} = 0.715.$$

c) Again referring to part (a), we get

$$P(40 \leq X \leq 50) = F_X(50) - F_X(40) = 0.5595 - 0.3712 = 0.1883.$$

8.128

a) We have $F_Y(y) = 0$ if $y < \beta$. For $y \geq \beta$,

$$F_Y(y) = P(Y \leq y) = P\big(\beta e^X \leq y\big) = P\big(X \leq \ln(y/\beta)\big) = 1 - e^{-\alpha \ln(y/\beta)} = 1 - (\beta/y)^\alpha.$$

b) Referring to part (a), we find that a PDF of Y is given by

$$f_Y(y) = \frac{d}{dy}\big(1 - (\beta/y)^\alpha\big) = \alpha\beta^\alpha / y^{\alpha+1}$$

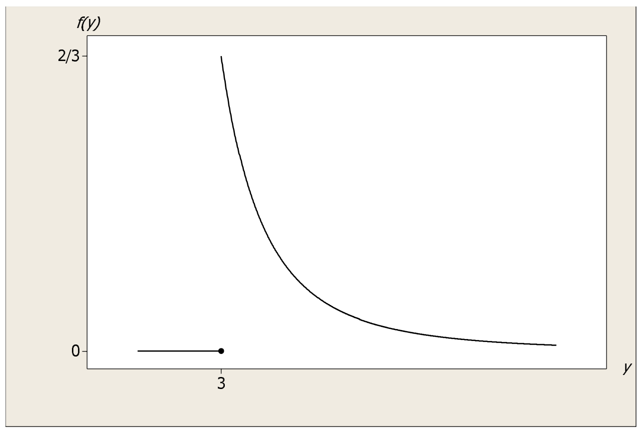
if $y > \beta$, and $f_Y(y) = 0$ otherwise.

8.129 Let Y denote loss amount in thousands of dollars.
a) From the solution to Exercise 8.128(b),

$$f_Y(y) = \begin{cases} 2 \cdot 3^2/y^3, & \text{if } y > 3; \\ 0, & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 18/y^3, & \text{if } y > 3; \\ 0, & \text{otherwise.} \end{cases}$$

Note: In the following graph, we use $f(y)$ instead of $f_Y(y)$.

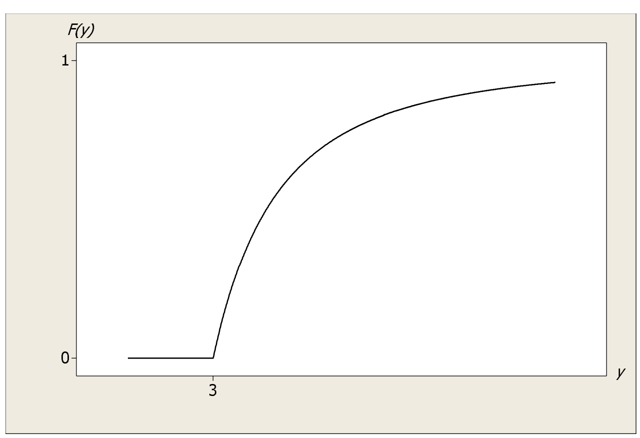


Also, from the solution to Exercise 8.128(a),

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 3; \\ 1 - (3/y)^2, & \text{if } y \geq 3. \end{cases}$$

$$= \begin{cases} 0, & \text{if } y < 3; \\ 1 - 9/y^2, & \text{if } y \geq 3. \end{cases}$$

Note: In the following graph, we use $F(y)$ instead of $F_Y(y)$.



b) Referring to part (a), we get

$$P(Y > 8) = 1 - P(Y \le 8) = 1 - F_Y(8) = 9/8^2 = 0.141.$$

c) Again referring to part (a), we get

$$P(4 \le Y \le 5) = F_Y(5) - F_Y(4) = 9/4^2 - 9/5^2 = 0.2025.$$

8.130

a) We have $F_Y(y) = 0$ if $y < 0$. For $y \ge 0$,

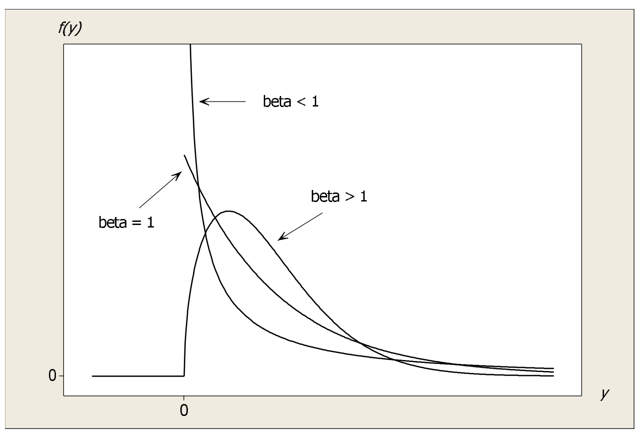
$$F_Y(y) = P(Y \le y) = P\big(X^{1/\beta} \le y\big) = P\left(X \le y^\beta\right) = 1 - e^{-\alpha y^\beta}.$$

b) Referring to part (a), we find that a PDF of Y is given by

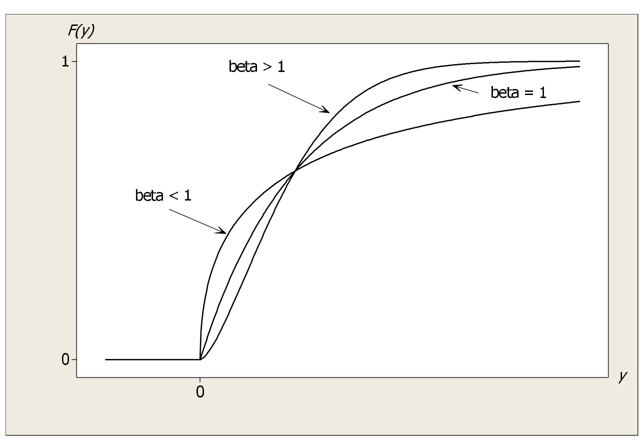
$$f_Y(y) = \frac{d}{dy}\left(1 - e^{-\alpha y^\beta}\right) = \alpha\beta y^{\beta-1} e^{-\alpha y^\beta}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

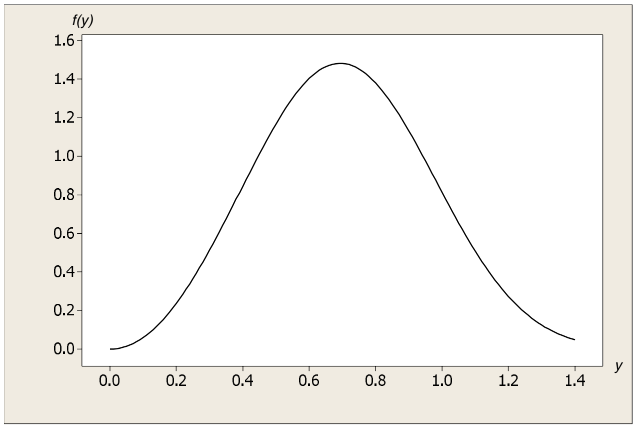
c) The graphs of the PDFs are as follows:



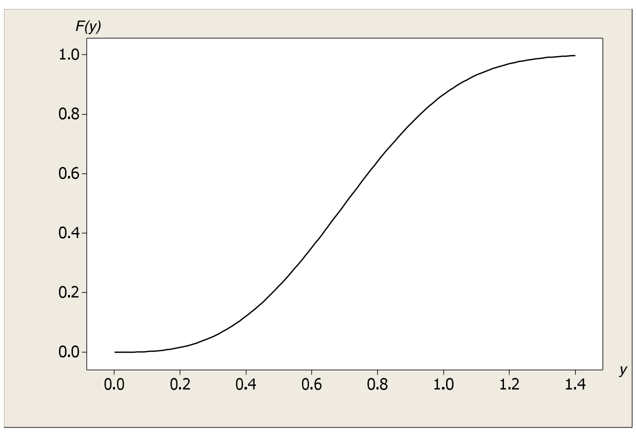
The graphs of the CDFs are as follows:



8.131 Let Y denote component lifetime, in years. Referring to the solution to Exercise 8.130, we get the following with $\alpha = 2$ and $\beta = 3$.
a) A PDF of Y is $f_Y(y) = 2 \cdot 3y^{3-1}e^{-2y^3} = 6y^2e^{-2y^3}$ if $y > 0$, and $f_Y(y) = 0$ otherwise. Its graph is:



The CDF of Y is $F_Y(y) = 0$ if $y \leq 0$, and $F_Y(y) = 1 - e^{-2y^3}$ if $y \geq 0$. Its graph is:



b) The probability that the component lasts at most 6 months is

$$P(Y \leq 0.5) = F_Y(0.5) = 1 - e^{-2(0.5)^3} = 0.221.$$

The probability that it lasts between 6 months and 1 year is

$$P(0.5 \leq Y \leq 1) = F_Y(1) - F_Y(0.5) = e^{-2(0.5)^3} - e^{-2 \cdot 1^3} = 0.643.$$

8.132
a) Answers will vary.
b) We see from Figure 8.19(b) on page 458 that the PDF of X is larger for values of x near 0 or 1 than for values near 1/2. Thus, X is more likely to be near 0 or 1 than near 1/2.
c) First note that

$$B(1/2, 1/2) = \frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)} = \left(\sqrt{\pi}\right)^2 = \pi.$$

For $0 \le x < 1$, we make the substitution $u = \sqrt{t}$ to get

$$F_X(x) = P(X \le x) = \int_0^x f_X(t)\,dt = \frac{1}{\pi}\int_0^x t^{1/2-1}(1-t)^{1/2-1}\,dt$$
$$= \frac{2}{\pi}\int_0^{\sqrt{x}} \frac{du}{\sqrt{1-u^2}} = \frac{2}{\pi}\arcsin\sqrt{x}.$$

d) We have

$$P(0.4 \le X \le 0.6) = F_X(0.6) - F_X(0.4) = \frac{2}{\pi}\arcsin\sqrt{0.6} - \frac{2}{\pi}\arcsin\sqrt{0.4} = 0.128$$

and

$$P(X \le 0.1 \text{ or } X \ge 0.9) = P(X \le 0.1) + P(X \ge 0.9) = F_X(0.1) + 1 - F_X(0.9)$$
$$= \frac{2}{\pi}\arcsin\sqrt{0.1} + 1 - \frac{2}{\pi}\arcsin\sqrt{0.9} = 0.410.$$

These two probabilities illustrate the fact alluded to in part (b), namely, that X is more likely to be near 0 or 1 than near 1/2.

Theory Exercises

8.133 Using integration by parts with $u = x^t$ and $dv = e^{-x}\,dx$ gives

$$\Gamma(t+1) = \int_0^\infty x^t e^{-x}\,dx = -x^t e^{-x}\Big|_0^\infty + \int_0^\infty t x^{t-1} e^{-x}\,dx$$
$$= 0 - 0 + t\int_0^\infty x^{t-1}e^{-x}\,dx = t\Gamma(t).$$

8.134 Using the already-established fact that $\Gamma(1) = 0! = (1-1)!$ and referring to Equation (8.43) on page 451, we proceed inductively to get

$$\Gamma(n+1) = n\Gamma(n) = n \cdot (n-1)! = n!.$$

Thus, $\Gamma(n) = (n-1)!$ for all $n \in \mathcal{N}$.

8.135 We have

$$\Gamma\big(0 + \tfrac{1}{2}\big) = \Gamma(1/2) = \sqrt{\pi} = \frac{(2\cdot 0)!}{0!\,2^{2\cdot 0}}\sqrt{\pi}.$$

Proceeding inductively and using Exercise 8.133, we get

$$\Gamma\big((n+1)+\tfrac{1}{2}\big) = \Gamma\Big(\big(n+\tfrac{1}{2}\big)+1\Big) = \big(n+\tfrac{1}{2}\big)\Gamma\big(n+\tfrac{1}{2}\big) = \big(n+\tfrac{1}{2}\big)\frac{(2n)!}{n!\,2^{2n}}\sqrt{\pi}$$
$$= (2n+1)\frac{(2n)!}{n!\,2^{2n+1}}\sqrt{\pi} = \frac{(2n+2)(2n+1)(2n)!}{2(n+1)n!\,2^{2n+1}}\sqrt{\pi} = \frac{\big(2(n+1)\big)!}{(n+1)!\,2^{2(n+1)}}\sqrt{\pi}.$$

Advanced Exercises

8.136

a) By the conditional probability rule, for $\Delta t > 0$,

$$P(T \le t + \Delta t \mid T > t) = \frac{P(t < T \le t + \Delta t)}{P(T > t)} = \frac{F_T(t+\Delta t) - F_T(t)}{1 - F_T(t)}.$$

Therefore,

$$\lim_{\Delta t \to 0} \frac{P(T \le t + \Delta t \mid T > t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\big(F_T(t+\Delta t) - F_T(t)\big)/\Delta t}{1 - F_T(t)}$$

$$= \frac{F_T'(t)}{1 - F_T(t)} = \frac{f_T(t)}{1 - F_T(t)} = h_T(t).$$

The left side of the previous equation gives the instantaneous risk of item failure at time t.

b) We have

$$h_T(u) = \frac{f_T(u)}{1 - F_T(u)} = \frac{F_T'(u)}{1 - F_T(u)} = -\frac{d}{du} \ln\big(1 - F_T(u)\big).$$

Integrating from 0 to t and using the fact that T is a positive continuous random variable, which implies that $F_T(0) = 0$, we get

$$\int_0^t h_T(u)\, du = -\ln\big(1 - F_T(t)\big).$$

The required result now follows easily.

c) Applying the conditional probability rule and referring to part (b) yields, for $s, t > 0$,

$$P(T > s + t \mid T > t) = \frac{P(T > s + t)}{P(T > t)} = \frac{1 - F_T(s+t)}{1 - F_T(t)} = \frac{e^{-\int_0^{s+t} h_T(u)\, du}}{e^{-\int_0^t h_T(u)\, du}}$$

$$= e^{-\left(\int_0^{s+t} h_T(u)\, du - \int_0^t h_T(u)\, du\right)} = e^{-\int_t^{s+t} h_T(u)\, du}.$$

d) Suppose that h_T is a decreasing function, then $\int_t^{s+t} h_T(u)\, du$ is a decreasing function of t. This, in turn, implies that $-\int_t^{s+t} h_T(u)\, du$ is an increasing function, so that $e^{-\int_t^{s+t} h_T(u)\, du}$ is an increasing function. Similarly, if h_T is an increasing function, then $e^{-\int_t^{s+t} h_T(u)\, du}$ is a decreasing function of t. In view of part (c), these two results imply that, if h_T is a decreasing function, then the item improves with age, whereas if h_T is an increasing function, then the item deteriorates with age.

e) The item neither improves nor deteriorates with age; that is, it is "memoryless."

8.137

a) Because of the lack-of-memory property of an exponential random variable, the hazard-rate function should be constant.

b) We have

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = \frac{\lambda e^{-\lambda t}}{1 - \big(1 - e^{-\lambda t}\big)} = \lambda.$$

Therefore, an item whose lifetime has an exponential distribution neither improves nor deteriorates with age; that is, it is "memoryless."

8.138 From the solution to Exercise 8.130, $F_T(t) = 1 - e^{-\alpha t^\beta}$ and $f_T(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^\beta}$ for $t > 0$. Therefore,

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = \frac{\alpha\beta t^{\beta-1} e^{-\alpha t^\beta}}{e^{-\alpha t^\beta}} = \alpha\beta t^{\beta-1}.$$

From this result, we see that the hazard function is decreasing, constant, or increasing when $\beta < 1$, $\beta = 1$, and $\beta > 1$, respectively. It now follows easily from Exercises 8.136(d) and (e) that the item improves with age, is memoryless, or deteriorates with age depending on whether $\beta < 1$, $\beta = 1$, and $\beta > 1$, respectively.

8.139 Referring to Exercise 8.136(b), we find that

$$f_T(t) = F_T'(t) = e^{-\int_0^t h_T(u)\,du} \frac{d}{dt}\left(\int_0^t h_T(u)\,du\right) = h_T(t)e^{-\int_0^t h_T(u)\,du}.$$

a) If $h_T(t) = \lambda$, then

$$f_T(t) = h_T(t)e^{-\int_0^t h_T(u)\,du} = \lambda e^{-\int_0^t \lambda\,du} = \lambda e^{-\lambda t},$$

so that $T \sim \mathcal{E}(\lambda)$.
b) If $h_T(t) = \alpha + \beta t$, then

$$f_T(t) = h_T(t)e^{-\int_0^t h_T(u)\,du} = (\alpha+\beta t)e^{-\int_0^t (\alpha+\beta u)\,du} = (\alpha+\beta t)e^{-(\alpha t+\beta t^2/2)}.$$

8.140 Let T_{I} and T_{II} denote the lifetimes of Items I and II, respectively.
a) Answers will vary. Many people guess that the probability for Item II is one-half of that for Item I but, as we will see in part (b), that is incorrect.
b) By assumption, $h_{T_{\mathrm{II}}} = 2h_{T_{\mathrm{I}}}$. Referring now to Exercise 8.136(c),

$$\begin{aligned} P(T_{\mathrm{II}} > t+s \mid T_{\mathrm{II}} > t) &= e^{-\int_t^{t+s} h_{T_{\mathrm{II}}}(u)\,du} = e^{-\int_t^{t+s} 2h_{T_{\mathrm{I}}}(u)\,du} \\ &= \left(e^{-\int_t^{t+s} h_{T_{\mathrm{I}}}(u)\,du}\right)^2 = \big(P(T_{\mathrm{I}} > t+s \mid T_{\mathrm{I}} > t)\big)^2. \end{aligned}$$

8.7 Functions of Continuous Random Variables

Basic Exercises

8.141 As $X \sim \mathcal{U}(0,1)$, its PDF is $f_X(x) = 1$ if $0 < x < 1$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0,1)$. Here $g(x) = c + dx$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = (y-c)/d$. Letting $Y = c + dX$, we see that the range of Y is the interval $(c, c+d)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X\big((y-c)/d\big)}{|d|} = \frac{1}{d}$$

if $c < y < c + d$, and $f_Y(y) = 0$ otherwise. So, $c + dX \sim \mathcal{U}(c, c+d)$.

8.142 As $X \sim \mathcal{E}(1)$, its PDF is $f_X(x) = e^{-x}$ if $x > 0$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0,\infty)$. Here $g(x) = bx$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = y/b$. Letting $Y = bX$, we see that the range of Y is the interval $(0,\infty)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(y/b)}{|b|} = \frac{e^{-y/b}}{b} = \frac{1}{b}e^{-\frac{1}{b}y}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. So, $bX \sim \mathcal{E}(1/b)$.

8.143 As $X \sim \mathcal{U}(0,1)$, its PDF is $f_X(x) = 1$ if $0 < x < 1$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0,1)$. Here $g(x) = -\ln x$, which is strictly decreasing and differentiable on the range of X.

The transformation method is appropriate. Note that $g^{-1}(y) = e^{-y}$. Letting $Y = -\ln X$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{|-1/x|} = \frac{f_X\left(e^{-y}\right)}{\left|-1/e^{-y}\right|} = e^{-y}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. So, $-\ln X \sim \mathcal{E}(1)$.

8.144 As $X \sim \mathcal{N}\left(0, 1/\alpha^2\right)$, its PDF is $f_X(x) = \left(\alpha/\sqrt{2\pi}\right) e^{-\alpha^2 x^2/2}$ if $-\infty < x < \infty$. The range of X is the interval $(-\infty, \infty)$. Here $g(x) = 1/x^2$, which is not monotone on the range of X. The transformation method is not appropriate.

8.145 As $X \sim \mathcal{N}\left(0, \sigma^2\right)$, its PDF is $f_X(x) = \left(1/\sqrt{2\pi}\sigma\right)e^{-x^2/2\sigma^2}$ if $-\infty < x < \infty$. The range of X is the interval $(-\infty, \infty)$. Here $g(x) = x^2$ for part (a) and $g(x) = x^2/\sigma^2$ for part (b). In either case, g is not monotone on the range of X. The transformation method is not appropriate.

8.146 As X has the beta distribution with parameters α and β, its PDF is

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1,$$

and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, 1)$. Here $g(x) = x^{-1} - 1$, which is strictly decreasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = 1/(y+1)$. Letting $Y = X^{-1} - 1$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{\left|-1/x^2\right|} = \frac{f_X\big(1/(y+1)\big)}{\left|-1/\big(1/(y+1)\big)^2\right|} \\ &= \frac{1}{(y+1)^2} \frac{1}{B(\alpha, \beta)} \left(\frac{1}{y+1}\right)^{\alpha-1} \left(1 - \frac{1}{y+1}\right)^{\beta-1} \\ &= \frac{1}{B(\alpha, \beta)} \frac{1}{(y+1)^{\alpha+1}} \frac{y^{\beta-1}}{(y+1)^{\beta-1}} = \frac{y^{\beta-1}}{B(\alpha, \beta)(y+1)^{\alpha+\beta}} \end{aligned}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

8.147 As $X \sim \mathcal{U}(0, 1)$, its PDF is $f_X(x) = 1$ if $0 < x < 1$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, 1)$. Here $g(x) = x^{1/\alpha}$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = y^\alpha$. Letting $Y = X^{1/\alpha}$, we see that the range of Y is the interval $(0, 1)$. Applying the transformation method now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{\left|(1/\alpha)x^{1/\alpha-1}\right|} = \frac{f_X\left(y^\alpha\right)}{\left|(1/\alpha)\left(y^\alpha\right)^{1/\alpha-1}\right|} \\ &= \frac{1}{(1/\alpha)y^{1-\alpha}} = \alpha y^{\alpha-1} = \frac{1}{B(\alpha, 1)} y^{\alpha-1}(1-y)^{1-1} \end{aligned}$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. So, $X^{1/\alpha}$ has the beta distribution with parameters α and 1.

8.148 As $X \sim \mathcal{E}(\alpha)$, its PDF is $f_X(x) = \alpha e^{-\alpha x}$ if $x > 0$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, \infty)$. Here $g(x) = \beta e^x$, which is strictly increasing and differentiable on the range of X.

The transformation method is appropriate. Note that $g^{-1}(y) = \ln(y/\beta)$. Letting $Y = \beta e^X$, we see that the range of Y is the interval (β, ∞). Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{|\beta e^x|} = \frac{f_X\big(\ln(y/\beta)\big)}{\left|\beta e^{\ln(y/\beta)}\right|} = \frac{\alpha e^{-\alpha \ln(y/\beta)}}{y} = \frac{\alpha (y/\beta)^{-\alpha}}{y} = \frac{\alpha\beta^\alpha}{y^{\alpha+1}}$$

if $y > \beta$, and $f_Y(y) = 0$ otherwise.

8.149 As $X \sim \mathcal{E}(\alpha)$, its PDF is $f_X(x) = \alpha e^{-\alpha x}$ if $x > 0$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, \infty)$. Here $g(x) = x^{1/\beta}$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = y^\beta$. Letting $Y = X^{1/\beta}$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{\left|(1/\beta)x^{1/\beta-1}\right|} = \frac{f_X\left(y^\beta\right)}{\left|(1/\beta)\left(y^\beta\right)^{1/\beta-1}\right|} = \frac{\alpha e^{-\alpha y^\beta}}{(1/\beta)y^{1-\beta}} = \alpha\beta y^{\beta-1} e^{-\alpha y^\beta}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

8.150 As $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, its PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

Letting $Y = a + bX$, we see that the range of Y is $(-\infty, \infty)$. Assume that $b < 0$. For $y \in \mathcal{R}$,

$$F_Y(y) = P(a + bX \le y) = P\big(X \ge (y-a)/b\big) = 1 - F_X\big((y-a)/b\big).$$

Differentiation now yields

$$\begin{aligned} f_Y(y) &= -f_X\big((y-a)/b\big) \cdot \frac{1}{b} = -\frac{1}{b}\frac{1}{\sqrt{2\pi}\,\sigma} e^{-((y-a)/b-\mu)^2/2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}\,|b|\sigma} e^{-(y-a-b\mu)^2/2b^2\sigma^2} = \frac{1}{\sqrt{2\pi}\,(|b|\sigma)} e^{-(y-(a+b\mu))^2/2(b\sigma)^2} \end{aligned}$$

for all $y \in \mathcal{R}$. Thus, $a + bX \sim \mathcal{N}\left(a + b\mu, (b\sigma)^2\right)$. A similar argument shows that this same result holds when $b > 0$.

8.151 Here $g(x) = a + bx$, which is strictly decreasing if $b < 0$ and strictly increasing if $b > 0$. Note that g is differentiable on all of $\mathcal{R}$. The transformation method is appropriate. We have $g^{-1}(y) = (y-a)/b$. Applying the transformation method now yields

$$f_{a+bX}(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X\big((y-a)/b\big)}{|b|} = \frac{1}{|b|} f_X\big((y-a)/b\big).$$

8.152 As $X \sim \mathcal{E}(1)$, its PDF is $f_X(x) = e^{-x}$ if $x > 0$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, \infty)$. Here $g(x) = 10x^{0.8}$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = (0.1y)^{1.25}$. Letting $Y = 10X^{0.8}$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{\left|8x^{-0.2}\right|} = \frac{f_X\left((0.1y)^{1.25}\right)}{\left|8\left((0.1y)^{1.25}\right)^{-0.2}\right|}$$

$$= \frac{e^{-(0.1y)^{1.25}}}{8(0.1y)^{-0.25}} = 0.125(0.1y)^{0.25} e^{-(0.1y)^{1.25}}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

8.153

a) Referring to Definition 8.11 on page 470, we find that, for $x \in \mathcal{R}$,

$$\frac{1}{\theta}\psi\left((x-\eta)/\theta\right) = \frac{1}{\theta}\frac{1}{\pi\left(1 + \left((x-\eta)/\theta\right)^2\right)} = \frac{1}{\theta\pi\left(1 + (x-\eta)^2/\theta^2\right)}$$

$$= \frac{\theta}{\pi\left(\theta^2 + (x-\eta)^2\right)} = f_X(x).$$

b) Let $Y = (X-\eta)/\theta = -\eta/\theta + (1/\theta)X$. Referring to part (a) and applying Equation (8.64) on page 469 with $a = -\eta/\theta$ and $b = 1/\theta$ yields

$$f_Y(y) = \frac{1}{|1/\theta|} f_X\left(\frac{y - (-\eta/\theta)}{1/\theta}\right) = \theta f_X(\theta y + \eta) = \theta \frac{1}{\theta}\psi\left(\frac{(\theta y + \eta) - \eta}{\theta}\right) = \psi(y).$$

Therefore, $(X-\eta)/\theta$ has the standard Cauchy distribution.

8.154

a) As $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, its PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

Here $g(x) = e^x$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = \ln y$. Letting $Y = e^X$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{|e^x|} = \frac{f_X(\ln y)}{e^{\ln y}}$$

$$= \frac{\left(\sqrt{2\pi}\,\sigma\right)^{-1} e^{-(\ln y - \mu)^2/2\sigma^2}}{y} = \frac{1}{\sqrt{2\pi}\,\sigma y} e^{-(\ln y - \mu)^2/2\sigma^2}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

b) The (natural) logarithm of the random variable is normally distributed.

8.155

a) As $X \sim \mathcal{U}(-\pi/2, \pi/2)$, its PDF is $f_X(x) = 1/\pi$ if $-\pi/2 < x < \pi/2$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(-\pi/2, \pi/2)$. Here $g(x) = \sin x$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = \arcsin y$. Letting $Y = \sin X$, we see that the range of Y is the interval $(-1, 1)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{|\cos x|} = \frac{1/\pi}{|\cos(\arcsin y)|} = \frac{1}{\pi\sqrt{1-y^2}}$$

if $-1 < y < 1$, and $f_Y(y) = 0$ otherwise.

b) As $X \sim \mathcal{U}(-\pi, \pi)$, its PDF is $f_X(x) = 1/(2\pi)$ if $-\pi < x < \pi$, and $f_X(x) = 0$ otherwise. Note that the transformation method is inappropriate here because $\sin x$ is not monotone on the interval $(-\pi, \pi)$. We instead use the CDF method. If $-1 \le y < 0$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\sin X \le y) = P(-\pi - \arcsin y \le X \le \arcsin y) \\ &= \frac{\big(\arcsin y - (-\pi - \arcsin y)\big)}{2\pi} = \frac{1}{2} + \frac{1}{\pi} \arcsin y. \end{aligned}$$

If $0 \le y < 1$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\sin X \le y) = P(X \le \arcsin y \ \text{ or } \ X \ge \pi - \arcsin y) \\ &= \frac{\big(\arcsin y - (-\pi)\big) + \big(\pi - (\pi - \arcsin y)\big)}{2\pi} = \frac{1}{2} + \frac{1}{\pi} \arcsin y. \end{aligned}$$

Thus,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < -1; \\ \dfrac{1}{2} + \dfrac{1}{\pi} \arcsin y, & \text{if } -1 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

Differentiation now shows that $f_Y(y) = \left(\pi\sqrt{1 - y^2}\right)^{-1}$ if $-1 < y < 1$, and $f_Y(y) = 0$ otherwise.

8.156 Let Θ denote the angle, in radians, made by the line segment connecting the center of the circle to the point chosen. Then $\Theta \sim \mathcal{U}(-\pi, \pi)$.
a) Let X denote the x-coordinate of the point chosen. Then $X = \cos\Theta$. For $-1 < x \le 1$,

$$\begin{aligned} F_X(x) &= P(X \le x) = P(\cos\Theta \le x) = 1 - P(\cos\Theta > x) \\ &= 1 - P(-\arccos x < \Theta < \arccos x) = 1 - \frac{2\arccos x}{2\pi} = 1 - \frac{1}{\pi}\arccos x. \end{aligned}$$

Differentiation now yields

$$f_X(x) = F_X'(x) = \frac{d}{dx}\left(1 - \frac{1}{\pi}\arccos x\right) = \frac{1}{\pi\sqrt{1 - x^2}}$$

if $-1 < x < 1$, and $f_X(x) = 0$ otherwise.
b) Let Y denote the y-coordinate of the point chosen. By symmetry, X and Y have the same distribution. Thus, a PDF of Y is given by $f_Y(y) = \left(\pi\sqrt{1 - y^2}\right)^{-1}$ if $-1 < y < 1$, and $f_Y(y) = 0$ otherwise.
c) Let Z denote the distance of the point chosen to the point $(0, 1)$. Then

$$Z = \sqrt{(X - 1)^2 + Y^2} = \sqrt{(X - 1)^2 + 1 - X^2} = \sqrt{2(1 - X)}.$$

We know that the range of X is the interval $(-1, 1)$. Let $g(x) = \sqrt{2(1 - x)}$, which is strictly decreasing on the range of X. The transformation method is appropriate. Note that $g^{-1}(z) = 1 - z^2/2$.

$$\begin{aligned} f_Z(z) &= \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{\left|-1/\sqrt{2(1 - x)}\right|} = \sqrt{2(1 - x)}\, f_X(x) \\ &= \sqrt{2\Big(1 - \big(1 - z^2/2\big)\Big)}\, f_X\left(1 - z^2/2\right) = \frac{z}{\pi\sqrt{1 - \left(1 - z^2/2\right)^2}} = \frac{2}{\pi\sqrt{4 - z^2}} \end{aligned}$$

if $0 < z < 2$, and $f_Z(z) = 0$ otherwise.

8.157
a) Recall that the range of a beta distribution is the interval $(0, 1)$. Because $a + (b - a) \cdot 0 = a$ and $a + (b - a) \cdot 1 = b$, we see that the range of Y is the interval (a, b).
b) As X has the beta distribution with parameters α and β, its PDF is

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1,$$

and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, 1)$. Here $g(x) = a + (b - a)x$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = (y - a)/(b - a)$. Applying the transformation method now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{b-a} = \frac{1}{b-a} f_X\big((y-a)/(b-a)\big) \\ &= \frac{1}{(b-a)B(\alpha, \beta)} \big((y-a)/(b-a)\big)^{\alpha-1} \big(1 - (y-a)/(b-a)\big)^{\beta-1} \\ &= \frac{1}{(b-a)B(\alpha, \beta)} \frac{(y-a)^{\alpha-1}}{(b-a)^{\alpha-1}} \frac{(b-y)^{\beta-1}}{(b-a)^{\beta-1}} = \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1} B(\alpha, \beta)} \end{aligned}$$

if $a < y < b$, and $f_Y(y) = 0$ otherwise.

8.158 We apply the CDF method. For $y < 0$,

$$F_Y(y) = P(Y \le y) = P(1/X \le y) = P(1/y \le X < 0) = F_X(0) - F_X(1/y).$$

Differentiation yields

$$f_Y(y) = \frac{1}{y^2} f_X(1/y) = \frac{1}{y^2 \pi \big(1 + (1/y)^2\big)} = \frac{1}{\pi(1 + y^2)}.$$

For $y \ge 0$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(1/X \le y) = P(X < 0 \text{ or } X \ge 1/y) \\ &= P(X < 0) + 1 - P(X < y) = F_X(0) + 1 - F_X(1/y). \end{aligned}$$

Differentiation yields

$$f_Y(y) = \frac{1}{y^2} f_X(1/y) = \frac{1}{y^2 \pi \big(1 + (1/y)^2\big)} = \frac{1}{\pi(1 + y^2)}.$$

Thus, we see that $f_Y(y) = 1/\big(\pi(1 + y^2)\big)$ for $-\infty < y < \infty$. In other words, $1/X \sim \mathcal{C}(0, 1)$.

8.159 The range of R is $(0, \infty)$. Here $g(r) = r^2$, which is strictly increasing and differentiable on the range of R. The transformation method is appropriate. Note that $g^{-1}(y) = \sqrt{y}$. Letting $Y = R^2$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(r)|} f_R(r) = \frac{f_R(r)}{|2r|} = \frac{f_R\left(\sqrt{y}\right)}{2\sqrt{y}} = \frac{\left(\sqrt{y}/\sigma^2\right) e^{-(\sqrt{y})^2/2\sigma^2}}{2\sqrt{y}} = \frac{1}{2\sigma^2} e^{-y/2\sigma^2}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. Thus, we see that $R^2 \sim \mathcal{E}\left(1/2\sigma^2\right)$.

8.160

a) As X has the chi-square distribution with n degrees of freedom, its PDF is

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \qquad x > 0,$$

and $f_X(x) = 0$ otherwise. The range of X is the interval $(0, \infty)$. Here $g(x) = \sigma\sqrt{x}$, which is strictly increasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = y^2/\sigma^2$. Letting $Y = \sigma\sqrt{X}$, we see that the range of Y is the interval $(0, \infty)$. Applying the transformation method now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{|g'(x)|} f_X(x) = \frac{f_X(x)}{\left|\sigma/2\sqrt{x}\right|} = \frac{f_X\left(y^2/\sigma^2\right)}{\left|\sigma/2\sqrt{y^2/\sigma^2}\right|} \\ &= \frac{\frac{1}{2^{n/2}\Gamma(n/2)}\left(y^2/\sigma^2\right)^{n/2-1} e^{-(y^2/\sigma^2)/2}}{\sigma^2/2y} = \frac{1}{2^{n/2-1}\Gamma(n/2)\sigma^n} y^{n-1} e^{-y^2/2\sigma^2} \end{aligned}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

b) Let $U \sim \mathcal{N}\left(0, \sigma^2\right)$ and set $V = |U|$. We want to find a PDF of V, which we do by applying the CDF method. For $v \geq 0$,

$$F_V(v) = P(V \leq v) = P(|U| \leq v) = P(-v \leq U \leq v) = F_U(v) - F_U(-v).$$

Differentiation now yields

$$f_V(v) = f_U(v) + f_U(-v) = 2f_U(v) = 2 \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-v^2/2\sigma^2} = \frac{\sqrt{2/\pi}}{\sigma} e^{-v^2/2\sigma^2}$$

if $v > 0$, and $f_V(v) = 0$ otherwise. Setting $n = 1$ in the PDF obtained in part (a) and using the fact that $\Gamma(1/2) = \sqrt{\pi}$, we get, for $y > 0$,

$$f_Y(y) = \frac{1}{2^{1/2-1}\Gamma(1/2)\sigma^1} y^{1-1} e^{-y^2/2\sigma^2} = \frac{\sqrt{2/\pi}}{\sigma} e^{-y^2/2\sigma^2} = f_V(y).$$

Thus, for $n = 1$, $\sigma\sqrt{X}$ has the same distribution as the absolute value of a $\mathcal{N}\left(0, \sigma^2\right)$ distribution.

c) Setting $n = 2$ in the PDF obtained in part (a) and using the fact that $\Gamma(1) = 1$, we get, for $y > 0$,

$$f_Y(y) = \frac{1}{2^{2/2-1}\Gamma(2/2)\sigma^2} y^{2-1} e^{-y^2/2\sigma^2} = \frac{y}{\sigma^2} e^{-y^2/2\sigma^2},$$

which is the PDF of a Rayleigh random variable with parameter σ^2.

d) Setting $n = 3$ in the PDF obtained in part (a) and using the fact that $\Gamma(3/2) = \sqrt{\pi}/2$, we get, for $y > 0$,

$$f_Y(y) = \frac{1}{2^{3/2-1}\Gamma(3/2)\sigma^3} y^{3-1} e^{-y^2/2\sigma^2} = \frac{\sqrt{2/\pi}}{\sigma^3} y^2 e^{-y^2/2\sigma^2},$$

which is the PDF of a Maxwell distribution with parameter σ^2.

8.161 As X has the standard Cauchy distribution, its PDF is

$$f_X(x) = \frac{1}{\pi\left(1 + x^2\right)}, \qquad -\infty < x < \infty.$$

The range of X is the interval $(-\infty, \infty)$. Here $g(x) = 1/(1+x^2)$, which is not monotone on the range of X. Thus, we apply the CDF method instead of the transformation method. Letting $Y = 1/(1+X^2)$, we note that the range of Y is the interval $(0, 1)$. For $0 \le y < 1$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P\left(\frac{1}{1+X^2} \le y\right) = P\left(X^2 \ge 1/y - 1\right) = 1 - P\left(X^2 < 1/y - 1\right) \\ &= 1 - P\left(-\sqrt{1/y-1} < X < \sqrt{1/y-1}\right) = 1 - F_X\left(\sqrt{1/y-1}\right) + F_X\left(-\sqrt{1/y-1}\right). \end{aligned}$$

Differentiation now yields

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{1/y-1}}\frac{1}{y^2} f_X\left(\sqrt{1/y-1}\right) + \frac{1}{2\sqrt{1/y-1}}\frac{1}{y^2} f_X\left(-\sqrt{1/y-1}\right) \\ &= \frac{1}{y^2\sqrt{1/y-1}} f_X\left(\sqrt{1/y-1}\right) = \frac{1}{y^2\sqrt{1/y-1}} \frac{1}{\pi\left(1+\left(\sqrt{1/y-1}\right)^2\right)} \\ &= \frac{1}{y^2\sqrt{1/y-1}}\frac{y}{\pi} = \frac{1}{\pi y\sqrt{1/y-1}} = \frac{1}{\pi\sqrt{y}\sqrt{1-y}} = \frac{1}{B(1/2, 1/2)} y^{-1/2}(1-y)^{-1/2} \end{aligned}$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Thus, $1/(1+X^2)$ has the beta distribution with parameters 1/2 and 1/2 or, equivalently, the arcsine distribution.

8.162

a) Because $U \sim \mathcal{U}(0, 1)$ and $1 - u$ is a linear transformation that takes the interval $(0, 1)$ to itself, we would guess that $1 - U \sim \mathcal{U}(0, 1)$.

b) As $U \sim \mathcal{U}(0, 1)$, its PDF is $f_U(u) = 1$ if $0 < u < 1$, and $f_U(u) = 0$ otherwise. The range of U is the interval $(0, 1)$. Here $g(u) = 1 - u$, which is strictly decreasing and differentiable on the range of U. The transformation method is appropriate. Note that $g^{-1}(y) = 1 - y$. Letting $Y = 1 - U$, we see that the range of Y is the interval $(0, 1)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(u)|} f_U(u) = \frac{f_U(u)}{|-1|} = f_U(1-y) = 1$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. So, $1 - U \sim \mathcal{U}(0, 1)$.

8.163 From Exercise 8.65 on page 434, we know that $F_X(x) = (x-a)/(b-a)$ if $a \le x < b$. Using this fact, we also see that $F_X^{-1}(y) = a + (b-a)y$ if $0 < y < 1$.

a) From Proposition 8.16(a), we know that $F_X(X) \sim \mathcal{U}(0, 1)$. However, $F_X(X) = (X-a)/(b-a)$ so that $(X-a)/(b-a) \sim \mathcal{U}(0, 1)$.

b) From Proposition 8.16(b), we know that $F_X^{-1}(U)$ has the same distribution as X, which is $\mathcal{U}(a, b)$. However, $F_X^{-1}(U) = a + (b-a)U$ so that $a + (b-a)U \sim \mathcal{U}(a, b)$.

8.164

a) Use a basic random number generator to obtain 5000 observations of a uniform random variable on the interval $(0, 1)$ and then transform those numbers by using the function $a + (b-a)u$.

b) Answers will vary.

c) Answers will vary.

8.165 Let $Y = \lfloor m + (n - m + 1)X \rfloor$. As $X \sim \mathcal{U}(0, 1)$, its CDF is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

In particular, the range of X is the interval $(0, 1)$. If follows that the range of Y is $S = \{m, m+1, \dots, n\}$. For $y \in S$,

$$\begin{aligned} p_Y(y) &= P(Y = y) = P\big(\lfloor m + (n-m+1)X \rfloor = y\big) = P\big(y \le m + (n-m+1)X < y+1\big) \\ &= P\left(\frac{y-m}{n-m+1} \le X < \frac{y+1-m}{n-m+1}\right) = F_X\left(\frac{y+1-m}{n-m+1}\right) - F_X\left(\frac{y-m}{n-m+1}\right) \\ &= \left(\frac{y+1-m}{n-m+1}\right) - \left(\frac{y-m}{n-m+1}\right) = \frac{1}{n-m+1}. \end{aligned}$$

Thus, Y has the discrete uniform distribution on $\{m, m+1, \dots, n\}$.

8.166
a) Use a basic random number generator to obtain 10,000 observations of a uniform random variable on the interval $(0, 1)$ and then transform those numbers by using the function $\lfloor m + (n-m+1)u \rfloor$.
b) Answers will vary.
c) Answers will vary.

Theory Exercises

8.167 We have $Z = (X - \mu)/\sigma = (-\mu/\sigma) + (1/\sigma)X$. Applying Proposition 8.15 with $a = -\mu/\sigma$ and $b = 1/\sigma$, we conclude that Z has the normal distribution with parameters

$$a + b\mu = -\frac{\mu}{\sigma} + \frac{1}{\sigma}\mu = 0 \qquad \text{and} \qquad b^2\sigma^2 = \left(\frac{1}{\sigma}\right)^2 \sigma^2 = 1.$$

8.168 To prove the univariate transformation theorem in the case where g is strictly increasing on the range of X, we apply the CDF method. In the proof, we use the notation g^{-1} for the inverse function of g, defined on the range of Y. For y in the range of Y,

$$F_Y(y) = P(Y \le y) = P\big(g(X) \le y\big) = P\big(X \le g^{-1}(y)\big) = F_X\big(g^{-1}(y)\big),$$

where the third equality follows from the fact that g is increasing. Taking the derivative with respect to y and applying the chain rule, we obtain

$$f_Y(y) = F_Y'(y) = F_X'\big(g^{-1}(y)\big)\frac{d}{dy}g^{-1}(y) = \frac{1}{g'\big(g^{-1}(y)\big)} f_X\big(g^{-1}(y)\big),$$

where, in the last equality, we used the calculus result $(d/dy)g^{-1}(y) = 1/g'\big(g^{-1}(y)\big)$. Because g is increasing, g' is positive. Letting x denote the unique real number in the range of X such that $g(x) = y$, we conclude that, for y in the range of Y, $f_Y(y) = f_X(x)/|g'(x)|$, as required.

8.169 We apply the CDF method. For y in the range of Y

$$F_Y(y) = P(Y \le y) = P\big(g(X) \le y\big) = P\Big(X \in g^{-1}\big((-\infty, y]\big)\Big) = \int_{g^{-1}((-\infty, y])} f_X(x)\,dx,$$

where the last equality is a result of the FPF for continuous random variables. Differentiation now yields

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\int_{g^{-1}((-\infty, y])} f_X(x)\,dx.$$

This formula is the continuous analogue of the one given for discrete random variables in Equation (5.51) on page 247. The PMFs and sum in the latter are replaced by PDFs and an integral in the former.

8.170 Let R denote the range of X. Assume first that g is strictly increasing on R. Referring to Exercise 8.169 and applying the first fundamental theorem of calculus and the chain rule, we get

$$f_Y(y) = \frac{d}{dy}\int_{g^{-1}((-\infty,y])} f_X(x)\,dx = \frac{d}{dy}\int_{\{x\in R: g(x)\le y\}} f_X(x)\,dx = \frac{d}{dy}\int_{\{x\in R: x\le g^{-1}(y)\}} f_X(x)\,dx$$
$$= f_X\big(g^{-1}(y)\big)\frac{d}{dy}g^{-1}(y) = f_X\big(g^{-1}(y)\big)\frac{1}{g'\big(g^{-1}(y)\big)} = \frac{1}{\big|g'\big(g^{-1}(y)\big)\big|}f_X\big(g^{-1}(y)\big),$$

where the last equality follows from the fact that g' is positive because g is strictly increasing. Next assume that g is strictly decreasing on R. Referring to Exercise 8.169 and applying the first fundamental theorem of calculus and the chain rule, we get

$$f_Y(y) = \frac{d}{dy}\int_{g^{-1}((-\infty,y])} f_X(x)\,dx = \frac{d}{dy}\int_{\{x\in R: g(x)\le y\}} f_X(x)\,dx = \frac{d}{dy}\int_{\{x\in R: x\ge g^{-1}(y)\}} f_X(x)\,dx$$
$$= -f_X\big(g^{-1}(y)\big)\frac{d}{dy}g^{-1}(y) = -f_X\big(g^{-1}(y)\big)\frac{1}{g'\big(g^{-1}(y)\big)} = \frac{1}{\big|g'\big(g^{-1}(y)\big)\big|}f_X\big(g^{-1}(y)\big),$$

where the last equality follows from the fact that g' is negative because g is strictly decreasing.

Advanced Exercises

8.171 Answers will vary. However, the idea is as follows. Let X_j denote the roundoff error for the jth number—that is, the difference between the jth number and its rounded value. The sum of the rounded numbers equals the rounded sum of the unrounded numbers if and only if $|X_1 + \cdots + X_{10}| \le 0.5$. We want to use simulation to estimate the probability of that event. We can reasonably assume, as we do, that $X_1, \ldots, X_{10}$ are independent and identically distributed random variables with common distribution $\mathcal{U}(-0.5, 0.5)$. Thus, for the estimate, we generate a large number, say, 10,000, of 10-tuples of observations of a $\mathcal{U}(-0.5, 0.5)$ random variable by using a basic random number generator (as explained in Exercise 8.164), and then sum each 10-tuple. The proportion of those sums that are at most 0.5 in magnitude provides the required probability estimate. We followed this procedure and obtained a probability estimate of 0.418.

8.172 For $0 < y < 1$, the set $A_y = \{x : F_X(x) = y\}$ is a nonempty bounded closed interval. Indeed, A_y is nonempty and bounded because F_X is continuous and $F_X(-\infty) = 0$ and $F_X(\infty) = 1$; A_y is closed because it equals the inverse image under the continuous function F_X of the closed set $\{y\}$; A_y is an interval because F_X is nondecreasing—if $x_1, x_2 \in A_y$ with $x_1 < x_2$, then, for each $x_1 < x < x_2$, we have

$$y = F_X(x_1) \le F_X(x) \le F_X(x_2) = y,$$

and hence $F_X(x) = y$, that is, $x \in A_y$. Now let r_y denote the right endpoint of A_y. Then we have that $\{F_X(X) \le y\} = \{X \le r_y\}$. Hence

$$P\big(F_X(X) \le y\big) = P(X \le r_y) = F_X(r_y) = y,$$

which shows that $F_X(X) \sim \mathcal{U}(0, 1)$.

8.173 Yes. Suppose to the contrary that X isn't a continuous random variable. Then there is a real number x_0 such that $P(X = x_0) > 0$. Because $\{X = x_0\} \subset \{F_X(X) = F_X(x_0)\}$, the domination principle implies that $P\big(F_X(X) = F_X(x_0)\big) \ge P(X = x_0) > 0$. This inequality shows that $F_X(X)$ isn't a continuous random variable and, consequently, couldn't be uniform on the interval $(0, 1)$.

8.174 Suppose that $0 < y < 1$. As we showed in the solution to Exercise 8.172, the set $\{ x : F_X(x) = y \}$ is a nonempty bounded closed interval. Consequently, we can define $F_X^{-1}(y) = \min\{ x : F_X(x) = y \}$. Let $W = F_X^{-1}(U)$. Then, for $w \in \mathcal{R}$, we have $\{W \le w\} = \{U \le F_X(w)\}$. Hence,

$$F_W(w) = P(W \le w) = P\big(U \le F_X(w)\big) = F_X(w),$$

which shows that W has the same CDF as X and, thus, the same probability distribution.

Review Exercises for Chapter 8

Basic Exercises

8.175 Your colleague should have written

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1; \\ 1/3, & \text{if } 1 \le x < 2; \\ 2/3, & \text{if } 2 \le x < 3; \\ 1, & \text{if } x \ge 3. \end{cases}$$

8.176
a) For $0 \le w < 1$,

$$F_W(w) = P(W \le w) = 1 - P(W > w) = 1 - \frac{(1-w)^2}{1} = 2w - w^2.$$

Thus,

$$F_W(w) = \begin{cases} 0, & \text{if } w < 1; \\ 2w - w^2, & \text{if } 0 \le w < 1; \\ 1, & w \ge 1. \end{cases}$$

b) Referring to the result of part (a), we get

$$f_W(w) = F_W'(w) = \frac{d}{dw}\left(2w - w^2\right) = 2 - 2w = 2(1-w)$$

if $0 < w < 1$, and $f_W(w) = 0$ otherwise.
c) First note that $B(1, 2) = \Gamma(1)\Gamma(2)/\Gamma(3) = 0!\,1!/2! = 1/2$. For $0 < w < 1$, we can then write

$$f_W(w) = 2(1-w) = \frac{1}{B(1,2)} w^{1-1}(1-w)^{2-1}.$$

Thus, W has the beta distribution with parameters 1 and 2.

8.177
a) Noting that X is a continuous random variable, we have

$$\begin{aligned} F_Y(y) = P(Y \le y) &= P\big(Y \le y, X \in \{0.2, 0.9\}\big) + P\big(Y \le y, X \notin \{0.2, 0.9\}\big) \\ &= 0 + P\big(Y \le y, X \notin \{0.2, 0.9\}\big) = P(Y \le y, X \ne 0.2, X \ne 0.9). \end{aligned}$$

This result shows that the CDF of Y does not depend on the values assigned to Y when $X = 0.2$ or $X = 0.9$. In particular, then, we can assume without loss of generality that $Y = X$ when $X \in \{0.2, 0.9\}$. Doing so, we have $F_Y(y) = 0$ if $y < 0.2$, and $F_Y(y) = 1$ if $y \ge 0.9$. Also, for $0.2 \le y < 0.9$,

$$F_Y(y) = P(Y \le y) = P(X \le y) = y.$$

Thus,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0.2; \\ y, & \text{if } 0.2 \le y < 0.9; \\ 1, & \text{if } y \ge 0.9. \end{cases}$$

b) Referring to part (a), we note that

$$P(Y = y) = F_Y(y) - F_Y(y-) = \begin{cases} 0.2, & \text{if } y = 0.2; \\ 0.1, & \text{if } y = 0.9; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, Y is not a discrete random variable because $\sum_y P(Y = y) = 0.3 < 1$. It is not a continuous random variable because, for instance, $P(Y = 0.2) \neq 0$.
c) From part (b), we know that $S = \{0.2, 0.9\}$ and that

$$P(Y \in S) = P(Y = 0.2) + P(Y = 0.9) = 0.2 + 0.1 = 0.3.$$

For $y \in S$,

$$P(Y = y \mid Y \in S) = \frac{P(Y = y, Y \in S)}{P(Y \in S)} = \frac{P(Y = y)}{0.3} = \begin{cases} 2/3, & \text{if } y = 0.2; \\ 1/3, & \text{if } y = 0.9, \end{cases}$$

and $P(Y = y \mid Y \in S) = 0$ otherwise. As $\sum_y P(Y = y \mid Y \in S) = 1$, the random variable $Y_{|\{Y \in S\}}$ is discrete. Its CDF is

$$F_{Y_{|\{Y \in S\}}}(y) = \begin{cases} 0, & \text{if } y < 0.2; \\ 2/3, & \text{if } 0.2 \le y < 0.9; \\ 1, & \text{if } y \ge 0.9. \end{cases}$$

d) Note that

$$P(Y \notin S) = 1 - P(Y \in S) = 1 - 0.3 = 0.7$$

and that

$$P(Y \le y, Y \notin S) = P(Y \le y) - P(Y \le y, Y \in S).$$

Referring to parts (a) and (c), we have

$$\begin{aligned} F_{Y_{|\{Y \notin S\}}}(y) &= P(Y \le y \mid Y \notin S) = \frac{P(Y \le y, Y \notin S)}{P(Y \notin S)} = \frac{1}{0.7}\big(P(Y \le y) - P(Y \le y, Y \in S)\big) \\ &= \frac{1}{0.7}\left(P(Y \le y) - \frac{0.3}{P(Y \in S)} P(Y \le y, Y \in S)\right) = \frac{1}{0.7}\big(F_Y(y) - 0.3 F_{Y_{|\{Y \in S\}}}(y)\big), \end{aligned}$$

or

$$F_{Y_{|\{Y \notin S\}}}(y) = \begin{cases} 0, & \text{if } y < 0.2; \\ (y - 0.2)/0.7, & \text{if } 0.2 \le y < 0.9; \\ 1, & \text{if } y \ge 0.9. \end{cases}$$

We see that $Y_{|\{Y \notin S\}} \sim \mathcal{U}(0.2, 0.9)$ and, in particular, is a continuous random variable.
e) From part (d),

$$F_W(y) = F_{Y_{|\{Y \notin S\}}}(y) = \frac{1}{0.7}\big(F_Y(y) - 0.3 F_{Y_{|\{Y \in S\}}}(y)\big) = \frac{1}{0.7}\big(F_Y(y) - 0.3 F_V(y)\big),$$

which means that $F_Y = 0.3 F_V + 0.7 F_W$. Thus, choose $a = 0.3$ and $b = 0.7$.
f) As X is a continuous random variable, the four required probabilities are identical and equal

$$F_X(0.9) - F_X(0.2) = 0.9 - 0.2 = 0.7.$$

Referring to part (a) and Proposition 8.2 on page 412, we find that the required probabilities are

$$\begin{aligned}
F_Y(0.9-) - F_Y(0.2-) &= 0.9 - 0 = 0.9, \\
F_Y(0.9-) - F_Y(0.2) &= 0.9 - 0.2 = 0.7, \\
F_Y(0.9) - F_Y(0.2) &= 1 - 0.2 = 0.8, \\
F_Y(0.9) - F_Y(0.2-) &= 1 - 0 = 1.
\end{aligned}$$

Referring to part (c) and Proposition 8.2, we find that the required probabilities are

$$\begin{aligned}
F_V(0.9-) - F_V(0.2-) &= 2/3 - 0 = 2/3, \\
F_V(0.9-) - F_V(0.2) &= 2/3 - 2/3 = 0, \\
F_V(0.9) - F_V(0.2) &= 1 - 2/3 = 1/3, \\
F_V(0.9) - F_V(0.2-) &= 1 - 0 = 1.
\end{aligned}$$

As W is a continuous random variable, the four required probabilities are identical and equal

$$F_W(0.9) - F_W(0.2) = 1 - 0 = 1.$$

8.178 Yes. For a CDF, F_X, we have $F_X(x) = P(X \le x) = 1 - P(X > x)$ for all $x \in \mathcal{R}$. Thus, the CDF can be obtained from the tail probabilities. And because the CDF determines the probability distribution of X (fourth bulleted item on page 407), so do the tail probabilities.

8.179 Let Y denote weekly gas sales, in gallons. Then $Y = \min\{X, m\}$.

a) For $0 \le y < m$,

$$P(Y > y) = P\big(\min\{X, m\} > y\big) = P(X > y) = e^{-\lambda y}.$$

Therefore,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ 1 - e^{-\lambda y}, & \text{if } 0 \le y < m; \\ 1, & \text{otherwise.} \end{cases}$$

b) From part (a), we see that

$$P(Y = m) = F_Y(m) - F_Y(m-) = 1 - \big(1 - e^{-\lambda m}\big) = e^{-\lambda m},$$

and $P(Y = y) = 0$ if $y \ne m$. Because $\sum_y P(Y = y) = e^{-\lambda m} < 1$, Y is not a discrete random variable and, because $P(Y = m) \ne 0$, Y is not a continuous random variable.

c) The gas station runs out of gas before the end of the week if and only if $\{X > m\}$, which, because $X \sim \mathcal{E}(\lambda)$, has probability $e^{-\lambda m}$.

8.180

a) f_X is discontinuous at $x = 0$, 1/2, and 1.

b) Because X is a continuous random variable, its CDF, F_X, is an everywhere continuous function, that is, F_X has no points of discontinuity.

c) We have

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{1/2} c\,dx + \int_{1/2}^{1} 2c\,dx = \frac{1}{2}c + \left(1 - \frac{1}{2}\right)(2c) = \frac{3}{2}c.$$

Hence, $c = 2/3$.

d) For $0 \le x < 1/2$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x (2/3)\,dt = 2x/3.$$

For $1/2 \le x < 1$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^{1/2} (2/3)\,dt + \int_{1/2}^{x} (4/3)\,dt = \frac{1}{3} + \left(x - \frac{1}{2}\right)\left(\frac{4}{3}\right) = (4x-1)/3.$$

Thus,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 2x/3, & \text{if } 0 \le x < 1/2; \\ (4x-1)/3, & \text{if } 1/2 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

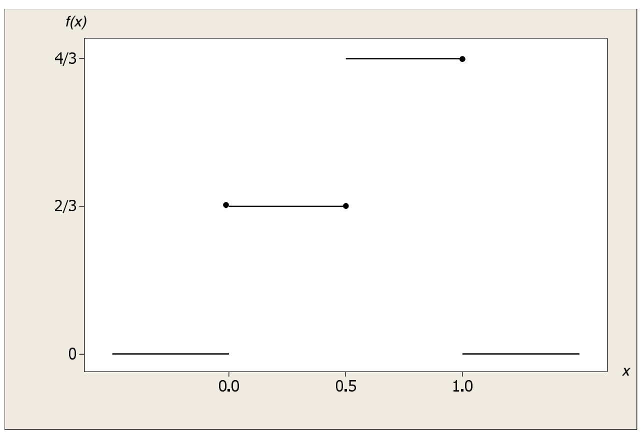
From this result, it follows easily that F_X is everywhere continuous. Note, in particular, that

$$F_X(0-) = 0 = (2 \cdot 0)/3 = F_X(0),$$
$$F_X(1/2-) = 2 \cdot (1/2)/3 = 1/3 = \big(4(1/2) - 1\big)/3 = F_X(1/2),$$

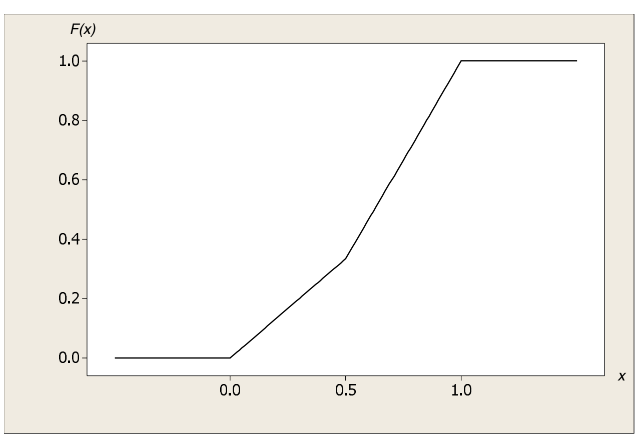
and

$$F_X(1-) = (4 \cdot 1 - 1)/3 = 1 = F_X(1).$$

e) A graph of f_X is as follows:



A graph of F_X is as follows:



8.181 A PDF must integrate to 1. Thus,

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^5 cx(5-x)\,dx = c\int_0^5 \left(5x - x^2\right)dx = 5^3c/6.$$

Thus, $c = 6/5^3 = 0.048$.

8.182 If there were a constant c such that cg is a PDF, then c would have to be positive. Then,

$$1 = \int_{-\infty}^{\infty} cg(x)\,dx = c\int_0^1 \frac{1}{x}\,dx = c \cdot \infty = \infty,$$

which is a contradiction.

8.183

a) We can write $f_X(x) = cx^{3-1}(1-x)^{2-1}$ if $0 < x < 1$, and $f_X(x) = 0$ otherwise. Because of the functional form of the PDF, we see that X must have the beta distribution with parameters 3 and 2.

b) We have

$$c = \frac{1}{B(3,2)} = \frac{\Gamma(3+2)}{\Gamma(3)\Gamma(2)} = \frac{\Gamma(5)}{\Gamma(3)\Gamma(2)} = \frac{4!}{2!\,1!} = 12.$$

c) We have

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 cx^2(1-x)\,dx = c\int_0^1 (x^2 - x^3)\,dx = c \cdot \left(\frac{1}{12}\right).$$

Thus, $c = 12$.

d) From Equation (8.55) on page 457, the CDF of X is

$$F_X(x) = \sum_{j=3}^{4} \binom{4}{j} x^j (1-x)^{4-j} = 4x^3(1-x) + x^4 = 4x^3 - 3x^4 = x^3(4-3x).$$

Therefore, $P(X \le 1/2) = (1/2)^3\bigl(4 - 3\cdot(1/2)\bigr) = 0.3125$.

e) Applying the conditional probability rule and using the results of part (d) yields

$$\begin{aligned} P(X \le 1/4 \mid X \le 1/2) &= \frac{P(X \le 1/4,\, X \le 1/2)}{P(X \le 1/2)} = \frac{P(X \le 1/4)}{P(X \le 1/2)} \\ &= \frac{(1/4)^3\bigl(4 - 3\cdot(1/4)\bigr)}{0.3125} = 0.1625. \end{aligned}$$

8.184

a) It would make sense for $x > 0$.

b) Because X^2 cannot be negative, we would necessarily have $f_{X^2}(x) = 0$ for all $x < 0$. Thus, the required constant is 0.

c) If f_X is an even function, then, making the substitution $u = -t$, we get

$$\begin{aligned} F_X(-x) &= \int_{-\infty}^{-x} f_X(t)\,dt = \int_x^{\infty} f_X(-u)\,du = \int_x^{\infty} f_X(u)\,du \\ &= P(X > x) = 1 - P(X \le x) = 1 - F_X(x). \end{aligned}$$

Hence, the quantity in brackets on the right of the display in part (a) can be simplified as follows:

$$F_X\left(\sqrt{x}\right) - F_X\left(-\sqrt{x}\right) = F_X\left(\sqrt{x}\right) - \Bigl(1 - F_X\left(\sqrt{x}\right)\Bigr) = 2F_X\left(\sqrt{x}\right) - 1.$$

8.185

a) We can write

$$f_Z(z) = 2z = \frac{2!}{1!\,0!}z = \frac{\Gamma(3)}{\Gamma(2)\Gamma(1)}z = \frac{1}{B(2,1)}z^{2-1}(1-z)^{1-1}.$$

Thus, Z has the beta distribution with parameters 2 and 1.

b) For $0 \le \theta < 2\pi$, the event $\{\Theta \le \theta\}$ occurs if and only if the center of the first spot lies in the angular sector $[0, \theta)$, which has area $\theta/2$. Hence, by Equation (8.5) on page 403,

$$F_\Theta(\theta) = P(\Theta \le \theta) = \frac{|\{\Theta \le \theta\}|}{\pi} = \frac{\theta/2}{\pi} = \frac{\theta}{2\pi}.$$

Thus,

$$F_\Theta(\theta) = \begin{cases} 0, & \text{if } \theta < 0; \\ \dfrac{\theta}{2\pi}, & \text{if } 0 \le \theta < 2\pi; \\ 1, & \text{if } \theta \ge 2\pi. \end{cases}$$

c) From the result of part (b), we see that

$$f_\Theta(\theta) = F'_\Theta(\theta) = \frac{d}{d\theta}\left(\frac{\theta}{2\pi}\right) = \frac{1}{2\pi}$$

if $0 < \theta < 2\pi$, and $f_\Theta(\theta) = 0$ otherwise. Thus $\Theta \sim \mathcal{U}(0, 2\pi)$.

8.186

a) As $N(t) \sim \mathcal{P}(3.8t)$, its mean is $3.8t$. Thus, on average, $3.8t$ calls arrive at the switchboard during the first t minutes.

b) Event $\{W_1 > t\}$ occurs if and only if the first phone call arrives after time t, which means that $N(t) = 0$. Similarly, event $\{W_1 \le t\}$ occurs if and only if the first phone call arrives on or before time t, which means that $N(t) \ge 1$.

c) Referring to part (b), we have

$$P(W_1 > t) = P\big(N(t) = 0\big) = e^{-3.8t}\frac{(3.8)^0}{0!} = e^{-3.8t}.$$

d) From Propostion 8.8 on page 433 and the result of part (c), we conclude that $W_1 \sim \mathcal{E}(3.8)$.

e) For $t > 0$,

$$\begin{aligned} P(W_3 \le t) &= P\big(N(t) \ge 3\big) = 1 - P\big(N(t) \le 2\big) \\ &= 1 - \sum_{j=0}^{2} e^{-3.8t}\frac{(3.8t)^j}{j!} = 1 - e^{-3.8t}\sum_{j=0}^{3-1}\frac{(3.8t)^j}{j!}. \end{aligned}$$

Referring now to the note directly preceding Example 8.16 and to Equation (8.49), we can conclude that $W_3 \sim \Gamma(3, 3.8)$.

f) Proceeding as in part (e), we have, for $t > 0$,

$$\begin{aligned} P(W_n \le t) &= P\big(N(t) \ge n\big) = 1 - P\big(N(t) \le n-1\big) \\ &= 1 - \sum_{j=0}^{n-1} e^{-3.8t}\frac{(3.8t)^j}{j!} = 1 - e^{-3.8t}\sum_{j=0}^{n-1}\frac{(3.8t)^j}{j!}. \end{aligned}$$

Thus, $W_n \sim \Gamma(n, 3.8)$.

8.187 Let X denote the computed area of the circle, so that $X = \pi D^2/4$.
a) For $\pi(d-\epsilon)^2/4 \le x < \pi(d+\epsilon)^2/4$,

$$F_X(x) = P(X \le x) = P\left(\pi D^2/4 \le x\right) = P\left(D \le 2\sqrt{x/\pi}\right) = F_D\left(2\sqrt{x/\pi}\right)$$

Differentiation now yields

$$f_X(x) = F_X'(x) = \frac{1}{\sqrt{\pi x}} f_D\left(2\sqrt{x/\pi}\right) = \frac{1}{\sqrt{\pi x}} \cdot \frac{1}{2\epsilon} = \frac{1}{2\epsilon\sqrt{\pi x}}$$

if $\pi(d-\epsilon)^2/4 \le x < \pi(d+\epsilon)^2/4$, and $f_X(x) = 0$ otherwise.
b) Note that, because D can't be negative, the normal distribution given for D must be an approximation. Thus, our calculation here is purely formal. For $x > 0$,

$$\begin{aligned} F_X(x) &= P(X \le x) = P\left(\pi D^2/4 \le x\right) \\ &= P\left(-2\sqrt{x/\pi} \le D \le 2\sqrt{x/\pi}\right) = F_D\left(2\sqrt{x/\pi}\right) - F_D\left(-2\sqrt{x/\pi}\right). \end{aligned}$$

Differentiation now yields

$$\begin{aligned} f_X(x) = F_X'(x) &= \frac{1}{\sqrt{\pi x}} f_D\left(2\sqrt{x/\pi}\right) + \frac{1}{\sqrt{\pi x}} f_D\left(-2\sqrt{x/\pi}\right) \\ &= \frac{1}{\sqrt{\pi x}}\left(\frac{1}{\sqrt{2\pi}\,\epsilon} e^{-\left(2\sqrt{x/\pi}-d\right)^2/2\epsilon^2} + \frac{1}{\sqrt{2\pi}\,\epsilon} e^{-\left(-2\sqrt{x/\pi}-d\right)^2/2\epsilon^2}\right) \\ &= \frac{1}{\pi\epsilon\sqrt{2x}}\left(e^{-\left(2\sqrt{x}-\sqrt{\pi}d\right)^2/2\pi\epsilon^2} + e^{-\left(2\sqrt{x}+\sqrt{\pi}d\right)^2/2\pi\epsilon^2}\right) \end{aligned}$$

if $x > 0$ and $f_X(x) = 0$ otherwise.

8.188
a) We have $P\big(N(t) = 0\big) = P(X > t) = e^{-6.9t}$.
b) We have

$$P\big(N(2.5) = 0 \mid N(2) = 0\big) = P(X > 2.5 \mid X > 2) = P(X > 0.5) = e^{-6.9\cdot 0.5} = 0.0317,$$

where, in the second equality, we used the lack-of-memory property of an exponential random variable.
c) Referring to the solution to part (b) yields

$$P\big(N(2.5) - N(2) = 0 \mid N(2) = 0\big) = P\big(N(2.5) = 0 \mid N(2) = 0\big) = 0.0317.$$

8.189
a) For $x > 0$,

$$F_X(x) = P(X \le x) = P\big(-\log_b U \le x\big) = P\left(U \ge b^{-x}\right) = 1 - F_U\left(b^{-x}\right).$$

Differentiation now yields,

$$f_X(x) = F_X'(x) = (\ln b)b^{-x} f_U\left(b^{-x}\right) = (\ln b)b^{-x} = (\ln b)e^{-(\ln b)x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Thus, $Y \sim \mathcal{E}(\ln b)$.
b) From algebra, $\log_b y = (\ln y)/(\ln b)$. Hence, $X = -\log_b U = -(\ln b)^{-1} \ln U$ which, by Example 8.22(b), is exponentially distributed with parameter $\ln b$.

8.190 Note that, given $X = 0$, we must have $Y = 0$. Thus, $P(4 < Y < 8 \mid X = 0) = 0$. Furthermore,

$$P(4 < Y < 8 \mid X = 1) = e^{-4/5} - e^{-8/5} \qquad \text{and} \qquad P(4 < Y < 8 \mid X > 1) = e^{-4/8} - e^{-8/8}.$$

Applying the law of total probability yields

$$\begin{aligned} P(4 < Y < 8) &= P(X = 0)P(4 < Y < 8 \mid X = 0) + P(X = 1)P(4 < Y < 8 \mid X = 1) \\ &\quad + P(X > 1)P(4 < Y < 8 \mid X > 1) \\ &= \frac{1}{2} \cdot 0 + \frac{1}{3}\left(e^{-4/5} - e^{-8/5}\right) + \frac{1}{6}\left(e^{-4/8} - e^{-8/8}\right) = 0.122. \end{aligned}$$

8.191
a) We can show that F^n is a CDF by proving that it satisfies the conditions of Proposition 8.1 on page 411. Alternatively, we can proceed as follows. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having common PDF f and set $U = \max\{X_1, \ldots, X_n\}$. Then

$$\begin{aligned} F_U(u) &= P(U \le u) = P\big(\max\{X_1, \ldots, X_n\} \le u\big) = P(X_1 \le u, \ldots, X_n \le u) \\ &= P(X_1 \le u) \cdots P(X_n \le u) = \big(F(u)\big)^n = F^n(u). \end{aligned}$$

Thus, F^n is the CDF of U and, in particular, is a CDF. The corresponding PDF is

$$f_U(u) = F_U'(u) = n\big(F(u)\big)^{n-1} F'(u) = nf(u)\big(F(u)\big)^{n-1}.$$

Thus, the PDF corresponding to F^n is nfF^{n-1}.
b) We can show that $1 - (1 - F)^n$ is a CDF by proving that it satisfies the conditions of Proposition 8.1 on page 411. Alternatively, we can proceed as follows. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having common PDF f and set $V = \min\{X_1, \ldots, X_n\}$. Then

$$\begin{aligned} F_V(v) &= P(V \le v) = 1 - P(V > v) = 1 - P\big(\min\{X_1, \ldots, X_n\} > v\big) \\ &= 1 - P(X_1 > v, \ldots, X_n > v) = 1 - P(X_1 > v) \cdots P(X_n > v) \\ &= 1 - \big(1 - P(X_1 \le v)\big) \cdots \big(1 - P(X_n \le v)\big) = 1 - \big(1 - F(v)\big)^n. \end{aligned}$$

Thus, $1 - (1 - F)^n$ is the CDF of V and, in particular, is a CDF. The corresponding PDF is

$$f_V(v) = F_V'(v) = -n\big(1 - F(v)\big)^{n-1}\big(-F'(v)\big) = nf(v)\big(1 - F(v)\big)^{n-1}.$$

Thus, the PDF corresponding to $1 - (1 - F)^n$ is $nf(1 - F)^{n-1}$.

8.192 Note that in both parts, $X(t)$ is a continuous random variable with a symmetric distribution. These facts imply that $F_{X(t)}(-x) = 1 - F_{X(t)}(x)$ for all $x \in \mathcal{R}$. Therefore, for $x > 0$,

$$F_{|X(t)|}(x) = P\big(|X(t)| \le x\big) = P\big(-x \le X(t) \le x\big) = F_{X(t)}(x) - F_{X(t)}(-x) = 2F_{X(t)}(x) - 1.$$

Differentiation yields $f_{|X(t)|}(x) = 2f_{X(t)}(x)$ if $x > 0$, and $f_{|X(t)|}(x) = 0$ otherwise.
a) In this case,

$$f_{|X(t)|}(x) = 2 \cdot \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{t}}\, e^{-x^2/2\sigma^2 t} = \frac{\sqrt{2/\pi t}}{\sigma}\, e^{-x^2/2\sigma^2 t}$$

if $x > 0$, and $f_{|X(t)|}(x) = 0$ otherwise.
b) In this case,

$$f_{|X(t)|}(x) = 2 \cdot \frac{1}{2t} = \frac{1}{t}$$

if $0 < x < t$, and $f_{|X(t)|}(x) = 0$ otherwise. Thus, $|X(t)| \sim \mathcal{U}(0, t)$.

8.193 We have $F_{X^+(t)}(x) = 0$ if $x < 0$ and, for $x \geq 0$,

$$F_{X^+(t)}(x) = P\big(\max\{X(t), 0\} \leq x\big) = P\big(X(t) \leq x\big) = F_{X(t)}(x).$$

a) Referring to Equation (8.35) on page 443, we see that, in this case, $F_{X^+(t)}(x) = 0$ if $x < 0$ and, for $x \geq 0$,

$$F_{X^+(t)}(x) = F_{X(t)}(x) = \Phi\left(\frac{x}{\sigma\sqrt{t}}\right).$$

b) In this case, $F_{X^+(t)}(x) = 0$ if $x < 0$,

$$F_{X^+(t)}(x) = F_{X(t)}(x) = \frac{x+t}{2t}$$

if $0 \leq x \leq t$, and $F_{X^+(t)}(x) = 1$ if $x \geq t$.
c) We have $P\big(X^+(t) = 0\big) = F_{X^+(t)}(0) - F_{X^+(t)}(0-)$. In part (a), this quantity equals $\Phi(0) - 0 = 1/2$, and, in part (b), it equals $1/2 - 0 = 1/2$. Thus, in either case, $X^+(t)$ is not a continuous random variable.

8.194
a) The range of V is the interval $(0, \infty)$. Here $g(v) = mv^2/2$, which is strictly increasing and differentiable on the range of V. The transformation method is appropriate. Note that $g^{-1}(k) = \sqrt{2k/m}$. We see that the range of K is the interval $(0, \infty)$. Applying the transformation method now yields

$$\begin{aligned} f_K(k) &= \frac{1}{|g'(v)|} f_V(v) = \frac{f_V(v)}{|mv|} = \frac{f_V\big(\sqrt{2k/m}\big)}{m\sqrt{2k/m}} \\ &= \frac{1}{m\sqrt{2k/m}}\, a\left(\sqrt{2k/m}\right)^2 e^{-bm\left(\sqrt{2k/m}\right)^2} = \frac{a\sqrt{2k}}{m^{3/2}}\, e^{-2bk} \end{aligned}$$

if $k > 0$, and $f_K(k) = 0$ otherwise.
b) From the result of part (a), for $k > 0$, the PDF of K has the form $c\sqrt{k}\, e^{-2bk} = ck^{3/2-1}e^{-(2b)k}$, where c is a constant. It follows that $K \sim \Gamma(3/2, 2b)$.
c) From parts (a) and (b),

$$\frac{a\sqrt{2}}{m^{3/2}} = c = \frac{(2b)^{3/2}}{\Gamma(3/2)} = \frac{(2b)^{3/2}}{\sqrt{\pi}/2}.$$

Hence, $a = (4bm)^{3/2}/\sqrt{\pi}$.

8.195
a) Because a normal random variable can be negative, whereas a gestation period can't.
b) The probability that a normal random variable with parameters $\mu = 266$ and $\sigma^2 = 256$ will be negative is essentially zero (2.30×10^{-62}).

8.196 Let X denote tee shot distance, in yards, for PGA players (circa 1999). Then $X \sim \mathcal{N}(272.2, 8.12^2)$. We apply Proposition 8.11 on page 443 and use Table I.
a) We have

$$\begin{aligned} P(260 \leq X \leq 280) &= \Phi\left(\frac{280 - 272.2}{8.12}\right) - \Phi\left(\frac{260 - 272.2}{8.12}\right) = \Phi(0.96) - \Phi(-1.50) \\ &= \Phi(0.96) - \big(1 - \Phi(1.50)\big) = 0.7647. \end{aligned}$$

Thus, 76.5% of PGA player tee shots go between 260 and 280 yards.

b) We have

$$P(X > 300) = 1 - P(X \le 300) = 1 - \Phi\left(\frac{300 - 272.2}{8.12}\right) = 1 - \Phi(3.42) = 1 - 0.9997 = 0.0003.$$

Thus, 0.03% of PGA player tee shots go more than 300 yards.

8.197 Let X denote height, in inches, of men in their 20s. Then $X \sim \mathcal{N}(69, 2.5^2)$. We apply Proposition 8.11 on page 443 and use Table I.
a) We have

$$\begin{aligned} P(68 \le X \le 73) &= \Phi\left(\frac{73-69}{2.5}\right) - \Phi\left(\frac{68-69}{2.5}\right) = \Phi(1.6) - \Phi(-0.4) \\ &= \Phi(1.6) - \big(1 - \Phi(0.4)\big) = 0.6006. \end{aligned}$$

b) We need to choose c so that

$$\begin{aligned} 0.80 = P(69 - c \le X \le 69 + c) &= \Phi\left(\frac{(69+c)-69}{2.5}\right) - \Phi\left(\frac{(69-c)-69}{2.5}\right) \\ &= \Phi(c/2.5) - \Phi(-c/2.5) = 2\Phi(c/2.5) - 1. \end{aligned}$$

Hence, $c = 2.5\Phi^{-1}(0.9) = 2.5 \cdot 1.28 = 3.2$. We interpret this by stating that 80% of men in their 20s have heights within 3.2 inches of the mean height of 69 inches.

8.198 Let X denote the number of the 100 adults selected who are in favor of strict gun control laws. Then $X \approx \mathcal{B}(100, 0.47)$, the approximate symbol reflecting the fact that the exact distribution is $\mathcal{H}(N, 100, 0.47)$, where N is the size of the adult population of the southwestern city in question. Note that $np = 100 \cdot 0.47 = 47$ and $np(1-p) = 100 \cdot 0.47 \cdot 0.53 = 24.91$. Thus, $X \approx \mathcal{N}(47, 24.91)$.
a) We have

$$P(X = 47) = p_X(47) \approx \frac{1}{\sqrt{2\pi \cdot 24.91}}\, e^{-(47-47)^2/(2\cdot 24.91)} = 0.0799.$$

b) We have

$$P(X = 46, 47, \text{ or } 48) = \sum_{x=46}^{48} p_X(x) \approx \frac{1}{\sqrt{2\pi \cdot 24.91}} \sum_{x=46}^{48} e^{-(x-47)^2/(2\cdot 24.91)} = 0.237.$$

8.199 Let X denote the number of the 300 Mexican households selected that are living in poverty. Then $X \approx \mathcal{B}(300, 0.26)$, the approximate symbol reflecting the fact that the exact distribution is actually $\mathcal{H}(N, 300, 0.26)$, where N is the number of Mexican households. Note that $np = 300 \cdot 0.26 = 78$ and $np(1-p) = 300 \cdot 0.26 \cdot 0.74 = 57.72$. Consequently, $X \approx \mathcal{N}(78, 57.72)$.
a) Here we want the probability that 26% of the Mexican households sampled are living in poverty. As 26% of 300 is 78, the required probability is

$$P(X = 78) = p_X(78) \approx \frac{1}{\sqrt{2\pi \cdot 57.72}}\, e^{-(78-78)^2/(2\cdot 57.72)} = 0.0525.$$

b) Here we want the probability that between 25% and 27%, inclusive, of the Mexican households sampled are living in poverty. As 25% of 300 is 75 and 27% of 300 is 81, the required probability is

$$P(75 \le X \le 81) = \sum_{x=75}^{81} p_X(x) \approx \frac{1}{\sqrt{2\pi \cdot 57.72}} \sum_{x=75}^{81} e^{-(x-78)^2/(2\cdot 57.72)} = 0.355.$$

8.200 We have $X \sim \mathcal{U}(0, 1)$ and $Y = \lfloor 100X \rfloor$. Note that the range of Y is the set $S = \{00, 01, \ldots, 99\}$. For $y \in S$,

$$\begin{aligned} p_Y(y) &= P(Y = y) = P(\lfloor 100X \rfloor = y) = P(y \le 100X < y + 1) \\ &= P\left(\frac{y}{100} \le X < \frac{y+1}{100}\right) = \frac{y}{100} - \frac{y-1}{100} = \frac{1}{100}. \end{aligned}$$

Thus, Y has the discrete uniform distribution on $\{00, 01, \ldots, 99\}$.

8.201 We have

$$\begin{aligned} \int_0^\infty x f_X(x)\,dx &= \int_0^\infty x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1}\, e^{-\lambda x}\,dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \cdot 1 = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\alpha\Gamma(\alpha)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda}. \end{aligned}$$

Also,

$$\begin{aligned} \int_0^\infty x^2 f_X(x)\,dx &= \int_0^\infty x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} \int_0^\infty \frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)} x^{(\alpha+2)-1}\, e^{-\lambda x}\,dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} \cdot 1 = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\lambda^{\alpha+2}} = \frac{\alpha(\alpha+1)}{\lambda^2}. \end{aligned}$$

8.202

a) We can write

$$f_X(x) = xe^{-x} = \frac{1}{(2-1)!} xe^{-x} = \frac{1^2}{\Gamma(2)} x^{2-1} e^{-1x}$$

if $x > 0$ and $f_X(x) = 0$ otherwise. Hence, $X \sim \Gamma(2, 1)$.

b) From part (a) and Equation (8.49) on page 453, the CDF of X is given by $F_X(x) = 0$ if $x < 0$, and

$$F_X(x) = 1 - e^{-x} \sum_{j=0}^{2-1} \frac{x^j}{j!} = 1 - e^{-x}(1 + x), \qquad x \ge 0.$$

c) Referring to the result of part (b), we find that the proportion of claims in which the claim amount exceeds 1000 is

$$P(X > 1) = 1 - F_X(1) = e^{-1}(1 + 1) = 2e^{-1}.$$

Hence the company should expect to pay $2e^{-1} \times 100 \approx 73.6$ claims.

Theory Exercises

8.203

a) Suppose that X is a symmetric random variable, that is, X and $-X$ have the same probability distribution. Then, for all $x \in \mathcal{R}$,

$$P(X \le -x) = P(-X \le -x) = P(X \ge x).$$

Conversely, suppose that $P(X \le -x) = P(X \ge x)$ for all $x \in \mathcal{R}$. Then

$$F_{-X}(x) = P(-X \le x) = P(X \ge -x) = P(X \le x) = F_X(x)$$

for all $x \in \mathcal{R}$. Hence, X and $-X$ have the same probability distribution, that is, X is a symmetric random variable.

b) From part (a), the complementation rule, and Equation (8.7) on page 412, we see that the following conditions are equivalent:

- X is a symmetric random variable.
- $P(X \le -x) = P(X \ge x)$ for all $x \in \mathcal{R}$.
- $P(X \le -x) = 1 - P(X < x)$ for all $x \in \mathcal{R}$.
- $F_X(-x) = 1 - F_X(x-)$ for all $x \in \mathcal{R}$.

The last bulleted item provides the required equivalent condition.
c) Referring to part (b) and Proposition 8.2(b) on page 412 yields

$$P(|X| \le x) = P(-x \le X \le x) = F_X(x) - F_X(-x-) = F_X(x) - \big(1 - F_X(x)\big) = 2F_X(x) - 1.$$

d) From part (a), $P(X \le 0) = P(X \ge 0)$. Thus,

$$1 = P(X < 0) + P(X \ge 0) \le P(X \le 0) + P(X \ge 0) = 2P(X \ge 0).$$

Therefore $P(X \ge 0) \ge 1/2$. Furthermore, $P(X \le 0) = P(X \ge 0) \ge 1/2$. Hence, 0 is a median of X.

8.204 Let $x \in \mathcal{R}$. Referring to Exercise 8.203(a) and making the substitution $u = -t$ yields

$$F_X(x) = P(X \le x) = P(X \ge -x) = \int_{-x}^{\infty} f_X(t)\,dt = \int_{-\infty}^{x} f_X(-u)\,du.$$

Applying Proposition 8.5 on page 422 now shows that the function g on $\mathcal{R}$ defined by $g(x) = f_X(-x)$ is a PDF of X. In other words, we can write $f_X(-x) = f_X(x)$ for all $x \in \mathcal{R}$, meaning that we can choose a PDF for X that is a symmetric function.

8.205
a) Referring to Exercise 8.203(a), we see that, for each $x \in \mathcal{R}$,

$$\begin{aligned} F_X(c-x) &= P(X \le c-x) = P(X - c \le -x) = P(X - c \ge x) = 1 - P(X - c < x) \\ &= 1 - P(X - c \le x) = 1 - P(X \le c + x) = 1 - F_X(c+x). \end{aligned}$$

Thus, a continuous random variable, X, is symmetric about c if and only if

$$F_X(c-x) = 1 - F_X(c+x), \qquad x \in \mathcal{R}.$$

b) We have

$$\frac{d}{dx}\big(F_X(c-x)\big) = -f_X(c-x) \qquad \text{and} \qquad \frac{d}{dx}\big(1 - F_X(c+x)\big) = -f_X(c+x).$$

Referring now to the equivalent condition in part (a), we see that a continuous random variable, X, with a PDF is symmetric about c if and only if we can write $f_X(c-x) = f_X(c+x)$ for all $x \in \mathcal{R}$, meaning that we can choose a PDF for X that is a symmetric function about c.

8.206 In answering each part, draw a graph of the PDF of the random variable in question and refer to Exercise 8.205(b).
a) X is symmetric about $(a+b)/2$.
b) X is not symmetric about any number c.
c) X is symmetric about μ.
d) X is not symmetric about any number c.
e) From the last bulleted item on page 457, we see that X is symmetric about some number c if and only if $\alpha = \beta$, in which case $c = 1/2$.

f) X is symmetric about $(a+b)/2$.
g) X is symmetric about η.

8.207
a) By defintion, $X-c$ is a symmetric random variable. Therefore, by Exercise 8.203(d), we know that 0 is a median of $X-c$. Consequently,

$$P(X \le c) = P(X - c \le 0) \ge \frac{1}{2} \qquad \text{and} \qquad P(X \ge c) = P(X - c \ge 0) \ge \frac{1}{2}.$$

Hence, c is a median of X.
b) Referring to part (a) and the results of Exercise 8.206 shows that the medians of the four specified distributions are $(a+b)/2$, μ, $(a+b)/2$, and η, respectively.

Advanced Exercises

8.208 In each case, we must show that the specified function satisfies the four properties of Proposition 8.1 on page 411.
a) Let $x \le y$. We make the substitution $u = t + (y-x)$ and use the fact that F is nondecreasing to conclude that

$$G(x) = \frac{1}{h}\int_x^{x+h} F(t)\,dt = \frac{1}{h}\int_y^{y+h} F\big(u-(y-x)\big)\,du \le \frac{1}{h}\int_y^{y+h} F(u)\,du = G(y).$$

Hence, G is nondecreasing.

We can write

$$G(x) = \frac{1}{h}\int_x^{x+h} F(t)\,dt = \frac{1}{h}\left(\int_0^{x+h} F(t)\,dt - \int_0^{x} F(t)\,dt\right).$$

From calculus, the two functions of x within the parentheses are everywhere continuous and, hence, so is their difference. Thus, G is everywhere continuous and, in particular, everywhere right continuous.

Let $\epsilon > 0$ be given. Because $F(-\infty) = 0$, we can choose M so that $F(t) < \epsilon$ whenever $t < M$. Then, for $x < M - h$,

$$G(x) = \frac{1}{h}\int_x^{x+h} F(t)\,dt \le \frac{1}{h}\int_x^{x+h} \epsilon\,dt = \epsilon.$$

Thus, $G(-\infty) = 0$.

Let $\epsilon > 0$ be given. Because $F(\infty) = 1$, we can choose M so that $F(t) > 1 - \epsilon$ whenever $t > M$. Then, for $x > M$,

$$G(x) = \frac{1}{h}\int_x^{x+h} F(t)\,dt \ge \frac{1}{h}\int_x^{x+h} (1-\epsilon)\,dt = 1 - \epsilon.$$

Thus, $G(\infty) = 1$.
b) Let $x \le y$. We make the substitution $u = t + (y-x)$ and use the fact that F is nondecreasing to conclude that

$$H(x) = \frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt = \frac{1}{2h}\int_{y-h}^{y+h} F\big(u-(y-x)\big)\,du \le \frac{1}{2h}\int_{y-h}^{y+h} F(u)\,du = H(y).$$

Hence, H is nondecreasing.

We can write

$$H(x) = \frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt = \frac{1}{2h}\left(\int_0^{x+h} F(t)\,dt - \int_0^{x-h} F(t)\,dt\right).$$

From calculus, the two functions of x within the parentheses are everywhere continuous and, hence, so is their difference. Thus, H is everywhere continuous and, in particular, everywhere right continuous.

Let $\epsilon > 0$ be given. Because $F(-\infty) = 0$, we can choose M so that $F(t) < \epsilon$ whenever $t < M$. Then, for $x < M - h$,

$$H(x) = \frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt \le \frac{1}{2h}\int_{x-h}^{x+h} \epsilon\,dt = \epsilon.$$

Thus, $H(-\infty) = 0$.

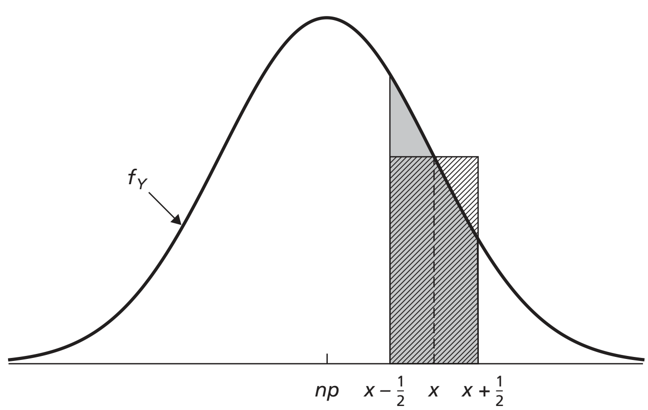
Let $\epsilon > 0$ be given. Because $F(\infty) = 1$, we can choose M so that $F(t) > 1 - \epsilon$ whenever $t > M$. Then, for $x > M + h$,

$$H(x) = \frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt \ge \frac{1}{2h}\int_{x-h}^{x+h} (1-\epsilon)\,dt = 1 - \epsilon.$$

Thus, $H(\infty) = 1$.

8.209

a) Consider the following graph:



The height of the dashed line is $f_Y(x)$, which also equals the area of the crosshatched rectangle. However, the area of the crosshatched rectangle approximately equals the shaded area under the normal curve, or $P(x - 1/2 < Y < x + 1/2)$.

b) Now let $Y \sim \mathcal{N}(np, np(1-p))$. Referring to Equation (8.39) on page 445, part (a), and Proposition 8.11 on page 443, we get

$$p_X(x) \approx f_Y(x) \approx P(x - 1/2 < Y < x + 1/2) = \Phi\left(\frac{x + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x - 1/2 - np}{\sqrt{np(1-p)}}\right),$$

for $x = 0, 1, \ldots, n$.

From the FPF and the preceding relation, we now conclude that

$$P(a \le X \le b) = \sum_{x=a}^{b} p_X(x) \approx \sum_{x=a}^{b} \left[\Phi\left(\frac{x + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x - 1/2 - np}{\sqrt{np(1-p)}}\right) \right]$$

$$= \Phi\left(\frac{b + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{np(1-p)}}\right).$$

c) Here $n = 50$ and $p = 0.4$. Hence,

$$P(X = 21) = P(21 \le X \le 21) \approx \Phi\left(\frac{21 + 1/2 - 50 \cdot 0.4}{\sqrt{50 \cdot 0.4 \cdot 0.6}}\right) - \Phi\left(\frac{21 - 1/2 - 50 \cdot 0.4}{\sqrt{50 \cdot 0.4 \cdot 0.6}}\right) = 0.1101,$$

where, for better results, we used statistical software, rather than Table I, to evaluate the two values of Φ. Also,

$$P(19 \le X \le 21) \approx \Phi\left(\frac{21 + 1/2 - 50 \cdot 0.4}{\sqrt{50 \cdot 0.4 \cdot 0.6}}\right) - \Phi\left(\frac{19 - 1/2 - 50 \cdot 0.4}{\sqrt{50 \cdot 0.4 \cdot 0.6}}\right) = 0.3350.$$

In Example 8.15, where we used the local De Moivre–Laplace theorem, the probabilities obtained were 0.110 and 0.336, respectively, or to four decimal places, 0.1105 and 0.3362.

d) Suppose that you want to use a normal distribution to approximate the probability that a binomial random variable takes a value among m consecutive nonnegative integers. Using the local De Moivre–Laplace theorem, m values of a normal PDF must be calculated and summed, whereas, using the integral De Moivre–Laplace theorem, only two values of the standard normal CDF must be calculated and subtracted. Clearly, then, when m is large, use of the integral De Moivre–Laplace theorem is preferable to that of the local De Moivre–Laplace theorem.

8.210 Answers will vary, but here are some possibilities.

a) If $X \sim \mathcal{U}(0, 1)$, then X has a unique pth quantile equal to p.

b) If $X \sim \mathcal{B}(1, 1 - p)$, then every number in $[0, 1]$ is a pth quantile of X.

8.211

a) Suppose that ξ_p is a pth quantile of X. Then

$$F_X(\xi_p-) = P(X < \xi_p) = 1 - P(X \ge \xi_p) \le 1 - (1 - p) = p$$

and $F_X(\xi_p) = P(X \le \xi_p) \ge p$. Conversely, suppose that $F_X(\xi_p-) \le p \le F_X(\xi_p)$. Then we have $P(X \le \xi_p) = F_X(\xi_p) \ge p$ and

$$P(X \ge \xi_p) = 1 - P(X < \xi_p) = 1 - F_X(\xi_p-) \ge 1 - p.$$

Thus, ξ_p is a pth quantile of X.

b) The conditions $P(X < \xi_p) \le p$ and $P(X > \xi_p) \le 1 - p$ are precisely the same as the conditions $F_X(\xi_p-) \le p$ and $F_X(\xi_p) \ge p$, respectively. Therefore, the equivalence follows from part (a).

8.212 Let $X = I_E$. We have

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - P(E), & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \qquad F_X(x-) = \begin{cases} 0, & \text{if } x \le 0; \\ 1 - P(E), & \text{if } 0 < x \le 1; \\ 1, & \text{if } x > 1. \end{cases}$$

We use the condition for being a pth quantile from Exercise 8.211(a). It follows easily that no real number less than 0 or greater than 1 can be a pth quantile of X. For other values of x, we consider the cases $x = 0$, $x = 1$, and $0 < x < 1$ separately.

We see that $x = 0$ is a pth quantile of X if and only if

$$0 = F_X(0-) \leq p \leq F_X(0) = 1 - P(E),$$

which is true if and only if $P(E) \leq 1 - p$. We see that $x = 1$ is a pth quantile of X if and only if

$$1 - P(E) = F_X(1-) \leq p \leq F_X(1) = 1,$$

which is true if and only if $P(E) \geq 1 - p$. If $0 < x < 1$, then x is a pth quantile of X if and only if

$$1 - P(E) = F_X(x-) \leq p \leq F_X(x) = 1 - P(E),$$

which is true if and only if $P(E) = 1 - p$.

Consequently, if $P(E) < 1 - p$, then 0 is the unique pth quantile of X; if $P(E) = 1 - p$, then all numbers between 0 and 1, inclusive, are pth quantiles of X; and if $P(E) > 1 - p$, then 1 is the unique pth quantile of X.

8.213

a) We know that $F_X(-\infty) = 0$ and $F_X(\infty) = 1$. Furthermore, because X is a continuous random variable, F_X is everywhere continuous. Hence, the intermediate value theorem guarantees the existence of an $x \in \mathcal{R}$ such that $F_X(x) = p$.

b) Because F_X is everywhere continuous, we have $F_X(x-) = F_X(x)$ for all $x \in \mathcal{R}$. Referring to Exercise 8.211(a), we conclude that ξ_p is a pth quantile of X if and only if $F_X(\xi_p) \leq p \leq F_X(\xi_p)$, that is, if and only if $F_X(\xi_p) = \xi_p$.

8.214 Let the range of X be the interval (a, b), where $-\infty \leq a < b \leq \infty$. Note that we have $F_X(a) = 0$ and $F_X(b) = 1$. Because X is a continuous random variable, its CDF is everywhere continuous. Applying the intermediate value theorem, we conclude that there is a number ξ_p, with $a < \xi_p < b$, such that $F_X(\xi_p) = p$ and, because F_X is strictly increasing on the range of X, we know that ξ_p is the unique such number. In view of Exercise 8.213(b), we can now conclude that $F_X^{-1}(p)$ is the unique pth quantile of X.

8.215

a) Here the CDF of X is $F_X(x) = x$ for $0 \leq x < 1$. Therefore, $F_X^{-1}(y) = y$ for $y \in (0, 1)$. Thus, the unique pth percentile of X is given by $\xi_p = F_X^{-1}(p) = p$.

b) Here the CDF of X is $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$. Hence, $F_X^{-1}(y) = -\lambda^{-1} \ln(1 - y)$ for $y \in (0, 1)$. Thus, the unique pth percentile of X is given by $\xi_p = F_X^{-1}(p) = -\lambda^{-1} \ln(1 - p)$.

8.216 Let $A_p = \{\, x : F_X(x) \geq p \,\}$ and note that, as $F_X(-\infty) = 0$ and $F_X(\infty) = 1$, A_p is nonempty and bounded below. Set $\xi_p = \inf A_p$. We claim that ξ_p is a pth quantile of X. Let $\{x_n\}_{n=1}^{\infty}$ be a decreasing sequence of elements of A_p such that $\lim_{n\to\infty} x_n = \xi_p$. Applying the right continuity of F_X, we get

$$P(X \leq \xi_p) = F_X(\xi_p) = \lim_{n\to\infty} F_X(x_n) \geq p.$$

Next let $\{x_n\}_{n=1}^{\infty}$ be an increasing sequence of real numbers such that $\lim_{n\to\infty} x_n = \xi_p$. For each $n \in \mathcal{N}$, we have $x_n < \xi_p$ and hence $x_n \notin A_p$ which, in turn, implies that $F_X(x_n) < p$. Therefore,

$$P(X < x_p) = F_X(\xi_p-) = \lim_{n\to\infty} F_X(x_n) \leq p.$$

Consequently,

$$P(X \geq \xi_p) = 1 - P(X < \xi_p) \geq 1 - p.$$

We have now shown that $P(X \leq \xi_p) \geq p$ and $P(X \geq \xi_p) \geq 1 - p$. Thus, ξ_p is a pth quantile of X.

8.217 Let X denote loss incurred. Then $X \sim \mathcal{E}(1/300)$. We want to find the 95th percentile of the random variable $X_{|X>100}$. In view of Exercise 8.213(b), this means that we need to determine x such that

$$0.95 = F_{X_{|X>100}}(x) = P(X \le x \mid X > 100),$$

or, equivalently, $P(X > x \mid X > 100) = 0.05$. Applying the lack-of-memory property of exponential random variables, we get, for $x > 100$,

$$P(X > x \mid X > 100) = P(X > x - 100) = e^{-(x-100)/300}.$$

Setting this last expression equal to 0.05 and solving for x, we get $x = 100 - 300 \ln(0.05) = 998.72$.

CHAPTER NINE

Instructor's Solutions Manual

Anna Amirdjanova
University of Michigan, Ann Arbor

Neil A. Weiss
Arizona State University

FOR

A Course In

Probability

Neil A. Weiss
Arizona State University

PEARSON
Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

Publisher: Greg Tobin
Editor-in-Chief: Deirdre Lynch
Associate Editor: Sara Oliver Gordus
Editorial Assistant: Christina Lepre
Production Coordinator: Kayla Smith-Tarbox
Senior Author Support/Technology Specialist: Joe Vetere
Compositor: Anna Amirdjanova and Neil A. Weiss
Accuracy Checker: Delray Schultz
Proofreader: Carol A. Weiss

Chapter 9

Jointly Continuous Random Variables

9.1 Joint Cumulative Distribution Functions

Basic Exercises

9.1 For $u, v \in \mathcal{R}$,

$$\begin{aligned} F_{U,V}(u,v) &= P(U \le u,\ V \le v) = P(a + bX \le u,\ c + dY \le v) \\ &= P\big(X \le (u-a)/b,\ Y \le (v-c)/d\big) = F_{X,Y}\big((u-a)/b, (v-c)/d\big). \end{aligned}$$

9.2 If not both u and v are positive, then $F_{U,V}(u,v) = 0$. For $u, v > 0$, we apply Proposition 9.1 on page 488 and use the fact that X and Y are continuous random variables to get

$$\begin{aligned} F_{U,V}(u,v) &= P(U \le u,\ V \le v) = P\big(X^2 \le u,\ Y^2 \le v\big) \\ &= P\big(-\sqrt{u} \le X \le \sqrt{u},\ -\sqrt{v} \le Y \le \sqrt{v}\big) \\ &= F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big) - F_{X,Y}\big(\sqrt{u}, -\sqrt{v}\big) - F_{X,Y}\big(-\sqrt{u}, \sqrt{v}\big) + F_{X,Y}\big(-\sqrt{u}, -\sqrt{v}\big). \end{aligned}$$

9.3
a) For $0 \le x < 1$, we use Proposition 9.2 and Example 9.1 to get

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y) = x.$$

Hence,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & x \ge 0. \end{cases}$$

b) $X \sim \mathcal{U}(0,1)$
c) By symmetry, Y has the same probability distribution as X and, hence, the same CDF. From part (b), we see that $Y \sim \mathcal{U}(0,1)$.

9.4 Let Ω denote the upper half of the unit disk.
a) Clearly, $F_X(x) = 0$ if $x < -1$, and $F_X(x) = 1$ if $x \ge 1$. For $-1 \le x < 1$,

$$\begin{aligned} F_X(x) &= P(X \le x) = \frac{|\{X \le x\}|}{|\Omega|} = \frac{\int_{-1}^{x} \sqrt{1-t^2}\,dt}{\pi/2} = \frac{2}{\pi} \cdot \frac{1}{2}\Big[t\sqrt{1-t^2} + \arcsin t\Big]_{-1}^{x} \\ &= \frac{1}{\pi}\left(x\sqrt{1-x^2} + \arcsin x + \frac{\pi}{2}\right) = \frac{1}{2} + \frac{1}{\pi}\left(x\sqrt{1-x^2} + \arcsin x\right). \end{aligned}$$

b) Clearly, $F_Y(y) = 0$ if $y < 0$, and $F_Y(y) = 1$ if $y \geq 1$. For $0 \leq y < 1$,

$$F_Y(y) = P(Y \leq y) = \frac{|\{Y \leq y\}|}{|\Omega|} = \frac{\int_0^y 2\sqrt{1-t^2}\,dt}{\pi/2}$$
$$= \frac{2}{\pi} \cdot \left[t\sqrt{1-t^2} + \arcsin t\right]_0^y = \frac{2}{\pi}\left(y\sqrt{1-y^2} + \arcsin y\right).$$

c) Clearly, $F_{X,Y}(x, y) = 0$ if $x < -1$ or $y < 0$, and $F_{X,Y}(x, y) = 1$ if $x \geq 1$ and $y \geq 1$. For the other possibilities, we consider five separate cases.

Case 1: $-1 \leq x < 1$ and $0 \leq y < \sqrt{1-x^2}$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{|\{X \leq x,\ Y \leq y\}|}{|\Omega|} = \frac{2}{\pi}\int_0^y \left(x + \sqrt{1-t^2}\right) dt$$
$$= \frac{2}{\pi}\left(xy + \frac{1}{2}\left(y\sqrt{1-y^2} + \arcsin y\right)\right) = \frac{1}{\pi}\left(2xy + y\sqrt{1-y^2} + \arcsin y\right).$$

Case 2: $-1 \leq x < 1$ and $y \geq 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{|\{X \leq x,\ Y \leq y\}|}{|\Omega|} = \frac{2}{\pi}\int_{-1}^x \sqrt{1-t^2}\,dt$$
$$= \frac{2}{\pi}\left(\frac{1}{2}\left(x\sqrt{1-x^2} + \arcsin x + \frac{\pi}{2}\right)\right) = \frac{1}{2} + \frac{1}{\pi}\left(x\sqrt{1-x^2} + \arcsin x\right).$$

Case 3: $0 \leq x < 1$ and $\sqrt{1-x^2} \leq y < 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{|\{X \leq x,\ Y \leq y\}|}{|\Omega|}$$
$$= \frac{2}{\pi}\left(\int_0^{\sqrt{1-x^2}} \left(x + \sqrt{1-t^2}\right) dt + \int_{\sqrt{1-x^2}}^y 2\sqrt{1-t^2}\,dt\right)$$
$$= \frac{1}{\pi}\left(x\sqrt{1-x^2} - \arcsin\sqrt{1-x^2} + 2y\sqrt{1-y^2} + 2\arcsin y\right).$$

Case 4: $-1 \leq x < 0$ and $\sqrt{1-x^2} \leq y < 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{|\{X \leq x,\ Y \leq y\}|}{|\Omega|}$$
$$= \frac{2}{\pi}\left(\int_0^{\sqrt{1-x^2}} \left(x + \sqrt{1-t^2}\right) dt\right) = \frac{1}{\pi}\left(x\sqrt{1-x^2} + \arcsin\sqrt{1-x^2}\right).$$

Case 5: $x \geq 1$ and $0 \leq y < 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{|\{X \leq x,\ Y \leq y\}|}{|\Omega|}$$
$$= \frac{2}{\pi}\left(\int_0^y 2\sqrt{1-t^2}\,dt\right) = \frac{2}{\pi}\left(y\sqrt{1-y^2} + \arcsin y\right).$$

d) Clearly, $F_X(x) = 0$ if $x < -1$, and $F_X(x) = 1$ if $x \geq 1$. For $-1 \leq x < 1$, we apply Proposition 9.2 and refer to Case 2 of part (c) to get

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) = \frac{1}{2} + \frac{1}{\pi}\left(x\sqrt{1-x^2} + \arcsin x\right),$$

which agrees with the answer obtained in part (a).
e) Clearly, $F_Y(y) = 0$ if $y < 0$, and $F_Y(y) = 1$ if $y \geq 1$. For $0 \leq y < 1$, we apply Proposition 9.2 and refer to Case 5 of part (c) to get

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x, y) = \frac{2}{\pi}\left(y\sqrt{1-y^2} + \arcsin y\right),$$

which agrees with the answer obtained in part (b).

9.5 Here the sample space is $\Omega = \{(0,0), (1,0), (0,1), (1,1)\}$.
a) Clearly, $F_{X,Y}(x, y) = 0$ if $x < 0$ or $y < 0$, and $F_{X,Y}(x, y) = 1$ if $x \geq 1$ and $y \geq 1$. For the other possibilities, we consider three separate cases.

Case 1: $0 \leq x < 1$ and $0 \leq y < 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{N(\{X \leq x,\ Y \leq y\})}{N(\Omega)} = \frac{N(\{(0,0)\})}{4} = \frac{1}{4}.$$

Case 2: $0 \leq x < 1$ and $y \geq 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{N(\{X \leq x,\ Y \leq y\})}{N(\Omega)} = \frac{N(\{(0,0), (0,1)\})}{4} = \frac{2}{4} = \frac{1}{2}.$$

Case 3: $x \geq 1$ and $0 \leq y < 1$.

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{N(\{X \leq x,\ Y \leq y\})}{N(\Omega)} = \frac{N(\{(0,0), (1,0)\})}{4} = \frac{2}{4} = \frac{1}{2}.$$

Thus,

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ 1/4, & \text{if } 0 \leq x < 1 \text{ and } 0 \leq y < 1; \\ 1/2, & \text{if } 0 \leq x < 1 \text{ and } y \geq 1, \text{ or if } x \geq 1 \text{ and } 0 \leq y < 1; \\ 1, & \text{if } x \geq 1 \text{ and } y \geq 1. \end{cases}$$

b) For $0 \leq x < 1$, we apply Proposition 9.2 and refer to Case 2 of part (a) to get

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) = \frac{1}{2}.$$

Thus,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1/2, & \text{if } 0 \leq x < 1; \\ 1, & \text{if } x \geq 1. \end{cases}$$

c) The CDF obtained in part (b) is that of a Bernoulli random variable with parameter 1/2. Hence, X has that probability distribution.

d) For $0 \le y < 1$, we apply Proposition 9.2 and refer to Case 3 of part (a) to get

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x, y) = \frac{1}{2}.$$

Thus,

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ 1/2, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

This CDF is that of a Bernoulli random variable with parameter 1/2. Hence, Y has that probability distribution.

e) From parts (a), (b), and (d), we see that $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all $x, y \in \mathcal{R}$.

9.6 By assumption, the marginal PMFs of X and Y are as follows:

$$p_X(x) = \begin{cases} 1-p, & \text{if } x = 0; \\ p, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad p_Y(y) = \begin{cases} 1-p, & \text{if } y = 0; \\ p, & \text{if } y = 1; \\ 0, & \text{otherwise.} \end{cases}$$

a) Using the marginal PMFs of X and Y, the assumed independence, and Proposition 6.10 on page 292, we get the joint PMF of X and Y:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) = \begin{cases} (1-p)^2, & \text{if } x = 0 \text{ and } y = 0; \\ p(1-p), & \text{if } x = 0 \text{ and } y = 1, \text{ or if } x = 1 \text{ and } y = 0; \\ p^2, & \text{if } x = 1 \text{ and } y = 1. \end{cases}$$

b) From the FPF for two discrete random variables, Proposition 6.3 on page 265,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y) = \sum_{s\le x,\, t\le y}\sum p_{X,Y}(s, t).$$

Clearly, $F_{X,Y}(x, y) = 0$ if $x < 0$ or $y < 0$, and $F_{X,Y}(x, y) = 1$ if $x \ge 1$ and $y \ge 1$. For the other possibilities, we consider three separate cases.

Case 1: $0 \le x < 1$ and $0 \le y < 1$.

$$F_{X,Y}(x, y) = \sum_{s\le x,\, t\le y}\sum p_{X,Y}(s, t) = (1-p)^2.$$

Case 2: $0 \le x < 1$ and $y \ge 1$.

$$F_{X,Y}(x, y) = \sum_{s\le x,\, t\le y}\sum p_{X,Y}(s, t) = (1-p)^2 + p(1-p) = 1-p.$$

Case 3: $x \ge 1$ and $0 \le y < 1$.

$$F_{X,Y}(x, y) = \sum_{s\le x,\, t\le y}\sum p_{X,Y}(s, t) = (1-p)^2 + p(1-p) = 1-p.$$

Thus,

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ (1-p)^2, & \text{if } 0 \le x < 1 \text{ and } 0 \le y < 1; \\ 1-p, & \text{if } 0 \le x < 1 \text{ and } y \ge 1, \text{ or if } x \ge 1 \text{ and } 0 \le y < 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

c) For $0 \le x < 1$, we apply Proposition 9.2 and refer to Case 2 of part (b) to get

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) = 1 - p.$$

Thus,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - p, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

Of course, Y has the same CDF as X, because these two random variables have the same probability distribution.

d) Applying the FPF for one discrete random variable, Proposition 5.2 on page 190, to the PMF of X, obtained at the beginning of the solution to this exercise, yields

$$F_X(x) = P(X \le x) = \sum_{s\le x} p_X(s) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - p, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

Of course, Y has the same CDF as X, because these two random variables have the same probability distribution.

e) Using the assumed independence of X and Y, and the result of part (d) yields

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \le x,\ Y \le y) = P(X \le x)P(Y \le y) \\ &= F_X(x)F_Y(y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ (1 - p)^2, & \text{if } 0 \le x < 1 \text{ and } 0 \le y < 1; \\ 1 - p, & \text{if } 0 \le x < 1 \text{ and } y \ge 1, \text{ or if } x \ge 1 \text{ and } 0 \le y < 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases} \end{aligned}$$

f) The answers in parts (b) and (e) are identical.

9.7

a) From Exercise 8.70(a), the CDFs of X and Y are

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-\lambda x}, & \text{if } x \ge 0. \end{cases} \quad \text{and} \quad F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ 1 - e^{-\mu y}, & \text{if } y \ge 0. \end{cases}$$

b) Referring to Definition 6.4 on page 291 and then to part (a), we get

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \le x,\ Y \le y) = P(X \le x)P(Y \le y) \\ &= F_X(x)F_Y(y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ 1 - e^{-\lambda x} - e^{-\mu y} + e^{-\lambda x - \mu y}, & \text{if } x \ge 0 \text{ and } y \ge 0. \end{cases} \end{aligned}$$

9.8 Suppose that F is a joint CDF, say, of the random variables X and Y. Then, by Proposition 9.1 on page 488,

$$\begin{aligned} P(0 < X \le 1,\ 0 < Y \le 1) &= F_{X,Y}(1, 1) - F_{X,Y}(1, 0) - F_{X,Y}(0, 1) + F_{X,Y}(0, 0) \\ &= (1 - e^{-2}) - (1 - e^{-1}) - (1 - e^{-1}) + (1 - e^{-0}) \\ &= 2e^{-1} - e^{-2} - 1 < 0, \end{aligned}$$

which is impossible. Hence, F is not a joint CDF.

9.9

a) We have

$$F_V(v) = P(V \le v) = P\big(\max\{X, Y\} \le v\big) = P(X \le v,\ Y \le v) = F_{X,Y}(v, v).$$

b) Applying the general addition rule, we get

$$\begin{aligned}F_U(u) &= P(U \le u) = P\big(\min\{X, Y\} \le u\big) = P(\{X \le u\} \cup \{Y \le u\}) \\ &= P(X \le u) + P(Y \le u) - P(X \le u,\ Y \le u) = F_X(u) + F_Y(u) - F_{X,Y}(u, u).\end{aligned}$$

c) From part (b) and Proposition 9.2 on page 490,

$$F_U(u) = F_X(u) + F_Y(u) - F_{X,Y}(u, u) = F_{X,Y}(u, \infty) + F_{X,Y}(\infty, u) - F_{X,Y}(u, u).$$

9.10
a) We have

$$\begin{aligned}F_V(v) &= P(V \le v) = P\big(\max\{X_1, \dots, X_m\} \le v\big) \\ &= P(X_1 \le v, \dots, X_m \le v) = F_{X_1,\dots,X_m}(v, \dots, v).\end{aligned}$$

b) Applying the inclusion-exclusion principle yields

$$\begin{aligned}F_U(u) &= P(U \le u) = P\big(\min\{X_1, \dots, X_m\} \le u\big) = P\left(\bigcup_{n=1}^{m}\{X_n \le u\}\right) \\ &= \sum_{k=1}^{m} P(X_k \le u) - \sum_{k_1<k_2} P(X_{k_1} \le u,\ X_{k_2} \le u) + \cdots \\ &\quad + (-1)^{n+1} \sum_{k_1<k_2<\cdots<k_n} P(X_{k_1} \le u, \dots, X_{k_n} \le u) \\ &\quad + \cdots + (-1)^{m+1} P(X_1 \le u, \dots, X_m \le u) \\ &= \sum_{k=1}^{m} F_{X_k}(u) - \sum_{k_1<k_2} F_{X_{k_1},X_{k_2}}(u, u) + \cdots \\ &\quad + (-1)^{n+1} \sum_{k_1<k_2<\cdots<k_n} F_{X_{k_1},\dots,X_{k_n}}(u, \dots, u) \\ &\quad + \cdots + (-1)^{m+1} F_{X_1,\dots,X_m}(u, \dots, u).\end{aligned}$$

Theory Exercises

9.11
a) Suppose that $x_1 < x_2$. Then $\{X \le x_1,\ Y \le y\} \subset \{X \le x_2,\ Y \le y\}$. Consequently, by the domination principle,

$$F_{X,Y}(x_1, y) = P(X \le x_1,\ Y \le y) \le P(X \le x_2,\ Y \le y) = F_{X,Y}(x_2, y).$$

Thus, $F_{X,Y}$ is nondecreasing as a function of its first variable. Similarly, it is also nondecreasing as a function of its second variable.
b) Suppose that $x \in \mathcal{R}$ and let $\{x_n\}_{n=1}^{\infty}$ be any decreasing sequence of real numbers that converges to x. For each $n \in \mathcal{N}$, set $A_n = \{X \le x_n,\ Y \le y\}$. Then $A_1 \supset A_2 \supset \cdots$ and $\bigcap_{n=1}^{\infty} A_n = \{X \le x,\ Y \le y\}$.

Applying the continuity property of probability (Proposition 2.11 on page 74) gives

$$\lim_{n\to\infty} F_{X,Y}(x_n, y) = \lim_{n\to\infty} P(X \le x_n,\ Y \le y) = \lim_{n\to\infty} P(A_n)$$
$$= P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(X \le x,\ Y \le y) = F_{X,Y}(x, y).$$

Hence $F_{X,Y}$ is everywhere right-continuous as a function of its first variable. Similarly, it is everywhere right-continuous as a function of its second variable.
c) Let $\{y_n\}_{n=1}^{\infty}$ be any decreasing sequence of real numbers such that $\lim_{n\to\infty} y_n = -\infty$. For each $n \in \mathcal{N}$, set $A_n = \{X \le x,\ Y \le y_n\}$. Then $A_1 \supset A_2 \supset \cdots$ and $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Applying the continuity property of probability yields

$$\lim_{n\to\infty} F_{X,Y}(x, y_n) = \lim_{n\to\infty} P(X \le x,\ Y \le y_n) = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\emptyset) = 0.$$

Therefore, $\lim_{y\to-\infty} F_{X,Y}(x, y) = 0$.
d) Let $\{x_n\}_{n=1}^{\infty}$ be any decreasing sequence of real numbers such that $\lim_{n\to\infty} x_n = -\infty$. For each $n \in \mathcal{N}$, set $A_n = \{X \le x_n,\ Y \le y\}$. Then $A_1 \supset A_2 \supset \cdots$ and $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Applying the continuity property of probability yields

$$\lim_{n\to\infty} F_{X,Y}(x_n, y) = \lim_{n\to\infty} P(X \le x_n,\ Y \le y) = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\emptyset) = 0.$$

Therefore, $\lim_{x\to-\infty} F_{X,Y}(x, y) = 0$.
e) Let $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ be any increasing sequences of real numbers such that $\lim_{n\to\infty} x_n = \infty$ and $\lim_{n\to\infty} y_n = \infty$. For $n \in \mathcal{N}$, set $A_n = \{X \le x_n,\ Y \le y_n\}$. Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \Omega$. Applying the continuity property of probability gives

$$\lim_{n\to\infty} F_{X,Y}(x_n, y_n) = \lim_{n\to\infty} P(X \le x_n,\ Y \le y_n) = \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\Omega) = 1.$$

Therefore, $\lim_{x,y\to\infty} F_{X,Y}(x, y) = 1$.

9.12 Let

$$C = \{\, (x_1, x_2, x_3) : a_i < x_i \le b_i,\ i = 1, 2, 3 \,\}$$

and

$$H_{u_1,u_2,u_3} = \{\, (x_1, x_2, x_3) : x_i \le u_i,\ i = 1, 2, 3 \,\}.$$

Note that

$$F_{X_1,X_2,X_3}(u_1, u_2, u_3) = P\big((X_1, X_2, X_3) \in H_{u_1,u_2,u_3}\big).$$

We have

$$H_{b_1,b_2,b_3} = C \cup \big(H_{b_1,b_2,a_3} \cup H_{b_1,a_2,b_3} \cup H_{a_1,b_2,b_3}\big),$$

and C is disjoint with the union of the other three sets on the right. Also observe that

$$H_{b_1,b_2,a_3} \cap H_{b_1,a_2,b_3} = H_{b_1,a_2,a_3}, \quad H_{b_1,b_2,a_3} \cap H_{a_1,b_2,b_3} = H_{a_1,b_2,a_3}, \quad H_{b_1,a_2,b_3} \cap H_{a_1,b_2,b_3} = H_{a_1,a_2,b_3},$$

and

$$H_{b_1,b_2,a_3} \cap H_{b_1,a_2,b_3} \cap H_{a_1,b_2,b_3} = H_{a_1,a_2,a_3}.$$

Applying the inclusion-exclusion principle for three events, Equation (2.17) on page 71, we get

$$\begin{aligned} P\big((X_1, X_2, X_3) \in H_{b_1,b_2,b_3}\big) = {} & P\big((X_1, X_2, X_3) \in C\big) + P\big((X_1, X_2, X_3) \in H_{b_1,b_2,a_3}\big) \\ & + P\big((X_1, X_2, X_3) \in H_{b_1,a_2,b_3}\big) + P\big((X_1, X_2, X_3) \in H_{a_1,b_2,b_3}\big) \\ & - P\big((X_1, X_2, X_3) \in H_{b_1,a_2,a_3}\big) - P\big((X_1, X_2, X_3) \in H_{a_1,b_2,a_3}\big) \\ & - P\big((X_1, X_2, X_3) \in H_{a_1,a_2,b_3}\big) + P\big((X_1, X_2, X_3) \in H_{a_1,a_2,a_3}\big). \end{aligned}$$

Rearranging terms, we get

$$\begin{aligned} P(a_i < X_i \le b_i, i = 1, 2, 3) = {} & F_{X_1,X_2,X_3}(b_1, b_2, b_3) - F_{X_1,X_2,X_3}(b_1, b_2, a_3) \\ & - F_{X_1,X_2,X_3}(b_1, a_2, b_3) - F_{X_1,X_2,X_3}(a_1, b_2, b_3) \\ & + F_{X_1,X_2,X_3}(b_1, a_2, a_3) + F_{X_1,X_2,X_3}(a_1, b_2, a_3) \\ & + F_{X_1,X_2,X_3}(a_1, a_2, b_3) - F_{X_1,X_2,X_3}(a_1, a_2, a_3). \end{aligned}$$

9.13 Let $\{y_n\}_{n=1}^{\infty}$ be any increasing sequence of real numbers such that $\lim_{n\to\infty} y_n = \infty$. For $n \in \mathcal{N}$, set $A_n = \{X \le x,\ Y \le y_n\}$. Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \{X \le x\}$. Applying the continuity property of probability gives

$$\lim_{n\to\infty} F_{X,Y}(x, y_n) = \lim_{n\to\infty} P(X \le x,\ Y \le y_n) = \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(X \le x) = F_X(x).$$

Thus, $\lim_{y\to\infty} F_{X,Y}(x, y) = F_X(x)$. Similarly, we find that $\lim_{x\to\infty} F_{X,Y}(x, y) = F_Y(y)$.

Advanced Exercises

9.14 The sample space is $\Omega = \{\, (x, y) \in \mathcal{R}^2 : y = x, 0 \le x \le 1 \,\}$.
a) For $0 \le x < 1$, observe that $\{X \le x\} = \{\, (s, y) \in \mathcal{R}^2 : y = s, 0 \le s \le x \,\}$. Consequently,

$$F_X(x) = P(X \le x) = \frac{|\{X \le x\}|}{|\Omega|} = \frac{\sqrt{2}x}{\sqrt{2}} = x.$$

Thus,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

b) From part (a), we see that $X \sim \mathcal{U}(0, 1)$.
c) We note that $P(Y = X) = 1$ and, therefore,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y) = P\big(X \le \min\{x, y\}\big) = F_X\big(\min\{x, y\}\big).$$

Consequently, in view of part (a),

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ x, & \text{if } 0 \le x < 1 \text{ and } y \ge x; \\ y, & \text{if } 0 \le y < 1 \text{ and } x \ge y; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

d) Referring to part (c), we see that, for $0 \le x < 1$,

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

This result agrees with the one obtained in part (a).

e) By symmetry, Y has the same probability distribution as X. Thus,

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

and $Y \sim \mathcal{U}(0, 1)$.

9.15

a) Let $x, y \in \mathcal{R}$. Applying Definition 6.4 with $A = (-\infty, x]$ and $B = (-\infty, y]$ yields,

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \le x,\ Y \le y) = P(X \in A,\ Y \in B) \\ &= P(X \in A)P(Y \in B) = P(X \le x)P(Y \le y) = F_X(x)F_Y(y). \end{aligned}$$

b) Let U and V be independent random variables with the same marginal probability distributions as X and Y, respectively. Applying part (a) to U and V, we obtain

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) = F_U(x)F_V(y) = F_{U,V}(x, y), \qquad x, y \in \mathcal{R}.$$

Thus, the joint CDF of X and Y is the same as that of U and V. Consequently, because the joint CDF of two random variables completely determines their joint probability distribution, X and Y must have the same joint probability distribution as U and V. Thus, for all subsets A and B of real numbers,

$$\begin{aligned} P(X \in A,\ Y \in B) &= P\big((X, Y) \in A \times B\big) = P\big((U, V) \in A \times B\big) = P(U \in A,\ V \in B) \\ &= P(U \in A)P(V \in B) = P(X \in A)P(Y \in B). \end{aligned}$$

Hence, X and Y are independent random variables.

9.16

a) Let $x_1, \dots, x_m \in \mathcal{R}$. Applying Definition 6.5 with $A_j = (-\infty, x_j]$, for $j = 1, 2, \dots, m$, yields,

$$\begin{aligned} F_{X_1,\dots,X_m}(x_1, \dots, x_m) &= P(X_1 \le x_1, \dots, X_m \le x_m) = P(X_1 \in A_1, \dots, X_m \in A_m) \\ &= P(X_1 \in A_1) \cdots P(X_m \in A_m) = P(X_1 \le x_1) \cdots P(X_m \le x_m) \\ &= F_{X_1}(x_1) \cdots F_{X_m}(x_m). \end{aligned}$$

b) Let $Y_1, \dots, Y_m$ be independent random variables with the same marginal probability distributions as $X_1, \dots, X_m$, respectively. Applying part (a) to $Y_1, \dots, Y_m$, we obtain, for $x_1, \dots, x_m \in \mathcal{R}$,

$$F_{X_1,\dots,X_m}(x_1, \dots, x_m) = F_{X_1}(x_1) \cdots F_{X_m}(x_m) = F_{Y_1}(x_1) \cdots F_{Y_m}(x_m) = F_{Y_1,\dots,Y_m}(x_1, \dots, x_m).$$

Thus, the joint CDF of $X_1, \dots, X_m$ is the same as that of $Y_1, \dots, Y_m$. Consequently, because the joint CDF of m random variables completely determines their joint probability distribution, $X_1, \dots, X_m$ must have the same joint probability distribution as $Y_1, \dots, Y_m$. Thus, for all subsets $A_1, \dots, A_m$ of real numbers,

$$\begin{aligned} P(X_1 \in A_1, \dots, X_m \in A_m) &= P\big((X_1, \dots, X_m) \in A_1 \times \cdots \times A_m\big) \\ &= P\big((Y_1, \dots, Y_m) \in A_1 \times \cdots \times A_m\big) = P(Y_1 \in A_1, \dots, Y_m \in A_m) \\ &= P(Y_1 \in A_1) \cdots P(Y_m \in A_m) = P(X_1 \in A_1) \cdots P(X_m \in A_m). \end{aligned}$$

Hence, $X_1, \dots, X_m$ are independent random variables.

9.17 For $x < y$,

$$\begin{aligned} P(X > x,\ Y \le y) &= P\big(\min\{X_1, \dots, X_n\} > x,\ \max\{X_1, \dots, X_n\} \le y\big) \\ &= P(x < X_1 \le y, \dots, x < X_n \le y) \\ &= P(x < X_1 \le y) \cdots P(x < X_n \le y) = \big(F(y) - F(x)\big)^n. \end{aligned}$$

Using the preceding equation and the law of partitions gives

$$P(Y \le y) = P(X > x,\ Y \le y) + P(X \le x,\ Y \le y) = \big(F(y) - F(x)\big)^n + F_{X,Y}(x, y).$$

Noting that $P(Y \le y) = \big(F(y)\big)^n$, we now obtain

$$F_{X,Y}(x, y) = \big(F(y)\big)^n - \big(F(y) - F(x)\big)^n, \qquad x < y.$$

9.18 Let $x_{k_1}, \dots, x_{k_j} \in \mathcal{R}$. For each $i \notin \{k_1, \dots, k_j\}$, let $x_1^{(i)}, x_2^{(i)}, \dots$ be any increasing sequence of real numbers such that $\lim_{n\to\infty} x_n^{(i)} = \infty$. Also, for each $n \in \mathcal{N}$, let

$$y_n^{(i)} = \begin{cases} x_i, & \text{if } i \in \{k_1, \dots, k_j\}; \\ x_n^{(i)}, & \text{if } i \notin \{k_1, \dots, k_j\}. \end{cases} \qquad \text{and} \qquad A_n = \left\{X_1 \le y_n^{(1)}, \dots, X_m \le y_n^{(m)}\right\}.$$

Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^\infty A_n = \{X_{k_1} \le x_{k_1}, \dots, X_{k_j} \le x_{k_j}\}$. Applying the continuity property of probability gives

$$\begin{aligned}
\lim_{\substack{x_i \to \infty \\ i \notin \{k_1,\dots,k_j\}}} F_{X_1,\dots,X_m}(x_1, \dots, x_m) &= \lim_{n\to\infty} F_{X_1,\dots,X_m}\left(y_n^{(1)}, \dots, y_n^{(m)}\right) \\
&= \lim_{n\to\infty} P\left(X_1 \le y_n^{(1)}, \dots, X_m \le y_n^{(m)}\right) = \lim_{n\to\infty} P(A_n) \\
&= P\left(\bigcup_{n=1}^\infty A_n\right) = P(X_{k_1} \le x_{k_1}, \dots, X_{k_j} \le x_{k_j}) \\
&= F_{X_{k_1},\dots,X_{k_j}}(x_{k_1}, \dots, x_{k_j}).
\end{aligned}$$

9.2 Introducing Joing Probability Density Functions

Basic Exercises

9.19 The domination principle and the definition of a continuous random variable give

$$0 \le P(X = x,\ Y = y) \le P(X = x) = 0,$$

which implies that $P(X = x,\ Y = y) = 0$.

9.20 From the solution to Exercise 9.1,

$$F_{U,V}(u, v) = F_{X,Y}\big((u-a)/b, (v-c)/d\big).$$

Applying Proposition 9.3 on page 496 and the two-variable chain rule now gives

$$\begin{aligned}
f_{U,V}(u, v) &= \frac{\partial^2}{\partial u\, \partial v} F_{U,V}(u, v) = \frac{\partial^2}{\partial u\, \partial v} F_{X,Y}\big((u-a)/b, (v-c)/d\big) \\
&= \frac{\partial}{\partial u}\left(\frac{\partial}{\partial v} F_{X,Y}\big((u-a)/b, (v-c)/d\big)\right) = \frac{\partial}{\partial u}\left(\frac{1}{d} \cdot \frac{\partial}{\partial y} F_{X,Y}\big((u-a)/b, (v-c)/d\big)\right) \\
&= \frac{1}{d}\left(\frac{\partial}{\partial u}\left(\frac{\partial}{\partial y} F_{X,Y}\big((u-a)/b, (v-c)/d\big)\right)\right) \\
&= \frac{1}{bd} \cdot \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}\big((u-a)/b, (v-c)/d\big) = \frac{1}{bd} f_{X,Y}\big((u-a)/b, (v-c)/d\big).
\end{aligned}$$

9.21 From the solution to Exercise 9.2,

$$F_{U,V}(u,v) = F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big) - F_{X,Y}\big(\sqrt{u}, -\sqrt{v}\big) - F_{X,Y}\big(-\sqrt{u}, \sqrt{v}\big) + F_{X,Y}\big(-\sqrt{u}, -\sqrt{v}\big).$$

Note that, by the two-variable chain rule and Proposition 9.3 on page 496,

$$\begin{aligned}\frac{\partial^2}{\partial u\,\partial v} F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big) &= \frac{\partial}{\partial u}\left(\frac{\partial}{\partial v} F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big)\right) = \frac{\partial}{\partial u}\left(\frac{1}{2\sqrt{v}} \cdot \frac{\partial}{\partial y} F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big)\right) \\ &= \frac{1}{2\sqrt{v}}\left(\frac{\partial}{\partial u}\left(\frac{\partial}{\partial y} F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big)\right)\right) = \frac{1}{4\sqrt{uv}} \cdot \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}\big(\sqrt{u}, \sqrt{v}\big) \\ &= \frac{1}{4\sqrt{uv}} f_{X,Y}\big(\sqrt{u}, \sqrt{v}\big).\end{aligned}$$

Using similar reasoning for the other three terms on the right of the first display in this solution, we find that

$$\begin{aligned}f_{U,V}(u,v) &= \frac{\partial^2}{\partial u\,\partial v} F_{U,V}(u,v) \\ &= \frac{1}{4\sqrt{uv}}\Big(f_{X,Y}\big(\sqrt{u}, \sqrt{v}\big) + f_{X,Y}\big(\sqrt{u}, -\sqrt{v}\big) \\ &\qquad + f_{X,Y}\big(-\sqrt{u}, \sqrt{v}\big) + f_{X,Y}\big(-\sqrt{u}, -\sqrt{v}\big)\Big).\end{aligned}$$

9.22

a) We have

$$\begin{aligned}P(1/2 \le X \le 3/4,\ 1/4 \le Y \le 3/4) &= \big|\{1/2 \le X \le 3/4,\ 1/4 \le Y \le 3/4)\}\big| \\ &= \big|[1/2, 3/4] \times [1/4, 3/4]\big| = (3/4 - 1/2) \cdot (3/4 - 1/4) \\ &= 0.125.\end{aligned}$$

b) From Proposition 9.1 and the joint CDF of X and Y, both on page 488, we see that

$$\begin{aligned}P(1/2 \le X \le 3/4,\ 1/4 \le Y \le 3/4) &= P(1/2 < X \le 3/4,\ 1/4 < Y \le 3/4) \\ &= F_{X,Y}(3/4, 3/4) - F_{X,Y}(3/4, 1/4) \\ &\quad - F_{X,Y}(1/2, 3/4) + F_{X,Y}(1/2, 1/4) \\ &= (3/4)(3/4) - (3/4)(1/4) - (1/2)(3/4) + (1/2)(1/4) \\ &= 0.125.\end{aligned}$$

c) From Definition 9.2 on page 495 and the joint PDF of X and Y on page 497,

$$\begin{aligned}P(1/2 \le X \le 3/4,\ 1/4 \le Y \le 3/4) &= \int_{1/2}^{3/4}\int_{1/4}^{3/4} f_{X,Y}(x,y)\,dx\,dy \\ &= \int_{1/2}^{3/4}\int_{1/4}^{3/4} 1\,dx\,dy = (3/4 - 1/2) \cdot (3/4 - 1/4) \\ &= 0.125.\end{aligned}$$

d) First note that, by the complementation rule and the continuity of the random variables, we have

$$P(X > 0.6 \text{ or } Y < 0.2) = 1 - P(0 < X \le 0.6,\ 0.2 \le Y \le 1).$$

Now we proceed as in parts (a)–(c):

$$\begin{aligned}P(0 < X \le 0.6,\ 0.2 \le Y \le 1) &= \left|\{0 < X \le 0.6,\ 0.2 \le Y \le 1\}\right| \\ &= \left|(0, 0.6] \times [0.2, 1]\right| = (0.6 - 0) \cdot (1 - 0.2) = 0.48.\end{aligned}$$

Alternatively,

$$\begin{aligned}P(0 < X \le 0.6,\ 0.2 \le Y \le 1) &= P(0 < X \le 0.6,\ 0.2 < Y \le 1) \\ &= F_{X,Y}(0.6, 1) - F_{X,Y}(0.6, 0.2) \\ &\quad - F_{X,Y}(0, 1) + F_{X,Y}(0, 0.2) \\ &= (0.6)(1) - (0.6)(0.2) - (0)(1) + (0)(0.2) = 0.48\end{aligned}$$

Alternatively,

$$\begin{aligned}P(0 < X \le 0.6,\ 0.2 \le Y \le 1) &= P(0 \le X \le 0.6,\ 0.2 \le Y \le 1) \\ &= \int_0^{0.6} \int_{0.2}^{1} f_{X,Y}(x, y)\, dx\, dy \\ &= \int_0^{0.6} \int_{0.2}^{1} 1\, dx\, dy = (0.6 - 0) \cdot (1 - 0.2) = 0.48.\end{aligned}$$

Thus, in all cases,

$$P(X > 0.6 \text{ or } Y < 0.2) = 1 - P(0 < X \le 0.6,\ 0.2 \le Y \le 1) = 1 - 0.48 = 0.52.$$

9.23

a) From the solution to Exercise 9.4(c), we see that the mixed partial of $F_{X,Y}(x, y)$ is 0 unless (x, y) is in the upper half of the unit disk. In that case,

$$\frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y}\left(\frac{1}{\pi}\left(2xy + y\sqrt{1 - y^2} + \arcsin y\right)\right) = \frac{2}{\pi}.$$

Thus,

$$f_{X,Y}(x, y) = \begin{cases} 2/\pi, & \text{if } -1 < x < 1 \text{ and } 0 < y < \sqrt{1 - x^2}; \\ 0, & \text{otherwise.} \end{cases}$$

b) Because the appropriate probability model is a geometric probability model—a point is being selected at random from the upper half of the unit disk.

9.24

a) From the solution to Exercise 9.7(b),

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ 1 - e^{-\lambda x} - e^{-\mu y} + e^{-\lambda x - \mu y}, & \text{if } x \ge 0 \text{ and } y \ge 0. \end{cases}$$

Therefore,

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) = \begin{cases} \lambda\mu e^{-\lambda x - \mu y}, & \text{if } x > 0 \text{ and } y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

b) As $X \sim \mathcal{E}(\lambda)$ and $Y \sim \mathcal{E}(\mu)$,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} \mu e^{-\mu y}, & \text{if } y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Because $e^{-\lambda x} e^{-\mu y} = e^{-\lambda x - \mu y}$, we see that $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all $x, y \in \mathcal{R}$. In words, the joint PDF of X and Y equals the product of their marginal PDFs.

c) Answers will vary, but you should not be surprised at the multiplicative relationship between the joint PDF and the marginal PDFs, as the multiplication rule has previously applied in many contexts of independence.

9.25 Recall that the CDf and PDF of an $\mathcal{E}(\lambda)$ distribution are given by

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, from the general result obtained in Example 9.4, which begins on page 497, we see that

$$f_{X,Y}(x, y) = n(n-1) f(x) f(y) \big(F(y) - F(x)\big)^{n-2} = n(n-1)\lambda^2 e^{-\lambda(x+y)} \left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2},$$

if $0 < x < y$, and $f_{X,Y}(x, y) = 0$ otherwise.

9.26

a) We have, in view of Definition 9.2 on page 495,

$$\begin{aligned} P(1/4 < X < 3/4,\ 1/2 < Y < 1) &= P(1/4 \leq X \leq 3/4,\ 1/2 \leq Y \leq 1) \\ &= \int_{1/4}^{3/4} \int_{1/2}^{1} f_{X,Y}(x, y)\, dx\, dy = \int_{1/4}^{3/4} \int_{1/2}^{1} (x + y)\, dx\, dy \\ &= \int_{1/4}^{3/4} \left(\int_{1/2}^{1} (x + y)\, dy \right) dx = \frac{5}{16}. \end{aligned}$$

b) Clearly, $F_{X,Y}(x, y) = 0$ if $x < 0$ or $y < 0$, and $F_{X,Y}(x, y) = 1$ if $x \geq 1$ and $y \geq 1$. For $0 \leq x < 1$ and $0 \leq y < 1$,

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\, ds\, dt = \int_{0}^{x} \int_{0}^{y} (s + t)\, ds\, dt \\ &= \int_{0}^{x} \left(\int_{0}^{y} (s + t)\, dt \right) ds = \int_{0}^{x} \left(sy + y^2/2 \right) ds = \frac{1}{2} xy(x + y). \end{aligned}$$

For $0 \leq x < 1$ and $y \geq 1$,

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\, ds\, dt = \int_{0}^{x} \int_{0}^{1} (s + t)\, ds\, dt \\ &= \int_{0}^{x} \left(\int_{0}^{1} (s + t)\, dt \right) ds = \int_{0}^{x} (s + 1/2)\, ds = \frac{1}{2} x(x + 1). \end{aligned}$$

By symmetry, for $x \geq 1$ and $0 \leq y < 1$, we have $F_{X,Y}(x, y) = y(y + 1)/2$. Thus,

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ xy(x + y)/2, & \text{if } 0 \leq x < 1 \text{ and } 0 \leq y < 1; \\ x(x + 1)/2, & \text{if } 0 \leq x < 1 \text{ and } y \geq 1; \\ y(y + 1)/2, & \text{if } x \geq 1 \text{ and } 0 \leq y < 1; \\ 1, & \text{if } x \geq 1 \text{ and } y \geq 1. \end{cases}$$

c) We have, in view of Proposition 9.1 on page 488,

$$
\begin{aligned}
P(1/4 < X < 3/4,\ 1/2 < Y < 1) &= P(1/4 < X \le 3/4,\ 1/2 < Y \le 1) \\
&= F_{X,Y}(3/4, 1) - F_{X,Y}(3/4, 1/2) \\
&\quad - F_{X,Y}(1/4, 1) + F_{X,Y}(1/4, 1/2) \\
&= (3/4)(3/4 + 1)/2 - (3/4)(1/2)(3/4 + 1/2)/2 \\
&\quad - (1/4)(1/4 + 1)/2 + (1/4)(1/2)(1/4 + 1/2)/2,
\end{aligned}
$$

which equals 5/16. This answer agrees with the one obtained in part (a).

Theory Exercises

9.27 Applying the second fundamental theorem of calculus twice and referring to Proposition 9.1 on page 488, we get, for all real numbers, a, b, c, d, with $a < b$ and $c < d$,

$$
\begin{aligned}
\int_a^b \int_c^d \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y)\, dx\, dy &= \int_c^d \left(\int_a^b \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y)\, dx \right) dy \\
&= \int_c^d \left(\frac{\partial}{\partial y} F_{X,Y}(b, y) - \frac{\partial}{\partial y} F_{X,Y}(a, y) \right) dy \\
&= \int_c^d \left(\frac{\partial}{\partial y} F_{X,Y}(b, y) \right) dy - \int_c^d \left(\frac{\partial}{\partial y} F_{X,Y}(a, y) \right) dy \\
&= \big(F_{X,Y}(b, d) - F_{X,Y}(b, c) \big) - \big(F_{X,Y}(a, d) - F_{X,Y}(a, c) \big) \\
&= F_{X,Y}(b, d) - F_{X,Y}(b, c) - F_{X,Y}(a, d) + F_{X,Y}(a, c) \\
&= P(a < X \le b,\ c < Y \le d) = P(a \le X \le b,\ c \le Y \le d).
\end{aligned}
$$

Therefore, from Definition 9.2 on page 495, $\partial^2 F_{X,Y} \big/ \partial x\, \partial y$ is a joint PDF of X and Y. We can use a similar argument to show that $\partial^2 F_{X,Y} \big/ \partial y\, \partial x$ is a joint PDF of X and Y, or, alternatively, we can use the fact that $\partial^2 F_{X,Y} \big/ \partial y\, \partial x = \partial^2 F_{X,Y} \big/ \partial x\, \partial y$.

9.28 Suppose that there is a nonnegative function f defined on $\mathcal{R}^2$ such that Equation (9.21) on page 499 holds. Then $F_{X,Y}$ is everywhere continuous. For $a < b$ and $c < d$, Proposition 9.1 on page 488 gives

$$
\begin{aligned}
P(a \le X \le b,\ c \le Y \le d) &= F_{X,Y}(b, d) - F_{X,Y}(b, c) - F_{X,Y}(a, d) + F_{X,Y}(a, c) \\
&= \int_{-\infty}^b \int_{-\infty}^d f(s, t)\, ds\, dt - \int_{-\infty}^b \int_{-\infty}^c f(s, t)\, ds\, dt \\
&\quad - \int_{-\infty}^a \int_{-\infty}^d f(s, t)\, ds\, dt + \int_{-\infty}^a \int_{-\infty}^c f(s, t)\, ds\, dt \\
&= \int_{-\infty}^b \left(\int_{-\infty}^d f(s, t)\, dt - \int_{-\infty}^c f(s, t)\, dt \right) ds \\
&\quad - \int_{-\infty}^a \left(\int_{-\infty}^d f(s, t)\, dt - \int_{-\infty}^c f(s, t)\, dt \right) ds
\end{aligned}
$$

Therefore,

$$P(a \le X \le b,\ c \le Y \le d) = \int_{-\infty}^{b} \left(\int_{c}^{d} f(s,t)\,dt \right) ds - \int_{-\infty}^{a} \left(\int_{c}^{d} f(s,t)\,dt \right) ds$$
$$= \int_{a}^{b} \left(\int_{c}^{d} f(s,t)\,dt \right) ds = \int_{a}^{b} \int_{c}^{d} f(s,t)\,ds\,dt.$$

Hence, by Definition 9.2 on page 495, f is a joint PDF of X and Y.

Conversely, suppose that X and Y have a joint PDF, $f_{X,Y}$. Then, by Definition 9.2, $f_{X,Y}$ is a nonnegative function defined on $\mathcal{R}^2$ such that, for all $a < b$ and $c < d$,

$$P(a \le X \le b,\ c \le Y \le d) = \int_{a}^{b} \int_{c}^{d} f_{X,Y}(s,t)\,ds\,dt.$$

Let $x, y \in \mathcal{R}$ and set $A_n = \{-n \le X \le x,\ -n \le Y \le y\}$ for each $n \in \mathcal{N}$. Then we have $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \{X \le x,\ Y \le y\}$. Applying the continuity property of a probability measure and the previous display yields

$$F_{X,Y}(x,y) = P(X \le x,\ Y \le y) = P\left(\bigcup_{n=1}^{\infty} A_n \right) = \lim_{n\to\infty} P(A_n)$$
$$= \lim_{n\to\infty} P(-n \le X \le x,\ -n \le Y \le y) = \lim_{n\to\infty} \int_{-n}^{x} \int_{-n}^{y} f_{X,Y}(s,t)\,ds\,dt$$
$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)\,ds\,dt.$$

Thus $f_{X,Y}$ is a nonnegative function defined on $\mathcal{R}^2$ that satisfies Equation (9.21).

Equation (9.22) is a consequence of the first fundamental theorem of calculus. It shows again that a joint PDF of two continuous random variables—when it exists—is essentially a mixed second-order partial derivative of the joint CDF of the two random variables.

9.29

a) A nonnegative function $f_{X_1,\ldots,X_m}$ is said to be a *joint probability density function* of $X_1, \ldots, X_m$ if, for all real numbers a_i and b_i, with $a_i < b_i$ for $1 \le i \le m$,

$$P(a_1 \le X_1 \le b_1, \ldots, a_m \le X_m \le b_m) = \int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)\,dx_1 \cdots dx_m.$$

b) Let $X_1, \ldots, X_m$ be continuous random variables defined on the same sample space. If the partial derivatives of $F_{X_1,\ldots,X_m}$, up to and including those of the mth-order, exist and are continuous except possibly at (portions of) a finite number of $(m-1)$-dimensional planes parallel to the coordinate axes, then $X_1, \ldots, X_m$ have a joint PDF, which we can take to be

$$f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \frac{\partial^m}{\partial x_{k_1} \cdots \partial x_{k_m}} F_{X_1,\ldots,X_m}(x_1, \ldots, x_m)$$

at continuity points of the partials, and $f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = 0$ otherwise, where $k_1, \ldots, k_m$ is any permutation of $1, \ldots, m$.

c) Random variables $X_1, \ldots, X_m$ defined on the same sample space have a joint PDF if and only if there is a nonnegative function f defined on $\mathcal{R}^m$ such that, for all $x_1, \ldots, x_m \in \mathcal{R}$,

$$F_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f(u_1, \ldots, u_m)\,du_1 \cdots du_m.$$

In this case, f is a joint PDF of $X_1, \ldots, X_m$, and

$$f(x_1, \ldots, x_m) = \frac{\partial^m}{\partial x_1 \cdots \partial x_m} F_{X_1,\ldots,X_m}(x_1, \ldots, x_m)$$

at all points $(x_1, \ldots, x_m)$ where this mth-order mixed partial exists.

Advanced Exercises

9.30

a) In the solution to Exercise 9.14, we showed that both X and Y are $\mathcal{U}(0, 1)$ random variables and, hence, they are continuous random variables.

b) In the solution to Exercise 9.14(c), we found that

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ x, & \text{if } 0 \le x < 1 \text{ and } y \ge x; \\ y, & \text{if } 0 \le y < 1 \text{ and } x \ge y; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

It follows that the second-order mixed partials equal 0, wherever they exist. Hence, X and Y couldn't possibly have a joint PDF. Another way to show this fact relies on Proposition 9.6 on page 502 of Section 9.3, the FPF for two random variables with a joint PDF. Applying this result and recalling that $P(Y = X) = 1$, we see that, if X and Y had a joint PDF, then

$$1 = P(Y = X) = \iint\limits_{y=x} f_{X,Y}(x, y)\,dx\,dy = \int_{-\infty}^{\infty} \left(\int_x^x f_{X,Y}(x, y)\,dy \right) dx = 0,$$

which is impossible.

9.31 Suppose that X and Y are independent. Then, by Exercise 9.15(a), for $x, y \in \mathcal{R}$,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) = \left(\int_{-\infty}^{x} f_X(s)\,ds \right) \left(\int_{-\infty}^{y} f_Y(t)\,dt \right) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(s) f_Y(t)\,ds\,dt.$$

Consequently, by Proposition 9.4 on page 499, the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_Y(y)$ is a joint PDF of X and Y.

Conversely, suppose that the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_Y(y)$ is a joint PDF of X and Y. Then, for $x, y \in \mathcal{R}$,

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(s) f_Y(t)\,ds\,dt = \left(\int_{-\infty}^{x} f_X(s)\,ds \right) \left(\int_{-\infty}^{y} f_Y(t)\,dt \right) = F_X(x)F_Y(y).$$

Hence, by Exercise 9.15(b), X and Y are independent.

9.32 Suppose that $X_1, \ldots, X_m$ are independent. Then, by Exercise 9.16, for $x_1, \ldots, x_m \in \mathcal{R}$,

$$\begin{aligned} F_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = F_{X_1}(x_1) \cdots F_{X_m}(x_m) &= \left(\int_{-\infty}^{x_1} f_{X_1}(u_1)\,du_1 \right) \cdots \left(\int_{-\infty}^{x_m} f_{X_m}(u_m)\,du_m \right) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_{X_1}(u_1) \cdots f_{X_m}(u_m)\,du_1 \cdots du_m. \end{aligned}$$

Consequently, by the m-variate version of Proposition 9.4 on page 499, the function f defined on $\mathcal{R}^m$ by $f(x_1, \ldots, x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$ is a joint PDF of $X_1, \ldots, X_m$.

Conversely, suppose that the function f defined on $\mathcal{R}^m$ by $f(x_1, \ldots, x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$ is a joint PDF of $X_1, \ldots, X_m$. Then, for $x_1, \ldots, x_m \in \mathcal{R}$,

$$\begin{aligned} F_{X_1,\ldots,X_m}(x_1, \ldots, x_m) &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_{X_1}(u_1) \cdots f_{X_m}(u_m)\, du_1 \cdots du_m. \\ &= \left(\int_{-\infty}^{x_1} f_{X_1}(u_1)\, du_1 \right) \cdots \left(\int_{-\infty}^{x_m} f_{X_m}(u_m)\, du_m \right) = F_{X_1}(x_1) \cdots F_{X_m}(x_m). \end{aligned}$$

Hence, by Exercise 9.16, $X_1, \ldots, X_m$ are independent.

