Introduction to Probability

Detailed Solutions to Exercises

David F. Anderson

Timo Seppäläinen

Benedek Valkó

Contents

Preface	1
Solutions to Chapter 1	3
Solutions to Chapter 2	29
Solutions to Chapter 3	59
Solutions to Chapter 4	91
Solutions to Chapter 5	113
Solutions to Chapter 6	129
Solutions to Chapter 7	155
Solutions to Chapter 8	167
Solutions to Chapter 9	197
Solutions to Chapter 10	205
Solutions to the Appendix	235

Preface

This collection of solutions is for reference for the instructors who use our book. The authors firmly believe that the best way to master new material is via problem solving. Having all the detailed solutions readily available would undermine this process. Hence, we ask that instructors not distribute this document to the students in their courses.

The authors welcome comments and corrections to the solutions. A list of corrections and clarifications to the textbook is updated regularly at the website `https://www.math.wisc.edu/asv/`

Solutions to Chapter 1

1.1. One sample space is

$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\} = \{(i, j) : i, j \in \{1, \dots, 6\}\},$$

where we view order as mattering. Note that $\#\Omega = 6^2 = 36$. Since all outcomes are equally likely, we take $P(\omega) = \frac{1}{36}$ for each $\omega \in \Omega$. The event A is

$$A = \left\{ \begin{array}{ccccc} (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ & (2,3), & (2,4), & (2,5), & (2,6) \\ & & (3,4), & (3,5), & (3,6) \\ & & & (4,5), & (4,6) \\ & & & & (5,6) \end{array} \right\} = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}, i < j\},$$

and

$$P(A) = \frac{\#A}{\#\Omega} = \frac{15}{36}.$$

One way to count the number of elements in A without explicitly writing them out is to note that for a first roll of $i \in \{1, 2, 3, 4, 5\}$, there are only $6 - i$ allowable rolls for the second. Hence,

$$\#A = \sum_{i=1}^{5}(6 - i) = 5 + 4 + 3 + 2 + 1 = 15.$$

1.2. (a) Since Bob has to choose exactly two options, Ω consists of the 2-element subsets of the set $\{\texttt{cereal}, \texttt{eggs}, \texttt{fruit}\}$:

$$\Omega = \{\{\texttt{cereal}, \texttt{eggs}\}, \{\texttt{cereal}, \texttt{fruit}\}, \{\texttt{eggs}, \texttt{fruit}\}\}$$

The items in Bob's breakfast do not come in any particular order, hence the outcomes are sets instead of ordered pairs.

(b) The two outcomes in the event A are $\{\texttt{cereal}, \texttt{eggs}\}$ and $\{\texttt{cereal}, \texttt{fruit}\}$. In symbols,

$$A = \{\text{Bob's breakfast includes cereal}\} = \{\{\texttt{cereal}, \texttt{eggs}\}, \{\texttt{cereal}, \texttt{fruit}\}\}.$$

1.3. (a) This is a Cartesian product where the first factor covers the outcome of the coin flip ($\{H, T\}$ or $\{0, 1\}$, depending on how you want to encode heads and tails) and the second factor represents the outcome of the die. Hence

$$\Omega = \{0, 1\} \times \{1, 2, \dots, 6\} = \{(i, j) : i = 0 \text{ or } 1 \text{ and } j \in \{1, 2, \dots, 6\}\}.$$

(b) Now we need a larger Cartesian product space because the outcome has to contain the coin flip and die roll of each person. Let c_i be the outcome of the coin flip of person i, and let d_i be the outcome of the die roll of person i. Index i runs from 1 to 10 (one index value for each person). Each $c_i \in \{0, 1\}$ and each $d_i \in \{1, 2, \dots, 6\}$. Here are various ways of writing down the sample space:

$$\begin{aligned}\Omega &= (\{0, 1\} \times \{1, 2, \dots, 6\})^{10} \\ &= \{(c_1, d_1, c_2, d_2, \dots, c_{10}, d_{10}) : \text{each } c_i \in \{0, 1\} \text{ and each } d_i \in \{1, 2, \dots, 6\}\} \\ &= \{(c_i, d_i)_{1 \le i \le 10} : \text{each } c_i \in \{0, 1\} \text{ and each } d_i \in \{1, 2, \dots, 6\}\}.\end{aligned}$$

The last formula illustrates the use of indexing to shorten the writing of the 20-tuple of all outcomes. The number of elements is $\#\Omega = 2^{10} \cdot 6^{10} = 12^{10} =$ 61,917,364,224.

(c) If nobody rolled a five, then each die outcome d_i comes from the set $\{1, 2, 3, 4, 6\}$ that has 5 elements. Hence the number of these outcomes is $2^{10} \cdot 5^{10} = 10^{10}$. To get the number of outcomes where at least 1 person rolls a five, subtract the number of outcomes where no one rolls a 5 from the total: $12^{10} - 10^{10} =$ 51,917,364,224.

1.4. (a) This is an example of sampling with replacement, where order matters. Thus, the sample space is

$$\Omega = \{\omega = (x_1, x_2, x_3) : x_i \in \{\text{states in the U.S.}\}\}.$$

In other words, each sample point is a 3-tuple or ordered triple of U.S. states.

The problem statement contains the assumption that every day each state is equally likely to be chosen. Since $\#\Omega = 50^3$ =125,000, each sample point ω has equal probability $P\{\omega\} = 50^{-3} = \frac{1}{125{,}000}$. This specifies the probability measure completely because then the probability of any event A comes from the formula $P(A) = \frac{\#A}{125{,}000}$.

(b) The 3-tuple (Wisconsin, Minnesota, Florida) is a particular outcome, and hence as explained above,

$$P((\text{Wisconsin, Minnesota, Florida})) = \tfrac{1}{50^3}.$$

(c) The number of ways to have Wisconsin come on Monday and Tuesday, but not Wednesday is $1 \cdot 1 \cdot 49$, with similar expressions for the other combinations. Since there is only 1 way for Wisconsin to come each of the three days, we see the total number of positive outcomes is

$$1 \cdot 1 \cdot 49 + 1 \cdot 49 \cdot 1 + 49 \cdot 1 \cdot 1 + 1 = 3 \cdot 49 + 1 = 148.$$

Thus

$$\begin{aligned}&P(\text{Wisconsin's flag hung at least two of the three days}) \\ &\qquad = \frac{3 \cdot 49 + 1}{50^3} = \frac{37}{31250} = 0.001184.\end{aligned}$$

1.5. (a) There are two natural sample spaces we can choose, depending upon whether or not we want to let order matter.

If we let the order of the numbers matter, then we may choose

$$\Omega_1 = \{(x_1, \dots, x_5) : x_i \in \{1, \dots, 40\}, x_i \neq x_j \text{ if } i \neq j\},$$

the set of ordered 5-tuples of distinct elements from the set $\{1, 2, 3, \dots, 40\}$. In this case $\#\Omega_1 = 40 \cdot 39 \cdot 38 \cdot 37 \cdot 36$ and $P_1(\omega) = \frac{1}{\#\Omega_1}$ for each $\omega \in \Omega_1$.

If we do not let order matter, then we take

$$\Omega_2 = \{\{x_1, \dots, x_5\} : x_i \in \{1, 2, 3, \dots, 40\}, x_i \neq x_j \text{ if } i \neq j\},$$

the set of 5-element subsets of the set $\{1, 2, 3, \dots, 40\}$. In this case $\#\Omega_2 = \binom{40}{5}$ and $P_2(\omega) = \frac{1}{\#\Omega_2}$ for each $\omega \in \Omega_2$.

(b) The correct calculation for this question depends on which sample space was chosen in part (a).

When order matters, we imagine filling the positions of the 5-tuple with three even and two odd numbers. There are $\binom{5}{3}$ ways to choose the positions of the three even numbers. The remaining two positions are for the two odd numbers. We fill these positions in order, separately for the even and odd numbers. There are $20 \cdot 19 \cdot 18$ ways to choose the even numbers and $20 \cdot 19$ ways to choose the odd numbers. This gives

$$P(\text{exactly three numbers are even}) = \frac{\binom{5}{3} \cdot 20 \cdot 19 \cdot 18 \cdot 20 \cdot 19}{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36} = \frac{475}{1443}.$$

When order does not matter, we choose sets. There are $\binom{20}{3}$ ways to choose a set of three even numbers between 1 and 40, and $\binom{20}{2}$ ways to choose a set of two odd numbers. Therefore, the probability can be computed as

$$P(\text{exactly three numbers are even}) = \frac{\binom{20}{3} \cdot \binom{20}{2}}{\binom{40}{5}} = \frac{475}{1443}.$$

1.6. We give two solutions, first with an ordered sample, and then without order.

(a) Label the three green balls 1, 2, and 3, and label the yellow balls 4, 5, 6, and 7. We imagine picking the balls in order, and hence take

$$\Omega = \{(i, j) : i, j \in \{1, 2, \dots, 7\}, i \neq j\},$$

the set of ordered pairs of distinct elements from the set $\{1, 2, \dots, 7\}$. The event of two different colored balls is,

$$A = \{(i, j) : (i \in \{1, 2, 3\} \text{ and } j \in \{4, \dots, 7\}) \textbf{ or } (i \in \{4, \dots, 7\} \text{ and } j \in \{1, 2, 3\})\}.$$

(b) We have $\#\Omega = 7 \cdot 6 = 42$ and $\#A = 3 \cdot 4 + 4 \cdot 3 = 24$. Thus,

$$P(A) = \frac{24}{42} = \frac{4}{7}.$$

Alternatively, we could have chosen a sample space in which order does not matter. In this case the size of the sample space is $\binom{7}{2}$. There are $\binom{3}{1}$ ways to choose one of the green balls and $\binom{4}{1}$ ways to choose one yellow ball. Hence, the probability is computed as

$$P(A) = \frac{\binom{3}{1}\binom{4}{1}}{\binom{7}{2}} = \frac{4}{7}.$$

1.7. (a) Label the balls 1 through 7, with the green balls labeled 1, 2 and 3, and the yellow balls labeled 4, 5, 6 and 7. Let

$$\Omega = \{(i, j, k) : i, j, k \in \{1, 2, \dots, 7\}, i \neq j, j \neq k, i \neq k\},$$

which captures the idea that order matters for this problem. Note that $\#\Omega = 7 \cdot 6 \cdot 5$. There are exactly

$$3 \cdot 4 \cdot 2 = 24$$

ways to choose first a green ball, then a yellow ball, and then a green ball. Thus the desired probability is

$$P(\text{green, yellow, green}) = \frac{24}{7 \cdot 6 \cdot 5} = \frac{4}{35}.$$

(b) We can use the same reasoning as in the previous part, by accounting for all the different orders in which the colors can come:

$$\begin{aligned} P(2 \text{ greens and one yellow}) &= P(\text{green, green, yellow}) \\ &\quad P(\text{green, yellow, green}) + P(\text{yellow, green, green}) \\ &= \frac{3 \cdot 2 \cdot 4 + 3 \cdot 4 \cdot 2 + 4 \cdot 3 \cdot 2}{7 \cdot 6 \cdot 5} = \frac{72}{210} = \frac{12}{35}. \end{aligned}$$

Alternatively, since this question does not require ordering the sample of balls, we can take

$$\Omega = \{\{i, j, k\} : i, j, k \in \{1, 2, \dots, 7\}, i \neq j, j \neq k, i \neq k\},$$

the set of 3-element subsets of the set $\{1, 2, \dots, 7\}$. Now $\#\Omega = \binom{7}{3}$. There are $\binom{3}{2}$ ways to choose 2 green balls from the 3 green balls, and $\binom{4}{1}$ ways to choose one yellow ball from the 4 yellow balls. So the desired probability is

$$P(2 \text{ greens and one yellow}) = \frac{\binom{3}{2} \cdot \binom{4}{1}}{\binom{7}{3}} = \frac{12}{35}.$$

1.8. (a) Label the letters from 1 to 14 so that the first 5 are `E`s, the next 4 are `A`s, the next 3 are `N`s and the last 2 are `B`s.

Our Ω consists of (ordered) sequences of four distinct elements:

$$\Omega = \{(a_1, a_2, a_3, a_4) : a_i \neq a_j, a_i \in \{1, 2, \dots, 14\}\}.$$

The size of Ω is $14 \cdot 13 \cdot 12 \cdot 11 = 24024$. (Because we can choose a_1 14 different ways, then a_2 13 different ways and so on.)

The event C consists of sequences (a_1, a_2, a_3, a_4) consisting of two numbers between 1 and 5, one between 6 and 9 and one between 10 and 12. We can count these by constructing such a sequence step-by-step: we first choose the positions of the two `E`s: we can do that $\binom{4}{2} = 6$ ways. Then we choose a first `E` out of the 5 choices and place it to the first chosen position. Then we choose the second `E` out of the remaining 4 and place it to the second (remaining) chosen position. Then we choose the `A` out of the 4 choices, and its position (there are 2 possibilities left), Finally we choose the letter `N` out of the 3 choices and place it in the remaining position (we only have one possibility here). In each step the number of choices did not depend on the previous choices so we can just multiply the numbers together to get $6 \cdot 5 \cdot 4 \cdot 4 \cdot 2 \cdot 3 \cdot 1 = 2880$.

The probability of C is

$$P(C) = \frac{\#C}{\#\Omega} = \frac{2880}{24024} = \frac{120}{1001}.$$

(b) As before, we label the letters from 1 to 14 so that the first 5 are `E`s, the next 4 are `A`s, the next 3 are `N`s and the last 2 are `B`s. Our Ω is the set of unordered samples of size 4, or in other words: all subsets of $\{1, 2, \ldots, 14\}$ of size 4:

$$\Omega = \{\{a_1, a_2, a_3, a_4\} : a_i \neq a_j, a_i \in \{1, 2, \ldots, 14\}\}.$$

The size of Ω is $\binom{14}{4} = 1001$.

The event C is that $\{a_1, a_2, a_3, a_4\}$ has two numbers between 1 and 5, one between 6 and 9 and one between 10 and 12. The number of ways we can choose such a set is $\binom{5}{2}\binom{4}{1}\binom{3}{1} = 120$. (Because we can choose the two `E`s out of 5 possibilities, the single `A` out of 4 possibilities and the single `N` out of 3 possibilities.)

This gives

$$P(C) = \frac{\#C}{\#\Omega} = \frac{120}{1001},$$

the same as in part (a).

1.9. We model the point at which the stick is broken as being chosen uniformly at random along the length of the stick, which we take to be L (in some arbitrary units). Thus, $\Omega = [0, L]$. The event we care about is $A = \{\omega \in \Omega : \omega \leq L/5 \text{ or } \omega \geq 4L/5\}$. Hence, since the two events are mutually exclusive,

$$P(A) = P\{\omega \in [0, L] : \omega \leq L/5\} + P\{\omega \in [0, L] : \omega \geq 4L/5\} = \frac{L/5}{L} + \frac{L/5}{L} = \frac{2}{5}.$$

1.10. (a) Since the outcome of the experiment is the number of times we roll the die (as in Example 1.16), we take

$$\Omega = \{\infty, 1, 2, 3, \ldots\}.$$

Element k in Ω means that it took k rolls to see the first four. Element ∞ means that four never appeared.

Next we deduce the probability measure P on Ω. Since Ω is a discrete sample space (countably infinite), P is determined by giving the probabilities of all the individual sample points.

For an integer $k \geq 1$, we have

$$P(k) = P\{\text{needed } k \text{ rolls}\} = P\{\text{no fours in the first } k-1 \text{ rolls, then a } 4\}.$$

Each roll has 6 outcomes so the total number of outcomes from k rolls is 6^k. Each roll can fail to be a four in 5 ways. Hence by taking the ratio of the number of favorable outcomes over the total number of outcomes,

$$P(k) = P\{\text{no fours in the first } k-1 \text{ rolls, then a } 4\} = \frac{5^{k-1} \cdot 1}{6^k} = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

To complete the specification of the measure P, we find the value $P(\infty)$. Since the outcomes are mutually exclusive,

$$
\begin{aligned}
1 = P(\Omega) &= P(\infty) + \sum_{k=1}^{\infty} P(k) \\
&= P(\infty) + \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} \\
\text{(reindex)} \qquad &= P(\infty) + \frac{1}{6} \sum_{j=0}^{\infty} \left(\frac{5}{6}\right)^{j} \\
\text{(geometric series)} \qquad &= P(\infty) + \frac{1}{6} \cdot \frac{1}{1 - 5/6} \\
&= P(\infty) + 1.
\end{aligned}
$$

Thus, $P(\infty) = 0$.

(b) We already deduced above that

$$P(\text{the number four never appears}) = P(\infty) = 0.$$

Here is an alternative solution.

$$P(\text{the number four never appears}) \le P(\text{no fours in the first } n \text{ rolls}) = \left(\frac{5}{6}\right)^n.$$

Since $(\frac{5}{6})^n \to 0$ as $n \to \infty$ and the inequality holds for any n, the probability on the left must be zero.

1.11. The sample space Ω that represents the dartboard itself is a square of side length 20 inches. We can assume that the center of the board is at the origin. The event A, that the dart hits within 2 inches of the center, is then the subset of Ω described by $A = \{x : |x| \le 2\}$. Probability is now proportional to area, and so

$$P(A) = \frac{\text{area of } A}{\text{area of the board}} = \frac{\pi \cdot 2^2}{20^2} = \frac{\pi}{100}.$$

1.12. The sample space and probability measure for this experiment were described in the solution to Exercise 1.10: $P(k) = (\frac{5}{6})^{k-1}\frac{1}{6}$ for positive integers k.

(a) $P(\text{need at most 3 rolls}) = P(1) + P(2) + P(3) = \frac{1}{6}\big(1 + \frac{5}{6} + (\frac{5}{6})^2\big) = \frac{91}{216}$.

(b)

$$
\begin{aligned}
P(\text{even number of rolls}) &= \sum_{m=1}^{\infty} P(2m) = \sum_{m=1}^{\infty} (\tfrac{5}{6})^{2m-1}\tfrac{1}{6} = \tfrac{1}{5} \sum_{m=1}^{\infty} (\tfrac{25}{36})^m \\
&= \tfrac{1}{5} \cdot \frac{\frac{25}{36}}{1 - \frac{25}{36}} = \tfrac{5}{11}.
\end{aligned}
$$

1.13. (a) Imagine selecting one student uniformly at random from the school. Thus, Ω is the set of students and each outcome is equally likely. Let W be the subset of Ω consisting of those students who wear a watch. Let B be the subset of students who wear a bracelet. We are told that

$$P(W^c B^c) = 0.6, \quad P(W) = 0.25, \quad P(B) = 0.30.$$

We are asked for $P(W \cup B)$. By de Morgan (or a Venn Diagram) we have

$$P(W \cup B) = 1 - P((W \cup B)^c) = 1 - P(W^c B^c) = 1 - 0.6 = 0.4.$$

(b) We want $P(W \cap B)$. We have

$$P(W \cap B) = P(W) + P(B) - P(W \cup B) = 0.25 + 0.30 - 0.4 = 0.15.$$

1.14. From the inclusion-exclusion principle we get

$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.4 + 0.7 - P(AB) = 1.1 - P(AB).$$

Rearranging this we get $P(AB) = 1.1 - P(A \cup B)$.

Since $P(A \cup B)$ is a probability, it is at most 1, so

$$P(AB) = 1.1 - P(A \cup B) \geq 1.1 - 1 = 0.1.$$

On the other hand, $B \subset A \cup B$ so $P(A \cup B) \geq P(B) = 0.7$ which gives

$$P(AB) = 1.1 - P(A \cup B) \leq 1.1 - 0.7 = 0.4.$$

Putting these together we get $0.1 \leq P(AB) \leq 0.4$.

1.15. (a) The event that one of the colors does not appear is $W \cup G \cup R$. If we use the inclusion-exclusion principle then

$$P(W \cup G \cup R) = P(W) + P(G) + P(R) - P(WG) - P(GR) - P(RW) + P(WGR).$$

We compute each term on the right-hand side. Note that the we can label the 4 balls so that we can differentiate between the 2 red balls. This way the three draws lead to equally likely outcomes, each with probability $\frac{1}{4^3}$.

We have

$$P(W) = P(\text{each pick is green or red}) = \frac{3^3}{4^3}$$

and similarly $P(G) = \frac{3^3}{4^3}$ and $P(R) = \frac{2^3}{4^3}$. Also:

$$P(WG) = P(\text{each pick is red}) = \frac{2^3}{4^3}$$

and similarly $P(GR) = \frac{1}{4^3}$ and $P(RW) = \frac{1}{4^3}$. Finally, $P(WGR) = 0$, since it is not possible to have none of the colors in the sample.

Putting everything together:

$$P(W \cup G \cup R) = \frac{1}{4^3}(3^3 + 3^3 + 2^3 - 2^3 - 1 - 1) = \frac{13}{16}.$$

(b) The complement of the event is $\{$all three colors appear$\}$. Let us count how many different ways we can get such an outcome. We have 2 choices to decide which red ball will show up, while there is only one possibility for the green and the white. Then there are $3! = 6$ different ways we can order the three colors. This gives $2 \cdot 6 = 12$ possibilities. Thus

$$P(\text{all three colors appear}) = \frac{12}{4^3} = \frac{3}{16}$$

from which

$$P(\text{one of the colors does not appear}) = 1 - P(\text{all three colors appear}) = \frac{13}{16}.$$

1.16. If we see only heads, I win \$5. If we see 4 heads, I win \$3. If we see 3 heads, I win \$1. If we see 2 heads, I "win" -\$1. If we see 1 heads, I "win" -\$3. Finally, if we see 0 heads, then I "win" -\$5. Thus, the possible values of X are $\{-5, -3, -1, 1, 3, 5\}$. The sample space for the 5 coin flips is $\Omega = \{(x_1, \dots, x_5) : x_i \in \{H, T\}\}$ with $\#\Omega = 2^5$. Each individual outcome $(x_1, \dots, x_5)$ of five flips has probability 2^{-5}.

Let $k \in \{0, 1, \dots, 5\}$. To calculate the probability of exactly k heads we need to count how many five-flip outcomes yield exactly k heads. The answer is $\binom{5}{k}$, the number of ways of specifying which of the five flips are heads. Hence

$$
\begin{aligned}
&P(\text{precisely } k \text{ heads}) \\
&\quad = \frac{\#\text{ ways to select } k \text{ slots from the 5 for the } k \text{ heads}}{2^5} = \binom{5}{k} 2^{-5}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
P(X = -5) &= P(0 \text{ heads}) = 2^{-5} \\
P(X = -3) &= P(1 \text{ heads}) = 5 \cdot 2^{-5} \\
P(X = -1) &= P(2 \text{ heads}) = \binom{5}{2} \cdot 2^{-5} \\
P(X = 1) &= P(3 \text{ heads}) = \binom{5}{3} \cdot 2^{-5} \\
P(X = 3) &= P(4 \text{ heads}) = \binom{5}{4} \cdot 2^{-5} \\
P(X = 5) &= P(5 \text{ heads}) = \binom{5}{5} 2^{-5}.
\end{aligned}
$$

1.17. (a) Possible values of Z are $\{0, 1, 2\}$.

$$
\begin{aligned}
p_Z(0) &= P(Z = 0) = \frac{\binom{4}{2}}{\binom{7}{2}} = \tfrac{2}{7}, \\
p_Z(1) &= P(Z = 1) = \frac{\binom{4}{1}\binom{3}{1}}{\binom{7}{2}} = \tfrac{4}{7}, \\
p_Z(2) &= P(Z = 2) = \frac{\binom{3}{2}}{\binom{7}{2}} = \tfrac{1}{7}.
\end{aligned}
$$

(b) Possible values of W are $\{0, 1, 2\}$.

$$
\begin{aligned}
p_W(0) &= P(W = 0) = \frac{4 \cdot 4}{7 \cdot 7} = \frac{16}{49}, \\
p_W(1) &= P(W = 1) = \frac{4 \cdot 3 + 3 \cdot 4}{7 \cdot 7} = \frac{24}{49}, \\
p_W(2) &= P(W = 2) = \frac{3 \cdot 4}{7 \cdot 7} = \frac{9}{49}.
\end{aligned}
$$

1.18. The possible values of X are $\{3, 4, 5\}$ as these are the possible lengths of the words. The probability mass function is

$$P(X = 3) = P(\text{we chose one of the letters of ARE}) = \frac{3}{16}$$
$$P(X = 4) = P(\text{we chose one of the letters of SOME or DOGS}) = \frac{8}{16} = \frac{1}{2}$$
$$P(X = 5) = P(\text{we chose one of the letters of BROWN}) = \frac{5}{16}.$$

1.19. The possible values of X are 5 and 1. For the probability mass function we need $P(X = 1)$ and $P(X = 5)$. From the wording of the problem

$$P(X = 5) = P(\text{dart lands within 2 inches of the center}).$$

We may assume that the position of the dart is chosen uniformly from the disk of radius 6 inches, and hence we may compute the probability above as the ratio of the area of the disk of radius 2 to the area of the entire disk of radius 6:

$$P(\text{dart lands within 2 inches of the center}) = \frac{\pi 2^2}{\pi 6^2} = \frac{1}{9}.$$

Since $P(X = 5) + P(X = 1) = 1$, we get $P(X = 1) = 1 - P(X = 5) = \frac{8}{9}$.

1.20. (a) One appropriate sample space is

$$\Omega = \{1, \dots, 6\}^4 = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, \dots, 6\}\}.$$

Note that $\#\Omega = 6^4 = 1296$. Since it is reasonable to assume that all outcomes are equally likely, we set

$$P(\omega) = \frac{1}{\#\Omega} = \frac{1}{1296}.$$

(b) To find $P(A)$ and $P(B)$ we count to find $\#A$ and $\#B$, that is, the number of outcomes in these events.

Begin with the easy observation: there is only one way for there to be four fives, namely $(5, 5, 5, 5)$. There are 5 ways to get three fives in the pattern $(5, 5, 5, X)$, one for each $X \in \{1, 2, 3, 4, 6\}$. Similarly, there are 5 ways to have three fives in each of the patterns $(5, 5, X, 5), (5, X, 5, 5)$ and $(X, 5, 5, 5)$. Thus, there are a total of $5+5+5+5 = 20$ ways to have three fives. A slicker way to calculate this would be to note that there are $\binom{4}{1} = 4$ ways to choose which roll is not five, and for each not-five we have 5 choices, thus altoghether $4 \cdot 5 = 20$.

Continuing this logic, we see that the number of ways to have precisely two fives is:

$$(\#\text{ways to choose the not-five rolls}) \cdot 5 \cdot 5 = \binom{4}{2} \cdot 5 \cdot 5 = 150.$$

Thus,

$$P(A) = \frac{\#A}{\#\Omega} = \frac{1 + 20 + 150}{1296} = \frac{171}{1296} = \frac{19}{144}.$$

Similarly,

$$P(B) = \frac{\#B}{\#\Omega} = \frac{\binom{4}{4} \cdot 5^4 + \binom{4}{3} \cdot 5^3}{1296} = \frac{1125}{1296} = \frac{125}{144}.$$

(c) $A \cup B = \Omega$. Since A and B are disjoint we should have $1 = P(\Omega) = P(A \cup B) = P(A) + P(B)$, which agrees with the above.

1.21. (a) Number the black chips 1, 2, 3, the red chips 4 and 5, and the green chips 6 and 7. Then, let the sample space be $\Omega = \{(x_1, x_2, x_3) : x_i \in \{1, \dots, 7\}, x_i \neq x_j \text{ for } i \neq j\}$, where the entry x_i represents our ith draw. Note that elements of this Ω are equally likely and that there are precisely $7 \cdot 6 \cdot 5 = 210$ such elements.

To compute $P(A)$ we count the number of ways we can get three different colored chips for our three choices. We can choose a black chip, a red chip and a green chip in $3 \cdot 2 \cdot 2 = 12$ different ways. For each such choice we can order the three chips $3! = 6$ ways. Thus $\#A = 12 \cdot 6 = 72$ and $P(A) = \frac{\#A}{\#\Omega} = \frac{72}{210} = \frac{12}{35}$.

(b) Use the same labels for the chips as in part (a). Our sample space is

$$\Omega = \{\{x_1, x_2, x_3\} : x_i \in \{1, \dots, 7\}, x_i \neq x_j \text{ for } i \neq j\}.$$

Note that the sample points are now subsets of size 3 instead of ordered triples, and to indicate this the notation changed from (x_1, x_2, x_3) to $\{x_1, x_2, x_3\}$. We have $\#\Omega = \binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{3!} = 35$. $\#A = 3 \cdot 2 \cdot 2 = 12$, the number of ways to choose one of three black chips, one of two red chips and one of two green chips. Thus $P(A) = \frac{\#A}{\#\Omega} = \frac{12}{35}$. The answer is the same as in part (a), as it should be.

1.22. (a) The sample space is the set of 52 cards. We can represent the cards with numbers from 1 to 52, or with their names. Since each outcome is equally likely, $P\{\omega\} = \frac{1}{52}$ for any fixed card ω. For any subset A of cards we have $P(A) = \frac{\#A}{52}$.

(b) An event is a subset of the sample space Ω. In part (a) we saw that for an event A we have $P(A) = \frac{\#A}{52}$. So the desired event must have three elements. Any such set will work, for example $\{\heartsuit 2, \heartsuit 3, \heartsuit K\}$. In words, this is the event that the chosen card is the two of hearts, the three of hearts or the king of hearts.

(c) By part (a), if $P(A) = \frac{1}{5}$ then $\frac{\#A}{52} = \frac{1}{5}$ which forces $\#A = \frac{52}{5}$. Since $\frac{52}{5}$ is not an integer, there cannot be a subset with this many elements. Consequently this probability space has no event with probability $1/5$.

1.23. (a) You win if the prize is behind door 1. Probability $\frac{1}{3}$.

(b) You win if the prize is behind door 2 or 3. Probability $\frac{2}{3}$.

1.24. Choose door 3 and commit to switch. Then probability of winning is $p_1 + p_2$.

1.25. (a) Since there are 5 restaurants with at least one friend out of 6 total restaurants, this probability is $\frac{5}{6}$.

(b) She has 7 friends in total. 3 of them are at a restaurant alone and 4 of them are at a restaurant with somebody else. Thus the probability that she calls a friend at a restaurant with 2 friends present is $\frac{4}{7}$.

1.26. This is sampling without replacement for it would make no sense to put the same person twice on the committee. We are choosing 4 out of 15. We can do this with order (there is a first pick, a second pick, etc) or without order (we choose the subset of 4). It does not matter which approach we choose. But once we have chosen a method, or calculations have to be consistent. If we work with order then we have $15 \cdot 14 \cdot 13 \cdot 12$ possible outcomes, while if we work without order then we have $\binom{15}{4}$ choices. Each computation boils down to counting the number of favorable outcomes and then dividing by the total number of outcomes.

(a) Without order: we can choose the two men $\binom{10}{2}$ ways and the two women $\binom{5}{2}$ ways. Thus te number of favorable outcomes is $\binom{10}{2} \cdot \binom{5}{2}$ and the probability is $\frac{\binom{10}{2} \cdot \binom{5}{2}}{\binom{15}{4}} = \frac{30}{91}$.
With order: we can choose the two men $10 \cdot 9$ different ways and the two women $5 \cdot 4$ different ways. We also have to choose which two positions out of the 4 belong to men, and there are $\binom{4}{2}$ choices for that. Thus the number of favorable outcomes is $10 \cdot 9 \cdot 5 \cdot 4 \cdot \binom{4}{2}$ and the probability is $\frac{10 \cdot 9 \cdot 5 \cdot 4 \cdot \binom{4}{2}}{15 \cdot 14 \cdot 13 \cdot 12} = \frac{30}{91}$.
We got the same answer, but the computation without order was quicker.

(b) Without order: we want to count the number of committees that have both Bob and Jane. We need to choose two additional members out of the remaining 13: we can do that $\binom{13}{2}$ different ways. Thus the probability that both Bob and Jane are on the committee is $\frac{\binom{13}{2}}{\binom{15}{4}} = \frac{2}{35}$.
With order: choose Bob's position among the 4 members (4 choices), then Jane's position among the remaining 3 places (3 choices), and finally choose two other members for the remaining two places ($13 \cdot 12$ choices). This gives $\frac{4 \cdot 3 \cdot 13 \cdot 12}{15 \cdot 14 \cdot 13 \cdot 12} = \frac{2}{35}$.

(c) Without order: we need to choose 3 additional members besides Bob, out of the 13 possibilities (since Jane cannot be chosen). This gives $\binom{13}{3}$ choices and the corresponding probability is $\frac{\binom{13}{3}}{\binom{15}{4}} = \frac{22}{105}$.
With order: we choose Bob's position (4 choices) and the 3 additional members ($13 \cdot 12 \cdot 11$ choices). This gives $\frac{4 \cdot 13 \cdot 12 \cdot 11}{15 \cdot 14 \cdot 13 \cdot 12} = \frac{22}{105}$.

1.27. (a) The colors do not matter for this part. So we can set up our sample space as follows: $\Omega = \{(x_1, \dots, x_7) : x_i \in \{1, \dots, 7\}, x_i \neq x_j \text{ for } i \neq j\}$. Ω is the set of all permutations of the numbers $1, 2, 3 \dots, 7$ and $\#\Omega = 7!$.

For $1 \leq i \leq 7$ we need to compute the probability of the event

$$A_i = \{\text{the } i\text{th draw is the number } 5\}.$$

For a given i we count the number of elements in A_i. We can construct all elements of A_i by first placing 5 in the ith position, and then distributing the remaining 6 numbers among the remaining 6 positions. We can do this 6! different ways: there are 6 choices for the number in the first available position, 5 choices for the next available position, and so on. Thus $\#A_i = 6!$ (the same for each i), and thus for all $1 \leq i \leq 7$ we get

$$P(A_i) = \frac{\#A_i}{\#\Omega} = \frac{6!}{7!} = \frac{1}{7}.$$

(b) Assume that the three black chips are labeled by $a_1 < a_2 < a_3$. We can use the same sample space as in part (a). We need to compute the probability of the event B_i that the ith pick is black. Again we may assume $1 \leq i \leq 7$. For a given i we can construct all elements of B_i as follows: we pick one of the black chips (a_1, a_2 or a_3) and place it in position i. (We have three choices for that.) Then we distribute the remaining 6 numbers among the remaining 6 places. (There are 6! ways we can do that.) Thus for any $1 \le i \le 7$ we get $\#B_i = 3 \cdot 6!$ and then

$$P(B_i) = \frac{\#B_i}{\#\Omega} = \frac{3 \cdot 6!}{7!} = \frac{3}{7}.$$

1.28. Assume that both m and n are at least 1 so the problem is not trivial.

(a) Sampling without replacement. We can compute the answer using either an ordered or an unordered sample. It helps to assume that the balls are labeled (e.g. by numbering them from 1 to $m+n$), although the actual labeling will not play a role in the computation.

With an ordered sample we have $(m+n)(m+n-1)$ outcomes (we have $m+n$ choices for the first pick and $m+n-1$ choices for the second). The favorable outcomes can be counted by considering green-green and yellow-yellow pairs separately: their number is $m(m-1)+n(n-1)$. The answer is the ratio of the number of favorable outcomes to the total number of outcomes,

$$P\{(\text{g,g}) \text{ or } (\text{y,y})\} = \frac{m(m-1)+n(n-1)}{(m+n)(m+n-1)}.$$

The unordered sample calculation gives the same answer:

$$P\{\text{a set of two greens or a set of two yellows}\} = \frac{\binom{m}{2}+\binom{n}{2}}{\binom{m+n}{2}} = \frac{m(m-1)+n(n-1)}{(m+n)(m+n-1)}.$$

Note: for integers $0 \le k < \ell$, the convention is $\binom{k}{\ell} = 0$. This makes the answers above correct even if m or n or both are 1.

(b) Sampling with replacement. Now the sample has to be ordered (there is a first pick and a second pick). The total number of outcomes is $(m+n)^2$, and the number of favorable outcomes (again counting the green-green and yellow-yellow pairs separately) is m^2+n^2. This gives

$$P\{(\text{g,g}) \text{ or } (\text{y,y})\} = \frac{m^2+n^2}{(m+n)^2}.$$

(c) We simplify the inequality through a sequence of equivalences, by cancelling factors, multiplying away the denominators, and then cancelling some more.

$$\begin{aligned}
& \text{answer to (a)} \;<\; \text{answer to (b)} \\
\iff\quad & \frac{m(m-1)+n(n-1)}{(m+n)(m+n-1)} < \frac{m^2+n^2}{(m+n)^2} \\
\iff\quad & \frac{m(m-1)+n(n-1)}{m+n-1} < \frac{m^2+n^2}{m+n} \\
\iff\quad & (m(m-1)+n(n-1))(m+n) < (m^2+n^2)(m+n-1) \\
\iff\quad & (m^2-m+n^2-n)(m+n) < (m^2+n^2)(m+n) - m^2 - n^2 \\
\iff\quad & (-m-n)(m+n) < -m^2-n^2 \\
\iff\quad & (m+n)^2 > m^2+n^2 \\
\iff\quad & 2mn > 0.
\end{aligned}$$

The last inequality is *always true* for positive m or n. Since the last inequality is equivalent to the first one, the first one is also *always true*.

The conclusion we take from this is that if you want to maximize your chances of getting two of the same color, you want to sample with replacement rather than without replacement. Intuitively this should be obvious: once you remove a ball, you have diminished the chances of drawing another one of the same color.

1.29. (a) Label the liberals 1 through 7 and the conservatives 8 through 13. We do not care about order, so

$$\Omega = \{\{x_1, x_2, x_3, x_4, x_5\} : x_i \in \{1, \dots, 13\}, x_i \neq x_j \text{ if } i \neq j\},$$

in other words the set of 5-element subsets of the set $\{1, 2, \dots, 13\}$. Note that $\#\Omega = \binom{13}{5}$. The event A is

$$\begin{aligned} A &= \{\text{more conservatives than liberals}\} \\ &= \{\{x_1, x_2, x_3, x_4, x_5\} \in \Omega : \text{ at least three elements in } \{8, \dots, 13\}\}. \end{aligned}$$

(b) Let A_3, A_4, A_5 be the events that there are three, four, and five conservatives, respectively, chosen for the committee. Then $A = A_3 \cup A_4 \cup A_5$ and these are mutually exclusive events. By counting the number of ways we can choose conservatives and liberals, we have

$$P(A_3) = \frac{\binom{6}{3} \cdot \binom{7}{2}}{\binom{13}{5}} = \frac{140}{429}$$

$$P(A_4) = \frac{\binom{6}{4} \cdot \binom{7}{1}}{\binom{13}{5}} = \frac{35}{429}$$

$$P(A_5) = \frac{\binom{6}{5} \cdot \binom{7}{0}}{\binom{13}{5}} = \frac{2}{429}.$$

Thus,

$$P(A) = P(A_1) + P(A_2) + P(A_3) = \frac{140}{429} + \frac{35}{429} + \frac{2}{429} = \frac{59}{143}.$$

1.30. First a solution that imagines that the rooks are labeled, for example numbered 1 through 8, and places the rooks on the chessboard in order. There are 64 squares on the chessboard, hence the total number of ways to place 8 rooks in order is $64 \cdot 63 \cdot 62 \cdots 57$.