9.33

a) Event $\{X_{(k)} \le x\}$ occurs if and only if the kth smallest of $X_1, \ldots, X_n$ is at most x, that is, if and only if at least k of the X_js fall in the interval $(-\infty, x]$. Let Y denote the number of X_js that fall in the interval $(-\infty, x]$. Because $X_1, \ldots, X_n$ are independent and have common CDF, F, we see that $Y \sim \mathcal{B}\big(n, F(x)\big)$. Therefore,

$$F_{X_{(k)}}(x) = P(X_{(k)} \le x) = P(Y \ge k) = \sum_{j=k}^{n} \binom{n}{j} \big(F(x)\big)^j \big(1 - F(x)\big)^{n-j}.$$

b) Referring to part (a), we differentiate $F_{X_{(k)}}$ to obtain

$$\begin{aligned} f_{X_{(k)}}(x) &= \sum_{j=k}^{n} \binom{n}{j} \Big(jf(x)\big(F(x)\big)^{j-1}\big(1 - F(x)\big)^{n-j} - (n-j)f(x)\big(F(x)\big)^j\big(1 - F(x)\big)^{n-j-1} \Big) \\ &= \sum_{j=k}^{n} \frac{n!}{(j-1)!\,(n-j)!} f(x)\big(F(x)\big)^{j-1}\big(1 - F(x)\big)^{n-j} \\ &\quad - \sum_{j=k}^{n-1} \frac{n!}{j!\,(n-j-1)!} f(x)\big(F(x)\big)^{j}\big(1 - F(x)\big)^{n-j-1} \\ &= \sum_{j=k}^{n} \frac{n!}{(j-1)!\,(n-j)!} f(x)\big(F(x)\big)^{j-1}\big(1 - F(x)\big)^{n-j} \\ &\quad - \sum_{j=k+1}^{n} \frac{n!}{(j-1)!\,(n-j)!} f(x)\big(F(x)\big)^{j-1}\big(1 - F(x)\big)^{n-j} \\ &= \frac{n!}{(k-1)!\,(n-k)!} f(x)\big(F(x)\big)^{k-1}\big(1 - F(x)\big)^{n-k}. \end{aligned}$$

c) We argue as in Example 9.4(a), except we partition the real line into three intervals using the numbers x and $x + \Delta x$. We classify each of the n observations, $X_1, \ldots, X_n$, by the interval in which it falls. Because the n observations are independent, we can apply the multinomial distribution with parameters n and p_1, p_2, p_3, where

$$p_1 = F(x), \qquad p_2 = F(x + \Delta x) - F(x) \approx f(x)\Delta x, \qquad p_3 = 1 - F(x + \Delta x) \approx 1 - F(x).$$

For event $\{x \le X_{(k)} \le x + \Delta x\}$ to occur, the number of observations that fall in the three intervals must be $k - 1$, 1, and $n - k$, respectively, where we have ignored higher order differentials. Referring now to the PMF of a multinomial distribution, Proposition 6.7 on page 277, we conclude that

$$\begin{aligned}
P(x \le X_{(k)} \le x + \Delta x) &\approx \binom{n}{k-1,\ 1,\ n-k} p_1^{k-1} p_2^1 p_3^{n-k} \\
&\approx \frac{n!}{(k-1)!\,1!\,(n-k)!} \big(F(x)\big)^{k-1} \big(f(x)\Delta x\big)^1 \big(1 - F(x)\big)^{n-k} \\
&= \frac{n!}{(k-1)!\,(n-k)!} f(x) \big(F(x)\big)^{k-1} \big(1 - F(x)\big)^{n-k} \Delta x.
\end{aligned}$$

Therefore, heuristically,

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} f(x) \big(F(x)\big)^{k-1} \big(1 - F(x)\big)^{n-k},$$

which agrees with the result obtained formally in part (b).
d) For a $\mathcal{U}(0, 1)$ distribution,

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from part (b),

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} x^{k-1} (1-x)^{n-k} = \frac{1}{B(k, n-k+1)} x^{k-1} (1-x)^{n-k}$$

if $0 < x < 1$, and $f_{X_{(k)}}(x) = 0$ otherwise. We see that $X_{(k)}$ has the beta distribution with parameters k and $n - k + 1$.
e) For an $\mathcal{E}(\lambda)$ distribution,

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-\lambda x}, & \text{if } x \ge 0. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from part (b),

$$\begin{aligned}
f_{X_{(k)}}(x) &= \frac{n!}{(k-1)!\,(n-k)!} \lambda e^{-\lambda x} \left(1 - e^{-\lambda x}\right)^{k-1} \left(e^{-\lambda x}\right)^{n-k} \\
&= \frac{n!}{(k-1)!\,(n-k)!} \lambda e^{-\lambda(n-k+1)x} \left(1 - e^{-\lambda x}\right)^{k-1}
\end{aligned}$$

if $x > 0$, and $f_{X_{(k)}}(x) = 0$ otherwise.

9.34
a) By symmetry, all $n!$ permutations of $X_1, \ldots, X_n$ are equally likely to be the order statistics. Hence, by Exercise 9.32,

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n!\, f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = n!\, f(x_1) \cdots f(x_n)$$

if $x_1 < \cdots < x_n$, and $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = 0$ otherwise.

b) Let $x_1 < \cdots < x_n$. We argue as in Example 9.4(a), except we partition the real line into $2n+1$ intervals by using the numbers x_k and $x_k + \Delta x_k$, for $k = 1, 2, \ldots, n$. We classify each of the n observations, $X_1, \ldots, X_n$, by the interval in which it falls. Because the n observations are independent, we can apply the multinomial distribution with parameters n and $p_1, p_2, \ldots, p_{2n}, p_{2n+1}$, where

$$p_1 = F(x_1), \qquad p_2 = f(x_1)\Delta x_1, \qquad \ldots, \qquad p_{2n} = f(x_n)\Delta x_n, \qquad p_{2n+1} = 1 - F(x_n).$$

For event $\bigcap_{k=1}^{n}\{x_k \le X_{(k)} \le x_k + \Delta x_k\}$ to occur, the number of observations that fall in the $2n+1$ intervals must be 0, 1, ..., 1, 0, respectively, where we have ignored higher order differentials. Referring now to the PMF of a multinomial distribution, Proposition 6.7 on page 277, we conclude that

$$\begin{aligned}
P\left(\bigcap_{k=1}^{n}\{x_k \le X_{(k)} \le x_k + \Delta x_k\}\right) &\approx \binom{n}{0,\ 1, \ldots, 1,\ 0} p_1^0 p_2^1 \cdots p_{2n}^1 p_{2n+1}^0 \\
&= n!\big(F(x_1)\big)^0\big(f(x_1)\Delta x_1\big)^1 \cdots \big(f(x_n)\Delta x_n\big)^1\big(1 - F(x_n)\big)^0 \\
&= n!\, f(x_1)\Delta x_1 \cdots f(x_n)\Delta x_n \\
&= n!\, f(x_1) \cdots f(x_n)\Delta x_1 \cdots \Delta x_n.
\end{aligned}$$

Therefore, heuristically,

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n!\, f(x_1) \cdots f(x_n)$$

if $x_1 < \cdots < x_n$, and $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = 0$ otherwise, which agrees with the result obtained in part (a).

c) For a $\mathcal{U}(0, 1)$ distribution,

$$f(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from part (b),

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n!\, f(x_1) \cdots f(x_n) = n!$$

if $0 < x_1 < \cdots < x_n < 1$, and $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = 0$ otherwise.

For an $\mathcal{E}(\lambda)$ distribution,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from part (b),

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n!\, f(x_1) \cdots f(x_n) = n!\big(\lambda e^{-\lambda x_1}\big) \cdots \big(\lambda e^{-\lambda x_n}\big) = n!\lambda^n e^{-\lambda(x_1 + \cdots + x_n)}$$

if $0 < x_1 < \cdots < x_n$, and $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = 0$ otherwise.

9.35

a) Let $x < y$. We argue as in Example 9.4(a), where we partition the real line into five intervals using the numbers x, $x + \Delta x$, y, and $y + \Delta y$. We classify each of the n observations, $X_1, \ldots, X_n$, by the interval in which it falls. Because the n observations are independent, we can apply the multinomial distribution with parameters n and $p_1, \ldots, p_5$, where, approximately,

$$p_1 = F(x), \qquad p_2 = f(x)\Delta x, \qquad p_3 = F(y) - F(x), \qquad p_4 = f(y)\Delta y, \qquad p_5 = 1 - F(y).$$

For event $\{x \le X_{(j)} \le x + \Delta x,\ y \le X_{(k)} \le y + \Delta y\}$ to occur, the number of observations that fall in the five intervals must be $j-1$, 1, $k-j-1$, 1, and $n-k$, respectively, where we have ignored higher order differentials. Referring now to the PMF of a multinomial distribution, Proposition 6.7 on page 277, we conclude that

$$\begin{aligned}
&P(x \le X_{(j)} \le x + \Delta x,\ y \le X_{(k)} \le y + \Delta y) \\
&\quad \approx \binom{n}{j-1,\ 1,\ k-j-1,\ 1,\ n-k} p_1^{j-1} p_2^1 p_3^{k-j-1} p_4^1 p_5^{n-k} \\
&\quad = \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} f(x) f(y) \big(F(x)\big)^{j-1} \big(F(y) - F(x)\big)^{k-j-1} \big(1 - F(y)\big)^{n-k} \Delta x \Delta y.
\end{aligned}$$

Therefore, heuristically,

$$\begin{aligned}
f_{X_{(j)},X_{(k)}}(x, y) = &\frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} f(x) f(y) \\
&\times \big(F(x)\big)^{j-1} \big(F(y) - F(x)\big)^{k-j-1} \big(1 - F(y)\big)^{n-k}
\end{aligned}$$

if $x < y$, and $f_{X_{(j)},X_{(k)}}(x, y) = 0$ otherwise.

b) Let $x < y < z$. We argue as in Example 9.4(a), except we partition the real line into seven intervals using the numbers $x, x + \Delta x, y, y + \Delta y, z$, and $z + \Delta z$. We classify each of the n observations, $X_1, \ldots, X_n$, by the interval in which it falls. Because the n observations are independent, we can apply the multinomial distribution with parameters n and $p_1, \ldots, p_7$, where, approximately,

$$p_1 = F(x), \qquad p_2 = f(x)\Delta x, \qquad p_3 = F(y) - F(x), \qquad p_4 = f(y)\Delta y,$$
$$p_5 = F(z) - F(y), \qquad p_6 = f(z)\Delta z, \qquad p_7 = 1 - F(z).$$

For event $\{x \le X_{(i)} \le x + \Delta x,\ y \le X_{(j)} \le y + \Delta y,\ z \le X_{(k)} \le z + \Delta z\}$ to occur, the number of observations that fall in the seven intervals must be $i-1, 1, j-i-1, 1, k-j-1, 1$, and $n-k$, respectively, where we have ignored higher order differentials. Referring now to the PMF of a multinomial distribution, Proposition 6.7 on page 277, we conclude that

$$\begin{aligned}
&P(x \le X_{(i)} \le x + \Delta x,\ y \le X_{(j)} \le y + \Delta y,\ z \le X_{(k)} \le z + \Delta z) \\
&\quad \approx \binom{n}{i-1,\ 1,\ j-i-1,\ 1,\ k-j-1,\ 1,\ n-k} p_1^{i-1} p_2^1 p_3^{j-i-1} p_4^1 p_5^{k-j-1} p_6^1 p_7^{n-k} \\
&\quad = \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(n-k)!} f(x) f(y) f(z) \\
&\qquad \times \big(F(x)\big)^{i-1} \big(F(y) - F(x)\big)^{j-i-1} \big(F(z) - F(y)\big)^{k-j-1} \big(1 - F(z)\big)^{n-k} \Delta x \Delta y \Delta z.
\end{aligned}$$

Therefore, heuristically,

$$\begin{aligned}
f_{X_{(i)},X_{(j)},X_{(k)}}(x, y, z) = &\frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(n-k)!} f(x) f(y) f(z) \\
&\times \big(F(x)\big)^{i-1} \big(F(y) - F(x)\big)^{j-i-1} \big(F(z) - F(y)\big)^{k-j-1} \big(1 - F(z)\big)^{n-k}
\end{aligned}$$

if $x < y < z$, and $f_{X_{(i)},X_{(j)},X_{(k)}}(x, y, z) = 0$ otherwise.

c) For a $\mathcal{U}(0, 1)$ distribution,

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from part (a),

$$f_{X_{(j)},X_{(k)}}(x, y) = \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} x^{j-1}(y-x)^{k-j-1}(1-y)^{n-k}$$

if $0 < x < y < 1$, and $f_{X_{(j)},X_{(k)}}(x, y) = 0$ otherwise. And from part (b),

$$\begin{aligned} f_{X_{(i)},X_{(j)},X_{(k)}}(x, y, z) &= \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(n-k)!} \\ &\quad \times x^{i-1}(y-x)^{j-i-1}(z-y)^{k-j-1}(1-z)^{n-k} \end{aligned}$$

if $0 < x < y < z < 1$, and $f_{X_{(i)},X_{(j)},X_{(k)}}(x, y, z) = 0$ otherwise.

For an $\mathcal{E}(\lambda)$ distribution,

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-\lambda x}, & \text{if } x \ge 0. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, from part (a),

$$\begin{aligned} f_{X_{(j)},X_{(k)}}(x, y) &= \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} \left(\lambda e^{-\lambda x}\right)\left(\lambda e^{-\lambda y}\right) \\ &\quad \times \left(1 - e^{-\lambda x}\right)^{j-1}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{k-j-1}\left(e^{-\lambda y}\right)^{n-k} \\ &= \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} \lambda^2 e^{-\lambda(x+(n-k+1)y)} \\ &\quad \times \left(1 - e^{-\lambda x}\right)^{j-1}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{k-j-1} \end{aligned}$$

if $0 < x < y$, and $f_{X_{(j)},X_{(k)}}(x, y) = 0$ otherwise. And from part (b),

$$\begin{aligned} f_{X_{(i)},X_{(j)},X_{(k)}}(x, y, z) &= \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(n-k)!} \left(\lambda e^{-\lambda x}\right)\left(\lambda e^{-\lambda y}\right)\left(\lambda e^{-\lambda z}\right) \\ &\quad \times \left(1 - e^{-\lambda x}\right)^{i-1}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{j-i-1}\left(e^{-\lambda y} - e^{-\lambda z}\right)^{k-j-1}\left(e^{-\lambda z}\right)^{n-k} \\ &= \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(n-k)!} \lambda^3 e^{-\lambda(x+y+(n-k+1)z)} \\ &\quad \times \left(1 - e^{-\lambda x}\right)^{i-1}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{j-i-1}\left(e^{-\lambda y} - e^{-\lambda z}\right)^{k-j-1} \end{aligned}$$

if $0 < x < y < z$, and $f_{X_{(i)},X_{(j)},X_{(k)}}(x, y, z) = 0$ otherwise.

9.3 Properties of Joint Probability Density Functions

Basic Exercises

9.36 By the FPF, we have, for each $x \in \mathcal{R}$,

$$P(X = x) = P\big((X, Y) \in \{x\} \times \mathcal{R}\big) = \iint_{(s,t)\in\{x\}\times\mathcal{R}} f_{X,Y}(s, t)\, ds\, dt$$

$$= \int_{-\infty}^{\infty} \left(\int_x^x f_{X,Y}(s, t)\, ds \right) dt = \int_{-\infty}^{\infty} 0\, dt = 0.$$

Thus, X is a continuous random variable. Similarly, we can show that Y is a continuous random variable.

9.37 Clearly, property (a) is satisfied. For property (b),

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)\, dx\, dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(y)\, dy \right) f(x)\, dx$$

$$= \int_{-\infty}^{\infty} 1 \cdot f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

9.38

a) We note that $g(x, y) \ge 0$ for all $(x, y) \in \mathcal{R}^2$, and that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, dx\, dy = \int_0^{\infty} \left(\int_0^x e^{-\lambda x}\, dy \right) dx = \int_0^{\infty} x e^{-\lambda x}\, dx = \frac{1}{\lambda^2}.$$

Therefore, the corresponding joint PDF is

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda x}, & \text{if } 0 < y < x; \\ 0, & \text{otherwise.} \end{cases}$$

b) We note that $g(x, y) \ge 0$ for all $(x, y) \in \mathcal{R}^2$. For the double integral over $\mathcal{R}^2$ to exist, we must have $\alpha > -1$, in which case

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, dx\, dy = \int_0^1 \left(\int_0^y (y - x)^{\alpha}\, dx \right) dy = \int_0^1 \frac{y^{\alpha+1}}{\alpha + 1}\, dy = \frac{1}{(\alpha + 1)(\alpha + 2)}.$$

Therefore, the corresponding joint PDF is

$$f(x, y) = \begin{cases} (\alpha + 1)(\alpha + 2)(y - x)^{\alpha}, & \text{if } 0 < x < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

c) As $g(x, y) = x - y < 0$ for $0 < x < y < 1$, constructing a joint PDF from g is not possible.

d) We note that $g(x, y) \ge 0$ for all $(x, y) \in \mathcal{R}^2$, and that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, dx\, dy = \int_0^1 \left(\int_0^1 (x + y)\, dx \right) dy = \int_0^1 (1/2 + y)\, dy = 1.$$

Therefore, g itself is a joint PDF.

e) We note that $g(x, y) \ge 0$ for all $(x, y) \in \mathcal{R}^2$, but that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, dx\, dy = \int_0^{\infty} \left(\int_0^{\infty} (x + y)\, dx \right) dy = \infty.$$

Therefore, constructing a joint PDF from g is not possible.

9.39 Note that the region in which the PDF is nonzero is the interior of the triangle with vertices (0, 0), (50, 0), and (0, 50). We want $P(X > 20,\ Y > 20)$, which, from the FPF, equals

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = \int_{20}^{30}\left(\int_{20}^{50-x} c(50 - x - y)\,dy\right)dx,$$

which is the expression in (b). It is easy to see that none of the other expressions provide the required probability.

9.40

a) Because the point is selected at random, we can take a joint PDF to be a positive constant on S—that is, $f_{X,Y}(x, y) = c$ for $(x, y) \in S$ and $f_{X,Y}(x, y) = 0$ otherwise. Because a joint PDF must integrate to 1, we have

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = \iint\limits_{S} c\,dx\,dy = c|S|,$$

where $|S|$ denotes the area of S. Hence $c = 1/|S|$ and, consequently, $f_{X,Y}(x, y) = 1/|S|$ for $(x, y) \in S$, and $f_{X,Y}(x, y) = 0$ otherwise.

b) From the FPF and part (a),

$$P\big((X, Y) \in A\big) = \iint\limits_{A} f_{X,Y}(x, y)\,dx\,dy = \iint\limits_{A\cap S} \frac{1}{|S|}\,dx\,dy = \frac{|A \cap S|}{|S|}.$$

9.41 Let S denote the unit square and note that $|S| = 1$.

a) From Exercise 9.40(a),

$$f_{X,Y}(x, y) = \begin{cases} 1/|S|, & \text{if } (x, y) \in S; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 1, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) In this special case of bivariate uniform random variables, using the bivariate uniform model from Exercise 9.40 is a far easier way to obtain a joint PDF than by the CDF method used in Examples 9.1 and 9.3.

c) Using the joint PDF of X and Y and applying the FPF, we get

$$\begin{aligned} P(|X - Y| \le 1/4) &= \iint\limits_{|x-y|\le 1/4} f_{X,Y}(x, y)\,dx\,dy \\ &= \int_0^{1/4}\left(\int_0^{x+1/4} 1\,dy\right)dx + \int_{1/4}^{3/4}\left(\int_{x-1/4}^{x+1/4} 1\,dy\right)dx + \int_{3/4}^{1}\left(\int_{x-1/4}^{1} dy\right)dx \\ &= 3/32 + 1/4 + 3/32 = 7/16. \end{aligned}$$

Alternatively, let $A = \{\,(x, y) : |x - y| \le 1/4\,\}$ and set $B = S \setminus A$. Using Exercise 9.40(b), we get

$$P(|X - Y| \le 1/4) = P\big((X, Y) \in A\big) = \frac{|A \cap S|}{|S|} = 1 - |B| = 1 - 2\cdot\frac{1}{2}(3/4)^2 = 7/16.$$

9.42 Let S denote the upper half of the unit disk and note that $|S| = \pi/2$.

a) From Exercise 9.40(a),

$$f_{X,Y}(x, y) = \begin{cases} 1/|S|, & \text{if } (x, y) \in S; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 2/\pi, & \text{if } -1 < x < 1 \text{ and } 0 < y < \sqrt{1 - x^2}; \\ 0, & \text{otherwise.} \end{cases}$$

b) In this special case of bivariate uniform random variables, using the bivariate uniform model from Exercise 9.40 is a far easier way to obtain a joint PDF than by the CDF method used in Exercises 9.4 and 9.23.
c) Let T denote the interior of the specified triangle. Using the joint PDF of X and Y and applying the FPF, we get

$$P\big((X,Y)\in T\big) = \iint_T f_{X,Y}(x,y)\,dx\,dy = \int_0^1 \left(\int_{y-1}^{1-y} (2/\pi)\,dx\right) dy = \frac{2}{\pi}\cdot 1 = 0.637.$$

Alternatively, using Exercise 9.40(b), we get

$$P\big((X,Y)\in T\big) = \frac{|T\cap S|}{|S|} = \frac{|T|}{|S|} = \frac{1}{\pi/2} = 0.637.$$

d) Let $A = \{\,(x,y) : |x| \ge 1/2\,\}$. Using the joint PDF of X and Y and the FPF, we get

$$\begin{aligned} P\big((X,Y)\in A^c\big) &= \iint_{A^c} f_{X,Y}(x,y)\,dx\,dy = \iint_{A^c\cap S} (2/\pi)\,dx\,dy \\ &= \frac{4}{\pi}\int_0^{1/2}\left(\int_0^{\sqrt{1-x^2}} 1\,dy\right) dx = 0.609. \end{aligned}$$

Hence, by the complementation rule, $P\big((X,Y)\in A\big) = 1 - 0.609 = 0.391.$

9.43
a) Because the point is selected at random, we can take a joint PDF to be a positive constant on S—that is, $f_{X_1,\dots,X_m}(x_1,\dots,x_m) = c$ for $(x_1,\dots,x_m)\in S$ and $f_{X_1,\dots,X_m}(x_1,\dots,x_m) = 0$ otherwise. Because a joint PDF must integrate to 1, we have

$$1 = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,\dots,X_m}(x_1,\dots,x_m)\,dx_1\cdots dx_m = \int\cdots\int_S c\,dx_1\cdots dx_m = c|S|,$$

where $|S|$ denotes the m-dimensional volume of S. Hence $c = 1/|S|$ and, consequently, we have $f_{X_1,\dots,X_m}(x_1,\dots,x_m) = 1/|S|$ for $(x_1,\dots,x_m)\in S$, and $f_{X_1,\dots,X_m}(x_1,\dots,x_m) = 0$ otherwise.
b) From the FPF and part (a),

$$\begin{aligned} P\big((X_1,\dots,X_m)\in A\big) &= \int\cdots\int_A f_{X_1,\dots,X_m}(x_1,\dots,x_m)\,dx_1\cdots dx_m \\ &= \int\cdots\int_{A\cap S} \frac{1}{|S|}\,dx_1\cdots dx_m = \frac{|A\cap S|}{|S|}. \end{aligned}$$

9.44 Let S denote the unit cube and note that $|S| = 1$.
a) From Exercise 9.43(a),

$$f_{X_1,\dots,X_m}(x_1,\dots,x_m) = \begin{cases} 1/|S|, & \text{if } (x_1,\dots,x_m)\in S; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 1, & \text{if } 0 < x, y, z < 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) From the FPF and part (a),

$$P(Z = \max\{X, Y, Z\}) = P(Z > X,\ Z > Y) = \iiint\limits_{z>x,\, z>y} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz$$
$$= \int_0^1 \left(\int_0^z \left(\int_0^z 1\,dx\right) dy\right) dz = \int_0^1 \left(\int_0^z z\,dy\right) dz = \int_0^1 z^2\,dz = \frac{1}{3}.$$

c) By symmetry, each of X, Y, and Z are equally likely to be the largest of the three. Thus,

$$1 = P(X = \max\{X, Y, Z\}) + P(Y = \max\{X, Y, Z\}) + P(Z = \max\{X, Y, Z\})$$
$$= 3P(Z = \max\{X, Y, Z\}).$$

Thus, $P(Z = \max\{X, Y, Z\}) = 1/3$.
d) Let A denote the sphere of radius 1/4 centered at $(1/2, 1/2, 1/2)$. Then, by Exercise 9.43(b),

$$P\big((X, Y, Z) \in A\big) = \frac{|A \cap S|}{|S|} = |A| = \frac{4}{3}\pi \left(\frac{1}{4}\right)^3 = 0.0654.$$

e) From the FPF and part (a),

$$P(X + Y > Z) = \iiint\limits_{x+y>z} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz = \int_0^1 \left(\int_0^1 \left(\int_0^{\min\{x+y,1\}} 1\,dz\right) dy\right) dx$$
$$= \int_0^1 \left(\int_0^{1-x} \left(\int_0^{x+y} 1\,dz\right) dy\right) dx + \int_0^1 \left(\int_{1-x}^1 \left(\int_0^1 1\,dz\right) dy\right) dx$$
$$= \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

9.45
a) Here we want $P(X > Y)$. From the FPF,

$$P(X > Y) = \iint\limits_{x>y} f_{X,Y}(x, y)\,dx\,dy = \int_0^{1/2} \left(\int_y^{1-y} 6(1 - x - y)\,dx\right) dy = \frac{1}{2}.$$

Alternatively, we can use the fact that the joint PDF of X and Y is symmetric in x and y to conclude that

$$1 = P(X = Y) + P(X > Y) + P(X < Y) = 0 + P(X > Y) + P(X < Y) = 2P(X > Y),$$

so that $P(X > Y) = 1/2$.
b) Here we want $P(X < 0.2)$. From the FPF

$$P(X < 0.2) = \iint\limits_{x<0.2} f_{X,Y}(x, y)\,dx\,dy = \int_0^{0.2} \left(\int_0^{1-x} 6(1 - x - y)\,dy\right) dx = 0.488.$$

9.46 Let T denote the electrical unit's lifetime.
a) In a parallel system, $T = \max\{X, Y\}$. Then, for $t \ge 0$,

$$\begin{aligned} F_T(t) &= P(T \le t) = P(\max\{X, Y\} \le t) = P(X \le t,\ Y \le t) \\ &= \iint\limits_{x \le t,\ y \le t} f_{X,Y}(x, y)\,dx\,dy = \int_0^t \int_0^t \lambda\mu e^{-(\lambda x + \mu y)}\,dx\,dy \\ &= (1 - e^{-\lambda t})(1 - e^{-\mu t}) = 1 - e^{-\lambda t} - e^{-\mu t} + e^{-(\lambda+\mu)t}. \end{aligned}$$

Taking the derivative yields

$$f_T(t) = \lambda e^{-\lambda t} + \mu e^{-\mu t} - (\lambda + \mu)e^{-(\lambda+\mu)t}, \qquad t > 0.$$

b) In a series system, $T = \min\{X, Y\}$. Then, for $t \ge 0$,

$$F_T(t) = P(T \le t) = 1 - P(T > t) = 1 - P(\min\{X, Y\} > t) = 1 - P(X > t,\ Y > t).$$

Referring now to the result of Example 9.5(a) on page 503, we see that $F_T(t) = 1 - e^{-(\lambda+\mu)t}$ for $t \ge 0$. Differentiating gives

$$f_T(t) = (\lambda + \mu)e^{-(\lambda+\mu)t}, \qquad t > 0.$$

Hence, $T \sim \mathcal{E}(\lambda + \mu)$.
c) The event exactly one of the components is functioning at time t is $\{X > t,\ Y \le t\} \cup \{X \le t,\ Y > t\}$, where we note that these two events are mutually exclusive. Now, by the FPF,

$$\begin{aligned} P(X > t,\ Y \le t) &= \iint\limits_{x > t,\ y \le t} f_{X,Y}(x, y)\,dx\,dy = \int_0^t \left(\int_t^\infty \lambda\mu e^{-(\lambda x + \mu y)}\,dx \right) dy \\ &= \int_0^t \left(\int_t^\infty \lambda e^{-\lambda x}\,dx \right) \mu e^{-\mu y}\,dy = e^{-\lambda t}\left(1 - e^{-\mu t}\right) = e^{-\lambda t} - e^{-(\lambda+\mu)t}. \end{aligned}$$

Likewise, we find that $P(X \le t,\ Y > t) = e^{-\mu t} - e^{-(\lambda+\mu)t}$. Thus, the probability that exactly one of the components is functioning at time t is

$$\left(e^{-\lambda t} - e^{-(\lambda+\mu)t}\right) + \left(e^{-\mu t} - e^{-(\lambda+\mu)t}\right) = e^{-\lambda t} + e^{-\mu t} - 2e^{-(\lambda+\mu)t}.$$

9.47 From the FPF,

$$P(Y = X) = \iint\limits_{y=x} f_{X,Y}(x, y)\,dx\,dy = \int_{-\infty}^\infty \left(\int_x^x f_{X,Y}(x, y)\,dy \right) dx = \int_{-\infty}^\infty 0\,dx = 0.$$

9.48 Recall that the event that component B is the first to fail is given by $\{Y < X\}$.
a) By the FPF,

$$\begin{aligned} P(Y < X) &= \iint\limits_{y < x} f_{X,Y}(x, y)\,dx\,dy = \int_0^\infty \left(\int_y^\infty \lambda\mu e^{-(\lambda x + \mu y)}\,dx \right) dy \\ &= \lambda\mu \int_0^\infty e^{-\mu y} \left(\int_y^\infty e^{-\lambda x}\,dx \right) dy = \frac{\mu}{\lambda + \mu}. \end{aligned}$$

b) We have

$$\begin{aligned} P(Y < X) &= 1 - P(Y \ge X) = 1 - \big(P(X < Y) + P(X = Y)\big) \\ &= 1 - P(X < Y) - 0 = 1 - \frac{\lambda}{\lambda + \mu} = \frac{\mu}{\lambda + \mu}. \end{aligned}$$

c) Let U and V denote the lifetimes of components B and A, respectively. Then U and V have joint PDF

$$f_{U,V}(u, v) = \mu\lambda\, e^{-(\mu u+\lambda v)}, \qquad u, v > 0,$$

and $f_{U,V}(u, v) = 0$ otherwise. Applying the result of Example 9.5(b) with the roles of λ and μ interchanged, we deduce that

$$P(U < V) = \frac{\mu}{\mu + \lambda}.$$

In other words, the probability is $\mu/(\lambda + \mu)$ that component B is the first to fail.

9.49 For an $\mathcal{E}(\lambda)$ distribution,

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

a) Referring now to Equation (9.20) yields

$$f_{X,Y}(x, y) = n(n-1)\big(\lambda e^{-\lambda x}\big)\big(\lambda e^{-\lambda y}\big)\left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2} = n(n-1)\lambda^2 e^{-\lambda(x+y)}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2}$$

if $0 < x < y$, and $f_{X,Y}(x, y) = 0$ otherwise.

b) Applying the FPF, the result from part (a), and making the substitution $u = e^{-\lambda x} - e^{-\lambda y}$, we get

$$\begin{aligned} P(Y > X + 1) &= \iint\limits_{y>x+1} f_{X,Y}(x, y)\, dx\, dy \\ &= \int_0^\infty \left(\int_{x+1}^\infty n(n-1)\lambda^2 e^{-\lambda(x+y)} \left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2} dy \right) dx \\ &= n(n-1)\lambda \int_0^\infty \left(\int_{e^{-\lambda x}(1-e^{-\lambda})}^{e^{-\lambda x}} u^{n-2}\, du \right) e^{-\lambda x}\, dx \\ &= n\lambda \left(1 - \left(1 - e^{-\lambda}\right)^{n-1}\right) \int_0^\infty e^{-n\lambda x}\, dx = 1 - \left(1 - e^{-\lambda}\right)^{n-1}. \end{aligned}$$

c) Referring now to Equation (9.26) yields,

$$\begin{aligned} f_R(r) &= n(n-1) \int_{-\infty}^\infty \left(\lambda e^{-\lambda x}\right)\left(\lambda e^{-\lambda(r+x)}\right)\left(e^{-\lambda x} - e^{-\lambda(r+x)}\right)^{n-2} dx \\ &= n(n-1)\lambda^2 e^{-\lambda r}\left(1 - e^{-\lambda r}\right)^{n-2} \int_0^\infty e^{-n\lambda x}\, dx = (n-1)\lambda e^{-\lambda r}\left(1 - e^{-\lambda r}\right)^{n-2} \end{aligned}$$

if $r > 0$, and $f_R(r) = 0$ otherwise.

d) Using the (univariate) FPF, referring to part (c), and making the substitution $u = 1 - e^{-\lambda r}$ yields

$$\begin{aligned} P(Y > X + 1) = P(R > 1) &= \int_{r>1} f_R(r)\, dr = \int_1^\infty (n-1)\lambda e^{-\lambda r}\left(1 - e^{-\lambda r}\right)^{n-2} \\ &= (n-1)\int_{1-e^{-\lambda}}^1 u^{n-2}\, du = 1 - \left(1 - e^{-\lambda}\right)^{n-1}. \end{aligned}$$

Theory Exercises

9.50 The property in Propostion 9.5(a) is part of the definition of a joint PDF, as given in Definition 9.2 on page 495. We still need to verify the property in Proposition 9.5(b). Intuitively, it reflects the fact that the probability that X and Y take on some real numbers equals 1. To prove it mathematically, we let $A_n = \{-n \le X \le n,\ -n \le Y \le n\}$ for each $n \in \mathcal{N}$. Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \Omega$. Applying the definition of a joint PDF, the continuity property of a probability measure, and the certainty property of a probability measure gives

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy &= \lim_{n\to\infty}\int_{-n}^{n}\int_{-n}^{n} f_{X,Y}(x,y)\,dx\,dy \\ &= \lim_{n\to\infty} P(-n \le X \le n,\ -n \le Y \le n) \\ &= \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\Omega) = 1.\end{aligned}$$

9.51 A joint PDF, $f_{X_1,\ldots,X_m}$, of m continuous random variables, $X_1, \ldots, X_m$, satisfies the following two properties:

a) $f_{X_1,\ldots,X_m}(x_1,\ldots,x_m) \ge 0$ for all $(x_1,\ldots,x_m) \in \mathcal{R}^m$

b) $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,\ldots,X_m}(x_1,\ldots,x_m)\,dx_1\cdots dx_m = 1$

9.52 Suppose that $X_1, \ldots, X_m$ are continuous random variables with a joint PDF. Then, for any subset, $A \subset \mathcal{R}^m$,

$$P\big((X_1,\ldots,X_m) \in A\big) = \int\cdots\int_A f_{X_1,\ldots,X_m}(x_1,\ldots,x_m)\,dx_1\cdots dx_m.$$

In words, the probability that m continuous random variables take a value in a specified subset of $\mathcal{R}^m$ can be obtained by integrating a joint PDF of the random variables over that subset.

9.53

a) We use the FPF, make the substitution $u = x + y$, and then interchange the order of integration to get

$$\begin{aligned}F_{X+Y}(z) = P(X+Y \le z) &= \iint_{x+y\le z} f_{X,Y}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,dy\right)dx \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f_{X,Y}(x,u-x)\,du\right)dx = \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f_{X,Y}(x,u-x)\,dx\right)du.\end{aligned}$$

Therefore, from Proposition 8.5 on page 422,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx.$$

b) We use the FPF, make the substitution $u = x - y$, and then interchange the order of integration to get

$$\begin{aligned}F_{X-Y}(z) = P(X-Y \le z) &= \iint_{x-y\le z} f_{X,Y}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\left(\int_{x-z}^{\infty} f_{X,Y}(x,y)\,dy\right)dx \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f_{X,Y}(x,x-u)\,du\right)dx = \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f_{X,Y}(x,x-u)\,dx\right)du.\end{aligned}$$

Therefore, from Proposition 8.5 on page 422,

$$f_{X-Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, x-z)\,dx.$$

Advanced Exercises

9.54 No! The random variables X and Y in Exercise 9.30 are continuous random variables but, clearly, $P(Y = X) = 1$.

9.55 No! The random variables X and Y in Exercise 9.30 are continuous random variables but cannot have a joint PDF because $P(Y = X) = 1 \neq 0$.

9.56

a) Let X_1, X_2, and X_3 denote the positions of the three gas stations. By assumption, these three random variables are a random sample of size 3 from a $\mathcal{U}(0, 1)$ distribution. Therefore, from Exercise 9.34(c), a joint PDF of $X_{(1)}$, $X_{(2)}$, and $X_{(3)}$ is $f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3) = 3! = 6$ if $0 < x_1 < x_2 < x_3 < 1$, and $f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3) = 0$ otherwise. The event that no two of the gas stations are less than 1/3 mile apart can be expressed as $\{X_{(2)} \geq X_{(1)} + 1/3,\ X_{(3)} \geq X_{(2)} + 1/3\}$. From the FPF,

$$
\begin{aligned}
P(X_{(2)} \geq X_{(1)} + 1/3,\ X_{(3)} \geq X_{(2)} + 1/3) &= \iiint\limits_{\substack{x_2 \geq x_1 + 1/3 \\ x_3 \geq x_2 + 1/3}} f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 \\
&= 6 \int_0^{1/3} \left(\int_{x_1+1/3}^{2/3} \left(\int_{x_2+1/3}^{1} dx_3 \right) dx_2 \right) dx_1 \\
&= 6 \cdot \frac{1}{162} = 0.0370.
\end{aligned}
$$

b) We must have $2d < 1$, or $d < 1/2$.

c) We evaluate the following integrals by first making the substitution $u = 1 - d - x_2$ and then the substitution $v = 1 - 2d - x_1$, thus:

$$
\begin{aligned}
&P(X_{(2)} \geq X_{(1)} + d,\ X_{(3)} \geq X_{(2)} + d) \\
&\quad = \iiint\limits_{\substack{x_2 \geq x_1 + d \\ x_3 \geq x_2 + d}} f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 = 6 \int_0^{1-2d} \left(\int_{x_1+d}^{1-d} \left(\int_{x_2+d}^{1} dx_3 \right) dx_2 \right) dx_1 \\
&\quad = 6 \int_0^{1-2d} \left(\int_{x_1+d}^{1-d} (1 - d - x_2)\, dx_2 \right) dx_1 = 6 \int_0^{1-2d} \left(\int_0^{1-2d-x_1} u\, du \right) dx_1 \\
&\quad = 3 \int_0^{1-2d} (1 - 2d - x_1)^2\, dx_1 = 3 \int_0^{1-2d} v^2\, dv \\
&\quad = (1 - 2d)^3.
\end{aligned}
$$

9.57

a) We must have $(n - 1)d < 1$, or $d < 1/(n - 1)$.

b) Let $X_1, \ldots, X_n$ denote the positions of the n gas stations, which, by assumption are a random sample of size n from a $\mathcal{U}(0, 1)$ distribution. Therefore, from Exercise 9.34(c), a joint PDF of $X_{(1)}, \ldots, X_{(n)}$ is $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n!$ if $0 < x_1 < \cdots < x_n < 1$, and $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = 0$ otherwise. The event that no two of the gas stations are less than d mile apart is $\bigcap_{k=1}^{n-1} \{X_{(k+1)} \geq X_{(k)} + d\}$. From the FPF, we have

$$
\begin{aligned}
P\Bigg(\bigcap_{k=1}^{n-1}\{X_{(k+1)} \geq X_{(k)} + d\}\Bigg)
&= \underset{x_{k+1}\geq x_k+d,\, k=1,\ldots,n-1}{\int\cdots\int} f_{X_{(1)},\ldots,X_{(n)}}(x_1,\ldots,x_n)\,dx_1\cdots dx_n \\
&= n!\int_0^{1-(n-1)d}\left(\int_{x_1+d}^{1-(n-2)d}\left(\cdots\int_{x_{n-1}+d}^{1} dx_n\right)\cdots dx_2\right)dx_1. \\
&= n!\int_0^{1-(n-1)d}\left(\int_{x_1+d}^{1-(n-2)d}\left(\cdots\int_{x_{n-2}+d}^{1-d}(1-d-x_{n-1})\,dx_{n-1}\right)\cdots dx_2\right)dx_1.
\end{aligned}
$$

Now we proceed as in Exercise 9.56(c). We evaluate the integrals in the last expression of the previous display by making the successive substitutions $u_{n-k} = 1 - kd - x_{n-k}$ for $k = 1, 2, \ldots, n-1$. The result is $\big(1-(n-1)d\big)^n$.

9.58

a) That $h_{X,Y}$ is a nonnegative function on $\mathcal{R}^2$ is part of the definition of a joint PDF/PMF.

b) For each $n \in \mathcal{N}$, let $A_n = \{-n \leq X \leq n,\ Y = y\}$. Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \{Y = y\}$. Hence, by the continuity property of a probability measure,

$$
\begin{aligned}
\int_{-\infty}^{\infty} h_{X,Y}(x,y)\,dx = \lim_{n\to\infty}\int_{-n}^{n} h_{X,Y}(x,y)\,dx &= \lim_{n\to\infty} P(-n \leq X \leq n,\ Y = y) \\
&= \lim_{n\to\infty} P(A_n) = P\Bigg(\bigcup_{n=1}^{\infty} A_n\Bigg) = P(Y = y).
\end{aligned}
$$

The required result now follows from the fact that $\sum_y P(Y = y) = 1$.

9.59

a) Clearly, $h_{X,Y}$ is a nonnegative function on $\mathcal{R}^2$. Also,

$$
\begin{aligned}
\int_{-\infty}^{\infty}\sum_y h_{X,Y}(x,y)\,dx &= \int_0^{\infty}\Bigg(\sum_{y=0}^{\infty}\frac{x^y}{y!}\Bigg)\lambda e^{-(1+\lambda)x}\,dx \\
&= \int_0^{\infty} e^x \lambda e^{-(1+\lambda)x}\,dx = \int_0^{\infty}\lambda e^{-\lambda x}\,dx = 1.
\end{aligned}
$$

b) For $y = 0, 1, \ldots,$

$$
\begin{aligned}
p_Y(y) = P(Y = y) &= P(-\infty < X < \infty,\ Y = y) = \int_{-\infty}^{\infty} h_{X,Y}(x,y)\,dx \\
&= \int_0^{\infty}\lambda\, e^{-(1+\lambda)x}\frac{x^y}{y!}\,dx = \frac{\lambda}{y!}\int_0^{\infty} x^{(y+1)-1}e^{-(1+\lambda)x}\,dx = \frac{\lambda}{y!}\,\frac{\Gamma(y+1)}{(1+\lambda)^{y+1}} \\
&= \frac{\lambda}{(1+\lambda)^{y+1}} = \frac{(1/\lambda)^y}{(1+1/\lambda)^{y+1}}.
\end{aligned}
$$

Thus Y has the Pascal distribution (as defined in Exercise 7.26) with parameter $1/\lambda$.

c) For $x > 0$, we have, in view of the law of partitions,

$$F_X(x) = P(X \le x) = \sum_{y=0}^{\infty} P(X \le x,\ Y = y) = \sum_{y=0}^{\infty} \left(\int_{-\infty}^{x} h_{X,Y}(s, y)\, ds \right)$$
$$= \sum_{y=0}^{\infty} \left(\int_{0}^{x} \lambda\, e^{-(1+\lambda)s} \frac{s^y}{y!}\, ds \right) = \int_0^x \left(\sum_{y=0}^{\infty} \frac{s^y}{y!} \right) \lambda\, e^{-(1+\lambda)s}\, ds$$
$$= \int_0^x e^s \lambda\, e^{-(1+\lambda)s}\, ds = \int_0^x \lambda e^{-\lambda s}\, ds.$$

Thus, by Proposition 8.5 on page 422, we have $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$ and $f_X(x) = 0$ otherwise. Hence, $X \sim \mathcal{E}(\lambda)$.

9.60 Clearly, h is a nonnegative function on $\mathcal{R}^2$. Also,

$$\int_{-\infty}^{\infty} \sum_y h(x, y)\, dx = \int_{-\infty}^{\infty} \sum_y f(x) p(y)\, dx = \int_{-\infty}^{\infty} \left(\sum_y p(y) \right) f(x)\, dx$$
$$= \int_{-\infty}^{\infty} 1 \cdot f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

9.61 Clearly, h is a nonnegative function on $\mathcal{R}^2$. Also,

$$\int_{-\infty}^{\infty} \sum_y h(x, y)\, dx = \int_{-\infty}^{\infty} \sum_y f(x) p_x(y)\, dx = \int_{-\infty}^{\infty} \left(\sum_y p_x(y) \right) f(x)\, dx$$
$$= \int_{-\infty}^{\infty} 1 \cdot f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

9.62 Clearly, h is a nonnegative function on $\mathcal{R}^2$. Also,

$$\int_{-\infty}^{\infty} \sum_y h(x, y)\, dx = \int_{-\infty}^{\infty} \sum_y f_y(x) p(y)\, dx = \sum_y \left(\int_{-\infty}^{\infty} f_y(x)\, dx \right) p(y)$$
$$= \sum_y 1 \cdot p(y) = \sum_y p(y) = 1.$$

9.4 Marginal and Conditional Probability Density Functions

Basic Exercises

9.63 From Example 9.7, we have

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{if } 0 < x, y < 1; \\ 0, & \text{otherwise.} \end{cases} \quad f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases} \quad f_Y(y) = \begin{cases} 1, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

a) Let $0 < x < 1$. Then, for $0 < y < 1$,

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

Hence, for $0 < x < 1$,

$$f_{Y \mid X}(y \mid x) = \begin{cases} 1, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}.$$

In other words, $Y_{|X=x} \sim \mathcal{U}(0, 1)$ for each $0 < x < 1$. By symmetry, we conclude that $X_{|Y=y} \sim \mathcal{U}(0, 1)$ for each $0 < y < 1$.

b) From part (a), we conclude that the conditional PDFs of Y given $X = x$ are identical to each other and to the marginal PDF of Y. Thus, in this case, knowing the value of the random variable X imparts no information about the distribution of Y; in other words, knowing the value of X doesn't affect the probability distribution of Y.

9.64

a) In this case, we know from Exercise 9.63 that $X_{|Y=y} \sim \mathcal{U}(0, 1)$ for each $0 < y < 1$. Consequently, we have $P(X > 0.9 \mid Y = 0.8) = 0.1$.

b) For $0 < y < 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_0^1 (x + y)\, dx = 0.5 + y.$$

Then, for $0 < x < 1$,

$$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{x + y}{0.5 + y}.$$

Therefore, from (the conditional version of) the FPF,

$$P(X > 0.9 \mid Y = 0.8) = \int_{0.9}^{\infty} f_{X \mid Y}(x \mid 0.8)\, dx = \frac{1}{0.5 + 0.8} \int_{0.9}^{1} (x + 0.8)\, dx = 0.135.$$

c) For $0 < y < 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_0^1 \frac{3}{2} \left(x^2 + y^2\right) dx = \frac{1}{2} + \frac{3}{2} y^2.$$

Then, for $0 < x < 1$,

$$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{3\left(x^2 + y^2\right)}{1 + 3y^2}.$$

Therefore, from (the conditional version of) the FPF,

$$P(X > 0.9 \mid Y = 0.8) = \int_{0.9}^{\infty} f_{X \mid Y}(x \mid 0.8)\, dx = \frac{3}{1 + 3 \cdot (0.8)^2} \int_{0.9}^{1} \left(x^2 + y^2\right) dx = 0.159.$$

9.65

a) For $-1 < x < 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_0^{\sqrt{1-x^2}} (2/\pi)\, dy = (2/\pi)\sqrt{1 - x^2}.$$

Hence,

$$f_X(x) = \begin{cases} (2/\pi)\sqrt{1 - x^2}, & \text{if } -1 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) For $0 < y < 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} (2/\pi)\, dx = (4/\pi)\sqrt{1-y^2}.$$

Hence,

$$f_Y(y) = \begin{cases} (4/\pi)\sqrt{1-y^2}, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

c) Let $-1 < x < 1$. Then, for $0 < y < \sqrt{1-x^2}$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{2/\pi}{(2/\pi)\sqrt{1-x^2}} = \frac{1}{\sqrt{1-x^2}}.$$

Thus, $Y_{|X=x} \sim \mathcal{U}\left(0, \sqrt{1-x^2}\right)$ for $-1 < x < 1$.
d) Let $0 < y < 1$. Then, for $-\sqrt{1-y^2} < x < \sqrt{1-y^2}$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{2/\pi}{(4/\pi)\sqrt{1-y^2}} = \frac{1}{2\sqrt{1-y^2}}.$$

Thus, $X_{|Y=y} \sim \mathcal{U}\left(-\sqrt{1-y^2}, \sqrt{1-y^2}\right)$ for $0 < y < 1$.

9.66
a) We have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{3}{8}(|x| + |y|)\, dy = \frac{3}{4}\int_0^{\sqrt{1-x^2}} (|x| + y)\, dy \\ &= \frac{3}{4}\left(|x|\sqrt{1-x^2} + \frac{1-x^2}{2}\right) = \frac{3}{8}\left(1 + 2|x|\sqrt{1-x^2} - x^2\right) \end{aligned}$$

if $-1 < x < 1$, and $f_X(x) = 0$ otherwise.
b) By symmetry, X and Y have the same probability distribution. Thus,

$$f_Y(y) = \begin{cases} \dfrac{3}{8}\left(1 + 2|y|\sqrt{1-y^2} - y^2\right), & \text{if } -1 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

c) Referring to the results of parts (a) and (b), we see that, for $-1 < y < 1$,

$$f_{Y\,|\,X}(y\,|\,0) = \frac{f_{X,Y}(0, y)}{f_X(0)} = \frac{\frac{3}{8}|y|}{\frac{3}{8}} = |y|.$$

Hence, by the FPF,

$$P(|Y| > 1/2\,|\,X = 0) = \int_{|y|>1/2} f_{Y\,|\,X}(y\,|\,0)\, dy = \int_{1/2<|y|<1} |y|\, dy = 2\int_{1/2}^{1} y\, dy = 3/4.$$

9.67 For $0 < x < 1$,

$$f_X(x) = n(1-x)^{n-1} = \frac{n!}{0!\,(n-1)!}x^{1-1}(1-x)^{n-1}$$
$$= \frac{\Gamma(1+n)}{\Gamma(1)\Gamma(n)}x^{1-1}(1-x)^{n-1} = \frac{1}{B(1,n)}x^{1-1}(1-x)^{n-1}.$$

Thus, X has the beta distribution with parameters 1 and n.

9.68

a) This problem concerns conditional probability, but it does not involve a conditional PDF because the event conditioned on is not of the form $\{X = x\}$.

b) In this case,

$$F(t) = \begin{cases} 0, & \text{if } t < 0; \\ t, & \text{if } 0 \le t < 1; \\ 1, & \text{if } t \ge 1. \end{cases}$$

Applying the first displayed equation in the solution to Example 9.4(b) (see page 498) yields

$$P(X > 0.1,\ Y \le 0.7) = (0.7 - 0.1)^n = (0.6)^n$$

and

$$P(X > 0.1) = P(X > 0.1,\ Y \le 1) = (1 - 0.1)^n = (0.9)^n.$$

Therefore, by the conditional probability rule,

$$P(Y \le 0.7 \mid X > 0.1) = \frac{P(X > 0.1,\ Y \le 0.7)}{P(X > 0.1)} = \frac{(0.6)^n}{(0.9)^n} = (2/3)^n.$$

9.69

a) Yes, there is a difference. With method II, we have $X \sim \mathcal{U}(0,1)$. However, with method I, it's clear that values of X near 1 are more likely than values of X near 0; thus, $X \not\sim \mathcal{U}(0,1)$.

b) For method I, $f_{X,Y}(x,y) = 1\big/\frac{1}{2} = 2$ if $(x,y) \in T$, and $f_{X,Y}(x,y) = 0$ otherwise. For method II, we know that $X \sim \mathcal{U}(0,1)$ and $Y_{|X=x} \sim \mathcal{U}(0,x)$. Therefore, by the general multiplication rule,

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

if $(x,y) \in T$, and $f_{X,Y}(x,y) = 0$ otherwise. These results show that the joint PDFs differ for the two methods and, hence, provide formal justification for the answer to part(a).

c) For method I,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^x 2\,dy = 2x$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise; thus, X has the beta distribution with parameters 2 and 1. Also,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_y^1 2\,dx = 2(1-y)$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise; thus, Y has the beta distribution with parameters 1 and 2.

For method II, we already know that $X \sim \mathcal{U}(0,1)$. To obtain the marginal PDF of Y, we proceed as follows:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_y^1 \frac{1}{x}\,dx = -\ln y$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise.

d) For method I, we have, for $0 < x < 1$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}$$

if $0 < y < x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise; thus, $Y_{|X=x} \sim \mathcal{U}(0, x)$ for $0 < x < 1$. Also, for $0 < y < 1$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y}$$

if $y < x < 1$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise; thus, $X_{|Y=y} \sim \mathcal{U}(y, 1)$ for $0 < y < 1$.

For method II, we already know that $Y_{|X=x} \sim \mathcal{U}(0, x)$ for $0 < x < 1$; this conditional distribution is the same as that for method I. Also, for $0 < y < 1$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1/x}{-\ln y} = -\frac{1}{x \ln y},$$

if $y < x < 1$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise; this conditional distribution differs from that for method I.

9.70

a) We have

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\max\{X_1, \dots, X_n\} \le y) = P(X_1 \le y, \dots, X_n \le y) \\ &= P(X_1 \le y) \cdots P(X_n \le y) = \big(F(y)\big)^n. \end{aligned}$$

Differentiation now yields

$$f_Y(y) = F_Y'(y) = n\big(F(y)\big)^{n-1} f(y) = nf(y)\big(F(y)\big)^{n-1}.$$

b) Making the substitution $u = F(y) - F(x)$, we get

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{-\infty}^{y} n(n-1) f(x) f(y)\big(F(y) - F(x)\big)^{n-2}\,dx \\ &= n(n-1) f(y) \int_{-\infty}^{y} f(x)\big(F(y) - F(x)\big)^{n-2}\,dx = n(n-1) f(y) \int_0^{F(y)} u^{n-2}\,du \\ &= nf(y)\big(F(y)\big)^{n-1}. \end{aligned}$$

c) In this case,

$$F(y) = \begin{cases} 0, & \text{if } y < 0; \\ y, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases} \qquad \text{and} \qquad f(y) = \begin{cases} 1, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, from part (a) or (b),

$$f_Y(y) = n \cdot 1 \cdot y^{n-1} = \frac{1}{B(n,1)} y^{n-1}(1-y)^{1-1}$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Thus, Y has the beta distribution with parameters n and 1.

9.71

a) Because the lifetimes of the components, $X_1, \dots, X_n$, are independent and identically distributed exponential random variables with common parameter λ, they can be considered a random sample of size n from an exponential distribution with parameter λ.

b) Let F and f denote the common CDF and PDF, respectively, of $X_1, \dots, X_n$. Then

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-\lambda x}, & \text{if } x \ge 0. \end{cases} \qquad \text{and} \qquad f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

The time at which the first failure occurs is $X = \min\{X_1, \ldots, X_n\}$. From Example 9.9, we know a PDF of X in the general case. Applying that to the present situation yields

$$f_X(x) = nf(x)\bigl(1 - F(x)\bigr)^{n-1} = n\lambda e^{-\lambda x}\left(e^{-\lambda x}\right)^{n-1} = n\lambda e^{-n\lambda x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Thus, we see that $X \sim \mathcal{E}(n\lambda)$.
c) The system lifetime is $Y = \max\{X_1, \ldots, X_n\}$. Referring now to the result of Exercise 9.70(b) yields

$$f_Y(y) = nf(y)\bigl(F(y)\bigr)^{n-1} = n\lambda e^{-\lambda y}\bigl(1 - e^{-\lambda y}\bigr)^{n-1}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.
d) We have

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{n(n-1)f(x)f(y)\left(F(y) - F(x)\right)^{n-2}}{nf(x)\left(1 - F(x)\right)^{n-1}}$$
$$= \frac{(n-1)f(y)\left(F(y) - F(x)\right)^{n-2}}{\left(1 - F(x)\right)^{n-1}}.$$

Substituting the exponential PDF and CDF yields, for $x > 0$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{(n-1)\lambda e^{-\lambda y}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2}}{\left(e^{-\lambda x}\right)^{n-1}} = (n-1)\lambda e^{-\lambda(y-x)}\left(1 - e^{-\lambda(y-x)}\right)^{n-2}$$

if $y > x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise.
e) Given that the time of the first component failure is x, system lifetime is at the failure of the remaining $n-1$ components. By the lack-of-memory property of the exponential distribution, given that a component is still working at time x, its residual lifetime is exponential with parameter λ. Thus, given that the time of the first failure is at x, system lifetime is that of an $(n-1)$-component system starting at time x. Applying the result of part (c) with $n-1$ replacing n and $y-x$ replacing y, we find that, for $x > 0$,

$$f_{Y\,|\,X}(y\,|\,x) = (n-1)\lambda e^{-\lambda(y-x)}\left(1 - e^{-\lambda(y-x)}\right)^{n-2}$$

if $y > x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise.
f) We have, for $y > 0$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{n(n-1)\left(\lambda e^{-\lambda x}\right)\left(\lambda e^{-\lambda y}\right)\left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2}}{n\lambda e^{-\lambda y}\left(1 - e^{-\lambda y}\right)^{n-1}}$$
$$= \frac{(n-1)\lambda e^{-\lambda x}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2}}{\left(1 - e^{-\lambda y}\right)^{n-1}}$$

if $0 < x < y$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise.

9.72
a) Referring to Exercise 9.71(d) and making the substitution $u = 1 - e^{-(y-1)}$, we find that

$$P(Y > 4\,|\,X = 1) = \int_{y>4} f_{Y\,|\,X}(y\,|\,1)\,dy = \int_4^\infty 9e^{-(y-1)}\left(1 - e^{-(y-1)}\right)^8\,dy$$
$$= 9\int_{1-e^{-3}}^1 u^8\,du = 1 - \left(1 - e^{-3}\right)^9 = 0.368.$$

b) Referring to Exercise 9.71(f) and making the substitution $u = e^{-x} - e^{-4}$, we find that

$$P(X > 1 \mid Y = 4) = \int_{x>1} f_{X \mid Y}(x \mid 4)\,dx = \int_1^4 \frac{9e^{-x}\left(e^{-x} - e^{-4}\right)^8}{\left(1 - e^{-4}\right)^9}\,dx$$

$$\frac{9}{\left(1 - e^{-4}\right)^9}\int_0^{e^{-1}-e^{-4}} u^8\,du = \frac{\left(e^{-1} - e^{-4}\right)^9}{\left(1 - e^{-4}\right)^9} = \left(\frac{e^{-1} - e^{-4}}{1 - e^{-4}}\right)^9 = 0.0000920.$$

9.73 Note that, for $y > 1$,

$$f_{X,Y}(2, y) = \frac{2}{2^2(2-1)}y^{-(2\cdot2-1)/(2-1)} = \frac{1}{2}y^{-3}.$$

Therefore,

$$f_X(2) = \int_{-\infty}^{\infty} f_{X,Y}(2, y)\,dy = \frac{1}{2}\int_1^{\infty} y^{-3}\,dy = \frac{1}{4}$$

and, hence,

$$f_{Y \mid X}(y \mid 2) = \frac{f_{X,Y}(2, y)}{f_X(2)} = \frac{(1/2)y^{-3}}{1/4} = 2y^{-3}$$

if $y > 1$, and $f_{Y \mid X}(y \mid 2) = 0$ otherwise. Consequently,

$$P(1 \le Y \le 3 \mid X = 2) = \int_{1 \le y \le 3} f_{Y \mid X}(y \mid 2)\,dy = \int_1^3 2y^{-3}\,dy = \frac{8}{9}.$$

9.74 From the general multiplication rule,

$$f_{X,Y}(x, y) = f_X(x)f_{Y \mid X}(y \mid x) = 1 \cdot 1 = 1$$

if $0 < x < 1$ and $x < y < x + 1$, and $f_{X,Y}(x, y) = 0$ otherwise. Thus,

$$P(Y > 0.5) = \iint_{y>0.5} f_{X,Y}(x, y)\,dx\,dy = \int_0^{0.5}\left(\int_{0.5}^{x+1} 1\,dy\right)dx + \int_{0.5}^{1}\left(\int_x^{x+1} 1\,dy\right)dx$$

$$= \int_0^{0.5}(x + 0.5)\,dx + \int_{0.5}^1 1\,dx = \frac{7}{8}.$$

9.75 By assumption,

$$f_{Y \mid \Lambda}(y \mid \lambda) = \begin{cases} \lambda e^{-\lambda y}, & \text{if } y > 0; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_\Lambda(\lambda) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}, & \text{if } \lambda > 0; \\ 0, & \text{otherwise.} \end{cases}$$

a) From the general multiplication rule,

$$f_{\Lambda,Y}(\lambda, y) = f_\Lambda(\lambda)f_{Y \mid \Lambda}(y \mid \lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \cdot \lambda e^{-\lambda y} = \frac{(\beta\lambda)^\alpha}{\Gamma(\alpha)}e^{-(\beta+y)\lambda}$$

if $\lambda > 0$ and $y > 0$, and $f_{\Lambda,Y}(\lambda, y) = 0$ otherwise. Thus, $f_Y(y) = 0$ if $y \le 0$ and, for $y > 0$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{\Lambda,Y}(\lambda, y)\,d\lambda = \int_0^{\infty}\frac{(\beta\lambda)^\alpha}{\Gamma(\alpha)}e^{-(\beta+y)\lambda}\,d\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^{\infty}\lambda^\alpha e^{-(\beta+y)\lambda}\,d\lambda$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{(\beta+y)^{\alpha+1}}\int_0^{\infty}\frac{(\beta+y)^{\alpha+1}}{\Gamma(\alpha+1)}\lambda^{(\alpha+1)-1}e^{-(\beta+y)\lambda}\,d\lambda = \frac{\alpha\beta^\alpha}{(\beta+y)^{\alpha+1}}.$$

b) For $y > 0$,

$$f_{\Lambda\,|\,Y}(\lambda\,|\,y) = \frac{f_{\Lambda,Y}(\lambda, y)}{f_Y(y)} = \frac{\dfrac{(\beta\lambda)^\alpha}{\Gamma(\alpha)} e^{-(\beta+y)\lambda}}{\dfrac{\alpha\beta^\alpha}{(\beta+y)^{\alpha+1}}}$$

$$= \frac{(\beta+y)^{\alpha+1}}{\alpha\Gamma(\alpha)} \lambda^\alpha e^{-(\beta+y)\lambda} = \frac{(\beta+y)^{\alpha+1}}{\Gamma(\alpha+1)} \lambda^{(\alpha+1)-1} e^{-(\beta+y)\lambda}$$

if $\lambda > 0$, and $f_{\Lambda\,|\,Y}(\lambda\,|\,y) = 0$ otherwise. Thus, $\Lambda_{|Y=y} \sim \Gamma(\alpha+1, \beta+y)$ for $y > 0$.

9.76

a) By assumption,

$$f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}, \qquad f_{Y\,|\,X}(y\,|\,x) = \begin{cases} 1/x, & \text{if } 0 < y < x < 1; \\ 0, & \text{otherwise.} \end{cases},$$

and

$$f_{Z\,|\,X,Y}(z\,|\,x, y) = \begin{cases} 1/y, & \text{if } 0 < z < y < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, by the general multiplication rule in the form of Equation (9.36) on page 519,

$$f_{X,Y,Z}(x, y, z) = f_X(x) f_{Y\,|\,X}(y\,|\,x) f_{Z\,|\,X,Y}(z\,|\,x, y) = 1 \cdot \frac{1}{x} \cdot \frac{1}{y} = \frac{1}{xy}$$

if $0 < z < y < x < 1$, and $f_{X,Y,Z}(x, y, z) = 0$ otherwise.

b) We have

$$f_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\,dx\,dy = \int_z^1 \left(\int_z^x \frac{1}{xy}\,dy\right) dx = \int_z^1 \left(\int_z^x \frac{dy}{y}\right) \frac{1}{x}\,dx$$

$$= \int_z^1 \frac{1}{x} \ln(x/z)\,dx = \int_0^{-\ln z} u\,du = \frac{1}{2}(\ln z)^2$$

if $0 < z < 1$, and $f_Z(z) = 0$ otherwise, where we obtained the penultimate integral by making the substitution $u = \ln(x/z)$.

c) From parts (a) and (b), we have, for $0 < z < 1$,

$$f_{X,Y\,|\,Z}(x, y\,|\,z) = \frac{f_{X,Y,Z}(x, y, z)}{f_Z(z)} = \frac{\dfrac{1}{xy}}{\dfrac{1}{2}(\ln z)^2} = \frac{2}{xy(\ln z)^2}$$

if $z < y < x < 1$ and $f_{X,Y\,|\,Z}(x, y\,|\,z) = 0$ otherwise.

Theory Exercises

9.77 Clearly, $f_{Y\,|\,X}(\cdot\,|\,x)$ is a nonnegative function on $\mathcal{R}$. Also, in view of Definition 9.3 on page 515 and Equation (9.27) on page 510,

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

9.78
a) For all $x, y \in \mathcal{R}$,

$$0 \leq P\big(X \in A^c \cap (-\infty, x],\ Y \leq y\big) \leq P(X \in A^c) = \int_{A^c} f_X(x)\,dx = 0,$$

where the last equality follows from the fact that $f_X(x) = 0$ for all $x \in A^c$. Consequently, we have shown that $P\big(X \in A^c \cap (-\infty, x],\ Y \leq y\big) = 0$ for all $x, y \in \mathcal{R}$.
b) From the law of partitions and part (a),

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x,\ Y \leq y) = P\big(X \in A \cap (-\infty, x],\ Y \leq y\big) + P\big(X \in A^c \cap (-\infty, x],\ Y \leq y\big) \\ &= P\big(X \in A \cap (-\infty, x],\ Y \leq y\big) + 0 = P\big(X \in A \cap (-\infty, x],\ Y \leq y\big). \end{aligned}$$

c) Recall from Definition 9.3 on page 515 that, if $f_X(x) = 0$, then we define $f_{Y\,|\,X}(y\,|\,x) = 0$ for all $y \in \mathcal{R}$. Using that convention and part (b), we get, for all $x, y \in \mathcal{R}$,

$$\begin{aligned} F_{X,Y}(x, y) &= P\big(X \in A \cap (-\infty, x],\ Y \leq y\big) \\ &= \int_{A \cap (-\infty, x]} \int_{-\infty}^{y} f_{X,Y}(s, t)\,ds\,dt = \int_{A \cap (-\infty, x]} \int_{-\infty}^{y} f_X(s) f_{Y\,|\,X}(t\,|\,s)\,ds\,dt \\ &= \int_{A \cap (-\infty, x]} \int_{-\infty}^{y} f_X(s) f_{Y\,|\,X}(t\,|\,s)\,ds\,dt + \int_{A^c \cap (-\infty, x]} \int_{-\infty}^{y} f_X(s) f_{Y\,|\,X}(t\,|\,s)\,ds\,dt \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(s) f_{Y\,|\,X}(t\,|\,s)\,ds\,dt. \end{aligned}$$

d) From part (c) and Proposition 9.4 on page 499, we conclude that the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_{Y\,|\,X}(y\,|\,x)$ is a joint PDF of X and Y.

9.79
a) Applying Proposition 9.7 on page 510 and the general multiplication rule yields

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{-\infty}^{\infty} f_{Y\,|\,X}(y\,|\,x) f_X(x)\,dx.$$

Therefore, the function f defined on $\mathcal{R}$ by $f(y) = \int_{-\infty}^{\infty} f_{Y\,|\,X}(y\,|\,x) f_X(x)\,dx$ is a PDF of the random variable Y.
b) Applying the FPF, part (a), and Equation (9.33) on page 517, we obtain, for all $A \subset \mathcal{R}^2$,

$$\begin{aligned} P(Y \in A) &= \int_A f_Y(y)\,dy = \int_A \left(\int_{-\infty}^{\infty} f_{Y\,|\,X}(y\,|\,x) f_X(x)\,dx\right) dy \\ &= \int_{-\infty}^{\infty} \left(\int_A f_{Y\,|\,X}(y\,|\,x)\,dy\right) f_X(x)dx = \int_{-\infty}^{\infty} P(Y \in A\,|\,X = x) f_X(x)\,dx. \end{aligned}$$

9.80 Let $X_1, \ldots, X_m$ be continuous random variables with a joint PDF. Then the function f defined on $\mathcal{R}^m$ by

$$f(x_1, \ldots, x_m) = f_{X_1}(x_1) f_{X_2\,|\,X_1}(x_2\,|\,x_1) \cdots f_{X_m\,|\,X_1,\ldots,X_{m-1}}(x_m\,|\,x_1, \ldots, x_{m-1})$$

is a joint PDF of $X_1, \ldots, X_m$.

Advanced Exercises

9.81 As usual, we use $|\cdot|$ to denote, in the extended sense, length in $\mathcal{R}$ and area in $\mathcal{R}^2$.
a) From Proposition 9.7 on page 510,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = \int_{S_x} \frac{1}{|S|}\,dy = \frac{|S_x|}{|S|}.$$

Thus, $f_X(x) = |S_x|/|S|$ if $|S_x| > 0$, and $f_X(x) = 0$ otherwise. Similarly, we find that $f_Y(y) = |S^y|/|S|$ if $|S^y| > 0$, and $f_Y(y) = 0$ otherwise.
b) Suppose that $f_X(x) > 0$, that is, $|S_x| > 0$. From the solution to part (a),

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{1/|S|}{|S_x|/|S|} = \frac{1}{|S_x|},$$

if $y \in S_x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. Consequently, $Y_{|X=x} \sim \mathcal{U}(S_x)$. Similarly, $X_{|Y=y} \sim \mathcal{U}(S^y)$ for all y such that $|S^y| > 0$.
c) No, as Example 9.8 on page 511 shows.
d) Suppose that there exist subsets A and B of $\mathcal{R}$ such that $S = A \times B$. Observe that $|S| = |A||B|$ and, in particular, both $|A|$ and $|B|$ must be positive. Also observe that $S_x = B$ if $x \in A$, and $S_x = \emptyset$ otherwise. Therefore, from the solution to part (a),

$$f_X(x) = \frac{|S_x|}{|S|} = \frac{|B|}{|S|} = \frac{|B|}{|A||B|} = \frac{1}{|A|}$$

if $x \in A$, and $f_X(x) = 0$ otherwise. Hence, $X \sim \mathcal{U}(A)$. Similarly, we find that $Y \sim \mathcal{U}(B)$.

9.82 No! The random variables X and Y from Exercise 9.30 are each $\mathcal{U}(0, 1)$ random variables and, hence, each have a PDF. However, they don't have a joint PDF.

9.83 We first observe that, from the definition of a joint PDF/PMF and the continuity property of a probability measure,

$$P(X \le x,\ Y = y) = \int_{-\infty}^{x} h_{X,Y}(s, y)\,ds \quad \text{and} \quad P(-\infty < X < \infty,\ Y = y) = \int_{-\infty}^{\infty} h_{X,Y}(x, y)\,dx.$$

a) From the law of partitions,

$$\begin{aligned} F_X(x) = P(X \le x) &= \sum_y P(X \le x,\ Y = y) \\ &= \sum_y \int_{-\infty}^{x} h_{X,Y}(s, y)\,ds = \int_{-\infty}^{x} \left(\sum_y h_{X,Y}(s, y) \right) ds. \end{aligned}$$

Therefore, from Proposition 8.5 on page 422, we have $f_X(x) = \sum_y h_{X,Y}(x, y)$.
b) We have

$$p_Y(y) = P(Y = y) = P(-\infty < X < \infty,\ Y = y) = \int_{-\infty}^{\infty} h_{X,Y}(x, y)\,dx.$$

9.84 From Exercise 9.83(a),

$$f_X(x) = \sum_y h_{X,Y}(x, y) = \sum_{y=0}^{1} \frac{1}{2} e^{-x} = e^{-x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Thus, $X \sim \mathcal{E}(1)$. Also, from Exercise 9.83(b),

$$p_Y(y) = P(Y = y) = \int_{-\infty}^{\infty} h_{X,Y}(x, y)\,dx = \int_0^{\infty} \frac{1}{2} e^{-x}\,dx = \frac{1}{2}$$

if $y \in \{0, 1\}$, and $p_Y(y) = 0$ otherwise. Thus, Y has the Bernoulli distribution with parameter 1/2 or, equivalently, $Y \sim \mathcal{B}(1, 1/2)$.

9.85
a) If $f_X(x) > 0$, we define the conditional PMF of Y given $X = x$ by

$$p_{Y\,|\,X}(y\,|\,x) = \frac{h_{X,Y}(x, y)}{f_X(x)}.$$

If $f_X(x) = 0$, we define $p_{Y\,|\,X}(y\,|\,x) = 0$ for all $y \in \mathcal{R}$ but, in that case, we don't refer to $p_{Y\,|\,X}$ as a conditional PMF.
b) If $p_Y(y) > 0$, we define the conditional PDF of X given $Y = y$ by

$$f_{X\,|\,Y}(x\,|\,y) = \frac{h_{X,Y}(x, y)}{p_Y(y)}.$$

If $p_Y(y) = 0$, we define $f_{X\,|\,Y}(x\,|\,y) = 0$ for all $x \in \mathcal{R}$ but, in that case, we don't refer to $f_{X\,|\,Y}$ as a conditional PDF.

9.86
a) We have

$$f_X(x) = \sum_y h_{X,Y}(x, y) = \sum_{y=0}^{\infty} \lambda e^{-(1+\lambda)x} \frac{x^y}{y!} = \lambda e^{-(1+\lambda)x} \sum_{y=0}^{\infty} \frac{x^y}{y!} = \lambda e^{-(1+\lambda)x} e^x = \lambda e^{-\lambda x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Thus, $X \sim \mathcal{E}(\lambda)$.

Also, we have

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} h_{X,Y}(x, y)\,dx = \int_0^{\infty} \lambda e^{-(1+\lambda)x} \frac{x^y}{y!}\,dx = \frac{\lambda}{y!} \int_0^{\infty} e^{-(1+\lambda)x} x^y\,dx \\ &= \frac{\lambda}{y!} \cdot \frac{\Gamma(y+1)}{(1+\lambda)^{y+1}} \int_0^{\infty} \frac{(1+\lambda)^{y+1}}{\Gamma(y+1)} x^{(y+1)-1}\, e^{-(1+\lambda)x}\,dx = \frac{\lambda}{y!} \cdot \frac{y!}{(1+\lambda)^{y+1}} \cdot 1 \\ &= \frac{(1/\lambda)^y}{(1+1/\lambda)^{y+1}} \end{aligned}$$

if $y = 0, 1, \ldots$, and $p_Y(y) = 0$ otherwise. Thus, Y has the Pascal distribution with parameter $1/\lambda$.
b) For $y = 0, 1, \ldots,$

$$f_{X\,|\,Y}(x\,|\,y) = \frac{h_{X,Y}(x, y)}{p_Y(y)} = \frac{\lambda e^{-(1+\lambda)x} \dfrac{x^y}{y!}}{\dfrac{(1/\lambda)^y}{(1+1/\lambda)^{y+1}}} = \frac{(1+\lambda)^{y+1}}{\Gamma(y+1)} x^{(y+1)-1}\, e^{-(1+\lambda)x}$$

if $x > 0$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise. Thus, $X_{|Y=y} \sim \Gamma(y+1, 1+\lambda)$ for $y = 0, 1, \ldots.$

Also, for $x > 0$,

$$p_{Y\,|\,X}(y\,|\,x) = \frac{h_{X,Y}(x,y)}{f_X(x)} = \frac{\lambda e^{-(1+\lambda)x}\dfrac{x^y}{y!}}{\lambda e^{-\lambda x}} = e^{-x}\frac{x^y}{y!}$$

if $y = 0, 1, \ldots,$ and $p_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. Thus, $Y_{|X=x} \sim \mathcal{P}(x)$ for $x > 0$.

9.87 By assumption, $Y_{|X=\lambda} \sim \mathcal{P}(\lambda)$ for $\lambda > 0$, and $X \sim \Gamma(\alpha, \beta)$.

a) From the general multiplication rule for a joint PDF/PMF,

$$h_{X,Y}(\lambda, y) = f_X(\lambda)p_{Y\,|\,X}(y\,|\,\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \cdot e^{-\lambda}\frac{\lambda^y}{y!} = \frac{\beta^\alpha\lambda^{y+\alpha-1}}{y!\,\Gamma(\alpha)}e^{-(\beta+1)\lambda}$$

if $\lambda > 0$ and $y = 0, 1, \ldots,$ and $h_{X,Y}(\lambda, y) = 0$ otherwise.

b) We have, for $y = 0, 1, \ldots,$

$$\begin{aligned}
p_Y(y) &= \int_{-\infty}^{\infty} h_{X,Y}(\lambda, y)\,d\lambda = \int_0^\infty \frac{\beta^\alpha\lambda^{y+\alpha-1}}{y!\,\Gamma(\alpha)}e^{-(\beta+1)\lambda}\,d\lambda \\
&= \frac{\beta^\alpha}{y!\,\Gamma(\alpha)} \cdot \frac{\Gamma(y+\alpha)}{(\beta+1)^{y+\alpha}}\int_0^\infty \frac{(\beta+1)^{y+\alpha}}{\Gamma(y+\alpha)}\lambda^{(y+\alpha)-1}e^{-(\beta+1)\lambda}\,d\lambda \\
&= \frac{\beta^\alpha\lambda^{y+\alpha-1}}{y!\,\Gamma(\alpha)} \cdot \frac{\Gamma(y+\alpha)}{(\beta+1)^{y+\alpha}} \cdot 1 = \frac{\Gamma(y+\alpha)}{y!\,\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{(\beta+1)^{y+\alpha}}.
\end{aligned}$$

However, from Equation (8.43) on page 451,

$$\begin{aligned}
\frac{\Gamma(y+\alpha)}{y!\,\Gamma(\alpha)} &= \frac{(\alpha+y-1)(\alpha+y-2)\cdots\alpha\Gamma(\alpha)}{y!\,\Gamma(\alpha)} \\
&= \frac{(-1)^y(-\alpha)(-\alpha-1)\cdots(-\alpha-y+1)}{y!} = (-1)^y\binom{-\alpha}{y}.
\end{aligned}$$

Also,

$$\frac{\beta^\alpha}{(\beta+1)^{y+\alpha}} = \left(\frac{\beta}{\beta+1}\right)^\alpha \cdot \frac{1}{(\beta+1)^y} = \left(\frac{\beta}{\beta+1}\right)^\alpha\left(1 - \frac{\beta}{\beta+1}\right)^y.$$

It now follows that

$$p_Y(y) = \binom{-\alpha}{y}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{\beta}{\beta+1} - 1\right)^y$$

if $y = 0, 1, \ldots,$ and $p_Y(y) = 0$ otherwise. Hence, referring to page 243, we see that Y has the Pascal distribution with parameters α and $\beta/(\beta+1)$.

c) For $y = 0, 1, \ldots,$

$$\begin{aligned}
f_{X\,|\,Y}(\lambda\,|\,y) &= \frac{h_{X,Y}(\lambda, y)}{p_Y(y)} = \frac{\dfrac{\beta^\alpha\lambda^{y+\alpha-1}}{y!\,\Gamma(\alpha)}e^{-(\beta+1)\lambda}}{\dbinom{-\alpha}{y}\left(\dfrac{\beta}{\beta+1}\right)^\alpha\left(\dfrac{\beta}{\beta+1} - 1\right)^y} \\
&= \frac{\dfrac{\beta^\alpha\lambda^{y+\alpha-1}}{y!\,\Gamma(\alpha)}e^{-(\beta+1)\lambda}}{\dfrac{\Gamma(y+\alpha)\beta^\alpha}{y!\,\Gamma(\alpha)(\beta+1)^{y+\alpha}}} = \frac{(\beta+1)^{y+\alpha}}{\Gamma(y+\alpha)}\lambda^{(y+\alpha)-1}e^{-(\beta+1)\lambda}
\end{aligned}$$

for $\lambda > 0$, and $f_{X\,|\,Y}(\lambda\,|\,y) = 0$ otherwise. Hence, $X_{|Y=y} \sim \Gamma(y+\alpha, \beta+1)$ for $y = 0, 1, \ldots.$

9.5 Independent Continuous Random Variables

Basic Exercises

9.88
a) Given that $X = 0$, the possible values of Y are between -1 and 1. However, given that $X = 0.5$, the possible values of Y are between $-\sqrt{3}/2$ and $\sqrt{3}/2$. In particular, knowing the value of X affects the probability distribution of Y.
b) From Example 9.8,

$$f_Y(y) = \begin{cases} (2/\pi)\sqrt{1-y^2}, & \text{if } -1 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

In particular, we note that Y does not have a uniform distribution. However, from Example 9.10, we know that $Y_{|X=x} \sim \mathcal{U}\big(-\sqrt{1-x^2}, \sqrt{1-x^2}\big)$ for $-1 < x < 1$. It now follows from Proposition 9.10(a) on page 527 that X and Y are not independent random variables.
c) Let g denote the joint PDF of X and Y in Equation (9.39) on page 524, and let f denote the product of the marginals of X and Y. We find that, within the unit disk, the second-order mixed partials exist. In particular, they exist at $(0, 0)$. If f were also a joint PDF of X and Y, then, according to Equation (9.22) on page 499, we would have $f(0, 0) = g(0, 0)$. But that isn't true as $f(0, 0) = 4/\pi^2$ and $g(0, 0) = 1/\pi$.

9.89 From Example 9.11 on page 516, each conditional PDF of Y given $X = x$ is a PDF of Y. Independence of X and Y now follows from Proposition 9.10(a) on page 527.

9.90 By assumption,

$$f_X(x) = \begin{cases} 1/10, & \text{if } -5 < x < 5; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} 1/10, & \text{if } -5 < y < 5; \\ 0, & \text{otherwise.} \end{cases}$$

As X and Y are independent, a joint PDF of X and Y is equal to the product of the two marginals:

$$f_{X,Y}(x, y) = \begin{cases} 1/100, & \text{if } -5 < x < 5 \text{ and } -5 < y < 5; \\ 0, & \text{otherwise.} \end{cases}$$

We can use the FPF to evaluate $P(|X - Y| > 5)$. Alternatively, we note that, in this case, X and Y are the x and y coordinates, respectively, of a point selected at random from the square

$$\Omega = \{\, (x, y) : -5 < x < 5, -5 < y < 5 \,\}.$$

Thus,

$$P(|X - Y| > 5) = \frac{\big|\{|X - Y| > 5\}\big|}{|\Omega|} = \frac{\big|\{\, (x, y) \in \Omega : |x - y| > 5 \,\}\big|}{|\Omega|} = \frac{25}{100} = \frac{1}{4}.$$

9.91 We have

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

and, from Example 9.7 on page 510,

$$f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} 1, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, we see that the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_Y(y)$ is a joint PDF of X and Y. Consequently, by Proposition 9.9 on page 523, X and Y are independent random variables.

9.92 From the solution to Exercise 9.65,

$$f_Y(y) = \begin{cases} (4/\pi)\sqrt{1-y^2}, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

In particular, we note that Y does not have a uniform distribution. However, we also know from the solution to Exercise 9.65 that $Y_{|X=x} \sim \mathcal{U}\big(0, \sqrt{1-x^2}\big)$ for $-1 < x < 1$. It now follows from Proposition 9.10(a) on page 527 that X and Y are not independent random variables.

9.93 By assumption, $Y_{|X=x} \sim \mathcal{N}\big(\rho x, 1-\rho^2\big)$ for each $x \in \mathcal{R}$. From Example 9.13(b) on page 517, $Y \sim \mathcal{N}(0, 1)$. From Proposition 9.10(a), X and Y are independent if and only if $Y_{|X=x}$ is a PDF of Y for each $x \in \mathcal{R}$, which is the case, if and only if $\rho = 0$.

9.94 We have

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

if $-1 < x < 1$ and $-1 < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise. The roots of the quadratic equation $x^2 + Xx + Y = 0$ are real if and only if $X^2 - 4Y \ge 0$. Applying the FPF now yields

$$P\left(X^2 - 4Y \ge 0\right) = \iint\limits_{x^2-4y\ge 0} f_{X,Y}(x, y)\,dx\,dy = \int_{-1}^{1}\left(\int_{-1}^{x^2/4} \frac{1}{4}\,dy\right)dx = \frac{13}{24}.$$

9.95

a) We have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = \int_{-\infty}^{\infty} g(x)h(y)\,dy = \left(\int_{-\infty}^{\infty} h(y)\,dy\right) g(x).$$

b) We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{-\infty}^{\infty} g(x)h(y)\,dx = \left(\int_{-\infty}^{\infty} g(x)\,dx\right) h(y).$$

c) From Proposition 8.6(b) on page 423 and part (a),

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty}\left(\left(\int_{-\infty}^{\infty} h(y)\,dy\right) g(x)\right)dx = \left(\int_{-\infty}^{\infty} g(x)\,dx\right)\left(\int_{-\infty}^{\infty} h(y)\,dy\right).$$

d) By assumption and parts (a)–(c),

$$\begin{aligned} f_{X,Y}(x, y) = g(x)h(y) &= \left(\int_{-\infty}^{\infty} g(s)\,ds\right)\left(\int_{-\infty}^{\infty} h(t)\,dt\right) g(x)h(y) \\ &= \left(\left(\int_{-\infty}^{\infty} h(t)\,dt\right) g(x)\right)\left(\left(\int_{-\infty}^{\infty} g(s)\,ds\right) h(y)\right) = f_X(x) f_Y(y). \end{aligned}$$

Hence, from Proposition 9.9 on page 523, X and Y are independent random variables.

9.96 No, not necessarily. From Exercise 9.95(c),

$$\int_{-\infty}^{\infty} g(x)\,dx = \left(\int_{-\infty}^{\infty} h(y)\,dy\right)^{-1}.$$

Referring to Proposition 8.6(b) on page 423 and Exercise 9.95(a), we see that g is a PDF of X if and only if $\int_{-\infty}^{\infty} h(y)\,dy = 1$, which, in turn, implies that h is a PDF of Y by Exercise 9.95(b). Thus, we see that g

and h are marginal PDFs of X and Y, respectively, if and only if $\int_{-\infty}^{\infty} h(y)\,dy = 1$ or, equivalently, if and only if $\int_{-\infty}^{\infty} g(x)\,dx = 1$.

9.97

a) From Equation (9.20) on page 498, a joint PDF of X and Y is

$$f_{X,Y}(x, y) = 2 \cdot (2-1) f(x) f(y) \big(F(y) - F(x)\big)^{2-2} = 2f(x)f(y)$$

if $x < y$, and $f_{X,Y}(x, y) = 0$ otherwise.
b) No. Although at first glance it might appear that the joint PDF in part (a) can be factored into a function of x alone and a function of y alone, that isn't the case. Indeed,

$$f_{X,Y}(x, y) = 2f(x)f(y)I_{\{(x,y):x<y\}}(x, y), \qquad x, y \in \mathcal{R}.$$

The function $I_{\{(x,y):x<y\}}$ can't be factored into a function of x alone and y alone.
c) No, X and Y are not independent. This fact can be seen in several ways. One way is to note that, if X and Y were independent random variables, then Proposition 9.10(a) on page 527 would imply that each conditional PDF of Y given $X = x$ does not depend on x. However, from Example 9.9, which begins on page 512,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{n(n-1)f(x)f(y)\big(F(y) - F(x)\big)^{n-2}}{nf(x)\big(1 - F(x)\big)^{n-1}}$$
$$= \frac{(n-1)f(y)\big(F(y) - F(x)\big)^{n-2}}{\big(1 - F(x)\big)^{n-1}}$$

if $y > x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. As we see, the conditional PDF of Y given $X = x$ does in fact depend on x.

9.98

a) Clearly, property (a) of Proposition 9.5 is satisfied. For property (b), we apply Proposition 8.6(b) on page 423 twice to obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x)f_Y(y)\,dx\,dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f_Y(y)\,dy\right) f_X(x)\,dx$$
$$= \int_{-\infty}^{\infty} 1 \cdot f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

b) From Proposition 9.7 on page 510,

$$f_U(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_{-\infty}^{\infty} f_X(x)f_Y(y)\,dy = f_X(x)\int_{-\infty}^{\infty} f_Y(y)\,dy = f_X(x).$$

Thus, f_X is a PDF of U. Similarly, f_Y is a PDF of V.
c) By construction, f is a joint PDF of U and V. However, from part (b),

$$f(x, y) = f_X(x)f_Y(y) = f_U(x)f_V(y).$$

Thus, from Proposition 9.9 on page 523, U and V are independent random variables.
d) We know that U and V are independent and have the same marginal distributions as X and Y, respectively. If marginal PDFs determined a joint PDF, then X and Y would have the same joint PDF as U and V. This, in turn, would imply that X and Y are independent, which isn't the case.

9.99

a) Because, by the general multiplication rule for the PDF of two continuous random variables, Equation (9.34) on page 517, the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_{Y\,|\,X}(y\,|\,x)$ is a joint PDF of X and Y. Symbolically, we have $f_{X,Y}(x, y) = f_X(x) f_{Y\,|\,X}(y\,|\,x)$.

b) A condition is that X and Y are independent. For, then, as shown by Proposition 9.9 on page 523, a joint PDF is determined from the marginal PDFs, namely, as the product of the marginal PDFs. Symbolically, we have $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

9.100

a) From Equation (9.20) on page 498 with $n = 2$, a joint PDF of X and Y is $f_{X,Y}(x, y) = 2f(x)f(y)$ if $0 < x < y$, and $f_{X,Y}(x, y) = 0$ otherwise. Thus, for $x > 0$ and $r > 0$,

$$F_{X,R}(x, r) = P(X \le x,\ R \le r) = P(X \le x,\ Y - X \le r) = \iint\limits_{\substack{s \le x \\ t-s \le r}} f_{X,Y}(s, t)\, ds\, dt$$
$$= \int_0^x \left(\int_s^{s+r} 2f(s)f(t)\, dt \right) ds = \int_0^x \left(\int_0^r 2f(s)f(u+s)\, du \right) ds.$$

Therefore, from Proposition 9.4 on page 499,

$$f_{X,R}(x, r) = \begin{cases} 2f(x)f(r+x), & \text{if } x > 0 \text{ and } r > 0; \\ 0, & \text{otherwise.} \end{cases}$$

b) Applying Proposition 9.7 on page 510, the result of part (a), and making the substitution $u = r + x$, we get

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,R}(x, r)\, dr = \int_0^{\infty} 2f(x)f(r+x)\, dr = 2f(x) \int_0^{\infty} f(r+x)\, dr$$
$$= 2f(x) \int_x^{\infty} f(u)\, du = 2f(x)\big(1 - F(x)\big)$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Also,

$$f_R(r) = \int_{-\infty}^{\infty} f_{X,R}(x, r)\, dx = \int_0^{\infty} 2f(x)f(r+x)\, dx = 2\int_0^{\infty} f(x)f(r+x)\, dx$$

if $r > 0$, and $f_R(r) = 0$ otherwise.

c) Assume $X_k \sim \mathcal{E}(\lambda)$ for $k = 1, 2$. Then, from part (a),

$$f_{X,R}(x, r) = 2f(x)f(r+x) = 2\lambda e^{-\lambda x} \lambda e^{-\lambda(r+x)} = 2\lambda^2 e^{-\lambda(r+2x)}$$

if $x > 0$ and $r > 0$, and $f_{X,R}(x, r) = 0$ otherwise. Also, from part (b),

$$f_X(x) = 2f(x)\big(1 - F(x)\big) = 2\lambda e^{-\lambda x} e^{-\lambda x} = 2\lambda e^{-2\lambda x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Moreover,

$$f_R(r) = 2\int_0^{\infty} f(x)f(r+x)\, dx = 2\int_0^{\infty} \lambda e^{-\lambda x} \lambda e^{-\lambda(r+x)}\, dx = 2\lambda^2 e^{-\lambda r} \int_0^{\infty} e^{-2\lambda x}\, dx = \lambda e^{-\lambda r}$$

if $r > 0$, and $f_R(r) = 0$ otherwise. Because $2\lambda^2 e^{-\lambda(r+2x)} = \left(2\lambda e^{-2\lambda x}\right)\left(\lambda e^{-\lambda r}\right)$, it follows from Proposition 9.9 on page 523 that X and R are independent random variables.

d) No! The random variables X and R aren't independent, for example, if the common probability distribution of X_1 and X_2 is uniform on the interval (0, 1). In that case,

$$f_{X,R}(x,r) = 2f(x)f(r+x) = 2$$

if $0 < r < 1 - x$ and $0 < x < 1$, and $f_{X,R}(x,r) = 0$ otherwise. Also,

$$f_X(x) = 2f(x)\big(1 - F(x)\big) = 2(1-x)$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise. Moreover,

$$f_R(r) = 2\int_0^\infty f(x)f(r+x)\,dx = 2\int_0^{1-r} 1\,dx = 2(1-r)$$

if $0 < r < 1$, and $f_R(r) = 0$ otherwise.

We can now proceed in several ways to show that, in this case, X and R are not independent. One way is to argue as in Example 9.14 on page 524, except use the region $S = \{\,(x,r) : 1-x < r < 1,\ 0 < x < 1\,\}$. Indeed, let f be the function defined on $\mathcal{R}^2$ by $f(x,r) = f_X(x)f_R(r)$. Then

$$f(x,r) = \begin{cases} 4(1-x)(1-r), & \text{if } 0 < x < 1 \text{ and } 0 < r < 1; \\ 0, & \text{otherwise.} \end{cases}$$

We see from the joint PDF of X and R that the probability is 0 that the random point (X, R) falls in S. If f were a joint PDF of X and R, then its double integral over S would also be 0, which it clearly isn't. Hence, f is not a joint PDF of X and R, which, by Proposition 9.9, means that X and R aren't independent.

9.101 Let X and Y denote the bid amounts by the two insurers. We want to find $P(|X - Y| < 20)$. By assumption, X and Y are independent $\mathcal{U}(2000, 2200)$ random variables. Thus, a joint PDF of X and Y is given by

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{200}\cdot\frac{1}{200} = \frac{1}{40{,}000}$$

if $2000 < x < 2200$ and $2000 < y < 2200$, and $f_{X,Y}(x,y) = 0$ otherwise. From the FPF and complementation rule,

$$\begin{aligned} P(|X-Y| < 20) &= 1 - P(|X-Y| \ge 20) = 1 - \iint\limits_{|x-y|\ge 20} f_{X,Y}(x,y)\,dx\,dy \\ &= 1 - 2\int_{2020}^{2200}\left(\int_{2000}^{x-20}\frac{1}{40{,}000}\,dy\right)dx = 1 - \frac{1}{20{,}000}\int_{2020}^{2200}(x-2020)\,dx \\ &= 1 - 0.81 = 0.19. \end{aligned}$$

Note that we can also obtain the required probability by using the fact that we have a uniform probability model on the square $S = (2000, 2200) \times (2000, 2200)$. Then,

$$\begin{aligned} P(|X-Y| < 20) &= \frac{\big|\{|X-Y| < 20\}\big|}{|S|} = \frac{\big|\{\,(x,y) : |x-y| < 20\,\} \cap S\big|}{|S|} \\ &= \frac{|S| - \big|\{\,(x,y) : |x-y| \ge 20\,\} \cap S\big|}{|S|} = 1 - 2\cdot\frac{(180)(180)/2}{(200)(200)} = 0.19. \end{aligned}$$

9.102 Let X and Y denote the waiting times for the first claim by a good driver and bad driver, respectively. By assumption, X and Y are independent random variables with $X \sim \mathcal{E}(1/6)$ and $Y \sim \mathcal{E}(1/3)$.
a) We have

$$\begin{aligned} P(X \le 3,\ Y \le 2) &= P(X \le 3)P(Y \le 2) = \left(1 - e^{-(1/6)\cdot 3}\right)\left(1 - e^{-(1/3)\cdot 2}\right) \\ &= 1 - e^{-2/3} - e^{-1/2} + e^{-7/6} = 0.191. \end{aligned}$$

b) By independence,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \left((1/6)e^{-(1/6)x}\right)\left((1/3)e^{-(1/3)y}\right) = \frac{1}{18}e^{-(1/6)x}e^{-(1/3)y}$$

if $x > 0$ and $y > 0$, and $f_{X,Y}(x, y) = 0$ otherwise. Applying the FPF now gives

$$\begin{aligned} P(X < Y) &= \iint_{x<y} f_{X,Y}(x, y)\, dx\, dy = \frac{1}{18}\int_0^\infty \left(\int_x^\infty e^{-(1/3)y}\, dy\right) e^{-(1/6)x}\, dx \\ &= \frac{1}{6}\int_0^\infty e^{-(1/3)x}e^{-(1/6)x}\, dx = \frac{1}{6}\int_0^\infty e^{-(1/2)x}\, dx = \frac{1}{3}. \end{aligned}$$

c) Recalling that X and Y are independent random variables, we get

$$P(Y > 4 \mid X = 4) = P(Y > 4) = e^{-(1/3)\cdot 4} = 0.264.$$

9.103
a) By symmetry, each of the 3! possible orderings of X, Y, and Z are equally likely. Consequently, we have $P(X < Y < Z) = 1/3! = 1/6$.
b) From independence,

$$f_{X,Y,Z}(x, y, z) = f_X(x) f_Y(y) f_Z(z) = f(x) f(y) f(z).$$

Applying the FPF and making the successive substitutions $u = F(y)$ and $v = F(z)$, we get

$$\begin{aligned} P(X < Y < Z) &= \iiint_{x<y<z} f_{X,Y,Z}(x, y, z)\, dx\, dy\, dz \\ &= \int_{-\infty}^\infty \left(\int_{-\infty}^z \left(\int_{-\infty}^y f(x)\, dx\right) f(y)\, dy\right) f(z)\, dz \\ &= \int_{-\infty}^\infty \left(\int_{-\infty}^z F(y) f(y)\, dy\right) f(z)\, dz = \int_{-\infty}^\infty \left(\int_0^{F(z)} u\, du\right) f(z)\, dz \\ &= \frac{1}{2}\int_{-\infty}^\infty \left(F(z)\right)^2 f(z)\, dz = \frac{1}{2}\int_0^1 v^2\, dv = \frac{1}{6}. \end{aligned}$$

c) By symmetry, each of the $n!$ possible orderings of $X_1, \ldots, X_n$ are equally likely. Consequently, we have $P\left(X_{i_1} < X_{i_2} < \cdots < X_{i_n}\right) = 1/n!$.

9.104
a) By assumption, $(X, Y) \sim \mathcal{U}(R)$; that is,

$$f_{X,Y}(x, y) = \frac{1}{(b-a)(d-c)}$$

if $(x, y) \in R$, and $f_{X,Y}(x, y) = 0$ otherwise. Also,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = \int_c^d \frac{dy}{(b-a)(d-c)} = \frac{1}{b-a}$$

if $a < x < b$, and $f_X(x) = 0$ otherwise. Thus, $X \sim \mathcal{U}(a, b)$. Moreover,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_a^b \frac{dx}{(b-a)(d-c)} = \frac{1}{d-c}$$

if $c < y < d$, and $f_Y(y) = 0$ otherwise. Thus, $Y \sim \mathcal{U}(c, d)$. It now follows that

$$f_X(x) f_Y(y) = \left(\frac{1}{b-a}\right)\left(\frac{1}{d-c}\right) = \frac{1}{(b-a)(d-c)}$$

if $(x, y) \in R$, and $f_X(x) f_Y(y) = 0$ otherwise. Consequently, we see that the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_Y(y)$ is a joint PDF of X and Y. Hence, by Proposition 9.9 on page 523, X and Y are independent random variables.
b) Let U and V be independent $\mathcal{U}(0, 1)$ random variables. From Proposition 8.16(b), we know that $X = a + (b-a)U \sim \mathcal{U}(a, b)$ and $Y = c + (d-c)V \sim \mathcal{U}(c, d)$. Furthermore, from Proposition 6.9 on page 291, X and Y are independent random variables. It now follows from part (a) that the random point $\big(a + (b-a)U, c + (d-c)V\big) \sim \mathcal{U}(R)$. Thus, to simulate the random selection of a point from R, use a basic random number generator to obtain two independent uniform numbers between 0 and 1, say, u and v, and then calculate the point $\big(a + (b-a)u, c + (d-c)v\big)$.

9.105 By assumption, the positions of the three gas stations, X, Y, and Z, are independent $\mathcal{U}(0, 1)$ random variables. Hence,

$$f_{X,Y,Z}(x, y, z) = f_X(x) f_Y(y) f_Z(z) = 1 \cdot 1 \cdot 1 = 1$$

if $0 < x < y < z < 1$, and $f_{X,Y,Z}(x, y, z) = 0$ otherwise.
a) The required probability is, by symmetry, $3!$ times $P(Y > X + 1/3,\ Z > Y + 1/3)$. We evaluate the required probability by using the FPF and by first making the substitution $u = 2/3 - y$ and then the substitution $v = 1/3 - x$, thus:

$$\begin{aligned}
&6P(Y > X + 1/3,\ Z > Y + 1/3) \\
&\quad = 6 \iiint\limits_{\substack{y > x+1/3 \\ z > y+1/3}} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz = 6\int_0^{1/3}\left(\int_{x+1/3}^{2/3}\left(\int_{y+1/3}^{1} 1\,dz\right)dy\right)dx \\
&\quad = 6\int_0^{1/3}\left(\int_{x+1/3}^{2/3} (2/3 - y)\,dy\right)dx = 6\int_0^{1/3}\left(\int_0^{1/3-x} u\,du\right)dx \\
&\quad = 3\int_0^{1/3} (1/3 - x)^2\,dx = 3\int_0^{1/3} v^2\,dv = \frac{1}{27}.
\end{aligned}$$

b) We must have $2d \le 1$, or $d \le 1/2$.
c) For $d \le 1/2$, the required probability is, by symmetry, 3! times $P(Y > X + d,\ Z > Y + d)$. We evaluate the required probability by using the FPF and by first making the substitution $u = 1 - d - y$ and then the substitution $v = 1 - 2d - x$, thus:

$$\begin{aligned}
&P(Y > X + d,\ Z > Y + d)\\
&= \iiint\limits_{\substack{y>x+d\\ z>y+d}} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz = 6\int_0^{1-2d}\left(\int_{x+d}^{1-d}\left(\int_{y+d}^{1} 1\,dz\right)dy\right)dx\\
&= 6\int_0^{1-2d}\left(\int_{x+d}^{1-d}(1-d-y)\,dy\right)dx = 6\int_0^{1-2d}\left(\int_0^{1-2d-x} u\,du\right)dx\\
&= 3\int_0^{1-2d}(1-2d-x)^2\,dx = 3\int_0^{1-2d} v^2\,dv\\
&= (1-2d)^3.
\end{aligned}$$

9.106
a) Let $X = X_k$ and let $Y = \min_{j\ne k}\{X_j,\ 1 \le j \le n\}$. From Proposition 6.13 on page 297, X and Y are independent and, from Example 9.9(a) on page 512, $f_Y(y) = (n-1)f(y)\bigl(1 - F(y)\bigr)^{n-2}$. We evaluate the required probability by using the FPF and by first making the substitution $u = 1 - F(y)$ and then the substitution $v = 1 - F(x)$, thus:

$$\begin{aligned}
P(X_k = \min\{X_1, \dots, X_n\}) &= P(X < Y)\\
&= \iint\limits_{x<y} f_{X,Y}(x, y)\,dx\,dy = \iint\limits_{x<y} f_X(x)f_Y(y)\,dx\,dy\\
&= \iint\limits_{x<y} f(x)(n-1)f(y)\bigl(1 - F(y)\bigr)^{n-2}\,dx\,dy\\
&= \int_{-\infty}^{\infty} f(x)\left(\int_x^{\infty}(n-1)f(y)\bigl(1 - F(y)\bigr)^{n-2}\,dy\right)dx\\
&= \int_{-\infty}^{\infty} f(x)\left(\int_0^{1-F(x)}(n-1)u^{n-2}\,du\right)dx\\
&= \int_{-\infty}^{\infty} f(x)\bigl(1 - F(x)\bigr)^{n-1}\,dx = \int_0^1 v^{n-1}\,dv\\
&= \frac{1}{n}.
\end{aligned}$$

b) By symmetry, each of the X_js, of which there are n, is equally likely to be the minimum of the random sample. Hence, the probability that the minimum is X_k equals $1/n$.

9.107
a) Let $X = X_k$ and let $Y = \max_{j\ne k}\{X_j,\ 1 \le j \le n\}$. From Proposition 6.13 on page 297, X and Y are independent and, from Exercise 9.70(a) on page 520, $f_Y(y) = (n-1)f(y)\bigl(F(y)\bigr)^{n-2}$. We evaluate

the required probability by using the FPF and by first making the substitution $u = F(y)$ and then the substitution $v = F(x)$, thus:

$$\begin{aligned}
P(X_k = \max\{X_1, \dots, X_n\}) &= P(X > Y) \\
&= \iint\limits_{x>y} f_{X,Y}(x, y)\, dx\, dy = \iint\limits_{x>y} f_X(x) f_Y(y)\, dx\, dy \\
&= \iint\limits_{x>y} f(x)(n-1) f(y) \left(F(y)\right)^{n-2}\, dx\, dy \\
&= \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{x} (n-1) f(y) \left(F(y)\right)^{n-2}\, dy \right) dx \\
&= \int_{-\infty}^{\infty} f(x) \left(\int_{0}^{F(x)} (n-1) u^{n-2}\, du \right) dx \\
&= \int_{-\infty}^{\infty} f(x) \left(F(x)\right)^{n-1}\, dx = \int_0^1 v^{n-1}\, dv \\
&= \frac{1}{n}.
\end{aligned}$$

b) By symmetry, each of the X_js, of which there are n, is equally likely to be the maximum of the random sample. Hence, the probability that the maximum is X_k equals $1/n$.

Theory Exercises

9.108 Continuous random variables, $X_1, \dots, X_m$, with a joint PDF are independent if and only if the function f defined on $\mathcal{R}^m$ by $f(x_1, \dots, x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$ is a joint PDF of $X_1, \dots, X_m$. We write this condition symbolically as

$$f_{X_1,\dots,X_m}(x_1, \dots, x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m), \qquad x_1, \dots, x_m \in \mathcal{R}.$$

To prove this result, we proceed as in the proof of the bivariate case. Suppose that $X_1, \dots, X_m$ are independent random variables. Then, for each $x_1, \dots, x_m \in \mathcal{R}$,

$$\begin{aligned}
F_{X_1,\dots,X_m}(x_1, \dots, x_m) &= P(X_1 \le x_1, \dots, X_m \le x_m) = P(X_1 \le x_1) \cdots P(X_m \le x_m) \\
&= \left(\int_{-\infty}^{x_1} f_{X_1}(s_1)\, ds_1 \right) \cdots \left(\int_{-\infty}^{x_m} f_{X_m}(s_m)\, ds_m \right) \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_{X_1}(s_1) \cdots f_{X_m}(s_m)\, ds_1 \cdots ds_m.
\end{aligned}$$

Consequently, from the multivariate analogue of Proposition 9.4 on page 499 [see Exercise 9.29(c)], the function f defined on $\mathcal{R}^m$ by $f(x_1, \dots, x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$ is a joint PDF of $X_1, \dots, X_m$.

Conversely, suppose that the function f defined on $\mathcal{R}^m$ by $f(x_1, \dots, x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$ is a joint PDF of $X_1, \dots, X_m$. Let $A_1, \dots, A_m$ be any m subsets of real numbers. Applying the FPFs for multivariate and univariate continuous random variables, we get

$$
\begin{aligned}
&P(X_1 \in A_1, \ldots, X_m \in A_m) \\
&\quad = P\big((X_1, \ldots, X_m) \in A_1 \times \cdots \times A_m\big) = \int \cdots \int_{A_1 \times \cdots \times A_m} f(x_1, \ldots, x_m)\, dx_1 \cdots dx_m \\
&\quad = \int \cdots \int_{A_1 \times \cdots \times A_m} f_{X_1}(x_1) \cdots f_{X_m}(x_m)\, dx_1 \cdots dx_m = \int_{A_m} \left(\cdots \left(\int_{A_1} f_{X_1}(x_1)\, dx_1 \right) \cdots \right) f_{X_m}(x_m)\, dx_m \\
&\quad = \left(\int_{A_1} f_{X_1}(x_1)\, dx_1 \right) \cdots \left(\int_{A_m} f_{X_m}(x_m)\, dx_m \right) = P(X_1 \in A_1) \cdots P(X_m \in A_m).
\end{aligned}
$$

Thus $X_1, \ldots, X_m$ are independent random variables.

9.109

a) Using independence,

$$
\begin{aligned}
F_{X,Y}(x, y) &= P(X \le x,\ Y \le y) = P(X \le x)P(Y \le y) \\
&= \left(\int_{-\infty}^{x} f_X(s)\, ds \right) \left(\int_{-\infty}^{y} f_Y(t)\, dt \right) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(s) f_Y(t)\, ds\, dt.
\end{aligned}
$$

Hence, from Proposition 9.4 on page 499, the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_Y(y)$ is a joint PDF of X and Y.

b) No. Let X and Y be the x and y coordinates, respectively, of a point selected at random from the diagonal of the unit square, that is, from $\{\,(x, y) \in \mathcal{R}^2 : y = x,\ 0 < x < 1\,\}$. Then both X and Y are uniform on the interval (0, 1) and hence each has a PDF. However, X and Y doesn't have a joint PDF because $P(X = Y) = 1 > 0$.

c) No. In the petri-dish illustration of Example 9.14 on page 524, X and Y have a joint PDF but aren't independent.

9.110

a) Using the FPF and independence, and then making the substitution $u = y + x$, we get

$$
\begin{aligned}
F_{X+Y}(z) &= P(X + Y \le z) = \iint_{x+y \le z} f_{X,Y}(x, y)\, dx\, dy = \iint_{x+y \le z} f_X(x) f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_Y(y)\, dy \right) f_X(x)\, dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(u - x)\, du \right) f_X(x)\, dx \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_X(x) f_Y(u - x)\, dx \right) du.
\end{aligned}
$$

Therefore, from Proposition 8.5 on page 422,

$$
f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx.
$$

b) We can proceed as in part (a) to find the PDF of $X - Y$. Alternatively, we note that

$$F_{-Y}(y) = P(-Y \le y) = P(Y \ge -y) = 1 - P(Y < -y) = 1 - F_Y(-y)$$

so that $f_{-Y}(y) = f_Y(-y)$. Now applying the result of part (a) with Y replaced by $-Y$, we get

$$f_{X-Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x - z)\, dx.$$

Advanced Exercises

9.111 We begin by noting that $|S| = |A||B|$; $S_x = B$ for $x \in A$, and $S_x = \emptyset$ otherwise; $S^y = A$ for $y \in B$, and $S^y = \emptyset$ otherwise. From Exercise 9.40, a joint PDF of X and Y is

$$f_{X,Y}(x, y) = \frac{1}{|S|} = \frac{1}{|A||B|}, \qquad x \in A,\ y \in B,$$

and $f_{X,Y}(x, y) = 0$ otherwise. From Exercise 9.81,

$$f_X(x) = \frac{|S_x|}{|S|} = \frac{|B|}{|A||B|} = \frac{1}{|A|}, \qquad x \in A,$$

and $f_X(x) = 0$ otherwise; and

$$f_Y(y) = \frac{|S^y|}{|S|} = \frac{|A|}{|A||B|} = \frac{1}{|B|}, \qquad y \in B,$$

and $f_Y(y) = 0$ otherwise. From the three displays, we conclude that the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x) f_Y(y)$ is a joint PDF of X and Y. Hence, by Proposition 9.9 on page 523, the random variables X and Y are independent.

9.112 Suppose that X and Y are independent. Then,

$$P(a \le X \le b,\ Y = y) = P(a \le X \le b) P(Y = y) = \left(\int_a^b f_X(x)\, dx \right) p_Y(y) = \int_a^b f_X(x) p_Y(y)\, dx.$$

Thus the function h defined on $\mathcal{R}^2$ by $h(x, y) = f_X(x) p_Y(y)$ is a joint PDF/PMF of X and Y. Conversely, suppose that the function h defined on $\mathcal{R}^2$ by $h(x, y) = f_X(x) p_Y(y)$ is a joint PDF/PMF of X and Y. Let A and B be two subsets of real numbers. Then

$$\begin{aligned} P(X \in A,\ Y \in B) &= \sum_{y \in B} P(X \in A,\ Y = y) = \sum_{y \in B} \int_A h_{X,Y}(x, y)\, dx \\ &= \sum_{y \in B} \int_A f_X(x) p_Y(y)\, dx = \left(\int_A f_X(x)\, dx \right) \left(\sum_{y \in B} p_Y(y) \right) \\ &= P(X \in A) P(Y \in B). \end{aligned}$$

Hence, X and Y are independent random variables.

9.113 From Exercise 9.59, $X \sim \mathcal{E}(\lambda)$ and Y has the Pascal distribution (as defined in Exercise 7.26) with parameter $1/\lambda$. Thus,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad p_Y(y) = \begin{cases} \dfrac{(1/\lambda)^y}{(1 + 1/\lambda)^{y+1}}, & y = 0, 1, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

It appears that the function h defined on $\mathcal{R}^2$ by $h(x, y) = f_X(x)p_Y(y)$ is not a joint PDF/PMF of X and Y and, hence, in view of Exercise 9.112, that X and Y are not independent random variables. To verify, note, for instance, that

$$\begin{aligned} P(X > 1,\ Y = 0) &= \int_1^\infty h_{X,Y}(x, 0)\,dx = \int_1^\infty \lambda e^{-(1+\lambda)x}\,dx \\ &= \frac{\lambda}{1+\lambda}e^{-(1+\lambda)} = e^{-1}e^{-\lambda}\frac{1}{1+1/\lambda} \\ &\neq e^{-\lambda}\frac{1}{1+1/\lambda} = P(X > 1)P(Y = 0). \end{aligned}$$

9.114 From Exercise 9.84, $X \sim \mathcal{E}(1)$ and $Y \sim \mathcal{B}(1, 1/2)$. Hence,

$$f_X(x) = \begin{cases} e^{-x}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad p_Y(y) = \begin{cases} 1/2, & \text{if } y = 0 \text{ or } 1; \\ 0, & \text{otherwise.} \end{cases}$$

Note that

$$f_X(x)p_Y(y) = \begin{cases} \dfrac{1}{2}e^{-x}, & \text{if } x > 0 \text{ and } y = 0 \text{ or } 1; \\ 0, & \text{otherwise.} \end{cases} = h_{X,Y}(x, y).$$

Thus, the function h defined on $\mathcal{R}^2$ by $h(x, y) = f_X(x)p_Y(y)$ is a joint PDF/PMF of X and Y and, hence, in view of Exercise 9.112, X and Y are independent random variables.

9.6 Functions of Two or More Continuous Random Variables

Basic Exercises

9.115

a) By independence,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2} \cdot \frac{1}{\sqrt{2\pi}\sigma}e^{-y^2/2\sigma^2} = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$$

for all $x, y \in \mathcal{R}$. For $r > 0$, we use the FPF and change to polar coordinates to get

$$\begin{aligned} F_R(r) &= P(R \le r) = P\left(\sqrt{X^2 + Y^2} \le r\right) = \iint\limits_{\sqrt{x^2+y^2} \le r} f_{X,Y}(x, y)\,dx\,dy \\ &= \frac{1}{2\pi\sigma^2}\iint\limits_{\sqrt{x^2+y^2} \le r} e^{-(x^2+y^2)/2\sigma^2}\,dx\,dy = \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\left(\int_0^r e^{-u^2/2\sigma^2}u\,du\right)d\theta \\ &= \frac{1}{\sigma^2}\int_0^r e^{-u^2/2\sigma^2}u\,du. \end{aligned}$$

Therefore, from Propostion 8.5 on page 422,

$$f_R(r) = \frac{r}{\sigma^2}e^{-r^2/2\sigma^2}$$

if $r > 0$, and $f_R(r) = 0$ otherwise.

b) Because $X \sim \mathcal{N}(0, \sigma^2)$, we know that $X/\sigma \sim \mathcal{N}(0, 1)$ and, hence, that X^2/σ^2 has the chi-square distribution with one degree of freedom, as does Y^2/σ^2. Referring now to the second bulleted item on page 537, we conclude that $W = (X^2 + Y^2)/\sigma^2$ has the chi-square distribution with two degrees of freedom:

$$f_W(w) = \frac{(1/2)^{(2/2)}}{\Gamma(2/2)} w^{(2/2)-1} e^{-w/2} = \frac{1}{2} e^{-w/2}$$

if $w > 0$, and $f_W(w) = 0$ otherwise. Noting that $R = \sigma\sqrt{W}$, we apply the univariate transformation theorem with $g(w) = \sigma\sqrt{w}$ to conclude that

$$f_R(r) = \frac{1}{|g'(w)|} f_W(w) = \frac{1}{\sigma/2\sqrt{w}} \cdot \frac{1}{2} e^{-w/2} = \frac{2\sqrt{r^2/\sigma^2}}{\sigma} \cdot \frac{1}{2} e^{-r^2/2\sigma^2} = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}$$

if $r > 0$, and $f_R(r) = 0$ otherwise.

9.116

a) Let $Z = X + Y$. Referring to Equation (9.44) on page 534 and then to Equation (9.20), we get

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx = n(n-1) \int_{-\infty}^{z/2} f(x) f(z - x) \big(F(z - x) - F(x)\big)^{n-2}\, dx.$$

Noting that $M = Z/2$, we apply the univariate transformation theorem with $g(z) = z/2$ to conclude that

$$\begin{aligned} f_M(m) &= \frac{1}{|g'(z)|} f_Z(z) = \frac{1}{|1/2|} f_Z(z) = 2 f_Z(2m) \\ &= 2n(n-1) \int_{-\infty}^{m} f(x) f(2m - x) \big(F(2m - x) - F(x)\big)^{n-2}\, dx. \end{aligned}$$

b) For a $\mathcal{U}(0, 1)$ distribution,

$$F(t) = \begin{cases} 0, & \text{if } t < 0; \\ t, & \text{if } 0 \le t < 1; \\ 1, & \text{if } t \ge 1. \end{cases} \qquad \text{and} \qquad f(t) = \begin{cases} 1, & \text{if } 0 < t < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Referring now to the formula for f_M in part (a), we note that the integrand is nonzero if and only if $0 < x < 1$ and $0 < 2m - x < 1$. Keeping in mind that, in this case, $0 < m < 1$, we see that the integrand is nonzero if and only if $x > \max\{2m - 1, 0\}$. For convenience, let $a = \max\{2m - 1, 0\}$. Then, upon making the substitution $u = m - x$, we get

$$\begin{aligned} f_M(m) &= 2n(n-1) \int_a^m \big((2m - x) - x\big)^{n-2}\, dx \\ &= 2^{n-1} n(n-1) \int_0^{m-a} u^{n-2}\, du = n2^{n-1}(m - a)^{n-1}. \end{aligned}$$

Now,

$$m - a = m - \max\{2m - 1, 0\} = \begin{cases} m, & \text{if } m < 1/2; \\ 1 - m, & \text{if } m > 1/2. \end{cases}$$

Therefore,

$$f_M(m) = \begin{cases} n(2m)^{n-1}, & \text{if } 0 < m < /2; \\ n(2 - 2m)^{n-1}, & \text{if } 1/2 < m < 1; \\ 0, & \text{otherwise.} \end{cases}$$

9.117 Let Y denote system lifetime, so that $Y = \max\{X_1, \ldots, X_n\}$. Then $F_Y(y) = 0$ if $y < 0$ and, for $y > 0$,

$$F_Y(y) = P(Y \le y) = P(\max\{X_1, \ldots, X_n\} \le y) = P(X_1 \le y, \ldots, X_n \le y)$$
$$= P(X_1 \le y) \cdots P(X_n \le y) = \prod_{k=1}^{n} \left(1 - e^{-\lambda_k y}\right).$$

Differentiation now yields

$$f_Y(y) = F_Y'(y) = \sum_{k=1}^{n} \left(\prod_{j \ne k} \left(1 - e^{-\lambda_j y}\right) \lambda_k e^{-\lambda_k y} \right) = \left(\prod_{k=1}^{n} \left(1 - e^{-\lambda_k y}\right) \right) \left(\sum_{k=1}^{n} \frac{\lambda_k e^{-\lambda_k y}}{1 - e^{-\lambda_k y}} \right)$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

9.118 Let X and Y denote the lifetimes of the original component and the spare, respectively. By assumption, both X and Y have an $\mathcal{E}(\lambda)$ distribution and, furthermore, we can assume that they are independent random variables. Thus,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda y} = \lambda^2 e^{-\lambda(x+y)}$$

if $x > 0$ and $y > 0$, and $f_{X,Y}(x, y) = 0$ otherwise. We want to find the CDF of the random variable $R = X/(X + Y)$. For $0 < r < 1$,

$$F_R(r) = P(R \le r) = P\left(\frac{X}{X+Y} \le r\right) = P\left(Y \ge (1-r)X/r\right) = \iint_{y \ge (1-r)x/r} f_{X,Y}(x, y)\, dx\, dy$$
$$= \int_0^\infty \left(\int_{(1-r)x/r}^\infty \lambda^2 e^{-\lambda(x+y)}\, dy \right) dx = \int_0^\infty \left(\int_{(1-r)x/r}^\infty \lambda e^{-\lambda y}\, dy \right) \lambda e^{-\lambda x}\, dx$$
$$= \lambda \int_0^\infty e^{-\lambda(1-r)x/r} e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda/r)x}\, dx = r.$$

Thus, $R \sim \mathcal{U}(0, 1)$.

9.119 From the fact that X and Y are independent $\mathcal{E}(\lambda)$ random variables, we get, for $z > 0$,

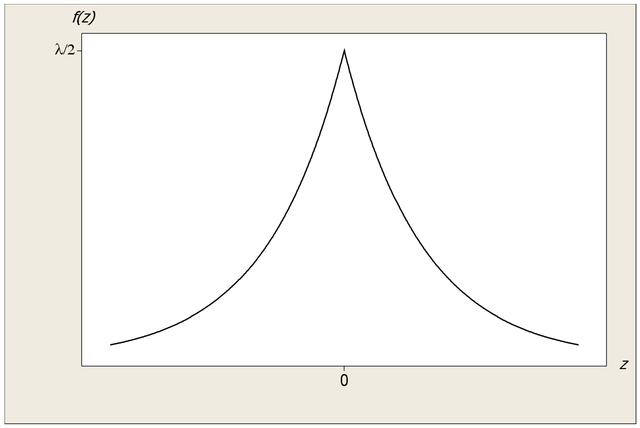
$$F_{X-Y}(z) = P(X - Y \le z) = \iint_{x-y \le z} f_{X,Y}(x, y)\, dx\, dy = \iint_{x-y \le z} f_X(x) f_Y(y)\, dx\, dy$$
$$= \int_0^\infty \left(\int_0^{y+z} \lambda e^{-\lambda x}\, dx \right) \lambda e^{-\lambda y}\, dy = \int_0^\infty \left(1 - e^{-\lambda(y+z)}\right) \lambda e^{-\lambda y}\, dy$$
$$= \int_0^\infty \lambda e^{-\lambda y}\, dy - \lambda e^{-\lambda z} \int_0^\infty e^{-2\lambda z}\, dy = 1 - \frac{1}{2} e^{-\lambda z}.$$

Therefore, $f_{X-Y}(z) = (\lambda/2) e^{-\lambda z}$ if $z > 0$. We can do a similar computation to determine $f_{X-Y}(z)$ when $z < 0$. Alternatively, we can use symmetry, as discussed in Exercise 8.204 on page 482. Thus, for $z < 0$, we have

$$f_{X-Y}(z) = f_{X-Y}(-z) = (\lambda/2) e^{-\lambda(-z)} = (\lambda/2) e^{-\lambda|z|}.$$

Consequently, $f_{X-Y}(z) = (\lambda/2) e^{-\lambda|z|}$ for all $z \in \mathcal{R}$.

Note: In the following graph, we use $f(z)$ instead of $f_{X-Y}(z)$.



9.120

a) From the FPF,

$$P(X+Y \le 1) = \iint_{x+y\le 1} f_{X,Y}(x,y)\,dx\,dy = \int_0^1 \left(\int_0^{1-x} e^{-(x+y)}\,dy\right) dx$$
$$= \int_0^1 \left(\int_0^{1-x} e^{-y}\,dy\right) e^{-x}\,dx = \int_0^1 \left(1 - e^{-(1-x)}\right) e^{-x}\,dx$$
$$= \int_0^1 \left(e^{-x} - e^{-1}\right) dx = 1 - 2e^{-1}.$$

b) We note that $f_{X,Y}(x,y) = e^{-x}e^{-y}$ for $x > 0$ and $y > 0$, and $f_{X,Y}(x,y) = 0$ otherwise. It follows easily that X and Y are independent $\mathcal{E}(1)$ random variables or, equivalently, independent $\Gamma(1,1)$ random variables. Therefore, by Proposition 9.13, $X+Y \sim \Gamma(2,1)$. Applying now Equation (8.49) with $r=2$ and $\lambda = 1$ gives

$$P(X+Y \le 1) = F_{X+Y}(1) = 1 - e^{-1}(1+1) = 1 - 2e^{-1}.$$

9.121 Let X and Y denote the times, in days, until the next basic and deluxe policy claims, respectively. By assumption, X and Y are independent $\mathcal{E}(1/2)$ and $\mathcal{E}(1/3)$ random variables.

a) From the FPF and independence,

$$P(Y < X) = \iint_{y<x} f_{X,Y}(x,y)\,dx\,dy = \iint_{y<x} f_X(x) f_Y(y)\,dx\,dy$$
$$= \int_0^\infty \left(\int_y^\infty (1/2)e^{-x/2}\,dx\right)(1/3)e^{-y/3}\,dy = \frac{1}{3}\int_0^\infty e^{-y/2}e^{-y/3}\,dy$$
$$= \frac{1}{3}\int_0^\infty e^{-5y/6}\,dy = \frac{2}{5}.$$

b) From Proposition 9.11(b),

$$P(Y < X) = P\big(Y = \min\{X,Y\}\big) = \frac{1/3}{1/2 + 1/3} = \frac{2}{5}.$$

9.122 If follows from Exercise 8.73 on page 435 that $2T_1 \sim \mathcal{E}(1/2)$. Applying now Equation (9.46) on page 534 yields

$$f_{2T_1+T_2}(t) = \int_{-\infty}^{\infty} f_{2T_1}(s) f_{T_2}(t-s)\,ds = \int_0^t (1/2)e^{-s/2}e^{-(t-s)}\,ds = \frac{1}{2}e^{-t}\int_0^t e^{s/2}\,ds = e^{-t/2} - e^{-t}$$

if $t > 0$, and $f_{2T_1+T_2}(t) = 0$ otherwise.

9.123 From the form of the PDFs, we see that $X \sim \Gamma(\alpha+1, \lambda)$ and $Y \sim \Gamma(\beta+1, \lambda)$. As X and Y are independent, we can now apply Proposition 9.13 on page 537 to conclude that $X + Y \sim \Gamma(\alpha+\beta+2, \lambda)$.

9.124 We apply Proposition 9.14 on page 540.
a) $X + Y \sim \mathcal{N}(2+3, 1+2) \sim \mathcal{N}(5, 3)$
b) $X - Y \sim \mathcal{N}(2-3, 1+2) \sim \mathcal{N}(-1, 3)$
c) $2 - 3X + 4Y \sim \mathcal{N}(2 - 3\cdot 2 + 4\cdot 3, 3^2\cdot 1 + 4^2\cdot 2) \sim \mathcal{N}(8, 41)$

9.125
a) From Proposition 9.14 on page 540, $X_1 + \cdots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$. The standardized version is therefore,

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}},$$

which, by Proposition 8.10 on page 441, has the standard normal distribution.
b) We have

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n,$$

which, by Proposition 9.14, has a normal distribution with parameters

$$\frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = \mu \qquad \text{and} \qquad \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}.$$

9.126 Let X and Y denote the two measurements, and set $M = (X+Y)/2$. By assumption, we have $X - h \sim \mathcal{N}\big(0, (0.0056h)^2\big)$ and $Y - h \sim \mathcal{N}\big(0, (0.0044h)^2\big)$. From Proposition 9.14 on page 540, then, $X \sim \mathcal{N}\big(h, (0.0056h)^2\big)$ and $Y \sim \mathcal{N}\big(h, (0.0044h)^2\big)$. Again, from Proposition 9.14,

$$M = \frac{1}{2}X + \frac{1}{2}Y \sim \mathcal{N}\left(\frac{1}{2}h + \frac{1}{2}h, \frac{1}{4}(0.0056h)^2 + \frac{1}{4}(0.0044h)^2\right) \sim \mathcal{N}\Big(h, 0.00001268h^2\Big).$$

Therefore,

$$\begin{aligned} P(|M-h| \le 0.005h) &= P(h - 0.005h \le M \le h + 0.005h) \\ &= \Phi\left(\frac{(h+0.005h)-h}{\sqrt{0.00001268}\,h}\right) - \Phi\left(\frac{(h-0.005h)-h}{\sqrt{0.00001268}\,h}\right) \\ &= 2\Phi\left(\frac{0.005}{\sqrt{0.00001268}}\right) - 1 = 0.840. \end{aligned}$$

9.127 Let n be the number of light bulbs purchased and let $X_1, \ldots, X_n$ be their lifetimes. By assumption, $X_1, \ldots, X_n$ are independent $\mathcal{N}(3, 1)$ random variables. The total lifetime of all n bulbs is $S_n = X_1 + \cdots + X_n$, which, from Proposition 9.14 on page 540, has an $\mathcal{N}(3n, n)$ distribution. We want to choose the smallest n such that $P(S_n \geq 40) \geq 0.9772$. Thus,

$$0.9772 \leq P(S_n \geq 40) = 1 - \Phi\left(\frac{40 - 3n}{\sqrt{n}}\right),$$

or $\big((40 - 3n)/\sqrt{n}\big) \leq \Phi^{-1}(0.0228) = -2.00$, or $3n - 2\sqrt{n} - 40 \geq 0$. However, the term on the left of this last inequality factors as $\big(\sqrt{n} - 4\big)\big(3\sqrt{n} + 10\big)$. Thus, $\sqrt{n} = 4$, or $n = 16$.

9.128

a) From Proposition 6.13 on page 297 and the definition of a lognormal random variable, we know that $\ln X_1, \ldots, \ln X_m$ are independent normal random variables. Therefore, from Proposition 9.14 on page 540,

$$\ln\left(\prod_{j=1}^{m} X_j^{a_j}\right) = \sum_{j=1}^{m} a_j \ln X_j$$

is normally distributed. Hence, $\prod_{j=1}^{m} X_j^{a_j}$ is lognormal.

b) The required result follows immediately from part (a) by setting $a_j = 1$ for all $j = 1, 2, \ldots, m$.

9.129 Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be the lifetimes of the n type A batteries and m type B batteries, respectively. We can reasonably assume that $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ are independent random variables. Set $X = \min\{X_1, \ldots, X_n\}$ and $Y = \min\{Y_1, \ldots, Y_m\}$. The problem is to determine $P(X < Y)$. From Proposition 6.13 on page 297, X and Y are independent random variables, and, from Proposition 9.11(a) on page 532, $X \sim \mathcal{E}(n\lambda)$ and $Y \sim \mathcal{E}(m\mu)$. Therefore, from Proposition 9.11(b),

$$P(X < Y) = P\big(X = \min\{X, Y\}\big) = \frac{n\lambda}{n\lambda + m\mu}.$$

9.130 Let X and Y denote annual premiums and annual claims, respectively. We know that $X \sim \mathcal{E}(1/2)$, $Y \sim \mathcal{E}(1)$, and X and Y are independent random variables. Referring to Equation (9.49) on page 540, we get

$$\begin{aligned} f_{Y/X}(z) &= \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) = \int_0^{\infty} x(1/2)e^{-x/2}e^{-xz}\,dx = \frac{1}{2}\int_0^{\infty} x^{2-1}e^{-(1/2+z)x}\,dx \\ &= \frac{1}{2}\frac{\Gamma(2)}{(1/2+z)^2}\int_0^{\infty}\frac{(1/2+z)^2}{\Gamma(2)}x^{2-1}e^{-(1/2+z)x}\,dx = \frac{1}{2(1/2+z)^2}\cdot 1 = \frac{2}{(1+2z)^2} \end{aligned}$$

if $z > 0$, and $f_{Y/X}(z) = 0$ otherwise.

9.131

a) Answers will vary.

b) Answers will vary.

c) From Example 9.19 on page 535, the sum of two independent $\mathcal{U}(0, 1)$ random variables has a $\mathcal{T}(0, 2)$ distribution. Hence, we would expect the histogram to be shaped roughly like an isosceles triangle whose base connects the two points $(0, 0)$ and $(2, 0)$.

d) Answers will vary.

9.132

a) Answers will vary.

b) Answers will vary.

c) From Example 9.22 on page 539, the sum of independent $\mathcal{N}(-2, 9)$ and $\mathcal{N}(3, 16)$ random variables has an $\mathcal{N}(1, 25)$ distribution. Hence, we would expect the histogram to be shaped roughly like that of a

normal curve with parameters $\mu = 1$ and $\sigma^2 = 25$. In particular, the histogram should be roughly bell shaped, centered at 1, and close to the horizontal axis at $1 \pm 3 \cdot 5$ (i.e., at -14 and 16).
d) Answers will vary.

9.133

a) We apply Equation (9.49) on page 540 to get

$$\begin{aligned}
f_{Y/X}(z) &= \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)\,dx = \int_0^{\infty} x \left(\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\right)\left(\frac{\lambda^\beta}{\Gamma(\beta)} (xz)^{\beta-1} e^{-\lambda(xz)}\right) dx \\
&= \frac{\lambda^\alpha \lambda^\beta z^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^{\infty} x^{\alpha+\beta-1} e^{-\lambda(z+1)x}\,dx = \frac{\lambda^\alpha \lambda^\beta z^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\big(\lambda(z+1)\big)^{\alpha+\beta}} \\
&= \frac{1}{B(\alpha, \beta)} \frac{z^{\beta-1}}{(z+1)^{\alpha+\beta}}
\end{aligned}$$

if $z > 0$, and $f_{Y/X}(z) = 0$ otherwise.
b) Let $U = Y/X$. We have $Z = X/(X+Y) = 1\big/(1 + Y/X) = 1/(1+U)$. Applying the univariate transformation theorem with $g(u) = 1/(1+u)$ and referring to part (a), we get

$$\begin{aligned}
f_Z(z) &= \frac{1}{|g'(u)|} f_U(u) = \frac{1}{\big|-1/(1+u)^2\big|} f_U(u) = (1+u)^2 f_U(u) = \big(1 + (1/z - 1)\big)^2 f_U(1/z - 1) \\
&= \frac{1}{z^2} \frac{1}{B(\alpha, \beta)} \frac{(1/z - 1)^{\beta-1}}{\big((1/z - 1) + 1\big)^{\alpha+\beta}} = \frac{1}{B(\alpha, \beta)} z^{\alpha-1}(1-z)^{\beta-1}
\end{aligned}$$

if $0 < z < 1$, and $f_Z(z) = 0$ otherwise. Thus, Z has the beta distribution with parameters α and β.
c) In Exercise 9.118, X and Y are independent $\mathcal{E}(\lambda)$ random variables or, equivalently, $\Gamma(1, \lambda)$ random variables. Applying part (b), with $\alpha = \beta = 1$, we find that $f_Z(z) = 1/B(1,1) = 1$ if $0 < z < 1$, and $f_Z(z) = 0$ otherwise. Thus, in this case, $Z = X/(X+Y) \sim \mathcal{U}(0, 1)$.

9.134

a) Let $U = W + X$. From Example 9.19 on page 535, $U \sim \mathcal{T}(0, 2)$ and, from Proposition 6.13 on page 297, U and Y are independent. Therefore, from Equation (9.46) on page 534,

$$f_{U+Y}(t) = \int_{-\infty}^{\infty} f_U(u) f_Y(t-u)\,du.$$

The integrand is nonzero if and only if $0 < u < 2$ and $0 < t - u < 1$; that is, if and only if $0 < u < 2$ and $t - 1 < u < t$. Therefore,

$$f_{U+Y}(t) = \int_{\max\{0, t-1\}}^{\min\{2, t\}} f_U(u)\,du.$$

Now,

$$\max\{0, t-1\} = \begin{cases} 0, & \text{if } t \le 1; \\ t-1, & \text{if } t > 1. \end{cases} \qquad \text{and} \qquad \min\{2, t\} = \begin{cases} t, & \text{if } t \le 2; \\ 2, & \text{if } t > 2. \end{cases}$$

Consequently,

$$f_{U+Y}(t) = \begin{cases} \int_0^t f_U(u)\,du, & \text{if } 0 < t \le 1; \\ \int_{t-1}^t f_U(u)\,du, & \text{if } 1 < t \le 2; \\ \int_{t-1}^2 f_U(u)\,du, & \text{if } 2 < t \le 3. \end{cases} = \begin{cases} \int_0^t u\,du, & \text{if } 0 < t \le 1; \\ \int_{t-1}^1 u\,du + \int_1^t (2-u)\,du, & \text{if } 1 < t \le 2; \\ \int_{t-1}^2 (2-u)\,du, & \text{if } 2 < t \le 3. \end{cases}$$

Evaluating the three integrals, we get

$$f_{W+X+Y}(t) = \begin{cases} t^2/2, & \text{if } 0 < t \le 1; \\ -t^2 + 3t - 3/2, & \text{if } 1 < t \le 2; \\ t^2/2 - 3t + 9/2, & \text{if } 2 < t \le 3; \\ 0, & \text{otherwise.} \end{cases}$$

b) Let $V = W + X + Y$, whose PDF is given in part (a). From Proposition 6.13 on page 297, V and Z are independent. Therefore, from Equation (9.46),

$$f_{V+Z}(t) = \int_{-\infty}^{\infty} f_V(v) f_Y(t - v)\, dv.$$

The integrand is nonzero if and only if $0 < v < 3$ and $0 < t - v < 1$; that is, if and only if $0 < u < 3$ and $t - 1 < v < t$. Therefore,

$$f_{V+Z}(t) = \int_{\max\{0,t-1\}}^{\min\{3,t\}} f_V(v)\, dv.$$

Now,

$$\max\{0, t-1\} = \begin{cases} 0, & \text{if } t \le 1; \\ t - 1, & \text{if } t > 1. \end{cases} \qquad \text{and} \qquad \min\{3, t\} = \begin{cases} t, & \text{if } t \le 3; \\ 3, & \text{if } t > 3. \end{cases}$$

Consequently,

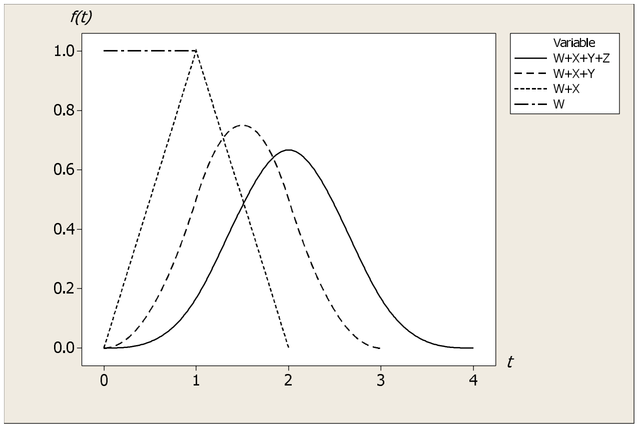
$$f_{V+Z}(t) = \begin{cases} \int_0^t f_V(v)\, dv, & \text{if } 0 < t \le 1; \\ \int_{t-1}^t f_V(v)\, dv, & \text{if } 1 < t \le 2; \\ \int_{t-1}^t f_V(v)\, dv, & \text{if } 2 < t \le 3; \\ \int_{t-1}^3 f_V(v)\, dv, & \text{if } 3 < t \le 4. \end{cases}$$

$$= \begin{cases} \int_0^t (v^2/2)\, dv, & \text{if } 0 < t \le 1; \\ \int_{t-1}^1 (v^2/2)\, dv + \int_1^t (-v^2 + 3v - 3/2)\, dv, & \text{if } 1 < t \le 2; \\ \int_{t-1}^2 (-v^2 + 3v - 3/2)\, dv + \int_2^t (v^2/2 - 3v + 9/2)\, dv, & \text{if } 2 < t \le 3; \\ \int_{t-1}^3 (v^2/2 - 3v + 9/2)\, dv, & \text{if } 3 < t \le 4. \end{cases}$$

Evaluating the four integrals, we get

$$f_{W+X+Y+Z}(t) = \begin{cases} t^3/6, & \text{if } 0 < t \le 1; \\ -t^3/2 + 2t^2 - 2t + 2/3, & \text{if } 1 < t \le 2; \\ t^3/2 - 4t^2 + 10t - 22/3, & \text{if } 2 < t \le 3; \\ -t^3/6 + 2t^2 - 8t + 32/3, & \text{if } 3 < t \le 4; \\ 0, & \text{otherwise.} \end{cases}$$

c) Following are graphs of the PDFs of W, $W + X$, $W + X + Y$, and $W + X + Y + Z$.



Theory Exercises

9.135

a) Using the FPF and making the substitution $u = xy$, we get

$$
\begin{aligned}
F_{XY}(z) = P(XY \le z) &= \iint_{xy \le z} f_{X,Y}(x, y)\,dx\,dy \\
&= \int_{-\infty}^{0} \left(\int_{z/x}^{\infty} f_{X,Y}(x, y)\,dy \right) dx + \int_{0}^{\infty} \left(\int_{-\infty}^{z/x} f_{X,Y}(x, y)\,dy \right) dx \\
&= \int_{-\infty}^{0} \frac{1}{x} \left(\int_{z}^{-\infty} f_{X,Y}(x, u/x)\,du \right) dx + \int_{0}^{\infty} \frac{1}{x} \left(\int_{-\infty}^{z} f_{X,Y}(x, u/x)\,du \right) dx \\
&= \int_{-\infty}^{0} \left(-\frac{1}{x} \right) \left(\int_{-\infty}^{z} f_{X,Y}(x, u/x)\,du \right) dx + \int_{0}^{\infty} \frac{1}{x} \left(\int_{-\infty}^{z} f_{X,Y}(x, u/x)\,du \right) dx \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{0} \frac{1}{|x|} f_{X,Y}(x, u/x)\,dx + \int_{0}^{\infty} \frac{1}{|x|} f_{X,Y}(x, u/x)\,dx \right) du. \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}(x, u/x)\,dx \right) du.
\end{aligned}
$$

Therefore, from Proposition 8.5 on page 422,

$$
f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}(x, z/x)\,dx.
$$

b) If X and Y are independent, then we can replace the joint PDF in the preceding display by the product of the marginal PDFs, thus:

$$
f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(z/x)\,dx.
$$

c) From part (b),

$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(z/x)\, dx = \int_z^1 \frac{1}{x} \cdot 1 \cdot 1\, dx = -\ln z$$

if $0 < z < 1$, and $f_{XY}(z) = 0$ otherwise.

9.136 Proceeding by induction, we assume that Proposition 9.13 holds for $m-1$ and must establish the result for m. From Proposition 6.13 on page 297 and the induction assumption, $X_1 + \cdots + X_{m-1}$ and X_m are independent random variables with distributions $\Gamma(\alpha_1 + \cdots + \alpha_{m-1}, \lambda)$ and $\Gamma(\alpha_m, \lambda)$, respectively. Hence, from Example 9.20(a) on page 536,

$$\begin{aligned} X_1 + \cdots + X_m &= (X_1 + \cdots + X_{m-1}) + X_m \\ &\sim \Gamma\big((\alpha_1 + \cdots + \alpha_{m-1}) + \alpha_m, \lambda\big) = \Gamma(\alpha_1 + \cdots + \alpha_m, \lambda). \end{aligned}$$

9.137 Proceeding by induction, we assume that Proposition 9.14 holds for $m-1$ and must establish the result for m. From Proposition 6.13 on page 297, the induction assumption, and Proposition 8.15 on page 468, the random variables $a + b_1 X_1 + \cdots + b_{m-1} X_{m-1}$ and $b_m X_m$ are independent $\mathcal{N}\big(a + b_1\mu_1 + \cdots + b_{m-1}\mu_{m-1}, b_1^2\sigma_1^2 + \cdots + b_{m-1}^2\sigma_{m-1}^2\big)$ and $\mathcal{N}\big(b_m\mu_m, b_m^2\sigma_m^2\big)$, respectively. Hence, from Example 9.22 on page 539,

$$\begin{aligned} &a + b_1 X_1 + \cdots + b_m X_m \\ &\quad = (a + b_1 X_1 + \cdots + b_{m-1} X_{m-1}) + b_m X_m \\ &\quad \sim \mathcal{N}\Big(\big(a + b_1\mu_1 + \cdots + b_{m-1}\mu_{m-1}\big) + b_m\mu_m, \big(b_1^2\sigma_1^2 + \cdots + b_{m-1}^2\sigma_{m-1}^2\big) + b_m^2\sigma_m^2\Big). \\ &\quad = \mathcal{N}\big(a + b_1\mu_1 + \cdots + b_m\mu_m, b_1^2\sigma_1^2 + \cdots + b_m^2\sigma_m^2\big). \end{aligned}$$

Advanced Exercises

9.138 Let $W = X + Y + Z$. Applying the FPF yields, for $w > 0$,

$$\begin{aligned} F_W(w) = P(W \le w) = P(X + Y + Z \le w) &= \iiint\limits_{x+y+z\le w} f_{X,Y,Z}(x, y, z)\, dx\, dy\, dz \\ &= \int_0^w \left(\int_0^{w-x} \left(\int_0^{w-x-y} \frac{6}{(1+x+y+z)^4}\, dz \right) dy \right) dx \\ &= \int_0^w \left(\int_0^{w-x} \left(\frac{2}{(1+x+y)^3} - \frac{2}{(1+w)^3} \right) dy \right) dx \\ &= \int_0^w \left(\frac{1}{(1+x)^2} - \frac{1}{(1+w)^2} - \frac{2(w-x)}{(1+w)^3} \right) dx \\ &= 1 - \left(\frac{1}{1+w} + \frac{w}{(1+w)^2} + \frac{w^2}{(1+w)^3} \right) = \left(\frac{w}{1+w} \right)^3. \end{aligned}$$

Differentiation now yields

$$f_{X+Y+Z}(w) = \frac{3w^2}{(1+w)^4}$$

if $w > 0$, and $f_{X+Y+Z}(w) = 0$ otherwise.

9.139

a) From Equation (9.44) on page 534 and Equation (9.35),

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\left(x^2-2\rho x(z-x)+(z-x)^2\right)/2(1-\rho^2)}\,dx.$$

Performing some algebra and completing the square yields

$$x^2 - 2\rho x(z-x) + (z-x)^2 = 2(1+\rho)(x-z/2)^2 + (1-\rho)z^2/2.$$

Hence,

$$\frac{x^2 - 2\rho x(z-x) + (z-x)^2}{2(1-\rho)^2} = \frac{(x-z/2)^2}{1-\rho} + \frac{z^2}{4(1+\rho)} = \frac{(x-z/2)^2}{2(1-\rho)/2} + \frac{z^2}{4(1+\rho)}.$$

Therefore, using the fact that a PDF must integrate to 1, we get

$$\begin{aligned} f_{X+Y}(z) &= e^{-z^2/4(1+\rho)} \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-(x-z/2)^2/\left(2(1-\rho)/2\right)}\,dx \\ &= e^{-z^2/4(1+\rho)} \frac{1}{2\pi\sqrt{1-\rho^2}} \sqrt{2\pi(1-\rho)/2} = \frac{1}{\sqrt{2\pi}\sqrt{2(1+\rho)}}\, e^{-z^2/2\left(2(1+\rho)\right)}. \end{aligned}$$

Thus, we see that $X+Y \sim \mathcal{N}\big(0, 2(1+\rho)\big)$.

b) From Equation (9.48) on page 540 and Equation (9.35),

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_{X,Y}(x, xz)\,dx = \int_{-\infty}^{\infty} |x| \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\left(x^2-2\rho x\cdot(xz)+(xz)^2\right)/2(1-\rho^2)}\,dx.$$

However,

$$x^2 - 2\rho x\cdot(xz) + (xz)^2 = \left(1 - 2\rho z + z^2\right)x^2.$$

Therefore, using symmetry and making the substitution $u = \left(1-2\rho z+z^2\right)x^2/2(1-\rho^2)$, we get

$$\begin{aligned} f_{Y/X}(z) &= \frac{2}{2\pi\sqrt{1-\rho^2}} \int_0^{\infty} x\, e^{-\left(1-2\rho z+z^2\right)x^2/2(1-\rho^2)}\,dx = \frac{1}{\pi\sqrt{1-\rho^2}} \frac{1-\rho^2}{1-2\rho z+z^2} \int_0^{\infty} e^{-u}\,du \\ &= \frac{\sqrt{1-\rho^2}}{\pi\left(1-2\rho z+z^2\right)} = \frac{\sqrt{1-\rho^2}}{\pi\left((1-\rho^2)+(z-\rho)^2\right)}. \end{aligned}$$

Referring now to Definition 8.11 on page 470, we conclude that Y/X has a Cauchy distribution with parameters ρ and $\sqrt{1-\rho^2}$.

9.140 Denote by I_n the interarrival time between the $(n-1)$st and nth customer. By assumption, $I_1, I_2, \ldots$ are independent exponential random variables with common parameter λ. Now let W_n denote the time at which the nth customer arrives. Because $W_n = I_1 + \cdots + I_n$, Proposition 9.13 on page 537 implies that $W_n \sim \Gamma(n, \lambda)$. Next, we observe that $\{N(t) = n\} = \{W_n \le t,\ W_{n+1} > t\}$. Consequently, by Equation (8.49) on page 453, for each nonnegative integer n,

$$\begin{aligned} P\big(N(t) = n\big) &= P(W_n \le t,\ W_{n+1} > t) = P(W_n \le t) - P(W_{n+1} \le t) \\ &= \left(1 - e^{-\lambda t}\sum_{j=0}^{n-1}\frac{(\lambda t)^j}{j!}\right) - \left(1 - e^{-\lambda t}\sum_{j=0}^{n}\frac{(\lambda t)^j}{j!}\right) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}. \end{aligned}$$

Thus, $N(t) \sim \mathcal{P}(\lambda t)$.

9.141 We begin by finding the PDF of $\sqrt{Y/\nu}$, which we call X. The required result can be obtained by using either the CDF method or the univariate transformation method. The result is

$$f_X(x) = \frac{\nu^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)} x^{\nu-1}\, e^{-\nu x^2/2}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. For convenience, let c denote the constant in the right of the preceding display. Because Z and Y are independent random variables, so are Z and X, by Proposition 6.13 on page 297. Therefore, we can now apply Equation (9.49) on page 540 and make the substitution $u = x^2$ to obtain

$$\begin{aligned}
f_T(t) = f_{Z/X}(t) &= \int_{-\infty}^{\infty} |x| f_X(x) f_Z(xt)\, dx = \frac{c}{\sqrt{2\pi}} \int_0^{\infty} x x^{\nu-1} e^{-\nu x^2/2} e^{-(xt)^2/2}\, dx \\
&= \frac{c}{\sqrt{2\pi}} \int_0^{\infty} x x^{\nu-1} e^{-(\nu+t^2)x^2/2}\, dx = \frac{c}{2\sqrt{2\pi}} \int_0^{\infty} u^{(\nu+1)/2-1}\, e^{-(\nu+t^2)u/2}\, du \\
&= \frac{1}{2\sqrt{2\pi}} \frac{\nu^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)} \frac{\Gamma\big((\nu+1)/2\big)}{\big((\nu+t^2)/2\big)^{(\nu+1)/2}} = \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \big(1 + t^2/\nu\big)^{-(\nu+1)/2}.
\end{aligned}$$

9.142

a) Let $U = X/2$ and $V = Y/2$. Using either the CDF method or the transformation method, we easily find that

$$f_U(u) = \frac{2}{\pi\left(1+4u^2\right)} \qquad \text{and} \qquad f_V(v) = \frac{2}{\pi\left(1+4v^2\right)}.$$

We want to show that the random variable $W = U + V$ has the standard Cauchy distribution. From Equation (9.46) on page 534,

$$f_W(w) = \int_{-\infty}^{\infty} f_U(u) f_V(w-u)\, du = \frac{4}{\pi^2} \int_{-\infty}^{\infty} \frac{du}{\left(1+4u^2\right)\left(1+4(w-u)^2\right)}.$$

To evaluate the integral, we break the integrand into two parts by using partial fractions:

$$\frac{1}{\left(1+4u^2\right)\left(1+4(w-u)^2\right)} = \frac{A+Bu}{1+4u^2} + \frac{C+Du}{1+4(w-u)^2}.$$

Clearing fractions and solving for A, B, C, and D yields

$$B = \frac{1}{2w\left(1+w^2\right)}, \quad A = \frac{1}{2}wB, \quad C = \frac{3}{2}wB, \quad D = -B.$$

Now we make the substitutions $v = 2u$ and $v = 2(u-w)$, respectively, in the two integrals and note that $A + (B/2)v + C + Dw + (D/2)v = \left(1+w^2\right)/2$ to obtain

$$\begin{aligned}
f_W(w) &= \frac{4}{\pi^2}\left(\frac{1}{2}\int_{-\infty}^{\infty} \frac{A+(B/2)v}{1+v^2}\, dv + \frac{1}{2}\int_{-\infty}^{\infty} \frac{C+Dw+(D/2)v}{1+v^2}\, dv\right) \\
&= \frac{2}{\pi^2}\int_{-\infty}^{\infty} \frac{A+(B/2)v+C+Dw+(D/2)v}{1+v^2}\, dv = \frac{1}{\pi\left(1+w^2\right)} \int_{-\infty}^{\infty} \frac{dv}{\pi\left(1+v^2\right)} \\
&= \frac{1}{\pi\left(1+w^2\right)}.
\end{aligned}$$

Hence, we see that W—that is, $(X+Y)/2$—has the standard Cauchy distribution.

b) If $c = 0$ or $c = 1$, the result is trivial. So, assume that $0 < c < 1$. For convenience, let $a = c$ and $b = 1 - c$, and set $U = aX$ and $V = bY$. Using either the CDF method or the transformation method, we easily find that

$$f_U(u) = \frac{a}{\pi\left(a^2 + u^2\right)} \qquad \text{and} \qquad f_V(v) = \frac{b}{\pi\left(b^2 + v^2\right)}.$$

We want to show that the random variable $W = U + V$ has the standard Cauchy distribution. From Equation (9.46) on page 534,

$$f_W(w) = \int_{-\infty}^{\infty} f_U(u) f_V(w-u)\,du = \frac{ab}{\pi^2} \int_{-\infty}^{\infty} \frac{du}{\left(a^2+u^2\right)\left(b^2+(w-u)^2\right)}.$$

At this point, we proceed as in part (a) to obtain the required result that W—that is, $(X+Y)/2$—has the standard Cauchy distribution.
c) We proceed by induction. Thus, assume that $(X_1 + \cdots + X_n)/n$ has the standard Cauchy distribution. By Proposition 6.13 on page 297, the random variables $X_1 + \cdots + X_n$ and X_{n+1} are independent. Now,

$$\frac{X_1 + \cdots + X_{n+1}}{n+1} = \left(\frac{n}{n+1}\right)\frac{X_1 + \cdots + X_n}{n} + \left(\frac{1}{n+1}\right) X_{n+1}.$$

Noting that $n/(n+1) + 1/(n+1) = 1$, we conclude from part (b) that $(X_1 + \cdots + X_{n+1})/(n+1)$ has the standard Cauchy distribution.

9.143 By Proposition 6.13 on page 297, the random variables X and $Y + Z$ are independent. Hence, by Equation (9.46) on page 534,

$$f_{X+Y+Z}(w) = f_{X+(Y+Z)}(w) \int_{-\infty}^{\infty} f_X(x) f_{Y+Z}(w-x)\,dx.$$

9.144
a) We apply the multivariate version of the FPF and make the substitution $u = x_1 + \cdots + x_m$ to obtain, for $w \in \mathcal{R}$,

$$\begin{aligned}
F_{X_1+\cdots+X_m}(w) &= P(X_1 + \cdots + X_m \le w) = \underset{x_1+\cdots+x_m \le w}{\int \cdots \int} f_{X_1,\ldots,X_m}(x_1,\ldots,x_m)\,dx_1 \cdots dx_m \\
&= \int_{-\infty}^{\infty}\left(\cdots \int_{-\infty}^{\infty}\left(\int_{-\infty}^{w-(x_1+\cdots+x_{m-1})} f_{X_1,\ldots,X_m}(x_1,\ldots,x_m)dx_m\right)dx_{m-1}\cdots\right)dx_1 \\
&= \int_{-\infty}^{\infty}\left(\cdots \int_{-\infty}^{\infty}\left(\int_{-\infty}^{w} f_{X_1,\ldots,X_m}(x_1,\ldots,x_{m-1},u-(x_1+\cdots+x_{m-1}))du\right)dx_{m-1}\cdots\right)dx_1 \\
&= \int_{-\infty}^{w}\left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_m}(x_1,\ldots,x_{m-1},u-x_1-\cdots-x_{m-1})\,dx_1\cdots dx_{m-1}\right)du.
\end{aligned}$$

The required result now follows from Proposition 8.5 on page 422.
b) For $k = 1, 2, \ldots, m-1$, we have

$$\begin{aligned}
&f_{X_1+\cdots+X_m}(w) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_m}\left(x_1,\ldots,x_{k-1}, w-\left(\textstyle\sum_{j\ne k} x_j\right), x_{k+1},\ldots,x_m\right)dx_1\cdots dx_{k-1}\,dx_{k+1}\cdots dx_m.
\end{aligned}$$

c) Because of independence, a joint PDF of $X_1, \ldots, X_m$ is the product of the marginal PDFs. Thus, for part (a), we can write

$$f_{X_1+\cdots+X_m}(w) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \big(f_{X_1}(x_1) \cdots f_{X_{m-1}}(x_{m-1}) f_{X_m}(w - x_1 - \cdots - x_{m-1})\big) dx_1 \cdots dx_{m-1}$$

$$= \int_{-\infty}^{\infty} \left(\cdots \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f_{X_{m-1}}(x_{m-1}) f_{X_m}(w - x_1 - \cdots - x_{m-1})\, dx_{m-1} \right) \right.$$

$$\left. \times\, f_{X_{m-2}}(x_{m-2}) dx_{m-2} \cdots \right) f_{X_1}(x_1)\, dx_1,$$

and similarly for the $m-1$ formulas in part (b).

d) For part (a),

$$f_{X+Y+Z}(w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, w - x - y)\, dx\, dy.$$

For part (b),

$$f_{X+Y+Z}(w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(w - y - z, y, z)\, dy\, dz$$

and

$$f_{X+Y+Z}(w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x, w - x - z, z)\, dx\, dz.$$

For part (c),

$$f_{X+Y+Z}(w) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f_Y(y) f_Z(w - x - y)\, dy \right) f_X(x)\, dx,$$

$$f_{X+Y+Z}(w) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f_X(w - y - z) f_Z(z)\, dz \right) f_Y(y)\, dy,$$

and

$$f_{X+Y+Z}(w) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f_Y(w - x - z) f_Z(z)\, dz \right) f_X(x)\, dx.$$

9.145 By Proposition 6.13 on page 297, the random variables $X_1 + \cdots + X_{m-1}$ and X_m are independent. Hence, by Equation (9.46) on page 534,

$$f_{X_1+\cdots+X_m}(w) = f_{(X_1+\cdots+X_{m-1})+X_m}(w) = \int_{-\infty}^{\infty} f_{X_1+\cdots+X_{m-1}}(x) f_{X_m}(w - x)\, dx.$$

9.7 Multivariate Transformation Theorem

Basic Exercises

9.146

a) Let X and Y denote the inspection times for the first and second engineers, respectively. By assumption, X and Y are independent random variables with distributions $\Gamma(\alpha, \lambda)$ and $\Gamma(\beta, \lambda)$, respectively.

Hence a joint PDF of X and Y is given by

$$f_{X,Y}(x,y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \cdot \frac{\lambda^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y} = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1}\, e^{-\lambda(x+y)}$$

if $x > 0$ and $y > 0$, and $f_{X,Y}(x,y) = 0$ otherwise. We note that $S = X/(X+Y)$ and $T = X+Y$ and, furthermore, that the range of S is the interval $(0,1)$ and the range of T is the interval $(0,\infty)$. As in Example 9.25, the Jacobian determinant of the transformation $s = x/(x+y)$ and $t = x+y$ is $J(x,y) = 1/(x+y)$, and the inverse transformation is $x = st$ and $y = (1-s)t$. Applying the bivariate transformation theorem, we obtain a joint PDF of S and T:

$$\begin{aligned} f_{S,T}(s,t) &= \frac{1}{|J(x,y)|} f_{X,Y}(x,y) = \big(st + (1-s)t\big) \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} (st)^{\alpha-1} \big((1-s)t\big)^{\beta-1}\, e^{-\lambda\left(st+(1-s)t\right)} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} t^{\alpha+\beta-1} e^{-\lambda t} \end{aligned}$$

if $0 < s < 1$ and $t > 0$, and $f_{S,T}(s,t) = 0$ otherwise.

b) To obtain a PDF of S, we apply Proposition 9.7 on page 510 to get

$$\begin{aligned} f_S(s) &= \int_{-\infty}^{\infty} f_{S,T}(s,t)\,dt = \int_0^\infty \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} t^{\alpha+\beta-1} e^{-\lambda t}\,dt \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} \int_0^\infty t^{\alpha+\beta-1} e^{-\lambda t}\,dt \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} \frac{\Gamma(\alpha+\beta)}{\lambda^{\alpha+\beta}} = \frac{1}{B(\alpha,\beta)} s^{\alpha-1}(1-s)^{\beta-1} \end{aligned}$$

if $0 < s < 1$, and $f_S(s) = 0$ otherwise. Hence, S has the beta distribution with parameters α and β.

c) To obtain a PDF of T, we again apply Proposition 9.7 to get

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_{S,T}(s,t)\,ds = \int_0^1 \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} t^{\alpha+\beta-1} e^{-\lambda t}\,ds \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha+\beta-1} e^{-\lambda t} \int_0^1 s^{\alpha-1}(1-s)^{\beta-1}\,ds \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha+\beta-1} e^{-\lambda t} B(\alpha,\beta) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} t^{\alpha+\beta-1} e^{-\lambda t} \end{aligned}$$

if $t > 0$, and $f_T(t) = 0$ otherwise. Hence, $T \sim \Gamma(\alpha+\beta, \lambda)$.

d) Yes. From parts (b) and (c),

$$\begin{aligned} f_S(s) f_T(t) &= \frac{1}{B(\alpha,\beta)} s^{\alpha-1}(1-s)^{\beta-1} \cdot \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} t^{\alpha+\beta-1} e^{-\lambda t} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} t^{\alpha+\beta-1} e^{-\lambda t} \end{aligned}$$

if $0 < s < 1$ and $t > 0$, and $f_S(s) f_T(t) = 0$ otherwise. Thus, in view of part (a), the function f defined on $\mathcal{R}^2$ by $f(s,t) = f_S(s) f_T(t)$ is a joint PDF of S and T. Consequently, S and T are independent random variables.

9.147
a) Let X and Y denote the inspection times for the first and second engineers, respectively. By assumption, X and Y are independent random variables with distributions $\mathcal{E}(\lambda)$ and $\mathcal{E}(\mu)$, respectively. Hence a joint PDF of X and Y is given by

$$f_{X,Y}(x, y) = \lambda e^{-\lambda x} \cdot \mu e^{-\mu y} = \lambda\mu e^{-(\lambda x+\mu y)}$$

if $x > 0$ and $y > 0$, and $f_{X,Y}(x, y) = 0$ otherwise. We note that $S = X/(X + Y)$ and $T = X + Y$ and, furthermore, that the range of S is the interval $(0, 1)$ and the range of T is the interval $(0, \infty)$. As in Example 9.25, the Jacobian determinant of the transformation $s = x/(x + y)$ and $t = x + y$ is $J(x, y) = 1/(x + y)$, and the inverse transformation is $x = st$ and $y = (1 - s)t$. Applying the bivariate transformation theorem, we obtain a joint PDF of S and T:

$$f_{S,T}(s, t) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = \big(st + (1 - s)t\big)\lambda\mu e^{-(\lambda(st)+\mu(1-s)t)} = \lambda\mu t e^{-(\mu+(\lambda-\mu)s)t}$$

if $0 < s < 1$ and $t > 0$, and $f_{S,T}(s, t) = 0$ otherwise.
b) To obtain a PDF of S, we apply Proposition 9.7 on page 510 to get

$$f_S(s) = \int_{-\infty}^{\infty} f_{S,T}(s, t)\,dt = \int_0^{\infty} \lambda\mu t e^{-(\mu+(\lambda-\mu)s)t}\,dt = \lambda\mu \int_0^{\infty} t^{2-1} e^{-(\mu+(\lambda-\mu)s)t}\,dt$$
$$= \lambda\mu \frac{\Gamma(2)}{\big(\mu + (\lambda - \mu)s\big)^2} = \frac{\lambda\mu}{\big(\mu + (\lambda - \mu)s\big)^2}$$

if $0 < s < 1$, and $f_S(s) = 0$ otherwise.
c) To obtain a PDF of T, we again apply Proposition 9.7 to get

$$f_T(t) = \int_{-\infty}^{\infty} f_{S,T}(s, t)\,ds = \int_0^1 \lambda\mu t e^{-(\mu+(\lambda-\mu)s)t}\,ds = \lambda\mu t e^{-\mu t} \int_0^1 e^{-((\lambda-\mu)t)s}\,ds$$
$$= \frac{\lambda\mu t e^{-\mu t}}{(\lambda - \mu)t}\left(1 - e^{-(\lambda-\mu)t}\right) = \frac{\lambda\mu}{\lambda - \mu}\left(e^{-\mu t} - e^{-\lambda t}\right)$$

if $t > 0$, and $f_T(t) = 0$ otherwise.
d) No. We can see that S and T are not independent random variables in several ways. One way is to note that, for $0 < s < 1$ and $t > 0$,

$$f_{T\,|\,S}(t\,|\,s) = \frac{f_{S,T}(s, t)}{f_S(s)} = \frac{\lambda\mu t e^{-(\mu+(\lambda-\mu)s)t}}{\lambda\mu\big/\big(\mu + (\lambda - \mu)s\big)^2} = \big(\mu + (\lambda - \mu)s\big)^2 t e^{-(\mu+(\lambda-\mu)s)t},$$

which clearly depends on s because $\lambda \neq \mu$. Consequently, from Proposition 9.10 on page 527, we conclude that S and T are not independent.

9.148 Let $U = X$ (the "dummy" variable) and $V = Y/X$. The Jacobian determinant of the transformation $u = x$ and $v = y/x$ is

$$J(x, y) = \begin{vmatrix} 1 & 0 \\ -y/x^2 & 1/x \end{vmatrix} = \frac{1}{x}.$$

Solving the equations $u = x$ and $v = y/x$ for x and y, we obtain the inverse transformation $x = u$ and $y = uv$. Hence, by the bivariate transformation theorem,

$$f_{U,V}(u, v) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = |x| f_{X,Y}(x, y) = |u| f_{X,Y}(u, uv).$$

Therefore,

$$f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u, v)\,du = \int_{-\infty}^{\infty} |u| f_{X,Y}(u, uv)\,du.$$

In other words,

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_{X,Y}(x, xz)\,dx,$$

which is Equation (9.48).

9.149

a) Let X and Y denote the x and y coordinates, respectively, of the first spot to appear. Recall that $f_{X,Y}(x, y) = 1/\pi$ if (x, y) is in the unit disk, and $f_{X,Y}(x, y) = 0$ otherwise. We have $X = R\cos\Theta$ and $Y = R\sin\Theta$. Hence, the inverse transformation is $x = r\cos\theta$ and $y = r\sin\theta$. The Jacobian determinant of this inverse transformation is

$$J(r, \theta) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Recalling that the Jacobian determinant of this inverse transformation is the reciprocal of that of the transformation itself, we conclude from the bivariate transformation theorem that

$$f_{R,\Theta}(r, \theta) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = |J(r, \theta)| f_{X,Y}(x, y) = \frac{r}{\pi}$$

if $0 < r < 1$ and $0 < \theta < 2\pi$, and $f_{R,\Theta}(r, \theta) = 0$ otherwise.

b) We have

$$f_R(r) = \int_{-\infty}^{\infty} f_{R,\Theta}(r, \theta)\,d\theta = \int_0^{2\pi} (r/\pi)\,d\theta = 2r$$

if $0 < r < 1$, and $f_R(r) = 0$ otherwise. Thus, R has the beta distribution with parameters 2 and 1. Also,

$$f_\Theta(\theta) = \int_{-\infty}^{\infty} f_{R,\Theta}(r, \theta)\,dr = \int_0^1 (r/\pi)\,dr = \frac{1}{2\pi}$$

if $0 < \theta < 2\pi$, and $f_\Theta(\theta) = 0$ otherwise. Thus, $\Theta \sim \mathcal{U}(0, 2\pi)$.

c) From parts (a) and (b), we see that the function f defined on $\mathcal{R}^2$ by $f(r, \theta) = f_R(r) f_\Theta(\theta)$ is a joint PDF of R and Θ. Consequently, R and Θ are independent random variables.

9.150

a) We have

$$f_{U,V}(u, v) = f_U(u) f_V(v) = 1 \cdot 1 = 1$$

if $0 < u < 1$ and $0 < v < 1$, and $f_{U,V}(u, v) = 0$ otherwise. The Jacobian determinant of the transformation $x = a + (b - a)u$ and $y = c + (d - c)v$ is

$$J(x, y) = \begin{vmatrix} b - a & 0 \\ 0 & d - c \end{vmatrix} = (b - a)(d - c).$$

We note that the ranges of X and Y are the intervals (a, b) and (c, d), respectively. Solving the equations $x = a + (b - a)u$ and $y = c + (d - c)v$ for u and v, we obtain the inverse transformation $u = (x - a)/(b - a)$ and $v = (y - c)/(d - c)$. Hence, by the bivariate transformation theorem,

$$f_{X,Y}(x, y) = \frac{1}{|J(u, v)|} f_{U,V}(u, v) = \frac{1}{(b - a)(d - c)}$$

if $a < x < b$ and $c < y < d$, and $f_{X,Y}(x, y) = 0$ otherwise. Thus, the random point (X, Y) has the uniform distribution on the rectangle $(a, b) \times (c, d)$.

b) First obtain two (independent) numbers, say, u and v, from a basic random number generator. Then $\big(a + (b-a)u, c + (d-c)v\big)$ is a point selected at random from the rectangle $(a, b) \times (c, d)$.

9.151

a) From Equation (9.20), a joint PDF of X and Y is

$$f_{X,Y}(x, y) = n(n-1)f(x)f(y)\big(F(y) - F(x)\big)^{n-2}$$

if $x < y$, and $f_{X,Y}(x, y) = 0$ otherwise. The Jacobian determinant of the transformation $r = y - x$ and $m = (x + y)/2$ is

$$J(x, y) = \begin{vmatrix} -1 & 1 \\ 1/2 & 1/2 \end{vmatrix} = -1.$$

Solving the equations $r = y - x$ and $m = (x + y)/2$ for x and y, we obtain the inverse transformation $x = m - r/2$ and $y = m + r/2$. Therefore, by the bivariate transformation theorem, a joint PDF of R and M is

$$f_{R,M}(r, m) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = f_{X,Y}(m - r/2, m + r/2).$$

Thus, for $r > 0$ and $-\infty < m < \infty$,

$$f_{R,M}(r, m) = n(n-1)f(m - r/2)f(m + r/2)\big(F(m + r/2) - F(m - r/2)\big)^{n-2},$$

and $f_{R,M}(r, m) = 0$ otherwise.

b) To obtain a marginal PDF of R, we apply Proposition 9.7 on page 510 to the joint PDF of R and M that we just found. Making the substitution $x = m - r/2$, we get

$$\begin{aligned} f_R(r) &= \int_{-\infty}^{\infty} f_{R,M}(r, m)\, dm \\ &= n(n-1)\int_{-\infty}^{\infty} f(m - r/2)f(m + r/2)\big(F(m + r/2) - F(m - r/2)\big)^{n-2}\, dm \\ &= n(n-1)\int_{-\infty}^{\infty} f(x)f(r + x)\big(F(r + x) - F(x)\big)^{n-2}\, dx \end{aligned}$$

if $r > 0$, and $f_R(r) = 0$ otherwise. This result agrees with the one found in Example 9.6.

c) To obtain a marginal PDF of M, we again apply Proposition 9.7 to the joint PDF of R and M. Making the substitution $x = m - r/2$, we get, for $-\infty < m < \infty$,

$$\begin{aligned} f_M(m) &= \int_{-\infty}^{\infty} f_{R,M}(r, m)\, dr \\ &= n(n-1)\int_{0}^{\infty} f(m - r/2)f(m + r/2)\big(F(m + r/2) - F(m - r/2)\big)^{n-2}\, dr \\ &= 2n(n-1)\int_{-\infty}^{m} f(x)f(2m - x)\big(F(2m - x) - F(x)\big)^{n-2}\, dx. \end{aligned}$$

This result agrees with the one found in Exercise 9.116.

d) No, R and M are not independent random variables. This fact can be seen in several ways. One way is to note from parts (a) and (b) that the conditional PDF of M given $R = r$ depends on r. Thus, from Proposition 9.10 on page 527, R and M are not independent.

9.152 For all $x, y \in \mathcal{R}$,

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.$$

Let $U = X^2 + Y^2$ and $V = Y/X$, and observe that the ranges of U and V are the intervals $(0, \infty)$ and $(-\infty, \infty)$, respectively. We want to find a joint PDF of U and V. Observe, however, that the transformation $u = x^2 + y^2$ and $v = y/x$ does not satisfy the conditions of the bivariate transformation theorem because it is not one-to-one on the range of (X, Y), which is $\mathcal{R}^2$. Indeed, each two points (x, y) and $(-x, -y)$ are transformed into the same (u, v) point. Note, though, that the joint density values of each two such points are identical. Consequently,

$$f_{U,V}(u, v) = 2 \cdot \frac{1}{|J(x, y)|} f_{X,Y}(x, y).$$

The Jacobian determinant of the transformation $u = x^2 + y^2$ and $v = y/x$ is

$$J(x, y) = \begin{vmatrix} 2x & 2y \\ -y/x^2 & 1/x \end{vmatrix} = 2\left(1 + \frac{y^2}{x^2}\right) = \frac{2}{x^2}\left(x^2 + y^2\right).$$

Noting that $x^2 = u/(1 + v^2)$, we now get

$$\begin{aligned} f_{U,V}(u, v) &= 2 \cdot \frac{x^2}{2\left(x^2 + y^2\right)} \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} = \frac{u/(1 + v^2)}{u} \frac{1}{2\pi\sigma^2} e^{-u/2\sigma^2} \\ &= \frac{1}{1 + v^2} \frac{1}{2\pi\sigma^2} e^{-u/2\sigma^2} = \left(\frac{1}{2\sigma^2} e^{-u/2\sigma^2}\right)\left(\frac{1}{\pi\left(1 + v^2\right)}\right) \end{aligned}$$

if $u > 0$ and $-\infty < v < \infty$, and $f_{U,V}(u, v) = 0$ otherwise. We see that the joint PDF of U and V can be factored into a function of u alone and a function of v alone. Therefore, from Exercise 9.95 on page 528, U and V are independent random variables, that is, $X^2 + Y^2$ and Y/X are independent. We note in passing that the factorization also shows that $X^2 + Y^2 \sim \mathcal{E}\left(1/2\sigma^2\right)$ and $Y/X \sim \mathcal{C}(0, 1)$.

9.153 We have

$$f_{X,Y}(x, y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \cdot \frac{\lambda^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y} = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)}$$

if $x > 0$ and $y > 0$, and $f_{X,Y}(x, y) = 0$ otherwise. Let $U = X + Y$ and $V = Y/X$, and observe that the ranges of both U and V are $(0, \infty)$. We want to find a joint PDF of U and V. The Jacobian determinant of the transformation $u = x + y$ and $v = y/x$ is

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ -y/x^2 & 1/x \end{vmatrix} = \frac{x + y}{x^2}.$$

Solving the equations $u = x + y$ and $v = y/x$ for x and y, we determine the inverse transformation to be $x = u/(1 + v)$ and $y = uv/(1 + v)$. Therefore, by the bivariate transformation theorem, a joint PDF of U and V is

$$f_{U,V}(u, v) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = \frac{x^2}{x+y} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)}$$
$$= \frac{u^2/(1+v)^2}{u} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{u}{1+v}\right)^{\alpha-1} \left(\frac{uv}{1+v}\right)^{\beta-1} e^{-\lambda u}$$
$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \frac{u^{\alpha+\beta-1} v^{\beta-1}}{(1+v)^{\alpha+\beta}} e^{-\lambda u} = \left(\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} u^{\alpha+\beta-1} e^{-\lambda u}\right)\left(\frac{1}{B(\alpha, \beta)} \frac{v^{\beta-1}}{(1+v)^{\alpha+\beta}}\right)$$

if $u > 0$ and $v > 0$, and $f_{U,V}(u, v) = 0$ otherwise. We see that the joint PDF of U and V can be factored into a function of u alone and a function of v alone. Therefore, from Exercise 9.95 on page 528, U and V are independent random variables, that is, $X + Y$ and Y/X are independent.

9.154 Let $U = X + Y$ and $V = X - Y$.
a) We want to find a joint PDF of U and V. The Jacobian determinant of the transformation $u = x + y$ and $v = x - y$ is

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2.$$

Solving the equations $u = x + y$ and $v = x - y$ for x and y, we determine the inverse transformation to be $x = (u + v)/2$ and $y = (u - v)/2$. Therefore, by the bivariate transformation theorem, a joint PDF of U and V is

$$f_{U,V}(u, v) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = \frac{1}{2} f_{X,Y}\big((u+v)/2, (u-v)/2\big).$$

b) In this case,

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu)^2/2\sigma^2}$$
$$= \frac{1}{2\pi\sigma^2} e^{-\left((x-\mu)^2+(y-\mu)^2\right)/2\sigma^2} = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2-2\mu(x+y)+2\mu^2\right)/2\sigma^2}.$$

Recalling that $x = (u + v)/2$ and $y = (u - v)/2$, we find that

$$x^2 + y^2 - 2\mu(x + y) + 2\mu^2 = \frac{1}{2}\left((u - 2\mu)^2 + v^2\right).$$

Hence, from part (a),

$$f_{U,V}(u, v) = \frac{1}{2} \frac{1}{2\pi\sigma^2} e^{-\left((u-2\mu)^2+v^2\right)/4\sigma^2} = \left(\frac{1}{2\pi\sigma} e^{-(u-2\mu)^2/4\sigma^2}\right)\left(\frac{1}{2\pi\sigma} e^{-v^2/4\sigma^2}\right).$$

We see that the joint PDF of U and V can be factored into a function of u alone and a function of v alone. Therefore, from Exercise 9.95 on page 528, U and V are independent random variables, that is, $X + Y$ and $X - Y$ are independent. We note in passing that the factorization also shows that $X + Y \sim \mathcal{N}\left(2\mu, 2\sigma^2\right)$ and $X - Y \sim \mathcal{N}\left(0, 2\sigma^2\right)$.

c) In this case, $f_{X,Y}(x, y) = 1 \cdot 1 = 1$ if $0 < x < 1$ and $0 < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise. Recalling that $x = (u + v)/2$ and $y = (u - v)/2$, we find from part (a) that

$$f_{U,V}(u, v) = \frac{1}{2}$$

if $(u, v) \in S$, and $f_{U,V}(u, v) = 0$ otherwise, where S is the square with vertices $(0, 0)$, $(1, -1)$, $(2, 0)$, and $(1, 1)$. We can now show in several ways that U and V (i.e., $X + Y$ and $X - Y$) are not independent random variables. One way is to note that the range of V is the interval $(-1, 1)$, whereas the range of $V_{|U=1/2}$ is the interval $(-1/2, 1/2)$. Hence, by Proposition 9.10 on page 527, U and V are not independent.

9.155 Let $X(t)$ and $Y(t)$ denote the x and y coordinates, respectively, of the particle at time t. By assumption, $X(t)$ and $Y(t)$ are independent $\mathcal{N}\big(0, \sigma^2 t\big)$ random variables. Hence, their joint PDF is, for all $(x, y) \in \mathcal{R}^2$,

$$f_{X(t),Y(t)}(x, y) = \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{t}}\, e^{-x^2/2\sigma^2 t} \cdot \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{t}}\, e^{-y^2/2\sigma^2 t} = \frac{1}{2\pi\sigma^2 t} e^{-\left(x^2+y^2\right)/2\sigma^2 t}.$$

a) We have $X(t) = R(t)\cos\big(\Theta(t)\big)$ and $Y(t) = R(t)\sin\big(\Theta(t)\big)$. Hence, the inverse transformation is $x = r\cos\theta$ and $y = r\sin\theta$. The Jacobian determinant of this inverse transformation is

$$J(r, \theta) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Recalling that the Jacobian determinant of this inverse transformation is the reciprocal of that of the transformation itself, we conclude from the bivariate transformation theorem that

$$\begin{aligned} f_{R(t),\Theta(t)}(r, \theta) &= \frac{1}{|J(x, y)|} f_{X(t),Y(t)}(x, y) = |J(r, \theta)|\, f_{X(t),Y(t)}(x, y) \\ &= r\frac{1}{2\pi\sigma^2 t} e^{-\left(x^2+y^2\right)/2\sigma^2 t} = \frac{r}{2\pi\sigma^2 t} e^{-r^2/2\sigma^2 t} \end{aligned}$$

if $r > 0$ and $0 < \theta < 2\pi$, and $f_{R(t),\Theta(t)}(r, \theta) = 0$ otherwise.

b) We can apply Proposition 9.7 on page 510 to obtain the marginal PDFs of $R(t)$ and $\Theta(t)$. Alternatively, we can and do use Exercises 9.95 and 9.96 on page 528. From part (a), we can write

$$f_{R(t),\Theta(t)}(r, \theta) = \left(\frac{r}{\sigma^2 t} e^{-r^2/2\sigma^2 t}\right)\left(\frac{1}{2\pi}\right)$$

if $r > 0$ and $0 < \theta < 2\pi$, and $f_{R(t),\Theta(t)}(r, \theta) = 0$ otherwise. It now follows immediately from the aforementioned exercises that

$$f_{R(t)}(r) = \begin{cases} \dfrac{r}{\sigma^2 t} e^{-r^2/2\sigma^2 t}, & \text{if } r > 0; \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad f_{\Theta(t)}(\theta) = \begin{cases} \dfrac{1}{2\pi}, & \text{if } 0 < \theta < 2\pi; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we see that $\Theta(t) \sim \mathcal{U}(0, 2\pi)$ and, referring to Exercise 8.159 on page 475, that $R(t)$ has the Rayleigh distribution with parameter $\sigma^2 t$.

c) It follows immediately from part (b) and either Exercise 9.95(d) or Proposition 9.9 on page 523 that $R(t)$ and $\Theta(t)$ are independent random variables.

9.156

a) A marginal PDF of X and all conditional PDFs of Y given $X = x$ determine a joint PDF of X and Y because of the general multiplication rule for the joint PDF of two continuous random variables, Equation (9.34) on page 517.

b) We have

$$f_{X,Y}(x, y) = \frac{1}{|T|} = \frac{1}{(1/2) \cdot 1 \cdot 1} = 2$$

if $(x, y) \in T$, and $f_{X,Y}(x, y) = 0$ otherwise. Also,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_0^x 2\, dy = 2x$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise. Furthermore, for $0 < x < 1$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}$$

if $0 < y < x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise.

c) It follows from part (b) that $F_X(x) = x^2$ if $0 < x < 1$ and, hence, $F_X^{-1}(u) = \sqrt{u}$. Therefore, from Proposition 8.16(b), $\sqrt{U}$ has the same probability distribution as X. It also follows from part (b) that, for each $0 < x < 1$, $F_{Y\,|\,X}(y\,|\,x) = y/x$ if $0 < y < x$ and, hence, $F_{Y\,|\,X}^{-1}(v) = xv$. Therefore, from Proposition 8.16(b), xV has the same probability distribution as $Y_{|X=x}$. However, because of the independence of U and V, $\sqrt{U}V_{|\sqrt{U}=x}$ has the same probability distribution as xV and, hence, as $Y_{|X=x}$.

d) From part (c), $\sqrt{U}$ and X have the same probability distributions and, hence, the same marginal PDFs. Also, from part (c), for each x, the conditional distribution of $\sqrt{U}V$ given $\sqrt{U} = x$ is the same as the conditional distribution of Y given $X = x$; in other words, for each x, the conditional PDF of $\sqrt{U}V$ given $\sqrt{U} = x$ is the same as the conditional PDF of Y given $X = x$. Consequently, from part (a), $\sqrt{U}$ and $\sqrt{U}V$ have the same joint PDF and, hence, the same joint probability distribution as X and Y.

e) We first note that, because U and V are independent $\mathcal{U}(0, 1)$ random variables, we have $f_{U,V}(u, v) = 1$ if $0 < u < 1$ and $0 < v < 1$, and $f_{U,V}(u, v) = 0$ otherwise. Let $R = \sqrt{U}$ and $S = \sqrt{U}V$. The Jacobian determinant of the transformation $r = \sqrt{u}$ and $s = \sqrt{u}v$ is

$$J(u, v) = \begin{vmatrix} 1/\left(2\sqrt{u}\right) & 0 \\ v/\left(2\sqrt{u}\right) & \sqrt{u} \end{vmatrix} = \frac{1}{2}.$$

Solving the equations $r = \sqrt{u}$ and $s = \sqrt{u}v$ for u and v, we obtain the inverse transformation $u = r^2$ and $v = s/r$. It follows that $0 < u < 1$ and $0 < v < 1$ if and only if $0 < r < 1$ and $0 < s < r$. Thus, by the bivariate transformation theorem,

$$f_{R,S}(r, s) = \frac{1}{|J(u, v)|} f_{U,V}(u, v) = 2 \cdot 1 = 2$$

if $(r, s) \in T$, and $f_{R,S}(r, s) = 0$ otherwise. Consequently, in view of part (a), we see that R and S—that is, $\sqrt{U}$ and $\sqrt{U}V$—have the same joint probability distribution as X and Y.

f) Use a basic random number generator to obtain two independent uniform numbers between 0 and 1, say, u and v, and then calculate the point $\left(\sqrt{u}, \sqrt{u}v\right)$.

g) Consider random variables X and Y that satisfy the following properties:

- X and Y are continuous random variables with a joint PDF.
- The conditions of Proposition 8.16 on page 471 hold for the random variable X and for the random variable $Y_{|X=x}$ for each x in the range of X.
- For each x in the range of X, the equation $u = F_X(x)$ can be solved explicitly for x in terms of u and, for each (x, y) in the range of (X, Y), the equation $v = F_{Y\,|\,X}(y\,|\,x)$ can be solved explicitly for y in terms of x and v.

Then we can proceed in an analogous manner to that in parts (a)–(f) to simulate an observation of the random point (X, Y).

9.157 From the Box–Müller transformation, if U and V are independent $\mathcal{U}(0,1)$ random variables, then $\sqrt{-2\ln U}\cos(2\pi V)$ has the standard normal distribution. Therefore, from Proposition 8.15 on page 468, the random variable $X = \mu + \sigma\sqrt{-2\ln U}\cos(2\pi V)$ has the normal distribution with parameters μ and σ^2. So, to simulate an observation of a normal random variable with parameters μ and σ^2, we first obtain two (independent) numbers, say, u and v, from a basic random number generator. The number $x = \mu + \sigma\sqrt{-2\ln u}\cos(2\pi v)$ is then a simulated observation of a normal random variable with parameters μ and σ^2.

9.158
a) Answers will vary, but here is the procedure: First obtain 10,000 pairs (a total of 20,000 numbers) from a basic random number generator. Then, for each pair, (u, v), compute $x = 266 + 16\sqrt{-2\ln u}\cos(2\pi v)$. The 10,000 x-values thus obtained will constitute a simulation of 10,000 human gestation periods.
b) The histogram should be shaped roughly like a normal curve with parameters $\mu = 266$ and $\sigma = 16$.
c) Answers will vary.

9.159
a) We have, for all $z_1, z_2 \in \mathcal{R}$,

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}\cdot\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2} = \frac{1}{2\pi}e^{-(z_1^2+z_2^2)/2}.$$

The Jacobian determinant of the transformation $x = z_1$ and $y = \rho z_1 + \sqrt{1-\rho^2}\,z_2$ is

$$J(z_1,z_2) = \begin{vmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{vmatrix} = \sqrt{1-\rho^2}.$$

Solving the equations $x = z_1$ and $y = \rho z_1 + \sqrt{1-\rho^2}\,z_2$, we obtain the inverse transformation $z_1 = x$ and $z_2 = (y - \rho x)/\sqrt{1-\rho^2}$. It follows that $z_1^2 + z_2^2 = \left(x^2 - 2\rho xy + y^2\right) / \left(1-\rho^2\right)$. Thus, by the bivariate transformation theorem,

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{|J(z_1,z_2)|}f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{\sqrt{1-\rho^2}}\frac{1}{2\pi}e^{-(z_1^2+z_2^2)/2} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\left(x^2-2\rho xy+y^2\right)/2\left(1-\rho^2\right)} \end{aligned}$$

for all $x, y \in \mathcal{R}$.
b) Let ρ be a specified real number with $-1 < \rho < 1$. Obtain two (independent) numbers, say, u and v, from a basic random number generator, and then set

$$x = \sqrt{-2\ln u}\cos(2\pi v) \qquad \text{and} \qquad y = \rho\sqrt{-2\ln u}\cos(2\pi v) + \sqrt{1-\rho^2}\sqrt{-2\ln u}\sin(2\pi v).$$

From the Box–Müller transformation and part (a), the pair (x, y) is a simulated observation of a bivariate normal distribution with joint PDF given by Equation (9.35).

Theory Exercises

9.160 Let $X_1, \ldots, X_m$ be continuous random variables with a joint PDF. Suppose that $g_1, \ldots, g_m$ are real-valued functions of m real variables defined on the range of $(X_1, \ldots, X_m)$ and that they and their first partial derivatives are continuous on it. Further suppose that the m-dimensional transformation defined on the range of $(X_1, \ldots, X_m)$ by $y_j = g_j(x_1, \ldots, x_m)$, for $1 \le j \le m$, is one-to-one. Then a joint PDF of the random variables $Y_1 = g_1(X_1, \ldots, X_m), \ldots, Y_m = g_m(X_1, \ldots, X_m)$ is

$$f_{Y_1,\ldots,Y_m}(y_1, \ldots, y_m) = \frac{1}{|J(x_1, \ldots, x_m)|} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)$$

for $(y_1, \ldots, y_m)$ in the range of $(Y_1, \ldots, Y_m)$, where $(x_1, \ldots, x_m)$ is the unique point in the range of $(X_1, \ldots, X_m)$ such that $g_j(x_1, \ldots, x_m) = y_j$ for $1 \le j \le m$. Here, $J(x_1, \ldots, x_m)$ is the Jacobian determinant of the transformation:

$$J(x_1, \ldots, x_m) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1}(x_1, \ldots, x_m) & \ldots & \dfrac{\partial g_1}{\partial x_m}(x_1, \ldots, x_m) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_m}{\partial x_1}(x_1, \ldots, x_m) & \ldots & \dfrac{\partial g_m}{\partial x_m}(x_1, \ldots, x_m) \end{vmatrix}.$$

Proof: For $(y_1, \ldots, y_m)$ in the range of $(Y_1, \ldots, Y_m)$, we have, from the multivariate version of the FPF, that

$$\begin{aligned} F_{Y_1,\ldots,Y_m}(y_1, \ldots, y_m) &= P(Y_1 \le y_1, \ldots, Y_m \le y_m) \\ &= P\big(g_1(X_1, \ldots, X_m) \le y_1, \ldots, g_m(X_1, \ldots, X_m) \le y_m\big) \\ &= \underset{g_j(x_1,\ldots,x_m) \le y_j,\ 1 \le j \le m}{\int \cdots \int} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)\, dx_1 \cdots dx_m. \end{aligned}$$

Making the transformation $s_j = g_j(x_1, \ldots, x_m)$, $1 \le j \le m$, and applying from calculus the change of variable formula for multiple integrals, we get

$$F_{Y_1,\ldots,Y_m}(y_1, \ldots, y_m) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_m} \frac{1}{|J(x_1, \ldots, x_m)|} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)\, ds_1 \cdots ds_m,$$

where, for each point $(s_1, \ldots, s_m)$, the point $(x_1, \ldots, x_m)$ is the unique point in the range of $(X_1, \ldots, X_m)$ such that $g_j(x_1, \ldots, x_m) = s_j$ for $1 \le j \le m$. The required result now follows from the multivariate version of Proposition 9.4 on page 499.

9.161 We apply Exercise 9.160 with $m = n$ and $g_i(x_1, \ldots, x_n) = \sum_{j=1}^n a_{ij}x_j$ for $1 \le i \le n$. Because the matrix $\mathbf{A}$ is nonsingular, the transformation $y_i = \sum_{j=1}^n a_{ij}x_j$, for $1 \le i \le n$, is one-to-one. For each i and j,

$$\frac{\partial g_i}{\partial x_j}(x_1, \ldots, x_n) = \frac{\partial}{\partial x_j}\left(\sum_{k=1}^n a_{ik}x_k\right) = a_{ij}.$$

Thus, the Jacobian determinant of the transformation $y_i = \sum_{j=1}^n a_{ij}x_j$, $1 \le i \le n$, is $\det \mathbf{A}$. Applying the multivariate transformation theorem now yields

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \frac{1}{|J(x_1, \ldots, x_n)|} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{1}{|\det \mathbf{A}|} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

for $(y_1, \ldots, y_n)$ in the range of $(Y_1, \ldots, Y_n)$, where $(x_1, \ldots, x_n)$ is the unique point in the range of $(X_1, \ldots, X_n)$ such that $\sum_{j=1}^n a_{ij}x_j = y_i$ for $1 \le i \le n$. Letting $\mathbf{x}$ and $\mathbf{y}$ denote the column vectors of $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, respectively, we have $\mathbf{y} = \mathbf{A}\mathbf{x}$ and, hence, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$.

Advanced Exercises

9.162

a) We apply Exercise 9.161 with

$$a_{ij} = \begin{cases} 1, & \text{if } j \le i; \\ 0, & \text{otherwise.} \end{cases}$$

From this result, we see that $\det \mathbf{A} = 1$. Now,

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = \lambda e^{-\lambda x_1}\cdots\lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda(x_1+\cdots+x_n)}$$

if $x_i \ge 0$ for $1 \le i \le n$, and $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 0$ otherwise. Therefore,

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \frac{1}{|\det \mathbf{A}|} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{1}{1}\lambda^n e^{-\lambda(x_1+\cdots+x_n)} = \lambda^n e^{-\lambda y_n}$$

if $0 < y_1 < \cdots < y_n$, and $f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = 0$ otherwise.

b) We note that the inverse transformation of the transformation in part (a) is $x_1 = y_1$ and $x_i = y_i - y_{i-1}$ for $2 \le i \le n$. Consequently, by referring to part (a), we see that the random variables defined by $X_1 = Y_1$ and $X_i = Y_i - Y_{i-1}$, for $2 \le i \le n$, must be independent, each with an $\mathcal{E}(\lambda)$ distribution. In particular, their joint PDF is the one given for $X_1, \ldots, X_n$ in part (a).

9.163 Let $X(t)$, $Y(t)$, and $Z(t)$ denote the x, y, and z coordinates, respectively, of the particle at time t. By assumption, $X(t)$, $Y(t)$, and $Z(t)$ are independent $\mathcal{N}\big(0, \sigma^2 t\big)$ random variables. Hence, their joint PDF is, for all $(x, y, z) \in \mathcal{R}^3$,

$$\begin{aligned} f_{X(t),Y(t),Z(t)}(x,y,z) &= \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{t}}\, e^{-x^2/2\sigma^2 t} \cdot \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{t}}\, e^{-y^2/2\sigma^2 t} \cdot \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{t}}\, e^{-z^2/2\sigma^2 t} \\ &= \frac{1}{\big(\sqrt{2\pi}\,\sigma\sqrt{t}\big)^3}\, e^{-(x^2+y^2+z^2)/2\sigma^2 t}. \end{aligned}$$

a) Recalling the definition of spherical coordinates, we have that $X(t) = P(t)\sin\big(\Phi(t)\big)\cos\big(\Theta(t)\big)$, $Y(t) = P(t)\sin\big(\Phi(t)\big)\sin\big(\Theta(t)\big)$, and $Z(t) = P(t)\cos\big(\Phi(t)\big)$. Hence, the inverse transformation is given by $x = \rho\sin\phi\cos\theta$, $y = \rho\sin\phi\sin\theta$, and $z = \rho\cos\phi$. The Jacobian determinant of the inverse transformation is

$$J(\rho,\phi,\theta) = \begin{vmatrix} \sin\phi\cos\theta & \sin\phi\sin\theta & \cos\phi \\ \rho\cos\phi\cos\theta & \rho\cos\phi\sin\theta & -\rho\sin\phi \\ -\rho\sin\phi\sin\theta & \rho\sin\phi\cos\theta & 0 \end{vmatrix} = \rho^2\sin\phi.$$

Recalling that the Jacobian determinant of an inverse transformation is the reciprocal of that of the transformation itself, we conclude from the multivariate transformation theorem that

$$\begin{aligned} f_{P(t),\Phi(t),\Theta(t)}(\rho,\phi,\theta) &= \frac{1}{|J(x,y,z)|} f_{X(t),Y(t),Z(t)}(x,y,z) = |J(\rho,\phi,\theta)|\, f_{X(t),Y(t),Z(t)}(x,y,z) \\ &= \rho^2\sin\phi \frac{1}{\big(\sqrt{2\pi}\,\sigma\sqrt{t}\big)^3}\, e^{-(x^2+y^2+z^2)/2\sigma^2 t} = \frac{\rho^2\sin\phi}{\big(\sqrt{2\pi}\,\sigma\sqrt{t}\big)^3}\, e^{-\rho^2/2\sigma^2 t} \end{aligned}$$

if $\rho > 0$, $0 < \phi < \pi$, and $0 < \theta < 2\pi$, and $f_{P(t),\Phi(t),\Theta(t)}(\rho,\phi,\theta) = 0$ otherwise.

b) We can apply the trivariate analogue of Proposition 9.7 on page 510 to determine the marginal PDFs of $P(t)$, $\Phi(t)$, and $\Theta(t)$. Alternatively, we can and do use the trivariate analogues of Exercises 9.95

and 9.96 on page 528. From part (a), we can write

$$f_{P(t),\Phi(t),\Theta(t)}(\rho,\phi,\theta) = \left(\frac{\sqrt{2/\pi}}{\left(\sigma\sqrt{t}\right)^3}\rho^2\, e^{-\rho^2/2\sigma^2 t}\right)\left(\frac{1}{2}\sin\phi\right)\left(\frac{1}{2\pi}\right)$$

if $\rho > 0$, $0 < \phi < \pi$, and $0 < \theta < 2\pi$, and $f_{P(t),\Phi(t),\Theta(t)}(\rho,\phi,\theta) = 0$ otherwise. It now follows immediately from the aforementioned exercises that

$$f_{P(t)}(\rho) = \begin{cases} \dfrac{\sqrt{2/\pi}}{\left(\sigma\sqrt{t}\right)^3}\rho^2\, e^{-\rho^2/2\sigma^2 t}, & \text{if } \rho > 0; \\ 0, & \text{otherwise.} \end{cases}$$

$$f_{\Phi(t)}(\phi) = \begin{cases} \dfrac{1}{2}\sin\phi, & \text{if } 0 < \phi < \pi; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_{\Theta(t)}(\theta) = \begin{cases} \dfrac{1}{2\pi}, & \text{if } 0 < \theta < 2\pi; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we see that $\Theta(t) \sim \mathcal{U}(0, 2\pi)$ and, referring to Exercise 8.160(d) on page 475, that $P(t)$ has the Maxwell distribution with parameter $\sigma^2 t$.

c) It follows immediately from part (b) and either the trivariate version of Exercise 9.95(d) or Proposition 9.9 on page 523 that $P(t)$, $\Phi(t)$, and $\Theta(t)$ are independent random variables.

9.164 By assumption, we have, for $y > 0$ and $z \in \mathcal{R}$,

$$f_Y(y) = \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2} \qquad \text{and} \qquad f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}.$$

Because Y and Z are independent, a joint PDF of Y and Z is the product of the marginals, or

$$f_{Y,Z}(y,z) = f_Y(y) f_Z(z) = \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2},$$

or

$$f_{Y,Z}(y,z) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)\sqrt{2\pi}}\; y^{\nu/2-1} e^{-(y+z^2)/2},$$

if $y > 0$ and $-\infty < z < \infty$, and $f_{Y,Z}(y,z) = 0$ otherwise.

We now let $T = Z/\sqrt{Y/\nu}$ and $U = Y$ (a "dummy variable"). The Jacobian determinant of the transformation $t = z/\sqrt{y/\nu}$ and $u = y$ is

$$J(y,z) = \begin{vmatrix} -\dfrac{z}{2\sqrt{y^3/\nu}} & \dfrac{1}{\sqrt{y/\nu}} \\ 1 & 0 \end{vmatrix} = -\frac{1}{\sqrt{y/\nu}}.$$

Solving the equations $t = z/\sqrt{y/\nu}$ and $u = y$ for y and z, we obtain the inverse transformation $y = u$ and $z = t\sqrt{u/\nu}$. Therefore, by the bivariate transformation theorem, a joint PDF of T and U is

$$\begin{aligned} f_{T,U}(t,u) &= \frac{1}{|J(y,z)|} f_{Y,Z}(y,z) = \frac{1}{|-1/\sqrt{y/\nu}|} f_{Y,Z}(y,z) \\ &= \sqrt{y/\nu}\, f_{Y,Z}(y,z) = \sqrt{u/\nu}\, f_{Y,Z}\big(u, t\sqrt{u/\nu}\big). \end{aligned}$$

Noting that $y + z^2 = u\left(1 + t^2/\nu\right)$, we get

$$
\begin{aligned}
f_{T,U}(t,u) &= \sqrt{u/\nu}\,\frac{1}{2^{\nu/2}\Gamma(\nu/2)\sqrt{2\pi}}\,u^{\nu/2-1}e^{-u\left(1+t^2/\nu\right)/2} \\
&= \frac{1}{2^{(\nu+1)/2}\Gamma(\nu/2)\sqrt{\nu\pi}}\,u^{(\nu+1)/2-1}e^{-u\left(1+t^2/\nu\right)/2},
\end{aligned}
$$

if $-\infty < t < \infty$ and $u > 0$, and $f_{T,U}(t,u) = 0$ otherwise.

To obtain a PDF of T, we apply Proposition 9.7 on page 510 and then use the fact that the integral of a gamma PDF is 1:

$$
\begin{aligned}
f_T(t) &= \int_{-\infty}^{\infty} f_{T,U}(t,u)\,du = \int_0^{\infty} \frac{1}{2^{(\nu+1)/2}\Gamma(\nu/2)\sqrt{\nu\pi}}\,u^{(\nu+1)/2-1}e^{-u\left(1+t^2/\nu\right)/2}\,du \\
&= \frac{1}{2^{(\nu+1)/2}\Gamma(\nu/2)\sqrt{\nu\pi}}\,\frac{\Gamma\big((\nu+1)/2\big)}{\Big(\big(1+t^2/\nu\big)/2\Big)^{(\nu+1)/2}} \\
&\quad\times \int_0^{\infty} \frac{\Big(\big(1+t^2/\nu\big)/2\Big)^{(\nu+1)/2}}{\Gamma\big((\nu+1)/2\big)}\,u^{(\nu+1)/2-1}e^{-u\left(1+t^2/\nu\right)/2}\,du \\
&= \frac{1}{2^{(\nu+1)/2}\Gamma(\nu/2)\sqrt{\nu\pi}}\,\frac{\Gamma\big((\nu+1)/2\big)}{\Big(\big(1+t^2/\nu\big)/2\Big)^{(\nu+1)/2}}.
\end{aligned}
$$

Therefore,

$$
f_T(t) = \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left(1+t^2/\nu\right)^{-(\nu+1)/2}, \qquad \infty < t < \infty.
$$

9.165 We use the bivariate transformation theorem. By assumption,

$$
\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= f_{X_1}(x_1)\cdot f_{X_2}(x_2) = \frac{(1/2)^{\nu_1/2}}{\Gamma(\nu_1/2)}x_1^{\nu_1/2-1}\,e^{-x_1/2}\cdot\frac{(1/2)^{\nu_2/2}}{\Gamma(\nu_2/2)}x_2^{\nu_2/2-1}\,e^{-x_2/2} \\
&= \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)}x_1^{\nu_1/2-1}x_2^{\nu_2/2-1}e^{-(x_1+x_2)/2}
\end{aligned}
$$

if $x_1 > 0$ and $x_2 > 0$, and $f_{X_1,X_2}(x_1,x_2) = 0$ otherwise. For convenience, set

$$
a = \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \qquad\text{and}\qquad b = \frac{\nu_1}{\nu_2}.
$$

Let $W = (X_1/\nu_1)\big/(X_2/\nu_2) = (1/b)(X_1/X_2)$ and $V = X_2$ (the "dummy variable"). The Jacobian determinant of the transformation $w = (1/b)(x_1/x_2)$ and $v = x_2$ is

$$
J(x_1,x_2) = \begin{vmatrix} \dfrac{1}{bx_2} & -\dfrac{x_1}{bx_2^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{bx_2}.
$$

Solving the equations $w = (1/b)(x_1/x_2)$ and $v = x_2$ for x_1 and x_2, we get the inverse transformation $x_1 = bvw$ and $x_2 = v$. Thus, by the bivariate transformation theorem,

$$\begin{aligned} f_{W,V}(w, v) &= \frac{1}{|J(x_1, x_2)|} f_{X_1,X_2}(x_1, x_2) = bv f_{X_1,X_2}(bvw, v) \\ &= abv(bvw)^{\nu_1/2-1} v^{\nu_2/2-1} e^{-v(1+bw)/2} = ab^{\nu_1/2} w^{\nu_1/2-1} v^{(\nu_1+\nu_2)/2-1} e^{-v(1+bw)/2} \end{aligned}$$

if $w > 0$ and $v > 0$, and $f_{W,V}(w, v) = 0$ otherwise. Therefore,

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} f_{W,V}(w, v)\, dv = ab^{\nu_1/2} w^{\nu_1/2-1} \int_0^{\infty} v^{(\nu_1+\nu_2)/2-1} e^{-v(1+bw)/2}\, dv \\ &= ab^{\nu_1/2} w^{\nu_1/2-1} \frac{\Gamma\big((\nu_1+\nu_2)/2\big)}{\big((1+bw)/2\big)^{(\nu_1+\nu_2)/2}} \\ &= \frac{(\nu_1/\nu_2)^{\nu_1/2}\Gamma\big((\nu_1+\nu_2)/2\big)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \frac{w^{\nu_1/2-1}}{\big(1+(\nu_1/\nu_2)w\big)^{(\nu_1+\nu_2)/2}} \end{aligned}$$

if $w > 0$, and $f_W(w) = 0$ otherwise.

Review Exercises for Chapter 9

Basic Exercises

9.166

a) Applying first the complementation rule and then the general addition rule, we get

$$\begin{aligned} P(X > x,\ Y > y) &= 1 - P(\{X \le x\} \cup \{Y \le y\}) \\ &= 1 - \big(P(X \le x) + P(Y \le y) - P(X \le x,\ Y \le y)\big) \\ &= 1 - F_X(x) - F_Y(y) + F_{X,Y}(x, y). \end{aligned}$$

b) In view of the result of part (a) and Proposition 9.2 on page 490,

$$\begin{aligned} P(X > x,\ Y > y) &= 1 - F_X(x) - F_Y(y) + F_{X,Y}(x, y) \\ &= 1 - F_{X,Y}(x, \infty) - F_{X,Y}(\infty, y) + F_{X,Y}(x, y). \end{aligned}$$

9.167

a) We have

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ xy, & \text{if } 0 \le x < 1 \text{ and } 0 \le y < 1; \\ x, & \text{if } 0 \le x < 1 \text{ and } y \ge 1; \\ y, & \text{if } x \ge 1 \text{ and } 0 \le y < 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

Therefore, from Proposition 9.1 on page 488,

$$\begin{aligned} P(X > 0.3,\ Y > 0.4) &= P(0.3 < X \le 1,\ 0.4 < Y \le 1) \\ &= F_{X,Y}(1, 1) - F_{X,Y}(1, 0.4) - F_{X,Y}(0.3, 1) + F_{X,Y}(0.3, 0.4) \\ &= 1 - 0.4 - 0.3 + (0.3)(0.4) = 0.42. \end{aligned}$$

b) We have

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \qquad \text{and} \qquad F_Y(y) = \begin{cases} 0, & \text{if } y < 0; \\ y, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

Therefore, from Exercise 9.166(a),

$$\begin{aligned} P(X > 0.3,\ Y > 0.4) &= 1 - F_X(0.3) - F_Y(0.4) + F_{X,Y}(0.3, 0.4) \\ &= 1 - 0.3 - 0.4 + (0.3)(0.4) = 0.42. \end{aligned}$$

c) In view of Exercise 9.166(b) and Proposition 9.2 on page 490,

$$\begin{aligned} P(X > 0.3,\ Y > 0.4) &= 1 - F_{X,Y}(0.3, \infty) - F_{X,Y}(\infty, 0.4) + F_{X,Y}(0.3, 0.4) \\ &= 1 - F_X(0.3) - F_Y(0.4) + F_{X,Y}(0.3, 0.4) \\ &= 1 - 0.3 - 0.4 + (0.3)(0.4) = 0.42. \end{aligned}$$

d) We have

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, by the FPF,

$$\begin{aligned} P(X > 0.3,\ Y > 0.4) &= \iint\limits_{x>0.3,\ y>0.4} f_{X,Y}(x, y)\, dx\, dy = \int_{0.3}^{1} \int_{0.4}^{1} 1\, dx\, dy \\ &= (1 - 0.3)(1 - 0.4) = 0.42. \end{aligned}$$

9.168

a) The sample space, Ω, for this random experiment is the rectangle $(a, b) \times (c, d)$. Because we are selecting a point at random, use of a geometric probability model is appropriate. Thus, for each event E,

$$P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{(b-a)(d-c)},$$

where $|E|$ denotes the area (in the extended sense) of the set E. The random variables X and Y are defined on Ω by $X(x, y) = x$ and $Y(x, y) = y$.

To find the joint CDF of X and Y, we begin by observing that the range of (X, Y) is Ω. Therefore, if either $x < a$ or $y < c$, we have $F_{X,Y}(x, y) = 0$; and, if $x > b$ and $y > d$, we have $F_{X,Y}(x, y) = 1$. We must still find $F_{X,Y}(x, y)$ in all other cases, of which we can consider three.

Case 1: $a \le x < b$ and $c \le y < d$

In this case, $\{X \le x,\ Y \le y\}$ is the event that the randomly selected point lies in the rectangle with vertices (a, c), (x, c), (x, y), and (a, y), specifically, $(a, x] \times (c, y]$. Hence,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y) = \frac{|\{X \le x,\ Y \le y\}|}{(b-a)(d-c)} = \frac{|(a, x] \times (c, y]|}{(b-a)(d-c)} = \frac{(x-a)(y-c)}{(b-a)(d-c)}.$$

Case 2: $a \le x < b$ and $y \ge d$

In this case, $\{X \le x,\ Y \le y\}$ is the event that the randomly selected point lies in the rectangle with vertices (a, c), (x, c), (x, d), and (a, d), specifically, $(a, x] \times (c, d)$. Hence,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y) = \frac{|\{X \le x,\ Y \le y\}|}{(b-a)(d-c)} = \frac{|(a, x] \times (c, d]|}{(b-a)(d-c)} = \frac{x-a}{b-a}.$$

Case 3: $x \geq b$ and $c \leq y < d$

In this case, $\{X \leq x,\ Y \leq y\}$ is the event that the randomly selected point lies in the rectangle with vertices (a, c), (b, c), (b, y), and (a, y), specifically, $(a, b) \times (c, y]$. Thus,

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \frac{|\{X \leq x,\ Y \leq y\}|}{(b-a)(d-c)} = \frac{|(a, b] \times (c, y]|}{(b-a)(d-c)} = \frac{y-c}{d-c}.$$

We conclude that the joint CDF of the random variables X and Y is

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < a \text{ or } y < c; \\ \dfrac{(x-a)(y-c)}{(b-a)(d-c)}, & \text{if } a \leq x < b \text{ and } c \leq y < d; \\ \dfrac{x-a}{b-a}, & \text{if } a \leq x < b \text{ and } y \geq d; \\ \dfrac{y-c}{d-c}, & \text{if } x \geq b \text{ and } c \leq y < d; \\ 1, & \text{if } x \geq b \text{ and } y \geq d. \end{cases}$$

b) Referring to Proposition 9.2 on page 490 and to part (a), we find that

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < a; \\ \dfrac{x-a}{b-a}, & \text{if } a \leq x < b; \\ 1, & \text{if } x \geq b. \end{cases}$$

and

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } y < c; \\ \dfrac{y-c}{d-c}, & \text{if } c \leq y < d; \\ 1, & \text{if } y \geq d. \end{cases}$$

c) From the form of the marginal CDFs obtained in part (b), we see that $X \sim \mathcal{U}(a, b)$ and $Y \sim \mathcal{U}(c, d)$.
d) From parts (a) and (b), we find that the joint CDF of X and Y equals the product of the marginal CDFs of X and Y.
e) From part (d), we have $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all $(x, y) \in \mathcal{R}^2$. Thus,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) = \int_{-\infty}^{x} f_X(s)\,ds \cdot \int_{-\infty}^{y} f_Y(t)\,dt = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(s)f_Y(t)\,ds\,dt.$$

Therefore, from Proposition 9.4, the function f defined on $\mathcal{R}^2$ by $f(x, y) = f_X(x)f_Y(y)$ is a joint PDF of X and Y. Now applying Proposition 9.9 on page 523, we conclude that X and Y are independent random variables.

9.169

a) Taking the mixed partial of the joint CDF found in the solution to Exercise 9.168(a), we find that

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y) = \begin{cases} \dfrac{1}{(b-a)(d-c)}, & \text{if } a < x < b \text{ and } c < y < d; \\ 0, & \text{otherwise.} \end{cases}$$

b) Because a point is being selected at random from the rectange $R = (a, b) \times (c, d)$, a joint PDF of X and Y should be a nonzero constant, say, k, on R, and zero elsewhere. As a joint PDF must integrate to 1, we have

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = \int_{a}^{b} \int_{c}^{d} k\, dx\, dy = k(b-a)(d-c).$$

Hence, $k = \big((b-a)(d-c)\big)^{-1}$, so that

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{(b-a)(d-c)}, & \text{if } a < x < b \text{ and } c < y < d; \\ 0, & \text{otherwise.} \end{cases}$$

9.170 Suppose that F is the joint CDF of the random variables X and Y. Then, from Proposition 9.1 on page 488,

$$\begin{aligned} P(-1 < X \le 1,\ -1 < Y \le 1) &= F(1, 1) - F(1, -1) - F(-1, 1) + F(-1, -1) \\ &= 1 - 1 - 1 + 0 = -1, \end{aligned}$$

which is impossible because a probability cannot be negative. Hence, F can't be the joint CDF of two random variables.

9.171 Completing the square relative to the variable x, we get

$$x^2 + xy + y^2 = x^2 + xy + \frac{1}{4}y^2 - \frac{1}{4}y^2 + y^2 = (x + 1/2)^2 + (3/4)y^2.$$

Then,

$$e^{-2\left(x^2+xy+y^2\right)/3} = e^{-2\left((x+1/2)^2+(3/4)y^2\right)/3} = e^{-y^2/2} e^{-2\left((x+1/2)^2\right)/3} = e^{-y^2/2} e^{-\left(x-(-1/2)\right)^2 \big/ 2\left(\sqrt{3}/2\right)^2}$$

Consequently,

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, dx\, dy &= \int_{-\infty}^{\infty} e^{-y^2/2} \left(\int_{-\infty}^{\infty} e^{-\left(x-(-1/2)\right)^2 \big/ 2\left(\sqrt{3}/2\right)^2} dx \right) dy \\ &= \int_{-\infty}^{\infty} \left(\sqrt{2\pi}\sqrt{3}/2\right) e^{-y^2/2}\, dy = \sqrt{2\pi}\sqrt{3}/2 \int_{-\infty}^{\infty} e^{-y^2/2}\, dy \\ &= \left(\sqrt{2\pi}\sqrt{3}/2\right)\left(\sqrt{2\pi}\right) = \sqrt{3}\pi. \end{aligned}$$

Hence, g is the "variable form" of a joint PDF of two random variables. The actual joint PDF is the function f defined on $\mathcal{R}^2$ by

$$f(x, y) = \frac{1}{\sqrt{3}\pi} g(x, y) = \frac{1}{\sqrt{3}\pi} e^{-2\left(x^2+xy+y^2\right)/3}.$$

9.172 We want to determine $P(X + Y \ge 1)$. We can obtain that probability by using the bivariate FPF or we can use Equation (9.44) on page 534 and the univariate FPF. Let's use the bivariate FPF:

$$\begin{aligned} P(X + Y \ge 1) &= \iint\limits_{x+y\ge 1} f_{X,Y}(x, y)\, dx\, dy = \frac{1}{4} \int_0^1 \left(\int_{1-x}^{2} (2x + 2 - y)\, dy \right) dx \\ &= \frac{1}{8} \int_0^1 \left(6x + 1 + 5x^2\right) dx = \frac{17}{24}. \end{aligned}$$

9.173

a) Referring to Proposition 9.7 on page 510, we get

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_x^{\infty} e^{-y}\, dy = e^{-x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Hence, $X \sim \mathcal{E}(1)$. Also,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_0^{y} e^{-y}\, dx = ye^{-y}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. Hence, $Y \sim \Gamma(2, 1)$.

b) We can show that X and Y are not independent random variables in several ways. One way is to note that, for $x > 0$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}$$

if $y > x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. As $f_{Y\,|\,X}(y\,|\,x)$ is not a PDF of Y, Proposition 9.10(a) on page 527 shows that X and Y are not independent.

c) From Proposition 9.12 on page 534, we have, for $z > 0$,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\, dy.$$

The integrand is nonzero if and only if $0 < z - y < y$, that is, if and only if $z/2 < y < z$. Hence,

$$f_{X+Y}(z) = \int_{z/2}^{z} e^{-y}\, dy = e^{-z/2} - e^{-z} = e^{-z}\big(e^{z/2} - 1\big)$$

if $z > 0$, and $f_{X+Y}(z) = 0$ otherwise.

d) For $y > 0$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y}$$

if $0 < x < y$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise. Hence, $X_{|Y=y} \sim \mathcal{U}(0, y)$ for $y > 0$.

e) We have

$$P(Y < 5) = \int_{-\infty}^{5} f_Y(y)\, dy = \int_0^{5} ye^{-y}\, dy = 1 - 6e^{-5}$$

and

$$P(X > 3,\ Y < 5) = \iint\limits_{x>3,\ y<5} f_{X,Y}(x, y)\, dx\, dy = \int_3^5 \left(\int_x^5 e^{-y}\, dy\right) dx$$

$$= \int_3^5 \left(e^{-x} - e^{-5}\right) dx = e^{-3} - 3e^{-5}.$$

Hence,

$$P(X > 3\,|\,Y < 5) = \frac{P(X > 3,\ Y < 5)}{P(Y < 5)} = \frac{e^{-3} - 3e^{-5}}{1 - 6e^{-5}} = 0.0308.$$

f) From part (d), $X_{|Y=5} \sim \mathcal{U}(0, 5)$. Hence,

$$P(X > 3\,|\,Y = 5) = (5 - 3)/5 = 0.4.$$

9.174

a) Jane will catch the bus if and only if she arrives no later than the bus does and the bus arrives no later than 5 minutes after Jane does, which is the event $\{Y \le X \le Y+5\}$. Thus, by the FPF and independence, the required probability is

$$P(Y \le X \le Y+5) = \iint\limits_{y \le x \le y+5} f_{X,Y}(x,y)\,dx\,dy = \iint\limits_{y \le x \le y+5} f_X(x) f_Y(y)\,dx\,dy.$$

b) Let time be measured in minutes after 7:00 A.M.. Then $X \sim \mathcal{U}(0,30)$ and $Y \sim \mathcal{U}(5,20)$. Hence, from part (a),

$$P(Y \le X \le Y+5) = \iint\limits_{y \le x \le y+5} f_X(x) f_Y(y)\,dx\,dy = \int_{-\infty}^{\infty}\left(\int_y^{y+5} f_X(x)\,dx\right) f_Y(y)dy$$

$$= \int_5^{20}\left(\int_y^{y+5} \frac{1}{30}\,dx\right)\frac{1}{15}\,dy = \frac{1}{450}\int_5^{20} 5\,dy = \frac{1}{6}.$$

9.175

a) We have

$$P(A_n) = P(a < X_1 \le b, \dots, a < X_n \le b) = P(\min\{X_1,\dots,X_n\} > a,\ \max\{X_1,\dots,X_n\} \le b)$$
$$= P(X > a,\ Y \le b) = \big(F(b) - F(a)\big)^n.$$

b) The event that any particular X_k falls in the interval $(a,b]$ has probability $F(b) - F(a)$. Therefore, by independence, the number of X_ks that fall in the interval $(a,b]$ has the binomial distribution with parameters n and $F(b) - F(a)$. Hence,

$$P(A_k) = \binom{n}{k}\big(F(b) - F(a)\big)^k\big(1 - F(b) + F(a)\big)^{n-k}, \qquad k = 0, 1, \dots, n.$$

c)

k	$P(A_k)$
0	0.0081
1	0.0756
2	0.2646
3	0.4116
4	0.2401

d)

k	$P(A_k)$
0	0.1194
1	0.3349
2	0.3522
3	0.1646
4	0.0289

9.176

a) Because a point is selected at random, a geometric probability model is appropriate. The sample space is S, the unit sphere, which has volume $4\pi/3$. Letting $|\cdot|$ denote three-dimensional volume, we have, for $E \subset \mathcal{R}^3$,

$$P(E) = \frac{|E \cap S|}{|S|} = \frac{|E \cap S|}{4\pi/3} = \frac{3}{4\pi}|E \cap S|.$$

b) Because a point is being selected at random from S, a joint PDF of X, Y, and Z should be a nonzero constant, say, k, on S, and zero elsewhere. As a joint PDF must integrate to 1, we have

$$1 = \iiint\limits_{\mathcal{R}^3} f_{X,Y,Z}(x,y,z)\,dx\,dy\,dz = \iiint\limits_{S} k\,dx\,dy\,dz = k|S| = \frac{4\pi}{3}k.$$

Hence, $k = 3/4\pi$, so that

$$f_{X,Y,Z}(x, y, z) = \begin{cases} \dfrac{3}{4\pi}, & \text{if } (x, y, z) \in S; \\ 0, & \text{otherwise.} \end{cases}$$

c) From the FPF and part (b),

$$P(E) = P\big((X, Y, Z) \in E\big) = \iiint\limits_{E} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz = \iiint\limits_{E\cap S} \frac{3}{4\pi}\,dx\,dy\,dz = \frac{3}{4\pi}|E \cap S|.$$

d) We note that, for $-1 < x < 1$, the intersection of the plane $s = x$ with the unit sphere is, in y-z space, the disk $D_x = \{\,(y, z) : y^2 + z^2 < 1 - x^2\,\}$. Thus,

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\,dy\,dz = \iint\limits_{D_x} \frac{3}{4\pi}\,dy\,dz = \frac{3}{4\pi}|D_x| = \frac{3}{4\pi}\pi\big(1 - x^2\big) = \frac{3}{4}\big(1 - x^2\big)$$

if $-1 < x < 1$, and $f_X(x) = 0$ otherwise. Here, $|\cdot|$ denotes two-dimensional volume, that is, area. By symmetry, Y and Z have the same marginal PDFs as X.
e) We note that, for $x^2 + y^2 < 1$, the intersection of the planes $s = x$ and $t = y$ with the unit sphere is, in z space, the interval $I_{x,y} = \left\{\, z : -\sqrt{1 - x^2 - y^2} < z < \sqrt{1 - x^2 - y^2}\,\right\}$. Hence,

$$f_{X,Y}(x, y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\,dz = \int_{I_{x,y}} \frac{3}{4\pi}\,dz = \frac{3}{4\pi}\cdot 2\sqrt{1 - x^2 - y^2} = \frac{3}{2\pi}\sqrt{1 - x^2 - y^2}$$

if $x^2 + y^2 < 1$, and $f_{X,Y}(x, y) = 0$ otherwise. By symmetry, the joint PDF of X and Z and the joint PDF of Y and Z are the same as the joint PDF of X and Y.
f) From parts (b) and (e), we have, for $x^2 + y^2 < 1$,

$$f_{Z\,|\,X,Y}(z\,|\,x, y) = \frac{f_{X,Y,Z}(x, y, z)}{f_{X,Y}(x, y)} = \frac{3/4\pi}{(3/2\pi)\sqrt{1 - x^2 - y^2}} = \frac{1}{2\sqrt{1 - x^2 - y^2}}$$

if $z \in I_{x,y}$, and $f_{Z\,|\,X,Y}(z\,|\,x, y) = 0$ otherwise. Hence, for $x^2 + y^2 < 1$, $Z_{|X=x,Y=y} \sim \mathcal{U}(I_{x,y})$.
g) From parts (b) and (d), we have, for $-1 < x < 1$,

$$f_{Y,Z\,|\,X}(y, z\,|\,x) = \frac{f_{X,Y,Z}(x, y, z)}{f_X(x)} = \frac{3/4\pi}{(3/4)(1 - x^2)} = \frac{1}{\pi(1 - x^2)}$$

if $(y, z) \in D_x$, and $f_{Y,Z\,|\,X}(y, z\,|\,x) = 0$ otherwise. Hence, for $-1 < x < 1$, $(Y, Z)_{|X=x} \sim \mathcal{U}(D_x)$.
h) We can see in several ways that X, Y, and Z are not independent random variables. One way is to note from part (f) that $f_{Z\,|\,X,Y}(z\,|\,x, y)$ is not a PDF of Z, as it depends on x and y.

9.177 Let R denote the distance of the point chosen from the origin. Note that, for $0 < r < 1$, the event $\{R \le r\}$ occurs if and only if the point chosen lies in the sphere, S_r, of radius r centered at the origin. Hence, from Exercise 9.176(a), we have, for $0 < r < 1$,

$$F_R(r) = P(R \le r) = \frac{3}{4\pi}|S_r| = \frac{3}{4\pi}\frac{4\pi r^3}{3} = r^3.$$

Differentiation now yields

$$f_R(r) = \begin{cases} 3r^2, & \text{if } 0 < r < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, R has the beta distribution with parameters 3 and 1.

9.178 Let S and T denote the lifetimes of the two components. Then system lifetime is $Z = \min\{S, T\}$. From the FPF,

$$\begin{aligned} P(Z \le 0.5) &= P(\min\{S, T\} \le 0.5) = P(S \le 0.5 \text{ or } T \le 0.5) \\ &= P(S > 0.5,\ T \le 0.5) + P(S \le 0.5) \\ &= \int_0^{0.5} \left(\int_{0.5}^1 f(s,t)\,ds \right) dt + \int_0^1 \left(\int_0^{0.5} f(s,t)\,ds \right) dt. \end{aligned}$$

Hence, the correct answer is (e).

9.179 We must have $0 \le Y \le X \le 1$. Hence,

$$f_{X,Y}(x,y) = \begin{cases} 2(x+y), & \text{if } 0 < y < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$f_X(0.10) = \int_{-\infty}^{\infty} f_{X,Y}(0.10, y)\,dy = \int_0^{0.10} 2(0.10 + y)\,dy = 0.03.$$

Therefore,

$$f_{Y\,|\,X}(y\,|\,0.10) = \frac{f_{X,Y}(0.10, y)}{f_X(0.10)} = \frac{2(0.10+y)}{0.03} = \frac{200}{3}(0.10 + y)$$

if $0 < y < 0.10$, and $f_{Y\,|\,X}(y\,|\,0.10) = 0$ otherwise. Consequently,

$$P(Y < 0.05\,|\,X = 0.10) = \int_{-\infty}^{0.05} f_{Y\,|\,X}(y\,|\,0.10)\,dy = \int_0^{0.05} \frac{200}{3}(0.10 + y)\,dy = 0.417.$$

9.180 The sample space is D. Because a point is selected at random, a geometric probability model is appropriate. Thus, for each event E,

$$P(E) = \frac{|E|}{|D|} = \frac{|E|}{\sqrt{2}}$$

where $|\cdot|$ denotes length (in the extended sense).

a) Clearly, $F_{X,Y}(x,y) = 0$ if $x < 0$ or $y < 0$, and $F_{X,Y}(x,y) = 1$ if $x \ge 1$ and $y \ge 1$. Now suppose that $0 \le y < \min\{x, 1\}$. Then

$$F_{X,Y}(x,y) = P(X \le x,\ Y \le y) = \frac{|\{\,(s,t) \in D : 0 < t < y\,\}|}{\sqrt{2}} = \frac{\sqrt{2}y}{\sqrt{2}} = y.$$

By symmetry, for $0 \le x < \min\{1, y\}$, we have $F_{X,Y}(x,y) = x$. Hence,

$$F_{X,Y}(x,y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ x, & \text{if } 0 \le x < \min\{1, y\}; \\ y, & \text{if } 0 \le y < \min\{1, x\}; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

b) Referring first to Proposition 9.2 on page 490 and then to part (a), we get

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

By symmetry, Y has the same CDF as X.

c) Referring to part (b), we have

$$f_X(x) = F'_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, $X \sim \mathcal{U}(0, 1)$. By symmetry, $Y \sim \mathcal{U}(0, 1)$.
d) Suppose to the contrary that X and Y have a joint PDF. Then, by the FPF,

$$1 = P\big((X, Y) \in D\big) = \iint_D f_{X,Y}(x, y)\,dx\,dy = \int_0^1 \left(\int_x^x f_{X,Y}(x, y)\,dy \right) dx = \int_0^1 0\,dx = 0,$$

Thus, we have a contradiction, which means that X and Y don't have a joint PDF.

9.181 Let X, Y, and Z be the annual losses due to storm, fire, and theft, respectively, Then, by assumption, X, Y, and Z are independent $\mathcal{E}(1)$, $\mathcal{E}(2/3)$, and $\mathcal{E}(5/12)$ random variables. Let $W = \max\{X, Y, Z\}$. We want to find $P(W > 3)$. Applying first the complementation rule and then independence, we get

$$\begin{aligned} P(W > 3) &= 1 - P(W \le 3) = 1 - P\big(\max\{X, Y, Z\} \le 3\big) = 1 - P(X \le 3,\ Y \le 3,\ Z \le 3) \\ &= 1 - P(X \le 3)P(Y \le 3)P(Z \le 3) = 1 - \left(1 - e^{-1\cdot 3}\right)\left(1 - e^{-(2/3)\cdot 3}\right)\left(1 - e^{-(5/12)\cdot 3}\right) \\ &= 0.414. \end{aligned}$$

9.182 Let X, Y, and Z denote the arrival times, which, by assumption are independent $\mathcal{U}(-5, 5)$ random variables.
a) The probability that all three people arrive before noon is

$$P(X < 0,\ Y < 0,\ Z < 0) = P(X < 0)P(Y < 0)P(Z < 0) = \left(\frac{0 - (-5)}{10}\right)^3 = \frac{1}{8}.$$

b) The probability that any particular one of the three persons arrives after 12:03 P.M. is 0.2. Let W denote the number of the three people that arrive after 12:03 P.M.. Then $W \sim \mathcal{B}(3, 0.2)$. Thus,

$$P(W = 1) = \binom{3}{1}(0.2)^1(1 - 0.2)^{3-1} = 0.384.$$

c) A joint PDF of X, Y, and Z is

$$f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z) = \frac{1}{10} \cdot \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{1000}$$

if $-5 < x, y, z < 5$, and $f_{X,Y,Z}(x, y, z) = 0$ otherwise. By symmetry, the required probability equals

$$\begin{aligned} 3P(Z \ge X + 3,\ Z \ge Y + 3) &= 3 \iiint\limits_{z \ge x+3,\ z \ge y+3} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz \\ &= \frac{3}{1000} \int_{-2}^5 \left(\int_{-5}^{z-3} \left(\int_{-5}^{z-3} 1\,dy \right) dx \right) dz \\ &= \frac{3}{1000} \int_{-2}^5 \left(\int_{-5}^{z-3} (z + 2)\,dx \right) dz = \frac{3}{1000} \int_{-2}^5 (z + 2)^2\,dz \\ &= \frac{3}{1000} \int_0^7 u^2\,du = 0.343. \end{aligned}$$

9.183
a) We have

$$|C| = \iint_C 1\,dx\,dy = \int_{-\infty}^{\infty}\left(\int_0^{f(x)} 1\,dy\right)dx = \int_{-\infty}^{\infty} f(x)\,dx = 1,$$

where the last equality follows from the fact that f is a PDF. Therefore, a joint PDF of X and Y is $f_{X,Y}(x, y) = 1/|C| = 1$ if $(x, y) \in C$, and $f_{X,Y}(x, y) = 0$ otherwise.
b) From Proposition 9.7 on page 510,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = \int_0^{f(x)} 1\,dy = f(x).$$

c) For fixed y, let $A(y) = \{x : f(x) \geq y\}$. Then, from Proposition 9.7,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{A(y)} 1\,dx = |A(y)|,$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.
d) Referring to parts (a) and (b), we find that, for $f(x) > 0$,

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

if $0 < y < f(x)$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. Hence, $Y_{|X=x} \sim \mathcal{U}\big(0, f(x)\big)$ for all x with $f(x) > 0$.
e) Referring to parts (a) and (c), we find that, for $|A(y)| > 0$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{1}{|A(y)|}$$

if $x \in A(y)$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise. Hence, $X_{|Y=y} \sim \mathcal{U}\big(A(y)\big)$ for all $y > 0$ with $|A(y)| > 0$.

9.184 Let X and Y be two independent observations of the random variable. We want to determine $P(X \leq 1,\ Y \leq 1)$. By independence, and upon making the substitution $u = e^x$, we get

$$P(X \leq 1,\ Y \leq 1) = P(X \leq 1)P(Y \leq 1) = \left(\int_{-\infty}^{1} \frac{2}{\pi\left(e^{-x} + e^x\right)}\,dx\right)^2$$
$$= \left(\frac{2}{\pi}\int_0^e \frac{du}{1 + u^2}\right)^2 = \left(\frac{2}{\pi}\arctan e\right)^2 = 0.602.$$

9.185 Let X and Y denote claim amount and processing time, respectively. By assumption,

$$f_X(x) = \begin{cases} (3/8)x^2, & \text{if } 0 < x < 2; \\ 0, & \text{otherwise.} \end{cases}$$

and, for $0 < x < 2$,

$$f_{Y\,|\,X}(y\,|\,x) = \begin{cases} 1/x, & \text{if } x < y < 2x; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, from the general multiplication rule, Equation (9.34) on page 517,

$$f_{X,Y}(x, y) = f_X(x) f_{Y\,|\,X}(y\,|\,x) = (3/8)x^2 \cdot \frac{1}{x} = (3/8)x$$

if $x < y < 2x$ and $0 < x < 2$, and $f_{X,Y}(x, y) = 0$ otherwise. Consequently, from Proposition 9.7 on page 510, we have, for $0 < y < 2$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{y/2}^{y} (3/8)x\,dx = \frac{9}{64}y^2$$

and, for $2 < y < 4$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{y/2}^{2} (3/8)x\,dx = \frac{3}{64}(16 - y^2).$$

Therefore,

$$P(1 \le Y \le 3) = \int_1^3 f_Y(y),\ dy = \int_1^2 \frac{9}{64}y^2\,dy + \int_2^3 \frac{3}{64}(16 - y^2)\,dy = 0.781.$$

9.186 We have

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2h} f_Y(y)$$

if $-h < x < h$, and $f_{X,Y}(x, y) = 0$ otherwise.

a) Applying the bivariate FPF and making the substitution $u = z - x$, we get

$$F_{X+Y}(z) = P(X + Y \le z) = \iint\limits_{x+y\le z} f_{X,Y}(x, y)\,dx\,dy = \frac{1}{2h}\int_{-h}^{h}\left(\int_{-\infty}^{z-x} f_Y(y)\,dy\right)dx$$

$$= \frac{1}{2h}\int_{-h}^{h} F_Y(z - x)\,dx = \frac{1}{2h}\int_{z-h}^{z+h} F_Y(u)\,du.$$

b) Differentiating the result in part (a) yields,

$$f_{X+Y}(z) = F'_{X+Y}(z) = \frac{1}{2h}\big(F_Y(z + h)\cdot 1 - F_Y(z - h)\cdot 1\big)$$

$$= \frac{1}{2h}P(z - h < Y \le z + h) = \frac{1}{2h}\int_{z-h}^{z+h} f_Y(u)\,du.$$

c) Applying first Proposition 9.12 and then making the substitution $u = z - x$, we get

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\,dx = \frac{1}{2h}\int_{-h}^{h} f_Y(z - x)\,dx = \frac{1}{2h}\int_{z-h}^{z+h} f_Y(u)\,du.$$

d) From part (c) and the FEF,

$$F_{X+Y}(z) = \int_{-\infty}^{z} f_{X+Y}(t)\,dt = \int_{-\infty}^{z}\left(\frac{1}{2h}\int_{t-h}^{t+h} f_Y(u)\,du\right)dt$$

$$= \frac{1}{2h}\int_{-\infty}^{z}\big(F_Y(t + h) - F_Y(t - h)\big)\,dt = \frac{1}{2h}\left(\int_{-\infty}^{z} F_Y(t + h)\,dt - \int_{-\infty}^{z} F_Y(t - h)\,dt\right)$$

$$= \frac{1}{2h}\left(\int_{-\infty}^{z+h} F_Y(u)\,du - \int_{-\infty}^{z-h} F_Y(u)\,du\right) = \frac{1}{2h}\int_{z-h}^{z+h} F_Y(u)\,du.$$

9.187 Let X denote the time, in minutes, that it takes the device to warm up, and let Y denote the lifetime, in minutes, of the device after it warms up. By assumption, X and Y are independent $\mathcal{U}(1, 3)$ and $\mathcal{E}(1/600)$ random variables, respectively, and $T = X + Y$. We have

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2} \cdot \frac{1}{600} e^{-y/600} = \frac{1}{1200} e^{-y/600}$$

if $1 < x < 3$ and $y > 0$, and $f_{X,Y}(x, y) = 0$ otherwise. From Proposition 9.12 on page 534, for $t > 1$,

$$f_T(t) = \int_{-\infty}^{\infty} f_{X,Y}(x, t - x)\, dx.$$

The integrand is nonzero if and only if $1 < x < 3$ and $t - x > 0$, that is, if and only if $1 < x < \min\{t, 3\}$. Thus,

$$\begin{aligned} f_T(t) &= \frac{1}{1200} \int_1^{\min\{t,3\}} e^{-(t-x)/600}\, dx = \frac{e^{-t/600}}{1200} \int_1^{\min\{t,3\}} e^{x/600}\, dx \\ &= \frac{e^{-t/600}}{2} \left(e^{\min\{t,3\}/600} - e^{1/600} \right) = \begin{cases} \frac{1}{2} \left(1 - e^{-(t-1)/600} \right), & 1 < t < 3; \\ \frac{1}{2} \left(e^{-(t-3)/600} - e^{-(t-1)/600} \right), & t > 3; \end{cases} \end{aligned}$$

and $f_T(t) = 0$ otherwise.

9.188 Let $X(t)$ and $Y(t)$ denote the x and y coordinates, respectively, at time t. By assumption, $X(t)$ and $Y(t)$ are independent $\mathcal{N}(0, \sigma^2 t)$ random variables. Thus, by Example 9.23 on page 541, the random variable $U(t) = Y(t)/X(t)$ has the standard Cauchy distribution. Now, we have that $\Theta(t) = \arctan U(t)$. Applying the univariate transformation theorem with $g(u) = \arctan u$, we get, for $-\pi/2 < \theta < \pi/2$,

$$f_{\Theta(t)}(\theta) = \frac{1}{|g'(u)|} f_U(u) = (1 + u^2) \cdot \frac{1}{\pi(1 + u^2)} = \frac{1}{\pi}.$$

Thus, $\Theta(t) \sim \mathcal{U}(-\pi/2, \pi/2)$.

9.189

a) Let Z denote the fraction of harmful impurities. Then $Z = XY$. Using the FPF, independence, and making the substitution $u = xy$, we get

$$\begin{aligned} F_Z(z) = P(XY \le z) &= \iint\limits_{xy \le z} f_{X,Y}(x, y)\, dx\, dy = \int_0^1 \left(\int_0^{z/x} f_Y(y)\, dy \right) f_X(x)\, dx \\ &= \int_0^1 \left(\int_0^z \frac{1}{x} f_Y(u/x)\, du \right) f_X(x)\, dx = \int_0^z \left(\int_0^1 \frac{1}{x} f_X(x) f_Y(u/x)\, dx \right) du. \end{aligned}$$

Therefore, from Proposition 8.5 on page 422,

$$f_Z(z) = \int_0^1 \frac{1}{x} f_X(x) f_Y(z/x)\, dx$$

if $0 < z < 1$, and $f_Z(z) = 0$ otherwise.

b) We have

$$f_X(x) = \begin{cases} 10, & \text{if } 0 < x < 0.1; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} 2, & \text{if } 0 < y < 0.5; \\ 0, & \text{otherwise.} \end{cases}$$

From part (a), then,

$$f_Z(z) = \int_0^1 \frac{1}{x} f_X(x) f_Y(z/x)\, dx = \int_{2z}^{0.1} \frac{1}{x}(10 \cdot 2)\, dx = 20\big(\ln(0.1) - \ln(2z)\big) = -20\ln(20z)$$

if $0 < z < 0.05$, and $f_Z(z) = 0$ otherwise. The PDF of Z is strictly decreasing on the range of Z, which, in particular, means that the fraction of harmful impurities is more likely to be small (close to 0) than large (close to 0.05).

9.190
a) From Proposition 9.7 on page 510,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_0^x 8xy\, dy = 4x^3$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise. Thus, X has the beta distribution with parameters 4 and 1.
b) From Proposition 9.7,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_y^1 8xy\, dx = 4y(1 - y^2)$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise.
c) Let Z denote the amount of the item remaining at the end of the day, so that $Z = X - Y$. Then, by the FPF, we have, for $0 < z < 1$,

$$\begin{aligned} F_Z(z) = P(Z \le z) = P(X - Y \le z) &= \iint\limits_{x-y \le z} f_{X,Y}(x, y)\, dx\, dy \\ &= \int_0^z \left(\int_0^x 8xy\, dy \right) dx + \int_z^1 \left(\int_{x-z}^x 8xy\, dy \right) dx \\ &= \int_0^z 4x^3\, dx + \int_z^1 (8x^2 z - 4xz^2)\, dx = \frac{1}{3}z^4 - 2z^2 + \frac{8}{3}z. \end{aligned}$$

Differentiation now yields

$$f_Z(z) = \frac{4}{3}z^3 - 4z + \frac{8}{3}$$

if $0 < z < 1$, and $f_Z(z) = 0$ otherwise.
d) By the FPF,

$$\begin{aligned} P(X - Y < X/3) = P(Y > 2X/3) &= \iint\limits_{y > 2x/3} f_{X,Y}(x, y)\, dx\, dy = \int_0^1 \left(\int_{2x/3}^x y\, dy \right) 8x\, dx \\ &= \int_0^1 \left(\frac{5}{18}x^2 \right) 8x\, dx = \frac{20}{9} \int_0^1 x^3\, dx = \frac{5}{9}. \end{aligned}$$

9.191 Let $X_1, \ldots, X_{25}$ denote the 25 randomly selected claims. We know that $X_k \sim \mathcal{N}(19.4, 25)$ for $1 \le k \le 25$, and we can reasonably assume that $X_1, \ldots, X_{25}$ are independent random variables.

Now, from Proposition 9.14 on page 540,

$$\begin{aligned}\bar{X}_{25} &= \frac{X_1 + \cdots + X_{25}}{25} = \frac{1}{25}X_1 + \cdots + \frac{1}{25}X_{25} \\ &\sim \mathcal{N}\left(\frac{1}{25}\cdot 19.4 + \cdots + \frac{1}{25}\cdot 19.4, \left(\frac{1}{25}\right)^2 \cdot 25 + \cdots + \left(\frac{1}{25}\right)^2 \cdot 25\right) = \mathcal{N}(19.4, 1).\end{aligned}$$

Therefore, from Proposition 8.11 on page 443,

$$P(\bar{X}_{25} > 20) = 1 - P(\bar{X}_{25} \le 20) = 1 - \Phi\left(\frac{20 - 19.4}{1}\right) = 0.274.$$

9.192 Let E denote the event that the total claim amount of Company B will exceed that of Company A. Also, let X and Y denote the claim amounts for Companies A and B, respectively, given a positive claim. By assumption, $X \sim \mathcal{N}(10, 4)$ and $Y \sim \mathcal{N}(9, 4)$, and X and Y are independent. First we find $P(Y > X)$. From the independence assumption and Proposition 9.14 on page 540, $Y - X \sim \mathcal{N}(-1, 8)$. Therefore, from Proposition 8.11 on page 443,

$$P(Y > X) = P(Y - X > 0) = 1 - P(Y - X \le 0) = 1 - \Phi\left(\frac{0 - (-1)}{\sqrt{8}}\right) = 0.362.$$

Next let

$A =$ event that no claim is made on either Company A or Company B,

$B =$ event that no claim is made on Company A but at least one claim is made on Company B,

$C =$ event that at least one claim is made on Company A but no claim is made on Company B,

$D =$ event that at least one claim is made on both Company A and Company B.

Applying the law of total probability now yields

$$\begin{aligned}P(E) &= P(A)P(E \mid A) + P(B)P(E \mid B) + P(C)P(E \mid C) + P(D)P(E \mid D) \\ &= (0.6)(0.7)\cdot 0 + (0.6)(0.3)\cdot 1 + (0.4)(0.7)\cdot 0 + (0.4)(0.3)\cdot 0.362 \\ &= 0.223.\end{aligned}$$

9.193 By assumption

$$f_I(i) = \begin{cases} 6i(1-i), & 0 < i < 1; \\ 0, & \text{otherwise.}\end{cases} \quad \text{and} \quad f_R(r) = \begin{cases} 2r, & 0 < r < 1; \\ 0, & \text{otherwise.}\end{cases}$$

a) For $0 < w < 1$, we use independence and apply the FPF to get

$$\begin{aligned}F_W(w) &= P(W \le w) = P\big(I^2 R \le w\big) = \iint_{i^2 r \le w} f_{I,R}(i,r)\,di\,dr = \iint_{i^2 r \le w} f_I(i) f_R(r)\,di\,dr \\ &= 1 - \int_{\sqrt{w}}^1 \left(\int_{w/i^2}^1 2r\,dr\right) 6i(1-i)\,di = 1 - \int_{\sqrt{w}}^1 \left(1 - \frac{w^2}{i^4}\right) 6i(1-i)\,di \\ &= 1 - 6\int_{\sqrt{w}}^1 \left(i - i^2 - \frac{w^2}{i^3} + \frac{w^2}{i^2}\right) dw = 1 - 6\left(\frac{1}{6} - \frac{1}{2}w^2 - w + \frac{4}{3}w^{3/2}\right) \\ &= 3w^2 + 6w - 8w^{3/2}.\end{aligned}$$

Differentiation now yields

$$f_W(w) = \begin{cases} 6w + 6 - 12w^{1/2}, & \text{if } 0 < w < 1; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} 6\left(1 - \sqrt{w}\right)^2, & \text{if } 0 < w < 1; \\ 0, & \text{otherwise.} \end{cases}$$

b) Let $W = I^2 R$ and $X = R$ (the "dummy variable"). The Jacobian determinant of the transformation $w = i^2 r$ and $x = r$ is

$$J(i, r) = \begin{vmatrix} 2ir & i^2 \\ 0 & 1 \end{vmatrix} = 2ir.$$

Solving the equations $w = i^2 r$ and $x = r$ for i and r, we obtain the inverse transformation $i = \sqrt{w/x}$ and $r = x$. Therefore, by the bivariate transformation theorem,

$$f_{W,X}(w, x) = \frac{1}{|J(i, r)|} f_{I,R}(i, r) = \frac{1}{2ir} 6i(1 - i) \cdot 2r = 6(1 - i) = 6\left(1 - \sqrt{w/x}\right)$$

if $0 < w < x < 1$, and $f_{W,X}(w, x) = 0$ otherwise. Applying Proposition 9.7 on page 510 now yields

$$f_W(w) = \int_{-\infty}^{\infty} f_{W,X}(w, x)\, dx = \int_w^1 6\left(1 - \sqrt{w/x}\right)\, dx = 6\left(1 - 2\sqrt{w} + w\right) = 6\left(1 - \sqrt{w}\right)^2$$

if $0 < w < 1$, and $f_W(w) = 0$ otherwise.

9.194 We have $S = cWD^3$. To obtain a PDF of S, we apply the bivariate transformation theorem with $S = cWD^3$ and $T = D$ (the "dummy variable"). The Jacobian determinant of the transformation $s = cwd^3$ and $t = d$ is

$$J(w, d) = \begin{vmatrix} cd^3 & 3cdw^2 \\ 0 & 1 \end{vmatrix} = cd^3.$$

Solving the equations $s = cwd^3$ and $t = d$ for w and d, we obtain the inverse transformation $w = s\big/\left(ct^3\right)$ and $d = t$. Hence, by the bivariate transformation theorem and the assumed independence of W and D,

$$f_{S,T}(s, t) = \frac{1}{|J(w, d)|} f_{W,D}(w, d) = \frac{1}{cd^3} f_W(w) f_D(d) = \frac{1}{ct^3} f_W\left(\frac{s}{ct^3}\right) f_D(t)$$

if $s > 0$ and $t > 0$, and $f_{S,T}(s, t) = 0$ otherwise. Consequently, by Proposition 9.7 on page 510,

$$f_S(s) = \int_{-\infty}^{\infty} f_{S,T}(s, t)\, dt = \int_0^{\infty} \frac{1}{ct^3} f_W\left(\frac{s}{ct^3}\right) f_D(t)\, dt$$

if $s > 0$, and $f_S(s) = 0$ otherwise.

9.195 Let $X = \sum_{k=1}^{m} X_k$ and $Y = \sum_{k=m+1}^{n} X_k$. From Proposition 6.13 on page 297, X and Y are independent random variables. Furthermore, from Proposition 9.14 on page 540, $X \sim \mathcal{N}\left(m\mu, m\sigma^2\right)$ and $Y \sim \mathcal{N}\left((n - m)\mu, (n - m)\sigma^2\right)$. Let $U = X + Y$ and $V = X$. The Jacobian determinant of the transformation $u = x + y$ and $v = x$ is

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

Solving the equations $u = x + y$ and $v = x$ for x and y, we obtain the inverse transformation $x = v$ and $y = u - v$. Hence, by the bivariate transformation, for all $u, v \in \mathcal{R}$,

$$\begin{aligned}
f_{U,V}(u, v) &= \frac{1}{|J(x, y)|} f_{X,Y}(x, y) = f_X(x) f_Y(y) = f_X(v) f_Y(u - v) \\
&= \left(\frac{1}{\sqrt{2\pi m}\,\sigma} e^{-(v-m\mu)^2/2m\sigma^2} \right) \cdot \left(\frac{1}{\sqrt{2\pi(n-m)}\,\sigma} e^{-\left((u-v)-(n-m)\mu\right)^2/2(n-m)\sigma^2} \right) \\
&= \frac{1}{2\pi \sqrt{n}\,\sigma \sqrt{m}\,\sigma \sqrt{1 - m/n}} e^{-\frac{1}{2} Q(u,v)},
\end{aligned}$$

where

$$Q(u, v) = \frac{1}{1 - m/n} \left\{ \left(\frac{u - n\mu}{\sqrt{n}\,\sigma} \right)^2 - 2\sqrt{\frac{m}{n}} \left(\frac{u - n\mu}{\sqrt{n}\,\sigma} \right) \left(\frac{v - m\mu}{\sqrt{m}\,\sigma} \right) + \left(\frac{v - m\mu}{\sqrt{m}\,\sigma} \right)^2 \right\}.$$

Theory Exercises

9.196 From the FPF, the general multiplication rule, and Equation (9.33) on page 517,

$$\begin{aligned}
P(X \in A,\ Y \in B) = P\left((X, Y) \in A \times B\right) &= \iint\limits_{A \times B} f_{X,Y}(x, y)\, dx\, dy = \int_A \left(\int_B f_{X,Y}(x, y)\, dy \right) dx \\
&= \int_A \left(\int_B f_X(x) f_{Y\,|\,X}(y\,|\,x)\, dy \right) dx = \int_A \left(\int_B f_{Y\,|\,X}(y\,|\,x)\, dy \right) f_X(x)\, dx \\
&= \int_A P(Y \in B \,|\, X = x) f_X(x)\, dx.
\end{aligned}$$

9.197

a) Applying the definition of a conditional PDF, the general multiplication rule, and Exercise 9.79(a), we get

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x) f_{Y\,|\,X}(y\,|\,x)}{\int_{-\infty}^{\infty} f_X(s) f_{Y\,|\,X}(y\,|\,s)\, ds}.$$

This result shows that we can obtain a conditional PDF of X given $Y = y$ when we know a marginal PDF of X and conditional PDFs of Y given $X = x$ for all $x \in \mathcal{R}$.

b) For $y > 0$,

$$\begin{aligned}
f_{X\,|\,Y}(x\,|\,y) &= \frac{f_X(x) f_{Y\,|\,X}(y\,|\,x)}{\int_{-\infty}^{\infty} f_X(s) f_{Y\,|\,X}(y\,|\,s)\, ds} = \frac{\lambda e^{-\lambda x} \lambda e^{-\lambda(y-x)}}{\int_0^y \lambda e^{-\lambda s} \lambda e^{-\lambda(y-s)}\, ds} \\
&= \frac{\lambda^2 e^{-\lambda y}}{\int_0^y \lambda^2 e^{-\lambda y}\, ds} = \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 e^{-\lambda y} \int_0^y 1\, ds} = \frac{1}{y},
\end{aligned}$$

if $0 < x < y$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise. Thus, $X_{|Y=y} \sim \mathcal{U}(0, y)$ for $y > 0$; that is, the posterior distribution of X under the condition that $Y = y$ is uniform on the interval $(0, y)$.

9.198

a) From Proposition 9.2 on page 490 and Proposition 8.1 on page 411,

$$\begin{aligned} F_U(u) &= \lim_{v\to\infty} F_{U,V}(u,v) = \lim_{v\to\infty} G(u,v) = \lim_{v\to\infty} \max\{F_X(u) + F_Y(v) - 1, 0\} \\ &= \max\{F_X(u) + 1 - 1, 0\} = \max\{F_X(u), 0\} = F_X(u). \end{aligned}$$

Thus, $F_U = F_X$. Similarly, $F_V = F_Y$.

b) Let S and T be random variables having H as their joint CDF. Again, from Proposition 9.2 and Proposition 8.1,

$$\begin{aligned} F_S(s) &= \lim_{t\to\infty} F_{S,T}(s,t) = \lim_{t\to\infty} H(s,t) = \lim_{t\to\infty} \min\{F_X(s), F_Y(t)\} \\ &= \min\{F_X(s), 1\} = F_X(s). \end{aligned}$$

Thus, $F_S = F_X$. Similarly, $F_T = F_Y$.

c) From Bonferroni's inequality, Exercise 2.74 on page 76,

$$F_{X,Y}(x,y) = P(X \le x,\ Y \le y) \ge P(X \le x) + P(Y \le y) - 1 = F_X(x) + F_Y(y) - 1.$$

Because, in addition, $F_{X,Y}(x,y) \ge 0$, we conclude that

$$F_{X,Y}(x,y) \ge \max\{F_X(x) + F_Y(y) - 1, 0\} = G(x,y).$$

Now, because $\{X \le x,\ Y \le y\} \subset \{X \le x\}$ and $\{X \le x,\ Y \le y\} \subset \{Y \le y\}$, the domination principle implies that

$$F_{X,Y}(x,y) = P(X \le x,\ Y \le y) \le \min\{P(X \le x), P(Y \le y)\} = \min\{F_X(x), F_Y(y)\} = H(x,y).$$

We have now shown that

$$G(x,y) \le F_{X,Y}(x,y) \le H(x,y),$$

or, in other words,

$$\max\{F_X(x) + F_Y(y) - 1, 0\} \le F_{X,Y}(x,y) \le \min\{F_X(x), F_Y(y)\}$$

for all $x, y \in \mathcal{R}$.

Advanced Exercises

9.199

a) We have

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \le x,\ Y \le y) = P(X \le x,\ 1 - X \le y) \\ &= P(1 - y \le X \le x) = \begin{cases} 0, & \text{if } x < 0,\ y < 0, \text{ or } x + y < 1; \\ x - (1-y), & \text{if } 0 < 1 - y \le x < 1; \\ 1 - (1-y), & \text{if } 0 \le y < 1 \text{ and } x \ge 1; \\ x - 0, & \text{if } 0 \le x < 1 \text{ and } y \ge 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases} \\ &= \begin{cases} 0, & \text{if } x < 0,\ y < 0, \text{ or } x + y < 1; \\ x + y - 1, & \text{if } 0 < 1 - y \le x < 1; \\ y, & \text{if } 0 \le y < 1 \text{ and } x \ge 1; \\ x, & \text{if } 0 \le x < 1 \text{ and } y \ge 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases} \end{aligned}$$

b) Referring to the result of part (a), we deduce from Proposition 9.2 on page 490 that

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

and

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y) = \begin{cases} 0, & \text{if } y < 0; \\ y, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

c) Differentiating the results in part (b) yields

$$f_X(x) = F_X'(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

and

$$f_Y(y) = F_Y'(y) = \begin{cases} 1, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we see that both X and Y have the uniform distribution on the interval (0, 1).

d) No, X and Y don't have a joint PDF. Suppose to the contrary that they do. Note that, as $Y = 1 - X$, we have $P(X + Y = 1) = 1$. Hence, from the FPF,

$$1 = P(X+Y=1) = \iint\limits_{x+y=1} f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_0^1 \left(\int_{1-x}^{1-x} f_{X,Y}(x,y)\,dy \right) dx = \int_0^1 0\,dx = 0,$$

which is impossible. Hence, X and Y don't have a joint PDF.

9.200

a) From Proposition 9.2 on page 490,

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y) = \begin{cases} 0, & \text{if } x < 0; \\ 1 - e^{-x}, & \text{if } x \ge 0. \end{cases}$$

and

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y) = \begin{cases} 0, & \text{if } y < 0; \\ 1/2, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

b) From part (a), we see that $X \sim \mathcal{E}(1)$ and $Y \sim \mathcal{B}(1, 1/2)$ (or, equivalently, Y has the Bernoulli distribution with parameter 1/2).

c) No, X and Y don't have a joint PDF. One way to see this fact is to note that Y is a discrete random variable. If two random variables have a joint PDF, then both must have a (marginal) PDF; in particular, both must be continuous random variables.

d) From part (a), we see that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $x, y \in \mathcal{R}$. Hence, from Exercise 9.15(b), we conclude that X and Y are independent random variables.

e) Applying first the result of part (d) and then that of parts (a) and (b), we get

$$\begin{aligned} P(3 < X < 5,\ Y = 0) &= P(3 < X < 5)P(Y=0) = \big(F_X(5) - F_X(3)\big)P(Y=0) \\ &= \Big(\big(1 - e^{-5}\big) - \big(1 - e^{-3}\big)\Big)P(Y=0) = \frac{1}{2}\left(e^{-3} - e^{-5}\right) \\ &= 0.0215. \end{aligned}$$

9.201

a) For $0 \le y < 1$,

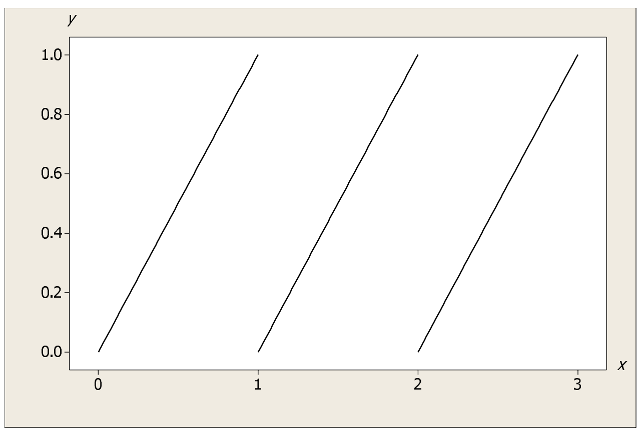
$$\begin{aligned} F_Y(y) &= P(Y \le y) = P\big(X - \lfloor X \rfloor \le y\big) \\ &= P(X \le y) + P(1 \le X \le y+1) + P(2 \le X \le y+2) \\ &= \frac{y}{3} + \frac{y}{3} + \frac{y}{3} = y. \end{aligned}$$

Therefore $Y \sim \mathcal{U}(0, 1)$.

b) The range of (X, Y) consists of following three line segments:

$$\{\,(x, y) : y = x,\ 0 < x < 1\,\},\ \ \{\,(x, y) : y = x - 1,\ 1 \le x < 2\,\},\ \ \{\,(x, y) : y = x - 2,\ 2 \le x < 3\,\}.$$

A graph of the range of (X, Y) is as follows:



c) Clearly, we have $F_{X,Y}(x, y) = 0$ if $x < 0$ or $y < 0$, and $F_{X,Y}(x, y) = 1$ if $x \ge 3$ and $y \ge 1$. From part (a), we see that, if $x \ge 3$ and $0 \le y < 1$,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y) = P(Y \le y) = F_Y(y) = y.$$

Also, if $0 \le x < 3$ and $y \ge 1$,

$$F_{X,Y}(x, y) = P(X \le x,\ Y \le y) = P(X \le x) = F_X(x) = x/3.$$

Furthermore, by referring to the graph obtained in part (b), we find that, for $0 \le x < 3$ and $0 \le y < 1$,

$$F_{X,Y}(x, y) = \begin{cases} x/3, & \text{if } 0 \le x < y; \\ y/3, & \text{if } y \le x < 1; \\ (x - 1 + y)/3, & \text{if } 1 \le x < y + 1; \\ 2y/3, & \text{if } y + 1 \le x < 2; \\ (x - 2 + 2y)/3, & \text{if } 2 \le x < y + 2; \\ y, & \text{if } y + 2 \le x < 3. \end{cases}$$

In summary, then,

$$F_{X,Y}(x,y)=\begin{cases}0, & \text{if } x<0 \text{ or } y<0;\\ x/3, & \text{if } 0\le x<y<1 \text{ or if } 0\le x<3 \text{ and } y\ge 1;\\ y/3, & \text{if } 0\le y\le x<1;\\ (x-1+y)/3, & \text{if } 1\le x<y+1<2;\\ 2y/3, & \text{if } 1\le y+1\le x<2;\\ (x-2+2y)/3, & \text{if } 2\le x<y+2<3;\\ y, & \text{if } 2\le y+2\le x<3 \text{ or if } x\ge 3 \text{ and } 0\le y<1;\\ 1, & \text{if } x\ge 3 \text{ and } y\ge 1.\end{cases}$$

9.202

a) We have $Z = X - Y = X - (X - \lfloor X\rfloor) = \lfloor X\rfloor$. Thus,

$$p_Z(z)=P(Z=z)=\begin{cases}P(z\le X<z+1), & \text{if } z=0,1,2;\\ 0, & \text{otherwise.}\end{cases}=\begin{cases}1/3, & \text{if } z=0,1,2;\\ 0, & \text{otherwise.}\end{cases}$$

Therefore, Z has the discrete uniform distribution on the set $\{0, 1, 2\}$.

b) If two random variables have a joint PDF, their difference is a continuous random variable (with a PDF). By part (a), $X - Y$ is a discrete random variable and, hence, it can't be a continuous random variable. Thus, X and Y can't have a joint PDF.

c) Let R denote the range of X and Y. We know that R is the union of the three line segments shown in the solution to Exercise 9.201(b), which we call L_1, L_2, and L_3, respectively. If X and Y had a joint PDF, we would have

$$\begin{aligned}1 &= P\big((X,Y)\in R\big)=\iint\limits_R f_{X,Y}(x,y)\,dx\,dy\\ &= \iint\limits_{L_1} f_{X,Y}(x,y)\,dx\,dy+\iint\limits_{L_2} f_{X,Y}(x,y)\,dx\,dy+\iint\limits_{L_3} f_{X,Y}(x,y)\,dx\,dy\\ &= \int_0^1\left(\int_x^x f_{X,Y}(x,y)\,dy\right)dx+\int_1^2\left(\int_{x-1}^{x-1} f_{X,Y}(x,y)\,dy\right)dx+\int_2^3\left(\int_{x-2}^{x-2} f_{X,Y}(x,y)\,dy\right)dx\\ &= 0+0+0=0,\end{aligned}$$

which is impossible. Hence, X and Y can't have a joint PDF.

9.203

a) Let M denote the sample median. For a sample of size $2n+1$, M is the $(n+1)$st order statistic. Hence, from the solution to Exercise 9.33(b),

$$\begin{aligned}f_M(m) &= \frac{(2n+1)!}{\big((n+1)-1\big)!\,\big((2n+1)-(n+1)\big)!}\,f(m)\big(F(m)\big)^{(n+1)-1}\big(1-F(m)\big)^{(2n+1)-(n+1)}\\ &= \frac{(2n+1)!}{(n!)^2}\,f(m)\big(F(m)\big)^n\big(1-F(m)\big)^n.\end{aligned}$$

b) In this case,

$$F_X(x)=\begin{cases}0, & \text{if } x<0;\\ x, & \text{if } 0\le x<1;\\ 1, & \text{if } x\ge 1.\end{cases} \qquad \text{and} \qquad f_X(x)=\begin{cases}1, & \text{if } 0<x<1;\\ 0, & \text{otherwise.}\end{cases}$$

Applying the result of part (a) with $n = 2$ now yields

$$f_M(m) = \frac{5!}{(2!)^2} 1 \cdot m^2(1-m)^2 = 30m^2(1-m)^2$$

if $0 < m < 1$, and $f_M(m) = 0$ otherwise. Thus, M has the beta distribution with parameters 3 and 3. We want to determine the probability that the sample median will be within $1/4$ of the distribution median of $1/2$. Applying the FPF yields

$$P(|M - 1/2| \le 1/4) = \int_{|m-1/2|\le 1/4} f_M(m)\, dm = \int_{1/4}^{3/4} 30m^2(1-m)^2 = 0.793.$$

9.204 For $i = 1, 2$, let L_i denote the event that employee i incurs a loss and let Y_i denote the loss amount in thousands of dollars for employee i. By assumption, L_1 and L_2 are independent events each having probability 0.4, Y_1 and Y_2 are independent random variables, and the conditional distribution of Y_i given event L_i occurs is uniform on the interval $(1, 5)$. First let's determine $P(Y_i > 2)$ for $i = 1, 2$. By the law of total probability,

$$P(Y_i > 2) = P(L_i)P(Y_i > 2 \mid L_i) + P(L_i^c)P(Y_1 > 2 \mid L_i^c) = 0.4 \cdot \frac{3}{4} + 0.6 \cdot 0 = 0.3.$$

Next let's determine $P(\{Y_1 > 2\} \cup \{Y_2 > 2\})$. We have, by the general addition rule and the independence of Y_1 and Y_2,

$$\begin{aligned} P(\{Y_1 > 2\} \cup \{Y_2 > 2\}) &= P(Y_1 > 2) + P(Y_2 > 2) - P(Y_1 > 2,\ Y_2 > 2) \\ &= P(Y_1 > 2) + P(Y_2 > 2) - P(Y_1 > 2)P(Y_2 > 2) \\ &= 0.3 + 0.3 - (0.3)^2 = 0.51. \end{aligned}$$

Now let's determine $P(Y_1 + Y_2 > 8)$. Again, by the law of total probability,

$$\begin{aligned} P(Y_1 + Y_2 > 8) &= P(L_1 \cap L_2)P(Y_1 + Y_2 > 8 \mid L_1 \cap L_2) \\ &\quad + P\big((L_1 \cap L_2)^c\big)P\big(Y_1 + Y_2 > 8 \mid (L_1 \cap L_2)^c\big). \end{aligned}$$

As the maximum loss of an employee is 5, $P\big(Y_1 + Y_2 > 8 \mid (L_1 \cap L_2)^c\big) = 0$. Also, by the independence of L_1 and L_2, we have $P(L_1 \cap L_2) = P(L_1)P(L_2) = 0.4 \cdot 0.4 = 0.16$. Moreover, by the independence and assumed conditional distributions of Y_1 and Y_2, $P(Y_1 + Y_2 > 8 \mid L_1 \cap L_2) = P(X_1 + X_2 > 8)$, where X_1 and X_2 are independent $\mathcal{U}(1, 5)$ random variables. Applying the FPF, we get

$$\begin{aligned} P(X_1 + X_2 > 8) &= \iint\limits_{x_1 + x_2 > 8} f_{X_1,X_2}(x_1, x_2)\, dx_1\, dx_2 = \int_3^5 \left(\int_{8-x_1}^5 \frac{1}{16}\, dx_2 \right) dx_1 \\ &= \frac{1}{16} \int_3^5 (x_1 - 3)\ dx_1 = 0.125. \end{aligned}$$

Thus,

$$P(Y_1 + Y_2 > 8) = 0.16 \cdot 0.125 + 0.84 \cdot 0 = 0.02.$$

a) Here we assume that "one" means a specified one, say, employee 1. The problem is then to determine the probability $P(Y_1 + Y_2 > 8 \mid Y_1 > 2)$. Observing that $\{Y_1 + Y_2 > 8\} \subset \{Y_1 > 2\}$ and applying the conditional probability rule, we get

$$P(Y_1 + Y_2 > 8 \mid Y_1 > 2) = \frac{P(Y_1 + Y_2 > 8,\ Y_1 > 2)}{P(Y_1 > 2)} = \frac{P(Y_1 + Y_2 > 8)}{P(Y_1 > 2)} = \frac{0.02}{0.3} = \frac{1}{15}.$$

b) Here we assume that "one" means at least one. The problem is then to determine the probability $P(Y_1 + Y_2 > 8 \mid \{Y_1 > 2\} \cup \{Y_2 > 2\})$. Noting that $\{Y_1 + Y_2 > 8\} \subset \{\{Y_1 > 2\} \cup \{Y_2 > 2\}\}$, we get

$$P(Y_1 + Y_2 > 8 \mid \{Y_1 > 2\} \cup \{Y_2 > 2\}) = \frac{P(Y_1 + Y_2 > 8)}{P(\{Y_1 > 2\} \cup \{Y_2 > 2\})} = \frac{0.02}{0.51} = \frac{2}{51}.$$

c) Here we assume that "one" means exactly one. The problem is then to determine the probability $P(Y_1 + Y_2 > 8 \mid E)$, where $E = \{Y_1 > 2,\ Y_2 \le 2\} \cup \{Y_1 \le 2,\ Y_2 > 2\}$. However, given that E occurs, the maximum total loss amount is \$7000. Hence, $P(Y_1 + Y_2 > 8 \mid E) = 0$; that is,

$$P(Y_1 + Y_2 > 8 \mid \{Y_1 > 2,\ Y_2 \le 2\} \cup \{Y_1 \le 2,\ Y_2 > 2\}) = 0.$$

9.205

a) Let F_X and F_Y denote the two marginal CDFs of F_α. Then, from Proposition 9.2 on page 490 and Proposition 8.1 on page 411,

$$\begin{aligned} F_X(x) &= \lim_{y\to\infty} F_\alpha(x, y) = \lim_{y\to\infty} G(x)H(y)\big[1 + \alpha\big(1 - G(x)\big)\big(1 - H(y)\big)\big] \\ &= G(x)\cdot 1 \cdot \big[1 + \alpha\big(1 - G(x)\big)(1 - 1)\big] = G(x) \end{aligned}$$

and

$$\begin{aligned} F_Y(y) &= \lim_{x\to\infty} F_\alpha(x, y) = \lim_{x\to\infty} G(x)H(y)\big[1 + \alpha\big(1 - G(x)\big)\big(1 - H(y)\big)\big] \\ &= 1 \cdot H(y)\big[1 + \alpha(1 - 1)\big(1 - H(y)\big)\big] = H(y). \end{aligned}$$

Thus, $F_X = G$ and $F_Y = H$.

b) In this case, we take

$$G(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \qquad \text{and} \qquad H(y) = \begin{cases} 0, & \text{if } y < 0; \\ y, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

Hence,

$$F_\alpha(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ xy\big(1 + \alpha(1 - x)(1 - y)\big), & \text{if } 0 \le x < 1 \text{ and } 0 \le y < 1; \\ y, & \text{if } x \ge 1 \text{ and } 0 \le y < 1; \\ x, & \text{if } 0 \le x < 1 \text{ and } y \ge 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

From part (a), each F_α has G and H as its marginal CDFs, which, in this case, are both CDFs of a uniform distribution on the interval (0, 1). Now, it is easy to see that, if $F_{\alpha_1} = F_{\alpha_2}$, then $\alpha_1 = \alpha_2$. Hence, we have an uncountably infinite number of different bivariate CDFs (one for each $\alpha \in [-1, 1]$) whose marginals are both CDFs of a uniform distribution on the interval (0, 1).

c) Taking the mixed partial of F_α in part (b), we get

$$f_\alpha(x, y) = \begin{cases} 1 + \alpha(1 - 2x)(1 - 2y), & \text{if } 0 \le x < 1 \text{ and } 0 \le y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

9.206

a) Taking the mixed partial of F_α, as defined in the statement of Exercise 9.205, yields

$$f_\alpha(x, y) = (1+\alpha)g(x)h(y) - 2\alpha g(x)h(y)\big(G(x) + H(y)\big) + 4\alpha g(x)h(y)G(x)H(y).$$

b) Let f_X and f_Y denote the two marginal CDFs of F_α. Applying Proposition 9.7 on page 510 and making the substitution $u = H(x)$ in the appropriate intergrals, we get

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_\alpha(x, y)\,dy \\
&= \int_{-\infty}^{\infty} \Big((1+\alpha)g(x)h(y) - 2\alpha g(x)h(y)\big(G(x) + H(y)\big) + 4\alpha g(x)h(y)G(x)H(y)\Big)dy \\
&= (1+\alpha)g(x)\int_{-\infty}^{\infty} h(y)\,dy - 2\alpha g(x)G(x)\int_{-\infty}^{\infty} h(y)\,dy - 2\alpha g(x)\int_{-\infty}^{\infty} h(y)H(y)\,dy \\
&\qquad + 4\alpha g(x)G(x)\int_{-\infty}^{\infty} h(y)H(y)\,dy \\
&= (1+\alpha)g(x) - 2\alpha g(x)G(x) - 2\alpha g(x)\int_0^1 u\,du + 4\alpha g(x)G(x)\int_0^1 u\,du \\
&= (1+\alpha)g(x) - 2\alpha g(x)G(x) - \alpha g(x) + 2\alpha g(x)G(x) = g(x).
\end{aligned}$$

Thus, $f_X = g$. Similarly, we find that $f_Y = h$.

c) From Exercise 9.205(a), we know that $F_X = G$ and $F_Y = H$. Therefore, we see that $f_X = g$ and $f_Y = h$.

d) From parts (a) and (b) and Proposition 9.9 on page 523, the corresponding random variables are independent if and only if

$$(1+\alpha)g(x)h(y) - 2\alpha g(x)h(y)\big(G(x) + H(y)\big) + 4\alpha g(x)h(y)G(x)H(y) = g(x)h(y)$$

or

$$(1+\alpha) - 2\alpha\big(G(x) + H(y)\big) + 4\alpha G(x)H(y) = 1.$$

Letting both x and y go to infinity in the previous display, we get, in view of Proposition 8.1 on page 411, that $(1+\alpha) - 2\alpha\cdot(1+1) + 4\alpha\cdot 1\cdot 1 = 1$, or $1+\alpha = 1$, which means that $\alpha = 0$.

e) In this case,

$$G(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \qquad \text{and} \qquad H(y) = \begin{cases} 0, & \text{if } y < 0; \\ y, & \text{if } 0 \le y < 1; \\ 1, & \text{if } y \ge 1. \end{cases}$$

and

$$g(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad h(y) = \begin{cases} 1, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, from part (a),

$$\begin{aligned}
f_\alpha(x, y) &= (1+\alpha)\cdot 1\cdot 1 - 2\alpha\cdot 1\cdot 1\cdot(x+y) + 4\alpha\cdot 1\cdot 1\cdot xy \\
&= 1 + \alpha - 2\alpha x - 2\alpha y + 4\alpha xy = 1 + \alpha(1-2x)(1-2y)
\end{aligned}$$

if $0 < x < 1$ and $0 < y < 1$, and $f_\alpha(x, y) = 0$ otherwise. This result is the same as the one found in Exercise 9.205(c).

9.207

a) Referring to Exercises 9.83 and 9.85 on page 522, we get

$$f_{X\,|\,Y}(x\,|\,y) = \frac{h_{X,Y}(x, y)}{p_Y(y)} = \frac{f_X(x)p_{Y\,|\,X}(y\,|\,x)}{\int_{-\infty}^{\infty} h_{X,Y}(s, y)\,ds} = \frac{f_X(x)p_{Y\,|\,X}(y\,|\,x)}{\int_{-\infty}^{\infty} f_X(s)p_{Y\,|\,X}(y\,|\,s)\,ds}.$$

This result shows that we can obtain a conditional PDF of X given $Y = y$ when we know a marginal PDF of X and conditional PMFs of Y given $X = x$ for all $x \in \mathcal{R}$.

b) Let Y denote the number of employees who take at least one sick day during the year. By assumption, $Y_{|\Pi=p} \sim \mathcal{B}(n, p)$ and Π has the beta distribution with parameters α and β. We want to determine the probability distribution of $\Pi_{|Y=k}$. Applying Bayes's rule, we get, for $0 < p < 1$,

$$\begin{aligned} f_{\Pi\,|\,Y}(p\,|\,k) &= \frac{f_\Pi(p)p_{Y\,|\,\Pi}(k\,|\,p)}{\int_{-\infty}^{\infty} f_\Pi(s)p_{Y\,|\,\Pi}(k\,|\,s)\,ds} \\ &= \frac{\dfrac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}\cdot\dbinom{n}{k}p^k(1-p)^{n-k}}{\displaystyle\int_0^1 \frac{1}{B(\alpha,\beta)}s^{\alpha-1}(1-s)^{\beta-1}\cdot\binom{n}{k}s^k(1-s)^{n-k}\,ds} \\ &= \frac{p^{\alpha+k-1}(1-p)^{\beta+n-k-1}}{\int_0^1 s^{\alpha+k-1}(1-s)^{\beta+n-k-1}\,ds}. \end{aligned}$$

Using the fact that a beta PDF integrates to 1, we conclude that

$$f_{\Pi\,|\,Y}(p\,|\,k) = \frac{1}{B(\alpha+k, \beta+n-k)}p^{\alpha+k-1}(1-p)^{\beta+n-k-1},$$

if $0 < p < 1$, and $f_{\Pi\,|\,Y}(p\,|\,k) = 0$ otherwise. Thus, given that the number of employees who take at least one sick day during the year is k, the posterior probability distribution of Π is the beta distribution with parameters $\alpha + k$ and $\beta + n - k$.

9.208

a) Intuitively, as $n \to \infty$, we would expect $U_n \to 0$ and $V_n \to 1$, so that $(U_n, V_n) \to (0, 1)$. Thus, we would expect the joint probability distribution of U_n and V_n to converge to that of the discrete random vector that equals $(0, 1)$ with probability 1.

b) The common CDF of $X_1, X_2, \ldots,$ is given by

$$F(t) = \begin{cases} 0, & \text{if } t < 0; \\ t, & \text{if } 0 \le t < 1; \\ 1, & \text{if } t \ge 1. \end{cases}$$

Because $U_n \le V_n$, we have, for $x \ge y$,

$$F_n(x, y) = P(U_n \le x,\ V_n \le y) = P(V_n \le y) = \big(F(y)\big)^n.$$

From the preceding result and Equation (9.19), we find that

$$F_n(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 0; \\ y^n - (y-x)^n, & \text{if } 0 \le x < y < 1; \\ y^n, & \text{if } 0 \le y \le x < 1 \text{ or if } x \ge 1 \text{ and } 0 \le y < 1; \\ 1 - (1-x)^n, & \text{if } 0 \le x < 1 \text{ and } y \ge 1; \\ 1, & \text{if } x \ge 1 \text{ and } y \ge 1. \end{cases}$$

Consequently,

$$\lim_{n\to\infty} F_n(x, y) = \begin{cases} 0, & \text{if } x \le 0 \text{ or } y < 1; \\ 1, & \text{if } x > 0 \text{ and } y \ge 1. \end{cases}$$

c) We first recall that, if two events each have probability 1, then so does their intersection. Thus,

$$P(X = 0,\ Y = 1) = P(\{X = 0\} \cap \{Y = 1\}) = 1.$$

It follows that

$$F_{X,Y}(x, y) = \begin{cases} 0, & \text{if } x < 0 \text{ or } y < 1; \\ 1, & \text{if } x \ge 0 \text{ and } y \ge 1. \end{cases}$$

d) Yes, from parts (b) and (c), as $n \to \infty$, the joint CDF of U_n and V_n converges to that of the discrete random vector that equals (0, 1) with probability 1, except at the points in $\mathcal{R}^2$ at which the CDF of the latter is discontinuous. *Note:* As discussed in advanced probability courses, requiring convergence at every point of $\mathcal{R}^2$ is too strong a condition.

CHAPTER TEN

Instructor's Solutions Manual

Anna Amirdjanova
University of Michigan, Ann Arbor

Neil A. Weiss
Arizona State University

FOR

A Course In Probability

Neil A. Weiss
Arizona State University

PEARSON Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

Publisher: Greg Tobin
Editor-in-Chief: Deirdre Lynch
Associate Editor: Sara Oliver Gordus
Editorial Assistant: Christina Lepre
Production Coordinator: Kayla Smith-Tarbox
Senior Author Support/Technology Specialist: Joe Vetere
Compositor: Anna Amirdjanova and Neil A. Weiss
Accuracy Checker: Delray Schultz
Proofreader: Carol A. Weiss

Chapter 10

Expected Value of Continuous Random Variables

10.1 Expected Value of a Continuous Random Variable

Basic Exercises

10.1 From Exercise 8.41,

$$f_Z(z) = \begin{cases} 3z^2/r^3, & \text{if } 0 < z < r; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, from Definition 10.1 on page 565,

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} z f_Z(z)\,dz = \int_0^r z \cdot \frac{3z^2}{r^3}\,dz = \frac{3}{r^3}\int_0^r z^3\,dz = \frac{3}{4}r.$$

On average, we would expect the point chosen to be a distance of $3r/4$ from the center of the sphere.

10.2 From Exercise 8.43,

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathcal{R}.$$

In other words, X has the standard Cauchy distribution. From Example 10.5 on page 568, we know that a Cauchy distribution does not have finite expectation.

10.3 From Exercise 8.44,

$$f_Y(y) = \begin{cases} 2(h-y)/h^2, & \text{if } 0 < y < h; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, from Definition 10.1 on page 565,

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_0^h y\frac{2(h-y)}{h^2}\,dy = \frac{2}{h^2}\int_0^h (hy - y^2)\,dy = \frac{1}{3}h.$$

On average, the point chosen will be a distance of $h/3$ from the base of the triangle.

10.4
a) By symmetry, we would guess that $\mathcal{E}(X) = 0$.

b) From Example 9.8,

$$f_X(x) = \frac{2}{\pi}\sqrt{1 - x^2}$$

if $-1 < x < 1$, and $f_X(x) = 0$ otherwise. Hence, from Definition 10.1 on page 565,

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{2}{\pi}\int_{-1}^{1} \frac{x}{\sqrt{1 - x^2}}\,dx = 0,$$

where the last equation follows from the fact that $x\big/\sqrt{1 - x^2}$ is an odd function.
c) By symmetry, Y has the same probability distribution as X and, hence, the same expected value. Thus, $\mathcal{E}(Y) = 0$.

10.5 Obviously, X has finite expectation. Applying Definition 10.1 on page 565 yields

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{0}^{30} x \cdot \frac{1}{30}\,dx = \frac{1}{30}\int_{0}^{30} x\,dx = \frac{1}{30} \cdot \frac{(30)^2}{2} = 15.$$

The expected time that John waits for the train is 15 minutes, which we could have guessed in advance. On average, John waits 15 minutes for the train to arrive.

10.6
a) We have

$$f_X(x) = \begin{cases} 1/(b-a), & \text{if } a < x < b; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, from Definition 10.1 on page 565,

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{1}{b-a}\int_{a}^{b} x\,dx = \frac{1}{b-a}\frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

b) In Exercise 10.5, $X \sim \mathcal{U}(0, 30)$. Thus, from part (a), $\mathcal{E}(X) = (0 + 30)/2 = 15$.

10.7 From Example 10.2 on page 566, the expected value of an exponential random variable is the reciprocal of its parameter. Hence, the expected duration of a pause during a monologue is $1/1.4$, or $5/7$ second.

10.8 Let X denote the time, in minutes, that it takes a finisher to complete the race. We are given that $X \sim \mathcal{N}(61, 81)$. From Example 10.3 on page 567, the expected value of a normal random variable is its first parameter. Thus, $\mathcal{E}(X) = 61$ minutes. On average, it takes 61 minutes for a finisher to complete the race.

10.9 We have

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1}$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise. Referring to Equation (8.43) on page 451, we get

$$\begin{aligned} \mathcal{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{1}{B(\alpha, \beta)}\int_{0}^{1} x \cdot x^{\alpha-1}(1 - x)^{\beta-1}\,dx \\ &= \frac{1}{B(\alpha, \beta)}\int_{0}^{1} x^{(\alpha+1)-1}(1 - x)^{\beta-1}\,dx = \frac{1}{B(\alpha, \beta)} \cdot B(\alpha + 1, \beta) \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta}. \end{aligned}$$

10.10 Let X denote the proportion of these manufactured items that require service during the first 5 years of use. We know that X has the beta distribution with parameters $\alpha = 2$ and $\beta = 3$. Applying Exercise 10.9, we find that $\mathcal{E}(X) = 2/(2+3) = 2/5$. On average, 40% of the manufactured items require service during the first 5 years of use.

10.11 Some possible beta distributions follow. To obtain the mean, we apply Exercise 10.9, and to determine $P(X \geq \mu)$, we use Equation (8.55) on page 457.

a) $\alpha = \beta = 1$: $\mu_X = 1/(1+1) = 0.5$;

$$\begin{aligned} P(X \geq \mu_X) &= P(X \geq 0.5) = 1 - P(X < 0.5) \\ &= 1 - \sum_{j=1}^{1+1-1} \binom{1+1-1}{j} (0.5)^j (1-0.5)^{1+1-1-j} = 1 - \binom{1}{1}(0.5)^1(0.5)^0 \\ &= 1 - 0.5 = 0.5. \end{aligned}$$

b) $\alpha = 3, \beta = 1$: $\mu_X = 3/(3+1) = 0.75$;

$$\begin{aligned} P(X \geq \mu_X) &= P(X \geq 0.75) = 1 - P(X < 0.75) \\ &= 1 - \sum_{j=3}^{3+1-1} \binom{3+1-1}{j} (0.75)^j (1-0.75)^{3+1-1-j} = 1 - \binom{3}{3}(0.75)^3(0.25)^0 \\ &= 1 - (0.75)^3 > 1/2. \end{aligned}$$

c) $\alpha = 1, \beta = 3$: $\mu_X = 1/(1+3) = 0.25$;

$$\begin{aligned} P(X \geq \mu_X) &= P(X \geq 0.25) = 1 - P(X < 0.25) \\ &= 1 - \sum_{j=1}^{1+3-1} \binom{1+3-1}{j} (0.25)^j (1-0.25)^{1+3-1-j} \\ &= 1 - \binom{3}{1}(0.25)^1(0.75)^2 - \binom{3}{2}(0.25)^2(0.75)^1 - \binom{3}{3}(0.25)^3(0.75)^0 < 1/2. \end{aligned}$$

10.12 We have

$$f_X(x) = \frac{2}{b-a}\left(1 - \frac{|a+b-2x|}{b-a}\right)$$

if $a < x < b$, and $f_X(x) = 0$ otherwise. Applying Definition 10.1 on page 565 and making the substitutions $u = 1 - (a+b-2x)/(b-a)$ and $u = 1 - (2x-a-b)/(b-a)$, respectively, we get

$$\begin{aligned} \mathcal{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{2}{b-a}\int_a^b x\left(1 - \frac{|a+b-2x|}{b-a}\right)dx \\ &= \frac{2}{b-a}\int_a^{(a+b)/2} x\left(1 - \frac{a+b-2x}{b-a}\right)dx + \frac{2}{b-a}\int_{(a+b)/2}^b x\left(1 - \frac{2x-a-b}{b-a}\right)dx \\ &= \int_0^1 \left(a + \frac{b-a}{2}u\right)u\,du + \int_0^1 \left(b - \frac{b-a}{2}u\right)u\,du = (a+b)\int_0^1 u\,du = \frac{a+b}{2}. \end{aligned}$$

For a more elegant approach, see Exercise 10.20(b).

10.13 We know that $f_X(x) = c(1+x)^{-4}$ if $x > 0$, and $f_X(x) = 0$ otherwise. To determine c, we proceed as follows:

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = c\int_0^{\infty} \frac{dx}{(1+x)^4} = c\int_1^{\infty} \frac{du}{u^4} = \frac{1}{3}c.$$

Hence, $c = 3$. Consequently, making the substitution $u = 1 + x$ yields

$$\begin{aligned}\mathcal{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = 3\int_0^{\infty} \frac{x}{(1+x)^4}\,dx = 3\int_1^{\infty} \frac{u-1}{u^4}\,du \\ &= 3\int_1^{\infty}\left(\frac{1}{u^3} - \frac{1}{u^4}\right)du = 3\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{2}.\end{aligned}$$

10.14 Let X_1, X_2, X_3, and X_4 denote the bids. By assumption, the X_js are independent random variables with common CDF $F(x) = (1 + \sin \pi x)/2$ for $3/2 \le x < 5/2$. It follows that the common PDF of the X_js is $f(x) = (\pi \cos \pi x)/2$ if $3/2 < x < 5/2$, and $f(x) = 0$ otherwise. The accepted bid is $Y = \max\{X_1, X_2, X_3, X_4\}$. Thus, from Exercise 9.70,

$$f_Y(y) = 4f(y)\big(F(y)\big)^3 = 4\cdot\left(\frac{\pi}{2}\cos\pi y\right)\left(\frac{1}{2}(1+\sin\pi y)\right)^3 = \frac{\pi}{4}(\cos\pi y)(1+\sin\pi y)^3$$

if $3/2 < y < 5/2$, and $f_Y(y) = 0$ otherwise. Consequently,

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \frac{\pi}{4}\int_{3/2}^{5/2} y(\cos\pi y)(1+\sin\pi y)^3\,dy.$$

Therefore, the correct answer is (d).

10.15 We have

$$f_X(x) = \frac{\theta}{\pi\big(\theta^2 + (x-\eta)^2\big)}, \qquad -\infty < x < \infty.$$

Let Y be a random variable that has the standard Cauchy distribution. We first make the substitution $y = (x-\eta)/\theta$ and then apply the triangle inequality to get

$$\begin{aligned}\int_{-\infty}^{\infty} |x| f_X(x)\,dx &= \int_{-\infty}^{\infty} |\theta y + \eta| \frac{1}{\pi(1+y^2)}\,dy = \int_{-\infty}^{\infty} |\theta y + \eta| f_Y(y)\,dy \\ &\ge \int_{-\infty}^{\infty} (|\theta y| - |\eta|)\, f_Y(y)\,dy = \theta\int_{-\infty}^{\infty} |y| f_Y(y)\,dy - |\eta|\int_{-\infty}^{\infty} f_Y(y)\,dy \\ &= \theta\int_{-\infty}^{\infty} |y| f_Y(y)\,dy - |\eta| = \infty,\end{aligned}$$

where the last equality follows from Example 10.5 on page 568, which shows that a standard Cauchy random variable doesn't have finite expectation.

10.16 We first determine the two marginal PDFs:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = 2x \int_0^{x/2} y\,dy = x^3/4$$

if $0 < x < 2$, and $f_X(x) = 0$ otherwise. Also,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = 2y \int_{2y}^{2} x\,dx = 4y(1 - y^2)$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Therefore,

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^2 x\,\frac{x^3}{4}\,dx = \frac{2^5}{20} = \frac{8}{5}$$

and

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = 4\int_0^1 y^2(1 - y^2)\,dy = 4\left(\frac{1}{3} - \frac{1}{5}\right) = \frac{8}{15}.$$

10.17 Let us set $S = \{\,(t_1, t_2) : 0 < t_1 < 6, 0 < t_2 < 6, t_1 + t_2 < 10\,\}$. We know that $f_{T_1,T_2}(t_1, t_2) = c$ if $(t_1, t_2) \in S$, and $f_{T_1,T_2}(t_1, t_2) = 0$ otherwise. To determine c, we draw a graph of S and find that

$$|S| = \frac{1}{2} \cdot 10 \cdot 10 - 2 \cdot \frac{1}{2} \cdot 4 \cdot 4 = 34.$$

Therefore,

$$1 = \iint_S f_{T_1,T_2}(t_1, t_2)\,dt_1\,dt_2 = c \iint_S 1\,dt_1\,dt_2 = c|S| = 34c,$$

which means that $c = 1/34$. Now let T denote the time between a car accident and payment of the claim, so that $T = T_1 + T_2$. We note that the range of T is the interval $(0, 10)$. For t in the range of T, we have, from Proposition 9.12, $f_T(t) = \int_{-\infty}^{\infty} f_{T_1,T_2}(s, t - s)\,ds$. The integrand is nonzero if and only if $0 < s < 6$ and $0 < t - s < 6$, that is, if and only if $0 < s < 6$ and $t - 6 < s < t$, which is true if and only if $\max\{t - 6, 0\} < s < \min\{t, 6\}$. Now,

$$\min\{t, 6\} - \max\{t - 6, 0\} = \begin{cases} t - 0, & \text{if } t < 6; \\ 6 - (t - 6), & \text{if } t \geq 6. \end{cases} = \begin{cases} t, & \text{if } t < 6; \\ 12 - t, & \text{if } t \geq 6. \end{cases}$$

Hence,

$$f_T(t) = \int_{\max\{t-6,0\}}^{\min\{t,6\}} \frac{1}{34}\,ds = \frac{1}{34}\big(\min\{t, 6\} - \max\{t - 6, 0\}\big) = \begin{cases} t/34, & \text{if } 0 < t < 6; \\ (12 - t)/34, & \text{if } 6 < t < 10; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\mathcal{E}(T) = \int_{-\infty}^{\infty} t f_T(t)\,dt = \frac{1}{34}\int_0^6 t^2\,dt + \frac{1}{34}\int_6^{10} t(12 - t)\,dt = \frac{1}{34}(72 + 368/3) = 5.73.$$

10.18

a) On average X will equal $1/2$ and, on average, Y will equal $1/2$ of X. Hence, on average, Y will equal $1/4$. Thus, we guess that $\mathcal{E}(Y) = 1/4$.

b) By assumption,

$$f_X(x) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

and, for $0 < x < 1$,

$$f_{Y\,|\,X}(y\,|\,x) = \begin{cases} 1/x, & \text{if } 0 < y < x; \\ 0, & \text{otherwise.} \end{cases}$$

From the general multiplication rule and Proposition 9.7 on page 510,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{-\infty}^{\infty} f_X(x) f_{Y\,|\,X}(y\,|\,x)\,dx = \int_y^1 1 \cdot \frac{1}{x}\,dx = \int_y^1 \frac{dx}{x} = -\ln y$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Applying Definition 10.1 and using integration by parts, we now get

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = -\int_0^1 y \ln y\,dy = -\left[\frac{y^2}{2}\ln y - \frac{y^2}{4}\right]_0^1 = \frac{1}{4}.$$

Theory Exercises

10.19 Let f_X be any PDF of X. We will prove that the function g defined by $g(x) = f_X(x)$ if $|x| \le M$, and $g(x) = 0$ otherwise, is a PDF of X. To do so, we show that g satisfies Equation (8.15) on page 422. First note that, because $P(|X| \le M) = 1$, we have $F_X(x) = 0$ if $x \le -M$, and $F_X(x) = 1$ if $x \ge M$. Now we consider three cases.

Case 1: $x < -M$

In this case,

$$F_X(x) = 0 = \int_{-\infty}^{x} g(t)\,dt.$$

Case 2: $-M \le x < M$

In this case,

$$\begin{aligned} F_X(x) &= F_X(-M) + \big(F_X(x) - F_X(-M)\big) = 0 + \int_{-M}^{x} f_X(t)\,dt \\ &= \int_{-\infty}^{-M} g(t)\,dt + \int_{-M}^{x} g(t)\,dt = \int_{-\infty}^{x} g(t)\,dt. \end{aligned}$$

Case 3: $x \ge M$

In this case,

$$\begin{aligned} F_X(x) &= F_X(-M) + \big(F_X(M) - F_X(-M)\big) + \big(F_X(x) - F_X(M)\big) \\ &= 0 + \int_{-M}^{M} f_X(t)\,dt + (1 - 1) = \int_{-M}^{M} g(t)\,dt \\ &= \int_{-\infty}^{-M} g(t)\,dt + \int_{-M}^{M} g(t)\,dt + \int_{M}^{x} g(t)\,dt = \int_{-\infty}^{x} g(t)\,dt. \end{aligned}$$

We have shown that $F_X(x) = \int_{-\infty}^{x} g(t)\,dt$ for all $x \in \mathcal{R}$. Hence, g is a PDF of X. Thus, without loss of generality, we can assume that a PDF of X is such that $f_X(x) = 0$ for $|x| > M$. Now,

$$\int_{-\infty}^{\infty} |x| f_X(x)\,dx = \int_{-M}^{M} |x| f_X(x)\,dx \le M \int_{-M}^{M} f_X(x)\,dx = M P(|X| \le M) = M < \infty.$$

Hence X has finite expectation. Moreover,

$$|\,\mathcal{E}(X)| = \left| \int_{-\infty}^{\infty} x f_X(x)\,dx \right| \le \int_{-\infty}^{\infty} |x| f_X(x)\,dx \le M.$$

10.20

a) Define $g(x) = xf_X(c-x)$. Then, because f_X is symmetric about c, we have

$$g(-x) = -xf_X(c+x) = -xf_X(c-x) = -g(x).$$

Thus, g is an odd function. Consequently, making the substitution $y = c - x$, we get

$$\begin{aligned}\int_{-\infty}^{\infty} xf_X(x)\,dx &= \int_{-\infty}^{\infty} (c-y)f_X(c-y)\,dy = c\int_{-\infty}^{\infty} f_X(c-y)\,dy - \int_{-\infty}^{\infty} yf_X(c-y)\,dy \\ &= c\int_{-\infty}^{\infty} f_X(x)\,dx - \int_{-\infty}^{\infty} g(y)\,dy = c\cdot 1 - 0 = c.\end{aligned}$$

b) If $X \sim \mathcal{U}(a,b)$, then f_X is symmetric about $(a+b)/2$, so that, by part (a), we have $\mathcal{E}(X) = (a+b)/2$. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then f_X is symmetric about μ, so that, by part (a), we have $\mathcal{E}(X) = \mu$. If $X \sim \mathcal{T}(a,b)$, then f_X is symmetric about $(a+b)/2$, so that, by part (a), we have $\mathcal{E}(X) = (a+b)/2$.
c) If a PDF of X is an even function (i.e., symmetric about 0), then, by part (a), $\mathcal{E}(X) = 0$.

Advanced Exercises

10.21

a) We note that, as $xf_X(x) \geq 0$ for $x \geq 0$, the function G defined on $[0,\infty)$ by $G(x) = \int_0^x tf_X(t)\,dt$ is nondecreasing. If G is bounded, then $\lim_{x\to\infty} G(x) = U$, where $U = \sup\{\, G(x) : x \geq 0\,\}$. Hence,

$$\int_0^{\infty} xf_X(x)\,dx = \lim_{x\to\infty}\int_0^x tf_X(t)\,dt = \lim_{x\to\infty} G(x) = U.$$

If G is unbounded, then $\lim_{x\to\infty} G(x) = \infty$, so that

$$\int_0^{\infty} xf_X(x)\,dx = \lim_{x\to\infty}\int_0^x tf_X(t)\,dt = \lim_{x\to\infty} G(x) = \infty.$$

b) We note that, as $xf_X(x) \leq 0$ for $x \leq 0$, the function H defined on $(-\infty, 0]$ by $H(x) = \int_x^0 tf_X(t)\,dt$ is nondecreasing. If H is bounded, then $\lim_{x\to-\infty} H(x) = L$, where $L = \inf\{\, H(x) : x \leq 0\,\}$. Hence,

$$\int_{-\infty}^0 xf_X(x)\,dx = \lim_{x\to-\infty}\int_x^0 tf_X(t)\,dt = \lim_{x\to-\infty} H(x) = L.$$

If H is unbounded, then $\lim_{x\to\infty} H(x) = -\infty$, so that

$$\int_{-\infty}^0 xf_X(x)\,dx = \lim_{x\to-\infty}\int_x^0 tf_X(t)\,dt = \lim_{x\to-\infty} H(x) = -\infty.$$

c) We have

$$\begin{aligned}\int_{-\infty}^{\infty} |x|f_X(x)\,dx &= \int_0^{\infty} |x|f_X(x)\,dx + \int_{-\infty}^0 |x|f_X(x)\,dx \\ &= \int_0^{\infty} xf_X(x)\,dx - \int_{-\infty}^0 xf_X(x)\,dx = \mathcal{E}(X^+) + \mathcal{E}(X^-).\end{aligned}$$

By definition, X has finite expectation if and only if $\int_{-\infty}^{\infty} |x|f_X(x)\,dx < \infty$, which, as we see from the previous display, is the case if and only if both $\mathcal{E}(X^+)$ and $\mathcal{E}(X^-)$ are real numbers. In that case,

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} xf_X(x)\,dx = \int_0^{\infty} xf_X(x)\,dx + \int_{-\infty}^0 xf_X(x)\,dx = \mathcal{E}(X^+) - \mathcal{E}(X^-).$$

d) Let X be a random variable with PDF given by

$$f_X(x) = \begin{cases} \dfrac{2}{\pi(1+x^2)}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

We have $\mathcal{E}(X^-) = 0$ and, from the solution to Example 10.5 on page 568, we see that $\mathcal{E}(X^+) = \infty$. Thus, X has infinite expectation, and

$$\mathcal{E}(X) = \mathcal{E}(X^+) - \mathcal{E}(X^-) = \infty - 0 = \infty.$$

e) No, the standard Cauchy distribution does not have infinite expectation because, as is easily seen from the solution to Example 10.5, both $\mathcal{E}(X^+)$ and $\mathcal{E}(X^-)$ equal ∞.

10.22

a) We note that T has finite expectation if and only if $\int_{-\infty}^{\infty} |t|\left(1+t^2/\nu\right)^{-(\nu+1)/2}\,dt < \infty$. Using symmetry and making the substitution $u = 1 + t^2/\nu$, we find that

$$\int_{-\infty}^{\infty} \frac{|t|}{\left(1+t^2/\nu\right)^{(\nu+1)/2}}\,dt = 2\int_0^{\infty} \frac{t}{\left(1+t^2/\nu\right)^{(\nu+1)/2}}\,dt = \nu \int_1^{\infty} \frac{du}{u^{(\nu+1)/2}}.$$

From calculus, we know that the integral on the right of the preceding display converges (i.e., is finite) if and only if $(\nu+1)/2 > 1$, that is, if and only if $\nu > 1$.

b) We see that the PDF of T is an even function. Hence, when T has finite expectation, we see from Exercise 10.20(c) that $\mathcal{E}(T) = 0$.

10.2 Basic Properties of Expected Value

Basic Exercises

10.23

a) Applying, in turn, Proposition 8.10 on page 441 and Proposition 8.13 on page 466, we deduce that X^2/σ^2 has the chi-square distribution with 1 degree of freedom, as do both Y^2/σ^2 and Z^2/σ^2. Because X, Y, and Z are independent, so are X^2/σ^2, Y^2/σ^2, and Z^2/σ^2 (Proposition 6.13 on page 297). Applying the second bulleted item on page 537, we conclude that $W = (X^2+Y^2+Z^2)/\sigma^2$ has the chi-square distribution with three degrees of freedom:

$$f_W(w) = \frac{\left(\frac{1}{2}\right)^{3/2}}{\Gamma(3/2)} w^{3/2-1} e^{-w/2} = \frac{1}{\sqrt{2\pi}} w^{1/2} e^{-w/2}, \qquad w > 0,$$

and $f_W(w) = 0$ otherwise.

b) We note that $S = \sigma\sqrt{W}$. Thus, from part (a) and the FEF,

$$\begin{aligned} \mathcal{E}(S) = \mathcal{E}\left(\sigma\sqrt{W}\right) &= \int_{-\infty}^{\infty} \left(\sigma\sqrt{w}\right) f_W(w)\,dw = \frac{\sigma}{\sqrt{2\pi}} \int_0^{\infty} \sqrt{w}\cdot w^{1/2} e^{-w/2}\,dw \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_0^{\infty} w^{2-1} e^{-w/2}\,dw = \frac{\sigma}{\sqrt{2\pi}} \frac{\Gamma(2)}{(1/2)^2} = 2\sqrt{2/\pi}\,\sigma. \end{aligned}$$

10.24

a) To obtain a PDF of $W = |Z|$, we use the CDF method. For $w > 0$,

$$F_W(w) = P(W \le w) = P(|Z| \le w) = P(-w \le Z \le w) = \Phi(w) - \Phi(-w) = 2\Phi(w) - 1.$$

Differentiation now yields

$$f_W(w) = F'_W(w) = 2\phi(w) = 2 \cdot \frac{1}{\sqrt{2\pi}}\, e^{-w^2/2} = \sqrt{2/\pi}\, e^{-w^2/2}$$

if $w > 0$, and $f_W(w) = 0$ otherwise. Applying now Definition 10.1 on page 565 and making the substitution $u = w^2/2$, we get

$$\mathcal{E}(|Z|) = \mathcal{E}(W) = \int_{-\infty}^{\infty} w f_W(w)\, dw = \sqrt{2/\pi} \int_0^{\infty} w\, e^{-w^2/2}\, dw = \sqrt{2/\pi} \int_0^{\infty} e^{-u}\, du = \sqrt{2/\pi}.$$

b) From the FEF and symmetry, we get, upon making the substitution $u = z^2/2$,

$$\begin{aligned}\mathcal{E}(|Z|) &= \int_{-\infty}^{\infty} |z|\phi(z)\, dz = \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = 2\int_0^{\infty} z \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z\, e^{-z^2/2}\, dz = \sqrt{2/\pi} \int_0^{\infty} e^{-u}\, du = \sqrt{2/\pi}.\end{aligned}$$

10.25 Applying the FEF with $g(x, y) = x$, we get

$$\begin{aligned}\mathcal{E}(X) = \mathcal{E}\big(g(X, Y)\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x, y)\, dx\, dy \\ &= \int_0^2 \left(\int_0^{x/2} x \cdot 2xy\, dy\right) dx = 2\int_0^2 x^2 \left(\int_0^{x/2} y\, dy\right) dx = 2\int_0^2 x^2 \left(\frac{x^2}{8}\right) dx \\ &= \frac{1}{4}\int_0^2 x^4\, dx = \frac{1}{4} \cdot \frac{2^5}{5} = \frac{8}{5}.\end{aligned}$$

Applying the FEF with $g(x, y) = y$, we get

$$\begin{aligned}\mathcal{E}(Y) = \mathcal{E}\big(g(X, Y)\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x, y)\, dx\, dy \\ &= \int_0^2 \left(\int_0^{x/2} y \cdot 2xy\, dy\right) dx = 2\int_0^2 x \left(\int_0^{x/2} y^2\, dy\right) dx = 2\int_0^2 x \left(\frac{x^3}{24}\right) dx \\ &= \frac{1}{12}\int_0^2 x^4\, dx = \frac{1}{12} \cdot \frac{2^5}{5} = \frac{8}{15}.\end{aligned}$$

10.26 We know from Exercise 10.17 that

$$f_{T_1,T_2}(t_1, t_2) = \begin{cases} 1/34, & \text{if } 0 < t_1 < 6,\ 0 < t_2 < 6, \text{ and } t_1 + t_2 < 10; \\ 0, & \text{otherwise.} \end{cases}$$

From Proposition 10.2 on page 576 and symmetry,

$$\mathcal{E}(T_1 + T_2) = \mathcal{E}(T_1) + \mathcal{E}(T_2) = 2\mathcal{E}(T_1).$$

Applying the FEF with $g(t_1, t_2) = t_1$ yields

$$\begin{aligned}\mathcal{E}(T_1) = \mathcal{E}\big(g(T_1, T_2)\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t_1, t_2) f_{T_1,T_2}(t_1, t_2)\, dt_1\, dt_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_1 f_{T_1,T_2}(t_1, t_2)\, dt_1\, dt_2 \\ &= \frac{1}{34}\int_0^4 t_1 \left(\int_0^6 1\, dt_2\right) dt_1 + \frac{1}{34}\int_4^6 t_1 \left(\int_0^{10-t_1} 1\, dt_2\right) dt_1 \\ &= \frac{6}{34}\int_0^4 t_1\, dt_1 + \frac{1}{34}\int_4^6 t_1(10 - t_1)\, dt_1 = \frac{146}{51}.\end{aligned}$$

Hence, $\mathcal{E}(T_1 + T_2) = 2 \cdot (146/51) = 5.73$.