Next we place the rooks one by one so that none of them can capture any of the previously placed rooks. The first rook can go anywhere on the board and so has $8^2 = 64$ choices. Placing the first rook removes one row and one column from further consideration. Hence the second rook has $7^2 = 49$ options. The first two rooks remove two rows and two columns from further consideration. Thus the third rook has $6^2 = 36$ squares to choose from. The pattern continues. In total, there are $8^2 \cdot 7^2 \cdots 2^2 \cdot 1^2 = (8!)^2$ ways to place the rooks in order, subject to the restriction that no two rooks share a row or a column. The probability comes from the ratio:

$$P(\text{no two rooks can capture each other}) = \frac{(8!)^2}{64 \cdot 63 \cdot 62 \cdots 57} \approx 0.000009109.$$

A solution without order comes by erasing the labels of the rooks and only considering the *set* of squares they occupy. For the number of sets of 8 squares that share no row or column we can take the count $(8!)^2$ from the previous answer and divide it by the number of orderings of the rooks, namely $8!$. This leaves $(8!)^2/8! = 8!$ as the number of sets of 8 squares that share no row or column.

Alternately, pick the squares one column at a time. There are 8 choices for the square from the first column, 7 available squares in the second column, 6 in the third, and so on, to give $8!$ sets of 8 squares that share no row or column.

The total number of sets of 8 squares is $\binom{64}{8}$. So again

$$
\begin{aligned}
&P(\text{no two rooks can capture each other}) \\
&\qquad = \frac{8!}{\binom{64}{8}} = \frac{(8!)^2}{64 \cdot 63 \cdot 62 \cdots 57} \approx 0.000009109.
\end{aligned}
$$

1.31. (a) Number the cards in the deck $1, 2, \ldots, 52$, with the numbers $1, 2, 3, 4$ for the four aces, and the number 1 for the ace of spades. We sample two cards without replacement. We solve the problem without considering order. Thus we set our sample space to be

$$\Omega = \{\{x_1, x_2\} : x_1 \neq x_2, 1 \leq x_i \leq 52 \text{ for } i = 1, 2\},$$

the set of 2-element subsets of the set $\{1, 2, \ldots, 52\}$. We have $\#\Omega = \binom{52}{2} = \frac{52 \cdot 51}{2!} = 1326$.

We need to compute the probability of the event A that both of the chosen cards are aces and one of them is the ace of spades. Thus $A = \{\{1, 2\}, \{1, 3\}, \{1, 4\}\}$ and $\#A = 3$. From this we get $P(A) = \frac{\#A}{\#\Omega} = \frac{3}{1326} = \frac{1}{442}$.

(b) We use the same sample space as in part (a). We need to compute the probability of the event B that at least one of the chosen cards is an ace. It is a bit easier to compute the probability of the complement B^c: this is the event that none of the two chosen cards are aces. B^c is the collection of 2-element sets $\{x_1, x_2\} \in \Omega$ such that both $x_1 \geq 5$ and $x_2 \geq 5$. There are 48 cards that are not aces. The number of 2-element sets of such cards is $\binom{48}{2} = \frac{48 \cdot 47}{2!} = 1128$. Thus $\#B^c = 1128$ and $P(B^c) = \frac{\#B^c}{\#\Omega} = \frac{1128}{1326} = \frac{188}{221}$. Now we can compute $P(B)$ as $P(B) = 1 - P(B^c) = 1 - \frac{188}{221} = \frac{33}{221}$.

Here is an alternative solution with ordered samples of cards.

(a)
$$
\begin{aligned}
&P(\text{two aces and one of them the ace of spaces}) \\
&= P(\text{ace of spades, a different ace}) + P(\text{a different ace, ace of spades}) \\
&= \frac{1 \cdot 3}{52 \cdot 51} + \frac{3 \cdot 1}{52 \cdot 51} = \frac{6}{52 \cdot 51} = \frac{1}{26 \cdot 17} = \frac{1}{442}.
\end{aligned}
$$

(b)
$$
\begin{aligned}
&P(\text{at least one of the cards is an ace}) \\
&= P(\text{ace, ace}) + P(\text{ace, non-ace}) + P(\text{non-ace, ace}) \\
&= \frac{4 \cdot 3}{52 \cdot 51} + \frac{4 \cdot 48}{52 \cdot 51} + \frac{48 \cdot 4}{52 \cdot 51} = \frac{33}{221}.
\end{aligned}
$$

1.32. Here is one way to determine the number of ways to be dealt a full house. We take as our sample space the set of 5-element subsets of the deck of cards:

$$\Omega = \{\{x_1, \dots, x_5\} : x_i \in \{\text{deck of } 52\}, x_i \neq x_j \text{ if } i \neq j\}.$$

Note that $\#\Omega = \binom{52}{5}$.

Now count the number of ways to get a full house. First, choose the face value for the 3 cards that share a face value. There are 13 options. Then select 3 of the 4 suits for this face value. There are $\binom{4}{3}$ ways to do that. We now have the three of a kind selected. Next, choose another face value for the remaining two cards from the remaining 12 face values. Then select 2 of the 4 suits for this face value. There are $\binom{4}{2}$ ways to do that. By the multiplication rule we conclude that there are

$$13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$$

ways to be dealt a full house. Since there are a total of $\binom{52}{5}$ poker hands, the probability is

$$P(\text{full house}) = \frac{13 \cdot 12 \cdot \binom{4}{3}\binom{4}{2}}{\binom{52}{5}} \approx 0.00144.$$

1.33. We let our sample space be the set of ordered 5-tuples from the set $\{1, 2, 3, 4, 5, 6\}$:

$$\Omega = \{(x_1, \dots, x_5) : x_i \in \{1, \dots, 6\}\}.$$

This comes from sampling five times with replacement from $\{1, 2, 3, 4, 5, 6\}$, to produce an ordered sample. Note that $\#\Omega = 6^5$.

We count the number of 5-tuples that give a full house. First pick one of the six numbers (6 choices) for the face value that appears three times. Then pick another number (5 choices) for the face value that appears twice. Next, select 3 of the 5 rolls for the first number. There are $\binom{5}{3}$ ways to choose three slots from five. The remaining two positions are for the second number. (Here is an example: suppose we picked the numbers "4" and "6" and then positions $\{1, 3, 4\}$. Then our full house would be $(4, 6, 4, 4, 6)$.)

Thus there are $6 \cdot 5 \cdot \binom{5}{3}$ ways to roll a full house, and the probability is

$$P(\text{full house}) = \frac{6 \cdot 5 \cdot \binom{5}{3}}{6^5} \approx 0.03858.$$

1.34. Let the corners of the unit square be the points $(0, 0), (0, 1), (1, 1), (1, 0)$. The circle of radius of $1/3$ around the random point is completely within the square if and only if the random point lies within the smaller square with corners $(1/3, 1/3), (2/3, 1/3), (2/3, 2/3), (1/3, 2/3)$. The unit square has area one and the smaller square has area $1/9$. Consequently

$$\begin{aligned} &P(\text{the circle lies inside the unit square}) \\ &= \frac{\text{area of the smaller square}}{\text{area of original unit square}} = \frac{1/9}{1} = \frac{1}{9}. \end{aligned}$$

1.35. (a) Our sample space Ω is the set of points in the triangle with vertices $(0, 0)$, $(3, 0)$ and $(0, 3)$. The area of Ω is $\frac{3 \cdot 3}{2} = \frac{9}{2}$.

The event A describes the points in Ω with distance less than 1 from the y-axis. These are exactly the points in the trapezoid with vertices $(0,0)$, $(1,0)$, $(1,2)$, $(0,3)$. The area of A is $\frac{(3+2)\cdot 1}{2} = \frac{5}{2}$. Since we are choosing our point uniformly from Ω, we can compute $P(A)$ using the ratio of areas:

$$P(A) = \frac{\text{area of } A}{\text{area of } \Omega} = \frac{(\frac{5}{2})}{(\frac{9}{2})} = \frac{5}{9}.$$

(b) We use the same sample space as in part (a). The event B describes the set of points in Ω with distance more than 1 from the origin. The event B^c is the set of points that are in Ω and at most distance one from the origin. B^c is a quarter circle with center at $(0,0)$, radius 1, and corner points at $(1,0)$ and $(0,1)$. The area of B^c is $\frac{\pi}{4}$. Thus

$$P(B^c) = \frac{\text{area of } B^c}{\text{area of } \Omega} = \frac{(\frac{\pi}{4})}{(\frac{9}{2})} = \frac{\pi}{18}$$

and then

$$P(B) = 1 - P(B^c) = 1 - \frac{\pi}{18}.$$

1.36. (a) Since (X, Y) is a uniformly random point, probability is proportional to area:

$$\begin{aligned}
&P(a < X < b) \\
&= P(\text{point } (X,Y) \text{ lies in rectangle with vertices } (a,0),(b,0),(b,1),(a,1)) \\
&= \frac{\text{area of rectangle with vertices } (a,0),(b,0),(b,1),(a,1)}{\text{area of square with vertices } (0,0),(1,0),(1,1),(0,1)} \\
&= b - a.
\end{aligned}$$

Thus, X has a uniform distribution on $[0,1]$.

(b) The region of the xy plane defined by the inequality $|x - y| \le 1/4$ consists of the region between the lines $y = x - 1/4$ and $y = x + 1/4$. Intersecting this region with the unit square gives a region with an area of $7/16$. (Easiest to see by subtracting the complementary triangles from the unit square.) Thus, the desired probability is also $7/16$ since the unit square has an area of one.

1.37. (a) Let $B_k = \{\text{Mary wins on her } k\text{th roll and her } k\text{th roll is a six}\}$.

$$P(B_k) = \frac{(4\cdot 2)^{k-1}\cdot 4\cdot 1}{(6\cdot 6)^k} = \left(\frac{8}{36}\right)^{k-1}\frac{4}{36} = \left(\frac{2}{9}\right)^{k-1}\frac{1}{9}.$$

Then

$$P(\text{Mary wins and her last roll is a six}) = \sum_{k=1}^{\infty} P(B_k) = \sum_{k=1}^{\infty}\left(\frac{2}{9}\right)^{k-1}\frac{1}{9} = \frac{1}{7}.$$

(b) Let $A_k = \{\text{Mary wins on her } k\text{th roll}\}$.

$$P(A_k) = \frac{(2\cdot 4)^{k-1}\cdot 4}{(6\cdot 6)^{k-1}\cdot 6} = \left(\frac{2}{9}\right)^{k-1}\frac{2}{3}.$$

Then

$$P(\text{Mary wins}) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{\infty} \left(\frac{2}{9}\right)^{k-1} \frac{2}{3} = \frac{6}{7}.$$

(c) Suppose Peter starts. Then the game lasts an even number of rolls precisely when Mary wins. Thus the calculation is the same as in the example. Let $D_m = \{\text{the game lasts exactly } m \text{ rolls}\}$. Then for $k \geq 1$,

$$P(D_{2k}) = \frac{(4 \cdot 2)^{k-1} \cdot 4 \cdot 4}{(6 \cdot 6)^k} = \left(\frac{2}{9}\right)^{k-1} \frac{4}{9}$$

and

$$P(\text{the game lasts an even number of rolls}) = \sum_{k=1}^{\infty} P(D_{2k}) = \sum_{k=1}^{\infty} \left(\frac{2}{9}\right)^{k-1} \frac{4}{9} = \frac{4}{7}.$$

If Mary starts, then an even-roll game ends with Peter's roll. In this case

$$P(D_{2k}) = \frac{(2 \cdot 4)^{k-1} \cdot 2 \cdot 2}{(6 \cdot 6)^k} = \left(\frac{2}{9}\right)^{k-1} \frac{1}{9}$$

and

$$P(\text{the game lasts an even number of rolls}) = \sum_{k=1}^{\infty} P(D_{2k}) = \sum_{k=1}^{\infty} \left(\frac{2}{9}\right)^{k-1} \frac{1}{9} = \frac{1}{7}.$$

(d) Let again $D_m = \{\text{the game lasts exactly } m \text{ rolls}\}$. Suppose Peter starts. Then for $k \geq 1$

$$P(D_{2k}) = \frac{(4 \cdot 2)^{k-1} \cdot 4 \cdot 4}{(6 \cdot 6)^k} = \left(\frac{2}{9}\right)^{k-1} \frac{4}{9}$$

and

$$P(D_{2k-1}) = \frac{(4 \cdot 2)^{k-1} \cdot 2}{(6 \cdot 6)^{k-1} \cdot 6} = \left(\frac{2}{9}\right)^{k-1} \frac{1}{3}.$$

Next, for $j \geq 1$:

$$\begin{aligned}
P(\text{game lasts at most } 2j \text{ rolls}) &= \sum_{m=1}^{2j} P(D_m) = \sum_{k=1}^{j} P(D_{2k}) + \sum_{k=1}^{j} P(D_{2k-1}) \\
&= \sum_{k=1}^{j} \left(\frac{2}{9}\right)^{k-1} \frac{4}{9} + \sum_{k=1}^{j} \left(\frac{2}{9}\right)^{k-1} \frac{1}{3} = \sum_{k=1}^{j} \left(\frac{2}{9}\right)^{k-1} \frac{7}{9} = \sum_{i=0}^{j-1} \left(\frac{2}{9}\right)^{i} \frac{7}{9} \\
&= \frac{\frac{7}{9}(1 - (\frac{2}{9})^j)}{1 - \frac{2}{9}} = 1 - \left(\frac{2}{9}\right)^j
\end{aligned}$$

and

$$P(\text{game lasts at most } 2j-1 \text{ rolls}) = \sum_{m=1}^{2j-1} P(D_m) = \sum_{k=1}^{j-1} P(D_{2k}) + \sum_{k=1}^{j} P(D_{2k-1})$$
$$= \sum_{k=1}^{j-1}\left(\frac{2}{9}\right)^{k-1}\frac{4}{9} + \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{1}{3} = \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{4}{9} - \left(\frac{2}{9}\right)^{j-1}\frac{4}{9} + \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{1}{3}$$
$$= \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{7}{9} - \left(\frac{2}{9}\right)^{j-1}\frac{4}{9} = \sum_{i=0}^{j-1}\left(\frac{2}{9}\right)^{i}\frac{7}{9} - \left(\frac{2}{9}\right)^{j-1}\frac{4}{9}$$
$$= \frac{\frac{7}{9}(1-(\frac{2}{9})^j)}{1-\frac{2}{9}} - \left(\frac{2}{9}\right)^{j-1}\frac{4}{9} = 1 - \left(\frac{2}{9}\right)^{j} - \left(\frac{2}{9}\right)^{j-1}\frac{4}{9} = 1 - 3\left(\frac{2}{9}\right)^{j}.$$

Finally, suppose Mary starts. Then for $k \ge 1$

$$P(D_{2k}) = \frac{(2\cdot 4)^{k-1}\cdot 2\cdot 2}{(6\cdot 6)^k} = \left(\frac{2}{9}\right)^{k-1}\frac{1}{9}$$

and

$$P(D_{2k-1}) = \frac{(2\cdot 4)^{k-1}\cdot 4}{(6\cdot 6)^{k-1}\cdot 6} = \left(\frac{2}{9}\right)^{k-1}\frac{2}{3}.$$

Next, for $j \ge 1$:

$$P(\text{game lasts at most } 2j \text{ rolls}) = \sum_{m=1}^{2j} P(D_m) = \sum_{k=1}^{j} P(D_{2k}) + \sum_{k=1}^{j} P(D_{2k-1})$$
$$= \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{1}{9} + \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{2}{3} = \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{7}{9} = \sum_{i=0}^{j-1}\left(\frac{2}{9}\right)^{i}\frac{7}{9}$$
$$= \frac{\frac{7}{9}(1-(\frac{2}{9})^j)}{1-\frac{2}{9}} = 1 - \left(\frac{2}{9}\right)^{j}$$

and

$$P(\text{game lasts at most } 2j-1 \text{ rolls}) = \sum_{m=1}^{2j-1} P(D_m) = \sum_{k=1}^{j-1} P(D_{2k}) + \sum_{k=1}^{j} P(D_{2k-1})$$
$$= \sum_{k=1}^{j}\big[P(D_{2k}) + P(D_{2k-1})\big] - P(D_{2j}) = \sum_{k=1}^{j}\left(\frac{2}{9}\right)^{k-1}\frac{7}{9} - \left(\frac{2}{9}\right)^{j-1}\frac{1}{9}$$
$$= 1 - \left(\frac{2}{9}\right)^{j} - \left(\frac{2}{9}\right)^{j-1}\frac{1}{9} = 1 - \frac{3}{2}\left(\frac{2}{9}\right)^{j}.$$

We see that when Mary starts, the game tends to be over faster.

1.38. If the choice is to be uniformly random, then each integer has to have the same probability, say $P\{k\} = c$ for each integer k. If $c > 0$, choose an integer $n > 1/c$. Then by the additivity of probability over mutually exclusive alternatives,

$$P(\text{the outcome is between 1 and } n) = P\{1, 2, \dots, n\} = nc > 1.$$

Since total probability cannot exceed 1, it must be that $c = 0$ and so $P\{k\} = 0$ for each positive integer k. The total sample space Ω is the union of the sequence of singletons $\{k\}$ as k ranges over all positive integers. Hence again by the additivity axiom

$$1 = P(\Omega) = \sum_{k=1}^{\infty} P\{k\} = \sum_{k=1}^{\infty} 0 = 0.$$

We have a contradiction. Thus there cannot be a sample space and probability P that represents a uniformly chosen random positive integer.

1.39. (a) Define

$$\begin{aligned} A &= \text{the event that a portion of the bill was paid using cash,} \\ B &= \text{the event that a portion of the bill was paid using check,} \\ C &= \text{the event that a portion of the bill was paid using card.} \end{aligned}$$

Note that we know the following:

$$\begin{aligned} P(A) &= 0.78, \quad P(B) = 0.16, \quad P(C) = 0.26 \\ P(AC) &= 0.13, \quad P(AB) = 0.06, \quad P(BC) = 0.04 \\ P(ABC) &= 0.03. \end{aligned}$$

The probability that someone paid with cash only is now seen to be

$$\begin{aligned} P(A \cap (B \cup C)^c) &= P(A) - P(AB) - P(AC) + P(ABC) \\ &= 0.78 - 0.06 - 0.13 + 0.03 = 0.62. \end{aligned}$$

The probability that someone paid with check only is

$$\begin{aligned} P(B \cap (A \cup C)^c) &= P(B) - P(BC) - P(AB) + P(ABC) \\ &= 0.16 - 0.04 - 0.06 + 0.03 = 0.09. \end{aligned}$$

The probability that someone paid with card only is

$$\begin{aligned} P(C \cap (A \cup B)^c) &= P(C) - P(AC) - P(BC) + P(ABC) \\ &= 0.26 - 0.13 - 0.04 + 0.03 = 0.12. \end{aligned}$$

So the probability of the union of these three mutually disjoint sets is,

$$\begin{aligned} &P(\text{only one method of payment}) \\ &= P(\text{cash only}) + P(\text{check only}) + P(\text{card only}) \\ &= 0.62 + 0.09 + 0.12 = 0.83. \end{aligned}$$

(b) Define the event

$$D = \{\text{at least one bill was paid using two or more methods}\}.$$

Then D^c is the event that both bills were paid using only one method. By part (a), we know that there are 83 bills that were paid with only one method. Hence, since there are precisely $\binom{100}{2}$ ways to choose the two checks from the 100, and precisely $\binom{83}{2}$ ways to choose the two bills from the pool of 83, we have

$$P(D) = 1 - P(D^c) = 1 - \frac{\binom{83}{2}}{\binom{100}{2}} = 1 - \frac{83 \cdot 82}{100 \cdot 99} \approx 0.3125.$$

1.40. This is an application of inclusion-exclusion with four events. Below we use some hopefully self-evident summation notation to avoid writing out long sums.

$$\begin{aligned} &P(\text{at least one color is repeated exactly twice}) = P(G \cup R \cup Y \cup W) \\ &\quad = P(G) + P(R) + P(Y) + P(W) - \sum_{\substack{A,B \in \{G,R,Y,W\} \\ A \neq B}} P(AB) \\ &\quad + \sum_{\substack{A,B,C \in \{G,R,Y,W\} \\ A,B,C \text{ distinct}}} P(ABC) - P(GRYW) \end{aligned}$$

Next we derive the probabilities that appear in the equation above. The outcomes of this experiment are 4-tuples from the set $\{\texttt{green}, \texttt{red}, \texttt{yellow}, \texttt{white}\}$. The total number of 4-tuples is $4^4 = 256$.

$$P(G) = P(\text{exactly two greens}) = \frac{\binom{4}{2} \cdot 3 \cdot 3}{256} = \frac{27}{128}.$$

The numerator above is derived as follows: there are $\binom{4}{2}$ ways to pick the positions of the two greens in the 4-tuple. For both of the remaining two positions we have 3 colors to choose from. By the same reasoning, $P(G) = P(R) = P(Y) = P(W) = \frac{27}{128}$.

An event of type AB above means that the four draws yielded two balls of color `a` and two balls of color `b`, where `a` and `b` are two distinct particular colors. The number of 4-tuples in the event AB is $\binom{4}{2} = 6$. We can even list them easily. Here they are in lexicogaphic order:

$$\texttt{aabb}, \texttt{abab}, \texttt{abba}, \texttt{baab}, \texttt{baba}, \texttt{bbaa}.$$

Thus $P(AB) = 6/256 = 3/128$.

Events of the type ABC are empty because four draws cannot yield three different colors that each appear exactly twice. For the same reason $GRYW = \varnothing$.

Putting everything together gives

$$\begin{aligned} &P(\text{at least one color is repeated exactly twice}) \\ &\quad = 4 \cdot \frac{27}{128} - 6 \cdot \frac{3}{128} = \frac{45}{64} \approx 0.7031. \end{aligned}$$

1.41. Let A_1, A_2, A_3 be the events that person 1, 2, and 3 win no games, respectively. Then we want

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3),$$

where we used inclusion-exclusion. Since each person has a probability of $2/3$ of not winning each particular game, we have

$$P(A_i) = \left(\frac{2}{3}\right)^4,$$

for each $i \in \{1, 2, 3\}$. Event A_1A_2 is equivalent to saying that person 1 won all three games, and analogously for A_1A_3 and A_2A_3. Hence

$$P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = \left(\frac{1}{3}\right)^4.$$

Finally, we have $P(A_1A_2A_3) = 0$ because somebody had to win at least one game. Thus,

$$P(A_1 \cup A_2 \cup A_3) = 3 \cdot \left(\frac{2}{3}\right)^4 - 3 \cdot \left(\frac{1}{3}\right)^4 = \frac{5}{9}.$$

1.42. By inclusion-exclusion and the bound $P(A \cup B) \le 1$,

$$P(AB) = P(A) + P(B) - P(A \cup B) > 0.8 + 0.5 - 1 = 0.3.$$

1.43. For $n = 2$ we can use the inclusion exclusion to get

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2) \le P(A_1) + P(A_2).$$

From this we can get the statement step by step for larger and larger values of n. For $n = 3$ we can use the $n = 2$ statement twice, first for $A_1 \cup A_2$ and A_3:

$$P((A_1 \cup A_2) \cup A_3) \le P(A_1 \cup A_2) + P(A_3)$$

and then for A_1 and A_2:

$$P((A_1 \cup A_2) \cup A_3) \le P(A_1 \cup A_2) + P(A_3) \le P(A_1) + P(A_2) + P(A_3).$$

For general n one can do the same by repeating the procedure $n - 1$ times.

The last step of the proof can also be finished with mathematical induction. Here is the induction step. If the statement is assumed to be true for $n - 1$ then, first by the case of two events and then by the induction assumption,

$$\begin{aligned} P((A_1 \cup \cdots \cup A_{n-1}) \cup A_n) &\le P(A_1 \cup \cdots \cup A_{n-1}) + P(A_n) \\ &\le \sum_{k=1}^{n-1} P(A_k) + P(A_n) = \sum_{k=1}^{n} P(A_k). \end{aligned}$$

1.44. Let $\Omega = \{(i,j) : i,j \in \{1,\dots,6\}\}$ be the sample space of the two rolls of the two dice (order matters). Note that $\#\Omega = 36$. For $(i,j) \in \Omega$ we let $X = \max\{i,j\}$ and $Y = \min\{i,j\}$.

(a) The possible values of both X and Y are $\{1,\dots,6\}$.

(b) Note that $P(X \le 6) = 1$. $P(X \le 5)$ is the probability that both rolls yielded five or less. Then there are 5 possibilities for each die, and this event has a probability of

$$P(X \le 5) = \frac{5 \cdot 5}{36} = \frac{25}{36}.$$

Continuing in the same manner:

$$P(X \le 4) = \frac{4 \cdot 4}{36} = \frac{16}{36}, \qquad P(X \le 3) = \frac{3 \cdot 3}{36} = \frac{9}{36}$$
$$P(X \le 2) = \frac{2 \cdot 2}{36} = \frac{4}{36}, \qquad P(X \le 1) = \frac{1 \cdot 1}{36} = \frac{1}{36}.$$

We now have

$$\begin{aligned}
P(X=6) &= P(X\le 6) - P(X\le 5) = 1 - \frac{25}{36} = \frac{11}{36}\\
P(X=5) &= P(X\le 5) - P(X\le 4) = \frac{25}{36} - \frac{16}{36} = \frac{9}{36}\\
P(X=4) &= P(X\le 4) - P(X\le 3) = \frac{16}{36} - \frac{9}{36} = \frac{7}{36}\\
P(X=3) &= P(X\le 3) - P(X\le 2) = \frac{9}{36} - \frac{4}{36} = \frac{5}{36}\\
P(X=2) &= P(X\le 2) - P(X\le 1) = \frac{4}{36} - \frac{1}{36} = \frac{3}{36}\\
P(X=1) &= P(X\le 1) = \frac{1}{36}.
\end{aligned}$$

(c) We can use similar reasoning for the probabilities associated with Y:

$$\begin{aligned}
P(Y\ge 1) &= 1\\
P(Y\ge 2) &= \frac{\text{\# ways to roll only 2s or higher}}{36} = \frac{5^2}{36} = \frac{25}{36}\\
P(Y\ge 3) &= \frac{\text{\# ways to roll only 3s or higher}}{36} = \frac{4^2}{36} = \frac{16}{36}\\
P(Y\ge 4) &= \frac{\text{\# ways to roll only 4s or higher}}{36} = \frac{3^2}{36} = \frac{9}{36}\\
P(Y\ge 5) &= \frac{\text{\# ways to roll only 5s or higher}}{36} = \frac{2^2}{36} = \frac{4}{36}\\
P(Y\ge 6) &= \frac{\text{\# ways to roll only 6s or higher}}{36} = \frac{1^2}{36} = \frac{1}{36},
\end{aligned}$$

and using that $P(Y=k) = P(Y\ge k) - P(Y\ge k+1)$ we get

$$\begin{aligned}
P(Y=1) &= P(Y\ge 1) - P(Y\ge 2) = 1 - \frac{25}{36} = \frac{11}{36}\\
P(Y=2) &= P(Y\ge 2) - P(Y\ge 3) = \frac{25}{36} - \frac{16}{36} = \frac{9}{36}\\
P(Y=3) &= P(Y\ge 3) - P(Y\ge 4) = \frac{16}{36} - \frac{9}{36} = \frac{7}{36}\\
P(Y=4) &= P(Y\ge 4) - P(Y\ge 5) = \frac{9}{36} - \frac{4}{36} = \frac{5}{36}\\
P(Y=5) &= P(Y\ge 5) - P(Y\ge 6) = \frac{4}{36} - \frac{1}{36} = \frac{3}{36}\\
P(Y=6) &= P(Y\ge 6) = \frac{1}{36}.
\end{aligned}$$

1.45. The possible values of X are $4, 3, 2, 1, 0$, because you can win at most 4 dollars. The probability mass function is

$$\begin{aligned}
P(X=4) &= P(\text{the first six was rolled on the first roll}) = \frac{1}{6} \\
P(X=3) &= P(\text{the first six was rolled on the 2nd roll}) = \frac{5}{6^2} \\
P(X=2) &= P(\text{the first six was rolled on the 3rd roll}) = \frac{5^2}{6^3} \\
P(X=1) &= P(\text{the first six was rolled on the 4th roll}) = \frac{5^3}{6^4} \\
P(X=0) &= P(\text{no six was rolled in the first 4 rolls}) = \frac{5^4}{6^4}
\end{aligned}$$

You can check that these probabilities add up to 1, as they should.

1.46. To simplify the counting task we imagine that *all* four balls are drawn from the urn one by one, and then let X denote the number of red balls that come before the yellow. (This is subtly different from the setup of the problem which says that we stop drawing balls once we see the red. This distinction makes no difference for the value that X takes.) Number the red balls 1, 2 and 3, and number the yellow ball 4. Then the sample space is

$$\Omega = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, 2, 3, 4\} \text{ and } x_i \neq x_j \text{ if } i \neq j\}.$$

In other words, Ω is the set of all permutations of the numbers $1, 2, 3, 4$ and consequently $\#\Omega = 4! = 24$.

The possible values of X are $\{0, 1, 2, 3\}$. To compute the probabilities $P(X = k)$ we count the number of ways in which each event can take place.

$$P(X=0) = P(\text{yellow came first}) = \frac{1 \cdot 3 \cdot 2 \cdot 1}{24} = \frac{1}{4}.$$

The numerator equals the number of ways to choose one yellow (1) times the number of ways to choose the first red (3) times the number of ways to choose the second red (2) times the number of ways to choose the last red (1). By similar reasoning,

$$\begin{aligned}
P(X=1) &= P(\text{yellow came second}) = \frac{3 \cdot 1 \cdot 2 \cdot 1}{24} = \frac{1}{4} \\
P(X=2) &= P(\text{yellow came third}) = \frac{3 \cdot 2 \cdot 1 \cdot 1}{24} = \frac{1}{4} \\
P(X=3) &= P(\text{yellow came fourth}) = \frac{3 \cdot 2 \cdot 1 \cdot 1}{24} = \frac{1}{4}.
\end{aligned}$$

1.47. Since $\omega \in [0, 1]$, the random variable Z satisfies $Z(\omega) = e^\omega \in [1, e]$. Thus for $t < 1$ the event $\{Z \le t\}$ is empty and has probability $P(Z \le t) = 0$. If $t \ge e$ then $\{Z \le t\} = \Omega$ (in other words, $Z \le t$ is always true) and so $P(Z \le t) = 1$ for $t \ge e$.

For $1 \le t < e$ then we have this equality of events:

$$\{Z \le t\} = \{\omega : e^\omega \le t\} = \{\omega : \omega \le \ln t\}.$$

Since $0 \le \ln t < 1$, we have $P(\omega : \omega \le \ln t) = \ln t$. In summary,

$$P(Z \le t) = \begin{cases} 0 & \text{if } t < 0 \\ \ln t & \text{if } 0 \le t < e \\ 1 & \text{if } e \le t. \end{cases}$$

1.48. The first digit takes one of the values $0, 1, \dots, 9$, which then also form the range of Y. Since the range of Y is finite, Y must be a discrete random variable.

However, a subtlety having to do with real numbers has to be addressed. Namely, as it stands, the definition of $Y(\omega)$ is ambiguous for certain sample points ω. This is because $0.1 = 0.0\overline{9} = 0.0999\dots$, $0.2 = 0.1\overline{9} = 0.1999\dots$, and so on, up until $1.0 = 0.\overline{9} = 0.999....$ But there are only ten of these real numbers in $[0, 1]$ whose first digit after the decimal is not precisely defined. Since individual numbers have probability zero under a uniform draw from $[0, 1]$, we can ignore these ten sample points $\{0.1, 0.2, \dots, 1.0\}$ without affecting the probabilities of Y.

With the convention of the previous paragraph, for each $k \in \{0, 1, \dots, 9\}$, the event $\{Y = k\}$ is the same as the left-closed right-open interval $[\frac{k}{10}, \frac{k+1}{10})$. Thus

$$P(Y = k) = P\big([\tfrac{k}{10}, \tfrac{k+1}{10}) \big) = \frac{1}{10} \qquad \text{for each } k \in \{0, 1, \dots, 9\}.$$

1.49. (a) To answer the question with inclusion-exclusion, let $A_i = \{i\text{th draw is red}\}$. Then $B = \cup_{i=1}^{\ell} A_i$. To apply (1.20) we need the probabilities $P(A_{i_1} \cap \dots \cap A_{i_k})$ for each choice of indices $1 \le i_1 < \dots < i_k \le \ell$. To see how this goes, let us first derive the example

$$P(A_2 \cap A_5) = P(\text{the 2nd draw and 5th draw are red})$$

by counting favorable outcomes and total outcomes. Each of the ℓ draws comes from a set of n balls, so $\#\Omega = n^{\ell}$. The number of favorable outcomes is $n \cdot 3 \cdot n \cdot n \cdot 3 \cdot n \cdots n = n^{\ell-2} 3^2$ because the second and fifth draws are restricted to the 3 red balls, and the other $\ell - 2$ draws are unrestricted. This gives

$$P(A_2 \cap A_5) = \frac{n^{\ell-2} 3^2}{n^{\ell}} = \left(\frac{3}{n} \right)^2.$$

The same reasoning gives for any choice of k indices $1 \le i_1 < \dots < i_k \le \ell$

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{n^{\ell-k} 3^k}{n^{\ell}} = \left(\frac{3}{n} \right)^k.$$

Then

$$\begin{aligned} P(B) &= \sum_{k=1}^{\ell} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le \ell} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=1}^{\ell} (-1)^{k+1} \binom{\ell}{k} \left(\frac{3}{n} \right)^k = -\sum_{k=1}^{\ell} \binom{\ell}{k} \left(-\frac{3}{n} \right)^k \\ &= 1 - \sum_{k=0}^{\ell} \binom{\ell}{k} \left(-\frac{3}{n} \right)^k = 1 - \left(1 - \frac{3}{n} \right)^{\ell}. \end{aligned}$$

In the second to last equality above we added and subtracted the term for $k = 0$ which is 1. This enabled us to apply the binomial theorem (Fact D.2 in Appendix D).

(b) Let $B_k = \{\text{a red ball is seen exactly } k \text{ times}\}$ for $1 \le k \le \ell$. There are $\binom{\ell}{k}$ ways to decide which k of the ℓ draws produce the red ball. Thus there are altogether $\binom{\ell}{k} 3^k (n-3)^{\ell-k}$ ways to draw exactly k red balls. Then

$$P(B_k) = \frac{\binom{\ell}{k} 3^k (n-3)^{\ell-k}}{n^\ell} = \binom{\ell}{k} \left(\frac{3}{n}\right)^k \left(1 - \frac{3}{n}\right)^{\ell-k}$$

and then by the binomial theorem (add and subtract the $k = 0$ term)

$$\begin{aligned} P(B) &= \sum_{k=1}^{\ell} P(B_k) = \sum_{k=1}^{\ell} \binom{\ell}{k} \left(\frac{3}{n}\right)^k \left(1 - \frac{3}{n}\right)^{\ell-k} \\ &= \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\frac{3}{n}\right)^k \left(1 - \frac{3}{n}\right)^{\ell-k} - \left(1 - \frac{3}{n}\right)^{\ell} = 1 - \left(1 - \frac{3}{n}\right)^{\ell}. \end{aligned}$$

(c) The quickest solution comes by using the complement $B^c = \{\text{each draw is green}\}$.

$$P(B) = 1 - P(B^c) = 1 - \frac{(n-3)^\ell}{n^\ell} = 1 - \left(1 - \frac{3}{n}\right)^{\ell}.$$

Solutions to Chapter 2

2.1. We can set our sample space to be $\Omega = \{(a_1, a_2) : 1 \le a_i \le 6\}$. We have $\#\Omega = 36$ and each outcome is equally likely.

Denote by A the event that at least one number is even and by B the event that the sum is 8. Then we need $P(A|B)$ which can be computed from the definition as $P(A|B) = \frac{P(AB)}{P(B)}$.

We have $B = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$, and hence $P(B) = \frac{\#B}{\#\Omega} = \frac{5}{36}$. Moreover, $AB = \{(2,6), (4,4), (6,2)\}$ and hence $P(AB) = \frac{\#AB}{\#\Omega} = \frac{3}{36} = \frac{1}{12}$. Thus $P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{12}}{\frac{5}{36}} = \frac{3}{5}$.

Since the outcomes are equally likely, we can equivalently find the answer from $P(A|B) = \frac{\#AB}{\#B} = \frac{3}{5}$.

2.2. $A = \{\text{second flip is tails}\} = \{(H,T,H), (H,T,T), (T,T,H), (T,T,T)\}$,

$$B = \{\text{at most one tails}\} = \{(H,H,H), (H,H,T), (H,T,H), (T,H,H)\}.$$

Hence $AB = \{(H,T,H)\}$, and since we have equally likely outcomes,

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{\#AB}{\#B} = \tfrac{1}{4}.$$

2.3. We set the sample space as $\Omega = \{1, 2, \dots, 100\}$. We have $\#\Omega = 100$ and each outcome is equally likely.

Let A denote the event that the chosen number is divisible by 3 and B denote the event that at least one digit is equal to 5. Then

$$B = \{5, 15, 25, \dots, 95\} \cup \{50, 51, \dots, 59\}$$

and $\#B = 19$. (As there are 10 numbers with 5 as the last digit, 10 numbers with 5 at the tens place, and 55 was counted both times.) We also have

$$AB = \{15, 45, 51, 54, 57, 75\}, \quad \#AB = 6.$$

This gives $P(A|B) = \frac{P(AB)}{P(B)} = \frac{6/100}{19/100} = \frac{6}{19}$.

2.4. Let A be the event that we picked the ball labeled 5 and B the event that we picked the first urn. Then we have $P(B) = 1/2$, $P(B^c) = P(\text{we picked the second urn}) = 1/2$. Moreover, from the setup if the problem

$$P(A|B) = P(\text{we chose the number 5} \,|\, \text{we chose from the first urn}) = 0,$$

$$P(A|B^c) = P(\text{we chose the number 5} \,|\, \text{we chose from the second urn}) = \frac{1}{3}.$$

We compute $P(A)$ by conditioning on B and B^c:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = 0 \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

2.5. Let A be the event that we picked the number 2 and B the event that we picked the first urn. Then we have $P(B) = 1/5$, $P(B^c) = P(\text{we picked the second urn}) = 4/5$. Moreover, from the setup if the problem

$$P(A|B) = P(\text{we chose the number 2} \,|\, \text{we chose from the first urn}) = \frac{1}{3},$$

$$P(A|B^c) = P(\text{we chose the number 2} \,|\, \text{we chose from the second urn}) = \frac{1}{4}.$$

Then we can compute $P(A)$ by conditioning on B and B^c:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = \frac{1}{3} \cdot \frac{1}{5} + \frac{1}{4} \cdot \frac{4}{5} = \frac{4}{15}.$$

2.6. Define events

$$A = \{\text{Alice watches TV tomorrow}\} \quad \text{and} \quad B = \{\text{Betty watches TV tomorrow}\}.$$

(a) $P(AB) = P(A)P(B|A) = 0.6 \cdot 0.8 = 0.48$.