10.27 Let Y denote the policyholder's loss and let X denote the benefit paid. Noting that $X = \min\{10, Y\}$, we apply the FEF with $g(y) = \min\{10, y\}$ to get

$$\begin{aligned}\mathcal{E}(X) = \mathcal{E}\big(g(Y)\big) &= \int_{-\infty}^{\infty} g(y) f_Y(y)\, dy = \int_{-\infty}^{\infty} \min\{10, y\} f_Y(y)\, dy = 2\int_1^{\infty} \min\{10, y\} \cdot \frac{1}{y^3}\, dy \\ &= 2\int_1^{10} \frac{y}{y^3}\, dy + 2\int_{10}^{\infty} \frac{10}{y^3}\, dy = 2\int_1^{10} \frac{dy}{y^2} + 20\int_{10}^{\infty} \frac{dy}{y^3} = \frac{18}{10} + \frac{1}{10} = 1.9.\end{aligned}$$

10.28

a) We have $f_X(x) = 1$ if $1 < x < 2$, and $f_X(x) = 0$ otherwise. The range of X is the interval $(1, 2)$. Here $g(x) = 1/x$, which is strictly decreasing and differentiable on the range of X. The transformation method is appropriate. Note that $g^{-1}(y) = 1/y$. Letting $Y = 1/X$, we see that the range of Y is the interval $(1/2, 1)$. Applying the transformation method now yields

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{1}{\left|-1/x^2\right|} f_X(x) = x^2 \cdot 1 = \frac{1}{y^2}$$

if $1/2 < y < 1$, and $f_Y(y) = 0$ otherwise. Therefore, from Definition 10.1 on page 565,

$$\mathcal{E}(1/X) = \mathcal{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_{1/2}^{1} y \cdot \frac{1}{y^2}\, dy = \int_{1/2}^{1} \frac{dy}{y} = \ln 1 - \ln(1/2) = \ln 2.$$

b) Applying the FEF with $g(x) = 1/x$ yields

$$\mathcal{E}(1/X) = \int_{-\infty}^{\infty} (1/x) f_X(x)\, dx = \int_1^2 (1/x) \cdot 1\, dx = \int_1^2 \frac{dx}{x} = \ln 2 - \ln 1 = \ln 2.$$

c) Because $X \sim \mathcal{U}(1, 2)$, we know that $\mathcal{E}(X) = (1 + 2)/2 = 1.5$. Referring to either part (a) or part (b) shows that

$$\mathcal{E}(1/X) = \ln 2 \neq 2/3 = \frac{1}{1.5} = \frac{1}{\mathcal{E}(X)}.$$

Thus, in general, $\mathcal{E}(1/X) \neq 1/\mathcal{E}(X)$.

10.29 Let T denote the lifetime, in years, of the piece of equipment. We know that $T \sim \mathcal{E}(0.1)$. Let A be the amount paid. Then

$$A = \begin{cases} x, & \text{if } 0 < T \le 1; \\ 0.5x, & \text{if } 1 < T \le 3; \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$g(t) = \begin{cases} x, & \text{if } 0 < t \le 1; \\ 0.5x, & \text{if } 1 < t \le 3; \\ 0, & \text{otherwise.} \end{cases}$$

so that $A = g(T)$. Applying the FEF, we obtain

$$\begin{aligned}\mathcal{E}(A) = \mathcal{E}\big(g(T)\big) &= \int_{-\infty}^{\infty} g(t) f_T(t) = \int_0^1 x f_T(t)\,dt + \int_1^3 0.5x f_T(t)\,dt \\ &= xP(T \le 1) + 0.5xP(1 < T \le 3) = \Big(\big(1 - e^{-0.1}\big) + 0.5\big(e^{-0.1} - e^{-0.3}\big)\Big)x.\end{aligned}$$

Setting this last expression equal to 1000, we find that $x = 5644.23$.

10.30 From Example 9.4,

$$f_{X,Y}(x, y) = n(n-1)(y-x)^{n-2}$$

if $0 < x < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise.
a) Applying the FEF with $g(x, y) = x$ and making the substitution $u = y - x$, we get

$$\begin{aligned}\mathcal{E}(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x, y)\,dx\,dy = \int_0^1 \left(\int_x^1 x \cdot n(n-1)(y-x)^{n-2}\,dy\right)dx \\ &= n(n-1)\int_0^1 x\left(\int_0^{1-x} u^{n-2}\,du\right)dx = n\int_0^1 x^{2-1}(1-x)^{n-1}\,dx = nB(2, n) \\ &= n \cdot \frac{1}{(n+1)n} = \frac{1}{n+1}.\end{aligned}$$

b) Applying the FEF with $g(x, y) = y$ and making the substitution $u = y - x$, we get

$$\begin{aligned}\mathcal{E}(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x, y)\,dx\,dy = \int_0^1 \left(\int_0^y y \cdot n(n-1)(y-x)^{n-2}\,dx\right)dy \\ &= n(n-1)\int_0^1 y\left(\int_0^y u^{n-2}\,du\right)dy = n\int_0^1 y^n\,dy = \frac{n}{n+1}.\end{aligned}$$

c) Applying the FEF with $g(x, y) = y - x$ and making the substitution $u = y - x$, we get

$$\begin{aligned}\mathcal{E}(R) &= \mathcal{E}(Y - X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-x) f_{X,Y}(x, y)\,dx\,dy \\ &= \int_0^1 \left(\int_0^y (y-x) \cdot n(n-1)(y-x)^{n-2}\,dx\right)dy = n(n-1)\int_0^1 \left(\int_0^y u^{n-1}\,du\right)dy \\ &= (n-1)\int_0^1 y^n\,dy = \frac{n-1}{n+1}.\end{aligned}$$

d) Applying the FEF with $g(x, y) = (x + y)/2$ and making the substitution $u = y - x$, we get

$$\begin{aligned}
\mathcal{E}(M) = \mathcal{E}\big((X+Y)/2\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x+y}{2} f_{X,Y}(x, y)\,dx\,dy \\
&= \frac{1}{2}\int_0^1 \left(\int_0^y (x+y)\cdot n(n-1)(y-x)^{n-2}\,dx\right) dy \\
&= \frac{n(n-1)}{2}\int_0^1 \left(\int_0^y (2y-u)u^{n-2}\,du\right) dy \\
&= n(n-1)\int_0^1 y\left(\int_0^y u^{n-2}\,du\right) dy - \frac{n(n-1)}{2}\int_0^1 \left(\int_0^y u^{n-1}\,du\right) dy \\
&= n\int_0^1 y^n\,dy - \frac{(n-1)}{2}\int_0^1 y^n\,dy = \frac{1}{2}.
\end{aligned}$$

10.31

a) We have

$$\begin{aligned}
\mathcal{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x\cdot\frac{1}{B(1,n)}x^{1-1}(1-x)^{n-1}\,dx = n\int_0^1 x^{2-1}(1-x)^{n-1}\,dx \\
&= nB(2,n) = n\cdot\frac{1}{(n+1)n} = \frac{1}{n+1}.
\end{aligned}$$

b) We have

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_0^1 y\cdot\frac{1}{B(n,1)}y^{n-1}(1-y)^{1-1}\,dy = n\int_0^1 y^n\,dy = \frac{n}{n+1}.$$

c) Applying Equation (10.13) on page 576 and referring to parts (a) and (b), we get

$$\mathcal{E}(R) = \mathcal{E}(Y-X) = \mathcal{E}(Y) - \mathcal{E}(X) = \frac{n}{n+1} - \frac{1}{n+1} = \frac{n-1}{n+1}.$$

d) Applying Proposition 10.2 on page 576 and referring to parts (a) and (b), we get

$$\mathcal{E}(M) = \mathcal{E}\big((X+Y)/2\big) = \frac{1}{2}\big(\mathcal{E}(X)+\mathcal{E}(Y)\big) = \frac{1}{2}\left(\frac{1}{n+1}+\frac{n}{n+1}\right) = \frac{1}{2}.$$

10.32 Let T denote the payment. We have

$$T = \begin{cases} 0, & \text{if } X+Y \le 1; \\ X+Y-1, & \text{if } X+Y > 1. \end{cases}$$

Define

$$g(x, y) = \begin{cases} 0, & \text{if } x+y \le 1; \\ x+y-1, & \text{if } x+y > 1. \end{cases}$$

Then $T = g(X, Y)$. Applying the FEF now yields

$$\begin{aligned}
\mathcal{E}(T) = \mathcal{E}\big(g(X,Y)\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{1-x} 0 \cdot f_{X,Y}(x,y)\,dy + \int_{1-x}^{\infty}(x+y-1)\cdot f_{X,Y}(x,y)\,dy\right)dx \\
&= \int_0^1\left(\int_{1-x}^{1}(x+y-1)\cdot 2x\,dy\right)dx = 2\int_0^1 x\left(\int_{1-x}^{1}(x+y-1)\,dy\right)dx \\
&= 2\int_0^1 x\cdot\frac{x^2}{2}\,dx = \int_0^1 x^3\,dx = \frac{1}{4}.
\end{aligned}$$

Thus, the expected payment is \$250.

10.33

a) From Example 9.25(b), we know that the proportion of the total inspection time attributed to the first engineer is $\mathcal{U}(0, 1)$. Because the expected value of a $\mathcal{U}(0, 1)$ random variable is 1/2, we see that, on average, 50% of the total inspection time is attributed to the first engineer.

b) Let X and Y denote the proportions of the total inspection time attributed to the first and second engineers, respectively. Clearly, $X + Y = 1$. Because the inspection times for the two engineers have the same probability distribution, it follows, by symmetry, that X and Y have the same probability distribution and, hence, that $\mathcal{E}(X) = \mathcal{E}(Y)$. Using the linearity property of expected value and the fact that the expected value of a constant random variable is the constant, we get

$$1 = \mathcal{E}(1) = \mathcal{E}(X+Y) = \mathcal{E}(X) + \mathcal{E}(Y) = 2\mathcal{E}(X)\,.$$

Consequently, $\mathcal{E}(X) = 1/2$. On average, 50% of the total inspection time is attributed to the first engineer.

10.34

a) Let $U \sim \mathcal{U}(0, \ell)$. For $0 \le x < \ell/2$,

$$F_X(x) = P(X \le x) = P(U \le x \ \text{ or } \ U \ge \ell - x) = \frac{x}{\ell} + \frac{\ell-(\ell-x)}{\ell} = \frac{2x}{\ell} = \frac{x}{\ell/2}.$$

Hence, $X \sim \mathcal{U}(0, \ell/2)$.

b) Noting that $R = X/(\ell - X)$, we apply the FEF and make the substitution $u = \ell - x$ to get

$$\begin{aligned}
\mathcal{E}(R) = \mathcal{E}\left(\frac{X}{\ell - X}\right) &= \int_{-\infty}^{\infty}\frac{x}{\ell - x} f_X(x)\,dx = \int_0^{\ell/2}\frac{x}{\ell-x}\cdot\frac{1}{\ell/2}\,dx \\
&= \frac{2}{\ell}\int_{\ell/2}^{\ell}\frac{\ell-u}{u}\,du = 2\int_{\ell/2}^{\ell}\frac{du}{u} - \frac{2}{\ell}\int_{\ell/2}^{\ell} 1\,du = 2\ln 2 - 1.
\end{aligned}$$

c) The range of X is the interval $(0, \ell/2)$. To determine a PDF of R, we use the transformation method with $g(x) = x/(\ell - x)$, which is strictly increasing and differentiable on the range of X. Note that $g^{-1}(r) = r\ell/(1 + r)$ and that the range of R is the interval $(0, 1)$. Applying the transformation method now yields

$$f_R(r) = \frac{1}{|g'(x)|} f_X(x) = \frac{1}{\ell/(\ell-x)^2}\,\frac{1}{\ell/2} = \frac{2(\ell-x)^2}{\ell^2} = \frac{2}{\ell^2}\left(\ell - \frac{r\ell}{1+r}\right)^2 = \frac{2}{(1+r)^2}$$

if $0 < r < 1$, and $f_R(r) = 0$ otherwise.

d) From the definition of expected value and part (c), we find, upon making the substitution $u = 1 + r$,

$$\begin{aligned}\mathcal{E}(R) &= \int_{-\infty}^{\infty} r f_R(r)\,dr = \int_0^1 r \cdot \frac{2}{(1+r)^2}\,dr = 2\int_1^2 \frac{u-1}{u^2}\,du \\ &= 2\int_1^2 \left(\frac{1}{u} - \frac{1}{u^2}\right) du = 2\ln 2 - 1.\end{aligned}$$

e) The ratio of the longer segment to the shorter segment is $(\ell - X)/X$. Letting $g(x) = (\ell - x)/x$, we have

$$\int_{-\infty}^{\infty} |g(x)| f_X(x)\,dx = \int_{-\infty}^{\infty} \left|\frac{\ell - x}{x}\right| f_X(x)\,dx = \frac{2}{\ell}\int_0^{\ell/2} \frac{\ell - x}{x}\,dx = 2\int_0^{\ell/2} \frac{dx}{x} - 1 = \infty.$$

Hence, by the FEF, $(\ell - X)/X$ doesn't have finite expectation.

10.35
a) From Exercise 10.34(c), the ratio of the shorter piece to the longer piece doesn't depend on ℓ; thus, we can assume that $\ell = 2$. Then, from Exercise 10.34(a), $X \sim \mathcal{U}(0, 1)$. The idea is to use a basic random number generator to obtain a large number of, say, 10,000, uniform random numbers on the interval $(0, 1)$. For each such random number, u, compute $r = u/(2 - u)$, which gives 10,000 independent observations of the ratio of the smaller segment to the longer segment. By the long-run-average interpretation of expected value, the average of these 10,000 observations should approximately equal $\mathcal{E}(R)$, which, from Exercise 10.34(d), is $2\ln 2 - 1 \approx 0.386$.
b) Answers will vary, but you should get roughly 0.386.

10.36 The area of the triangle equals $1/2$. Thus,

$$f_{X,Y}(x, y) = \begin{cases} 2, & \text{if } 0 < y < x \text{ and } 0 < x < 1; \\ 0, & \text{otherwise.}\end{cases}$$

a) Applying the FEF, we get

$$\begin{aligned}\mathcal{E}(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x, y)\,dx\,dy = \int_0^1 \left(\int_0^x x \cdot 2\,dy\right) dx = 2\int_0^1 x^2\,dx = \frac{2}{3}, \\ \mathcal{E}(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x, y)\,dx\,dy = \int_0^1 \left(\int_0^x y \cdot 2\,dy\right) dx = \int_0^1 x^2\,dx = \frac{1}{3},\end{aligned}$$

and

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x, y)\,dx\,dy = \int_0^1 \left(\int_0^x xy \cdot 2\,dy\right) dx = \int_0^1 x^3\,dx = \frac{1}{4}.$$

Therefore,

$$\mathcal{E}(XY) = \frac{1}{4} \neq \frac{2}{9} = \frac{2}{3} \cdot \frac{1}{3} = \mathcal{E}(X)\,\mathcal{E}(Y);$$

that is, $\mathcal{E}(XY) \neq \mathcal{E}(X)\,\mathcal{E}(Y)$.
b) Yes, we can conclude that X and Y aren't independent random variables. Indeed, by Proposition 10.4 on page 576, if X and Y were independent random variables, then we would have $\mathcal{E}(XY) = \mathcal{E}(X)\,\mathcal{E}(Y)$, which, by part (a) isn't the case.

10.37 The area of the triangle equals 1. Thus,

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{if } -x < y < x \text{ and } 0 < x < 1; \\ 0, & \text{otherwise.}\end{cases}$$

a) By symmetry, $\mathcal{E}(Y) = 0$, so that $\mathcal{E}(X)\,\mathcal{E}(Y) = 0$. Applying the FEF, we get

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy = \int_0^1 \left(\int_{-x}^{x} xy \cdot 1\,dy\right) dx$$
$$= \int_0^1 x \left(\int_{-x}^{x} y\,dy\right) dx = \int_0^1 x \cdot 0\,dx = 0.$$

Therefore, $\mathcal{E}(XY) = \mathcal{E}(X)\,\mathcal{E}(Y)$.
b) No, we cannot conclude that X and Y are independent random variables. Independence of random variables is not a necessary condition for the expected value of a product to equal the product of the expected values; in other words, it's possible for the expected value of a product to equal the product of the expected values without the random variables being independent.
c) No, X and Y aren't independent random variables. We can see this fact in many ways. One way is to note that the range of Y is the interval $(-1, 1)$, whereas, the range of $Y_{|X=1/2}$ is the interval $(-1/2, 1/2)$.

10.38 Let X, Y, and Z denote the three claim amounts. We know that these three random variables all have f as their PDF, and we can assume that they are independent. The largest of the three claim amounts is $\max\{X, Y, Z\}$, which we denote by W.
a) We first note that the CDF corresponding to f is

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_1^x \frac{3}{t^4}\,dt = 1 - \frac{1}{x^3}$$

if $x \geq 1$, and $F(x) = 0$ otherwise. From Exercise 9.70(a),

$$f_W(w) = 3f(w)\big(F(w)\big)^2 = 3 \cdot \frac{3}{w^4}\left(1 - \frac{1}{w^3}\right)^2 = \frac{9}{w^4}\left(1 - \frac{1}{w^3}\right)^2$$

if $w > 1$, and $f_W(w) = 0$ otherwise. Hence, by Definition 10.1 on page 565,

$$\mathcal{E}(W) = \int_{-\infty}^{\infty} w f_W(w)\,dw = \int_1^{\infty} w \cdot \frac{9}{w^4}\left(1 - \frac{1}{w^3}\right)^2 dw = 9\int_1^{\infty}\left(\frac{1}{w^3} - \frac{2}{w^6} + \frac{1}{w^9}\right) dw = \frac{81}{40}.$$

Hence, the expected value of the largest of the three claims is \$(81/40) thousand, or \$2025.
b) From the FEF, independence, and symmetry,

$$\mathcal{E}(W) = \mathcal{E}(\max\{X, Y, Z\}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\{x, y, z\} f_{X,Y,Z}(x, y, z)\,dx\,dy\,dz$$
$$= 6 \iiint_{x>y>z} x f_X(x) f_Y(y) f_Z(z)\,dx\,dy\,dz = 6 \cdot 27 \int_1^{\infty}\left(\int_z^{\infty}\left(\int_y^{\infty} x \cdot \frac{1}{x^4}\,dx\right)\frac{1}{y^4}\,dy\right)\frac{1}{z^4}\,dz$$
$$= 3 \cdot 27 \int_1^{\infty}\left(\int_z^{\infty} \frac{1}{y^6}\,dy\right)\frac{1}{z^4}\,dz = \frac{3 \cdot 27}{5}\int_1^{\infty} \frac{1}{z^9}\,dz = \frac{3 \cdot 27}{5 \cdot 8} = \frac{81}{40}.$$

Hence, the expected value of the largest of the three claims is \$(81/40) thousand, or \$2025.

10.39 Let X_1 and X_2 denote the arrival times of the two people. By assumption, X_1 and X_2 are independent random variables with common PDF f. The arrival times of the first and second persons to arrive are $X = \min\{X_1, X_2\}$ and $Y = \max\{X_1, X_2\}$, respectively. The expected amount of time that the first person to arrive waits for the second person to arrive is $\mathcal{E}(Y - X)$.

a) Noting that $Y - X = |X_1 - X_2|$, applying the FEF, and using independence and symmetry, we get

$$\begin{aligned}
\mathcal{E}(Y - X) = \mathcal{E}(|X_1 - X_2|) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x_1 - x_2| f_{X_1,X_2}(x_1, x_2)\,dx_1\,dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x_1 - x_2| f(x_1) f(x_2)\,dx_1\,dx_2 = 2\iint\limits_{x_1<x_2} (x_2 - x_1) f(x_1) f(x_2)\,dx_1\,dx_2 \\
&= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y} (y - x) f(x)\,dx\right) dy.
\end{aligned}$$

b) Applying the FEF and referring to Equation (9.20) on page 498, we get

$$\begin{aligned}
\mathcal{E}(Y - X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - x) f_{X,Y}(x, y)\,dx\,dy \\
&= \iint\limits_{x<y} (y - x) \cdot 2f(x) f(y)\,dx\,dy = 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y} (y - x) f(x)\,dx\right) dy.
\end{aligned}$$

c) From Example 9.4 and Exercise 9.70,

$$f_X(x) = 2f(x)\big(1 - F(x)\big) \qquad \text{and} \qquad f_Y(y) = 2f(y)F(y).$$

Hence, by the linearity property of expected value,

$$\begin{aligned}
\mathcal{E}(Y - X) = \mathcal{E}(Y) - \mathcal{E}(X) &= \int_{-\infty}^{\infty} y f_Y(y)\,dy - \int_{-\infty}^{\infty} x f_X(x)\,dx \\
&= 2\int_{-\infty}^{\infty} y f(y) F(y)\,dy - 2\int_{-\infty}^{\infty} x f(x)\big(1 - F(x)\big)\,dx \\
&= 2\int_{-\infty}^{\infty} y f(y)\left(\int_{-\infty}^{y} f(x)\,dx\right) dy - 2\int_{-\infty}^{\infty} x f(x)\left(\int_{x}^{\infty} f(y)\,dy\right) dx \\
&= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y} y f(x)\,dx\right) dy - 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y} x f(x)\,dx\right) dy \\
&= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y} (y - x) f(x)\,dx\right) dy.
\end{aligned}$$

d) From Example 9.6, $f_{Y-X}(r) = 2\int_{-\infty}^{\infty} f(t) f(r + t)\,dt$ if $r > 0$, and $f_{Y-X}(r) = 0$ otherwise. Therefore, by the definition of expected value, and upon making the successive substitutions $y = r + t$ and $x = y - r$, we get

$$\begin{aligned}
\mathcal{E}(Y - X) &= \int_{-\infty}^{\infty} r f_{Y-X}(r)\,dr = 2\int_{0}^{\infty} r\left(\int_{-\infty}^{\infty} f(t) f(r + t)\,dt\right) dr \\
&= 2\int_{0}^{\infty} r\left(\int_{-\infty}^{\infty} f(y - r) f(y)\,dy\right) dr = 2\int_{-\infty}^{\infty} f(y)\left(\int_{0}^{\infty} r f(y - r)\,dr\right) dy \\
&= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y} (y - x) f(x)\,dx\right) dy.
\end{aligned}$$

10.40

a) In this case, $f(t) = 1/2\ell$ if $0 < t < 2\ell$, and $f(t) = 0$ otherwise. Hence, from Exercise 10.39,

$$\begin{aligned}\mathcal{E}(Y-X) &= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y}(y-x)f(x)\,dx\right)dy = 2\int_0^{2\ell}\frac{1}{2\ell}\left(\int_0^y (y-x)\cdot\frac{1}{2\ell}\,dx\right)dy \\ &= \frac{1}{2\ell^2}\int_0^{2\ell}\left(\int_0^y (y-x)\,dx\right)dy = \frac{1}{4\ell^2}\int_0^{2\ell} y^2\,dy = \frac{2}{3}\ell.\end{aligned}$$

b) In this case,

$$f(t) = \begin{cases}\dfrac{2}{2\ell}\left(1-\dfrac{|2\ell-2t|}{2\ell}\right), & \text{if } 0 < t < 2\ell; \\ 0, & \text{otherwise.}\end{cases} = \begin{cases} t/\ell^2, & \text{if } 0 < t < \ell; \\ (2\ell - t)/\ell^2, & \text{if } \ell < t < 2\ell; \\ 0, & \text{otherwise.}\end{cases}$$

Hence, from Exercise 10.39,

$$\begin{aligned}\mathcal{E}(Y-X) &= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y}(y-x)f(x)\,dx\right)dy = 2\int_0^{2\ell} f(y)\left(\int_0^y (y-x)f(x)\,dx\right)dy \\ &= \frac{2}{\ell^4}\Bigg(\int_0^{\ell} y\left(\int_0^y (y-x)\cdot x\,dx\right)dy + \int_{\ell}^{2\ell}(2\ell-y)\left(\int_0^{\ell}(y-x)\cdot x\,dx\right)dy \\ &\quad + \int_{\ell}^{2\ell}(2\ell-y)\left(\int_{\ell}^{y}(y-x)\cdot(2\ell-x)\,dx\right)dy\Bigg) \\ &= \frac{2}{\ell^4}\left(\frac{\ell^5}{30}+\frac{\ell^5}{6}+\frac{\ell^5}{30}\right) = \frac{7}{15}\ell.\end{aligned}$$

10.41

a) Because of the lack-of-memory property of the exponential distribution, the amount of time that the first person to arrive waits for the second person to arrive has the same exponential distribution as that of the common arrival-time distribution of the two people, namely, exponential with mean ℓ.

b) From part (a), we have $\mathcal{E}(Y-X) = \ell$.

c) We know that the common arrival-time distribution is $\mathcal{E}(\lambda)$, where $\lambda = 1/\ell$. Hence, from Exercise 10.39,

$$\begin{aligned}\mathcal{E}(Y-X) &= 2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{y}(y-x)f(x)\,dx\right)dy = 2\int_0^{\infty}\lambda e^{-\lambda y}\left(\int_0^y (y-x)\lambda e^{-\lambda x}\,dx\right)dy \\ &= 2\int_0^{\infty}\lambda e^{-\lambda y}\left(y\int_0^y \lambda e^{-\lambda x}\,dx - \int_0^y x\lambda e^{-\lambda x}\,dx\right)dy \\ &= 2\int_0^{\infty}\lambda e^{-\lambda y}\left(y\left(1-e^{-\lambda y}\right) - \frac{1}{\lambda}\left(1-e^{-\lambda y}-\lambda y e^{-\lambda y}\right)\right)dy \\ &= 2\int_0^{\infty}\left(\lambda y e^{-\lambda y} - e^{-\lambda y} + e^{-2\lambda y}\right)dy = 2\left(\frac{1}{\lambda}-\frac{1}{\lambda}+\frac{1}{2\lambda}\right) = \frac{1}{\lambda} = \ell.\end{aligned}$$

d) The answers obtained in parts (b) and (c) are identical.

10.42 Let T denote the time at which the device is replaced, so that $T = \min\{X, t_r\}$. Observe that, for $t > 0$,

$$P(T > t) = P(\min\{X, t_r\} > t) = \begin{cases} P(X > t), & \text{if } t < t_r; \\ 0, & \text{if } t \geq t_r. \end{cases}$$

a) Recall from the first bulleted item on page 578 that Proposition 10.5 holds for any type of random variable. Hence,

$$\mathcal{E}(T) = \int_0^\infty P(T > t)\,dt = \int_0^{t_r} P(X > t)\,dt = \int_0^{t_r} \big(1 - F_X(t)\big)\,dt.$$

b) Let Y denote the cost of replacement. Note that $Y = c_1$ if $X > t_r$, and $Y = c_2$ if $X \leq t_r$. Hence, Y is a discrete random variable, and we have

$$\begin{aligned} \mathcal{E}(Y) &= \sum_y y p_Y(y) = c_1 P(X > t_r) + c_2 P(X \leq t_r) \\ &= c_1\big(1 - F_X(t_r)\big) + c_2 F_X(t_r) = c_1 + (c_2 - c_1)F_X(t_r). \end{aligned}$$

c) In this case, $X \sim \mathcal{E}(0.001)$, so that $F_X(x) = 1 - e^{-0.001x}$ for $x \geq 0$. Hence, from part (b),

$$\mathcal{E}(Y) = 50 + (80 - 50)\left(1 - e^{-0.001 \cdot 950}\right) = \$68.40.$$

d) In this case, $X \sim \mathcal{U}(900, 1100)$, so that $F_X(x) = (x - 900)/200$ for $900 \leq x < 1100$. Consequently, from part (b),

$$\mathcal{E}(Y) = 50 + (80 - 50)\left(\frac{950 - 900}{200}\right) = \$57.50.$$

10.43

a) Let E denote the event that the component fails immediately when turned on. Applying the law of total probability, we find that, for $x \geq 0$,

$$\begin{aligned} F_X(x) &= P(X \leq x) = P\big(E\big)P\big(X \leq x \mid E\big) + P\big(E^c\big)P\big(X \leq x \mid E^c\big) \\ &= p \cdot 1 + (1 - p) \cdot F(x) = p + (1 - p)F(x). \end{aligned}$$

Hence,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ p + (1 - p)F(x), & \text{if } x \geq 0. \end{cases}$$

b) Recall from the first bulleted item on page 578 that Proposition 10.5 holds for any type of random variable. Hence, from part (a),

$$\begin{aligned} \mathcal{E}(X) &= \int_0^\infty P(X > x)\,dx = \int_0^\infty \big(1 - F_X(x)\big)\,dx \\ &= \int_0^\infty \Big(1 - \big(p + (1 - p)F(x)\big)\Big)\,dx = (1 - p)\int_0^\infty \big(1 - F(x)\big)\,dx. \end{aligned}$$

c) In this case, $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$. Hence, from part (b),

$$\mathcal{E}(X) = (1 - p)\int_0^\infty e^{-\lambda x}\,dx = \frac{1 - p}{\lambda}.$$

Theory Exercises

10.44

a) We apply the triangle inequality and use the fact that X has finite expectation to conclude that

$$\begin{aligned}\int_{-\infty}^{\infty} |a+bx| f_X(x)\,dx &\le \int_{-\infty}^{\infty} (|a|+|b||x|) f_X(x)\,dx \\ &= |a|\int_{-\infty}^{\infty} f_X(x)\,dx + |b|\int_{-\infty}^{\infty} |x| f_X(x)\,dx \\ &= |a| + |b|\int_{-\infty}^{\infty} |x| f_X(x)\,dx < \infty.\end{aligned}$$

Hence, from the FEF, $a+bX$ has finite expectation, and

$$\begin{aligned}\mathcal{E}(a+bX) &= \int_{-\infty}^{\infty} (a+bx) f_X(x)\,dx = \int_{-\infty}^{\infty} a f_X(x)\,dx + \int_{-\infty}^{\infty} bx f_X(x)\,dx \\ &= a\int_{-\infty}^{\infty} f_X(x)\,dx + b\int_{-\infty}^{\infty} x f_X(x)\,dx = a + b\mathcal{E}(X)\,.\end{aligned}$$

b) Let f_{Y-X} be any PDF of $Y-X$. We will prove that the function g, defined by $g(z)=0$ if $z<0$, and $g(z)=f_{Y-X}(z)$ if $z\ge 0$, is a PDF of $Y-X$. To do so, we show that g satisfies Equation (8.15) on page 422. First note that, because $X \le Y$, we have $P(Y-X\ge 0)=1$. Hence, for $z<0$,

$$F_{Y-X}(z) = P(Y-X\le z) \le P(Y-X<0) = 1-P(Y-X\ge 0) = 1-1 = 0.$$

Thus, $F_{Y-X}(z)=0$ if $z<0$. As F_{Y-X} is everywhere continuous, it follows that, in fact, $F_{Y-X}(z)=0$ if $z\le 0$. Now we consider two cases.

Case 1: $z<0$

In this case,

$$F_{Y-X}(z) = 0 = \int_{-\infty}^{z} 0\,dt = \int_{-\infty}^{z} g(t)\,dt.$$

Case 2: $z\ge 0$

In this case,

$$\begin{aligned}F_{Y-X}(z) &= F_{Y-X}(0) + \big(F_{Y-X}(z) - F_{Y-X}(0)\big) = 0 + \int_0^z f_{Y-X}(t)\,dt \\ &= \int_{-\infty}^{0} g(t)\,dt + \int_0^z g(t)\,dt = \int_{-\infty}^{z} g(t)\,dt.\end{aligned}$$

We have shown that $F_{Y-X}(z) = \int_{-\infty}^{z} g(t)\,dt$ for all $z\in\mathcal{R}$. Hence, g is a PDF of $Y-X$. Thus, without loss of generality, we can assume that a PDF of $Y-X$ is such that $f_{Y-X}(z)=0$ for $z<0$. Applying the linearity property of expected value for continuous random variables, we conclude that $Y-X$ has finite expectation and

$$\mathcal{E}(Y)-\mathcal{E}(X) = \mathcal{E}(Y-X) = \int_{-\infty}^{\infty} z f_{Y-X}(z)\,dz = \int_0^{\infty} z f_{Y-X}(z)\,dz \ge 0.$$

Thus, $\mathcal{E}(X)\le\mathcal{E}(Y)$, as required.

10.45

a) We have

$$\begin{aligned}
\int_{-\infty}^{\infty} |x| f_X(x)\,dx &= \int_{-\infty}^{0} -x f_X(x)\,dx + \int_{0}^{\infty} x f_X(x)\,dx \\
&= \int_{-\infty}^{0} \left(\int_{x}^{0} 1\,dy \right) f_X(x)\,dx + \int_{0}^{\infty} \left(\int_{0}^{x} 1\,dy \right) f_X(x)\,dx \\
&= \int_{-\infty}^{0} \left(\int_{-\infty}^{y} f_X(x)\,dx \right) dy + \int_{0}^{\infty} \left(\int_{y}^{\infty} f_X(x)\,dx \right) dy \\
&= \int_{-\infty}^{0} P(X < y)\,dy + \int_{0}^{\infty} P(X > y)\,dy \\
&= \int_{0}^{\infty} P(X < -y)\,dy + \int_{0}^{\infty} P(X > y)\,dy = \int_{0}^{\infty} P(|X| > y)\,dy.
\end{aligned}$$

Thus, we see that X has finite expectation if and only if $\int_0^\infty P(|X| > x)\,dx < \infty$. In that case, by applying the same argument as the one just given, but without absolute values, we find that

$$\mathcal{E}(X) = \int_0^\infty P(X > x)\,dx - \int_0^\infty P(X < -x)\,dx.$$

b) We have

$$P(X > x) = \begin{cases} 1, & \text{if } x < -1; \\ (3-x)/4, & \text{if } -1 \le x < 3; \\ 0, & \text{if } x \ge 3. \end{cases} \quad \text{and} \quad P(X < -x) = \begin{cases} 1, & \text{if } x < -3; \\ (-x+1)/4, & \text{if } -3 \le x < 1; \\ 0, & \text{if } x \ge 1. \end{cases}$$

Hence, by part (a),

$$\begin{aligned}
\mathcal{E}(X) &= \int_0^\infty P(X > x)\,dx - \int_0^\infty P(X < -x)\,dx \\
&= \int_0^3 \frac{3-x}{4}\,dx - \int_0^1 \frac{-x+1}{4}\,dx = \frac{9}{8} - \frac{1}{8} = 1.
\end{aligned}$$

10.46

a) Because X and $-X$ have the same probability distribution, they have the same expected value. Hence, from the linearity property of expected value, $\mathcal{E}(X) = \mathcal{E}(-X) = -\mathcal{E}(X)$, which implies that $\mathcal{E}(X) = 0$.

b) From part (a) and the linearity property of expected value,

$$0 = \mathcal{E}(X - c) = \mathcal{E}(X) - \mathcal{E}(c) = \mathcal{E}(X) - c.$$

Hence, $\mathcal{E}(X) = c$.

c) Using the symmetry of f_X about c and making the successive substitutions $u = c - t$ and $t = c + u$, we get

$$\begin{aligned}
F_{c-X}(x) = P(c - X \le x) = P(X \ge c - x) = \int_{c-x}^{\infty} f_X(t)\,dt = \int_{-\infty}^{x} f_X(c-u)\,du \\
= \int_{-\infty}^{x} f_X(c+u)\,du = \int_{-\infty}^{c+x} f_X(t)\,dt = P(X \le c + x) = P(X - c \le x) = F_{X-c}(x).
\end{aligned}$$

Therefore, $c - X$ has the same probability distribution as $X - c$, which means that $X - c$ is a symmetric random variable; that is, X is symmetric about c.
d) We note that a $\mathcal{U}(a, b)$ random variable has a PDF that is symmetric about $(a+b)/2$, and, therefore, by parts (c) and (b) has mean $(a+b)/2$. Likewise, we observe that a $\mathcal{N}(\mu, \sigma^2)$ random variable has a PDF that is symmetric about μ, and, therefore, has mean μ. Also, we note that a $\mathcal{T}(a, b)$ random variable has a PDF that is symmetric about $(a+b)/2$, and, therefore, has mean $(a+b)/2$.

Advanced Exercises

10.47 From Exercise 9.33(d), $X_{(k)}$ has the beta distribution with parameters $\alpha = k$ and $\beta = n - k + 1$. Hence, from Table 10.1 on page 568, we have $\mathcal{E}(X_{(k)}) = k/(n+1)$.

10.48
a) From Proposition 9.11(a) on page 532 or Exercise 9.33(e), $X_{(1)} = \min\{X_1, \dots, X_n\}$ has the exponential distribution with parameter $n\lambda$. Hence, we have $\mathcal{E}(X_{(1)}) = 1/n\lambda$.
b) For convenience, we let $X = X_{(j-1)}$ and $Y = X_{(j)}$. From Exercise 9.35(c), a joint PDF of X and Y is

$$f_{X,Y}(x, y) = \frac{n!}{(j-2)!\,(n-j)!}\lambda^2 e^{-\lambda x} e^{-(n-j+1)\lambda y}(1 - e^{-\lambda x})^{j-2}, \qquad 0 < x < y.$$

Set $c = n!\big/\big((j-2)!\,(n-j)!\big)$. Then, for $z > 0$, we get, upon making the substitution $u = 1 - e^{-\lambda x}$,

$$\begin{aligned}
f_{Y-X}(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x, x+z)\,dx = c\int_0^{\infty} \lambda^2 e^{-\lambda x} e^{-(n-j+1)\lambda(x+z)}(1 - e^{-\lambda x})^{j-2}\,dx \\
&= c\lambda e^{-(n-j+1)\lambda z}\int_0^{\infty} \lambda e^{-\lambda x} e^{-(n-j+1)\lambda x}(1 - e^{-\lambda x})^{j-2}\,dx \\
&= c\lambda e^{-(n-j+1)\lambda z}\int_0^1 u^{j-2}(1-u)^{n-j+1}\,du \\
&= c\lambda e^{-(n-j+1)\lambda z}\int_0^1 u^{(j-1)-1}(1-u)^{(n-j+2)-1}\,du = c\lambda e^{-(n-j+1)\lambda z} B(j-1, n-j+2) \\
&= \frac{n!}{(j-2)!\,(n-j)!}\,\frac{(j-2)!\,(n-j+1)!}{n!}\lambda e^{-(n-j+1)\lambda z} = (n-j+1)\lambda e^{-(n-j+1)\lambda z}.
\end{aligned}$$

Hence, we see that $X_{(j)} - X_{(j-1)} \sim \mathcal{E}\big((n-j+1)\lambda\big)$.
c) We have $X_{(k)} = X_{(1)} + \sum_{j=2}^{k}\big(X_{(j)} - X_{(j-1)}\big)$. Therefore, from the linearity property of expected value and parts (a) and (b),

$$\begin{aligned}
\mathcal{E}(X_{(k)}) &= \mathcal{E}(X_{(1)}) + \sum_{j=2}^{k}\mathcal{E}\big(X_{(j)} - X_{(j-1)}\big) = \frac{1}{n\lambda} + \sum_{j=2}^{k}\frac{1}{(n-j+1)\lambda} \\
&= \frac{1}{n\lambda} + \sum_{i=n-k+1}^{n-1}\frac{1}{i\lambda} = \frac{1}{\lambda}\sum_{j=n-k+1}^{n}\frac{1}{j}.
\end{aligned}$$

10.49
a) We can consider the lifetimes of the n components a random sample of size n from an exponential distribution with parameter λ. The lifetime of the unit in a series system is then $X_{(1)}$. From Exercise 10.48(a), we conclude that the expected lifetime of the unit is $1/n\lambda$.

b) Again, we can consider the lifetimes of the n components a random sample of size n from an exponential distribution with parameter λ. The lifetime of the unit in a parallel system is then $X_{(n)}$. From Exercise 10.48(c), we conclude that the expected lifetime of the unit is $(1/\lambda)\sum_{j=1}^{n} 1/j$.
c) This result follows from part (b) and the relation $\sum_{j=1}^{n}(1/j) \sim \ln n$, where, here, the symbol $\sim$ means that the ratio of the two sides approach 1 as $n \to \infty$.

10.50 Consider a parallel system of n components whose lifetimes, $X_1, \ldots, X_n$, are independent $\mathcal{E}(\lambda)$ random variables. The kth order statistic, $X_{(k)}$, represents the time of the kth component failure. Thus, the random variable $X_{(j)} - X_{(j-1)}$ represents the time between the $(j-1)$st and jth component failure. When the $(j-1)$st component failure occurs, there remain $(n-j+1)$ working components. Because of the lack-of-memory property of the exponential distribution, the remaining elapsed time until failure for each of these $(n-j+1)$ working components still has the exponential distribution with parameter λ. Hence the time until the next failure, $X_{(j)} - X_{(j-1)}$, has the probability distribution of the minimum of $(n-j+1)$ independent exponential random variables with parameter λ, which, by Proposition 9.11(a) on page 532, is the exponential distribution with parameter $(n-j+1)\lambda$.

10.3 Variance, Covariance, and Correlation

Basic Exercises

10.51 Let X be the current annual cost and let Y be the annual cost with the 20% increase. Then $Y = 1.2X$.
a) By the linearity property of expected value, we have $\mathcal{E}(Y) = 1.2\mathcal{E}(X)$, so there is a 20% increase in the mean annual cost.
b) From Equation (10.27) on page 586, we have $\mathrm{Var}(Y) = \mathrm{Var}(1.2X) = 1.44\,\mathrm{Var}(X)$, so there is a 44% increase in the variance of the annual cost.
c) Referring to part (b), we have $\sigma_Y = \sqrt{\mathrm{Var}(Y)} = \sqrt{1.44\,\mathrm{Var}(X)} = 1.2\sigma_X$, so there is a 20% increase in the standard deviation of the annual cost.

10.52
a) Because $\mathcal{E}(X) = 0$, we have

$$\mathrm{Var}(X) = \mathcal{E}\big(X^2\big) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{2}{\pi}\int_{-1}^{1} x^2\sqrt{1-x^2}\,dx$$
$$= \frac{4}{\pi}\int_{0}^{1} x^2\sqrt{1-x^2}\,dx = \frac{4}{\pi}\cdot\frac{\pi}{16} = \frac{1}{4}.$$

b) By symmetry, Y has the same probability distribution as X and, hence, the same variance. Consequently, $\mathrm{Var}(Y) = \mathrm{Var}(X) = 1/4$.
c) As $\mathcal{E}(X) = 0$, we have

$$\mathrm{Cov}(X, Y) = \mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x, y)\,dx\,dy$$
$$= \frac{1}{\pi}\int_{-1}^{1}\left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y\,dy\right) x\,dx = \frac{1}{\pi}\int_{-1}^{1} 0\cdot x\,dx = 0.$$

d) No, as noted in the text, random variables can be uncorrelated without being independent.
e) No, X and Y are not independent random variables, as we discovered in Example 9.14 on page 524.

10.53 We have

$$\mathcal{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{1}{30}\int_0^{30} x^2\,dx = 300.$$

Therefore,

$$\operatorname{Var}(X) = \mathcal{E}(X^2) - \big(\mathcal{E}(X)\big)^2 = 300 - (15)^2 = 75.$$

10.54
a) We have

$$\mathcal{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

Recalling that $\mathcal{E}(X) = (a+b)/2$, we get

$$\begin{aligned}\operatorname{Var}(X) &= \mathcal{E}(X^2) - \big(\mathcal{E}(X)\big)^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{a^2+2ab+b^2}{4}\right)\\ &= \frac{1}{12}\left(4b^2+4ab+4a^2-3a^2-6ab-3b^2\right) = \frac{(b-a)^2}{12}.\end{aligned}$$

b) In Exercise 10.53, $X \sim \mathcal{U}(0, 30)$. Therefore, from part (a), $\operatorname{Var}(X) = (30-0)^2/12 = 75$.

10.55 From Example 10.13 on page 587, we know that the variance of an exponential random variable is the reciprocal of the square of its parameter. Hence, the standard deviation of the duration of a pause during a monologue is $\sqrt{1/(1.4)^2} = 1/1.4$, or about 0.714 seconds.

10.56 Let X denote the time, in minutes, that it takes a finisher to complete the race. By assumption, $X \sim \mathcal{N}(61, 81)$. From Example 10.12 on page 586, we know that the variance of a normal random variable is its second parameter. Thus, $\sigma_X = \sqrt{\operatorname{Var}(X)} = \sqrt{81} = 9$ minutes.

10.57 We have

$$\begin{aligned}\mathcal{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{1}{B(\alpha,\beta)}\int_0^1 x^2 \cdot x^{\alpha-1}(1-x)^{\beta-1}\,dx\\ &= \frac{1}{B(\alpha,\beta)}\int_0^1 x^{(\alpha+2)-1}(1-x)^{\beta-1}\,dx = \frac{1}{B(\alpha,\beta)}\cdot B(\alpha+2,\beta) = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.\end{aligned}$$

Recalling that $\mathcal{E}(X) = \alpha/(\alpha+\beta)$, we get

$$\operatorname{Var}(X) = \mathcal{E}(X^2) - \big(\mathcal{E}(X)\big)^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

10.58 Let X denote the proportion of these manufactured items that requires service during the first 5 years of use. By assumption, X has the beta distribution with parameters $\alpha = 2$ and $\beta = 3$. Applying Exercise 10.57 gives

$$\sigma_X = \sqrt{\operatorname{Var}(X)} = \sqrt{\frac{2\cdot 3}{(2+3)^2(2+3+1)}} = \sqrt{0.04} = 0.2.$$

10.59

a) From Exercise 10.17, a PDF of $T = T_1 + T_2$ is

$$f_T(t) = \begin{cases} t/34, & \text{if } 0 < t < 6; \\ (12 - t)/34, & \text{if } 6 < t < 10; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\mathcal{E}(T^2) = \int_{-\infty}^{\infty} t^2 f_T(t)\,dt = \frac{1}{34}\int_0^6 t^3\,dt + \frac{1}{34}\int_6^{10} t^2(12 - t)\,dt = \frac{1}{34}(324 + 960) = \frac{642}{17.}$$

Recalling from the solution to Exercise 10.17 that $\mathcal{E}(T) = 292/51$, we get

$$\sigma_T = \sqrt{\operatorname{Var}(T)} = \sqrt{\frac{642}{17} - \left(\frac{292}{51}\right)^2} = 2.23.$$

b) Applying the FEF with $g(t_1, t_2) = t_1^2$ yields

$$\begin{aligned} \mathcal{E}(T_1^2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_1^2 f_{T_1,T_2}(t_1, t_2)\,dt_1\,dt_2 \\ &= \frac{1}{34}\int_0^4 t_1^2\left(\int_0^6 1\,dt_2\right)dt_1 + \frac{1}{34}\int_4^6 t_1^2\left(\int_0^{10-t_1} 1\,dt_2\right)dt_1 \\ &= \frac{6}{34}\int_0^4 t_1^2\,dt_1 + \frac{1}{34}\int_4^6 t_1^2(10 - t_1)\,dt_1 = \frac{64}{17} + \frac{370}{51} = \frac{562}{51}. \end{aligned}$$

Recalling from the solution to Exercise 10.26 that $\mathcal{E}(T_1) = 146/51$, we get

$$\operatorname{Var}(T_1) = \frac{562}{51} - \left(\frac{146}{51}\right)^2 = \frac{7346}{(51)^2}.$$

By symmetry, we have $\operatorname{Var}(T_2) = \operatorname{Var}(T_1)$. We also need to compute $\operatorname{Cov}(T_1, T_2)$. Applying the FEF with $g(t_1, t_2) = t_1 t_2$ yields

$$\begin{aligned} \mathcal{E}(T_1 T_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_1 t_2 f_{T_1,T_2}(t_1, t_2)\,dt_1\,dt_2 \\ &= \frac{1}{34}\int_0^4 t_1\left(\int_0^6 t_2\,dt_2\right)dt_1 + \frac{1}{34}\int_4^6 t_1\left(\int_0^{10-t_1} t_2\,dt_2\right)dt_1 \\ &= \frac{18}{34}\int_0^4 t_1\,dt_1 + \frac{1}{68}\int_4^6 t_1(10 - t_1)^2\,dt_1 = \frac{72}{17} + \frac{185}{51} = \frac{401}{51}. \end{aligned}$$

Hence,

$$\operatorname{Cov}(T_1, T_2) = \mathcal{E}(T_1 T_2) - \mathcal{E}(T_1)\,\mathcal{E}(T_2) = \frac{401}{51} - \left(\frac{146}{51}\right)^2 = -\frac{865}{(51)^2}.$$

Consequently,

$$\operatorname{Var}(T_1 + T_2) = \operatorname{Var}(T_1) + \operatorname{Var}(T_2) + 2\operatorname{Cov}(T_1, T_2) = 2 \cdot \frac{7346}{(51)^2} - 2 \cdot \frac{865}{(51)^2} = \frac{12{,}962}{(51)^2}.$$

Therefore,

$$\sigma_{T_1+T_2} = \sqrt{\operatorname{Var}(T_1 + T_2)} = \sqrt{\frac{12{,}962}{(51)^2}} = 2.23.$$

10.60 In the solution to Exercise 10.18(b), we discovered that $\mathcal{E}(Y) = 1/4$ and that $f_Y(y) = -\ln y$ if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Therefore,

$$\mathcal{E}(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy = -\int_0^1 y^2 \ln y\,dy = \frac{y^3}{9}(1 - 3\ln y)\Big]_0^1 = \frac{1}{9}.$$

Hence,

$$\operatorname{Var}(Y) = \mathcal{E}(Y^2) - \big(\mathcal{E}(Y)\big)^2 = \frac{1}{9} - \left(\frac{1}{4}\right)^2 = \frac{7}{144}.$$

10.61

a) Clearly, $X + Y = 1$. Hence, by Equation (10.27) on page 586,

$$\operatorname{Var}(Y) = \operatorname{Var}(1 - X) = \operatorname{Var}\big(1 + (-1)X\big) = (-1)^2 \operatorname{Var}(X) = \operatorname{Var}(X).$$

b) From properties of variance and part (a), we get

$$0 = \operatorname{Var}(1) = \operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y) = 2\operatorname{Var}(X) + 2\operatorname{Cov}(X, Y).$$

Hence, $\operatorname{Cov}(X, Y) = -\operatorname{Var}(X)$.

c) From properties of covariance,

$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(X, 1 - X) = \operatorname{Cov}(X, 1) - \operatorname{Cov}(X, X) = 0 - \operatorname{Var}(X) = -\operatorname{Var}(X).$$

d) From parts (a) and (b),

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{-\operatorname{Var}(X)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(X)}} = -\frac{\operatorname{Var}(X)}{\operatorname{Var}(X)} = -1.$$

e) As $Y = 1 - X = 1 + (-1)X$, we conclude from Proposition 7.16(d) on page 372 that $\rho(X, Y) = -1$.

10.62 Let X denote the lifetime of the machine and let Y denote the age of the machine at the time of replacement. We want to determine the standard deviation of the random variable Y. Observe that

$$Y = \begin{cases} X, & \text{if } X \le 4; \\ 4, & \text{if } X > 4. \end{cases}$$

Now, set

$$g(x) = \begin{cases} x, & \text{if } x \le 4; \\ 4, & \text{if } x > 4. \end{cases}$$

Then $Y = g(X)$ and, hence, from the FEF,

$$\begin{aligned} \mathcal{E}(Y^r) &= \mathcal{E}\Big(\big(g(X)\big)^r\Big) = \int_{-\infty}^{\infty} \big(g(x)\big)^r f_X(x)\,dx \\ &= \int_0^4 x^r f_X(x)\,dx + \int_4^{\infty} 4^r f_X(x)\,dx = \int_0^4 x^r f_X(x)\,dx + 4^r P(X > 4). \end{aligned}$$

a) In this case, we assume that $X \sim \mathcal{U}(0, 6)$. Then

$$f_X(x) = \begin{cases} 1/6, & \text{if } 0 < x < 6; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\mathcal{E}(Y) = \frac{1}{6}\int_0^4 x\,dx + 4 \cdot \frac{2}{6} = \frac{8}{3}$$

and

$$\mathcal{E}\left(Y^2\right) = \frac{1}{6}\int_0^4 x^2\,dx + 4^2 \cdot \frac{2}{6} = \frac{80}{9}.$$

Hence,

$$\sigma_Y = \sqrt{\operatorname{Var}(Y)} = \sqrt{\mathcal{E}\left(Y^2\right) - \left(\mathcal{E}(Y)\right)^2} = \sqrt{\frac{80}{9} - \left(\frac{8}{3}\right)^2} = 1.333.$$

b) In this case, we assume that $X \sim \mathcal{T}(0, 6)$. Then

$$f_X(x) = \begin{cases} x/9, & \text{if } 0 < x < 3; \\ (6-x)/9, & \text{if } 3 < x < 6; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\mathcal{E}(Y) = \int_0^3 x \cdot \frac{x}{9}\,dx + \int_3^4 x \cdot \frac{6-x}{9}\,dx + 4P(X > 4) = 1 + \frac{26}{27} + \frac{8}{9} = \frac{77}{27}.$$

and

$$\mathcal{E}\left(Y^2\right) = \int_0^3 x^2 \cdot \frac{x}{9}\,dx + \int_3^4 x^2 \cdot \frac{6-x}{9}\,dx + 16P(X > 4) = \frac{9}{4} + \frac{121}{36} + \frac{32}{9} = \frac{165}{18}.$$

Hence,

$$\sigma_Y = \sqrt{\operatorname{Var}(Y)} = \sqrt{\mathcal{E}\left(Y^2\right) - \left(\mathcal{E}(Y)\right)^2} = \sqrt{\frac{165}{18} - \left(\frac{77}{27}\right)^2} = 1.017.$$

c) In this case, $X \sim \mathcal{E}(1/3)$. Then

$$f_X(x) = \begin{cases} (1/3)e^{-x/3}, & \text{if } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\mathcal{E}(Y) = \frac{1}{3}\int_0^4 xe^{-x/3}\,dx + 4P(X > 4) = 1.1548 + 1.0544 = 2.2092$$

and

$$\mathcal{E}\left(Y^2\right) = \frac{1}{3}\int_0^4 x^2e^{-x/3}\,dx + 16P(X > 4) = 2.7114 + 4.2176 = 6.9290.$$

Hence,

$$\sigma_Y = \sqrt{\operatorname{Var}(Y)} = \sqrt{\mathcal{E}\left(Y^2\right) - \left(\mathcal{E}(Y)\right)^2} = \sqrt{6.9290 - (2.2092)^2} = 1.431.$$

10.63
a) We begin by noting that

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = c\int_0^1 x\left(\int_0^1 1\,dy\right)dx = c\int_0^1 x\,dx = \frac{c}{2}.$$

So, we see that $c = 2$. From the FEF,

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x, y)\,dx\,dy = 2\int_0^1 x^2\left(\int_0^1 y\,dy\right)dx = \int_0^1 x^2\,dx = \frac{1}{3},$$

$$\mathcal{E}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{X,Y}(x, y)\,dx\,dy = 2\int_0^1 x^2\left(\int_0^1 1\,dy\right)dx = 2\int_0^1 x^2\,dx = \frac{2}{3},$$

and

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf_{X,Y}(x, y)\,dx\,dy = 2\int_0^1 x\left(\int_0^1 y\,dy\right)dx = \int_0^1 x\,dx = \frac{1}{2}.$$

Hence,

$$\operatorname{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \frac{1}{3} - \frac{2}{3}\cdot\frac{1}{2} = 0.$$

b) We can write

$$f_{X,Y}(x, y) = \big(2xI_{(0,1)}(x)\big)\big(I_{(0,1)}(y)\big), \qquad (x, y) \in \mathcal{R}^2.$$

Therefore, from Exercise 9.95, X and Y are independent random variables. Consequently, from Proposition 7.14 on page 368, we have $\operatorname{Cov}(X, Y) = 0$.

10.64 Let X denote the repair costs and let Y denote the insurance payment in the event that the automobile is damaged. We know that $X \sim \mathcal{U}(0, 1500)$. Note that

$$Y = \begin{cases} 0, & \text{if } X \le 250; \\ X - 250, & \text{if } X > 250. \end{cases}$$

Set

$$g(x) = \begin{cases} 0, & \text{if } x \le 250; \\ x - 250, & \text{if } x > 250. \end{cases}$$

Then $Y = g(X)$. By the FEF,

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx = \frac{1}{1500}\int_{250}^{1500}(x - 250)\,dx = 520.833,$$

and

$$\mathcal{E}\big(Y^2\big) = \int_{-\infty}^{\infty}\big(g(x)\big)^2 f_X(x)\,dx = \frac{1}{1500}\int_{250}^{1500}(x - 250)^2\,dx = 434027.778.$$

Hence,

$$\sigma_Y = \sqrt{\operatorname{Var}(Y)} = \sqrt{434027.778 - (520.833)^2} = \$403.44.$$

10.65 By assumption, $\mu_X = \mu_Y = 100$ and $(X - 100)(Y - 100) < 0$. Therefore, by the monotonicity property of expected value,

$$\operatorname{Cov}(X, Y) = \mathcal{E}\big((X - \mu_X)(Y - \mu_Y)\big) = \mathcal{E}\big((X - 100)(Y - 100)\big) < 0.$$

Now applying Equation (10.32) on page 589, we get

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y) < \operatorname{Var}(X) + \operatorname{Var}(Y).$$

Thus, (b) is the correct relationship.

10.66 No, the converse is not true. For instance, consider the random variables X and Y in Exercise 10.37. Because $\mathcal{E}(XY) = \mathcal{E}(X)\,\mathcal{E}(Y)$, we have

$$\operatorname{Cov}(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = 0.$$

However, as noted in the solution of that exercise, these two random variables are not independent.

10.67

a) In Example 10.11, we found that

$$\operatorname{Var}(X) = \operatorname{Var}(Y) = \frac{n}{(n+2)(n+1)^2}$$

and, in Example 10.14,

$$\operatorname{Cov}(X,Y) = \frac{1}{(n+2)(n+1)^2}.$$

Hence,

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}} = \frac{\dfrac{1}{(n+2)(n+1)^2}}{\sqrt{\dfrac{n}{(n+2)(n+1)^2}\cdot\dfrac{n}{(n+2)(n+1)^2}}} = \frac{1}{n}.$$

b) Answers will vary, but following is one possible explanation. As $n \to \infty$, we would expect that $\min\{X_1,\dots,X_n\} \to 0$ and $\max\{X_1,\dots,X_n\} \to 1$, and the correlation between two constant random variables equals 0.

c) Recalling that $\operatorname{Var}(Y) = \operatorname{Var}(X)$ and using properties of covariance, we find that

$$\begin{aligned}
\operatorname{Cov}(R,M) &= \operatorname{Cov}\big(Y-X,(X+Y)/2\big) = \frac{1}{2}\operatorname{Cov}(Y-X,X+Y) \\
&= \frac{1}{2}\big(\operatorname{Cov}(Y,X) + \operatorname{Cov}(Y,Y) - \operatorname{Cov}(X,X) - \operatorname{Cov}(X,Y)\big) \\
&= \frac{1}{2}\big(\operatorname{Cov}(X,Y) + \operatorname{Var}(Y) - \operatorname{Var}(X) - \operatorname{Cov}(X,Y)\big) = 0.
\end{aligned}$$

Hence, $\rho(R,M) = 0$; the range and midrange are uncorrelated.

d) No. Although independent random variables are uncorrelated, it is not necessarily the case that uncorrelated random variables are independent. See, for instance, Exercise 10.66.

e) No, R and M are not independent random variables. We can see this fact in several ways. One way is to note that the range of M is the interval $(0,1)$. However, given that, say, $r = 1/2$, the range of M is the interval $(1/4, 3/4)$.

10.68

a) From the linearity property of expected value,

$$\begin{aligned}
\mathcal{E}(\bar{X}_n) &= \mathcal{E}\left(\frac{1}{n}(X_1+\cdots+X_n)\right) = \frac{1}{n}\big(\mathcal{E}(X_1+\cdots+X_n)\big) \\
&= \frac{1}{n}\big(\mathcal{E}(X_1)+\cdots+\mathcal{E}(X_n)\big) = \frac{1}{n}(\mu+\cdots+\mu) = \frac{1}{n}(n\mu) = \mu.
\end{aligned}$$

b) Using the independence of $X_1, \dots, X_n$ and Equations (10.27) and (10.34) on pages 586 and 590, respectively, we get

$$\begin{aligned}\operatorname{Var}(\bar{X}_n) &= \operatorname{Var}\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{1}{n^2}\left(\operatorname{Var}(X_1 + \cdots + X_n)\right) \\ &= \frac{1}{n^2}\left(\operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)\right) = \frac{1}{n^2}\left(\sigma^2 + \cdots + \sigma^2\right) = \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}.\end{aligned}$$

c) Because $X_1, \dots, X_n$ are independent random variables each having variance σ^2, we know that $\operatorname{Cov}(X_k, X_j) = \sigma^2$ if $k = j$, and $\operatorname{Cov}(X_k, X_j) = 0$ if $k \neq j$. Applying properties of covariance and referring to part (b), we get

$$\begin{aligned}\operatorname{Cov}(\bar{X}_n, X_j - \bar{X}_n) &= \operatorname{Cov}(\bar{X}_n, X_j) - \operatorname{Cov}(\bar{X}_n, \bar{X}_n) \\ &= \operatorname{Cov}\left(\frac{1}{n}\sum_{k=1}^{n} X_k, X_j\right) - \operatorname{Var}(\bar{X}_n) \\ &= \frac{1}{n}\sum_{k=1}^{n}\operatorname{Cov}(X_k, X_j) - \operatorname{Var}(\bar{X}_n) = \frac{1}{n}\sigma^2 - \frac{\sigma^2}{n} = 0.\end{aligned}$$

d) The mean of the sample mean equals the distribution mean; the variance of the sample mean equals the ratio of the distribution variance to the sample size; and the sample mean is uncorrelated with each deviation from the sample mean.

10.69 For $n \in \mathcal{N}$,

$$\begin{aligned}\mathcal{E}(X^n) &= \int_{-\infty}^{\infty} x^n f_X(x)\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^n \cdot x^{\alpha-1}e^{-\lambda x}\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{(\alpha+n)-1}e^{-\lambda x}\,dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+n)}{\lambda^{\alpha+n}} = \frac{(\alpha+n-1)(\alpha+n-2)\cdots\alpha\Gamma(\alpha)}{\lambda^n\Gamma(\alpha)} = \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)}{\lambda^n}.\end{aligned}$$

10.70

a) From the FEF and upon making the substitution $z = (x-\mu)/\sigma$, we find that, for $n \in \mathcal{N}$,

$$\mathcal{E}\left((X-\mu)^{2n+1}\right) = \int_{-\infty}^{\infty}(x-\mu)^{2n+1}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{\sigma^{2n+1}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^{2n+1}e^{-z^2/2}\,dz = 0,$$

where the last equation follows from the fact that $z^{2n+1}e^{-z^2/2}$ is an odd function.

b) From the FEF and upon first making the substitution $z = (x-\mu)/\sigma$ and then making the substitution $u = z^2/2$, we find that, for $n \in \mathcal{N}$,

$$\begin{aligned}\mathcal{E}\left((X-\mu)^{2n}\right) &= \int_{-\infty}^{\infty}(x-\mu)^{2n}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{\sigma^{2n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^{2n}e^{-z^2/2}\,dz \\ &= \frac{2\sigma^{2n}}{\sqrt{2\pi}}\int_0^\infty z^{2n}e^{-z^2/2}\,dz = \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\int_0^\infty u^{n-1/2}e^{-u}\,du \\ &= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\int_0^\infty u^{(n+1/2)-1}e^{-u}\,du = \frac{2^n\sigma^{2n}}{\sqrt{\pi}} \cdot \Gamma(n+1/2) = \frac{2^n\sigma^{2n}}{\sqrt{\pi}} \cdot \frac{(2n)!}{n!\,2^{2n}}\sqrt{\pi} \\ &= \frac{(2n)!}{n!\,2^n}\sigma^{2n},\end{aligned}$$

where the penultimate equality is due to Equation (8.46).

10.71

a) Applying properties of covariance, we obtain

$$\begin{aligned}\operatorname{Cov}(aX+bY, aX-bY) &= a^2 \operatorname{Cov}(X,X) - ab \operatorname{Cov}(X,Y) + ba \operatorname{Cov}(Y,X) - b^2 \operatorname{Cov}(Y,Y) \\ &= a^2 \operatorname{Var}(X) - ab \operatorname{Cov}(X,Y) + ab \operatorname{Cov}(X,Y) - b^2 \operatorname{Var}(Y) \\ &= a^2 \operatorname{Var}(X) - b^2 \operatorname{Var}(Y)\,.\end{aligned}$$

b) If $\operatorname{Var}(Y) = \operatorname{Var}(X)$, then, from part (a),

$$\operatorname{Cov}(X+Y, X-Y) = 1^2 \operatorname{Var}(X) - 1^2 \operatorname{Var}(Y) = \operatorname{Var}(X) - \operatorname{Var}(X) = 0.$$

Thus, X and Y are uncorrelated random variables.

Theory Exercises

10.72

a) Because X and $-X$ have the same probability distribution, they have the same moments. Hence, if X has a finite odd moment of order r, then

$$\mathcal{E}(X^r) = \mathcal{E}\big((-X)^r\big) = \mathcal{E}\big((-1)^r X^r\big) = \mathcal{E}\big(-1X^r\big) = -\mathcal{E}(X^r)\,,$$

which implies that $\mathcal{E}(X^r) = 0$.

b) Suppose that X is symmetric about c. From Exercise 10.46(b), we know that $\mu_X = c$. And because $X - c$ is symmetric, part (a) shows that

$$\mathcal{E}\big((X-\mu_X)^r\big) = \mathcal{E}\big((X-c)^r\big) = 0,$$

for all odd positive integers r for which X has a finite moment of that order. In other words, all odd central moments of X equal 0.

10.73 We have

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^n f_X(x)\,dx = \int_0^{\infty} x^n f_X(x)\,dx &= \int_0^{\infty}\left(\int_0^x ny^{n-1}\,dy\right) f_X(x)\,dx \\ &= n\int_0^{\infty}\left(\int_y^{\infty} f_X(x)\,dx\right) y^{n-1}\,dy = n\int_0^{\infty} y^{n-1} P(X>y)\,dy.\end{aligned}$$

It now follows from the FEF that X has a finite nth moment if and only if $\int_0^\infty x^{n-1} P(X>x)\,dx < \infty$, and, in that case, the FEF and the previous display show that

$$\mathcal{E}(X^n) = \int_0^{\infty} x^n f_X(x)\,dx = n\int_0^{\infty} x^{n-1} P(X>x)\,dx.$$

10.74

a) Let σ^2 denote the common variance of the X_js. Applying Chebyshev's inequality and Exercise 10.68, we get

$$\begin{aligned}P\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| < \epsilon\right) &= P\left(\left|\bar{X}_n - \mu_{\bar{X}_n}\right| < \epsilon\right) = 1 - P\left(\left|\bar{X}_n - \mu_{\bar{X}_n}\right| \geq \epsilon\right) \\ &\geq 1 - \frac{\operatorname{Var}(\bar{X}_n)}{\epsilon^2} = 1 - \frac{\sigma^2}{n\epsilon^2}.\end{aligned}$$

It now follows that

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|<\epsilon\right)=1.$$

b) In words, the weak law of large numbers states that, when n is large, the average value of $X_1, \ldots, X_n$ is likely to be near μ.
c) The weak law of large numbers provides a mathematically precise version of the long-run-average interpretation of expected value.

Advanced Exercises

10.75
a) From the linearity property of expected value

$$\mathcal{E}(Y)=\mathcal{E}\left(\sum_{j=1}^{n} c_j X_j\right)=\sum_{j=1}^{n} c_j\mathcal{E}(X_j)=\left(\sum_{j=1}^{n} c_j\right)\mu.$$

Hence, Y is an unbiased estimator of μ if and only if $\sum_{j=1}^{n} c_j = 1$.
b) By independence and Equation (10.27) on page 586,

$$\operatorname{Var}(Y)=\operatorname{Var}\left(\sum_{j=1}^{n} c_j X_j\right)=\sum_{j=1}^{n} c_j^2\operatorname{Var}(X_j)=\sum_{j=1}^{n} c_j^2\sigma_j^2.$$

We want to choose the c_js that minimize $\operatorname{Var}(Y)$ and also provide an unbiased estimator. To do so, we use the method of Lagrange multipliers. Here the objective and constraint functions are

$$f(c_1,\ldots,c_n)=\sum_{j=1}^{n} c_j^2\sigma_j^2 \qquad \text{and} \qquad g(c_1,\ldots,c_n)=\sum_{j=1}^{n} c_j - 1,$$

respectively. The gradient equation yields $2c_j\sigma_j^2=\lambda$, or $c_j=\lambda/2\sigma_j^2$, for $1\le j\le n$. Substituting these values of c_j into the equation $g(c_1,\ldots,c_n)=0$ gives $\lambda = 2\big/\sum_{k=1}^{n}\left(1/\sigma_k^2\right)$. Therefore,

$$c_j=\frac{2\big/\sum_{k=1}^{n}\left(1/\sigma_k^2\right)}{2\sigma_j^2}=\left(\sum_{k=1}^{n}\frac{1}{\sigma_k^2}\right)^{-1}\cdot\frac{1}{\sigma_j^2}, \qquad 1\le j\le n.$$

c) Unbiasedness is a desirable property because it means that, on average, the estimator will give the actual value of the unknown parameter; minimum variance is a desirable property because it means that, on average, the estimator will give the closest estimate of (least error for) the value of the unknown parameter.
d) Let σ^2 denote the common variance of the X_js. From part (b), the best linear unbiased estimator of μ is obtained by choosing the c_js as follows:

$$c_j=\left(\sum_{k=1}^{n}\frac{1}{\sigma_k^2}\right)^{-1}\cdot\frac{1}{\sigma_j^2}=\left(\sum_{k=1}^{n}\frac{1}{\sigma^2}\right)^{-1}\cdot\frac{1}{\sigma^2}=\left(\frac{n}{\sigma^2}\right)^{-1}\cdot\frac{1}{\sigma^2}=\frac{1}{n}, \qquad 1\le j\le n.$$

Therefore, in this case, the best linear unbiased estimator of μ is

$$Y=\sum_{j=1}^{n} c_j X_j=\frac{1}{n}\sum_{j=1}^{n} X_j=\bar{X}_n.$$

10.76

a) We note that T has finite variance if and only if $\int_{-\infty}^{\infty} t^2\bigl(1+t^2/\nu\bigr)^{-(\nu+1)/2}\,dt < \infty$. Using symmetry and making the substitution $u = 1 + t^2/\nu$, we find that

$$\int_{-\infty}^{\infty} \frac{t^2}{\bigl(1+t^2/\nu\bigr)^{(\nu+1)/2}}\,dt = 2\int_{0}^{\infty} \frac{t^2}{\bigl(1+t^2/\nu\bigr)^{(\nu+1)/2}}\,dt = \nu^{3/2}\int_{1}^{\infty} \frac{\sqrt{u-1}}{u^{(\nu+1)/2}}\,du.$$

Noting that the integrand on the right of the preceding display is asymptotic to $u^{-\nu/2}$ as $u \to \infty$ and recalling from calculus that $\int_1^\infty u^{-p}\,du$ converges (i.e., is finite) if and only if $p > 1$, we see that T has finite variance if and only if $\nu/2 > 1$, that is, if and only if $\nu > 2$.

b) Using integration by parts, we find that

$$\int_{0}^{\infty} \frac{t^2}{\bigl(1+t^2/\nu\bigr)^{(\nu+1)/2}}\,dt = \frac{\nu}{\nu-1}\int_{0}^{\infty} \frac{dt}{\bigl(1+t^2/\nu\bigr)^{(\nu-1)/2}}.$$

Making the substitution $u = \sqrt{(\nu-2)/\nu}\,t$ and using the fact that a t PDF with $\nu - 2$ degrees of freedom integrates to 1, we get

$$\begin{aligned}
\int_{0}^{\infty} \frac{dt}{\bigl(1+t^2/\nu\bigr)^{(\nu-1)/2}} &= \sqrt{\frac{\nu}{\nu-2}}\int_{0}^{\infty} \frac{du}{\bigl(1+u^2/(\nu-2)\bigr)^{((\nu-2)+1)/2}} \\
&= \frac{1}{2}\sqrt{\frac{\nu}{\nu-2}}\int_{-\infty}^{\infty} \frac{du}{\bigl(1+u^2/(\nu-2)\bigr)^{((\nu-2)+1)/2}} \\
&= \frac{1}{2}\sqrt{\frac{\nu}{\nu-2}}\cdot\frac{\sqrt{(\nu-2)\pi}\,\Gamma\bigl((\nu-2)/2\bigr)}{\Gamma\bigl((\nu-1)/2\bigr)} \\
&= \frac{1}{2}\sqrt{\nu\pi}\,\frac{\Gamma\bigl((\nu-2)/2\bigr)}{\Gamma\bigl((\nu-1)/2\bigr)}.
\end{aligned}$$

Recalling from Exercise 10.22 that $\mathcal{E}(T) = 0$, we conclude that

$$\begin{aligned}
\operatorname{Var}(T) = \mathcal{E}\bigl(T^2\bigr) &= \frac{\Gamma\bigl((\nu+1)/2\bigr)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\int_{-\infty}^{\infty} \frac{t^2}{\bigl(1+t^2/\nu\bigr)^{(\nu+1)/2}}\,dt \\
&= 2\frac{\Gamma\bigl((\nu+1)/2\bigr)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\int_{0}^{\infty} \frac{t^2}{\bigl(1+t^2/\nu\bigr)^{(\nu+1)/2}}\,dt \\
&= 2\frac{\Gamma\bigl((\nu+1)/2\bigr)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\cdot\frac{\nu}{\nu-1}\cdot\frac{1}{2}\sqrt{\nu\pi}\,\frac{\Gamma\bigl((\nu-2)/2\bigr)}{\Gamma\bigl((\nu-1)/2\bigr)} \\
&= 2\frac{\bigl((\nu-1)/2\bigr)\Gamma\bigl((\nu-1)/2\bigr)}{\bigl((\nu-2)/2\bigr)\Gamma\bigl((\nu-2)/2\bigr)}\cdot\frac{\nu}{\nu-1}\cdot\frac{1}{2}\frac{\Gamma\bigl((\nu-2)/2\bigr)}{\Gamma\bigl((\nu-1)/2\bigr)} \\
&= \frac{\nu}{\nu-2}.
\end{aligned}$$

10.4 Conditional Expectation

Basic Exercises

10.77 We know that $Y_{|X=x} \sim \mathcal{U}(0, x)$ for $0 < x < 1$.
a) Recall that for a $\mathcal{U}(a, b)$ distribution, the mean and variance are $(a+b)/2$ and $(b-a)^2/12$, respectively. Hence, for $0 < x < 1$,

$$\mathcal{E}(Y \mid X = x) = x/2 \qquad \text{and} \qquad \operatorname{Var}(Y \mid X = x) = x^2/12.$$

b) Referring to part (a) and applying the law of total expectation, we get

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\left(\frac{X}{2}\right) = \frac{1}{2}\mathcal{E}(X) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Referring to part (a) and applying the law of total variance, we get

$$\begin{aligned}\operatorname{Var}(Y) &= \mathcal{E}\big(\operatorname{Var}(Y \mid X)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\left(\frac{X^2}{12}\right) + \operatorname{Var}\left(\frac{X}{2}\right) \\ &= \frac{1}{12}\mathcal{E}(X^2) + \frac{1}{4}\operatorname{Var}(X) = \frac{1}{12}\Big(\operatorname{Var}(X) + \big(\mathcal{E}(X)\big)^2\Big) + \frac{1}{4}\operatorname{Var}(X) \\ &= \frac{1}{3}\operatorname{Var}(X) + \frac{1}{12}\big(\mathcal{E}(X)\big)^2 = \frac{1}{3} \cdot \frac{1}{12} + \frac{1}{12} \cdot \frac{1}{4} = \frac{7}{144}.\end{aligned}$$

10.78 For $0 < x < 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_x^{x+1} 2x\, dy = 2x.$$

Therefore,

$$f_{Y \mid X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{2x}{2x} = 1$$

if $x < y < x + 1$, and $f_{Y \mid X}(y \mid x) = 0$ otherwise. So, we see that $Y_{|X=x} \sim \mathcal{U}(x, x+1)$ for $0 < x < 1$. Consequently,

$$\mathcal{E}(Y \mid X = x) = \frac{x + (x+1)}{2} = x + \frac{1}{2}$$

if $0 < x < 1$, and $\mathcal{E}(Y \mid X = x) = 0$ otherwise. Also,

$$\operatorname{Var}(Y \mid X = x) = \frac{\big((x+1) - x\big)^2}{12} = \frac{1}{12}$$

if $0 < x < 1$, and $\operatorname{Var}(Y \mid X = x) = 0$ otherwise.

10.79 We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_{2y}^{2} 2xy\,dx = 4y(1 - y^2).$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Therefore, for $0 < y < 1$,

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{2xy}{4y(1 - y^2)} = \frac{x}{2(1 - y^2)}$$

if $2y < x < 2$, and $f_{X\,|\,Y}(x\,|\,y) = 0$ otherwise. Thus,

$$\mathcal{E}(X\,|\,Y = y) = \int_{-\infty}^{\infty} x f_{X\,|\,Y}(x\,|\,y)\,dx = \frac{1}{2(1 - y^2)} \int_{2y}^{2} x^2\,dx = \frac{4(1 - y^3)}{3(1 - y^2)}$$

if $0 < y < 1$, and $\mathcal{E}(X\,|\,Y = y) = 0$ otherwise. Applying the law of total expectation, we now find that

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} \mathcal{E}(X\,|\,Y = y)\, f_Y(y)\,dy = \int_0^1 \frac{4(1 - y^3)}{3(1 - y^2)} \cdot 4y(1 - y^2)\,dy = \frac{16}{3} \int_0^1 \left(y - y^4\right) dy = \frac{8}{5}.$$

10.80

a) We have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = \int_0^{\infty} \lambda\mu e^{-(\lambda x + \mu y)}\,dy = \lambda e^{-\lambda x} \int_0^{\infty} \mu e^{-\mu y}\,dy = \lambda e^{-\lambda x}$$

if $x > 0$, and $f_X(x) = 0$ otherwise. Thus, for $x > 0$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\lambda\mu e^{-(\lambda x + \mu y)}}{\lambda e^{-\lambda x}} = \mu e^{-\mu y}$$

if $y > 0$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. We observe that $Y_{|X=x} \sim \mathcal{E}(\mu)$ for $x > 0$. It now follows that

$$\mathcal{E}(Y\,|\,X = x) = \frac{1}{\mu} \qquad \text{and} \qquad \operatorname{Var}(Y\,|\,X = x) = \frac{1}{\mu^2}$$

if $x > 0$, and $\mathcal{E}(Y\,|\,X = x) = 0$ and $\operatorname{Var}(Y\,|\,X = x) = 0$ otherwise.

b) They don't depend on x because X and Y are independent random variables.

c) Referring to part (a) yields

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y\,|\,X)\big) = \mathcal{E}\left(\frac{1}{\mu}\right) = \frac{1}{\mu}.$$

d) Referring to part (a) yields

$$\operatorname{Var}(Y) = \mathcal{E}\big(\operatorname{Var}(Y\,|\,X)\big) + \operatorname{Var}\big(\mathcal{E}(Y\,|\,X)\big) = \mathcal{E}\left(\frac{1}{\mu^2}\right) + \operatorname{Var}\left(\frac{1}{\mu}\right) = \frac{1}{\mu^2} + 0 = \frac{1}{\mu^2}.$$

10.81

a) Because X and Y are independent, knowing the value of X doesn't affect the probability distribution of Y. Hence, $\mathcal{E}(Y\,|\,X = x) = \mathcal{E}(Y)$ for each possible value x of X.

b) Because X and Y are independent, f_Y is a PDF of $Y_{|X=x}$ for each possible value x of X. Hence,

$$\mathcal{E}(Y\,|\,X = x) = \int_{-\infty}^{\infty} y f_{Y\,|\,X}(y\,|\,x)\,dy = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \mathcal{E}(Y)\,.$$

c) Because X and Y are independent, knowing the value of X doesn't affect the probability distribution of Y. Hence, $\operatorname{Var}(Y\,|\,X = x) = \operatorname{Var}(Y)$ for each possible value x of X.

d) Because X and Y are independent, f_Y is a PDF of $Y_{|X=x}$ for each possible value x of X. Therefore, from Definition 10.3, part (b), and the conditional form of the FEF,

$$\begin{aligned}\operatorname{Var}(Y \mid X = x) &= \mathcal{E}\Big(\big(Y - \mathcal{E}(Y \mid X = x)\big)^2 \mid X = x\Big) = \mathcal{E}\Big(\big(Y - \mathcal{E}(Y)\big)^2 \mid X = x\Big) \\ &= \int_{-\infty}^{\infty} \big(y - \mathcal{E}(Y)\big)^2 f_{Y \mid X}(y \mid x)\, dy = \int_{-\infty}^{\infty} \big(y - \mathcal{E}(Y)\big)^2 f_Y(y)\, dy = \operatorname{Var}(Y).\end{aligned}$$

e) Because the expected value of a constant is that constant, we have, in view of part (b), that

$$\mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\big(\mathcal{E}(Y)\big) = \mathcal{E}(Y).$$

f) Because the expected value of a constant is that constant and the variance of a constant equals 0, we have, in view of parts (b) and (d), that

$$\mathcal{E}\big(\operatorname{Var}(Y \mid X)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\big(\operatorname{Var}(Y)\big) + \operatorname{Var}\big(\mathcal{E}(Y)\big) = \operatorname{Var}(Y) + 0 = \operatorname{Var}(Y).$$

10.82
a) By assumption $Y_{|X=x} \sim \mathcal{N}\big(\rho x, 1 - \rho^2\big)$. Thus, $\mathcal{E}(Y|X = x) = \rho x$ and $\operatorname{Var}(Y \mid X = x) = 1 - \rho^2$. Consequently, $\mathcal{E}(Y|X) = \rho X$ and $\operatorname{Var}(Y \mid X) = 1 - \rho^2$.
b) By assumption, $X \sim \mathcal{N}(0, 1)$ and, so, $\mathcal{E}(X) = 0$ and $\operatorname{Var}(X) = 1$. Consequently, by referring to part (a) and applying the laws of total expectation and total variance, we get

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}(\rho X) = \rho \mathcal{E}(X) = 0$$

and

$$\begin{aligned}\operatorname{Var}(Y) &= \mathcal{E}\big(\operatorname{Var}(Y \mid X)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\big(1 - \rho^2\big) + \operatorname{Var}(\rho X) \\ &= \big(1 - \rho^2\big) + \rho^2 \operatorname{Var}(X) = \big(1 - \rho^2\big) + \rho^2 = 1.\end{aligned}$$

c) In Example 9.13(b), we found that $Y \sim \mathcal{N}(0, 1)$. Thus, if we used the PDF of Y to find its mean and variance, we would obtain $\mathcal{E}(Y) = 0$ and $\operatorname{Var}(Y) = 1$, which is what we got in part (b).

10.83 Starting from now, let X denote the amount of time, in hours, until the bus arrives and let Y denote the number of passengers at the bus stop when the bus arrives. By assumption, $X \sim \text{Beta}(2, 1)$ and $Y_{|X=x} \sim \mathcal{P}(\lambda x)$. Hence, $\mathcal{E}(Y \mid X = x) = \lambda x$ and $\operatorname{Var}(Y \mid X = x) = \lambda x$. Applying the law of total expectation and recalling that the mean and variance of a $\text{Beta}(\alpha, \beta)$ distribution are $\alpha/(\alpha + \beta)$ and $\alpha\beta/(\alpha + \beta)^2(\alpha + \beta + 1)$, respectively, we get

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}(\lambda X) = \lambda \mathcal{E}(X) = \frac{2}{3}\lambda.$$

Also, by the law of total variance,

$$\begin{aligned}\operatorname{Var}(Y) &= \mathcal{E}\big(\operatorname{Var}(Y \mid X)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}(\lambda X) + \operatorname{Var}(\lambda X) \\ &= \lambda \mathcal{E}(X) + \lambda^2 \operatorname{Var}(X) = \frac{2}{3}\lambda + \frac{1}{18}\lambda^2 = \frac{\lambda(\lambda + 12)}{18}.\end{aligned}$$

10.84
a) Because X and Y are independent, observing X provides no information about Y. Consequently, the best predictor of Y, based on observing X, is just $\mathcal{E}(Y)$.
b) By the prediction theorem, the best predictor of Y, based on observing X, is $\mathcal{E}(Y \mid X)$. However, because X and Y are independent, we have $\mathcal{E}(Y \mid X) = \mathcal{E}(Y)$, as shown in Exercise 10.81.

10.85
a) Observing X determines Y, namely, $Y = g(X)$. So, clearly, $g(X)$ is the best predictor of Y, based on observing X.
b) By the prediction theorem, the best possible predictor of Y, based on observing X, is $\mathcal{E}(Y \mid X)$. Using the assumption that $Y = g(X)$ and applying Equation (10.49) on page 602, with $k(x) = g(x)$ and $\ell(x, y) = 1$, we get

$$\mathcal{E}(Y \mid X) = \mathcal{E}\big(g(X) \mid X\big) = \mathcal{E}\big(g(X) \cdot 1 \mid X\big) = g(X)\mathcal{E}(1 \mid X) = g(X) \cdot 1 = g(X).$$

Hence, the best predictor of Y, based on observing X, is $g(X)$.

10.86 From the prediction theorem (Proposition 10.8 on page 602) and the law of total expectation (Proposition 10.6 on page 599), we find that the minimum mean square error is

$$\mathcal{E}\Big(\big(Y - \mathcal{E}(Y \mid X)\big)^2\Big) = \mathcal{E}\bigg(\mathcal{E}\Big(\big(Y - \mathcal{E}(Y \mid X)\big)^2 \mid X\Big)\bigg) = \mathcal{E}\big(\operatorname{Var}(Y \mid X)\big).$$

10.87 Let X denote the division point of the interval (0, 1) and let L denote the length of the subinterval that contains c. Note that $X \sim \mathcal{U}(0, 1)$.
a) We have

$$\mathcal{E}(L \mid X = x) = \begin{cases} 1 - x, & \text{if } 0 < x < c; \\ x, & \text{if } c \le x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, by the law of total expectation,

$$\mathcal{E}(L) = \int_{-\infty}^{\infty} \mathcal{E}(L \mid X = x)\, f_X(x)\, dx = \int_0^c (1 - x) \cdot 1\, dx + \int_c^1 x \cdot 1\, dx = 1/2 + c - c^2.$$

b) Referring to part (a), we want to maximize the function $f(c) = 1/2 + c - c^2$ for $0 < c < 1$. Differentiating f and setting the result equal to 0, we find that $1/2$ is the value of c that maximizes the expected length of the interval that contains c.

10.88 Let Y denote the number of defects per yard. We know that $Y_{|\Lambda=\lambda} \sim \mathcal{P}(\lambda)$. Therefore,

$$\mathcal{E}(Y \mid \Lambda = \lambda) = \lambda \qquad \text{and} \qquad \operatorname{Var}(Y \mid \Lambda = \lambda) = \lambda.$$

a) In this case, $\Lambda \sim \mathcal{U}(0, 3)$, so that

$$\mathcal{E}(\Lambda) = \frac{0 + 3}{2} = 1.5 \qquad \text{and} \qquad \operatorname{Var}(\Lambda) = \frac{(3 - 0)^2}{12} = 0.75.$$

Hence, by the laws of total expectation and total variance,

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid \Lambda)\big) = \mathcal{E}(\Lambda) = 1.5$$

and

$$\operatorname{Var}(Y) = \mathcal{E}\big(\operatorname{Var}(Y \mid \Lambda)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid \Lambda)\big) = \mathcal{E}(\Lambda) + \operatorname{Var}(\Lambda) = 1.5 + 0.75 = 2.25.$$

b) In this case, $\Lambda \sim \mathcal{T}(0, 3)$, so that

$$\mathcal{E}(\Lambda) = \frac{0 + 3}{2} = 1.5 \qquad \text{and} \qquad \operatorname{Var}(\Lambda) = \frac{(3 - 0)^2}{24} = 0.375.$$

Hence, by the laws of total expectation and total variance,

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid \Lambda)\big) = \mathcal{E}(\Lambda) = 1.5$$

and

$$\operatorname{Var}(Y) = \mathcal{E}\big(\operatorname{Var}(Y \mid \Lambda)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid \Lambda)\big) = \mathcal{E}(\Lambda) + \operatorname{Var}(\Lambda) = 1.5 + 0.375 = 1.875.$$

10.89 Let Y denote the lifetime of the selected device. Let $X = k$ if the device is of Type k. For $k = 1, 2$, we have $\mathcal{E}(Y \mid X = k) = \mu_k$ and $\operatorname{Var}(Y \mid X = k) = \sigma_k^2$. Thus,

$$\mathcal{E}(Y \mid X) = (2 - X)\mu_1 + (X - 1)\mu_2 = 2\mu_1 - \mu_2 + (\mu_2 - \mu_1)X$$

and

$$\operatorname{Var}(Y \mid X) = (2 - X)\sigma_1^2 + (X - 1)\sigma_2^2 = 2\sigma_1^2 - \sigma_2^2 + (\sigma_2^2 - \sigma_1^2)X.$$

Moreover, observing that we can write $X = 1 + W$, where $W \sim \mathcal{B}(1, 1 - p)$, we see that

$$\mathcal{E}(X) = 1 + \mathcal{E}(W) = 1 + (1 - p) = 2 - p \qquad \text{and} \qquad \operatorname{Var}(X) = \operatorname{Var}(W) = p(1 - p).$$

Hence, by the laws of total expectation and total variance,

$$\begin{aligned}\mathcal{E}(Y) &= \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\big(2\mu_1 - \mu_2 + (\mu_2 - \mu_1)X\big) = 2\mu_1 - \mu_2 + (\mu_2 - \mu_1)\mathcal{E}(X) \\ &= 2\mu_1 - \mu_2 + (\mu_2 - \mu_1)(2 - p) = p\mu_1 + (1 - p)\mu_2\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}(Y) &= \mathcal{E}\big(\operatorname{Var}(Y \mid X)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid X)\big) \\ &= \mathcal{E}\Big(2\sigma_1^2 - \sigma_2^2 + (\sigma_2^2 - \sigma_1^2)X\Big) + \operatorname{Var}\big(2\mu_1 - \mu_2 + (\mu_2 - \mu_1)X\big) \\ &= 2\sigma_1^2 - \sigma_2^2 + (\sigma_2^2 - \sigma_1^2)\mathcal{E}(X) + (\mu_2 - \mu_1)^2 \operatorname{Var}(X) \\ &= 2\sigma_1^2 - \sigma_2^2 + (\sigma_2^2 - \sigma_1^2)(2 - p) + (\mu_2 - \mu_1)^2 p(1 - p) \\ &= p\sigma_1^2 + (1 - p)\sigma_2^2 + p(1 - p)(\mu_2 - \mu_1)^2.\end{aligned}$$

10.90

a) Let Y denote the total claim amount. We can write $Y = \sum_{k=1}^{N} Y_k$, where N is the number of claims and $Y_1, Y_2, \ldots,$ are independent Beta(1, 2) random variables. Note that $N \sim \mathcal{B}(32, 1/6)$ and that $\mathcal{E}(Y_j) = 1/3$ and $\operatorname{Var}(Y_j) = 1/18$. From Examples 7.26 and 7.29 on pages 383 and 387, respectively,

$$\mathcal{E}(Y) = \frac{1}{3}\mathcal{E}(N) = \frac{1}{3} \cdot 32 \cdot \frac{1}{6} = \frac{16}{9}$$

and

$$\operatorname{Var}(Y) = \frac{1}{18}\mathcal{E}(N) + \left(\frac{1}{3}\right)^2 \operatorname{Var}(N) = \frac{1}{18} \cdot 32 \cdot \frac{1}{6} + \frac{1}{9} \cdot 32 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{8}{27} + \frac{40}{81} = \frac{64}{81}.$$

b) Let Y denote the total claim amount. We can write $Y = \sum_{k=1}^{32} Z_k$, where Z_k denotes the claim amount for risk k. Let $X_k = 1$ if risk k makes a claim and let $X_k = 0$ otherwise. Note that each $X_k \sim \mathcal{B}(1, 1/6)$ and, therefore, $\mathcal{E}(X_k) = 1/6$ and $\operatorname{Var}(X_k) = 5/36$. Now, given $X_k = 1$, $Z_k \sim \text{Beta}(1, 2)$, whereas, given $X_k = 0$, $Z_k = 0$. Thus,

$$\mathcal{E}(Z_k \mid X_k = x) = \begin{cases} 1/3, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad \operatorname{Var}(Z_k \mid X_k = x) = \begin{cases} 1/18, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$\mathcal{E}(Z_k \mid X_k) = \frac{1}{3}X_k \qquad \text{and} \qquad \operatorname{Var}(Z_k \mid X_k) = \frac{1}{18}X_k.$$

Applying the laws of total expectation and total variance, we get

$$\mathcal{E}(Z_k) = \mathcal{E}\big(\mathcal{E}(Z_k \mid X_k)\big) = \mathcal{E}\left(\frac{1}{3}X_k\right) = \frac{1}{3}\mathcal{E}(X_k) = \frac{1}{3}\cdot\frac{1}{6} = \frac{1}{18}$$

and

$$\begin{aligned}\operatorname{Var}(Z_k) &= \mathcal{E}\big(\operatorname{Var}(Z_k \mid X_k)\big) + \operatorname{Var}\big(\mathcal{E}(Z_k \mid X_k)\big) = \mathcal{E}\left(\frac{1}{18}X_k\right) + \operatorname{Var}\left(\frac{1}{3}X_k\right) \\ &= \frac{1}{18}\mathcal{E}(X_k) + \frac{1}{9}\operatorname{Var}(X_k) = \frac{1}{18}\cdot\frac{1}{6} + \frac{1}{9}\cdot\frac{5}{36} \\ &= \frac{2}{81}.\end{aligned}$$

Therefore,

$$\mathcal{E}(Y) = \mathcal{E}\left(\sum_{k=1}^{32} Z_k\right) = \sum_{k=1}^{32}\mathcal{E}(Z_k) = 32\cdot\frac{1}{18} = \frac{16}{9}$$

and, because the Z_ks are independent,

$$\operatorname{Var}(Y) = \operatorname{Var}\left(\sum_{k=1}^{32} Z_k\right) = \sum_{k=1}^{32}\operatorname{Var}(Z_k) = 32\cdot\frac{2}{81} = \frac{64}{81}.$$

10.91 We use units of thousands of dollars. Let Y denote the claim payment and let X denote the amount of damage if there is partial damage to the car. By assumption, f is a PDF of X. Also, let $W = 0$ if there is no damage, $W = 1$ if there is partial damage, and $W = 2$ if there is a total loss. Now, we have $Y_{|W=0} = 0$, $Y_{|W=1} = \max\{0, X-1\}$, and $Y_{|W=2} = 14$. Applying the FEF for continuous random variables yields

$$\begin{aligned}\mathcal{E}(Y \mid W = 1) &= \mathcal{E}\big(\max\{0, X-1\}\big) = \int_{-\infty}^{\infty}\max\{0, x-1\}f(x)\,dx \\ &= \int_0^1 0\cdot f(x)\,dx + \int_1^{15}(x-1)f(x)\,dx = 0.5003\int_1^{15}(x-1)\,e^{-x/2}\,dx \\ &= 1.2049.\end{aligned}$$

Therefore, by the law of total probability and the FEF for discrete random variables,

$$\begin{aligned}\mathcal{E}(Y) &= \mathcal{E}\big(\mathcal{E}(Y \mid W)\big) = \sum_w \mathcal{E}(Y \mid W = w)\,p_W(w) \\ &= 0\cdot p_W(0) + 1.2049\cdot p_W(1) + 14\cdot p_W(2) \\ &= 1.2049\cdot 0.04 + 14\cdot 0.02 = 0.328196.\end{aligned}$$

The expected claim payment is \$328.20.

Theory Exercises

10.92

a) From the bivariate form of the FEF and the general multiplication rule for continuous random variables,

$$\begin{aligned}
\mathcal{E}\big(\mathcal{E}(Z \mid X, Y)\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{E}(Z \mid X = x, Y = y)\, f_{X,Y}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} z f_{Z \mid X,Y}(z \mid x, y)\, dz\right) f_{X,Y}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} z \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y) f_{Z \mid X,Y}(z \mid x, y)\, dx\, dy\right) dz \\
&= \int_{-\infty}^{\infty} z \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\, dx\, dy\right) dz \\
&= \int_{-\infty}^{\infty} z f_Z(z)\, dz \\
&= \mathcal{E}(Z)\,.
\end{aligned}$$

b) From the multivariate form of the FEF and the general multiplication rule for continuous random variables,

$$\begin{aligned}
&\mathcal{E}\big(\mathcal{E}(X_m \mid X_1, \dots, X_{m-1})\big) \\
&\quad= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{E}(X_m \mid X_1 = x_1, \dots, X_{m-1} = x_{m-1})\, f_{X_1,\dots,X_{m-1}}(x_1, \dots, x_{m-1})\, dx_1 \cdots dx_{m-1} \\
&\quad= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x_m f_{X_m \mid X_1,\dots,X_{m-1}}(x_m \mid x_1, \dots, x_{m-1})\, dx_m\right) f_{X_1,\dots,X_{m-1}}(x_1, \dots, x_{m-1}) \\
&\qquad \times\, dx_1 \cdots dx_{m-1} \\
&\quad= \int_{-\infty}^{\infty} x_m \bigg(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_{m-1}}(x_1, \dots, x_{m-1}) f_{X_m \mid X_1,\dots,X_{m-1}}(x_m \mid x_1, \dots, x_{m-1}) \\
&\qquad \times\, dx_1 \cdots dx_{m-1}\bigg) dx_m \\
&\quad= \int_{-\infty}^{\infty} x_m \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_m}(x_1, \dots, x_m)\, dx_1 \cdots dx_{m-1}\right) dx_m \\
&\quad= \int_{-\infty}^{\infty} x_m f_{X_m}(x_m)\, dx_m \\
&\quad= \mathcal{E}(X_m)\,.
\end{aligned}$$

10.93

a) Let $X = \sum_n n I_{A_n}$. Observe that $\{X = n\} = A_n$ for all $n \in \mathcal{N}$. We see that X is a discrete random variable with PMF given by $p_X(x) = P(A_n)$ for $x \in \mathcal{N}$, and $p_X(x) = 0$ otherwise. Applying the law of total expectation and the FEF for discrete random variables, we get

$$\begin{aligned}\mathcal{E}(Y) &= \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \sum_x \mathcal{E}(Y \mid X = x)\, p_X(x)\\ &= \sum_n \mathcal{E}(Y \mid X = n)\, P(X = n) = \sum_n \mathcal{E}(Y \mid A_n)\, P(A_n).\end{aligned}$$

b) Let B be an event and set $Y = I_B$. Applying part (a) yields

$$P(B) = \mathcal{E}(I_B) = \mathcal{E}(Y) = \sum_n \mathcal{E}(Y \mid A_n)\, P(A_n) = \sum_n \mathcal{E}(I_B \mid A_n)\, P(A_n) = \sum_n P(B \mid A_n) P(A_n),$$

which is the law of total probability.

Advanced Exercises

10.94 Let Y denote the lifetime of the component and let E be the event that the component fails immediately. Note that $\mathcal{E}(Y \mid E) = 0$. Applying Exercise 10.93(a) with the partition E and E^c, we obtain

$$\mathcal{E}(Y) = \mathcal{E}\big(Y \mid E\big)\, P\big(E\big) + \mathcal{E}\big(Y \mid E^c\big)\, P\big(E^c\big) = 0 \cdot p + \mathcal{E}\big(Y \mid E^c\big)\,(1-p) = (1-p)\mathcal{E}\big(Y \mid E^c\big).$$

By assumption, the conditional distribution of Y given E^c has CDF F. Consequently, by Proposition 10.5 on page 577, we have $\mathcal{E}\big(Y \mid E^c\big) = \int_0^\infty \big(1 - F(y)\big)\, dy$. Therefore,

$$\mathcal{E}(Y) = (1-p) \int_0^\infty \big(1 - F(y)\big)\, dy.$$

Note: We could also solve this problem with the law of total expectation by conditioning on the random variable $X = I_E$.

10.95

a) Let $G(a, b) = \mathcal{E}\Big(\big(Y - (a + bX)\big)^2\Big)$. We want to find real numbers a and b that minimize G. For convenience, set $\rho = \rho(X, Y)$. Now,

$$\begin{aligned}\frac{\partial G}{\partial a}(a, b) &= \mathcal{E}\Big(2\big(Y - (a + bX)\big)\cdot(-1)\Big) = -2\left(\mathcal{E}(Y) - a - b\mathcal{E}(X)\right)\\ \frac{\partial G}{\partial b}(a, b) &= \mathcal{E}\Big(2\big(Y - (a + bX)\big)\cdot(-X)\Big) = -2\Big(\mathcal{E}(XY) - a\mathcal{E}(X) - b\mathcal{E}\big(X^2\big)\Big).\end{aligned}$$

Setting these two partial deriviatives equal to 0 and solving for a and b, we find that

$$a = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X \qquad \text{and} \qquad b = \rho\frac{\sigma_Y}{\sigma_X}.$$

Thus, the best linear predictor of Y based on observing X is

$$\left(\mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X\right) + \rho\frac{\sigma_Y}{\sigma_X}X = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X).$$

b) Referring to part (a), we have, for the optimal a and b,

$$\begin{aligned}\big(Y - (a + bX)\big)^2 &= \left(Y - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)\right)\right)^2 = \left((Y - \mu_Y) - \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)\right)^2\\ &= (Y - \mu_Y)^2 - 2\rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)(Y - \mu_Y) + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}(X - \mu_X)^2.\end{aligned}$$

Therefore, the minimum mean square error for linear prediction is

$$\begin{aligned}\mathcal{E}&\left((Y-\mu_Y)^2 - 2\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)(Y-\mu_Y) + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}(X-\mu_X)^2\right)\\ &= \operatorname{Var}(Y) - 2\rho\frac{\sigma_Y}{\sigma_X}\operatorname{Cov}(X,Y) + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\operatorname{Var}(X)\\ &= \sigma_Y^2 - 2\rho\frac{\sigma_Y}{\sigma_X}\rho\sigma_X\sigma_Y + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 = \left(1-\rho^2\right)\sigma_Y^2.\end{aligned}$$

10.96

a) In Exercise 10.67(a), we found that $\rho = \rho(X,Y) = 1/n$. Applying now Exercise 10.95(a) and the equations referred to in the statement of this exercise, we find that

$$a = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X = \frac{n}{n+1} - \frac{1}{n}\frac{\sqrt{n/(n+2)(n+1)^2}}{\sqrt{n/(n+2)(n+1)^2}}\frac{1}{n+1} = \frac{n^2-1}{n(n+1)} = \frac{n-1}{n}$$

and

$$b = \rho\frac{\sigma_Y}{\sigma_X} = \frac{1}{n}\frac{\sqrt{n/(n+2)(n+1)^2}}{\sqrt{n/(n+2)(n+1)^2}} = \frac{1}{n}.$$

Thus, the best linear predictor of Y, based on observing X, is

$$a + bx = \frac{n-1}{n} + \frac{1}{n}x = 1 - \frac{1-x}{n}.$$

b) The two predictors are the same; that is, the best linear predictor is also the best predictor.
c) The two predictors are identical because the best predictor is linear and, hence, in this case, the best predictor is also the best linear predictor.
d) As we have seen, in this case, the best linear predictor is also the best predictor. Thus, the minimum mean square error for linear prediction, which, by Exercise 10.95(b), is $\left(1-\rho^2\right)\sigma_Y^2$, must also be the minimum mean square error for prediction, which, by Exercise 10.86, is $\mathcal{E}\left(\operatorname{Var}(Y \mid X)\right)$. To verify explicitly that these two quantities are equal, we proceed as follows. In Example 10.19 on page 600, we found that

$$\mathcal{E}\left(\operatorname{Var}(Y \mid X)\right) = \frac{n-1}{n(n+1)(n+2)}.$$

Furthermore,

$$\left(1-\rho^2\right)\sigma_Y^2 = \left(1 - \left(\frac{1}{n}\right)^2\right)\frac{n}{(n+2)(n+1)^2} = \frac{n\left(n^2-1\right)}{n^2(n+2)(n+1)^2} = \frac{n-1}{n(n+1)(n+2)}.$$

10.97

a) We have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 \left(2 - \frac{6}{5}x - \frac{4}{5}y\right)dy = \frac{8}{5} - \frac{6}{5}x$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise. Hence, for $0 < x < 1$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2 - \frac{6}{5}x - \frac{4}{5}y}{\frac{8}{5} - \frac{6}{5}x} = \frac{5-3x-2y}{4-3x}$$

if $0 < y < 1$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise.

Consequently,

$$\begin{aligned}\mathcal{E}(Y \mid X = x) &= \int_{-\infty}^{\infty} y f_{Y \mid X}(y \mid x)\, dy = \frac{1}{4-3x}\int_0^1 y \cdot (5 - 3x - 2y)\, dy \\ &= \frac{1}{4-3x}\int_0^1 \left(5y - 3xy - 2y^2\right) dy = \frac{1}{4-3x}\left(\frac{11}{6} - \frac{3}{2}x\right) \\ &= \frac{11 - 9x}{6(4-3x)}\end{aligned}$$

if $0 < x < 1$, and $\mathcal{E}(Y \mid X = x) = 0$ otherwise. We can now conclude from the prediction theorem that, if an employee took X hundred hours of sick leave last year, the best prediction for the number of hundreds of hours, Y, that he will take this year is

$$\mathcal{E}(Y \mid X) = \frac{11 - 9X}{6(4 - 3X)}.$$

b) We see from Exercise 10.95 that to find the best linear predictor of Y, based on observing X, we must determine $\mu_X, \mu_Y, \sigma_X, \sigma_Y$, and $\rho = \rho(X, Y)$. From the FEF,

$$\begin{aligned}\mathcal{E}(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x, y)\, dx\, dy = \int_0^1 x\left(\int_0^1 \left(2 - \frac{6}{5}x - \frac{4}{5}y\right) dy\right) dx \\ &= \int_0^1 x \cdot \left(\frac{8}{5} - \frac{6}{5}x\right) dx = \int_0^1 \left(\frac{8}{5}x - \frac{6}{5}x^2\right) dx = \frac{2}{5}\end{aligned}$$

and

$$\begin{aligned}\mathcal{E}(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x, y)\, dx\, dy = \int_0^1 y\left(\int_0^1 \left(2 - \frac{6}{5}x - \frac{4}{5}y\right) dx\right) dy \\ &= \int_0^1 y \cdot \left(\frac{7}{5} - \frac{4}{5}y\right) dy = \int_0^1 \left(\frac{7}{5}y - \frac{4}{5}y^2\right) dy = \frac{13}{30}\end{aligned}$$

and

$$\begin{aligned}\mathcal{E}(X^2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{X,Y}(x, y)\, dx\, dy = \int_0^1 x^2\left(\int_0^1 \left(2 - \frac{6}{5}x - \frac{4}{5}y\right) dy\right) dx \\ &= \int_0^1 x^2 \cdot \left(\frac{8}{5} - \frac{6}{5}x\right) dx = \int_0^1 \left(\frac{8}{5}x^2 - \frac{6}{5}x^3\right) dx = \frac{7}{30}\end{aligned}$$

and

$$\begin{aligned}\mathcal{E}(Y^2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f_{X,Y}(x, y)\, dx\, dy = \int_0^1 y^2\left(\int_0^1 \left(2 - \frac{6}{5}x - \frac{4}{5}y\right) dx\right) dy \\ &= \int_0^1 y^2 \cdot \left(\frac{7}{5} - \frac{4}{5}y\right) dy = \int_0^1 \left(\frac{7}{5}y^2 - \frac{4}{5}y^3\right) dy = \frac{4}{15}\end{aligned}$$

and

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)\,dx\,dy = \int_0^1 x\left(\int_0^1 y\cdot\left(2-\frac{6}{5}x-\frac{4}{5}y\right)dy\right)dx$$
$$= \int_0^1 x\left(\int_0^1 \left(2y-\frac{6}{5}xy-\frac{4}{5}y^2\right)dy\right)dx = \int_0^1 x\cdot\left(\frac{11}{15}-\frac{3}{5}x\right)dx$$
$$= \int_0^1 \left(\frac{11}{15}x-\frac{3}{5}x^2\right)dx = \frac{1}{6}.$$

It now follows that $\mu_X = 2/5$, $\mu_Y = 13/30$, $\sigma_X^2 = 11/150$, $\sigma_Y^2 = 71/900$, and $\operatorname{Cov}(X,Y) = -1/150$. Therefore, from Exercise 10.95(a), the best linear predictor of the number of hundreds of hours, Y, an employee will take this year if he took X hundred hours last year is

$$\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X) = \frac{13}{30} - \frac{1}{11}\left(X-\frac{2}{5}\right) = \frac{31}{66} - \frac{1}{11}X.$$

c) The results of parts (a) and (b) are different because the best predictor is not linear.
d) Using the best predictor and the best linear predictor, the estimates for the number of hundreds of hours of sick leave an employee will take this year if he took 10 hours (1/10 hundred hours) last year are

$$\frac{11-9\cdot\frac{1}{10}}{6\left(4-3\cdot\frac{1}{10}\right)} = \frac{101}{222} \approx 0.455$$

and

$$\frac{31}{66} - \frac{1}{11}\cdot\frac{1}{10} = \frac{76}{165} \approx 0.461,$$

respectively. In other words, for an employee who took 10 hours of sick leave last year, the best prediction for the number of hours he will take this year is about 45.5 hours and the best linear prediction is about 46.1 hours.

10.98
a) Assuming finite expectations, we define

$$\mathcal{E}(Y \mid X = x) = \sum_y y p_{Y\,|\,X}(y \mid x)$$

if $f_X(x) > 0$, and $\mathcal{E}(Y \mid X = x) = 0$ otherwise; and we define

$$\mathcal{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X\,|\,Y}(x \mid y)\,dx$$

if $p_Y(y) > 0$, and $\mathcal{E}(X \mid Y = y) = 0$ otherwise.
b) From the FEF for continuous random variables and part (a),

$$\mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \int_{-\infty}^{\infty} \mathcal{E}(Y \mid X = x)\, f_X(x)\,dx = \int_{-\infty}^{\infty}\left(\sum_y y p_{Y\,|\,X}(y \mid x)\right) f_X(x)\,dx$$
$$= \int_{-\infty}^{\infty}\left(\sum_y y\,\frac{h_{X,Y}(x,y)}{f_X(x)}\right) f_X(x)\,dx = \sum_y y\left(\int_{-\infty}^{\infty} h_{X,Y}(x,y)\,dx\right)$$
$$= \sum_y y p_Y(y) = \mathcal{E}(Y).$$

From the FEF for discrete random variables and part (a),

$$\begin{aligned}\mathcal{E}\big(\mathcal{E}(X \mid Y)\big) &= \sum_y \mathcal{E}(X \mid Y = y)\, p_Y(y) = \sum_y \left(\int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y)\, dx\right) p_Y(y) \\ &= \sum_y \left(\int_{-\infty}^{\infty} x \, \frac{h_{X,Y}(x, y)}{p_Y(y)}\, dx\right) p_Y(y) = \int_{-\infty}^{\infty} x \left(\sum_y h_{X,Y}(x, y)\right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x)\, dx = \mathcal{E}(X)\,.\end{aligned}$$

10.99
a) In Exercise 9.86(b), we showed that $Y_{|X=x} \sim \mathcal{P}(x)$ for $x > 0$ and that $X_{|Y=y} \sim \Gamma(y+1, 1+\lambda)$ for $y = 0, 1, \ldots.$ Consequently, we have $\mathcal{E}(Y \mid X = x) = x$ if $x > 0$, and equals 0 otherwise; and $\mathcal{E}(X \mid Y = y) = (y+1)/(1+\lambda)$ if $y = 0, 1, \ldots,$ and equals 0 otherwise.
b) In Exercise 9.86(a), we showed that $X \sim \mathcal{E}(\lambda)$. Hence, from part (a) and the law of total expectation,

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}(X) = \frac{1}{\lambda}.$$

We also showed in Exercise 9.86(a) that Y has the Pascal distribution, as defined in Exercise 7.26 on page 336, with parameter $1/\lambda$. From that exercise, $\mathcal{E}(Y) = 1/\lambda$. Hence, from part (a) and the law of total expectation,

$$\mathcal{E}(X) = \mathcal{E}\big(\mathcal{E}(X \mid Y)\big) = \mathcal{E}\left(\frac{Y+1}{1+\lambda}\right) = \frac{\mathcal{E}(Y)+1}{1+\lambda} = \frac{1/\lambda + 1}{1+\lambda} = \frac{1}{\lambda}.$$

10.100
a) By assumption $Y_{|X=x} \sim \mathcal{P}(x)$ for $x > 0$. Hence, $\mathcal{E}(Y \mid X = x) = x$ if $x > 0$, and equals 0 otherwise.
b) From the law of total expectation and part (a),

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}(X) = \frac{\alpha}{\beta}.$$

c) In Exercise 9.87(b), we found that Y has the Pascal distribution, as given by Equation (5.50) on page 243, with parameters α and $\beta/(\beta+1)$. For convenience, set $p = \beta/(\beta+1)$. Referring to the binomial series, Equation (5.45) on page 241, we get

$$\begin{aligned}\mathcal{E}(Y) &= \sum_y y p_Y(y) = \sum_{y=0}^{\infty} y \binom{-\alpha}{y} p^{\alpha} (p-1)^y = p^{\alpha}(p-1) \sum_{y=0}^{\infty} y \binom{-\alpha}{y} (p-1)^{y-1} \\ &= p^{\alpha}(p-1) \frac{d}{dp} \left(\sum_{y=0}^{\infty} \binom{-\alpha}{y} (p-1)^y\right) = p^{\alpha}(p-1) \frac{d}{dp} \big(1 + (p-1)\big)^{-\alpha} \\ &= p^{\alpha}(p-1) \frac{d}{dp} p^{-\alpha} = \alpha p^{\alpha}(1-p) p^{-\alpha-1} = \alpha\left(\frac{1}{p} - 1\right) = \frac{\alpha}{\beta}.\end{aligned}$$

d) In Exercise 9.87(c), we found that $X_{|Y=y} \sim \Gamma(y+\alpha, \beta+1)$ for $y = 0, 1, \ldots.$ Hence, we see that $\mathcal{E}(X \mid Y = y) = (y+\alpha)/(\beta+1)$ if $y = 0, 1, \ldots,$ and equals 0 otherwise.
e) From the law of total expectation and part (d),

$$\mathcal{E}(X) = \mathcal{E}\big(\mathcal{E}(X \mid Y)\big) = \mathcal{E}\left(\frac{Y+\alpha}{\beta+1}\right) = \frac{\mathcal{E}(Y)+\alpha}{\beta+1} = \frac{\alpha/\beta + \alpha}{\beta+1} = \frac{\alpha}{\beta}.$$

10.5 The Bivariate Normal Distribution

Basic Exercises

10.101 We have $\mu_X = -2$, $\sigma_X^2 = 4$, $\mu_Y = 3$, $\sigma_Y^2 = 2$, and $\rho = -0.7$. Moreover,

$$\operatorname{Cov}(X, Y) = \rho\sigma_X\sigma_Y = (-0.7) \cdot 2 \cdot \sqrt{2} = -1.4\sqrt{2}.$$

From the linearity property of expected value,

$$\mu_W = \mathcal{E}(W) = \mathcal{E}(-2X + 4Y - 12) = -2\mathcal{E}(X) + 4\mathcal{E}(Y) - 12 = -2 \cdot (-2) + 4 \cdot 3 - 12 = 4.$$

Also, from properties of variance and covariance,

$$\begin{aligned}\sigma_W^2 &= \operatorname{Var}(W) = \operatorname{Var}(-2X + 4Y - 12) = \operatorname{Var}(-2X + 4Y) \\ &= \operatorname{Var}(-2X) + \operatorname{Var}(4Y) + 2\operatorname{Cov}(-2X, 4Y) = 4\operatorname{Var}(X) + 16\operatorname{Var}(Y) - 16\operatorname{Cov}(X, Y) \\ &= 4 \cdot 4 + 16 \cdot 2 - 16 \cdot (-1.4\sqrt{2}) = 48 + 22.4\sqrt{2}.\end{aligned}$$

Hence, $\sigma_W = \sqrt{48 + 22.4\sqrt{2}} = 8.93$.

10.102

a) Because f_1 and f_2 are nonnegative functions, so is f. Moreover,

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{f_1(x, y) + f_2(x, y)}{2}\right)dx\,dy \\ &= \frac{1}{2}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x, y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_2(x, y)\,dx\,dy\right) \\ &= \frac{1}{2}(1 + 1) = 1.\end{aligned}$$

b) From Proposition 10.10 on page 616, we know that the marginals of f_1 are both $\mathcal{N}(0, 1)$, as are those of f_2. Then

$$\begin{aligned}\int_{-\infty}^{\infty} f(x, y)\,dy &= \int_{-\infty}^{\infty}\left(\frac{f_1(x, y) + f_2(x, y)}{2}\right)dy = \frac{1}{2}\left(\int_{-\infty}^{\infty} f_1(x, y)\,dy + \int_{-\infty}^{\infty} f_2(x, y)\,dy\right) \\ &= \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2} + \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\right) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}.\end{aligned}$$

Similarly, we find that

$$\int_{-\infty}^{\infty} f(x, y)\,dx = \frac{1}{\sqrt{2\pi}}\,e^{-y^2/2}.$$

Hence, both marginals of f are standard normal PDFs.

10.103

a) The level curves of the joint PDF of X and Y are obtained by setting $f_{X,Y}(x, y)$ equal to a positive constant or, equivalently, they are the curves of the form $Q(x, y) - k = 0$, where k is a constant. This second-order equation can be written in the form

$$\frac{1}{\sigma_X^2}x^2 - \frac{2\rho}{\sigma_X\,\sigma_Y}xy + \frac{1}{\sigma_Y^2}y^2 + ax + by + c = 0,$$

where a, b, and c are constants. The discriminant of this equation is

$$\Delta = \left(\frac{2\rho}{\sigma_X\,\sigma_Y}\right)^2 - 4 \cdot \frac{1}{\sigma_X^2} \cdot \frac{1}{\sigma_Y^2} = \frac{4}{\sigma_X^2\,\sigma_Y^2}\big(\rho^2 - 1\big).$$

Because $-1 < \rho < 1$, we see that $\Delta < 0$ and, consequently, we know from analytic geometry that the level curves are ellipses (or circles).

b) From analytic geometric, the level-curve ellipses are parallel to the coordinate axes if and only if the coefficient of the xy-term is 0, that is, if and only if $\rho = 0$. In view of Proposition 10.9 on page 613, this latter condition is satisfied if and only if X and Y are independent random variables.

10.104 We have

$$\mu_X = 120, \qquad \sigma_X = 20, \qquad \mu_Y = 82, \qquad \sigma_Y = 15, \qquad \rho = 0.6.$$

a) From the previous display, we find that

$$\begin{aligned} 2\sigma_X\sigma_Y\sqrt{1-\rho^2} &= 2 \cdot 20 \cdot 15\sqrt{1-0.6^2} = 480, \\ 2\big(1-\rho^2\big) &= 2\big(1-0.6^2\big) = 2(0.64) = 32/25, \\ 2\rho &= 2 \cdot 0.6 = 6/5. \end{aligned}$$

Therefore, from Definition 10.4 on page 610,

$$f_{X,Y}(x, y) = \frac{1}{480\pi}\, e^{-\frac{25}{32}\left\{\left(\frac{x-120}{20}\right)^2 - \frac{6}{5}\left(\frac{x-120}{20}\right)\left(\frac{y-82}{15}\right) + \left(\frac{y-82}{15}\right)^2\right\}}$$

b) From Proposition 10.10(a) on page 616, $X \sim \mathcal{N}(120, 400)$. Hence, the systolic blood pressures for this population are normally distributed with mean 120 and standard deviation 20.

c) From Proposition 10.10(b), $Y \sim \mathcal{N}(82, 225)$. Hence, the diastolic blood pressures for this population are normally distributed with mean 82 and standard deviation 15.

d) From Proposition 10.10(c), $\rho(X, Y) = \rho = 0.6$. Hence, the correlation between the systolic and diastolic blood pressures for this population is 0.6. We see that there is a moderate positive correlation between the two types of blood pressures. In particular, people with high systolic blood pressures tend to have higher diastolic blood pressures than people with low systolic blood pressures.

e) We have

$$\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 82 + 0.6 \cdot \frac{15}{20}(x - 120) = 28 + 0.45x$$

and

$$\sigma_Y^2\big(1-\rho^2\big) = 225\big(1-0.6^2\big) = 144.$$

Consequently, from Proposition 10.10(d), $Y_{|X=x} \sim \mathcal{N}(28 + 0.45x, 144)$. The diastolic blood pressures for people in the population who have a systolic blood pressure of x are normally distributed with mean $28 + 0.45x$ and standard deviation 12.

f) From part (e), we see that, in this case, the regression equation is $y = 28 + 0.45x$. Hence, the predicted diastolic blood pressure for an individual in this population whose systolic blood pressure is 123 equals $28 + 0.45 \cdot 123$, or 83.35.

g) We have

$$\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 120 + 0.6 \cdot \frac{20}{15}(y - 82) = 54.4 + 0.8y$$

and

$$\sigma_X^2\bigl(1 - \rho^2\bigr) = 400\bigl(1 - 0.6^2\bigr) = 256.$$

Thus, from Proposition 10.10(e), $X_{|Y=y} \sim \mathcal{N}(54.4 + 0.8y, 256)$. The systolic blood pressures for people in the population who have a diastolic blood pressure of y are normally distributed with mean $54.4 + 0.8y$ and standard deviation 16. For a diastolic blood pressure of 86, we have $54.4 + 0.8 \cdot 86 = 123.2$. Consequently,

$$P(X > 130 \mid Y = 86) = 1 - \Phi\left(\frac{130 - 123.2}{16}\right) = 1 - \Phi(0.425) = 0.335.$$

10.105 Throughout this exercise, nitrate concentrations are in units of micromoles per liter and arsenic concentrations are in units of nanomoles per liter. We have

$$\mu_X = 49.8, \qquad \sigma_X = 30.7, \qquad \mu_Y = 2.26, \qquad \sigma_Y = 1.18, \qquad \rho = 0.732.$$

a) From the previous display, we find that

$$\begin{aligned} 2\sigma_X\sigma_Y\sqrt{1-\rho^2} &= 2 \cdot 30.7 \cdot 1.18\sqrt{1 - 0.732^2} = 49.362, \\ 2\bigl(1-\rho^2\bigr) &= 2\bigl(1 - 0.732^2\bigr) = 0.928, \\ 2\rho &= 2 \cdot 0.732 = 1.464. \end{aligned}$$

Therefore, from Definition 10.4 on page 610,

$$f_{X,Y}(x, y) = \frac{1}{49.362\pi}\, e^{-1.077\left\{\left(\frac{x-49.8}{30.7}\right)^2 - 1.464\left(\frac{x-49.8}{30.7}\right)\left(\frac{y-2.26}{1.18}\right) + \left(\frac{y-2.26}{1.18}\right)^2\right\}}$$

b) From Proposition 10.10(a) on page 616, $X \sim \mathcal{N}(49.8, 30.7^2)$. Nitrate concentration is normally distributed with mean 49.8 and standard deviation 30.7 .
c) From Proposition 10.10(b), $Y \sim \mathcal{N}(2.26, 1.18^2)$. Arsenic concentration is normally distributed with mean 2.26 and standard deviation 1.18.
d) From Proposition 10.10(c), $\rho(X, Y) = \rho = 0.732$. Hence, the correlation between nitrate concentration and arsenic concentration is 0.732. A relatively high positive correlation exists between nitrate concentration and arsenic concentration.
e) We have

$$\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 2.26 + 0.732 \cdot \frac{1.18}{30.7}(x - 49.8) = 0.859 + 0.028x$$

and

$$\sigma_Y^2\bigl(1 - \rho^2\bigr) = 1.18^2\bigl(1 - 0.732^2\bigr) = 0.646.$$

So, from Proposition 10.10(d), $Y_{|X=x} \sim \mathcal{N}(0.859 + 0.028x, 0.646)$. Given a nitrate concentration of x, arsenic concentrations are normally distributed with mean $0.859 + 0.028x$ and standard deviation 0.804.
f) From part (e), we see that, in this case, the regression equation is $y = 0.859 + 0.028x$, whose graph is a straight line with slope 0.028 and y-intercept 0.859. Hence, the predicted arsenic concentration for a nitrate concentration of x is $0.859 + 0.028x$.
g) From part (f), we see that the predicted arsenic concentration for a nitrate concentration of 60 is $0.859 + 0.028 \cdot 60$, or 2.539.

10.106
a) From Example 10.22(b), the regression equation for predicting the height of an eldest son from the height of his mother is $y = 34.9 + 0.537x$. Hence, the predicted height of the eldest son of a mother who is $5'\,8''$ (i.e., 68 inches) tall is $34.9 + 0.537 \cdot 68$, or about 71.4 inches.
b) We have

$$\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 63.7 + 0.5 \cdot \frac{2.7}{2.9}(y - 69.1) \approx 31.533 + 0.466y.$$

Consequently, the predicted height of the mother of an eldest son who is $5'\,8''$ (i.e., 68 inches) tall is $31.533 + 0.466 \cdot 68$, or about 63.2 inches.
c) From Example 10.22, the heights of eldest sons of mothers who are $5'\,2''$ (i.e., 62 inches) tall are normally distributed with mean 68.2 inches and standard deviation $\sqrt{6.31}$ inches. Hence,

$$P(Y > 72 \mid X = 62) = 1 - \Phi\left(\frac{72 - 68.2}{\sqrt{6.31}}\right) = 1 - \Phi(1.51) = 0.065.$$

d) From Table I on page A-39, the probability that a normally distributed random variable takes a value within 1.96 standard deviations of its mean is 0.95. Hence, from Example 10.22, the required prediction interval is $68.2 \pm 1.96 \cdot \sqrt{6.31}$, or $(63.277'', 73.123'')$.

10.107
a) We know that we can express X and Y in the form given in Equations (10.55), where Z_1 and Z_2 are independent standard normal random variables. Consequently, a linear combination of X and Y is also a linear combination of Z_1 and Z_2 plus a constant (which may be 0). Referring now to Proposition 9.14 on page 540, we conclude that any nonzero linear combination of X and Y is also normally distributed.
b) We have

$$\mathcal{E}(X + Y) = \mathcal{E}(X) + \mathcal{E}(Y) = \mu_X + \mu_Y$$

and

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y) = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y.$$

Hence, from part (a), $X + Y \sim \mathcal{N}\big(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y\big)$.

Also, we have

$$\mathcal{E}(Y - X) = \mathcal{E}(Y) - \mathcal{E}(X) = \mu_Y - \mu_X$$

and

$$\begin{aligned}\operatorname{Var}(Y - X) &= \operatorname{Var}\big(Y + (-1)X\big) = \operatorname{Var}(Y) + \operatorname{Var}\big((-1)X\big) + 2\operatorname{Cov}\big(Y, (-1)X\big)\\ &= \operatorname{Var}(Y) + (-1)^2\operatorname{Var}(X) - 2\operatorname{Cov}(Y, X) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X, Y)\\ &= \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y.\end{aligned}$$

Hence, from part (a), $Y - X \sim \mathcal{N}\big(\mu_Y - \mu_X, \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y\big)$.

10.108 From Example 10.21,

$$\mu_X = 63.7, \quad \sigma_X = 2.7, \quad \mu_Y = 69.1, \quad \sigma_Y = 2.9, \quad \rho = 0.5.$$

a) From Exercise 10.107(b), $X + Y \sim \mathcal{N}\big(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y\big)$, which, in view of Proposition 8.15 on page 468, implies that

$$\frac{X + Y}{2} \sim \mathcal{N}\left(\frac{\mu_X + \mu_Y}{2}, \frac{\sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y}{4}\right) = \mathcal{N}(66.4, 5.883).$$

b) From Exercise 10.107(b),

$$Y - X \sim \mathcal{N}\left(\mu_Y - \mu_X, \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y\right) = \mathcal{N}(5.4, 7.870).$$

c) Here we want $\mathcal{E}(Y - X)$, which, by part (b), equals 5.4 inches.
d) We have, in view of part (b),

$$\begin{aligned} P(Y > X) &= P(Y - X > 0) = 1 - P(Y - X \le 0) = 1 - \Phi\left(\frac{0 - 5.4}{\sqrt{7.870}}\right) \\ &= 1 - \Phi(-1.925) = 0.973. \end{aligned}$$

e) Here we want $P(Y > 1.1X) = P(Y - 1.1X > 0)$. We have

$$\mathcal{E}(Y - 1.1X) = \mathcal{E}(Y) - 1.1\mathcal{E}(X) = 69.1 - 1.1 \cdot 63.7 = -0.97.$$

and

$$\begin{aligned} \operatorname{Var}(Y - 1.1X) &= \operatorname{Var}(Y) + (-1.1)^2 \operatorname{Var}(X) + 2 \cdot (-1.1) \operatorname{Cov}(X, Y) \\ &= \operatorname{Var}(Y) + 1.21 \operatorname{Var}(X) - 2.2\rho\sigma_X\sigma_Y = 8.618. \end{aligned}$$

Thus, from Exercise 10.107(a), $Y - 1.1X \sim \mathcal{N}(-0.97, 8.618)$. Consequently,

$$\begin{aligned} P(Y > 1.1X) &= P(Y - 1.1X > 0) = 1 - P(Y - 1.1X \le 0) \\ &= 1 - \Phi\left(\frac{0 - (-0.97)}{\sqrt{8.618}}\right) = 1 - \Phi(0.330) = 0.371. \end{aligned}$$

10.109 We have $\mu_X = 0$, $\sigma_X = 1$, $\mu_Y = 0$, $\sigma_Y = 1$. Hence,

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}.$$

From the FEF and symmetry,

$$\mathcal{E}\big(\max\{X, Y\}\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\{x, y\} f_{X,Y}(x, y)\, dx\, dy = 2 \iint_{y>x} y f_{X,Y}(x, y)\, dx\, dy.$$

To evaluate the integral, we rotate the coordinate axes $\pi/4$ radians. That is, we set

$$x = \frac{1}{\sqrt{2}}u - \frac{1}{\sqrt{2}}v \qquad \text{and} \qquad y = \frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v.$$

The Jacobian determinant of the inverse transformation is identically 1. Moreover, simple algebra shows that

$$\frac{1}{2(1 - \rho^2)}\left(x^2 - 2\rho xy + y^2\right) = \frac{1}{2}\left(\frac{u^2}{1 + \rho} + \frac{v^2}{1 - \rho}\right).$$

Therefore,

$$
\begin{aligned}
\mathcal{E}\big(\max\{X, Y\}\big) &= \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^\infty \left(\int_{-\infty}^\infty \frac{1}{\sqrt{2}}(u+v)\, e^{-\frac{1}{2}\left(\frac{u^2}{1+\rho}+\frac{v^2}{1-\rho}\right)} du \right) dv \\
&= \frac{1}{\pi\sqrt{2(1-\rho^2)}} \int_0^\infty e^{-\frac{v^2}{2(1-\rho)}} \left(\int_{-\infty}^\infty u\, e^{-\frac{u^2}{2(1+\rho)}} du + v \int_{-\infty}^\infty e^{-\frac{u^2}{2(1+\rho)}} du \right) dv \\
&= \frac{1}{\pi\sqrt{2(1-\rho^2)}} \int_0^\infty e^{-\frac{v^2}{2(1-\rho)}} \left(0 + v\sqrt{2\pi(1+\rho)} \right) dv \\
&= \frac{1}{\sqrt{\pi(1-\rho)}} \int_0^\infty v\, e^{-\frac{v^2}{2(1-\rho)}}\, dv = \frac{1-\rho}{\sqrt{\pi(1-\rho)}} \int_0^\infty e^{-w}\, dw = \sqrt{(1-\rho)/\pi}.
\end{aligned}
$$

10.110 To simulate bivariate normal random variables with specified means, variances, and correlation given by μ_1 and μ_2, σ_1^2 and σ_2^2, and ρ, respectively, proceed as follows. First simulate independent standard normal random variables, Z_1 and Z_2, by applying the Box-Müller transformation. Then set

$$
\begin{aligned}
X &= \mu_1 + \sigma_1 Z_1 \\
Y &= \mu_2 + \rho\sigma_2 Z_1 + \sqrt{1-\rho^2}\,\sigma_2\, Z_2.
\end{aligned}
$$

Referring to Equations (10.54) and (10.55) on page 608, we see that $(X, Y) \sim \mathcal{BVN}\big(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho\big)$.

Theory Exercises

10.111

a) We have

$$\mu_Y = \mathcal{E}(Y) = \mathcal{E}(\beta_0 + \beta_1 X + \epsilon) = \beta_0 + \beta_1\mathcal{E}(X) + \mathcal{E}(\epsilon) = \beta_0 + \beta_1\mathcal{E}(X) + 0 = \beta_0 + \beta_1\mu_X.$$

b) As X and ϵ are independent random variables, we have

$$\sigma_Y^2 = \operatorname{Var}(Y) = \operatorname{Var}(\beta_0 + \beta_1 X + \epsilon) = \beta_1^2 \operatorname{Var}(X) + \operatorname{Var}(\epsilon) = \beta_1^2\sigma_X^2 + \sigma^2.$$

c) As X and ϵ are independent random variables, we have

$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(X, \beta_0 + \beta_1 X + \epsilon) = \beta_1 \operatorname{Cov}(X, X) + \operatorname{Cov}(X, \epsilon) = \beta_1 \operatorname{Var}(X) + 0 = \beta_1\sigma_X^2.$$

Consequently,

$$\rho = \rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\beta_1\sigma_X^2}{\sqrt{\sigma_X^2\sigma_Y^2}} = \beta_1\frac{\sigma_X}{\sigma_Y}.$$

d) From parts (c) and (b),

$$\sigma_Y^2\big(1-\rho^2\big) = \sigma_Y^2\left(1 - \beta_1^2\frac{\sigma_X^2}{\sigma_Y^2}\right) = \sigma_Y^2 - \beta_1^2\sigma_X^2 = \sigma^2.$$

e) From parts (a) and (c),

$$\beta_0 = \mu_Y - \beta_1\mu_X = \mu_Y - \left(\rho\frac{\sigma_Y}{\sigma_X}\right)\mu_X.$$

f) This result just re-expresses the one from part (c).

g) From Equation (10.49) on page 602 with $k(x) = \beta_0 + \beta_1 x$ and $\ell(x, y) = 1$, and from the independence of X and ϵ, we get

$$\begin{aligned}\mathcal{E}(Y \mid X) &= \mathcal{E}(\beta_0 + \beta_1 X + \epsilon \mid X) = \mathcal{E}(\beta_0 + \beta_1 X \mid X) + \mathcal{E}(\epsilon \mid X) \\ &= (\beta_0 + \beta_1 X)\mathcal{E}(1 \mid X) + \mathcal{E}(\epsilon) = (\beta_0 + \beta_1 X) \cdot 1 + 0 \\ &= \beta_0 + \beta_1 X.\end{aligned}$$

Hence, $\mathcal{E}(Y \mid X = x) = \beta_0 + \beta_1 x$.

h) From Equation (10.49) on page 602, and from the independence of X and ϵ, we get

$$\begin{aligned}\mathcal{E}\big(Y^2 \mid X\big) &= \mathcal{E}\big((\beta_0 + \beta_1 X + \epsilon)^2 \mid X\big) = \mathcal{E}\big((\beta_0 + \beta_1 X)^2 + 2(\beta_0 + \beta_1 X)\epsilon + \epsilon^2 \mid X\big) \\ &= \mathcal{E}\big((\beta_0 + \beta_1 X)^2 \mid X\big) + 2\mathcal{E}\big((\beta_0 + \beta_1 X)\epsilon \mid X\big) + \mathcal{E}\big(\epsilon^2 \mid X\big) \\ &= (\beta_0 + \beta_1 X)^2 + 2(\beta_0 + \beta_1 X)\mathcal{E}(\epsilon \mid X) + \mathcal{E}\big(\epsilon^2 \mid X\big) \\ &= (\beta_0 + \beta_1 X)^2 + 2(\beta_0 + \beta_1 X)\mathcal{E}(\epsilon) + \mathcal{E}\big(\epsilon^2\big) \\ &= (\beta_0 + \beta_1 X)^2 + 2(\beta_0 + \beta_1 X) \cdot 0 + \sigma^2 \\ &= (\beta_0 + \beta_1 X)^2 + \sigma^2.\end{aligned}$$

Hence, we see that $\mathcal{E}\big(Y^2 \mid X = x\big) = (\beta_0 + \beta_1 x)^2 + \sigma^2$. Therefore, from part (g),

$$\operatorname{Var}(Y \mid X = x) = \mathcal{E}\big(Y^2 \mid X = x\big) - \big(\mathcal{E}(Y \mid X = x)\big)^2 = (\beta_0 + \beta_1 x)^2 + \sigma^2 - (\beta_0 + \beta_1 x)^2 = \sigma^2.$$

10.112

a) As X and ϵ are independent normally distributed random variables and $Y = \beta_0 + \beta_1 X + \epsilon$, it follows from Proposition 9.14 on page 540 that Y is normally distributed.

b) From the argument leading up to Definition 10.4 on page 610, we know that two random variables, X and Y, that can be written in the form of Equations (10.55), where Z_1 and Z_2 are independent standard normal random variables, have a bivariate normal distribution. Set $Z_1 = (X - \mu_X)/\sigma_X$ and $Z_2 = \epsilon/\sigma$. Then Z_1 and Z_2 are independent standard normal random variables. Now, we have $X = \mu_X + \sigma_X Z_1$ and $Y = \beta_0 + \beta_1 X + \epsilon = \beta_0 + \beta_1(\mu_X + \sigma_X Z_1) + \sigma Z_2$. Therefore,

$$\begin{aligned}X &= \mu_X + \sigma_X Z_1 + 0 \cdot Z_2 \\ Y &= (\beta_0 + \beta_1 \mu_X) + (\beta_1 \sigma_X) Z_1 + \sigma Z_2,\end{aligned}$$

Hence, X and Y have a bivariate normal distribution.

c) From part (b) and Proposition 10.10(d) on page 616, we know that $Y_{|X=x}$ is normally distributed. Moreover, from parts (g) and (h) of Exercise 10.111, $\mathcal{E}(Y \mid X = x) = \beta_0 + \beta_1 x$ and $\operatorname{Var}(Y \mid X = x) = \sigma^2$. Therefore, $Y_{|X=x} \sim \mathcal{N}\left(\beta_0 + \beta_1 x, \sigma^2\right)$.

Advanced Exercises

10.113

a) Clearly, $f(x, y) \geq 0$ for all $(x, y) \in \mathcal{R}^2$. Thus, it remains to show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$. Now,

$$\begin{aligned}\iint\limits_A f(x, y)\,dx\,dy &= \frac{1}{\pi}\int_{-\infty}^{-c} e^{-x^2/2}\left(\int_0^{\infty} e^{-y^2/2}\,dy\right)dx \\ &= \frac{1}{\pi}\frac{\sqrt{2\pi}}{2}\int_{-\infty}^{-c} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-c} e^{-x^2/2}\,dx.\end{aligned}$$

Similarly,

$$\iint_C f(x, y)\,dx\,dy = \frac{1}{\sqrt{2\pi}} \int_c^{\infty} e^{-x^2/2}\,dx.$$

Also,

$$\begin{aligned}\iint_B f(x, y)\,dx\,dy &= \frac{1}{\pi} \int_{-c}^{c} e^{-x^2/2} \left(\int_{-\infty}^{0} e^{-y^2/2}\,dy \right) dx \\ &= \frac{\sqrt{2\pi}}{2\pi} \int_{-c}^{c} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-x^2/2}\,dx.\end{aligned}$$

Hence,

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy &= \iint_A f(x, y)\,dx\,dy + \iint_B f(x, y)\,dx\,dy + \iint_C f(x, y)\,dx\,dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-c} e^{-x^2/2}\,dx + \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-x^2/2}\,dx + \frac{1}{\sqrt{2\pi}} \int_{c}^{\infty} e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\,dx = 1.\end{aligned}$$

b) We have

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \begin{cases} \dfrac{1}{\pi} e^{-x^2/2} \displaystyle\int_0^{\infty} e^{-y^2/2}\,dy, & \text{if } |x| > c; \\ \dfrac{1}{\pi} e^{-x^2/2} \displaystyle\int_{-\infty}^{0} e^{-y^2/2}\,dy, & \text{if } |x| < c. \end{cases} = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}.$$

Thus, $X \sim \mathcal{N}(0, 1)$. At this point, we must take $c = \Phi^{-1}(3/4)$. Then,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx = \begin{cases} \dfrac{1}{\pi} e^{-y^2/2} \displaystyle\int_{-c}^{c} e^{-x^2/2}\,dx, & \text{if } y < 0; \\ \dfrac{1}{\pi} e^{-y^2/2} \displaystyle\int_{-\infty}^{-c} e^{-x^2/2}\,dx + \dfrac{1}{\pi} e^{-y^2/2} \displaystyle\int_{c}^{\infty} e^{-x^2/2}\,dx, & \text{if } y > 0. \end{cases}$$

$$= \begin{cases} \dfrac{2\sqrt{2\pi}}{\pi}\, e^{-y^2/2} \displaystyle\int_0^{c} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx, & \text{if } y < 0; \\ \dfrac{2\sqrt{2\pi}}{\pi}\, e^{-y^2/2} \displaystyle\int_c^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx, & \text{if } y > 0. \end{cases}$$

$$= \begin{cases} \dfrac{2\sqrt{2}}{\sqrt{\pi}}\, e^{-y^2/2} \big(\Phi(c) - \Phi(0)\big), & \text{if } y < 0; \\ \dfrac{2\sqrt{2}}{\sqrt{\pi}}\, e^{-y^2/2} \big(1 - \Phi(c)\big) & \text{if } y > 0. \end{cases} = \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}.$$

Thus, $Y \sim \mathcal{N}(0, 1)$.

c) Note that, for each $y \in \mathcal{R}$, the function $g_y(x) = xe^{-x^2/2}$ is odd. Therefore,

$$\begin{aligned}\iint_A xyf(x,y)\,dx\,dy &= \int_0^\infty \frac{1}{\pi} ye^{-y^2/2}\left(\int_{-\infty}^{-c} g_y(x)\,dx\right)dy \\ &= \int_0^\infty \frac{1}{\pi} ye^{-y^2/2}\left(-\int_c^\infty g_y(x)\,dx\right)dy = -\iint_C xyf(x,y)\,dx\,dy\end{aligned}$$

and

$$\iint_B xyf(x,y)\,dx\,dy = \int_{-\infty}^0 \frac{1}{\pi} ye^{-y^2/2}\left(\int_{-c}^{c} g_y(x)\,dx\right)dy = \int_{-\infty}^0 \frac{1}{\pi} ye^{-y^2/2}\cdot 0\,dy = 0.$$

Hence,

$$\begin{aligned}\mathcal{E}(XY) &= \int_{-\infty}^\infty \int_{-\infty}^\infty xyf(x,y)\,dx\,dy \\ &= \iint_A xyf(x,y)\,dx\,dy + \iint_B xyf(x,y)\,dx\,dy + \iint_C xyf(x,y)\,dx\,dy \\ &= -\iint_C xyf(x,y)\,dx\,dy + 0 + \iint_C xyf(x,y)\,dx\,dy = 0.\end{aligned}$$

Therefore, in view of part (b),

$$\operatorname{Cov}(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = 0 - 0\cdot 0 = 0.$$

Thus, X and Y are uncorrelated.
d) We can show that X and Y aren't independent in various ways. One way is to note that

$$P(|X| < c,\ Y > 0) = 0 \neq \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \big(2\Phi(c) - 1\big)\cdot P(Y > 0) = P(|X| < c)P(Y > 0).$$

e) From part (c), X and Y are uncorrelated. If they were bivariate normal, then, by Proposition 10.9 on page 613, they would be independent, which, by part (d), they aren't. Hence, X and Y are not bivariate normal random variables.

10.114
a) From Proposition 10.10 on page 616, $Y \sim \mathcal{N}(0,1)$. Therefore,

$$\begin{aligned}F_{X\,|\,Y>0}(x) &= P(X \le x \,|\, Y > 0) = \frac{P(X \le x,\ Y > 0)}{P(Y > 0)} \\ &= \frac{\int_{-\infty}^x \int_0^\infty f_{X,Y}(s,y)\,ds\,dy}{1/2} = 2\int_{-\infty}^x \left(\int_0^\infty f_{X,Y}(s,y)\,dy\right)ds.\end{aligned}$$

Therefore, from Proposition 8.5 on page 422,

$$f_{X\,|\,Y>0}(x) = 2\int_0^\infty f_{X,Y}(x,y)\,dy.$$

b) From Proposition 10.10(e), $X_{|Y=y} \sim \mathcal{N}\left(\rho y, 1-\rho^2\right)$. We now get, by referring to part (a), applying the general multiplication rule, and making the substitution $u = y^2$,

$$\begin{aligned}
\mathcal{E}(X \mid Y > 0) &= \int_{-\infty}^{\infty} x f_{X \mid Y>0}(x)\,dx = \int_{-\infty}^{\infty} x \left(2\int_0^{\infty} f_{X,Y}(x, y)\,dy\right) dx \\
&= 2\int_0^{\infty} \left(\int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y)\,dx\right) f_Y(y)\,dy = 2\int_0^{\infty} \rho y f_Y(y)\,dy \\
&= \frac{2\rho}{\sqrt{2\pi}} \int_0^{\infty} y\, e^{-y^2/2}\,dy = \rho\sqrt{2/\pi} \int_0^{\infty} e^{-u}\,du = \rho\sqrt{2/\pi}.
\end{aligned}$$

10.115

a) Recalling that the covariance of a random variable with itself equals the variance of the random variable, we see that the entries of $\mathbf{\Sigma}$ provide the four possible covariances involving X and Y. For that reason, $\mathbf{\Sigma}$ is called the covariance matrix of X and Y.

b) The covariance matrix is symmetric because $\operatorname{Cov}(Y, X) = \operatorname{Cov}(X, Y)$. To show that it's positive definite, we first note that

$$\begin{aligned}
\mathbf{x}'\mathbf{\Sigma}\mathbf{x} &= [\,x \quad y\,]\begin{bmatrix} \operatorname{Var}(X) & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(Y, X) & \operatorname{Var}(Y) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = [\,x \quad y\,]\begin{bmatrix} \operatorname{Cov}(X, X) & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(Y, X) & \operatorname{Cov}(Y, Y) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\
&= x^2 \operatorname{Cov}(X, X) + yx \operatorname{Cov}(Y, X) + xy \operatorname{Cov}(X, Y) + y^2 \operatorname{Cov}(Y, Y) \\
&= \operatorname{Cov}(xX + yY, xX + yY) = \operatorname{Var}(xX + yY) \geq 0.
\end{aligned}$$

Thus, $\mathbf{\Sigma}$ is nonnegative definite. Now suppose that $\mathbf{x} \neq \mathbf{0}$. Then $xX + yY$ is a nonzero linear combination of X and Y and, hence, by Exercise 10.107(a), is normally distributed. In particular, then, we have $P(xX + yY = c) = 0$ for all real numbers c. It now follows from Proposition 7.7 on page 354 that $\mathbf{x}'\mathbf{\Sigma}\mathbf{x} = \operatorname{Var}(xX + yY) > 0$. Hence, $\mathbf{\Sigma}$ is positive definite.

c) We have

$$\begin{aligned}
\det \mathbf{\Sigma} &= \operatorname{Var}(X)\operatorname{Var}(Y) - \operatorname{Cov}(X, Y)\operatorname{Cov}(Y, X) = \sigma_X^2\sigma_Y^2 - \left(\operatorname{Cov}(X, Y)\right)^2 \\
&= \sigma_X^2\sigma_Y^2 - \rho^2\sigma_X^2\sigma_Y^2 = \left(1-\rho^2\right)\sigma_X^2\sigma_Y^2.
\end{aligned}$$

Therefore, $(\det \mathbf{\Sigma})^{1/2} = \sigma_X\sigma_Y\sqrt{1-\rho^2}$. Now, for convenience, set $\sigma_{XY} = \operatorname{Cov}(X, Y)$. We then have

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}^{-1} = \frac{1}{\det \mathbf{\Sigma}}\begin{bmatrix} \sigma_Y^2 & -\sigma_{XY} \\ -\sigma_{XY} & \sigma_X^2 \end{bmatrix}.$$

Consequently,

$$\begin{aligned}
(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= \frac{1}{\left(1-\rho^2\right)\sigma_X^2\sigma_Y^2}\Big((x-\mu_X)^2\sigma_Y^2 - 2(x-\mu_X)(y-\mu_Y)\sigma_{XY} + (y-\mu_Y)^2\sigma_X^2\Big) \\
&= \frac{1}{\left(1-\rho^2\right)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\} \\
&= Q(x, y).
\end{aligned}$$

Referring now to Definition 10.4 on page 610, we see that

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2}Q(x,y)} = \frac{1}{2\pi(\det \mathbf{\Sigma})^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

d) Let X be a normal random variable. Define the following 1×1 vectors and matrix: $\mathbf{x} = [x]$, $\boldsymbol{\mu} = [\mu_X]$, and $\boldsymbol{\Sigma} = [\mathrm{Var}(X)]$. Then

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_X}\, e^{-(x-\mu_X)^2/2\sigma_X^2} = \frac{1}{\sqrt{2\pi}\big(\mathrm{Var}(X)\big)^{1/2}}\, e^{-\frac{1}{2}(x-\mu_X)\left(\mathrm{Var}(X)\right)^{-1}(x-\mu_X)}$$

$$= \frac{1}{\sqrt{2\pi}(\det \boldsymbol{\Sigma})^{1/2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

e) Define $\boldsymbol{\Sigma} = \big[\mathrm{Cov}\big(X_i, X_j\big)\big]$, an $m \times m$ matrix, and let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_m} \end{bmatrix}.$$

Then, in view of parts (c) and (d), an educated guess for the form of the joint PDF of an m-variate normal distribution is

$$f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \frac{1}{(2\pi)^{m/2}(\det \boldsymbol{\Sigma})^{1/2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

Review Exercises for Chapter 10

Basic Exercises

10.116 Let $T \sim \mathcal{T}(-1, 1)$. Then $f_T(t) = 1 - |t|$ if $|t| < 1$, and $f_T(t) = 0$ otherwise. By symmetry, we have $\mathcal{E}(T) = 0$. Also,

$$\mathrm{Var}(T) = \mathcal{E}\big(T^2\big) = \int_{-1}^{1} t^2 f_T(t)\, dt = 2\int_0^1 t^2(1-t)\, dt = \frac{1}{6}.$$

Now suppose that $X \sim \mathcal{T}(a, b)$. From Exercise 8.125, we can write $X = (a+b)/2 + (b-a)T/2$, where $T \sim \mathcal{T}(-1, 1)$. Therefore,

$$\mathcal{E}(X) = \mathcal{E}\left(\frac{a+b}{2} + \frac{b-a}{2}T\right) = \frac{a+b}{2} + \frac{b-a}{2}\mathcal{E}(T) = \frac{a+b}{2} + \frac{b-a}{2}\cdot 0 = \frac{a+b}{2},$$

and

$$\mathrm{Var}(X) = \mathrm{Var}\left(\frac{a+b}{2} + \frac{b-a}{2}T\right) = \left(\frac{b-a}{2}\right)^2 \mathrm{Var}(T) = \frac{(b-a)^2}{4}\cdot\frac{1}{6} = \frac{(b-a)^2}{24}.$$

10.117
a) Let $g(x) = \beta e^x$. We have

$$\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx = \int_0^{\infty} \beta e^x \alpha e^{-\alpha x}\, dx = \alpha\beta \int_0^{\infty} e^{-(\alpha-1)x}\, dx.$$

Applying the FEF, we conclude that Y has finite expectation if and only if $\alpha > 1$.
b) In Exercise 8.128(b), we found that $f_Y(y) = \alpha\beta^\alpha / y^{\alpha+1}$ if $y > \beta$, and $f_Y(y) = 0$ otherwise. Hence, by the definition of expected value,

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_{\beta}^{\infty} y \frac{\alpha\beta^\alpha}{y^{\alpha+1}}\, dy = \alpha\beta^\alpha \int_\beta^\infty y^{-\alpha}\, dy = \alpha\beta^\alpha \cdot \frac{1}{(\alpha-1)\beta^{\alpha-1}} = \frac{\alpha\beta}{\alpha-1}.$$

c) From the FEF,

$$\mathcal{E}(Y) = \mathcal{E}\big(\beta e^X\big) = \int_0^\infty \beta e^x f_X(x)\,dx = \int_0^\infty \beta e^x \alpha e^{-\alpha x}\,dx = \alpha\beta \int_0^\infty e^{-(\alpha-1)x}\,dx = \frac{\alpha\beta}{\alpha-1}.$$

d) Clearly, $P(Y > y) = 1$ if $y \le \beta$. For $y > \beta$,

$$P(Y > y) = \int_y^\infty f_Y(t)\,dt = \int_y^\infty \frac{\alpha\beta^\alpha}{t^{\alpha+1}}\,dt = \alpha\beta^\alpha \int_y^\infty t^{-(\alpha+1)}\,dt = \frac{\alpha\beta^\alpha}{\alpha y^\alpha} = \frac{\beta^\alpha}{y^\alpha}.$$

Therefore, from Proposition 10.5 on page 577,

$$\begin{aligned}\mathcal{E}(Y) &= \int_0^\infty P(Y > y)\,dy = \int_0^\beta 1\,dy + \int_\beta^\infty \frac{\beta^\alpha}{y^\alpha}\,dy \\ &= \beta + \beta^\alpha \int_\beta^\infty y^{-\alpha}\,dy = \beta + \frac{\beta^\alpha}{(\alpha-1)\beta^{\alpha-1}} = \frac{\alpha\beta}{\alpha-1}.\end{aligned}$$

e) Let Y denote loss amount. Then Y has the Pareto distribution with parameters $\alpha = 2$ and $\beta = 3$. Therefore,

$$\mathcal{E}(Y) = \frac{\alpha\beta}{\alpha-1} = \frac{2\cdot 3}{2-1} = 6.$$

Hence, the expected loss amount is \$6000.