(b) Intuitively, the answer must be the same 0.48 as in part (a) because Betty cannot watch TV unless Alice is also watching. Mathematically, this says that $P(B|A^c) = 0$. Then by the law of total probability,

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = 0.8 \cdot 0.6 + 0 \cdot 0.4 = 0.48.$$

(c) $P(AB^c) = P(A) - P(AB) = 0.6 - 0.48 = 0.12$. Or, by conditioning and using the outcome of Exercise 2.7(a),

$$P(AB^c) = P(A)P(B^c|A) = P(A)\big(1 - P(B|A)\big) = 0.6 \cdot 0.2 = 0.12.$$

2.7. (a) By definition $P(A^c|B) = \frac{P(A^cB)}{P(B)}$. We have $A^cB \cup AB = B$, and the two sets on the left are disjoint, so $P(A^cB) + P(AB) = P(B)$, and $P(A^cB) = P(B) - P(AB)$. This gives

$$P(A^c|B) = \frac{P(A^cB)}{P(B)} = \frac{P(B) - P(AB)}{P(B)} = 1 - \frac{P(AB)}{P(B)} = 1 - P(A|B).$$

(b) From part (a) we have $P(A^c|B) = 1 - P(A|B) = 0.4$. Then $P(A^cB) = P(A^c|B)P(B) = 0.4 \cdot 0.5 = 0.2$.

2.8. Let A_1, A_2, A_3 denote the events that the first, second and third cards are queen, king and ace, respectively. We need to compute $P(A_1A_2A_3)$. One could do this by counting favorable outcomes. But conditional probabilities provide an easier way because then we can focus on picking one card at a time. We just have to keep track of how earlier picks influence the probabilities of the later picks.

We have $P(A_1) = \frac{4}{52} = \frac{1}{13}$ since there are 52 equally likely choices for the first pick and four of them are queens. The conditional probability $P(A_2 \mid A_1)$ must reflect the fact that one queen has been removed from the deck and is no longer a possible outcome. Since the outcomes are still equally likely, the conditional probability of getting a king for the second pick is $\frac{4}{51}$. Similarly, when we compute $P(A_3 \mid A_1 A_2)$ we can assume that we pick a card out of 50 (with one queen and one king removed) and thus the conditional probability of picking an ace will be $\frac{4}{50} = \frac{2}{25}$. Thus the probability of $A_1 A_2 A_3$ is given by

$$P(A_1 A_2 A_3) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_2 A_1) = \tfrac{1}{13} \cdot \tfrac{4}{51} \cdot \tfrac{2}{25} = \tfrac{8}{16{,}575}.$$

2.9. Let C be the event that we chose the ball 3 and D the event that we chose from the second urn. Then we have

$$P(D) = \frac{4}{5}, \quad P(D^c) = \frac{1}{5}, \quad P(C|D) = \frac{1}{4}, \quad P(C|D^c) = \frac{1}{3}.$$

We need to compute $P(D|C)$, which we can do using Bayes' formula:

$$P(D|C) = \frac{P(C|D)P(D)}{P(C|D)P(D) + P(C|D^c)P(D^c)} = \frac{\frac{1}{4} \cdot \frac{4}{5}}{\frac{1}{4} \cdot \frac{4}{5} + \frac{1}{3} \cdot \frac{1}{5}} = \frac{3}{4}.$$

2.10. Define events:

$$A = \{\text{outcome of the roll is } 4\} \quad \text{and} \quad B_k = \{\text{the } k\text{-sided die is picked}\}.$$

Then

$$P(B_6|A) = \frac{P(A \cap B_6)}{P(A)} = \frac{P(A|B_6)P(B_6)}{P(A|B_4)P(B_4) + P(A|B_6)P(B_6) + P(A|B_{12})P(B_{12})}$$
$$= \frac{\frac{1}{6} \cdot \frac{1}{3}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{1}{3} + \frac{1}{12} \cdot \frac{1}{3}} = \frac{1}{3}.$$

2.11. Let A be the event that the chosen customer is reckless. Let B be the event that the chosen customer has an accident. We know the following:

$$P(A) = 0.2, \quad P(A^c) = 0.8, \quad P(B|A) = 0.04, \quad \text{and } P(B|A^c) = 0.01.$$

The probability asked for is $P(A^c|B)$. Using Bayes' formula we get

$$P(A^c|B) = \frac{P(B|A^c)P(A^c)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{0.01 \cdot 0.80}{0.04 \cdot 0.2 + 0.01 \cdot 0.80} = \frac{1}{2}.$$

2.12. (a) $A = \{X \text{ is even}\}$, $B = \{X \text{ is divisible by } 5\}$. $\#A = 50$, $\#B = 20$ and $AB = \{10, 20, \dots, 100\}$ so $\#AB = 10$. Thus

$$P(A)P(B) = \tfrac{50}{100} \cdot \tfrac{20}{100} = \tfrac{1}{10} \quad \text{and} \quad P(AB) = \tfrac{10}{100} = \tfrac{1}{10}.$$

This shows $P(A)P(B) = P(AB)$ and verifies the independence of A and B.

(b) $C = \{X \text{ has two digits}\} = \{10, 11, 12, \dots, 99\}$ and $\#C = 90$.

$$D = \{X \text{ is divisible by } 3\} = \{3, 6, 9, 12, \dots, 99\} \text{ and } \#D = 33.$$

$CD = \{12, 15, \dots, 99\}$ and $\#C = 30$. Thus

$$P(C)P(D) = \tfrac{90}{100} \cdot \tfrac{33}{100} \approx 0.297 \quad \text{and} \quad P(CD) = \tfrac{30}{100} = \tfrac{3}{10}.$$

This shows $P(C)P(D) \neq P(CD)$ and verifies that C and D are not independent.

(c) $E = \{X \text{ is a prime}\} = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97\}$,

and $\#E = 25$.

$$F = \{X \text{ has a digit } 5\} = \{5, 15, 25, \dots, 95\} \cup \{50, 51, \dots, 59\}$$

and $\#F = 19$. $EF = \{5, 53, 59\}$ and $\#EF = 3$. We have

$$P(E)P(F) = \tfrac{25}{100} \cdot \tfrac{19}{100} = 0.0475 \quad \text{and} \quad P(EF) = \tfrac{3}{100}.$$

This shows $P(E)P(F) \neq P(EF)$ and verifies that E and F are not independent.

2.13. We need to check whether or not we have

$$P(AB) = P(A)P(B).$$

We know that $P(A)P(B) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$. We also know that $A = AB \cup AB^c$ and that the events AB and AB^c are disjoint. Thus,

$$\frac{1}{3} = P(A) = P(AB) + P(AB^c) = P(AB) + \frac{2}{9}.$$

Thus,

$$P(AB) = \frac{1}{3} - \frac{2}{9} = \frac{1}{9} = P(A)P(B),$$

so A and B are independent.

2.14. Since $P(AB) = P(\varnothing) = 0$ and independence requires $P(A)P(B) = P(AB)$, disjoint events A and B are independent if and only if at least one of them has probability zero.

2.15. Number the days by 1,2,3,4,5 starting from Monday. Let $X_i = 1$ if Ramona catches her bus on day i and $X_i = 0$ if she misses it. Then we need to compute $P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0)$. By assumption, the events $\{X_1 = 1\}$, $\{X_2 = 1\}$, $\{X_3 = 0\}$, $\{X_4 = 1\}$, $\{X_5 = 0\}$ are independent from each other, and $P(X_i = 1) = \frac{9}{10}$ and $P(X_i = 0) = \frac{1}{10}$. Thus

$$\begin{aligned}
&P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0) \\
&\quad = P(X_1 = 1)P(X_2 = 1)P(X_3 = 0)P(X_4 = 1)P(X_5 = 0) \\
&\quad = \frac{9}{10} \cdot \frac{9}{10} \cdot \frac{1}{10} \cdot \frac{9}{10} \cdot \frac{1}{10} = \frac{729}{10000}.
\end{aligned}$$

2.16. Let us label heads as 0 and tails as 1. The sample space is

$$\Omega = \{(s_1, s_2, s_3) : \text{each } s_i \in \{0, 1\}\},$$

the set of ordered triples of zeros and ones. $\#\Omega = 8$ and so for equally likely outcomes we have $P(\omega) = 1/8$ for each $\omega \in \Omega$. The events and their probabilities we need for answering the question of independence are

$$
\begin{aligned}
P(A_1) &= P\{(0,0,0),(0,0,1),(0,1,0),(0,1,1)\} = \tfrac{4}{8} = \tfrac{1}{2},\\
P(A_2) &= P\{(0,1,0),(0,1,1),(1,0,0),(1,0,1)\} = \tfrac{4}{8} = \tfrac{1}{2},\\
P(A_3) &= P\{(0,1,1),(1,0,1),(1,1,0),(0,0,0)\} = \tfrac{4}{8} = \tfrac{1}{2},\\
P(A_1A_2) &= \{(0,1,0),(0,1,1)\} = \tfrac{2}{8} = \tfrac{1}{4} = \tfrac{1}{2}\cdot\tfrac{1}{2} = P(A_1)P(A_2),\\
P(A_1A_3) &= \{(0,1,1),(0,0,0)\} = \tfrac{2}{8} = \tfrac{1}{4} = \tfrac{1}{2}\cdot\tfrac{1}{2} = P(A_1)P(A_3),\\
P(A_2A_3) &= \{(0,1,1),(1,0,1)\} = \tfrac{2}{8} = \tfrac{1}{4} = \tfrac{1}{2}\cdot\tfrac{1}{2} = P(A_1)P(A_3),\\
P(A_1A_2A_3) &= \{(0,1,1)\} = \tfrac{1}{8} = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = P(A_1)P(A_2)P(A_3).
\end{aligned}
$$

All the four possible combinations of more than two events from A_1, A_2, A_3 satisfy the product identity. Hence independence of A_1, A_2, A_3 has been verified.

2.17. We have $AB \cup C = ABC^c \cup C$, and the events ABC^c and C are disjoint. Thus $P(AB \cup C) = P(ABC^c) + P(C)$. Since A, B, C are mutually independent, this is also true for A, B, C^c. Thus

$$P(ABC^c) = P(A)P(B)P(C^c) = \frac{1}{2}\cdot\frac{1}{3}\cdot\Big(1-\frac{1}{4}\Big) = \frac{1}{8},$$

From this we get

$$P(AB \cup C) = P(ABC^c) + P(C) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}.$$

Here is another solution: by inclusion-exclusion $P(AB\cup C) = P(AB)+P(C)-P(ABC)$. Because of independence

$$P(AB) = P(A)P(B) = \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{6}, \qquad P(ABC) = P(A)P(B)P(C) = \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{4} = \frac{1}{24}.$$

Thus

$$P(AB \cup C) = P(AB) + P(C) - P(ABC) = \frac{1}{6} + \frac{1}{4} - \frac{1}{24} = \frac{3}{8}.$$

2.18. There are 90 numbers to choose from and so each outcome has probability $\frac{1}{90}$.

(a) From enumerating the possible values of X, we see that $P(X = k) = \frac{1}{9}$ for each $k \in \{1, 2, \ldots, 9\}$. (For example, the event $\{X = 3\} = \{30, 31, \ldots, 39\}$ has 10 outcomes from the 90 total.) For Y we have $P(Y = \ell) = \frac{1}{10}$ for each $\ell \in \{0, 1, 2, \ldots, 9\}$. (For example, the event $\{Y = 3\} = \{13, 23, 33, \ldots, 93\}$ has 9 outcomes from the 90 total.)

The intersection $\{X = k, Y = \ell\}$ contains exactly one number from the 90 outcomes, namely $10k + \ell$. (For example $\{X = 3, Y = 5\} = \{35\}$). Thus for each pair (k, ℓ) of possible values,

$$P(X = k, Y = \ell) = P\{10k + \ell\} = \tfrac{1}{90} = \tfrac{1}{9}\cdot\tfrac{1}{10} = P(X = k)P(Y = \ell).$$

Thus we have checked that X and Y are independent.

(b) To show that independence fails, we need to find only one case where the product property $P(X = k, Z = m) = P(X = k)P(Z = m)$ fails. Let's take an extreme case. The smallest possible value for Z is 1 that comes only from the outcome 10, since the sum of the digits is $1 + 0 = 1$. (Formally, since Z is a function on Ω, $Z(10) = 1 + 0 = 1$.) And so $P(Z = 1) = P\{10\} = \frac{1}{90}$. If we take $X = 2$, we cannot get $Z = 1$. Here is the precise derivation:

$$P(X = 2, Z = 1) = P(\{20, 21, \ldots, 29\} \cap \{10\}) = P(\varnothing) = 0.$$

Since $P(X = 2)P(Z = 1) = \frac{1}{9} \cdot \frac{1}{90} = \frac{1}{810} \neq 0$, we have shown that X and Z are not independent.

2.19. (a) If we draw with replacement then we have 7^2 equally likely outcomes for the two picks. Counting the favorable outcomes gives

$$\begin{aligned} P(X_1 = 4) &= \frac{1 \cdot 7}{7 \cdot 7} = \frac{1}{7} \\ P(X_2 = 5) &= \frac{7 \cdot 1}{7 \cdot 7} = \frac{1}{7} \\ P(X_1 = 4, X_2 = 5) &= \frac{1}{7 \cdot 7} = \frac{1}{49}. \end{aligned}$$

(b) If we draw without replacement then we have $7 \cdot 6$ equally likely outcomes for the two picks. Counting the favorable outcomes gives

$$\begin{aligned} P(X_1 = 4) &= \frac{1 \cdot 6}{7 \cdot 6} = \frac{1}{7} \\ P(X_2 = 5) &= \frac{6 \cdot 1}{7 \cdot 6} = \frac{1}{7} \\ P(X_1 = 4, X_2 = 5) &= \frac{1}{7 \cdot 6} = \frac{1}{42}. \end{aligned}$$

(c) The answer to part (b) showed that $P(X_1 = 4)P(X_2 = 5) \neq P(X_1 = 4, X_2 = 5)$. This proves that X_1 and X_2 are not independent when drawing without replacement.

Part (a) showed that the *events* $\{X_1 = 4\}$ and $\{X_2 = 5\}$ are independent when drawing with replacement, but this is not enough for proving that the *random variables* X_1 and X_2 are independent. Independence of random variables requires checking $P(X_1 = a)P(X_2 = b) = P(X_1 = a, X_2 = b)$ for all possible choices of a and b. (This can be done and so independence of X_1 and X_2 does actually hold here.)

2.20. (a) Let S_5 denote the number of threes in the first five rolls. Then

$$P(S_5 \leq 2) = \sum_{k=0}^{2} \binom{5}{k} \left(\tfrac{1}{6}\right)^k \left(\tfrac{5}{6}\right)^{5-k}.$$

(b) Let N be the number of rolls needed to see the first three. Then from the p.m.f. of a geometric random variable,

$$P(N > 4) = \sum_{k=5}^{\infty} \left(\tfrac{5}{6}\right)^{k-1} \left(\tfrac{1}{6}\right) = \left(\tfrac{5}{6}\right)^4.$$

Equivalently,

$$P(N > 4) = P(\text{no three in the first four rolls}) = \left(\tfrac{5}{6}\right)^4.$$

(c) We can approach this in a couple different ways. By using the independence of the rolls,

$$\begin{aligned}&P(5 \le N \le 20)\\ &= P(\text{no three in the first four rolls, at least one three in rolls 5–20})\\ &= \left(\tfrac{5}{6}\right)^4\left(1 - \left(\tfrac{5}{6}\right)^{16}\right) = \left(\tfrac{5}{6}\right)^4 - \left(\tfrac{5}{6}\right)^{20}.\end{aligned}$$

Equivalently, thinking of the roll at which the first three comes,

$$\begin{aligned}P(5 \le N \le 20) &= P(N \ge 5) - P(N \ge 21)\\ &= \sum_{k=5}^{\infty}\left(\tfrac{5}{6}\right)^{k-1}\left(\tfrac{1}{6}\right) - \sum_{k=21}^{\infty}\left(\tfrac{5}{6}\right)^{k-1}\left(\tfrac{1}{6}\right)\\ &= \left(\tfrac{5}{6}\right)^4 - \left(\tfrac{5}{6}\right)^{20}.\end{aligned}$$

2.21. (a) Let S be the number of problems she gets correct. Then $S \sim \text{Bin}(4, 0.8)$ and

$$\begin{aligned}P(\text{Jane gets an A}) = P(S \ge 3) &= P(S = 3) + P(S = 4)\\ &= \binom{4}{3}(0.8)^3(0.2) + (0.8)^4\\ &= 0.8192.\end{aligned}$$

(b) Let S_2 be the number of problems Jane gets correct out of the last three. Then $S_2 \sim \text{Bin}(3, 0.8)$. Let $X_1 \sim \text{Bern}(0.8)$ model whether or not she gets the first problem correct. By assumption, S_2 and X_1 are independent. We have

$$\begin{aligned}P(S \ge 3 \,|\, X_1 = 1) &= \frac{P(S \ge 3, X_1 = 1)}{P(X_1 = 1)}\\ &= \frac{P(S_2 \ge 2, X_1 = 1)}{P(X_1 = 1)} = \frac{P(S_2 \ge 2)P(X_1 = 1)}{P(X_1 = 1)}.\end{aligned}$$

The last equality followed by the independence of S_2 and X_1. Hence,

$$P(S \ge 3|X_1 = 1) = P(S_2 \ge 2) = \binom{3}{2}(0.8)^2(0.2) + (0.8)^3 = 0.896.$$

2.22. (a) Let us encode the possible events in a single round as

$$A_R = \{\text{Annie chooses rock}\}, \quad A_P = \{\text{Annie chooses paper}\}$$
$$\text{and} \quad A_S = \{\text{Annie chooses scissors}\}$$

and similarly B_R, B_P and B_S for Bill. Then, using the independence of the players' choices,

$$\begin{aligned}&P(\text{Ann wins the round}) = P(A_R B_S) + P(A_P B_R) + P(A_S B_P)\\ &\quad = P(A_R)P(B_S) + P(A_P)P(B_R) + P(A_S)P(B_P)\\ &\quad = \tfrac{1}{3}\cdot\tfrac{1}{3} + \tfrac{1}{3}\cdot\tfrac{1}{3} + \tfrac{1}{3}\cdot\tfrac{1}{3} = \tfrac{1}{3}.\end{aligned}$$

Conceptually quicker than enumerating cases would be to notice that no matter what Ann chooses, the probability that Bill makes a losing choice is $\frac{1}{3}$.

Hence by the law of total probability, Ann's probability of winning must be $\frac{1}{3}$. Here is the calculation:

$$\begin{aligned}
P(\text{Ann wins the round}) &= P(\text{Ann wins the round} \mid A_R)P(A_R) \\
&\quad + P(\text{Ann wins the round} \mid A_P)P(A_P) \\
&\quad + P(\text{Ann wins the round} \mid A_S)P(A_S) \\
&= \tfrac{1}{3} \cdot P(A_R) + \tfrac{1}{3} \cdot P(A_P) + \tfrac{1}{3} \cdot P(A_S) \\
&= \tfrac{1}{3} \cdot \big(P(A_R) + P(A_P) + P(A_S)\big) = \tfrac{1}{3}.
\end{aligned}$$

(b) By the independence of the outcomes of different rounds,

$$\begin{aligned}
&P(\text{Ann's first win happens in the fourth round}) \\
&= P(\text{Ann does not win any of the first three rounds,} \\
&\qquad\qquad \text{Ann wins the fourth round}) \\
&= \tfrac{2}{3} \cdot \tfrac{2}{3} \cdot \tfrac{2}{3} \cdot \tfrac{1}{3} = \tfrac{8}{81}.
\end{aligned}$$

(c) Again by the independence of the outcomes of different rounds,

$$P(\text{Ann does not win any of the first four rounds}) = \left(\tfrac{2}{3}\right)^4 = \tfrac{16}{81}.$$

2.23. Whether there is an accident on a given day can be treated as the outcome of a trial (where success means that there is at least one accident). The success probability is $p = 1 - 0.95 = 0.05$ and the failure probability is 0.95.

(a) The probability of no accidents at this intersection during the next 7 days is the probability that the first seven trials failed, which is $(1-p)^7 = 0.95^7 \approx 0.6983$.

(b) There are 30 days in September. Let X be the number of days that have at least one accident. X counts the number of 'successes' among 30 trials, so $X \sim \text{Bin}(30, 0.05)$. Using the probability mass function of the binomial we get

$$P(X = 2) = \binom{30}{2} 0.05^2 0.95^{28} \approx 0.2586.$$

(c) Let N denote the number of days we have to wait for the next accident, or equivalently, the number of trials needed for the first success. N has geometric distribution with parameter $p = 0.05$. We need to compute $P(4 < N \le 10)$. The event $\{4 < N \le 10\}$ is the same as $\{N \in \{5, 6, 7, 8, 9, 10\}\}$. Using the probability mass function of the geometric distribution,

$$\begin{aligned}
P(4 < N \le 10) &= \sum_{k=5}^{10} P(N = k) = \sum_{k=5}^{10} (1-p)^{k-1} p = \sum_{k=5}^{10} 0.95^{k-1} 0.05 \\
&\approx 0.2158.
\end{aligned}$$

Here is an alternative solution. Note that

$$\begin{aligned}
P(4 < N \le 10) &= P(N \le 10) - P(N \le 4) \\
&= (1 - P(N > 10)) - (1 - P(N > 4)) \\
&= P(N > 4) - P(N > 10).
\end{aligned}$$

For any positive integer k the event $\{N > k\}$ is the same as having k failures in the first k trials. By part (a) the probability of this is $(1-p)^k$, which gives $P(N > k) = (1-p)^k = 0.95^k$ and then

$$\begin{aligned} P(4 < N \le 10) &= P(N > 4) - P(N > 10) = (1-p)^4 - (1-p)^{10} \\ &= 0.95^4 - 0.95^{10} \approx 0.2158. \end{aligned}$$

2.24. (a) X is hypergeometric with parameters $(6, 4, 3)$.

(b) The probability mass function of X is

$$P(X = k) = \frac{\binom{4}{k}\binom{2}{3-k}}{\binom{6}{3}} \qquad \text{for } k \in \{0, 1, 2, 3\},$$

with the convention that $\binom{a}{k} = 0$ for integers $k > a \ge 0$. In particular, $P(X = 0) = 0$ because with only 2 men available, a team of 3 cannot consist of men alone.

2.25. Define events: $A = \{\text{first roll is a three}\}$, $B = \{\text{second roll is a four}\}$, $D_i = \{\text{the die has } i \text{ sides}\}$. Assume that A and B are independent, given D_i, for each $i = 4, 6, 12$.

$$\begin{aligned} P(AB) &= \sum_{i=4,6,12} P(AB|D_i)P(D_i) = \sum_{i=4,6,12} P(A|D_i)P(B|D_i)P(D_i) \\ &= \big((\tfrac{1}{4})^2 + (\tfrac{1}{6})^2 + (\tfrac{1}{12})^2\big) \cdot \tfrac{1}{3}. \end{aligned}$$

$$P(D_6|AB) = \frac{P(AB|D_6)P(D_6)}{P(AB)} = \frac{(\frac{1}{6})^2 \cdot \frac{1}{3}}{\big((\frac{1}{4})^2 + (\frac{1}{6})^2 + (\frac{1}{12})^2\big) \cdot \frac{1}{3}} = \tfrac{2}{7}.$$

2.26.

$$P((AB) \cap (CD)) = P(ABCD) = P(A)P(B)P(C)P(D) = P(AB)P(CD).$$

The very first equality is set algebra, namely, the associativity of intersection. This can be taken as intuitively obvious, or verified from the definition of intersection and common sense logic:

$$\begin{aligned} \omega \in (AB) \cap (CD) &\iff \omega \in AB \ \text{ and } \ \omega \in CD \\ &\iff \big(\omega \in A \ \text{ and } \ \omega \in B\big) \ \text{ and } \ \big(\omega \in C \ \text{ and } \ \omega \in D\big) \\ &\iff \omega \in A \ \text{ and } \ \omega \in B \ \text{ and } \ \omega \in C \ \text{ and } \ \omega \in D \\ &\iff \omega \in ABCD. \end{aligned}$$

Then we used the product rule first for all four events A, B, C, D, and then separately for the pairs A, B and C, D.

2.27. (a) First introduce the necessary events. Let A be the event that we picked Urn I. Then A^c is the event that we picked Urn II. Let B_1 the event that we picked a green ball. Then

$$P(A) = P(A^c) = \frac{1}{2}, \qquad P(B_1|A) = \frac{1}{3}, \qquad P(B_1|A^c) = \frac{2}{3}.$$

$P(B_1)$ is computed from the law of total probability:

$$P(B_1) = P(B_1|A)P(A) + P(B_1|A^c)P(A^c) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{2}.$$

(b) The two experiments are identical and independent. Thus the probability of picking green both times is the square of the probability from part (a): $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

(c) Let B_2 be the event that we picked a green ball in the second draw. The events B_1, B_2 are conditionally independent given A (and given A^c), since we are sampling with replacement from the same urn. Thus

$$P(B_2|A) = \frac{1}{3}, \quad P(B_2|A^c) = \frac{2}{3},$$
$$P(B_1B_2|A) = P(B_1|A)P(B_2|A), \quad P(B_1B_2|A^c) = P(B_1|A^c)P(B_2|A^c).$$

From this we get

$$\begin{aligned} P(B_1B_2) &= P(B_1B_2|A)P(A) + P(B_1B_2|A^c)P(A^c) \\ &= P(B_1|A)P(B_2|A)P(A) + P(B_1|A^c)P(B_2|A^c)P(A^c) \\ &= (\tfrac{1}{3})^2 \tfrac{1}{2} + (\tfrac{2}{3})^2 \tfrac{1}{2} = \tfrac{5}{18}. \end{aligned}$$

(d) The probability of getting a green from the first urn is $\frac{1}{3}$ and the probability of getting a green from the second urn is $\frac{2}{3}$. Since the picks are independent, the probability of both picks being green is $\frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$.

2.28. (a) The number of aces I get in the first game is hypergeometric with parameters $(52, 4, 13)$.

(b) The number of games in which I receive at least one ace during the evening is binomial with parameters $(50, 1 - (\binom{48}{13}/\binom{52}{13}))$.

(c) The number of games in which all my cards are from the same suit is binomial with parameters $(50, \binom{52}{13}^{-1})$.

(d) The number of spades I receive in the 5th game is hypergeometric with parameters $(52, 13, 13)$.

2.29. Let E_1, E_2, E_3, N be the events that Uncle Bob hits a single, double, triple, or not making it on base, respectively. These events form a partition of our sample space. We also define S as the event Uncle Bob scores in this turn at bat. By the law of total probability we have

$$\begin{aligned} P(S) &= P(SE_1) + P(SE_2) + P(SE_3) + P(SN) \\ &= P(S|E_1)P(E_1) + P(S|E_2)P(E_2) + P(S|E_3)P(E_3) + P(S|N)P(N) \\ &= 0.2 \cdot 0.35 + 0.3 \cdot 0.25 + 0.4 \cdot 0.1 + 0 \cdot 0.3 \\ &= 0.185. \end{aligned}$$

2.30. Identical twins have the same gender. We assume that identical twins are equally likely to be boys or girls. Fraternal twins are also equally likely to be boys or girls, but independently of each other. Thus fraternal twins are two girls with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Let I be the event that the twins are identical, F the event that the twins are fraternal.

(a) $P(\text{two girls}) = P(\text{two girls} \,|\, I)P(I) + P(\text{two girls} \,|\, F)P(F) = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{3}$.

(b) $P(I \,|\, \text{two girls}) = \dfrac{P(\text{two girls} \,|\, I)P(I)}{P(\text{two girls})} = \dfrac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{3}} = \dfrac{1}{2}$.

2.31. (a) The sample space is

$$\Omega = \{(g,b),(b,g),(b,b),(g,g)\},$$

and the probability measure is simply

$$P(g,b) = P(b,g) = P(b,b) = P(g,g) = \frac{1}{4},$$

since we assume that each outcome is equally likely.

(b) Let A be the event that there is a girl in the family. Let B be the event that there is a boy in the family. Note that the question is asking for $P(B|A)$. Begin to solve by noting that

$$A = \{(g,b),(b,g),(g,g)\} \text{ and } P(A) = \frac{3}{4}.$$

Similarly,

$$B = \{(g,b),(b,g),(b,b)\} \text{ and } P(B) = \frac{3}{4}.$$

Finally, we have

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(\{(g,b),(b,g)\})}{3/4} = \frac{2/4}{3/4} = \frac{2}{3}.$$

(c) Let $C = \{(g,b),(g,g)\}$ be the event that the first child is a girl. B is as above. We want $P(B|C)$. Since $P(C) = 1/2$ we have

$$P(B|C) = \frac{P(BC)}{P(C)} = \frac{P\{(g,b)\}}{1/2} = \frac{1/4}{1/2} = \frac{1}{2}.$$

2.32. (a) The sample space is

$$\Omega = \{(b,b,b),(b,b,g),(b,g,b),(b,g,g),(g,b,b),(g,b,g),(g,g,b),(g,g,g)\},$$

and each sample point has probability $\frac{1}{8}$ since we assume all outcomes equally likely.

(b) Let $A = \{(b,g,g),(g,b,g),(g,g,b),(g,g,g)\}$ be the event that there are at least two girls in the family. Let

$$B = \{(b,b,b),(b,b,g),(b,g,b),(b,g,g),(g,b,b),(g,b,g),(g,g,b)\}$$

be the event that there is a boy in the family.

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(\{(b,g,g),(g,b,g),(g,g,b)\})}{P\{(b,g,g),(g,b,g),(g,g,b),(g,g,g)\}} = \frac{3/8}{4/8} = \frac{3}{4}.$$

(c) Let $C = \{(g,g,b),(g,g,g)\}$ be the event that the first two children are girls. B is as above. We want $P(B|C)$. We have

$$P(B|C) = \frac{P(BC)}{P(C)} = \frac{P\{(g,g,b)\}}{P\{(g,g,b),(g,g,g)\}} = \frac{1}{2}.$$

2.33. (a) Let B_k be the event that we choose urn k and let A be the event that we chose a red ball. Then

$$P(B_k) = \tfrac{1}{5}, \qquad P(A|B_k) = \tfrac{k}{10}, \qquad \text{for } 1 \le k \le 5.$$

By conditioning on the urn we chose and using (2.7) we get

$$P(A) = \sum_{k=1}^{5} P(A \mid B_k)P(B_k) = \sum_{k=1}^{5} \tfrac{k}{10} \cdot \tfrac{1}{5} = \tfrac{1+2+3+4+5}{50} = \tfrac{3}{10}.$$

(b)

$$P(B_k \mid A) = \frac{P(A|B_k)P(B_k)}{\sum_{k=1}^{5} P(A \mid B_k)P(B_k)} = \frac{\frac{k}{10} \cdot \frac{1}{5}}{\frac{3}{10}} = \frac{k}{15}.$$

2.34. Since the urns are interchangeable, we can put the marked ball in urn 1. There are three ways to arrange the two unmarked balls. Let case i for $i \in \{0, 1, 2\}$ denote the situation where we put i unmarked balls together with the marked ball, and the remaining $2 - i$ unmarked balls in the other urn. Let M denote the event that your friend draws the marked ball, and A_j the event that she chooses urn j, $j = 1, 2$. Since $P(M|A_2) = 0$, we get the following probabilities.

Case 0: $P(M) = P(M|A_1)P(A_1) = 1 \cdot \frac{1}{2} = \frac{1}{2}$.

Case 1: $P(M) = P(M|A_1)P(A_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Case 2: $P(M) = P(M|A_1)P(A_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$.

So (a) you would put all the balls in one urn (Case 2) while (b) she would put the marked ball in one urn and the other balls in the other urn.

(c) The situation is analogous. If we put k unmarked balls together with the marked ball in urn 1, then

$$P(M) = P(M|A_1)P(A_1) = \tfrac{1}{k+1} \cdot \tfrac{1}{2} = \tfrac{1}{2(k+1)}.$$

Hence to minimize the chances of drawing the marked ball, put all the balls in one urn, and to maximize the chances of drawing the marked ball, put the marked ball in one urn and all the unmarked balls in the other.

2.35. Let A be the event that the first card is a queen and B the event that the second card is a spade. Note that A and B are not independent, and there is no immediate way to compute $P(B|A)$. We can compute $P(AB)$ by counting favorable outcomes. Let Ω be the collection of all ordered pairs drawn without replacement from 52 cards. $\#\Omega = 52 \cdot 51$ and all outcomes are equally likely. We can break up AB into the union of the following two disjoint events:

$C = \{$first card is queen of spades, second is a spade$\}$,

$D = \{$first card is a queen but not a spade, the second card is a spade$\}$.

We have $\#C = 12$, as we can choose the second card 12 different ways. We have $\#D = 3 \cdot 13 = 39$ as the first card can be any of the three non-spade queens, and the second card can be any of the 13 spades. Thus $\#AB = \#C + \#D = 12 + 39 = 51$ and we get $P(AB) = \frac{\#AB}{\#\Omega} = \frac{51}{52 \cdot 51} = \frac{1}{52}$.

2.36. Let A_j be the event that a j-sided die was chosen and B the event that a six was rolled.

(a) By the law of total probability,

$$\begin{aligned} P(B) &= P(B|A_4)P(A_4) + P(B|A_6)P(A_6) + P(B|A_{12})P(A_{12}) \\ &= 0 \cdot \tfrac{7}{12} + \tfrac{1}{6} \cdot \tfrac{3}{12} + \tfrac{1}{12} \cdot \tfrac{2}{12} = \tfrac{1}{18}. \end{aligned}$$

(b)

$$P(A_6|B) = \frac{P(B|A_6)P(A_6)}{P(B)} = \frac{\frac{1}{6} \cdot \frac{3}{12}}{\frac{1}{18}} = \tfrac{3}{4}.$$

2.37. (a) Let S, E, T, and W be the events that the six, eight, ten, and twenty sided die is chosen. Let X be the outcome of the roll. Then

$$\begin{aligned} P(X = 6) &= P(X = 6|S)P(S) + P(X = 6|E)P(E) \\ &\quad + P(X = 6|T)P(T) + P(X = 6|W)P(W) \\ &= \frac{1}{6} \cdot \frac{1}{10} + \frac{1}{8} \cdot \frac{2}{10} + \frac{1}{10} \cdot \frac{3}{10} + \frac{1}{20} \cdot \frac{4}{10} \\ &= \frac{11}{120}. \end{aligned}$$

(b) We want

$$P(W|X = 7) = \frac{P(W, X = 7)}{P(X = 7)} = \frac{P(X = 7|W)P(W)}{P(X = 7)}.$$

Following part (a), we have

$$\begin{aligned} P(X = 7) &= P(X = 7|S)P(S) + P(X = 7|E)P(E) \\ &\quad + P(X = 7|T)P(T) + P(X = 7|W)P(W) \\ &= 0 \cdot \frac{1}{10} + \frac{1}{8} \cdot \frac{2}{10} + \frac{1}{10} \cdot \frac{3}{10} + \frac{1}{20} \cdot \frac{4}{10} = \frac{3}{40}. \end{aligned}$$

Thus,

$$P(W|X = 7) = \frac{(1/20) \cdot (4/10)}{(3/40)} = \frac{4}{15}.$$

2.38. Let R denote the event that the chosen letter is `R` and let A_i be the event that the ith word of the sentence is chosen.

(a) $P(R) = \sum_{i=1}^{4} P(R|A_i)P(A_i) = 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{5} \cdot \frac{1}{4} = \frac{2}{15}$.

(b) $P(X = 3) = \frac{1}{4}$, $P(X = 4) = \frac{1}{2}$, $P(X = 5) = \frac{1}{4}$.

(c) $P(X = 3 \,|\, X > 3) = 0$.

$$P(X = 4 \,|\, X > 3) = \frac{P(\{X = 4\} \cap \{X > 3\})}{P(X > 3)} = \frac{P(X = 4)}{P(X = 4) + P(X = 5)} = \frac{\left(\frac{1}{2}\right)}{\left(\frac{3}{4}\right)} = \frac{2}{3}.$$

$$P(X = 5 \,|\, X > 3) = \frac{P(\{X = 5\} \cap \{X > 3\})}{P(X > 3)} = \frac{\left(\frac{1}{4}\right)}{\left(\frac{3}{4}\right)} = \frac{1}{3}.$$

(d) Use below that $R \cap A_1 = R \cap A_2 = A_3 \cap \{X > 3\} = \varnothing$.

$$\begin{aligned}
P(R \,|\, X > 3) &= \sum_{i=1}^{4} P(RA_i|X > 3) = P(RA_3|X > 3) + P(RA_4|X > 3)\\
&= \frac{P(R \cap A_3 \cap \{X > 3\})}{P(X > 3)} + \frac{P(R \cap A_4 \cap \{X > 3\})}{P(X > 3)}\\
&= \frac{P(R \cap A_4)}{P(X = 4) + P(X = 5)} = \frac{P(R \,|\, A_4)P(A_4)}{P(X = 4) + P(X = 5)}\\
&= \frac{\frac{1}{5} \cdot \frac{1}{4}}{\frac{1}{2} + \frac{1}{4}} = \tfrac{1}{15}.
\end{aligned}$$

(e)

$$P(A_4 \,|\, R) = \frac{P(R \,|\, A_4)P(A_4)}{P(R)} = \frac{\frac{1}{5} \cdot \frac{1}{4}}{\frac{2}{15}} = \tfrac{3}{8}.$$

2.39. (a) Let B_i the event that we chose the ith word ($i = 1, \dots, 8$). Events $B_1, \dots, B_8$ form a partition of the sample space and $P(B_i) = \frac{1}{8}$ for each i. Let A be the event that we chose the letter `O`. Then $P(A|B_3) = \frac{1}{5}$, $P(A|B_4) = \frac{1}{3}$, $P(A|B_6) = \frac{1}{4}$ with all other $P(A|B_i) = 0$. This gives

$$P(A) = \sum_{i=1}^{8} P(A|B_i)P(B_i) = \frac{1}{8}\left(\frac{1}{5} + \frac{1}{3} + \frac{1}{4}\right) = \frac{47}{480}.$$

(b) The length of the chosen word can be 3, 4, 5 or 6, so the range of X is the set $\{3, 4, 5, 6\}$. For each of the possible value x we have to find the probability $P(X = x)$.

$$\begin{aligned}
p_X(3) &= P(X = 3) = P(\text{we chose the 1st, the 4th or the 7th word})\\
&= P(B_1 \cup B_4 \cup B_7) = \frac{3}{8},\\
p_X(4) &= P(X = 4) = P(\text{we chose the 6th or the 8th word}) = P(B_6 \cup B_8) = \frac{2}{8},\\
p_X(5) &= P(X = 5) = P(\text{we chose the 2nd or the 3rd word}) = P(B_2 \cup B_3) = \frac{2}{8},\\
p_X(6) &= P(X = 6) = P(\text{we chose the 5th word}) = P(B_5) = \frac{1}{8}.
\end{aligned}$$

Note that the probabilities add up to 1, as they should.

2.40. (a) For $i \in \{1, 2, 3, 4\}$ let A_i be the event that the student scores i on the test. Let M be the event that the student becomes a math major.

$$P(M) = \sum_{i=1}^{4} P(M|A_i)P(A_i) = 0 \cdot 0.1 + \tfrac{1}{5} \cdot 0.2 + \tfrac{1}{3} \cdot 0.6 + \tfrac{3}{7} \cdot 0.1 \approx 0.2829.$$

(b)

$$P(A_4|M) = \frac{P(M|A_4)P(A_4)}{P(M)} = \frac{\frac{3}{7} \cdot 0.1}{\frac{1}{5} \cdot 0.2 + \frac{1}{3} \cdot 0.6 + \frac{3}{7} \cdot 0.1} \approx 0.1515.$$

2.41. Introduce the following events:

$$B = \{\text{the phone is not defective}\}, \quad A = \{\text{the phone comes from factory II}\}.$$

Then A^c is the event that the phone is from factory I. We know that

$$P(A) = 0.4 = \frac{2}{5}, \quad P(A^c) = 0.6 = \frac{3}{5}, \quad P(B^c|A) = 0.2 = \frac{1}{5}, \quad P(B^c|A^c) = 0.1 = \frac{1}{10}.$$

Note that this also gives

$$P(B|A) = 1 - P(B^c|A) = \frac{4}{5}, \qquad P(B|A^c) = 1 - P(B^c|A^c) = \frac{9}{10}.$$

We need to compute $P(A|B)$. By Bayes' formula,

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{\frac{4}{5} \cdot \frac{2}{5}}{\frac{4}{5} \cdot \frac{2}{5} + \frac{9}{10} \cdot \frac{3}{5}} = \frac{16}{16+27} \\ &= \frac{16}{43} \approx 0.3721. \end{aligned}$$

2.42. Let R be the event that the transferred ball was red, and W the event that the transferred ball was white. Let V be the event that a white ball was drawn from urn B. Then $P(R) = \frac{1}{3}$ and $P(W) = \frac{2}{3}$. If a red ball was transferred, then the new composition of urn B is 2 red and 1 white, while if a white ball was transferred, then the new composition of urn B is 1 red and 2 white. Putting all this together gives the following calculation.

$$\begin{aligned} P(W|V) &= \frac{P(WV)}{P(V)} = \frac{P(V|W)P(W)}{P(V|W)P(W) + P(V|R)P(R)} \\ &= \frac{\frac{2}{3} \cdot \frac{2}{3}}{\frac{2}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3}} = \frac{4}{5}. \end{aligned}$$

2.43. (a) Let A_1 be the event that the first sample had two balls of the same color. If we imagine that the draws are done one at a time in order then there are $5 \cdot 4$ possible outcomes. Counting the green-green and yellow-yellow cases separately we get that $3 \cdot 2 + 2 \cdot 1$ of those outcomes have two balls of the same color. Thus

$$P(A_1) = \frac{3 \cdot 2 + 2 \cdot 1}{5 \cdot 4} = \frac{2}{5}.$$

(b) Let A_2 be the event that the second sample had two balls of the same color. We have $P(A_2|A_1) = 1$, since if the first sample had two balls of the same color then this must be true for the second one. Furthermore, $P(A_2|A_1^c) = \frac{1}{2}$, because if we sample twice with replacement from an urn containing one yellow and one green ball, then $1/2$ is the probability that the second draw has the same color as the first one. (Or, dividing the number of favorable outcomes by the total, $\frac{1 \cdot 1 + 1 \cdot 1}{2 \cdot 2} = \frac{1}{2}$.) From part (a) we know that $P(A_1) = \frac{2}{5}$ and $P(A_1^c) = \frac{3}{5}$. Altogether this gives

$$P(A_2) = P(A_2|A_1)P(A_1) + P(A_2|A_1^c)P(A_1^c) = 1 \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} = \frac{7}{10}.$$

(c) Using the already computed probabilities:

$$P(A_1|A_2) = \frac{P(A_2|A_1)P(A_1)}{P(A_2)} = \frac{1 \cdot \frac{2}{5}}{\frac{7}{10}} = \frac{4}{7}.$$

2.44. Let A_i be the event that bin i was chosen ($i = 1, 2$) and Y_j the event that draw j ($j = 1, 2$) is yellow.

(a)

$$\begin{aligned} P(A_1|Y_1) &= \frac{P(Y_1|A_1)P(A_1)}{P(Y_1|A_1)P(A_1) + P(Y_1|A_2)P(A_2)} \\ &= \frac{\frac{4}{10}\cdot\frac{1}{2}}{\frac{4}{10}\cdot\frac{1}{2} + \frac{4}{7}\cdot\frac{1}{2}} = \frac{14}{34} \approx 0.4118. \end{aligned}$$

(b) This question asks for the conditional probability of A_1, given that two draws with replacement from the chosen urn yield yellow. We assume that draws with replacement from the same urn are independent. This translates into conditional independence of Y_1 and Y_2, given A_i.

$$\begin{aligned} P(A_1|Y_1Y_2) &= \frac{P(Y_1Y_2|A_1)P(A_1)}{P(Y_1Y_2|A_1)P(A_1) + P(Y_1Y_2|A_2)P(A_2)} \\ &= \frac{P(Y_1|A_1)P(Y_1|A_1)P(A_1)}{P(Y_1|A_1)P(Y_1|A_1)P(A_1) + P(Y_1|A_2)P(Y_1|A_2)P(A_2)} \\ &= \frac{\frac{4}{10}\cdot\frac{4}{10}\cdot\frac{1}{2}}{\frac{4}{10}\cdot\frac{4}{10}\cdot\frac{1}{2} + \frac{4}{7}\cdot\frac{4}{7}\cdot\frac{1}{2}} = \frac{196}{596} \approx 0.3289. \end{aligned}$$

2.45. (a) Let B, G, and O be the events that a 7-year-old like the Bears, Packers, and some other team, respectively. We are given the following:

$$P(B) = 0.10, \quad P(G) = 0.75, \quad P(O) = 0.15.$$

Let A be the event that the 7-year-old goes to a game. Then we have

$$P(A|B) = 0.01, \quad P(A|G) = 0.05, \quad P(A|O) = 0.005.$$

$P(A)$ is computed from the law of total probability:

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|G)P(G) + P(A|O)P(O) \\ &= 0.01\cdot 0.1 + 0.05\cdot 0.75 + 0.005\cdot 0.15 = 0.03925. \end{aligned}$$

(b) Using the result of (a) (or Bayes' formula directly):

$$P(G|A) = \frac{P(AG)}{P(A)} = \frac{P(A|G)P(G)}{P(A)} = \frac{0.05\cdot 0.75}{0.03925} = \frac{0.0375}{0.03925} \approx 0.9554.$$

2.46. A sample point is an ordered triple (x, y, z) where x is the number drawn from box A, y is the number drawn from box B, and z the number drawn from box C. All $6 \cdot 12 \cdot 4 = 288$ outcomes are equally likely, so we can solve these problems by counting.

(a) The number of outcomes with exactly two 1s is

$$1\cdot 1\cdot 3 + 1\cdot 11\cdot 1 + 5\cdot 1\cdot 1 = 19.$$

The number of outcomes with a 1 from box A and exactly two 1s is

$$1\cdot 1\cdot 3 + 1\cdot 11\cdot 1 = 14.$$

Thus

$$P(\text{ball 1 from } A \,|\, \text{exactly two 1s}) = \frac{P(\text{ball 1 from } A \text{ and exactly two 1s})}{P(\text{exactly two 1s})}$$
$$= \frac{14/288}{19/288} = \frac{14}{19}.$$

(b) There are three sample points whose sum is 21: $(6, 12, 3), (6, 11, 4), (5, 12, 4)$. Two of these have 12 drawn from B. Hence the answer is $2/3$. Here is the formal calculation.

$$P(\text{ball 12 from } B \,|\, \text{sum of balls 21}) = \frac{P(\text{ball 12 from } B \text{ and sum of balls 21})}{P(\text{sum of balls 21})}$$
$$= \frac{P\{(6, 12, 3), (5, 12, 4)\}}{P\{(6, 12, 3), (6, 11, 4), (5, 12, 4)\}} = \frac{2/288}{3/288} = \frac{2}{3}.$$

2.47. Define random variables X and Y and event S:

$X =$ total number of patients for whom the drug is effective

$Y =$ number of patients for whom the drug is effective, excluding your friends

$S =$ trial is a success for your two friends.

We need to find

$$P(S|X = 55) = \frac{P(S \cap \{X = 55\})}{P(X = 55)}.$$

Note that $X \sim \text{Bin}(80, p)$, and thus $P(X = 55) = \binom{80}{55} p^{55}(1 - p)^{25}$. Moreover, $S \cap \{X = 55\} = S \cap \{Y = 53\}$. The events S and $\{Y = 53\}$ are independent, as S depends on the trial outcomes for your friends, and Y on the trial outcomes of the other patients. Thus

$$P(S \cap \{X = 55\}) = P(S \cap \{Y = 53\}) = P(S)P(Y = 53).$$

We have $P(S) = p^2$ and $P(Y = 53) = \binom{78}{53} p^{53}(1-p)^{25}$, as $Y \sim \text{Bin}(78, p)$. Collecting everything:

$$P(S|X = 55) = \frac{P(S \cap \{X = 55\})}{P(X = 55)} = \frac{p^2 \cdot \binom{78}{53} p^{53}(1 - p)^{25}}{\binom{80}{55} p^{55}(1 - p)^{25}} = \frac{\binom{78}{53}}{\binom{80}{55}}$$
$$= \frac{297}{632} \approx 0.4699.$$

2.48. Define events $G = \{\text{Kevin is guilty}\}$, $A = \{\text{DNA match}\}$. Before the DNA evidence $P(G) = 1/100,000$. After the DNA match

$$P(G|A) = \frac{P(A|G)P(G)}{P(A|G)P(G) + P(A|G^c)P(G^c)} = \frac{1 \cdot \frac{1}{100{,}000}}{1 \cdot \frac{1}{100{,}000} + \frac{1}{10{,}000} \cdot \frac{99{,}999}{100{,}000}}$$
$$= \frac{1}{1 + 10 - 10^{-4}} \approx \frac{1}{11}.$$

2.49. (a) The given numbers are nonnegative, so we just need to check that $\sum_{k=0}^{\infty} P(X = k) = 1$:

$$\sum_{k=0}^{\infty} P(X = k) = \frac{4}{5} + \sum_{k=1}^{\infty} \tfrac{1}{10} \cdot \left(\tfrac{2}{3}\right)^k = \frac{4}{5} + \frac{\frac{1}{10} \cdot \frac{2}{3}}{1 - \frac{2}{3}} = 1.$$

(b) For $k \geq 1$, by changing the summation index from j to $i = j - k$:

$$P(X \geq k) = \sum_{j=k}^{\infty} \tfrac{1}{10} \cdot \left(\tfrac{2}{3}\right)^j = \tfrac{1}{10} \cdot \left(\tfrac{2}{3}\right)^k \sum_{i=0}^{\infty} \left(\tfrac{2}{3}\right)^i = \tfrac{1}{10} \cdot \left(\tfrac{2}{3}\right)^k \frac{1}{1-\frac{2}{3}} = \tfrac{1}{5} \left(\tfrac{2}{3}\right)^{k-1}.$$

Thus again for $k \geq 1$,

$$\begin{aligned} P(X \geq k \,|\, X \geq 1) &= \frac{P(\{X \geq k\} \cap \{X \geq 1\})}{P(X \geq 1)} = \frac{P(X \geq k)}{P(X \geq 1)} \\ &= \frac{\frac{1}{5}\left(\frac{2}{3}\right)^{k-1}}{\frac{1}{5}} = \left(\tfrac{2}{3}\right)^{k-1}. \end{aligned}$$

The numerator simplified because $\{X \geq k\} \subset \{X \geq 1\}$. The answer shows that conditional on $X \geq 1$, X has Geom$(\frac{1}{3})$ distribution.

2.50. (a)

$$\begin{aligned} P(A|D) &= \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)} \\ &= \frac{p \cdot \frac{1}{3}}{p \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{p}{1+p}. \end{aligned}$$

(b)

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} = \frac{1 \cdot \frac{1}{3}}{(p+1) \cdot \frac{1}{3}} = \frac{1}{1+p}.$$

If the guard is equally likely to name either B or C when both of them are slated to die, then A has not gained anything (his probability of pardon is still $\frac{1}{3}$) but C's chances of pardon have increased to $\frac{2}{3}$. In the extreme case where the guard would never name B unless he had to ($p = 0$), C is now sure to be pardoned.

2.51. Since $C \subset B$ we have $B \cup C = B$ and thus $A \cup B \cup C = A \cup B$. Then

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

Since A and B are independent we have $P(AB) = P(A)P(B)$. This gives

$$P(A \cup B \cup C) = P(A) + P(B) - P(A)P(B) = 1/2 + 1/4 - 1/8 = 5/8.$$

2.52. Yes, A, B, and C are mutually independent. There are four equations to check:

(i) $P(AB) = P(A)P(B)$
(ii) $P(AC) = P(A)P(C)$
(iii) $P(BC) = P(B)P(C)$
(iv) $P(ABC) = P(A)P(B)P(C)$.

(i) comes from inclusion-exclusion:

$$P(AB) = P(A) + P(B) - P(A \cup B) = 0.06 = P(A)P(B).$$

(ii) comes from $P(AC) = P(C) - P(A^cC) = 0.03 = P(A)P(C)$. (iii) is given. Finally, (iv) comes from using inclusion-exclusion once more and the previous computations:

$$\begin{aligned}P(ABC) &= P(A \cup B \cup C) - P(A) - P(B) - P(C) \\ &\qquad + P(AB) + P(AC) + P(BC) \\ &= 0.006 = P(A)P(B)P(C).\end{aligned}$$

2.53. (a) If the events are disjoint then

$$P(A \cup B) = P(A) + P(B) = 0.3 + 0.6 = 0.9.$$

(b) If the events are independent then

$$\begin{aligned}P(A \cup B) &= P(A) + P(B) - P(AB) = P(A) + P(B) - P(A)P(B) \\ &= 0.3 + 0.6 - 0.3 \cdot 0.6 = 0.72.\end{aligned}$$

2.54. (a) It is possible. We use the fact that $A = AB \cup AB^c$ and that these are mutually exclusive:

$$\begin{aligned}P(A) &= P(AB) + P(AB^c) = P(A|B)P(B) + P(A|B^c)P(B^c) \\ &= \frac{1}{3}P(B) + \frac{1}{3}P(B^c) = \frac{1}{3}(P(B) + P(B^c)) = \frac{1}{3}.\end{aligned}$$

(b) A and B are independent. By part (a) and the given information,

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

from which $P(AB) = P(A)P(B)$ and independence has been verified. (Note that the value $\frac{1}{3}$ was not needed for this conclusion.)

2.55. (a) Since Peter throws the first dart, in order for Mary to win Peter must fail once more than she does.

$$\begin{aligned}P(\text{Mary wins}) &= \sum_{k=1}^{\infty} P(\text{Mary wins on her } k\text{th throw}) \\ &= \sum_{k=1}^{\infty} ((1-p)(1-r))^{k-1}(1-p)r = \frac{(1-p)r}{1-(1-p)(1-r)} \\ &= \frac{(1-p)r}{p+r-pr}.\end{aligned}$$

(b) The possible values of X are the nonnegative integers.

$$P(X = 0) = P(\text{Peter wins on his first throw}) = p.$$

For $k \geq 1$,

$$\begin{aligned}P(X = k) &= P(\text{Mary wins on her } k\text{th throw}) \\ &\qquad + P(\text{Peter wins on his } (k+1)\text{st throw}) \\ &= ((1-p)(1-r))^{k-1}(1-p)r + ((1-p)(1-r))^k p \\ &= ((1-p)(1-r))^{k-1}(1-p)(p+r-pr).\end{aligned}$$

We check that the values for $k \geq 1$ add up to $1-$(the value at $k = 0$):

$$\sum_{k=1}^{\infty}((1-p)(1-r))^{k-1}(1-p)(p+r-pr) = \frac{(1-p)(p+r-pr)}{1-(1-p)(1-r)} = 1-p.$$

This is not one of our named distributions.

(c) For $k \geq 1$,

$$\begin{aligned} P(X = k \mid \text{Mary wins}) &= \frac{P(\text{Mary wins on her } k\text{th throw})}{P(\text{Mary wins})} \\ &= \frac{((1-p)(1-r))^{k-1}(1-p)r}{\frac{(1-p)r}{p+r-pr}} \\ &= ((1-p)(1-r))^{k-1}(p+r-pr). \end{aligned}$$

Thus given that Mary wins, $X \sim \text{Geom}(p+r-pr)$.

2.56. Suppose $P(A) = 0$. Then for any B, $AB \subset A$ implies $P(AB) = 0$. We also have $P(A)P(B) = 0 \cdot P(B) = 0$. Thus $P(AB) = 0 = P(A)P(B)$ and independence of A and B has been verified.

Suppose $P(A) = 1$. Then $P(A^c) = 0$ and the previous case gives the independence of A^c and B, from which follows the independence of A and B. Alternatively, we can prove this case by first observing that $P(AB) = P(B) - P(A^cB) = P(B) - 0 = P(B)$ and then $P(A)P(B) = 1 \cdot P(B) = P(B)$. Again $P(AB) = P(A)P(B)$ has been verified.

2.57. (a) Let E_1 be the event that the first component functions. Let E_2 be the event that the second component functions. Let S be the event that the entire system functions. $S = E_1 \cap E_2$ since both components must function in order for the whole system to be operational. By the assumption that each component acts independently, we have

$$P(S) = P(E_1 \cap E_2) = P(E_1)P(E_2).$$

Next we find the probabilities $P(E_1)$ and $P(E_2)$.

Let X_i be a Bernoulli random variable taking the value 1 if the ith element of the first component is working. The information given is that $P(X_i = 1) = 0.95$, $P(X_i = 0) = 0.05$ and $X_1, \ldots, X_8$ are mutually independent. Similarly, let Y_i be a Bernoulli random variable taking the value 1 if the ith element of the second component is working. Then $P(Y_i = 1) = 0.90$, $P(Y_i = 0) = 0.1$ and $Y_1, \ldots, Y_4$ are mutually independent. Let $X = \sum_{i=1}^{8} X_i$ give the total number of working elements in component number one and $Y = \sum_{i=1}^{4} Y_i$ the total number of working elements in component number 2. Then $X \sim \text{Bin}(8, 0.95)$ and $Y \sim \text{Bin}(4, 0.90)$, and X and Y are independent (by the assumption that the components behave independently). We have

$$\begin{aligned}P(E_1) &= P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8)\\ &= \binom{8}{6}(0.95)^6(0.05)^2 + \binom{8}{7}(0.95)^7(0.05)^1 + \binom{8}{8}(0.95)^8(0.05)^0\\ &= 0.9942117,\end{aligned}$$

and

$$\begin{aligned}P(E_2) &= P(Y \geq 3) = P(Y = 3) + P(Y = 4)\\ &= \binom{4}{3}(0.9)^3(0.1) + (0.9)^4\\ &= 0.9477.\end{aligned}$$

Thus,

$$P(S) = P(E_1)P(E_2) = 0.9942117 \cdot 0.9477 \approx 0.9422.$$

(b) We look for $P(E_2^c \mid S^c)$. We have

$$P(E_2^c|S^c) = \frac{P(E_2^c S^c)}{P(S^c)} = \frac{P(E_2^c)}{1 - P(S)},$$

where we used that $E_2^c \subset S^c$. (If the first component does not work, then the system does not work; mathematically a consequence of de Morgan's law: $S^c = E_1^c \cup E_2^c$.) Thus,

$$P(E_2^c|S^c) = \frac{1 - P(E_2)}{1 - P(S)} = \frac{1 - 0.9477}{1 - 0.9422} \approx 0.9048.$$

2.58. (a) It is enough to show that any two of them are pairwise independent since the argument is the same for any such pair. We show that $P(AB) = P(A)P(B)$. Let

$$\Omega = \{(a, b, c) : a, b, c \in \{1, 2, \ldots, 365\}\} \implies \#\Omega = 365^3.$$

We have by counting the possibilities

$$\#AB = \{\text{all three have same birthday}\} = 365 \cdot 1 \cdot 1 \implies P(AB) = \frac{1}{365^2}.$$

Also,

$$\#A = \{\text{Alex and Betty have the same birthday}\} = 365 \cdot 1 \cdot 365,$$

where we counted as follows: there are 365 ways for Alex to have a birthday, then only once choice for Betty, and then another 365 ways for Conlin. Thus,

$$P(A) = \frac{365^2}{365^3} = \frac{1}{365}.$$

Similarly, $P(B) = \frac{1}{365}$ and so,

$$P(AB) = P(A)P(B).$$

(b) The events are not independent. Note that $ABC = AB$ and so,

$$P(ABC) = P(AB) = \frac{1}{365^2} \neq P(A)P(B)P(C) = \frac{1}{365^3}.$$

2.59. Define events: $B = \{\text{the bus functions}\}$, $T = \{\text{the train functions}\}$, and $S = \{\text{no storm}\}$. The event that travel is possible is $(B \cup T) \cap S = BS \cup TS$. We calculate the probability with inclusion-exclusion and independence:

$$\begin{aligned}
P(BS \cup TS) &= P(BS) + P(TS) - P(BTS) \\
&= P(B)P(S) + P(T)P(S) - P(B)P(T)P(S) \\
&= \tfrac{8}{10} \cdot \tfrac{19}{20} + \tfrac{9}{10} \cdot \tfrac{19}{20} - \tfrac{8}{10} \cdot \tfrac{9}{10} \cdot \tfrac{19}{20} = \tfrac{931}{1000}.
\end{aligned}$$

2.60. (a) $P(AB^c) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)\big(1 - P(B)\big) = P(A)P(B^c)$.

(b) Apply first de Morgan and then inclusion-exclusion:

$$\begin{aligned}
P(A^cC^c) &= 1 - P(A \cup C) = 1 - P(A) - P(C) + P(AC) \\
&= 1 - P(A) - P(C) + P(A)P(C) \\
&= \big(1 - P(A)\big)\big(1 - P(C)\big) = P(A^c)P(C^c).
\end{aligned}$$

(c) $P(AB^cC) = P(AC) - P(ABC) = P(A)P(C) - P(A)P(B)P(C) = P(A)\big(1 - P(B)\big)P(C = P(A)P(B^c)P(C)$.

(d) Again first de Morgan and then inclusion-exclusion:

$$\begin{aligned}
&P(A^cB^cC^c) = 1 - P(A \cup B \cup C) \\
&= 1 - P(A) - P(B) - P(C) + P(AB) + P(AC) + P(BC) - P(ABC) \\
&= 1 - P(A) - P(B) - P(C) + P(A)P(B) + P(A)P(C) + P(B)P(C) \\
&\qquad\qquad - P(A)P(B)P(C) \\
&= \big(1 - P(A)\big)\big(1 - P(B)\big)\big(1 - P(C)\big) \\
&= P(A^c)P(B^c)P(C^c).
\end{aligned}$$

2.61. (a) Treat each draw as a trial: green is success, red is failure. By counting favorable outcomes, the probability of success is $p = \frac{3}{7}$ for each draw. Because we draw with replacement the outcomes are independent. Thus the number of greens in the 9 picks is the number of successes in 9 trials, hence a $\text{Bin}(9, \frac{3}{7})$ distribution. Using the probability mass function of the binomial distribution gives

$$\begin{aligned}
P(X \geq 1) &= 1 - P(X = 0) = 1 - (1 - p)^9 \approx 0.9935, \\
P(X \leq 5) &= \sum_{k=0}^{5} P(X = k) = \sum_{k=0}^{5} \binom{9}{k} p^k (1 - p)^{9-k} \approx 0.8653.
\end{aligned}$$

(b) N is the number of trials needed for the first success, and so has geometric distribution with parameter $p = \frac{3}{7}$. The probability mass function of the geometric distribution gives

$$P(N \leq 9) = \sum_{k=1}^{9} P(N = k) = \sum_{k=1}^{9} p(1 - p)^{k-1} \approx 0.9935.$$

(c) We have $P(X \geq 1) = P(N \leq 9)$. We can check this by using the geometric sum formula to get

$$\sum_{k=1}^{9} p(1-p)^{k-1} = p\frac{1-(1-p)^9)}{1-(1-p)} = 1-(1-p)^9.$$

Here is another way to see this, without any algebra. Imagine that we draw balls with replacement infinitely many times. Think of X as the number of green balls in the first 9 draws. N is still the number of draws needed for the first green. Now if $X \geq 1$, then we have at least one green within the first 9 draws, which means that the first green draw happened within the first 9 draws. Thus $X \geq 1$ implies $N \leq 9$. But this works in the opposite direction as well: if $N \leq 9$ then the first green draw happened within the first 9 draws, which means that we must have at least one green within the first 9 picks. Thus $N \leq 9$ implies $X \geq 1$. This gives the equality of event: $\{X \geq 1\} = \{N \leq 9\}$, and hence the probabilities must agree as well.

2.62. Regard the drawing of three marbles as one trial, with success probability p given by

$$p = P(\text{all three marbles blue}) = \frac{\binom{9}{3}}{\binom{13}{3}} = \frac{7 \cdot 8 \cdot 9 \cdot 10}{10 \cdot 11 \cdot 12 \cdot 13} = \frac{42}{143}.$$

$X \sim \text{Bin}(20, \frac{42}{143})$. The probability mass function is

$$P(X = k) = \binom{20}{k}\left(\tfrac{42}{143}\right)^k\left(\tfrac{101}{143}\right)^{20-k} \quad \text{for } k = 0, 1, 2, \dots, 20.$$

2.63. The number of heads in n coin flips has distribution $\text{Bin}(n, 1/2)$. Thus the probability of winning if we choose to flip n times is

$$f_n = P(n \text{ flips yield exactly 2 heads}) = \binom{n}{2}\frac{1}{2^n} = \frac{n(n-1)}{2^{n+1}}.$$

We want to find the n which maximizes f_n. Let us compare f_n and f_{n+1}. We have

$$f_n < f_{n+1} \iff \frac{n(n-1)}{2^{n+1}} < \frac{(n+1)n}{2^{n+2}} \iff 2(n-1) < n+1 \iff n < 3.$$

Similarly, $f_n > f_{n+1}$ if and only if $n > 3$, and $f_3 = f_4$. Thus

$$f_2 < f_3 = f_4 > f_5 > f_6 > \dots.$$

This means that the maximum happens at $n = 3$ and $n = 4$, and the probability of winning at those values is $f_3 = f_4 = \frac{3 \cdot 2}{2^4} = \frac{3}{8}$.

2.64. Let X be the number of correct answers. X is the number of successes in 20 independent trials with success probability $p + \frac{1}{2}r$.

$$P(X \geq 19) = P(X = 19) + P(X = 20) = 20\left(p + \tfrac{1}{2}r\right)^{19}\left(q + \tfrac{1}{2}r\right) + \left(p + \tfrac{1}{2}r\right)^{20}.$$

2.65. Let A be the event that at least one die lands on a 4 and B be the event that all three dice land on different numbers. Our sample space is the set of all triples (a_1, a_2, a_3) with $1 \leq a_i \leq 6$. All outcomes are equally likely and there are 216 outcomes. We need $P(A|B) = \frac{P(AB)}{P(B)}$. There are $6 \cdot 5 \cdot 4 = 120$ elements in B. To count the elements of AB, we first consider A^cB. This is the set of triples where the three numbers are distinct and none of them is a 4. So $\#A^cB = 5 \cdot 4 \cdot 3 = 60$. Then $\#AB = \#B - \#A^cB = 120 - 60 = 60$ and

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{60}{216}}{\frac{120}{216}} = \frac{1}{2}.$$

2.66. Let

$$f_n = P(n \text{ die rolls give exactly two sixes}) = \binom{n}{2}\left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right)^{n-2} = \frac{n(n-1)5^{n-2}}{2 \cdot 6^n}.$$

Next,

$$\begin{aligned} f_n < f_{n+1} \quad &\Longleftrightarrow \quad \frac{n(n-1)5^{n-2}}{2 \cdot 6^n} < \frac{(n+1)n5^{n-1}}{2 \cdot 6^{n+1}} \quad \Longleftrightarrow \quad 6(n-1) < 5(n+1) \\ &\Longleftrightarrow \quad n < 11. \end{aligned}$$

By reversing the inequalities we get the equivalence

$$f_n > f_{n+1} \quad \Longleftrightarrow \quad n > 11.$$

By complementing the two equivalences, we get

$$\begin{aligned} f_n = f_{n+1} \quad &\Longleftrightarrow \quad f_n \geq f_{n+1} \text{ and } f_n \leq f_{n+1} \\ &\Longleftrightarrow \quad n \geq 11 \text{ and } n \leq 11 \quad \Longleftrightarrow \quad n = 11. \end{aligned}$$

Putting all these facts together we conclude that the probability of two sixes is maximized by $n = 11$ and $n = 12$ and for these two values of n, that probability is

$$\frac{11 \cdot 10 \cdot 5^9}{2 \cdot 6^{11}} \approx 0.2961.$$

2.67. Since $\{X = n+k\} \subset \{X > n\}$ for $k \geq 1$, we have

$$P(X = n+k|X > n) = \frac{P(X = n+k, X > n)}{P(X > n)} = \frac{P(X = n+k)}{P(X > n)} = \frac{(1-p)^{n+k-1}p}{P(X > n)}.$$

Evaluate the denominator:

$$\begin{aligned} P(X > n) &= \sum_{k=n+1}^{\infty} P(X = k) = \sum_{k=n+1}^{\infty} (1-p)^{k-1}p \\ &= p(1-p)^n \sum_{k=0}^{\infty} (1-p)^k = p(1-p)^n \cdot \frac{1}{1-(1-p)} = (1-p)^n. \end{aligned}$$

Thus,

$$\begin{aligned} P(X = n+k|X > n) &= \frac{(1-p)^{n+k-1}p}{P(X > n)} = \frac{(1-p)^{n+k-1}p}{(1-p)^n} \\ &= (1-p)^{k-1}p = P(X = k). \end{aligned}$$

2.68. For $k \geq 1$, the assumed memoryless property gives

$$P(X = k) = P(X = k+1 \,|\, X > 1) = \frac{P(X = k+1)}{P(X > 1)}$$

which we convert into $P(X = k+1) = P(X > 1)P(X = k)$. Now let $m \ge 2$, and apply this repeatedy to $k = m-1, m-2, \dots, 2$:

$$\begin{aligned} P(X = m) &= P(X > 1)P(X = m-1) = P(X > 1)^2 P(X = m-2) \\ &= \cdots = P(X > 1)^{m-1} P(X = 1). \end{aligned}$$

Set $p = P(X = 1)$. Then it follows that $P(X = m) = (1-p)^{m-1}p$ for all $m \ge 1$ ($m = 1$ by definition of p, $m \ge 2$ by the calculation above). In other words, $X \sim \text{Geom}(p)$.

2.69. We assume that the successive flips of a given coin are independent. This gives us the conditional independence:

$$\begin{aligned} &P(A_1A_2 \,|\, F) = P(A_1 \,|\, F)\, P(A_2 \,|\, F), \quad P(A_1A_2 \,|\, M) = P(A_1 \,|\, M)\, P(A_2 \,|\, M), \\ &\qquad \text{and} \ \ P(A_1A_2 \,|\, H) = P(A_1 \,|\, H)\, P(A_2 \,|\, H). \end{aligned}$$

The solution comes by the law of total probability:

$$\begin{aligned} P(A_1A_2) &= P(A_1A_2 \,|\, F)\, P(F) + P(A_1A_2 \,|\, M)\, P(M) + P(A_1A_2 \,|\, H)\, P(H) \\ &= P(A_1 \,|\, F)P(A_2 \,|\, F)P(F) + P(A_1 \,|\, B)P(A_2 \,|\, B)P(B) \\ &= \tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{90}{100} + \tfrac{3}{5} \cdot \tfrac{3}{5} \cdot \tfrac{9}{100} + \tfrac{9}{10} \cdot \tfrac{9}{10} \cdot \tfrac{1}{100} = \tfrac{2655}{10{,}000}. \end{aligned}$$

Now $\frac{2655}{10{,}000} \ne (\frac{513}{1000})^2$ which says that $P(A_1A_2) \ne P(A_1)P(A_2)$. In other words, A_1 and A_2 are *not* independent without the conditioning on the type of coin. The intuitive reason is that the first flip gives us information about the coin we hold, and thereby alters our expectations about the second flip.

2.70. The relevant probabilities: $P(A) = P(B) = 2p(1-p)$ and

$$P(AB) = P\{(\mathtt{T},\mathtt{H},\mathtt{T}), (\mathtt{H},\mathtt{T},\mathtt{H})\} = p^2(1-p) + p(1-p)^2 = p(1-p).$$

Thus A and B are independent if and only if

$$\begin{aligned} &\big(2p(1-p)\big)^2 = p(1-p) \iff 4p^2(1-p)^2 - p(1-p) = 0 \\ &\quad\iff p(1-p)\big(4p(1-p) - 1\big) = 0 \\ &\quad\iff p = 0 \ \text{ or } \ 1-p = 0 \ \text{ or } \ 4p(1-p) - 1 = 0 \iff p \in \{0, \tfrac{1}{2}, 1\}. \end{aligned}$$

Note that cancelling $p(1-p)$ from the very first equation misses the solutions $p = 0$ and $p = 1$.

2.71. Let $F = \{\text{coin is fair}\}$, $B = \{\text{coin is biased}\}$ and $A_k = \{k\text{th flip is tails}\}$. We assume that conditionally on F, the events A_k are independent, and similarly conditionally on B. Let $D_n = A_1 \cap A_2 \cap \cdots \cap A_n = \{\text{the first } n \text{ flips are all tails}\}$.

(a)

$$\begin{aligned} P(B|D_n) &= \frac{P(D_n|B)P(B)}{P(D_n|B)P(B) + P(D_n|F)P(F)} = \frac{(\frac{3}{5})^n \frac{1}{10}}{(\frac{3}{5})^n \frac{1}{10} + (\frac{1}{2})^n \frac{9}{10}} \\ &= \frac{(\frac{3}{5})^n}{(\frac{3}{5})^n + 9(\frac{1}{2})^n}. \end{aligned}$$

In particular, $P(B|D_1) = \frac{2}{17}$ and $P(B|D_2) = \frac{4}{29}$.

(b)

$$\frac{(\frac{3}{5})^{24}}{(\frac{3}{5})^{24}+9(\frac{1}{2})^{24}} \approx 0.898$$

while

$$\frac{(\frac{3}{5})^{25}}{(\frac{3}{5})^{25}+9(\frac{1}{2})^{25}} \approx 0.914,$$

so 25 flips are needed.

(c)

$$\begin{aligned} P(A_{n+1}|D_n) = \frac{P(D_{n+1})}{P(D_n)} &= \frac{P(D_{n+1}|B)P(B)+P(D_{n+1}|F)P(F)}{P(D_n|B)P(B)+P(D_n|F)P(F)} \\ &= \frac{(\frac{3}{5})^{n+1}\frac{1}{10}+(\frac{1}{2})^{n+1}\frac{9}{10}}{(\frac{3}{5})^{n}\frac{1}{10}+(\frac{1}{2})^{n}\frac{9}{10}}. \end{aligned}$$

(d) Intuitively speaking, an unending sequence of tails would push the probability of a biased coin to 1, and hence the probability of the next tails is $3/5$. For a rigorous calculation we take the limit of the previous answer:

$$\lim_{n\to\infty} P(A_{n+1}|D_n) = \lim_{n\to\infty} \frac{(\frac{3}{5})^{n+1}\frac{1}{10}+(\frac{1}{2})^{n+1}\frac{9}{10}}{(\frac{3}{5})^{n}\frac{1}{10}+(\frac{1}{2})^{n}\frac{9}{10}} = \lim_{n\to\infty} \frac{\frac{3}{5}+\frac{9}{2}(\frac{5}{6})^{n+1}}{1+9(\frac{5}{6})^{n}} = \frac{3}{5}.$$

2.72. The sample space for n trials is the same, regardless of the probabilities, namely the space of ordered n-tuples of zeros and ones:

$$\Omega = \{\omega = (s_1,\dots,s_n) : \text{each } s_i \text{ equals 0 or 1}\}.$$

By independence, the probability of a sample point $\omega = (s_1,\dots,s_n)$ is obtained by multiplying together a factor p_i for each $s_i = 1$ and $1-p_i$ for each $s_i = 0$. We can express this in a single formula as follows:

$$P\{(s_1,\dots,s_n)\} = \prod_{i=1}^{n}\big(p_i^{s_i}(1-p_i)^{1-s_i}\big).$$

2.73. Let X be the number of blond customers at the pancake place. The population of the town is 500, and 100 of them are blond. We may assume that the visitors are chosen randomly from the population, which means that we take a sample of size 14 without replacement from the population. X denotes the number of blonds among this sample. This is exactly the setup for the hypergeometric distribution and $X \sim \text{Hypergeom}(500, 100, 14)$. (Because the total population size is $N = 500$, the number of blonds is $N_A = 100$ and we take a sample of $n = 14$.) We can now use the probability mass function of the hypergeometric distribution to answer the two questions.

(a)

$$P(\text{exactly 10 blonds}) = P(X=10) = \frac{\binom{100}{10}\binom{400}{4}}{\binom{500}{14}} \approx 0.00003122.$$

(b)

$$\begin{aligned} P(\text{at most 2 blonds}) = P(X \le 2) &= \sum_{k=0}^{2} P(X=k) = \sum_{k=0}^{2}\frac{\binom{100}{k}\binom{400}{14-k}}{\binom{500}{14}} \\ &\approx 0.4458. \end{aligned}$$

2.74. Define events: $D = \{\text{Steve is a drug user}\}$, $A_1 = \{\text{Steve fails the first drug test}\}$ and $A_2 = \{\text{Steve fails the second drug test}\}$. Assume that Steve is no more or less likely to be a drug user than a random person from the company, so $P(D) = 0.01$. The data about the reliability of the tests tells us that $P(A_i|D) = 0.99$ and $P(A_i|D^c) = 0.02$ for $i = 1, 2$, and conditional independence $P(A_1A_2|D) = P(A_1|D)P(A_2|D)$ and also the same under conditioning on D^c.