10.118

a) From the FEF,

$$\begin{aligned}\mathcal{E}(Y) &= \mathcal{E}\big(X^{1/\beta}\big) = \int_{-\infty}^\infty x^{1/\beta} f_X(x)\,dx = \int_0^\infty x^{1/\beta}\alpha\, e^{-\alpha x}\,dx \\ &= \alpha \int_0^\infty x^{(1/\beta+1)-1}\, e^{-\alpha x}\,dx = \alpha\cdot\frac{\Gamma(1/\beta+1)}{\alpha^{(1/\beta+1)}} = \frac{\Gamma(1+1/\beta)}{\alpha^{1/\beta}}.\end{aligned}$$

Also,

$$\begin{aligned}\mathcal{E}\big(Y^2\big) &= \mathcal{E}\Big(\big(X^{1/\beta}\big)^2\Big) = \int_{-\infty}^\infty \Big(x^{1/\beta}\Big)^2 f_X(x)\,dx = \int_0^\infty x^{2/\beta}\alpha\, e^{-\alpha x}\,dx \\ &= \alpha \int_0^\infty x^{(2/\beta+1)-1}\, e^{-\alpha x}\,dx = \alpha\cdot\frac{\Gamma(2/\beta+1)}{\alpha^{(2/\beta+1)}} = \frac{\Gamma(1+2/\beta)}{\alpha^{2/\beta}}.\end{aligned}$$

Therefore,

$$\operatorname{Var}(Y) = \mathcal{E}\big(Y^2\big) - \big(\mathcal{E}(Y)\big)^2 = \frac{\Gamma(1+2/\beta)}{\alpha^{2/\beta}} - \left(\frac{\Gamma(1+1/\beta)}{\alpha^{1/\beta}}\right)^2 = \frac{\Gamma(1+2/\beta) - \big(\Gamma(1+1/\beta)\big)^2}{\alpha^{2/\beta}}.$$

b) Let Y denote the lifetime of the component. Then Y has the Weibull distribution with parameters $\alpha = 2$ and $\beta = 3$. Hence, from part (a),

$$\mathcal{E}(Y) = \frac{\Gamma(1+1/\beta)}{\alpha^{1/\beta}} = \frac{\Gamma(4/3)}{2^{1/3}}.$$

The mean lifetime is approximately 0.71 year. Also,

$$\operatorname{Var}(Y) = \frac{\Gamma(1+2/\beta) - \big(\Gamma(1+1/\beta)\big)^2}{\alpha^{2/\beta}} = \frac{\Gamma(5/3) - \big(\Gamma(4/3)\big)^2}{2^{2/3}}.$$

The standard deviation of the lifetime is approximately $\sqrt{0.066} = 0.26$ year.

10.119

a) From the FEF,

$$\begin{aligned}\mathcal{E}(Y) = \mathcal{E}\big(e^X\big) &= \int_{-\infty}^{\infty} e^x f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\left((x-\mu)^2/2\sigma^2 - x\right)}.\end{aligned}$$

Doing some algebra and completing the square we find that

$$(x-\mu)^2/2\sigma^2 - x = \frac{\left(x - (\mu+\sigma^2)\right)^2}{2\sigma^2} - \left(\mu + \frac{\sigma^2}{2}\right).$$

Because the PDF of a $\mathcal{N}\left(\mu+\sigma^2, \sigma^2\right)$ distribution integrates to 1, we conclude that $\mathcal{E}(Y) = e^{\mu+\sigma^2/2}$. Also,

$$\begin{aligned}\mathcal{E}\big(Y^2\big) = \mathcal{E}\Big(\big(e^X\big)^2\Big) &= \int_{-\infty}^{\infty} \left(e^x\right)^2 f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} e^{2x} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\left((x-\mu)^2/2\sigma^2 - 2x\right)}.\end{aligned}$$

Doing some algebra and completing the square we find that

$$(x-\mu)^2/2\sigma^2 - 2x = \frac{\left(x - (\mu+2\sigma^2)\right)^2}{2\sigma^2} - 2\left(\mu+\sigma^2\right).$$

Because the PDF of a $\mathcal{N}\left(\mu+2\sigma^2, \sigma^2\right)$ distribution integrates to 1, we conclude that $\mathcal{E}\big(Y^2\big) = e^{2(\mu+\sigma^2)}$. Consequently,

$$\operatorname{Var}(Y) = \mathcal{E}\big(Y^2\big) - \big(\mathcal{E}(Y)\big)^2 = e^{2(\mu+\sigma^2)} - \left(e^{\mu+\sigma^2/2}\right)^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right).$$

b) Let Y denote claim amount. We know that Y has the lognormal distribution with parameters $\mu = 7$ and $\sigma^2 = 0.25$. Therefore, from part (a),

$$\mathcal{E}(Y) = e^{\mu+\sigma^2/2} = e^{7.125}.$$

The mean claim amount is approximately \$1242.65. Also,

$$\operatorname{Var}(Y) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) = e^{14.25}\left(e^{0.25} - 1\right).$$

The standard deviation of claim amounts is approximately $\sqrt{438{,}584.796} = \$662.26$.

10.120

a) For the solution to this part, see Example 10.7(b) on page 574, where we found that $\mathcal{E}(S) = 2\sqrt{2/\pi}\,\sigma$.

b) Let $W = \left(X^2+Y^2+Z^2\right)/\sigma^2$. Then, from Propositions 8.15, 8.13, and 9.13 on pages 468, 466, and 537, respectively,

$$W = \frac{X^2+Y^2+Z^2}{\sigma^2} = \left(\frac{X}{\sigma}\right)^2 + \left(\frac{Y}{\sigma}\right)^2 + \left(\frac{Z}{\sigma}\right)^2 \sim \Gamma\left(\frac{3}{2}, \frac{1}{2}\right).$$

Noting that $S = \sigma\sqrt{W}$, we apply the FEF to get

$$\mathcal{E}(S) = \mathcal{E}\left(\sigma\sqrt{W}\right) = \int_{-\infty}^{\infty} \sigma\sqrt{w}\, f_W(w)\, dw = \int_0^{\infty} \sigma\sqrt{w}\, \frac{(1/2)^{3/2}}{\Gamma(3/2)} w^{3/2-1}\, e^{-w/2}\, dw$$
$$= \frac{\sigma}{2^{3/2}\Gamma(3/2)} \int_0^{\infty} w^{2-1}\, e^{-w/2}\, dw = \frac{\sigma}{2^{3/2}\Gamma(3/2)} \cdot \frac{\Gamma(2)}{(1/2)^2} = 2\sqrt{2/\pi}\,\sigma.$$

c) For the solution to this part, see Example 10.7(a) on page 574, where we found that $\mathcal{E}(S) = 2\sqrt{2/\pi}\,\sigma$.

10.121
a) We have

$$\mathcal{E}(Y) = \mathcal{E}\big(X(1-X)\big) = \int_{-\infty}^{\infty} x(1-x) f_X(x)\, dx = \int_0^1 x(1-x)\, dx$$
$$= \int_0^1 x^{2-1}(1-x)^{2-1}\, dx = B(2,2) = \frac{1}{6}.$$

Also,

$$\mathcal{E}(XY) = \mathcal{E}\big(X \cdot X(1-X)\big) = \mathcal{E}\big(X^2(1-X)\big) = \int_{-\infty}^{\infty} x^2(1-x) f_X(x)\, dx = \int_0^1 x^2(1-x)\, dx$$
$$= \int_0^1 x^{3-1}(1-x)^{2-1}\, dx = B(3,2) = \frac{1}{12}.$$

b) We have

$$\mathcal{E}(X^n) = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = \int_0^1 x^n\, dx = \frac{1}{n+1}.$$

Therefore,

$$\mathcal{E}(Y) = \mathcal{E}\big(X(1-X)\big) = \mathcal{E}\big(X - X^2\big) = \mathcal{E}(X) - \mathcal{E}\big(X^2\big) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

and

$$\mathcal{E}(XY) = \mathcal{E}\big(X \cdot X(1-X)\big) = \mathcal{E}\big(X^2 - X^3\big) = \mathcal{E}\big(X^2\big) - \mathcal{E}\big(X^3\big) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

c) Referring to part (b), we find that

$$\mathcal{E}(XY) = \frac{1}{12} = \frac{1}{2} \cdot \frac{1}{6} = \mathcal{E}(X)\,\mathcal{E}(Y).$$

Clearly, X and Y are not independent because, in fact, Y is a function of X.
d) Two random variables can be uncorrelated without being independent.

10.122 The area of the triangle is $1/2$ and, hence, $f_{X,Y}(x, y) = 2$ if (x, y) is in the triangle, and equals 0 otherwise.
a) We have

$$\mathcal{E}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\, dx\, dy = 2\int_0^1 x \left(\int_0^{1-x} 1\, dy\right) dx = 2\int_0^1 x(1-x)\, dx = \frac{1}{3}.$$

By symmetry, $\mathcal{E}(Y) = 1/3$.

b) We have

$$\mathcal{E}(X)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{X,Y}(x,y)\,dx\,dy = 2\int_0^1 x^2\left(\int_0^{1-x} 1\,dy\right)dx = 2\int_0^1 x^2(1-x)\,dx = \frac{1}{6}.$$

Hence, $\operatorname{Var}(X) = (1/6) - (1/3)^2 = 1/18$. By symmetry, $\operatorname{Var}(Y) = 1/18$.
c) Examining the triangle over which X and Y are uniformly distributed, we see that large values (close to 1) of X correspond to small values (close to 0) of Y. Thus, it appears that X and Y are negatively correlated, that is, that the sign of the correlation coefficient is negative.
d) We have

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy = 2\int_0^1 x\left(\int_0^{1-x} y\,dy\right)dx = \int_0^1 x(1-x)^2\,dx = \frac{1}{12}.$$

Thus,

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{1/12 - (1/3)(1/3)}{\sqrt{(1/18)(1/18)}} = -\frac{1}{2}.$$

10.123 Let X denote the manufacturer's annual loss and let Y denote the amount of that loss not paid by the insurance policy. Then

$$Y = \begin{cases} X, & \text{if } X \le 2; \\ 2, & \text{if } X > 2. \end{cases}$$

Let

$$g(x) = \begin{cases} x, & \text{if } x \le 2; \\ 2, & \text{if } x > 2. \end{cases}$$

Then $Y = g(X)$ and, hence, by the FEF,

$$\begin{aligned}\mathcal{E}(Y) = \mathcal{E}\big(g(X)\big) &= \int_{-\infty}^{\infty} g(x) f(x)\,dx = \int_{0.6}^{2} x\cdot 2.5(0.6)^{2.5}x^{-3.5}\,dx + \int_2^{\infty} 2\cdot 2.5(0.6)^{2.5}x^{-3.5}\,dx \\ &= 2.5(0.6)^{2.5}\int_{0.6}^{2} x^{-2.5}\,dx + 5(0.6)^{2.5}\int_2^{\infty} x^{-3.5}\,dx = 0.934.\end{aligned}$$

10.124 Note that the time it takes to complete the two successive tasks is Y.
a) We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^y 6e^{-(x+2y)}\,dx = 6e^{-2y}\int_0^y e^{-x}\,dx = 6e^{-2y}\left(1 - e^{-y}\right)$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. Hence, by the definition of expected value,

$$\begin{aligned}\mathcal{E}(Y) &= \int_{-\infty}^{\infty} y f_Y(y)\,dy = 6\int_0^{\infty} ye^{-2y}\left(1-e^{-y}\right)dy = 6\left(\int_0^{\infty} ye^{-2y}\,dy - \int_0^{\infty} ye^{-3y}\,dy\right) \\ &= 6\left(\int_0^{\infty} y^{2-1}e^{-2y}\,dy - \int_0^{\infty} y^{2-1}e^{-3y}\,dy\right) = 6\left(\frac{\Gamma(2)}{2^2} - \frac{\Gamma(2)}{3^2}\right) = \frac{5}{6}.\end{aligned}$$

b) Applying the FEF, we get

$$\begin{aligned}
\mathcal{E}(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dx\,dy = 6\int_0^{\infty} y\left(\int_0^y e^{-(x+2y)}\,dx\right)dy \\
&= 6\int_0^{\infty} ye^{-2y}\left(\int_0^y e^{-x}\,dx\right)dy = 6\int_0^{\infty} ye^{-2y}\left(1-e^{-y}\right)dy \\
&= 6\left(\int_0^{\infty} ye^{-2y}\,dy - \int_0^{\infty} ye^{-3y}\,dy\right) = 6\left(\int_0^{\infty} y^{2-1}e^{-2y}\,dy - \int_0^{\infty} y^{2-1}e^{-3y}\,dy\right) \\
&= 6\left(\frac{\Gamma(2)}{2^2} - \frac{\Gamma(2)}{3^2}\right) = \frac{5}{6}.
\end{aligned}$$

10.125 Let X denote printer lifetime, in years. Then $X \sim \mathcal{E}(1/2)$. Also, let Y denote the refund amount, in hundreds of dollars, for one printer. We have,

$$Y = \begin{cases} 2, & \text{if } X \le 1; \\ 1, & \text{if } 1 < X \le 2; \\ 0, & \text{if } X > 2. \end{cases}$$

Define

$$g(x) = \begin{cases} 2, & \text{if } x \le 1; \\ 1, & \text{if } 1 < x \le 2; \\ 0, & \text{if } x > 2. \end{cases}$$

Then $Y = g(X)$ and, hence, by the FEF,

$$\begin{aligned}
\mathcal{E}(Y) = \mathcal{E}\big(g(X)\big) &= \int_{-\infty}^{\infty} g(x) f_X(x)\,dx = \int_0^1 2\cdot f_X(x)\,dx + \int_1^2 1\cdot f_X(x)\,dx \\
&= 2P(0 < X \le 1) + P(1 < X \le 2) = 2\left(1-e^{-1/2}\right) + \left(e^{-1/2}-e^{-1}\right) \\
&= 2 - e^{-1/2} - e^{-1}.
\end{aligned}$$

Therefore, the expected refund, in hundreds of dollars, for 100 printers is $100\left(2-e^{-1/2}-e^{-1}\right)$, or about \$10,255.90.

10.126 Let X denote the time to failure, in years, for the device. We know that $X \sim \mathcal{E}(1/3)$. Let Y denote the time to discovery of the device's failure. Then

$$Y = \begin{cases} 2, & \text{if } X \le 2; \\ X, & \text{if } X > 2. \end{cases}$$

Let

$$g(x) = \begin{cases} 2, & \text{if } x \le 2; \\ x, & \text{if } x > 2. \end{cases}$$

Then $Y = g(X)$ and, hence, by the FEF,

$$\begin{aligned}
\mathcal{E}(Y) = \mathcal{E}\big(g(X)\big) &= \int_{-\infty}^{\infty} g(x) f_X(x)\,dx = \int_0^2 2\cdot(1/3)e^{-x/3}\,dx + \int_2^{\infty} x\cdot(1/3)e^{-x/3}\,dx \\
&= 2P(0 < X \le 2) + \frac{1}{3}\int_2^{\infty} x\,e^{-x/3}\,dx = 2\left(1-e^{-2/3}\right) + \frac{1}{3}\cdot 15e^{-2/3} = 2 + 3e^{-2/3}.
\end{aligned}$$

The expected time to discovery of the device's failure is about 3.54 years.

10.127 Let X denote the loss, in thousands of dollars. Then $X \sim \mathcal{U}(0, 1)$. The expected payment without a deductible is $\mathcal{E}(X) = 0.5$. Let Y denote the payment with a deductible of c thousand dollars. Then

$$Y = \begin{cases} 0, & \text{if } X \le c; \\ X - c, & \text{if } X > c. \end{cases}$$

From the FEF,

$$\mathcal{E}(Y) = \int_{-\infty}^{c} 0 \cdot f_X(x)\,dx + \int_{c}^{\infty} (x-c) f_X(x)\,dx = \int_{c}^{1} (x-c)\,dx = (1-c)^2/2.$$

We want to choose c so that $(1-c)^2/2 = 0.25 \cdot 0.5$, which means that $c = 0.5$. Thus, the deductible should be set at \$500.

10.128 Let X and Y denote the lifetimes, in years, of the two generators. By assumption, both of these random variables have an $\mathcal{E}(1/10)$ distribution, and we assume, as reasonable, that they are independent random variables. The total time that the two generators produce electricity is $X + Y$. We have

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) = \frac{1}{(1/10)^2} + \frac{1}{(1/10)^2} = 200.$$

Consequently, the standard deviation of the total time that the generators produce electricity is $\sqrt{200}$, or about 14.1 years.

10.129 Let X and Y be independent standard normal random variables. Then X and Y have finite moments of all orders and, in particular, finite expectation. From Example 9.23 on page 541, the random variable Y/X has the standard Cauchy distribution, which, by Example 10.5 on page 568, doesn't have finite expectation.

10.130 Let X denote loss amount and let Y denote processing time. We know that

$$f_X(x) = \begin{cases} 3x^2/8, & \text{if } 0 < x < 2; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_{Y\,|\,X}(y\,|\,x) = \begin{cases} 1/x, & \text{if } x < y < 2x; \\ 0, & \text{otherwise.} \end{cases}$$

a) From the general multiplication rule,

$$f_{X,Y}(x,y) = f_X(x) f_{Y\,|\,X}(y\,|\,x) = \frac{3x^2}{8} \cdot \frac{1}{x} = \frac{3x}{8}$$

if $0 < x < y < 2x < 4$, and $f_{X,Y}(x,y) = 0$ otherwise. Hence,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{y/2}^{\min\{y,2\}} \frac{3x}{8}\,dx = \begin{cases} 9y^2/64, & \text{if } 0 < y < 2; \\ 3\left(16 - y^2\right)/64, & \text{if } 2 < y < 4. \end{cases}$$

and $f_Y(y) = 0$ otherwise. Consequently,

$$\begin{aligned} \mathcal{E}(Y) &= \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_0^2 y \cdot \frac{9y^2}{64}\,dy + \int_2^4 y \cdot \frac{3\left(16-y^2\right)}{64}\,dy \\ &= \frac{9}{64}\int_0^2 y^3\,dy + \frac{3}{64}\int_2^4 y\left(16-y^2\right)\,dy = \frac{9}{64}\cdot 4 + \frac{3}{64}\cdot 36 = \frac{9}{4}. \end{aligned}$$

b) Applying the law of total expectation and recalling that $Y_{|X=x} \sim \mathcal{U}(x, 2x)$ for $0 < x < 2$, we get

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y\,|\,X)\big) = \mathcal{E}\left(\frac{3}{2}X\right) = \frac{3}{2}\mathcal{E}(X) = \frac{3}{2}\int_0^2 x \cdot \frac{3x^2}{8}\,dx = \frac{9}{16}\int_0^2 x^3\,dx = \frac{9}{4}.$$

10.131 Let X and Y denote the remaining lifetimes of the husband and wife, respectively. By assumption, X and Y are independent $\mathcal{U}(0, 40)$ random variables. The problem is to find $\operatorname{Cov}(X,\ \max\{X, Y\})$. Now $f_{X,Y}(x, y) = (1/40) \cdot (1/40) = 1/1600$ if $0 < x < 40$ and $0 < y < 40$, and equals 0 otherwise. We know that $\mathcal{E}(X) = 20$. From the FEF and symmetry,

$$\mathcal{E}\big(\max\{X, Y\}\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\{x, y\} f_{X,Y}(x, y)\, dx\, dy = \frac{2}{1600}\int_0^{40} x \left(\int_0^x 1\, dy\right) dx = \frac{80}{3}.$$

Also,

$$\begin{aligned}
\mathcal{E}\big(X \max\{X, Y\}\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \max\{x, y\} f_{X,Y}(x, y)\, dx\, dy \\
&= \frac{1}{1600}\int_0^{40} x^2 \left(\int_0^x 1\, dy\right) dx + \frac{1}{1600}\int_0^{40} x \left(\int_x^{40} y\, dy\right) dx \\
&= \frac{1}{1600}\int_0^{40} x^3\, dx + \frac{1}{3200}\int_0^{40} x \left(1600 - x^2\right)\, dx = 400 + 200 = 600.
\end{aligned}$$

Consequently,

$$\operatorname{Cov}\big(X,\ \max\{X, Y\}\big) = \mathcal{E}\big(X \max\{X, Y\}\big) - \mathcal{E}(X)\,\mathcal{E}\big(\max\{X, Y\}\big) = 600 - 20 \cdot \frac{80}{3} = \frac{200}{3}.$$

10.132

a) Because the PDF of X is symmetric about 0, all odd moments of X equal 0. In particular, then, we have $\mathcal{E}(X) = \mathcal{E}\big(X^3\big) = 0$. Therefore,

$$\mathcal{E}(XY) = \mathcal{E}\big(X \cdot X^2\big) = \mathcal{E}\big(X^3\big) = 0 = \mathcal{E}(X)\,\mathcal{E}(Y)\,.$$

b) No, although independence of two random variables is a sufficient condition for having the expected value of a product equal the product of the expected values, it is not a necessary condition.

c) No, X and Y are not independent random variables; in fact, Y is a function of X.

10.133 For positive x and y, we can write

$$f_{X,Y}(x, y) = \frac{1}{64} y\, e^{-(x+y)/4} = \frac{1}{4}e^{-x/4} \cdot \frac{1}{16} y\, e^{-y/4}.$$

This result shows that $X \sim \mathcal{E}(1/4)$, $Y \sim \Gamma(2, 1/4)$, and that X and Y are independent random variables.

a) Let T denote the lifetime of the device. Then $T = \max\{X, Y\}$. Referring to Equation (8.49) on page 453, we see that, for $t \geq 0$,

$$\begin{aligned}
F_T(t) &= P(T \leq t) = P\big(\max\{X, Y\} \leq t\big) = P(X \leq t,\ Y \leq t) = P(X \leq t)P(Y \leq t) \\
&= \left(1 - e^{-t/4}\right)\left(1 - e^{-t/4} - (t/4)e^{-t/4}\right) = 1 - 2e^{-t/4} + e^{-t/2} - (t/4)e^{-t/4} + (t/4)e^{-t/2}.
\end{aligned}$$

Differentiation now yields

$$f_T(t) = \frac{1}{4}e^{-t/4} - \frac{1}{4}e^{-t/2} + \frac{1}{16}te^{-t/4} - \frac{1}{8}te^{-t/2}$$

if $t > 0$, and $f_T(t) = 0$ otherwise.

Therefore,

$$\begin{aligned}
\mathcal{E}(T) &= \int_{-\infty}^{\infty} t f_T(t)\,dt \\
&= \frac{1}{4}\int_0^{\infty} te^{-t/4}\,dt - \frac{1}{4}\int_0^{\infty} te^{-t/2}\,dt + \frac{1}{16}\int_0^{\infty} t^2e^{-t/4}\,dt - \frac{1}{8}\int_0^{\infty} t^2e^{-t/2}\,dt \\
&= \frac{1}{4}\cdot\frac{\Gamma(2)}{(1/4)^2} - \frac{1}{4}\cdot\frac{\Gamma(2)}{(1/2)^2} + \frac{1}{16}\cdot\frac{\Gamma(3)}{(1/4)^3} - \frac{1}{8}\cdot\frac{\Gamma(3)}{(1/2)^3} = 9.
\end{aligned}$$

Thus, the expected lifetime of the device is 9000 hours.

b) Applying the FEF, we get

$$\begin{aligned}
\mathcal{E}(T) = \mathcal{E}\big(\max\{X, Y\}\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \max\{x, y\} f_{X,Y}(x, y)\,dx\,dy \\
&= \frac{1}{4}\int_0^{\infty} x\,e^{-x/4}\left(\int_0^x \frac{1}{16}y\,e^{-y/4}\,dy\right)dx + \frac{1}{16}\int_0^{\infty} y^2\,e^{-y/4}\left(\int_0^y \frac{1}{4}e^{-x/4}\,dx\right)dy \\
&= \frac{1}{4}\int_0^{\infty} x\,e^{-x/4}\left(1 - e^{-x/4} - (x/4)e^{-x/4}\right)dx + \frac{1}{16}\int_0^{\infty} y^2\,e^{-y/4}\left(1 - e^{-y/4}\right)dy \\
&= \frac{1}{4}\int_0^{\infty} xe^{-x/4}\,dx - \frac{1}{4}\int_0^{\infty} xe^{-x/2}\,dx - \frac{1}{16}\int_0^{\infty} x^2e^{-x/2}\,dx \\
&\quad + \frac{1}{16}\int_0^{\infty} y^2e^{-y/4}\,dy - \frac{1}{16}\int_0^{\infty} y^2e^{-y/2}\,dy \\
&= \frac{1}{4}\cdot\frac{\Gamma(2)}{(1/4)^2} - \frac{1}{4}\cdot\frac{\Gamma(2)}{(1/2)^2} - \frac{1}{16}\cdot\frac{\Gamma(3)}{(1/2)^3} + \frac{1}{16}\cdot\frac{\Gamma(3)}{(1/4)^3} - \frac{1}{16}\cdot\frac{\Gamma(3)}{(1/2)^3} = 9.
\end{aligned}$$

Thus, the expected lifetime of the device is 9000 hours.

c) Let W denote the cost of operating the device until its failure. Then $W = 65 + 3X + 5Y$. Recalling that $X \sim \mathcal{E}(1/4)$ and $Y \sim \Gamma(2, 1/4)$, we get

$$\mathcal{E}(W) = \mathcal{E}(65 + 3X + 5Y) = 65 + 3\mathcal{E}(X) + 5\mathcal{E}(Y) = 65 + 3\cdot\frac{1}{1/4} + 5\cdot\frac{2}{1/4} = 117.$$

Thus, the expected cost of operating the device until its failure is \$117.

d) Referring to part (c) and using the independence of X and Y, we get

$$\operatorname{Var}(W) = \operatorname{Var}(65 + 3X + 5Y) = 9\operatorname{Var}(X) + 25\operatorname{Var}(Y) = 9\cdot\frac{1}{(1/4)^2} + 25\cdot\frac{2}{(1/4)^2} = 944.$$

Thus, the standard deviation of the cost of operating the device until its failure is $\sqrt{944}$, or about \$30.72.

10.134

a) We have

$$f_X(x) = \begin{cases} 1/12, & \text{if } 0 < x < 12; \\ 0, & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_{Y\,|\,X}(y\,|\,x) = \begin{cases} 1/x, & \text{if } 0 < y < x; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, from the general multiplication rule,

$$f_{X,Y}(x, y) = f_X(x) f_{Y\,|\,X}(y\,|\,x) = \frac{1}{12}\cdot\frac{1}{x} = \frac{1}{12x}$$

if $0 < y < x < 12$, and $f_{X,Y}(x, y) = 0$ otherwise.

We know that $\mathcal{E}(X) = 6$. Applying the FEF yields

$$\mathcal{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x, y)\,dx\,dy = \frac{1}{12}\int_0^{12}\frac{1}{x}\left(\int_0^x y\,dy\right)dx = \frac{1}{24}\int_0^{12} x\,dx = 3$$

and

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x, y)\,dx\,dy = \frac{1}{12}\int_0^{12} x\cdot\frac{1}{x}\left(\int_0^x y\,dy\right)dx = \frac{1}{24}\int_0^{12} x^2\,dx = 24.$$

Consequently,

$$\operatorname{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = 24 - 6\cdot 3 = 6.$$

b) Recalling that $X \sim \mathcal{U}(0, 12)$ and $Y_{|X=x} \sim \mathcal{U}(0, x)$, we apply the law of total probability to get

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\left(\frac{X}{2}\right) = \frac{1}{2}\mathcal{E}(X) = \frac{1}{2}\cdot 6 = 3$$

and, also referring to Equation (10.49) on page 602,

$$\begin{aligned}\mathcal{E}(XY) &= \mathcal{E}\big(\mathcal{E}(XY \mid X)\big) = \mathcal{E}\big(X\mathcal{E}(Y \mid X)\big) = \mathcal{E}\left(X\cdot\frac{X}{2}\right) = \frac{1}{2}\mathcal{E}\big(X^2\big)\\ &= \frac{1}{2}\Big(\operatorname{Var}(X) + \big(\mathcal{E}(X)\big)^2\Big) = \frac{1}{2}\left(\frac{(12-0)^2}{12} + 6^2\right) = 24.\end{aligned}$$

Consequently,

$$\operatorname{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = 24 - 6\cdot 3 = 6.$$

10.135

a) From the FEF,

$$\begin{aligned}\mathcal{E}(Y^r) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^r f_{X,Y}(x, y)\,dx\,dy = 8\int_0^1 x\left(\int_0^x y^{r+1}\,dy\right)dx\\ &= \frac{8}{r+2}\int_0^1 x^{r+3}\,dx = \frac{8}{r+2}\cdot\frac{1}{r+4} = \frac{8}{(r+2)(r+4)}.\end{aligned}$$

Therefore,

$$\mathcal{E}(Y) = \frac{8}{(1+2)(1+4)} = \frac{8}{15}$$

and

$$\operatorname{Var}(Y) = \mathcal{E}\big(Y^2\big) - \big(\mathcal{E}(Y)\big)^2 = \frac{8}{(2+2)(2+4)} - \left(\frac{8}{15}\right)^2 = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}.$$

b) We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = 8y\int_y^1 x\,dx = 4y\left(1 - y^2\right)$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. From the FEF,

$$\begin{aligned}\mathcal{E}(Y^r) &= \int_{-\infty}^{\infty} y^r f_Y(y)\,dy = 4\int_0^1 y^{r+1}\left(1 - y^2\right)dy = 4\int_0^1 \left(y^{r+1} - y^{r+3}\right)dy\\ &= 4\left(\frac{1}{r+2} - \frac{1}{r+4}\right) = \frac{8}{(r+2)(r+4)}.\end{aligned}$$

Therefore,

$$\mathcal{E}(Y) = \frac{8}{(1+2)(1+4)} = \frac{8}{15}$$

and

$$\operatorname{Var}(Y) = \mathcal{E}\big(Y^2\big) - \big(\mathcal{E}(Y)\big)^2 = \frac{8}{(2+2)(2+4)} - \left(\frac{8}{15}\right)^2 = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}.$$

c) We have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = 8x \int_0^x y\,dy = 4x^3$$

if $0 < x < 1$, and $f_X(x) = 0$ otherwise. We see that $X \sim \text{Beta}(4, 1)$. Therefore,

$$\mathcal{E}(X) = \frac{4}{4+1} = \frac{4}{5} \qquad \text{and} \qquad \operatorname{Var}(X) = \frac{4 \cdot 1}{(4+1)^2(4+1+1)} = \frac{2}{75}.$$

Now, for $0 < x < 1$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2}$$

if $0 < y < x$, and $f_{Y\,|\,X}(y\,|\,x) = 0$ otherwise. Therefore,

$$\mathcal{E}(Y^r \mid X = x) = \int_{-\infty}^{\infty} y^r f_{Y\,|\,X}(y\,|\,x)\,dy = \frac{2}{x^2}\int_0^x y^{r+1}\,dy = \frac{2x^r}{r+2}.$$

Hence,

$$\mathcal{E}(Y \mid X) = \frac{2}{3}X \qquad \text{and} \qquad \operatorname{Var}(Y \mid X) = \frac{1}{2}X^2 - \left(\frac{2}{3}X\right)^2 = \frac{1}{18}X^2.$$

Applying now the laws of total expectation and total variance, we get

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y\,|\,x)\big) = \mathcal{E}\left(\frac{2}{3}X\right) = \frac{2}{3}\mathcal{E}(X) = \frac{2}{3}\cdot\frac{4}{5} = \frac{8}{15}$$

and

$$\begin{aligned}
\operatorname{Var}(Y) &= \mathcal{E}\big(\operatorname{Var}(Y \mid X)\big) + \operatorname{Var}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}\left(\frac{1}{18}X^2\right) + \operatorname{Var}\left(\frac{2}{3}X\right) = \frac{1}{18}\mathcal{E}\big(X^2\big) + \frac{4}{9}\operatorname{Var}(X) \\
&= \frac{1}{18}\Big(\operatorname{Var}(X) + \big(\mathcal{E}(X)\big)^2\Big) + \frac{4}{9}\operatorname{Var}(X) = \frac{1}{18}\cdot\big(\mathcal{E}(X)\big)^2 + \frac{1}{2}\cdot\operatorname{Var}(X) \\
&= \frac{1}{18}\cdot\left(\frac{4}{5}\right)^2 + \frac{1}{2}\cdot\frac{2}{75} = \frac{11}{225}.
\end{aligned}$$

10.136

a) Referring to Exercise 10.135(c), we conclude from the prediction theorem, Proposition 10.8 on page 602, that the best predictor of the amount sold, given that the amount stocked is 0.5, is

$$\mathcal{E}(Y \mid X = 0.5) = \frac{2}{3}\cdot 0.5 = \frac{1}{3}.$$

b) Here we want $\mathcal{E}(X \mid Y = 0.4)$. Referring to Exercise 10.135, we see that

$$f_{X \mid Y}(x \mid 0.4) = \frac{f_{X,Y}(x, 0.4)}{f_Y(0.4)} = \frac{8x \cdot (0.4)}{4 \cdot (0.4)\left(1 - (0.4)^2\right)} = \frac{50x}{21}$$

if $0.4 < x < 1$, and $f_{X \mid Y}(x \mid 0.4) = 0$ otherwise. Therefore,

$$\mathcal{E}(X \mid Y = 0.4) = \int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid 0.4)\, dx = \frac{50}{21} \int_{0.4}^{1} x^2\, dx = \frac{26}{35}.$$

c) The amount left unsold is $X - Y$, so here we want $\mathcal{E}(Y \mid X - Y = 0.1)$. We can obtain that conditional expectation as follows. First we use the bivariate transformation theorem to show that

$$f_{Y,X-Y}(u, v) = \begin{cases} 8u(u+v), & \text{if } u > 0,\ v > 0, \text{ and } u + v < 1; \\ 0, & \text{otherwise.} \end{cases}$$

From that, we get

$$f_{X-Y}(v) = \int_{-\infty}^{\infty} f_{Y,X-Y}(u, v)\, du = 8 \int_{0}^{1-v} u(u+v)\, du = \frac{4}{3}(1-v)^2(2+v)$$

if $0 < v < 1$, and $f_{X-Y}(v) = 0$ otherwise. Therefore,

$$f_{X-Y}(0.1) = \frac{4}{3}(1 - 0.1)^2(2 + 0.1) = \frac{567}{250}$$

and

$$f_{Y \mid X-Y}(u \mid 0.1) = \frac{f_{Y,X-Y}(u, 0.1)}{f_{X-Y}(0.1)} = \frac{250}{567} \cdot 8u(u + 0.1) = \frac{2000}{567} u(u + 0.1)$$

if $0 < u < 0.9$, and $f_{Y \mid X-Y}(u \mid 0.1) = 0$ otherwise. Consequently,

$$\mathcal{E}(Y \mid X - Y = 0.1) = \int_{-\infty}^{\infty} u f_{Y \mid X-Y}(u \mid 0.1)\, du = \frac{2000}{567} \int_{0}^{0.9} u^2(u + 0.1)\, du = \frac{93}{140}.$$

10.137 For $n \geq 2$, we have $X_{n \mid X_{n-1} = x_{n-1}} \sim \mathcal{U}(0, x_{n-1})$ and, hence, $\mathcal{E}(X_n \mid X_{n-1} = x_{n-1}) = x_{n-1}/2$. Thus, by the law of total expectation,

$$\mathcal{E}(X_n) = \mathcal{E}\big(\mathcal{E}(X_n \mid X_{n-1})\big) = \mathcal{E}\left(\frac{X_{n-1}}{2}\right) = \frac{1}{2}\mathcal{E}(X_{n-1}).$$

Consequently,

$$\mathcal{E}(X_n) = \mathcal{E}(X_1)\left(\prod_{k=2}^{n} \frac{\mathcal{E}(X_k)}{\mathcal{E}(X_{k-1})}\right) = \frac{1}{2} \cdot \left(\frac{1}{2}\right)^{n-1} = \frac{1}{2^n}.$$

10.138

a) We have

$$\begin{aligned} 2\sigma_X \sigma_Y \sqrt{1 - \rho^2} &= 2 \cdot 1.4 \cdot 31.2 \sqrt{1 - (-0.9)^2} = 87.36\sqrt{0.19}, \\ 2\left(1 - \rho^2\right) &= 2\left(1 - (-0.9)^2\right) = 2(0.19) = 0.38, \\ 2\rho &= 2 \cdot (-0.9) = -0.18. \end{aligned}$$

Therefore, from Definition 10.4 on page 610,

$$f_{X,Y}(x, y) = \frac{1}{87.36\sqrt{0.19}\,\pi}\, e^{-2.63\left\{\left(\frac{x-5.3}{1.4}\right)^2 + 0.18\left(\frac{x-5.3}{1.4}\right)\left(\frac{y-88.6}{31.2}\right) + \left(\frac{y-88.6}{31.2}\right)^2\right\}}.$$

b) From Proposition 10.10(a) on page 616,

$$X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right) = \mathcal{N}\left(5.3, 1.4^2\right).$$

The ages of Orions are normally distributed with a mean of 5.3 years and a standard deviation of 1.4 years.
c) From Proposition 10.10(b),

$$Y \sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right) = \mathcal{N}\left(88.6, 31.2^2\right).$$

The prices of Orions are normally distributed with a mean of \$8860 and a standard deviation of \$3120.
d) From Proposition 10.10(c), $\rho(X, Y) = \rho = -0.9$. There is a strong negative correlation between age and price of Orions.
e) We have

$$\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 88.6 - 0.9 \cdot \frac{31.2}{1.4}(x - 5.3) = 194.90 - 20.06x$$

and

$$\sigma_Y^2\left(1 - \rho^2\right) = 31.2^2\left(1 - (-0.9)^2\right) = 184.95.$$

Hence, from Proposition 10.10(d), $Y_{|X=x} \sim \mathcal{N}(194.90 - 20.06x, 184.95)$. The prices of x-year-old Orions are normally distributed with a mean of \$$(19490 - 2006x)$ and a standard deviation of \$1360.
f) From part (e), the regression equation is $y = 194.90 - 20.06x$.
g) Referring to part (f), the predicted price of a 3-year-old Orion is $y = 194.90 - 20.06 \cdot 3 = 134.72$, or \$13,472.

10.139 We have

$$\mu_X = 3.0, \qquad \sigma_X = 0.5, \qquad \mu_Y = 2.7, \qquad \sigma_Y = 0.6, \qquad \rho = 0.5.$$

a) From the preceding display,

$$\begin{aligned} 2\sigma_X\sigma_Y\sqrt{1-\rho^2} &= 2 \cdot 0.5 \cdot 0.6\sqrt{1-(0.5)^2} = 0.3\sqrt{3}, \\ 2\left(1-\rho^2\right) &= 2\left(1-(0.5)^2\right) = 2(0.75) = 1.5, \\ 2\rho &= 2 \cdot (0.5) = 1. \end{aligned}$$

Therefore, from Definition 10.4 on page 610,

$$f_{X,Y}(x, y) = \frac{1}{0.3\sqrt{3}\,\pi}\, e^{-\frac{2}{3}\left\{\left(\frac{x-3.0}{0.5}\right)^2 - \left(\frac{x-3.0}{0.5}\right)\left(\frac{y-2.7}{0.6}\right) + \left(\frac{y-2.7}{0.6}\right)^2\right\}}.$$

b) We have

$$\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 2.7 + 0.5 \cdot \frac{0.6}{0.5}(x - 3.0) = 0.9 + 0.6x$$

Hence, the predicted cumulative GPA at the end of the sophomore year for an ASU student whose high-school GPA was 3.5 is $0.9 + 0.6 \cdot 3.5 = 3.0$.
c) We have

$$\sigma_Y^2\left(1 - \rho^2\right) = 0.6^2\left(1 - (0.5)^2\right) = 0.27.$$

Hence, from Proposition 10.10(d) on page 616 and part (b), $Y_{|X=3.5} \sim \mathcal{N}(3.0, 0.27)$. Referring to Table I on page A-39, we see that the probability a normally distributed random variable takes a value within 1.645 standard deviations of its mean is 0.90. Consequently, the required prediction interval is $3.0 \pm 1.645 \cdot \sqrt{0.27}$, or $(2.15, 3.85)$.

10.140 Let X denote the weekly demand, in hundreds of pounds, for unbleached enriched whole wheat flour and let Y denote the weekly profit, in dollars, due to sales of the flour. We know that $X \sim \text{Beta}(2, 2)$. Note that, because the maximum value of a beta random variable is 1, we can assume that $a < 1$. Now, the amount sold by the end of the week is $\min\{X, a\}$. Therefore,

$$Y = 21 \min\{X, a\} - 10a.$$

We want to choose a to maximize $\mathcal{E}(Y)$. Let

$$g(x) = 21 \min\{x, a\} - 10a = \begin{cases} 21x - 10a, & \text{if } x < a; \\ 11a, & \text{if } x \geq a. \end{cases}$$

Then $Y = g(X)$ and, hence, from the FEF and Equation (8.55) on page 457,

$$\begin{aligned} \mathcal{E}(Y) = \mathcal{E}\big(g(X)\big) &= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx = \int_0^a (21x - 10a) \cdot 6x(1-x)\, dx + \int_a^1 11a \cdot 6x(1-x)\, dx \\ &= 6 \int_0^a x(21x - 10a)(1-x)\, dx + 11a P(X > a) = \left(12a^3 - \frac{23}{2}a^4\right) + 11a\left(1 - 3a^2 + 2a^3\right) \\ &= \frac{21}{2}a^4 - 21a^3 + 11a. \end{aligned}$$

Differentiating the last term in the above display and setting the result equal to 0, we get

$$42a^3 - 63a^2 + 11 = 0.$$

Solving this equation, we find that $a \approx 0.51588$. Thus, to maximize expected profit, the manager should stock about 51.6 pounds of the flour at the beginning of each week.

10.141

a) From Example 7.8 on page 343, we know that, in general, $\bar{X}_n$ is an unbiased estimator of the common mean of the random variables $X_1, \ldots, X_n$. In this case, that common mean is

$$\frac{(\theta - 0.5) + (\theta + 0.5)}{2} = \theta.$$

Hence, $\bar{X}_n$ is an unbiased estimator of θ.

b) Observe that if $T \sim \mathcal{U}(\theta - 0.5, \theta + 0.5)$, then $T + 0.5 - \theta \sim \mathcal{U}(0, 1)$. Let $X = \min\{X_1, \ldots, X_n\}$ and $Y = \max\{X_1, \ldots, X_n\}$, and note that $M = (X + Y)/2$. Also, let $U_j = X_j + 0.5 - \theta$ for $1 \leq j \leq n$. We see that $U_1, \ldots, U_n$ is a random sample from a $\mathcal{U}(0, 1)$ distribution. Now let $U = \min\{U_1, \ldots, U_n\}$, $V = \max\{U_1, \ldots, U_n\}$, and set $W = (U + V)/2$. Then

$$\begin{aligned} X &= \min\{X_1, \ldots, X_n\} = \min\{U_1 + \theta - 0.5, \ldots, U_n + \theta - 0.5\} = U + \theta - 0.5, \\ Y &= \max\{X_1, \ldots, X_n\} = \max\{U_1 + \theta - 0.5, \ldots, U_n + \theta - 0.5\} = V + \theta - 0.5, \\ M &= (X + Y)/2 = W + \theta - 0.5. \end{aligned}$$

Referring Example 10.14(c) on page 590, we find that

$$\mathcal{E}(M) = \mathcal{E}(W + \theta - 0.5) = \mathcal{E}(W) + \theta - 0.5 = 0.5 + \theta - 0.5 = \theta.$$

Hence, M is an unbiased estimator of θ.

c) We have $\text{Var}\big(X_j\big) = \text{Var}\big(U_j - 0.5 + \theta\big) = \text{Var}\big(U_j\big) = 1/12$. Hence,

$$\text{Var}\big(\bar{X}_n\big) = \text{Var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\sum_{j=1}^{n} \text{Var}\big(X_j\big) = \frac{1}{12n}.$$

Also, from Example 10.14(c),

$$\operatorname{Var}(M) = \operatorname{Var}(W + \theta - 0.5) = \operatorname{Var}(W) = \frac{1}{2(n+2)(n+1)}.$$

Now,

$$\frac{\operatorname{Var}(M)}{\operatorname{Var}(\bar{X}_n)} = \frac{1/2(n+2)(n+1)}{1/12n} = \frac{6n}{(n+2)(n+1)}.$$

However, $(n+2)(n+1) - 6n = n^2 - 3n + 2 = (n-2)(n-1)$. Hence, $\operatorname{Var}(M) = \operatorname{Var}(\bar{X}_n)$ if $n = 1$ or $n = 2$, and $\operatorname{Var}(M) < \operatorname{Var}(\bar{X}_n)$ if $n \geq 3$.

d) Although, from parts (a) and (b), both M and $\bar{X}_n$ are unbiased estimators of θ, we see from part (c) that M is a better estimator because its variance is smaller than (or, in case $n = 1$ or $n = 2$, equal to) that of $\bar{X}_n$. Thus, on average, M will be closer to (or, in case $n = 1$ or $n = 2$, no further from) θ than $\bar{X}_n$.

Theory Exercises

10.142 We have, for all t,

$$P(X > t) = 1 - F_X(t) \geq 1 - F_Y(t) = P(Y > t)$$

and

$$P(X < t) = F_X(t-) \leq F_Y(t-) = P(Y < t).$$

Therefore, from Exercise 10.45(a) on page 581,

$$\begin{aligned} \mathcal{E}(X) &= \int_0^\infty P(X > t)\,dt - \int_0^\infty P(X < -t)\,dt \\ &\geq \int_0^\infty P(Y > t)\,dt - \int_0^\infty P(Y < -t)\,dt = \mathcal{E}(Y). \end{aligned}$$

10.143

a) If $\mathcal{E}(Y^2) = 0$, then $P(Y = 0) = 1$ and the terms on both sides of Schwarz's inequality equal 0. So, assume that $\mathcal{E}(Y^2) \neq 0$ and let $h(t) = \mathcal{E}\big((X + tY)^2\big)$. We have

$$0 \leq h(t) = \mathcal{E}\big(X^2 + 2XYt + t^2Y^2\big) = \mathcal{E}(X^2) + \big(2\mathcal{E}(XY)\big)\,t + \mathcal{E}(Y^2)\,t^2.$$

We see that h is a nonnegative quadratic polynomial and, hence, must have at most one real root. This, in turn, implies that

$$\big(2\mathcal{E}(XY)\big)^2 - 4 \cdot \mathcal{E}(Y^2)\,\mathcal{E}(X^2) \leq 0,$$

from which Schwarz's inequality immediately follows.

b) Applying Schwarz's inequality to $X - \mu_X$ and $Y - \mu_Y$, we obtain

$$\big(\operatorname{Cov}(X,Y)\big)^2 = \Big(\mathcal{E}\big((X-\mu_X)(Y-\mu_Y)\big)\Big)^2 \leq \mathcal{E}\big((X-\mu_X)^2\big)\,\mathcal{E}\big((Y-\mu_Y)^2\big) = \operatorname{Var}(X)\operatorname{Var}(Y).$$

Consequently,

$$\big(\rho(X,Y)\big)^2 = \frac{\big(\operatorname{Cov}(X,Y)\big)^2}{\operatorname{Var}(X)\operatorname{Var}(Y)} \leq 1,$$

from which we conclude that $|\rho(X,Y)| \leq 1$.

10.144 Let $Z = Y - X$. By assumption, Z is a positive random variable, that is, $P(Z > 0) = 1$. For $n \in \mathcal{N}$, let $A_n = \{Z \geq 1/n\}$. Then $A_1 \subset A_2 \subset \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \{Z > 0\}$. Applying the continuity property of a probability measure, we get $\lim_{n\to\infty} P(A_n) = P(Z > 0) = 1$. In particular, then, we have $P(Z \geq 1/n) > 0$ for sufficiently large n. Choose $n_0 \in \mathcal{N}$ such that $P(Z \geq 1/n_0) > 0$ and let $0 < z_0 < 1/n_0$. Then,

$$P(Z > z_0) \geq P(Z \geq 1/n_0) > 0.$$

Consequently, from Proposition 10.5 on page 577,

$$\mathcal{E}(Y) - \mathcal{E}(X) = \mathcal{E}(Z) = \int_0^\infty P(Z > z)\,dz \geq \int_0^{z_0} P(Z > z)\,dz \geq z_0 P(Z > z_0) > 0.$$

Hence, $\mathcal{E}(X) < \mathcal{E}(Y)$. We note that this result holds true if only $X \leq Y$ and $P(X < Y) > 0$.

10.145

a) Because the range of X is contained in I, it follows from the monotonicity property of expected value that $\mu_X \in I$. From Taylor's formula, for each $x \in I$, there is a number ξ_x between μ_X and x (and, hence, in I) such that

$$g(x) = g(\mu_X) + g'(\mu_X)(x - \mu_X) + \frac{g''(\xi_x)}{2}(x - \mu_X)^2.$$

Thus, because $g'' \geq 0$ on I, we have $g(x) \geq g(\mu_X) + g'(\mu_X)(x - \mu_X)$ for all $x \in I$ and, because the range of X is contained in I, we conclude that

$$g(X) \geq g(\mu_X) + g'(\mu_X)(X - \mu_X).$$

Consequently, by the monotonicity property of expected value,

$$\begin{aligned}\mathcal{E}\big(g(X)\big) &\geq \mathcal{E}\big(g(\mu_X) + g'(\mu_X)(X - \mu_X)\big) = g(\mu_X) + g'(\mu_X)\mathcal{E}(X - \mu_X)\\ &= g(\mu_X) + g'(\mu_X) \cdot 0 = g\left(\mathcal{E}(X)\right).\end{aligned}$$

b) Applying part (a) to the function $-g$ gives

$$-g\left(\mathcal{E}(X)\right) = (-g)\left(\mathcal{E}(X)\right) \leq \mathcal{E}\big((-g)(X)\big) = \mathcal{E}\big(-g(X)\big) = -\mathcal{E}\big(g(X)\big).$$

Therefore, $g\left(\mathcal{E}(X)\right) \geq \mathcal{E}\big(g(X)\big)$.

c) Let X denote the yield of the equity investment, and let E_{cd} and E_e denote the expected utility of the certificate of deposit and equity investment, respectively. Then

$$E_{cd} = u(r) \qquad \text{and} \qquad E_e = \mathcal{E}\big(u(X)\big).$$

First suppose that the investor uses the utility function $u(x) = \sqrt{x}$. This function is concave on $(0, \infty)$ because its second derivative is negative thereon. Hence, by part (b),

$$E_{cd} = u(r) = u\left(\mathcal{E}(X)\right) \geq \mathcal{E}\big(u(X)\big) = E_e.$$

Consequently, in this case, the investor will choose the certificate of deposit. Next suppose that the investor uses the utility function $u(x) = x^{3/2}$. This function is convex on $(0, \infty)$ because its second derivative is positive thereon. Hence, by part (a),

$$E_{cd} = u(r) = u\left(\mathcal{E}(X)\right) \leq \mathcal{E}\big(u(X)\big) = E_e.$$

Consequently, in this case, the investor will choose the equity investment.

Advanced Exercises

10.146

a) We begin by observing that $\mu_Y = \mathcal{E}(X) + \mathcal{E}(Z) = \mu$ and $\sigma_Y^2 = \operatorname{Var}(X) + \operatorname{Var}(Z) = \sigma^2 + 1$. Also,

$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(X, X + Z) = \operatorname{Cov}(X, X) + \operatorname{Cov}(X, Z) = \operatorname{Var}(X) + 0 = \operatorname{Var}(X)$$

so that

$$\rho = \rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\operatorname{Var}(X)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\sigma}{\sqrt{\sigma^2+1}}.$$

It follows that

$$1 - \rho^2 = 1 - \frac{\sigma^2}{\sigma^2+1} = \frac{1}{\sigma^2+1}$$

and

$$2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2} = 2\pi\sigma\sqrt{\sigma^2+1}\cdot\frac{1}{\sqrt{\sigma^2+1}} = 2\pi\sigma.$$

Now, let $U = X$ and $Y = X + Z$. The Jacobian determinant of the transformation $u = x$ and $y = x + z$ is

$$J(x, z) = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1.$$

Solving the equations $u = x$ and $y = x + z$ for x and z, we obtain the inverse transformation $x = u$ and $z = y - u$. Therefore, by the bivariate transformation theorem and the independence of X and Z,

$$f_{U,Y}(u, y) = \frac{1}{|J(x, z)|} f_{X,Y}(x, z) = \frac{1}{1} f_X(x) f_Z(z) = f_X(u) f_Z(y - u).$$

In other words,

$$\begin{aligned} f_{X,Y}(x, y) = f_X(x) f_Z(y - x) &= \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-(y-x)^2/2} \\ &= \frac{1}{2\pi\sigma}\, e^{-\left((x-\mu)^2+\sigma^2(y-x)^2\right)/2\sigma^2} = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\left((x-\mu)^2+\sigma^2(y-x)^2\right)/2\sigma^2}. \end{aligned}$$

However,

$$(y - x)^2 = \big((y - \mu) - (x - \mu)\big)^2 = (x - \mu)^2 - 2(x - \mu)(y - \mu) + (y - \mu)^2$$

and from that fact we get

$$\begin{aligned} &\frac{(x-\mu)^2 + \sigma^2(y-x)^2}{\sigma^2} \\ &\quad = (\sigma^2+1)\left\{\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{2\sigma}{\sqrt{\sigma^2+1}}\left(\frac{x-\mu}{\sigma}\right)\left(\frac{y-\mu}{\sqrt{\sigma^2+1}}\right) + \left(\frac{y-\mu}{\sqrt{\sigma^2+1}}\right)^2\right\} \\ &\quad = \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}. \end{aligned}$$

Therefore, X and Y are bivariate normal random variables.

b) Let Z_1 be the standardized version of X and let $Z_2 = Z$. Then Z_1 and Z_2 are independent standard normal random variables and

$$X = \mu + \sigma Z_1 + 0 \cdot Z_2$$
$$Y = \mu + \sigma Z_1 + 1 \cdot Z_2.$$

Therefore, from the argument used after Equations (10.55), we conclude that X and Y are bivariate normal random variables.
c) By the prediction theorem, the best predictor is $\mathcal{E}(X \mid Y)$. From part (a) or (b) and Proposition 10.10(e) on page 616, we deduce that

$$\mathcal{E}(X \mid Y) = \mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = \mu + \frac{\sigma}{\sqrt{1+\sigma^2}} \frac{\sigma}{\sqrt{1+\sigma^2}}(Y - \mu) = \frac{\mu + \sigma^2 Y}{1+\sigma^2}.$$

10.147 Let X denote the height, in inches, of the 25-year-old man, let Y denote the reported height, in inches, and let Z be the measurement error, in inches. We know that X and Z are independent random variables and that $X \sim \mathcal{N}\left(69, 2.5^2\right)$, $Z \sim \mathcal{N}(0, 1)$, and $Y = X + Z$. Applying Exercise 10.146(c) with $\mu = 69$, $\sigma = 2.5$, and $Y = 72$, we find that the best predictor of the man's height is

$$\frac{\mu + \sigma^2 Y}{1+\sigma^2} = \frac{69 + 2.5^2 \cdot 72}{1 + 2.5^2} = 71.6 \text{ inches.}$$

10.148 Let F and f denote the common CDF and PDF, respectively, of the X_js. We first note that $\{N > n\} = \bigcap_{j=2}^{n}\{X_j \le X_1\}$. Therefore, for $n \ge 2$,

$$\begin{aligned}
P(N > n) &= P(X_2 \le X_1, \dots, X_n \le X_1) \\
&= \underset{x_j \le x_1,\ 2 \le j \le n}{\int \cdots \int} f_{X_1,\dots,X_n}(x_1, \dots, x_n)\, dx_1 \cdots dx_n \\
&= \underset{x_j \le x_1,\ 2 \le j \le n}{\int \cdots \int} f(x_1) \cdots f(x_n)\, dx_1 \cdots dx_n \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x_1} f(x_2)\, dx_2 \cdots \int_{-\infty}^{x_1} f(x_n)\, dx_n \right) f(x_1)\, dx_1 \\
&= \int_{-\infty}^{\infty} \big(F(x_1)\big)^{n-1} f(x_1)\, dx_1 = \int_0^1 u^{n-1}\, du = \frac{1}{n},
\end{aligned}$$

where in the penultimate equation we made the substitution $u = F(x_1)$. Recalling that $\sum_{n=1}^{\infty} n^{-1} = \infty$, we conclude from Proposition 7.6 on page 347 that N doesn't have finite expectation.

10.149 Let $S_n = X_1 + \cdots + X_n$. We begin by observing that, because the X_js are positive random variables, $\{N > n\} = \{S_n \le 1\}$. We claim that, for $n \in \mathcal{N}$,

$$P(S_n \le x) = \frac{x^n}{n!}, \qquad 0 \le x \le 1. \qquad (*)$$

We use induction. Let $0 \le x \le 1$. Because $X_1 \sim \mathcal{U}(0, 1)$, $P(S_1 \le x) = P(X_1 \le x) = x$. Thus Equation $(*)$ holds for $n = 1$. Assume now that Equation $(*)$ holds for n. As $X_1, X_2, \dots$ are independent random variables, Proposition 6.13 on page 297 implies that S_n and X_{n+1} are also independent random variables. Hence, by Equation (9.46) on page 534, we have, for $0 < y < 1$,

$$f_{S_{n+1}}(y) = \int_{-\infty}^{\infty} f_{S_n}(u) f_{X_{n+1}}(y-u)\,du = \int_0^y f_{S_n}(u)\cdot 1\,du = P(S_n \le y) = \frac{y^n}{n!}.$$

Therefore, for $0 \le x \le 1$,

$$P(S_{n+1} \le x) = \int_0^x f_{S_{n+1}}(y)\,dy = \int_0^x \frac{y^n}{n!}\,dy = \frac{x^{n+1}}{(n+1)!},$$

which is Equation $(*)$ for $n+1$. In particular, we have shown that

$$P(N > n) = P(S_n \le 1) = \frac{1}{n!}, \qquad n \in \mathcal{N}.$$

a) Applying Proposition 7.6 on page 347, we get

$$\mathcal{E}(N) = \sum_{n=0}^{\infty} P(N > n) = \sum_{n=0}^{\infty} \frac{1}{n!} = e.$$

b) We have

$$P(N = n) = P(N > n-1) - P(N > n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}, \qquad n \in \mathcal{N}.$$

Hence,

$$\begin{aligned}\mathcal{E}\big(N^2\big) &= \sum_{n=1}^{\infty} n^2 P(N=n) = \sum_{n=2}^{\infty} n^2 P(N=n)\\ &= \sum_{n=2}^{\infty} n(n-1)P(N=n) + \sum_{n=2}^{\infty} nP(N=n) = \sum_{n=2}^{\infty} n(n-1)P(N=n) + \mathcal{E}(N).\end{aligned}$$

From part (a), $\mathcal{E}(N) = e$. Also,

$$n(n-1)P(N=n) = n(n-1)\frac{n-1}{n!} = \frac{n-1}{(n-2)!} = \frac{n-2}{(n-2)!} + \frac{1}{(n-2)!}.$$

Thus,

$$\sum_{n=2}^{\infty} n(n-1)P(N=n) = \sum_{n=3}^{\infty} \frac{1}{(n-3)!} + \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = 2e.$$

So, $\mathcal{E}\big(N^2\big) = 3e$ and, consequently, $\mathrm{Var}(N) = 3e - e^2 = (3-e)e$.

10.150 Use a basic random number generator to choose a random (uniform) number, u_1, between 0 and 1. Next choose another such random number, u_2. If $u_1 + u_2 > 1$, set $N_1 = 2$; otherwise, choose another such random number, u_3. If $u_1 + u_2 + u_3 > 1$, set $N_1 = 3$; otherwise, choose another random number, u_4, and so forth. The integer N_1 thus obtained has the same probability distribution as the random variable N in Exercise 10.149. Now repeat this entire procedure a large number of times, say, 10,000, to obtain integers $N_1, N_2, \ldots, N_{10{,}000}$. These integers represent 10,000 independent observations of the random variable N. Hence, by the long-run-average interpretation of expected value and Exercise 10.149(a),

$$\frac{N_1 + \cdots + N_{10{,}000}}{10{,}000} \approx \mathcal{E}(N) = e.$$

Note: For those readers who have access to Minitab statistical software, we have provided the following macro to perform the simulation:

```
MACRO
ESIM
MCOLUMN C D
MCONSTANT s x n i j k
BRIEF 2
GPAUSE
NOTE
NOTE This macro simulates e.
NOTE How many observations should be used for the simulation?
SET C;
FILE 'TERMINAL';
NOBS 1.
COPY C n
DO i=1:n
 LET s=0
 LET j=0
  WHILE s<=1
   RANDOM 1 C;
    UNIFORM 0 1.
   COPY C x
   LET s=s+x
   LET j=j+1
  ENDWHILE
 LET D(i)=j
ENDDO
NOTE The estimate of e is:
MEAN D
BRIEF 2     #DEFAULT
GPAUSE 0    #DEFAULT
ENDMACRO
```

10.151

a) From the FEF,

$$
\begin{aligned}
R(c) = \mathcal{E}(|Y-c|) = \int_{-\infty}^{\infty} |y-c| f_Y(y)\,dy &= \int_{-\infty}^{c} (c-y) f_Y(y)\,dy + \int_{c}^{\infty} (y-c) f_Y(y)\,dy \\
&= cF_Y(c) - \int_{-\infty}^{c} y f_Y(y)\,dy + \int_{c}^{\infty} y f_Y(y)\,dy - c\big(1 - F_Y(c)\big) \\
&= 2cF_Y(c) - \int_{-\infty}^{c} y f_Y(y)\,dy + \int_{c}^{\infty} y f_Y(y)\,dy - c.
\end{aligned}
$$

Consequently

$$R'(c) = 2cf_Y(c) + 2F_Y(c) - cf_Y(c) - cf_Y(c) - 1 = 2F_Y(c) - 1.$$

Therefore, $R'(c) = 0$ if and only if $F_Y(c) = 1/2$. In other words, the risk is minimized when c is any median, M, of Y. Thus, for proportional error, the minimum value of the risk is

$$R(M) = \mathcal{E}(|Y - M|),$$

which, in this context, is called the *minimum mean absolute error.*

b) From the FEF,

$$R(c) = \mathcal{E}\big((Y-c)^2\big) = \int_{-\infty}^{\infty} (y-c)^2 f_Y(y)\,dy.$$

Consequently,

$$R'(c) = 2\int_{-\infty}^{\infty} (y-c) f_Y(y)\,dy = 2\mathcal{E}(Y-c) = 2\left(\mathcal{E}(Y)-c\right).$$

Therefore, $R'(c) = 0$ if and only if $c = \mu_Y$. Thus, for squared error, the minimum risk is

$$R(\mu_Y) = \mathcal{E}\big((Y-\mu_Y)^2\big) = \mathrm{Var}(Y),$$

which, in this context, is also called the *minimum mean square error.*

10.152

a) From the law of total probability, we have, for $0 < w < 1$,

$$\begin{aligned} F_W(w) = P(W \le w) &= P(Y=0)P(W \le w \mid Y=0) + P(Y=1)P(W \le w \mid Y=1) \\ &= \frac{1}{2}P(X/2 \le w) + \frac{1}{2}P\left((X+1)/2 \le w\right) = \frac{1}{2}F_X(2w) + \frac{1}{2}F_X(2w-1). \end{aligned}$$

Differentiation now yields

$$f_W(w) = f_X(2w) + f_X(2w-1) = \begin{cases} 1+0, & \text{if } 0 < w < 1/2; \\ 0+1, & \text{if } 1/2 < w < 1. \end{cases}$$

Consequently, $f_W(w) = 1$ if $0 < w < 1$, and $f_W(w) = 0$ otherwise. Thus, $W \sim \mathcal{U}(0,1)$.

b) Given that $Y = 0$, $W = X/2$; and, given that $Y = 1$, $W = (X+1)/2$. Therefore,

$$\mathcal{E}(W \mid Y=0) = \mathcal{E}\left(\frac{X}{2}\right) = \frac{1}{2}\mathcal{E}(X) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

and

$$\mathcal{E}(W \mid Y=1) = \mathcal{E}\left(\frac{X+1}{2}\right) = \mathcal{E}\left(\frac{X}{2}\right) + \mathcal{E}\left(\frac{1}{2}\right) = \frac{1}{2}\mathcal{E}(X) + \frac{1}{2} = \frac{3}{4}.$$

c) Recalling that $Y \sim \mathcal{B}(1, 0.5)$, so that $P(Y=0) = P(Y=1) = 1/2$, we see, from part (b), that

$$p_{\mathcal{E}(W \mid Y)}(z) = \begin{cases} 1/2, & \text{if } z = 1/4 \text{ or } z = 3/4; \\ 0, & \text{otherwise.} \end{cases}$$

d) For convenience, set $Z = \mathcal{E}(W \mid Y)$. From part (a) and the law of total expectation,

$$\mu_Z = \mathcal{E}\big(\mathcal{E}(W \mid Y)\big) = \mathcal{E}(W) = \frac{1}{2}.$$

From part (c), we then see that $P\left((Z-\mu_Z)^2 = 1/16\right) = 1$. Hence,

$$\mathrm{Var}\big(\mathcal{E}(W \mid Y)\big) = \mathrm{Var}(Z) = \mathcal{E}\Big((Z-\mu_Z)^2\Big) = \frac{1}{16}.$$

e) Given that $Y = 0$, $W = X/2$; and, given that $Y = 1$, $W = (X+1)/2$. Therefore,

$$\mathrm{Var}(W \mid Y=0) = \mathrm{Var}\left(\frac{X}{2}\right) = \frac{1}{4}\,\mathrm{Var}(X)$$

and

$$\mathrm{Var}(W \mid Y=1) = \mathrm{Var}\left(\frac{X+1}{2}\right) = \mathrm{Var}\left(\frac{X}{2}+\frac{1}{2}\right) = \mathrm{Var}\left(\frac{X}{2}\right) = \frac{1}{4}\,\mathrm{Var}(X).$$

f) From part (e), we see that $P\left(\operatorname{Var}(W \mid Y) = \operatorname{Var}(X)/4\right) = 1$. Therefore, from part (a), the law of total variance, and part (d),

$$\operatorname{Var}(X) = \operatorname{Var}(W) = \mathcal{E}\big(\operatorname{Var}(W \mid Y)\big) + \operatorname{Var}\big(\mathcal{E}(W \mid Y)\big) = \frac{1}{4}\operatorname{Var}(X) + \frac{1}{16}.$$

Consequently, $\operatorname{Var}(X) = 1/12$.

10.153

a) Let ξ be a random variable representing the number of direct offspring of an individual. Then

$$\mu = \mathcal{E}(\xi) = \sum_x x p_\xi(x) = \sum_{j=1}^{\infty} j p_j$$

and

$$\sigma^2 = \sum_x x^2 p_\xi(x) - \left(\sum_x x p_\xi(x)\right)^2 = \sum_{j=1}^{\infty} j^2 p_j - \left(\sum_{j=1}^{\infty} j p_j\right)^2.$$

b) Given that $X_{n-1} = k$, there are k individuals in generation $n-1$. Let $\xi_1, \ldots, \xi_k$ denote the number of direct offspring of those k individuals, respectively. Then,

$$\mathcal{E}(X_n \mid X_{n-1} = k) = \mathcal{E}\left(\sum_{i=1}^{k} \xi_i\right) = \sum_{i=1}^{k} \mathcal{E}(\xi_i) = k\mu.$$

Therefore, $\mathcal{E}(X_n \mid X_{n-1}) = X_{n-1}\mu$ and, hence, by the law of total expectation

$$\mathcal{E}(X_n) = \mathcal{E}\big(\mathcal{E}(X_n \mid X_{n-1})\big) = \mathcal{E}(X_{n-1}\mu) = \mu\mathcal{E}(X_{n-1}).$$

Consequently,

$$\mathcal{E}(X_n) = \mathcal{E}(X_0)\left(\prod_{k=1}^{n} \frac{\mathcal{E}(X_k)}{\mathcal{E}(X_{k-1})}\right) = 1 \cdot \mu^n = \mu^n.$$

c) Referring to part (b) and using the fact that the individuals of any given generation act independently of one another, we get

$$\operatorname{Var}(X_n \mid X_{n-1} = k) = \operatorname{Var}\left(\sum_{i=1}^{k} \xi_i\right) = \sum_{i=1}^{k} \operatorname{Var}(\xi_i) = k\sigma^2.$$

Therefore, $\operatorname{Var}(X_n \mid X_{n-1}) = X_{n-1}\sigma^2$ and, hence, by the law of total variance,

$$\begin{aligned}\operatorname{Var}(X_n) &= \mathcal{E}\big(\operatorname{Var}(X_n \mid X_{n-1})\big) + \operatorname{Var}\big(\mathcal{E}(X_n \mid X_{n-1})\big) = \mathcal{E}\big(X_{n-1}\sigma^2\big) + \operatorname{Var}(X_{n-1}\mu) \\ &= \sigma^2\mathcal{E}(X_{n-1}) + \mu^2\operatorname{Var}(X_{n-1}) = \sigma^2\mu^{n-1} + \mu^2\operatorname{Var}(X_{n-1}).\end{aligned}$$

The preceding result gives a nonhomogeneous first-order linear difference equation, which can be solved by standard methods. Alternatively, we can proceed in the following way:

$$\begin{aligned}\operatorname{Var}(X_1) &= \sigma^2\mu^{1-1} + \mu^2\operatorname{Var}(X_0) = \sigma^2 + \mu^2 \cdot 0 = \sigma^2, \\ \operatorname{Var}(X_2) &= \sigma^2\mu^{2-1} + \mu^2\operatorname{Var}(X_1) = \sigma^2\mu + \mu^2\sigma^2 = \sigma^2\mu(1+\mu), \\ \operatorname{Var}(X_3) &= \sigma^2\mu^{3-1} + \mu^2\operatorname{Var}(X_2) = \sigma^2\mu^2 + \mu^2\left(\sigma^2\mu(1+\mu)\right) = \sigma^2\mu^2\left(1+\mu+\mu^2\right),\end{aligned}$$

and, hence, establish the pattern

$$\operatorname{Var}(X_n) = \sigma^2\mu^{n-1}\sum_{k=0}^{n-1}\mu^k,$$

which can be formally verified by using induction. In any case, the required result now follows.

d) Let E denote the event that there are no direct offspring for some generation. Then $\pi_0 = P(E)$. Event E occurs if and only if the first generation contains 0 individuals (i.e., the original individual has 0 direct offspring), which has probability p_0, or the first generation contains $j \geq 1$ individuals (i.e., the original individual has j direct offspring), which has probability p_j, and there are no direct offspring for some subsequent generation for each of the j members of the first generation. By independence, this latter event has probability $\big(P(E)\big)^j = \pi_0^j$. Hence, by the law of total probability,

$$\pi_0 = P(E) = \sum_{j=0}^{\infty} P(X_1 = j)P(E \mid X_1 = j) = p_0 \cdot 1 + \sum_{j=1}^{\infty} p_j \pi_0^j = \sum_{j=0}^{\infty} p_j \pi_0^j.$$

10.154 For $n \in \mathcal{N}$, we have, by the (multivariate) law of total expectation and the martingale property,

$$\mathcal{E}(X_{n+1}) = \mathcal{E}\big(\mathcal{E}(X_{n+1} \mid Y_1, \dots, Y_n)\big) = \mathcal{E}(X_n).$$

Therefore, $\mathcal{E}(X_n) = \mathcal{E}(X_1)$ for all $n \in \mathcal{N}$.

10.155 From Proposition 6.13 on page 297, because $\xi_1, \xi_2, \dots, \xi_{n+1}$ are independent, so are X_j and ξ_{n+1} for each $j = 1, 2, \dots, n$. Hence,

$$\begin{aligned}\mathcal{E}(X_{n+1} \mid X_1, \dots, X_n) &= \mathcal{E}(\xi_{n+1} + X_n \mid X_1, \dots, X_n) \\ &= \mathcal{E}(\xi_{n+1} \mid X_1, \dots, X_n) + \mathcal{E}(X_n \mid X_1, \dots, X_n) \\ &= \mathcal{E}(\xi_{n+1}) + X_n = 0 + X_n = X_n.\end{aligned}$$

Therefore, $\{X_n\}_{n=1}^{\infty}$ is a martingale.

10.156

a) Because the number of individuals in the $(n+1)$st generation depends on the number in the antecedent generations only through the number in the nth generation, we have $X_{n+1 \mid X_1, \dots, X_n} = X_{n+1 \mid X_n}$. Referring to the solution of Exercise 10.153(b), we now conclude that

$$\mathcal{E}(X_{n+1} \mid X_1, \dots, X_n) = \mathcal{E}(X_{n+1} \mid X_n) = X_n \mu = X_n \cdot 1 = X_n.$$

Therefore, $\{X_n\}_{n=1}^{\infty}$ is a martingale.

b) Let $U_n = X_n/\mu^n$ for each $n \in \mathcal{N}$. We need to show that $\{U_n\}_{n=1}^{\infty}$ is a martingale. Referring to part (a), we see that $U_{n+1 \mid U_1, \dots, U_n} = U_{n+1 \mid U_n}$. Hence, from the solution of Exercise 10.153(b),

$$\begin{aligned}\mathcal{E}(U_{n+1} \mid U_1 = u_1, \dots, U_n = u_n) &= \mathcal{E}(U_{n+1} \mid U_n = u_n) = \mathcal{E}\big(X_{n+1}/\mu^{n+1} \mid X_n/\mu^n = u_n\big) \\ &= \frac{1}{\mu^{n+1}}\mathcal{E}\big(X_{n+1} \mid X_n = \mu^n u_n\big) = \frac{1}{\mu^{n+1}}\mu^n u_n \cdot \mu = u_n.\end{aligned}$$

Thus, $\mathcal{E}(U_{n+1} \mid U_1, \dots, U_n) = U_n$ and, therefore, $\{U_n\}_{n=1}^{\infty}$ is a martingale.

10.157 We need to show that $\mathcal{E}(X_{n+1} \mid Y_1, \dots, Y_n) = X_n$ for all $n \in \mathcal{N}$; that is,

$$\mathcal{E}\big(\mathcal{E}(X \mid Y_1, \dots, Y_{n+1}) \mid Y_1, \dots, Y_n\big) = \mathcal{E}(X \mid Y_1, \dots, Y_n), \qquad n \in \mathcal{N}.$$

For convenience, set $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ and $\mathbf{y}_n = (y_1, \dots, y_n)$. From the multivariate version of Equation (10.48) on page 601, if $g\colon \mathcal{R}^{n+1} \to \mathcal{R}$, then

$$\mathcal{E}\big(g(\mathbf{Y}_n, Y_{n+1}) \mid \mathbf{Y}_n = \mathbf{y}_n\big) = \int_{-\infty}^{\infty} g(\mathbf{y}_n, y_{n+1}) f_{Y_{n+1} \mid \mathbf{Y}_n}(y_{n+1} \mid \mathbf{y}_n)\, dy_{n+1}.$$

Now, let $g(\mathbf{y}_n, y_{n+1}) = \mathcal{E}(X \mid \mathbf{Y}_n = \mathbf{y}_n, Y_{n+1} = y_{n+1})$. Then

$$\begin{aligned}
&\mathcal{E}\big(\mathcal{E}(X \mid \mathbf{Y}_n, Y_{n+1}) \mid \mathbf{Y}_n = \mathbf{y}_n\big) \\
&= \mathcal{E}\big(g(\mathbf{Y}_n, Y_{n+1}) \mid \mathbf{Y}_n = \mathbf{y}_n\big) = \int_{-\infty}^{\infty} g(\mathbf{y}_n, y_{n+1}) f_{Y_{n+1} \mid \mathbf{Y}_n}(y_{n+1} \mid \mathbf{y}_n)\, dy_{n+1} \\
&= \int_{-\infty}^{\infty} \mathcal{E}(X \mid \mathbf{Y}_n = \mathbf{y}_n, Y_{n+1} = y_{n+1}) f_{Y_{n+1} \mid \mathbf{Y}_n}(y_{n+1} \mid \mathbf{y}_n)\, dy_{n+1} \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X \mid \mathbf{Y}_n, Y_{n+1}}(x \mid \mathbf{y}_n, y_{n+1})\, dx \right) f_{Y_{n+1} \mid \mathbf{Y}_n}(y_{n+1} \mid \mathbf{y}_n)\, dy_{n+1} \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X \mid \mathbf{Y}_n, Y_{n+1}}(x \mid \mathbf{y}_n, y_{n+1}) f_{Y_{n+1} \mid \mathbf{Y}_n}(y_{n+1} \mid \mathbf{y}_n)\, dy_{n+1} \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X, Y_{n+1} \mid \mathbf{Y}_n}(x, y_{n+1} \mid \mathbf{y}_n)\, dy_{n+1} \right) dx = \int_{-\infty}^{\infty} x f_{X \mid \mathbf{Y}_n}(x \mid \mathbf{y}_n)\, dx \\
&= \mathcal{E}(X \mid \mathbf{Y}_n = \mathbf{y}_n)\,.
\end{aligned}$$

Thus,

$$\mathcal{E}\big(\mathcal{E}(X \mid Y_1, \dots, Y_{n+1}) \mid Y_1 = y_1, \dots Y_n = y_n\big) = \mathcal{E}(X \mid Y_1 = y_1, \dots Y_n = y_n)\,,$$

or, equivalently,

$$\mathcal{E}\big(\mathcal{E}(X \mid Y_1, \dots, Y_{n+1}) \mid Y_1, \dots Y_n\big) = \mathcal{E}(X \mid Y_1, \dots Y_n)\,,$$

as required.

10.158 For convenience, set $S_n = \sum_{j=1}^{n} Y_j$, which is the number of red balls that have been added to the urn after n draws. Now,

$$\mathcal{E}(Y_{n+1} \mid Y_1, \dots, Y_n) = P(R_{n+1} \mid Y_1, \dots, Y_n) = \frac{r + S_n}{n + r + b} = X_n.$$

Therefore, in view of the multivariate version of Equation (10.49) on page 602,

$$\begin{aligned}
\mathcal{E}(X_{n+1} \mid Y_1, \dots, Y_n) &= \mathcal{E}\left(\frac{r + S_{n+1}}{n + 1 + r + b} \mid Y_1, \dots, Y_n \right) = \mathcal{E}\left(\frac{r + S_n + Y_{n+1}}{n + 1 + r + b} \mid Y_1, \dots, Y_n \right) \\
&= \frac{\mathcal{E}(r + S_n + Y_{n+1} \mid Y_1, \dots, Y_n)}{n + 1 + r + b} \\
&= \frac{\mathcal{E}(r + S_n \mid Y_1, \dots, Y_n) + \mathcal{E}(Y_{n+1} \mid Y_1, \dots, Y_n)}{n + 1 + r + b} = \frac{r + S_n + X_n}{n + 1 + r + b} \\
&= \frac{(n + r + b)X_n + X_n}{n + 1 + r + b} = X_n.
\end{aligned}$$

BETA

Chapter 11

Generating Functions and Limit Theorems

11.1 Moment Generating Functions

Basic Exercises

11.1 We note that Y can take only the values 0 and 1. Moreover, using independence, we get

$$\begin{aligned} p_Y(1) &= P(Y=1) = P(X_1X_2X_3=1) = P(X_1=1,\ X_2=1,\ X_3=1) \\ &= P(X_1=1)P(X_2=1)P(X_3=1) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}. \end{aligned}$$

Thus, we see that $Y \sim \mathcal{B}(1, 8/27)$. Hence, from Table 11.1 on page 634, we have

$$M_Y(t) = \frac{19}{27} + \frac{8}{27}e^t, \qquad t \in \mathcal{R}.$$

11.2

a) We have, for all $t \in \mathcal{R}$,

$$M_X(t) = \sum_x e^{tx} p_X(x) = \sum_{x=0}^{\infty} e^{tx}\, e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}.$$

b) From part (a),

$$M_X'(t) = \lambda e^t e^{\lambda(e^t-1)}$$

and, consequently,

$$M_X''(t) = \left(\lambda e^t\right)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}.$$

Referring now to Proposition 11.2 on page 635, we get

$$\mathcal{E}(X) = M_X'(0) = \lambda e^0 e^{\lambda(e^0-1)} = \lambda e^0 e^{\lambda(1-1)} = \lambda.$$

Also,

$$\mathcal{E}\left(X^2\right) = M_X''(0) = \lambda^2 \cdot 1 + \lambda \cdot 1 \cdot 1 = \lambda^2 + \lambda,$$

so that

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

c) From Proposition 11.4 on page 636 and part (a),

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1)}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}.$$

Consequently, from the uniqueness property of MGFs (Proposition 11.3 on page 636) and part (a), we have $X + Y \sim \mathcal{P}(\lambda + \mu)$.
d) Suppose that $X_j \sim \mathcal{P}(\lambda_j)$, for $1 \le j \le m$, are independent. Arguing as in part (c), we find that

$$M_{X_1+\cdots+X_m}(t) = M_{X_1}(t)\cdots M_{X_m}(t) = e^{\lambda_1(e^t-1)}\cdots e^{\lambda_m(e^t-1)} = e^{(\lambda_1+\cdots+\lambda_m)(e^t-1)}.$$

Hence, from the uniqueness property of MGFs and part (a), $X_1 + \cdots + X_m \sim \mathcal{P}(\lambda_1 + \cdots + \lambda_m)$.

11.3
a) Using a geometric series, we get

$$M_X(t) = \sum_x e^{tx}p_X(x) = \sum_{x=1}^{\infty} e^{tx}p(1-p)^{x-1} = \frac{p}{1-p}\sum_{x=1}^{\infty}\big(e^t(1-p)\big)^x$$
$$= \frac{p}{1-p}\cdot\frac{e^t(1-p)}{1-e^t(1-p)} = \frac{pe^t}{1-(1-p)e^t}.$$

This result holds (i.e., the MGF is defined) if and only if $e^t(1-p) < 1$ or, equivalently, if and only if $t < -\ln(1-p)$.
b) Let $q = 1 - p$. From part (a),

$$M_X'(t) = \frac{(1-qe^t)pe^t + pe^tqe^t}{(1-qe^t)^2} = \frac{pe^t}{(1-qe^t)^2}$$

and, consequently,

$$M_X''(t) = \frac{(1-qe^t)^2pe^t + pe^t\cdot 2(1-qe^t)qe^t}{(1-qe^t)^4} = \frac{pe^t(1+qe^t)}{(1-qe^t)^3}.$$

Referring now to Proposition 11.2 on page 635, we get

$$\mathcal{E}(X) = M_X'(0) = \frac{pe^0}{(1-qe^0)^2} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Also,

$$\mathcal{E}\big(X^2\big) = M_X''(0) = \frac{pe^0(1+qe^0)}{(1-qe^0)^3} = \frac{p(1+q)}{(1-q)^3} = \frac{2-p}{p^2}$$

so that

$$\mathrm{Var}(X) = \mathcal{E}\big(X^2\big) - \big(\mathcal{E}(X)\big)^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

11.4
a) We know that $M_X(0) = 1$. For $t \ne 0$,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}f_X(x)\,dx = \int_a^b e^{tx}\frac{1}{b-a}\,dx = \frac{e^{bt}-e^{at}}{(b-a)t}.$$

b) We have

$$M_X(t) - M_X(0) = \frac{e^{bt}-e^{at}}{(b-a)t} - 1 = \frac{e^{bt}-e^{at}-(b-a)t}{(b-a)t}.$$

Applying L'Hôpital's rule, we now get

$$\begin{aligned}\mathcal{E}(X) = M_X'(0) &= \lim_{t\to 0}\frac{M_X(t) - M_X(0)}{t} = \lim_{t\to 0}\frac{e^{bt} - e^{at} - (b-a)t}{(b-a)t^2} \\ &= \lim_{t\to 0}\frac{be^{bt} - ae^{at} - (b-a)}{2(b-a)t} = \lim_{t\to 0}\frac{b^2e^{bt} - a^2e^{at}}{2(b-a)} = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.\end{aligned}$$

Next we determine $\mathcal{E}\left(X^2\right) = M_X''(0)$. To that end, we first refer to part (a) to get that, for $t \neq 0$,

$$M_X'(t) = \frac{1}{b-a}\frac{t\left(be^{bt} - ae^{at}\right) - \left(e^{bt} - e^{at}\right)}{t^2} = \frac{t\left(be^{bt} - ae^{at}\right) - \left(e^{bt} - e^{at}\right)}{(b-a)t^2}.$$

Therefore,

$$\begin{aligned}M_X'(t) - M_X'(0) &= \frac{t\left(be^{bt} - ae^{at}\right) - \left(e^{bt} - e^{at}\right)}{(b-a)t^2} - \frac{a+b}{2} \\ &= \frac{2t\left(be^{bt} - ae^{at}\right) - 2\left(e^{bt} - e^{at}\right) - \left(b^2 - a^2\right)t^2}{2(b-a)t^2}.\end{aligned}$$

Applying L'Hôpital's rule three times, we find that

$$\mathcal{E}\left(X^2\right) = M_X''(0) = \lim_{t\to 0}\frac{M_X'(t) - M_X'(0)}{t} = \frac{4\left(b^3 - a^3\right)}{12(b-a)} = \frac{a^2 + ab + b^2}{3}.$$

Consequently,

$$\begin{aligned}\operatorname{Var}(X) &= \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{4\left(a^2 + ab + b^2\right) - 3\left(a^2 + 2ab + b^2\right)}{12} = \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

An alternative approach to finding $\mathcal{E}(X)$ and $\mathcal{E}\left(X^2\right)$ is to use the exponential series, Equation (5.26) on page 222, to expand the MGF of X in a power series. We have

$$e^{bt} - e^{at} = \sum_{n=0}^{\infty}\frac{(bt)^n}{n!} - \sum_{n=0}^{\infty}\frac{(at)^n}{n!} = \sum_{n=1}^{\infty}\frac{(b^n - a^n)t^n}{n!}.$$

Therefore,

$$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t} = \sum_{n=1}^{\infty}\frac{(b^n - a^n)}{n!(b-a)}t^{n-1} = 1 + \frac{b^2 - a^2}{2(b-a)}t + \frac{b^3 - a^3}{6(b-a)}t^2 + \sum_{n=4}^{\infty}\frac{(b^n - a^n)}{n!(b-a)}t^{n-1}.$$

Differentiation now yields

$$M_X'(t) = \frac{b^2 - a^2}{2(b-a)} + \frac{b^3 - a^3}{3(b-a)}t + \sum_{n=4}^{\infty}\frac{(n-1)(b^n - a^n)}{n!(b-a)}t^{n-2}$$

and

$$M_X''(t) = \frac{b^3 - a^3}{3(b-a)} + \sum_{n=4}^{\infty}\frac{(n-1)(n-2)(b^n - a^n)}{n!(b-a)}t^{n-3}.$$

From the preceding two displays, it follows immediately that

$$\mathcal{E}(X) = M_X'(0) = \frac{a+b}{2} \qquad \text{and} \qquad \mathcal{E}\big(X^2\big) = M_X''(0) = \frac{a^2+ab+b^2}{3}.$$

11.5 Let X_j, $1 \le j \le 3$, denote the losses for city j. Then the combined losses for the three cities can be represented as $Y = X_1 + X_2 + X_3$. Because the X_js are independent, we can apply the multiplication property of MGFs, Proposition 11.4 on page 636, to conclude that

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t)M_{X_3}(t) = (1-2t)^{-3}(1-2t)^{-2.5}(1-2t)^{-4.5} = (1-2t)^{-10}.$$

To determine the third moment of Y, we can, in view of Proposition 11.2 on page 635, first obtain $M_Y'''(t)$ and then set $t = 0$. Doing so, we get

$$\mathcal{E}\big(X^3\big) = M_Y'''(0) = M_Y'''(t)\big|_{t=0} = 10{,}560(1-2t)^{-13}\big|_{t=0} = 10{,}560.$$

Alternatively, we note from Table 11.1 on page 634 and the uniqueness property of MGFs that Y has the chi-square distribution with 20 degrees of freedom or, equivalently, $Y \sim \Gamma(10, 1/2)$. Therefore, from Exercise 10.69,

$$\mathcal{E}\big(X^3\big) = \frac{10 \cdot 11 \cdot 12}{(1/2)^3} = 10{,}560.$$

11.6 No, there is no open interval containing 0 for which M_X is defined. If there were such an interval, then, in view of Proposition 11.2 on page 635, X would have moments of all orders. But, in fact, X does not even have a mean (first moment).

11.7
a) From Example 11.4 on page 632, $M_X(t) = \big(\lambda/(\lambda-t)\big)^{\alpha}$ and $M_Y(t) = \big(\lambda/(\lambda-t)\big)^{\beta}$. Therefore, from the multiplication property of MGFs,

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha}\left(\frac{\lambda}{\lambda-t}\right)^{\beta} = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha+\beta}.$$

Hence, from Example 11.4 and the uniqueness property of MGFs, we see that $X + Y \sim \Gamma(\alpha+\beta, \lambda)$.
b) Let $X_j \sim \Gamma(\alpha_j, \lambda)$, for $1 \le j \le m$, be independent random variables. From Example 11.4, we have $M_{X_j}(t) = \big(\lambda/(\lambda-t)\big)^{\alpha_j}$. Therefore, from the multiplication property of MGFs,

$$M_{X_1+\cdots+X_m}(t) = M_{X_1}(t)\cdots M_{X_m}(t) = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1}\cdots\left(\frac{\lambda}{\lambda-t}\right)^{\alpha_m} = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1+\cdots+\alpha_m}.$$

Thus, from Example 11.4 and the uniqueness property of MGFs, $X_1 + \cdots + X_m \sim \Gamma(\alpha_1 + \cdots + \alpha_m, \lambda)$.

11.8
a) From Exercise 8.154,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma y} e^{-(\ln y-\mu)^2/2\sigma^2}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. Therefore, upon making the substitution $x = \ln y$, we get

$$\begin{aligned}\mathcal{E}(Y^n) &= \int_{-\infty}^{\infty} y^n f_Y(y)\,dy = \frac{1}{\sqrt{2\pi}\,\sigma}\int_0^{\infty} y^{n-1}e^{-(\ln y-\mu)^2/2\sigma^2}\,dy \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\big(e^x\big)^n\, e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\left((x-\mu)^2/2\sigma^2-nx\right)}.\end{aligned}$$

Doing some algebra and completing the square we find that

$$(x-\mu)^2/2\sigma^2 - nx = \frac{\left(x-(\mu+n\sigma^2)\right)^2}{2\sigma^2} - \left(n\mu + n^2\sigma^2/2\right).$$

As the PDF of a $\mathcal{N}\left(\mu+n\sigma^2, \sigma^2\right)$ distribution integrates to 1, we conclude that $\mathcal{E}(Y^n) = e^{n\mu+n^2\sigma^2/2}$.
b) Let $X = \ln Y$, so that $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$. From the FEF,

$$\mathcal{E}\left(Y^n\right) = \mathcal{E}\left(\left(e^X\right)^n\right) = \mathcal{E}\left(e^{nX}\right) = \int_{-\infty}^{\infty} e^{nx} f_X(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{nx} e^{-(x-\mu)^2/2\sigma^2}\,dx.$$

The integral on the right of the preceding display is the same as the one that we evaluated in part (a). Hence, we conclude that $\mathcal{E}(Y^n) = e^{n\mu+n^2\sigma^2/2}$.
c) Let $X = \ln Y$, so that $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Referring to Example 11.5 on page 633, we find that

$$\mathcal{E}\left(Y^n\right) = \mathcal{E}\left(\left(e^X\right)^n\right) = \mathcal{E}\left(e^{nX}\right) = M_X(n) = e^{\mu n+\sigma^2 n^2/2} = e^{n\mu+n^2\sigma^2/2}.$$

d) Making the substitution $x = \ln y$, we get

$$\begin{aligned}\int_{-\infty}^{\infty} e^{ty} f_Y(y)\,dy &= \int_0^{\infty} e^{ty}\frac{1}{\sqrt{2\pi}\,\sigma y}\, e^{-(\ln y-\mu)^2/2\sigma^2}\,dy \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{te^x} e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{te^x - g(x)}\,dx,\end{aligned}$$

where, for convenience, we have set $g(x) = (x-\mu)^2/2\sigma^2$. For $t \le 0$,

$$\int_{-\infty}^{\infty} e^{te^x-g(x)}\,dx \le \int_{-\infty}^{\infty} e^{-g(x)}\,dx = \sqrt{2\pi}\,\sigma < \infty.$$

Now suppose that $t > 0$. Using L'Hôpital's rule, we find that $g(x)/te^x \to 0$ as $x \to \infty$ and, hence, that $te^x - g(x) > te^x/2 > x$ for sufficiently large x, say, for $x \ge M$. Therefore,

$$\int_{-\infty}^{\infty} e^{te^x-g(x)}\,dx \ge \int_M^{\infty} e^{te^x-g(x)}\,dx \ge \int_M^{\infty} e^x\,dx = \infty.$$

Referring now to the FEF, we conclude that $M_Y(t)$ is defined if and only if $t \le 0$.
e) No. As we have shown in parts (c) and (d), a lognormal random variable has moments of all orders, but its MGF is not defined in any open interval containing 0.

11.9
a) We have

$$\lim_{n\to\infty}\left(-1+\frac{1}{n}\right) = -1 \qquad \text{and} \qquad \lim_{n\to\infty}\left(1-\frac{1}{n}\right) = 1.$$

Thus, heuristically, we would say that the probability distribution of a random variable X to which $\{X_n\}_{n=1}^{\infty}$ converges in distribution is the discrete uniform distribution on $\{-1, 1\}$.
b) Let Y have the discrete uniform distribution on $\{a, b\}$. Then

$$M_Y(t) = \sum_y e^{ty} p_Y(y) = \frac{1}{2}e^{at} + \frac{1}{2}e^{bt} = \frac{1}{2}\left(e^{at}+e^{bt}\right).$$

Now, let X have the discrete uniform distribution on $\{-1, 1\}$. Then, by the continuity of the exponential function, we have, for all $t \in \mathcal{R}$,

$$\lim_{n\to\infty} M_{X_n}(t) = \lim_{n\to\infty} \frac{1}{2}\left(e^{(-1+1/n)t} + e^{(1-1/n)t}\right) = \frac{1}{2}\lim_{n\to\infty}\left(e^{(-1+1/n)t} + e^{(1-1/n)t}\right)$$
$$= \frac{1}{2}\left(e^{-t} + e^{t}\right) = \frac{1}{2}\left(e^{(-1)t} + e^{(1)t}\right) = M_X(t).$$

Hence, from the continuity theorem (Proposition 11.5 on page 638), we conclude that $\{X_n\}_{n=1}^{\infty}$ converges in distribution to X.

c) Let Y have the discrete uniform distribution on $\{a, b\}$. Then

$$F_Y(y) = P(Y \le y) = \begin{cases} 0, & \text{if } y < a; \\ 1/2, & \text{if } a \le y < b; \\ 1, & \text{if } y \ge b. \end{cases}$$

Thus, if X has the discrete uniform distribution on $\{-1, 1\}$,

$$F_X(x) = \begin{cases} 0, & \text{if } x < -1; \\ 1/2, & \text{if } -1 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

Observe that F_X is continuous except at -1 and 1.

Now, we have

$$F_{X_n}(x) = \begin{cases} 0, & \text{if } x < -1 + 1/n; \\ 1/2, & \text{if } -1 + 1/n \le x < 1 - 1/n; \\ 1, & \text{if } x \ge 1 - 1/n. \end{cases}$$

To determine $\lim_{n\to\infty} F_{X_n}(x)$, we consider three cases.

Case 1: $x \le -1$

In this case, we have $x < -1 + 1/n$ for all $n \in \mathcal{N}$, so that $F_{X_n}(x) = 0$ for all $n \in \mathcal{N}$. Consequently, we see that $\lim_{n\to\infty} F_{X_n}(x) = 0$.

Case 2: $-1 < x < 1$

In this case, for sufficiently large n, we have $-1 + 1/n < x < 1 - 1/n$, so that $F_{X_n}(x) = 1/2$. Consequently, we see that $\lim_{n\to\infty} F_{X_n}(x) = 1/2$.

Case 3: $x \ge 1$

In this case, we have $x > 1 - 1/n$ for all $n \in \mathcal{N}$, so that $F_{X_n}(x) = 1$ for all $n \in \mathcal{N}$. Consequently, we see that $\lim_{n\to\infty} F_{X_n}(x) = 1$.

We have now shown that

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0, & \text{if } x \le -1; \\ 1/2, & \text{if } -1 < x < 1; \\ 1, & \text{if } x \ge 1. \end{cases}$$

This limit agrees with $F_X(x)$ except when $x = -1$. In particular, then, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all $x \in \mathcal{C}_{F_X}$. Thus, by definition, $\{X_n\}_{n=1}^{\infty}$ converges in distribution to X.

11.10

a) Let $\epsilon > 0$ be small. The probability that a $\mathcal{U}(0, 1)$ random variable takes a value less than ϵ equals ϵ, which is positive. Consequently, by Proposition 4.6 on page 155, eventually one of the X_js will be less than ϵ. Heuristically, then, we would expect U_n to converge in distribution to 0.

b) Let U be the random variable that equals 0 with probability 1. Then,

$$F_U(u) = \begin{cases} 0, & \text{if } u < 0; \\ 1, & \text{if } u \geq 0. \end{cases}$$

From Example 9.9(c) on page 512, we find that

$$F_{U_n}(u) = \begin{cases} 0, & \text{if } u < 0; \\ 1 - (1-u)^n, & \text{if } 0 \leq u < 1; \\ 1, & \text{if } u \geq 1. \end{cases}$$

It follows that

$$\lim_{n\to\infty} F_{U_n}(u) = \begin{cases} 0, & \text{if } u \leq 0; \\ 1, & \text{if } u > 0. \end{cases}$$

This limit agrees with $F_U(u)$ except when $u = 0$, which is a point of discontinuity of F_U. In particular, then, $\lim_{n\to\infty} F_{U_n}(u) = F_U(u)$ for all $u \in \mathcal{C}_{F_U}$. Hence, by definition, $\{U_n\}_{n=1}^{\infty}$ converges in distribution to U.

c) Let $\epsilon > 0$ be small. The probability that a $\mathcal{U}(0, 1)$ random variable takes a value greater than $1 - \epsilon$ equals ϵ, which is positive. Consequently, by Proposition 4.6 on page 155, eventually one of the X_js will be greater than $1 - \epsilon$. Heuristically, then, we would expect V_n to converge in distribution to 1. Let V be the random variable that equals 1 with probability 1. Then,

$$F_V(v) = \begin{cases} 0, & \text{if } v < 1; \\ 1, & \text{if } v \geq 1. \end{cases}$$

From Exercise 9.70, we find that

$$F_{V_n}(v) = \begin{cases} 0, & \text{if } v < 0; \\ v^n, & \text{if } 0 \leq v < 1; \\ 1, & \text{if } v \geq 1. \end{cases}$$

It follows that

$$\lim_{n\to\infty} F_{V_n}(v) = \begin{cases} 0, & \text{if } v < 1; \\ 1, & \text{if } v \geq 1. \end{cases}$$

This limit agrees everywhere with $F_V(v)$. In particular, then, $\lim_{n\to\infty} F_{V_n}(v) = F_V(v)$ for all $v \in \mathcal{C}_{F_V}$. Hence, by definition, $\{V_n\}_{n=1}^{\infty}$ converges in distribution to V.

11.11

a) We have

$$\begin{aligned} M_{\bar{X}_n}(t) &= \mathcal{E}\left(e^{t\bar{X}_n}\right) = \mathcal{E}\left(e^{t\frac{X_1+\cdots+X_n}{n}}\right) = \mathcal{E}\left(e^{\frac{t}{n}(X_1+\cdots+X_n)}\right) \\ &= M_{X_1+\cdots+X_n}(t/n) = M_{X_1}(t/n)\cdots M_{X_n}(t/n) = \left(M(t/n)\right)^n . \end{aligned}$$

b) Referring to part (a), we see that

$$M'_{\bar{X}_n}(t) = n\left(M(t/n)\right)^{n-1} M'(t/n)(1/n) = M'(t/n)\left(M(t/n)\right)^{n-1} .$$

Therefore,

$$\mathcal{E}(\bar{X}_n) = M'_{\bar{X}_n}(0) = M'(0)\left(M(0)\right)^{n-1} = \mu \cdot 1^{n-1} = \mu.$$

Also,

$$\begin{aligned} M''_{\bar{X}_n}(t) &= M'(t/n)(n-1)\left(M(t/n)\right)^{n-2} M'(t/n)(1/n) + M''(t/n)(1/n)\left(M(t/n)\right)^{n-1} \\ &= \frac{n-1}{n}\left(M'(t/n)\right)^2\left(M(t/n)\right)^{n-2} + \frac{1}{n}M''(t/n)\left(M(t/n)\right)^{n-1} . \end{aligned}$$

Therefore,

$$\mathcal{E}\left(\bar{X}_n^2\right) = M''_{\bar{X}_n}(0) = \frac{n-1}{n}\left(M'(0)\right)^2\left(M(0)\right)^{n-2} + \frac{1}{n}M''(0)\left(M(0)\right)^{n-1}$$
$$= \frac{n-1}{n}\mu^2 \cdot 1^{n-2} + \frac{1}{n}\left(\sigma^2+\mu^2\right)\cdot 1^{n-1} = \mu^2 + \frac{1}{n}\sigma^2.$$

Hence,

$$\operatorname{Var}\left(\bar{X}_n\right) = \mathcal{E}\left(\bar{X}_n^2\right) - \left(\mathcal{E}\left(\bar{X}_n\right)\right)^2 = \mu^2 + \frac{1}{n}\sigma^2 - \mu^2 = \frac{\sigma^2}{n}.$$

c) Set $\psi_n(t) = \ln M_{\bar{X}_n}(t)$. Then,

$$\psi_n(t) = \ln M_{\bar{X}_n}(t) = \ln\left(M(t/n)\right)^n = n\ln M(t/n).$$

Applying L'Hôpital's rule, we now get

$$\lim_{n\to\infty}\psi_n(t) = \lim_{n\to\infty}\frac{\ln M(t/n)}{1/n} = \lim_{n\to\infty}\frac{-\left(t/n^2\right)M'(t/n)\big/M(t/n)}{-1/n^2}$$
$$= \lim_{n\to\infty} tM'(t/n)\big/M(t/n) = tM'(0)/M(0) = \mu t.$$

Therefore, $\lim_{n\to\infty} M_{\bar{X}_n}(t) = e^{\mu t}$.

d) Let X be the random variable that equals μ with probability 1. Then $M_X(t) = \mathcal{E}\left(e^{tX}\right) = \mathcal{E}\left(e^{t\mu}\right) = e^{\mu t}$. It now follows from part (c) and the continuity theorem, Proposition 11.5 on page 638, that $\left\{\bar{X}_n\right\}_{n=1}^{\infty}$ converges in distribution to μ.

Theory Exercises

11.12

a) Let X be a bounded discrete random variable, say, $P(|X| \le M) = 1$. Then $p_X(x) = 0$ for $|x| > M$. Therefore, for each $t \in \mathcal{R}$,

$$\sum_x |e^{tx}|p_X(x) = \sum_x e^{tx}p_X(x) \le \sum_x e^{|t||x|}p_X(x) = \sum_{|x|\le M} e^{|t||x|}p_X(x)$$
$$\le e^{|t|M}\sum_{|x|\le M} p_X(x) = e^{|t|M}P(|X|\le M) = e^{|t|M} < \infty.$$

Hence, from the FEF for discrete random variables, e^{tX} has finite expectation for all $t \in \mathcal{R}$; in other words, the MGF of X is defined for all $t \in \mathcal{R}$.

b) Let X be a bounded continuous random variable with a PDF, say, $P(|X| \le M) = 1$. From Exercise 10.19, we can assume that $f_X(x) = 0$ for $|x| > M$. Therefore, for each $t \in \mathcal{R}$,

$$\int_{-\infty}^{\infty}|e^{tx}|f_X(x)\,dx = \int_{-\infty}^{\infty}e^{tx}f_X(x)\,dx \le \int_{-\infty}^{\infty}e^{|t||x|}f_X(x)\,dx = \int_{|x|\le M}e^{|t||x|}f_X(x)\,dx$$
$$\le e^{|t|M}\int_{|x|\le M}f_X(x)\,dx = e^{|t|M}P(|X|\le M) = e^{|t|M} < \infty.$$

Hence, from the FEF for continuous random variables, e^{tX} has finite expectation for all $t \in \mathcal{R}$; in other words, the MGF of X is defined for all $t \in \mathcal{R}$.

c) We know that the MGF of a bounded random variable is defined for all $t \in \mathcal{R}$. Therefore, from Proposition 11.2 on page 635, a bounded random variable has moments of all orders.

11.13 Let $X_1, \ldots, X_m$ be independent random variables. We claim that

$$M_{X_1+\cdots+X_m} = M_{X_1} \cdots M_{X_m}.$$

Proceeding by induction, assume that the result is true for m. Now suppose that $X_1, \ldots, X_{m+1}$ are independent random variables. From Proposition 6.13 on page 297, the random variables $X_1 + \cdots + X_m$ and X_{m+1} are independent. Hence, from the first displayed equation in Proposition 11.4 and the induction assumption,

$$\begin{aligned} M_{X_1+\cdots+X_{m+1}} &= M_{(X_1+\cdots+X_m)+X_{m+1}} = M_{X_1+\cdots+X_m} M_{X_{m+1}} \\ &= \big(M_{X_1} \cdots M_{X_m}\big) M_{X_{m+1}} = M_{X_1} \cdots M_{X_{m+1}}, \end{aligned}$$

as required.

11.14
a) From Example 11.3 on page 631, we have $M_{X_n}(t) = \big(p_n e^t + 1 - p_n\big)^n$. Set $\psi_n(t) = \ln M_{X_n}(t)$. Then,

$$\begin{aligned} \psi_n(t) &= n \ln\big(p_n e^t + 1 - p_n\big) = n \ln\big(1 + p_n(e^t - 1)\big) \\ &= n \ln\big(1 + (\lambda/n)(e^t - 1)\big) = \lambda(e^t - 1)\, \frac{\ln\big(1 + (\lambda/n)(e^t - 1)\big)}{(\lambda/n)(e^t - 1)}. \end{aligned}$$

Applying the hint, we conclude that $\lim_{n\to\infty} \psi_n(t) = \lambda(e^t - 1)$, from which the required result immediately follows.
b) Let $X \sim \mathcal{P}(\lambda)$. From Table 11.1 on page 634, we have $M_X(t) = e^{\lambda(e^t-1)}$. Referring now to part (a) and the continuity theorem, we conclude that $\{X_n\}_{n=1}^{\infty}$ converges in distribution to X, that is, to a Poisson distribution with parameter λ.
c) Let $X \sim \mathcal{P}(\lambda)$ and let x be a nonnegative integer. The discontinuities of F_X occur at the nonnegative integers. Consequently, $x - 1/2$ and $x + 1/2$ are continuity points of F_X. We have

$$p_{X_n}(x) = P(X_n = x) = P(x - 1/2 < X_n \le x + 1/2) = F_{X_n}(x + 1/2) - F_{X_n}(x - 1/2).$$

Referring now to part (b), we conclude that

$$\begin{aligned} \lim_{n\to\infty} p_{X_n}(x) &= \lim_{n\to\infty} \big(F_{X_n}(x + 1/2) - F_{X_n}(x - 1/2)\big) \\ &= F_X(x + 1/2) - F_X(x - 1/2) = p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}. \end{aligned}$$

d) From part (c), we have, for large n,

$$p_{X_n}(x) \approx e^{-\lambda} \frac{\lambda^x}{x!} = e^{-np_n} \frac{(np_n)^x}{x!}, \qquad x = 0, 1, \ldots, n.$$

Because $p_n = \lambda/n$, we see that p_n is small when n is large. Using p instead of p_n in the previous display, we get the informal statement of Proposition 5.7.

11.15
a) We know that $\mathcal{E}(X_n) = \lambda_n$ and $\operatorname{Var}(X_n) = \lambda_n$. Therefore,

$$Y_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}} = -\sqrt{\lambda_n} + X_n/\sqrt{\lambda_n}.$$

Applying Proposition 11.1 on page 633 and recalling the MGF of a Poisson random variable, we get that

$$M_{Y_n}(t) = e^{-\sqrt{\lambda_n}\,t} M_{X_n}\left(t/\sqrt{\lambda_n}\right) = e^{-\sqrt{\lambda_n}\,t} e^{\lambda_n\left(e^{t/\sqrt{\lambda_n}}-1\right)} = e^{\lambda_n\left(e^{t/\sqrt{\lambda_n}}-1-t/\sqrt{\lambda_n}\right)}.$$

Taking logarithms gives

$$\ln M_{Y_n}(t) = \lambda_n\left(e^{t/\sqrt{\lambda_n}} - 1 - t/\sqrt{\lambda_n}\right) = t^2\left(\frac{e^{t/\sqrt{\lambda_n}} - 1 - t/\sqrt{\lambda_n}}{\left(t/\sqrt{\lambda_n}\right)^2}\right).$$

Applying the result given in the hint yields $\lim_{n\to\infty} \ln M_{Y_n}(t) = t^2/2$. Thus, $M_{Y_n}(t) \to e^{t^2/2}$ as $n \to \infty$. Now, let $Y \sim \mathcal{N}(0, 1)$. From Example 11.5 on page 633, we know that $M_Y(t) = e^{t^2/2}$. Therefore, from the continuity theorem, $\{Y_n\}_{n=1}^{\infty}$ converges in distribution to Y, that is, to a standard normal random variable.

b) Let $Y = (X-\lambda)/\sqrt{\lambda}$. Observing that Φ is everywhere continuous, we conclude from part (a) that $F_Y(y) \approx \Phi(y)$ for all $y \in \mathcal{R}$. Now let x be a nonnegative integer. Then

$$\begin{aligned}
p_X(x) = P(X = x) &= P\left(X \le x + \tfrac{1}{2}\right) - P\left(X \le x - \tfrac{1}{2}\right) \\
&= P\left(Y \le \frac{x + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - P\left(Y \le \frac{x - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) \\
&= F_Y\left(\frac{x + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - F_Y\left(Y_n \le \frac{x - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) \\
&\approx \Phi\left(\frac{x + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{x - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right).
\end{aligned}$$

Advanced Exercises

11.16

a) From calculus, we have $a_n = M_X^{(n)}(0)/n!$ for $n = 0, 1, \ldots$. Consequently, in view of the moment generation property of MGFs, we have $\mathcal{E}(X^n) = M_X^{(n)}(0) = n!\,a_n$ for each $n \in \mathcal{N}$.

b) Applying the geometric series to the MGF of an exponential random variable, we get

$$M_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - t/\lambda} = \sum_{n=0}^{\infty}\left(\frac{t}{\lambda}\right)^n = \sum_{n=0}^{\infty}\left(\frac{1}{\lambda}\right)^n t^n, \qquad t < \lambda.$$

Hence, in view of part (a), we have $\mathcal{E}(X^n) = n!\,\lambda^{-n}$ for all $n \in \mathcal{N}$.

c) Applying the exponential series to the MGF of a $\mathcal{N}\left(0, \sigma^2\right)$ random variable, Y, we get

$$M_Y(t) = e^{\sigma^2 t^2/2} = \sum_{n=0}^{\infty}\frac{\left(\sigma^2 t^2/2\right)^n}{n!} = \sum_{n=0}^{\infty}\frac{\sigma^{2n}}{n!\,2^n}t^{2n}.$$

Hence, from part (a),

$$\mathcal{E}\left(Y^{2n}\right) = (2n)!\,\frac{\sigma^{2n}}{n!\,2^n} = \frac{(2n)!\,\sigma^{2n}}{n!\,2^n}, \qquad n \in \mathcal{N}.$$

Now let $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Then $X - \mu \sim \mathcal{N}\left(0, \sigma^2\right)$ and, so, the previous display shows that the central moments of X are given by

$$\mathcal{E}\big((X-\mu)^{2n}\big) = \frac{(2n)!\,\sigma^{2n}}{n!\,2^n}, \qquad n \in \mathcal{N}.$$

11.17 As for real-valued random variables, a complex-valued random variable Z has finite expectation if and only if $\mathcal{E}(|Z|) < \infty$, where, in this case, $|z|$ denotes the modulus of a complex number z. Because $\left|e^{itX}\right| = 1$, we have $\mathcal{E}\left(\left|e^{itX}\right|\right) = \mathcal{E}(1) = 1 < \infty$. Consequently, e^{itX} has finite expectation for all $t \in \mathcal{R}$.

11.18
Analogue of Proposition 11.1: Let X be a random variable and let a and b be real numbers. Then the CHF of the random variable $a + bX$ is given by $\phi_{a+bX}(t) = e^{iat}\phi_X(bt)$. For the proof, we have

$$\phi_{a+bX}(t) = \mathcal{E}\left(e^{it(a+bX)}\right) = \mathcal{E}\left(e^{iat}e^{itbX}\right) = e^{iat}\mathcal{E}\left(e^{i(bt)X}\right) = e^{iat}\phi_X(bt).$$

Analogue of Proposition 11.2: If X has finite nth moment, then $\mathcal{E}(X^n) = (-i)^n\phi_X^{(n)}(0)$. For the proof, we first note that

$$\phi_X'(t) = \frac{d}{dt}\phi_X(t) = \frac{d}{dt}\mathcal{E}\left(e^{itX}\right) = \mathcal{E}\left(\frac{\partial}{\partial t}e^{itX}\right) = \mathcal{E}\left(iXe^{itX}\right) = i\mathcal{E}\left(Xe^{itX}\right).$$

Proceeding inductively, we obtain the relationship $\phi_X^{(n)}(t) = i^n\mathcal{E}\big(X^ne^{itX}\big)$. In particular, then,

$$\phi_X^{(n)}(0) = i^n\mathcal{E}\left(X^ne^{i\cdot 0\cdot X}\right) = i^n\mathcal{E}\big(X^n\big),$$

or, equivalently, $\mathcal{E}(X^n) = (-i)^n\phi_X^{(n)}(0)$.

Analogue of Proposition 11.4: If $X_1, \ldots, X_m$ are independent random variables, then

$$\phi_{X_1+\cdots+X_m} = \phi_{X_1}\cdots\phi_{X_m}.$$

For the proof, we first note that, from (the complex-valued version of) Proposition 6.13 on page 297, $e^{itX_1}, \ldots, e^{itX_m}$ are also independent random variables. Therefore, for each $t \in \mathcal{R}$,

$$\begin{aligned}\phi_{X_1+\cdots+X_m}(t) &= \mathcal{E}\left(e^{it(X_1+\cdots+X_m)}\right) = \mathcal{E}\left(e^{itX_1+\cdots+itX_m}\right) = \mathcal{E}\left(e^{itX_1}\cdots e^{itX_m}\right)\\ &= \mathcal{E}\left(e^{itX_1}\right)\cdots\mathcal{E}\left(e^{itX_m}\right) = \phi_{X_1}(t)\cdots\phi_{X_m}(t).\end{aligned}$$

In other words, $\phi_{X_1+\cdots+X_m} = \phi_{X_1}\cdots\phi_{X_m}$.

11.19
a) Formally, we have

$$\phi_X(t) = \mathcal{E}\left(e^{itX}\right) = \mathcal{E}\left(e^{(it)X}\right) = M_X(it).$$

b) Referring to part (a) and Table 11.1 on page 634, we get the following:

- If $X \sim \mathcal{B}(n, p)$, then
$$\phi_X(t) = M_X(it) = \left(pe^{it} + 1 - p\right)^n.$$
- If $X \sim \mathcal{P}(\lambda)$, then
$$\phi_X(t) = M_X(it) = e^{\lambda(e^{it}-1)}.$$

- If $X \sim \mathcal{G}(p)$, then
$$\phi_X(t) = M_X(it) = \frac{pe^{it}}{1-(1-p)e^{it}}.$$
- If $X \sim \mathcal{U}(a, b)$, then
$$\phi_X(t) = M_X(it) = \frac{e^{ibt} - e^{iat}}{(b-a)it}.$$
- If $X \sim \mathcal{E}(\lambda)$, then
$$\phi_X(t) = M_X(it) = \frac{\lambda}{\lambda - it}.$$
- If $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then
$$\phi_X(t) = M_X(it) = e^{\mu it + \sigma^2 (it)^2/2} = e^{i\mu t - \sigma^2 t^2/2}.$$

c) We apply the results from part (b) and the following formula, which we obtained in Exercise 11.18:
$$\mathcal{E}\big(X^n\big) = (-i)^n \phi_X^{(n)}(0).$$

- $X \sim \mathcal{B}(n, p)$: We have $M_X(it) = \left(pe^{it} + 1 - p\right)^n$. Hence,
$$\phi_X'(t) = n\left(pe^{it} + 1 - p\right)^{n-1} pe^{it} i = inpe^{it}\left(pe^{it} + 1 - p\right)^{n-1}$$
and
$$\begin{aligned}\phi_X''(t) &= inpe^{it}(n-1)\left(pe^{it} + 1 - p\right)^{n-2} pe^{it} i + inpe^{it} i \left(pe^{it} + 1 - p\right)^{n-1} \\ &= -n(n-1)p^2 e^{2it}\left(pe^{it} + 1 - p\right)^{n-2} - npe^{it}\left(pe^{it} + 1 - p\right)^{n-1}.\end{aligned}$$
Therefore,
$$\mathcal{E}(X) = (-i)^1 \phi_X'(0) = -i(inp) = np.$$
Also,
$$\mathcal{E}\big(X^2\big) = (-i)^2 \phi_X''(0) = -\big(-n(n-1)p^2 - np\big) = n(n-1)p^2 + np$$
and, hence,
$$\operatorname{Var}(X) = \mathcal{E}\big(X^2\big) - \big(\mathcal{E}(X)\big)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

- $X \sim \mathcal{P}(\lambda)$: We have $\phi_X(t) = e^{\lambda(e^{it}-1)}$. Hence,
$$\phi_X'(t) = e^{\lambda(e^{it}-1)} \lambda e^{it} i = i\lambda e^{it} e^{\lambda(e^{it}-1)} = i\lambda e^{it}\phi_X(t)$$
and
$$\phi_X''(t) = i\lambda e^{it}\phi_X'(t) + i\lambda e^{it} i \phi_X(t) = i\lambda e^{it}\phi_X'(t) - \lambda e^{it}\phi_X(t).$$
Therefore,
$$\mathcal{E}(X) = (-i)^1 \phi_X'(0) = -i(i\lambda) = \lambda.$$
Also,
$$\mathcal{E}\big(X^2\big) = (-i)^2 \phi_X''(0) = -\big(i\lambda\phi_X'(0) - \lambda\big) = -(i\lambda i\lambda - \lambda) = \lambda^2 + \lambda$$
and, hence,
$$\operatorname{Var}(X) = \mathcal{E}\big(X^2\big) - \big(\mathcal{E}(X)\big)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

- $X \sim \mathcal{G}(p)$: For convenience, let $q = 1 - p$. We have $\phi_X(t) = pe^{it}/\left(1 - qe^{it}\right)$. Hence,

$$\phi'_X(t) = \frac{\left(1 - qe^{it}\right) pe^{it} i - pe^{it}\left(-qe^{it} i\right)}{\left(1 - qe^{it}\right)^2} = \frac{ipe^{it}}{\left(1 - qe^{it}\right)^2}$$

and

$$\phi''_X(t) = \frac{\left(1 - qe^{it}\right)^2 ipe^{it} i - ipe^{it} \cdot 2\left(1 - qe^{it}\right)\left(-qe^{it} i\right)}{\left(1 - qe^{it}\right)^4} = -\frac{pe^{it}\left(1 + qe^{it}\right)}{\left(1 - qe^{it}\right)^3}.$$

Therefore,

$$\mathcal{E}(X) = (-i)^1 \phi'_X(0) = -i \cdot \frac{ip}{p^2} = \frac{1}{p}.$$

Also,

$$\mathcal{E}\left(X^2\right) = (-i)^2 \phi''_X(0) = -\left(-\frac{p(2-p)}{p^3}\right) = \frac{2-p}{p^2}$$

and, hence,

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

- $X \sim \mathcal{U}(a, b)$: We have $\phi_X(t) = \left(e^{ibt} - e^{iat}\right)\big/\left((b-a)it\right)$. Here, as one possible method of solution, we expand $\phi_X(t)$ in a power series. First, we note that

$$\begin{aligned} e^{ibt} - e^{iat} &= \sum_{n=0}^{\infty} \frac{(ibt)^n}{n!} - \sum_{n=0}^{\infty} \frac{(iat)^n}{n!} = \sum_{n=1}^{\infty} \frac{(it)^n\left(b^n - a^n\right)}{n!} \\ &= it(b-a) + \frac{(it)^2\left(b^2 - a^2\right)}{2} + \frac{(it)^3\left(b^3 - a^3\right)}{6} + \sum_{n=4}^{\infty} \frac{(it)^n\left(b^n - a^n\right)}{n!}. \end{aligned}$$

Consequently,

$$\phi_X(t) = \frac{e^{ibt} - e^{iat}}{(b-a)it} = 1 + \frac{i(a+b)}{2} t - \frac{a^2 + ab + b^2}{6} t^2 + \sum_{n=4}^{\infty} \frac{i^{n-1}\left(b^n - a^n\right)}{n!(b-a)} t^{n-1}.$$

Hence,

$$\phi'_X(t) = \frac{i(a+b)}{2} - \frac{a^2 + ab + b^2}{3} t + \sum_{n=4}^{\infty} \frac{i^{n-1}(n-1)\left(b^n - a^n\right)}{n!(b-a)} t^{n-2}$$

and

$$\phi''_X(t) = -\frac{a^2 + ab + b^2}{3} + \sum_{n=4}^{\infty} \frac{i^{n-1}(n-1)(n-2)\left(b^n - a^n\right)}{n!(b-a)} t^{n-3}.$$

Therefore,

$$\mathcal{E}(X) = (-i)^1 \phi'_X(0) = -i \cdot \frac{i(a+b)}{2} = \frac{a+b}{2}.$$

Also,

$$\mathcal{E}\left(X^2\right) = (-i)^2 \phi''_X(0) = -\left(-\frac{a^2 + ab + b^2}{3}\right) = \frac{a^2 + ab + b^2}{3}$$

and, so,

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

- $X \sim \mathcal{E}(\lambda)$: We have $\phi_X(t) = \lambda/(\lambda - it)$. Hence,

$$\phi_X'(t) = \lambda\left(-\frac{1}{(\lambda - it)^2}(-i)\right) = \frac{i\lambda}{(\lambda - it)^2}$$

and

$$\phi_X''(t) = i\lambda\left(\frac{-2}{(\lambda - it)^3}(-i)\right) = \frac{-2\lambda}{(\lambda - it)^3}$$

Therefore,

$$\mathcal{E}(X) = (-i)^1\phi_X'(0) = -i \cdot \frac{i\lambda}{\lambda^2} = \frac{1}{\lambda}.$$

Also,

$$\mathcal{E}\left(X^2\right) = (-i)^2\phi_X''(0) = -\left(-\frac{2\lambda}{\lambda^3}\right) = \frac{2}{\lambda^2}$$

and, hence,

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

- $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$: We have $\phi_X(t) = e^{i\mu t - \sigma^2 t^2/2}$. Hence,

$$\phi_X'(t) = \left(i\mu - \sigma^2 t\right)\phi_X(t)$$

and

$$\phi_X''(t) = \left(i\mu - \sigma^2 t\right)\phi_X'(t) - \sigma^2\phi_X(t).$$

Therefore,

$$\mathcal{E}(X) = (-i)^1\phi_X'(0) = -i \cdot i\mu = \mu.$$

Also,

$$\mathcal{E}\left(X^2\right) = (-i)^2\phi_X''(0) = -\left(i\mu\phi_X'(0) - \sigma^2\right) = -\left(i\mu \cdot i\mu - \sigma^2\right) = \mu^2 + \sigma^2$$

and, hence,

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \mu^2 + \sigma^2 - (\mu)^2 = \sigma^2.$$

11.20 Recall that a standard Cauchy random variable has a PDF given by

$$f_X(x) = \frac{1}{\pi\left(1 + x^2\right)}, \qquad x \in \mathcal{R}.$$

a) We have

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{e^{itx}}{\pi\left(1 + x^2\right)}, \qquad t \in \mathcal{R}.$$

Let $\mathbb{C}$ denote the set of complex numbers. Define $g_t \colon \mathbb{C} \to \mathbb{C}$ by $g_t(z) = e^{itz}\Big/\left(\pi\left(1 + z^2\right)\right)$. First assume that $t > 0$. Let $R > 1$ and let C_R denote the upper half of the circle of radius R centered at the origin. Then

$$\int_{C_R} g_t(z)\,dz = \int_{-R}^{R} g_t(x)\,dx + i\int_0^{\pi} g_t\left(Re^{i\theta}\right)Re^{i\theta}\,d\theta.$$

Now, for $0 < \theta < \pi$,

$$\left|e^{itRe^{i\theta}}\right| = \left|e^{itR(\cos\theta + i\sin\theta)}\right| = \left|e^{itR\cos\theta}e^{-tR\sin\theta}\right| = e^{-tR\sin\theta} \le 1,$$

where the inequality follows from the fact that $t > 0$. Also,

$$\begin{aligned}\left|1 + R^2e^{2i\theta}\right| &= \left|1 + R^2\cos(2\theta) + iR^2\sin(2\theta)\right| \\ &= \left(\left(1 + R^2\cos(2\theta)\right)^2 + \left(R^2\sin(2\theta)\right)^2\right)^{1/2} = \left(1 + 2R^2\cos(2\theta) + R^4\right)^{1/2} \\ &\ge \left(1 - 2R^2 + R^4\right)^{1/2} = R^2 - 1.\end{aligned}$$

Therefore,

$$|g_t(Re^{i\theta})| = \left|\frac{e^{itRe^{i\theta}}}{\pi\left(1 + R^2e^{2i\theta}\right)}\right| \le \frac{1}{\pi\left(R^2 - 1\right)}$$

and, consequently,

$$\left|\int_0^\pi g_t\left(Re^{i\theta}\right)Re^{i\theta}\,d\theta\right| \le R\int_0^\pi \left|g_t\left(Re^{i\theta}\right)\right|d\theta \le \frac{R}{R^2 - 1}.$$

As the term on the right of the preceding display converges to 0 as $R \to \infty$, we conclude that

$$\begin{aligned}\lim_{R\to\infty}\int_{C_R} g_t(z)\,dz &= \lim_{R\to\infty}\int_{-R}^{R} g_t(x)\,dx + \lim_{R\to\infty} i\int_0^\pi g_t\left(Re^{i\theta}\right)Re^{i\theta}\,d\theta \\ &= \int_{-\infty}^{\infty} g_t(x)\,dx + 0 = \phi_X(t).\end{aligned}$$

On the other hand, the function g_t is analytic in the upper half plane, except for a simple pole at $z = i$. The residue there is

$$\lim_{z\to i}(z - i)g_t(z) = \lim_{z\to i}\frac{e^{itz}}{\pi(i + z)} = \frac{e^{-t}}{2\pi i}.$$

Hence, by the residue theorem,

$$\int_{C_R} g_t(z)\,dz = 2\pi i\cdot\frac{e^{-t}}{2\pi i} = e^{-t}.$$

Thus, we have shown that $\phi_X(t) = e^{-t}$ if $t > 0$. By using an entirely similar argument, but with C_R the lower half of the circle of radius R centered at the origin, we can show that $\phi_X(t) = e^t$ if $t < 0$. Consequently, $\phi_X(t) = e^{-|t|}$ for all $t \in \mathcal{R}$.

b) Using the independence of the X_js and referring to part (a), we get

$$\begin{aligned}\phi_{\bar{X}_n}(t) &= \mathcal{E}\left(e^{it\bar{X}_n}\right) = \mathcal{E}\left(e^{it\frac{X_1+\cdots+X_n}{n}}\right) = \mathcal{E}\left(e^{i\frac{t}{n}(X_1+\cdots+X_n)}\right) = \phi_{X_1+\cdots+X_n}(t/n) \\ &= \phi_{X_1}(t/n)\cdots\phi_{X_n}(t/n) = \left(e^{-|t/n|}\right)^n = \left(e^{-|t|/n}\right)^n = e^{-|t|}.\end{aligned}$$

Hence, from the uniqueness theorem for characteristic functions and part (a), we conclude that $\bar{X}_n$ has the standard Cauchy distribution.

c) Let $X_1, \ldots, X_n$ represent n independent observations of a random variable X. The long-run-average interpretation of expected value is that, for large n, we have $\bar{X}_n \approx \mathcal{E}(X)$. In particular, this interpretation suggests that $\bar{X}_n$ will approach a single number, namely, $\mathcal{E}(X)$, as n increases without bound. Now, from part (b), if X has the standard Cauchy distribution, so does $\bar{X}_n$ for all n. But, then, $\bar{X}_n$ won't approach a single number as n increases without bound. This result, however, does not violate the long-run-average interpretation of expected value because, as we discovered in Example 10.5 on page 568, a random variable with the standard Cauchy distribution does not have a mean.

d) We recall from Exercise 8.153(b) that, if $X \sim \mathcal{C}(\eta, \theta)$, then $Y = (X - \eta)/\theta \sim \mathcal{C}(0, 1)$. From part (a), we have $\phi_Y(t) = e^{-|t|}$. Referring now to Exercise 11.18, we conclude that

$$\phi_X(t) = \phi_{\eta+\theta Y}(t) = e^{i\eta t}\phi_Y(\theta t) = e^{i\eta t}e^{-|\theta t|} = e^{i\eta t}e^{-\theta|t|} = e^{i\eta t-\theta|t|}.$$

11.2 Joint Moment Generating Functions

Basic Exercises

11.21

a) We have

$$p_{X,Y}(x, y) = \begin{cases} 1/4, & \text{if } (x, y) = (0, 0), (1, 0), (0, 1), \text{ or } (1, 1); \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} M_{X,Y}(s, t) &= \sum_{(x,y)}\sum e^{sx+ty}p_{X,Y}(x, y) = e^{s\cdot 0+t\cdot 0}\cdot\frac{1}{4} + e^{s\cdot 1+t\cdot 0}\cdot\frac{1}{4} + e^{s\cdot 0+t\cdot 1}\cdot\frac{1}{4} + e^{s\cdot 1+t\cdot 1}\cdot\frac{1}{4} \\ &= \frac{1}{4} + \frac{1}{4}e^s + \frac{1}{4}e^t + \frac{1}{4}e^{s+t} = \frac{1}{4}\left(1 + e^s + e^t + e^{s+t}\right). \end{aligned}$$

b) Referring to part (a), we get

$$\frac{\partial M_{X,Y}}{\partial s}(s, t) = \frac{1}{4}\left(e^s + e^{s+t}\right), \qquad \frac{\partial M_{X,Y}}{\partial t}(s, t) = \frac{1}{4}\left(e^t + e^{s+t}\right),$$

$$\frac{\partial^2 M_{X,Y}}{\partial s^2}(s, t) = \frac{1}{4}\left(e^s + e^{s+t}\right), \qquad \frac{\partial^2 M_{X,Y}}{\partial t^2}(s, t) = \frac{1}{4}\left(e^t + e^{s+t}\right),$$

$$\frac{\partial^2 M_{X,Y}}{\partial s\partial t}(s, t) = \frac{1}{4}e^{s+t}.$$

Therefore, from Proposition 11.6 on page 643,

$$\mathcal{E}(X) = \frac{\partial M_{X,Y}}{\partial s}(0, 0) = \frac{1}{4}\cdot 2 = \frac{1}{2}, \qquad \mathcal{E}(Y) = \frac{\partial M_{X,Y}}{\partial t}(0, 0) = \frac{1}{4}\cdot 2 = \frac{1}{2},$$

$$\mathcal{E}\left(X^2\right) = \frac{\partial^2 M_{X,Y}}{\partial s^2}(0, 0) = \frac{1}{4}\cdot 2 = \frac{1}{2}, \qquad \mathcal{E}\left(Y^2\right) = \frac{\partial^2 M_{X,Y}}{\partial t^2}(0, 0) = \frac{1}{4}\cdot 2 = \frac{1}{2},$$

$$\mathcal{E}(XY) = \frac{\partial^2 M_{X,Y}}{\partial s\partial t}(0, 0) = \frac{1}{4}\cdot 1 = \frac{1}{4}.$$

Also,

$$\operatorname{Var}(X) = \mathcal{E}\big(X^2\big) - \big(\mathcal{E}(X)\big)^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4},$$

$$\operatorname{Var}(Y) = \mathcal{E}\big(Y^2\big) - \big(\mathcal{E}(Y)\big)^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4},$$

$$\operatorname{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{2} = 0.$$

c) From Proposition 11.7 on page 645 and part (a),

$$M_X(s) = M_{X,Y}(s, 0) = \frac{1}{4}\left(1 + e^s + 1 + e^s\right) = \frac{1}{2}\left(1 + e^s\right)$$

and

$$M_Y(t) = M_{X,Y}(0, t) = \frac{1}{4}\left(1 + 1 + e^t + e^t\right) = \frac{1}{2}\left(1 + e^t\right).$$

d) From parts (a) and (c),

$$M_{X,Y}(s, t) = \frac{1}{4}\left(1 + e^s + e^t + e^{s+t}\right) = \left(\frac{1}{2}\left(1 + e^s\right)\right)\left(\frac{1}{2}\left(1 + e^t\right)\right) = M_X(s) M_Y(t).$$

Hence, from Proposition 11.9 on page 646, X and Y are independent random variables.
e) From part (c) and Table 11.1 on page 634, we see that both X and Y have the $\mathcal{B}(1, 1/2)$ distribution. Equivalently, they both have the Bernoulli distribution with parameter $1/2$ or the discrete uniform distribution on $\{0, 1\}$.

11.22
a) We have $f_{X,Y}(x, y) = 1$ if $0 < x < 1$ and $0 < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise. Therefore, if $s \neq 0$ and $t \neq 0$,

$$\begin{aligned}
M_{X,Y}(s, t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{sx+ty} f_{X,Y}(x, y)\, dx\, dy = \int_0^1\int_0^1 e^{sx+ty} \cdot 1\, dx\, dy \\
&= \int_0^1 e^{sx}\left(\int_0^1 e^{ty}\, dy\right) dx = \int_0^1 e^{sx}\left(\frac{e^t - 1}{t}\right) dx = \left(\frac{e^t - 1}{t}\right)\int_0^1 e^{sx}\, dx \\
&= \left(\frac{e^t - 1}{t}\right)\left(\frac{e^s - 1}{s}\right) = \frac{e^{s+t} - e^s - e^t + 1}{st}.
\end{aligned}$$

The other three cases are handled similarly, and we get

$$M_{X,Y}(s, t) = \begin{cases} \dfrac{e^{s+t} - e^s - e^t + 1}{st}, & \text{if } s \neq 0 \text{ and } t \neq 0; \\ \dfrac{e^s - 1}{s}, & \text{if } s \neq 0 \text{ and } t = 0; \\ \dfrac{e^t - 1}{t}, & \text{if } s = 0 \text{ and } t \neq 0; \\ 1, & \text{if } s = 0 \text{ and } t = 0. \end{cases}$$

b) There are several ways to solve this problem. We will use the most straightforward approach. From part (a), we have, for $s \neq 0$,

$$M_{X,Y}(s,0) - M_{X,Y}(0,0) = \frac{e^s - 1}{s} - 1 = \frac{e^s - 1 - s}{s}.$$

Therefore, from Proposition 11.6 on page 643,

$$\mathcal{E}(X) = \frac{\partial M_{X,Y}}{\partial s}(0,0) = \lim_{s\to 0} \frac{M_{X,Y}(s,0) - M_{X,Y}(0,0)}{s} = \lim_{s\to 0} \frac{e^s - 1 - s}{s^2} = \frac{1}{2},$$

where the last equality follows by two applications of L'Hôpital's rule. By symmetry, $\mathcal{E}(Y) = 1/2$.

Again referring to part (a), we get, for $s \neq 0$,

$$\frac{\partial M_{X,Y}}{\partial s}(s,0) - \frac{\partial M_{X,Y}}{\partial s}(0,0) = \frac{se^s - e^s + 1}{s^2} - \frac{1}{2} = \frac{2se^s - 2e^s + 2 - s^2}{2s^2}.$$

Hence, from Proposition 11.6,

$$\begin{aligned}\mathcal{E}(X^2) &= \frac{\partial^2 M_{X,Y}}{\partial s^2}(0,0) = \frac{\partial\left(\dfrac{\partial M_{X,Y}}{\partial s}\right)}{\partial s}(0,0) = \lim_{s\to 0} \frac{\dfrac{\partial M_{X,Y}}{\partial s}(s,0) - \dfrac{\partial M_{X,Y}}{\partial s}(0,0)}{s} \\ &= \lim_{s\to 0} \frac{2se^s - 2e^s + 2 - s^2}{2s^3} = \frac{1}{3},\end{aligned}$$

where the last equality follows by three applications of L'Hôpital's rule. Thus,

$$\operatorname{Var}(X) = \mathcal{E}(X^2) - \big(\mathcal{E}(X)\big)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

By symmetry, $\operatorname{Var}(Y) = 1/12$.

Again referring to part (a), we get, for $s \neq 0$ and $t \neq 0$,

$$\begin{aligned}M_{X,Y}(s,t) - M_{X,Y}(s,0) &= \frac{e^{s+t} - e^s - e^t + 1}{st} - \frac{e^s - 1}{s} \\ &= \frac{e^s - 1}{s}\left(\frac{e^t - 1}{t} - 1\right) = \frac{e^s - 1}{s}\left(\frac{e^t - 1 - t}{t}\right).\end{aligned}$$

Therefore,

$$\frac{\partial M_{X,Y}}{\partial t}(s,0) = \lim_{t\to 0} \frac{M_{X,Y}(s,t) - M_{X,Y}(s,0)}{t} = \frac{e^s - 1}{s} \lim_{t\to 0} \frac{e^t - 1 - t}{t^2} = \frac{e^s - 1}{2s},$$

where the the last equality follows by two applications of L'Hôpital's rule. Consequently,

$$\frac{\partial M_{X,Y}}{\partial t}(s,0) - \frac{\partial M_{X,Y}}{\partial t}(0,0) = \frac{e^s - 1}{2s} - \frac{1}{2} = \frac{e^s - 1 - s}{2s}$$

and, hence, from Proposition 11.6,

$$\mathcal{E}(XY) = \frac{\partial^2 M_{X,Y}}{\partial s \partial t}(0,0) = \frac{\partial\left(\dfrac{\partial M_{X,Y}}{\partial t}\right)}{\partial s}(0,0) = \lim_{s\to 0} \frac{\dfrac{\partial M_{X,Y}}{\partial t}(s,0) - \dfrac{\partial M_{X,Y}}{\partial t}(0,0)}{s}$$

$$= \lim_{s\to 0} \frac{e^s - 1 - s}{2s^2} = \frac{1}{4},$$

where the the last equality follows by two applications of L'Hôpital's rule. Thus,

$$\operatorname{Cov}(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \frac{1}{4} - \frac{1}{2}\cdot\frac{1}{2} = 0.$$

c) Referring to Proposition 11.7 on page 645 and part (a), we get

$$M_X(s) = M_{X,Y}(s,0) = \begin{cases} \big(e^s - 1\big)/s, & \text{if } s \neq 0; \\ 1, & \text{if } s = 0. \end{cases}$$

and

$$M_Y(t) = M_{X,Y}(0,t) = \begin{cases} \big(e^t - 1\big)/t, & \text{if } t \neq 0; \\ 1, & \text{if } t = 0. \end{cases}$$

d) From parts (c) and (a),

$$M_X(s)M_Y(t) = \begin{cases} \big(e^s - 1\big)/s, & \text{if } s \neq 0; \\ 1, & \text{if } s = 0. \end{cases} \cdot \begin{cases} \big(e^t - 1\big)/t, & \text{if } t \neq 0; \\ 1, & \text{if } t = 0. \end{cases}$$

$$= \begin{cases} \left(\dfrac{e^s-1}{s}\right)\cdot\left(\dfrac{e^t-1}{t}\right), & \text{if } s \neq 0 \text{ and } t \neq 0; \\ \dfrac{e^s-1}{s}, & \text{if } s \neq 0 \text{ and } t = 0; \\ \dfrac{e^t-1}{t}, & \text{if } s = 0 \text{ and } t \neq 0; \\ 1, & \text{if } s = 0 \text{ and } t = 0. \end{cases}$$

$$= \begin{cases} \dfrac{e^{s+t} - e^s - e^t + 1}{st}, & \text{if } s \neq 0 \text{ and } t \neq 0; \\ \dfrac{e^s-1}{s}, & \text{if } s \neq 0 \text{ and } t = 0; \\ \dfrac{e^t-1}{t}, & \text{if } s = 0 \text{ and } t \neq 0; \\ 1, & \text{if } s = 0 \text{ and } t = 0. \end{cases}$$

$$= M_{X,Y}(s,t).$$

Therefore, from Proposition 11.9 on page 646, X and Y are independent random variables.
e) From part (c) and Table 11.1 on page 634, we see that both X and Y have the $\mathcal{U}(0,1)$ distribution.

11.23 Let X and Y be as in Exercise 11.22. From parts (d) and (e) of that exercise, we know that X and Y are independent $\mathcal{U}(0,1)$ random variables. Consequently, from Example 9.19, $X + Y \sim \mathcal{T}(0,2)$.

However, from Proposition 11.10 on page 647 and part (a) of Exercise 11.22, we have, for $t \neq 0$,

$$M_{X+Y}(t) = M_{X,Y}(t,t) = \frac{e^{t+t} - e^t - e^t - 1}{t \cdot t} = \frac{e^{2t} - 2e^t - 1}{t^2}$$
$$= \left(\frac{2e^t}{t^2}\right)\left(\frac{e^t + e^{-t}}{2} - 1\right) = \frac{2e^t}{t^2}(\cosh t - 1).$$

Hence, the MGF of a triangular distribution on the interval (0, 2) is

$$M(t) = \begin{cases} \dfrac{2e^t}{t^2}(\cosh t - 1), & \text{if } t \neq 0; \\ 1, & \text{if } t = 0. \end{cases}$$

11.24

a) We have

$$\frac{\partial \psi}{\partial t}(s,t) = \frac{\partial \ln M_{X,Y}}{\partial t}(s,t) = \frac{1}{M_{X,Y}(s,t)} \cdot \frac{\partial M_{X,Y}}{\partial t}(s,t) = \big(M_{X,Y}(s,t)\big)^{-1} \frac{\partial M_{X,Y}}{\partial t}(s,t).$$

Consequently,

$$\frac{\partial^2 \psi}{\partial s \partial t}(s,t) = \big(M_{X,Y}(s,t)\big)^{-1} \frac{\partial^2 M_{X,Y}}{\partial s \partial t}(s,t) - \big(M_{X,Y}(s,t)\big)^{-2} \frac{\partial M_{X,Y}}{\partial s}(s,t) \frac{\partial M_{X,Y}}{\partial t}(s,t).$$

Setting $s = t = 0$, we conclude, in view of Proposition 11.6 on page 643, that

$$\frac{\partial^2 \psi}{\partial s \partial t}(0,0) = \big(M_{X,Y}(0,0)\big)^{-1} \frac{\partial^2 M_{X,Y}}{\partial s \partial t}(0,0) - \big(M_{X,Y}(0,0)\big)^{-2} \frac{\partial M_{X,Y}}{\partial s}(0,0) \frac{\partial M_{X,Y}}{\partial t}(0,0)$$
$$= 1^{-1}\mathcal{E}(XY) - 1^{-2}\mathcal{E}(X)\,\mathcal{E}(Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \operatorname{Cov}(X,Y)\,.$$

b) From Example 11.10 on page 642, we see that

$$\psi(s,t) = \ln M_{X,Y}(s,t) = \mu_X s + \mu_Y t + \frac{1}{2}\left(\sigma_X^2 s^2 + 2\rho\sigma_X\sigma_Y st + \sigma_Y^2 t^2\right).$$

From this result, we get, successively,

$$\frac{\partial \psi}{\partial t}(s,t) = \mu_Y + \frac{1}{2}\left(2\rho\sigma_X\sigma_Y s + 2\sigma_Y^2 t\right) = \mu_Y + \rho\sigma_X\sigma_Y s + \sigma_Y^2 t$$

and

$$\frac{\partial^2 \psi}{\partial s \partial t}(s,t) = \frac{\partial \left(\dfrac{\partial \psi}{\partial t}\right)}{\partial s}(s,t) = \rho\sigma_X\sigma_Y = \operatorname{Cov}(X,Y)\,.$$

11.25

a) From Proposition 11.7 on page 645,

$$M_X(s) = M_{X,Y}(s,0) = \left(pe^s + qe^0\right)^n = \left(pe^s + 1 - p\right)^n.$$

Referring now to Table 11.1 on page 634, we see that $X \sim \mathcal{B}(n,p)$. Similarly, we find that $Y \sim \mathcal{B}(n,q)$.

b) From part (a), we see that

$$M_X(s)M_Y(t) = \left(pe^s + q\right)^n\left(p + qe^t\right)^n = \left((pe^s + q)(p + qe^t)\right)^n$$
$$= \left(p^2e^s + pqe^{s+t} + pq + q^2e^t\right)^n \neq \left(pe^s + qe^t\right)^n = M_{X,Y}(s,t).$$

Hence, from Proposition 11.9 on page 646, X and Y are not independent random variables.

c) Recalling that $p + q = 1$, we find, in view of Proposition 11.10 on page 647, that

$$M_{X+Y}(t) = M_{X,Y}(t,t) = \left(pe^t + qe^t\right)^n = \left(e^t\right)^n = e^{nt}.$$

However, e^{nt} is the MGF of a random variable that equals n with probability 1. Thus, by the uniqueness property of MGFs, $P(X + Y = n) = 1$. In other words, $X + Y$ is a discrete random variable with $p_{X+Y}(n) = 1$ and $p_{X+Y}(z) = 0$ if $z \neq n$.

11.26

a) Let a and b be two real numbers, not both 0. Referring to Equation (11.15) on page 642, we get

$$\begin{aligned} M_{aX+bY}(t) &= \mathcal{E}\left(e^{t(aX+bY)}\right) = \mathcal{E}\left(e^{(at)X+(bt)Y}\right) = M_{X,Y}(at, bt) \\ &= e^{\mu_X(at)+\mu_Y(bt)+\frac{1}{2}\left(\sigma_X^2(at)^2+2\rho\sigma_X\sigma_Y(at)(bt)+\sigma_Y^2(bt)^2\right)} \\ &= e^{(a\mu_X+b\mu_Y)t+\left(a^2\sigma_X^2+2ab\rho\sigma_X\sigma_Y+b^2\sigma_Y^2\right)t^2/2}. \end{aligned}$$

Referring now to Table 11.1 on page 634, we conclude that

$$aX + bY \sim \mathcal{N}\left(a\mu_X + b\mu_Y, a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2\right).$$

b) We know that the mean and variance of a normal random variable are its μ and σ^2 parameters, respectively. Hence, from part (a),

$$\mathcal{E}(aX + bY) = a\mu_X + b\mu_Y \qquad \text{and} \qquad \operatorname{Var}(aX + bY) = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2.$$

11.27

a) We have

$$\begin{aligned} M_{X,Y}(s,t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{sx+ty} f_{X,Y}(x,y)\,dx\,dy \\ &= \iint\limits_{x>y>0} e^{sx+ty}(x-y)e^{-(x+y)}\,dx\,dy + \iint\limits_{y>x>0} e^{sx+ty}(y-x)e^{-(x+y)}\,dx\,dy. \end{aligned}$$

Upon making the substitution $u = x - y$ and using the fact that the mean of an exponential distribution is the reciprocal of its parameter, we get

$$\begin{aligned} \iint\limits_{x>y>0} e^{sx+ty}(x-y)e^{-(x+y)}\,dx\,dy &= \int_0^\infty e^{-(1-t)y}\left(\int_y^\infty e^{-(1-s)x}(x-y)\,dx\right)dy \\ &= \int_0^\infty e^{-(1-t)y}\left(\int_0^\infty ue^{-(1-s)(u+y)}\,du\right)dy \\ &= \frac{1}{1-s}\int_0^\infty e^{-(2-s-t)y}\left(\int_0^\infty u(1-s)e^{-(1-s)u}\,du\right)dy \\ &= \frac{1}{(1-s)^2}\int_0^\infty e^{-(2-s-t)y}\,dy = \frac{1}{(1-s)^2(2-s-t)}. \end{aligned}$$

Similarly, we find that

$$\iint\limits_{y>x>0} e^{sx+ty}(y-x)e^{-(x+y)}\,dx\,dy = \frac{1}{(1-t)^2(2-s-t)}.$$

Hence,

$$M_{X,Y}(s,t) = \frac{1}{2-s-t}\left(\frac{1}{(1-s)^2} + \frac{1}{(1-t)^2}\right), \qquad s,t < 1.$$

b) From part (a), we find that

$$\frac{\partial M_{X,Y}}{\partial s}(s,t) = (2-s-t)^{-1}\left(2(1-s)^{-3}\right) + (2-s-t)^{-2}\left((1-s)^{-2} + (1-t)^{-2}\right),$$

$$\begin{aligned}\frac{\partial^2 M_{X,Y}}{\partial s^2}(s,t) &= (2-s-t)^{-1}\left(6(1-s)^{-4}\right) + (2-s-t)^{-2}\left(2(1-s)^{-3}\right)\\ &\quad + (2-s-t)^{-2}\left(2(1-s)^{-3}\right) + 2(2-s-t)^{-3}\left((1-s)^{-2} + (1-t)^{-2}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 M_{X,Y}}{\partial t\,\partial s}(s,t) &= (2-s-t)^{-2}\left(2(1-s)^{-3}\right) + (2-s-t)^{-2}\left(2(1-t)^{-3}\right)\\ &\quad + 2(2-s-t)^{-3}\left((1-s)^{-2} + (1-t)^{-2}\right).\end{aligned}$$

Therefore, from Proposition 11.6 on page 643,

$$\mathcal{E}(X) = \frac{\partial M_{X,Y}}{\partial s}(0,0) = 2^{-1}\cdot 2\cdot 1^{-3} + 2^{-2}\left(1^{-2} + 1^{-2}\right) = \frac{3}{2},$$

$$\mathcal{E}\left(X^2\right) = \frac{\partial^2 M_{X,Y}}{\partial s^2}(0,0) = 2^{-1}\cdot 6\cdot 1^{-4} + 2\cdot 2^{-2}\cdot 2\cdot 1^{-3} + 2\cdot 2^{-3}\left(1^{-2} + 1^{-2}\right) = \frac{9}{2},$$

and

$$\mathcal{E}(XY) = \frac{\partial^2 M_{X,Y}}{\partial t\,\partial s}(0,0) = 2^{-2}\cdot 2\cdot 1^{-3} + 2^{-2}\cdot 2\cdot 1^{-3} + 2\cdot 2^{-3}\left(1^{-2} + 1^{-2}\right) = \frac{3}{2}.$$

By symmetry, $\mathcal{E}(Y) = 3/2$. Also,

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - (\mathcal{E}(X))^2 = \frac{9}{2} - \left(\frac{3}{2}\right)^2 = \frac{9}{4}$$

and

$$\operatorname{Cov}(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \frac{3}{2} - \frac{3}{2}\cdot\frac{3}{2} = -\frac{3}{4}.$$

By symmetry, $\operatorname{Var}(Y) = 9/4$.
c) From Proposition 11.7 on page 645 and part (a),

$$M_X(s) = M_{X,Y}(s,0) = \frac{1}{2-s}\left(1 + \frac{1}{(1-s)^2}\right), \qquad s < 1,$$

and

$$M_Y(t) = M_{X,Y}(0,t) = \frac{1}{2-t}\left(1 + \frac{1}{(1-t)^2}\right), \qquad t < 1.$$

d) From parts (a) and (c),

$$\begin{aligned} M_X(s)M_Y(t) &= \frac{1}{2-s}\left(1+\frac{1}{(1-s)^2}\right)\cdot\frac{1}{2-t}\left(1+\frac{1}{(1-t)^2}\right) \\ &= \frac{1}{(2-s)(2-t)}\left(1+\frac{1}{(1-s)^2}\right)\left(1+\frac{1}{(1-t)^2}\right) \\ &\neq \frac{1}{2-s-t}\left(\frac{1}{(1-s)^2}+\frac{1}{(1-t)^2}\right) = M_{X,Y}(s,t). \end{aligned}$$

Therefore, from Proposition 11.9 on page 646, X and Y are not independent random variables.
e) From Proposition 11.10 on page 647 and part (a),

$$M_{X+Y}(t) = M_{X,Y}(t,t) = \frac{1}{2-t-t}\left(\frac{1}{(1-t)^2}+\frac{1}{(1-t)^2}\right) = \frac{1}{(1-t)^3}.$$

Referring now to Table 11.1 on page 634, we see that the MGF of $X+Y$ is that of a $\Gamma(3,1)$ distribution. Hence, by the uniqueness property of MGFs, we conclude that $X+Y \sim \Gamma(3,1)$.