(a)

$$P(D|A_1) = \frac{P(A_1|D)P(D)}{P(A_1|D)P(D) + P(A_1|D^c)P(D^c)} = \frac{\frac{99}{100} \cdot \frac{1}{100}}{\frac{99}{100} \cdot \frac{1}{100} + \frac{2}{100} \cdot \frac{99}{100}} = \frac{1}{3}$$

(b)

$$\begin{aligned} P(A_2|A_1) &= \frac{P(A_1A_2)}{P(A_1)} = \frac{P(A_1A_2|D)P(D) + P(A_1A_2|D^c)P(D^c)}{P(A_1|D)P(D) + P(A_1|D^c)P(D^c)} \\ &= \frac{\left(\frac{99}{100}\right)^2 \cdot \frac{1}{100} + \left(\frac{2}{100}\right)^2 \cdot \frac{99}{100}}{\frac{99}{100} \cdot \frac{1}{100} + \frac{2}{100} \cdot \frac{99}{100}} = \frac{103}{300} \approx 0.3433. \end{aligned}$$

(c)

$$\begin{aligned} P(D|A_1A_2) &= \frac{P(A_1A_2|D)P(D)}{P(A_1A_2|D)P(D) + P(A_1A_2|D^c)P(D^c)} \\ &= \frac{\left(\frac{99}{100}\right)^2 \cdot \frac{1}{100}}{\left(\frac{99}{100}\right)^2 \cdot \frac{1}{100} + \left(\frac{2}{100}\right)^2 \cdot \frac{99}{100}} = \frac{99}{103} \approx 0.9612. \end{aligned}$$

2.75. We introduce the following events:

$$\begin{aligned} A &= \{\text{the store gets its phones from factory II}\}, \\ B_i &= \{\text{the } i\text{th phone is defective}\}, \quad i = 1, 2. \end{aligned}$$

Then A^c is the event that the phone is from factory I. We know that

$$P(A) = 0.4 = \frac{2}{5}, \quad P(A^c) = 0.6 = \frac{3}{5}, \quad P(B_i|A) = 0.2 = \frac{1}{5}, \quad P(B_i|A^c) = 0.1 = \frac{1}{10}.$$

We need to compute $P(A|B_1B_2)$. By Bayes' theorem,

$$P(A|B_1B_2) = \frac{P(B_1B_2|A) \cdot P(A)}{P(B_1B_2|A)P(A) + P(B_1B_2|A^c)P(A^c)}.$$

We may assume that conditionally on A the events B_1 and B_2 are independent. This means that given that the store gets its phones from factory II, the defectiveness of the phones stocked there are independent. We may also assume that conditionally on A^c the events B_1 and B_2 are independent. Then

$$P(B_1B_2|A) = P(B_1|A)P(B_2|A) = (\tfrac{1}{5})^2, \quad P(B_1B_2|A^c) = P(B_1|A^c)P(B_2|A^c) = (\tfrac{1}{10})^2$$

and

$$P(A|B_1B_2) = \frac{(\frac{1}{5})^2 \cdot \frac{2}{5}}{(\frac{1}{5})^2 \cdot \frac{2}{5} + (\frac{1}{10})^2 \cdot \frac{3}{5}} = \frac{8}{11} \approx 0.7273.$$

2.76. Let A_2 be the event that the second test comes back positive. Take now $P(D) = \frac{96}{494} \approx 0.194$ as the prior. Then

$$P(D|A_2) = \frac{P(A_2|D)P(D)}{P(A_2|D)P(D) + P(A_2|D^c)P(D^c)}$$
$$= \frac{\frac{96}{100} \cdot \frac{96}{494}}{\frac{96}{100} \cdot \frac{96}{494} + \frac{2}{100} \cdot \frac{398}{494}} = \frac{2304}{2503} \approx 0.9205.$$

2.77. By definition $P(A|B) = \frac{P(AB)}{P(B)}$. Since $AB \subset B$, we have $P(AB) \le P(B)$ and thus $P(A|B) = \frac{P(AB)}{P(B)} \le 1$. Furthermore, $P(A|B) = \frac{P(AB)}{P(B)} \ge 0$ because $P(B) > 0$ and $P(AB) \ge 0$. The property $0 \le P(A|B) \le 1$.

To check $P(\Omega \,|\, B) = 1$ note that $\Omega \cap B = B$, and so

$$P(\Omega \,|\, B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Similarly, $\varnothing \cap B = \varnothing$, thus

$$P(\varnothing \,|\, B) = \frac{P(\varnothing \cap B)}{P(B)} = \frac{P(\varnothing)}{P(B)} = \frac{0}{P(B)} = 0.$$

Finally, if we have a pairwise disjoint sequence $\{A_i\}$ then $\{BA_i\}$ are also pairwise disjoint, and their union is $(\cup_{i=1}^{\infty} A_i) \cap B$. This gives

$$P(\cup_{i=1}^{\infty} A_i \,|\, B) = \frac{P\left((\cup_{i=1}^{\infty} A_i) \cap B)\right)}{P(B)} = \frac{P\left(\cup_{i=1}^{\infty} A_i B)\right)}{P(B)}$$
$$= \frac{\sum_{i=1}^{\infty} P\left(A_i B)\right)}{P(B)} = \sum_{i=1}^{\infty} \frac{P\left(A_i B)\right)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B).$$

2.78. Define events $D = \{A \text{ happens before } B\}$ and

$$D_n = \{\text{neither } A \text{ nor } B \text{ happens in trials } 1, \dots, n-1, \\ \text{and } A \text{ happens in trial } n\}.$$

Then D is the union of the pairwise disjoint events $\{D_n\}_{1 \le n < \infty}$. This statement uses the assumption that A and B are disjoint. Without that assumption we would have to add to D_n the condition that B^c happens in trial n.

$$P(D) = \sum_{n=1}^{\infty} P(D_n) = \sum_{n=1}^{\infty} \bigl(1 - P(A \cup B)\bigr)^{n-1} P(A)$$
$$= \frac{P(A)}{P(A \cup B)} = P(A \,|\, A \cup B).$$

2.79. Following the text, we consider

$$\Omega = \{(x_1, \dots, x_{23}) : x_i \in \{1, \dots, 365\}\},$$

which is the set of possible birthday combinations for 23 people. Note that $\#\Omega = 365^{23}$. Next, note that there are exactly

$$365 \cdot 364 \cdot \dots \cdot (365 - 21) \cdot 22 = 22 \cdot \prod_{k=0}^{21} (365 - k)$$

ways to choose the first 22 birthdays to be all different and the twenty-third to be one of the first 22. Thus, the desired probability is

$$\frac{22 \cdot \prod_{k=0}^{21}(365-k)}{365^{23}} \approx 0.0316.$$

2.80. Assume that birth months of distinct people are independent and that for any particular person each month is equally likely. Then we are asking that a sample of seven items with replacement from a set of 12 produces no repetitions. The probability is

$$\frac{12 \cdot 11 \cdot 10 \cdots 6}{12^7} = \frac{385}{3456} \approx 0.1114.$$

2.81. Let A_n be the event that there is a match among the birthdays of the chosen n Martians. Then

$$P(A_n) = 1 - P(\text{all } n \text{ birthdays are distinct}) = 1 - \frac{669 \cdot 668 \cdots (669-(n-1))}{669^n}$$

To estimate the product we use $1 - x \simeq e^{-x}$ to get

$$\begin{aligned}\frac{669 \cdot 668 \cdots (669-(n-1))}{669^n} &= \prod_{k=0}^{n-1}\left(1 - \frac{k}{669}\right) \approx \prod_{k=0}^{n-1} e^{-\frac{k}{669}} \\ &= e^{-\frac{1}{669}\sum_{k=0}^{n-1} k} = e^{-\frac{1}{669}\frac{n(n-1)}{2}} \approx e^{-\frac{n^2}{2\cdot 669}}\end{aligned}$$

Thus $P(A_n) \approx 1 - e^{-\frac{n^2}{2\cdot 669}}$. Now solving the inequality $P(A_n) \geq 0.9$:

$$1 - e^{-\frac{n^2}{2\cdot 669}} \geq 0.9 \iff \frac{n^2}{2 \cdot 669} \geq -\ln(1-0.9) \iff n \geq \sqrt{2 \cdot 669 \ln 10} \simeq 55.5.$$

This would suggest $n = 56$.

In fact this is correct: the actual numerical values are $P(A_{56}) \simeq 0.9064$ and $P(A_{55}) \simeq 0.8980$.

Solutions to Chapter 3

3.1. (a) The random variable X takes the values $1, 2, 3, 4$ and 5. Collecting the probabilities corresponding to the values that are at most 3 we get

$$P(X \le 3) = P(X = 1) + P(X = 2) + P(X = 3) = p_X(1) + p_X(2) + p_X(3) = \frac{1}{7} + \frac{1}{14} + \frac{3}{14} = \frac{3}{7}.$$

(b) Now we have to collect the probabilities corresponding to the values which are less than 3:

$$P(X < 3) = P(X = 1) + P(X = 2) = p_X(1) + p_X(2) = \frac{1}{7} + \frac{1}{14} = \frac{3}{14}.$$

(c) First we use the definition of conditional probability to get

$$P(X < 4.12 \,|\, X > 1.638) = \frac{P(X < 4.12 \text{ and } X > 1.638)}{P(X > 1.638)}.$$

We have $P(X < 4.12 \text{ and } X > 1.638) = P(1.638 < X < 4.12)$. The possible values of X between 1.638 and 4.12 are $2, 3$ and 4. Thus

$$P(X < 4.12 \text{ and } X > 1.638) = p_X(2) + p_X(3) + p_X(4) = \frac{1}{14} + \frac{3}{14} + \frac{2}{7} = \frac{4}{7}.$$

Similarly,

$$P(X > 1.638) = p_X(2) + p_X(3) + p_X(4) + p_X(5) = \frac{1}{14} + \frac{3}{14} + \frac{2}{7} + \frac{2}{7} = \frac{6}{7}.$$

From this we get

$$P(X < 4.12 \,|\, X > 1.638) = \frac{\frac{4}{7}}{\frac{6}{7}} = \frac{2}{3}.$$

3.2. (a) We must have that the probability mass function sums to one. Hence, we require

$$1 = \sum_{k=1}^{6} p(k) = c\,(1 + 2 + 3 + 4 + 5 + 6) = 21c.$$

Thus, $c = \frac{1}{21}$.

(b) The probability that X is odd is

$$P(X \in \{1,3,5\}) = p(1) + p(3) + p(5) = \frac{1}{21}(1+3+5) = \frac{9}{21} = \frac{3}{7}.$$

3.3. (a) We need to check that f is non-negative and that it integrates to 1 on $\mathbb{R}$. The non-negativity follows from the definition. For the integral we can compute

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} 3e^{-3x}dx = -e^{-3x}\big|_{x=0}^{x=\infty} = \lim_{x\to\infty}(-e^{-3x}) - (-e^0) = 0 - (-1) = 1.$$

In the first step we used the formula for $f(x)$, and the fact that it is equal to 0 for $x \le 0$.
(b) Using the definition of the probability density function we get

$$P(-1 < X < 1) = \int_{-1}^{1} f(x)dx = \int_0^1 3e^{-3x}dx = -e^{-3x}\big|_{x=0}^{x=1} = 1 - e^{-3}.$$

(c) Using the definition of the probability density function again we get

$$P(X < 5) = \int_{-\infty}^{5} f(x)dx = \int_0^5 3e^{-3x}dx = -e^{-3x}\big|_{x=0}^{x=5} = 1 - e^{-15}.$$

(d) From the definition of conditional probability

$$P(2 < X < 4 \,|\, X < 5) = \frac{P(2 < X < 4 \text{ and } X < 5)}{P(X < 5)}.$$

We have $P(2 < X < 4 \text{ and } X < 5) = P(2 < X < 4)$. Similar to the previous parts:

$$P(2 < X < 4) = \int_2^4 f(x)dx = \int_2^4 3e^{-3x}dx = -e^{-3x}\big|_{x=2}^{x=4} = e^{-6} - e^{-15}.$$

Using the result of part (c):

$$P(2 < X < 4 \,|\, X < 5) = \frac{P(2 < X < 4)}{P(X < 5)} = \frac{e^{-6} - e^{-15}}{1 - e^{-15}}.$$

3.4. (a) The density of X is $\frac{1}{6}$ on $[4, 10]$ and zero otherwise. Hence,

$$P(X < 6) = P(4 < X < 6) = \frac{6-4}{6} = \frac{1}{3}.$$

(b)

$$\begin{aligned} P(|X - 7| > 1) &= P(X - 7 > 1) + P(X - 7 < -1) = P(X > 8) + P(X < 6) \\ &= \frac{10-8}{6} + \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

(c) For $4 \le t \le 6$ we have

$$P(X < t \,|\, X < 6) = \frac{P(X < t, X < 6)}{P(X < 6)} = \frac{P(X < t)}{1/3} = 3 \cdot \frac{t-4}{6} = \frac{t-4}{2}.$$

3.5. The possible values of a discrete random variable are exactly the values where the c.d.f. jumps. In this case these are the values $1, 4/3, 3/2$ and $9/5$. The corresponding probabilities are equal to the size of corresponding jumps:

$$\begin{aligned}
p_X(1) &= \tfrac{1}{3} - 0 = \tfrac{1}{3},\\
p_X(4/3) &= \tfrac{1}{2} - \tfrac{1}{3} = \tfrac{1}{6},\\
p_X(3/2) &= \tfrac{3}{4} - \tfrac{1}{2} = \tfrac{1}{4},\\
p_X(9/5) &= 1 - \tfrac{3}{4} = \tfrac{1}{4}.
\end{aligned}$$

3.6. For the random variable in Exercise 3.1, we may use (3.13). For $s \in (-\infty, \infty)$,

$$F(s) = P(X \le s) = \begin{cases} 0, & s < 1 \\ \frac{1}{7}, & 1 \le s < 2 \\ \frac{3}{14}, & 2 \le s < 3 \\ \frac{6}{14}, & 3 \le s < 4 \\ \frac{10}{14}, & 4 \le s < 5 \\ 1, & 5 \le s. \end{cases}$$

For the random variable in Exercise 3.3, we may use (3.15). For $s \le 0$ we have that

$$P(X \le s) = 0,$$

whereas for $s > 0$ we have

$$P(X \le s) = \int_0^s 3e^{-3x} dx = 1 - e^{-3s}.$$

3.7. (a) If $P(a \le X \le b) = 1$ then $F(y) = 0$ for $y < a$ and $F(y) = 1$ for $y \ge b$. From the definition of F we see that $a = \sqrt{2}$ and $b = \sqrt{3}$ gives the smallest such interval.

(b) Since X is continuous, $P(X = 1.6) = 0$. We can also see this directly from F:

$$P(X = 1.6) = F(1.6) - \lim_{x \to 1.6-} F(x) = F(1.6) - F(1.6-).$$

Since $F(x)$ is continuous at $x = 1.6$ (actually, it is continuous everywhere), we have $F(1.6-) = F(1.6)$ and this gives $P(X = 1.6) = 0$ again.

(c) Because X is continuous, we have $P(1 \le X \le 3/2) = P(1 < X \le 3/2)$. We also have

$$\begin{aligned}
P(1 \le X \le 3/2) &= P(1 < X \le 3/2) = P(X \le 3/2) - P(X \le 1)\\
&= F(3/2) - F(1) = ((\tfrac{3}{2})^2 - 2) - 0 = \tfrac{9}{4} - 2 = \tfrac{1}{4}.
\end{aligned}$$

We used $1 < \sqrt{2} \le 3/2 \le \sqrt{3}$ when we evaluated $F(3/2) - F(1)$.

(d) Since F is continuous, and it is differentiable apart from finitely many points ($\sqrt{2}$ and $\sqrt{3}$), we can just differentiate it to get the probability density function:

$$f(x) = F'(x) = \begin{cases} 2x & \text{if } \sqrt{2} < x < \sqrt{3} \\ 0 & \text{otherwise.} \end{cases}$$

We chose 0 for the value of f at $\sqrt{2}$ and $\sqrt{3}$, but the actual values are not important.

3.8. (a) We have

$$E[X] = \sum_{k=1}^{5} k p_X(k) = 1 \cdot \frac{1}{7} + 2 \cdot \frac{1}{14} + 3 \cdot \frac{3}{14} + 4 \cdot \frac{2}{7} + 5 \cdot \frac{2}{7} = \frac{7}{2}.$$

(b) We have

$$E[|X-2|] = \sum_{k=1}^{5} |k-2| p_X(k) = 1 \cdot \frac{1}{7} + 0 \cdot \frac{1}{14} + 1 \cdot \frac{3}{14} + 2 \cdot \frac{2}{7} + 3 \cdot \frac{2}{7} = \frac{25}{14}.$$

3.9. (a) Since X is continuous, we can compute its mean as

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} x \cdot 3e^{-3x} dx.$$

Using integration by parts we can evaluate the last integral to get $E[X] = \frac{1}{3}$.
(b) e^{2X} is a function of X, and X is continuous, so we can compute $E[e^{2X}]$ as follows:

$$E[e^{2X}] = \int_{-\infty}^{\infty} e^{2x} f(x) dx = \int_{0}^{\infty} e^{2x} \cdot 3e^{-3x} dx = \int_{0}^{\infty} 3e^{-x} dx = 3.$$

3.10. (a) The random variable $|X|$ takes values 0 and 1 with probabilities

$$P(|X|=0) = P(X=0) = \tfrac{1}{3} \quad \text{and} \quad P(|X|=1) = P(X=1) + P(X=-1) = \tfrac{2}{3}.$$

Then the definition of expectation gives

$$E[|X|] = 0 \cdot P(|X|=0) + 1 \cdot P(|X|=1) = \tfrac{2}{3}.$$

(b) Applying formula (3.24):

$$\begin{aligned} E[|X|] &= \sum_k |k| \, P(X=k) = 1 \cdot P(X=-1) + 0 \cdot P(X=0) + 1 \cdot P(X=1) \\ &= \tfrac{1}{2} + \tfrac{1}{6} = \tfrac{2}{3}. \end{aligned}$$

3.11. By (3.25) we have

$$E[(Y-1)^2] = \int_{-\infty}^{\infty} (x-1)^2 f(x) \, dx = \int_{1}^{2} (x-1)^2 \cdot \tfrac{2}{3} x \, dx = \tfrac{7}{18}.$$

The interval of integration changed from $(-\infty, \infty)$ to $[1,2]$ since $f(x) = 0$ outside $[1,2]$.

3.12. The expectation is

$$E[X] = \sum_{n=1}^{\infty} n P(X=n) = \sum_{n=1}^{\infty} n \cdot \frac{6}{\pi^2} \frac{1}{n^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n},$$

which is infinite by the conclusion of Example D.5 (using $\gamma = 1$ in that example).

3.13. (a) We need to find an m for which $P(X \ge m) \ge 1/2$ and $P(X \le m) \ge 1/2$. For X from Exercise 3.1 we have

$$P(X \le 3) = \tfrac{3}{7}, \quad P(X \le 4) = \tfrac{5}{7} \quad P(X \le 5) = 1$$

and

$$P(X \ge 3) = \tfrac{11}{14}, \quad P(X \ge 4) = \tfrac{3}{7}, \quad P(X \ge 5) = \tfrac{2}{7}.$$

From this we get that $m = 4$ works as the median, but any number that is larger or smaller than 4 is not a median.

For X from Exercise 3.3 we have

$$P(X \le m) = 1 - e^{-3m}, \text{ and } P(X \ge m) = e^{-3m} \qquad \text{if } m \ge 0$$

and $P(X \le m) = 0$, $P(X \ge m) = 1$ for $m < 0$. From this we get that the median m satisfies $e^{-3m} = 1/2$, which leads to $m = \ln(2)/3$.

(b) We need $P(X \le q) \ge 0.9$ and $P(X \ge q) \ge 0.1$. Since X is continuous, we must have $P(X \le q) + P(X \ge q) = 1$ and hence $P(X \le q) = 0.9$ and $P(X \ge q) = 0.1$. Using the calculations from part (a) we see that $e^{-3m} = 0.1$ from which $q = \ln(10)/3$.

3.14. The mean of the random variable X from Exercise 3.1 is

$$E[X] = \sum_{k=1}^{5} k p_X(k) = 1 \cdot \frac{1}{7} + 2 \cdot \frac{1}{14} + 3 \cdot \frac{3}{14} + 4 \cdot \frac{2}{7} + 5 \cdot \frac{2}{7} = \frac{7}{2}.$$

The second moment is

$$E[X^2] = \sum_{k=1}^{5} k^2 p_X(k) = 1^2 \cdot \frac{1}{7} + 2^2 \cdot \frac{1}{14} + 3^2 \cdot \frac{3}{14} + 4^2 \cdot \frac{2}{7} + 5^2 \cdot \frac{2}{7} = \frac{197}{14}.$$

Therefore, the variance is

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{197}{14} - \left(\frac{7}{2}\right)^2 = \frac{51}{28}.$$

Now let X be the random variable from Exercise 3.3. The mean is

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot 3e^{-3x} dx = \frac{1}{3},$$

which follows from an application of integration by parts. The second moment is

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^{\infty} x^2 \cdot 3e^{-3x} dx = \frac{2}{9},$$

where the integral is calculated using two rounds of integration by parts. Thus, the variance is

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{2}{9} - \left(\frac{1}{3}\right)^2 = \frac{1}{9}.$$

3.15. (a) We have

$$E[3X + 2] = 3E[X] + 2 = 3 \cdot 3 + 2 = 11.$$

(b) We know that $\mathrm{Var}(X) = E[X^2] - E[X]^2$. Rearranging the terms gives

$$E[X^2] = \mathrm{Var}(X) + E[X]^2 = 4 + 3^2 = 13.$$

(c) Expanding the square gives

$$E[(2X+3)^2] = E[4X^2 + 12X + 9] = 4E[X^2] + 12E[X] + 9 = 4 \cdot 13 + 12 \cdot 3 + 9 = 97,$$

where we also used the result of part (b).

(d) We have $\mathrm{Var}(4X - 2) = 4^2\, \mathrm{Var}(X) = 4^2 \cdot 4 = 64$.

3.16. The expectation of Z is

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z)dz = \int_1^2 z \cdot \frac{1}{7}dz + \int_5^7 z \cdot \frac{3}{7}dz = \frac{1}{7} \cdot \frac{4-1}{2} + \frac{3}{7} \cdot \frac{49-25}{2} = \frac{75}{14}.$$

The second moment is

$$\begin{aligned} E[Z^2] &= \int_{-\infty}^{\infty} z^2 f_Z(z)dz = \int_1^2 z^2 \cdot \frac{1}{7}dz + \int_5^7 z^2 \cdot \frac{3}{7}dz \\ &= \frac{1}{7} \cdot \frac{8-1}{3} + \frac{3}{7} \cdot \frac{7^3 - 5^3}{3} = \frac{661}{21}. \end{aligned}$$

Hence, the variance is

$$\mathrm{Var}(Z) = E[Z^2] - (E[Z])^2 = \frac{661}{21} - \left(\frac{75}{14}\right)^2 = \frac{1633}{588}.$$

3.17. If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable. We will reduce each question to a probability involving the standard normal random variable Z. Recall that $P(Z < x) = \Phi(x)$ and $P(Z > x) = 1 - \Phi(x)$. The numerical values of Φ can be looked up using the table in Appendix E.

(a)

$$\begin{aligned} P(X > 3.5) &= P\left(\frac{X-\mu}{\sigma} > \frac{3.5-\mu}{\sigma}\right) \\ &= P(Z > \tfrac{5.5}{\sqrt{7}}) = 1 - \Phi(\tfrac{5.5}{\sqrt{7}}) \\ &\approx 1 - \Phi(2.08) \approx 1 - 0.9812 = 0.0188. \end{aligned}$$

(b)

$$\begin{aligned} P(-2.1 < X < -1.9) &= P\left(\frac{-2.1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{-1.9-\mu}{\sigma}\right) \\ &= P(\tfrac{-0.1}{\sqrt{7}} < Z < \tfrac{0.1}{\sqrt{7}}) = \Phi(\tfrac{0.1}{\sqrt{7}}) - \Phi(-\tfrac{0.1}{\sqrt{7}}) \\ &= \Phi(\tfrac{0.1}{\sqrt{7}}) - (1 - \Phi(\tfrac{0.1}{\sqrt{7}})) = 2\Phi(\tfrac{0.1}{\sqrt{7}}) - 1 \\ &\approx 2\Phi(0.04) - 1 \approx 2 \cdot 0.516 - 1 = 0.032. \end{aligned}$$

(c)

$$\begin{aligned} P(X < 2) &= P\left(\frac{X-\mu}{\sigma} < \frac{2-\mu}{\sigma}\right) \\ &= P(Z < \tfrac{4}{\sqrt{7}}) = \Phi(\tfrac{4}{\sqrt{7}}) \\ &\approx \Phi(1.51) \approx 0.9345. \end{aligned}$$

(d)

$$\begin{aligned} P(X < -19) &= P\left(\frac{X-\mu}{\sigma} < \frac{-10-\mu}{\sigma}\right) \\ &= P(Z < \tfrac{-8}{\sqrt{7}}) = \Phi(-\tfrac{8}{\sqrt{7}}) = 1 - \Phi(-\tfrac{8}{\sqrt{7}}) \\ &\approx 1 - \Phi(3.02) \approx 1 - 0.9987 = 0.0013. \end{aligned}$$

(e)

$$\begin{aligned} P(X > 4) &= P\left(\frac{X-\mu}{\sigma} > \frac{4-\mu}{\sigma}\right) \\ &= P(Z > \tfrac{6}{\sqrt{7}}) = 1 - \Phi(\tfrac{6}{\sqrt{7}}) \\ &\approx 1 - \Phi(2.27) \approx 1 - 0.9884 = 0.0116. \end{aligned}$$

3.18. If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable. Recall that the values $P(Z < x) = \Phi(x)$ can be looked up using the table in Appendix E.

(a)

$$\begin{aligned} P(2 < X < 6) &= P\left(\frac{2-3}{2} < \frac{X-3}{2} < \frac{6-3}{2}\right) = P(-\tfrac{1}{2} < Z < \tfrac{3}{2}) \\ &= P(Z < 1.5) - P(Z < -.5) = \Phi(1.5) - \Phi(-0.5) \\ &= \Phi(1.5) - (1 - \Phi(0.5)) = 0.9332 - (1 - 0.6915) = .6247. \end{aligned}$$

(b) We need c so that

$$0.33 = P(X > c) = P\left(\frac{X-3}{2} > \frac{c-3}{2}\right) = 1 - \Phi\left(\frac{c-3}{2}\right).$$

Hence, we need c satisfying $\Phi\left(\frac{c-3}{2}\right) = 0.67$. Checking the table in Appendix E, we conclude that $\Phi(z) = 0.67$ is solved by $z = 0.44$. Hence,

$$\frac{c-3}{2} = 0.44 \iff c = 3.88.$$

(c) We have that

$$E[X^2] = \mathrm{Var}(X) + (E[X])^2 = 4 + 3^2 = 13.$$

3.19. From the definition of the c.d.f. we have

$$\begin{aligned} F(2) = P(Z \le 2) &= P(Z=0) + P(Z=1) + P(Z=2) \\ &= \binom{10}{0}\left(\tfrac{1}{3}\right)^0\left(\tfrac{2}{3}\right)^{10} + \binom{10}{1}\left(\tfrac{1}{3}\right)^1\left(\tfrac{2}{3}\right)^9 + \binom{10}{2}\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)^8 \\ &= \frac{2^{10} + 10\cdot 2^9 + 45\cdot 2^8}{3^{10}} \approx 0.299. \end{aligned}$$

The solution for $F(8)$ can be done the same way:

$$F(8) = P(Z \le 8) = \sum_{i=0}^{8}\binom{10}{i}\left(\tfrac{1}{3}\right)^i\left(\tfrac{2}{3}\right)^{10-i}.$$

There is another way which involves fewer terms:

$$\begin{aligned} F(8) = P(Z \le 8) = 1 - P(Z \ge 9) &= 1 - \left(\binom{10}{9}\left(\tfrac{1}{3}\right)^9\left(\tfrac{2}{3}\right)^1 + \binom{10}{10}\left(\tfrac{1}{3}\right)^{10}\left(\tfrac{2}{3}\right)^0\right) \\ &= 1 - \frac{21}{3^{10}} \approx 0.9996. \end{aligned}$$

3.20. We must show that $Y \sim \text{Unif}[0,c]$. We find the cumulative function. For any $t \in (-\infty, \infty)$ we have

$$F_Y(t) = P(Y \le t) = P(c - X \le t) = P(c - t \le X) = \begin{cases} 0, & t < 0 \\ \frac{c-(c-t)}{c} = \frac{t}{c}, & 0 \le t < c \\ 1, & c \le t. \end{cases}$$

which is the cumulative distribution function for a $\text{Unif}[0,c]$ random variable.

3.21. (a) The number of heads out of 2 coin flips can be 0, 1 or 2. These are the possible values of X. The possible outcomes of the experiment are $\{HH, HT, TH, TT\}$, and each one of these has a probability $\frac{1}{4}$. We can compute the probability mass function of X by identifying the events $\{X = 0\}$, $\{X = 1\}$, $\{X = 2\}$ and computing the corresponding probabilities:

$$\begin{aligned} p_X(0) &= P(X = 0) = P(\{TT\}) = \tfrac{1}{4} \\ p_X(1) &= P(X = 1) = P(\{HT, TH\}) = \tfrac{2}{4} = \tfrac{1}{2} \\ p_X(2) &= P(X = 2) = P(\{HH\}) = \tfrac{1}{4}. \end{aligned}$$

(b) Using the probability mass function from (a):

$$P(X \ge 1) = P(X = 1) + P(X = 2) = p_X(1) + p_X(2) = \tfrac{3}{4}$$

and

$$P(X > 1) = P(X = 2) = p_X(2) = \tfrac{1}{4}.$$

(c) Since X is a discrete random variable, we can compute the expectation as

$$E[X] = \sum_k k p_X(k) = 0 \cdot p_X(0) + 1 p_X(1) + 2 \cdot p_X(2) = \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1.$$

For the variance we need to compute $E[X^2]$:

$$E[X^2] = \sum_k k^2 p_X(k) = 0 \cdot p_X(0) + 1 p_X(1) + 4 \cdot p_X(2) = \tfrac{1}{2} + 4 \cdot \tfrac{1}{4} = \tfrac{3}{2}.$$

This gives

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \tfrac{3}{2} - 1 = \tfrac{1}{2}.$$

3.22. (a) The random variable X is binomially distributed with parameters $n = 3$ and $p = \frac{1}{2}$. Thus, the possible values of X are $\{0, 1, 2, 3\}$ and the probability mass function is

$$P(X = 0) = \frac{1}{2^3}, \quad P(X = 1) = 3 \cdot \frac{1}{2^3}, \quad P(X = 2) = 3 \cdot \frac{1}{2^3}, \quad P(X = 3) = \frac{1}{2^3}.$$

(b) We have

$$P(X \ge 1) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{3 + 3 + 1}{8} = \frac{7}{8},$$

and

$$P(X > 1) = P(X = 2) + P(X = 3) = \frac{3 + 1}{8} = \frac{1}{2}.$$

(c) The mean is

$$E[X] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2}.$$

The second moment is

$$E[X] = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = \frac{24}{8} = 3.$$

Hence, the variance is

$$\text{Var}(X) = E[X^2] - (E[X])^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$$

3.23. (a) The possible values for the profit (in dollars) are $0 - 1 = -1, 2 - 1 = 1, 100 - 1 = 99$ and $7000 - 1 = 6999$. The probability mass function can be computed as follows:

$$P(X = -1) = P(\text{the randomly chosen player was not a winner}) = \frac{10000 - 100}{10000} = \frac{99}{100},$$
$$P(X = 1) = P(\text{the randomly chosen player was one of the 80 who won \$2}) = \frac{80}{10000} = \frac{1}{125},$$
$$P(X = 99) = P(\text{the randomly chosen player was one of the 19 who won \$100}) = \frac{19}{10000},$$
$$P(X = 6999) = P(\text{the randomly chosen player was the one who won \$7000}) = \frac{1}{10000}.$$

(b)

$$P(X \geq 100) = P(X = 6999) = \frac{1}{10000}.$$

(c) Since X is discrete, we can find its expectation as

$$E[X] = \sum_k kP(X = k) = -1 \cdot \frac{99}{100} + 1 \cdot \frac{1}{125} + 99 \cdot \frac{19}{10000} + 6999 \cdot \frac{1}{10000} = -0.094.$$

For the variance we need $E[X^2]$:

$$E[X^2] = \sum_k k^2 P(X = k) = 1 \cdot \frac{99}{100} + 1 \cdot \frac{1}{125} + 99^2 \cdot \frac{19}{10000} + 6999^2 \cdot \frac{1}{10000} = 4918.22.$$

From this we get

$$\text{Var}(X) = E[X^2] - (E[X])^2 \approx 4918.21.$$

3.24. (a) We have

$$P(X \geq 2) = P(X = 2) + P(X = 3) = \frac{2}{7} + \frac{4}{7} = \frac{6}{7}.$$

(b) We have

$$E\left(\frac{1}{1+X}\right) = \frac{1}{1+1} \cdot \frac{1}{7} + \frac{1}{1+2} \cdot \frac{2}{7} + \frac{1}{1+3} \cdot \frac{4}{7} = \frac{13}{47}.$$

3.25. (a) If f is a pdf then $\int_{-\infty}^{\infty} f(x)dx = 1$. We have

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_1^3 (x^2 - b)dx = x^3/3 - bx \Big|_{x=1}^{x=3} = \frac{26}{3} - 2b.$$

This gives $b = \frac{23}{6}$. However, $x^2 - \frac{23}{6}$ is negative for $1 \le x < \sqrt{\frac{23}{6}} \approx 1.96$ which shows that the function f cannot be a pdf.
(b) We need $b \ge 0$, otherwise the function is zero everywhere. The $\cos x$ function is non-negative on $[-\pi/2, \pi/2]$, but then it goes below 0. Thus if g is a pdf then $b \le \pi/2$. Computing the integral of g on $(-\infty, \infty)$ we get

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-b}^{b} \cos(x)dx = 2\sin(b).$$

There is exactly one solution for $2\sin(b) = 1$ in the interval $(0, \pi/2]$, this is $b = \arcsin(1/2) = \pi/6$. For this choice of b the function g is a pdf.

3.26. (a) We require that the probability mass function sum to one. Hence,

$$1 = \sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} \frac{c}{k(k+1)}.$$

The sum can be computed in the following way:

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{c}{k(k+1)} &= \lim_{M\to\infty} \sum_{k=1}^{M} \frac{c}{k(k+1)} = c \lim_{M\to\infty} \sum_{k=1}^{M} \left(\frac{1}{k} - \frac{1}{k+1}\right) \\ &= c \lim_{M\to\infty} \left(1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{M} - \frac{1}{M+1}\right) \\ &= c \lim_{M\to\infty} \left(1 - \frac{1}{M+1}\right) = c.\end{aligned}$$

Combining the above shows that $c = 1$.

(b) Turning to the expectation,

$$E(X) = \sum_{k=1}^{\infty} k \frac{1}{k(k+1)} = \sum_{k=2}^{\infty} \frac{1}{k} = \infty,$$

by the conclusion of Example D.5.

3.27. (a) By collecting the possible values of X that are at least 2 we get

$$P(X \ge 2) = P(X = 2) + P(X = 3) + P(X = 4) = \tfrac{1}{5} + \tfrac{1}{5} + \tfrac{1}{5} = \tfrac{3}{5}.$$

(b) We have

$$P(X \le 3 | X \ge 2) = \frac{P(X \le 3 \text{ and } X \ge 2)}{P(X \ge 2)} = \frac{P(2 \le X \le 3)}{P(X \ge 2)}.$$

We already computed $P(X \ge 2) = \frac{3}{5}$ in (a). Similarly,

$$P(2 \le X \le 3) = P(X = 2) + P(X = 3) = \tfrac{2}{5},$$

and

$$P(X \le 3 | X \ge 2) = \frac{P(2 \le X \le 3)}{P(X \ge 2)} = \frac{2/5}{3/5} = \frac{2}{3}.$$

(c) We need to compute $E[X]$ and $E[X^2]$. Since X is discrete:

$$E[X] = \sum_k kP(X = k) = 1 \cdot \tfrac{2}{5} + 2 \cdot \tfrac{1}{5} + 3 \cdot \tfrac{1}{5} + 4 \cdot \tfrac{1}{5} = \tfrac{11}{5},$$

and

$$E[X^2] = \sum_k k^2 P(X = k) = 1 \cdot \tfrac{2}{5} + 4 \cdot \tfrac{1}{5} + 9 \cdot \tfrac{1}{5} + 16 \cdot \tfrac{1}{5} = \tfrac{31}{5}.$$

This leads to

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \tfrac{34}{25}.$$

3.28. (a) The possible values of X are 1, 2, and 3. Since there are three boxes with nice prizes, we have

$$P(X = 1) = \frac{3}{5}.$$

Next, for $X = 2$, we must first choose a box that does not have a good prize (two choices) followed by one that does (three choices). Hence,

$$P(X = 2) = \frac{2 \cdot 3}{5 \cdot 4} = \frac{3}{10}.$$

Similarly,

$$P(X = 3) = \frac{2 \cdot 1 \cdot 3}{5 \cdot 4 \cdot 3} = \frac{1}{10}.$$

(b) The expectation is

$$E[X] = 1 \cdot \frac{3}{5} + 2 \cdot \frac{3}{10} + 3 \cdot \frac{1}{10} = \frac{3}{2}.$$

(c) The second moment is

$$E[X^2] = 1^2 \cdot \frac{3}{5} + 2^2 \cdot \frac{3}{10} + 3^2 \cdot \frac{1}{10} = \frac{27}{10}.$$

Hence, the variance is

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{27}{10} - \left(\frac{3}{2}\right)^2 = \frac{9}{20}.$$

(d) Let W be the gain or loss in this game. Then

$$W = 100(2 - X) = 200 - 100X = \begin{cases} 100, & \text{if } X = 1 \\ 0, & \text{if } X = 2 \\ -100, & \text{if } X = 3. \end{cases}$$

Thus, by Fact 3.52,

$$E[W] = E[200 - 100X] = 200 - 100E[X] = 200 - 100 \cdot \frac{3}{2} = 50.$$

3.29. The possible values of X are the possible class sizes: $17, 21, 24, 28$. We can compute the corresponding probabilities by computing the probability of choosing a student from that class:

$$p_X(17) = \tfrac{17}{90},\ p_X(21) = \tfrac{21}{90} = \tfrac{7}{30},\ p_X(24) = \tfrac{24}{90} = \tfrac{4}{15},\ p_X(28) = \tfrac{28}{90} = \tfrac{14}{45}.$$

From this we can compute $E[X]$:

$$E[X] = \sum_k k P(X = k) = 17 \cdot \tfrac{17}{90} + 21 \cdot \tfrac{7}{30} + 24 \cdot \tfrac{4}{15} + 28 \cdot \tfrac{14}{45} = \tfrac{209}{9}.$$

For the variance we need $E[X^2]$:

$$E[X^2] = \sum_k k^2 P(X = k) = 17^2 \cdot \tfrac{17}{90} + 21^2 \cdot \tfrac{7}{30} + 24^2 \cdot \tfrac{4}{15} + 28^2 \cdot \tfrac{14}{45} = 555.$$

Then the variance is

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1274}{81}.$$

3.30. (a) The probability mass function is found by utilizing Fact 2.6. We have

$$\begin{aligned} P(X=0) &= P(\text{hit on first shot}) = \frac{1}{2} \\ P(X=1) &= P(\text{miss on first, then hit}) \\ &= P(\text{hit on second}|\text{miss on first})P(\text{miss on first}) = \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}. \end{aligned}$$

Continuing,

$$\begin{aligned} P(X=2) &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{4} = \frac{1}{12} \\ P(X=3) &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{1}{5} = \frac{1}{20} \\ P(X=4) &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{4}{5} = \frac{1}{5}. \end{aligned}$$

(b) The expected value of X, the number of misses, is

$$E[X] = 0\cdot\frac{1}{2} + 1\cdot\frac{1}{6} + 2\cdot\frac{1}{12} + 3\cdot\frac{1}{20} + 4\cdot\frac{1}{5} = \frac{77}{60}.$$

3.31. (a) We must have $1 = \int_{-\infty}^{\infty} f(x)dx$. So, we solve:

$$1 = \int_1^\infty cx^{-4}dx = \frac{c}{3}$$

which gives $c = 3$.