Theory Exercises

11.28
a) We have

$$M_{aX+bY}(t) = \mathcal{E}\Big(e^{t(aX+bY)}\Big) = \mathcal{E}\Big(e^{(at)X+(bt)Y}\Big) = M_{X,Y}(at, bt).$$

b) We have

$$\begin{aligned} M_{a+bX,c+dY}(s,t) &= \mathcal{E}\Big(e^{s(a+bX)+t(c+dY)}\Big) = \mathcal{E}\Big(e^{as+ct}e^{(bs)X+(dt)Y}\Big) \\ &= e^{as+ct}\mathcal{E}\Big(e^{(bs)X+(dt)Y}\Big) = e^{as+ct}M_{X,Y}(bs, dt). \end{aligned}$$

c) We have

$$\begin{aligned} M_{aX+bY,cX+dY}(s,t) &= \mathcal{E}\Big(e^{s(aX+bY)+t(cX+dY)}\Big) = \mathcal{E}\Big(e^{(as+ct)X+(bs+dt)Y}\Big) \\ &= M_{X,Y}(as+ct, bs+dt). \end{aligned}$$

11.29
Analogue of Proposition 11.6: Suppose that the joint MGF of the random variables $X_1, \dots, X_m$ is defined in some open m-dimensional rectangle containing the origin. Then $X_1, \dots, X_m$ have individual and joint moments of all orders, and we have

$$\mathcal{E}\Big(X_1^{j_1}\cdots X_m^{j_m}\Big) = \frac{\partial^{j_1+\cdots+j_m} M_{X_1,\dots,X_m}}{\partial t_1^{j_1}\cdots\partial t_m^{j_m}}(0,\dots,0),$$

for each choice of nonnegative integers j_k $(1 \le k \le m)$.

Analogue of Proposition 11.7: Let $X_1, \dots, X_m$ be random variables defined on the same sample space. Then

$$M_{X_j}(t_j) = M_{X_1,\dots,X_m}(0,\dots,0,t_j,0,\dots,0), \qquad 1 \le j \le m.$$

Thus, for each $1 \le j \le m$, the marginal MGF of X_j is obtained by setting $t_k = 0$ for $k \neq j$ in the joint MGF.

Analogue of Proposition 11.8: If the joint MGF of $X_1, \ldots, X_m$ equals the joint MGF of $Y_1, \ldots, Y_m$ in some open m-dimensional rectangle containing the origin, then $X_1, \ldots, X_m$ have the same joint probability distribution as $Y_1, \ldots, Y_m$.

Analogue of Proposition 11.9: Let $X_1, \ldots, X_m$ be random variables defined on the same sample space. Then $X_1, \ldots, X_m$ are independent if and only if

$$M_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = M_{X_1}(t_1) \cdots M_{X_m}(t_m).$$

In words, m random variables are independent if and only if their joint MGF equals the product of their marginal MGFs.

Analogue of Proposition 11.10: Let $X_1, \ldots, X_m$ be random variables defined on the same sample space. Then

$$M_{X_1+\cdots+X_m}(t) = M_{X_1,\ldots,X_m}(t, \ldots, t).$$

11.30 The joint MGF of $X_{k_1}, \ldots, X_{k_j}$ is obtained by setting $t_i = 0$ for all $i \notin \{k_1, \ldots, k_j\}$ in the joint MGF of $X_1, \ldots, X_m$.

Advanced Exercises

11.31

a) From Exercise 11.28(c) and Equation (11.15) on page 642,

$$\begin{aligned} M_{X+Y,X-Y}(s,t) = M_{X,Y}(s+t, s-t) &= e^{\mu_X(s+t)+\mu_Y(s-t)+\frac{1}{2}\left(\sigma_X^2(s+t)^2+2\rho\sigma_X\sigma_Y(s+t)(s-t)+\sigma_Y^2(s-t)^2\right)} \\ &= e^{(\mu_X+\mu_Y)s+(\mu_X-\mu_Y)t+\frac{1}{2}\left((\sigma_X^2+2\rho\sigma_X\sigma_Y+\sigma_Y^2)s^2+2(\sigma_X^2-\sigma_Y^2)st+(\sigma_X^2-2\rho\sigma_X\sigma_Y+\sigma_Y^2)t^2\right)}. \end{aligned}$$

Thus, again referring to Equation (11.15), we see that

$$(X, Y) \sim \mathcal{BVN}\left(\mu_X + \mu_Y, \sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2, \mu_X - \mu_Y, \sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2, \rho^*\right),$$

where

$$\rho^* = \rho(X+Y, X-Y) = \frac{\sigma_X^2 - \sigma_Y^2}{\sqrt{\left(\sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2\right)\left(\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2\right)}}.$$

b) From part (a) and Proposition 10.9 on page 613, $X + Y$ and $X - Y$ are independent if and only if they are uncorrelated. Hence, $X + Y$ and $X - Y$ are independent if and only if $\rho^* = 0$, that is, if and only if $\operatorname{Var}(X) = \operatorname{Var}(Y)$.

11.32 Recall from Proposition 6.7 on page 277 that

$$p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \binom{n}{x_1, \ldots, x_m} p_1^{x_1} \cdots p_m^{x_m},$$

if $x_1, \ldots, x_m$ are nonnegative integers whose sum is n, and $p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = 0$ otherwise.

a) We have

$$\begin{aligned} M_{X_1,\ldots,X_m}(t_1,\ldots,t_m) &= \sum_{(x_1,\ldots,x_m)}\!\!\cdots\sum e^{t_1x_1+\cdots+t_mx_m}\, p_{X_1,\ldots,X_m}(x_1,\ldots,x_m) \\ &= \sum_{x_1+\cdots+x_m=n}\!\!\cdots\sum e^{t_1x_1+\cdots+t_mx_m}\binom{n}{x_1,\ldots,x_m} p_1^{x_1}\cdots p_m^{x_m} \\ &= \sum_{x_1+\cdots+x_m=n}\!\!\cdots\sum \binom{n}{x_1,\ldots,x_m}\left(p_1e^{t_1}\right)^{x_1}\cdots\left(p_me^{t_m}\right)^{x_m} \\ &= \left(p_1e^{t_1}+\cdots+p_me^{t_m}\right)^n, \end{aligned}$$

where the last equality follows from the multinomial theorem, as presented in Exercise 3.49 on page 109.
b) From Exercise 11.29, we have, for each $1 \le j \le m$,

$$M_{X_j}(t_j) = M_{X_1,\ldots,X_m}(0,\ldots,0,t_j,0,\ldots,0) = \left(p_je^{t_j} + \sum_{k\neq j} p_k\right)^n = \left(p_je^{t_j}+1-p_j\right)^n.$$

Hence, from Table 11.1, we see that $X_j \sim \mathcal{B}(n, p_j)$.
c) We have, for each $1 \le j \le m$,

$$\frac{\partial M_{X_1,\ldots,X_m}}{\partial t_j}(t_1,\ldots,t_m) = n\left(p_1e^{t_1}+\cdots+p_me^{t_m}\right)^{n-1} p_je^{t_j}$$

and, for $k \neq \ell$,

$$\frac{\partial^2 M_{X_1,\ldots,X_m}}{\partial t_k \partial t_\ell}(t_1,\ldots,t_m) = n(n-1)\left(p_1e^{t_1}+\cdots+p_me^{t_m}\right)^{n-2} p_ke^{t_k}\,p_\ell e^{t_\ell}.$$

Hence, from Exercise 11.29,

$$\mathcal{E}\big(X_j\big) = \frac{\partial M_{X_1,\ldots,X_m}}{\partial t_j}(0,\ldots,0) = n\cdot 1^{n-1}\cdot p_j\cdot 1 = np_j$$

and, for $k \neq \ell$,

$$\mathcal{E}(X_kX_\ell) = \frac{\partial^2 M_{X_1,\ldots,X_m}}{\partial t_k \partial t_\ell}(0,\ldots,0) = n(n-1)\cdot 1^{n-2}p_k\cdot 1\cdot p_\ell\cdot 1 = n(n-1)p_kp_\ell.$$

Consequently,

$$\operatorname{Cov}(X_k, X_\ell) = \mathcal{E}(X_kX_\ell) - \mathcal{E}(X_k)\,\mathcal{E}(X_\ell) = n(n-1)p_kp_\ell - np_k\cdot np_\ell = -np_kp_\ell,$$

which agrees with Equation (∗) on page 375.
d) If $k = \ell$, we know that $\rho(X_k, X_\ell) = 1$. We can use either Exercise 11.29 or part (b) to show that $\operatorname{Var}\big(X_j\big) = np_j(1-p_j)$. Therefore, from part (c), if $k \neq \ell$,

$$\rho(X_k, X_\ell) = \frac{\operatorname{Cov}(X_k, X_\ell)}{\sqrt{\operatorname{Var}(X_k)\cdot\operatorname{Var}(X_\ell)}} = \frac{-np_kp_\ell}{\sqrt{np_k(1-p_k)\cdot np_\ell(1-p_\ell)}} = -\sqrt{\frac{p_kp_\ell}{(1-p_k)(1-p_\ell)}}.$$

11.33 The result of this exercise is needed for the proof of Proposition 11.9 on page 646. First we show that X and U have the same marginal distributions:

$$P(X \in A) = P(X \in A,\ Y \in \mathcal{R}) = P(U \in A,\ V \in \mathcal{R}) = P(U \in A).$$

Likewise, Y and V have the same marginal distributions. Using the independence of U and V, we get

$$P(X \in A,\ Y \in B) = P(U \in A,\ V \in B) = P(U \in A)P(V \in B) = P(X \in A)P(Y \in B).$$

Thus, X and Y are independent random variables.

11.34
a) To begin, we note that

$$\mu_X s + \mu_Y t = [\,\mu_X \quad \mu_Y\,]\begin{bmatrix} s \\ t \end{bmatrix} = \boldsymbol{\mu}'\mathbf{t}$$

and

$$\begin{aligned}
\sigma_X^2 s^2 + 2\rho\sigma_X\sigma_Y st + \sigma_Y^2 t^2 &= \operatorname{Var}(X)\, s^2 + 2\operatorname{Cov}(X, Y)\, st + \operatorname{Var}(Y)\, t^2 \\
&= [\,s \quad t\,]\begin{bmatrix} s\operatorname{Var}(X) + t\operatorname{Cov}(X, Y) \\ s\operatorname{Cov}(Y, X) + t\operatorname{Var}(Y) \end{bmatrix} \\
&= [\,s \quad t\,]\begin{bmatrix} \operatorname{Var}(X) & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(Y, X) & \operatorname{Var}(Y) \end{bmatrix}\begin{bmatrix} s \\ t \end{bmatrix} = \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}.
\end{aligned}$$

Therefore, in view of Equation (11.15) on page 642,

$$M_{X,Y}(s, t) = e^{\mu_X s + \mu_Y t + \frac{1}{2}(\sigma_X^2 s^2 + 2\rho\sigma_X\sigma_Y st + \sigma_Y^2 t^2)} = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}.$$

b) Let X be a normal random variable, say, $X \sim \mathcal{N}(\mu, \sigma^2)$. Define the following 1×1 vectors and matrix: $\mathbf{t} = [t]$, $\boldsymbol{\mu} = [\mu] = [\mu_X]$, and $\boldsymbol{\Sigma} = [\sigma^2] = [\operatorname{Var}(X)]$. Then

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} = e^{t\mu_X + \frac{1}{2}t\operatorname{Var}(X)t} = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}.$$

c) Define $\boldsymbol{\Sigma} = \left[\operatorname{Cov}(X_i, X_j)\right]$, an $m \times m$ matrix, and let

$$\mathbf{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_m} \end{bmatrix}.$$

Then, in view of parts (a) and (b), an educated guess for the matrix form of the joint MGF of an m-variate normal distribution is

$$M_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}.$$

11.35 As for real-valued random variables, a complex-valued random variable Z has finite expectation if and only if $\mathcal{E}(|Z|) < \infty$, where, in this case, $|z|$ denotes the modulus of a complex number z. Because $\left|e^{i(sX+tY)}\right| = 1$, we have $\mathcal{E}\left(\left|e^{i(sX+tY)}\right|\right) = \mathcal{E}(1) = 1 < \infty$. Consequently, $e^{i(sX+tY)}$ has finite expectation for all $s, t \in \mathcal{R}$.

11.36
Analogue of Proposition 11.6: If (X, Y) has a joint moment of order (j, k), then

$$\mathcal{E}\left(X^j Y^k\right) = (-i)^{j+k}\frac{\partial^{j+k}\phi_{X,Y}}{\partial s^j\,\partial t^k}(0, 0).$$

For the proof, we note that, by interchanging the order of differentiation and expectation, we get

$$\begin{aligned}\frac{\partial^{j+k}\phi_{X,Y}}{\partial s^j \partial t^k}(s,t) &= \frac{\partial^{j+k}\mathcal{E}\big(e^{i(sX+tY)}\big)}{\partial s^j \partial t^k} = \mathcal{E}\left(\frac{\partial^{j+k} e^{i(sX+tY)}}{\partial s^j \partial t^k}\right) \\ &= \mathcal{E}\Big(i^{j+k} X^j Y^k e^{i(sX+tY)}\Big) = i^{j+k}\mathcal{E}\Big(X^j Y^k e^{i(sX+tY)}\Big).\end{aligned}$$

Setting $s = t = 0$, we now get the required result.

Analogue of Proposition 11.7: Let X and Y be random variables defined on the same sample space. Then

$$\phi_X(s) = \phi_{X,Y}(s,0) \qquad \text{and} \qquad \phi_Y(t) = \phi_{X,Y}(0,t).$$

Thus, the marginal CHF of X is obtained by setting $t = 0$ in the joint CHF, and the marginal CHF of Y is obtained by setting $s = 0$ in the joint CHF. For the proof, we have

$$\phi_X(s) = \mathcal{E}\Big(e^{isX}\Big) = \mathcal{E}\Big(e^{i(sX+0\cdot Y)}\Big) = \phi_{X,Y}(s,0).$$

Arguing similarly, we find that $\phi_Y(t) = \phi_{X,Y}(0,t)$.

Analogue of Proposition 11.9: Let X and Y be random variables defined on the same sample space. Then X and Y are independent if and only if

$$\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t), \qquad s,t \in \mathcal{R}. \tag{$*$}$$

In words, two random variables are independent if and only if their joint CHF equals the product of their marginal CHFs. For the proof, first suppose that X and Y are independent random variables. Proposition 6.9 on page 291 implies that e^{isX} and e^{itY} are also independent random variables for each $s, t \in \mathcal{R}$. Now using the fact that, for independent random variables, the expected value of the product equals the product of the expected values, we conclude that

$$\phi_{X,Y}(s,t) = \mathcal{E}\Big(e^{i(sX+tY)}\Big) = \mathcal{E}\Big(e^{isX}e^{itY}\Big) = \mathcal{E}\Big(e^{isX}\Big)\,\mathcal{E}\Big(e^{itY}\Big) = \phi_X(s)\phi_Y(t).$$

Hence, Equation $(*)$ holds.

Conversely, suppose that Equation $(*)$ holds. Let U and V be independent random variables with the same (marginal) probability distributions as X and Y, respectively. Then, by Equation $(*)$ and what we just proved,

$$\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t) = \phi_U(s)\phi_V(t) = \phi_{U,V}(s,t).$$

Hence, the joint CHF of X and Y is the same as that of U and V. Therefore, by the uniqueness property of joint CHFs, the joint probability distribution of X and Y is the same as that of U and V. Because U and V are independent random variables, it now follows from Exercise 11.33 that X and Y are also independent random variables.

Analogue of Proposition 11.10: Let X and Y be random variables defined on the same sample space. Then

$$\phi_{X+Y}(t) = \phi_{X,Y}(t,t).$$

For the proof, we have

$$\phi_{X+Y}(t) = \mathcal{E}\Big(e^{it(X+Y)}\Big) = \mathcal{E}\Big(e^{i(tX+tY)}\Big) = \phi_{X,Y}(t,t).$$

11.37

a) Formally,

$$\phi_{X,Y}(s,t) = \mathcal{E}\Big(e^{i(sX+tY)}\Big) = \mathcal{E}\Big(e^{isX+itY}\Big) = \mathcal{E}\Big(e^{(is)X+(it)Y}\Big) = M_{X,Y}(is,it).$$

b) From part (a) and Equation (11.15) on page 642,

$$\begin{aligned}\phi_{X,Y}(s,t) &= M_{X,Y}(is,it) = e^{\mu_X(is)+\mu_Y(it)+\frac{1}{2}\left(\sigma_X^2(is)^2+2\rho\sigma_X\sigma_Y(is)(it)+\sigma_Y^2(it)^2\right)} \\ &= e^{i\left(\mu_X s+\mu_Y t\right)-\frac{1}{2}\left(\sigma_X^2 s^2+2\rho\sigma_X\sigma_Y st+\sigma_Y^2 t^2\right)}.\end{aligned}$$

c) Applying elementary calculus and the result from part (b), we get

$$\begin{aligned}\frac{\partial\phi_{X,Y}}{\partial s}(s,t) &= \left(i\mu_X - \frac{1}{2}\left(2\sigma_X^2 s + 2\rho\sigma_X\sigma_Y t\right)\right)\phi_{X,Y}(s,t); \\ \frac{\partial\phi_{X,Y}}{\partial t}(s,t) &= \left(i\mu_Y - \frac{1}{2}\left(2\rho\sigma_X\sigma_Y s + 2\sigma_Y^2 t\right)\right)\phi_{X,Y}(s,t); \\ \frac{\partial^2\phi_{X,Y}}{\partial s^2}(s,t) &= \left(\left(i\mu_X - \frac{1}{2}\left(2\sigma_X^2 s + 2\rho\sigma_X\sigma_Y t\right)\right)^2 - \sigma_X^2\right)\phi_{X,Y}(x,y); \\ \frac{\partial^2\phi_{X,Y}}{\partial t^2}(s,t) &= \left(\left(i\mu_Y - \frac{1}{2}\left(2\rho\sigma_X\sigma_Y s + 2\sigma_Y^2 t\right)\right)^2 - \sigma_Y^2\right)\phi_{X,Y}(s,t); \\ \frac{\partial^2\phi_{X,Y}}{\partial s\partial t}(s,t) &= \left(\left(i\mu_Y - \frac{1}{2}\left(2\rho\sigma_X\sigma_Y s + 2\sigma_Y^2 t\right)\right)\left(i\mu_X - \frac{1}{2}\left(2\sigma_X^2 s + 2\rho\sigma_X\sigma_Y t\right)\right)\right. \\ &\qquad \left. - \rho\sigma_X\sigma_Y\right)\phi_{X,Y}(s,t).\end{aligned}$$

Referring to Exercise 11.36 and noting that $\phi_{X,Y}(0,0) = 1$, we conclude that

$$\begin{aligned}\mathcal{E}(X) &= \mathcal{E}\left(X^1Y^0\right) = (-i)^{1+0}\frac{\partial\phi_{X,Y}}{\partial s}(0,0) = (-i)\cdot i\mu_X = \mu_X; \\ \mathcal{E}(Y) &= \mathcal{E}\left(X^0Y^1\right) = (-i)^{0+1}\frac{\partial\phi_{X,Y}}{\partial t}(0,0) = (-i)\cdot i\mu_Y = \mu_Y; \\ \mathcal{E}\left(X^2\right) &= \mathcal{E}\left(X^2Y^0\right) = (-i)^{2+0}\frac{\partial^2\phi_{X,Y}}{\partial s^2}(0,0) = (-1)\left((i\mu_X)^2 - \sigma_X^2\right) = \mu_X^2 + \sigma_X^2; \\ \mathcal{E}\left(Y^2\right) &= \mathcal{E}\left(X^0Y^2\right) = (-i)^{0+2}\frac{\partial^2\phi_{X,Y}}{\partial t^2}(0,0) = (-1)\left((i\mu_Y)^2 - \sigma_Y^2\right) = \mu_Y^2 + \sigma_Y^2; \\ \mathcal{E}(XY) &= \mathcal{E}\left(X^1Y^1\right) = (-i)^{1+1}\frac{\partial^2\phi_{X,Y}}{\partial s\partial t}(0,0) = (-1)\left((i\mu_Y)(i\mu_X) - \rho\sigma_X\sigma_Y\right) = \mu_X\mu_Y + \rho\sigma_X\sigma_Y.\end{aligned}$$

From these equations, $\mathcal{E}(X) = \mu_X$, $\mathcal{E}(Y) = \mu_Y$,

$$\begin{aligned}\mathrm{Var}(X) &= \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = \mu_X^2 + \sigma_X^2 - (\mu_X)^2 = \sigma_X^2, \\ \mathrm{Var}(Y) &= \mathcal{E}\left(Y^2\right) - \left(\mathcal{E}(Y)\right)^2 = \mu_Y^2 + \sigma_Y^2 - (\mu_Y)^2 = \sigma_Y^2,\end{aligned}$$

and

$$\mathrm{Cov}(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \mu_X\mu_Y + \rho\sigma_X\sigma_Y - \mu_X\mu_Y = \rho\sigma_X\sigma_Y.$$

11.38

a) The joint CHF of $X_1, \ldots, X_m$ is defined by

$$\phi_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = \mathcal{E}\Big(e^{i(t_1X_1+\cdots+t_mX_m)}\Big), \qquad t_1, \ldots, t_m \in \mathcal{R}.$$

b) *Analogue of Proposition 11.6:* If $(X_1, \ldots, X_m)$ has a joint moment of order $(j_1, \ldots, j_m)$, then

$$\mathcal{E}\Big(X_1^{j_1} \cdots X_m^{j_m}\Big) = (-i)^{j_1+\cdots+j_m} \frac{\partial^{j_1+\cdots+j_m}\phi_{X_1,\ldots,X_m}}{\partial t_1^{j_1} \cdots \partial t_m^{j_m}}(0, \ldots, 0).$$

Analogue of Proposition 11.7: Let $X_1, \ldots, X_m$ be random variables defined on the same sample space. Then

$$\phi_{X_j}(t_j) = \phi_{X_1,\ldots,X_m}(0, \ldots, 0, t_j, 0, \ldots, 0), \qquad 1 \le j \le m.$$

Thus, for each $1 \le j \le m$, the marginal CHF of X_j is obtained by setting $t_k = 0$ for $k \ne j$ in the joint CHF.

Analogue of Proposition 11.8: If the joint CHF of $X_1, \ldots, X_m$ equals the joint CHF of $Y_1, \ldots, Y_m$, then $X_1, \ldots, X_m$ have the same joint probability distribution as $Y_1, \ldots, Y_m$.

Analogue of Proposition 11.9: Let $X_1, \ldots, X_m$ be random variables defined on the same sample space. Then $X_1, \ldots, X_m$ are independent if and only if

$$\phi_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = \phi_{X_1}(t_1) \cdots \phi_{X_m}(t_m).$$

In words, m random variables are independent if and only if their joint CHF equals the product of their marginal CHFs.

Analogue of Proposition 11.10: Let $X_1, \ldots, X_m$ be random variables defined on the same sample space. Then

$$\phi_{X_1+\cdots+X_m}(t) = \phi_{X_1,\ldots,X_m}(t, \ldots, t).$$

11.3 Laws of Large Numbers

Basic Exercises

11.39 In this context, the strong law of large numbers states that $\lim_{n\to\infty}(X_1 + \cdots + X_n)/n = 1/2$ with probability 1, not that the limit always equals 1/2. In other words, the outcomes for which the average value doesn't converge to 1/2 constitute an event that has zero probability.

11.40 For each $j \in \mathcal{N}$, let $Y_j = X_j^r$. As $X_1, X_2, \ldots$ are independent random variables, all having the same probability distribution as X, we have that $Y_1, Y_2, \ldots$ are independent random variables, all having the same probability distribution as X^r. By assumption, X^r has finite mean. Because $\mathcal{E}(Y_j) = \mathcal{E}(X^r)$ for all $j \in \mathcal{N}$, the weak law of large numbers implies that

$$\lim_{n\to\infty} P\left(\left|\frac{X_1^r + \cdots + X_n^r}{n} - \mathcal{E}(X^r)\right| < \epsilon\right) = \lim_{n\to\infty} P\left(\left|\frac{Y_1 + \cdots + Y_n}{n} - \mathcal{E}(X^r)\right| < \epsilon\right) = 1,$$

for each $\epsilon > 0$. In other words, the rth sample moment is a consistent estimator of the rth moment.

11.41

a) For $n \in \mathcal{N}$,

$$Y_n = s\prod_{j=1}^{n} \frac{Y_j}{Y_{j-1}} = s\prod_{j=1}^{n} X_j.$$

Consequently,

$$\ln\left(\frac{Y_n}{s}\right)^{1/n} = \frac{1}{n}\ln\left(\frac{Y_n}{s}\right) = \frac{1}{n}\ln\left(\prod_{j=1}^{n} X_j\right) = \frac{1}{n}\sum_{j=1}^{n}\ln X_j = \frac{\ln X_1 + \cdots + \ln X_n}{n}.$$

Applying the strong law of large numbers, Theorem 11.2 on page 654, to the sequence of independent and identically distributed random variables $\ln X_1, \ln X_2, \ldots,$ we get

$$\lim_{n\to\infty}\ln\left(\frac{Y_n}{s}\right)^{1/n} = \lim_{n\to\infty}\frac{\ln X_1 + \cdots + \ln X_n}{n} = \mathcal{E}(\ln X),$$

with probability 1. Consequently, with probability 1,

$$\lim_{n\to\infty}\left(\frac{Y_n}{s}\right)^{1/n} = \lim_{n\to\infty} e^{\ln\left(\frac{Y_n}{s}\right)^{1/n}} = e^{\mathcal{E}(\ln X)}.$$

b) From part (a), we have, for large n, that $(Y_n/s)^{1/n} \approx e^{\mathcal{E}(\ln X)}$ or, equivalently,

$$Y_n = s\left(\left(\frac{Y_n}{s}\right)^{1/n}\right)^n \approx s\left(e^{\mathcal{E}(\ln X)}\right)^n = se^{n\mathcal{E}(\ln X)}.$$

c) We consider each distribution given for X in turn.

- $X \sim \mathcal{U}(0, 1)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x)\cdot 1\,dx = \int_0^1 \ln x\,dx = -1.$$

  Hence, with probability 1,

$$\lim_{n\to\infty}\left(\frac{Y_n}{s}\right)^{1/n} = e^{-1},$$

  and we have $Y_n \approx se^{-n}$ for large n.

- $X \sim \text{Beta}(2, 1)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x)\cdot 2x^{2-1}(1-x)^{1-1}\,dx = 2\int_0^1 x\ln x\,dx = 2\cdot\left(-\frac{1}{4}\right) = -\frac{1}{2}.$$

  Hence, with probability 1,

$$\lim_{n\to\infty}\left(\frac{Y_n}{s}\right)^{1/n} = e^{-1/2},$$

  and we have $Y_n \approx se^{-n/2}$ for large n.

- $X \sim \text{Beta}(1, 2)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x)\cdot 2x^{1-1}(1-x)^{2-1}\,dx = 2\int_0^1 (1-x)\ln x\,dx = 2\cdot\left(-\frac{3}{4}\right) = -\frac{3}{2}.$$

Hence, with probability 1,

$$\lim_{n\to\infty}\left(\frac{Y_n}{s}\right)^{1/n} = e^{-3/2},$$

and we have $Y_n \approx se^{-3n/2}$ for large n.

- $X \sim \text{Beta}(2, 2)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x)\cdot 6x^{2-1}(1-x)^{2-1}\,dx = 6\int_0^1 x(1-x)\ln x\,dx = 6\cdot\left(-\frac{5}{36}\right) = -\frac{5}{6}.$$

Hence, with probability 1,

$$\lim_{n\to\infty}\left(\frac{Y_n}{s}\right)^{1/n} = e^{-5/6},$$

and we have $Y_n \approx se^{-5n/6}$ for large n.

11.42 For each $x \in \mathcal{R}$ and $j \in \mathcal{N}$, let $Y_j(x) = I_{\{X_j \le x\}}$. Then $Y_1(x), Y_2(x), \dots$ are independent Bernoulli random variables with common parameter $F_X(x)$ and, hence, with common mean $F_X(x)$. Noting that $\hat{F}_n(x) = \big(Y_1(x) + \cdots + Y_n(x)\big)/n$, the strong law of large numbers yields

$$\lim_{n\to\infty}\hat{F}_n(x) = \lim_{n\to\infty}\frac{Y_1(x)+\cdots+Y_n(x)}{n} = F_X(x),$$

with probability 1. Thus, for large n, the CDF of X can be approximated by the corresponding empirical distribution function.

11.43

a) Because g is Riemann integrable, it is bounded, say, by M. Thus,

$$\int_{-\infty}^{\infty}|g(x)|f_X(x)\,dx = \int_0^1 |g(x)|\cdot 1\,dx = \int_0^1 |g(x)|\,dx \le \int_0^1 M\,dx = M < \infty.$$

Thus, by the FEF for continuous random variables, $g(X)$ has finite expectation.

b) The random variables $g(X_1), g(X_2), \dots$ all have the same distribution as the random variable $g(X)$ of part (a). From Proposition 6.13 on page 297, they are also independent. Moreover, by part (a), they have finite mean $\mathcal{E}\big(g(X)\big)$. Therefore, by the strong law of large numbers and the FEF,

$$\lim_{n\to\infty}\frac{g(X_1)+\cdots+g(X_n)}{n} = \mathcal{E}\big(g(X)\big) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx = \int_0^1 g(x)\cdot 1\,dx = \int_0^1 g(x)\,dx,$$

with probability 1.

c) For a large $n \in \mathcal{N}$, use a basic random number generator to obtain n (independent) uniform numbers between 0 and 1—say, $x_1, \dots, x_n$. In view of the result in part (b), we then have

$$\int_0^1 g(x)\,dx \approx \frac{g(x_1)+\cdots+g(x_n)}{n}.$$

d) Answers will vary. We used statistical software to obtain 1000 uniform numbers between 0 and 1. Next we cubed each of those numbers. Then we took the average value of the 1000 cubed numbers. The result we obtained was 0.248, which is quite close to the exact value of 0.25 for $\int_0^1 x^3\,dx$.

e) Answers will vary. Here we want to estimate

$$P(0 \le Z \le 1) = \int_0^1 \phi(x)\,dx = \int_0^1 \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx.$$

We used statistical software to obtain 1000 uniform numbers between 0 and 1. Next we computed the value of the standard normal PDF for each of those numbers. Then we took the average of the 1000 values of the standard normal PDF. The result we got was 0.3414, which is quite close to the value of 0.3413 obtained by using Table I.

f) Answers will vary. Repeat the processes used in parts (d) and (e), except use 10,000 uniform numbers instead of 1000 uniform numbers.

11.44

a) As we know, if $X \sim \mathcal{U}(0, 1)$, then $Y = a + (b - a)X \sim \mathcal{U}(a, b)$. Because g is Riemann integrable on $[a, b]$, it is bounded thereon and, hence, by an argument similar to that given in the solution to Exercise 11.43(a), $g(Y)$ has finite expectation. Let $X_1, X_2, \ldots$ be independent $\mathcal{U}(0, 1)$ random variables, and set $Y_j = a + (b - a)X_j$ for $j \in \mathcal{N}$. Then $Y_1, Y_2, \ldots$ are independent $\mathcal{U}(a, b)$ random variables and, consequently, $g(Y_1), g(Y_2), \ldots$ are independent and identically distributed random variables, each with the same distribution as $g(Y)$. Applying the strong law of large numbers and the FEF, we get

$$\lim_{n\to\infty} \frac{g(Y_1) + \cdots + g(Y_n)}{n} = \mathcal{E}\big(g(Y)\big) = \int_{-\infty}^{\infty} g(y) f_Y(y)\, dy = \frac{1}{b-a}\int_a^b g(y)\, dy$$

with probability 1. From this result, we see that the following procedure can be applied. Use a basic random number generator to obtain n (independent) uniform numbers between 0 and 1—say, $x_1, \ldots, x_n$, where n is large. Set $y_j = a + (b - a)x_j$ for $j = 1, 2, \ldots, n$. Then

$$\frac{g(y_1) + \cdots + g(y_n)}{n} \approx \frac{1}{b-a}\int_a^b g(x)\, dx,$$

or, equivalently,

$$\int_a^b g(x)\, dx \approx (b-a)\left(\frac{g(y_1) + \cdots + g(y_n)}{n}\right).$$

b) Answers will vary. We used statistical software to obtain 1000 uniform numbers between 0 and 1. Next we transformed those numbers by using the relation $y = 1 + 2x$ and then cubed each of the transformed numbers. Finally, we took the average value of the 1000 cubed transformed numbers and multiplied that average by 2 $(= 3 - 1)$. The result we obtained was 20.265, which is reasonably close to the exact value of 20 for $\int_1^3 x^3\, dx$.

c) Answers will vary. Here we want to estimate

$$P(-1 \le Z \le 2) = \int_{-1}^{2} \phi(x)\, dx = \int_{-1}^{2} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

We used statistical software to obtain 1000 uniform numbers between 0 and 1. Next we transformed those numbers by using the relation $y = -1 + 3x$ and then computed the value of the standard normal PDF for each of the transformed numbers. Finally, we took the average of the 1000 values of the standard normal PDF of the transformed numbers and multiplied that average by 3 $\big(= 2 - (-1)\big)$. The result we got was 0.8274, which is reasonably close to the value of 0.8185 obtained by using Table I.

d) Answers will vary. Repeat the processes used in parts (b) and (c), except use 10,000 uniform numbers instead of 1000 uniform numbers.

11.45

a) To begin, we recall that the PDF of a standard Cauchy random variable is

$$f(x) = \frac{1}{\pi\big(1 + x^2\big)}, \qquad x \in \mathcal{R}.$$

For convenience, set $\bar{X}_n = (X_1 + \cdots + X_n)/n$. From Exercise 11.20(b), the random variable $\bar{X}_n$ has the standard Cauchy distribution for each $n \in \mathcal{N}$. Hence, by the FPF,

$$P\big(|\bar{X}_n - c| < \epsilon\big) = \int_{c-\epsilon}^{c+\epsilon} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi}\big(\arctan(c+\epsilon) - \arctan(c-\epsilon)\big),$$

for all $n \in \mathcal{N}$. As the right side of the previous display doesn't depend on n, it also is the limit as $n \to \infty$ of the left side; that is,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - c\right| < \epsilon\right) = \frac{1}{\pi}\big(\arctan(c+\epsilon) - \arctan(c-\epsilon)\big).$$

b) The weak law of large numbers states that, if $X_1, X_2, \ldots$ are independent and identically distributed random variables with finite expectation, μ, then, for each $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \epsilon\right) = 1.$$

Thus, for such random variables, the limit in Expression $(*)$ equals 1 when $c = \mu$. However, the weak law of large numbers doesn't apply to standard Cauchy random variables because they don't have finite expectation.

11.46

b) The mean of a $\mathcal{U}(0, 1)$ distribution is $1/2$. Hence, from the strong law of large numbers, we would expect the average value of the 10,000 numbers obtained in part (a) to be roughly $1/2$.
c) Answers will vary. We performed the simulation and got an average value of 0.506, which is quite close to $1/2$.

11.47

b) The mean of a $\mathcal{N}(0, 1)$ distribution is 0. Hence, from the strong law of large numbers, we would expect the average value of the 10,000 numbers obtained in part (a) to be roughly 0.
c) Answers will vary. We performed the simulation and got an average value of -0.006, which is quite close to 0.

11.48

a) We first note that the CDF of a standard Cauchy random variable is

$$F(x) = \int_{-\infty}^{x} \frac{dt}{\pi\big(1+t^2\big)} = \frac{1}{2} + \frac{1}{\pi}\arctan x, \qquad x \in \mathcal{R}.$$

Solving for x in the equation $y = 1/2 + (\arctan x)/\pi$, we find that $F^{-1}(y) = \tan\big(\pi(2y-1)/2\big)$. Hence, from Proposition 8.16(b), if $U \sim \mathcal{U}(0, 1)$, then $F^{-1}(U) = \tan\big(\pi(2U-1)/2\big) \sim \mathcal{C}(0, 1)$. Consequently, to obtain 10,000 observations of a standard Cauchy random variable, we first use a basic random number generator to get 10,000 uniform numbers between 0 and 1 and then transform each of those numbers by using the relation $x = \tan\big(\pi(2u-1)/2\big)$.
b) No, we cannot guess the average value of the 10,000 numbers obtained in part (a). Indeed, by Exercise 11.20, the average value of 10,000 observations of a standard Cauchy random variable also has the standard Cauchy distribution.

c) Answers will vary. We performed the simulation in part (a) 20 times and obtained the following 20 averages:

−0.168	0.380	−2.459	−4.208	−1.730
2.720	0.062	1.586	−0.935	3.034
−0.643	0.746	−0.794	−1.433	2.097
−0.153	0.119	0.645	0.309	−0.850

The strong law of large numbers states that, if $X_1, X_2, \ldots$ are independent and identically distributed random variables with finite expectation, μ, then,

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu,$$

with probability 1. Thus, for large n, the average value of n independent observations of a random variable X with finite expectation, μ, should be roughly equal to μ. In particular, if we repeat the experiment of making n independent observations of X, the average values should all be close to each other. As we see from the preceding table, this phenomenon does not occur for a standard Cauchy random variable. This is not a problem, however, as the strong law of large numbers doesn't apply to a standard Cauchy random variable because it doesn't have finite expectation.

Theory Exercises

11.49 Let $S_n = X_1 + \cdots + X_n$. From properties of expected value,

$$\mathcal{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\mathcal{E}(S_n) = \frac{1}{n}\mathcal{E}(X_1 + \cdots + X_n) = \frac{\mathcal{E}(X_1) + \cdots + \mathcal{E}(X_n)}{n}.$$

Hence, by Chebyshev's inequality, Proposition 7.11 on page 360, for each $\epsilon > 0$ and $n \in \mathcal{N}$,

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mathcal{E}(X_1) + \cdots + \mathcal{E}(X_n)}{n}\right| \geq \epsilon\right)$$
$$= P\left(\left|\frac{S_n}{n} - \mathcal{E}\left(\frac{S_n}{n}\right)\right| \geq \epsilon\right) \leq \frac{\operatorname{Var}(S_n/n)}{\epsilon^2} = \frac{\operatorname{Var}(S_n)}{n^2\epsilon^2} = \frac{1}{\epsilon^2}\left(\frac{\operatorname{Var}(X_1 + \cdots + X_n)}{n^2}\right).$$

By assumption, the last term in the preceding display converges to 0 as $n \to \infty$. The required result now follows from the complementation rule.

11.50 Let M be such that $\operatorname{Var}(X_j) \leq M$ for all $j \in \mathcal{N}$. From Equation (10.31) on page 589 and the uncorrelatedness of the X_js, we get

$$\operatorname{Var}(X_1 + \cdots + X_n) = \sum_{k=1}^{n} \operatorname{Var}(X_k) + 2\sum_{i<j}\sum \operatorname{Cov}(X_i, X_j) = \sum_{k=1}^{n} \operatorname{Var}(X_k) \leq nM.$$

Hence,

$$0 \leq \lim_{n\to\infty} \frac{\operatorname{Var}(X_1 + \cdots + X_n)}{n^2} \leq \lim_{n\to\infty} \frac{nM}{n^2} = M \lim_{n\to\infty} \frac{1}{n} = 0.$$

Thus, $\lim_{n\to\infty} \operatorname{Var}(X_1 + \cdots + X_n)/n^2 = 0$. The required result now follows from Markov's weak law of large numbers, as presented in Exercise 11.49.

Advanced Exercises

11.51

a) Using the independence of the X_js, we get

$$\begin{aligned}\phi_n(t) &= \mathcal{E}\Big(e^{it(X_1+\cdots+X_n)/n}\Big) = \mathcal{E}\Big(e^{i(t/n)X_1+\cdots+i(t/n)X_n}\Big) = \mathcal{E}\Big(e^{i(t/n)X_1}\cdots e^{i(t/n)X_n}\Big) \\ &= \mathcal{E}\Big(e^{i(t/n)X_1}\Big)\cdots\mathcal{E}\Big(e^{i(t/n)X_n}\Big) = \phi_{X_1}(t/n)\cdots\phi_{X_n}(t/n) = \big(\phi(t/n)\big)^n .\end{aligned}$$

b) Referring to part (a), we have

$$\ln\phi_n(t) = \ln\big(\phi(t/n)\big)^n = n\ln\phi(t/n).$$

Applying L'Hôpital's rule and referring to Exercise 11.18, we find that

$$\begin{aligned}\lim_{n\to\infty}\ln\phi_n(t) &= \lim_{n\to\infty}\frac{\ln\phi(t/n)}{1/n} = \lim_{n\to\infty}\frac{-\big(\phi(t/n)\big)^{-1}\phi'(t/n)\,\big(t/n^2\big)}{-1/n^2} \\ &= \big(\phi(0)\big)^{-1}\phi'(0)t = i\mu t,\end{aligned}$$

for all $t\in\mathcal{R}$.

c) Noting that $e^{i\mu t}$ is the CHF of the constant random variable μ, we deduce from the continuity theorem for CHFs (the analogue of Proposition 11.5 on page 638) that $(X_1+\cdots+X_n)/n$ converges in distribution to μ.

d) Let X be the random variable that equals μ (with probability 1). Then,

$$F_X(x) = \begin{cases} 0, & \text{if } x<\mu; \\ 1, & \text{if } x\ge\mu. \end{cases}$$

Note that the only discontinuity point of F_X is at μ. Let F_n denote the CDF of $\bar{X}_n=(X_1+\cdots+X_n)/n$. From part (c), we know that $F_n(x)\to F_X(x)$ as $n\to\infty$ for all $x\neq\mu$. Now let $\epsilon>0$. Then, for each $n\in\mathcal{N}$,

$$P\big(|\bar{X}_n-\mu|<\epsilon\big) \ge P\big(\mu-\epsilon/2<\bar{X}_n\le\mu+\epsilon/2\big) = F_n(\mu+\epsilon/2)-F_n(\mu-\epsilon/2).$$

Therefore,

$$\begin{aligned}1 &\ge \lim_{n\to\infty}P\big(|\bar{X}_n-\mu|<\epsilon\big) \ge \lim_{n\to\infty}\big(F_n(\mu+\epsilon/2)-F_n(\mu-\epsilon/2)\big) \\ &= F_X(\mu+\epsilon/2)-F_X(\mu-\epsilon/2) = 1-0 = 1.\end{aligned}$$

Hence, $\lim_{n\to\infty}P\big(|\bar{X}_n-\mu|<\epsilon\big)=1$, which is Equation (11.26) on page 651.

11.52 As usual, we let $\bar{X}_n=(X_1+\cdots+X_n)/n$.

a) For convenience, set

$$E=\Big\{\lim_{n\to\infty}\bar{X}_n=\mu\Big\} \qquad \text{and} \qquad E_{mn}=\Big\{\big|\bar{X}_n-\mu\big|<1/m\Big\}.$$

We have $\omega\in E$ if and only if $\lim_{n\to\infty}\bar{X}_n(\omega)=\mu$, which, by the definition of convergence for a sequence of real numbers, happens if and only if for each $m\in\mathcal{N}$, there is an $N\in\mathcal{N}$ such that $\big|\bar{X}_n(\omega)-\mu\big|<1/m$ for all $n\ge N$. However, this last condition means that $\omega\in\bigcap_{m=1}^{\infty}\big(\bigcup_{N=1}^{\infty}\big(\bigcap_{n=N}^{\infty}E_{mn}\big)\big)$. Hence, we have proved that $E=\bigcap_{m=1}^{\infty}\big(\bigcup_{N=1}^{\infty}\big(\bigcap_{n=N}^{\infty}E_{mn}\big)\big)$, as required.

b) Let E and E_{mn} be as in part (a), and set $A_{mn}=\bigcap_{k=n}^{\infty}E_{mk}$. We note that $A_{m1}\subset A_{m2}\subset\cdots$ and, therefore, by the continuity property of a probability measure,

$$P\bigg(\bigcup_{n=1}^{\infty}A_{mn}\bigg)=\lim_{n\to\infty}P(A_{mn}).$$

We want to prove that $P(E) = 1$ if and only if

$$\lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < \epsilon\}\right) = 1, \tag{$*$}$$

for each $\epsilon > 0$. From part (a), we know that $E = \bigcap_{m=1}^{\infty}\left(\bigcup_{n=1}^{\infty} A_{mn}\right)$. It follows that $P(E) = 1$ if and only if $P\left(\bigcup_{n=1}^{\infty} A_{mn}\right) = 1$ for all $m \in \mathcal{N}$ or, equivalently, if and only if $\lim_{n\to\infty} P(A_{mn}) = 1$ for all $m \in \mathcal{N}$; that is,

$$\lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < 1/m\}\right) = 1, \tag{$**$}$$

for all $m \in \mathcal{N}$. Hence, we need to show that Equation $(*)$ holds for each $\epsilon > 0$ if and only if Equation $(**)$ holds for all $m \in \mathcal{N}$.

First suppose that Equation $(*)$ holds for each $\epsilon > 0$. Let $m \in \mathcal{N}$. Taking $\epsilon = 1/m$, we see that Equation $(**)$ holds. Conversely, suppose that Equation $(**)$ holds for all $m \in \mathcal{N}$. Let $\epsilon > 0$. Choose $M \in \mathcal{N}$ such that $1/M \le \epsilon$. Then $\{|\bar{X}_k - \mu| < 1/M\} \subset \{|\bar{X}_k - \mu| < \epsilon\}$ for all $k \in \mathcal{N}$ and, hence,

$$1 = \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < 1/M\}\right) \le \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < \epsilon\}\right) \le 1.$$

Consequently, Equation $(*)$ holds for each $\epsilon > 0$.

c) Clearly, for each $\epsilon > 0$,

$$\{|\bar{X}_n - \mu| < \epsilon\} \supset \bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < \epsilon\}, \qquad n \in \mathcal{N}.$$

Therefore, for each $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\{|\bar{X}_n - \mu| < \epsilon\}\right) \ge \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < \epsilon\}\right).$$

From part (b),

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1 \quad \text{if and only if} \quad \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty}\{|\bar{X}_k - \mu| < \epsilon\}\right) = 1 \text{ for each } \epsilon > 0.$$

Thus, if $P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$, then $\lim_{n\to\infty} P\left(\{|\bar{X}_n - \mu| < \epsilon\}\right) = 1$ for each $\epsilon > 0$. In other words, the strong law of large numbers implies the weak law of large numbers.

d) From the strong law of large numbers, we have $P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$. As we showed in part (b), this equation is equivalent to

$$P\left(\bigcup_{N=1}^{\infty}\left(\bigcap_{n=N}^{\infty}\{|\bar{X}_n - \mu| < \epsilon\}\right)\right) = 1,$$

for each $\epsilon > 0$. For convenience, let A denote the event in the previous display. Then $P(A) = 1$. If $\omega \in A$, then there is an $N \in \mathcal{N}$ such that $|\bar{X}_n(\omega) - \mu| < \epsilon$ for all $n \ge N$. Therefore, $|\bar{X}_n(\omega) - \mu| \ge \epsilon$ occurs for only finitely many n. In other words, with probability 1, deviations of the average value of $X_1, \ldots, X_n$ from μ in excess of any specified amount occur for only finitely many n.

11.53

a) The number x is a normal number if and only if the relative frequency of occurrence of each decimal digit in x is $1/10$.

b) Let $D = \{0, 1, \ldots, 9\}$. For $k_1, \ldots, k_{n-1}, k \in D$, we have

$$\{X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k\} = \left[\frac{k_1}{10} + \cdots + \frac{k_{n-1}}{10^{n-1}} + \frac{k}{10^n}, \frac{k_1}{10} + \cdots + \frac{k_{n-1}}{10^{n-1}} + \frac{k+1}{10^n}\right).$$

Therefore,

$$\begin{aligned} &P\left(X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k\right) \\ &\quad = \left|\left[\frac{k_1}{10} + \cdots + \frac{k_{n-1}}{10^{n-1}} + \frac{k}{10^n}, \frac{k_1}{10} + \cdots + \frac{k_{n-1}}{10^{n-1}} + \frac{k+1}{10^n}\right)\right| = \frac{1}{10^n}. \end{aligned}$$

In the hint, the events in the union are mutually exclusive. Thus, for each $k \in D$,

$$\begin{aligned} P(X_n = k) &= P\left(\bigcup_{k_1=0}^{9} \cdots \bigcup_{k_{n-1}=0}^{9} \{X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k\}\right) \\ &= \sum_{k_1=0}^{9} \cdots \sum_{k_{n-1}=0}^{9} P\left(X_1 = k_1, \ldots, X_{n-1} = k_{n-1}, X_n = k\right) \\ &= \sum_{k_1=0}^{9} \cdots \sum_{k_{n-1}=0}^{9} \frac{1}{10^n} = 10^{n-1} \cdot \frac{1}{10^n} = \frac{1}{10}. \end{aligned}$$

Hence, each X_n has the discrete uniform distribution on D. Moreover, from what we have already established, we see that, for each $n \in \mathcal{N}$ and $k_1, \ldots, k_n \in D$,

$$P\left(X_1 = k_1, \ldots, X_n = k_n\right) = \frac{1}{10^n} = \underbrace{\frac{1}{10} \cdots \frac{1}{10}}_{n \text{ times}} = P(X_1 = k_1) \cdots P(X_n = k_n).$$

Therefore, $X_1, X_2, \ldots$ are independent random variables.

c) Let $k \in D$. For each $j \in \mathcal{N}$, let $Y_j = 1_{\{X_j = k\}}$ and note that $n_k = \sum_{j=1}^{n} Y_j$. Because $X_1, X_2, \ldots$ are independent random variables, each having the discrete uniform distribution on D, the random variables $Y_1, Y_2, \ldots$ are independent, each having the Bernoulli distribution with parameter $1/10$, which is also the mean of such a distribution. Therefore, by the strong law of large numbers,

$$\lim_{n \to \infty} \frac{n_k}{n} = \lim_{n \to \infty} \frac{Y_1 + \cdots + Y_n}{n} = \frac{1}{10},$$

with probability 1. In other words, with probability 1, a randomly selected number from the interval $[0, 1]$ is a normal number.

11.54

a) Each real number between 0 and 1 has a binary expansion and, except for numbers of the form $m/2^n$, the expansion is unique. For definiteness, we take the unique terminating expansion for numbers of the form $m/2^n$. Let x denote a randomly selected number from the interval $[0, 1]$ and define $X_n(x) = x_n$, where x_n is the nth binary digit of x. Proceeding as in Exercise 11.53(b), we can show that the random variables $X_1, X_2, \ldots$ are independent and identically distributed, all having the discrete uniform distribution on the set $\{0, 1\}$. Now, for a single toss of a balanced coin, let us associate the binary digit 1 with the outcome of a head and the binary digit 0 with the outcome of a tail. If we select a number at random from the interval $[0, 1]$, then its binary expansion represents an outcome of repeatedly tossing a balanced coin and observing at each toss (digit) whether the result is a head (1) or a tail (0).

b) Let $x \in [0, 1]$. For each $n \in \mathcal{N}$ and $k \in \{0, 1\}$, use $n_k(x)$ to denote the number of the first n binary digits of x that equal k. Now let x denote a randomly selected number from the interval $[0, 1]$ and let $k \in \{0, 1\}$. For each $j \in \mathcal{N}$, let $Y_j = 1_{\{X_j = k\}}$ and note that $n_k = \sum_{j=1}^{n} Y_j$. Because $X_1, X_2, \ldots$ are independent random variables, each having the discrete uniform distribution on $\{0, 1\}$, the random variables $Y_1, Y_2, \ldots$ are independent, each having the Bernoulli distribution with parameter $1/2$, which is also the mean of such a distribution. Therefore, by the strong law of large numbers,

$$\lim_{n\to\infty} \frac{n_k}{n} = \lim_{n\to\infty} \frac{Y_1 + \cdots + Y_n}{n} = \frac{1}{2},$$

with probability 1. In terms of tossing a balanced coin repeatedly, this result states that, in the long run, half of the tosses will be heads (and half will be tails).

11.4 The Central Limit Theorem

Note: For this section, we have used statistical software to obtain any required standard normal probabilities. If you use Table I instead, your answers may differ somewhat from those shown here.

Basic Exercises

11.55 Let X denote the number of people who don't take the flight. Then $X \sim \mathcal{B}(42, 0.16)$ and, hence,

$$p_X(x) = \binom{42}{x}(0.16)^x(0.84)^{42-x}, \qquad x = 0, 1, \ldots, 42.$$

As $n = 42$ and $p = 0.16$, we have $np = 6.72$ and $np(1-p) = 5.6448$. Hence, in view of Proposition 11.11 on page 661, for integers a and b, with $0 \le a \le b \le 42$,

$$P(a \le X \le b) \approx \Phi\left(\frac{b + \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) - \Phi\left(\frac{a - \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right).$$

a) We have

$$P(X = 5) = p_X(5) = \binom{42}{5}(0.16)^5(0.84)^{37} = 0.1408.$$

Also,

$$P(X = 5) = P(5 \le X \le 5) \approx \Phi\left(\frac{5 + \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) - \Phi\left(\frac{5 - \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) = 0.1288.$$

b) We have

$$P(9 \le X \le 12) = \sum_{x=9}^{12} p_X(x) = \sum_{x=9}^{12} \binom{42}{x}(0.16)^x(0.84)^{42-x} = 0.2087.$$

Also,

$$P(9 \le X \le 12) \approx \Phi\left(\frac{12 + \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) - \Phi\left(\frac{9 - \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) = 0.2194.$$

c) We have

$$P(X \ge 1) = 1 - P(X = 0) = 1 - p_X(0) = 1 - \binom{42}{0}(0.16)^0(0.84)^{42-0} = 0.9993.$$

Also,

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - \left(\Phi\left(\frac{0 + \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) - \Phi\left(\frac{0 - \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) \right) = 0.9968.$$

d) We have

$$P(X \leq 2) = \sum_{x=0}^{2} p_X(x) = \sum_{x=0}^{2} \binom{42}{x} (0.16)^x (0.84)^{42-x} = 0.0266.$$

Also,

$$P(X \leq 2) = P(0 \leq X \leq 2) \approx \Phi\left(\frac{2 + \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) - \Phi\left(\frac{0 - \frac{1}{2} - 6.72}{\sqrt{5.6448}}\right) = 0.0367.$$

11.56

a) See the solution to Exercise 8.108.

b) The following table compares the results of all three methods of obtaining the probabilities. Each number in parentheses gives the absolute percentage error in each case.

	Binomial probability	Local normal approximation	Integral normal approximation
$P(X = 5)$	0.1408	0.1292 (8.2%)	0.1288 (8.5%)
$P(9 \leq X \leq 12)$	0.2087	0.2181 (4.5%)	0.2194 (5.1%)
$P(X \geq 1)$	0.9993	0.9969 (0.2%)	0.9968 (0.3%)
$P(X \leq 2)$	0.0266	0.0357 (34.2%)	0.0367 (38.0%)

c) The customary rule of thumb for using the normal approximation to the binomial distribution is that both np and $n(1-p)$ are 5 or greater. In this case, $np = 6.72$ and $n(1-p) = 35.28$, so the rule of thumb is met. However, it is barely met, which explains some of the large errors, as shown in the table in part (b).

11.57 Let T_n denote the number of times that the gambler wins in n bets. Then the gambler is ahead after n bets if and only if $T_n > n/2$. We know that $T_n \sim \mathcal{B}(n, 9/19)$. As $p = 9/19$, Proposition 11.11 on page 661, shows that, for even n,

$$\begin{aligned}
P(X > n/2) &= 1 - P(X \leq n/2) = 1 - P(0 \leq X \leq n/2) \\
&\approx 1 - \left(\Phi\left(\frac{n/2 + \frac{1}{2} - n \cdot \frac{9}{19}}{\sqrt{n \cdot \frac{9}{19} \cdot \frac{10}{19}}}\right) - \Phi\left(\frac{0 - \frac{1}{2} - n \cdot \frac{9}{19}}{\sqrt{n \cdot \frac{9}{19} \cdot \frac{10}{19}}}\right) \right) \\
&\approx 1 - \left(\Phi\left(\frac{\frac{n+1}{2} - \frac{9}{19}n}{\sqrt{\frac{90}{361}n}}\right) - \Phi\left(\frac{-\frac{1}{2} - \frac{9}{19}n}{\sqrt{\frac{90}{361}n}}\right) \right).
\end{aligned}$$

The following answers are given to three significant digits.

a) Here $n = 100$, so the required probability approximately equals

$$1 - \left(\Phi\left(\frac{\frac{100+1}{2} - \frac{9}{19} \cdot 100}{\sqrt{\frac{90}{361} \cdot 100}} \right) - \Phi\left(\frac{-\frac{1}{2} - \frac{9}{19} \cdot 100}{\sqrt{\frac{90}{361} \cdot 100}} \right) \right)$$
$$= 1 - \Phi(0.627185) + \Phi(-9.58697) = 1 - 0.734731 + 0.000000$$
$$= 0.265.$$

b) Here $n = 1000$, so the required probability approximately equals

$$1 - \left(\Phi\left(\frac{\frac{1000+1}{2} - \frac{9}{19} \cdot 1000}{\sqrt{\frac{90}{361} \cdot 1000}} \right) - \Phi\left(\frac{-\frac{1}{2} - \frac{9}{19} \cdot 1000}{\sqrt{\frac{90}{361} \cdot 1000}} \right) \right)$$
$$= 1 - \Phi(1.69833) + \Phi(-30.0317) = 1 - 0.955278 + 0.000000$$
$$= 0.0447.$$

c) Here $n = 5000$, so the required probability approximately equals

$$1 - \left(\Phi\left(\frac{\frac{5000+1}{2} - \frac{9}{19} \cdot 5000}{\sqrt{\frac{90}{361} \cdot 5000}} \right) - \Phi\left(\frac{-\frac{1}{2} - \frac{9}{19} \cdot 5000}{\sqrt{\frac{90}{361} \cdot 5000}} \right) \right)$$
$$= 1 - \Phi(3.74094) + \Phi(-67.0962) = 1 - 0.9999083 + 0.0000000$$
$$= 0.0000917.$$

11.58 Let $X_1, \ldots, X_{100}$ denote the lengths of the 100 steps taken by the boy. We know that $X_1, \ldots, X_{100}$ are independent and identically distributed random variables with $\mu = 0.95$ and $\sigma = 0.08$. The distance stepped off by the boy is $D = X_1 + \cdots + X_{100}$. Note that

$$n\mu = 100 \cdot 0.95 = 95 \qquad \text{and} \qquad \sigma\sqrt{n} = 0.08 \cdot \sqrt{100} = 0.8.$$

a) Applying Equation (11.42) on page 666, we get

$$P(|D - 100| \le 4) = P(96 \le D \le 104) \approx \Phi\left(\frac{104 - 95}{0.8}\right) - \Phi\left(\frac{96 - 95}{0.8}\right)$$
$$= \Phi(11.25) - \Phi(1.25) = 0.106.$$

b) Again applying Equation (11.42), we get

$$P(|D - 100| \le 6) = P(94 \le D \le 106) \approx \Phi\left(\frac{106 - 95}{0.8}\right) - \Phi\left(\frac{94 - 95}{0.8}\right)$$
$$= \Phi(13.75) - \Phi(-1.25) = 0.894.$$

11.59 Let Y denote the lifetime of a single battery. As $Y \sim \mathcal{N}(30, 5^2)$,

$$P(Y > 25) = 1 - P(Y \le 25) = 1 - \Phi\left(\frac{25 - 30}{5}\right) = 1 - \Phi(-1) = 0.841.$$

Now let X denote the number of the 500 batteries that last longer than 25 hours. Then $X \sim \mathcal{B}(500, 0.841)$. We want to determine $P(X \geq 400)$. Applying the integral De Moivre–Laplace theorem, we have

$$\begin{aligned} P(X \geq 400) &= P(400 \leq X \leq 500) \\ &\approx \Phi\left(\frac{500 + \frac{1}{2} - 500 \cdot 0.841}{\sqrt{500 \cdot 0.841 \cdot (1 - 0.841)}}\right) - \Phi\left(\frac{400 - \frac{1}{2} - 500 \cdot 0.841}{\sqrt{500 \cdot 0.841 \cdot (1 - 0.841)}}\right) \\ &= \Phi(9.78382) - \Phi(-2.56825) = 0.995. \end{aligned}$$

11.60 Let $X_1, \ldots, X_{100}$ denote the checkout times, in minutes, for 100 customers. Then $X_1, \ldots, X_{100}$ are independent and identically distributed random variables with $\mu = 3.5$ and $\sigma = 1.5$.
a) Let T denote the time, in minutes, required to check out 100 customers. Then $T = X_1 + \cdots + X_{100}$. Referring to Equation (11.42) on page 666, we get

$$\begin{aligned} P(T \geq 360) &= 1 - P(0 \leq T < 360) \approx 1 - \left(\Phi\left(\frac{360 - 100 \cdot 3.5}{1.5\sqrt{100}}\right) - \Phi\left(\frac{0 - 100 \cdot 3.5}{1.5\sqrt{100}}\right)\right) \\ &= 1 - \big(\Phi(0.66667) - \Phi(-23.33333)\big) = 0.252. \end{aligned}$$

b) The average checkout time for 100 customers is $T/100$. Again referring to Equation (11.42), we get

$$\begin{aligned} P(T/100 < 3.4) &= P(T < 340) = P(0 \leq T < 340) \\ &\approx \Phi\left(\frac{340 - 100 \cdot 3.5}{1.5\sqrt{100}}\right) - \Phi\left(\frac{0 - 100 \cdot 3.5}{1.5\sqrt{100}}\right) \\ &= \Phi(-0.66667) - \Phi(-23.33333) = 0.252. \end{aligned}$$

11.61
a) Let $\bar{X}_{160}$ denote the average income of 160 randomly chosen households. From the central limit theorem, in the form of the fourth bulleted item on page 667, we know that $\bar{X}_{160}$ is approximately normally distributed with mean \$35,216 and standard deviation $3134/\sqrt{160}$. Thus,

$$\begin{aligned} P\left(\bar{X}_{160} < 34{,}600\right) &= P\left(0 \leq \bar{X}_{160} < 34{,}600\right) \approx \Phi\left(\frac{34{,}600 - 35{,}216}{3134/\sqrt{160}}\right) - \Phi\left(\frac{0 - 35{,}216}{3134/\sqrt{160}}\right) \\ &= \Phi(-2.48623) - \Phi(-142.13500) = 0.00646. \end{aligned}$$

b) No; the independence assumption of the central limit theorem isn't strictly satisfied when sampling is without replacement.
c) It is permissible to use the central limit theorem to solve part (a) because the population size is large relative to the sample size and, hence, there is little difference between sampling without and with replacement. In the latter case, the independence assumption of the central limit theorem is satisfied (as are its other assumptions).

11.62 For convenience, we work in units of thousands. We first observe that the claim amount of each individual policy has an exponential distribution with parameter 1. Let $X_1, \ldots, X_{100}$ denote the claim amounts for the 100 policies. Then $X_1, \ldots, X_{100}$ are independent and identically distributed $\mathcal{E}(1)$ random variables.
a) Let T denote the total claim amount of the 100 policies. Then $T = X_1 + \cdots + X_{100}$. From the first bulleted item on page 537, we know that $T \sim \Gamma(100, 1)$.

b) The expected claim amount per policy is $\mathcal{E}(X_j) = 1/1 = 1$. Because the premium for each policy is set at 0.1 over the expected claim amount, the premium is 1.1 per policy. Hence, the total premium for 100 policies is $100 \cdot 1.1 = 110$. Therefore, in view of part (a), the exact probability that the insurance company will have claims exceeding the premiums collected is

$$P(T > 110) = \int_{110}^{\infty} \frac{1^{100}}{\Gamma(100)} x^{100-1} e^{-1 \cdot x}\, dx = \frac{1}{99!} \int_{110}^{\infty} x^{99} e^{-x}\, dx.$$

In view of Equation (8.49) on page 453, an alternative expression is

$$P(T > 110) = 1 - P(T \le 110) = 1 - F_T(110) = e^{-110} \sum_{j=0}^{99} \frac{(110)^j}{j!}.$$

c) We have $\mu = 1/1 = 1$ and $\sigma^2 = 1/1^2 = 1$. Hence, by the central limit theorem,

$$\begin{aligned} P(T > 110) &= 1 - P(T \le 110) = 1 - P(0 \le T \le 110) \\ &\approx 1 - \left(\Phi\left(\frac{110 - 100 \cdot 1}{1\sqrt{100}} \right) - \Phi\left(\frac{0 - 100 \cdot 1}{1\sqrt{100}} \right) \right) \\ &= 1 - \big(\Phi(1) - \Phi(-10)\big) = 0.159. \end{aligned}$$

d) Using statistical software, we obtained 0.158 for the probability that a $\Gamma(100, 1)$ random variable will exceed 110. We see that this probability and the approximate probability of 0.159 found in part (c) are almost identical. Thus, the normal approximation here is excellent.

11.63 Let T denote the total number of hours that a randomly selected person watches movies or sporting events during a 3-month period. Then $T = X + Y$. We have

$$\mathcal{E}(T) = \mathcal{E}(X + Y) = \mathcal{E}(X) + \mathcal{E}(Y) = 50 + 20 = 70$$

and

$$\operatorname{Var}(T) = \operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y) = 50 + 30 + 2 \cdot 10 = 100.$$

Now let $T_1, \ldots, T_{100}$ denote the number of hours that the 100 randomly selected people watch movies or sporting events during the 3-month period. Then $T_1, \ldots, T_{100}$ are (basically) independent and identically distributed random variables with common mean $\mu = 70$ and common variance $\sigma^2 = 100$. Hence, from the central limit theorem,

$$\begin{aligned} P(T_1 + \cdots + T_{100} \le 7100) &= P(0 \le T_1 + \cdots + T_{100} \le 7100) \\ &\approx \Phi\left(\frac{7100 - 100 \cdot 70}{10\sqrt{100}} \right) - \Phi\left(\frac{0 - 100 \cdot 70}{10\sqrt{100}} \right) \\ &= \Phi(1) - \Phi(-70) = 0.841. \end{aligned}$$

11.64 Let X denote the number of the 10,000 screws that are not within tolerance specifications. Then $X \sim \mathcal{B}(10{,}000, 0.05)$. We want to choose r as small as possible so that $P(X > r) \le 0.01$.
a) We have $\mu_X = 10{,}000 \cdot 0.05 = 500$ and $\sigma_X^2 = 10{,}000 \cdot 0.05 \cdot 0.95 = 475$. In view of Chebyshev's inequality, we get, for $x > 500$, that

$$P(X > x) \le P(|X - 500| \ge x - 500) \le \frac{\sigma_X^2}{(x - 500)^2} = \frac{475}{(x - 500)^2}.$$

Setting this last expression equal to 0.01 and solving for x, we find that $x = 500 + \sqrt{47{,}500} \approx 717.94$. Hence, using Chebyshev's inequality, we choose $r = 718$.
b) Here, because $n = 10{,}000$ is so large, we can use the central limit theorem for binomial random variables in the form of Equation (11.34) on page 661:

$$\begin{aligned} P(X > x) &= 1 - P(X \le x) = 1 - P(0 \le X \le x) \\ &\approx 1 - \left(\Phi\left(\frac{x - 500}{\sqrt{475}}\right) - \Phi\left(\frac{0 - 500}{\sqrt{475}}\right)\right) \\ &\approx 1 - \Phi\left(\frac{x - 500}{\sqrt{475}}\right). \end{aligned}$$

Setting this last expression equal to 0.01 and solving for x gives $x = 500 + \sqrt{475}\,\Phi^{-1}(0.99) \approx 550.70$. Hence, using the central limit theorem, we choose $r = 551$.
c) Using statistical software, we find that $P(X > 550) = 0.0110571$ and $P(X > 551) = 0.0098269$. Thus, the smallest integer r for which $P(X > r) \le 0.01$ is $r = 551$. This result is the same as the one obtained in part (b) by using the central limit theorem. We see that the result obtained in part (a) by using Chebyshev's rule is quite crude (a vast overestimate), which is not surprising.

11.65 Let $X_1, \ldots, X_{250}$ denote the lifetimes, in months, of the 250 previously used compressors. We know that $\sigma = 40$ and we assume as reasonable that $X_1, \ldots, X_{250}$ are independent and identically distributed random variables. From Equation (11.43) on page 668,

$$P\left(\left|\bar{X}_{250} - \mu\right| \le 5\right) \approx 2\Phi\left(\frac{\sqrt{250}\cdot 5}{40}\right) - 1 = 2\Phi(1.97642) - 1 = 0.952.$$

11.66
a) We have

$$Y_n = s\prod_{j=1}^{n} \frac{Y_j}{Y_{j-1}} = s\prod_{j=1}^{n} X_j.$$

Taking natural logarithms, we get $\ln Y_n = \ln s + \sum_{j=1}^{n} \ln X_j$. Assume that X isn't constant and that $\ln X$ has finite variance. Set $\mu = \mathcal{E}(\ln X)$ and $\sigma^2 = \operatorname{Var}(\ln X)$. Applying the central limit theorem, we have, for large n, that $\sum_{j=1}^{n} \ln X_j$ is approximately normal with mean $n\mu$ and variance $n\sigma^2$. Therefore, for large n, the random variable $\ln Y_n$ is approximately normal with mean $\ln s + n\mu$ and variance $n\sigma^2$. Consequently, for large n, the random variable Y_n has approximately a lognormal distribution with parameters $\ln s + n\mathcal{E}(\ln X)$ and $n\operatorname{Var}(\ln X)$.
b) We consider each distribution given for X in turn.

- $X \sim \mathcal{U}(0, 1)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x)\cdot 1\,dx = \int_0^1 \ln x\,dx = -1.$$

Also,

$$\mathcal{E}\left(\left(\ln X\right)^2\right) = \int_0^1 (\ln x)^2 \cdot 1\, dx = \int_0^1 (\ln x)^2\, dx = 2.$$

Hence, $\mathrm{Var}(\ln X) = 2 - (-1)^2 = 1$. Consequently, Y_n has approximately a lognormal distribution with parameters $\ln s - n$ and n.

- $X \sim \mathrm{Beta}(2, 1)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x) \cdot 2x^{2-1}(1-x)^{1-1}\, dx = 2\int_0^1 x \ln x\, dx = 2 \cdot \left(-\frac{1}{4}\right) = -\frac{1}{2}.$$

Also,

$$\mathcal{E}\left(\left(\ln X\right)^2\right) = \int_0^1 (\ln x)^2 \cdot 2x^{2-1}(1-x)^{1-1}\, dx = 2\int_0^1 x(\ln x)^2\, dx = 2 \cdot \left(\frac{1}{4}\right) = \frac{1}{2}.$$

Hence, $\mathrm{Var}(\ln X) = 1/2 - (-1/2)^2 = 1/4$. Consequently, Y_n has approximately a lognormal distribution with parameters $\ln s - n/2$ and $n/4$.

- $X \sim \mathrm{Beta}(1, 2)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x) \cdot 2x^{1-1}(1-x)^{2-1}\, dx = 2\int_0^1 (1-x) \ln x\, dx = 2 \cdot \left(-\frac{3}{4}\right) = -\frac{3}{2}.$$

Also,

$$\mathcal{E}\left(\left(\ln X\right)^2\right) = \int_0^1 (\ln x)^2 \cdot 2x^{1-1}(1-x)^{2-1}\, dx = 2\int_0^1 (1-x)(\ln x)^2\, dx = 2 \cdot \left(\frac{7}{4}\right) = \frac{7}{2}.$$

Hence, $\mathrm{Var}(\ln X) = 7/2 - (-3/2)^2 = 5/4$. Consequently, Y_n has approximately a lognormal distribution with parameters $\ln s - 3n/2$ and $5n/4$.

- $X \sim \mathrm{Beta}(2, 2)$: In this case,

$$\mathcal{E}(\ln X) = \int_0^1 (\ln x) \cdot 6x^{2-1}(1-x)^{2-1}\, dx = 6\int_0^1 x(1-x) \ln x\, dx = 6 \cdot \left(-\frac{5}{36}\right) = -\frac{5}{6}.$$

Also,

$$\mathcal{E}\left(\left(\ln X\right)^2\right) = \int_0^1 (\ln x)^2 \cdot 6x^{2-1}(1-x)^{2-1}\, dx = 6\int_0^1 x(1-x)(\ln x)^2\, dx = 6 \cdot \left(\frac{19}{108}\right) = \frac{19}{18}.$$

Hence, $\mathrm{Var}(\ln X) = 19/18 - (-5/6)^2 = 13/36$. Consequently, Y_n has approximately a lognormal distribution with parameters $\ln s - 5n/6$ and $13n/36$.

11.67

a) We apply Equation (11.47) on page 671. We have $n = 51$, $\sigma = 2.4$, and the sample mean of the data is $\bar{x}_{51} = 15.051$. Also, $\gamma = 0.90$ and, so,

$$\Phi^{-1}\left(\frac{1}{2}(1+\gamma)\right) = \Phi^{-1}\left(\frac{1}{2} \cdot 1.90\right) = \Phi^{-1}(0.95) = 1.645.$$

Consequently, a 90% confidence interval for the mean depth of all subterranean coruro burrows has endpoints $15.051 \pm 1.645 \cdot 2.4/\sqrt{51}$, or $(14.50, 15.60)$.

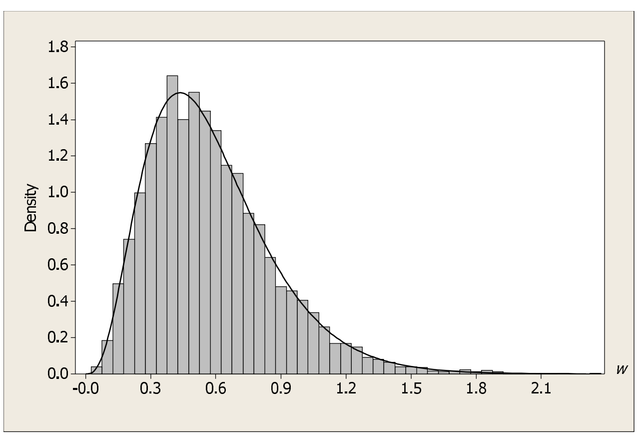
b) We can be 90% confident that the mean depth of all subterranean coruro burrows is somewhere between 14.50 cm and 15.60 cm.

c) Here $\gamma = 0.99$, so

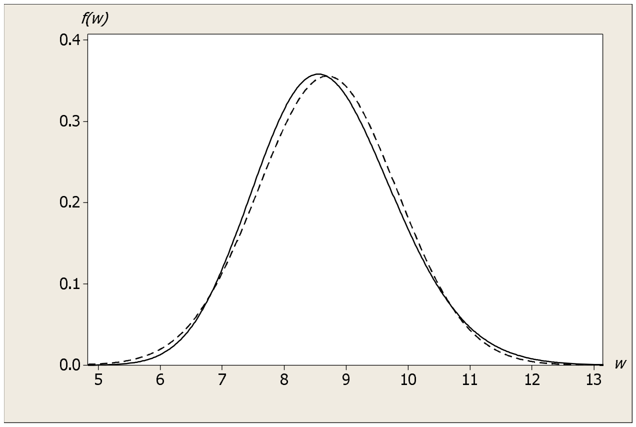
$$\Phi^{-1}\left(\frac{1}{2}(1+\gamma)\right) = \Phi^{-1}\left(\frac{1}{2} \cdot 1.99\right) = \Phi^{-1}(0.995) = 2.576.$$

Consequently, a 99% confidence interval for the mean depth of all subterranean coruro burrows has endpoints $15.051 \pm 2.576 \cdot 2.4/\sqrt{51}$, or (14.19, 15.92).
d) If you want to be more confident that μ lies in your confidence interval, then you must settle for a greater-length interval.

11.68 Let I_1 denote the elapsed time until the arrival of the first patient and, for $j \geq 2$, let I_j denote the elapsed time between the arrivals of the $(j-1)$st and jth patients. Then $I_1, I_2, \ldots$ are independent and identically distributed $\mathcal{E}(6.9)$ random variables. Moreover, we have $W_n = I_1 + \cdots + I_n$ for each $n \in \mathcal{N}$.
a) From the first bulleted item on page 537, we see that $W_n \sim \Gamma(n, 6.9)$ for each $n \in \mathcal{N}$.
b) Answers will vary, but here is the procedure: We know that $W_4 = I_1 + I_2 + I_3 + I_4$, where the I_js are independent $\mathcal{E}(6.9)$ random variables. From Proposition 8.16 or Example 8.22(b) on page 472, we find that, if $U \sim \mathcal{U}(0, 1)$, then $-(\ln U)/6.9 \sim \mathcal{E}(6.9)$. Hence, to use a basic random number generator to get 5000 observations of W_4, proceed as follows: First obtain 5000 four-tuples of uniform numbers between 0 and 1. Next transform each of the 20,000 numbers so obtained by applying the function $-(\ln u)/6.9$. Then take the sum of the four numbers in each of the 5000 transformed four-tuples.
c) No, we would not expect a histogram of the 5000 observations from part (b) to be roughly bell shaped. Rather, as seen by referring to part (a), we would expect it to have roughly the shape of the PDF of a $\Gamma(4, 6.9)$ random variable, which is right skewed.
d) Answers will vary, but here is what we obtained. The superimposed curve is the PDF of a $\Gamma(4, 6.9)$ random variable.



e) For the solution corresponding to part (b), we proceed as in that part, except that we obtain 5000 60-tuples instead of 5000 four-tuples. Here, though, we would expect a histogram of the 5000 observations of W_{60} to be bell shaped because of the central limit theorem. Indeed, we have $W_{60} = I_1 + \cdots + I_{60}$, where $I_1, \ldots, I_{60}$ are independent and identically distributed random variables, all having the $\mathcal{E}(6.9)$ distribution. Hence, by the central limit theorem, W_{60} is approximately normally distributed; its exact distribution is $\Gamma(60, 6.9)$, as we see by referring to part (a). The solid curve in the following figure is of the PDF of a $\Gamma(60, 6.9)$ random variable. The histogram that you obtain of the 5000 observations of W_{60} should be shaped roughly like that PDF. For comparison, we have also included a plot of the PDF of a normal random variable having the same mean and variance as a $\Gamma(60, 6.9)$ random variable. That plot is the dashed curve in the figure.



11.69 Let X and Y denote the times, in seconds, that it takes George and Julia, respectively, to make a double decaf skim milk latte. We are assuming that $\mu_X = \mu_Y$ and that $\sigma_X = \sigma_Y = 4$, and we will also assume that X and Y are independent random variables. Let $D = X - Y$. Then

$$\mathcal{E}(D) = \mathcal{E}(X - Y) = \mathcal{E}(X) - \mathcal{E}(Y) = 0$$

and

$$\operatorname{Var}(D) = \operatorname{Var}(X - Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) = 2 \cdot 4^2 = 32.$$

Let $X_1, \ldots, X_{40}$ and $Y_1, \ldots, Y_{40}$ denote the times, in seconds, that it takes George and Julia, respectively, to make the 40 lattes. We want to compute $P\big(\bar{X}_{40} - \bar{Y}_{40} > 3\big)$. Now,

$$\begin{aligned}\bar{X}_{40} - \bar{Y}_{40} &= \frac{X_1 + \cdots + X_{40}}{40} - \frac{Y_1 + \cdots + Y_{40}}{40} \\ &= \frac{(X_1 - Y_1) + \cdots + (X_{40} - Y_{40})}{40} \\ &= \frac{D_1 + \cdots + D_{40}}{40} = \bar{D}_{40},\end{aligned}$$

where $D_j = X_j - Y_j$. The random variables $D_1, \ldots, D_{40}$ are independent and identically distributed with mean 0 and variance 32. Hence, by the central limit theorem, in the form of the fourth bulleted item on page 667,

$$\begin{aligned}P\big(\bar{X}_{40} - \bar{Y}_{40} > 3\big) &= P\big(\bar{D}_{40} > 3\big) = 1 - P\big(\bar{D}_{40} \le 3\big) \\ &\approx 1 - \Phi\left(\frac{3 - 0}{\sqrt{32}/\sqrt{40}}\right) = 1 - \Phi(3.3541) \\ &= 0.000398.\end{aligned}$$

11.70

a) Let X_j $(1 \le j \le 20)$ denote the lifetime, in hours, of bulb j. We know that $X_1, \ldots, X_{20}$ are independent random variables, each having the $\mathcal{E}(1/300)$ distribution. Let T denote the total time that the 20 bulbs last, so that $T = X_1 + \cdots + X_{20}$. Recalling that the mean and variance of an $\mathcal{E}(1/300)$ random variable are 300 and 300^2, respectively, we can apply the central limit theorem to conclude that

$$\begin{aligned} P(T \le 7000) &= P(0 \le T \le 7000) \\ &\approx \Phi\left(\frac{7000 - 20 \cdot 300}{300\sqrt{20}}\right) - \Phi\left(\frac{0 - 20 \cdot 300}{300\sqrt{20}}\right) \\ &= \Phi(0.745356) - \Phi(-4.472136) = 0.772. \end{aligned}$$

b) From the first bulleted item on page 537, we see that $T \sim \Gamma(20, 1/300)$. Using statistical software, we get that $P(T \le 7000) = 0.783$. The discrepancy between the exact value of 0.783 and the value of 0.772 obtained by using the normal approximation is due to the fact that an exponential distribution is highly skewed and that here $n = 20$, which is relatively small. Nonetheless, that absolute percentage error resulting from using the normal approximation is only 1.4%.

Theory Exercises

11.71

a) The quantity $\Phi^{-1}\left(\frac{1}{2}(1+\gamma)\right) \cdot \sigma/\sqrt{n}$ appearing in Expression (11.47) is called the margin of error (or maximum error of the estimate) for the estimate of μ because we are $100\gamma\%$ confident that our error in estimating μ by $\bar{x}_n$ is at most $\Phi^{-1}\left(\frac{1}{2}(1+\gamma)\right) \cdot \sigma/\sqrt{n}$.

b) We want to find the smallest integer, n, such that

$$\Phi^{-1}\left(\frac{1}{2}(1+\gamma)\right) \cdot \frac{\sigma}{\sqrt{n}} = E.$$

Solving for n in the preceding equation, we obtain

$$n = \left\lceil \left(\frac{\Phi^{-1}\left(\frac{1}{2}(1+\gamma)\right) \cdot \sigma}{E}\right)^2 \right\rceil.$$

Advanced Exercises

11.72

a) Referring to the solution to Exercise 11.67(a), we see that, for a 90% confidence interval, the margin of error is

$$\Phi^{-1}\left(\frac{1}{2}(1+0.90)\right) \cdot \frac{2.4}{\sqrt{51}} = \Phi^{-1}(0.95) \cdot \frac{2.4}{\sqrt{51}} = 1.645 \cdot \frac{2.4}{\sqrt{51}} = 0.55$$

and, referring to the solution to Exercise 11.67(c), we see that, for a 99% confidence interval, the margin of error is

$$\Phi^{-1}\left(\frac{1}{2}(1+0.99)\right) \cdot \frac{2.4}{\sqrt{51}} = \Phi^{-1}(0.995) \cdot \frac{2.4}{\sqrt{51}} = 2.576 \cdot \frac{2.4}{\sqrt{51}} = 0.87.$$

Note that these margins of error can also be obtained by taking half the lengths of the confidence intervals found in parts (a) and (c), respectively, of Exercise 11.67.

b) Applying the result of Exercise 11.71(b), we find that, for a confidence coefficient of 0.90, the required sample size for a 0.5 cm margin of error is

$$n = \left\lceil \left(\frac{\Phi^{-1}\left(\frac{1}{2}(1+0.90)\right) \cdot 2.4}{0.5} \right)^2 \right\rceil = \left\lceil \left(\frac{1.645 \cdot 2.4}{0.5} \right)^2 \right\rceil = 63$$

and that, for a confidence coefficient of 0.99, the required sample size for a 0.5 cm margin of error is

$$n = \left\lceil \left(\frac{\Phi^{-1}\left(\frac{1}{2}(1+0.99)\right) \cdot 2.4}{0.5} \right)^2 \right\rceil = \left\lceil \left(\frac{2.576 \cdot 2.4}{0.5} \right)^2 \right\rceil = 153.$$

11.73 Let $X_1, \dots, X_{44}$ denote the purchase amounts, in dollars, for 44 sales with the new strategy. We assume that these random variables are independent and identically distributed with standard deviation \$11 (as with the old strategy). If the mean purchase amount has not changed, then, by the central limit theorem,

$$P(\bar{X}_{44} \geq 73) \approx 1 - \Phi\left(\frac{73 - 68}{11/\sqrt{44}} \right) = 1 - \Phi(3.01511) = 0.00128.$$

Thus, if the mean purchase amount has not changed, then the chances are only about 0.1% that 44 sales would average \$73 or greater. Consequently, either an extremely unlikely event has occurred or there has been an increase in the mean purchase amount. The most reasonable conclusion is that the latter has occurred and, hence, that the new strategy is effective.

11.74 Let $X_1, X_2, \dots$ be independent and identically distributed random variables, each having the Poisson distribution with parameter $\lambda = 1$. In particular, then, the common mean and variance of these random variables is $\mu = 1$ and $\sigma^2 = 1$, respectively. From Proposition 6.20(b) on page 311, we know that $X_1 + \cdots + X_n \sim \mathcal{P}(n)$ for each $n \in \mathcal{N}$. Hence,

$$P(X_1 + \cdots + X_n \leq n) = \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!} = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

Therefore,

$$\lim_{n\to\infty} P(X_1 + \cdots + X_n \leq n) = \lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

On the other hand, by the central limit theorem,

$$\lim_{n\to\infty} P(X_1 + \cdots + X_n \leq n) = \lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n \cdot 1}{1 \cdot \sqrt{n}} \leq 0 \right) = \Phi(0) = \frac{1}{2}.$$

The required result now follows.

11.75

a) From Proposition 11.2 on page 635, the assumption that the common MGF of $X_1, X_2, \dots$ is defined in some open interval containing 0 implies that these random variables have moments of all orders.

b) From Exercise 11.18, if a random variable X has finite nth moment, then $\mathcal{E}(X^n) = (-i)^n \phi_X^{(n)}(0)$. And, from Exercise 11.19(b), if $X \sim \mathcal{N}(0, 1)$, then $\phi_X(t) = e^{-t^2/2}$. To prove the central limit theorem by using characteristic functions, we first show that, if X is a random variable with finite variance, then

$$\lim_{t\to 0} \frac{\ln \phi_X(t) - i\mu_X t}{t^2} = -\frac{\sigma_X^2}{2}. \qquad (*)$$

We recall that $\phi_X(0) = 1$, and we have

$$\phi'_X(0) = i\mathcal{E}(X) = i\mu_X \qquad \text{and} \qquad \phi''_X(0) = -\mathcal{E}\left(X^2\right) = -\mu_X^2 - \sigma_X^2.$$

Applying L'Hôpital's rule twice, we get

$$\begin{aligned}
\lim_{t\to 0} \frac{\ln\phi_X(t) - i\mu_X t}{t^2} &= \lim_{t\to 0} \frac{\phi'_X(t)/\phi_X(t) - i\mu_X}{2t} = \lim_{t\to 0} \frac{\phi'_X(t) - i\mu_X\phi_X(t)}{2t\phi_X(t)} \\
&= \lim_{t\to 0} \frac{\phi'_X(t) - i\mu_X\phi_X(t)}{2t} = \lim_{t\to 0} \frac{\phi''_X(t) - i\mu_X\phi'_X(t)}{2} \\
&= \frac{\phi''_X(0) - i\mu_X\phi'_X(0)}{2} = \frac{-\mu_X^2 - \sigma_X^2 - i\mu_X \cdot i\mu_X}{2} = -\frac{\sigma_X^2}{2}.
\end{aligned}$$

Now, let $X_1, X_2, \ldots$ be independent and identically distributed random variables with common finite mean μ and common finite nonzero variance σ^2. For convenience, set $S_n = X_1 + \cdots + X_n$ and let Y_n denote the standardized version of $X_1 + \cdots + X_n$. Then

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = -n\frac{\mu}{\sigma\sqrt{n}} + \frac{1}{\sigma\sqrt{n}}S_n.$$

Applying Exercise 11.18, we get

$$\phi_{Y_n}(t) = e^{-in\frac{\mu}{\sigma\sqrt{n}}t}\phi_{S_n}\left(t/\sigma\sqrt{n}\right) = e^{-in\frac{\mu}{\sigma\sqrt{n}}t}\left(\phi\left(t/\sigma\sqrt{n}\right)\right)^n = \left(e^{-\frac{i\mu}{\sigma\sqrt{n}}t}\phi\left(t/\sigma\sqrt{n}\right)\right)^n,$$

where ϕ is the common CHF of $X_1, X_2, \ldots$. Taking natural logarithms yields

$$\ln\phi_{Y_n}(t) = n\left(\ln\phi\left(t/\sigma\sqrt{n}\right) - \frac{i\mu}{\sigma\sqrt{n}}t\right) = \frac{t^2}{\sigma^2} \cdot \frac{\ln\phi\left(t/\sigma\sqrt{n}\right) - i\mu \cdot \left(t/\sigma\sqrt{n}\right)}{\left(t/\sigma\sqrt{n}\right)^2}.$$

It follows from Equation $(*)$ that

$$\lim_{n\to\infty} \ln\phi_{Y_n}(t) = \frac{t^2}{\sigma^2} \cdot \left(-\frac{\sigma^2}{2}\right) = -\frac{t^2}{2},$$

or, equivalently,

$$\lim_{n\to\infty} \phi_{Y_n}(t) = e^{-t^2/2}.$$

As the function on the right of this last display is the CHF of a standard normal random variable, the central limit theorem now follows from the continuity theorem for characteristic functions.

c) A proof of the central limit theorem by using characteristic functions is preferable to that by using moment generating functions for the following reason: By using characteristic functions, we need only assume moments up to order 2 (i.e., finite variance), whereas, by using moment generating functions, we must assume that the common MGF is defined in some open interval containing 0, which, in particular, assumes moments of all orders, as we noted in part (a).

Review Exercises for Chapter 11

Basic Exercises

11.76 Let $X \sim \Gamma(\alpha, \lambda)$. We have

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha = \lambda^\alpha (\lambda - t)^{-\alpha}.$$

Consequently,

$$M'_X(t) = \alpha\lambda^\alpha(\lambda - t)^{-\alpha-1} \qquad \text{and} \qquad M''_X(t) = (\alpha+1)\alpha\lambda^\alpha(\lambda - t)^{-\alpha-2}.$$

Therefore, from Proposition 11.2 on page 635,

$$\mathcal{E}(X) = M'_X(0) = \alpha\lambda^\alpha(\lambda - 0)^{-\alpha-1} = \frac{\alpha}{\lambda}$$

and

$$\mathcal{E}(X^2) = M''_X(0) = (\alpha+1)\alpha\lambda^\alpha(\lambda - 0)^{-\alpha-2} = \frac{(\alpha+1)\alpha}{\lambda^2}.$$

Also,

$$\text{Var}(X) = \mathcal{E}(X^2) - \big(\mathcal{E}(X)\big)^2 = \frac{(\alpha+1)\alpha}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}.$$

11.77 For convenience, let $c = 2500$.
a) We have $M_X(t) = (1 - ct)^{-4}$. Then

$$M'_X(t) = 4c(1 - ct)^{-5} \qquad \text{and} \qquad M''_X(t) = 20c^2(1 - ct)^{-6}.$$

Referring to Proposition 11.2 on page 635 and setting $t = 0$, we get

$$\mathcal{E}(X) = M'_X(0) = 4c(1 - 0)^{-5} = 4c$$

and

$$\mathcal{E}(X^2) = M''_X(0) = 20c^2(1 - 0)^{-6} = 20c^2.$$

Hence,

$$\text{Var}(X) = \mathcal{E}(X^2) - \big(\mathcal{E}(X)\big)^2 = 20c^2 - (4c)^2 = 4c^2,$$

and, therefore, $\sigma_X = \sqrt{4c^2} = 2c = 5000$.
b) We have

$$M_X(t) = \frac{1}{(1 - ct)^4} = \left(\frac{1/c}{1/c - t}\right)^4.$$

From Table 11.1 on page 634, we see that M_X is the MGF of a $\Gamma(4, 1/c)$ random variable. Hence, by the uniqueness property of MGFs, X has that distribution. Referring now to the result obtained in Exercise 11.76, we get

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{\frac{4}{(1/c)^2}} = 2 \cdot c = 5000.$$

11.78

a) Let $T \sim \mathcal{T}(-1, 1)$. From Equation (8.57) on page 460, we have $f_T(x) = 1 - |x|$ if $-1 < x < 1$, and $f_T(x) = 0$ otherwise. Therefore,

$$\begin{aligned} M_T(t) &= \int_{-\infty}^{\infty} e^{tx} f_T(x)\,dx = \int_{-1}^{1} (1 - |x|)e^{tx}\,dx = \int_{-1}^{0} (1 + x)e^{tx}\,dx + \int_{0}^{1} (1 - x)e^{tx}\,dx \\ &= \int_{0}^{1} (1 - x)e^{-tx}\,dx + \int_{0}^{1} (1 - x)e^{tx}\,dx. \end{aligned}$$

We know that $M_T(0) = 1$. From calculus, for $t \neq 0$,

$$\int_0^1 (1 - x)e^{tx}\,dx = \int_0^1 e^{tx}\,dx - \int_0^1 xe^{tx}\,dx = \frac{e^t - 1}{t} - \frac{te^t - e^t + 1}{t^2} = \frac{e^t - 1 - t}{t^2}.$$

Replacing t by $-t$, we get

$$\int_0^1 (1 - x)e^{-tx}\,dx = \frac{e^{-t} - 1 + t}{t^2}.$$

Consequently, for $t \neq 0$,

$$M_T(t) = \frac{e^{-t} - 1 + t}{t^2} + \frac{e^t - 1 - t}{t^2} = \frac{\left(e^t + e^{-t}\right) - 2}{t^2} = \frac{2(\cosh t - 1)}{t^2}.$$

b) Let $X \sim \mathcal{T}(a, b)$. Then we can write

$$X = \frac{a + b}{2} + \frac{b - a}{2} T,$$

where $T \sim \mathcal{T}(-1, 1)$. For convenience, set $c = (a + b)/2$ and $d = (b - a)/2$. From Proposition 11.1 on page 633 and part (a), we have, for $t \neq 0$,

$$\begin{aligned} M_X(t) &= M_{c+dT}(t) = e^{ct} M_T(dt) = e^{ct} \cdot \frac{2\big(\cosh(dt) - 1\big)}{(dt)^2} \\ &= \frac{2e^{(a+b)t/2}}{\Big(\big((b - a)/2\big)t\Big)^2} \left(\cosh\Big(\big((b - a)/2\big)t\Big) - 1\right) \\ &= \frac{8e^{(a+b)t/2}}{(b - a)^2 t^2} \left(\cosh \frac{(b - a)t}{2} - 1\right). \end{aligned}$$

c) Let X and Y be independent $\mathcal{U}(0, 1)$ random variables. From Table 11.1 on page 634, we know that $M_X(t) = M_Y(t) = \left(e^t - 1\right)/t$ for $t \neq 0$. Therefore, from Proposition 11.4 on page 636,

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) = \left(\frac{e^t - 1}{t}\right) \cdot \left(\frac{e^t - 1}{t}\right) = \frac{e^{2t} - 2e^t + 1}{t^2} \\ &= \frac{e^t}{t^2}\left(e^t + e^{-t} - 2\right) = \frac{2e^t}{t^2}(\cosh t - 1). \end{aligned}$$

Letting $a = 0$ and $b = 2$ in the result of part (b) gives the same function as the one at the end of the preceding display. Hence, $X + Y$ has the same MGF as a $\mathcal{T}(0, 2)$ random variable and, so, by the uniqueness property of MGFs, we conclude that $X + Y \sim \mathcal{T}(0, 2)$.

11.79

a) We want to use the MGF of T to obtain its mean and variance. To do so, we need to determine $M_T'(0)$ and $M_T''(0)$. We can accomplish this task in several ways. One way is to use the Taylor series expansion of the hyperbolic cosine, namely, $\cosh t = \sum_{n=0}^{\infty} t^{2n}/(2n)!$. From this expansion and Exercise 11.78(a), we get

$$M_T(t) = \frac{2(\cosh t - 1)}{t^2} = \frac{2}{t^2}\sum_{n=1}^{\infty}\frac{t^{2n}}{(2n)!} = 1 + \frac{t^2}{12} + \sum_{n=3}^{\infty}\frac{2t^{2n-2}}{(2n)!}.$$

Therefore,

$$M_T'(t) = \frac{t}{6} + \sum_{n=3}^{\infty}\frac{4(n-1)t^{2n-3}}{(2n)!} \qquad \text{and} \qquad M_T''(t) = \frac{1}{6} + \sum_{n=3}^{\infty}\frac{4(n-1)(2n-3)t^{2n-4}}{(2n)!}.$$

Consequently,

$$\mathcal{E}(T) = M_T'(0) = 0$$

and

$$\operatorname{Var}(T) = \mathcal{E}\big(T^2\big) - \big(\mathcal{E}(T)\big)^2 = \mathcal{E}\big(T^2\big) = M_T''(0) = \frac{1}{6}.$$

b) As we noted in Exercise 11.78(b), we can write

$$X = \frac{a+b}{2} + \frac{b-a}{2}T,$$

where $T \sim \mathcal{T}(-1, 1)$. From part (a), then,

$$\mathcal{E}(X) = \mathcal{E}\left(\frac{a+b}{2} + \frac{b-a}{2}T\right) = \frac{a+b}{2} + \frac{b-a}{2}\mathcal{E}(T) = \frac{a+b}{2} + \frac{b-a}{2}\cdot 0 = \frac{a+b}{2}$$

and

$$\operatorname{Var}(X) = \operatorname{Var}\left(\frac{a+b}{2} + \frac{b-a}{2}T\right) = \left(\frac{b-a}{2}\right)^2 \operatorname{Var}(T) = \left(\frac{b-a}{2}\right)^2 \cdot \frac{1}{6} = \frac{(b-a)^2}{24}.$$

11.80

a) Using Equation (5.44), making the substitution $j = x - r$, and applying Equation (5.45), we get

$$\begin{aligned} M_X(t) &= \sum_x e^{tx} p_X(x) = \sum_{x=r}^{\infty} e^{tx}\binom{-r}{x-r} p^r (p-1)^{x-r} = \sum_{j=0}^{\infty} e^{t(j+r)}\binom{-r}{j} p^r (p-1)^j \\ &= \left(pe^t\right)^r \sum_{j=0}^{\infty}\binom{-r}{j}\left((p-1)e^t\right)^j = \left(pe^t\right)^r \left(1 + (p-1)e^t\right)^{-r} = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r. \end{aligned}$$

For this result to make sense, we need to have $(1-p)e^t < 1$ or, equivalently, $t < -\ln(1-p)$.

b) For convenience, set

$$q = 1 - p \qquad \text{and} \qquad g(t) = \frac{pe^t}{1-qe^t}.$$

We have

$$\begin{aligned} M_X'(t) &= r\left(g(t)\right)^{r-1} g'(t) = r\left(g(t)\right)^{r-1} \cdot \frac{\left(1-qe^t\right)pe^t + pe^t qe^t}{(1-qe^t)^2} \\ &= \frac{r}{1-qe^t}\left(g(t)\right)^r = \frac{r}{1-qe^t} M_X(t) \end{aligned}$$

and

$$M_X''(t) = \frac{r}{1-qe^t} M_X'(t) + \frac{rqe^t}{(1-qe^t)^2} M_X(t).$$

Consequently, by the moment generation property of MGFs,

$$\mathcal{E}(X) = M_X'(0) = \frac{r}{1-q} M_X(0) = \frac{r}{p} \cdot 1 = \frac{r}{p}$$

and

$$\mathcal{E}\big(X^2\big) = M_X''(0) = \frac{r}{1-q} M_X'(0) + \frac{rq}{(1-q)^2} M_X(0) = \frac{r}{p} \cdot \frac{r}{p} + \frac{r(1-p)}{p^2} \cdot 1 = \frac{r^2}{p^2} + \frac{r(1-p)}{p^2}.$$

Also,

$$\operatorname{Var}(X) = \mathcal{E}\big(X^2\big) = \big(\mathcal{E}(X)\big)^2 = \frac{r^2}{p^2} + \frac{r(1-p)}{p^2} - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}.$$

11.81

a) From Proposition 11.4 on page 636 and Exercise 11.80(a),

$$M_{X_1+\cdots+X_m}(t) = M_{X_1}(t) \cdots M_{X_m}(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^{r_1} \cdots \left(\frac{pe^t}{1-(1-p)e^t}\right)^{r_m}$$
$$= \left(\frac{pe^t}{1-(1-p)e^t}\right)^{r_1+\cdots+r_m}.$$

We see that the MGF of $X_1 + \cdots + X_m$ is that of a random variable having the negative binomial distribution with parameters $r_1 + \cdots + r_m$ and p. Hence, by the uniqueness property of MGFs, we conclude that $X_1 + \cdots + X_m \sim \mathcal{NB}(r_1 + \cdots + r_m, p)$.

b) As we know, a geometric random variable with parameter p is also a negative binomial random variable with parameters 1 and p. Hence, $X_1, \ldots, X_m$ are also independent random variables with $X_j \sim \mathcal{NB}(1, p)$ for $1 \le j \le m$. Consequently, from part (a), we have $X_1 + \cdots + X_m \sim \mathcal{NB}(m, p)$.

11.82 Let $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and $Y \sim \mathcal{N}\left(\nu, \tau^2\right)$. As X and Y are independent random variables, Proposition 6.9 on page 291 implies that aX and bY are also independent random variables. Thus, from Propositions 11.4 and 11.1 on pages 636 and 633, respectively, and Table 11.1 on page 634,

$$M_{aX+bY}(t) = M_{aX}(t) M_{bY}(t) = M_X(at) M_Y(bt)$$
$$= e^{\mu(at)+\sigma^2(at)^2/2} e^{\nu(bt)+\tau^2(bt)^2/2} = e^{(a\mu+b\nu)t+\left(a^2\sigma^2+b^2\tau^2\right)t^2/2}.$$

Referring again to Table 11.1 and applying the uniqueness property of MGFs (Proposition 11.3 on page 636), we conclude that $aX + bY \sim \mathcal{N}\left(a\mu + b\nu, a^2\sigma^2 + b^2\tau^2\right)$.

11.83

a) As both $-1/n$ and $1/n$ converge to 0 as $n \to \infty$, heuristically, $\{X_n\}_{n=1}^{\infty}$ converges in distribution to the constant random variable $X = 0$; that is, $p_X(0) = 1$ and $p_X(x) = 0$ if $x \neq 0$.

b) We know that $M_{X_n}(0) = 1$ for all $n \in \mathcal{N}$. From Table 11.1 on page 634, we see that, for $t \neq 0$,

$$M_{X_n}(t) = \frac{e^{t/n} - e^{-t/n}}{2t/n}.$$

Applying L'Hôpital's rule gives

$$\lim_{n\to\infty} M_{X_n}(t) = \lim_{n\to\infty} \frac{e^{t/n} - e^{-t/n}}{2t/n} = \lim_{n\to\infty} \frac{-\left(t/n^2\right)\left(e^{t/n} + e^{-t/n}\right)}{-2t/n^2} = \frac{2}{2} = 1.$$

Now, let X be the constant random variable equal to 0. Then, $M_X(t) = \mathcal{E}\big(e^{t\cdot 0}\big) = 1$ for all $t \in \mathcal{R}$. Therefore, we have shown that $M_{X_n}(t) \to M_X(t)$ as $n \to \infty$ for all $t \in \mathcal{R}$.

c) We have

$$F_{X_n}(x) = \begin{cases} 0, & \text{if } x < -1/n; \\ (nx+1)/2, & \text{if } -1/n \le x < 1/n; \\ 1, & \text{if } x \ge 1/n. \end{cases} \qquad \text{and} \qquad F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1, & \text{if } x \ge 0. \end{cases}$$

We note that F_X is continuous except at $x = 0$. Now, if $x < 0$, then, for sufficiently large n, we have $x < -1/n$ and, hence, $F_{X_n}(x) = 0$. Therefore,

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} 0 = 0 = F_X(x).$$

If $x > 0$, then, for sufficiently large n, we have $x \ge 1/n$ and, hence, $F_{X_n}(x) = 1$. Therefore,

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} 1 = 1 = F_X(x).$$

We have shown that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ for all $x \in \mathcal{R}$ at which F_X is continuous.

11.84

a) We have $M_{X_n}(t) = \mathcal{E}\big(e^{t\cdot 1/n}\big) = e^{t/n}$ and $M_X(t) = \mathcal{E}\big(e^{t\cdot 0}\big) = 1$. Consequently, for each $t \in \mathcal{R}$,

$$\lim_{n\to\infty} M_{X_n}(t) = \lim_{n\to\infty} e^{t/n} = 1 = M_X(t).$$

Therefore, by the continuity theorem, Proposition 11.5 on page 638, we have that $\{X_n\}_{n=1}^{\infty}$ converges in distribution to X. Now, we have $P(X_n = 0) = 0$ for all $n \in \mathcal{N}$ and $P(X = 0) = 1$. Therefore,

$$\lim_{n\to\infty} P(X_n = 0) = 0 \neq 1 = P(X = 0).$$

b) We have

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1, & \text{if } x \ge 0. \end{cases}$$

To resolve your classmate's confusion, just note that 0 is not a point of continuity of F_X.

11.85

a) We have

$$\begin{aligned} M_{X,Y}(s,t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{sx+ty} f_{X,Y}(x,y)\,dx\,dy = \int_0^{\infty}\left(\int_x^{\infty} e^{sx+ty}\cdot 6e^{-(x+2y)}\,dy\right)dx \\ &= 6\int_0^{\infty} e^{-(1-s)x}\left(\int_x^{\infty} e^{-(2-t)y}\,dy\right)dx = \frac{6}{2-t}\int_0^{\infty} e^{-(1-s)x}e^{-(2-t)x}\,dx \\ &= \frac{6}{2-t}\int_0^{\infty} e^{-(3-s-t)x}\,dx = \frac{6}{(2-t)(3-s-t)} \end{aligned}$$

if $t < 2$ and $s + t < 3$.

b) From Proposition 11.7 on page 645 and part (a),

$$M_X(s) = M_{X,Y}(s,0) = \frac{6}{(2-0)(3-s-0)} = \frac{3}{3-s}.$$

Referring now to Table 11.1 on page 634, we see that $X \sim \mathcal{E}(3)$.

c) The time required to complete the second task is $Y - X$. From part (a),

$$M_{Y-X}(t) = \mathcal{E}\left(e^{(Y-X)t}\right) = \mathcal{E}\left(e^{(-t)X+tY}\right) = M_{X,Y}(-t, t) = \frac{6}{(2-t)\big(3-(-t)-t\big)} = \frac{2}{2-t}.$$

Referring now to Table 11.1, we see that $Y - X \sim \mathcal{E}(2)$.
d) The times required to complete the two tasks are X and $Y - X$. From parts (a)–(c),

$$\begin{aligned} M_{X,Y-X}(s,t) &= \mathcal{E}\left(e^{sX+t(Y-X)}\right) = \mathcal{E}\left(e^{(s-t)X+tY}\right) = \frac{6}{(2-t)\big(3-(s-t)-t\big)} \\ &= \frac{6}{(2-t)(3-s)} = \frac{3}{3-s}\cdot\frac{2}{2-t} = M_X(s)M_{Y-X}(t). \end{aligned}$$

Hence, from Proposition 11.9 on page 646, X and $Y - X$ are independent random variables.
e) The time required to complete the two successive tasks is Y. From Proposition 11.7 and part (a),

$$M_Y(t) = M_{X,Y}(0,t) = \frac{6}{(2-t)(3-0-t)} = \frac{6}{6-5t+t^2}.$$

Hence,

$$M_Y'(t) = -\frac{6}{\left(6-5t+t^2\right)^2}\cdot(-5+2t) = \frac{6(5-2t)}{\left(6-5t+t^2\right)^2}.$$

Consequently, $\mathcal{E}(Y) = M_Y'(0) = 30/36 = 5/6$.

11.86
a) We have

$$\begin{aligned} M_{X,Y}(s,t) &= \sum_{(x,y)}\sum e^{sx+ty}p_{X,Y}(x,y) \\ &= \frac{4}{9}\left(e^{s\cdot 0+t\cdot 0}2^{-(0+0)} + e^{s\cdot 1+t\cdot 0}2^{-(1+0)} + e^{s\cdot 0+t\cdot 1}2^{-(0+1)} + e^{s\cdot 1+t\cdot 1}2^{-(1+1)}\right) \\ &= \frac{4}{9}\left(1 + \frac{1}{2}e^s + \frac{1}{2}e^t + \frac{1}{4}e^{s+t}\right) = \frac{1}{9}\left(4 + 2e^s + 2e^t + e^{s+t}\right). \end{aligned}$$

b) We have

$$\frac{\partial M_{X,Y}}{\partial s}(s,t) = \frac{1}{9}\left(2e^s + e^{s+t}\right), \quad \frac{\partial^2 M_{X,Y}}{\partial s^2}(s,t) = \frac{1}{9}\left(2e^s + e^{s+t}\right), \quad \frac{\partial^2 M_{X,Y}}{\partial t\,\partial s}(s,t) = \frac{1}{9}e^{s+t}.$$

Therefore, from Proposition 11.6 on page 643,

$$\mathcal{E}(X) = \frac{\partial M_{X,Y}}{\partial s}(0,0) = \frac{1}{3}, \quad \mathcal{E}\left(X^2\right) = \frac{\partial^2 M_{X,Y}}{\partial s^2}(0,0) = \frac{1}{3}, \quad \mathcal{E}(XY) = \frac{\partial^2 M_{X,Y}}{\partial t\,\partial s}(0,0) = \frac{1}{9}.$$

Thus, $\mathcal{E}(X) = 1/3$ and $\mathrm{Var}(X) = 1/3 - (1/3)^2 = 2/9$. By symmetry, Y has the same mean and variance as X. Also,

$$\mathrm{Cov}(X,Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \frac{1}{9} - \frac{1}{3}\cdot\frac{1}{3} = 0.$$

c) From Proposition 11.7 on page 645 and part (a),

$$M_X(s) = M_{X,Y}(s,0) = \frac{1}{9}\left(4 + 2e^s + 2e^0 + e^{s+0}\right) = \frac{1}{3}e^s + \frac{2}{3}.$$

Referring now to Table 11.1 on page 634 and applying the uniqueness property of MGFs, we see that $X \sim \mathcal{B}(1, 1/3)$ or, equivalently, X has the Bernoulli distribution with parameter $1/3$. By symmetry, Y has the same probability distribution as X.

d) From Proposition 11.10 on page 647 and part (a),

$$\begin{aligned} M_{X+Y}(t) = M_{X,Y}(t,t) &= \frac{1}{9}\left(4 + 2e^t + 2e^t + e^{t+t}\right) = \frac{1}{9}\left(4 + 4e^t + e^{2t}\right) \\ &= \frac{1}{9}\left(2 + e^t\right)^2 = \left(\frac{1}{3}e^t + \frac{2}{3}\right)^2. \end{aligned}$$

Referring now to Table 11.1 on page 634 and applying the uniqueness property of MGFs, we see that $X + Y \sim \mathcal{B}(2, 1/3)$.

e) From parts (a) and (c),

$$\begin{aligned} M_{X,Y}(s,t) &= \frac{1}{9}\left(4 + 2e^s + 2e^t + e^{s+t}\right) = \frac{1}{9}\left(2 + e^s\right)\left(2 + e^t\right) \\ &= \left(\frac{1}{3}e^s + \frac{2}{3}\right)\left(\frac{1}{3}e^t + \frac{2}{3}\right) = M_X(s)M_Y(t). \end{aligned}$$

Hence, from Proposition 11.9 on page 646, X and Y are independent random variables.

11.87

a) We have

$$\begin{aligned} M_{X,Y}(s,t) &= \sum\sum_{(x,y)} e^{sx+ty} p_{X,Y}(x,y) \\ &= \frac{4}{7}\left(e^{s\cdot 0+t\cdot 0}2^{-(0+0)} + e^{s\cdot 1+t\cdot 0}2^{-(1+0)} + e^{s\cdot 1+t\cdot 1}2^{-(1+1)}\right) \\ &= \frac{4}{7}\left(1 + \frac{1}{2}e^s + \frac{1}{4}e^{s+t}\right) = \frac{1}{7}\left(4 + 2e^s + e^{s+t}\right). \end{aligned}$$

b) We have

$$\frac{\partial M_{X,Y}}{\partial s}(s,t) = \frac{1}{7}\left(2e^s + e^{s+t}\right), \quad \frac{\partial M_{X,Y}}{\partial t}(s,t) = \frac{1}{7}e^{s+t},$$

$$\frac{\partial^2 M_{X,Y}}{\partial s^2}(s,t) = \frac{1}{7}\left(2e^s + e^{s+t}\right), \quad \frac{\partial^2 M_{X,Y}}{\partial t^2}(s,t) = \frac{1}{7}e^{s+t}, \quad \frac{\partial^2 M_{X,Y}}{\partial s\,\partial t}(s,t) = \frac{1}{7}e^{s+t}.$$

Therefore, from Proposition 11.6 on page 643,

$$\mathcal{E}(X) = \frac{\partial M_{X,Y}}{\partial s}(0,0) = \frac{3}{7}, \quad \mathcal{E}(Y) = \frac{\partial M_{X,Y}}{\partial t}(0,0) = \frac{1}{7},$$

$$\mathcal{E}\left(X^2\right) = \frac{\partial^2 M_{X,Y}}{\partial s^2}(0,0) = \frac{3}{7}, \quad \mathcal{E}\left(Y^2\right) = \frac{\partial^2 M_{X,Y}}{\partial t^2}(0,0) = \frac{1}{7}, \quad \mathcal{E}(XY) = \frac{\partial^2 M_{X,Y}}{\partial s\,\partial t}(0,0) = \frac{1}{7}.$$

Hence, $\mathcal{E}(X) = 3/7$, $\operatorname{Var}(X) = 3/7 - (3/7)^2 = 12/49$, $\mathcal{E}(Y) = 1/7$, $\operatorname{Var}(Y) = 1/7 - (1/7)^2 = 6/49$, and

$$\operatorname{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y) = \frac{1}{7} - \frac{3}{7} \cdot \frac{1}{7} = \frac{4}{49}.$$

c) From Proposition 11.7 on page 645 and part (a),

$$M_X(s) = M_{X,Y}(s, 0) = \frac{1}{7}\left(4 + 2e^s + e^{s+0}\right) = \frac{3}{7}e^s + \frac{4}{7}.$$

Referring now to Table 11.1 on page 634 and applying the uniqueness property of MGFs, we see that $X \sim \mathcal{B}(1, 3/7)$ or, equivalently, X has the Bernoulli distribution with parameter $3/7$. Also,

$$M_Y(t) = M_{X,Y}(0, t) = \frac{1}{7}\left(4 + 2e^0 + e^{0+t}\right) = \frac{1}{7}e^t + \frac{6}{7}.$$

Referring again to Table 11.1 and applying the uniqueness property of MGFs, we see that $Y \sim \mathcal{B}(1, 1/7)$ or, equivalently, Y has the Bernoulli distribution with parameter $1/7$.

d) From Proposition 11.10 on page 647 and part (a),

$$M_{X+Y}(t) = M_{X,Y}(t, t) = \frac{1}{7}\left(4 + 2e^t + e^{t+t}\right) = \frac{1}{7}\left(4 + 2e^t + e^{2t}\right) = \frac{4}{7} + \frac{2}{7}e^t + \frac{1}{7}e^{2t}.$$

Applying the uniqueness property of MGFs, we conclude that

$$p_{X+Y}(z) = \begin{cases} 4/7, & \text{if } z = 0; \\ 2/7, & \text{if } z = 1; \\ 1/7, & \text{if } z = 2; \\ 0, & \text{otherwise.} \end{cases}$$

e) We claim that $X + Y$ does not have a binomial distribution. Suppose to the contrary that it does, say, $X + Y \sim \mathcal{B}(2, p)$. Then

$$\frac{4}{7} = p_{X+Y}(0) = \binom{2}{0} p^0 (1 - p)^{2-0} = (1 - p)^2$$

and

$$\frac{1}{7} = p_{X+Y}(2) = \binom{2}{2} p^2 (1 - p)^{2-2} = p^2.$$

Therefore,

$$1 = p + (1 - p) = \sqrt{\frac{1}{7}} + \sqrt{\frac{4}{7}} = \frac{3}{\sqrt{7}},$$

or $\sqrt{7} = 3$, which is not the case. Hence, $X + Y$ does not have a binomial distribution.

f) From parts (a) and (c),

$$M_{X,Y}(s, t) = \frac{1}{7}\left(4 + 2e^s + e^{s+t}\right) = \frac{4}{7} + \frac{2}{7}e^s + \frac{1}{7}e^{s+t}$$

and

$$M_X(s)M_Y(t) = \left(\frac{3}{7}e^s + \frac{4}{7}\right)\left(\frac{1}{7}e^t + \frac{6}{7}\right) = \frac{24}{49} + \frac{18}{49}e^s + \frac{4}{49}e^t + \frac{3}{49}e^{s+t}.$$

Thus, $M_{X,Y}(s, t) \neq M_X(s)M_Y(t)$ and, hence, from Proposition 11.9 on page 646, X and Y are not independent random variables.

11.88

a) Answers will vary.

b) As we know, the mean of a $\mathcal{P}(3)$ random variable is 3. So, in view of the law of large numbers, we would expect the average value of the 10,000 numbers obtained in part (a) to be roughly 3.
c) Answers will vary, but should be close to 3. We got 2.9747.

11.89
a) As g is Riemann integrable on $[a, b]$, it is bounded thereon, say, by M. Define $h(x) = g(x)/f(x)$ if $x \in [a, b]$, and $h(x) = 0$ otherwise. Then

$$\int_{-\infty}^{\infty} |h(x)| f(x)\,dx = \int_a^b \left|\frac{g(x)}{f(x)}\right| f(x)\,dx = \int_a^b |g(x)|\,dx \le M \int_a^b 1\,dx = (b-a)M < \infty.$$

Consequently, we conclude from the FEF that the random variable $h(X)$ has finite expectation. However, $P\big(g(X)/f(X) = h(X)\big) \ge P\big(X \in [a, b]\big) = 1$. Hence, $g(X)/f(X) = h(X)$ with probability 1. In particular, then, $g(X)/f(X)$ has the same probability distribution as $h(X)$. Thus, $g(X)/f(X)$ has finite expectation.
b) Let X be as in part (a) and let h be as in the solution to part (a). From the FEF and the result of part (a),

$$\mathcal{E}\big(h(X)\big) = \int_{-\infty}^{\infty} h(x) f(x)\,dx = \int_a^b \frac{g(x)}{f(x)} f(x)\,dx = \int_a^b g(x)\,dx.$$

Also, from the solution to part (a), $g(X)/f(X)$ has the same probability distribution as $h(X)$ and, hence, the same mean. Because $X_1, X_2, \ldots$ are independent and identically distributed random variables, so are the random variables $g(X_1)/f(X_1), g(X_2)/f(X_2), \ldots$. Applying the strong law of large numbers to these latter random variables, we get

$$\lim_{n\to\infty} \hat{I}_n(f, g) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \frac{g(X_k)}{f(X_k)} = \mathcal{E}\left(\frac{g(X)}{f(X)}\right) = \int_a^b g(x)\,dx,$$

with probability 1.
c) We have

$$\int_{-\infty}^{\infty} \big(h(x)\big)^2 f(x)\,dx = \int_a^b \left(\frac{g(x)}{f(x)}\right)^2 f(x)\,dx = \int_a^b \frac{\big(g(x)\big)^2}{f(x)}\,dx.$$

Hence, from the FEF, $\big(h(X)\big)^2$ has finite expectation if and only if

$$\int_a^b \frac{\big(g(x)\big)^2}{f(x)}\,dx < \infty. \qquad (*)$$

We note that $\hat{I}_n(f, g)$ has finite variance if and only if $g(X)/f(X)$ has finite second moment, which is the case, if and only if $\big(h(X)\big)^2$ has finite expectation. Thus, we have shown that $\hat{I}_n(f, g)$ has finite variance if and only if Relation $(*)$ holds. In that case,

$$\begin{aligned}\operatorname{Var}\Big(\hat{I}_n(f, g)\Big) &= \operatorname{Var}\left(\frac{1}{n}\sum_{k=1}^{n} \frac{g(X_k)}{f(X_k)}\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{k=1}^{n} \frac{g(X_k)}{f(X_k)}\right) = \frac{1}{n^2}\cdot n\operatorname{Var}\big(h(X)\big) \\ &= \frac{1}{n}\left(\mathcal{E}\Big(\big(h(X)\big)^2\Big) - \Big(\mathcal{E}\big(h(X)\big)\Big)^2\right) = \frac{1}{n}\left(\int_a^b \frac{\big(g(x)\big)^2}{f(x)}\,dx - \left(\int_a^b g(x)\,dx\right)^2\right).\end{aligned}$$

d) Suppose $a = 0$, $b = 1$, and f is the PDF of a $\mathcal{U}(0, 1)$ random variable. Then $f(x) = 1$ if $0 < x < 1$, and $f(x) = 0$ otherwise. However, because we can change the value of a PDF at a finite number of points without affecting the distribution of the corresponding random variable, we can take $f(x) = 1$ if $x \in [0, 1]$, and $f(x) = 0$ if $x \notin [0, 1]$. In this case, the result of part (b) becomes

$$\lim_{n\to\infty} \hat{I}_n(f, g) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{g(X_k)}{f(X_k)} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k) = \int_0^1 g(x)\,dx,$$

with probability 1, which is the result of the Monte Carlo integration technique presented in Exercise 11.43. Referring to part (c), we see that, in this case,

$$\operatorname{Var}\Big(\hat{I}_n(f, g)\Big) = \frac{1}{n}\left(\int_0^1 \big(g(x)\big)^2\,dx - \left(\int_0^1 g(x)\,dx\right)^2\right).$$

11.90 Let X denote the number of the 100 sampled items that are defective. Then $X \sim \mathcal{B}(100, 0.08)$. We want $P(X \geq 12)$. Note that $np = 100 \cdot 0.08 = 8$ and $n(1-p) = 100 \cdot 0.92 = 92$, both of which are 5 or greater. Applying the integral De Moivre–Laplace theorem, Proposition 11.11 on page 661, we get

$$\begin{aligned} P(X \geq 12) = P(12 \leq X \leq 100) &\approx \Phi\left(\frac{100 + \frac{1}{2} - 100 \cdot 0.08}{\sqrt{100 \cdot 0.08 \cdot 0.92}}\right) - \Phi\left(\frac{12 - \frac{1}{2} - 100 \cdot 0.08}{\sqrt{100 \cdot 0.08 \cdot 0.92}}\right) \\ &= \Phi(34.09595) - \Phi(1.29012) = 0.0985. \end{aligned}$$

11.91 Let X denote the number of the 10,000 sampled males who have Fragile X Syndrome. Then we have $X \sim \mathcal{B}(10{,}000, 1/1500)$. Note that $np = 10{,}000/1500 = 20/3$ and that

$$\sqrt{np(1-p)} = \sqrt{10{,}000 \cdot \frac{1}{1500} \cdot \frac{1499}{1500}} = \frac{\sqrt{1499}}{15}.$$

Applying the integral De Moivre–Laplace theorem, we get

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - \frac{20}{3}}{\sqrt{1499}/15}\right) - \Phi\left(\frac{a - \frac{1}{2} - \frac{20}{3}}{\sqrt{1499}/15}\right), \qquad 0 \leq a \leq b \leq 10{,}000.$$

a) We have

$$\begin{aligned} P(X > 7) = P(8 \leq X \leq 10{,}000) &\approx \Phi\left(\frac{10{,}000 + \frac{1}{2} - \frac{20}{3}}{\sqrt{1499}/15}\right) - \Phi\left(\frac{8 - \frac{1}{2} - \frac{20}{3}}{\sqrt{1499}/15}\right) \\ &= \Phi(3871.88585) - \Phi(0.32286) = 0.373. \end{aligned}$$

b) We have

$$\begin{aligned} P(X \leq 10) = P(0 \leq X \leq 10) &\approx \Phi\left(\frac{10 + \frac{1}{2} - \frac{20}{3}}{\sqrt{1499}/15}\right) - \Phi\left(\frac{0 - \frac{1}{2} - \frac{20}{3}}{\sqrt{1499}/15}\right) \\ &= \Phi(1.48514) - \Phi(-2.77656) = 0.928. \end{aligned}$$

11.92

b) We have

Probability	Local	Integral	Poisson
$P(X > 7)$	0.375	0.373	0.352
$P(X \leq 10)$	0.930	0.928	0.923

11.93 Let X denote the number of the 1500 cases in which the new oral vaccine is effective. Then we have $X \sim \mathcal{B}(1500, 0.8)$. Note that $np = 1500 \cdot 0.8 = 1200$ and that

$$\sqrt{np(1-p)} = \sqrt{1500 \cdot 0.8 \cdot 0.2} = 4\sqrt{15}.$$

Applying the integral De Moivre–Laplace theorem, we get

$$P(a \le X \le b) \approx \Phi\left(\frac{b + \frac{1}{2} - 1200}{4\sqrt{15}}\right) - \Phi\left(\frac{a - \frac{1}{2} - 1200}{4\sqrt{15}}\right), \qquad 0 \le a \le b \le 1500.$$

a) We have

$$P(X = 1225) = P(1225 \le X \le 1225) \approx \Phi\left(\frac{1225 + \frac{1}{2} - 1200}{4\sqrt{15}}\right) - \Phi\left(\frac{1225 - \frac{1}{2} - 1200}{4\sqrt{15}}\right)$$
$$= \Phi(1.64602) - \Phi(1.58147) = 0.00701.$$

b) We have

$$P(X \ge 1175) = P(1175 \le X \le 1500) \approx \Phi\left(\frac{1500 + \frac{1}{2} - 1200}{4\sqrt{15}}\right) - \Phi\left(\frac{1175 - \frac{1}{2} - 1200}{4\sqrt{15}}\right)$$
$$= \Phi(19.39719) - \Phi(-1.64602) = 0.950.$$

c) We have

$$P(1150 \le X \le 1250) \approx \Phi\left(\frac{1250 + \frac{1}{2} - 1200}{4\sqrt{15}}\right) - \Phi\left(\frac{1150 - \frac{1}{2} - 1200}{4\sqrt{15}}\right)$$
$$= \Phi(3.25976) - \Phi(-3.25976) = 0.999.$$

11.94 For convenience, set $S = X_1 + X_2 + X_3$.
a) The following table provides the 27 possible outcomes and their sums:

(X_1, X_2, X_3)	S	(X_1, X_2, X_3)	S	(X_1, X_2, X_3)	S
(1, 1, 1)	3	(2, 1, 1)	4	(3, 1, 1)	5
(1, 1, 2)	4	(2, 1, 2)	5	(3, 1, 2)	6
(1, 1, 3)	5	(2, 1, 3)	6	(3, 1, 3)	7
(1, 2, 1)	4	(2, 2, 1)	5	(3, 2, 1)	6
(1, 2, 2)	5	(2, 2, 2)	6	(3, 2, 2)	7
(1, 2, 3)	6	(2, 2, 3)	7	(3, 2, 3)	8
(1, 3, 1)	5	(2, 3, 1)	6	(3, 3, 1)	7
(1, 3, 2)	6	(2, 3, 2)	7	(3, 3, 2)	8
(1, 3, 3)	7	(2, 3, 3)	8	(3, 3, 3)	9

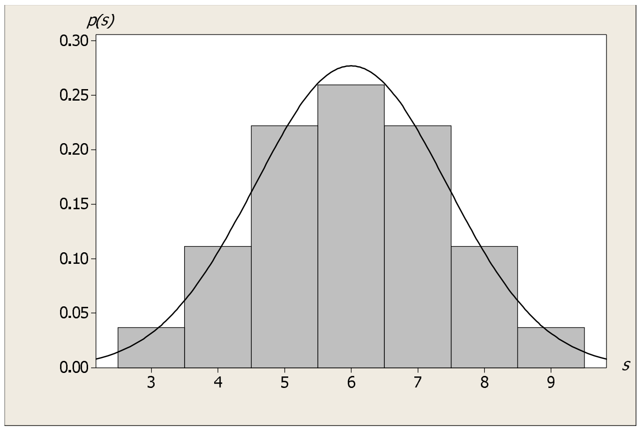
Because each of the 27 outcomes is equally likely, we obtain the following probability distribution of the random variable S.

s	3	4	5	6	7	8	9
$p_S(s)$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{7}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{1}{27}$

b) From the FPF and part (a),

$$P(S \le 7) = \sum_{s \le 7} p_S(s) = \frac{1}{27} + \frac{3}{27} + \frac{6}{27} + \frac{7}{27} + \frac{6}{27} = \frac{23}{27}.$$

c) In the following graph, we use $p(s)$ instead of $p_S(s)$. For comparison purposes, we have superimposed the appropriate normal curve.



As we see, the probability histogram is quite bell shaped and reasonably resembles the superimposed normal curve.

d) We have

$$\mu = \mathcal{E}(X_j) = \sum_x x p_{X_j}(x) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = 2$$

and

$$\sigma^2 = \mathcal{E}\left(X_j^2\right) - \left(\mathcal{E}(X_j)\right)^2 = \sum_x x^2 p_{X_j}(x) - 2^2 = 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} + 3^2 \cdot \frac{1}{3} - 4 = \frac{2}{3}.$$

Applying the normal approximation with a continuity correction, we get

$$\begin{aligned} P(S \le 7) = P(3 \le S \le 7) &\approx \Phi\left(\frac{7 + \frac{1}{2} - 3 \cdot 2}{\sqrt{2/3}\sqrt{3}}\right) - \Phi\left(\frac{0 - \frac{1}{2} - 3 \cdot 2}{\sqrt{2/3}\sqrt{3}}\right) \\ &= \Phi(1.06066) - \Phi(-4.59619) = 0.856. \end{aligned}$$

This approximate probability compares quite favorably with the exact probability of $23/27 = 0.852$ found in part (b).

11.95 Let $X_1, \ldots, X_{140}$ denote the scores of the 140 people. We assume that these random variables are independent and identically distributed with common mean 73 and common standard deviation 5. In view of the central limit theorem, in the form of the fourth bulleted item on page 667, we have

$$\begin{aligned} P\left(\bar{X}_{140} > 74\right) &= 1 - P\left(\bar{X}_{140} \le 74\right) = 1 - P\left(0 \le \bar{X}_{140} \le 74\right) \\ &\approx 1 - \left(\Phi\left(\frac{74 - 73}{5/\sqrt{140}}\right) - \Phi\left(\frac{0 - 73}{5/\sqrt{140}}\right)\right) \\ &= 1 - \left(\Phi(2.36643) - \Phi(-172.74953)\right) = 0.00898. \end{aligned}$$

11.96 Let X_j $(1 \le j \le 20)$ denote the lifetime, in hours, of bulb j. Also, let Y_j $(1 \le j \le 19)$ denote the replacement time, in hours, to install bulb $j+1$ after bulb j burns out. We know that $X_1, \ldots, X_{20}$ are independent and identically distributed random variables, each with mean 300 and standard deviation 300, and we assume that $Y_1, \ldots, Y_{19}$ are independent and identically distributed random variables, each with mean 0.5 and standard deviation 0.1, and independent of the X_js. Let T denote the total time that the 20 bulbs last, so that $T = X_1 + \cdots + X_{20} + Y_1 + \cdots + Y_{19}$. From the central limit theorem,

$$X_1 + \cdots + X_{20} \approx \mathcal{N}\left(20 \cdot 300, 20 \cdot 300^2\right) \qquad \text{and} \qquad Y_1 + \cdots + Y_{19} \approx \mathcal{N}\left(19 \cdot 0.5, 19 \cdot 0.1^2\right).$$

Therefore, from the independence of $X_1 + \cdots + X_{20}$ and $Y_1 + \cdots + Y_{19}$, we conclude that T is approximately normally distributed with

$$\mu_T = 20 \cdot 300 + 19 \cdot 0.5 = 6009.5 \qquad \text{and} \qquad \sigma_T^2 = 20 \cdot 300^2 + 19 \cdot 0.1^2 = 1{,}800{,}000.19.$$

Consequently,

$$\begin{aligned} P(T \le 7000) &= P(0 \le T \le 7000) \\ &\approx \Phi\left(\frac{7000 - 6009.5}{\sqrt{1{,}800{,}000.19}}\right) - \Phi\left(\frac{0 - 6009.5}{\sqrt{1{,}800{,}000.19}}\right) \\ &= \Phi(0.73828) - \Phi(-4.47922) = 0.770. \end{aligned}$$

11.97 All units are in years. We assume that the difference between a true age and its rounded age is uniformly distributed on the interval from -2.5 to 2.5. Now, for $1 \le j \le 48$, let X_j and Y_j denote the age and rounded age, respectively, of the jth person. We assume that $X_1, \ldots, X_{48}$ are independent and identically distributed, that $Y_1, \ldots, Y_{48}$ are independent and identically distributed, and that the X_js and Y_js are independent. Let $D_j = X_j - Y_j$. By assumption, $D_j \sim \mathcal{U}(-2.5, 2.5)$ and, from Proposition 6.13 on page 297, we conclude that $D_1, \ldots, D_{48}$ are independent. We want to find the approximate value of $P\left(\left|\bar{X}_{48} - \bar{Y}_{48}\right| \le 0.25\right)$. However,

$$\bar{X}_{48} - \bar{Y}_{48} = \frac{1}{48}\sum_{j=1}^{n} X_j - \frac{1}{48}\sum_{j=1}^{n} Y_j = \frac{1}{48}\sum_{j=1}^{n}(X_j - Y_j) = \frac{1}{48}\sum_{j=1}^{n} D_j = \bar{D}_{48}.$$

Because $D_j \sim \mathcal{U}(-2.5, 2.5)$, we have $\mu = \mathcal{E}\left(D_j\right) = 0$ and $\sigma^2 = \operatorname{Var}\left(D_j\right) = 5^2/12$. Applying the central limit theorem, in the form of Equation (11.43) on page 668, we have

$$P\left(\left|\bar{X}_{48} - \bar{Y}_{48}\right| \le 0.25\right) = P\left(\left|\bar{D}_{48} - 0\right| \le 0.25\right) \approx 2\Phi\left(\frac{\sqrt{48} \cdot 0.25}{5/\sqrt{12}}\right) - 1 = 2\Phi(1.2) - 1 = 0.770.$$

11.98 Let X_j $(1 \le j \le 1250)$ denote the number of claims filed by policyholder j during the 1-year period. We know that $X_1, \ldots, X_{1250}$ are independent random variables, each having the Poisson distribution with mean (parameter) 2.

a) The total number of claims during the 1-year period is $T = X_1 + \cdots + X_{1250}$, which, by Proposition 6.20(b) on page 311, has a $\mathcal{P}(2500)$ distribution. Hence,

$$P(2450 \le T \le 2600) = e^{-2500} \sum_{k=2450}^{2600} \frac{(2500)^k}{k!}.$$

b) We have $\mu = \mathcal{E}(X_j) = 2$ and $\sigma^2 = \mathrm{Var}(X_j) = 2$. Applying the central limit theorem, we get

$$\begin{aligned} P(2450 \le T \le 2600) &= P(2450 \le X_1 + \cdots + X_{1250} \le 2600) \\ &\approx \Phi\left(\frac{2600 - 1250 \cdot 2}{\sqrt{2}\sqrt{1250}}\right) - \Phi\left(\frac{2450 - 1250 \cdot 2}{\sqrt{2}\sqrt{1250}}\right) \\ &= \Phi(2) - \Phi(-1) = 0.819. \end{aligned}$$

11.99 As $X_1, X_2, \ldots$ are independent and identically distributed random variables, so are the random variables $\ln X_1, \ln X_2, \ldots$. By assumption, these latter random variables have finite nonzero variance. Noting that $\ln\left(\prod_{j=1}^n X_j\right) = \sum_{j=1}^n \ln X_j$, we conclude from the central limit theorem that, for large n, the random variable $\ln\left(\prod_{j=1}^n X_j\right)$ has approximately a normal distribution or, equivalently, $\prod_{j=1}^n X_j$ has approximately a lognormal distribution.

11.100 We apply Equation (11.47) on page 671 with $n = 30$, $\bar{x}_{30} = 27.97$, $\sigma = 10.04$, and $\gamma = 0.90$. A 90% confidence interval for μ is from

$$27.97 - \Phi^{-1}(0.95) \cdot \frac{10.04}{\sqrt{30}} \qquad \text{to} \qquad 27.97 + \Phi^{-1}(0.95) \cdot \frac{10.04}{\sqrt{30}}$$

or from 24.955 to 30.985. We can be 90% confident that the mean commute time in Washington, D.C., is somewhere between 24.955 and 30.985 minutes.

Theory Exercises

11.101 We know that $X_1, X_2, \ldots$ are independent Bernoulli random variables with common parameter p and, hence, with common finite mean p. As usual, we let $\bar{X}_n = (X_1 + \cdots + X_n)/n$.
a) From the strong law of large numbers, Theorem 11.2 on page 654, $\lim_{n\to\infty} \bar{X}_n = p$ with probability 1. Let

$$E = \left\{ \omega : \lim_{n\to\infty} \bar{X}_n(\omega) = p \right\}.$$

Then $P(E) = 1$. Let $\omega \in E$. Because f is continuous at p and $\lim_{n\to\infty} \bar{X}_n(\omega) = p$, we conclude that

$$\lim_{n\to\infty} f\left(\bar{X}_n(\omega)\right) = f\left(\lim_{n\to\infty} \bar{X}_n(\omega)\right) = f(p).$$

Because the preceding result holds for all $\omega \in E$, we have $\lim_{n\to\infty} f\left(\bar{X}_n\right) = f(p)$ with probability 1.
b) Let $S_n = X_1 + \cdots + X_n$. Then $S_n \sim \mathcal{B}(n, p)$. Applying the FEF with $g(x) = f(x/n)$ yields

$$\begin{aligned} \mathcal{E}\left(f\left(\frac{X_1 + \cdots + X_n}{n}\right)\right) &= \mathcal{E}\big(f(S_n/n)\big) = \mathcal{E}\big(g(S_n)\big) = \sum_x g(x) p_{S_n}(x) \\ &= \sum_x f(x/n) p_{S_n}(x) = \sum_{k=0}^n f(k/n) \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

c) Suppose that f is bounded by M on [0, 1]. By the triangle inequality,

$$|f(x) - f(y)| \le |f(x)| + |f(y)| \le M + M = 2M, \qquad x, y \in [0, 1].$$

Now, let $\epsilon > 0$ be given. Because f is uniformly continuous on [0, 1], we can choose a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/2$ whenever $x, y \in [0, 1]$ and $|x - y| < \delta$. Then we choose $N \in \mathcal{N}$ such that, if $n \ge N$, then $M/2n\delta^2 < \epsilon/2$.

Noting that $\sum_{k=0}^{n} p_{S_n}(k) = 1$, the triangle inequality yields

$$\begin{aligned}
\left|f(p) - B_n(p)\right| &= \left|f(p) - \sum_{k=0}^{n} f(k/n)\binom{n}{k}p^k(1-p)^{n-k}\right| = \left|f(p) - \sum_{k=0}^{n} f(k/n)p_{S_n}(k)\right| \\
&= \left|\sum_{k=0}^{n}\big(f(p) - f(k/n)\big)p_{S_n}(k)\right| \le \sum_{k=0}^{n}\left|f(p) - f(k/n)\right| p_{S_n}(k) \\
&= \sum_{|k/n-p|<\delta}\left|f(p) - f(k/n)\right| p_{S_n}(k) + \sum_{|k/n-p|\ge\delta}\left|f(p) - f(k/n)\right| p_{S_n}(k).
\end{aligned}$$

However, for all $0 \le p \le 1$,

$$\sum_{|k/n-p|<\delta}\left|f(p) - f(k/n)\right| p_{S_n}(k) \le \sum_{|k/n-p|<\delta}\frac{\epsilon}{2}p_{S_n}(k) \le \frac{\epsilon}{2}\sum_{k=0}^{n} p_{S_n}(k) = \frac{\epsilon}{2}.$$

Moreover, by the FPF and Chebyshev's inequality, we have, for all $0 \le p \le 1$ and $n \ge N$,

$$\begin{aligned}
\sum_{|k/n-p|\ge\delta}\left|f(p) - f(k/n)\right| p_{S_n}(k) &\le 2M \sum_{|k/n-p|\ge\delta} p_{S_n}(k) = 2M\,P\left(\left|\frac{S_n}{n} - p\right| \ge \delta\right) \\
&\le 2M\frac{\mathrm{Var}(S_n/n)}{\delta^2} = 2M\frac{\mathrm{Var}(S_n)}{n^2\delta^2} = 2M\frac{np(1-p)}{n^2\delta^2} \\
&= 2M\frac{p(1-p)}{n\delta^2} \le \frac{M}{2n\delta^2} < \frac{\epsilon}{2},
\end{aligned}$$

where, in the penultimate inequality, we used the fact that $p(1-p) \le 1/4$ for all $p \in [0, 1]$.

Therefore, for all $0 \le p \le 1$ and $n \ge N$,

$$\begin{aligned}
\left|f(p) - B_n(p)\right| &\le \sum_{|k/n-p|<\delta}\left|f(p) - f(k/n)\right| p_{S_n}(k) + \sum_{|k/n-p|\ge\delta}\left|f(p) - f(k/n)\right| p_{S_n}(k) \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Hence, we have shown that $\lim_{n\to\infty} B_n(x) = f(x)$ uniformly for $x \in [0, 1]$.

11.102 From the complementation rule,

$$P(|X - X_n| < \epsilon) + P(|X - X_n| \ge \epsilon) = 1.$$

Thus, we see that $\{X_n\}_{n=1}^{\infty}$ converges in probability to X if and only if $\lim_{n\to\infty} P(|X - X_n| < \epsilon) = 1$ for each $\epsilon > 0$.

a) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with common finite mean μ. The weak law of large numbers states that, for each $\epsilon > 0$, $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$. This statement is equivalent to saying that $\{X_n\}_{n=1}^{\infty}$ converges in probability to μ (the constant random variable μ).

b) Let X be a random variable and let $X_1, X_2, \ldots$ be independent and identically distributed random variables, all having the same distribution as X. A statistic $T_n = g(X_1, \ldots, X_n)$ is a consistent estimator of an unknown parameter θ of the distribution of X if, for each $\epsilon > 0$, $\lim_{n\to\infty} P(|T_n - \theta| < \epsilon) = 1$. This statement is equivalent to saying that $\{T_n\}_{n=1}^{\infty}$ converges in probability to θ.

c) Let $\epsilon > 0$. By the triangle inequality, if $|a| < \epsilon/2$ and $|b| < \epsilon/2$, then $|a+b| < \epsilon$. Thus,

$$\begin{aligned}\{|(X+Y)-(X_n+Y_n)| \ge \epsilon\} &= \{|(X-X_n)+(Y-Y_n)| \ge \epsilon\}\\ &\subset \{|X-X_n| \ge \epsilon/2\} \cup \{|Y-Y_n| \ge \epsilon/2\}.\end{aligned}$$

Therefore, by Boole's inequality,

$$P(|(X+Y)-(X_n+Y_n)| \ge \epsilon) \le P(|X-X_n| \ge \epsilon/2) + P(|Y-Y_n| \ge \epsilon/2).$$

By assumption, both terms on the right of the previous relation converge to 0 as $n \to \infty$ and, consequently, so does the term on the left. Hence, $\{X_n+Y_n\}_{n=1}^{\infty}$ converges in probability to $X+Y$.

11.103 Let $c_n = n/(n-1)$ and note that $\lim_{n\to\infty} c_n = 1$. Now,

$$\begin{aligned}S_n^2 &= \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j-\bar{X}_n\right)^2 = c_n \cdot \frac{1}{n}\sum_{j=1}^{n}\left(X_j^2 - 2\bar{X}_n X_j + \bar{X}_n^2\right)\\ &= c_n\left(\frac{1}{n}\sum_{j=1}^{n}X_j^2 - 2\bar{X}_n \cdot \frac{1}{n}\sum_{j=1}^{n}X_j + \bar{X}_n^2\right) = c_n\left(\frac{1}{n}\sum_{j=1}^{n}X_j^2 - \bar{X}_n^2\right).\end{aligned}$$

From Exercises 11.40 and 11.102(b), the first term within the parentheses on the right of the previous display converges in probability to $\mathcal{E}\left(X^2\right)$ and, from Example 11.17 and Exercises 11.102(a) and (c), the second term within the parentheses on the right of the previous display converges in probability to μ_X^2. Hence, from Exercise 11.102(c),

$$\lim_{n\to\infty} c_n\left(\frac{1}{n}\sum_{j=1}^{n}X_j^2 - \bar{X}_n^2\right) = 1 \cdot \left(\mathcal{E}\left(X^2\right) - \mu_X^2\right) = \sigma_X^2,$$

in probability. Therefore, $\left\{S_n^2\right\}_{n=1}^{\infty}$ converges in probability to σ_X^2 or, equivalently, S_n^2 is a consistent estimator of σ_X^2.

Advanced Exercises

11.104 Applying the binomial series, as presented in Equation (5.45) on page 241, we find that

$$M_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^{\alpha} = \left(1-\frac{t}{\lambda}\right)^{-\alpha} = \sum_{n=0}^{\infty}\binom{-\alpha}{n}\left(-\frac{t}{\lambda}\right)^{n} = \sum_{n=0}^{\infty}\frac{(-1)^n}{\lambda^n}\binom{-\alpha}{n}t^n = \sum_{n=0}^{\infty}a_n t^n.$$

Therefore, from the moment generation property of MGFs,

$$\mathcal{E}\left(X^n\right) = M_X^{(n)}(0) = n!\,a_n = n! \cdot \frac{(-1)^n}{\lambda^n}\binom{-\alpha}{n} = \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)}{\lambda^n}.$$

11.105

a) We have, for $|t| < 1$,

$$\begin{aligned}M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = \frac{1}{2}\int_{-\infty}^{\infty} e^{tx}e^{-|x|}\,dx = \frac{1}{2}\left(\int_{-\infty}^{0} e^{(t+1)x}\,dx + \int_{0}^{\infty} e^{(t-1)x}\,dx\right)\\ &= \frac{1}{2}\left(\frac{1}{1+t} + \frac{1}{1-t}\right) = \frac{1}{1-t^2}.\end{aligned}$$

b) Applying the geometric series, we find that, for $|t| < 1$,

$$M_X(t) = \frac{1}{1-t^2} = \sum_{n=0}^{\infty} \left(t^2\right)^n = \sum_{n=0}^{\infty} t^{2n}.$$

Let a_n denote the coefficient of t^n in the preceding power series expansion of M_X. From the moment generation property of MGFs,

$$\mathcal{E}\left(X^n\right) = M_X^{(n)}(0) = n!\, a_n = \begin{cases} n!, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

11.106

a) From the exponential series,

$$M_Z(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{\left(t^2/2\right)^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!\, 2^n} t^{2n}.$$

Let a_n denote the coefficient of t^n in the preceding power series expansion of M_Z. From the moment generation property of MGFs,

$$\mathcal{E}\left(Z^n\right) = M_Z^{(n)}(0) = n!\, a_n = \begin{cases} \dfrac{n!}{(n/2)!\, 2^{n/2}}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

b) We know that $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Hence, from part (a),

$$\frac{\mathcal{E}\left((X-\mu)^n\right)}{\sigma^n} = \mathcal{E}\left(\left(\frac{X-\mu}{\sigma}\right)^n\right) = \begin{cases} \dfrac{n!}{(n/2)!\, 2^{n/2}}, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

Consequently,

$$\mathcal{E}\left((X-\mu)^n\right) = \begin{cases} \dfrac{n!}{(n/2)!\, 2^{n/2}}\, \sigma^n, & \text{if } n \text{ is even;} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

c) We can write $X = \sigma Z + \mu$, where $Z \sim \mathcal{N}(0, 1)$. Applying the binomial theorem and referring to part (a), we get

$$\begin{aligned} \mathcal{E}(X^n) = \mathcal{E}\left((\sigma Z + \mu)^n\right) &= \mathcal{E}\left(\sum_{k=0}^{n} \binom{n}{k} (\sigma Z)^k \mu^{n-k}\right) = \sum_{k=0}^{n} \binom{n}{k} \sigma^k \mu^{n-k} \mathcal{E}\left(Z^k\right) \\ &= \sum_{\substack{k=0 \\ k \text{ even}}}^{n} \binom{n}{k} \sigma^k \mu^{n-k} \frac{k!}{(k/2)!\, 2^{k/2}} = \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \sigma^{2j} \mu^{n-2j} \frac{(2j)!}{j!\, 2^j} \\ &= n!\, \mu^n \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{\sigma^{2j}}{(n-2j)!\, j!\, \left(2\mu^2\right)^j}. \end{aligned}$$

11.107

a) We have

$$K_X'(t) = \frac{d}{dt} \ln M_X(t) = \frac{M_X'(t)}{M_X(t)} \qquad \text{and} \qquad K_X''(t) = \frac{M_X(t) M_X''(t) - \left(M_X'(t)\right)^2}{\left(M_X(t)\right)^2}.$$

Therefore, by the moment generating property of MGFs,

$$\kappa_1 = K_X'(0) = \frac{M_X'(0)}{M_X(0)} = \frac{\mathcal{E}(X)}{1} = \mathcal{E}(X)$$

and

$$\kappa_2 = K_X''(0) = \frac{M_X(0)M_X''(0) - \big(M_X'(0)\big)^2}{\big(M_X(0)\big)^2} = \frac{1 \cdot \mathcal{E}\big(X^2\big) - \big(\mathcal{E}(X)\big)^2}{1^2} = \mathcal{E}\big(X^2\big) - \big(\mathcal{E}(X)\big)^2 = \mathrm{Var}(X)\,.$$

b) From calculus, we know that

$$a_n = \frac{K_X^{(n)}(0)}{n!} = \frac{\kappa_n}{n!},$$

so that $\kappa_n = n!\,a_n$.

c) We consider each distribution given for X in turn.

- $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$: In this case, $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$, so that

  $$K_X(t) = \ln M_X(t) = \ln e^{\mu t + \sigma^2 t^2/2} = \mu t + \frac{\sigma^2 t^2}{2} = \mu t + \frac{\sigma^2}{2} t^2.$$

  From part (b), we conclude that

  $$\kappa_1 = 1!\,a_1 = 1 \cdot \mu = \mu \qquad \text{and} \qquad \kappa_2 = 2!\,a_2 = 2 \cdot \frac{\sigma^2}{2} = \sigma^2,$$

  and $\kappa_n = 0$ if $n \geq 3$.

- $X \sim \mathcal{P}(\lambda)$: In this case, $M_X(t) = e^{\lambda(e^t - 1)}$, so that

  $$K_X(t) = \ln M_X(t) = \ln e^{\lambda(e^t-1)} = \lambda\left(e^t - 1\right) = \lambda \sum_{n=1}^{\infty} \frac{t^n}{n!} = \sum_{n=1}^{\infty} \frac{\lambda}{n!} t^n.$$

  From part (b), we conclude that

  $$\kappa_n = n!\,a_n = n!\,\frac{\lambda}{n!} = \lambda, \qquad n \in \mathcal{N}.$$

- $X \sim \mathcal{E}(\lambda)$: In this case, $M_X(t) = \lambda/(\lambda - t)$, so that

  $$K_X(t) = \ln M_X(t) = \ln\left(\frac{\lambda}{\lambda - t}\right) = \ln\,(1 - t/\lambda)^{-1} = -\ln(1 - t/\lambda)$$
  $$= -\sum_{n=1}^{\infty} (-1)^{n-1} \frac{(-t/\lambda)^n}{n} = \sum_{n=1}^{\infty} \frac{1}{n\lambda^n} t^n.$$

  From part (b), we conclude that

  $$\kappa_n = n!\,a_n = n!\,\frac{1}{n\lambda^n} = \frac{(n-1)!}{\lambda^n}, \qquad n \in \mathcal{N}.$$

11.108

a) From the solution to Exercise 10.149, we know that $P(N = n) = 1/(n-1)! - 1/n!$. Applying the FEF and then the exponential series, we get

$$M_N(t) = \mathcal{E}\left(e^{tN}\right) = \sum_{n=1}^{\infty} e^{tn} P(N = n) = \sum_{n=1}^{\infty} \frac{e^{tn}}{(n-1)!} - \sum_{n=1}^{\infty} \frac{e^{tn}}{n!}$$
$$= e^t \sum_{n=0}^{\infty} \frac{e^{tn}}{n!} - \sum_{n=1}^{\infty} \frac{e^{tn}}{n!} = e^t e^{e^t} - \left(e^{e^t} - 1\right) = 1 + e^{e^t}\left(e^t - 1\right).$$

b) We have

$$M_N'(t) = e^{e^t} e^t + e^{e^t} e^t (e^t - 1) = e^{e^t + 2t} \qquad \text{and} \qquad M_N''(t) = M_N'(t)\left(e^t + 2\right).$$

Therefore, from the moment generating property of MGFs,

$$\mathcal{E}(X) = M_N'(0) = e^{1+0} = e \qquad \text{and} \qquad \mathcal{E}\left(X^2\right) = M_N''(0) = M_N'(0)\left(e^0 + 2\right) = 3e.$$

Also,

$$\operatorname{Var}(X) = \mathcal{E}\left(X^2\right) - \left(\mathcal{E}(X)\right)^2 = 3e - e^2 = (3 - e)e.$$

11.109

a) Let $a, b \in \mathcal{R}$ with $a < b$. For n sufficiently large, $P(a < X_n < b) = (b-a)/2n$ and, hence,

$$\lim_{n\to\infty} P(a < X_n < b) = \lim_{n\to\infty} \frac{b-a}{2n} = 0.$$

Heuristically, then, $\{X_n\}_{n=1}^{\infty}$ does not converge in distribution to a random variable.

b) We have

$$F_{X_n}(x) = P(X_n \le x) = \begin{cases} 0, & \text{if } x < -n; \\ (x+n)/2n, & \text{if } -n \le x < n; \\ 1, & \text{if } x \ge n. \end{cases}$$

Now, let $x \in \mathcal{R}$. Then, for n sufficiently large, we have $-n \le x < n$ so that $F_n(x) = (x+n)/2n$. Consequently,

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \frac{x+n}{2n} = \frac{1}{2}, \qquad x \in \mathcal{R}.$$

c) We claim that $\{X_n\}_{n=1}^{\infty}$ does not converge in distribution to a random variable. Suppose to the contrary that $\{X_n\}_{n=1}^{\infty}$ converges in distribution to the random variable X. From Proposition 8.1(d) on page 411, we can choose $M \in \mathcal{R}$ such that $F_X(M) > 1/2$. Because F_X has only a countable number of points of discontinuity, we can choose $x_0 \in \mathcal{C}_{F_X}$ with $x_0 \ge M$. Then, from part (b) and the definition of convergence in distribution,

$$\frac{1}{2} = \lim_{n\to\infty} F_{X_n}(x_0) = F_X(x_0) \ge F_X(M) > \frac{1}{2},$$

which is impossible. Hence, as suggested in part (a), $\{X_n\}_{n=1}^{\infty}$ does not converge in distribution to a random variable.

11.110

a) We have

$$M_N(t) = \mathcal{E}\left(e^{tN}\right) = \mathcal{E}\left(\left(e^t\right)^N\right) = P_N\left(e^t\right).$$

b) From part (a),

$$M_N'(t) = e^t P_N'\left(e^t\right)$$

and

$$M_N''(t) = e^t \cdot e^t P_N''\left(e^t\right) + e^t \cdot P_N'\left(e^t\right) = e^{2t} P_N''\left(e^t\right) + e^t P_N'\left(e^t\right).$$

Applying the moment generating property of MGFs, Proposition 11.2 on page 635, we now obtain

$$\mathcal{E}(N) = M_N'(0) = e^0 P_N'\left(e^0\right) = P_N'(1)$$

and

$$\mathcal{E}\left(N^2\right) = M_N''(0) = e^{2\cdot 0} P_N''\left(e^0\right) + e^0 P_N'\left(e^0\right) = P_N''(1) + P_N'(1).$$

Thus,

$$\operatorname{Var}(N) = \mathcal{E}\left(N^2\right) - \left(\mathcal{E}(N)\right)^2 = P_N''(1) + P_N'(1) - \left(P_N'(1)\right)^2.$$

11.111

a) From the multiplication property of MGFs (Proposition 11.4 on page 636), $\mathcal{E}\left(e^{tS_N} \mid N\right) = \left(M_X(t)\right)^N$. Hence, by the law of total expectation,

$$M_{S_N}(t) = \mathcal{E}\left(e^{tS_N}\right) = \mathcal{E}\Big(\mathcal{E}\left(e^{tS_N} \mid N\right)\Big) = \mathcal{E}\Big(\left(M_X(t)\right)^N\Big) = P_N\left(M_X(t)\right).$$

b) Referring to part (a), we get

$$M_{S_N}'(t) = P_N'\left(M_X(t)\right) M_X'(t)$$

and

$$M_{S_N}''(t) = P_N'\left(M_X(t)\right) M_X''(t) + P_N''\left(M_X(t)\right)\left(M_X'(t)\right)^2.$$

Applying the moment generating property of MGFs and Exercise 11.110(b) now yields

$$\mathcal{E}(S_N) = M_{S_N}'(0) = P_N'\left(M_X(0)\right) M_X'(0) = P_N'(1) M_X'(0) = \mu\mathcal{E}(N)$$

and

$$\begin{aligned}\mathcal{E}\left(S_N^2\right) &= M_{S_N}''(0) = P_N'\left(M_X(0)\right) M_X''(0) + P_N''\left(M_X(0)\right)\left(M_X'(0)\right)^2 \\ &= P_N'(1)\mathcal{E}\left(X^2\right) + P_N''(1)\left(\mathcal{E}(X)\right)^2 = P_N'(1)\mathcal{E}\left(X^2\right) + P_N''(1)\left(\mathcal{E}(X)\right)^2 \\ &= \sigma^2 P_N'(1) + \mu^2\left(P_N''(1) + P_N'(1)\right) = \sigma^2\mathcal{E}(N) + \mu^2\left(P_N''(1) + P_N'(1)\right).\end{aligned}$$

Consequently,

$$\begin{aligned}\operatorname{Var}(S_N) &= \mathcal{E}\left(S_N^2\right) - \left(\mathcal{E}(S_N)\right)^2 = \sigma^2\mathcal{E}(N) + \mu^2\left(P_N''(1) + P_N'(1)\right) - \left(\mu P_N'(1)\right)^2 \\ &= \sigma^2\mathcal{E}(N) + \mu^2\left(P_N''(1) + P_N'(1)\right) - \mu^2\left(P_N'(1)\right)^2 \\ &= \sigma^2\mathcal{E}(N) + \mu^2\Big(P_N''(1) + P_N'(1) - \left(P_N'(1)\right)^2\Big) \\ &= \sigma^2\mathcal{E}(N) + \mu^2\operatorname{Var}(N).\end{aligned}$$

11.112 Let

$$X = \bar{X}_2 = \frac{X_1 + X_2}{2}, \qquad Y = X_1 - \bar{X}_2 = \frac{X_1 - X_2}{2}, \qquad Z = X_2 - \bar{X}_2 = \frac{X_2 - X_1}{2}.$$

Then,

$$sX + tY + uZ = \frac{sX_1 + sX_2}{2} + \frac{tX_1 - tX_2}{2} + \frac{uX_2 - uX_1}{2} = \frac{1}{2}(s + t - u)X_1 + \frac{1}{2}(s - t + u)X_2.$$

Applying the independence of X_1 and X_2 and then recalling that the MGF of a standard normal random variable is $e^{t^2/2}$, we get

$$\begin{aligned} M_{X,Y,Z}(s,t,u) &= \mathcal{E}\left(e^{sX+tY+uZ}\right) = \mathcal{E}\left(e^{\frac{1}{2}(s+t-u)X_1+\frac{1}{2}(s-t+u)X_2}\right) = \mathcal{E}\left(e^{\frac{1}{2}(s+t-u)X_1}e^{\frac{1}{2}(s-t+u)X_2}\right) \\ &= \mathcal{E}\left(e^{\frac{1}{2}(s+t-u)X_1}\right)\mathcal{E}\left(e^{\frac{1}{2}(s-t+u)X_2}\right) = e^{(s+t-u)^2/8}e^{(s-t+u)^2/8} = e^{s^2/8+(t-u)^2/8} \\ &= e^{s^2/8}e^{(t-u)^2/8}. \end{aligned}$$

First setting $t = u = 0$ in the preceding result and then setting $s = 0$, we conclude that

$$M_{X,Y,Z}(s,t,u) = e^{s^2/8}e^{(t-u)^2/8} = M_X(s)M_{Y,Z}(t,u).$$

Consequently, X and (Y, Z) are independent; that is, $\bar{X}_2$ is independent of the random variables $X_1 - \bar{X}_2$ and $X_2 - \bar{X}_2$.

11.113

a) Let M bound the variance of $X_1, X_2, \ldots$. We have

$$\begin{aligned} \operatorname{Var}\left(\sum_{k=1}^n X_k\right) &= \operatorname{Cov}\left(\sum_{k=1}^n X_k, \sum_{k=1}^n X_k\right) = \sum_{k=1}^n\sum_{j=1}^n \operatorname{Cov}(X_j, X_k) \\ &= 2\sum_{k=1}^n\sum_{j=1}^k \operatorname{Cov}(X_j, X_k) - \sum_{k=1}^n \operatorname{Cov}(X_k, X_k) \\ &= 2\sum_{k=1}^n\sum_{j=1}^k \operatorname{Cov}(X_j, X_k) - \sum_{k=1}^n \operatorname{Var}(X_k)\,. \end{aligned}$$

We note that $\sum_{k=1}^n \operatorname{Var}(X_k) \le nM$ and, hence, upon division by n^2, the second term on the right of the preceding display approaches 0 as $n \to \infty$. We claim that the same is true for the first term on the right of the preceding display. Indeed, let $s_n = n^{-1}\sum_{k=1}^n \operatorname{Cov}(X_k, X_n)$, $a_n = n$, and $b_n = \sum_{k=1}^n a_k = n(n+1)/2$. By assumption, $\lim_{n\to\infty} s_n = 0$. Now,

$$\frac{1}{n^2}\sum_{k=1}^n\sum_{j=1}^k \operatorname{Cov}(X_j, X_k) = \frac{1}{n^2}\sum_{k=1}^n a_k s_k = \frac{n(n+1)}{2n^2}\cdot\frac{1}{b_n}\sum_{k=1}^n a_k s_k.$$

Applying Toeplitz's lemma, we conclude that the term on the right of the preceding display converges to $(1/2)\cdot 0 = 0$ as $n \to \infty$. We have now shown that $\lim_{n\to\infty} \operatorname{Var}\left(\sum_{k=1}^n X_k\right)/n^2 = 0$. The required result now follows from Markov's weak law of large numbers, Exercise 11.49.

b) Let M bound the variance of $X_1, X_2, \ldots$. From the hint,

$$\left|\operatorname{Cov}(X_i, X_j)\right| \le \sqrt{\operatorname{Var}(X_i)\operatorname{Var}(X_j)} \le M, \qquad i, j \in \mathcal{N}.$$

Let $\epsilon > 0$. Choose $N \in \mathcal{N}$ so that $\left|\operatorname{Cov}(X_i, X_j)\right| < \epsilon$ if $|i - j| \ge N$. Then, for $n > N$,

$$\begin{aligned} \left|\frac{1}{n}\sum_{k=1}^n \operatorname{Cov}(X_k, X_n)\right| &= \left|\frac{1}{n}\sum_{k=0}^{n-1} \operatorname{Cov}(X_{n-k}, X_n)\right| = \left|\frac{1}{n}\sum_{k=0}^{N-1} \operatorname{Cov}(X_{n-k}, X_n) + \frac{1}{n}\sum_{k=N}^{n-1} \operatorname{Cov}(X_{n-k}, X_n)\right| \\ &\le \frac{1}{n}\sum_{k=0}^{N-1}\left|\operatorname{Cov}(X_{n-k}, X_n)\right| + \frac{1}{n}\sum_{k=N}^{n-1}\left|\operatorname{Cov}(X_{n-k}, X_n)\right| \le \frac{N}{n}M + \frac{n-N}{n}\epsilon. \end{aligned}$$

Consequently,

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{k=0}^{n}\operatorname{Cov}(X_{n-k}, X_n)\right| \le \lim_{n\to\infty}\left(\frac{N}{n}M + \frac{n-N}{n}\epsilon\right) = \epsilon.$$

As $\epsilon > 0$ was arbitrary, we conclude that Equation $(*)$ holds.

11.114

a) If $k = 0$,

$$\int_{-\pi}^{\pi} e^{ikt}\,dt = \int_{-\pi}^{\pi} e^{i\cdot 0\cdot t}\,dt = \int_{-\pi}^{\pi} 1\,dt = 2\pi.$$

If $k \neq 0$,

$$\begin{aligned}\int_{-\pi}^{\pi} e^{ikt}\,dt &= \frac{1}{ik}\Big[e^{ikt}\Big]_{-\pi}^{\pi} = \frac{1}{ik}\left(e^{ik\pi} - e^{-ik\pi}\right) = \frac{1}{ik}\big(\cos(k\pi) - \cos(-k\pi)\big)\\ &= \frac{1}{ik}\big(\cos(k\pi) - \cos(k\pi)\big) = 0.\end{aligned}$$

b) For $x \in \mathcal{Z}$, we have, in view of part (a),

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ixt}\phi_X(t)\,dt &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ixt}\left(\sum_{k\in\mathcal{Z}} e^{itk}p_X(k)\right)dt\\ &= \frac{1}{2\pi}\sum_{k\in\mathcal{Z}}\left(\int_{-\pi}^{\pi} e^{i(k-x)t}\,dt\right)p_X(k)\\ &= \frac{1}{2\pi}\cdot 2\pi p_X(x) = p_X(x).\end{aligned}$$

c) From the multiplication property of CHFs for independent random variables, we know that

$$\phi_{X_1+\cdots+X_m}(t) = \phi_{X_1}(t)\cdots\phi_{X_m}(t) = \underbrace{\phi(t)\cdots\phi(t)}_{m\text{ times}} = \big(\phi(t)\big)^m.$$

Applying part (b) to the random variable $X_1 + \cdots + X_m$, we conclude that

$$p_{X_1+\cdots+X_m}(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ixt}\phi_{X_1+\cdots+X_m}(t)\,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ixt}\big(\phi(t)\big)^m\,dt, \qquad x\in\mathcal{Z}.$$

11.115

a) From Exercise 11.19, we know that, if $X \sim \mathcal{N}\left(0,\sigma^2\right)$, then $\phi_X(t) = e^{-\sigma^2t^2/2}$. Thus, for all $t \in \mathcal{R}$,

$$e^{-\sigma^2t^2/2} = \phi_X(t) = \int_{-\infty}^{\infty} e^{itx}f_X(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{itx}e^{-x^2/2\sigma^2}\,dx.$$

Hence,

$$e^{-\sigma^2t^2/2} = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{itx}e^{-x^2/2\sigma^2}\,dx, \qquad t\in\mathcal{R}.$$

Interchanging the roles of x and t yields

$$e^{-\sigma^2x^2/2} = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{ixt}e^{-t^2/2\sigma^2}\,dt, \qquad x\in\mathcal{R}.$$

Now replacing x by $-x$ and σ by $1/\sigma$, we get

$$e^{-x^2/2\sigma^2} = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixt} e^{-\sigma^2 t^2/2}\, dt, \qquad x \in \mathcal{R}.$$

Hence, for all $x \in \mathcal{R}$,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} e^{-\sigma^2 t^2/2}\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi_X(t)\, dt.$$

b) By independence,

$$\phi_{X+cZ}(t) = \mathcal{E}\Big(e^{it(X+cZ)}\Big) = \mathcal{E}\Big(e^{itX} e^{ictZ}\Big) = \mathcal{E}\Big(e^{itX}\Big)\,\mathcal{E}\Big(e^{ictZ}\Big) = \phi_X(t)\phi_Z(ct) = \phi_X(t) e^{-c^2t^2/2}.$$

Now,

$$\big|\phi_X(t)\big| = \Big|\mathcal{E}\Big(e^{itX}\Big)\Big| \le \mathcal{E}\Big(\big|e^{itX}\big|\Big) = \mathcal{E}(1) = 1.$$

Therefore,

$$\int_{-\infty}^{\infty} \big|\phi_{X+cZ}(t)\big|\, dt = \int_{-\infty}^{\infty} \big|\phi_X(t) e^{-c^2t^2/2}\big|\, dt \le \int_{-\infty}^{\infty} e^{-c^2t^2/2}\, dt = \frac{\sqrt{2\pi}}{c} < \infty.$$

Thus, $X + cZ$ is a continuous random variable and, by Equation $(**)$,

$$f_{X+cZ}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iyt} \phi_{X+cZ}(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \phi_X(t) e^{-c^2t^2/2}\, dt, \qquad y \in \mathcal{R}.$$

11.116

a) We have

$$\phi_Y(t) = \int_{-\infty}^{\infty} e^{ity} f_Y(y)\, dy = \frac{1}{2} \int_{-\infty}^{\infty} e^{ity} e^{-|y|}\, dy = \frac{1}{2} \int_{-\infty}^{\infty} (\cos ty + i \sin ty) e^{-|y|}\, dy$$
$$= \frac{1}{2} \int_{-\infty}^{\infty} e^{-|y|} \cos ty\, dy + \frac{i}{2} \int_{-\infty}^{\infty} e^{-|y|} \sin ty\, dy.$$

The second integral on the right of the preceding display equals 0 because the integrand is an odd function. Because the integrand in the first integral on the right of the preceding display is an even function,

$$\phi_Y(t) = \int_0^{\infty} e^{-y} \cos ty\, dy = \frac{1}{1+t^2} \Big[e^{-y}(-\cos ty + t \sin ty) \Big]_0^{\infty} = \frac{1}{1+t^2}.$$

b) From part (a) and Equation $(**)$, we have

$$\frac{1}{2} e^{-|y|} = f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iyt} \phi_Y(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iyt} \frac{1}{1+t^2}\, dt, \qquad y \in \mathcal{R},$$

or, equivalently,

$$e^{-|y|} = \int_{-\infty}^{\infty} e^{-iyt} \frac{1}{\pi\big(1+t^2\big)}\, dt, \qquad y \in \mathcal{R}.$$

Replacing y by t and t by $-x$, we get

$$e^{-|t|} = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi\big(1+x^2\big)}\, dx = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx = \phi_X(t), \qquad t \in \mathcal{R}.$$

BETA

Chapter 12

Applications of Probability Theory

12.1 The Poisson Process

Basic Exercises

12.1 Let $X_j = 1$ if the jth Bernoulli trial results in success and let $X_j = 0$ otherwise. Then $X_1, X_2, \ldots$ are independent Bernoulli random variables with common mean p. On the one hand, as time is quantized into units of duration $\frac{1}{n}$, the number of successes that occur per unit of continuous time is $X_1 + \cdots + X_n$, which is to equal λ. On the other hand, by the strong law of large numbers, $X_1 + \cdots + X_n \approx np$ for large n. Hence, $np \approx \lambda$, or $p \approx \lambda/n$.

12.2

a) As one Bernoulli trial is performed at the end of each $1/n$ time-unit period (i.e., at times $1/n, 2/n, \ldots$), the number performed by time t equals the greatest integer in $t/(1/n)$. In other words, for $t > 0$, the quantity $\lfloor nt \rfloor$ represents the number of Bernoulli trials performed by time t.

b) Recall that $x - 1 < \lfloor x \rfloor \leq x$ for all $x \in \mathcal{R}$. Hence,

$$k - 1 = (\lfloor ns \rfloor + 1) - 1 = \lfloor ns \rfloor \leq ns < \lfloor ns \rfloor + 1 = (k-1) + 1 = k,$$

and, hence,

$$\frac{k-1}{n} \leq s < \frac{k}{n}.$$

Furthermore,

$$\begin{aligned} k + m - 1 &= (\lfloor ns \rfloor + 1) + (\lfloor nt \rfloor - \lfloor ns \rfloor) - 1 = \lfloor nt \rfloor \\ &\leq nt \\ &< \lfloor nt \rfloor + 1 = (\lfloor ns \rfloor + 1) + (\lfloor nt \rfloor - \lfloor ns \rfloor) = k + m, \end{aligned}$$

and, hence,

$$\frac{k+m-1}{n} \leq t < \frac{k+m}{n}.$$

c) For each $x \in \mathcal{R}$,

$$x - \frac{1}{n} = \frac{nx-1}{n} < \frac{\lfloor nx \rfloor}{n} \leq \frac{nx}{n} = x,$$

and, hence, $\lim_{n\to\infty} \lfloor nx \rfloor / n = x$. Consequently,

$$\lim_{n\to\infty} \frac{m}{n} = \lim_{n\to\infty} \frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} = \lim_{n\to\infty} \frac{\lfloor nt \rfloor}{n} - \lim_{n\to\infty} \frac{\lfloor ns \rfloor}{n} = t - s.$$

12.3

a) The number of patients that arrive between 4:00 A.M. and 5:00 A.M. is $N(5) - N(4)$, which has a Poisson distribution with parameter $6.9 \cdot (5 - 4) = 6.9$. Hence,

$$P\big(N(5) - N(4) = 7\big) = e^{-6.9}\frac{(6.9)^7}{7!} = 0.149.$$

b) The number of patients that arrive between 6:00 A.M. and 8:00 A.M. is $N(8) - N(6)$, which has a Poisson distribution with parameter $6.9 \cdot (8 - 6) = 13.8$. Hence,

$$P\big(N(8) - N(6) = 14\big) = e^{-13.8}\frac{(13.8)^{14}}{14!} = 0.106.$$

c) The number of patients that arrive between 4:00 A.M. and 5:00 A.M. and the number of patients that arrive between 6:00 A.M. and 8:00 A.M. are independent random variables because the two time intervals are disjoint. Referring now to parts (a) and (b), we get

$$\begin{aligned} P\big(N(5) - N(4) = 7,\ N(8) - N(6) = 14\big) &= P\big(N(5) - N(4) = 7\big)\, P\big(N(8) - N(6) = 14\big) \\ &= 0.149 \cdot 0.106 = 0.0158. \end{aligned}$$

d) From the general addition rule and parts (a)–(c),

$$\begin{aligned} &P\big(N(5) - N(4) = 7 \quad \text{or} \quad N(8) - N(6) = 14\big) \\ &\quad = P\big(N(5) - N(4) = 7\big) + P\big(N(8) - N(6) = 14\big) \\ &\qquad - P\big(N(5) - N(4) = 7,\ N(8) - N(6) = 14\big) \\ &\quad = 0.149 + 0.106 - 0.0158 = 0.239. \end{aligned}$$

e) The number of patients that arrive between 4:00 A.M. and 7:00 A.M. is $N(7) - N(4)$, which has a Poisson distribution with parameter $6.9 \cdot (7 - 4) = 20.7$. Hence,

$$P\big(N(7) - N(4) = 20\big) = e^{-20.7}\frac{(20.7)^{20}}{20!} = 0.0878.$$

f) Let $A = \{N(7) - N(4) = 20\}$, $B = \{N(8) - N(6) = 14\}$, and $C_k = \{N(7) - N(6) = k\}$. The problem is to determine $P(A \cap B)$. For convenience, let us set $X = N(6) - N(4)$, $Y = N(7) - N(6)$, and $Z = N(8) - N(7)$. We know that $X \sim \mathcal{P}(13.8)$, $Y \sim \mathcal{P}(6.9)$, and $Z \sim \mathcal{P}(6.9)$. Applying the law of partitions and using the independent-increments property of a Poisson process, we get

$$\begin{aligned} P(A \cap B) &= \sum_{k=0}^{\infty} P(A \cap B \cap C_k) = \sum_{k=0}^{14} P(X = 20 - k,\ Y = k,\ Z = 14 - k) \\ &= \sum_{k=0}^{14} P(X = 20 - k)P(Y = k)P(Z = 14 - k) \\ &= \sum_{k=0}^{14} e^{-13.8}\,\frac{(13.8)^{20-k}}{(20 - k)!} \cdot e^{-6.9}\,\frac{(6.9)^k}{k!} \cdot e^{-6.9}\,\frac{(6.9)^{14-k}}{(14 - k)!} \\ &= (6.9)^{14} e^{-27.6} \sum_{k=0}^{14} \frac{(13.8)^{20-k}}{(20 - k)!\,k!\,(14 - k)!}. \end{aligned}$$

g) Using a scientific calculator, we found that the probability in part (f) equals 0.0101.

h) Applying the general addition rule and referring to parts (e), (b), and (g), we get

$$\begin{aligned}
&P\big(N(7)-N(4)=20 \quad\text{or}\quad N(8)-N(6)=14\big)\\
&\quad = P\big(N(7)-N(4)=20\big)+P\big(N(8)-N(6)=14\big)\\
&\qquad - P\big(N(7)-N(4)=20,\ N(8)-N(6)=14\big)\\
&\quad = 0.0878+0.106-0.0101=0.184.
\end{aligned}$$

12.4
a) The number of patients that arrive either between 4:00 A.M. and 5:00 A.M. or between 6:00 A.M. and 8:00 A.M. is the sum of $N(5)-N(4)$ and $N(8)-N(6)$. From properties of a Poisson process, these two quantities are independent Poisson random variables with parameters 6.9 and 13.8, respectively. Therefore, from Proposition 6.20(b) on page 311, the sum of those two random variables has the Poisson distribution with parameter $6.9+13.8=20.7$.
b) The number of patients that arrive either between 4:00 A.M. and 7:00 A.M. or between 6:00 A.M. and 8:00 A.M. is just the number of patients that arrive between 4:00 A.M. and 8:00 A.M., which has the Poisson distribution with parameter $6.9\cdot 4=27.6$.

12.5 Let $N(t)$ denote the number of queries to the database by time t, where time is measured in minutes. Then $\{N(t):t\geq 0\}$ is a Poisson process with rate 4.
a) The number of queries that occur between times s and t is $N(t)-N(s)$, which has a Poisson distribution with parameter $4(t-s)$.
b) On average, 4 queries occur per minute, so that, on average, 8 queries occur in a 2-minute interval.
c) The number of queries that occur in a 2-minute interval has a Poisson distribution with parameter $4\cdot 2=8$. Hence, the probability that between three and five queries, inclusive, occur in a 2-minute interval is

$$\sum_{k=3}^{5} e^{-8}\frac{8^k}{k!}=e^{-8}\left(\frac{8^3}{3!}+\frac{8^4}{4!}+\frac{8^5}{5!}\right)=0.177.$$

d) From Proposition 12.2 on page 691, the elapsed time between successive queries has an exponential distribution with parameter 4.
e) Referring to part (d) and recalling that the mean of an exponential random variable is the reciprocal of its parameter, we see that, on average it takes $1/4$ minute, or 15 seconds, between successive queries.
f) From Proposition 12.1 on page 689, the time elapsed until the nth query has an Erlang distribution with parameters n and 4 or, equivalently, a $\Gamma(n,4)$ distribution.
g) Referring to part (f) and recalling that the mean of a $\Gamma(\alpha,\lambda)$ random variable is α/λ, we see that, on average it takes $n/4$ minutes for the nth query to occur.
h) The elapsed time between the third and fifth queries is W_5-W_3, which, from part (g), has mean equal to $5/4-3/4=1/2$. Therefore, on average, it takes 0.5 minute, or 30 seconds, between the third and fifth queries.
i) The number of queries that occur in the first minute is $N(1)$, which has a Poisson distribution with parameter $4\cdot 1=4$, and the number of queries that occur in the next two minutes is $N(3)-N(1)$, which has a Poisson distribution with parameter $4\cdot(3-1)=8$. Moreover, from properties of a Poisson process, these two Poisson random variables are independent. Hence,

$$\begin{aligned}
P\big(N(1)=5,\ N(3)-N(1)=7\big)&=P\big(N(1)=5\big)\,P\big(N(3)-N(1)=7\big)\\
&=e^{-4}\frac{4^5}{5!}\cdot e^{-8}\frac{8^7}{7!}=0.0218.
\end{aligned}$$

12.6 For convenience, set $q = 1 - p$.
a) Let m and n be nonnegative integers. Note that, if $N_1(t) = m$ and $N_2(t) = n$, then $N(t) = m + n$, because $N(t) = N_1(t) + N_2(t)$. Hence, by the general multiplication rule,

$$\begin{aligned} P\left(N_1(t) = m,\ N_2(t) = n\right) &= P\left(N_1(t) = m,\ N_2(t) = n,\ N(t) = m + n\right) \\ &= P\left(N_1(t) = m,\ N(t) = m + n\right) \\ &= P\left(N(t) = m + n\right) P\left(N_1(t) = m \mid N(t) = m + n\right) \\ &= e^{-\lambda t}\,\frac{(\lambda t)^{m+n}}{(m+n)!} P\left(N_1(t) = m \mid N(t) = m + n\right). \end{aligned}$$

Now, given that $N(t) = m + n$, the number of Type 1 customers that arrive by time t has the $\mathcal{B}(m + n,\, p)$ distribution. Therefore,

$$P\left(N_1(t) = m \mid N(t) = m + n\right) = \binom{m+n}{m} p^m (1-p)^{(m+n)-m} = \binom{m+n}{m} p^m q^n.$$

Consequently,

$$\begin{aligned} P\left(N_1(t) = m,\ N_2(t) = n\right) &= e^{-\lambda t}\,\frac{(\lambda t)^{m+n}}{(m+n)!}\binom{m+n}{m} p^m q^n = e^{-\lambda(p+q)t}\,\frac{(\lambda t)^{m+n}}{(m+n)!}\,\frac{(m+n)!}{m!\,n!} p^m q^n \\ &= e^{-p\lambda t} e^{-q\lambda t}\,\frac{(\lambda t)^m (\lambda t)^n p^m q^n}{m!\,n!} = \left(e^{-p\lambda t}\,\frac{(p\lambda t)^m}{m!}\right)\left(e^{-q\lambda t}\,\frac{(q\lambda t)^n}{n!}\right). \end{aligned}$$

It now follows from Exercise 6.60 that $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with parameters $p\lambda t$ and $q\lambda t$, respectively.
b) We first show that $\{N_1(t) : t \geq 0\}$ is a Poisson process. Because $0 = N(0) = N_1(0) + N_2(0)$ and $N_1(0)$ and $N_2(0)$ are nonnegative-integer valued, it follows that $N_1(0) = 0$. Now, suppose that $r \in \mathcal{N}$ and let $0 \leq t_1 < t_2 < \cdots < t_r$. The random variables $N_1(t_j) - N_1(t_{j-1})$, $2 \leq j \leq r$, give the number of Type 1 customers that arrive in the time intervals $(t_{j-1}, t_j]$, respectively. Relative to the Poisson process $\{N(t) : t \geq 0\}$, they depend only on $N(t_j) - N(t_{j-1})$, respectively. From the independent-increments property of a Poisson process, these latter random variables are independent and, hence, so are the random variables $N_1(t_j) - N_1(t_{j-1})$. Hence, the counting process $\{N_1(t) : t \geq 0\}$ has independent increments. Next, let $0 \leq s < t$. From the law of total probability, for each nonnegative integer k,

$$\begin{aligned} P\left(N_1(t) - N_1(s) = k\right) &= \sum_{n=k}^{\infty} P\left(N(t) - N(s) = n\right) P\left(N_1(t) - N_1(s) = k \mid N(t) - N(s) = n\right) \\ &= \sum_{n=k}^{\infty} e^{-\lambda(t-s)}\,\frac{\left(\lambda(t-s)\right)^n}{n!} \cdot \binom{n}{k} p^k q^{n-k} \\ &= e^{-p\lambda(t-s)}\,\frac{\left(p\lambda(t-s)\right)^k}{k!} \sum_{n=k}^{\infty} e^{-q\lambda(t-s)}\,\frac{\left(q\lambda(t-s)\right)^{n-k}}{(n-k)!} \\ &= e^{-p\lambda(t-s)}\,\frac{\left(p\lambda(t-s)\right)^k}{k!} \cdot 1 = e^{-p\lambda(t-s)}\,\frac{\left(p\lambda(t-s)\right)^k}{k!}. \end{aligned}$$

Hence, $N_1(t) - N_1(s) \sim \mathcal{P}\left(p\lambda(t-s)\right)$. Consequently, we have shown that $\{N_1(t) : t \geq 0\}$ is a Poisson process with rate $p\lambda$. Similarly, $\{N_2(t) : t \geq 0\}$ is a Poisson process with rate $q\lambda$. The fact that these two Poisson processes are independent follows from part (a).

12.7

a) Let $n_1, \ldots, n_m$ be any m nonnegative integers. Note that $N_j(t) = n_j$ for $1 \le j \le m$ implies that $N(t) = n_1 + \cdots + n_m$, because $N(t) = N_1(t) + \cdots + N_m(t)$. Hence, by the general multiplication rule,

$$
\begin{aligned}
&P\left(N_1(t) = n_1, \ldots, N_m(t) = n_m\right) \\
&\quad = P\left(N_1(t) = n_1, \ldots, N_m(t) = n_m,\ N(t) = n_1 + \cdots + n_m\right) \\
&\quad = P\left(N(t) = n_1 + \cdots + n_m\right) P\left(N_1(t) = n_1, \ldots, N_m(t) = n_m \mid N(t) = n_1 + \cdots + n_m\right) \\
&\quad = e^{-\lambda t} \frac{(\lambda t)^{n_1 + \cdots + n_m}}{(n_1 + \cdots + n_m)!} P\left(N_1(t) = n_1, \ldots, N_m(t) = n_m \mid N(t) = n_1 + \cdots + n_m\right).
\end{aligned}
$$

Given that $N(t) = n_1 + \cdots + n_m$, the numbers of times that specified events with attributes $a_1, \ldots, a_m$ occur by time t has the multinomial distribution with parameters $n_1 + \cdots + n_m$ and $p_1, \ldots, p_m$. Therefore, we have

$$
P\left(N_1(t) = n_1, \ldots, N_m(t) = n_m \mid N(t) = n_1 + \cdots + n_m\right) = \binom{n_1 + \cdots + n_m}{n_1, \ldots, n_m} p_1^{n_1} \cdots p_m^{n_m}.
$$

Consequently,

$$
\begin{aligned}
P\left(N_1(t) = n_1, \ldots, N_m(t) = n_m\right) &= e^{-\lambda t} \frac{(\lambda t)^{n_1 + \cdots + n_m}}{(n_1 + \cdots + n_m)!} \binom{n_1 + \cdots + n_m}{n_1, \ldots, n_m} p_1^{n_1} \cdots p_m^{n_m} \\
&= e^{-\lambda (p_1 + \cdots + p_m) t} \frac{(\lambda t)^{n_1 + \cdots + n_m}}{(n_1 + \cdots + n_m)!} \frac{(n_1 + \cdots + n_m)!}{n_1! \cdots n_m!} p_1^{n_1} \cdots p_m^{n_m} \\
&= \frac{e^{-p_1 \lambda t} \cdots e^{-p_m \lambda t} (\lambda t)^{n_1} \cdots (\lambda t)^{n_m} p_1^{n_1} \cdots p_m^{n_m}}{n_1! \cdots n_m!} \\
&= \left(e^{-p_1 \lambda t} \frac{(p_1 \lambda t)^{n_1}}{n_1!}\right) \cdots \left(e^{-p_m \lambda t} \frac{(p_m \lambda t)^{n_m}}{n_m!}\right).
\end{aligned}
$$

It now follows from Exercise 6.60 that $N_1(t), \ldots, N_m(t)$ are independent Poisson random variables with parameters $p_1 \lambda t, \ldots, p_m \lambda t$, respectively.

b) We first show that $\{N_1(t) : t \ge 0\}$ is a Poisson process. Because $0 = N(0) = N_1(0) + N_2(0)$ and $N_1(0)$ and $N_2(0)$ are nonnegative-integer valued, it follows that $N_1(0) = 0$. Now, suppose that $r \in \mathcal{N}$ and let $0 \le t_1 < t_2 < \cdots < t_r$. The random variables $N_1(t_j) - N_1(t_{j-1})$, $2 \le j \le r$, give the number of times that a specified event with attribute a_1 occurs in the time intervals $(t_{j-1}, t_j]$, respectively. Relative to the Poisson process $\{N(t) : t \ge 0\}$, they depend only on $N(t_j) - N(t_{j-1})$, respectively. From the independent-increments property of a Poisson process, these latter random variables are independent and, hence, so are the random variables $N_1(t_j) - N_1(t_{j-1})$. Hence, the counting process $\{N_1(t) : t \ge 0\}$ has independent increments. Next, let $0 \le s < t$. From the law of total probability, for each nonnegative integer k,

$$
\begin{aligned}
P\left(N_1(t) - N_1(s) = k\right) &= \sum_{n=k}^{\infty} P\left(N(t) - N(s) = n\right) P\left(N_1(t) - N_1(s) = k \mid N(t) - N(s) = n\right) \\
&= \sum_{n=k}^{\infty} e^{-\lambda (t-s)} \frac{\left(\lambda (t-s)\right)^n}{n!} \cdot P\left(N_1(t) - N_1(s) = k \mid N(t) - N(s) = n\right).
\end{aligned}
$$

However, given that $N(t) - N(s) = n$, the number of times that a specified event with attribute a_1 occurs in the time interval $(s, t]$ has the binomial distribution with parameters n and p_1. Setting $q_1 = 1 - p_1$, we now see that

$$\begin{aligned}
P\big(N_1(t) - N_1(s) = k\big) &= \sum_{n=k}^{\infty} e^{-\lambda(t-s)} \frac{\big(\lambda(t-s)\big)^n}{n!} \cdot \binom{n}{k} p_1^k q_1^{n-k} \\
&= \sum_{n=k}^{\infty} e^{-(p_1+q_1)\lambda(t-s)} \frac{\big(\lambda(t-s)\big)^k \big(\lambda(t-s)\big)^{n-k}}{n!} \frac{n!}{k!\,(n-k)!} p_1^k q_1^{n-k} \\
&= e^{-p_1\lambda(t-s)} \frac{\big(p_1\lambda(t-s)\big)^k}{k!} \sum_{n=k}^{\infty} e^{-q_1\lambda(t-s)} \frac{\big(q_1\lambda(t-s)\big)^{n-k}}{(n-k)!} \\
&= e^{-p_1\lambda(t-s)} \frac{\big(p_1\lambda(t-s)\big)^k}{k!} \sum_{j=0}^{\infty} e^{-q_1\lambda(t-s)} \frac{\big(q_1\lambda(t-s)\big)^{j}}{j!} \\
&= e^{-p_1\lambda(t-s)} \frac{\big(p_1\lambda(t-s)\big)^k}{k!} \cdot 1 = e^{-p_1\lambda(t-s)} \frac{\big(p_1\lambda(t-s)\big)^k}{k!}.
\end{aligned}$$

Hence, $N_1(t) - N_1(s) \sim \mathcal{P}\big(p_1\lambda(t-s)\big)$. Consequently, we have shown that $\{N_1(t) : t \ge 0\}$ is a Poisson process with rate $p_1\lambda$. Similarly, $\{N_2(t) : t \ge 0\}, \dots, \{N_m(t) : t \ge 0\}$ are Poisson processes with rates $p_2\lambda, \dots, p_m\lambda$, respectively. The fact that these m Poisson processes are independent follows from part (a).

12.8 Let $0 = t_0 < t_1 < \cdots < t_n$, and let $\Delta t_1, \dots, \Delta t_n$ represent small positive numbers. For convenience, set $E = \{t_1 \le W_1 \le t_1 + \Delta t_1, \dots, t_n \le W_n \le t_n + \Delta t_n\}$. Event E occurs if and only if no successes occur in the time interval from 0 to t_1, one success occurs in the time interval from t_1 to $t_1 + \Delta t_1$, no successes occur in the time interval from $t_1 + \Delta t_1$ to t_2, one success occurs in the time interval from t_2 to $t_2 + \Delta t_2, \dots,$ no successes occur in the time interval from $t_{n-1} + \Delta t_{n-1}$ to t_n, and at least one success occurs in the time interval from t_n to $t_n + \Delta t_n$. See Fig. 12.2 on page 690.

For $1 \le j \le n-1$, let A_j denote the event that (exactly) one success occurs in the time interval from t_j to $t_j + \Delta t_j$ and, for $1 \le j \le n$, let B_j denote the event that no successes occur in the time interval from $t_{j-1} + \Delta t_{j-1}$ to t_j, where we define $\Delta t_0 = 0$. Also, let C_n denote the event that at least one success occurs in the time interval from t_n to $t_n + \Delta t_n$. We note that, for $1 \le j \le n-1$,

$$P(A_j) = P\big(N(t_j + \Delta t_j) - N(t_j) = 1\big) = e^{-\lambda \Delta t_j} \frac{(\lambda \Delta t_j)^1}{1!} = e^{-\lambda \Delta t_j} \lambda \Delta t_j$$

and, for $1 \le j \le n$,

$$\begin{aligned}
P(B_j) &= P\big(N(t_j) - N(t_{j-1} + \Delta t_{j-1}) = 0\big) \\
&= e^{-\lambda(t_j - t_{j-1} - \Delta t_{j-1})} \frac{\big(\lambda(t_j - t_{j-1} - \Delta t_{j-1})\big)^0}{0!} = e^{-\lambda(t_j - t_{j-1} - \Delta t_{j-1})}.
\end{aligned}$$

Also,

$$\begin{aligned}
P(C_n) &= P\big(N(t_n + \Delta t_n) - N(t_n) \ge 1\big) = 1 - P\big(N(t_n + \Delta t_n) - N(t_n) = 0\big) \\
&= 1 - e^{-\lambda \Delta t_n} \frac{(\lambda \Delta t_n)^0}{0!} = 1 - e^{-\lambda \Delta t_n}.
\end{aligned}$$

Hence, by the independent-increments property of a Poisson process and Relation (5.18) on page 219,

$$
\begin{aligned}
P(E) &= P\left(\bigcap_{j=1}^{n-1} A_j \cap \bigcap_{j=1}^{n} B_j \cap C_n\right) = \prod_{j=1}^{n-1} P(A_j) \prod_{j=1}^{n} P(B_j) \cdot P(C_j) \\
&= \prod_{j=1}^{n-1} \left(e^{-\lambda\Delta t_j}\lambda\Delta t_j\right) \prod_{j=1}^{n} e^{-\lambda(t_j - t_{j-1} - \Delta t_{j-1})} \cdot \left(1 - e^{-\lambda\Delta t_n}\right) \\
&= \prod_{j=1}^{n-1} \left(e^{-\lambda\Delta t_j}\lambda\Delta t_j\right) \prod_{j=0}^{n-1} e^{-\lambda(t_{j+1} - t_j - \Delta t_j)} \cdot \left(1 - e^{-\lambda\Delta t_n}\right) \\
&= e^{-\lambda t_1}\left(\prod_{j=1}^{n-1} e^{-\lambda(t_{j+1} - t_j)}\right)\lambda^{n-1}\Delta t_1 \cdots \Delta t_{n-1}\left(1 - e^{-\lambda\Delta t_n}\right) \\
&= e^{-\lambda t_1} e^{-\lambda \sum_{j=1}^{n-1}(t_{j+1} - t_j)}\lambda^{n-1}\Delta t_1 \cdots \Delta t_{n-1}\left(1 - e^{-\lambda\Delta t_n}\right) \\
&= \lambda^{n-1} e^{-\lambda t_n}\Delta t_1 \cdots \Delta t_{n-1}\left(1 - e^{-\lambda\Delta t_n}\right) \\
&\approx \lambda^{n-1} e^{-\lambda t_n}\Delta t_1 \cdots \Delta t_{n-1} \cdot \lambda\Delta t_n \\
&= \lambda^{n} e^{-\lambda t_n}\Delta t_1 \cdots \Delta t_n.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
f_{W_1,\ldots,W_n}(t_1, \ldots, t_n)\Delta t_1 \cdots \Delta t_n &\approx P(t_1 \le W_1 \le t_1 + \Delta t_1, \ldots,\ t_n \le W_n \le t_n + \Delta t_n) \\
&= P(E) \\
&\approx \lambda^n e^{-\lambda t_n}\Delta t_1 \cdots \Delta t_n,
\end{aligned}
$$

and, therefore, Equation (12.9) holds.

12.9

a) First we use a basic random number generator to obtain n uniform numbers between 0 and 1, say, $u_1, \ldots, u_n$. Next we let $i_j = -\lambda^{-1}\ln u_j$ for $1 \le j \le n$. In view of Example 8.22(b), the numbers $i_1, \ldots, i_n$ represent n independent observations of an exponential random variable with parameter λ. Because of Proposition 12.2 on page 691, these n numbers can serve as a realization of the first n interarrival times of a Poisson process with rate λ. Now we set $w_k = i_1 + \cdots + i_k$ for $1 \le k \le n$. Then $w_1, \ldots, w_n$ represent a realization of the waiting times for the first n events. In this way, we obtain a simulation of a Poisson process with rate λ up to and including the occurrence of the nth event.

b) Answers will vary. The method, however, is to apply part (a) 10 times, each time with $n = 5$ and $\lambda = 6.9$. In each case, the graph of $N(t)$ against t is obtained from the following formula:

$$
N(t) = \begin{cases} 0, & \text{if } 0 \le t < w_1; \\ 1, & \text{if } w_1 \le t < w_2; \\ 2, & \text{if } w_2 \le t < w_3; \\ 3, & \text{if } w_3 \le t < w_4; \\ 4, & \text{if } w_4 \le t < w_5; \\ 5, & \text{if } t = w_5. \end{cases}
$$

12.10 From the conditional probability rule and properties of a Poisson process, we get, for $0 \le k \le n$,

$$
\begin{aligned}
P\,(N(s) = k \mid N(t) = n) &= \frac{P\,(N(s) = k,\ N(t) = n)}{P\,(N(t) = n)} = \frac{P\,(N(s) = k,\ N(t) - N(s) = n - k)}{P\,(N(t) = n)} \\
&= \frac{P\,(N(s) = k)\, P\,(N(t) - N(s) = n - k)}{P\,(N(t) = n)} \\
&= \frac{e^{-\lambda s}\,\dfrac{(\lambda s)^k}{k!} \cdot e^{-\lambda(t-s)}\,\dfrac{(\lambda(t-s))^{n-k}}{(n-k)!}}{e^{-\lambda t}\,\dfrac{(\lambda t)^n}{n!}} = \binom{n}{k} \frac{\lambda^k s^k \cdot \lambda^{n-k}(t-s)^{n-k}}{\lambda^n t^n} \\
&= \binom{n}{k} \left(\frac{s}{t}\right)^k \left(\frac{t-s}{t}\right)^{n-k} = \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}.
\end{aligned}
$$

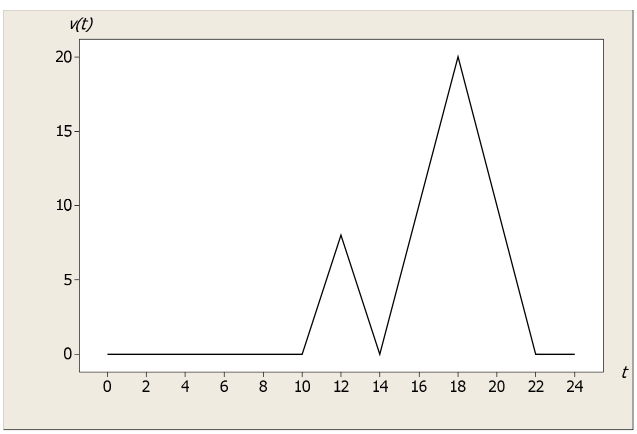
Hence, given that $N(t) = n$, we see that $N(s) \sim \mathcal{B}(n, s/t)$.

12.11

a) The intensity function, $\nu(t)$, is given by

$$
\nu(t) = \begin{cases} 0, & \text{if } 0 \le t < 10; \\ 4t - 40, & \text{if } 10 \le t < 12; \\ -4t + 56, & \text{if } 12 \le t < 14; \\ 5t - 70, & \text{if } 14 \le t < 18; \\ -5t + 110, & \text{if } 18 \le t < 22; \\ 0, & \text{if } 22 \le t < 24. \end{cases}
$$

In the following graph, we use $v(t)$ instead of $\nu(t)$:



b) The mean function is $\mu(t) = \int_0^t \nu(s)\, ds$. We note that μ is continuous, being an indefinite integral. We refer to part (a) and consider six cases.

Case 1: $0 \le t < 10$. In this case,

$$
\mu(t) = \int_0^t \nu(s)\, ds = \int_0^t 0\, ds = 0.
$$

Case 2: $10 \le t < 12$. In this case,

$$\mu(t) = \int_0^t \nu(s)\,ds = \int_0^{10} \nu(s)\,ds + \int_{10}^t \nu(s)\,ds = \mu(10-) + \int_{10}^t (4s - 40)\,ds = 2t^2 - 40t + 200.$$

Case 3: $12 \le t < 14$. In this case,

$$\mu(t) = \int_0^t \nu(s)\,ds = \int_0^{12} \nu(s)\,ds + \int_{12}^t \nu(s)\,ds = \mu(12-) + \int_{12}^t (-4s + 56)\,ds = -2t^2 + 56t - 376.$$

Case 4: $14 \le t < 18$. In this case,

$$\mu(t) = \int_0^t \nu(s)\,ds = \int_0^{14} \nu(s)\,ds + \int_{14}^t \nu(s)\,ds = \mu(14-) + \int_{14}^t (5s - 70)\,ds = \frac{5}{2}t^2 - 70t + 506.$$

Case 5: $18 \le t < 22$. In this case,

$$\mu(t) = \int_0^t \nu(s)\,ds = \int_0^{18} \nu(s)\,ds + \int_{18}^t \nu(s)\,ds = \mu(18-) + \int_{18}^t (-5s + 110)\,ds = -\frac{5}{2}t^2 + 110t - 1114.$$

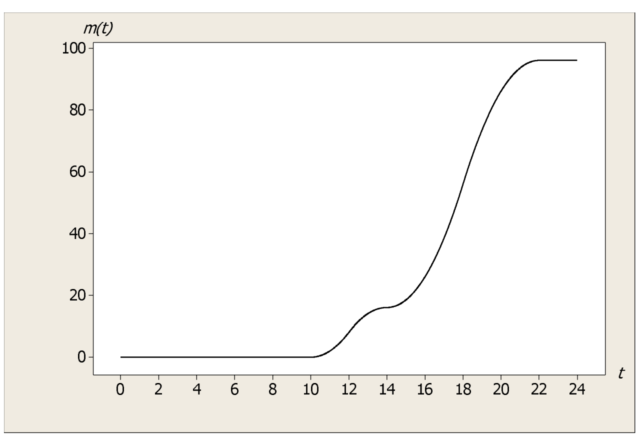
Case 6: $22 \le t < 24$. In this case,

$$\mu(t) = \int_0^t \nu(s)\,ds = \int_0^{22} \nu(s)\,ds + \int_{22}^t \nu(s)\,ds = \mu(22-) + \int_{22}^t 0\,ds = 96.$$

Hence, we have

$$\mu(t) = \begin{cases} 0, & 0 \le t < 10; \\ 2t^2 - 40t + 200, & 10 \le t < 12; \\ -2t^2 + 56t - 376, & 12 \le t < 14; \\ \frac{5}{2}t^2 - 70t + 506, & 14 \le t < 18; \\ -\frac{5}{2}t^2 + 110t - 1114, & 18 \le t < 22; \\ 96, & 22 \le t < 24. \end{cases}$$

In the following graph we use $m(t)$ instead of $\mu(t)$:



c) The number of customers that arrive at the restaurant per day is $N(24)$, which we know has a Poisson distribution with parameter $\mu(24)$. From part (b),

$$\mathcal{E}\big(N(24)\big) = \mu(24) = 96.$$

d) The number of customers that arrive between 11:00 A.M. and 7:00 P.M. is $N(19) - N(11)$, which we know has a Poisson distribution with parameter $\mu(19) - \mu(11)$. Referring to part (b), we get

$$\begin{aligned}\mathcal{E}\big(N(19) - N(11)\big) &= \mu(19) - \mu(11) \\ &= \Big(-(5/2)\cdot 19^2 + 110\cdot 19 - 1114\Big) - \Big(2\cdot 11^2 - 40\cdot 11 + 200\Big) \\ &= 71.5\end{aligned}$$

and

$$\sigma_{N(19)-N(11)} = \sqrt{\operatorname{Var}\big(N(19) - N(11)\big)} = \sqrt{\mu(19) - \mu(11)} = \sqrt{71.5} = 8.46.$$

e) The number of customers that arrive between 11:00 A.M. and 1:00 P.M. is $N(13) - N(11)$, which we know has a Poisson distribution with parameter $\mu(13) - \mu(11)$. From part (b),

$$\mu(13) - \mu(11) = \Big(-2\cdot 13^2 + 56\cdot 13 - 376\Big) - \Big(2\cdot 11^2 - 40\cdot 11 + 200\Big) = 12.$$

Hence,

$$P\Big(13 \le N(13) - N(11) \le 15\Big) = \sum_{k=13}^{15} e^{-12}\frac{12^k}{k!} = e^{-12}\left(\frac{12^{13}}{(13)!} + \frac{12^{14}}{(14)!} + \frac{12^{15}}{(15)!}\right) = 0.268.$$

12.12 Let $X_k = 1$ (the constant random variable equal to 1) for all $k \in \mathcal{N}$. Then, for $t \ge 0$,

$$Y(t) = \begin{cases} 0, & \text{if } N(t) = 0; \\ \displaystyle\sum_{k=1}^{N(t)} X_k, & \text{if } N(t) \ge 1. \end{cases} = \begin{cases} 0, & \text{if } N(t) = 0; \\ \displaystyle\sum_{k=1}^{N(t)} 1, & \text{if } N(t) \ge 1. \end{cases} = \begin{cases} N(t), & \text{if } N(t) = 0; \\ N(t), & \text{if } N(t) \ge 1. \end{cases} = N(t).$$

Therefore, a compound Poisson process does indeed generalize the concept of a (homogeneous) Poisson process.

12.13

a) Let λ denote the rate of the homogeneous Poisson process $\{N(t) : t \ge 0\}$. Now,

$$\mathcal{E}\big(Y(t)\,|\,N(t) = n\big) = \begin{cases} \mathcal{E}(0), & \text{if } n = 0; \\ \mathcal{E}\Big(\displaystyle\sum_{k=1}^{n} X_k\Big), & \text{if } n \ge 1. \end{cases} = \begin{cases} 0, & \text{if } n = 0; \\ n\mu, & \text{if } n \ge 1. \end{cases} = n\mu$$

and

$$\operatorname{Var}\big(Y(t)\,|\,N(t) = n\big) = \begin{cases} \operatorname{Var}(0), & \text{if } n = 0; \\ \operatorname{Var}\Big(\displaystyle\sum_{k=1}^{n} X_k\Big), & \text{if } n \ge 1. \end{cases} = \begin{cases} 0, & \text{if } n = 0; \\ n\sigma^2, & \text{if } n \ge 1. \end{cases} = n\sigma^2.$$

Hence, $\mathcal{E}\big(Y(t)\,|\,N(t)\big) = N(t)\mu$ and $\operatorname{Var}\big(Y(t)\,|\,N(t)\big) = N(t)\sigma^2$. Therefore, from the laws of total expectation and total variance,

$$\mu(t) = \mathcal{E}\big(Y(t)\big) = \mathcal{E}\Big(\mathcal{E}\big(Y(t)\,|\,N(t)\big)\Big) = \mathcal{E}\big(N(t)\mu\big) = \mu\mathcal{E}\big(N(t)\big) = \mu\lambda t$$

and

$$\begin{aligned}\sigma^2(t) = \operatorname{Var}\big(Y(t)\big) &= \mathcal{E}\Big(\operatorname{Var}\big(Y(t)\,|\,N(t)\big)\Big) + \operatorname{Var}\Big(\mathcal{E}\big(Y(t)\,|\,N(t)\big)\Big) = \mathcal{E}\Big(N(t)\sigma^2\Big) + \operatorname{Var}\big(N(t)\mu\big) \\ &= \sigma^2\mathcal{E}\big(N(t)\big) + \mu^2\operatorname{Var}\big(N(t)\big) = \sigma^2\lambda t + \mu^2\lambda t = \Big(\mu^2 + \sigma^2\Big)\lambda t.\end{aligned}$$

b) From the law of total expectation, Equation (10.49) on page 602, and part (a),

$$\begin{aligned}\mathcal{E}\big(N(t)Y(t)\big) &= \mathcal{E}\Big(\mathcal{E}\big(N(t)Y(t) \mid N(t)\big)\Big) = \mathcal{E}\Big(N(t)\mathcal{E}\big(Y(t) \mid N(t)\big)\Big) \\ &= \mathcal{E}\big(N(t)\mu N(t)\big) = \mu\mathcal{E}\Big(N(t)^2\Big) = \Big(\lambda t + (\lambda t)^2\Big)\,\mu.\end{aligned}$$

Referring again to part (a), we get

$$\begin{aligned}\rho\big(N(t), Y(t)\big) = \operatorname{Cov}\big(N(t), Y(t)\big) &= \frac{\mathcal{E}\big(N(t)Y(t)\big) - \mathcal{E}\big(N(t)\big)\,\mathcal{E}\big(Y(t)\big)}{\sqrt{\operatorname{Var}\big(N(t)\big)\operatorname{Var}\big(Y(t)\big)}} \\ &= \frac{\big(\lambda t + (\lambda t)^2\big)\,\mu - \lambda t \cdot \mu\lambda t}{\sqrt{\lambda t \cdot \big(\mu^2 + \sigma^2\big)\,\lambda t}} = \frac{\mu}{\sqrt{\mu^2 + \sigma^2}}.\end{aligned}$$

12.14 For each $k \in \mathcal{N}$, let X_k denote the amount deposited by customer k and let $Y(t)$ denote the total deposits by time t. In the notation of Exercise 12.13, $\lambda = 25.8$, $\mu = 574$ and $\sigma = 3167$.
a) Referring to the solution to Exercise 12.13(a), we have

$$\mathcal{E}\big(Y(t)\big) = \mu(t) = \mu\lambda t = 574 \cdot 25.8t = 14{,}809.20t.$$

b) Referring again to the solution to Exercise 12.13(a), we have

$$\sigma_{Y(t)} = \sqrt{\operatorname{Var}\big(Y(t)\big)} = \sqrt{\sigma^2(t)} = \sqrt{\big(\mu^2 + \sigma^2\big)\,\lambda t} = \sqrt{\big(574^2 + 3167^2\big) \cdot 25.8t} = 16{,}348.44\sqrt{t}.$$

Theory Exercises

12.15
a) We have

$$J(w_1, \ldots, w_n) = \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{vmatrix}.$$

We see that $J(w_1, \ldots, w_n)$ is the determinant of a lower triangular matrix and, hence, equals the product of the diagonal elements, all of which are always 1. Hence, $J(w_1, \ldots, w_n) = 1$ for all $w_1, \ldots, w_n \in \mathcal{R}$.
b) This result follows from the multivariate analogue of Exercise 9.95. Alternatively, for each $1 \le j \le n$, we have, for $t_j > 0$,

$$\begin{aligned}f_{I_j}(t_j) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{I_1,\ldots,I_n}(t_1, \ldots, t_n) \prod_{k \ne j} dt_k = \int_0^{\infty} \cdots \int_0^{\infty} \lambda e^{-\lambda t_1} \cdots \lambda e^{-\lambda t_n} \prod_{k \ne j} dt_k \\ &= \lambda e^{-\lambda t_j} \prod_{k \ne j} \int_0^{\infty} \lambda e^{-\lambda t_k}\, dt_k = \lambda e^{-\lambda t_j} \underbrace{1 \cdots 1}_{n-1 \text{ times}} = \lambda e^{-\lambda t_j}.\end{aligned}$$

Therefore, $I_j \sim \mathcal{E}(\lambda)$ for $1 \le j \le n$. Moreover,

$$f_{I_1,\ldots,I_n}(t_1, \ldots, t_n) = \lambda e^{-\lambda t_1} \cdots \lambda e^{-\lambda t_n} = f_{I_1}(t_1) \cdots f_{I_n}(t_n)$$

for all $t_1, \ldots, t_n > 0$, so that $I_1, \ldots, I_n$ are independent random variables.

c) The fact that $I_1, I_2, \ldots$ are independent and identically distributed exponential random variables with common parameter λ follows immediately from part (b) and the definition of independence for an infinite number of random variables as given in Definition 6.5 on page 296.

Advanced Exercises

12.16

a) For $0 \le s < t$,

$$\begin{aligned}
P\big(W_1 \le s \mid N(t) = 1\big) = P\big(N(s) \ge 1 \mid N(t) = 1\big) &= \frac{P\big(N(s) \ge 1,\ N(t) = 1\big)}{P\big(N(t) = 1\big)} \\
&= \frac{P\big(N(s) = 1,\ N(t) = 1\big)}{P\big(N(t) = 1\big)} = \frac{P\big(N(s) = 1,\ N(t) - N(s) = 0\big)}{P\big(N(t) = 1\big)} \\
&= \frac{P\big(N(s) = 1\big)\, P\big(N(t) - N(s) = 0\big)}{P\big(N(t) = 1\big)} \\
&= \frac{e^{-\lambda s}\,\dfrac{(\lambda s)^1}{1!} \cdot e^{-\lambda(t-s)}\,\dfrac{\big(\lambda(t-s)\big)^0}{0!}}{e^{-\lambda t}\,\dfrac{(\lambda t)^1}{1!}} \\
&= \frac{\lambda s}{\lambda t} = \frac{s}{t}.
\end{aligned}$$

Hence, $F_{W_1 \mid N(t)}(s \mid 1) = s/t$ if $0 \le s < t$, so that $W_{1 | N(t)=1} \sim \mathcal{U}(0, t)$.

b) Let $0 < s < t$ and let Δs be a small positive number. Given $N(t) = 1$, the event $\{s \le W_1 \le s + \Delta s\}$ occurs if and only if no successes occur in the time interval from 0 to s, exactly one success occurs in the time interval from s to $s + \Delta s$, and no successes occur in the time interval from $s + \Delta s$ to t. Hence, by properties of a Poisson process,

$$\begin{aligned}
&P\big(s \le W_1 \le s + \Delta s \mid N(t) = 1\big) \\
&\quad = P\big(N(s) = 0,\ N(s + \Delta s) - N(s) = 1,\ N(t) - N(s + \Delta s) = 0 \mid N(t) = 1\big) \\
&\quad = \frac{P\big(N(s) = 0,\ N(s + \Delta s) - N(s) = 1,\ N(t) - N(s + \Delta s) = 0\big)}{P\big(N(t) = 1\big)} \\
&\quad = \frac{P\big(N(s) = 0\big)\, P\big(N(s + \Delta s) - N(s) = 1\big)\, P\big(N(t) - N(s + \Delta s) = 0\big)}{P\big(N(t) = 1\big)} \\
&\quad = \frac{e^{-\lambda s}\,\dfrac{(\lambda s)^0}{0!} \cdot e^{-\lambda \Delta s}\,\dfrac{(\lambda \Delta s)^1}{1!} \cdot e^{-\lambda(t-s-\Delta s)}\,\dfrac{\big(\lambda(t - s - \Delta s)\big)^0}{0!}}{e^{-\lambda t}\,\dfrac{(\lambda t)^1}{1!}} = \frac{\lambda \Delta s}{\lambda t} = \frac{1}{t}\Delta s.
\end{aligned}$$

Therefore,

$$f_{W_1 \mid N(t)}(s \mid 1)\Delta s \approx P\big(s \le W_1 \le s + \Delta s \mid N(t) = 1\big) = \frac{1}{t}\Delta s, \qquad 0 < s < t.$$

Hence, $f_{W_1 \mid N(t)}(s \mid 1) = 1/t$ if $0 < s < t$, so that $W_{1|N(t)=1} \sim \mathcal{U}(0, t)$.

c) Let $0 \le s_1 < s_2 < t$. From the law of partitions, Proposition 2.8 on page 68, applied to a conditional probability, we get

$$\begin{aligned} P\big(W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big) = P\big(&N(s_1) = 0,\ W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big) \\ &+ P\big(N(s_1) = 1,\ W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big) \\ &+ P\big(N(s_1) = 2,\ W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big). \end{aligned}$$

The first probability is 0 because events $\{N(s_1) = 0\}$ and $\{W_1 \le s_1\}$ are mutually exclusive. For the second probability, we have

$$\begin{aligned} &P\big(N(s_1) = 1,\ W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big) \\ &\quad = P\big(N(s_1) = 1,\ N(s_2) - N(s_1) = 1 \mid N(t) = 2\big) \\ &\quad = \frac{P\big(N(s_1) = 1,\ N(s_2) - N(s_1) = 1,\ N(t) = 2\big)}{P\big(N(t) = 2\big)} \\ &\quad = \frac{P\big(N(s_1) = 1,\ N(s_2) - N(s_1) = 1,\ N(t) - N(s_2) = 0\big)}{P\big(N(t) = 2\big)} \\ &\quad = \frac{P\big(N(s_1) = 1\big)\, P\big(N(s_2) - N(s_1) = 1\big)\, P\big(N(t) - N(s_2) = 0\big)}{P\big(N(t) = 2\big)} \\ &\quad = \frac{e^{-\lambda s_1}\,\dfrac{(\lambda s_1)^1}{1!} \cdot e^{-\lambda(s_2 - s_1)}\,\dfrac{\big(\lambda(s_2 - s_1)\big)^1}{1!} \cdot e^{-\lambda(t - s_2)}\,\dfrac{\big(\lambda(t - s_2)\big)^0}{0!}}{e^{-\lambda t}\,\dfrac{(\lambda t)^2}{2!}} \\ &\quad = \frac{2\lambda s_1 \cdot \lambda(s_2 - s_1)}{\lambda^2 t^2} = \frac{2 s_1 (s_2 - s_1)}{t^2}. \end{aligned}$$

For the third probability, we have

$$\begin{aligned} &P\big(N(s_1) = 2,\ W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big) \\ &\quad = P\big(N(s_1) = 2 \mid N(t) = 2\big) = \frac{P\big(N(s_1) = 2,\ N(t) = 2\big)}{P\big(N(t) = 2\big)} \\ &\quad = \frac{P\big(N(s_1) = 2,\ N(t) - N(s_1) = 0\big)}{P\big(N(t) = 2\big)} = \frac{P\big(N(s_1) = 2\big)\, P\big(N(t) - N(s_1) = 0\big)}{P\big(N(t) = 2\big)} \\ &\quad \frac{e^{-\lambda s_1}\,\dfrac{(\lambda s_1)^2}{2!} \cdot e^{-\lambda(t - s_1)}\,\dfrac{\big(\lambda(t - s_1)\big)^0}{0!}}{e^{-\lambda t}\,\dfrac{(\lambda t)^2}{2!}} = \frac{\lambda^2 s_1^2}{\lambda^2 t^2} = \frac{s_1^2}{t^2}. \end{aligned}$$

Consequently,

$$P\big(W_1 \le s_1,\ W_2 \le s_2 \mid N(t) = 2\big) = \frac{2 s_1 (s_2 - s_1)}{t^2} + \frac{s_1^2}{t^2} = \frac{2 s_1 s_2 - s_1^2}{t^2}, \qquad 0 \le s_1 < s_2 < t.$$

Taking the mixed partial with respect to s_1 and s_2 yields

$$f_{W_1,W_2\,|\,N(t)}(s_1, s_2\,|\,2) = \frac{2}{t^2}, \qquad 0 < s_1 < s_2 < t.$$

Therefore, the conditional distribution of W_1 and W_2 given $N(t) = 2$ is the uniform distribution on the triangle $0 < s_1 < s_2 < t$.

d) Let $0 < s_1 < s_2 < t$ and let Δs_1 and Δs_2 represent small positive numbers. Given $N(t) = 2$, the event $\{s_1 \le W_1 \le s_1 + \Delta s_1,\ s_2 \le W_2 \le s_2 + \Delta s_2\}$ occurs if and only if no successes occur in the time interval from 0 to s_1, exactly one success occurs in the time interval from s_1 to $s_1 + \Delta s_1$, no successes occur in the time interval from $s_1 + \Delta s_1$ to s_2, exactly one success occurs in the time interval from s_2 to $s_2 + \Delta s_2$, and no successes occur in the time interval from $s_2 + \Delta s_2$ to t. Hence, by properties of a Poisson process, we get, by arguing as previously, that

$$\begin{aligned}
&P(s_1 \le W_1 \le s_1 + \Delta s_1,\ s_2 \le W_2 \le s_2 + \Delta s_2\,|\,N(t) = 2) \\
&\quad = \frac{e^{-\lambda s_1} \cdot e^{-\lambda \Delta s_1}\lambda \Delta s_1 \cdot e^{-\lambda(s_2 - s_1 - \Delta s_1)} \cdot e^{-\lambda \Delta s_2}\lambda \Delta s_2 \cdot e^{-\lambda(t - s_2 - \Delta s_2)}}{e^{-\lambda t}(\lambda t)^2/2} \\
&\quad = \frac{\lambda^2 e^{-\lambda t}\Delta s_1 \Delta s_2}{e^{-\lambda t}\lambda^2 t^2/2} = \frac{2}{t^2}\Delta s_1 \Delta s_2.
\end{aligned}$$

Therefore,

$$f_{W_1,W_2\,|\,N(t)}(s_1, s_2\,|\,2) = \frac{2}{t^2}, \qquad 0 < s_1 < s_2 < t.$$

Consequently, the conditional distribution of W_1 and W_2 given $N(t) = 2$ is the uniform distribution on the triangle $0 < s_1 < s_2 < t$.

e) Let $0 = s_0 < s_1 < \cdots < s_n < s_{n+1} = t$, and let $\Delta s_1, \ldots, \Delta s_n$ represent small positive numbers. For convenience, set $E = \{s_1 \le W_1 \le s_1 + \Delta s_1, \ldots, s_n \le W_n \le s_n + \Delta s_n\}$. Given $N(t) = n$, event E occurs if and only if no successes occur in the time interval from 0 to s_1, exactly one success occurs in the time interval from s_1 to $s_1 + \Delta s_1$, no successes occur in the time interval from $s_1 + \Delta s_1$ to s_2, exactly one success occurs in the time interval from s_2 to $s_2 + \Delta s_2, \ldots,$ no successes occur in the time interval from $s_{n-1} + \Delta s_{n-1}$ to s_n, exactly one success occurs in the time interval from s_n to $s_n + \Delta s_n$, and no successes occur in the time interval from $s_n + \Delta s_n$ to t.

For $1 \le j \le n$, let A_j denote the event that exactly one success occurs in the time interval from s_j to $s_j + \Delta s_j$ and, for $0 \le j \le n$, let B_j denote the event that no successes occur in the time interval from $s_j + \Delta s_j$ to s_{j+1}, where we define $\Delta s_0 = 0$. We note that, for $1 \le j \le n$,

$$P(A_j) = P\big(N(s_j + \Delta s_j) - N(s_j) = 1\big) = e^{-\lambda \Delta s_j}\,\frac{(\lambda \Delta s_j)^1}{1!} = e^{-\lambda \Delta s_j}\lambda \Delta s_j$$

and, for $0 \le j \le n$,

$$\begin{aligned}
P(B_j) &= P\big(N(s_{j+1}) - N(s_j + \Delta s_j) = 0\big) \\
&= e^{-\lambda(s_{j+1} - s_j - \Delta s_j)}\,\frac{\big(\lambda(s_{j+1} - s_j - \Delta s_j)\big)^0}{0!} = e^{-\lambda(s_{j+1} - s_j - \Delta s_j)}.
\end{aligned}$$

Hence, by the independent-increments property of a Poisson process,

$$P\left(E \mid N(t)=n\right) = \frac{P\left(E \cap \{N(t)=n\}\right)}{P\left(N(t)=n\right)} = \frac{P\left(B_0 \cap \bigcap_{j=1}^{n}(A_j \cap B_j)\right)}{P\left(N(t)=n\right)}$$

$$= \frac{P(B_0)\prod_{j=1}^{n} P(A_j)P(B_j)}{P\left(N(t)=n\right)} = \frac{e^{-\lambda s_1}\prod_{j=1}^{n}\left(e^{-\lambda \Delta s_j}\lambda \Delta s_j\right)e^{-\lambda(s_{j+1}-s_j-\Delta s_j)}}{e^{-\lambda t}(\lambda t)^n/n!}$$

$$= \frac{e^{-\lambda s_1}e^{-\lambda\sum_{j=1}^{n}(s_{j+1}-s_j)}\lambda^n \Delta s_1 \cdots \Delta s_n}{e^{-\lambda t}(\lambda t)^n/n!} = \frac{\lambda^n e^{-\lambda t}\Delta s_1 \cdots \Delta s_n}{e^{-\lambda t}(\lambda t)^n/n!}$$

$$= \frac{n!}{t^n}\Delta s_1 \cdots \Delta s_n.$$

Consequently,

$$f_{W_1,\ldots,W_n \mid N(t)}(s_1,\ldots,s_n \mid n) = \frac{n!}{t^n}, \qquad 0 < s_1 < \cdots < s_n < t,$$

and $f_{W_1,\ldots,W_n \mid N(t)}(s_1,\ldots,s_n \mid n) = 0$ otherwise.

f) From the result of Exercise 9.34, the conditional joint PDF found in part (e) is that of the joint PDF of the order statistics of a random sample of size n from a uniform distribution on the interval $(0, t)$. Hence, given $N(t) = n$, the times at which the n events occur, considered as unordered random variables, are independent and uniformly distributed on the interval $(0, t)$.

12.17

a) Given that the first success occurs at time s (i.e., $I_1 = s$), the elapsed time until the second success exceeds t (i.e., $I_2 > t$) if and only if no successes occur in the time interval $(s, s+t]$. Therefore,

$$\begin{aligned} P(I_2 > t \mid I_1 = s) &= P\left(N(s+t) - N(s) = 0 \mid I_1 = s\right) \\ &= P\left(N(I_1+t) - N(I_1) = 0 \mid I_1 = s\right). \end{aligned}$$

b) We have that $I_1 = s$ if and only if $N(u) = 0$ for $0 \le u < s$, and $N(s) = 1$. Hence, the value of I_1 depends only on the values of $\{N(u) : u \ge 0\}$ in the interval $0 \le u \le I_1$.

c) Applying in turn the facts that a Poisson process has independent increments, stationary increments, and $N(t) \sim \mathcal{P}(\lambda t)$, we conclude, in view of parts (a) and (b), that

$$\begin{aligned} P(I_2 > t \mid I_1 = s) &= P\left(N(I_1+t) - N(I_1) = 0 \mid I_1 = s\right) = P\left(N(I_1+t) - N(I_1) = 0\right) \\ &= P\left(N(t) = 0\right) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}. \end{aligned}$$

Referring now to Equation (9.33) on page 517, we get, for all $s, t > 0$,

$$\begin{aligned} P(I_1 > s,\ I_2 > t) &= \int_s^\infty \int_t^\infty f_{I_1,I_2}(u,v)\,du\,dv = \int_s^\infty \left(\int_t^\infty f_{I_2 \mid I_1}(v \mid u)\,dv\right) f_{I_1}(u)\,du \\ &= \int_s^\infty P(I_2 > t \mid I_1 = u) f_{I_1}(u)\,du = e^{-\lambda t}\int_s^\infty f_{I_1}(u)\,du = P(I_1 > s)e^{-\lambda t}. \end{aligned}$$

Letting $s \to -\infty$ yields $P(I_2 > t) = e^{-\lambda t}$ and, hence, that $P(I_1 > s,\ I_2 > t) = P(I_1 > s)P(I_2 > t)$. We therefore conclude that $I_2 \sim \mathcal{E}(\lambda)$ and that I_1 and I_2 are independent random variables. That $I_1 \sim \mathcal{E}(\lambda)$ is a consequence of Proposition 12.1 on page 689 and the fact that $I_1 = W_1$.

d) Given that $I_1 = s_1, \ldots, I_n = s_n$, the elapsed time until the $(n+1)$st success exceeds t (i.e., $I_n > t$) if and only if no successes occur in $(s_1 + \cdots + s_n, s_1 + \cdots + s_n + t]$. Therefore,

$$\begin{aligned} &P(I_{n+1} > t \mid I_1 = s_1, \ldots, I_n = s_n) \\ &\quad = P\big(N(s_1 + \cdots + s_n + t) - N(s_1 + \cdots + s_n) = 0 \mid I_1 = s_1, \ldots, I_n = s_n\big) \\ &\quad = P\big(N(I_1 + \cdots + I_n + t) - N(I_1 + \cdots + I_n) = 0 \mid I_1 = s_1, \ldots, I_n = s_n\big). \end{aligned}$$

Now, $I_1 = s_1, \ldots, I_n = s_n$ if and only if

$$\begin{array}{ll} N(u) = 0, & 0 \le u < s_1; \\ N(u) = 1, & s_1 \le u < s_1 + s_2; \\ \vdots & \vdots \\ N(u) = n - 1, & s_1 + \cdots + s_{n-1} \le u < s_1 + \cdots + s_n; \end{array}$$

and $N(s_1 + \cdots + s_n) = n$. Thus, the values of $I_1, \ldots, I_n$ depend only on the values of $\{N(u) : u \ge 0\}$ for $0 \le u \le I_1 + \cdots + I_n$. Applying in turn the facts that a Poisson process has independent increments and stationary increments, we conclude that

$$\begin{aligned} &P(I_{n+1} > t \mid I_1 = s_1, \ldots, I_n = s_n) \\ &\quad = P\big(N(I_1 + \cdots + I_n + t) - N(I_1 + \cdots + I_n) = 0 \mid I_1 = s_1, \ldots, I_n = s_n\big) \\ &\quad = P\big(N(I_1 + \cdots + I_n + t) - N(I_1 + \cdots + I_n) = 0\big) \\ &\quad = P\big(N(t) = 0\big) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}. \end{aligned}$$

e) We must show that, for each $n \in \mathcal{N}$, the random variables $I_1, \ldots, I_n$ are independent and identically distributed exponential random variables with common parameter λ. The validity of this statement for $n = 2$ has already been established in part (c). We now use mathematical induction. Referring to part (d) and noting that $e^{-\lambda t}$ does not depend on $s_1, \ldots, s_n$, it follows that I_{n+1} is independent of $I_1, \ldots, I_n$, and, furthermore, that $P(I_{n+1} > t) = e^{-\lambda t}$, which shows $I_{n+1} \sim \mathcal{E}(\lambda)$. Hence, from the induction assumption, $I_1, \ldots, I_{n+1}$ are independent and identically distributed exponential random variables with common parameter λ.

12.18 Suppose that $\{N(t) : t \ge 0\}$ is a Poisson process, either homogeneous or nonhomogeneous, with stationary increments. Then $\mu(t) - \mu(s)$ depends only on the difference between s and t; that is, there is a function g such that

$$\mu(t) - \mu(s) = g(t - s), \qquad 0 \le s < t < \infty.$$

Because $\mu(0) = 0$, setting $s = 0$ in the previous display shows that $\mu = g$; hence,

$$\mu(t) - \mu(s) = \mu(t - s), \qquad 0 \le s < t < \infty.$$

Letting $t = 2s$ shows that $\mu(2s) = 2\mu(s)$, and it follows by mathematical induction that $\mu(ps) = p\mu(s)$ for all $p \in \mathcal{N}$. Setting $s = 1/p$ in the preceding relation gives $\mu(1) = p\mu(1/p)$. Consequently, for all $p, q \in \mathcal{N}$,

$$\mu(p/q) = p\mu(1/q) = p\mu(1)/q = (p/q)\mu(1).$$

As μ is everywhere continuous (being the indefinite integral of ν), it now follows that $\mu(t) = \mu(1)t$ for all $t > 0$. Hence, μ is linear, which means that $\{N(t) : t \ge 0\}$ is a homogeneous Poisson process. Thus, a nonhomogeneous Poisson process doesn't have stationary increments.

12.19

a) By assumption, $Y_{|X=t} \sim \mathcal{P}(\lambda t)$ for $t > 0$. In particular, then, $\mathcal{E}(Y \mid X = t) = \lambda t$ for all $t > 0$. Applying the law of total expectation now yields

$$\mathcal{E}(Y) = \mathcal{E}\big(\mathcal{E}(Y \mid X)\big) = \mathcal{E}(\lambda X) = \lambda\mathcal{E}(X)\,.$$

b) Referring to Exercises 9.83(b) and 9.85(a), we get

$$p_Y(y) = \int_{-\infty}^{\infty} h_{X,Y}(x, y)\,dx = \int_{-\infty}^{\infty} f_X(x)p_{Y\,|\,X}(y\,|\,x)\,dx = \int_0^{\infty} e^{-\lambda x}\,\frac{(\lambda x)^y}{y!}\,f_X(x)\,dx$$

if $y = 0, 1, 2, \ldots$, and $p_Y(y) = 0$ otherwise.

c) From the definition of expected value and part (b),

$$\begin{aligned}\mathcal{E}(Y) &= \sum_Y y p_Y(y) = \sum_{y=0}^{\infty} y\left(\int_0^{\infty} e^{-\lambda x}\,\frac{(\lambda x)^y}{y!}\,f_X(x)\,dx\right) = \int_0^{\infty}\left(\sum_{y=0}^{\infty} y\,e^{-\lambda x}\,\frac{(\lambda x)^y}{y!}\right) f_X(x)\,dx \\ &= \int_0^{\infty} (\lambda x) f_X(x)\,dx = \lambda\int_0^{\infty} x f_X(x)\,dx = \lambda\mathcal{E}(X)\,,\end{aligned}$$

which agrees with the result of part (a).

d) When $X \sim \mathcal{E}(\mu)$, we have $\mathcal{E}(X) = 1/\mu$ and $f_X(x) = \mu e^{-\mu x}$ for $x > 0$. Hence, from parts (a) and (b),

$$\mathcal{E}(Y) = \lambda\mathcal{E}(X) = \frac{\lambda}{\mu}$$

and

$$\begin{aligned}p_Y(y) &= \int_0^{\infty} e^{-\lambda x}\,\frac{(\lambda x)^y}{y!}\mu e^{-\mu x}\,dx = \frac{\mu\lambda^y}{y!}\int_0^{\infty} x^{(y+1)-1}e^{-(\mu+\lambda)x}\,dx \\ &= \frac{\mu\lambda^y}{y!}\cdot\frac{\Gamma(y+1)}{(\mu+\lambda)^{y+1}} = \frac{(\lambda/\mu)^y}{(1+\lambda/\mu)^{y+1}}\end{aligned}$$

if $y = 0, 1, 2, \ldots$, and $p_Y(y) = 0$ otherwise. Thus, Y has the Pascal distribution with parameter λ/μ, as defined in Exercise 7.26.

12.2 Basic Queueing Theory

Basic Exercises

12.20

a) To reenter state n, the queueing system must first leave that state. Let $E_n(t)$ and $L_n(t)$ denote the numbers of times that the queueing system enters and leaves state n, respectively, by time t. If the queueing system is in state n at time t, then $E_n(t) = L_n(t) + 1$; otherwise, $E_n(t) = L_n(t)$. Hence, in each time interval $[0, t]$, the number of times the queueing system enters state n must equal or exceed by 1 the number of times the queueing system leaves state n.

b) The average numbers of times that the queueing system enters and leaves state n by time t equal $E_n(t)/t$ and $L_n(t)/t$, respectively. From part (a), we get

$$\frac{L_n(t)}{t} \le \frac{E_n(t)}{t} \le \frac{L_n(t)+1}{t}.$$

From these inequalities, we see that

$$\lim_{t\to\infty} \frac{L_n(t)}{t} \le \lim_{t\to\infty} \frac{E_n(t)}{t} \le \lim_{t\to\infty} \frac{L_n(t)+1}{t} = \lim_{t\to\infty} \frac{L_n(t)}{t},$$

and, hence, $\lim_{t\to\infty} E_n(t)/t = \lim_{t\to\infty} L_n(t)/t$. Thus, the average rates at which the queueing system enters and leaves state n are equal; that is, the rate-in = rate-out principle holds.

12.21

a) This drive-up station is an $M/M/1$ queue with $\lambda = 10$ and $\mu = 11$.
b) Because $\lambda < \mu$ ($10 < 11$), we know from Example 12.6 on page 703 that a steady-state distribution exists for the number of customers in this queueing system. And, from Equation (12.20) on page 704, the steady-state distribution is

$$P_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n = \left(1 - \frac{10}{11}\right)\left(\frac{10}{11}\right)^n = \frac{1}{11}\left(\frac{10}{11}\right)^n, \qquad n \ge 0.$$

c) From Equation (12.24) on page 707, the steady-state expected number of customers in the queueing system is

$$L = \frac{\lambda}{\mu - \lambda} = \frac{10}{11 - 10} = 10.$$

d) Let Y denote the steady-state number of customers in the queue. Then $Y = 0$ if $X = 0$ and $Y = X - 1$ if $X \ge 1$. Hence, from the FEF and parts (b) and (c),

$$\begin{aligned} L_q = \mathcal{E}(Y) &= \sum_{n=1}^{\infty}(n-1)P_n = \sum_{n=1}^{\infty} nP_n - \sum_{n=1}^{\infty} P_n \\ &= L - (1 - P_0) = 10 - \left(1 - \frac{1}{11}\right) = \frac{100}{11} \approx 9.09. \end{aligned}$$

12.22

a) This drive-up station is an $M/M/2$ queue with $\lambda = 10$ and $\mu = 11$.
b) Because $\lambda < 2\mu$ ($10 < 22$), we know from Example 12.6 on page 703 that a steady-state distribution exists for the number of customers in this queueing system. From Equation (12.22) on page 704,

$$P_0 = \left(\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!(1 - \lambda/s\mu)}\right)^{-1} = \left(\sum_{n=0}^{1} \frac{(10/11)^n}{n!} + \frac{(10/11)^2}{2(1 - 10/2\cdot 11)}\right)^{-1} = \frac{3}{8}.$$

And, from Equation (12.23) on page 705,

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0, & \text{if } 1 \le n \le s-1; \\[2ex] \dfrac{(\lambda/\mu)^n}{s!\,s^{n-s}} P_0, & \text{for } n \ge s. \end{cases} = \begin{cases} \dfrac{(10/11)^n}{n!} \cdot \dfrac{3}{8}, & \text{if } n = 1; \\[2ex] \dfrac{(10/11)^n}{2\cdot 2^{n-2}} \cdot \dfrac{3}{8}, & \text{if } n \ge 2. \end{cases} = \frac{3}{4}\left(\frac{5}{11}\right)^n, \qquad n \ge 1.$$

c) See the solution to Exercise 12.23(c), specifically, the first three columns in the table and the first two probability histograms.

d) From Equation (12.28) on page 708, the steady-state expected number of customers in the queueing system is

$$L = \frac{\lambda}{\mu} + \frac{(\lambda/s\mu)(\lambda/\mu)^s}{s!(1-\lambda/s\mu)^2} P_0 = \frac{10}{11} + \frac{(10/2\cdot 11)(10/11)^2}{2\cdot(1-10/2\cdot 11)^2}\cdot\frac{3}{8} \approx 1.15.$$

By having two tellers instead of one, we can reduce the steady-state expected number of customers in the queueing system from 10 to about 1.15.

e) Let Y denote the steady-state number of customers in the queue. Then we have $Y = 0$ if $X = 0$ or 1, and $Y = X - 2$ if $X \geq 2$. Hence, from the FEF and parts (b) and (d),

$$\begin{aligned} L_q = \mathcal{E}(Y) &= \sum_{n=2}^{\infty}(n-2)P_n = \sum_{n=2}^{\infty} nP_n - 2\sum_{n=2}^{\infty} P_n \\ &= (L - P_1) - 2(1 - P_0 - P_1) = L - 2 + 2P_0 + P_1 \approx 0.24. \end{aligned}$$

By having two tellers instead of one, we can reduce the steady-state expected number of customers in the queue from about 9.09 to about 0.24.

12.23

a) This drive-up station is an $M/M/3$ queue with $\lambda = 10$ and $\mu = 11$.

b) Because $\lambda < 3\mu$ $(10 < 33)$, we know from Example 12.6 on page 703 that a steady-state distribution exists for the number of customers in this queueing system. From Equation (12.22) on page 704,

$$P_0 = \left(\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!(1-\lambda/s\mu)}\right)^{-1} = \left(\sum_{n=0}^{2}\frac{(10/11)^n}{n!} + \frac{(10/11)^3}{6(1-10/3\cdot 11)}\right)^{-1} \approx 0.400.$$

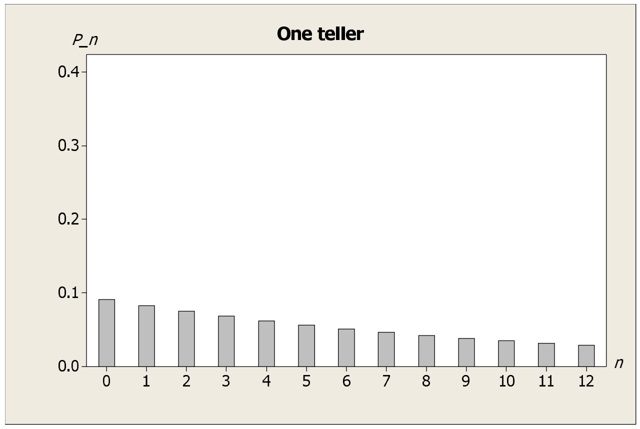
And, from Equation (12.23) on page 705,

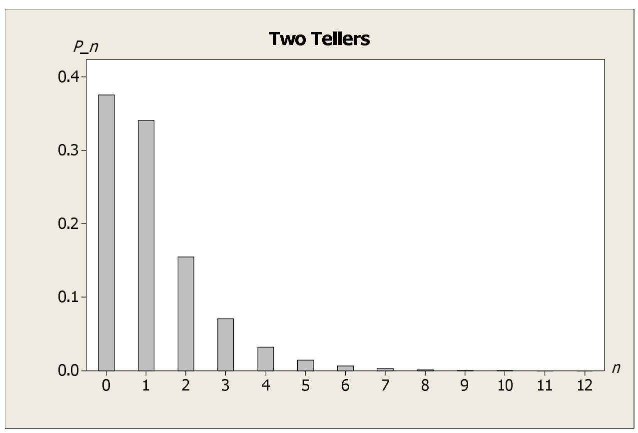
$$P_n = \begin{cases} \dfrac{(10/11)^n}{n!}P_0, & \text{if } 1 \leq n \leq 2; \\[2ex] \dfrac{(10/11)^n}{3!\,3^{n-3}}P_0, & \text{if } n \geq 3. \end{cases} \approx \begin{cases} 0.363, & \text{if } n = 1; \\[1ex] 1.799\left(\dfrac{10}{33}\right)^n, & \text{if } n \geq 2. \end{cases}$$

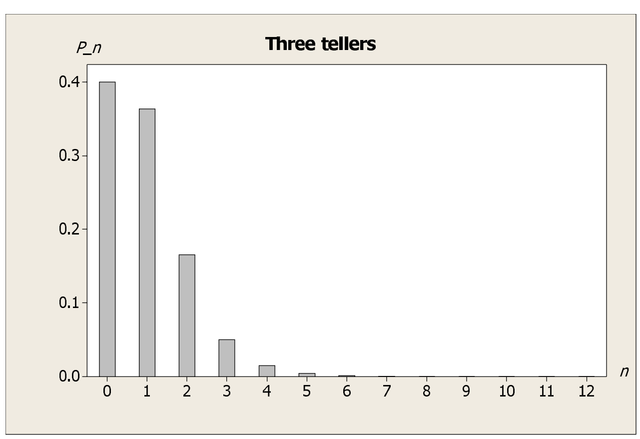
c) In the following table, we list the exact values of the steady-state probabilities to six decimal places from $n = 0$ to $n = 12$.

n	One teller	Two tellers	Three tellers
0	0.090909	0.375000	0.399684
1	0.082645	0.340909	0.363349
2	0.075131	0.154959	0.165159
3	0.068301	0.070436	0.050048
4	0.062092	0.032016	0.015166
5	0.056447	0.014553	0.004596
6	0.051316	0.006615	0.001393
7	0.046651	0.003007	0.000422
8	0.042410	0.001367	0.000128
9	0.038554	0.000621	0.000039
10	0.035049	0.000282	0.000012
11	0.031863	0.000128	0.000004
12	0.028966	0.000058	0.000001

The following three graphs provide probability histograms for the steady-state probabilities in the preceding table.







d) From Equation (12.28) on page 708, the steady-state expected number of customers in the queueing system is

$$L = \frac{\lambda}{\mu} + \frac{(\lambda/s\mu)(\lambda/\mu)^s}{s!(1-\lambda/s\mu)^2} P_0 = \frac{10}{11} + \frac{(10/3 \cdot 11)(10/11)^3}{3!(1-10/3 \cdot 11)^2} P_0 \approx 0.94.$$

Roughly, the steady-state expected number of customers in the three-teller system is 0.94 compared with 10 and 1.15 in the one-teller and two-teller systems, respectively.
e) Let Y denote the steady-state number of customers in the queue. Then we have $Y = 0$ if $X = 0$, 1, or 2, and $Y = X - 3$ if $X \geq 3$. Hence, from the FEF and parts (b) and (d),

$$\begin{aligned} L_q &= \mathcal{E}(Y) = \sum_{n=3}^{\infty} (n-3) P_n = \sum_{n=3}^{\infty} n P_n - 3 \sum_{n=3}^{\infty} P_n \\ &= (L - P_1 - 2P_2) - 3(1 - P_0 - P_1 - P_2) = L - 3 + 3P_0 + 2P_1 + P_2 \approx 0.03. \end{aligned}$$

Roughly, the steady-state expected number of customers in the queue in the three-teller system is 0.03 compared with 9.09 and 0.24 in the one-teller and two-teller systems, respectively.

12.24 As the maximum number of customers in an $M/M/s/N$ queueing system is N, having more than N servers would be pointless. Hence, we assume that $s \leq N$.

12.25
a) This barbershop is an $M/M/1/4$ queue with $\lambda = 3$ and $\mu = 2$.
b) We have

$$\lambda_n = \begin{cases} 3, & \text{if } 0 \leq n \leq 3; \\ 0, & \text{if } n \geq 4. \end{cases} \qquad \mu_n = 2, \quad \text{for } n \geq 1.$$

Hence,

$$\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} = \begin{cases} (3/2)^n, & \text{if } 1 \leq n \leq 4; \\ 0, & \text{if } n \geq 5. \end{cases}$$

Referring now to Proposition 12.3 on page 703, we find that

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right)\right)^{-1} = \left(1 + \sum_{n=1}^{4} (3/2)^n\right)^{-1} = \left(\frac{1-(3/2)^5}{1-3/2}\right)^{-1} = \frac{16}{211}$$

and, moreover, for $1 \leq n \leq 4$,

$$P_n = \left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right) P_0 = \left(\frac{3}{2}\right)^n \cdot \frac{16}{211}.$$

Hence,

$$P_n = \begin{cases} \dfrac{16}{211} \left(\dfrac{3}{2}\right)^n, & \text{if } 0 \leq n \leq 4; \\ 0, & \text{if } n \geq 5. \end{cases}$$

c) Referring to part (b), we get

$$L = \sum_{n=0}^{\infty} n P_n = \sum_{n=1}^{4} n \cdot \frac{16}{211} \left(\frac{3}{2}\right)^n = \frac{16}{211} \sum_{n=1}^{4} n \left(\frac{3}{2}\right)^n \approx 2.76.$$

12.26

a) From Equation (12.12) on page 700,

$$\lambda_n = \begin{cases} \lambda, & \text{if } 0 \le n \le N-1; \\ 0, & \text{if } n \ge N. \end{cases} \qquad \mu_n = \mu, \quad \text{for } n \ge 1.$$

Hence,

$$\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} = \begin{cases} (\lambda/\mu)^n, & \text{if } 1 \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases}$$

Therefore,

$$\sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right) = \sum_{n=1}^{N}(\lambda/\mu)^n < \infty,$$

so that, by Proposition 12.3 on page 703, an $M/M/1/N$ queueing system has a steady-state distribution.

b) Referring again to Proposition 12.3 and to part (a), we find that

$$P_0 = \left(1 + \sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right)\right)^{-1} = \left(\sum_{n=0}^{N}(\lambda/\mu)^n\right)^{-1} = \begin{cases} \dfrac{1}{N+1}, & \text{if } \lambda = \mu; \\ \dfrac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}}, & \text{if } \lambda \ne \mu. \end{cases}$$

and, moreover, for $1 \le n \le N$,

$$P_n = \left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right) P_0 = \left(\frac{\lambda}{\mu}\right)^n P_0.$$

Consequently, if $\lambda = \mu$,

$$P_n = \begin{cases} \dfrac{1}{N+1}, & \text{if } 0 \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases} \tag{$*$}$$

whereas, if $\lambda \ne \mu$,

$$P_n = \begin{cases} \dfrac{(1-\lambda/\mu)(\lambda/\mu)^n}{1-(\lambda/\mu)^{N+1}}, & \text{if } 0 \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases} \tag{$**$}$$

c) We consider two cases:

Case 1: $\lambda = \mu$. Then, in view of ($*$),

$$L = \sum_{n=0}^{\infty} n P_n = \sum_{n=1}^{N} n \frac{1}{N+1} = \frac{1}{N+1}\sum_{n=1}^{N} n = \frac{1}{N+1} \cdot \frac{N(N+1)}{2} = \frac{N}{2}.$$

Case 2: $\lambda \ne \mu$. For this case, we first note that, for $a \ne 1$,

$$\sum_{n=1}^{N} n a^n = a \sum_{n=1}^{N} n a^{n-1} = a \frac{d}{da}\left(\sum_{n=0}^{N} a^n\right) = a \frac{d}{da}\left(\frac{1-a^{N+1}}{1-a}\right)$$

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

For convenience, set $\rho = \lambda/\mu$. Then, in view of ($**$),

$$L = \sum_{n=0}^{\infty} n P_n = \frac{1-\lambda/\mu}{1-\rho^{N+1}} \sum_{n=1}^{N} n\rho^n = \frac{1-\rho}{1-\rho^{N+1}} \cdot \frac{\rho\left(1-\rho^{N+1}-(N+1)\rho^N(1-\rho)\right)}{(1-\rho)^2}$$
$$= \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} = \frac{\lambda/\mu}{1-\lambda/\mu} - \frac{(N+1)(\lambda/\mu)^{N+1}}{1-(\lambda/\mu)^{N+1}}.$$

d) In this case $\lambda = 3$, $\mu = 2$, and $N = 4$. Hence, by part (b), specifically, ($**$), the steady-state distribution for the number of customers in the barbershop is

$$P_n = \begin{cases} \dfrac{(1-3/2)(3/2)^n}{1-(3/2)^5}, & \text{if } 0 \le n \le 4; \\ 0, & \text{if } n \ge 5. \end{cases} = \begin{cases} \dfrac{16}{211}\left(\dfrac{3}{2}\right)^n, & \text{if } 0 \le n \le 4; \\ 0, & \text{if } n \ge 5. \end{cases}$$

From part (c), the steady-state expected number of customers in the barbershop is

$$L = \frac{3/2}{1-3/2} - \frac{5(3/2)^5}{1-(3/2)^5} = \frac{582}{211} \approx 2.76.$$

12.27
a) We have

$$\lambda_n = \lambda, \quad \text{for } n \ge 0. \qquad \mu_n = \begin{cases} \mu, & \text{if } 1 \le n \le 2; \\ \nu, & \text{if } n \ge 3. \end{cases}$$

Hence,

$$\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} = \begin{cases} (\lambda/\mu)^n, & \text{if } 1 \le n \le 2; \\ (\lambda/\mu)^2(\lambda/\nu)^{n-2}, & \text{if } n \ge 3. \end{cases} = \begin{cases} \lambda/\mu, & \text{if } n = 1; \\ (\lambda/\mu)^2(\lambda/\nu)^{n-2}, & \text{if } n \ge 2. \end{cases}$$

Also,

$$\sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right) = \lambda/\mu + \sum_{n=2}^{\infty}(\lambda/\mu)^2(\lambda/\nu)^{n-2} = \lambda/\mu + \frac{(\lambda/\mu)^2}{1-\lambda/\nu} = \lambda/\mu + \frac{(\lambda/\mu)^2\nu}{\nu-\lambda}.$$

Referring now to Proposition 12.3 on page 703, we find that

$$P_0 = \left(1 + \sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right)\right)^{-1} = \left(1 + \lambda/\mu + \frac{(\lambda/\mu)^2\nu}{\nu-\lambda}\right)^{-1}$$

and that

$$P_1 = (\lambda/\mu)P_0 \qquad \text{and} \qquad P_n = (\lambda/\mu)^2(\lambda/\nu)^{n-2}P_0, \quad n \ge 2.$$

b) Referring to part (a) and applying the formula $\sum_{n=1}^{\infty} na^{n-1} = 1/(1-a)^2$, we get

$$L_q = \sum_{n=1}^{\infty}(n-1)P_n = \sum_{n=2}^{\infty}(n-1)(\lambda/\mu)^2(\lambda/\nu)^{n-2}P_0 = (\lambda/\mu)^2 P_0 \sum_{n=2}^{\infty}(n-1)(\lambda/\nu)^{n-2}$$
$$= (\lambda/\mu)^2 P_0 \sum_{n=1}^{\infty} n(\lambda/\nu)^{n-1} = \frac{(\lambda/\mu)^2 P_0}{(1-\lambda/\nu)^2} = \left(\frac{\lambda\nu}{\mu(\nu-\lambda)}\right)^2 P_0.$$

c) From part (a), with $\lambda = 20$, $\mu = 16$, and $\nu = 24$,

$$P_0 = \left(1 + \lambda/\mu + \frac{(\lambda/\mu)^2\nu}{\nu - \lambda}\right)^{-1} = \left(1 + 20/16 + \frac{(20/16)^2 \cdot 24}{24 - 20}\right)^{-1} = \frac{8}{93}.$$

Applying now part (b) yields

$$L_q = \left(\frac{\lambda\nu}{\mu(\nu - \lambda)}\right)^2 P_0 = \left(\frac{20 \cdot 24}{16(24 - 20)}\right)^2 \cdot \frac{8}{93} = \frac{150}{31}.$$

The steady-state expected number of customers in the queue is roughly 4.8.

12.28 In the $M/M/2$ queue, service is at rate μ when one customer is in the system and at rate 2μ when two or more customers are in the system. In the $M/M/1$ queue, service is always at rate 2μ when one or more customers are in the system. Hence the $M/M/1$ queue is more efficient.

Note: Throughout the remaining exercises in this section, we let $\rho = \lambda/s\mu$ and $r = \lambda/\mu$. Of course, we have $\rho = r$ in the single-server case.

Theory Exercises

12.29

a) Let Y denote the steady-state number of customers in the queue and set $Q_n = P(Y = n)$. Then $Y = 0$ if $X = 0$ or 1, and $Y = X - 1$ if $X \geq 2$. Referring to Equation (12.20) on page 704, we get

$$Q_0 = P(Y = 0) = P(X = 0) + P(X = 1) = 1 - \rho + (1 - \rho)\rho = 1 - \rho^2 = 1 - (\lambda/\mu)^2$$

and, for $n \geq 1$,

$$Q_n = P(Y = n) = P(X = n + 1) = (1 - \rho)\rho^{n+1} = (1 - \lambda/\mu)(\lambda/\mu)^{n+1}.$$

Thus,

$$Q_n = \begin{cases} 1 - (\lambda/\mu)^2, & \text{if } n = 0; \\ (1 - \lambda/\mu)(\lambda/\mu)^{n+1}, & \text{if } n \geq 1. \end{cases}$$

b) From part (a) and the formula $\sum_{n=1}^{\infty} na^{n-1} = 1/(1 - a)^2$, we get

$$L_q = \mathcal{E}(Y) = \sum_{n=0}^{\infty} nQ_n = \sum_{n=1}^{\infty} n(1 - \rho)\rho^{n+1} = (1 - \rho)\rho^2 \sum_{n=1}^{\infty} n\rho^{n-1}$$
$$= \frac{(1 - \rho)\rho^2}{(1 - \rho)^2} = \frac{\rho^2}{1 - \rho} = \frac{(\lambda/\mu)^2}{1 - \lambda/\mu} = \frac{\lambda^2}{\mu(\mu - \lambda)}.$$

Alternatively, we can use the FEF and Equations (12.20) and (12.24) on pages 704 and 707, respectively, to get

$$L_q = \mathcal{E}(Y) = 0 \cdot P_0 + 0 \cdot P_1 + \sum_{n=2}^{\infty}(n - 1)P_n = \sum_{n=2}^{\infty} nP_n - \sum_{n=2}^{\infty} P_n$$
$$= (L - 1 \cdot P_1) - (1 - P_0 - P_1) = L - 1 + P_0 = \frac{\lambda}{\mu - \lambda} - 1 + 1 - \frac{\lambda}{\mu}$$
$$= \frac{\lambda^2}{\mu(\mu - \lambda)}.$$

12.30

a) As the service times are independent exponential random variables with common parameter μ, the time it takes for the server to serve n consecutive customers has a $\Gamma(n, \mu)$ distribution. Now, let $\mathcal{W}$ denote the steady-state waiting time in the queueing system for each individual customer. Because of the lack-of-memory property of an exponential random variable, we see that $\mathcal{W}_{|X=n} \sim \Gamma(n+1, \mu)$, where X denotes the number of customers in the queueing system when the individual customer arrives at the service facility. Applying the law of total probability and referring to Equation (12.20) on page 704, we find that, for $t \geq 0$,

$$
\begin{aligned}
P(\mathcal{W} > t) &= \sum_{n=0}^{\infty} P(X = n)P(\mathcal{W} > t \mid X = n) = \sum_{n=0}^{\infty} \left((1-\rho)\rho^n \int_t^{\infty} \frac{\mu^{n+1}}{\Gamma(n+1)} x^{(n+1)-1} \, e^{-\mu x} \, dx \right) \\
&= (1-\rho)\mu \int_t^{\infty} e^{-\mu x} \left(\sum_{n=0}^{\infty} \frac{(\lambda x)^n}{n!} \right) dx = (1-\rho)\mu \int_t^{\infty} e^{-\mu x} e^{\lambda x} \, dx \\
&= (1-\rho)\mu \int_t^{\infty} e^{-(\mu-\lambda)x} \, dx = \frac{(1-\rho)\mu}{\mu-\lambda} e^{-(\mu-\lambda)t} = e^{-(\mu-\lambda)t}.
\end{aligned}
$$

Thus, we see that $\mathcal{W}$ has the exponential distribution with parameter $\mu - \lambda$.

b) From part (a) and the fact that the mean of an exponential random variable is the reciprocal of its parameter, we have

$$
W = \mathcal{E}(\mathcal{W}) = \frac{1}{\mu - \lambda}.
$$

12.31

a) As the service times are independent exponential random variables with common parameter μ, the time it takes for the server to serve n consecutive customers has a $\Gamma(n, \mu)$ distribution. Now, let $\mathcal{W}_q$ denote the steady-state waiting time in the queue for each individual customer and let X denote the number of customers in the queueing system when the individual customer arrives at the service facility. Clearly, $\mathcal{W}_{q\,|X=0} = 0$ and, because of the lack-of-memory property of an exponential random variable, we have $\mathcal{W}_{q\,|X=n} \sim \Gamma(n, \mu)$ for $n \geq 1$. Applying the law of total probability and referring to Equation (12.20) on page 704, we find that, for $t \geq 0$,

$$
\begin{aligned}
P(\mathcal{W}_q > t) &= \sum_{n=0}^{\infty} P(X = n)P(\mathcal{W}_q > t \mid X = n) \\
&= P(X = 0)P(\mathcal{W}_q > t \mid X = 0) + \sum_{n=1}^{\infty} P(X = n)P(\mathcal{W}_q > t \mid X = n) \\
&= 0 + \sum_{n=1}^{\infty} \left((1-\rho)\rho^n \int_t^{\infty} \frac{\mu^n}{\Gamma(n)} x^{n-1} \, e^{-\mu x} \, dx \right) \\
&= (1-\rho)\lambda \int_t^{\infty} e^{-\mu x} \left(\sum_{n=1}^{\infty} \frac{(\lambda x)^{n-1}}{(n-1)!} \right) dx = (1-\rho)\lambda \int_t^{\infty} e^{-\mu x} e^{\lambda x} \, dx \\
&= (1-\rho)\lambda \int_t^{\infty} e^{-(\mu-\lambda)x} \, dx = \frac{(1-\rho)\lambda}{\mu-\lambda} e^{-(\mu-\lambda)t} = (\lambda/\mu) e^{-(\mu-\lambda)t}.
\end{aligned}
$$

Thus, we see that

$$F_{\mathcal{W}_q}(t) = \begin{cases} 0, & \text{if } t < 0; \\ 1 - (\lambda/\mu)e^{-(\mu-\lambda)t}, & \text{if } t \geq 0. \end{cases}$$

b) From Proposition 10.5 on page 577 and part (a),

$$W_q = \mathcal{E}(\mathcal{W}_q) = \int_0^\infty P(\mathcal{W}_q > t)\,dt = \frac{\lambda}{\mu}\int_0^\infty e^{-(\mu-\lambda)t}\,dt = \frac{\lambda}{\mu}\cdot\frac{1}{\mu-\lambda} = \frac{\lambda}{\mu(\mu-\lambda)}.$$

12.32 An arriving customer must join the queue if and only if $X \geq 1$. From Exercise 12.31, we know that $\mathcal{W}_{q\,|\,X=0} = 0$. Hence, for $t \geq 0$, we have $\{\mathcal{W}_q > t\} \subset \{X \geq 1\}$. Therefore, from Exercise 12.31 and Equation (12.20) on page 704, we have, for $t \geq 0$,

$$P(\mathcal{W}_q > t \mid X \geq 1) = \frac{P(\mathcal{W}_q > t,\ X \geq 1)}{P(X \geq 1)} = \frac{P(\mathcal{W}_q > t)}{P(X \geq 1)} = \frac{(\lambda/\mu)e^{-(\mu-\lambda)t}}{1-(1-\lambda/\mu)} = e^{-(\mu-\lambda)t}.$$

Consequently, the conditional distribution of the waiting time in the queue, given that the customer must join the queue, is exponential with parameter $\mu - \lambda$; the conditional expectation is, therefore, $1/(\mu-\lambda)$.

12.33

a) Let X and Y denote the steady-state number of customers in the queueing system and in the queue, respectively. We have $Y = 0$ if $X = 0$ or 1, and $Y = X - 1$, if $X \geq 2$. Hence, by the FEF,

$$\begin{aligned} L_q &= \mathcal{E}(Y) = 0\cdot P_0 + 0\cdot P_1 + \sum_{n=2}^{\infty}(n-1)P_n = \sum_{n=2}^{\infty} nP_n - \sum_{n=2}^{\infty} P_n \\ &= (L - 1\cdot P_1) - (1 - P_0 - P_1) = L - (1 - P_0). \end{aligned}$$

b) From part (a) and Equations (12.24) and (12.20),

$$L_q = L - (1 - P_0) = \frac{\lambda}{\mu-\lambda} - \big(1 - (1-\lambda/\mu)\big) = \frac{\lambda^2}{\mu(\mu-\lambda)}.$$

c) Let X and Y denote the steady-state number of customers in the queueing system and in the queue, respectively. We have $Y = 0$ if $X = 0, 1, \dots, s$, and $Y = X - s$, if $X \geq s + 1$. Hence, by the FEF,

$$\begin{aligned} L_q &= \mathcal{E}(Y) = \sum_{n=0}^{s} 0\cdot P_n + \sum_{n=s+1}^{\infty}(n-s)P_n = \sum_{n=s+1}^{\infty} nP_n - s\sum_{n=s+1}^{\infty} P_n \\ &= L - \sum_{n=1}^{s} nP_n - s\left(1 - \sum_{n=0}^{s} P_n\right) = L - s - \sum_{n=0}^{s} nP_n + s\sum_{n=0}^{s} P_n \\ &= L - \left(s - \sum_{n=0}^{s}(s-n)P_n\right). \end{aligned}$$

Advanced Exercises

12.34 As noted in Exercise 12.24, we can assume that $s \leq N$. Also, recall that we are letting $\rho = \lambda/s\mu$ and $r = \lambda/\mu$.

a) From Equation (12.13) on page 700,

$$\lambda_n = \begin{cases} \lambda, & \text{if } 0 \leq n \leq N-1; \\ 0, & \text{if } n \geq N. \end{cases} \qquad \mu_n = \begin{cases} n\mu, & \text{if } 1 \leq n \leq s-1; \\ s\mu, & \text{if } n \geq s. \end{cases}$$

Hence,

$$\prod_{k=1}^{n}\frac{\lambda_{k-1}}{\mu_k} = \begin{cases} r^n/n!, & \text{if } 1 \le n \le s-1; \\ (r^s/s!)\rho^{n-s}, & \text{if } s \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases} = \begin{cases} r^n/n!, & \text{if } 1 \le n \le s-1; \\ r^n/s!\,s^{n-s}, & \text{if } s \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases}$$

Therefore,

$$\sum_{n=1}^{\infty}\left(\prod_{k=1}^{n}\frac{\lambda_{k-1}}{\mu_k}\right) = \sum_{n=1}^{s-1}\frac{r^n}{n!} + \sum_{n=s}^{N}\frac{r^n}{s!\,s^{n-s}} < \infty,$$

so that, by Proposition 12.3 on page 703, an $M/M/s/N$ queueing system has a steady-state distribution.
b) Referring again to Proposition 12.3 and to part (a), we find that

$$P_0 = \left(1 + \sum_{n=1}^{\infty}\left(\prod_{k=1}^{n}\frac{\lambda_{k-1}}{\mu_k}\right)\right)^{-1} = \left(\sum_{n=0}^{s-1}\frac{r^n}{n!} + \sum_{n=s}^{N}\frac{r^n}{s!\,s^{n-s}}\right)^{-1} = \left(\sum_{n=0}^{s-1}\frac{r^n}{n!} + \frac{r^s}{s!}\sum_{n=s}^{N}\rho^{n-s}\right)^{-1}$$

and, moreover,

$$P_n = \left(\prod_{k=1}^{n}\frac{\lambda_{k-1}}{\mu_k}\right)P_0 = \begin{cases} \dfrac{r^n}{n!}P_0, & \text{if } 1 \le n \le s-1; \\ \dfrac{r^n}{s!\,s^{n-s}}P_0, & \text{if } s \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases} \tag{$*$}$$

We note that, if $\rho = 1$ (i.e., $\lambda = s\mu$), then

$$P_0 = \left(\sum_{n=0}^{s-1}\frac{r^n}{n!} + (N-s+1)\frac{r^s}{s!}\right)^{-1}, \tag{$**$}$$

whereas, if $\rho \neq 1$ (i.e., $\lambda \neq s\mu$), then

$$P_0 = \left(\sum_{n=0}^{s-1}\frac{r^n}{n!} + \frac{r^s}{s!}\cdot\frac{1-\rho^{N-s+1}}{1-\rho}\right)^{-1}. \tag{$***$}$$

c) Assuming that $\lambda < s\mu$ (i.e., $\rho < 1$) and referring to Equation $(***)$, we find that, in the limit, as $N \to \infty$,

$$P_0 = \left(\sum_{n=0}^{s-1}\frac{r^n}{n!} + \frac{r^s}{s!\,(1-\rho)}\right)^{-1},$$

which agrees with Equation (12.22) on page 704. Furthermore, referring to Equation $(*)$, we find that, in the limit, as $N \to \infty$,

$$P_n = \begin{cases} \dfrac{r^n}{n!}P_0, & \text{if } 1 \le n \le s-1; \\ \dfrac{r^n}{s!\,s^{n-s}}P_0, & \text{if } n \ge s. \end{cases}$$

which agrees with Equation (12.23) on page 705.
d) Let X and Y denote the steady-state number of customers in the queueing system and in the queue, respectively. We have $Y = 0$ if $X = 0, 1, \ldots, s$, and $Y = X - s$, if $X \ge s+1$. Hence, by the FEF

and Equation ($*$),

$$L_q = \mathcal{E}(Y) = \sum_{n=0}^{s} 0 \cdot P_n + \sum_{n=s+1}^{\infty} (n-s)P_n = \sum_{n=s+1}^{N} (n-s)P_n$$
$$= \frac{P_0}{s!} \sum_{n=s+1}^{N} (n-s)\frac{r^n}{s^{n-s}} = \frac{r^s P_0}{s!} \sum_{n=s+1}^{N} (n-s)\rho^{n-s} = \frac{r^s P_0}{s!} \sum_{n=1}^{N-s} n\rho^n.$$

We now consider two cases.

Case 1: $\rho = 1$. In this case,

$$L_q = \frac{r^s P_0}{s!} \sum_{n=1}^{N-s} n\rho^n = \frac{r^s P_0}{s!} \sum_{n=1}^{N-s} n = \frac{(N-s)(N-s+1)r^s}{2s!} P_0,$$

where P_0 is given by Equation ($**$).

Case 2: $\rho \neq 1$. In this case,

$$L_q = \frac{r^s P_0}{s!} \sum_{n=1}^{N-s} n\rho^n = \frac{\rho r^s P_0}{s!} \sum_{n=1}^{N-s} n\rho^{n-1} = \frac{\rho r^s P_0}{s!} \frac{d}{d\rho} \left(\sum_{n=0}^{N-s} \rho^n \right) = \frac{\rho r^s P_0}{s!} \frac{d}{d\rho} \left(\frac{1-\rho^{N-s+1}}{1-\rho} \right)$$
$$= \frac{\rho r^s \left(1 - \rho^{N-s+1} - (1-\rho)(N-s+1)\rho^{N-s}\right)}{s!\,(1-\rho)^2} P_0,$$

where P_0 is given by Equation ($***$).

12.35

a) In steady state, the proportion of time that the queueing system is in state n equals P_n and, when in state n, the arrival rate is λ_n. Hence, $\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n$.

b) For the $M/M/1$ and $M/M/s$ queues, we have $\lambda_n = \lambda$ for all $n \geq 0$. Hence, from part (a),

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{\infty} \lambda P_n = \lambda \sum_{n=0}^{\infty} P_n = \lambda \cdot 1 = \lambda.$$

c) For the $M/M/1/n$ and $M/M/s/N$ queues, we have $\lambda_n = \lambda$ if $0 \leq n \leq N-1$, and $\lambda_n = 0$ otherwise. Hence, from part (a) and the fact that $P_n = 0$ for $n \geq N+1$,

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{N-1} \lambda P_n = \lambda \sum_{n=0}^{N-1} P_n = \left(1 - \sum_{n=N}^{\infty} P_n\right) \lambda = (1 - P_N)\lambda.$$

12.36

a) Let $\mathcal{W}$ and $\mathcal{W}_q$ denote the steady-state waiting times in the queueing system and in the queue, respectively, for each individual customer. Also, let S denote the service time for each individual customer. Because the service rate is μ, we have $\mathcal{E}(S) = 1/\mu$. Noting that $\mathcal{W} = \mathcal{W}_q + S$, we conclude that

$$W = \mathcal{E}(\mathcal{W}) = \mathcal{E}\big(\mathcal{W}_q + S\big) = \mathcal{E}\big(\mathcal{W}_q\big) + \mathcal{E}(S) = W_q + \frac{1}{\mu}.$$

b) From Little's formulas and part (a),

$$L = \bar{\lambda} W = \bar{\lambda} \left(W_q + \frac{1}{\mu} \right) = \bar{\lambda} W_q + \bar{\lambda} \cdot \frac{1}{\mu} = L_q + \frac{\bar{\lambda}}{\mu}.$$

c) Little's formulas are $L = \bar{\lambda} W$ and $L_q = \bar{\lambda} W_q$. We can use these two formulas and the two from parts (a) and (b) to solve for any three of the quantities L, L_q, W, and W_q in terms of the other one.
d) From Exercise 12.35(b), we know that $\bar{\lambda} = \lambda$. Hence,

$$L_q = L - \frac{\bar{\lambda}}{\mu} = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\lambda^2}{\mu(\mu - \lambda)},$$
$$W = \frac{L}{\bar{\lambda}} = \frac{\lambda/(\mu - \lambda)}{\lambda} = \frac{1}{\mu - \lambda},$$
$$W_q = W - \frac{1}{\mu} = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}.$$

These three results agree with those obtained in Exercises 12.29(b), 12.30(b), and 12.31(b), respectively.

12.37 From Exercise 12.35(b), we know that $\bar{\lambda} = \lambda = 6.9$. Recalling that $\mu = 7.4$ and $L = 13.8$, we obtain the following results.
a) The expected number of patients waiting to be treated is

$$L_q = L - \frac{\bar{\lambda}}{\mu} = 13.8 - 6.9/7.4 \approx 12.9.$$

b) The expected waiting time until treatment commences is

$$W_q = \frac{L_q}{\bar{\lambda}} = \frac{13.8 - 6.9/7.4}{6.9} \approx 1.9 \text{ hr},$$

or about 112 minutes.
c) The expected elapsed time from arrival at the emergency room until treatment is completed is

$$W = \frac{L}{\bar{\lambda}} = 13.8/6.9 = 2 \text{ hr},$$

or 120 minutes.

12.38 From Exercise 12.35(b), we know that $\bar{\lambda} = \lambda = 6.9$. Recalling that $\mu = 7.4$ and $L \approx 1.2$, we obtain the following results.
a) The expected number of patients waiting to be treated is

$$L_q = L - \frac{\bar{\lambda}}{\mu} \approx 1.2 - 6.9/7.4 \approx 0.27.$$

b) The expected waiting time until treatment commences is

$$W_q = \frac{L_q}{\bar{\lambda}} \approx \frac{1.2 - 6.9/7.4}{6.9} \approx 0.039 \text{ hr},$$

or about 2.3 minutes.
c) The expected elapsed time from arrival at the emergency room until treatment is completed is

$$W = \frac{L}{\bar{\lambda}} \approx 1.2/6.9 \approx 0.17 \text{ hr},$$

or about 10.4 minutes.

d) By staffing two doctors instead of one, a drastic reduction occurs in the expected queue size and waiting times, as shown in the following table. Times are in minutes.

Quantity	One doctor	Two doctors
L	13.8	1.2
L_q	12.9	0.27
W_q	112	2.3
W	120	10.4

12.39 From Exercises 12.21–23, we have the following (approximate) results.

Quantity	One teller	Two tellers	Three tellers
L	10	1.15	0.94
L_q	9.09	0.24	0.03

From Exercise 12.35(b), we know that $\bar{\lambda} = \lambda = 10$. Applying Little's formulas in the form $W = L/\bar{\lambda}$ and $W_q = L_q/\bar{\lambda}$, we get the following (approximate) results.
a) The steady-state expected time, in minutes, that a customer spends in the queueing system, W:

One teller	Two tellers	Three tellers
60	6.9	5.6

b) The steady-state expected time, in minutes, that a customer waits before being served, W_q:

One teller	Two tellers	Three tellers
54.5	1.4	0.2

c) Referring to parts (a) and (b), we see that, by using two tellers instead of one, a drastic reduction occurs both in the expected time that a customer spends in the drive-up station and in the expected time a customer spends in the queue. Adding a third teller does reduce these two expected times somewhat further but would probably be considered an unjustified expense.

12.40
a) From Exercises 12.35(c) and 12.25(b),

$$\bar{\lambda} = (1 - P_4)\lambda = \left(1 - \frac{16}{211}(1.5)^4\right) \cdot 3 \approx 1.85.$$

Therefore, from Exercises 12.36(b) and 12.25(c),

$$L_q = L - \frac{\bar{\lambda}}{\mu} \approx 2.76 - \frac{1.85}{2} = 1.835.$$

b) From Little's formulas and part (a),

$$W = \frac{L}{\bar{\lambda}} \approx \frac{2.76}{1.85} \approx 1.49 \text{ hr},$$

or about 89.5 minutes.

c) From Little's formulas and part (a),

$$W_q = \frac{L_q}{\bar{\lambda}} \approx \frac{1.835}{1.85} \approx 0.99 \text{ hr},$$

or about 59.5 minutes.

12.3 The Multivariate Normal Distribution

Basic Exercises

12.41

a) First we note that $\mathbf{\Sigma_X}$ is symmetric because $\operatorname{Cov}(X_j, X_i) = \operatorname{Cov}(X_i, X_j)$ for all i and j. Applying the bilinearity property of covariance, Equation (7.40) on page 366, we find that for all $x_1, \ldots, x_m$,

$$\begin{aligned} 0 \le \operatorname{Var}\left(\sum_{j=1}^{m} x_j X_j\right) &= \operatorname{Cov}\left(\sum_{j=1}^{m} x_j X_j, \sum_{k=1}^{m} x_k X_k\right) \\ &= \sum_{j=1}^{m}\sum_{k=1}^{m} x_j x_k \operatorname{Cov}(X_j, X_k) = \mathbf{x}'\mathbf{\Sigma_X}\mathbf{x}. \end{aligned}$$

Hence, $\mathbf{x}'\mathbf{\Sigma_X}\mathbf{x} \ge 0$ for all $\mathbf{x} \in \mathcal{R}^m$. Consequently, $\mathbf{\Sigma_X}$ is nonnegative definite.

b) Answers will vary. However, note that, if $X_1, \ldots, X_m$ are uncorrelated random variables with nonzero variances, then, in view of part (a), if $\mathbf{x} \ne \mathbf{0}$,

$$\mathbf{x}'\mathbf{\Sigma_X}\mathbf{x} = \sum_{j=1}^{m}\sum_{k=1}^{m} x_j x_k \operatorname{Cov}(X_j, X_k) = \sum_{k=1}^{m} x_k^2 \operatorname{Var}(X_k) > 0.$$

Hence, $\mathbf{\Sigma_X}$ is positive definite.

c) Answers will vary. However, note that, if $X_1, \ldots, X_m$ are constant random variables, then $\mathbf{\Sigma_X} = \mathbf{0}$ and, hence, $\mathbf{\Sigma_X}$ is not positive definite.

12.42

a) For $1 \le i \le k$ and $1 \le j \le m$, let $Y_i = [\mathbf{Y}]_i$, $a_i = [\mathbf{a}]_i$, and $b_{ij} = [\mathbf{B}]_{ij}$. Then $Y_i = a_i + \sum_{j=1}^{m} b_{ij} X_j$. Therefore,

$$[\boldsymbol{\mu}_\mathbf{Y}]_i = \mu_{Y_i} = \mathcal{E}\left(a_i + \sum_{j=1}^{m} b_{ij} X_j\right) = a_i + \sum_{j=1}^{m} b_{ij}\mathcal{E}(X_j) = a_i + \sum_{j=1}^{m} b_{ij}\mu_{X_j} = [\mathbf{a} + \mathbf{B}\boldsymbol{\mu}_\mathbf{X}]_i.$$

Hence, $\boldsymbol{\mu}_\mathbf{Y} = \mathbf{a} + \mathbf{B}\boldsymbol{\mu}_\mathbf{X}$. Also,

$$\begin{aligned} [\mathbf{\Sigma_Y}]_{ij} &= \operatorname{Cov}(Y_i, Y_j) = \operatorname{Cov}\left(a_i + \sum_{n=1}^{m} b_{in} X_n, a_j + \sum_{p=1}^{m} b_{jp} X_p\right) \\ &= \sum_{n=1}^{m}\sum_{p=1}^{m} b_{in} \operatorname{Cov}(X_n, X_p)\, b_{jp} = \sum_{n=1}^{m}\sum_{p=1}^{m} [\mathbf{B}]_{in}[\mathbf{\Sigma_X}]_{np}[\mathbf{B}']_{pj} \\ &= [\mathbf{B}\mathbf{\Sigma_X}\mathbf{B}']_{ij}. \end{aligned}$$

Hence, $\mathbf{\Sigma_Y} = \mathbf{B}\mathbf{\Sigma_X}\mathbf{B}'$.

b) As $Z_1, \ldots, Z_m$ are independent standard normal random variables, we have $\boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{0}_m$ (the $m \times 1$ zero vector) and $\boldsymbol{\Sigma}_{\mathbf{Z}} = \mathbf{I}_m$ (the $m \times m$ identity matrix). Applying part (a) with $k = m$, $\mathbf{Y} = \mathbf{X}$, and $\mathbf{X} = \mathbf{Z}$, we find that

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{a} + \mathbf{B}\boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{a} + \mathbf{B} \cdot \mathbf{0}_m = \mathbf{a}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{Z}}\mathbf{B}' = \mathbf{B}\mathbf{I}_m\mathbf{B} = \mathbf{B}\mathbf{B}'.$$

12.43

a) We have

$$\begin{aligned}\mathcal{E}(Y_1) &= \mathcal{E}(3 + 2X_1 - 4X_2 + X_3) = 3 + 2\mathcal{E}(X_1) - 4\mathcal{E}(X_2) + \mathcal{E}(X_3) \\ &= 3 + 2 \cdot (-1) - 4 \cdot 0 + 5 = 6\end{aligned}$$

and

$$\mathcal{E}(Y_2) = \mathcal{E}(-1 + X_2 - 4X_3) = -1 + \mathcal{E}(X_2) - 4\mathcal{E}(X_3) = -1 + 0 - 4 \cdot 5 = -21.$$

Also, recalling that $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$ and $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$, we get

$$\begin{aligned}\operatorname{Var}(Y_1) &= \operatorname{Cov}(Y_1, Y_1) = \operatorname{Cov}(3 + 2X_1 - 4X_2 + X_3, 3 + 2X_1 - 4X_2 + X_3) \\ &= 4\operatorname{Cov}(X_1, X_1) - 8\operatorname{Cov}(X_1, X_2) + 2\operatorname{Cov}(X_1, X_3) \\ &\quad - 8\operatorname{Cov}(X_2, X_1) + 16\operatorname{Cov}(X_2, X_2) - 4\operatorname{Cov}(X_2, X_3) \\ &\quad + 2\operatorname{Cov}(X_3, X_1) - 4\operatorname{Cov}(X_3, X_2) + \operatorname{Cov}(X_3, X_3) \\ &= 4 \cdot 10 - 8 \cdot (-5) + 2 \cdot 2 - 8 \cdot (-5) + 16 \cdot 16 - 4 \cdot (-3) + 2 \cdot 2 - 4 \cdot (-3) + 4 \\ &= 412, \\ \operatorname{Cov}(Y_1, Y_2) &= \operatorname{Cov}(3 + 2X_1 - 4X_2 + X_3, -1 + X_2 - 4X_3) \\ &= 2\operatorname{Cov}(X_1, X_2) - 8\operatorname{Cov}(X_1, X_3) - 4\operatorname{Cov}(X_2, X_2) + 16\operatorname{Cov}(X_2, X_3) \\ &\quad + \operatorname{Cov}(X_3, X_2) - 4\operatorname{Cov}(X_3, X_3) \\ &= 2 \cdot (-5) - 8 \cdot 2 - 4 \cdot 16 + 16 \cdot (-3) + (-3) - 4 \cdot 4 \\ &= -157,\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}(Y_2) &= \operatorname{Cov}(Y_2, Y_2) = \operatorname{Cov}(-1 + X_2 - 4X_3, -1 + X_2 - 4X_3) \\ &= \operatorname{Cov}(X_2, X_2) - 4\operatorname{Cov}(X_2, X_3) - 4\operatorname{Cov}(X_3, X_2) + 16\operatorname{Cov}(X_3, X_3) \\ &= 16 - 4 \cdot (-3) - 4 \cdot (-3) + 16 \cdot 4 \\ &= 104.\end{aligned}$$

Consequently,

$$\boldsymbol{\mu}_{\mathbf{Y}} = \begin{bmatrix} 6 \\ -21 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_{\mathbf{Y}} = \begin{bmatrix} 412 & -157 \\ -157 & 104 \end{bmatrix}.$$

b) From the given information,

$$\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} -1 \\ 0 \\ 5 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 10 & -5 & 2 \\ -5 & 16 & -3 \\ 2 & -3 & 4 \end{bmatrix}.$$

c) We note that

$$\mathbf{Y} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 & -4 & 1 \\ 0 & 1 & -4 \end{bmatrix}\mathbf{X}.$$

Hence, from Exercise 12.42(a) and part (b),

$$\boldsymbol{\mu}_{\mathbf{Y}} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 & -4 & 1 \\ 0 & 1 & -4 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} 3 \\ -20 \end{bmatrix} = \begin{bmatrix} 6 \\ -21 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = \begin{bmatrix} 2 & -4 & 1 \\ 0 & 1 & -4 \end{bmatrix} \begin{bmatrix} 10 & -5 & 2 \\ -5 & 16 & -3 \\ 2 & -3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -4 & 1 \\ 1 & -4 \end{bmatrix}$$

$$= \begin{bmatrix} 42 & -77 & 20 \\ -13 & 28 & -19 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -4 & 1 \\ 1 & -4 \end{bmatrix} = \begin{bmatrix} 412 & -157 \\ -157 & 104 \end{bmatrix}$$

d) Although the methods used in parts (a) and (c) are essentially equivalent, doing the calculation with matrices, as in part (b), is simpler and provides a more organized way of proceeding.

12.44 From Proposition 12.5 on page 717, multivariate normal random variables are nonsingular if and only if their covariance matrix is nonsingular. We use this criterion in solving parts (a)–(d).
a) The determinant of this matrix equals 1 and, hence, is nonzero. Thus, this matrix is nonsingular, which implies that the multivariate normal random variables with this covariance matrix are nonsingular.
b) The determinant of this matrix equals 0 and, hence, the matrix is singular. Therefore, the multivariate normal random variables with this covariance matrix are singular.
c) The determinant of this matrix equals 49 and, hence, is nonzero. Thus, this matrix is nonsingular, which implies that the multivariate normal random variables with this covariance matrix are nonsingular.
d) The rows of this matrix are all multiples of one another. In particular, then, the matrix is singular. Therefore, the multivariate normal random variables with this covariance matrix are singular.

12.45 Let $Z_1, \dots, Z_m$ be independent standard normal random variables and let $\mathbf{R}$ be the $m \times m$ positive definite matrix such that $\mathbf{R}^2 = \boldsymbol{\Sigma}$. Set $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{RZ}$. From the argument leading to Definition 12.4, we know that $Y_1, \dots, Y_m$ are marginally normal random variables with $\boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{RR}' = \mathbf{R}^2 = \boldsymbol{\Sigma}$ and, furthermore, that

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{m/2}(\det \boldsymbol{\Sigma}_{\mathbf{Y}})^{1/2}}\, e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}})'\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}})} = \frac{1}{(2\pi)^{m/2}(\det \boldsymbol{\Sigma})^{1/2}}\, e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Thus, $Y_1, \dots, Y_m$ and $X_1, \dots, X_m$ have the same joint PDFs, which implies that they have the same univariate and bivariate marginals. Hence, the univariate marginal distributions of $X_1, \dots, X_m$ are all normal and, moreover, $\boldsymbol{\mu}_{\mathbf{X}} = \boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}_{\mathbf{Y}} = \boldsymbol{\Sigma}$.

12.46
a) Suppose that X has a nonsingular univariate normal distribution in the sense of Definition 12.4. Then

$$f_X(x) = \frac{1}{(2\pi)^{1/2}(\det \boldsymbol{\Sigma})^{1/2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

where $\boldsymbol{\mu}$ is a 1×1 vector and $\boldsymbol{\Sigma}$ is a 1×1 positive definite matrix. Thus, we can write $\boldsymbol{\mu} = [\mu]$ and $\boldsymbol{\Sigma} = [\sigma^2]$, where μ is a real number and σ is a positive real number. Therefore,

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = [x - \mu]'[\sigma^{-2}][x - \mu] = (x - \mu)\sigma^{-2}(x - \mu) = (x - \mu)^2/\sigma^2.$$

Noting that $\det \boldsymbol{\Sigma} = \sigma^2$, we conclude that

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\left(\sigma^2\right)^{1/2}} e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

Hence, X has a univariate normal distribution in the sense of Definition 8.6.

Conversely, suppose that X has a univariate normal distribution in the sense of Definition 8.6. Then

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2},$$

where μ and $\sigma > 0$ are real constants. Now, let $\boldsymbol{\mu} = [\mu]$ and $\boldsymbol{\Sigma} = [\sigma^2]$. We see that

$$(x-\mu)^2/\sigma^2 = (x-\mu)\sigma^{-2}(x-\mu) = [x-\mu]'[\sigma^{-2}][x-\mu] = (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}).$$

Noting that $\sigma^2 = \det \boldsymbol{\Sigma}$, we conclude that

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\left(\sigma^2\right)^{1/2}} e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{(2\pi)^{1/2}(\det \boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

where $\boldsymbol{\mu}$ is a 1×1 vector and $\boldsymbol{\Sigma}$ is a 1×1 positive definite matrix. Hence, X has a nonsingular univariate normal distribution in the sense of Definition 12.4.

b) Suppose that X and Y have a nonsingular bivariate normal distribution in the sense of Definition 12.4. Then

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^{2/2}(\det \boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

where $\boldsymbol{\mu}$ is a 2×1 vector and $\boldsymbol{\Sigma}$ is a 2×2 positive definite matrix. Let $\mathbf{X} = [X, Y]'$ and $\mathbf{x} = [x, y]'$. From Exercise 12.45, we know that

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} \operatorname{Cov}(X, X) & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(Y, X) & \operatorname{Cov}(Y, Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix},$$

where $\rho = \rho(X, Y)$. In particular, we see that

$$\det \boldsymbol{\Sigma} = \sigma_X^2\sigma_Y^2\left(1-\rho^2\right) \qquad \text{and} \qquad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}.$$

Therefore,

$$\begin{aligned}
&(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\
&\quad = \frac{1}{\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)}\left(\sigma_Y^2(x-\mu_X)^2 - 2\rho\sigma_X\sigma_Y(x-\mu_X)(y-\mu_Y) + \sigma_X^2(y-\mu_Y)^2\right) \\
&\quad = \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}.
\end{aligned}$$

Hence,

$$f_{X,Y}(x, y) = \frac{1}{2\pi\left(\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)\right)^{1/2}}\, e^{-\frac{1}{2}Q(x,y)} = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}Q(x,y)},$$

where

$$Q(x, y) = \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}.$$

Consequently, X and Y are bivariate normal random variables in the sense of Definition 10.4.

Conversely, suppose that X and Y are bivariate normal random variables in the sense of Definition 10.4. Then

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}Q(x,y)},$$

where

$$Q(x, y) = \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}.$$

Let

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Note that

$$\det\boldsymbol{\Sigma} = \sigma_X^2\sigma_Y^2\left(1-\rho^2\right) \qquad \text{and} \qquad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}$$

and that

$$\begin{aligned} Q(x, y) &= \frac{1}{\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)}\left(\sigma_Y^2(x-\mu_X)^2 - 2\rho\sigma_X\sigma_Y(x-\mu_X)(y-\mu_Y) + \sigma_X^2(y-\mu_Y)^2\right) \\ &= (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}). \end{aligned}$$

Hence,

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^{2/2}(\det\boldsymbol{\Sigma})^{1/2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.$$

Consequently, X and Y have a nonsingular bivariate normal distribution in the sense of Definition 12.4.

12.47

a) Let

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} \qquad \text{and} \qquad \mathbf{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}.$$

Then $\mathbf{X}'\mathbf{t} = t_1X_1 + \cdots + t_mX_m$ so that

$$M_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = \mathcal{E}\left(e^{t_1X_1+\cdots+t_mX_m}\right) = \mathcal{E}\left(e^{\mathbf{X}'\mathbf{t}}\right) = M_{\mathbf{X}}(\mathbf{t}).$$

b) From part (a),

$$M_{\mathbf{a}+\mathbf{B}\mathbf{X}}(\mathbf{t}) = \mathcal{E}\left(e^{(\mathbf{a}+\mathbf{B}\mathbf{X})'\mathbf{t}}\right) = \mathcal{E}\left(e^{\mathbf{a}'\mathbf{t}+\mathbf{X}'\mathbf{B}'\mathbf{t}}\right) = e^{\mathbf{a}'\mathbf{t}}\mathcal{E}\left(e^{\mathbf{X}'\mathbf{B}'\mathbf{t}}\right) = e^{\mathbf{a}'\mathbf{t}}\mathcal{E}\left(e^{\mathbf{X}'(\mathbf{B}'\mathbf{t})}\right) = e^{\mathbf{a}'\mathbf{t}}M_{\mathbf{X}}(\mathbf{B}'\mathbf{t}).$$

12.48
a) We use matrix notation. The Jacobian determinant of the transformation $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ is

$$J(\mathbf{x}) = \det \boldsymbol{\Sigma}^{-1/2} = (\det \boldsymbol{\Sigma})^{-1/2}.$$

Solving $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ for $\mathbf{x}$, we get the inverse transformation $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}$. Now, in terms of the inverse transformation,

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \Big(\big(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}\big) - \boldsymbol{\mu}\Big)' \boldsymbol{\Sigma}^{-1} \Big(\big(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}\big) - \boldsymbol{\mu}\Big) = \big(\boldsymbol{\Sigma}^{1/2}\mathbf{z}\big)'\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{\Sigma}^{1/2}\mathbf{z}\big)$$
$$= \mathbf{z}'\big(\boldsymbol{\Sigma}^{1/2}\big)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{z} = \mathbf{z}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{z} = \mathbf{z}'\mathbf{z}.$$

Hence, by the multivariate transformation theorem,

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{|J(\mathbf{x})|} f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(\det \boldsymbol{\Sigma})^{-1/2}} \frac{1}{(2\pi)^{m/2}(\det \boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} = \frac{1}{(2\pi)^{m/2}} e^{-\mathbf{z}'\mathbf{z}},$$

which, as we have seen, is the joint PDF of m independent standard normal random variables. Consequently, $Z_1, \dots, Z_m$ are independent standard normal random variables.
b) We can write $\mathbf{Z} = \mathbf{a} + \mathbf{B}\mathbf{X}$, where $\mathbf{a} = -\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\mathbf{B} = \boldsymbol{\Sigma}^{-1/2}$. Note that $\mathbf{B}$ is a symmetric matrix so that $\mathbf{B}' = \mathbf{B}$ and $\mathbf{a}' = \big(-\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}\big)' = -\boldsymbol{\mu}'\mathbf{B}$. From Exercise 12.47(b) and Proposition 12.4 on page 716,

$$M_{\mathbf{Z}}(\mathbf{t}) = M_{\mathbf{a}+\mathbf{B}\mathbf{X}}(\mathbf{t}) = e^{\mathbf{a}'\mathbf{t}} M_{\mathbf{X}}(\mathbf{B}'\mathbf{t}) = e^{-\boldsymbol{\mu}'\mathbf{B}\mathbf{t}} M_{\mathbf{X}}(\mathbf{B}\mathbf{t}) = e^{-\boldsymbol{\mu}'\mathbf{B}\mathbf{t}} e^{\boldsymbol{\mu}'(\mathbf{B}\mathbf{t}) + \frac{1}{2}(\mathbf{B}\mathbf{t})'\boldsymbol{\Sigma}(\mathbf{B}\mathbf{t})}$$
$$= e^{\frac{1}{2}\mathbf{t}'\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}\mathbf{t}} = e^{\frac{1}{2}\mathbf{t}'\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}\mathbf{t}} = e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\mathbf{t}} = e^{\frac{1}{2}\mathbf{t}'\mathbf{t}}.$$

Referring now to Lemma 12.1 on page 716, we conclude, in view of the uniqueness property of MGFs, that $Z_1, \dots, Z_m$ are independent standard normal random variables.
c) We can write $\mathbf{Z} = \mathbf{a} + \mathbf{B}\mathbf{X}$, where $\mathbf{a} = -\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\mathbf{B} = \boldsymbol{\Sigma}^{-1/2}$. Applying Proposition 12.7, we conclude that $\mathbf{Z}$ has the multivariate normal distribution with parameters

$$\mathbf{a} + \mathbf{B}\boldsymbol{\mu} = -\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = \mathbf{0} \qquad \text{and} \qquad \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}' = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \mathbf{I}.$$

Therefore, the joint MGF of $Z_1, \dots, Z_m$ is

$$M(Z)\mathbf{t} = e^{\mathbf{0}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\mathbf{I}\mathbf{t}} = e^{\frac{1}{2}\mathbf{t}'\mathbf{t}}.$$

Referring to Lemma 12.1, we see that this MGF is that of independent standard normal random variables. Hence, by the uniqueness property of MGFs, the random variables $Z_1, \dots, Z_m$ are independent standard normal.

12.49 In the univariate case, $\mathbf{Z} = [Z]$, $\boldsymbol{\Sigma} = [\sigma^2]$, $\mathbf{X} = [X]$, and $\boldsymbol{\mu} = [\mu]$. Hence, in that case, Equation (12.41) becomes

$$Z = \big(\sigma^2\big)^{-1/2}(X - \mu) = \frac{X - \mu}{\sigma}.$$

Consequently, Equation (12.41) is the multivariate analogue of standardizing.

12.50 Let $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$. From Exercise 12.48, we know that $Z_1, \dots, Z_m$ are independent standard normal random variables. Furthermore, because $\boldsymbol{\Sigma}^{-1/2}$ is symmetric,

$$(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) = \Big(\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\Big)' \Big(\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\Big)$$
$$= \mathbf{Z}'\mathbf{Z} = Z_1^2 + \cdots + Z_m^2.$$

Referring now to Proposition 8.13 on page 466 and the second bulleted item on page 537, we conclude that $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has the chi-square distribution with m degrees of freedom.

12.51 Choose an orthogonal matrix $\mathbf{B}$ such that $\mathbf{B}\mathbf{\Sigma_X}\mathbf{B}'$ is a diagonal matrix, say, $\mathbf{D}$. Let $\mathbf{Y} = \mathbf{BX}$. From Proposition 12.7 on page 719, we have

$$\mathbf{Y} \sim \mathcal{N}_m\left(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\mathbf{\Sigma_X}\mathbf{B}'\right) = \mathcal{N}_m\left(\mathbf{B}\boldsymbol{\mu}, \mathbf{D}\right).$$

Now, let $U_1, \ldots, U_m$ be independent normal random variables with $U_j \sim \mathcal{N}\left([\mathbf{B}\boldsymbol{\mu}]_j, [\mathbf{D}]_{jj}\right)$. Then, because $\mathbf{D}$ is a diagonal matrix, it follows from Equation (12.30) on page 713 that $\mathbf{Y}$ and $\mathbf{U}$ have the same mean vector and covariance matrix. Because the joint MGF of multivariate normal random variables is determined by the mean vector and covariance matrix, we conclude that $\mathbf{Y}$ and $\mathbf{U}$ have the same joint MGF and, hence, the same joint distribution. Thus, in particular, $Y_1, \ldots, Y_m$ are independent random variables.

12.52 Because $(\mathbf{BB}')' = (\mathbf{B}')'\mathbf{B}' = \mathbf{BB}'$, we see that $\mathbf{BB}'$ is symmetric. Suppose that $\mathbf{B}$ is an $m \times n$ matrix. For $\mathbf{x} \in \mathcal{R}^m$,

$$\mathbf{x}'(\mathbf{BB}')\mathbf{x} = (\mathbf{x}'\mathbf{B})(\mathbf{B}'\mathbf{x}) = (\mathbf{x}'\mathbf{B})(\mathbf{x}'\mathbf{B})' = \sum_{j=1}^{n}[\mathbf{x}'\mathbf{B}]_j^2 \geq 0.$$

Hence, $\mathbf{BB}'$ is nonnegative definite.

12.53

a) Let $a_1, \ldots, a_m$ be real numbers. Then

$$\sum_{i=1}^{m} a_i X_i = \sum_{i=1}^{m} a_i \left(\mu_i + \sum_{j=1}^{n} b_{ij} Z_j\right) = \sum_{i=1}^{m} a_i \mu_i + \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_i b_{ij}\right) Z_j.$$

Referring to Proposition 9.14 on page 540, we conclude that $\sum_{i=1}^{m} a_i X_i$ is normally distributed. Applying now Proposition 12.6 on page 718, we see that $X_1, \ldots, X_m$ are multivariate normal random variables.

b) We have

$$\mathcal{E}(X_i) = \mathcal{E}\left(\mu_i + \sum_{j=1}^{n} b_{ij} Z_j\right) = \mu_i + \sum_{j=1}^{n} b_{ij}\mathcal{E}(Z_j) = \mu_i.$$

Furthermore,

$$\operatorname{Cov}(X_i, X_j) = \operatorname{Cov}\left(\mu_i + \sum_{p=1}^{n} b_{ip} Z_p, \mu_j + \sum_{q=1}^{n} b_{jq} Z_q\right) = \sum_{p=1}^{n}\sum_{q=1}^{n} b_{ip} b_{jq} \operatorname{Cov}(Z_p, Z_q) = \sum_{p=1}^{n} b_{ip} b_{jp}.$$

Letting

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \ldots & b_{mn} \end{bmatrix},$$

we see that $\boldsymbol{\mu}_\mathbf{X} = \boldsymbol{\mu}$ and $\mathbf{\Sigma_X} = \mathbf{BB}'$.

c) Referring to part (b) and Proposition 12.5 on page 717, we see that a necessary and sufficient condition for $X_1, \ldots, X_m$ to be singular is that $\mathbf{BB}'$ be a singular matrix. We note in passing that, if $m = n$, then $\mathbf{BB}'$ is singular if and only if $\mathbf{B}$ is singular. Consequently, when $m = n$, $X_1, \ldots, X_m$ are singular if and only if $\mathbf{B}$ is singular.

12.54 Let $\mathbf{t}_i$ be the $m \times 1$ vector whose ith entry is t and has all other entries 0. Applying the multivariate analogue of Proposition 11.7 on page 645 and referring to Definition 12.5 on page 717, we get

$$M_{X_i}(t) = M_\mathbf{X}(\mathbf{t}_i) = e^{\boldsymbol{\mu}'\mathbf{t}_i + \frac{1}{2}\mathbf{t}_i'\mathbf{\Sigma}\mathbf{t}_i} = e^{[\boldsymbol{\mu}]_i t + \frac{1}{2}[\mathbf{\Sigma}]_{ii} t^2}.$$

Therefore, $X_i \sim \mathcal{N}([\boldsymbol{\mu}]_i, [\boldsymbol{\Sigma}]_{ii})$ and, from this result, we also conclude that $\boldsymbol{\mu}_{\mathbf{X}} = \boldsymbol{\mu}$. Now, let

$$Q(\mathbf{t}) = \boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t},$$

so that $M_{\mathbf{X}}(\mathbf{t}) = e^{Q(\mathbf{t})}$. We have

$$Q(\mathbf{t}) = \sum_{k=1}^{m} [\boldsymbol{\mu}]_k t_k + \frac{1}{2}\sum_{k=1}^{m}\sum_{\ell=1}^{m} [\boldsymbol{\Sigma}]_{k\ell} t_k t_\ell.$$

Using the symmetry of $\boldsymbol{\Sigma}$, we find that

$$\frac{\partial Q}{\partial t_j}(\mathbf{t}) = [\boldsymbol{\mu}]_j + \sum_{k=1}^{m} [\boldsymbol{\Sigma}]_{kj} t_k \qquad \text{and} \qquad \frac{\partial^2 Q}{\partial t_i \partial t_j}(\mathbf{t}) = [\boldsymbol{\Sigma}]_{ij}.$$

Furthermore,

$$\frac{\partial M_{\mathbf{X}}}{\partial t_j}(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{t})\frac{\partial Q}{\partial t_j}(\mathbf{t})$$

and, hence,

$$\frac{\partial^2 M_{\mathbf{X}}}{\partial t_i \partial t_j}(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{t})\frac{\partial^2 Q}{\partial t_i \partial t_j}(\mathbf{t}) + M_{\mathbf{X}}(\mathbf{t})\frac{\partial Q}{\partial t_i}(\mathbf{t})\frac{\partial Q}{\partial t_j}(\mathbf{t}).$$

Applying now the multivariate analogue of Proposition 11.6 on page 643, we find that

$$\begin{aligned}\mathcal{E}\big(X_i X_j\big) &= \frac{\partial^2 M_{\mathbf{X}}}{\partial t_i \partial t_j}(\mathbf{0}) = M_{\mathbf{X}}(\mathbf{0})\frac{\partial^2 Q}{\partial t_i \partial t_j}(\mathbf{0}) + M_{\mathbf{X}}(\mathbf{0})\frac{\partial Q}{\partial t_i}(\mathbf{0})\frac{\partial Q}{\partial t_j}(\mathbf{0}) \\ &= [\boldsymbol{\Sigma}]_{ij} + [\boldsymbol{\mu}]_i[\boldsymbol{\mu}]_j = [\boldsymbol{\Sigma}]_{ij} + \mathcal{E}(X_i)\,\mathcal{E}\big(X_j\big).\end{aligned}$$

Consequently, $\operatorname{Cov}\big(X_i, X_j\big) = [\boldsymbol{\Sigma}]_{ij}$, which means that $\boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}$.