(b) We have

$$P(0.5 < X < 1) = \int_{0.5}^1 f(x)dx = \int_{0.5}^1 0dx = 0.$$

(c) We have

$$P(0.5 < X < 2) = \int_{0.5}^2 f(x)dx = \int_1^2 3x^{-4}dx = -x^{-3}\Big|_{x=1}^2 = 1 - \frac{1}{8} = \frac{7}{8}.$$

(d) We have

$$P(2 < X < 4) = \int_2^4 f(x)dx = \int_2^4 3x^{-4}dx = -x^{-3}\Big|_{x=2}^4 = \frac{1}{8} - \frac{1}{64} = \frac{7}{64}.$$

(e) For $x < 1$ we have $F_X(x) = P(X \le x) = 0$. For $x \ge 1$ we have

$$F(x) = P(X \le x) = \int_1^x 3y^{-4}dy = -y^{-3}\Big|_{y=1}^x = 1 - \frac{1}{x^3}.$$

(f) We have

$$E(X) = \int_{-\infty}^\infty xf(x)dx = \int_1^\infty x\cdot 3x^{-4}dx = \frac{3x^{-2}}{-2}\Bigg|_{x=1}^{x=\infty} = 3/2,$$

and

$$E(X^2) = \int_{-\infty}^{\infty} x f(x)dx = \int_1^{\infty} x^2 \cdot 3x^{-4}dx = \frac{3x^{-1}}{-1}\Big|_{x=1}^{x=\infty} = 3.$$

From this we get

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 3 - \frac{9}{4} = 3/4.$$

(g) We have

$$E[5X^2+3X] = \int_{-\infty}^{\infty}(5x^2+3x)f(x)dx = \int_1^{\infty}(5x^2+3x)\cdot 3x^{-4}dx = -\frac{9}{2x^2} - \frac{15}{x}\Big|_{x=1}^{x=\infty} = \frac{39}{2}.$$

(h) We

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)dx = \int_1^{\infty} x^n \cdot 3x^{-4}dx.$$

Evaluating this integral for integer values of n we get

$$E(X^n) = \begin{cases} \infty, & n \geq 3 \\ \frac{3}{3-n}, & n \leq 2. \end{cases}$$

3.32. (a) We have

$$P(X > 10) = \int_{10}^{\infty} \frac{1}{2}x^{-3/2}dx = -x^{-1/2}\big|_{x=10}^{\infty} = \frac{1}{\sqrt{10}}.$$

(b) For $t < 1$, we have that $F_X(t) = P(X \leq t) = 0$. For $t \geq 1$ we have

$$P(X \leq t) = \int_1^t \frac{1}{2}x^{-3/2}dx = -x^{-1/2}\big|_{x=1}^{t} = 1 - \frac{1}{\sqrt{t}}.$$

(c) We have

$$E[X] = \int_1^{\infty} x \cdot \frac{1}{2}x^{-3/2}dx = \frac{1}{2}\int_1^{\infty} x^{-1/2}\,dx = \infty.$$

This last equality can be seen as follows:

$$\int_1^{\infty} x^{-1/2}\,dx = \lim_{b\to\infty}\int_1^b x^{-1/2}\,dx = 2\lim_{b\to\infty} x^{1/2}\big|_{x=1}^{x=b} = 2\lim_{b\to\infty}(\sqrt{b}-1) = \infty.$$

(d) We have

$$E[X^{1/4}] = \int_1^{\infty} \frac{1}{2}x^{1/4}x^{-3/2}dx = \frac{1}{2}\int_1^{\infty} x^{-5/4}dx = -4\cdot\frac{1}{2}\cdot x^{-1/4}\big|_{x=1}^{\infty} = 2.$$

3.33. (a) A probability density function must be nonnegative, and it has to integrate to 1. Thus $c \geq 0$ and we must have

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_1^2 \frac{1}{4}dx + \int_3^5 c\,dx = \frac{1}{4} + 2c.$$

This gives $c = \frac{3}{8}$.
(b) Since X has a probability density function we can compute the probability in question by integrating $f(x)$ on the interval $[\frac{3}{2}, 4]$:

$$P(\tfrac{3}{2} < X < 4) = \int_{3/2}^4 f(x)dx = \int_{3/2}^2 \frac{1}{4}dx + \int_3^4 c\,dx = \frac{1}{2}\cdot\frac{1}{4} + 1\cdot c = \frac{1}{2}.$$

(c) We can compute the expectation using the formula $E[X] = \int_{-\infty}^{\infty} xf(x)dx$ and evaluating the integral using he definition of f.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_1^2 x\frac{1}{4}\,dx + \int_3^5 x \cdot c\,dx$$
$$= \frac{x^2}{8}\Big|_{x=1}^{x=2} + \frac{cx^2}{2}\Big|_{x=3}^{x=5} = \frac{4}{8} - \frac{1}{8} + \frac{\frac{3}{8}\cdot 25}{2} - \frac{2\cdot 9}{2} = \frac{27}{8}.$$

3.34. (a) Since X is discrete, we can compute $E[g(X)]$ using the following formula:

$$E[g(X)] = \sum_k P(X=k)g(k) = \frac{1}{2}g(1) + \frac{1}{3}g(2) + \frac{1}{6}g(5).$$

Thus we will certainly have $E[g(X)] = \frac{1}{3}\ln 2 + \frac{1}{6}\ln 5$ if $g(1) = 0$, $g(2) = \ln 2$, $g(5) = \ln 5$. The function $g(x) = \ln x$ satisfies these requirements, thus $E[\ln(X)] = \frac{1}{3}\ln 2 + \frac{1}{6}\ln 5$.

(b) Based on the solution of part (a) there is a function g for which $g(1) = e^t$, $g(2) = 2e^{2t}$, $g(5) = 5e^{5t}$ then

$$E[g(X)] = \tfrac{1}{2}e^t + \tfrac{2}{3}e^{2t} + \tfrac{5}{6}e^{5t}.$$

The function $g(x) = xe^{xt}$ satisfies the requirements, so

$$E[Xe^{tX}] = \tfrac{1}{2}e^t + \tfrac{2}{3}e^{2t} + \tfrac{5}{6}e^{5t}.$$

(c) We need to find a function g for which

$$E[g(X)] = \frac{1}{2}g(1) + \frac{1}{3}g(2) + \frac{1}{6}g(5) = 2.$$

There are lots of functions that satisfy this requirement. The simplest choice is the constant function $g(x) = 2$, but for example the function $g(x) = x$ also works.

3.35.

$$E[X^4] = \sum_k k^4 P(X=k) = (-2)^4 P(X=-2) + 0^4 P(X=0) + 4^4 P(X=4)$$
$$= 16 \cdot \frac{1}{16} + 256 \cdot \frac{7}{64} = 29.$$

3.36. Since X is continuous, we can compute $E[X^4]$ as follows:

$$E[X^4] = \int_{-\infty}^{\infty} x^4 f(x)dx = \int_1^2 x^4 \cdot \frac{2}{x^2}dx = \int_1^2 2x^2 dx = \frac{2x^3}{3}\Big|_{x=1}^{x=2} = \frac{14}{3}.$$

3.37. (a) The cumulative distribution function $F(x)$ is continuous everywhere (even at $x = 0$) and it is differentiable everywhere except at $x = 0$. Thus we can get the probability density function by differentiating F.

$$f(x) = F'(x) = \begin{cases} (1+x)^{-2} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

(b) We have

$$P(2 < X < 3) = F(3) - F(2) = \frac{3}{4} - \frac{2}{3} = \frac{1}{12}.$$

We could also compute this probability by evaluating the integral $\int_2^3 f(x)dx$.
(c) Using the probability density function we can write

$$\begin{aligned}
E[(1+X)^2 e^{-2X}] &= \int_0^\infty f(x)(1+x)^2 e^{-2x} dx \\
&= \int_0^\infty (1+x)^2 e^{-2x}(1+x)^{-2} dx = \int_0^\infty e^{-2x} dx \\
&= -\frac{1}{2} e^{-2x} \Big|_{x=0}^{\infty} = \frac{1}{2}.
\end{aligned}$$

3.38. (a) Since Z is continuous and the pd.f. is given, we can compute its expectation as

$$E[Z] = \int_{-\infty}^{\infty} z f(z) dz = \int_{-1}^{1} z \cdot \tfrac{5}{2} z^4 dz = \tfrac{5}{12} z^6 \Big|_{z=-1}^{z=1} = 0.$$

(b) We have

$$P(0 < Z < 1/2) = \int_0^{1/2} f(z) dz = \int_0^{1/2} \tfrac{5}{2} z^4 dz = \tfrac{1}{2} z^5 \Big|_{z=0}^{z=1/2} = \tfrac{1}{2} \left(\tfrac{1}{2}\right)^5 = \tfrac{1}{64}.$$

(c) We have

$$P\{Z < \tfrac{1}{2} \,|\, Z > 0\} = \frac{P(Z < \frac{1}{2} \text{ and } Z > 0)}{P(Z > 0)} = \frac{P(0 < Z < 1/2)}{P(Z > 0)}.$$

The numerator is $\frac{1}{64}$. The denominator is

$$P(Z > 0) = \int_0^\infty f(z) dz = \int_0^1 \frac{5}{2} z^4 dz = \frac{z^5}{2} \Big|_{z=0}^{z=1} = 1/2.$$

Thus,

$$P\{Z < \tfrac{1}{2} \,|\, Z > 0\} = \frac{\frac{1}{64}}{1/2} = \frac{1}{32}.$$

(d) Since Z is continuous and the pd.f. is given, we can compute $E[Z^n]$ for $n \geq 1$ as follows

$$\begin{aligned}
E[Z^n] &= \int_{-\infty}^{\infty} z^n f(z) dz = \int_{-1}^{1} z^n \cdot \tfrac{5}{2} z^4 dz = \int_{-1}^{1} \tfrac{5}{2} z^{n+4} dz \\
&= \tfrac{5}{2(n+5)} z^{n+5} \Big|_{z=-1}^{z=1} = \tfrac{5}{2(n+5)} \left(1^{n+5} - (-1)^{n+5}\right) \\
&= \tfrac{5}{2(n+5)} \left(1 - (-1)^{n+5}\right).
\end{aligned}$$

Note that $(-1)^{n+5} = 1$ if n is odd and $(-1)^{n+5} = 1$ if n is even. Thus

$$E[Z^n] = \begin{cases} \frac{5}{n+5}, & \text{if } n \text{ is odd} \\ 0, & \text{if } n \text{ is even.} \end{cases}$$

3.39. (a) One possible example:

$$P(X=1) = \frac{1}{3}, \quad P(X=2) = \frac{3}{4} - \frac{1}{3} = \frac{5}{12}, \quad P(X=3) = 1 - P(X=1) - P(X=2) = \frac{1}{4}.$$

Then $F(1) = P(X \le 1) = P(X = 1) = \frac{1}{3}$,

$$F(2) = P(X \le 2) = P(X = 1) + P(X = 2) = \frac{3}{4}$$

and

$$F(3) = P(X \le 3) = P(X = 1) + P(X + 2) + P(X = 3) = 1.$$

(b) There are a number of possible solutions. Here is one that can be checked easily using part (a):

$$f(x) = \begin{cases} \frac{1}{3} & 0 \le x \le 1 \\ \frac{5}{12} & 1 < x \le 2 \\ \frac{1}{4} & 2 < x \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

3.40. Here is a continuous example: let $f(x) = \frac{1}{x^2}$ for $x \ge 1$ and 0 otherwise. This is a nonnegative function with $\int_{-\infty}^{\infty} f(x)dx = 1$, thus there is a random variable X with p.d.f. f. Then the cumulative distribution function of X is given by

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0, & \text{if } x < 1 \\ \int_1^x \frac{1}{y^2} dy = 1 - 1/x, & \text{if } x \ge 1. \end{cases}$$

In particular, $F(n) = 1 - \frac{1}{n}$ for each positive integer n.

3.41. We begin by deriving the probability $F(s) = P(X \le s)$ using the law of total probability. For $s \in (3, 4)$,

$$\begin{aligned} F(s) = P(X \le s) &= \sum_{k=1}^{6} P(X \le s \,|\, Y = k)P(Y = k) = \sum_{k=1}^{3} \frac{1}{6} + \sum_{k=4}^{6} \frac{s}{k} \cdot \frac{1}{6} \\ &= \frac{1}{2} + \frac{37s}{360}. \end{aligned}$$

We can find the density function f on the interval $(3, 4)$ by differentiating this.

Thus

$$f(s) = F'(s) = \tfrac{37}{360} \quad \text{for } s \in (3, 4).$$

3.42. (a) Note that $0 \le X \le 1$ so $F_X(x) = 1$ for $x \ge 1$ and $F_X(x) = 0$ for $x < 0$. For $0 \le x < 1$ the event $\{X \le x\}$ is the same as the event that the chosen point is in the trapezoid D_x with vertices $(0, 0)$, $(x, 0)$, $(x, 2 - x)$, $(0, 2)$. The area of this trapezoid is $\frac{1}{2}(2 + 2 - x)x$, while the area of D is $\frac{(2+1)1}{2} = \frac{3}{2}$. Thus

$$P(X \le x) = \frac{\text{area}(D_x)}{\text{area}(D)} = \frac{\frac{1}{2}(2 + 2 - x)x}{\frac{3}{2}} = \frac{4x}{3} - \frac{x^2}{3}.$$

Thus

$$F_X(x) = \begin{cases} 1, & \text{if } x \ge 1 \\ \frac{4x}{3} - \frac{x^2}{3}, & \text{if } 0 \le x < 1 \\ 0, & \text{if } x < 0. \end{cases}$$

To find F_Y we first note that $0 \le Y \le 2$ so $F_Y(y) = 1$ for $y \ge 2$ and $F_Y(y) = 0$ for $y < 0$.

For $0 \le y < 1$ the event $\{Y \le y\}$ is the same as the event that the chosen point is in the rectangle with vertices $(0,0)$, $(0,y)$, $(1,y)$, $(1,0)$. The area of this rectangle is y, so in that case $P(Y \le y) = \frac{y}{\frac{3}{2}} = \frac{2y}{3}$.

If $1 \le y < 2$ then the event $\{Y \le y\}$ is the same as the event that the chosen point in the region D_y with vertices $(0,0)$, $(0,y)$, $(2-y,y)$, $(1,1)$, $(1,0)$. The area of this region can be computed for example by subtracting the area of the triangle with vertices $(2,0)$, $(0,y)$, $(2-y,y)$ from the area of D, this gives $\frac{3}{2} - \frac{(2-y)^2}{2} = 2y - \frac{y^2}{2} - \frac{1}{2}$. Thus $P(Y \le y) = \frac{2y - \frac{y^2}{2} - \frac{1}{2}}{\frac{3}{2}} = \frac{1}{3}\left(4y - y^2 - 1\right)$

Thus we have

$$F_Y(y) = \begin{cases} 1, & \text{if } y \ge 2 \\ \frac{1}{3}\left(4y - y^2 - 1\right), & \text{if } 1 \le y < 2 \\ \frac{2y}{3}, & \text{if } 0 \le y < 1 \\ 0, & \text{if } x < 0. \end{cases}$$

(b) Both cumulative distribution functions found in part (a) are continuous everywhere, and differentiable everywhere apart from maybe a couple of points. Thus we can find f_X and f_Y by differentiating F_X and F_Y:

$$f_X(x) = \begin{cases} \frac{4}{3} - \frac{2x}{3}, & \text{if } 0 \le x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{3}\left(4 - 2y\right), & \text{if } 1 \le y < 2 \\ \frac{2}{3}, & \text{if } 0 \le y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

3.43. If (a,b) is a point in the square $[0,1]^2$ then the distances from the four sides are $a, b, 1-a, 1-b$ and the minimal distance is the minimum of these four numbers. Since $\min(a, 1-a) \le 1/2$, this minimal distance is at most $1/2$ (which can be achieved at $(a,b) = (1/2, 1/2)$), and at least 0. Thus the possible values of X are from the interval $[0, 1/2]$.

(a) We would like to compute $F(x) = P(X \le x)$ for all x. Because $0 \le X \le 1/2$, we have $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1/2$.

Denote the coordinates of the randomly chosen point by A and B. If $0 \le x \le 1/2$ then the set $\{X \le x\}^c = \{X > x\}$ is the same as the set

$$\{x < A, x < 1-A, x < B, 1-x < B\} = \{x < A < 1-x, x < B < 1-x\}.$$

This is the same as the point (A,B) being in the square $(x, 1-x)^2$ which has probability $(1-2x)^2$. Hence, for $0 \le x \le 1/2$ we have

$$F(x) = P(X \le x) = 1 - P(X > x) = 1 - (1-2x)^2 = 4x - 4x^2.$$

(b) Since the cumulative distribution function $F(x)$ that we found in part (a) is continuous, and it is differentiable apart from $x = 0$, we can find $f(x)$ just by differentiating $F(x)$. This means that $f(x) = 4 - 8x$ for $0 \le x \le 1/2$ and 0 otherwise.

3.44. (a) Let s be a real number. Let $\alpha = \arctan(r) \in (-\pi/2, \pi/2)$ be the angle corresponding to the slope s, this is the number $\alpha \in (-\pi/2, \pi/2)$ with $\tan(\alpha) = s$. The event that $\{S \le s\}$ is the same as the event that the uniformly chosen point is in the circular sector corresponding to the angles $-\pi/2$ and α and radius 1. The area of this circular sector is $\alpha + \pi/2$, while the area of the half disk is π. Thus

$$F_S(s) = P(S \le s) = \frac{\alpha + \pi/2}{\pi} = \frac{1}{2} + \frac{\arctan(s)}{\pi}.$$

(b) The c.d.f. found in part (a) is differentiable everywhere, hence the p.d.f. is equal to its derivative:

$$f_S(s) = \left(\frac{1}{2} + \frac{\arctan(s)}{\pi}\right)' = \frac{1}{\pi(1+s^2)}.$$

3.45. Let (X, Y) be the uniformly chosen point, then $S = \frac{Y}{X}$. We can disregard the case $X = 0$, as the probability of this is 0.

(a) We need to compute $F(s) = P(S \le s)$ for all s.

The slope S can be any nonnegative number, but it cannot be negative. Thus $F_S(s) = P(S \le s) = 0$ if $s < 0$.

If $0 \le s \le 1$ then the points $(x, y) \in [0,1]^2$ with $y/x \le s$ are exactly the points in the triangle with vertices $(0,0)$, $(1,0)$, $(1,s)$. The area of this triangle is $s/2$, hence for $0 \le s \le 1$ we have $F_S(s) = s/2$.

If $1 < s$ then the points $(x, y) \in [0,1]^2$ with $y/x \le s$ are either in the triangle with vertices $(0,0)$, $(1,0)$, $(1,1)$ or in the triangle with vertices $(0,0)$, $(1,1)$, $(1/s, 1)$. The area of the union of these triangles is $1/2 + \frac{1}{2}(1 - 1/s) = 1 - \frac{1}{2s}$, hence for $1 < s$ we have $F_S(s) = 1 - \frac{1}{2s}$.

To summarize:

$$F(s) = \begin{cases} 0 & s < 0 \\ \frac{1}{2}s & 0 < s \le 1 \\ 1 - \frac{1}{2s} & 1 < s \end{cases}.$$

(b) Since $F(s)$ is continuous everywhere and it is differentiable apart from $s = 0$, we can get the probability density function $f(s)$ just by differentiating F. This gives

$$f(s) = \begin{cases} 0 & s < 0 \\ \frac{1}{2} & 0 < s \le 1 \\ \frac{1}{2s^2} & 1 < s \end{cases}.$$

3.46. (a) The smaller piece cannot be larger than $\ell/2$, hence $0 \le X \le \ell/2$. Thus $F_X(x) = 0$ for $x < 0$ and $F_X(x) = 1$ for $x \ge \ell/2$.

For $0 \le x < \ell/2$ the event $\{X \le x\}$ is the same as the event that the chosen point where we break the stick in two is within x of one of the end points. The set of possible locations is thus the union of two intervals of length x, hence the probability of the uniformly chosen point to be in this set is $\frac{2 \cdot x}{\ell}$. Hence for $0 \le x < \ell/2$ we have $F_X(x) = \frac{2x}{\ell}$.

To summarize

$$F_X(x) = \begin{cases} 1 & \text{for } x \ge \ell/2 \\ \frac{2x}{\ell} & \text{for } 0 \le x < \ell/2 \\ 0 & \text{for } x < 0. \end{cases}$$

(b) The c.d.f. found in part (a) is continuous everywhere, and differentiable apart from $x = \ell/2$. Hence we can find the p.d.f. by differentiating it, which gives

$$f_X(x) = \begin{cases} \frac{2}{\ell} & \text{for } 0 \le x < \ell/2 \\ 0 & \text{otherwise.} \end{cases}$$

3.47. (a) We need to find $F(x) = P(X \le x)$ for all x. The X coordinate of a point in the triangle must be between 0 and 30, so $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x \ge 30$.

For $0 \le x < 30$ then the set of points in the triangle with $X \le x$ is the triangle with vertices $(0,0)$, $(x,0)$ and $(x, \frac{2}{3}x)$. The area of this triangle is $\frac{1}{3}x^2$, while the area of the original triangle is $\frac{20\cdot 30}{2} = 300$. This means that if $0 \le x < 30$ then $F(x) = \frac{\frac{1}{3}x^2}{300} = \frac{x^2}{900}$. Thus

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{900} & 0 \le x < 30 \\ 1 & x \ge 30 \end{cases}.$$

(b) Since $F(x)$ is continuous everywhere, and it is differentiable everywhere apart from $x = 30$ we can get the probability density function as $F'(x)$. This gives

$$f(x) = \begin{cases} \frac{x}{450} & 0 \le x < 30 \\ 0 & \text{otherwise} \end{cases}.$$

(c) Since X is absolutely continuous, we can compute $E[X]$ as

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

Using the solution from part (b):

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{30} x \frac{x}{450} dx = 20.$$

3.48. Denote the distance by R. The distance from the y-axis for a point in the triangle is at most 2, hence $0 \le R \le 2$. We first compute the c.d.f. of R. For $0 < r < 2$ the event $\{R < r\}$ is the same as the event that the chosen point is in the trapezoid with vertices $(0,0)$, $(r,0)$, $(r, 1 - r/2)$, $(0,1)$. The probability of this event can be computed by taking ratios of areas:

$$F_R(r) = P(R \le r) = \frac{\text{area(trapezoid)}}{\text{area(triangle)}} = \frac{\frac{r(1+1-r/2)}{2}}{\frac{2\cdot 1}{2}} = r - \frac{r^2}{4}.$$

For $r \ge 2$ we have $F_R(r) = P(R \le r) = 1$ and for $r \le 0$ we have $F_R(r) = P(R \le r) = 0$. The found c.d.f. is continuous everywhere and differentiable apart from $r = 0$. Thus we can find the probability density function by differentiation:

$$f_R(r) = (F_R(r))' = 1 - r/2, \qquad \text{if } 0 < r < 2,$$

and $f_R(r) = 0$ otherwise.

Thus R is a continuous random variable, and we can compute its expectation by evaluating the appropriate integral:

$$E[R] = \int_{-\infty}^{\infty} r f_R(R) dr = \int_0^2 r(1 - r/2) dr = \frac{2}{3}.$$

3.49. (a) The set of possible values for X is the interval $[0, 4]$. Thus $F(x) = P(X \leq x)$ is 0 for $x < 0$ and equal to 1 for $x \geq 4$. If $0 \leq x < 4$ then the set of points (X, Y) in the triangle with $X \leq x$ is the quadrilateral formed by the vertices $(0, 0)$, $(0, 2)$, $(x, x/4)$, $(x, 2 - x/4)$. This is actually a trapezoid, and its area can be readily computed as $\frac{(2-x/2+2)x}{2} = 2x - \frac{x^2}{4}$. (Another way is to integrate the function $2 - s/2$ on $(0, x)$.) The area of the triangle is $\frac{2\cdot 4}{2} = 4$ which means that $P(x \leq x) = \frac{1}{2}x - \frac{1}{16}x^2$ for $0 \leq x < 4$.

This gives the continuous cdf

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}x - \frac{1}{16}x^2 & 0 \leq x < 4 \\ 1 & x \geq 4 \end{cases}.$$

Differentiating this gives f:

$$f(x) = \begin{cases} \frac{1}{2} - \frac{1}{8}x & 0 < x < 4 \\ 0 & \text{otherwise} \end{cases}.$$

(b) Our goal now is to compute $f(x)$ directly. Since X takes values from $[0, 4]$, we can assume $0 < x < 4$. We would like to compute the probability $P(X \in (x, x + \varepsilon))$ for a small ε. The set of points (X, Y) in the triangle with $x \leq X \leq x + \varepsilon$ is the trapezoid formed by the points $(x, x/4)$, $(x, 2 - x/4)$, $(x + \varepsilon, \frac{x+\varepsilon}{4})$, $(x + \varepsilon, 2 - \frac{x+\varepsilon}{4})$. For ε small the area of this trapezoid will be close to $\varepsilon \cdot (2 - \frac{x}{2})$ (as the trapezoid is close to a rectangle with sides ε and $2 - \frac{x}{2}$). The area of the original triangle is 4, thus, for $0 < x < 4$ we have

$$P(X \in (x, x + \varepsilon)) \approx \varepsilon \cdot \frac{2 - \frac{x}{2}}{4}$$

which means that in this case $f(x) = \frac{1}{2} - \frac{1}{8}x$. For $x \leq 0$ and $x \geq 4$ we have $f(x) = 0$.

We can now compute the cumulative distribution function $F(x)$ using the formula $F(x) = \int_{-\infty}^{x} f(y)dy$.

For $x < 0$ we have $F(x) = \int_{-\infty}^{x} f(y)dy = 0$. For $x \geq 4$ we have

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_0^4 \frac{1}{2} - \frac{1}{8}y dy = 1.$$

Finally, for $0 \leq x < 4$ we have

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_0^x \frac{1}{2} - \frac{1}{8}y dy = \frac{1}{2}x - \frac{1}{16}x^2.$$

3.50. (a) For $\varepsilon < t < 9$ the event $\{t - \varepsilon < R < t\}$ is the event that the dart lands in the annulus (or ring) with radii $t - \epsilon$ and t. The area of this annulus is

$\pi(t^2 - (t-\varepsilon)^2)$, thus the corresponding probability is

$$P(t-\varepsilon < R < t) = \frac{\pi(t^2 - (t-\varepsilon)^2)}{9^2\pi} = \frac{1}{81}(t^2 - t^2 + 2\varepsilon t - \varepsilon^2) = \frac{2}{81}\varepsilon t - \frac{\varepsilon^2}{81}.$$

Taking the limit of $\varepsilon^{-1}P(t-\varepsilon < R < t)$ as $\varepsilon \to 0$ gives $\frac{2t}{81}$ for $0 < t < 9$. This is the probability density in $(0,9)$, and since R cannot be negative or larger than 9, the p.d.f. is 0 otherwise.

(b) The argument is similar to the one presented in part (a). If $\varepsilon < t < 9-\varepsilon$ then

$$P(t-\varepsilon < R < t+\varepsilon) = \frac{\pi((t+\varepsilon)^2 - (t-\varepsilon)^2)}{81\pi} = \frac{4t\varepsilon}{81}.$$

Hence $(2\varepsilon)^{-1}P(t-\varepsilon < R < t+\varepsilon) = \frac{2t}{81}$ (we don't even need to take a limit here). Thus the probability density function of R is $\frac{2t}{81}$ on $(0,9)$ and zero otherwise.

3.51. We have

$$E(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \sum_{k=1}^{\infty}\sum_{j=1}^{k}(1-p)^{k-1}p.$$

In the last sum we are summing for k, j with $1 \le j \le k$. If we reverse the order of summation, then k will go from j to ∞, while j goes from 1 to ∞:

$$\sum_{k=1}^{\infty}\sum_{j=1}^{k}(1-p)^{k-1}p = \sum_{j=1}^{\infty}\sum_{k=j}^{\infty}(1-p)^{k-1}p.$$

For a given positive integer j we have

$$\sum_{k=j}^{\infty}(1-p)^{k-1}p = p(1-p)^{j-1}\sum_{\ell=0}^{\infty}(1-p)^{\ell} = p(1-p)^{j-1}\frac{1}{1-(1-p)} = (1-p)^{j-1}.$$

where we introduced $k = j + \ell$ and evaluated the geometric sum. This gives

$$E(X) = \sum_{j=1}^{\infty}(1-p)^{j-1} = \sum_{i=0}^{\infty}(1-p)^{i} = \frac{1}{p}.$$

3.52. Using the hint we write

$$\sum_{k=1}^{\infty} P(X \ge k) = \sum_{k=1}^{\infty}\sum_{i=k}^{\infty} P(X = i).$$

Note that in the double sum we have $1 \le k \le i$. If we switch the order of the two summations (which is allowed, since each term is nonnegative) then k goes from 1 to i, and i goes from 1 to ∞:

$$\sum_{k=1}^{\infty}\sum_{i=k}^{\infty} P(X = i) = \sum_{i=1}^{\infty}\sum_{k=1}^{i} P(X = i).$$

Since $P(X = i)$ does not depend on k, we have $\sum_{k=1}^{i} P(X = i) = iP(X = i)$ and hence

$$\sum_{k=1}^{\infty} P(X \ge k) = \sum_{i=1}^{\infty}\sum_{k=1}^{i} P(X = i) = \sum_{i=1}^{\infty} iP(X = i).$$

Because X takes only nonnegative integers we have $E[X] = \sum_{i=0}^{\infty} iP(X = i)$, and since the $i = 0$ term is equal to zero we have $E[X] = \sum_{i=1}^{\infty} iP(X = i)$. This proves $E[X] = \sum_{k=1}^{\infty} P(X \geq k)$.

3.53. (a) Since X is discrete, taking values from $0, 1, 2, \ldots$, we can compute its expectation as follows:

$$E[X] = \sum_{k=0}^{\infty} kP(X = k) = 0 \cdot \frac{3}{4} + \sum_{k=1}^{\infty} k \cdot \tfrac{1}{2} \cdot (\tfrac{1}{3})^k = \tfrac{1}{2} \sum_{k=1}^{\infty} k \cdot (\tfrac{1}{3})^k$$

The infinite sum may be computed using the identity $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$ (which holds for $|x| < 1$, and follows from $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ by differentiation):

$$\sum_{k=1}^{\infty} k \cdot (\tfrac{1}{3})^k = \tfrac{1}{3} \sum_{k=1}^{\infty} k \cdot (\tfrac{1}{3})^{k-1} = \tfrac{1}{3} \frac{1}{(1-\frac{1}{3})^2} = \frac{3}{4},$$

which gives $E[X] = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$.

Another way to arrive to this solution would be to apply the approach outlined in Exercise 3.51.

(b) To compute $\mathrm{Var}(X)$ we need $E[X^2]$. It turns out that $E[X^2 - X] = E[X(X-1)]$ is easier to compute:

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)P(X = k) = \sum_{k=2}^{\infty} k(k-1) \cdot \tfrac{1}{2} \cdot (\tfrac{1}{3})^k.$$

Next we can use that for $|x| < 1$ we have $\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{1}{(1-x)^3}$. (This follows from $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ by differentiating twice.)

$$\sum_{k=2}^{\infty} k(k-1) \cdot \tfrac{1}{2} \cdot (\tfrac{1}{3})^k = \tfrac{1}{2} \cdot (\tfrac{1}{3})^2 \sum_{k=2}^{\infty} k(k-1) \cdot (\tfrac{1}{3})^{k-2} = \tfrac{1}{18} \cdot \frac{1}{(1-\frac{1}{3})^3} = \frac{3}{8}.$$

Thus $E[X(X-1)] = \frac{3}{8}$ and hence

$$E[X^2] = E[X(X-1) + X] = E[X(X-1)] + E[X] = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}$$

and

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = 3/4 - (3/8)^2 = \frac{39}{64}.$$

3.54. (a) We have $P(X \geq k) = (1-p)^{k-1}$. We can compute this by evaluating the geometric series

$$P(X \geq k) = \sum_{\ell=k}^{\infty} P(X = \ell) = \sum_{\ell=k}^{\infty} pq^{\ell-1}.$$

An easier way is to note that if X is the number of trials needed for the first success then $\{X \geq k\}$ is the event that the first $k-1$ trials are all failures, which has probability $(1-p)^{k-1}$.

(b) By Exercise 3.52 we have

$$E[X] = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{1}{1-q} = \frac{1}{p}.$$

3.55. We first find the probability mass function of Y. The possible values are $1, 2, 3, \dots$. Peter wins the game if Y is an odd number, and Mary wins the game if it is even. If $n \geq 0$ then

$$\begin{aligned}
&P(Y = 2n+1)\\
&= P(\text{Peter misses } n \text{ times, Mary misses } n \text{ times, Peter hits bullseye next})\\
&= (1-p)^n(1-r)^n p.
\end{aligned}$$

Similarly, for $n \geq 1$:

$$\begin{aligned}
&P(Y = 2n)\\
&= P(\text{Peter misses } n \text{ times, Mary misses } n-1 \text{ times, Mary hits bullseye next})\\
&= (1-p)^n(1-r)^{n-1} r.
\end{aligned}$$

Then

$$E[Y] = \sum_{k=1}^{\infty} kP(Y=k) = \sum_{n=0}^{\infty}(2n+1)(1-p)^n(1-r)^n p + \sum_{n=1}^{\infty} 2n(1-p)^n(1-r)^{n-1} r.$$

The evaluation of these sums is a bit lengthy, but in the end one just has to use the identities $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ and $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$, which holds for $|x| < 1$. To simplify notations a little bit, we introduce $s = (1-p)(1-r)$.

$$\begin{aligned}
\sum_{n=0}^{\infty}(2n+1)(1-p)^n(1-r)^n p &= \sum_{n=0}^{\infty}(2n+1)s^n p = \sum_{n=0}^{\infty} 2ns^n p + \sum_{n=0}^{\infty} s^n p\\
&= 2sp\sum_{n=1}^{\infty} ns^{n-1} + p\sum_{n=0}^{\infty} s^n\\
&= \frac{2sp}{(1-s)^2} + \frac{p}{1-s} = \frac{p(1+s)}{(1-s)^2}.
\end{aligned}$$

$$\begin{aligned}
\sum_{n=1}^{\infty} 2n(1-p)^n(1-r)^{n-1} r &= 2(1-p)r\sum_{n=1}^{\infty} n(1-p)^{n-1}(1-r)^{n-1}\\
&= 2(1-p)r\sum_{n=1}^{\infty} ns^{n-1} = \frac{2(1-p)r}{(1-s)^2}.
\end{aligned}$$

This gives

$$E[Y] = \frac{p(1+s) + 2(1-p)r}{(1-s)^2}.$$

Substituting back $s = (1-r)(1-p) = 1 - p - r + pr$:

$$E[Y] = \frac{p(1 + (1-p)(1-r)) + 2(1-p)r}{(p+r-pr)^2} = \frac{(2-p)(p+r-pr)}{(p+r-pr)^2} = \frac{2-p}{p+r-pr}.$$

For $r = p$ the random variable Y has geometric distribution with parameter p, and our formula gives $\frac{2-p}{2p-p^2} = \frac{1}{p}$, as it should.

3.56. Using the hint we compute $E[X(X-1)]$ first. Using the formula for the expectation of a function of a discrete random variable we get

$$E[X(X-1)] = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} = pq\sum_{k=1}^{\infty} k(k-1)q^{k-2} = pq\sum_{k=0}^{\infty} k(k-1)q^{k-2}.$$

(We used that $k(k-1)=0$ for $k=0$.) Note that $k(k-1)q^{k-2} = (q^k)''$ for $k \geq 2$, and the formula also works for $k=0$ and 1.

The identity $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ holds for $|x|<1$, and differentiating both sides we get

$$\left(\frac{1}{1-x}\right)'' = \frac{2}{(1-x)^3} = \left(\sum_{k=0}^{\infty} x^k\right)'' = \sum_{k=0}^{\infty} k(k-1)x^{k-2}.$$

(We are allowed to differentiate the series term by term for $|x|<1$.) Thus for $|x|<1$ we have $\sum_{k=0}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$ and thus

$$E[X(X-1)] = pq\sum_{k=0}^{\infty} k(k-1)q^{k-2} = pq \cdot \frac{2}{(1-q)^3} = \frac{2q}{p^2},$$

where we used $p+q=1$.

Then

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

where we used $p+q=1$ again.

3.57. We have $P(X=k) = p(1-p)^{k-1}$ for $k \geq 1$. Hence we can compute $E[\frac{1}{X}]$ using the following formula:

$$E[\tfrac{1}{X}] = \sum_{k=1}^{\infty} \frac{1}{k} p(1-p)^{k-1}.$$

In order to evaluate the infinite sum, we start with the identity $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ which holds for $|x|<1$, and then integrate both sides from 0 to y with $|y|<1$:

$$\int_0^y \frac{1}{1-x}dx = \int_0^y \sum_{k=0}^{\infty} x^k dy.$$

On the left side we have $\int_0^y \frac{1}{1-x}dx = \ln(\frac{1}{1-y})$. On the right side we integrate term by term to get

$$\int_0^y \sum_{k=0}^{\infty} x^k dy = \sum_{k=0}^{\infty} \frac{y^{k+1}}{k+1} = \sum_{n=1}^{\infty} \frac{y^n}{n}.$$

This gives the identity

$$\sum_{n=1}^{\infty} \frac{y^n}{n} = \ln(\tfrac{1}{1-y})$$

for $|y| < 1$. Using this with $y = 1 - p$:

$$
\begin{aligned}
E[\tfrac{1}{X}] = \sum_{k=1}^{\infty} \frac{1}{k} p(1-p)^{k-1} = \\
\frac{p}{1-p} \sum_{k=1}^{\infty} \frac{(1-p)^k}{k} = \frac{p}{1-p} \ln(\tfrac{1}{p})
\end{aligned}
$$

3.58. Using the formula for the expected value of a function of a discrete random variable we get

$$
E[X] = \sum_{k=0}^{n} \frac{1}{k+1} \binom{n}{k} p^k (1-p)^{n-k}.
$$

We have

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

where we used $(k+1) \cdot k! = (k+1)!$.