12.55 Referring to Proposition 12.8 on page 719, we get the following results:

$$X_1 \sim \mathcal{N}(2,3), \quad X_2 \sim \mathcal{N}(1,6), \quad X_3 \sim \mathcal{N}(0,6), \quad X_4 \sim \mathcal{N}(-1,2),$$

$$(X_1, X_2) \sim \mathcal{N}_2\left(\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}3 & 0\\0 & 6\end{bmatrix}\right), \quad (X_1, X_3) \sim \mathcal{N}_2\left(\begin{bmatrix}2\\0\end{bmatrix}, \begin{bmatrix}3 & -1\\-1 & 6\end{bmatrix}\right),$$

$$(X_1, X_4) \sim \mathcal{N}_2\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}\right), \quad (X_2, X_3) \sim \mathcal{N}_2\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}6 & 3\\3 & 6\end{bmatrix}\right)$$

$$(X_2, X_4) \sim \mathcal{N}_2\left(\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}6 & 2\\2 & 2\end{bmatrix}\right), \quad (X_3, X_4) \sim \mathcal{N}_2\left(\begin{bmatrix}0\\-1\end{bmatrix}, \begin{bmatrix}6 & 1\\1 & 2\end{bmatrix}\right)$$

$$(X_1, X_2, X_3) \sim \mathcal{N}_3\left(\begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}3 & 0 & -1\\0 & 6 & 3\\-1 & 3 & 6\end{bmatrix}\right), \quad (X_1, X_2, X_4) \sim \mathcal{N}_3\left(\begin{bmatrix}2\\1\\-1\end{bmatrix}, \begin{bmatrix}3 & 0 & 0\\0 & 6 & 2\\0 & 2 & 2\end{bmatrix}\right)$$

$$(X_1, X_3, X_4) \sim \mathcal{N}_3\left(\begin{bmatrix}2\\0\\-1\end{bmatrix}, \begin{bmatrix}3 & -1 & 0\\-1 & 6 & 1\\0 & 1 & 2\end{bmatrix}\right), \quad (X_2, X_3, X_4) \sim \mathcal{N}_3\left(\begin{bmatrix}1\\0\\-1\end{bmatrix}, \begin{bmatrix}6 & 3 & 2\\3 & 6 & 1\\2 & 1 & 2\end{bmatrix}\right)$$

Advanced Exercises

12.56 Let $\mathbf{X}_{\mathcal{P}}$ denote the column vector consisting of $X_{i_1}, \ldots, X_{i_m}$ and let $\mathbf{B}$ be the matrix obtained by permuting the rows of the $m \times m$ identity matrix according to the permutation $\mathcal{P}$. Then $\mathbf{X}_{\mathcal{P}} = \mathbf{BX}$.
a) From Proposition 12.7 on page 719, we know that

$$\mathbf{X}_{\mathcal{P}} = \mathbf{BX} \sim \mathcal{N}_m(\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}').$$

From linear algebra, $\mathbf{B}\boldsymbol{\mu}$ is the vector obtained by permuting the rows of $\boldsymbol{\mu}$ according to $\mathcal{P}$, and $\mathbf{B\Sigma B}'$ is the matrix obtained by permuting both the rows and columns of $\boldsymbol{\Sigma}$ according to $\mathcal{P}$. Thus, $\mathbf{B}\boldsymbol{\mu} = \boldsymbol{\mu}^*$ and $\mathbf{B\Sigma B}' = \boldsymbol{\Sigma}^*$ and, hence, $\mathbf{X}_{\mathcal{P}} \sim \mathcal{N}_m(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$.
b) We note that $\det \mathbf{B} = \pm 1$ and, hence, $\mathbf{B}$ is nonsingular. Now, from part (a), $\boldsymbol{\Sigma}^* = \mathbf{B\Sigma B}'$ and, consequently,

$$\det \boldsymbol{\Sigma}^* = \det(\mathbf{B\Sigma B}') = (\det \mathbf{B})(\det \boldsymbol{\Sigma})(\det \mathbf{B}') = (\det \mathbf{B})(\det \boldsymbol{\Sigma})(\det \mathbf{B}) = (\pm 1)^2 \det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}.$$

Because $X_1, \ldots, X_m$ are nonsingular, $\det \boldsymbol{\Sigma} \neq 0$ and, hence, $\det \boldsymbol{\Sigma}^* \neq 0$; thus, $\boldsymbol{\Sigma}^*$ is nonsingular. It now follows from Proposition 12.5 on page 717 that $X_{i_1}, \ldots, X_{i_m}$ are nonsingular multivariate random variables.

12.57
a) Write $\mathbf{t}' = [\mathbf{t}_1', \mathbf{t}_2']$, where $\mathbf{t}_1$ is $j \times 1$ and $\mathbf{t}_2$ is $k \times 1$. Now,

$$\boldsymbol{\mu}'\mathbf{t} = \boldsymbol{\mu}_1'\mathbf{t}_1 + \boldsymbol{\mu}_2'\mathbf{t}_2$$

and

$$\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} = \mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_2'\boldsymbol{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2 = \mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + 2\mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2,$$

where, in the last equality, we have used the fact that $\boldsymbol{\Sigma}_{21}' = \boldsymbol{\Sigma}_{12}$. Hence,

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} = e^{\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1} e^{\mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2} e^{\boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2}.$$

Consequently, from Exercise 11.30,

$$M_{\mathbf{X}_1}(\mathbf{t}_1) = M_{\mathbf{X}}(\mathbf{t}_1, \mathbf{0}) = e^{\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1}.$$

It now follows from the uniqueness property of MGFs that $\mathbf{X}_1 \sim \mathcal{N}_j(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.
b) Let j be a positive integer less than m. Consider a subcollection of $X_1, \ldots, X_m$ consisting of j random variables. Call the corresponding random vector $\mathbf{X}_j$. To determine the (marginal) distribution of $\mathbf{X}_j$, we first permute the components of $\mathbf{X}$ so that $\mathbf{X}_j$ comes first, applying as well the corresponding permutation to the mean vector and covariance matrix of $\mathbf{X}$. From Exercise 12.56(a), the resulting random vector, $\mathbf{Y}$, is m-dimensional multivariate normal. Next, we partition $\mathbf{Y}$ (and $\boldsymbol{\mu}_{\mathbf{Y}}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}$) so that $\mathbf{Y}_1 = \mathbf{X}_j$. Then, from part (a), $\mathbf{X}_j \sim \mathcal{N}_j(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_{11}$ are the mean vector and covariance matrix, respectively, of $\mathbf{Y}_1$.
c) From part (a),

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1} e^{\mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2} e^{\boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2}$$

and $M_{\mathbf{X}_1}(\mathbf{t}_1) = e^{\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1}$. Likewise, $M_{\mathbf{X}_2}(\mathbf{t}_2) = e^{\boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2}$. Now, from the multivariate analogue of Propostion 11.9 on page 646, $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2)$, which is the case if and only if

$$e^{\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1} e^{\mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2} e^{\boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2} = e^{\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1} e^{\boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2}.$$

Hence, we see that $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 = 0$ for all $\mathbf{t}_1$ and $\mathbf{t}_2$, which is true if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

12.58
a) Let $\mathbf{U} = \mathbf{X}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2$, $\mathbf{V} = \mathbf{X}_2$, and $\mathbf{W}' = [\mathbf{U}', \mathbf{V}']$. Then,

$$\mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} = \mathbf{B}\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \mathbf{B}\mathbf{X},$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_j & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_k \end{bmatrix}.$$

From Proposition 12.7 on page 719, $\mathbf{W} \sim \mathcal{N}_m\left(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\mathbf{\Sigma}\mathbf{B}'\right)$. However,

$$\mathbf{B}\boldsymbol{\mu} = \begin{bmatrix} \mathbf{I}_j & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_k \end{bmatrix}\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$

and

$$\mathbf{B}\mathbf{\Sigma}\mathbf{B}' = \begin{bmatrix} \mathbf{I}_j & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_k \end{bmatrix}\begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{I}_j & \mathbf{0} \\ -\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} & \mathbf{I}_k \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{bmatrix},$$

where we have used the fact that $\mathbf{\Sigma}_{12}' = \mathbf{\Sigma}_{21}$. Referring to Exercise 12.57, we see that $\mathbf{U}$ and $\mathbf{V}$ $(= \mathbf{X}_2)$ are independent random variables and that

$$\mathbf{U} \sim \mathcal{N}_j\left(\boldsymbol{\mu}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right).$$

It now follows that

$$\mathbf{X}_{1\,|\mathbf{X}_2=\mathbf{x}_2} = \left(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2 + \mathbf{U}\right)_{|\mathbf{X}_2=\mathbf{x}_2} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{x}_2 + \mathbf{U} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{x}_2 + \mathbf{I}_j\mathbf{U}.$$

Applying Proposition 12.7 again, we conclude that

$$\begin{aligned}\mathbf{X}_{1\,|\mathbf{X}_2=\mathbf{x}_2} &\sim \mathcal{N}_j\left(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{x}_2 + \mathbf{I}_j\left(\boldsymbol{\mu}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2\right), \mathbf{I}_j\left(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right)\mathbf{I}_j'\right) \\ &= \mathcal{N}_j\left(\boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right).\end{aligned}$$

b) Let j and k be positive integers with $j + k \leq m$. Consider a subcollection of $X_1, \ldots, X_m$ consisting of j random variables and another (mutually exclusive) subcollection of $X_1, \ldots, X_m$ consisting of k random variables. Call the corresponding random vectors $\mathbf{X}_j$ and $\mathbf{X}_k$, respectively. To determine the conditional distribution of $\mathbf{X}_j$ given $\mathbf{X}_k = \mathbf{x}_k$, we first permute the components of $\mathbf{X}$ so that $\mathbf{X}_j$ comes first, $\mathbf{X}_k$ comes second, and the remaining variables come third, applying as well the corresponding permutation to the mean vector and covariance matrix of $\mathbf{X}$. From Exercise 12.56(a), the resulting random vector, $\mathbf{Y}$, is m-dimensional multivariate normal. Next, we consider the random vector, $\mathbf{W}$, consisting of the first $j + k$ components of $\mathbf{Y}$, which, from Exercise 12.57(a) is $(j + k)$-dimensional multivariate normal. We can then apply part (a) to obtain the conditional distribution of $\mathbf{X}_j$ given $\mathbf{X}_k = \mathbf{x}_k$.
c) In the bivariate normal case,

$$\mathbf{X} = \begin{bmatrix} X \\ Y \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \qquad \text{and} \qquad \mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Then, from part (a),

$$\begin{aligned}X_{|Y=y} &\sim \mathcal{N}_1\left(\mu_X + \rho\sigma_X\sigma_Y\left(\sigma_Y^2\right)^{-1}(y-\mu_Y),\, \sigma_X^2 - \rho\sigma_X\sigma_Y\left(\sigma_Y^2\right)^{-1}\rho\sigma_X\sigma_Y\right)\\ &= \mathcal{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y),\, \sigma_X^2\left(1-\rho^2\right)\right),\end{aligned}$$

which is Proposition 10.10(e).

12.59

a) Here we have the following partition:

$$\mathbf{X} = \begin{bmatrix} [X_1] \\ \begin{bmatrix} X_2 \\ X_3 \\ X_4 \end{bmatrix} \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} [2] \\ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} [3] & [0 \quad -1 \quad 0] \\ \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} & \begin{bmatrix} 6 & 3 & 2 \\ 3 & 6 & 1 \\ 2 & 1 & 2 \end{bmatrix} \end{bmatrix}.$$

Simple calculations show that

$$\boldsymbol{\Sigma}_{22}^{-1} = \frac{1}{36}\begin{bmatrix} 11 & -4 & -9 \\ -4 & 8 & 0 \\ -9 & 0 & 27 \end{bmatrix}.$$

Therefore,

$$\begin{aligned}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) &= 2 + \frac{1}{36}[0 \quad -1 \quad 0]\begin{bmatrix} 11 & -4 & -9 \\ -4 & 8 & 0 \\ -9 & 0 & 27 \end{bmatrix}\begin{bmatrix} x_2 - 1 \\ x_3 - 0 \\ x_4 + 1 \end{bmatrix}\\ &= 2 + \frac{1}{36}[4 \quad -8 \quad 0]\begin{bmatrix} x_2 - 1 \\ x_3 - 0 \\ x_4 + 1 \end{bmatrix} = 2 + \frac{1}{9}(x_2 - 1) - \frac{2}{9}x_3\\ &= \frac{17}{9} + \frac{1}{9}x_2 - \frac{2}{9}x_3\end{aligned}$$

and

$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = 3 - \frac{1}{36}[0 \quad -1 \quad 0]\begin{bmatrix} 11 & -4 & -9 \\ -4 & 8 & 0 \\ -9 & 0 & 27 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = 3 - \frac{2}{9} = \frac{25}{9}.$$

Consequently, from Exercise 12.58(a),

$$X_{1|X_2=x_2, X_3=x_3, X_4=x_4} \sim \mathcal{N}\left(\frac{17}{9} + \frac{1}{9}x_2 - \frac{2}{9}x_3,\, \frac{25}{9}\right).$$

b) Here we have the following partition:

$$\mathbf{X} = \begin{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \\ \begin{bmatrix} X_3 \\ X_4 \end{bmatrix} \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 0 \\ -1 \end{bmatrix} \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} & \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix} \\ \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix} & \begin{bmatrix} 6 & 1 \\ 1 & 2 \end{bmatrix} \end{bmatrix}.$$

Simple calculations show that

$$\boldsymbol{\Sigma}_{22}^{-1} = \frac{1}{11}\begin{bmatrix} 2 & -1 \\ -1 & 6 \end{bmatrix}.$$

Therefore,

$$\begin{aligned}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) &= \begin{bmatrix}2\\1\end{bmatrix} + \frac{1}{11}\begin{bmatrix}-1 & 0\\ 3 & 2\end{bmatrix}\begin{bmatrix}2 & -1\\ -1 & 6\end{bmatrix}\begin{bmatrix}x_3 - 0\\ x_4 + 1\end{bmatrix}\\ &= \begin{bmatrix}2\\1\end{bmatrix} + \frac{1}{11}\begin{bmatrix}-2 & 1\\ 4 & 9\end{bmatrix}\begin{bmatrix}x_3 - 0\\ x_4 + 1\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix} + \frac{1}{11}\begin{bmatrix}-2x_3 + x_4 + 1\\ 4x_3 + 9x_4 + 9\end{bmatrix}\\ &= \begin{bmatrix}\frac{23}{11} - \frac{2}{11}x_3 + \frac{1}{11}x_4\\ \frac{20}{11} + \frac{4}{11}x_3 + \frac{9}{11}x_4\end{bmatrix}.\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} &= \begin{bmatrix}3 & 0\\ 0 & 6\end{bmatrix} - \frac{1}{11}\begin{bmatrix}-1 & 0\\ 3 & 2\end{bmatrix}\begin{bmatrix}2 & -1\\ -1 & 6\end{bmatrix}\begin{bmatrix}-1 & 3\\ 0 & 2\end{bmatrix}\\ &= \begin{bmatrix}3 & 0\\ 0 & 6\end{bmatrix} - \frac{1}{11}\begin{bmatrix}-2 & 1\\ 4 & 9\end{bmatrix}\begin{bmatrix}-1 & 3\\ 0 & 2\end{bmatrix} = \begin{bmatrix}3 & 0\\ 0 & 6\end{bmatrix} - \frac{1}{11}\begin{bmatrix}2 & -4\\ -4 & 30\end{bmatrix}\\ &= \begin{bmatrix}\frac{31}{11} & \frac{4}{11}\\ \frac{4}{11} & \frac{36}{11}\end{bmatrix}.\end{aligned}$$

Consequently, from Exercise 12.58(a),

$$(X_1, X_2)_{|X_3=x_3, X_4=x_4} \sim \mathcal{N}_2\left(\begin{bmatrix}\frac{23}{11} - \frac{2}{11}x_3 + \frac{1}{11}x_4\\ \frac{20}{11} + \frac{4}{11}x_3 + \frac{9}{11}x_4\end{bmatrix}, \begin{bmatrix}\frac{31}{11} & \frac{4}{11}\\ \frac{4}{11} & \frac{36}{11}\end{bmatrix}\right).$$

c) Here we have the following partition:

$$\mathbf{X} = \begin{bmatrix}\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix}\\ [X_4]\end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix}\begin{bmatrix}2\\1\\0\end{bmatrix}\\ [-1]\end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix}\begin{bmatrix}3 & 0 & -1\\ 0 & 6 & 3\\ -1 & 3 & 6\end{bmatrix} & \begin{bmatrix}0\\2\\1\end{bmatrix}\\ [0 \quad 2 \quad 1] & [2]\end{bmatrix}.$$

We have $\boldsymbol{\Sigma}_{22}^{-1} = 1/2$. Therefore,

$$\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) = \begin{bmatrix}2\\1\\0\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\2\\1\end{bmatrix}[x_4 + 1] = \begin{bmatrix}2\\1\\0\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\ 2(x_4+1)\\ x_4 + 1\end{bmatrix} = \begin{bmatrix}2\\ 2 + x_4\\ \frac{1}{2} + \frac{1}{2}x_4\end{bmatrix}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} &= \begin{bmatrix}3 & 0 & -1\\ 0 & 6 & 3\\ -1 & 3 & 6\end{bmatrix} - \frac{1}{2}\begin{bmatrix}0\\2\\1\end{bmatrix}[0 \quad 2 \quad 1] = \begin{bmatrix}3 & 0 & -1\\ 0 & 6 & 3\\ -1 & 3 & 6\end{bmatrix} - \frac{1}{2}\begin{bmatrix}0 & 0 & 0\\ 0 & 4 & 2\\ 0 & 2 & 1\end{bmatrix}\\ &= \begin{bmatrix}3 & 0 & -1\\ 0 & 4 & 2\\ -1 & 2 & 11/2\end{bmatrix}.\end{aligned}$$

Consequently, from Exercise 12.58(a),

$$(X_1, X_2, X_3)_{|X_4=x_4} \sim \mathcal{N}_2\left(\begin{bmatrix}2\\ 2 + x_4\\ \frac{1}{2} + \frac{1}{2}x_4\end{bmatrix}, \begin{bmatrix}3 & 0 & -1\\ 0 & 4 & 2\\ -1 & 2 & 11/2\end{bmatrix}\right).$$

12.4 Sampling Distributions

Basic Exercises

12.60 Noting that T is a symmetric random variable (i.e., f_T is an even function), we find that, for $t > 0$,

$$F_{T^2}(t) = P\left(T^2 \le t\right) = P\left(-\sqrt{t} \le T \le \sqrt{t}\right) = 2F_T\left(\sqrt{t}\right) - 1.$$

Therefore,

$$\begin{aligned} f_{T^2}(t) &= 2f_T\left(\sqrt{t}\right) \cdot \frac{1}{2\sqrt{t}} = t^{-1/2}\frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\,(1+t/\nu)^{-(\nu+1)/2} \\ &= \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\,\frac{t^{1/2-1}}{(1+t/\nu)^{(\nu+1)/2}} = \frac{(1/\nu)^{1/2}\Gamma\big((1+\nu)/2\big)}{\Gamma(1/2)\Gamma(\nu/2)}\,\frac{t^{1/2-1}}{\big(1+(1/\nu)t\big)^{(1+\nu)/2}}. \end{aligned}$$

Referring now to Equation (12.54) on page 727, we see that T^2 has the F-distribution with degrees of freedom 1 and ν.

12.61 Let $\chi^2 = (n-1)S_n^2/\sigma^2$. From Proposition 12.11 on page 725, we know that χ^2 has the chi-square distribution with $n-1$ degrees of freedom or, equivalently, $\chi^2 \sim \Gamma\left((n-1)/2, 1/2\right)$. Consequently,

$$\mathcal{E}\left(\chi^2\right) = \frac{(n-1)/2}{1/2} = n-1 \qquad \text{and} \qquad \operatorname{Var}\left(\chi^2\right) = \frac{(n-1)/2}{(1/2)^2} = 2(n-1).$$

a) We have

$$\mathcal{E}\left(S_n^2\right) = \mathcal{E}\left(\frac{\sigma^2}{n-1}\chi^2\right) = \frac{\sigma^2}{n-1}\mathcal{E}\left(\chi^2\right) = \frac{\sigma^2}{n-1}\cdot(n-1) = \sigma^2.$$

Therefore, S_n^2 is an unbiased estimator of σ^2.
b) We have

$$\operatorname{Var}\left(S_n^2\right) = \operatorname{Var}\left(\frac{\sigma^2}{n-1}\chi^2\right) = \frac{\sigma^4}{(n-1)^2}\operatorname{Var}\left(\chi^2\right) = \frac{\sigma^4}{(n-1)^2}\cdot 2(n-1) = \frac{2\sigma^4}{n-1}.$$

From part (a) and Chebyshev's inequality, for each $\epsilon > 0$,

$$P\left(\left|S_n^2 - \sigma^2\right| \ge \epsilon\right) = P\left(\left|S_n^2 - \mathcal{E}\left(S_n^2\right)\right| \ge \epsilon\right) \le \frac{\operatorname{Var}\left(S_n^2\right)}{\epsilon^2} = \frac{2\sigma^4}{(n-1)\epsilon^2},$$

which converges to 0 as $n \to \infty$. Consequently, from the complementation rule, for each $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|S_n^2 - \sigma^2\right| < \epsilon\right) = 1 - \lim_{n\to\infty} P\left(\left|S_n^2 - \sigma^2\right| \ge \epsilon\right) = 1 - 0 = 1.$$

In other words, S_n^2 is a consistent estimator of σ^2.

12.62
a) Referring to Exercise 12.61, we can write $S_n = \sigma\sqrt{\chi^2}/\sqrt{n-1}$, where $\chi^2 \sim \Gamma\left((n-1)/2, 1/2\right)$. Now, suppose that $Y \sim \Gamma(\alpha, \lambda)$. Then, from the FEF,

$$\begin{aligned} \mathcal{E}\left(\sqrt{Y}\right) &= \int_0^\infty \sqrt{y}\,\frac{\lambda^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\lambda y}\,dy = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty y^{\alpha+1/2-1}\,e^{-\lambda y}\,dy \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\,\frac{\Gamma(\alpha+1/2)}{\lambda^{\alpha+1/2}} = \frac{1}{\sqrt{\lambda}}\,\frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)}. \end{aligned}$$

Hence,

$$\mathcal{E}(S_n) = \mathcal{E}\left(\frac{\sigma}{\sqrt{n-1}}\sqrt{\chi^2}\right) = \frac{\sigma}{\sqrt{n-1}}\mathcal{E}\left(\sqrt{\chi^2}\right) = \frac{\sigma}{\sqrt{n-1}}\frac{1}{\sqrt{1/2}}\frac{\Gamma\big((n-1)/2+1/2\big)}{\Gamma\big((n-1)/2\big)}$$
$$= \left(\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma\big((n-1)/2\big)}\right)\sigma.$$

b) Here we apply Proposition 7.7 on page 354 (see also Exercise 7.81). From the solution to Exercise 12.61(b), we know that $\mathrm{Var}\big(S_n^2\big) > 0$. Hence, S_n^2 is not a constant random variable, which implies that S_n is not a constant random variable, which, in turn, implies that $\mathrm{Var}(S_n) > 0$. Referring to Exercise 12.61(a), we conclude that

$$0 < \mathrm{Var}(S_n) = \mathcal{E}\big(S_n^2\big) - \big(\mathcal{E}(S_n)\big)^2 = \sigma^2 - \big(\mathcal{E}(S_n)\big)^2,$$

which shows that $\mathcal{E}(S_n) < \sigma$. In particular, then, S_n is not an unbiased estimator of σ.

12.63

a) From Proposition 12.12 on page 726,

$$T_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

has the Student's t-distribution with $n-1$ degrees of freedom. Using the fact that a t-distribution is continuous and symmetric, we conclude that

$$P\left(-t_{(1+\gamma)/2,n-1} \le \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \le t_{(1+\gamma)/2,n-1}\right) = P\left(-t_{(1+\gamma)/2,n-1} \le T_n \le t_{(1+\gamma)/2,n-1}\right)$$
$$= F_{T_n}\big(t_{(1+\gamma)/2,n-1}\big) - F_{T_n}\big(-t_{(1+\gamma)/2,n-1}\big)$$
$$= 2F_{T_n}\big(t_{(1+\gamma)/2,n-1}\big) - 1 = 2\cdot\frac{1+\gamma}{2} - 1 = \gamma.$$

Solving for μ in the inequalities in the first term in the preceding display yields

$$P\Big(\bar{X}_n - t_{(1+\gamma)/2,n-1}\cdot S_n/\sqrt{n} \le \mu \le \bar{X}_n + t_{(1+\gamma)/2,n-1}\cdot S_n/\sqrt{n}\Big) = \gamma.$$

b) From part (a), chances are $100\gamma\%$ that the random interval from

$$\bar{X}_n - t_{(1+\gamma)/2,n-1}\cdot S_n/\sqrt{n} \qquad \text{to} \qquad \bar{X}_n + t_{(1+\gamma)/2,n-1}\cdot S_n/\sqrt{n}$$

contains μ. Hence, we can be $100\gamma\%$ confident that any such computed interval contains μ. In other words, the interval from

$$\bar{x}_n - t_{(1+\gamma)/2,n-1}\cdot s_n/\sqrt{n} \qquad \text{to} \qquad \bar{x}_n + t_{(1+\gamma)/2,n-1}\cdot s_n/\sqrt{n},$$

where $\bar{x}_n$ and s_n are the observed values of $\bar{X}_n$ and S_n, respectively, is a $100\gamma\%$ confidence interval for μ.

12.64

a) We have $\gamma = 0.95$ and $n = 16$, so that $(1+\gamma)/2 = 0.975$ and $n-1 = 15$. From the note, then, we have $t_{(1+\gamma)/2,n-1} = t_{0.975,15} = 2.13$. We also know that $\bar{x}_n = \bar{x}_{16} = 86.2$ and $s_n = s_{16} = 8.5$. Referring now to Exercise 12.63(b), we conclude that a 95% confidence interval for μ is from

$$86.2 - 2.13\cdot 8.5/\sqrt{16} \qquad \text{to} \qquad 86.2 - 2.13\cdot 8.5/\sqrt{16},$$

or from 81.7 mm Hg to 90.7 mm Hg.

b) We can be 95% confident that the mean arterial blood pressure for all children of diabetic mothers is somewhere between 81.7 mm Hg and 90.7 mm Hg.
c) From Proposition 12.12 on page 726, we must assume that arterial blood pressure for children of diabetic mothers is normally distributed. Technically, for the independence assumption in Proposition 12.12, we must, in addition, assume that the sampling is with replacement. However, because the sample size is small relative to the population size, sampling without replacement is also acceptable.

12.65
a) We assume that the current price for bananas is normally distributed. Technically, for the independence assumption in Proposition 12.12 on page 726, we must, in addition, assume that the sampling is with replacement. However, because the sample size is small relative to the population size, sampling without replacement is also acceptable. Here $n = 15$. If the mean retail price of bananas hasn't changed from that in 1998, then $\mu = 51.0$ and, from Proposition 12.12,

$$T_{15} = \frac{\bar{X}_{15} - 51.0}{S_{15}/\sqrt{15}}$$

has the Student's t-distribution with 14 degrees of freedom. In this case, the observed value of T_{15} is

$$t_{15} = \frac{\bar{x}_{15} - 51.0}{s_{15}/\sqrt{15}} = \frac{53.4 - 51.0}{3.5/\sqrt{15}} = 2.656.$$

Hence, using the symmetry of a Student's t-distribution, we get

$$P(|T_{15}| \geq |t_{15}|) = P(|T_{15}| \geq 2.656) = 2\left(1 - P(T_{15} \leq 2.656)\right) = 2\left(1 - F(2.656)\right),$$

where F is the CDF of the Student's t-distribution with 14 degrees of freedom.
b) Yes, we can reasonably conclude that the current mean retail price of bananas is different from the 1998 mean of 51.0 cents per pound. Indeed, there is less than a 2% chance of observing a value of the one-sample-t statistic at least as large in magnitude as that actually observed if the current mean retail price of bananas has not changed from that in 1998.

12.66
a) From Proposition 12.13 on page 728,

$$F_{n_1,n_2} = \frac{S_{n_1}^2/\sigma_1^2}{S_{n_2}^2/\sigma_2^2}$$

has the F-distribution with degrees of freedom $n_1 - 1$ and $n_2 - 1$. Using the fact that an F-distribution is continuous, we conclude that

$$\begin{aligned}
P&\left(f_{(1-\gamma)/2,n_1-1,n_2-1} \leq \frac{S_{n_1}^2/\sigma_1^2}{S_{n_2}^2/\sigma_2^2} \leq f_{(1+\gamma)/2,n_1-1,n_2-1}\right) \\
&= P\left(f_{(1-\gamma)/2,n_1-1,n_2-1} \leq F_{n_1,n_2} \leq f_{(1+\gamma)/2,n_1-1,n_2-1}\right) \\
&= F_{F_{n_1,n_2}}\left(f_{(1+\gamma)/2,n_1-1,n_2-1}\right) - F_{F_{n_1,n_2}}\left(f_{(1-\gamma)/2,n_1-1,n_2-1}\right) \\
&= \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma.
\end{aligned}$$

Solving for σ_1^2/σ_2^2 in the inequalities in the first term in the preceding display yields

$$P\left(\frac{S_{n_1}^2/S_{n_2}^2}{f_{(1+\gamma)/2,n_1-1,n_2-1}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_{n_1}^2/S_{n_2}^2}{f_{(1-\gamma)/2,n_1-1,n_2-1}}\right) = \gamma.$$

b) From part (a), chances are $100\gamma\%$ that the random interval from

$$\frac{S_{n_1}^2/S_{n_2}^2}{f_{(1+\gamma)/2,n_1-1,n_2-1}} \quad \text{to} \quad \frac{S_{n_1}^2/S_{n_2}^2}{f_{(1-\gamma)/2,n_1-1,n_2-1}}$$

contains σ_1^2/σ_2^2. Hence, we can be $100\gamma\%$ confident that any such computed interval contains σ_1^2/σ_2^2. In other words, the interval from

$$\frac{s_{n_1}^2/s_{n_2}^2}{f_{(1+\gamma)/2,n_1-1,n_2-1}} \quad \text{to} \quad \frac{s_{n_1}^2/s_{n_2}^2}{f_{(1-\gamma)/2,n_1-1,n_2-1}}$$

where $s_{n_1}^2$ and $s_{n_2}^2$ are the observed values of $S_{n_1}^2$ and $S_{n_2}^2$, respectively, is a $100\gamma\%$ confidence interval for σ_1^2/σ_2^2.

12.67

a) We have $\gamma = 0.95$ so that $(1-\gamma)/2 = 0.025$ and $(1+\gamma)/2 = 0.975$. Furthermore, $n_1 = n_2 = 10$, so that, from the note,

$$f_{(1-\gamma)/2,n_1-1,n_2-1} = f_{0.025,9,9} = 0.248 \qquad \text{and} \qquad f_{(1+\gamma)/2,n_1-1,n_2-1} = f_{0.975,9,9} = 4.026.$$

We also know that $s_{n_1}^2 = s_{10}^2 = (128.3)^2$ and $s_{n_2}^2 = s_{10}^2 = (199.7)^2$. Referring now to Exercise 12.66(b), we conclude that a 95% confidence interval for σ_1^2/σ_2^2 is from

$$\frac{(128.3/199.7)^2}{4.026} \quad \text{to} \quad \frac{(128.3/199.7)^2}{0.248},$$

or from 0.10 to 1.66.

b) We can be 95% confident that the ratio of the variances of tear strength for Brand A and Brand B vinyl floor coverings is somewhere between 0.10 and 1.66.

c) We assume that tear strengths for both Brand A and Brand B vinyl floor coverings are normally distributed.

12.68

a) From the FEF, we know that T has finite expectation if and only if $\int_{-\infty}^{\infty} |t|\left(1+t^2/\nu\right)^{-(\nu+1)/2}\,dt < \infty$. Using symmetry and making the substitution $u = 1 + t^2/\nu$, we find that

$$\int_{-\infty}^{\infty} \frac{|t|}{\left(1+t^2/\nu\right)^{(\nu+1)/2}}\,dt = 2\int_0^{\infty} \frac{t}{\left(1+t^2/\nu\right)^{(\nu+1)/2}}\,dt = \nu \int_1^{\infty} \frac{du}{u^{(\nu+1)/2}}.$$

From calculus, we know that the integral on the right of the preceding display converges (i.e., is finite) if and only if $(\nu+1)/2 > 1$, that is, if and only if $\nu > 1$.

b) We note that the PDF of T is an even function (i.e., it is symmetric about 0). Consequently, by Exercise 10.20(c), when T has finite expectation, we have $\mathcal{E}(T) = 0$.

c) We note that T has finite variance if and only if $\int_{-\infty}^{\infty} t^2\left(1+t^2/\nu\right)^{-(\nu+1)/2}\,dt < \infty$. Using symmetry and making the substitution $u = 1 + t^2/\nu$, we find that

$$\int_{-\infty}^{\infty} \frac{t^2}{\left(1+t^2/\nu\right)^{(\nu+1)/2}}\,dt = 2\int_0^{\infty} \frac{t^2}{\left(1+t^2/\nu\right)^{(\nu+1)/2}}\,dt = \nu^{3/2} \int_1^{\infty} \frac{\sqrt{u-1}}{u^{(\nu+1)/2}}\,du.$$

Noting that the integrand on the right of the preceding display is asymptotic to $u^{-\nu/2}$ as $u \to \infty$ and recalling from calculus that $\int_1^{\infty} u^{-p}\,du$ converges (i.e., is finite) if and only if $p > 1$, we see that T has finite variance if and only if $\nu/2 > 1$, that is, if and only if $\nu > 2$.

d) Using integration by parts, we find that

$$\int_0^\infty \frac{t^2}{\big(1+t^2/\nu\big)^{(\nu+1)/2}}\,dt = \frac{\nu}{\nu-1}\int_0^\infty \frac{dt}{\big(1+t^2/\nu\big)^{(\nu-1)/2}}.$$

Making the substitution $u = t\sqrt{(\nu-2)/\nu}$ and using the fact that a t PDF with $\nu-2$ degrees of freedom integrates to 1, we get

$$\begin{aligned}\int_0^\infty \frac{dt}{\big(1+t^2/\nu\big)^{(\nu-1)/2}} &= \sqrt{\frac{\nu}{\nu-2}}\int_0^\infty \frac{du}{\big(1+u^2/(\nu-2)\big)^{((\nu-2)+1)/2}}\\ &= \frac{1}{2}\sqrt{\frac{\nu}{\nu-2}}\int_{-\infty}^\infty \frac{du}{\big(1+u^2/(\nu-2)\big)^{((\nu-2)+1)/2}}\\ &= \frac{1}{2}\sqrt{\frac{\nu}{\nu-2}}\cdot\frac{\sqrt{(\nu-2)\pi}\,\Gamma\big((\nu-2)/2\big)}{\Gamma\big((\nu-1)/2\big)} = \frac{1}{2}\sqrt{\nu\pi}\,\frac{\Gamma\big((\nu-2)/2\big)}{\Gamma\big((\nu-1)/2\big)}.\end{aligned}$$

Recalling from part (b) that $\mathcal{E}(T)=0$, we conclude that

$$\begin{aligned}\operatorname{Var}(T) = \mathcal{E}\big(T^2\big) &= \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\int_{-\infty}^\infty \frac{t^2}{\big(1+t^2/\nu\big)^{(\nu+1)/2}}\,dt\\ &= 2\frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\int_0^\infty \frac{t^2}{\big(1+t^2/\nu\big)^{(\nu+1)/2}}\,dt\\ &= 2\frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\cdot\frac{\nu}{\nu-1}\cdot\frac{1}{2}\sqrt{\nu\pi}\,\frac{\Gamma\big((\nu-2)/2\big)}{\Gamma\big((\nu-1)/2\big)}\\ &= 2\frac{\big((\nu-1)/2\big)\Gamma\big((\nu-1)/2\big)}{\big((\nu-2)/2\big)\Gamma\big((\nu-2)/2\big)}\cdot\frac{\nu}{\nu-1}\cdot\frac{1}{2}\frac{\Gamma\big((\nu-2)/2\big)}{\Gamma\big((\nu-1)/2\big)}\\ &= \frac{\nu}{\nu-2}.\end{aligned}$$

12.69

a) From Proposition 12.9 on page 723, the random variable $Z = \big(\bar{X}_n-\mu\big)\big/\big(\sigma/\sqrt{n}\big)$ has the standard normal distribution and, hence, has mean 0 and variance 1.

b) From Proposition 12.12 on page 726, the random variable $T_n = \big(\bar{X}_n-\mu\big)\big/\big(s/\sqrt{n}\big)$ has the Student's t-distribution with $n-1$ degrees of freedom. Referring to Exercise 12.68, we see that T_n has finite mean if and only if $n \ge 3$, in which case, $\mathcal{E}(T_n)=0$, and that T_n has finite variance if and only if $n\ge 4$, in which case,

$$\operatorname{Var}(T_n) = \frac{n-1}{(n-1)-2} = \frac{n-1}{n-3}.$$

c) We see that the random variables in parts (a) and (b) both have mean 0 (when the mean of the latter exists). Futhermore, they asymptotically have the same variance because

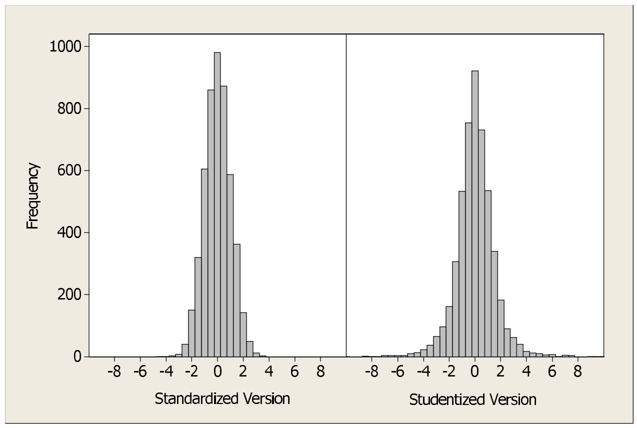
$$\lim_{n\to\infty}\operatorname{Var}(T_n) = \lim_{n\to\infty}\frac{n-1}{n-3} = 1 = \operatorname{Var}(Z)\,.$$

12.70

a) Answers will vary.

b) Answers will vary.

c) Answers will vary, but here is what we obtained:



d) The two histograms suggest that the distributions of both the standardized version of the sample mean and the studentized version of the sample mean are bell shaped and symmetric about 0.

e) The two histograms suggest that the distribution of the studentized version of the sample mean has more spread (larger variance) than that of the standardized version of the sample mean. Intuitively, this difference is due to the fact that the variation in the possible values of the standardized version is due solely to the variation of sample means, whereas that of the studentized version is due to the variation of both sample means and sample standard deviations. Mathematically, the difference in variation follows from Exercise 12.69; specifically, $\mathrm{Var}(T_4) = 3 > 1 = \mathcal{E}(Z)$.

f) The histograms would have the same look if the simulations were carried out with a different normal distribution because the distributions of Z and T_4 don't depend on μ and σ.

Theory Exercises

12.71

a) We have

$$Y_1 = \bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\left(X_1 + \cdots + X_n\right)$$

$$Y_2 = X_1 - \bar{X}_n = X_1 - \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\big((n-1)X_1 - X_2 - \cdots - X_n\big)$$

$$Y_3 = X_2 - \bar{X}_n = X_2 - \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\big(-X_1 + (n-1)X_2 - X_3 - \cdots - X_n\big)$$

$$\vdots$$

$$Y_n = X_{n-1} - \bar{X}_n = X_{n-1} - \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\big(-X_1 - \cdots - X_{n-2} + (n-1)X_{n-1} - X_n\big)$$

$$Y_{n+1} = X_n - \bar{X}_n = X_n - \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\big(-X_1 - \cdots - X_{n-1} + (n-1)X_n\big).$$

Hence, we can write $\mathbf{Y} = \mathbf{BX}$, where $\mathbf{B}$ is the $(n+1) \times n$ matrix given by

$$\mathbf{B} = \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ n-1 & -1 & -1 & \cdots & -1 & -1 \\ -1 & n-1 & -1 & \cdots & -1 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & n-1 & -1 \\ -1 & -1 & -1 & \cdots & -1 & n-1 \end{bmatrix}.$$

b) As $X_1, \ldots, X_n$ are independent random variables with common variance σ^2, they are multivariate normal random variables with $\mathbf{\Sigma_X} = \sigma^2 \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Applying Proposition 12.7 on page 719, we conclude that $\mathbf{Y}$ has the multivariate normal distribution with covariance matrix $\mathbf{B\Sigma_X B'} = \sigma^2\mathbf{BB'}$. Next we note that the sum of the first row of $\mathbf{B}$ equals 1 and that the sums of the other rows of $\mathbf{B}$ equal 0. Hence, the first row of $\mathbf{\Sigma_Y}$ is $[\sigma^2/n, 0, 0, \ldots, 0]$. We can thus partition $\mathbf{Y}$ and $\mathbf{\Sigma_Y}$ as follows:

$$\mathbf{Y} = \begin{bmatrix} [\,Y_1\,] \\ \begin{bmatrix} Y_2 \\ \vdots \\ Y_{n+1} \end{bmatrix} \end{bmatrix} \qquad \text{and} \qquad \mathbf{\Sigma_Y} = \begin{bmatrix} [\,\sigma^2/n\,] & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{bmatrix},$$

where $\mathbf{\Sigma}_{22}$ is the covariance matrix of $Y_2, \ldots, Y_{n+1}$. Referring to Exercise 12.57(c), we conclude that Y_1 and $Y_2, \ldots, Y_{n+1}$ are independent, that is, that $\bar{X}_n$ is independent of $X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n$.
c) From part (b) and the fact that S_n^2 is a function of $X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n$, we conclude that $\bar{X}_n$ and S_n^2 are independent random variables.

Advanced Exercises

12.72
a) Because $S_{n_1}^2$ and $S_{n_2}^2$ are both unbiased estimators of σ^2, we have

$$\mathcal{E}\left(cS_{n_1}^2 + (1-c)S_{n_2}^2\right) = c\mathcal{E}\left(S_{n_1}^2\right) + (1-c)\mathcal{E}\left(S_{n_2}^2\right) = c\sigma^2 + (1-c)\sigma^2 = \sigma.$$

Hence, $cS_{n_1}^2 + (1-c)S_{n_2}^2$ is an unbiased estimator of σ^2.
b) From the solution to Exercise 12.61(b), we have

$$\operatorname{Var}\left(S_{n_1}^2\right) = \frac{2\sigma^4}{n_1 - 1} \qquad \text{and} \qquad \operatorname{Var}\left(S_{n_1}^2\right) = \frac{2\sigma^4}{n_2 - 1}.$$

Therefore,

$$\operatorname{Var}\left(cS_{n_1}^2 + (1-c)S_{n_2}^2\right) = c^2 \operatorname{Var}\left(S_{n_1}^2\right) + (1-c)^2 \operatorname{Var}\left(S_{n_2}^2\right) = 2\sigma^4\left(\frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}\right).$$

Taking the derivative with respect to c of the term on the right of the preceding display, setting the result equal to 0, and solving for c, we find that $c = (n_1 - 1)/(n_1 + n_2 - 2)$.
c) Referring to part (b), we see that, among the estimators of σ^2 of the form $cS_{n_1}^2 + (1-c)S_{n_2}^2$, the one with minimum variance is

$$\frac{n_1 - 1}{n_1 + n_2 - 2}S_{n_1}^2 + \left(1 - \frac{n_1 - 1}{n_1 + n_2 - 2}\right)S_{n_2}^2 = \frac{(n_1 - 1)S_{n_1}^2 + (n_2 - 1)S_{n_2}^2}{n_1 + n_2 - 2},$$

which we denote S_p^2.

12.73

a) We begin by noting that

$$\frac{(n_1+n_2-2)S_p^2}{\sigma^2} = \frac{(n_1-1)S_{n_1}^2}{\sigma^2} + \frac{(n_2-1)S_{n_2}^2}{\sigma^2}.$$

From Proposition 12.11 on page 725, the two random variables in the sum on the right of the preceding display have the chi-square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, respectively. Furthermore, because the random samples are independent, those two random variables are independent. Hence, from the second bulleted item on page 537, $(n_1+n_2-2)S_p^2/\sigma^2$ has the chi-square distribution with $(n_1-1)+(n_2-1) = n_1+n_2-2$ degrees of freedom.

b) From previous results, we know that $\bar{X}_{n_1} \sim \mathcal{N}\left(\mu_1, \sigma^2/n_1\right)$ and $\bar{X}_{n_2} \sim \mathcal{N}\left(\mu_2, \sigma^2/n_2\right)$. Furthermore, because the random samples are independent, $\bar{X}_{n_1}$ and $\bar{X}_{n_2}$ are independent. Hence,

$$\mathcal{E}\left(\bar{X}_{n_1} - \bar{X}_{n_2}\right) = \mathcal{E}\left(\bar{X}_{n_1}\right) - \mathcal{E}\left(\bar{X}_{n_2}\right) = \mu_1 - \mu_2$$

and

$$\operatorname{Var}\left(\bar{X}_{n_1} - \bar{X}_{n_2}\right) = \operatorname{Var}\left(\bar{X}_{n_1}\right) + \operatorname{Var}\left(\bar{X}_{n_2}\right) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left((1/n_1) + (1/n_2)\right).$$

It follows that the random variable

$$Z = \frac{\left(\bar{X}_{n_1} - \bar{X}_{n_2}\right) - (\mu_1 - \mu_2)}{\sigma\sqrt{(1/n_1) + (1/n_2)}}$$

has the standard normal distribution. Now, let

$$Y = \frac{(n_1+n_2-2)S_p^2}{\sigma^2}.$$

From Proposition 12.10 on page 724, $\bar{X}_{n_1}$ and $S_{n_1}^2$ are independent, as are $\bar{X}_{n_2}$ and $S_{n_2}^2$. Thus, because the random samples are independent, $\bar{X}_{n_1}, S_{n_1}^2, \bar{X}_{n_2}$, and $S_{n_2}^2$ are independent random variables. Consequently, from Proposition 6.13 on page 297, Z and Y are independent. Referring now to part (a) and Exercise 9.164, we conclude that $Z\big/\sqrt{Y/(n_1+n_2-2)}$ has the Student's t-distribution with n_1+n_2-2 degrees of freedom. However,

$$\begin{aligned}
\frac{Z}{\sqrt{Y/(n_1+n_2-2)}} &= \frac{\left(\left(\bar{X}_{n_1} - \bar{X}_{n_2}\right) - (\mu_1-\mu_2)\right)\big/\sigma\sqrt{(1/n_1)+(1/n_2)}}{\sqrt{S_p^2/\sigma^2}} \\
&= \frac{\left(\bar{X}_{n_1} - \bar{X}_{n_2}\right) - (\mu_1-\mu_2)}{S_p\sqrt{(1/n_1)+(1/n_2)}} = T.
\end{aligned}$$

12.74 Using the fact that a t-distribution is continuous and symmetric and referring to Exercise 12.73, we conclude that

$$\begin{aligned}
&P\left(-t_{(1+\gamma)/2,n_1+n_2-2} \le \frac{\left(\bar{X}_{n_1} - \bar{X}_{n_2}\right) - (\mu_1-\mu_2)}{S_p\sqrt{(1/n_1)+(1/n_2)}} \le t_{(1+\gamma)/2,n_1+n_2-2}\right) \\
&\quad = P\left(-t_{(1+\gamma)/2,n_1+n_2-2} \le T \le t_{(1+\gamma)/2,n_1+n_2-2}\right) \\
&\quad = F_T\left(t_{(1+\gamma)/2,n_1+n_2-2}\right) - F_T\left(-t_{(1+\gamma)/2,n_1+n_2-2}\right) \\
&\quad = 2F_T\left(t_{(1+\gamma)/2,n_1+n_2-2}\right) - 1 = 2\cdot\frac{1+\gamma}{2} - 1 = \gamma.
\end{aligned}$$

Solving for $\mu_1 - \mu_2$ in the inequalities in the first term in the preceding display yields

$$P\Big(\big(\bar{X}_{n_1} - \bar{X}_{n_2}\big) - t_{(1+\gamma)/2, n_1+n_2-2} \cdot SE \le \mu_1 - \mu_2 \le \big(\bar{X}_{n_1} - \bar{X}_{n_2}\big) + t_{(1+\gamma)/2, n_1+n_2-2} \cdot SE\Big) = \gamma,$$

where $SE = S_p\sqrt{(1/n_1) + (1/n_2)}$. Therefore, chances are $100\gamma\%$ that the random interval with endpoints

$$\big(\bar{X}_{n_1} - \bar{X}_{n_2}\big) \pm t_{(1+\gamma)/2, n_1+n_2-2} \cdot S_p\sqrt{(1/n_1) + (1/n_2)}$$

contains $\mu_1 - \mu_2$. Hence, we can be $100\gamma\%$ confident that any such computed interval contains $\mu_1 - \mu_2$. In other words, the interval with endpoints

$$(\bar{x}_{n_1} - \bar{x}_{n_2}) \pm t_{(1+\gamma)/2, n_1+n_2-2} \cdot s_p\sqrt{(1/n_1) + (1/n_2)},$$

where $\bar{x}_{n_1}$, $\bar{x}_{n_2}$, and s_p are the observed values of $\bar{X}_{n_1}$, $\bar{X}_{n_2}$, and S_p, respectively, is a $100\gamma\%$ confidence interval for $\mu_1 - \mu_2$.

12.75

a) We assume that daily protein intakes of female vegetarians and female omnivores living in Taiwan are normally distributed with the same variance. Technically, we must, in addition, assume that the sampling is with replacement. However, because the sample sizes are small relative to the population sizes, sampling without replacement is also acceptable. Here $n_1 = 51$, $n_2 = 53$, $\bar{x}_{n_1} = 39.04$, $s_{n_1} = 18.82$, $\bar{x}_{n_2} = 49.92$, and $s_{n_2} = 18.97$. Hence,

$$s_p = \sqrt{\frac{(n_1 - 1)s_{n_1}^2 + (n_2 - 1)s_{n_2}^2}{n_1 + n_2 - 2}} = \sqrt{\frac{50 \cdot (18.82)^2 + 52 \cdot (18.97)^2}{102}} = 18.90.$$

Also, $\gamma = 0.99$ so that $(1 + \gamma)/2 = 0.995$. From the note, then, $t_{(1+\gamma)/2, n_1+n_2-2} = t_{0.995, 102} = 2.625$. Referring now to Exercise 12.74, we conclude that a 99% confidence interval for $\mu_1 - \mu_2$ has endpoints

$$(39.04 - 49.92) \pm 2.625 \cdot 18.90\sqrt{(1/51) + (1/53)},$$

or -20.61 g to -1.15 g.

b) From part (a), we can be 99% confident that the mean daily protein intake of female omnivores living in Taiwan exceeds that of female vegetarians living in Taiwan by somewhere between 1.15 g and 20.61 g. Thus, we can be quite confident that the mean daily protein intake of female omnivores living in Taiwan exceeds that of female vegetarians living in Taiwan.

12.76

a) We assume that both salary distributions are normally distributed with the same variance. Technically, we must, in addition, assume that the sampling is with replacement. However, because the sample sizes are small relative to the population sizes, sampling without replacement is also acceptable. For the samples obtained,

$$s_p = \sqrt{\frac{(n_1 - 1)s_{n_1}^2 + (n_2 - 1)s_{n_2}^2}{n_1 + n_2 - 2}} = \sqrt{\frac{29 \cdot (23.95)^2 + 34 \cdot (22.26)^2}{63}} = 23.05.$$

If, in fact, the two salary distributions have the same mean (i.e., $\mu_1 = \mu_2$), then the value of the pooled-t statistic is

$$t = \frac{(\bar{x}_{n_1} - \bar{x}_{n_2}) - (\mu_1 - \mu_2)}{s_p\sqrt{(1/n_1) + (1/n_2)}} = \frac{(57.48 - 66.39) - 0}{23.05\sqrt{(1/30) + (1/35)}} = -1.554.$$

The problem is to determine $P(|T| \ge |t|)$. Because a Student's t-distribution is symmetric about 0, we have

$$P(|T| \ge |t|) = P(|T| \ge 1.554) = 2 \cdot \big(1 - F_T(1.554)\big),$$

where T has the Student's t-distribution with 63 degrees of freedom.
b) No, the result does not provide reasonably strong evidence that a difference exists in mean annual salaries for faculty in public and private institutions. Indeed, if the mean annual salaries are the same, we would get a value of the pooled-t statistic at least as large in magnitude as the one actually observed more than 12% of the time.

Review Exercises for Chapter 12

Basic Exercises

12.77

Expected number of successes per trial:

In Bernoulli trials with success probability p, the number of successes per trial has the Bernoulli distribution with parameter p or, equivalently, the $\mathcal{B}(1, p)$ distribution. The expected value of such a random variable is p. Hence, the expected number of successes per trial equals p.

Number of trials until the rth success:

In Bernoulli trials with success probability p, the number of trials until the rth success has the negative binomial distribution with parameters r and p, that is, the $\mathcal{NB}(r, p)$ distribution, as we know from Proposition 5.13 on page 241.

Elapsed time between successes:

For a Poisson process with rate λ, the elapsed time between successes has the exponential distribution with parameter λ, that is, the $\mathcal{E}(\lambda)$ distribution, as we know from Proposition 12.2 on page 691.

Number of successes in $(s, t]$:

For a Poisson process with rate λ, the number of successes in the time interval $(s, t]$ has the Poisson distribution with parameter $\lambda(t - s)$, that is, the $\mathcal{P}\big(\lambda(t - s)\big)$ distribution, as we know from Definition 12.1 on page 688.

12.78

a) From Proposition 12.2 on page 691, the interarrival times are independent $\mathcal{E}(6.9)$ random variables, where time is measured in hours after midnight. Hence, the probability that there is no arrival-free span of 8 minutes or more up to the time when the third patient arrives is

$$\begin{aligned} P(I_1 < 8/60,\ I_2 < 8/60,\ I_3 < 8/60) &= P(I_1 < 8/60)P(I_2 < 8/60)P(I_3 < 8/60) \\ &= \left(1 - e^{-6.9\cdot(8/60)}\right)^3 = 0.218. \end{aligned}$$

b) No patient sees a doctor by 12:20 A.M. if and only if the first patient arrives after 12:17 A.M. This latter event has probability

$$P(I_1 > 17/60) = e^{-6.9\cdot(17/60)} = 0.142.$$

c) Let T denote the check-in time, in hours. Then, by assumption, $T \sim \mathcal{T}(1/30, 2/30)$. Let U denote the time when the first patient to arrive sees the doctor. We have $U = I_1 + T$ and, by assumption, I_1 and T are independent random variables. We want to determine $P(U > 20/60) = P(U > 1/3)$. To do so, we apply the bivariate FPF and independence to get

$$P(U > 1/3) = P(I_1 + T > 1/3) = \iint_{x+y>1/3} f_{I_1,T}(x, y)\,dx\,dy = \iint_{x+y>1/3} f_{I_1}(x) f_T(y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} \left(\int_{1/3-y}^{\infty} f_{I_1}(x)\,dx \right) f_T(y)\,dy = \int_{1/30}^{2/30} e^{-6.9(1/3-y)} f_T(y)\,dy$$
$$= e^{-2.3} \int_{1/30}^{2/30} e^{6.9y} f_T(y)\,dy = e^{-2.3} \int_{-\infty}^{\infty} e^{6.9y} f_T(y)\,dy = e^{-2.3} M_T(6.9).$$

Referring now to Exercise 11.78, we conclude that

$$P(U > 1/3) = e^{-2.3} \cdot \frac{8e^{(1/30+2/30)(6.9)/2}}{(2/30 - 1/30)^2(6.9)^2} \left(\cosh \frac{(2/30 - 1/30)(6.9)}{2} - 1 \right) = 0.412.$$

Of course, we can also obtain this result by direct double-integration of the joint PDF of I_1 and T over $\{ (x, y) : x + y > 1/3 \}$. However, the method we presented is far more efficient.

d) Let T be as in part (c), let S denote the time the first patient spends with a doctor, and let V denote the time that the first patient leaves the emergency room. We have $V = I_1 + T + S$ and, by assumption, I_1, T, and S are independent random variables. We know that $I_1 \sim \mathcal{E}(6.9)$, $T \sim \mathcal{T}(1/30, 2/30)$, and $S \sim \mathcal{U}(1/12, 2/12)$. Let $W = T + S$ so that $V = I_1 + W$. Note that $7/60 \le W \le 7/30$ and that $7/30 < 1/3$. We want to determine $P(V > 20/60) = P(V > 1/3)$. To do so, we apply the bivariate FPF and independence to get

$$P(V > 1/3) = P(I_1 + W > 1/3) = \iint_{x+y>1/3} f_{I_1,W}(x, y)\,dx\,dy = \iint_{x+y>1/3} f_{I_1}(x) f_W(y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} \left(\int_{1/3-y}^{\infty} f_{I_1}(x)\,dx \right) f_W(y)\,dy = \int_{7/60}^{7/30} e^{-6.9(1/3-y)} f_W(y)\,dy$$
$$= e^{-2.3} \int_{7/60}^{7/30} e^{6.9y} f_W(y)\,dy = e^{-2.3} \int_{-\infty}^{\infty} e^{6.9y} f_W(y)\,dy = e^{-2.3} M_W(6.9)$$
$$= e^{-2.3} M_{T+S}(6.9) = e^{-2.3} M_T(6.9) M_S(6.9).$$

Referring to the solution to part (c), we see that the product of the first two terms in the last expression in the preceding display equals 0.142 and, from Table 11.1 on page 634, the third term equals

$$M_S(6.9) = \frac{e^{(2/12)(6.9)} - e^{(1/12)(6.9)}}{(2/12 - 1/12)(6.9)} = 2.402.$$

Therefore,

$$P(V > 1/3) = 0.142 \cdot 2.402 = 0.341.$$

Of course, we can also obtain this result by direct triple-integration of the joint PDF of I_1, T, and S over $\{ (x, y, z) : x + y + z > 1/3 \}$. However, the method we presented is far more efficient.

12.79

a) We know that $N(t) \sim \mathcal{P}(\lambda t)$ and, hence,

$$\mathcal{E}\left(\frac{N(t)}{t}\right) = \frac{1}{t}\mathcal{E}\big(N(t)\big) = \frac{1}{t}\lambda t = \lambda.$$

b) From part (a) and Chebyshev's inequality, for each $\epsilon > 0$,

$$P\left(\left|\frac{N(t)}{t} - \lambda\right| \geq \epsilon\right) = P\left(\left|\frac{N(t)}{t} - \mathcal{E}\left(\frac{N(t)}{t}\right)\right| \geq \epsilon\right) \leq \frac{\operatorname{Var}\big(N(t)/t\big)}{\epsilon^2}$$
$$= \frac{\operatorname{Var}\big(N(t)\big)}{\epsilon^2 t^2} = \frac{\lambda t}{\epsilon^2 t^2} = \frac{\lambda}{\epsilon^2 t}.$$

As $t \to \infty$, the last term in the preceding display converges to 0. Hence, by the complementation rule,

$$\lim_{t\to\infty} P\left(\left|\frac{N(t)}{t} - \lambda\right| < \epsilon\right) = 1 - \lim_{t\to\infty} P\left(\left|\frac{N(t)}{t} - \lambda\right| \geq \epsilon\right) = 1 - 0 = 1.$$

12.80

a) For each $n \in \mathcal{N}$, we have $W_n = I_1 + \cdots + I_n$ and, by Proposition 12.2 on page 691, $I_1, I_2, \ldots$ are independent $\mathcal{E}(\lambda)$ random variables. Applying the strong law of large numbers, we conclude that

$$\lim_{n\to\infty} \frac{W_n}{n} = \lim_{n\to\infty} \frac{I_1 + \cdots + I_n}{n} = \mathcal{E}\big(I_j\big) = \frac{1}{\lambda},$$

with probability 1. Therefore,

$$\lim_{n\to\infty} \frac{n}{W_n} = \lim_{n\to\infty} \frac{1}{W_n/n} = \frac{1}{1/\lambda} = \lambda,$$

with probability 1.

b) We are given that $W_4 = 15$. Hence, from part (a),

$$\lambda \approx \frac{4}{W_4} = \frac{4}{15} \approx 0.267.$$

12.81 We assume that $\lambda < 2\mu$.

a) To begin, we note that, when only one customer is in the system, the service time is the minimum of the individual service times of the two servers. Hence, from Proposition 9.11(a) on page 532, the service time for a single customer in the system has an exponential distribution with parameter 2μ. Thus, we have a birth-and-death queueing system with $\lambda_n = \lambda$ for $n \geq 0$, and $\mu_n = 2\mu$ for $n \geq 1$. Hence, this queueing system behaves identically to an $M/M/1$ queue with the single server serving at rate 2μ. Referring now to Example 12.6 on page 703, specifically to Equation (12.20) on page 704, we conclude that $P_n = (1 - \lambda/2\mu)(\lambda/2\mu)^n$ for all $n \geq 0$.

b) Referring to part (a) and Example 12.8 on page 707, specifically, Equation (12.24) on that same page, we see that $L = \lambda/(2\mu - \lambda)$.

12.82

a) First assume that $s \leq t$. Applying properties of covariance and of a Poisson process, we find that

$$\begin{aligned}\operatorname{Cov}\big(N(s), N(t)\big) &= \operatorname{Cov}\big(N(s), N(t) - N(s) + N(s)\big)\\ &= \operatorname{Cov}\big(N(s), N(t) - N(s)\big) + \operatorname{Cov}\big(N(s), N(s)\big) = 0 + \operatorname{Var}\big(N(s)\big) = \lambda s.\end{aligned}$$

Interchanging the roles of s and t, we get that $\operatorname{Cov}\big(N(s), N(t)\big) = \lambda t$ if $t \leq s$. Consequently, we have shown that $\operatorname{Cov}\big(N(s), N(t)\big) = \min\{s, t\}$ for all $s, t \geq 0$.

b) Referring to part (a) and noting that $s \le s+t$, we get

$$\rho\big(N(s), N(s+t)\big) = \frac{\operatorname{Cov}\big(N(s), N(s+t)\big)}{\sqrt{\operatorname{Var}\big(N(s)\big)\operatorname{Var}\big(N(s+t)\big)}} = \frac{\lambda s}{\sqrt{\lambda s \cdot \lambda(s+t)}} = \sqrt{\frac{s}{s+t}}.$$

12.83 We measure time in hours after 7:45 A.M. Let $N(t)$ be the number of passengers at the bus stop at time t. By assumption, $\{N(t) : t \ge 0\}$ is a Poisson process with rate 60. We note that the bus arrives at time $T - 7.75$, which we denote by S. The number of passengers who are waiting at the bus stop when the bus arrives is $N(S)$, which we denote by X. The problem is to determine the mean and standard deviation of X. Now, let $Y = 4T - 32$. By assumption, $Y \sim \text{Beta}(2, 2)$. Therefore,

$$\mathcal{E}(Y) = \frac{2}{2+2} = \frac{1}{2} \qquad \text{and} \qquad \operatorname{Var}(Y) = \frac{2 \cdot 2}{(2+2)^2(2+2+1)} = \frac{1}{20}.$$

We have

$$S = T - 7.75 = \frac{Y+32}{4} - 7.75 = \frac{1}{4}(Y+1),$$

so that

$$\mathcal{E}(S) = \frac{1}{4}\big(\mathcal{E}(Y)+1\big) = \frac{1}{4}\left(\frac{1}{2}+1\right) = \frac{3}{8} \qquad \text{and} \qquad \operatorname{Var}(S) = \frac{1}{16}\operatorname{Var}(Y) = \frac{1}{16}\cdot\frac{1}{20} = \frac{1}{320}.$$

Now, because $\{N(t) : t \ge 0\}$ is independent of S, we have

$$\mathcal{E}(X \mid S) = \mathcal{E}\big(N(S) \mid S\big) = 60S \qquad \text{and} \qquad \operatorname{Var}(X \mid S) = \operatorname{Var}\big(N(S) \mid S\big) = 60S.$$

Consequently, by the laws of total expectation and total variance,

$$\mathcal{E}(X) = \mathcal{E}\big(\mathcal{E}(X \mid S)\big) = \mathcal{E}(60S) = 60\mathcal{E}(S) = 60 \cdot \frac{3}{8} = 22.5$$

and

$$\begin{aligned}\operatorname{Var}(X) &= \mathcal{E}\big(\operatorname{Var}(X \mid S)\big) + \operatorname{Var}\big(\mathcal{E}(X \mid S)\big) = \mathcal{E}(60S) + \operatorname{Var}(60S) \\ &= 60\mathcal{E}(S) + 3600\operatorname{Var}(S) = 60 \cdot \frac{3}{8} + 3600 \cdot \frac{1}{320} = 33.75.\end{aligned}$$

Hence, the mean number of passengers who are waiting at the bus stop when the bus arrives is 22.5 and the standard deviation of the number is $\sqrt{33.75} \approx 5.81$.

12.84

a) The counting process $\{N(t) : t \ge 0\}$ gives the total number of securities orders, both taxable and nontaxable, that have been received at the brokerage house. In particular, for each $t \ge 0$, $N(t)$ is the total number of securities orders that have been received at the brokerage house by time t.

b) We must show that the counting process $\{N(t) : t \ge 0\}$ satisfies the three conditions of Definition 12.1 on page 688 with $\lambda = \lambda_1 + \lambda_2$.

First condition:

As $\{N_1(t) : t \ge 0\}$ and $\{N_2(t) : t \ge 0\}$ are each Poisson processes, we have $N_1(0) = 0$ and $N_2(0) = 0$, so that $N(0) = N_1(0) + N_2(0) = 0$.

Second condition:

Let $r \in \mathcal{N}$ and $0 \le t_1 < t_2 < \cdots < t_r$. As $\{N_1(t) : t \ge 0\}$ and $\{N_2(t) : t \ge 0\}$ are Poisson processes, $N_1(t_j) - N_1(t_{j-1})$, $2 \le j \le r$, are independent random variables, as are $N_2(t_j) - N_2(t_{j-1})$, $2 \le j \le r$.

And, because $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are independent, we can apply Proposition 6.13 on page 297 to conclude that

$$\begin{aligned} N(t_j) - N(t_{j-1}) &= \big(N_1(t_j) + N_2(t_j)\big) - \big(N_1(t_{j-1}) + N_2(t_{j-1})\big) \\ &= \big(N_1(t_j) - N_1(t_{j-1})\big) + \big(N_2(t_j) - N_2(t_{j-1})\big), \qquad 2 \leq j \leq r, \end{aligned}$$

are independent random variables. Thus, $\{N(t) : t \geq 0\}$ has independent increments.

Third condition:

Let $0 \leq s < t < \infty$. As $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are Poisson processes with rates λ_1 and λ_2, respectively, we have that $N_1(t) - N_1(s) \sim \mathcal{P}\big(\lambda_1(t-s)\big)$ and $N_2(t) - N_2(s) \sim \mathcal{P}\big(\lambda_2(t-s)\big)$. And, because $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are independent, we can apply Proposition 6.20(b) on page 311 to conclude that

$$\begin{aligned} N(t) - N(s) &= \big(N_1(t) + N_2(t)\big) - \big(N_1(s) + N_2(s)\big) = \big(N_1(t) - N_1(s)\big) + \big(N_2(t) - N_2(s)\big) \\ &\sim \mathcal{P}\big(\lambda_1(t-s) + \lambda_2(t-s)\big) = \mathcal{P}\big((\lambda_1 + \lambda_2)(t-s)\big). \end{aligned}$$

We have, therefore, now shown that $\{N(t) : t \geq 0\}$ satisfies the three conditions given in Definition 12.1 with $\lambda = \lambda_1 + \lambda_2$ and, hence, that it is a Poisson process with rate $\lambda_1 + \lambda_2$.

c) Recall that, for a Poisson process, the interarrival times between events are independent and identically distributed exponential random variables with common parameter equal to the rate of the process. Therefore, because of the lack-of-memory property of an exponential random variable, starting at any specified time, the elapsed time until the next event occurs has that same exponential distribution. Now, the probability that a securities order is taxable equals the probability that, starting at a specified time, the next taxable security order arrives before the next nontaxable security order. Because $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are independent Poisson processes with rates λ_1 and λ_2, respectively, those two arrival times are independent exponential random variables with parameters λ_1 and λ_2, respectively. Hence, from Proposition 9.11(b), the probability that, starting at a specified time, the next taxable security order arrives before the next nontaxable security order is $\lambda_1/(\lambda_1 + \lambda_2)$.

d) Let

E = event two taxable orders are received before the first nontaxable order,

A = event the first order received is taxable, and

B = event the second order received is taxable.

We want to determine the probability of event E. Noting that $E = A \cap B$, we apply the general multiplication rule to conclude that $P(E) = P(A)P(B \mid A)$. From part (c), we know that $P(A) = \lambda_1/(\lambda_1 + \lambda_2)$. Now, because of the lack-of-memory property, given event A occurs, when that taxable order is received, both Poisson processes start anew, probabilistically speaking. Hence, from part (c), we have that $P(B \mid A) = \lambda_1/(\lambda_1 + \lambda_2)$. Thus,

$$P(E) = P(A)P(B \mid A) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2} = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2.$$

e) From the arguments used previously in this problem, we see that each securities order will be taxable with probability $\lambda_1/(\lambda_1 + \lambda_2)$ $\big($and, hence, nontaxable with probability $\lambda_2/(\lambda_1 + \lambda_2)\big)$, independent of the entire past history of both Poisson processes. In other words, we can consider the process of observing the type of each successive securities order a sequence of Bernoulli trials with success probability $p = \lambda_1/(\lambda_1 + \lambda_2)$, where a success denotes a taxable order. Using that interpretation, the probability that n_1 taxable orders are received before n_2 nontaxable orders are received equals the probability, in Bernoulli trials with success probability p, of n_1 successes occur before n_2 failures. However, n_1 successes occur before n_2 failures if and only if at least n_1 successes occur in the first $n_1 + n_2 - 1$ trials. This latter event has probability

$$\sum_{k=n_1}^{n_1+n_2-1} \binom{n_1+n_2-1}{k} p^k (1-p)^{n_1+n_2-1-k}.$$

Hence, the probability that n_1 taxable orders are received before n_2 nontaxable orders equals

$$\sum_{k=n_1}^{n_1+n_2-1} \binom{n_1+n_2-1}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n_1+n_2-1-k}.$$

12.85

a) The counting process $\{N(t) : t \geq 0\}$ gives the total number of event of all m types that occur by time t.
b) We must show that the counting process $\{N(t) : t \geq 0\}$ satisfies the three conditions of Definition 12.1 on page 688 with $\lambda = \lambda_1 + \cdots + \lambda_m$.

First condition:

Because $\{N_k(t) : t \geq 0\}$, $1 \leq k \leq m$, are Poisson processes, we have $N_k(0) = 0$ for all $1 \leq k \leq m$, so that $N(0) = N_1(0) + \cdots + N_m(0) = 0$.

Second condition:

Let $r \in \mathcal{N}$ and $0 \leq t_1 < t_2 < \cdots < t_r$. Because $\{N_k(t) : t \geq 0\}$, $1 \leq k \leq m$, are Poisson processes, the random variables $N_k(t_j) - N_k(t_{j-1})$, $2 \leq j \leq r$, are independent for each $1 \leq k \leq m$. And, as the m Poisson processes are independent, we can apply Proposition 6.13 on page 297 to conclude that

$$\begin{aligned} N(t_j) - N(t_{j-1}) &= \big(N_1(t_j) + \cdots + N_m(t_j)\big) - \big(N_1(t_{j-1}) + \cdots + N_m(t_{j-1})\big) \\ &= \big(N_1(t_j) - N_1(t_{j-1})\big) + \cdots + \big(N_m(t_j) - N_m(t_{j-1})\big), \qquad 2 \leq j \leq r, \end{aligned}$$

are independent random variables. Thus, $\{N(t) : t \geq 0\}$ has independent increments.

Third condition:

Let $0 \leq s < t < \infty$. As $\{N_k(t) : t \geq 0\}$, $1 \leq k \leq m$, are Poisson processes with rates λ_k, $1 \leq k \leq m$, respectively, we have that $N_k(t) - N_k(s) \sim \mathcal{P}\big(\lambda_k(t-s)\big)$ for each $1 \leq k \leq m$. And, as the m Poisson processes are independent, we can apply Proposition 6.20(b) on page 311 to conclude that

$$\begin{aligned} N(t) - N(s) &= \big(N_1(t) + \cdots + N_m(t)\big) - \big(N_1(s) + \cdots + N_m(s)\big) \\ &= \big(N_1(t) - N_1(s)\big) + \cdots + \big(N_m(t) - N_m(s)\big) \\ &\sim \mathcal{P}\big(\lambda_1(t-s) + \cdots + \lambda_m(t-s)\big) = \mathcal{P}\big((\lambda_1 + \cdots + \lambda_m)(t-s)\big). \end{aligned}$$

We have, therefore, now shown that $\{N(t) : t \geq 0\}$ satisfies the three conditions given in Definition 12.1 with $\lambda = \lambda_1 + \cdots + \lambda_m$ and, hence, that it is a Poisson process with rate $\lambda_1 + \cdots + \lambda_m$.
c) Recall that, for a Poisson process, the interarrival times between events are independent and identically distributed exponential random variables with common parameter equal to the rate of the process. Therefore, because of the lack-of-memory property of an exponential random variable, starting at any specified time, the elapsed time until the next event occurs has that same exponential distribution. Now, the probability an event occurring in the $\{N(t) : t \geq 0\}$ process is of Type j equals the probability that, starting at a specified time, the next Type j event occurs before the next occurrence of any of the other types of events. Because $\{N_k(t) : t \geq 0\}$, $1 \leq k \leq m$, are independent Poisson processes with rates λ_k, $1 \leq k \leq m$, respectively, the m occurrence times are independent exponential random variables with parameters λ_k, $1 \leq k \leq m$, respectively. Hence, from Proposition 9.11(b), the probability that, starting at a specified time, the next Type j event occurs before the next occurrence of any of the other types of events is $\lambda_j/(\lambda_1 + \cdots + \lambda_m)$.

12.86 We must show that the counting process $\{Y(t) : t \geq 0\}$ satisfies the three conditions of Definition 12.1 on page 688 with rate $p\lambda$.

First condition:

By definition, $Y(t) = 0$ if $N(t) = 0$. Hence, because, $N(0) = 0$, we have $Y(0) = 0$.

Second condition:

Let $r \in \mathcal{N}$ and $0 \leq t_1 < t_2 < \cdots < t_r$. For $2 \leq j \leq r$, we can write

$$Y(t_j) - Y(t_{j-1}) = I_{\{N(t_j)-N(t_{j-1})>0\}} \sum_{k=N(t_{j-1})+1}^{N(t_j)} X_k = I_{\{N(t_j)-N(t_{j-1})>0\}} \sum_{k=1}^{N(t_j)-N(t_{j-1})} X_{k+N(t_{j-1})}.$$

We note that, because $X_1, X_2, \ldots$ are independent random variables, which are also independent of $\{N(t) : t \geq 0\}$, the distribution of the sum on the right of the preceding display depends only on $N(t_j) - N(t_{j-1})$ and p. Hence, as $N(t_j) - N(t_{j-1})$, $2 \leq j \leq r$, are independent random variables, we conclude from Proposition 6.13 on page 297 that $Y(t_j) - Y(t_{j-1})$, $2 \leq j \leq r$, are independent random variables. Therefore, $\{Y(t) : t \geq 0\}$ has independent increments.

Third condition:

Let $0 \leq s < t < \infty$. Given that $N(t) - N(s) = n$, we see from the just-presented verification of the second condition, that $Y(t) - Y(s) \sim \mathcal{B}(n, p)$. Therefore, from the law of total probability, for each nonnegative integer k,

$$\begin{aligned}
P\big(Y(t) - Y(s) = k\big) &= \sum_{n=k}^{\infty} P\big(N(t) - N(s) = n\big)\, P\big(Y(t) - Y(s) = k \mid N(t) - N(s) = n\big) \\
&= \sum_{n=k}^{\infty} e^{-\lambda(t-s)} \frac{\big(\lambda(t-s)\big)^n}{n!} \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= e^{-p\lambda(t-s)} \frac{\big(p\lambda(t-s)\big)^k}{k!} \sum_{n=k}^{\infty} e^{-(1-p)\lambda(t-s)} \frac{\big((1-p)\lambda(t-s)\big)^{n-k}}{(n-k)!} \\
&= e^{-p\lambda(t-s)} \frac{\big(p\lambda(t-s)\big)^k}{k!} \cdot 1 = e^{-p\lambda(t-s)} \frac{\big(p\lambda(t-s)\big)^k}{k!}.
\end{aligned}$$

Hence, $Y(t) - Y(s) \sim \mathcal{P}\big(p\lambda(t-s)\big)$.

Consequently, we have now shown that the compound Poisson process $\{Y(t) : t \geq 0\}$ is, in this case, a Poisson process with rate $p\lambda$.

Note: An alternative method for solving this problem is to proceed as follows. Each time the specified event occurs, classify it as Type 1 if the corresponding X_k is 1 (which has probability p) and as Type 2 otherwise. We see the that $Y(t)$ equals the number of Type 1 events that occur by time t. Therefore, from Exercise 12.7, we conclude that $\{Y(t) : t \geq 0\}$ is a Poisson process with rate $p\lambda$.

12.87

a) The tacit assumption is that the capacity of the emergency room is infinite.

b) A better model would be an $M/M/1/N$ queueing system, because it would reflect the reality of a finite-capacity emergency room.

c) From Exercise 12.26, for $\lambda \neq \mu$, the steady-state distribution of the number of customers in the system for an $M/M/1/N$ queue is

$$P_n = \begin{cases} \dfrac{(1-\lambda/\mu)(\lambda/\mu)^n}{1-(\lambda/\mu)^{N+1}}, & \text{if } 0 \le n \le N; \\ 0, & \text{if } n \ge N+1. \end{cases}$$

Here we have $\lambda = 6.9$, $\mu = 7.4$, and $N = 40$. Then

$$\frac{\lambda}{\mu} = \frac{6.9}{7.4} \approx 0.93243 \qquad \text{and} \qquad \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}} = \frac{1-6.9/7.4}{1-(6.9/7.4)^{41}} \approx 0.07164.$$

Hence, the steady-state distribution of the number of patients in an emergency room with a capacity of 40 is

$$P_n \approx 0.07164 \cdot (0.93243)^n, \qquad 0 \le n \le 40,$$

and $P_n = 0$ otherwise. In contrast, the steady-state distribution of the number of patients in an emergency room with infinite capacity is, from Example 12.7,

$$P_n \approx 0.06757 \cdot (0.93243)^n, \qquad n \ge 0.$$

As we see, there is not much difference between the $M/M/1$ and $M/M/1/40$ queueing systems.
d) In this case, $N = 100$, so that

$$\frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}} = \frac{1-6.9/7.4}{1-(6.9/7.4)^{101}} \approx 0.06763.$$

Hence, the steady-state distribution of the number of patients in an emergency room with a capacity of 100 is

$$P_n \approx 0.06763 \cdot (0.93243)^n, \qquad 0 \le n \le 100,$$

and $P_n = 0$ otherwise. As we would expect, the steady-state distribution of this $M/M/1/100$ queueing system is even closer to that of the $M/M/1$ queueing system than is the steady-state distribution of the $M/M/1/40$ queueing system.

12.88
a) Measuring time in days, we note that the service facility is an $M/M/1$ queueing system with $\lambda = 5$ and $\mu = 24/4 = 6$. Let X denote the steady-state number of down machines and let T denote the total cost per day to the company due to down machines, including the cost of operating the service facility. Then $T = 247 + 50X$. From Equation (12.24) on page 707,

$$\mathcal{E}(X) = L = \frac{\lambda}{\lambda - \mu} = \frac{5}{6-5} = 5.$$

Hence,

$$\mathcal{E}(T) = \mathcal{E}(247 + 50X) = 247 + 50\mathcal{E}(X) = 247 + 50 \cdot 5 = 497.$$

b) In this case, the service facility is an $M/M/1/$ queueing system with $\lambda = 5$ and $\mu = 24/3 = 8$. Let X denote the steady-state number of down machines, c the daily operating cost for this new service facility, and T the total cost per day to the company due to down machines, including the cost of operating the service facility. Then $T = c + 50X$. From Equation (12.24),

$$\mathcal{E}(X) = L = \frac{\lambda}{\lambda - \mu} = \frac{5}{8-5} = \frac{5}{3}.$$

Hence,

$$\mathcal{E}(T) = \mathcal{E}(c + 50X) = c + 50\mathcal{E}(X) = c + 50 \cdot \frac{5}{3} = c + \frac{250}{3}.$$

Referring to part (a), we see that, for the new service facility to be economically feasible, we must have $c + 250/3 \le 497$, or $c \le 413.67$. Therefore, the maximum daily allowable operating cost for the new service facility is \$413.67.

12.89 We will compare the steady-state expected number of people in the queueing systems in the two configurations. With the current configuration, the queueing system consists of two separate $M/M/1$ queues. Measuring time in hours, the service rate for both of these $M/M/1$ queues is $\mu = 60/4 = 15$, whereas the arrival rates are 12 and 7. Hence, from Equation (12.24) on page 707,

$$L = \frac{12}{15 - 12} + \frac{7}{15 - 7} = \frac{39}{8} = 4.875.$$

With the photocopying facilities in the same room, we have an $M/M/2$ queueing system with service rate 15 and arrival rate $12 + 7 = 19$. Hence, from Equations (12.28) and (12.22) on pages 708 and 704, respectively,

$$L = \frac{19}{15} + \frac{(19/30)(19/15)^2}{2\,(1 - 19/30)^2} \cdot \left(1 + 19/15 + \frac{(19/15)^2}{2\,(1 - 19/30)}\right)^{-1} \approx 2.115.$$

Thus, with the current configuration, the expected number of people in the queueing system is 4.875, whereas, with the photocopying facilities in the same room, the expected number is roughly 2.115. We therefore see that the photocopying facilities would be more effectively used if they were in the same room.

12.90 Measuring time in hours, we have an $M/M/2/3$ queueing system with $\lambda = 25$ and $\mu = 60/6 = 10$. From Equation (12.13) on page 700,

$$\lambda_n = \begin{cases} 25, & \text{if } 0 \le n \le 2; \\ 0, & \text{if } n \ge 3. \end{cases} \qquad \mu_n = \begin{cases} 10, & \text{if } n = 1; \\ 20, & \text{if } n \ge 2. \end{cases}$$

For convenience, let $r = \lambda/\mu = 2.5$ and $\rho = \lambda/2\mu = 1.25$. Then

$$\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} = \begin{cases} r, & \text{if } n = 1; \\ r\rho^{n-1}, & \text{if } n = 2, 3; \\ 0, & \text{if } n \ge 4. \end{cases} = \begin{cases} r\rho^{n-1}, & \text{if } 1 \le n \le 3; \\ 0, & \text{if } n \ge 4. \end{cases}$$

Thus, from Equation (12.18) on page 703,

$$P_n = \begin{cases} r\rho^{n-1} P_0, & \text{if } 1 \le n \le 3; \\ 0, & \text{if } n \ge 4. \end{cases}$$

where

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right)\right)^{-1} = \left(1 + \sum_{n=1}^{3} r\rho^{n-1}\right)^{-1} = \left(1 + r\,\frac{1 - \rho^3}{1 - \rho}\right)^{-1}.$$

a) The (steady-state) expected number of customers in the gas station is

$$L = \sum_{n=0}^{\infty} nP_n = rP_0 \sum_{n=1}^{3} n\rho^{n-1} = r\left(1 + r\,\frac{1 - \rho^3}{1 - \rho}\right)^{-1} \left(1 + 2\rho + 3\rho^2\right) \approx 1.944.$$

b) A potential customer is lost when an arriving customer observes three customers in the gas station. The probability of that happening is

$$P_3 = r\rho^2 P_0 = r\rho^2 \left(1 + r\,\frac{1-\rho^3}{1-\rho}\right)^{-1} = 0.371.$$

Hence, 37.1% of potential customers are lost.
c) Both attendants are idle when and only when there are no customers in the gas station. The probability of that happening is

$$P_0 = \left(1 + r\,\frac{1-\rho^3}{1-\rho}\right)^{-1} \approx 0.095.$$

Hence, both attendants are idle 9.5% of the time.

12.91
a) Answers will vary, but one possibility is a self-service situation.
b) We have $\lambda_n = \lambda$ for all $n \ge 0$, and $\mu_n = n\mu$ for all $n \ge 1$.
c) For $n \ge 1$,

$$\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} = \prod_{k=1}^{n} \frac{\lambda}{k\mu} = \frac{(\lambda/\mu)^n}{n!}.$$

Therefore,

$$\sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right) = \sum_{n=1}^{\infty} \frac{(\lambda/\mu)^n}{n!} = e^{\lambda/\mu} - 1 < \infty.$$

Hence, from Propostion 12.3 on page 703, a steady-state distribution always exists. Furthermore,

$$P_0 = \left(1 + \sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right)\right)^{-1} = \left(1 + \left(e^{\lambda/\mu} - 1\right)\right)^{-1} = e^{-\lambda/\mu}$$

and, for $n \ge 1$,

$$P_n = \left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k}\right) P_0 = \frac{(\lambda/\mu)^n}{n!} e^{-\lambda/\mu}.$$

Consequently, we see that the steady-state number of customers in the queueing system has the Poisson distribution with parameter λ/μ.
d) Referring to part (c), we conclude that the expected number of customers in the queueing system equals λ/μ.
e) Because there are an infinite number of servers, there is never any queue. Thus, we have $L_q = W_q = 0$. Also, in this case, W is simply the expected time it takes a server to serve a customer, which is $1/\mu$.

12.92
a) Here the state of the system is the number of machines that are down. Clearly, $\mu_n = \mu$ for all $n \ge 1$, and $\lambda_n = 0$ for all $n \ge M$. Now suppose that $0 \le n \le M-1$. When n machines are down, there are $M - n$ machines that are working. Because the breakdown time for each machine has an exponential distribution with parameter λ, the time until the next machine breaks down has an exponential distribution with parameter $(M-n)\lambda$. Consequently, we have $\lambda_n = (M-n)\lambda$. Hence,

$$\lambda_n = \begin{cases} (M-n)\lambda, & \text{if } 0 \le n \le M-1; \\ 0, & \text{if } n \ge M. \end{cases} \qquad \mu_n = \mu, \quad n \ge 1.$$

b) Referring to part (a), we see that

$$\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} = \begin{cases} \displaystyle\prod_{k=1}^{n} \frac{\big(M-(k-1)\big)\lambda}{\mu}, & \text{if } 1 \le n \le M; \\ 0, & \text{if } n \ge M+1. \end{cases} = \begin{cases} (M)_n (\lambda/\mu)^n, & \text{if } 1 \le n \le M; \\ 0, & \text{if } n \ge M+1. \end{cases}$$

Hence, from Proposition 12.3 on page 703,

$$P_n = \begin{cases} (M)_n (\lambda/\mu)^n P_0, & \text{if } 1 \le n \le M; \\ 0, & \text{if } n \ge M+1. \end{cases}$$

where

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \left(\prod_{k=1}^{n} \frac{\lambda_{k-1}}{\mu_k} \right)\right)^{-1} = \left(\sum_{n=0}^{M} (M)_n (\lambda/\mu)^n \right)^{-1}.$$

c) Let $r = \lambda/\mu$. For $1 \le n \le M$,

$$nP_n = \big(M - (M-n)\big)P_n = MP_n - (M-n)P_n.$$

However,

$$(M-n)P_n = (M-n)(M)_n r^n P_0 = (M)_{n+1} r^n P_0 = \frac{1}{r}(M)_{n+1} r^{n+1} P_0 = \frac{1}{r} P_{n+1}.$$

Hence,

$$L = \sum_{n=0}^{\infty} nP_n = \sum_{n=1}^{M} MP_n - \sum_{n=1}^{M} (M-n)P_n = M\sum_{n=1}^{M} P_n - \frac{1}{r}\sum_{n=1}^{M} P_{n+1} = M\sum_{n=1}^{M} P_n - \frac{1}{r}\sum_{n=2}^{M} P_n$$

$$= M(1-P_0) - \frac{1}{r}(1 - P_0 - P_1) = M(1-P_0) - \frac{1}{r}(1 - P_0 - MrP_0) = M - \frac{\mu}{\lambda}(1-P_0).$$

Consequently, the steady-state expected number of working machines is

$$M - L = M - \left(M - \frac{\mu}{\lambda}(1-P_0)\right) = \frac{\mu}{\lambda}(1-P_0).$$

d) Here we have $M = 10$, $\lambda = 1/6$, and $\mu = 1/1.5 = 2/3$. From part (a),

$$\lambda_n = \begin{cases} (M-n)/6, & \text{if } 0 \le n \le 9; \\ 0, & \text{if } n \ge 10. \end{cases} \qquad \mu_n = 2/3, \quad n \ge 1.$$

Noting that $\lambda/\mu = (1/6)\big/(2/3) = 1/4$, we get, in view of part (b), that

$$P_n = \begin{cases} (10)_n 4^{-n} P_0, & \text{if } 1 \le n \le 10; \\ 0, & \text{if } n \ge 11. \end{cases}$$

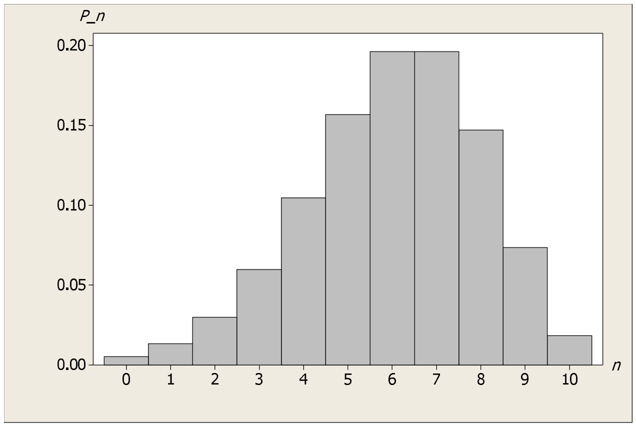
where

$$P_0 = \left(\sum_{n=0}^{10} (10)_n 4^{-n} \right)^{-1} \approx 0.00531.$$

The following table provides the steady-state distribution for the number of down machines:

n	0	1	2	3	4	5	6	7	8	9	10
P_n	0.00531	0.01327	0.02985	0.05971	0.10449	0.15674	0.19592	0.19592	0.14694	0.07347	0.01837

A probability histogram for this steady-state number of down machines is as follows:



From part (c), the expected number of working machines is

$$(\mu/\lambda)(1 - P_0) = 4(1 - P_0) \approx 3.98.$$

On average, roughly four machines are working (and six machines are down).

12.93 For convenience, let $r = \lambda/\mu$.

a) Referring to Equation (12.22) on page 704, we get

$$P_0 = \left(\sum_{n=0}^{2-1} \frac{r^n}{n!} + \frac{r^2}{2!(1 - r/2)}\right)^{-1} = \left(1 + r + \frac{r^2}{2(1 - r/2)}\right)^{-1}$$
$$= \left(1 + r + \frac{r^2}{2 - r}\right)^{-1} = \frac{2 - r}{2 + r} = \frac{2 - \lambda/\mu}{2 + \lambda/\mu} = \frac{2\mu - \lambda}{2\mu + \lambda}.$$

b) Referring to Equation (12.28) on page 708 and then to part (a), we get

$$L = r + \frac{(r/2)r^2}{2!(1 - r/2)^2} P_0 = r + \frac{r^3}{(2 - r)^2} \cdot \frac{2 - r}{2 + r} = r + \frac{r^3}{4 - r^2} = \frac{4r}{4 - r^2}$$
$$= \frac{4(\lambda/\mu)}{4 - (\lambda/\mu)^2} = \frac{4\lambda\mu}{4\mu^2 - \lambda^2} = \frac{4\lambda\mu}{(2\mu - \lambda)(2\mu + \lambda)}.$$

c) In this case, $\lambda = 6.9$ and $\mu = 7.4$. Hence, from parts (a) and (b),

$$P_0 = \frac{2 \cdot 7.4 - 6.9}{2 \cdot 7.4 + 6.9} \approx 0.364$$

and

$$L = \frac{4 \cdot 6.9 \cdot 7.4}{(2 \cdot 7.4 - 6.9)(2 \cdot 7.4 + 6.9)} \approx 1.2.$$

These two results agree with those found in Examples 12.7(b) and 12.9(b), respectively.

12.94 Let us call the three expected numbers L_1, L_2, and L_3.

a) Referring to Exercise 12.93(b), we have

$$L_1 = \frac{4\lambda\mu}{4\mu^2 - \lambda^2}.$$

b) The time it takes a customer to be served is, in this case, the minimum of two exponential random variables with parameter μ, which, as we know, has an exponential distribution with parameter 2μ. Hence, this system is an $M/M/1$ queue with arrival rate λ and service rate 2μ. Referring now to Equation (12.24) on page 707, we conclude that

$$L_2 = \frac{\lambda}{2\mu - \lambda}.$$

c) Each of these $M/M/1$ queueing systems has service rate μ and, from Exercise 12.6, each has arrival rate $\lambda/2$. Referring again to Equation (12.24), we conclude that

$$L_3 = 2 \cdot \frac{\lambda/2}{\mu - \lambda/2} = \frac{2\lambda}{2\mu - \lambda}.$$

d) To compare the three expected numbers, let $\alpha = \lambda/2\mu$. Then

$$L_1 = \frac{4\lambda\mu}{4\mu^2 - \lambda^2} = \frac{4\lambda\mu/4\mu^2}{\left(4\mu^2 - \lambda^2\right)/4\mu^2} = \frac{\lambda/\mu}{1 - (\lambda/2\mu)^2} = \frac{2\alpha}{(1-\alpha)(1+\alpha)},$$

$$L_2 = \frac{\lambda}{2\mu - \lambda} = \frac{\lambda/2\mu}{(2\mu - \lambda)/2\mu} = \frac{\lambda/2\mu}{1 - \lambda/2\mu} = \frac{\alpha}{1-\alpha},$$

and

$$L_3 = \frac{2\lambda}{2\mu - \lambda} = \frac{2\lambda/2\mu}{(2\mu - \lambda)/2\mu} = \frac{\lambda/\mu}{1 - \lambda/2\mu} = \frac{2\alpha}{1-\alpha}.$$

Because $\lambda < 2\mu$, we have $0 < \alpha < 1$. Hence,

$$\frac{\alpha}{1-\alpha} < \frac{2\alpha}{(1-\alpha)(1+\alpha)} < \frac{2\alpha}{1-\alpha},$$

that is, $L_2 < L_1 < L_3$. The efficiency of the three configurations are, from highest to lowest, the second, the first, and the third.

12.95 Let $\mu_i = [\boldsymbol{\mu}]_i$ for $1 \le i \le m$.
a) By convention, we can take $\mathbf{X} \sim \mathcal{N}_m(\boldsymbol{\mu}, \mathbf{0})$. Hence, from Definition 12.5 on page 717,

$$M_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = M_{\mathbf{X}}(\mathbf{t}) = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\mathbf{0}\mathbf{t}} = e^{\boldsymbol{\mu}'\mathbf{t} + 0} = e^{\boldsymbol{\mu}'\mathbf{t}} = e^{\sum_{j=1}^m \mu_j t_j}.$$

b) As $\operatorname{Var}(X_i) = \operatorname{Cov}(X_i, X_i) = [\mathbf{0}]_{ii} = 0$, Proposition 7.7 on page 354 implies that $P(X_i = \mu_i) = 1$; that is, X_i is the constant random variable μ_i. In other words,

$$p_{X_i}(x) = \begin{cases} 1, & \text{if } x = \mu_i; \\ 0, & \text{otherwise.} \end{cases} \qquad 1 \le i \le m.$$

From this result and the fact that a finite intersection of events with probability 1 has probability 1, we get

$$P(\mathbf{X} = \boldsymbol{\mu}) = P\big((X_1, \ldots, X_m) = (\mu_1, \ldots, \mu_m)\big) = P(X_1 = \mu_1, \ldots, X_m = \mu_m) = 1.$$

Thus, $(X_1, \ldots, X_m)$ is the constant vector $(\mu_1, \ldots, \mu_m)$. In other words,

$$p_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \begin{cases} 1, & \text{if } x_i = \mu_i \text{ for } 1 \le i \le m; \\ 0, & \text{otherwise.} \end{cases}$$

We note in passing that we can use the preceding result to obtain the result of part (a) as follows:

$$M_{X_1,\ldots,X_m}(t_1, \ldots, t_m) = M_{\mathbf{X}}(\mathbf{t}) = \mathcal{E}\left(e^{\mathbf{X}'\mathbf{t}}\right) = \sum_{\mathbf{x}} e^{\mathbf{x}'\mathbf{t}} P(\mathbf{X} = \mathbf{x}) = e^{\boldsymbol{\mu}'\mathbf{t}} = e^{\sum_{j=1}^m \mu_j t_j}.$$

12.96 From Proposition 12.7 on page 719, we know that $\mathbf{Y} \sim \mathcal{N}_m\big(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\big)$. Thus, it remains to show that $Y_1, \ldots, Y_m$ are nonsingular. Because $\mathbf{X}$ is nonsingular, Proposition 12.5 on page 717 implies that $\boldsymbol{\Sigma}$ is a nonsingular matrix, which, in turn, implies that $\det \boldsymbol{\Sigma} \neq 0$. And, because $\mathbf{B}$ is a nonsingular matrix, $\det \mathbf{B} \neq 0$. Hence,

$$\det \boldsymbol{\Sigma}_{\mathbf{Y}} = \det\big(\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\big) = \det \mathbf{B} \cdot \det \boldsymbol{\Sigma} \cdot \det \mathbf{B}' = (\det \mathbf{B})^2 \det \boldsymbol{\Sigma} \neq 0.$$

Therefore, $\boldsymbol{\Sigma}_{\mathbf{Y}}$ is a nonsingular matrix, which, by Proposition 12.5, implies that $Y_1, \ldots, Y_m$ are nonsingular multivariate normal random variables.

12.97 We have

$$\begin{aligned}\operatorname{Var}(Y_1) &= \operatorname{Var}(X_1 - 2X_2) = \operatorname{Cov}(X_1 - 2X_2, X_1 - 2X_2) \\ &= \operatorname{Cov}(X_1, X_1) - 2\operatorname{Cov}(X_1, X_2) - 2\operatorname{Cov}(X_2, X_1) + 4\operatorname{Cov}(X_2, X_2) \\ &= 1.0 - 2 \cdot 0.1 - 2 \cdot 0.1 + 4 \cdot 1.0 = 4.6,\end{aligned}$$

$$\begin{aligned}\operatorname{Var}(Y_2) &= \operatorname{Var}(X_3 - 3X_4) = \operatorname{Cov}(X_3 - 3X_4, X_3 - 3X_4) \\ &= \operatorname{Cov}(X_3, X_3) - 3\operatorname{Cov}(X_3, X_4) - 3\operatorname{Cov}(X_4, X_3) + 9\operatorname{Cov}(X_4, X_4) \\ &= 4.0 - 3 \cdot 2.4 - 3 \cdot 2.4 + 9 \cdot 4.0 = 25.6,\end{aligned}$$

and

$$\begin{aligned}\operatorname{Cov}(Y_1, Y_2) &= \operatorname{Cov}(X_1 - 2X_2, X_3 - 3X_4) \\ &= \operatorname{Cov}(X_1, X_3) - 3\operatorname{Cov}(X_1, X_4) - 2\operatorname{Cov}(X_2, X_3) + 6\operatorname{Cov}(X_2, X_4) \\ &= 0.4 - 3 \cdot 0.6 - 2 \cdot 1.0 + 6 \cdot 1.0 = 2.6.\end{aligned}$$

Consequently,

$$\rho(Y_1, Y_2) = \frac{\operatorname{Cov}(Y_1, Y_2)}{\sqrt{\operatorname{Var}(Y_1)\operatorname{Var}(Y_2)}} = \frac{2.6}{\sqrt{4.6 \cdot 25.6}} \approx 0.24.$$

12.98

a) Let $Y_1 = X_1$ and $Y_2 = X_2 - cX_1$. We can write $\mathbf{Y} = \mathbf{B}\mathbf{X}$, where

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ -c & 1 & 0 \end{bmatrix}.$$

From Proposition 12.7 on page 719, $\mathbf{Y}$ is bivariate normal. Hence, from Proposition 10.9 on page 613, Y_1 and Y_2 are independent if and only if they are uncorrelated. However,

$$\operatorname{Cov}(Y_1, Y_2) = \operatorname{Cov}(X_1, X_2 - cX_1) = \operatorname{Cov}(X_1, X_2) - c\operatorname{Cov}(X_1, X_1) = 1 - c \cdot 4 = 1 - 4c.$$

Hence, Y_1 and Y_2 are independent if and only if $c = 1/4$.

b) From Proposition 12.8 on page 719, X_1 and X_2 are bivariate normal with

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}.$$

c) From Proposition 12.6 on page 718 (or Proposition 12.7 on page 719), we know that $X_1 - 2X_2 + 3X_3$ is normally distributed. We have

$$\mathcal{E}(X_1 - 2X_2 + 3X_3) = \mathcal{E}(X_1) - 2\mathcal{E}(X_2) + 3\mathcal{E}(X_3) = -1 - 2 \cdot 0 + 3 \cdot 2 = 5$$

and

$$\begin{aligned}\operatorname{Var}(X_1 - 2X_2 + 3X_3) &= \operatorname{Cov}(X_1 - 2X_2 + 3X_3, X_1 - 2X_2 + 3X_3)\\ &= \operatorname{Cov}(X_1, X_1) - 2\operatorname{Cov}(X_1, X_2) + 3\operatorname{Cov}(X_1, X_3)\\ &\qquad - 2\operatorname{Cov}(X_2, X_1) + 4\operatorname{Cov}(X_2, X_2) - 6\operatorname{Cov}(X_2, X_3)\\ &\qquad + 3\operatorname{Cov}(X_3, X_1) - 6\operatorname{Cov}(X_3, X_2) + 9\operatorname{Cov}(X_3, X_3)\\ &= 4 - 2\cdot 1 + 3\cdot 2 - 2\cdot 1 + 4\cdot 1 - 6\cdot 0 + 3\cdot 2 + 9\cdot 3\\ &= 43.\end{aligned}$$

Hence, $X_1 - 2X_2 + 3X_3 \sim \mathcal{N}(5, 43)$.

d) Let $Y_1 = X_1 + X_2 - X_3$ and $Y_2 = X_1 - 2X_2$. We can write $\mathbf{Y} = \mathbf{BX}$, where

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix}.$$

Hence, by Proposition 12.7 on page 719, $\mathbf{Y} \sim \mathcal{N}_2\left(\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}'\right)$. Now,

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{B\Sigma B}' = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & -1 \\ -1 & 4 \end{bmatrix}.$$

We see that $\det \boldsymbol{\Sigma}_{\mathbf{Y}} \neq 0$ and, therefore, $\boldsymbol{\Sigma}_{\mathbf{Y}}$ is a nonsingular matrix. Applying Proposition 12.5 on page 717, we conclude that Y_1 and Y_2 are nonsingular bivariate normal random variables.

12.99

a) Referring to $\boldsymbol{\Sigma}$, we see that

$$\rho(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_2)}} = \frac{60}{\sqrt{400\cdot 25}} = 0.6,$$

$$\rho(X_1, X_3) = \frac{\operatorname{Cov}(X_1, X_3)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_3)}} = \frac{72}{\sqrt{400\cdot 36}} = 0.6,$$

$$\rho(X_2, X_3) = \frac{\operatorname{Cov}(X_2, X_3)}{\sqrt{\operatorname{Var}(X_2)\operatorname{Var}(X_3)}} = \frac{27}{\sqrt{25\cdot 36}} = 0.9.$$

b) From $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and Proposition 12.8 on page 719, we find that

$$X_1 \sim \mathcal{N}(150, 400), \qquad X_2 \sim \mathcal{N}(120, 25), \qquad X_3 \sim \mathcal{N}(80, 36).$$

c) Again from $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and Proposition 12.8 on page 719, we find that

$$(X_1, X_2) \sim \mathcal{N}_2\left(\begin{bmatrix} 150 \\ 120 \end{bmatrix}, \begin{bmatrix} 400 & 60 \\ 60 & 25 \end{bmatrix}\right),$$

$$(X_1, X_3) \sim \mathcal{N}_2\left(\begin{bmatrix} 150 \\ 80 \end{bmatrix}, \begin{bmatrix} 400 & 72 \\ 72 & 36 \end{bmatrix}\right),$$

$$(X_2, X_3) \sim \mathcal{N}_2\left(\begin{bmatrix} 120 \\ 80 \end{bmatrix}, \begin{bmatrix} 25 & 27 \\ 27 & 36 \end{bmatrix}\right).$$

d) Here we use the partition

$$\mathbf{X} = \begin{bmatrix} [\,X_1\,] \\ \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} [\,150\,] \\ \begin{bmatrix} 120 \\ 80 \end{bmatrix} \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} [\,400\,] & [\,60 \quad 72\,] \\ \begin{bmatrix} 60 \\ 72 \end{bmatrix} & \begin{bmatrix} 25 & 27 \\ 27 & 36 \end{bmatrix} \end{bmatrix}.$$

We see that

$$\boldsymbol{\Sigma}_{22}^{-1} = \frac{1}{171}\begin{bmatrix} 36 & -27 \\ -27 & 25 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) &= 150 + \frac{1}{171}[\,60 \quad 72\,]\begin{bmatrix} 36 & -27 \\ -27 & 25 \end{bmatrix}\begin{bmatrix} 130-120 \\ 90-80 \end{bmatrix} \\ &= 150 + \frac{1}{171}[\,216 \quad 180\,]\begin{bmatrix} 10 \\ 10 \end{bmatrix} = 150 + \frac{3960}{171} \\ &= 173.2 \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} &= 400 - \frac{1}{171}[\,60 \quad 72\,]\begin{bmatrix} 36 & -27 \\ -27 & 25 \end{bmatrix}\begin{bmatrix} 60 \\ 72 \end{bmatrix} \\ &= 400 - \frac{1}{171}[\,216 \quad 180\,]\begin{bmatrix} 60 \\ 72 \end{bmatrix} = 400 - \frac{25920}{171} \\ &= 248.4. \end{aligned}$$

Hence, from Exercise 12.58(a), we have $X_{1\,|\,X_2=130,\,X_3=90} \sim \mathcal{N}(173.2, 248.4)$.

e) Here we use the partition

$$\mathbf{X} = \begin{bmatrix} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \\ [\,X_1\,] \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \begin{bmatrix} 120 \\ 80 \end{bmatrix} \\ [\,150\,] \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \begin{bmatrix} 25 & 27 \\ 27 & 36 \end{bmatrix} & \begin{bmatrix} 60 \\ 72 \end{bmatrix} \\ [\,60 \quad 72\,] & [\,400\,] \end{bmatrix}.$$

We see that $\boldsymbol{\Sigma}_{22}^{-1} = 1/400$. Therefore,

$$\begin{aligned} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) &= \begin{bmatrix} 120 \\ 80 \end{bmatrix} + \frac{1}{400}\begin{bmatrix} 60 \\ 72 \end{bmatrix}[\,200 - 150\,] = \begin{bmatrix} 120 \\ 80 \end{bmatrix} + \frac{50}{400}\begin{bmatrix} 60 \\ 72 \end{bmatrix} \\ &= \begin{bmatrix} 127.5 \\ 89 \end{bmatrix} \end{aligned}$$

and

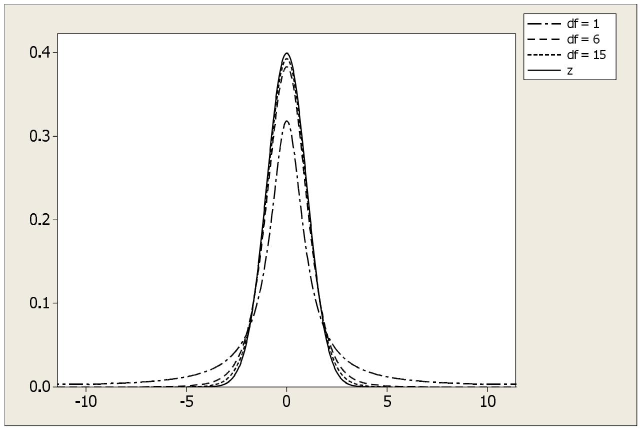
$$\begin{aligned} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} &= \begin{bmatrix} 25 & 27 \\ 27 & 36 \end{bmatrix} - \frac{1}{400}\begin{bmatrix} 60 \\ 72 \end{bmatrix}[\,60 \quad 72\,] = \begin{bmatrix} 25 & 27 \\ 27 & 36 \end{bmatrix} - \frac{1}{400}\begin{bmatrix} 3600 & 4320 \\ 4320 & 5184 \end{bmatrix} \\ &= \begin{bmatrix} 16 & 16.2 \\ 16.2 & 23.04 \end{bmatrix} \end{aligned}$$

Hence, from Exercise 12.58(a), we have

$$(X_2, X_3)_{|X_1=200} \sim \mathcal{N}_2\left(\begin{bmatrix} 127.5 \\ 89 \end{bmatrix}, \begin{bmatrix} 16 & 16.2 \\ 16.2 & 23.04 \end{bmatrix}\right).$$

12.100

a) The required graph is as follows:



b) All Student's t-distributions have greater spread than the standard normal distribution and, the greater the number of degrees of freedom, the more a Student's t-distribution resembles the standard normal distribution.

c) From the hint, as $n \to \infty$, we have $S_n \to \sigma$. Hence, for large n,

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \approx \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

From Proposition 12.12, the random variable on the left has the Student's t-distribution with $n-1$ degrees of freedom and, from Proposition 12.9, the random variable on the right has the standard normal distribution. Heuristically, then, a Student's t-distribution approaches the standard normal distribution as the number of degrees of freedom increases without bound.

12.101

a) Let U_1 and U_2 be independent random variables that have the chi-square distributions with degrees of freedom ν_1 and ν_2, respectively. From Lemma 12.2 on page 727, the random variable $(U_1/\nu_1)\big/(U_2/\nu_2)$ has the F-distribution with degrees of freedom ν_1 and ν_2. Hence, F has the same probability distribution as $(U_1/\nu_1)\big/(U_2/\nu_2)$. Consequently, $1/F$ has the same probability distribution as

$$\frac{1}{(U_1/\nu_1)\big/(U_2/\nu_2)} = (U_2/\nu_2)\big/(U_1/\nu_1),$$

which, again, by Lemma 12.2, is the F-distribution with degrees of freedom ν_2 and ν_1.

b) Let $g(w) = 1/w$ and let $Y = g(F)$. Note that $1/F = Y$. Now, g is defined, strictly decreasing, and differentiable on the range of F, and $g^{-1}(y) = 1/y$. We apply the univariate transformation theorem and Equation (12.54) on page 727 to conclude that a PDF of the random variable Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{|g'(w)|} f_F(w) = \frac{1}{\left|-1/w^2\right|} f_F(w) = w^2 f_F(w) = \frac{1}{y^2} f_F(1/y) \\ &= \frac{1}{y^2} \frac{(\nu_1/\nu_2)^{\nu_1/2}\Gamma\big((\nu_1+\nu_2)/2\big)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \frac{(1/y)^{\nu_1/2-1}}{\big(1+(\nu_1/\nu_2)(1/y)\big)^{(\nu_1+\nu_2)/2}} \end{aligned}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise.

However,

$$\frac{1}{y^2}\frac{(1/y)^{\nu_1/2-1}}{\big(1+(\nu_1/\nu_2)(1/y)\big)^{(\nu_1+\nu_2)/2}} = \frac{\big((\nu_2/\nu_1)y\big)^{(\nu_1+\nu_2)/2}}{\big((\nu_2/\nu_1)y\big)^{(\nu_1+\nu_2)/2}} \cdot \frac{1}{y^2}\frac{(1/y)^{\nu_1/2-1}}{\big(1+(\nu_1/\nu_2)(1/y)\big)^{(\nu_1+\nu_2)/2}}$$
$$= (\nu_2/\nu_1)^{(\nu_1+\nu_2)/2}\frac{y^{\nu_2/2-1}}{\big(1+(\nu_2/\nu_1)y\big)^{(\nu_2+\nu_1)/2}}$$

and

$$(\nu_2/\nu_1)^{(\nu_1+\nu_2)/2} \cdot \frac{(\nu_1/\nu_2)^{\nu_1/2}\Gamma\big((\nu_1+\nu_2)/2\big)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} = \frac{(\nu_2/\nu_1)^{\nu_2/2}\Gamma\big((\nu_2+\nu_1)/2\big)}{\Gamma(\nu_2/2)\Gamma(\nu_1/2)}.$$

Consequently,

$$f_Y(y) = \frac{(\nu_2/\nu_1)^{\nu_2/2}\Gamma\big((\nu_2+\nu_1)/2\big)}{\Gamma(\nu_2/2)\Gamma(\nu_1/2)}\frac{y^{\nu_2/2-1}}{\big(1+(\nu_2/\nu_1)y\big)^{(\nu_2+\nu_1)/2}}$$

if $y > 0$, and $f_Y(y) = 0$ otherwise. Thus, we see that $1/F$ $(= Y)$ has the F-distribution with degrees of freedom ν_2 and ν_1.

12.102 For convenience, set

$$a = \nu_1/\nu_2 \qquad \text{and} \qquad b = \frac{(\nu_1/\nu_2)^{\nu_1/2}\Gamma\big((\nu_1+\nu_2)/2\big)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}.$$

Let $g(w) = 1/(1+aw)$ and let $Y = g(F)$. We want to find and identify the probability distribution of Y. Now, g is defined, strictly decreasing, and differentiable on the range of F, and $g^{-1}(y) = (1-y)/ay$. We apply the univariate transformation theorem and Equation (12.54) on page 727 to conclude that a PDF of the random variable Y is

$$f_Y(y) = \frac{1}{|g'(w)|}f_F(w) = \frac{1}{\big|-a/(1+aw)^2\big|}f_F(w) = \frac{(1+aw)^2}{a}\frac{bw^{\nu_1/2-1}}{\big(1+aw\big)^{(\nu_1+\nu_2)/2}}$$
$$= \frac{b}{a}y^{(\nu_1+\nu_2)/2-2}\big((1-y)/ay\big)^{\nu_1/2-1} = \frac{b}{a^{\nu_1/2}}y^{\nu_2/2-1}(1-y)^{\nu_1/2-1}$$

if $0 < y < 1$, and $f_Y(y) = 0$ otherwise. Because a and b are constants, we can now conclude that Y has the beta distribution with parameters $\nu_2/2$ and $\nu_1/2$. However, as a check, we note that

$$\frac{b}{a^{\nu_1/2}} = \frac{(\nu_1/\nu_2)^{\nu_1/2}\Gamma\big((\nu_1+\nu_2)/2\big)\big/\Gamma(\nu_1/2)\Gamma(\nu_2/2)}{(\nu_1/\nu_2)^{\nu_1/2}} = \frac{\Gamma\big((\nu_2+\nu_1)/2\big)}{\Gamma(\nu_2/2)\Gamma(\nu_1/2)} = \frac{1}{B(\nu_2/2,\, \nu_1/2)}.$$

12.103

a) From Proposition 12.11 on page 725, we know that $\chi_n^2 = (n-1)S_n^2/\sigma^2$ has the chi-square distribution with $n-1$ degrees of freedom. Hence,

$$P\left(\chi^2_{(1-\gamma)/2,n-1} \le \frac{(n-1)S_n^2}{\sigma^2} \le \chi^2_{(1+\gamma)/2,n-1}\right) = P\left(\chi^2_{(1-\gamma)/2,n-1} \le \chi_n^2 \le \chi^2_{(1+\gamma)/2,n-1}\right)$$
$$= F_{\chi_n^2}\big(\chi^2_{(1+\gamma)/2,n-1}\big) - F_{\chi_n^2}\big(\chi^2_{(1-\gamma)/2,n-1}\big)$$
$$= \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma.$$

Solving for σ^2 in the inequalities in the first term of the preceding display yields

$$P\left(\frac{(n-1)S_n^2}{\chi^2_{(1+\gamma)/2,n-1}} \le \sigma^2 \le \frac{(n-1)S_n^2}{\chi^2_{(1-\gamma)/2,n-1}}\right) = \gamma.$$

b) From part (a) chances are $100\gamma\%$ that the random interval from

$$\frac{(n-1)S_n^2}{\chi^2_{(1+\gamma)/2,n-1}} \quad \text{to} \quad \frac{(n-1)S_n^2}{\chi^2_{(1-\gamma)/2,n-1}}$$

contains σ^2. Hence, we can be $100\gamma\%$ confident that any such computed interval contains σ^2. In other words, the interval from

$$\frac{(n-1)s_n^2}{\chi^2_{(1+\gamma)/2,n-1}} \quad \text{to} \quad \frac{(n-1)s_n^2}{\chi^2_{(1-\gamma)/2,n-1}},$$

where s_n^2 is the observed value of S_n^2, is a $100\gamma\%$ confidence interval for σ^2.

12.104

a) We have $\gamma = 0.95$ so that $(1-\gamma)/2 = 0.025$ and $(1+\gamma)/2 = 0.975$. Because $n = 12$,

$$\chi^2_{(1-\gamma)/2,n-1} = \chi^2_{0.025,11} = 3.816 \qquad \text{and} \qquad \chi^2_{(1+\gamma)/2,n-1} = \chi^2_{0.975,11} = 21.920.$$

From the data, $s_n^2 = 1.148$. Referring now to Exercise 12.103(b), we conclude that a 95% confidence interval for σ^2 is from

$$\frac{(12-1)\cdot 1.148}{21.920} \quad \text{to} \quad \frac{(12-1)\cdot 1.148}{3.816},$$

or from 0.58 to 3.31. Taking square roots, we find that a 95% confidence interval for the standard deviation of highway gas mileages for all cars of the model and year in question is from 0.76 mpg to 1.82 mpg.

b) We can be 95% confident that the standard deviation of highway gas mileages for all cars of the model and year in question is somewhere between 0.76 mpg and 1.82 mpg.

c) From Proposition 12.11 on page 725, we must assume that highway gas mileages for all cars of the model and year in question are normally distributed. Technically, for the independence assumption in Proposition 12.11, we must, in addition, assume that the sampling is with replacement. However, because the sample size is small relative to the population size, sampling without replacement is also acceptable.

12.105

a) We begin by noting that

$$\mu_D = \mathcal{E}(D) = \mathcal{E}(X-Y) = \mathcal{E}(X) - \mathcal{E}(Y) = \mu_X - \mu_Y.$$

Next we note that, if $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent random vectors, then the random variables $D_1, \ldots, D_n$, where $D_j = X_j - Y_j$ for $1 \le j \le n$, are also independent. Hence, from Proposition 12.12 on page 726,

$$T_n = \frac{\bar{D}_n - (\mu_X - \mu_Y)}{S_n/\sqrt{n}} = \frac{\bar{D}_n - \mu_D}{S_n/\sqrt{n}}$$

has the Student's t-distribution with $n-1$ degrees of freedom.

Using the fact that a t-distribution is continuous and symmetric, we conclude that

$$\begin{aligned}
P&\left(-t_{(1+\gamma)/2,n-1} \le \frac{\bar{D}_n - (\mu_X - \mu_Y)}{S_n/\sqrt{n}} \le t_{(1+\gamma)/2,n-1}\right) \\
&= P\left(-t_{(1+\gamma)/2,n-1} \le T_n \le t_{(1+\gamma)/2,n-1}\right) \\
&= F_{T_n}\left(t_{(1+\gamma)/2,n-1}\right) - F_{T_n}\left(-t_{(1+\gamma)/2,n-1}\right) \\
&= 2F_{T_n}\left(t_{(1+\gamma)/2,n-1}\right) - 1 = 2 \cdot \frac{1+\gamma}{2} - 1 = \gamma.
\end{aligned}$$

Solving for $\mu_X - \mu_Y$ in the inequalities in the first term in the preceding display yields

$$P\left(\bar{D}_n - t_{(1+\gamma)/2,n-1} \cdot S_n/\sqrt{n} \le \mu_X - \mu_Y \le \bar{D}_n + t_{(1+\gamma)/2,n-1} \cdot S_n/\sqrt{n}\right) = \gamma.$$

b) From part (a), chances are $100\gamma\%$ that the random interval from

$$\bar{D}_n - t_{(1+\gamma)/2,n-1} \cdot S_n/\sqrt{n} \qquad \text{to} \qquad \bar{D}_n + t_{(1+\gamma)/2,n-1} \cdot S_n/\sqrt{n}$$

contains $\mu_X - \mu_Y$. Thus, we can be $100\gamma\%$ confident that any such computed interval contains $\mu_X - \mu_Y$. In other words, the interval from

$$\bar{d}_n - t_{(1+\gamma)/2,n-1} \cdot s_n/\sqrt{n} \qquad \text{to} \qquad \bar{d}_n + t_{(1+\gamma)/2,n-1} \cdot s_n/\sqrt{n},$$

where $\bar{d}_n$ and s_n are the observed values of $\bar{D}_n$ and S_n, respectively, is a $100\gamma\%$ confidence interval for $\mu_X - \mu_Y$.

c) Yes, because a linear combination of multivariate normal random variables is normally distributed. In particular, then, if X and Y are bivariate normal random variables, then their difference, $D = X - Y$, which is a linear combination of X and Y, is normally distributed, as required.

12.106

a) We have $\gamma = 0.95$ so that $(1+\gamma)/2 = 0.975$. As $n = 10$, $t_{(1+\gamma)/2,n-1} = t_{0.975,9} = 2.262$. Also, from the data, $\bar{d}_n = 3.3$ and $s_n = 4.715$. Referring now to Exercise 12.105(b), we conclude that a 95% confidence interval for $\mu_X - \mu_Y$ is from

$$3.3 - 2.262 \cdot 4.715/\sqrt{10} \qquad \text{to} \qquad 3.3 + 2.262 \cdot 4.715/\sqrt{10},$$

or from -0.07 yr to 6.67 yr. We can be 95% confident that the difference between the mean age of married men and the mean age of married women is somewhere between -0.07 yr to 6.67 yr.

b) As we know, the mean of the difference between two random variables is the difference between the means of the two random variables. Consequently, in view of part (a), a 95% confidence interval for the mean age difference of married couples is from -0.07 yr to 6.67 yr. We can be 95% confident that the mean age difference of married couples is somewhere between -0.07 yr to 6.67 yr.

c) No, it's only required that the differences between the ages of married couples be normally distributed.

d) No, we can't necessarily conclude that the differences between the ages of married couples are normally distributed just because both the ages of married men and the ages of married women are normally distributed. Indeed, the difference between two normally distributed random variables is not always normally distributed. An example is provided by the random variables considered in Exercise 10.113.

Theory Exercises

12.107 Let $n_1 \le n_2 \le \cdots \le n_m$ be nonnegative integers. Using the independent-increments and distributional properties of a Poisson process, we get

$$
\begin{aligned}
&P\big(N(t_1) = n_1, \dots, N(t_m) = n_m\big) \\
&\quad = P\big(N(t_1) = n_1,\ N(t_2) - N(t_1) = n_2 - n_1, \dots, N(t_m) - N(t_{m-1}) = n_m - n_{m-1}\big) \\
&\quad = P\big(N(t_1) = n_1\big)P\big(N(t_2) - N(t_1) = n_2 - n_1\big)\cdots P\big(N(t_m) - N(t_{m-1}) = n_m - n_{m-1}\big) \\
&\quad = e^{-\lambda t_1}\frac{(\lambda t_1)^{n_1}}{n_1!}e^{-\lambda(t_2-t_1)}\frac{\big(\lambda(t_2-t_1)\big)^{n_2-n_1}}{(n_2-n_1)!}\cdots e^{-\lambda(t_m-t_{m-1})}\frac{\big(\lambda(t_m-t_{m-1})\big)^{n_m-n_{m-1}}}{(n_m-n_{m-1})!} \\
&\quad = e^{-\lambda t_m}\lambda^{n_m}\frac{t_1^{n_1}}{n_1!}\cdot\frac{(t_2-t_1)^{n_2-n_1}}{(n_2-n_1)!}\cdots\frac{(t_m-t_{m-1})^{n_m-n_{m-1}}}{(n_m-n_{m-1})!}.
\end{aligned}
$$

Hence, with the convention that $x_0 = 0$ and $t_0 = 0$, we have

$$
p_{N(t_1),\dots,N(t_m)}(x_1,\dots,x_m) = e^{-\lambda t_m}\lambda^{x_m}\prod_{k=1}^{m}\frac{(t_k - t_{k-1})^{x_k - x_{k-1}}}{(x_k - x_{k-1})!}
$$

if $x_1, \dots, x_m$ are nonnegative integers with $x_1 \le \cdots \le x_m$, and $p_{N(t_1),\dots,N(t_m)}(x_1,\dots,x_m) = 0$ otherwise.

12.108 We prove that (a) $\Rightarrow$ (d) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a).

(a) $\Rightarrow$ (d): This result is true because mutually independent random variables are necessarily pairwise independent.

(d) $\Rightarrow$ (c): Let $i \ne j$. From (d), we know that X_i and X_j are independent random variables and, hence, that $\mathcal{E}(X_i X_j) = \mathcal{E}(X_i)\,\mathcal{E}(X_j)$. Thus, $\operatorname{Cov}(X_i, X_j) = 0$ and, therefore, X_i and X_j are uncorrelated.

(c) $\Rightarrow$ (b): Because $\operatorname{Cov}(X_i, X_j) = 0$ for $i \ne j$, the off-diagonal entries of $\mathbf{\Sigma}$ all equal 0. Hence, $\mathbf{\Sigma}$ is a diagonal matrix.

(b) $\Rightarrow$ (a): Let $Y_1, \dots, Y_m$ be independent normal random variables such that $\mathcal{E}(Y_i) = \mathcal{E}(X_i)$ and $\operatorname{Var}(Y_i) = \operatorname{Var}(X_i)$ for all $1 \le i \le m$. As we know, because they are independent normal random variables, $Y_1, \dots, Y_m$ are multivariate normal random variables with a diagonal covariance matrix. By assumption, the covariance matrix of $X_1, \dots, X_m$ is also diagonal. Furthermore, because $\operatorname{Var}(Y_i) = \operatorname{Var}(X_i)$ for $1 \le i \le m$, the diagonal entries of the two covariance matrices are equal. Thus, $\mathbf{\Sigma_Y} = \mathbf{\Sigma_X}$. Also, because $\mathcal{E}(Y_i) = \mathcal{E}(X_i)$ for $1 \le i \le m$, we have $\boldsymbol{\mu}_\mathbf{Y} = \boldsymbol{\mu}_\mathbf{X}$. Hence, we have shown that $X_1, \dots, X_m$ and $Y_1, \dots, Y_m$ have the same mean vector and covariance matrix. As the mean vector and covariance matrix of multivariate normal random variables determine the MGF and, therefore, the joint distribution, we conclude that $X_1, \dots, X_m$ are independent random variables.

Advanced Exercises

12.109 Let $0 \le a < b$. Then, because $\lambda > 0$,

$$
0 < \lambda(b-a) = \mathcal{E}\big(N(b) - N(a)\big) = \sum_{n=0}^{\infty} nP\big(N(b) - N(a) = n\big) = \sum_{n=1}^{\infty} nP\big(N(b) - N(a) = n\big).
$$

Thus, there is an $n \in \mathcal{N}$ such that $P\big(N(b) - N(a) = n\big) > 0$ and, hence, $P\big(N(b) - N(a) \ge 1\big) > 0$. Now let $0 \le s < t < \infty$, let M be a positive real number, and let $N = \lceil M \rceil$. Consider the partition $s = t_0 < t_1 < \cdots < t_N = t$ of the interval $(s, t]$, where $t_j = s + j(t-s)/N$ for $0 \le j \le N$. Applying the domination principle and the assumed independent increments, we get

$$P\bigl(N(t) - N(s) \ge M\bigr) \ge P\bigl(N(t) - N(s) \ge N\bigr) \ge P\left(\bigcap_{j=1}^{N} \bigl\{N(t_j) - N(t_{j-1}) \ge 1\bigr\}\right)$$
$$= \prod_{j=1}^{N} P\bigl(N(t_j) - N(t_{j-1}) \ge 1\bigr) > 0.$$

Consequently, the random variable $N(t) - N(s)$ is unbounded.

12.110

a) Let $X \sim \mathcal{U}(0, t)$. For $1 \le j \le m$, let $E_j = \bigl\{X \in (t_{j-1}, t_j]\bigr\}$. We note that $E_1, \ldots, E_m$ are mutually exclusive and exhaustive events with probabilities $p_1, \ldots, p_m$, where

$$p_j = P(E_j) = P\bigl(X \in (t_{j-1}, t_j]\bigr) = \frac{t_j - t_{j-1}}{t}, \qquad 1 \le j \le m.$$

Thus, we see that, in n independent repetitions of the random experiment, $Y_1, \ldots, Y_m$ give the number of times that events $E_1, \ldots, E_m$ occur, respectively. Hence, $Y_1, \ldots, Y_m$ have the multinomial distribution with parameters n and $p_1, \ldots, p_m$. Referring now to Proposition 6.7 on page 277, we conclude that

$$p_{Y_1,\ldots,Y_m}(y_1, \ldots, y_m) = \binom{n}{y_1, \ldots, y_m} p_1^{y_1} \cdots p_m^{y_m} = \binom{n}{y_1, \ldots, y_m} \frac{(t_1 - t_0)^{y_1} \cdots (t_m - t_{m-1})^{y_m}}{t^n}$$

if $y_1, \ldots, y_m$ are nonnegative integers whose sum is n, and $p_{Y_1,\ldots,Y_m}(y_1, \ldots, y_m) = 0$ otherwise.

b) Let $n_1, \ldots, n_m$ be nonnegative integers whose sum is n. Note that, if $Z_1 = n_1, \ldots, Z_m = n_m$, then $N(t) = n$. Hence, from the conditional probability rule and the independent increments of a Poisson process, we get

$$\begin{aligned}
&P\bigl(Z_1 = n_1, \ldots, Z_m = n_m \mid N(t) = n\bigr) \\
&\quad = P\bigl(N(t_1) - N(t_0) = n_1, \ldots, N(t_m) - N(t_{m-1}) = n_m \mid N(t) = n\bigr) \\
&\quad = \frac{P\bigl(N(t_1) - N(t_0) = n_1, \ldots, N(t_m) - N(t_{m-1}) = n_m,\ N(t) = n\bigr)}{P\bigl(N(t) = n\bigr)} \\
&\quad = \frac{P\bigl(N(t_1) - N(t_0) = n_1, \ldots, N(t_m) - N(t_{m-1}) = n_m\bigr)}{P\bigl(N(t) = n\bigr)} \\
&\quad = \frac{P\bigl(N(t_1) - N(t_0) = n_1\bigr) \cdots P\bigl(N(t_m) - N(t_{m-1}) = n_m\bigr)}{P\bigl(N(t) = n\bigr)} \\
&\quad = \frac{e^{-\lambda(t_1 - t_0)} \dfrac{\bigl(\lambda(t_1 - t_0)\bigr)^{n_1}}{n_1!} \cdots e^{-\lambda(t_m - t_{m-1})} \dfrac{\bigl(\lambda(t_m - t_{m-1})\bigr)^{n_m}}{n_m!}}{e^{-\lambda t} \dfrac{(\lambda t)^n}{n!}} \\
&\quad = \binom{n}{n_1, \ldots, n_m} \frac{(t_1 - t_0)^{n_1} \cdots (t_m - t_{m-1})^{n_m}}{t^n}.
\end{aligned}$$

Hence,

$$p_{Z_1,\ldots,Z_m \mid N(t)}(z_1,\ldots,z_m \mid n) = \binom{n}{z_1,\ldots,z_m}\frac{(t_1-t_0)^{z_1}\cdots(t_m-t_{m-1})^{z_m}}{t^n}$$

if $z_1,\ldots,z_m$ are nonnegative integers whose sum is n, and $p_{Z_1,\ldots,Z_m}(z_1,\ldots,z_m) = 0$ otherwise.
c) The joint distributions in parts (a) and (b) are identical. Thus, for a Poisson process, given that exactly n events occur by time t, their times of occurrence, considered as unordered random variables, are the same as those obtained by selecting n times uniformly and independently from the interval $(0, t]$.

12.111 Conditions (1), (2), and (3) are part of the definition of a Poisson process. Applying L'Hôpital's rule, we find that

$$\lim_{h\to 0}\frac{e^{-\lambda h}-1+\lambda h}{h} = \lim_{h\to 0}\left(-\lambda e^{-\lambda h}+\lambda\right) = 0.$$

Hence, $e^{-\lambda h} - 1 + \lambda h = o(h)$ or, equivalently, $e^{-\lambda h} = 1 - \lambda h + o(h)$. Therefore,

$$P\big(N(h)=1\big) = e^{-\lambda h}\,\frac{(\lambda h)^1}{1!} = \lambda h - (\lambda h)^2 + \lambda h\cdot o(h).$$

Now,

$$\lim_{h\to 0}\frac{-(\lambda h)^2+\lambda h\cdot o(h)}{h} = \lim_{h\to 0}\Big(-\lambda^2 h + \lambda\cdot o(h)\Big) = \lim_{h\to 0}\left(-\lambda^2 h + \lambda h\cdot\frac{o(h)}{h}\right) = 0,$$

which means that $-(\lambda h)^2 + \lambda h\cdot o(h) = o(h)$. Hence, we see that $P\big(N(h)=1\big) = \lambda h + o(h)$, so that condition (4) holds. To verify condition (5), we first note that $P\big(N(h)=0\big) = e^{-\lambda h} = 1 - \lambda h + o(h)$. Hence, by the complementation rule,

$$\begin{aligned}P\big(N(h)\ge 2\big) &= 1 - P\big(N(h)=0\big) - P\big(N(h)=1\big)\\ &= 1 - \big(1-\lambda h+o(h)\big) - \big(\lambda h + o(h)\big) = o(h).\end{aligned}$$

12.112
a) Conditions (1) and (2) are the same as conditions (a) and (b) of the definition of a Poisson process (Definition 12.1 on page 688). Suppose that we have shown that

$$N(t) \sim \mathcal{P}(\lambda t), \qquad t > 0. \tag{$*$}$$

By condition (3), we know that $\{N(t) : t \ge 0\}$ has stationary increments. Thus, $N(t) - N(s)$ has the same probability distribution as $N(t-s) - N(0)$. But, by condition (1) and $(*)$,

$$N(t-s) - N(0) = N(t-s) \sim \mathcal{P}\big(\lambda(t-s)\big).$$

Hence, $N(t) - N(s) \sim \mathcal{P}\big(\lambda(t-s)\big)$, which is condition (c) of the definition of a Poisson process. Consequently, conditions (1)–(5) and $(*)$ imply that conditions (a)–(c) of Definition 12.1 hold; that is, that $\{N(t) : t \ge 0\}$ is a Poisson process with rate λ.
b) Applying the law of total probability and then using conditions (1), (2), and (3), we get

$$\begin{aligned}P_n(t+h) = P\big(N(t+h)=n\big) &= \sum_{k=0}^{\infty} P\big(N(t+h)=n \mid N(t)=k\big)P\big(N(t)=k\big)\\ &= \sum_{k=0}^{n} P\big(N(t+h)-N(t)=n-k \mid N(t)=k\big)P\big(N(t)=k\big)\\ &= \sum_{k=0}^{n-2} P\big(N(h)=n-k\big)P_k(t) + P\big(N(h)=1\big)P_{n-1}(t) + P\big(N(h)=0\big)P_n(t).\end{aligned}$$

Now, from conditions (4) and (5) and the complementation rule,

$$\begin{aligned}P\big(N(h)=0\big)&=1-P\big(N(h)=1\big)-P\big(N(h)\geq 2\big)\\&=1-\big(\lambda h+o(h)\big)-o(h)=1-\lambda h+o(h).\end{aligned}$$

Moreover,

$$0\leq\sum_{k=0}^{n-2}P\big(N(h)=n-k\big)P_k(t)\leq\sum_{k=0}^{n-2}P\big(N(h)=n-k\big)=\sum_{j=2}^{n}P\big(N(h)=j\big)\leq P\big(N(h)\geq 2\big).$$

Hence, by condition (5),

$$\sum_{k=0}^{n-2}P\big(N(h)=n-k\big)P_k(t)=o(h).$$

Therefore, we have shown that

$$\begin{aligned}P_n(t+h)&=o(h)+\big(\lambda h+o(h)\big)P_{n-1}(t)+\big(1-\lambda h+o(h)\big)P_n(t)\\&=\lambda hP_{n-1}(t)+(1-\lambda h)P_n(t)+o(h).\end{aligned}$$

c) We can write the result of part (b) as

$$P_n(t+h)-P_n(t)=-\lambda hP_n(t)+\lambda hP_{n-1}(t)+o(h).$$

Therefore,

$$P_n'(t)=\lim_{h\to 0}\frac{P_n(t+h)-P_n(t)}{h}=\lim_{h\to 0}\left(-\lambda P_n(t)+\lambda P_{n-1}(t)+\frac{o(h)}{h}\right)=-\lambda P_n(t)+\lambda P_{n-1}(t).$$

d) We want to prove that

$$P_n(t)=e^{-\lambda t}\frac{(\lambda t)^n}{n!},\qquad n\geq 0. \tag{$**$}$$

We proceed by mathematical induction. First we establish Equation $(**)$ for $n=0$. From the general multiplication rule and conditions (2) and (3),

$$\begin{aligned}P_0(t+h)&=P\big(N(t+h)=0\big)=P\big(N(t+h)=0,\ N(t)=0\big)\\&=P\big(N(t+h)=0\,|\,N(t)=0\big)P_0(t)=P\big(N(t+h)-N(t)=0\,|\,N(t)=0\big)P_0(t)\\&=P\big(N(h)=0\big)P_0(t)=\big(1-\lambda h+o(h)\big)P_0(t)=(1-\lambda h)P_0(t)+o(h).\end{aligned}$$

Thus,

$$P_0'(t)=\lim_{h\to 0}\frac{P_0(t+h)-P_0(t)}{h}=\lim_{h\to 0}\left(\frac{-\lambda hP_0(t)+o(h)}{h}\right)=-\lambda P_0(t).$$

From calculus, the solution to the differential equation $P_0'(t)=-\lambda P_0(t)$ is $P_0(t)=ce^{-\lambda t}$, where c is a constant. From condition (1), $P_0(0)=P\big(N(0)=0\big)=1$ and, hence, $c=1$. We have therefore established Equation $(**)$ for $n=0$. Assume now that Equation $(**)$ holds for $n-1$. We use part (c) to prove that it holds for n. By part (c) and the induction assumption,

$$P_n'(t)+\lambda P_n(t)=\lambda P_{n-1}(t)=\lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}=e^{-\lambda t}\frac{\lambda^n}{(n-1)!}t^{n-1}.$$

Multiplying both sides of the preceding result by $e^{\lambda t}$ yields

$$e^{\lambda t} P_n'(t) + \lambda e^{\lambda t} P_n(t) = \frac{\lambda^n}{(n-1)!} t^{n-1},$$

or, equivalently,

$$\frac{d}{dt}\left(e^{\lambda t} P_n(t)\right) = \frac{\lambda^n}{(n-1)!} t^{n-1}.$$

Integration now gives

$$e^{\lambda t} P_n(t) = \frac{\lambda^n}{(n-1)!}\frac{t^n}{n} + k = \frac{(\lambda t)^n}{n!} + k,$$

where k is a constant. Setting $t = 0$ and using condition (1) shows that $k = 0$. Therefore, we have

$$P_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!},$$

as required.

12.113

a) For each $n \in \mathcal{N}$, the nth interarrival interval is $[W_{n-1}, W_n]$, which has length $W_n - W_{n-1} = I_n$. As we know, I_n has the exponential distribution with parameter λ and, therefore, has mean $1/\lambda$. Hence, we might guess that $\mathcal{E}(Z)$, the expected length of the interarrival interval that contains t, equals $1/\lambda$. But, note that the interarrival interval that contains t is $[W_{N(t)}, W_{N(t)+1}]$, which has length $W_{N(t)+1} - W_{N(t)} = I_{N(t)+1}$. Thus, in fact, $Z = I_{N(t)+1}$ and, due to the random subscript in $I_{N(t)+1}$, the mean of Z may not be $1/\lambda$. In other words, our guess may be incorrect.
b) From the lack-of-memory property of the exponential distribution, it follows that the distribution of X is the same as that of I_1, which is exponential with parameter λ. Thus, $\mathcal{E}(X) = 1/\lambda$.
c) Let $0 \le y < t$. We note that $Y > y$ if and only if no events occur in the interval from $t - y$ to t. Consequently,

$$P(Y > y) = P\big(N(t) - N(t-y) = 0\big) = e^{-\lambda y}.$$

Hence, the CDF of Y is given by

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

Applying Proposition 10.5, we find that

$$\mathcal{E}(Y) = \int_0^\infty P(Y > y)\,dy = \int_0^t e^{-\lambda y}\,dy = \frac{1}{\lambda}\left(1 - e^{-\lambda t}\right).$$

d) We note that $Z = X + Y$. Hence, from parts (b) and (c),

$$\mathcal{E}(Z) = \mathcal{E}(X+Y) = \mathcal{E}(X) + \mathcal{E}(Y) = \frac{1}{\lambda} + \frac{1}{\lambda}\left(1 - e^{-\lambda t}\right) = \frac{1}{\lambda}\left(2 - e^{-\lambda t}\right).$$

This result shows that our guess in part (a) is incorrect except when $t = 0$. Moreover, for large t, we see that $\mathcal{E}(Z)$ is roughly $2/\lambda$, twice that of the guess made in part (a).
e) From part (d), the expected length of the interarrival interval that contains t is, for large t, roughly $2/\lambda$, which is twice $1/\lambda$, the common expected length of the interarrival intervals.

12.114

a) At time t, $N(t)$ events have occurred and, thus, the time of occurrence of the $N(t)$th event must be at or before t; hence, $W_{N(t)} \le t$. At time t, the $\big(N(t)+1\big)$st event has yet to occur; hence, $t < W_{N(t)+1}$. Consequently, we have $W_{N(t)} \le t < W_{N(t)+1}$.

b) We can write $W_n = I_1 + \cdots + I_n$. As $I_1, I_2, \ldots$ are independent and identically distributed $\mathcal{E}(\lambda)$ random variables, which have common mean $1/\lambda$, the strong law of large numbers implies that

$$\lim_{n\to\infty} \frac{W_n}{n} = \lim_{n\to\infty} \frac{I_1 + \cdots + I_n}{n} = \frac{1}{\lambda},$$

with probability 1.

c) We note that $N(t)$ is integer valued and nondecreasing as a function of t. Let $E = \{\lim_{t\to\infty} N(t) = \infty\}$ and, for each n, $M \in \mathcal{N}$, let $E_{Mn} = \{N(n) \geq M\}$. It follows that

$$E = \bigcap_{M=1}^{\infty}\left(\bigcup_{n=1}^{\infty} E_{Mn}\right).$$

Fix $M \in \mathcal{N}$. Then $E_{M1} \subset E_{M2} \subset \cdots$. Therefore, by the continuity property of a probability measure (Proposition 2.11 on page 74),

$$P\left(\bigcup_{n=1}^{\infty} E_{Mn}\right) = \lim_{n\to\infty} P(E_{Mn}).$$

However, because $N(t) \sim \mathcal{P}(\lambda t)$, we have by the complementation rule and FPF,

$$P(E_{Mn}) = P\big(N(n) \geq M\big) = 1 - P\big(N(n) < M\big) = 1 - \sum_{k=0}^{M-1} e^{-\lambda n}\frac{(\lambda n)^k}{k!}.$$

Therefore, by L'Hôpital's rule,

$$\lim_{n\to\infty} P(E_{Mn}) = \lim_{n\to\infty}\left(1 - \sum_{k=0}^{M-1} e^{-\lambda n}\frac{(\lambda n)^k}{k!}\right) = 1 - \sum_{k=0}^{M-1}\frac{\lambda^k}{k!}\left(\lim_{n\to\infty}\frac{n^k}{e^{\lambda n}}\right) = 1 - 0 = 1.$$

It now follows that $P\left(\bigcup_{n=1}^{\infty} E_{Mn}\right) = 1$ for each $M \in \mathcal{N}$ and, consequently, because a countable intersection of events with probability 1 has probability 1, we have that

$$P\left(\lim_{t\to\infty} N(t) = \infty\right) = P(E) = P\left(\bigcap_{M=1}^{\infty}\left(\bigcup_{n=1}^{\infty} E_{Mn}\right)\right) = 1.$$

d) Applying parts (b) and (c), we see that, with probability 1,

$$\lim_{t\to\infty}\frac{W_{N(t)}}{N(t)} = \frac{1}{\lambda}$$

and that

$$\lim_{t\to\infty}\frac{W_{N(t)+1}}{N(t)} = \lim_{t\to\infty}\frac{N(t)+1}{N(t)}\frac{W_{N(t)+1}}{N(t)+1} = 1\cdot\frac{1}{\lambda} = \frac{1}{\lambda}.$$

From part (a),

$$\frac{W_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{W_{N(t)+1}}{N(t)}.$$

In view of the preceding three displays, we conclude that, with probability 1,

$$\frac{1}{\lambda} = \lim_{t\to\infty}\frac{W_{N(t)}}{N(t)} \leq \lim_{t\to\infty}\frac{t}{N(t)} \leq \lim_{t\to\infty}\frac{W_{N(t)+1}}{N(t)} = \frac{1}{\lambda}.$$

Hence, $\lim_{t\to\infty} N(t)/t = \lambda$ with probability 1.

e) The result of part (d) shows that, with probability 1, the rate of occurrence of the specified event is λ.

12.115 *Note:* In this exercise, we use the standard conventions $\lambda_{-1} = 0$ and $\mu_0 = 0$.
a) To begin, suppose that $X \sim \mathcal{E}(a)$ so that $P(X > h) = e^{-ah}$. Referring to the hint in Exercise 12.111, we see that

$$P(X > h) = 1 - ah + o(h) \qquad \text{and} \qquad P(X \le h) = ah + o(h).$$

Now suppose that $Y \sim \mathcal{E}(b)$ and is independent of X. Then

$$P(X \le h,\ Y \le h) = \big(ah + o(h)\big)\big(bh + o(h)\big) = o(h),$$

$$P(X \le h,\ Y > h) = \big(ah + o(h)\big)\big(1 - bh + o(h)\big) = ah + o(h),$$

and

$$P(X > h,\ Y > h) = \big(1 - ah + o(h)\big)\big(1 - bh + o(h)\big) = 1 - ah - bh + o(h).$$

Next, recall that, for a birth-and-death queueing system, successive interarrival times are independent exponential random variables, successive service times are independent exponential random variables, and arrival and service times are independent. Therefore, from the lack-of-memory property of an exponential random variable and the preceding three displays, we get the following results: If the current state of the system is k, then, in a time interval of length h,

- the probability of more than one transition (arrival or departure) is $o(h)$,
- the probability of one arrival and no departures is $\lambda_k h + o(h)$,
- the probability of no arrivals and one departure is $\mu_k h + o(h)$, and
- the probability of no arrivals and no departures is $1 - \lambda_k h - \mu_k h + o(h)$.

Therefore, from the law of total probability and the bulleted list, we have, for $n \ge 0$,

$$\begin{aligned}
P_n(t+h) &= P\big(X(t+h) = n\big) = \sum_{k=0}^{\infty} P\big(X(t+h) = n \mid X(t) = k\big) P\big(X(t) = k\big) \\
&= \sum_{k=0}^{n-2} P\big(X(t+h) = n \mid X(t) = k\big) P_k(t) + \big(\lambda_{n-1} h + o(h)\big) P_{n-1}(t) \\
&\qquad + \big(1 - \lambda_n h - \mu_n h + o(h)\big) P_n(t) + \big(\mu_{n+1} h + o(h)\big) P_{n+1}(t) \\
&\qquad + \sum_{k=n+2}^{\infty} P\big(X(t+h) = n \mid X(t) = k\big) P_k(t) \\
&= \lambda_{n-1} h P_{n-1}(t) + \big(1 - (\lambda_n + \mu_n) h\big) P_n(t) + \mu_{n+1} h P_{n+1}(t) + o(h).
\end{aligned}$$

b) We can write the result of part (a) as

$$P_n(t+h) - P_n(t) = \lambda_{n-1} h P_{n-1}(t) - (\lambda_n + \mu_n) h P_n(t) + \mu_{n+1} h P_{n+1}(t) + o(h).$$

Hence, for $n \ge 0$,

$$\begin{aligned}
P_n'(t) &= \lim_{h \to 0} \left(\frac{P_n(t+h) - P_n(t)}{h} \right) \\
&= \lim_{h \to 0} \left(\lambda_{n-1} P_{n-1}(t) - (\lambda_n + \mu_n) P_n(t) + \mu_{n+1} P_{n+1}(t) + \frac{o(h)}{h} \right) \\
&= \lambda_{n-1} P_{n-1}(t) - (\lambda_n + \mu_n) P_n(t) + \mu_{n+1} P_{n+1}(t).
\end{aligned}$$

c) Assuming that a steady-state distribution, $\{P_n : n \geq 0\}$, exists, we have

$$\lim_{t\to\infty} P_n(t) = P_n, \qquad n \geq 0.$$

Heuristically, then, for $n \geq 0$,

$$\lim_{t\to\infty} P_n'(t) = \lim_{t\to\infty}\left(\frac{d}{dt}P_n(t)\right) = \frac{d}{dt}\left(\lim_{t\to\infty} P_n(t)\right) = \frac{d}{dt}P_n = 0.$$

Thus, in view of part (b), we have, for $n \geq 0$,

$$\begin{aligned} 0 &= \lim_{t\to\infty} P_n'(t) = \lim_{t\to\infty}\big(\lambda_{n-1}P_{n-1}(t) - (\lambda_n + \mu_n)P_n(t) + \mu_{n+1}P_{n+1}(t)\big) \\ &= \lambda_{n-1}\lim_{t\to\infty} P_{n-1}(t) - (\lambda_n + \mu_n)\lim_{t\to\infty} P_n(t) + \mu_{n+1}\lim_{t\to\infty} P_{n+1}(t) \\ &= \lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n + \mu_{n+1}P_{n+1}, \end{aligned}$$

or, equivalently,

$$\lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} = (\lambda_n + \mu_n)P_n, \qquad n \geq 0,$$

which are the balance equations.

12.116 We begin by showing that (univariate and multivariate) marginals of nonsingular normal random variables are nonsingular normal random variables. From Proposition 12.8 on page 719, such marginals are multivariate normal. They are nonsingular because, as we know from Chapter 9, marginals of random variables with a joint PDF also have a joint PDF.

Now we proceed to the problem at hand. Because $\mathbf{B}$ has rank $k \leq m$, linear algebra tells us that we can add $m - k$ rows to the bottom of $\mathbf{B}$ to form an $m \times m$ nonsingular matrix $\mathbf{A}$. Let us also add any $m - k$ numbers to the bottom of $\mathbf{a}$ to form an $m \times 1$ vector $\mathbf{b}$. From Exercise 12.96, we know that $\mathbf{W} = \mathbf{b} + \mathbf{AX}$ has a nonsingular multivariate normal distribution. Hence, the random vector consisting of the first k rows of $\mathbf{W}$, which is $\mathbf{Y} = \mathbf{a} + \mathbf{BX}$, has a nonsingular multivariate normal distribution.

12.117 For an $M/M/\infty$ queueing system, we have $\lambda_n = \lambda$ for all $n \geq 0$. Hence,

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \lambda \sum_{n=0}^{\infty} P_n = \lambda.$$

We know that $L = \lambda/\mu$. Hence, from Exercise 12.36(b),

$$L_q = L - \frac{\bar{\lambda}}{\mu} = \frac{\lambda}{\mu} - \frac{\lambda}{\mu} = 0.$$

The expected number of customers in the queue is 0, which makes sense because, for an $M/M/\infty$ queue, there is never any queue due to the infinite number of servers. Now applying Little's formulas, we get

$$W = \frac{L}{\bar{\lambda}} = \frac{\lambda/\mu}{\lambda} = \frac{1}{\mu} \qquad \text{and} \qquad W_q = \frac{L_q}{\bar{\lambda}} = \frac{0}{\lambda} = 0.$$

The expected time that a customer spends in the queueing system is $1/\mu$, which makes sense because, for an $M/M/\infty$ queue, the time that a customer spends in the queueing system is just the duration of service. Also, the expected time a customer waits in the queue is 0, which makes sense because, for an $M/M/\infty$ queue, there is never any queue due to the infinite number of servers.

12.118

a) Recall that $\bar{\lambda}$ represents the long-run-average arrival rate. On average, there are L down machines and, hence, $M - L$ up machines. As each up machine goes down at an exponential rate λ, independently of the other up machines, we conclude from Proposition 9.11 on page 532 that, on average, the arrival rate is $(M - L)\lambda$.

b) Referring to Exercises 12.92(a) and 12.92(b), we get that

$$\begin{aligned}\bar{\lambda} &= \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{M} (M-n)\lambda P_n = M\lambda \sum_{n=0}^{M} P_n - \lambda \sum_{n=0}^{M} n P_n \\ &= M\lambda \sum_{n=0}^{\infty} P_n - \lambda \sum_{n=0}^{\infty} n P_n = M\lambda \cdot 1 - \lambda L = (M-L)\lambda.\end{aligned}$$

c) From part (b), we have $\bar{\lambda} = (10 - 6.02)/6 = 0.663$. Hence, from Exercise 12.36(b),

$$L_q = L - \frac{\bar{\lambda}}{\mu} = 6.02 - \frac{0.663}{2/3} = 5.03.$$

Hence, from Little's formulas,

$$W = \frac{L}{\bar{\lambda}} = \frac{6.02}{0.663} = 9.08 \text{ hr} \qquad \text{and} \qquad W_q = \frac{L_q}{\bar{\lambda}} = \frac{5.03}{0.663} = 7.59 \text{ hr}.$$

12.119

a) The number of customers in service, S, equals the number of customers in the system, X, less the number of customers in the queue, Q. Hence, from Exercise 12.36(b),

$$\mathcal{E}(S) = \mathcal{E}(X - Q) = \mathcal{E}(X) - \mathcal{E}(Q) = L - L_q = \frac{\bar{\lambda}}{\mu}.$$

b) Let B_j denote the event that server j is busy. By symmetry, $P(B_j)$ does not depend on j, and we denote that probability p_b. Note that p_b equals the probability that a specified server is busy. Referring now to part (a) and observing that $S = \sum_{j=1}^{s} I_{B_j}$, we conclude that

$$\frac{\bar{\lambda}}{\mu} = \mathcal{E}(S) = \mathcal{E}\left(\sum_{j=1}^{s} I_{B_j}\right) = \sum_{j=1}^{s} \mathcal{E}\big(I_{B_j}\big) = \sum_{j=1}^{s} P(B_j) = sp_b.$$

Consequently, $p_b = \bar{\lambda}/s\mu$.

c) For a single-server queue, the server is busy if and only if $X \geq 1$. Hence, from part (b),

$$\frac{\bar{\lambda}}{\mu} = p_b = P(X \geq 1) = 1 - P(X = 0) = 1 - P_0.$$

Hence, $P_0 = 1 - \bar{\lambda}/\mu$.

d) From Exercise 12.35(b), $\bar{\lambda} = \lambda$ for an $M/M/1$ queue. Hence, from part (c), $P_0 = 1 - \lambda/\mu$.