Then

$$
\begin{aligned}
E[X] &= \sum_{k=0}^{n} \frac{1}{n+1} \binom{n+1}{k+1} p^k (1-p)^{n-k} \\
&= \frac{1}{p(n+1)} \sum_{k=0}^{n} \binom{n+1}{k+1} p^{k+1} (1-p)^{n+1-(k+1)} \\
&= \frac{1}{p(n+1)} \sum_{\ell=1}^{n+1} \binom{n+1}{\ell} p^{\ell} (1-p)^{n+1-\ell}.
\end{aligned}
$$

Adding and removing the $\ell = 0$ term to the sum and using the binomial theorem yields

$$
\begin{aligned}
E[X] &= \frac{1}{p(n+1)} \sum_{\ell=1}^{n+1} \binom{n+1}{\ell} p^{\ell} (1-p)^{n+1-\ell} \\
&= \frac{1}{p(n+1)} \left(\sum_{\ell=0}^{n+1} \binom{n+1}{\ell} p^{\ell} (1-p)^{n+1-\ell} - (1-p)^{n+1} \right) \\
&= \frac{1}{p(n+1)} (1 - (1-p)^{n+1}).
\end{aligned}
$$

3.59. (a) Using the solution for Example 1.38 we see that the following function works:

$$g(r) = \begin{cases} 10 & \text{if} \quad 0 \le r \le 1, \\ 5 & \text{if} \quad 1 < r \le 3, \\ 2 & \text{if} \quad 3 < r \le 6, \\ 1 & \text{if} \quad 6 < r \le 9, \\ 0 & \text{otherwise.} \end{cases}$$

Since $0 \le R \le 9$ we could have defined g any way we like it outside $[0, 9]$. (b) The probability mass function for X is given by

$$p_X(10) = \frac{1}{81}, \quad p_X(5) = \frac{8}{81}, \quad p_X(2) = \frac{27}{81}, \quad p_X(1) = \frac{45}{81}.$$

Thus the expectation is

$$E[X] = 10 \cdot \frac{1}{81} + 5 \cdot \frac{8}{81} + 2 \cdot \frac{27}{81} + 1 \cdot \frac{45}{81} = \frac{149}{81}$$

(c) Using the result of Example 3.19 we see that the probability density $f_R(r)$ of R is $\frac{2r}{81}$ for $0 \le< r \le 9$ and zero otherwise. We can now compute the expectation of $X = g(R)$ as follows:

$$\begin{aligned} E[X] = E[g(R)] &= \int_{-\infty}^{\infty} g(r) f_R(r) dr \\ &= \int_0^1 10 \cdot \frac{2r}{81} dr + \int_1^3 5 \cdot \frac{2r}{81} dr + \int_3^6 2 \cdot \frac{2r}{81} dr + \int_6^9 1 \cdot \frac{2r}{81} dr \\ &= \frac{149}{81}. \end{aligned}$$

3.60. (a) Let p_X be the probability mass function of X. Then

$$\begin{aligned} E[u(X) + v(X)] &= \sum_k p_X(k)(u(k) + v(k)) = \sum_k p_X(k) u(k) + \sum_k p_X(k) v(k) \\ &= E[u(X)] + E[v(X)]. \end{aligned}$$

The first step is the expectation of a function of a discrete random variable. In the second step we broke the sum into two parts. (This actually requires care in case of infinitely many terms. It is a valid step in this case because u and v are bounded and hence all the sums involved are finite.) In the last step we again used the formula for the expected value of a function of a discrete random variable.

(b) Suppose that the probability density function of X is f. Then

$$\begin{aligned} E[u(X) + v(X)] &= \int_{-\infty}^{\infty} f(x)(u(x) + v(x)) dx = \int_{-\infty}^{\infty} f(x) u(x) dx + \int_{-\infty}^{\infty} f(x) v(x) dx \\ &= E[u(X)] + E[v(X)]. \end{aligned}$$

The first step is the formula for the expectation of a function of a continuous random variable. In the second step we rewrote the integral of a sum as the sum of the integrals. (This is a valid step because u and v are bounded and thus all the integrals involved are finite.) In the last step we again used the formula for the expected value of a function of a continuous random variable.

3.61. (a) Note that the range of X is $[0, M]$. Thus, we know that

$$F_X(s) = 0 \text{ if } s < 0, \quad \text{and} \quad F(s) = 1 \text{ if } s > M.$$

Next, for $s \in [0, M]$ we have

$$F_X(s) = P(X \le s) = \int_0^s 2(M - x)/M^2 dx = \frac{2s}{M} - \frac{s^2}{M^2}.$$

(b) We have

$$Y = \begin{cases} X & \text{if } X \in [0, M/2] \\ M/2 & \text{if } X \in (M/2, M] \end{cases}.$$

(c) For $y < M/2$ we have that $\{Y \le y\} = \{X \le y\}$ and so,

$$P(Y \le y) = P(X \le y) = F_X(y) = \frac{2y}{M} - \frac{y^2}{M^2}.$$

Since $\{Y = M/2\} = \{X \ge M/2\}$ we have

$$\begin{aligned} P(Y = M/2) = P(X \ge M/2) &= 1 - P(X < M/2) \\ &= 1 - P(X \le M/2) \\ &= 1 - F_X(M/2) = 1 - \left[\frac{2(M/2)}{M} - \frac{(M/2)^2}{M^2}\right] \\ &= 1 - (1 - \frac{1}{4}) = \frac{1}{4}. \end{aligned}$$

Since Y is at most $M/2$, for $y > M/2$ we have

$$P(Y \le y) = P(Y \le M/2) = 1.$$

Putting this all together yields

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{2y}{M} - \frac{y^2}{M^2} & 0 \le y < M/2 \\ 1 & y \ge M/2 \end{cases}.$$

(d) We have

$$P(Y < M/2) = \lim_{y \to \frac{M}{2}-} F_Y(y) = \frac{3}{4}.$$

Another way to see this is by noticing that

$$P(Y < M/2) = 1 - P(Y \ge M/2) = 1 - P(Y = M/2) = 1 - \frac{1}{4} = \frac{3}{4}.$$

(e) Y cannot be continuous, as $P(Y = M/2) = \frac{1}{4} > 0$. But it cannot be discrete either, as there are no other values which Y takes with positive probability. Thus there is no density, nor is there a probability mass function.

3.62. From the set-up we know $F(s) = 0$ for $s < 0$ because negative values have no probability and $F(s) = 1$ for $s \ge 3/4$ because the boy is sure to be inside by time

3/4. For values $0 \le s < 3/4$ the probability $P(X \le s)$ comes from the uniform distribution and hence equals s, the length of the interval $[0, s]$. To summarize,

$$F(s) = \begin{cases} 0, & s < 0 \\ s, & 0 \le s < 3/4 \\ 1, & s \ge 3/4. \end{cases}$$

In particular, we have a jump in F that gives the probability for the value 3/4:

$$P(X = \tfrac{3}{4}) = F(\tfrac{3}{4}) - F\big((\tfrac{3}{4})-\big) = 1 - \tfrac{3}{4} = \tfrac{1}{4}.$$

This reflects the fact that, left to his own devices, the boy would come in after time 3/4 with probability 1/4. This option is removed by the mother's call and so all this probability concentrates on the value 3/4.

3.63. (a) We have $E[X] = \sum_k k p_X(k)$. Because X is symmetric, we must have $P(X = k) = P(X = -k)$ for all k. Thus we can write the sum as

$$E[X] = \sum_k k p_X(k) = 0 \cdot p_X(0) + \sum_{k>0} k p_X(k) + (-k) p_X(-k) = \sum_{k>0} k(p_X(k) - p_X(-k)) = 0$$

since each term is 0.
(b) The solution is similar in the continuous case. We have

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x f(x) dx + \int_{-\infty}^{0} x f(x) dx \\ &= \int_0^{\infty} x f(x) dx + \int_0^{\infty} -x f(-x) dx \\ &= \int_0^{\infty} x(f(x) - f(-x)) dx = 0. \end{aligned}$$

3.64. For the continuous random variable first recall that $\int_1^\infty \frac{1}{x^\alpha} dx = \infty$ if $\alpha \le 1$, and $\int_1^\infty \frac{1}{x^\alpha} dx = \frac{1}{\alpha - 1} < \infty$ if $\alpha > 1$.

Now set

$$f(x) = \begin{cases} \frac{2}{x^3}, & \text{if } x \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since $f(x) \ge 0$ and $\int_{-\infty}^{\infty} f(x) dx = 2 \int_1^\infty \frac{2}{x^3} dx = 1$, the function f is a probability density function. Let X be a continuous random variable with probability density function equal to f. Then

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x^k f(x) dx = \int_1^\infty x \frac{2}{x^3} dx = 2 \int_1^\infty \frac{1}{x^2} dx = 2 < \infty \\ E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_1^\infty x^2 \frac{3}{x^3} dx = 3 \int_1^\infty \frac{1}{x} dx = \infty. \end{aligned}$$

For the discrete example recall that $\sum_{k=1}^\infty \frac{1}{k^\alpha} < \infty$ if $\alpha > 1$ and $\sum_{k=1}^\infty \frac{1}{k^\alpha} = \infty$ for $\alpha \le 1$. Consider the discrete random variable X with probability mass function

$$P(X = k) = \frac{C}{k^3}, \qquad k = 1, 2, \ldots$$

with $C = \frac{1}{\sum_{k=1}^\infty \frac{1}{k^3}}$. Since $0 < \sum_{k=1}^\infty \frac{1}{k^3} < \infty$, this is indeed a probability mass function. Moreover, we have

$$E[X] = \sum_{k=1}^\infty kP(X = k) = \sum_{k=1}^\infty k \cdot \frac{C}{k^3} = C \sum_{k=1}^\infty \frac{1}{k^2} < \infty.$$

and

$$E[X^2] = \sum_{k=1}^\infty k^2 P(X = k) = \sum_{k=1}^\infty k^2 \cdot \frac{C}{k^3} = C \sum_{k=1}^\infty \frac{1}{k} = \infty.$$

3.65. (a) We have $\mathrm{Var}(2X + 1) = 2^2 \mathrm{Var}(X) = 4 \cdot 3 = 12$.
(b) We have

$$E[(3X - 4)^2\,] = E[9X^2 - 24X + 16] = 9E[X^2] - 24E[X] + 16.$$

We know that $\mathrm{Var}(X) = E[X^2] - E[X]^2$, so $E[X^2] = \mathrm{Var}(X) + E[X]^2 = 3 + 2^2 = 7$. Thus

$$E[(3X - 4)^2\,] = 9E[X^2] - 24E[X] + 16 = 9 \cdot 7 - 24 \cdot 2 + 16 = 31.$$

3.66. We can express X as $X = \sqrt{3}Y + 8$ where $Y \sim \mathcal{N}(0, 1)$. Then

$$0.15 = P(X > \alpha) = P(\sqrt{3}Y + 8 > \alpha) = P(Y > \tfrac{\alpha-8}{\sqrt{3}}) = 1 - \Phi(\tfrac{\alpha-8}{\sqrt{3}}).$$

Using the table in Appendix E we get that if $\Phi(\frac{\alpha-8}{\sqrt{3}}) = 0.85$ then $\frac{\alpha-8}{\sqrt{3}} \approx 1.04$. From this we get

$$\alpha \approx \sqrt{3}1.04 + 8 \approx 9.8.$$

3.67. (a) We have

$$E[Z^3] = \int_{-\infty}^\infty x^3 \varphi(x) dx = \int_{-\infty}^\infty x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Note that the function $g(x) = x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is odd: $g(x) = g(-x)$. Thus if the integral is finite then it must be equal to 0, as the values on the positive and negative half lines cancel each other out. The fact that the integral is finite follows from the fact that x^3 grows a lot slower than $e^{\frac{x^2}{2}}$. (Or you can evaluate the integral on the positive and negative half lines separately by integration by parts.)
(b) We can express X as $X = \sigma Y + \mu$ where $Y \sim \mathcal{N}(0, 1)$. Then

$$\begin{aligned} E[X^3] &= E[(\sigma Y + \mu)^3] = E[\sigma^3 Y^3 + 3\sigma^2 \mu Y^2 + 3\sigma\mu^2 Y + \mu^3] \\ &= \sigma^3 E[Y^3] + 3\sigma^2 \mu E[Y^2] + 3\sigma\mu^2 E[Y] + \mu^3. \end{aligned}$$

We have $E[Y] = E[Y^3] = 0$ and $E[Y^2] = 1$. Thus

$$E[X^3] = \sigma^3 E[Y^3] + 3\sigma^2 \mu E[Y^2] + 3\sigma\mu^2 E[Y] + \mu^3 = 3\sigma^2\mu + \mu^3$$

3.68. (a) Since the p.d.f. of Z is $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, we have

$$E[Z^4] = \int_{-\infty}^\infty \varphi(x) x^4 dx = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} x^4 dx.$$

We can evaluate the integral using integration by parts noting that $e^{-\frac{x^2}{2}}x = (-e^{-\frac{x^2}{2}})'$:

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}x^4dx &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}x\cdot x^3dx\\ &= \frac{1}{\sqrt{2\pi}}(-e^{-\frac{x^2}{2}})\cdot x^3\big|_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}(-e^{-\frac{x^2}{2}})\cdot 3x^2\\ &= 3\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}x^2dx = 3.\end{aligned}$$

We used that $\lim_{x\to\infty} e^{-\frac{x^2}{2}}x^3 = 0$ (and the same for $x\to-\infty$), and that $\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}x^2dx = E[Z^2] = 1$.

Hence $E[Z^4] = 3$.

(b) We can express X as $X = \sigma Y + \mu$ where $Y \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned}E[X^4] &= E[(\sigma Y+\mu)^4]\\ &= E[\sigma^4Y^4 + 4\sigma^3\mu Y^3 + 6\sigma^2\mu^2Y^2 + 4\sigma Y\mu^3 + \mu^4]\\ &= \sigma^4E[Y^4] + 4\sigma^3\mu E[Y^3] + 6\sigma^2\mu^2E[Y^2] + 4\sigma\mu^3E[Y] + \mu^4.\end{aligned}$$

We know that $E[Y] = 0$, $E[Y^2] = 1$. By part (a) we have $E[Y^4] = 3$ and by the previous problem we have $E[Y^3] = 0$. Substituting these in the previous expression we get

$$E[X^4] = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4.$$

3.69. Denote the nth moment $E[Z^n]$ by m_n. It can be computed as

$$m_n = \int_{-\infty}^{\infty}x^n\varphi(x)dx = \int_{-\infty}^{\infty}x^n\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

We have seen that $m_1 = E[Z] = 0$ and $m_2 = E[Z^2] = 1$.

Suppose first that $n = 2k+1$ is an odd number. Then the function x^{2k+1} is odd and hence the function $x^{2k+1}\varphi(x)$ is odd as well. If the integral is finite then the contribution of the positive and negative half lines in $\int_{-\infty}^{\infty}x^{2k+1}\varphi(x)dx$ cancel each other out and thus $m_{2k+1} = 0$. The fact that the integral is finite follows from the fact that for any fixed n x^n grows a lot slower than $e^{\frac{x^2}{2}}$.

For $n = 2k \geq 2$ we see that $x^n\varphi(x)$ is even, and thus (if the integrals are finite) we have

$$m_{2k} = \int_{-\infty}^{\infty}x^{2k}\varphi(x)dx = 2\int_0^{\infty}x^{2k}\varphi(x)dx$$

Using integration by parts with the functions x^{2k-1} and $x\varphi(x) = \left(-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)' = (-\varphi(x))'$ we get

$$\begin{aligned}\int_0^{\infty}x^{2k}\varphi(x)dx &= -x^{2k-1}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Big|_{x=0}^{x=\infty} + \int_0^{\infty}(2k-1)x^{2k-2}\varphi(x)dx\\ &= (2k-1)\int_0^{\infty}x^{2k-2}\varphi(x)dx.\end{aligned}$$

Here the boundary term at ∞ disappears because $x^n e^{-\frac{x^2}{2}} \to 0$ for any $n \geq 0$ as $x \to \infty$. The integration by parts reduced the exponent of x by 2, and multiplying both sides by 2 gives

$$m_{2k} = (2k-1)m_{2k-2}.$$

Repeating this step we get

$$\begin{aligned} m_{2k} &= (2k-1)m_{2k-2} = (2k-1)(2k-3)m_{2k-4} = \cdots = (2k-1)(2k-3)\cdots 3m_2 \\ &= (2k-1)(2k-3)\cdots 1. \end{aligned}$$

The final answer is the product of positive odd numbers not larger than $2k$, which is sometimes denoted by $(2k-1)!!$. It can also be computed as

$$(2k-1)(2k-3)\cdots 1 = \frac{2k\cdot(2k-1)\cdot(2k-2)\cdots 2\cdot 1}{(2k)(2k-2)\cdots 2} = \frac{(2k)!}{2^k\cdot k(k-1)\cdots 1} = \frac{(2k)!}{2^k k!}.$$

Thus we get

$$m_n = E[Z^n] = \begin{cases} 0, & \text{if } n = 2k+1 \\ \frac{(2k)!}{2^k k!}, & \text{if } n = 2k. \end{cases}$$

3.70. We assume $a \neq 0$, otherwise Y is not random.

We have seen in (3.42) that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $F_X(x) = P(X \leq x) = \Phi(\frac{x-\mu}{\sigma})$. Let us compute the cumulative distribution function of $Y = aX + b$. We have

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y).$$

If $a > 0$ then

$$F_Y(y) = P(aX + b \leq y) = P(X \leq \tfrac{y-b}{a}) = F_X(\tfrac{y-b}{a}) = \Phi\left(\frac{\frac{y-b}{a} - \mu}{\sigma}\right).$$

We have

$$\frac{\frac{y-b}{a} - \mu}{\sigma} = \frac{y - (a\mu + b)}{a\sigma}$$

thus $F_Y(y) = \Phi\left(\frac{y-(a\mu+b)}{a\sigma}\right)$. By (3.42) this is exactly the c.d.f. of a $\mathcal{N}(a\mu+b,\, a^2\sigma^2)$ distributed random variable, so $Y \sim \mathcal{N}(a\mu + b,\, a^2\sigma^2)$.

If $a < 0$ then

$$F_Y(y) = P(aX + b \leq y) = P(X \geq \tfrac{y-b}{a}) = 1 - F_X(\tfrac{y-b}{a}) = 1 - \Phi\left(\frac{\frac{y-b}{a} - \mu}{\sigma}\right).$$

Using $1 - \Phi(x) = \Phi(-x)$ and the computation above we get

$$F_Y(y) = \Phi\left(-\frac{\frac{y-b}{a} - \mu}{\sigma}\right) = \Phi\left(\frac{y-(a\mu+b)}{(-a)\sigma}\right) = \Phi\left(\frac{y-(a\mu+b)}{|a|\sigma}\right).$$

This is exactly the c.d.f. of a $\mathcal{N}(a\mu + b,\, a^2\sigma^2)$ distributed random variable, so $Y \sim \mathcal{N}(a\mu + b,\, a^2\sigma^2)$ in this case as well.

3.71. We define noon to be time zero. Let $X \sim \text{N}(0,36)$ model the arrival time of the bus in minutes (since the standard deviation is 6). Thus, $X = 6Z$ where $Z \sim \text{N}(0,1)$. The question is then:

$$\begin{aligned} P(X > 5) &= P(6Z > 5) = P(Z > 5/6) \\ &= 1 - \Phi(0.83) \approx 1 - 0.7967 = 0.2033. \end{aligned}$$

3.72. Define the random variable X as the number of points made on one swing of an axe. Note that X is a discrete random variable taking values $\{0, 5, 10, 15\}$ and its expected value can be computed as

$$E[X] = \sum_k kP(X = k) = 0P(X = 0) + 5P(X = 5) + 10P(X = 10) + 15P(X = 15).$$

From the point system given in the problem we have

$$\begin{aligned} P(X = 5) =& P(-20 \le Y \le -10) + P(10 \le Y \le 20) = 2P(10 \le Y \le 20) \\ P(X = 10) =& P(-10 \le Y \le -3) + P(3 \le Y \le 10) = 2P(3 \le Y \le 10) \\ P(X = 15) =& P(-3 \le Y \le 3) = 2P(0 \le Y \le 3). \end{aligned}$$

Since $Y \sim \mathcal{N}(0, 100)$ the random variable $Z = \frac{Y}{\sqrt{100}} = \frac{Y}{10}$ has standard normal distribution. Hence

$$\begin{aligned} P(X = 5) =& 2P(1 \le Z \le 2) = 2(\Phi(2) - \Phi(1)) \approx 2(.9772 - .8413) \approx 0.2718 \\ P(X = 10) =& 2P(0.3 \le Z \le 1) = 2(\Phi(1) - \Phi(0.3)) \approx 2(.8413 - 0.6179) \approx 0.4468 \\ P(X = 15) =& 2P(0 \le Z \le 0.3) = 2\Phi(0.3) - 1 \approx 2(0.6179) - 1 \approx 0.2358. \end{aligned}$$

Thus the expected value of X is

$$\begin{aligned} E[X] =& 0P(X = 0) + 5P(X = 5) + 10P(X = 10) + 15P(X = 15) \\ \approx& 5(0.2718) + 10(0.4468) + 15(0.2358) = 9.364. \end{aligned}$$

3.73. The answer is no. Although $xf_Y(x)$ is an odd function, which suggests that $E[Y] = \int_{-\infty}^{\infty} xf_Y(x)dx = 0$, this is incorrect. The problem is that $\int_0^{\infty} xf_Y(x)dx = \infty$ and $\int_{-\infty}^{0} xf_Y(x)dx = -\infty$ and hence the integral on $(-\infty, \infty)$ is not defined.

3.74. There are lots of ways to construct such a random variable. Here we will use the fact that $\int_1^{\infty} \frac{1}{x^\alpha} dx = \infty$ if $\alpha \le 1$, and $\int_1^{\infty} \frac{1}{x^\alpha} dx = \frac{1}{\alpha - 1} < \infty$ if $\alpha > 1$.

Now let

$$f(x) = \begin{cases} \frac{k+1}{x^{k+2}}, & \text{if } x \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since $f(x) \ge 0$ and $\int_{-\infty}^{\infty} f(x)dx = (k+1)\int_1^{\infty} \frac{k+1}{x^{k+2}} dx = 1$, the function f is a probability density function. Let X be a continuous random variable with probability density function equal to f. Then

$$\begin{aligned} E[X^k] &= \int_{-\infty}^{\infty} x^k f(x)dx = \int_1^{\infty} x^k \frac{k+1}{x^{k+2}} dx = (k+1)\int_1^{\infty} \frac{1}{x^2} dx = k+1 < \infty \\ E[X^{k+1}] &= \int_{-\infty}^{\infty} x^{k+1} f(x)dx = \int_1^{\infty} x^{k+1} \frac{k+1}{x^{k+2}} dx = (k+1)\int_1^{\infty} \frac{1}{x} dx = \infty. \end{aligned}$$

Solutions to Chapter 4

4.1. Let S be the number of students born in January. Then S is distributed as $\text{Bin}(1200, p)$, where p is the probability of a birthday being in January. We use the normal approximation for $P(S > 130)$:

$$P(S > 130) = P\left(\frac{S - 1200 \cdot p}{\sqrt{1200p(1-p)}} > \frac{130 - 1200 \cdot p}{\sqrt{1200p(1-p)}}\right) \approx 1 - \Phi\left(\frac{130 - 1200 \cdot p}{\sqrt{1200p(1-p)}}\right).$$

(a) Here $p = \frac{1}{12}$, and we get

$$P(S > 130) \approx 1 - \Phi\left(\frac{130 - 1200 \cdot p}{\sqrt{1200p(1-p)}}\right) \approx 1 - \Phi(3.13) \approx 0.0009.$$

(b) Here $p = \frac{31}{365}$, and we get

$$P(S > 130) \approx 1 - \Phi\left(\frac{130 - 1200 \cdot p}{\sqrt{1200p(1-p)}}\right) \approx 1 - \Phi(2.91) \approx 0.0018.$$

4.2. Let S be the number of hands with a single pair that are observed in 1000 poker hands. Then $S \sim \text{Bin}(n, p)$ where $n = 1000$ and p is the probability of getting a single pair in a poker hand of 5 cards. We take $p = 0.42$, which is the approximate success probability given in the exercise.

To approximate $P(S \geq 450)$ we use the normal approximation. With $p = 0.42$, $np(1-p) = 243.6$ so we can feel confident about using this method.

We have $E[S] = np = 420$ and $\sqrt{\text{Var}(S)} = \sqrt{243.6}$. Then

$$\begin{aligned} P(S \geq 450) &= P\left(\frac{S - 420}{\sqrt{243.6}} \geq \frac{450 - 420}{\sqrt{243.6}}\right) \\ &\approx P\left(\frac{S - 420}{\sqrt{243.6}} \geq 1.92\right) \approx P(Z \geq 1.92), \end{aligned}$$

where $Z \sim N(0,1)$. Hence,

$$P(S \geq 450) \approx P(Z \geq 1.92) = 1 - \Phi(1.92) \approx 1 - 0.9726 = 0.0274$$

4.3. Let S be the number of die rolls that are multiples of 3, that is, 3 or 6. Then $S \sim \text{Bin}(n,p)$ with $n = 300$ and $p = \frac{1}{3}$. We need to approximate $P(S = 100)$ for which we use the normal approximation with continuity correction.

$$\begin{aligned} P(S = 100) = P(99.5 \leq S \leq 100.5) &= P\left(-\frac{0.5}{\sqrt{200/3}} \leq \frac{S-100}{\sqrt{200/3}} \leq \frac{0.5}{\sqrt{200/3}}\right) \\ &\approx \Phi\left(\frac{0.5}{\sqrt{200/3}}\right) - \Phi\left(-\frac{0.5}{\sqrt{200/3}}\right) = 2\Phi\left(\frac{0.5}{\sqrt{200/3}}\right) - 1 \\ &\approx 2\Phi(0.06) - 1 \approx 0.0478. \end{aligned}$$

4.4. Let S_n be the number of times the roll is 3, 4, 5 or 6 in the first n rolls. Then $X_n = 2S_n + (n - S_n) = S_n + n$ and $S_n \sim \text{Bin}(n, \frac{2}{3})$. We have $E(S_{90}) = 60$ and $\text{Var}(S_{90}) = 90 \cdot \frac{2}{3} \cdot \frac{1}{3} = 20$. Then normal approximation gives

$$\begin{aligned} P(X_{90} \geq 160) = P(S_{90} \geq 70) &= P\left(\frac{S_{90}-60}{\sqrt{20}} \geq \frac{70-60}{\sqrt{20}}\right) = P\left(\frac{S_{90}-60}{\sqrt{20}} \geq \sqrt{5}\right) \\ &\approx 1 - \Phi(2.24) \approx 1 - 0.9875 = 0.0125. \end{aligned}$$

4.5. $X_n = 2S_n + (n - S_n) = S_n + n$ and $S_n \sim \text{Bin}(n, \frac{2}{3})$.

(a) Use below the inequality $\frac{2}{3} - 0.6 \geq 0.05$.

$$\begin{aligned} \lim_{n\to\infty} P(X_n > 1.6n) &= P(S_n > 0.6n) = P(S_n - \tfrac{2}{3}n > -(\tfrac{2}{3} - 0.6)n) \\ &\geq P(S_n - \tfrac{2}{3}n > -0.05n) \geq P(\left|S_n - \tfrac{2}{3}n\right| < 0.05n) \to 1 \end{aligned}$$

where the last limit is from the LLN.

(b) This time use $0.7 - \frac{2}{3} > 0.03$.

$$\begin{aligned} \lim_{n\to\infty} P(X_n > 1.7n) &= P(S_n > 0.7n) = P(S_n - \tfrac{2}{3}n > (0.7 - \tfrac{2}{3})n) \\ &\leq P(S_n - \tfrac{2}{3}n > 0.03n) \leq P(\left|S_n - \tfrac{2}{3}n\right| > 0.03n) \to 0. \end{aligned}$$

The last limit comes from taking complements in the LLN.

4.6. Let n be the size of the sample and S_n the number of positive answers in the sample. Then $\widehat{p} = \frac{S_n}{n}$ and we need $P(|\widehat{p} - p| \leq 0.02) \geq 0.95$.

We have seen in Section 4.3 that $P(|\widehat{p} - p| > \varepsilon)$ can be approximated as

$$\begin{aligned} P(|\hat{p} - p| < \varepsilon) &= P(-\varepsilon < \hat{p} - p < \varepsilon) = P(-\varepsilon < \frac{S_n - np}{n} < \varepsilon) \\ &= P(-\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} < \frac{S_n - np}{\sqrt{n}\sqrt{p(1-p)}} < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) \\ &\approx 2\Phi(\tfrac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) - 1. \end{aligned}$$

Moreover, since $\sqrt{p(1-p)} \leq 1/2$, we have the bound

$$P(|\hat{p} - p| < \varepsilon) \geq 2\Phi(2\varepsilon\sqrt{n}) - 1.$$

Here we have $\varepsilon = 0.02$ and need $2\Phi(2\varepsilon\sqrt{n}) - 1 \geq 0.95$. This leads to $\Phi(2\varepsilon\sqrt{n}) \geq 0.975$ which, by the table of Φ-values, is satisfied if $2\varepsilon\sqrt{n} \geq 1.96$. Solving this inequality gives

$$n \geq \frac{1.96^2}{4\varepsilon^2} = 2401.$$

Thus the size of the sample should be at least 2401.

4.7. Now $n = 1,000$ and take $S_n \sim \text{Bin}(n, p)$, where p is unknown. We estimate p with $\hat{p} = S_n/1000 = 457/1000 = .457$. For the 95% confidence interval we need to find $\varepsilon > 0$ such that

$$P(|\hat{p} - p| < \varepsilon) \geq 0.95.$$

Then the confidence interval is $(0.457 - \varepsilon, 0.457 + \varepsilon)$.

Repeating again the normal approximation procedure: gives

$$\begin{aligned} P(|\hat{p} - p| < \varepsilon) &= P(-\varepsilon < \hat{p} - p < \varepsilon) = P(-\varepsilon < \frac{S_n - np}{n} < \varepsilon) \\ &= P(-\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} < \frac{S_n - np}{\sqrt{n}\sqrt{p(1-p)}} < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) \\ &\approx 2\Phi(\tfrac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) - 1. \end{aligned}$$

Note that $\sqrt{p(1-p)} \leq 1/2$ on the interval $[0, 1]$, from which we conclude that

$$2\Phi(\tfrac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) - 1 \geq 2\Phi(2\varepsilon\sqrt{n}) - 1,$$

and so

$$P(|\hat{p} - p| < \varepsilon) \geq 2\Phi(2\varepsilon\sqrt{n}) - 1.$$

Hence, we just need to find $\epsilon > 0$ satisfying

$$2\Phi(2\varepsilon\sqrt{n}) - 1 = 0.95 \implies \Phi(2\varepsilon\sqrt{n}) = 0.975 \implies 2\varepsilon\sqrt{n} \approx 1.96.$$

Thus, take

$$\varepsilon = \frac{1.96}{2\sqrt{1000}} \approx 0.031$$

and the confidence interval is

$$(0.457 - 0.031, 0.457 + 0.031).$$

4.8. We have $n =$1,000,000 trials with an unknown success probability p. To find a 99.9% confidence interval we need an $\varepsilon > 0$ so that $P(|\hat{p} - p| < \varepsilon) \geq 0.999$, where $\hat{p}$ is the fraction of positive outcomes. We have seen in Section 4.3 that $P(|\hat{p} - p| < \varepsilon)$ can be estimated using the normal approximation as

$$P(|\hat{p} - p| < \varepsilon) \approx 2\Phi(\tfrac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) - 1 \geq 2\Phi(2\varepsilon\sqrt{n}) - 1.$$

We need $2\Phi(2\varepsilon\sqrt{n}) - 1 \geq 0.999$ which means $\Phi(2\varepsilon\sqrt{n}) \geq 0.9995$ and so approximately $2\varepsilon\sqrt{n} \geq 3.32$. (Since 0.9995 appears several times in our table, other values instead of 3.32 are also acceptable.) This gives

$$\varepsilon \geq \frac{3.32}{2\sqrt{n}} \approx 0.00166$$

with n =1,000,000. We had 180,000 positive outcomes, so $\hat{p} = 0.18$. Thus our confidence interval is $(0.18 - 0.00166, 0.18 + 0.00166) = (0.17834, 0.18166)$.

If we choose 3.28 from the table for the solution of $\Phi(x) = 0.9995$ then we get $(0.17836, 0.18164)$ instead.

4.9. If $X \sim \text{Poisson}(\lambda)$ with $\lambda = 10$ then

$$P(X \geq 7) = 1 - \sum_{k=0}^{6} P(X = k) = 1 - \sum_{k=0}^{6} \frac{\lambda^k}{k!} e^{-\lambda} \approx 0.8699,$$

and

$$\begin{aligned} P(X \leq 13 \,|\, X \geq 7) &= \frac{P(X \leq 13 \text{ and } X \geq 7)}{P(X \geq 7)} = \frac{\sum_{k=7}^{13} \frac{\lambda^k}{k!} e^{-\lambda}}{1 - \sum_{k=0}^{6} \frac{\lambda^k}{k!} e^{-\lambda}} \\ &\approx \frac{0.7343}{0.8699} \approx 0.844. \end{aligned}$$

4.10. It is reasonable to assume that the hockey player has a number of scoring chances per game, but only a few of them result in goals. Hence the number of goals in a given game corresponds to counting rare events, which means that it is reasonable to approximate this random number with a Poisson(λ) distributed random variable. Then the probability of scoring at least one goal would be $1 - e^{-\lambda}$ (since $e^{-\lambda}$ is the probability of no goals). Using the setup of the problem we have $1 - e^{-\lambda} \approx 0.5$ which gives $\lambda \approx \ln(2) \approx 0.6931$. We estimate the probability that the player scores exactly 3 goals. Using the Poisson probability mass function and our estimate on λ gives

$$P(\text{exactly 3 goals}) = \frac{\lambda^3}{3!} e^{-\lambda} \approx 0.028.$$

Thus we would expect the player to get a hat-trick in about 2.8% of his games.

Equally valid is the answer where we estimate the probability of scoring at least 3 goals:

$$\begin{aligned} P(\text{at least 3 goals}) = 1 - P(\text{at most 2 goals}) &= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2!} e^{-\lambda} \\ &= 1 - \tfrac{1}{2}\big(1 + \ln 2 + \tfrac{1}{2}(\ln 2)^2\big) \approx 0.033. \end{aligned}$$

Both calculations give the answer of roughly 3 percent.

4.11. We assume that typos are rare events that do not strongly depend on each other. Hence the number of typos on a given page should be well-approximated by a Poisson random variable with parameter $\lambda = 6$, since that is the average number of typos per page.

Let X be the number of errors on page 301. We now have

$$P(X \geq 4) = 1 - P(X \leq 3) \approx 1 - \sum_{k=0}^{3} e^{-6} \frac{6^k}{k!} = 0.8488.$$

4.12. The probability density function $f_T(x)$ of T is $\lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise. Thus $E[T^3]$ can be evaluated as

$$E[T^3] = \int_{-\infty}^{\infty} f_T(x) x^3 dx = \int_0^{\infty} \lambda x^3 e^{-\lambda x} dx.$$

To compute the integral we use integration by parts with $\lambda e^{-\lambda x} = (-e^{-\lambda x})'$:

$$\int_0^{\infty} \lambda x^3 e^{-\lambda x} dx = -x^3 e^{-\lambda x}\Big|_{x=0}^{x=\infty} - \int_0^{\infty} 3x^2(-e^{-\lambda x}) dx = \int_0^{\infty} 3x^2 e^{-\lambda x} dx.$$

Note that $-x^3 e^{-\lambda x}\big|_{x=0}^{x=\infty} = 0$ because $\lim_{x\to\infty} x^3 e^{-\lambda x} = 0$. To evaluate $\int_0^\infty 3x^2 e^{-\lambda x} dx$ we can integrate by parts twice more, or we can quote equation (4.18) from the text to get

$$\int_0^{\infty} 3x^2 e^{-\lambda x} dx = \frac{3}{\lambda} \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{3}{\lambda} \cdot \frac{2}{\lambda^2} = \frac{6}{\lambda^3}.$$

Thus $E[T^3] = \frac{6}{\lambda^3}$.

4.13. The probability density function of T is $f_T(x) = \frac{1}{3} e^{-\frac{1}{3}x}$ for $x \geq 0$, and zero otherwise. The cumulative distribution function is $F_T(x) = 1 - e^{-\frac{1}{3}x}$ for $x \geq 0$, and zero otherwise. From this we can compute

$$\begin{aligned}
P(T > 3) &= 1 - F_T(3) = e^{-1}, \\
P(1 \leq T < 8) &= F_T(8) - F_T(1) = e^{-1/3} - e^{-8/3}, \\
P(T > 4 \,|\, T > 1) &= \frac{P(T > 4 \text{ and } T > 1)}{P(T > 1)} = \frac{P(T > 4)}{P(T > 1)} \\
&= \frac{1 - F_T(4)}{1 - F_T(1)} = \frac{e^{-4/3}}{e^{-1/3}} = e^{-1}.
\end{aligned}$$

$P(T > 4 \,|\, T > 1)$ can also be computed using the memoryless property of the exponential:

$$P(T > 4 \,|\, T > 1) = P(T > 3) = 1 - F_T(3) = e^{-1}.$$

4.14. (a) Denote the lifetime of the lightbulb by T. Since T is exponentially distributed with expected value 1000 we have $T \sim \text{Exp}(\lambda)$ with $\lambda = \frac{1}{1000}$. The cumulative distribution function of T is then $F_T(t) = 1 - e^{-\lambda t}$ for $t > 0$ and 0 otherwise. Hence

$$P(T > 2000) = 1 - P(T \leq 2000) = 1 - F_T(2000) = e^{-2000\cdot\lambda} = e^{-2}.$$

(b) We need to compute $P(T > 2000|T > 500)$ where we used the notation of part (a). By the memoryless property $P(T > 2000|T > 500) = P(T > 1500)$. Using the steps in part (a) we get

$$P(T > 1500) = 1 - F_T(1500) = e^{-1500\cdot\lambda} = e^{-\frac{3}{2}}.$$

4.15. Let N be the Poisson process of arrival times of meteors. Let 11 PM correspond to the origin on the time line.

(a) Using the fact that $N([0,1])$, the number of meteors within the first hour, has Poisson(4) distribution, we get

$$\begin{aligned} P(N([0,1]) > 2) &= 1 - \sum_{k=0}^{2} P(N([0,1] = k) \\ &= 1 - \sum_{k=0}^{2} e^{-4}\frac{4^k}{k!} \approx 0.7619. \end{aligned}$$

(b) Using the independent increment property we get that $N([0,1])$ and $N([1,4])$ are independent. Moreover, $N([0,1]) \sim \text{Poisson}(4)$ and $N([1,4]) \sim \text{Poisson}(3 \cdot 4)$, which gives

$$\begin{aligned} P(N([0,1]) = 0, N([1,4]) \geq 10) &= P(N([0,1]) = 0) \cdot P(N([1,4]) \geq 10) \\ &= P(N([0,1]) = 0) \cdot (1 - P(N([1,4]) < 10)) \\ &= e^{-4} \cdot \left(1 - \sum_{k=0}^{9} e^{-12}\frac{12^k}{k!}\right) \\ &\approx 0.01388. \end{aligned}$$

(c) Using the independent increment property again:

$$\begin{aligned} P(N([0,1]) = 0 \mid N([0,4]) = 13) &= \frac{P(N([0,1]) = 0, N([0,4]) = 13)}{P(N([0,4]) = 13)} \\ &= \frac{P(N([0,1]) = 0, N([1,4]) = 13)}{P(N([0,4]) = 13)} \\ &= \frac{P(N([0,1]) = 0) \cdot P(N([1,4]) = 13)}{P(N([0,4]) = 13)} \\ &= \frac{e^{-4} \cdot e^{-12}12^{13}/13!}{e^{-16}16^{13}/13!} \\ &= \left(\frac{3}{4}\right)^{13} \\ &\approx 0.02376. \end{aligned}$$

4.16. (a) Denote by S the number of random numbers starting with the digit 1. Note that a number in the interval $[1.5, 4.8]$ starts with 1 if and only if it is in the interval $[1.5, 2)$. The probability that a uniformly chosen number from the interval $[1.5, 4.8]$ is in $[1.5, 2)$ is equal to $p = \frac{0.5}{4.8-1.5} = \frac{5}{33}$. Assuming that the 500 numbers are chosen independently, the distribution of S is binomial with parameters $n = 500$ and p.

To estimate $P(S < 65)$ we use normal approximation. Note that $E[S] = np = 500 \cdot \frac{5}{33} \approx 75.7576$ and $\text{Var}(S) = np(1-p) \approx 64.2792$. Hence

$$\begin{aligned} P(S < 65) &= P\left(\frac{S - 75.7576}{\sqrt{64.2792}} < \frac{65 - 75.7576}{\sqrt{64.2792}}\right) \approx P\left(\frac{S - 75.7576}{\sqrt{64.2792}} < -1.34\right) \\ &\approx \Phi(-1.34) = 1 - \Phi(1.34) \approx 1 - 0.9099 = 0.0901. \end{aligned}$$

Note that $P(S < 65) = P(S \leq 64)$. Using 64 instead of 65 in the calculation above gives $1 - \Phi(1.47) \approx 0.0708$. If we use the continuity correction then we need to use 64.5 instead of 65 which gives $1-\Phi(1.4)\approx 0.0808$. The actual probability (evaluated numerically) is 0.0778.

(b) We proceed similarly as in part (a). The probability that a given uniformly chosen number from $[1.5, 4.8]$ starts with 3 is $q=\frac{1}{3.3}=\frac{10}{33}$. If we denote the number of such numbers among the 500 random numbers by T then $T\sim \mathrm{Bin}(n,q)$ with $n=500$.

Then

$$\begin{aligned}P(T>160)&=P\left(\frac{T-nq}{\sqrt{nq(1-q)}}>\frac{160-nq}{\sqrt{nq(1-q)}}\right)\approx P\left(\frac{T-nq}{\sqrt{nq(1-q)}}>0.83\right)\\&\approx 1-\Phi(0.83)\approx 1-0.7967=0.2033.\end{aligned}$$

Again, since $P(T>160)=P(T\geq 161)$, we could have done the computation with 161 instead of 160, which would give $1-\Phi(0.92)\approx 0.1788$. If we use the continuity correction then we replace 160 with 160.5 in the calculation above which leads to $1-\Phi(0.87)\approx 0.1922$. The actual probability (evaluated numerically) is 0.1906.

4.17. The probability of rolling two ones is $\frac{1}{36}$. Denote the number of snake eyes out of 10,000 rolls by X. Then $X\sim \mathrm{Bin}(n,p)$ with $n=$10,000 and $p=\frac{1}{36}$. The expectation and variance are

$$np=\frac{2500}{9}\approx 277.78,\qquad np(1-p)=\frac{21{,}875}{81}\approx 270.06.$$

Using the normal approximation:

$$\begin{aligned}P(280\leq X\leq 300)&=P\left(\frac{280-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\leq\frac{X-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\leq\frac{300-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\right)\\&=P\left(\frac{4}{5\sqrt{35}}\leq\frac{X-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\leq\frac{8}{\sqrt{35}}\right)\\&\approx\Phi(\tfrac{8}{\sqrt{35}})-\Phi(\tfrac{4}{5\sqrt{35}})\approx\Phi(1.35)-\Phi(0.135)\\&\approx 0.9115-0.5537=0.3578\end{aligned}$$

(For $\Phi(0.135)$ we used the average of $\Phi(0.13)$ and $\Phi(0.14)$.)

With continuity correction:

$$\begin{aligned}P(279.5\leq X\leq 300.5)&=P\left(\frac{279.5-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\leq\frac{X-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\leq\frac{300.5-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\right)\\&=P\left(0.105\leq\frac{X-\frac{2500}{9}}{\sqrt{\frac{21{,}875}{81}}}\leq 1.38\right)\\&\approx\Phi(1.38)-\Phi(0.105)\approx 0.9162-0.5418\\&=0.3744.\end{aligned}$$

The exact probability can be computed using a computer:

$$P(280 \le X \le 300) = \sum_{k=280}^{300} \binom{10{,}000}{k} (\tfrac{1}{36})^k (\tfrac{35}{36})^{10{,}000-k} \approx 0.3699.$$

4.18. The probability of hitting the bullseye with a given dart is $p = \frac{\pi 1^1}{\pi 5^2} = \frac{1}{25}$. Denoting the number of bullseyes among the 2000 throws by S we get $S \sim \text{Bin}(n, p)$ with $n = 2000$.

Using the normal approximation,

$$\begin{aligned} P(S \ge 100) &= P\left(\frac{S - np}{\sqrt{np(1-p)}} \ge \frac{100 - np}{\sqrt{np(1-p)}} \right) = P\left(\frac{S - np}{\sqrt{np(1-p)}} \ge \frac{20}{8\sqrt{6/5}} \right) \\ &\approx P\left(\frac{S - np}{\sqrt{np(1-p)}} \ge 2.28 \right) \\ &\approx 1 - \Phi(2.28) \approx 1 - 0.9887 = 0.0113 \end{aligned}$$

With continuity correction we need to replace 100 with 99.5 in the calculation above. This way we get $1 - \Phi(2.225) \approx 0.01305$ (using linear approximation for $\Phi(2.225)$). The actual probability (evaluated numerically) is 0.0153.

4.19. Let X be number of people in the sample who prefer cereal A. We may approximate the distribution of X with a $\text{Bin}(n, p)$ distribution with $n = 100$, $p = 0.2$. (This is an approximation, because the true distribution is hypergeometric.) The expectation and variance are $np = 20$ and $np(1-p) = 16$. Since the variance is large enough, it is reasonable to use the normal approximation to estimate $P(X \ge 25)$:

$$\begin{aligned} P(X \ge 25) &= P\left(\frac{X - 20}{\sqrt{16}} \ge \frac{25 - 20}{\sqrt{16}} \right) \\ &\approx P(Z > 1.25) = 1 - \Phi(1.25) \approx 1 - 0.8944 = 0.1056, \end{aligned}$$

If we use the continuity correction then we get

$$\begin{aligned} P(X \ge 25) = P(X > 24.5) &= P\left(\frac{X - 20}{\sqrt{16}} \ge \frac{24.5 - 20}{\sqrt{16}} \right) \\ &\approx P(Z > 1.125) = 1 - \Phi(1.125) \approx 1 - 0.8697 = 0.1303. \end{aligned}$$

(We approximated $\Phi(1.125)$ as the average of $\Phi(1.12)$ and $\Phi(1.13)$.

Using a computer one can also compute the exact probability

$$P(X \ge 25) = \sum_{k=25}^{100} \binom{100}{k} (0.2)^k (0.8)^{100-k} \approx 0.1313.$$

4.20. Let X be the number of heads. Then $10{,}000 - X$ is the number of tails and $|X - (10{,}000 - X)| = |2X - 10{,}000|$ is the difference between the number of heads and number of tails. We need to estimate

$$P(|2X - 10{,}000| \le 100) = P(4950 \le X \le 5050).$$

Since $X \sim \text{Bin}(10{,}000, \frac{1}{2})$, we may use normal approximation to do that:

$$\begin{aligned}
&P(4950 \le X \le 5050) \\
&= P\left(\frac{4950 - 10{,}000 \cdot \frac{1}{2}}{\sqrt{10{,}000 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \le \frac{X - 10{,}000 \cdot \frac{1}{2}}{\sqrt{10{,}000 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \le \frac{5050 - 10{,}000 \cdot \frac{1}{2}}{\sqrt{10{,}000 \cdot \frac{1}{2} \cdot \frac{1}{2}}}\right) \\
&= P\left(-1 \le \frac{X - 10{,}000 \cdot \frac{1}{2}}{\sqrt{10{,}000 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \le 1\right) \\
&\approx 2\Phi(1) - 1 \approx 0.6826.
\end{aligned}$$

4.21. Let X_n be the number of games won out of the first n games. Then $X_n \sim \text{Bin}(n, p)$ with $p = \frac{1}{20}$. The amount of money won in the first n games is then $W_n = 10X_n - (n - X_n) = 11X_n - n$. We have

$$P(W_n > -100) = P(11X_n - n > -100) = P(X_n > \tfrac{n-100}{11}).$$

We apply the normal approximation to this probability.

For $n = 200$ (using the continuity correction):

$$\begin{aligned}
P(W_{200} > -100) &= P(X_{200} > \tfrac{100}{11}) = P(X_{200} \ge 10) \\
&= P(X_{200} > 9.5) = P(\tfrac{X_{200}-10}{\sqrt{9.5}} > \tfrac{-0.5}{\sqrt{9.5}}) \\
&\approx 1 - \Phi(-0.16) = \Phi(0.16) \approx 0.5636.
\end{aligned}$$

For $n = 300$ (using the continuity correction):

$$\begin{aligned}
P(W_{300} > -100) &= P(X_{300} > \tfrac{200}{11}) = P(X_{200} \ge 19) \\
&= P(X_{300} > 18.5) = P(\tfrac{X_{300}-15}{\sqrt{14.25}} > \tfrac{3.5}{\sqrt{14.25}}) \\
&\approx 1 - \Phi(0.93) \approx 0.1762.
\end{aligned}$$

Note that the variance in the $n = 200$ case is 9.5, which is slightly below 10, so the normal approximation is not fully justified. In this case $np^2 = 1/2$, so the Poisson approximation is not guaranteed to work either. The Poisson approximation is

$$P(W_{200} > -100) = P(X_{200} > \tfrac{100}{11}) = P(X_{200} \ge 10) \approx 1 - \sum_{k=0}^{9} e^{-10} \frac{10^k}{k!} \approx 0.5421.$$

The true probability (computed using binomial distribution) is approximately 0.5453, so the Poisson approximation is actually pretty good.

4.22. Let S be the number of times we flipped heads among the first 400 steps. Then $S \sim \text{Bin}(400, \frac{1}{2})$ and the position of the game piece on the board is $Y = S - (400 - S) = 2S - 400$. We need to estimate

$$P(|Y| \le 10) = P(|2S - 400| \le 10) = P(-10 \le 2S - 400 \le 10) = P(195 \le S \le 205).$$

Using the normal approximation (with $E[S] = 400 \cdot \frac{1}{2} = 200$ and $\mathrm{Var}(S) = 400 \cdot \frac{1}{2} \cdot \frac{1}{2} = 100$):

$$\begin{aligned} P(195 \le S \le 205) &= P(\tfrac{195-200}{10} \le \tfrac{S-200}{10} \le \tfrac{205-200}{10}) = P(-\tfrac{1}{2} \le \tfrac{S-200}{10} \le \tfrac{1}{2}) \\ &\approx P(-\tfrac{1}{2} \le Z \le \tfrac{1}{2}) = 2\Phi(1/2) - 1 \approx 2 \cdot 0.6915 - 1 = 0.383. \end{aligned}$$

With the continuity correction we get

$$\begin{aligned} P(195 \le S \le 205) &= P(194.5 < S < 205.5) \approx P(-0.55 \le Z \le 0.55) \\ &= 2\Phi(0.55) - 1 \approx 2 \cdot 0.7088 - 1 = 0.4176. \end{aligned}$$

4.23. Let $X \sim \mathcal{N}(1200, 10{,}000)$ be the lifetime of a single car battery. With $Z \sim \mathcal{N}(0,1)$, X has the same distribution as $1200 + 100Z$. Then

$$\begin{aligned} P(X \le 1100) &= P(1200 + 100Z \le 1100) \\ &= P(Z \le -1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587. \end{aligned}$$

Now let W be the number of car batteries, in a batch of 100, whose lifetimes are less than 1100 hours. Note that $W \sim \mathrm{Bin}(100, 0.1587)$ with an approximate variance of $100 \cdot 0.1587 \cdot 0.8413 = 13.35$. Using a normal approximation, we have

$$\begin{aligned} P(W \ge 20) &= P\left(\frac{W - 100 \cdot 0.1587}{\sqrt{100 \cdot 0.1587 \cdot 0.8413}} \ge \frac{20 - 100 \cdot 0.1587}{\sqrt{100 \cdot 0.1587 \cdot 0.8413}}\right) \approx P(Z \ge 1.13) \\ &= 1 - \Phi(1.13) = 1 - 0.8708 \\ &= 0.1292. \end{aligned}$$

4.24. (a) Let $S_{n,i}, i = 1, 2, \ldots, 6$ be the number of times we rolled the number i among the first n rolls. The probability of each number between 1 and 6 is $1/6$, so the law of large numbers states that for any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P\big(\big|\tfrac{S_{n,4}}{n} - \tfrac{1}{6}\big| < \varepsilon\big) = 1.$$

Using $\varepsilon = \frac{17}{100} - \frac{1}{6} = \frac{1}{300}$ and taking complements we get

$$\lim_{n\to\infty} P\big(\big|\tfrac{S_{n,4}}{n} - \tfrac{1}{6}\big| \ge \varepsilon\big) = 0.$$

But

$$P\big(\big|\tfrac{S_{n,4}}{n} - \tfrac{1}{6}\big| \ge \varepsilon\big) \ge P\big(\tfrac{S_{n,4}}{n} \ge \tfrac{1}{6} + \varepsilon\big) = P\big(\tfrac{S_{n,4}}{n} \ge \tfrac{17}{100}\big),$$

thus if $P(\big|\frac{S_{n,4}}{n} - \frac{1}{6}\big| \ge \varepsilon)$ converges to zero then so does $P(\frac{S_{n,4}}{n} \ge \frac{17}{100})$.

(b) Let $B_{n,i}, i = 1, \ldots, 6$ be the event that after n rolls the frequency of the number i is between 16% and 17%. Then $A_n = \cap_{i=1}^6 B_{n,i}$. Note that $A_n^c = \cup_{i=1}^6 B_{n,i}^c$, and

$$(*) \qquad P(A_n^c) = P(\cup_{i=1}^6 B_{n,i}^c) \le \sum_{i=1}^6 P(B_{n,i}^c).$$

(Exercise 1.43 proved this subadditivity relation.) We would like to show that for large enough n we have $P(A_n) \ge 0.999$. This is equivalent to $P(A_n^c) < 0.001$. If we could show that there is a K so that for $n \ge K$ we have $P(B_{n,i}^c) < \frac{0.001}{6}$ for each $1 \le i \le 6$, then the bound $(*)$ implies $P(A_n^c) < 0.001$ and thereby $P(A_n) \ge 0.999$.

Begin again with the statement given by the law of large numbers: for any $\varepsilon > 0$ and $1 \le i \le 6$ we have

$$\lim_{n\to\infty} P(\left|\tfrac{S_{n,i}}{n} - \tfrac{1}{6}\right| < \varepsilon) = 1.$$

Take $\varepsilon = \frac{17}{100} - \frac{1}{6} = \frac{1}{300}$. Then we have

$$\begin{aligned} P(\left|\tfrac{S_{n,i}}{n} - \tfrac{1}{6}\right| < \varepsilon) &= P(\tfrac{1}{6} - \varepsilon < \tfrac{S_{n,i}}{n} < \tfrac{1}{6} + \varepsilon) \\ &= P(\tfrac{49}{300} < \tfrac{S_{n,i}}{n} < \tfrac{17}{100}) \\ &\le P(\tfrac{16}{100} < \tfrac{S_{n,i}}{n} < \tfrac{17}{100}) = P(B_{n,i}). \end{aligned}$$

Since $P(\left|\frac{S_{n,i}}{n} - \frac{1}{6}\right| < \varepsilon)$ converges to 1, so does $P(B_{n,i})$ for each $1 \le i \le 6$. By this convergence there exists $K > 0$ so that $P(B_{n,i}) > 1 - \frac{0.001}{6}$ for each $1 \le i \le 6$ and all $n \ge K$. This gives $P(B_{n,i}^c) = 1 - P(B_{n,i}) < \frac{0.001}{6}$ for each $1 \le i \le 6$. As argued above, this implies that $P(A_n) \ge 0.999$ for all $n \ge K$.

4.25. Let S_n be the number of interviewed people that prefer cereal to bagels for breakfast. If the population is large, we can assume that sampling from the population with replacement or without replacement does not make a big difference, therefore we assume $S_n \sim$Bin(n, p). In this case, $n = 81$. As usual, the estimate of p will be

$$\hat{p} = \frac{S_n}{n}.$$

We want to find $q \in [0, 1]$ such that

$$P(|\hat{p} - p| < 0.05) = P\left(\left|\frac{S_n}{n} - p\right| < 0.05\right) \ge q$$

If $Z \sim$ N(0,1), we have that

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - p\right| < 0.05\right) &= P\left(-\frac{0.05\sqrt{n}}{p(1-p)} < \frac{S_n - np}{\sqrt{np(1-p)}} < \frac{0.05\sqrt{n}}{p(1-p)}\right) \\ &\approx P\left(-\frac{0.05\sqrt{n}}{p(1-p)} < Z < \frac{0.05\sqrt{n}}{p(1-p)}\right) \\ &\ge P(-2 \cdot 0.05\sqrt{n} < Z < 2 \cdot 0.05\sqrt{n}) \\ &= \Phi(2 \cdot 0.05\sqrt{n}) - \Phi(-2 \cdot 0.05\sqrt{n}) = 2\Phi(2 \cdot 0.05\sqrt{n}) - 1 \\ &= 2\Phi(0.9) - 1 \approx 2 \cdot 0.8159 - 1 = 0.6318. \end{aligned}$$

Therefore, the true p lies in the interval $(\hat{p} - 0.05, \hat{p} + 0.05)$ with probability greater than or equal to 0.6318. Note that this is not a very high confidence level.

4.26. Let S be the number of interviewed people that prefer whole milk to skim milk. Then $S \sim \text{Bin}(n, p)$ with $n = 100$. Our estimate for p is $\widehat{p} = \frac{S}{n}$. The event $p \in (\widehat{p} - 0.1\,, \widehat{p} + 0.1)$ is the same as $|\frac{S}{n} - p| < 0.1$. To estimate the probability of this event we use normal approximation:

$$\begin{aligned} P(|S/n - p| < 0.1) &= P(-\tfrac{0.1\sqrt{n}}{\sqrt{p(1-p)}} < \tfrac{S-np}{\sqrt{np(1-p)}} < \tfrac{0.1\sqrt{n}}{\sqrt{p(1-p)}}) \\ &\approx 2\Phi(\tfrac{0.1\sqrt{n}}{\sqrt{p(1-p)}}) - 1 \ge 2\Phi(0.2\sqrt{n}) - 1, \end{aligned}$$

where we used $p(1-p) \le 1/4$ in the last step.

Since $n = 100$ we have

$$2\Phi(0.2\sqrt{n}) - 1 = 2\Phi(2) - 1 \approx 2 \cdot 0.9772 - 1 = 0.9544.$$

Thus the interval $(\widehat{p} - 0.1\,,\, \widehat{p} + 0.1)$ corresponds to 95.44% confidence.

4.27. We need to find n so that

$$P(-\tfrac{1}{10} \le \tfrac{X}{n} - p \le \tfrac{1}{10}) \ge 0.9.$$

Using normal approximation:

$$\begin{aligned} P(-\tfrac{1}{10} \le \tfrac{X}{n} - p \le \tfrac{1}{10}) &\le P(-\tfrac{\sqrt{n}}{10\sqrt{p(1-p)}} \le \tfrac{X-np}{\sqrt{np(1-p)}} \le \tfrac{\sqrt{n}}{10\sqrt{p(1-p)}}) \\ &\approx 2\Phi(\tfrac{\sqrt{n}}{10\sqrt{p(1-p)}}) - 1 \end{aligned}$$

We need

$$2\Phi(\tfrac{\sqrt{n}}{10\sqrt{p(1-p)}}) - 1 \ge 0.9 \Leftrightarrow \Phi(\tfrac{\sqrt{n}}{10\sqrt{p(1-p)}}) \ge 0.95$$

which holds if

$$\frac{\sqrt{n}}{10\sqrt{p(1-p)}} \ge 1.645$$

(using linear interpolation in the table). This yields

$$n \ge 1.645^2 \cdot 100p(1-p).$$

We know that $p(1-p) \le 1/4$, so if $n \ge 1.645^2 \cdot 100 \cdot \frac{1}{4} = 67.65$ then our inequality will hold. Thus n should be at least 68.

4.28. For $p = 1$ the maximum is at n (since the p.m.f. is 1 there), and for $p = 0$ it is not (as the p.m.f. is 0 there). From this point we will assume $0 < p < 1$.

Denote by $f(k)$ the p.m.f. of the $\mathrm{Bin}(n,p)$ distribution at k. Then for $0 \le k \le n-1$ we have

$$\begin{aligned} \frac{f(k+1)}{f(k)} &= \frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{\frac{n!}{(k+1)!(n-k-1)!}p^{k+1}(1-p)^{n-k-1}}{\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}} \\ &= \frac{(n-k)p}{(k+1)(1-p)} \end{aligned}$$

Then $f(k+1) \ge f(k)$ if and only if $(n-k)p \ge (k+1)(1-p)$, which is equivalent to $k \le p(n+1) - 1$. This means that if $n - 1 \le p(n+1) - 1$ then we have $f(0) \le f(1) \le \cdots \le f(n-1) \le f(n)$. If $n-1 > p(n+1)-1$ then $f(n-1) > f(n)$. Thus the maximum is at n if $n-1 \le p(n+1)-1$ which is equivalent to $p \ge 1 - \frac{1}{n+1}$.

To summarize: the p.m.f. of the $\mathrm{Bin}(n,p)$ distribution has its maximum at n if $p \ge 1 - \frac{1}{n+1}$.

4.29. If $P(S_n = k) > 0$ then $|k|$ cannot be bigger than n, and the parity of n and k must be the same. (Otherwise the random walker cannot get from 0 to k in exactly n steps.)

Assume now that $|k| \le n$ and that $n - k = 2a$ with a being an integer. The random walker ends up at $k = n - 2a$ after n steps exactly if it takes $n - a$ up steps and a down steps. The probability of this is the same that a $\mathrm{Bin}(n,p)$ random variable is equal to $n-a$, which is $\binom{n}{n-a}p^{n-a}(1-p)^a$. Since $n-a=\frac{n+k}{2}$ and $a=\frac{n-k}{2}$, we get that for $|k|\le n$ and $n-k$ even we have

$$P(S_n=k)=\binom{n}{\frac{n+k}{2}}p^{\frac{n+k}{2}}(1-p)^{\frac{n-k}{2}},$$

otherwise $P(S_n=k)$ is zero.

4.30. Let $f(k)$ be the probability mass function of a Poisson(λ) random variable at k. Then for $k\ge 0$ we have

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

This means that $f(k+1)>f(k)$ exactly if $\lambda>k+1$ or $\lambda-1>k$, and $f(k+1)<f(k)$ exactly if $\lambda-1<k$.

If λ is not an integer then let $k^*=\lfloor\lambda\rfloor$ be the integer part of λ (the largest integer smaller than λ). By the arguments above we have

$$f(0)<f(1)<\dots<f(k^*)>f(k^*+1)>f(k^*+2)>\dots$$

If λ is a positive integer then

$$f(0)<f(1)<\dots<f(\lambda-1)=f(\lambda)>f(\lambda+1)>f(\lambda+2)>\dots$$

In both cases f is increasing and then decreasing.

4.31. We have

$$\begin{aligned}E\left[\frac{1}{1+X}\right]&=\sum_{k=0}^{\infty}\frac{1}{k+1}e^{-\mu}\frac{\mu^k}{k!}=\frac{1}{\mu}\sum_{k=0}^{\infty}e^{-\mu}\frac{\mu^{k+1}}{(k+1)!}\\&=\frac{1}{\mu}\sum_{\ell=1}^{\infty}e^{-\mu}\frac{\mu^\ell}{\ell!}=\frac{1-e^{-\mu}}{\mu}.\end{aligned}$$

We introduced $\ell=k+1$ and used $\sum_{\ell=1}^{\infty}e^{-\mu}\frac{\mu^\ell}{\ell!}=1-e^{-\mu}$.

4.32. (a) We can compute $E[g(Y)]$ with the formula $\sum_{k=0}^{\infty}g(k)P(Y=k)$. Thus

$$E[Y(Y-1)\cdots(Y-n+1)]=\sum_{k=0}^{\infty}k(k-1)\cdots(k-n+1)\frac{\mu^k}{k!}e^{-\mu}.$$

Note that $k(k-1)\cdots(k-n+1)=0$ for $k=0,1,\dots,n-1$. Thus we can start the sum at $k=n$:

$$E[Y(Y-1)\cdots(Y-n+1)]=\sum_{k=n}^{\infty}k(k-1)\cdots(k-n+1)\frac{\mu^k}{k!}e^{-\mu}.$$

Moreover, for $k\ge n$ the product $k(k-1)\cdots(k-n+1)$ is exactly the product of the first n factors in $k!=k(k-1)(k-2)\cdots 1$, hence

$$E[Y(Y-1)\cdots(Y-n+1)]=\sum_{k=n}^{\infty}\frac{\mu^k}{(k-n)!}e^{-\mu}.$$

Introducing $\ell = k - n$ we can rewrite the sum as

$$\sum_{k=n}^{\infty} \frac{\mu^k}{(k-n)!} e^{-\mu} = \sum_{\ell=0}^{\infty} \frac{\mu^{\ell+n}}{\ell!} e^{-\mu} = \mu^n \sum_{\ell=0}^{\infty} \frac{\mu^\ell}{\ell!} e^{-\mu} = \mu^n.$$

(The last step follows from $\sum_{\ell=0}^{\infty} \frac{\mu^\ell}{\ell!} e^{-\mu} = 1$.) Thus the nth factorial moment of Y is μ^n.

(b) We can compute $E[Y^3]$ by expressing it in terms of factorial moments of Y and then using part (a). We have

$$\begin{aligned} y^3 &= y(y-1)(y-2) + 3y^2 - 2y \\ &= y(y-1)(y-2) + 3y(y-1) + y. \end{aligned}$$

Thus

$$\begin{aligned} E[Y^3] &= \sum_{k=0}^{\infty} k^3 \frac{\mu^k}{k!} e^{-\mu} \\ &= \sum_{k=0}^{\infty} k(k-1)(k-2) \frac{\mu^k}{k!} e^{-\mu} + 3 \sum_{k=0}^{\infty} k(k-1) \frac{\mu^k}{k!} e^{-\mu} + \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} \\ &= \mu^3 + 3\mu^2 + \mu. \end{aligned}$$

4.33. Let X denote the number of calls on a given day. According to our assumption this is a Poisson(λ) random variable with some parameter λ, and our goal is to find λ. (Since the parameter is the same as the expected value.) We are given that $P(X = 0) = 0.005$, which gives $e^{-\lambda} = 0.005$ and $\lambda = -\log(0.005) \approx 5.298$.

4.34. We can assume that each taxi has a small probability of getting into an accident on a given day, independently of the others. Since there are a large number of taxis, the number of accidents on a given week could be well approximated with a Poisson(μ) distributed random variable. There are on average 3 accidents a week, thus it is reasonable to choose $\mu = 3$. Then the probability of having 2 accidents next week is given by $\frac{3^2}{2!} e^{-3} = \frac{9}{2} e^{-3}$.

4.35. The probability of getting all heads or all tails after flipping a coin ten times is $p = 2^{-9}$. The distribution of X is $\text{Bin}(n, p)$ with $n = 365$.

(a)

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - (1 - 2^{-9})^{365} - 365 \cdot 2^{-9}(1 - 2^{-9})^{364}.$$

(b) Since $np = 365 \cdot 2^{-9} \approx 0.7129$ and $np^2 < 0.0014$, the Poisson approximation is appropriate.

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) \approx 1 - e^{-0.7129} - 0.71 e^{-0.7129} \approx 0.1603.$$

4.36. Assume that we invite n guests and let X denote the number of guests with the same birth day as mine. We need to find n so that $P(X \geq 1) \geq 2/3$. If we disregard leap years, and assume that the birth days are chosen uniformly and independently, then X has binomial distribution with parameters n and $p = \frac{1}{365}$. We have $P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - \frac{1}{365})^n$. Solving $1 - (1 - \frac{1}{365})^n \geq 2/3$ gives $n \geq \frac{\ln(3)}{\ln(1 - \frac{1}{365})} \approx 400.444$ which means that we should invite at least 401 guests.

Note that we can approximate the $\mathrm{Bin}(n, \frac{1}{365})$ distribution with a $\mathrm{Poisson}(\frac{n}{365})$ distributed random variable Y. Then $P(X \geq 1) \approx P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-\frac{n}{365}}$. To get $1 - e^{-\frac{n}{365}} \geq 2/3$ we need $n \geq 365 \ln 3 \approx 400.993$ which also gives $n \geq 401$.

4.37. Since there are lots of scoring chances, but only a few of them result goals, it is reasonable to model the number of goals in a given game by a $\mathrm{Poisson}(\lambda)$ random variable. Then the percentage of games with no goals should be close to the probability of this $\mathrm{Poisson}(\lambda)$ random variable being zero, which is $e^{-\lambda}$. Thus

$$0.0816 = e^{-\lambda} \quad \rightsquigarrow \quad \lambda = -\log(0.0816) \approx 2.506$$

The percentage of games where exactly one goal was scored should be close to $\lambda e^{-\lambda} = 0.2045$ or 20.45%.

(Note: in reality 77 of the 380 games ended with one goal which gives 20.26%. The Poisson approximation gives an extremely precise estimate!)

4.38. Note that X is a Bernoulli random variable with success probability p, and $Y \sim \mathrm{Poisson}(p)$. We need to show that for any subset A of $\{0, 1, \dots\}$ we have

$$|P(X \in A) - P(Y \in A)| \leq p^2.$$

This looks hard, as there are lots of subsets of $\{0, 1, \dots\}$. Let us start with the subsets $\{0\}$ and $\{1\}$. In these cases

$$P(X \in A) - P(Y \in A) = P(X = k) - P(Y = k) = \begin{cases} 1 - p - e^{-p}, & \text{if } k = 0 \\ p - pe^{-p}, & \text{if } k = 1. \end{cases}$$

We have $1 - p \leq e^{-p}$. This can be shown by noting that the function e^{-x} is convex, and hence its tangent line at $x = 0$ (the line $1 - x$) must always be below the graph. Integrating this inequality on $[0, p]$ and then rearranging it gives $0 \leq e^{-p} + p - 1 \leq \frac{p^2}{2}$. We also get $0 \leq p - pe^{-p} = p(1 - e^{-p}) \leq p^2$.

This gives

$$-\frac{p^2}{2} \leq P(X = 0) - P(Y = 0) \leq 0, \qquad 0 \leq P(X = 1) - P(Y = 1) \leq p^2.$$

Now consider a general subset A of $\{0, 1, \dots\}$. We consider four cases.
Case 1: A does not contain 0 or 1. In this case $P(X \in A) = 0$ and

$$P(Y \in A) \leq P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - e^{-p}(1 + p).$$

Hence $P(X \in A) - P(Y \in A) = -P(Y \in A)$ and

$$|P(X \in A) - P(Y \in A)| \leq 1 - e^{-p}(1 + p) \leq 1 - (1 - p)(1 + p) = p^2.$$

Case 2: A contains both 0 and 1. In this case $P(X \in A) = 1$ and

$$1 \geq P(Y \in A) \geq P(Y \leq 1) = e^{-p}(1 + p).$$

Hence $P(X \in A) - P(Y \in A) = -P(Y \in A)$ and

$$|P(X \in A) - P(Y \in A)| \leq 1 - e^{-p}(1 + p) \leq 1 - (1 - p)(1 + p) = p^2.$$

Case 3: A contains 0 but not 1. In this case $P(X \in A) = 1 - p$ and

$$P(Y \in A) \geq P(Y = 0) = e^{-p}$$
$$P(Y \in A) \leq P(Y = 0) + P(Y \geq 2) = 1 - P(Y = 1) = 1 - pe^{-p}.$$

This gives

$$1 - p - (1 - pe^{-p}) \leq P(X \in A) - P(Y \in A) \leq 1 - p - e^{-p}.$$

We have seen that $-\frac{p^2}{2} \leq 1 - p - e^{-p} \leq 0$ and we also have

$$1 - p - (1 - pe^{-p}) = -p(1 - e^{-p}) \geq -p^2.$$

Thus

$$-p^2 \leq P(X \in A) - P(Y \in A) \leq -p^2/2$$

and $|P(X \in A) - P(Y \in A)| \leq p^2$. **Case 4: A contains 1 but not 0.** This case can be handled similarly as Case 3. Or we could note that A^c contains 0 but not 1, and thus by Case 3 we have $|P(X \in A^c) - P(Y \in A^c)| \leq p^2$. But

$$|P(X \in A) - P(Y \in A)| = |(1 - P(X \in A^c)) - (1 - P(Y \in A^c)| = |P(X \in A^c) - P(Y \in A^c)|$$

hence we get $|P(X \in A) - P(Y \in A)| \leq p^2$ in this case as well.

We checked all possible cases, and we have shown that Fact 4.20 holds for $n = 1$ every time.

4.39. Let X be the number of wheat cents among Cassandra's

(a) We have $X \sim \text{Bin}(n, p)$ with $n = 400$ and $p = \frac{1}{350}$. Thus

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (\tfrac{349}{350})^{400} - 400 \cdot \tfrac{1}{350} \cdot (\tfrac{349}{350})^{399}$$

We could also write this as

$$P(X \geq 2) = \sum_{k=2}^{400} \binom{400}{k} (\tfrac{1}{350})^k \cdot (\tfrac{349}{350})^{400-k}$$

(b) Since np^2 is small, the Poisson approximation is appropriate with parameter $\mu = np = \frac{8}{7}$. Then

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) \approx 1 - e^{-\frac{8}{7}} - \tfrac{8}{7} e^{-\frac{8}{7}} \approx 0.3166$$

4.40. Let X denote the number of times the number one appears in the sample. Then $X \sim \text{Bin}(111, \frac{1}{10})$. We need to approximate $P(X \leq 3)$. Using normal approximation gives

$$\begin{aligned} P(X \leq 3) &= P\left(\frac{X - 111 \cdot \frac{1}{10}}{\sqrt{111 \cdot \frac{1}{10} \cdot \frac{9}{10}}} \leq \frac{3 - 111 \cdot \frac{1}{10}}{\sqrt{111 \cdot \frac{1}{10} \cdot \frac{9}{10}}} \right) \\ &\approx P\left(\frac{X - 111 \cdot \frac{1}{10}}{\sqrt{111 \cdot \frac{1}{10} \cdot \frac{9}{10}}} \leq -2.56 \right) \\ &\approx \Phi(-2.56) = 1 - \Phi(2.56) \approx 1 - 0.9948 = 0.0052. \end{aligned}$$

If we use the continuity correction then we have to repeat the calculation above starting from $P(X \leq 3) = P(X < 2.5)$ which gives the approximation $\Phi(-2.72) \approx 0.0033$.

For the Poisson approximation we approximate X with a random variable $Y \sim \text{Poisson}(\frac{111}{10})$. Then

$$\begin{aligned} P(X \le 3) &\approx P(Y \le 3) = P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\ &= e^{-11.1}(1 + 11.1 + \frac{11.1^2}{2} + \frac{11.1^3}{6}) \approx 0.004559. \end{aligned}$$

The variance of X is $\frac{999}{100}$ which is almost 10, hence it is not that surprising that the normal approximation is pretty accurate (especially with continuity correction).

Since $np^2 = 111 \cdot (\frac{1}{10})^2 = 1.11$ is not very small, we cannot expect the Poisson approximation to be very precise, although it is still quite accurate.

4.41. Let X be the number of sixes. Then $X \sim \text{Bin}(n, p)$ with $n = 72$ and $p = 1/6$.

$$P(X = 3) = \binom{72}{3}(\tfrac{1}{6})^3(\tfrac{5}{6})^{69} \approx 0.00095.$$

The Poisson approximation would compare X with a Poisson(μ) random variable with $\mu = np = 12$:

$$P(X = 3) \approx e^{-12}\frac{12^3}{3!} \approx 0.0018.$$

For the normal approximation we need the continuity correction:

$$\begin{aligned} P(X = 3) = P(2.5 \le X \le 3.5) &= P\left(\tfrac{2.5-12}{\sqrt{10}} \le \tfrac{X-12}{\sqrt{10}} \le \tfrac{3.5-12}{\sqrt{10}}\right) \\ &\approx \Phi(-2.69) - \Phi(-3.0) = \Phi(3.0) - \Phi(2.69) \approx 0.9987 - 0.9964 = 0.0023. \end{aligned}$$

4.42. (a) Let X be the number of mildly defective gadgets in the box. Then $X \sim \text{Bin}(n, p)$ with $n = 100$ and $p = 0.2 = \frac{1}{5}$. We have

$$P(A) = P(X < 15) = \sum_{k=0}^{14}\binom{100}{k}(1/5)^k(4/5)^{100-k}.$$

(b) We have $np(1 - p) = 16 > 10$ and $np^2 = 4$. This suggests that the normal approximation is more appropriate than the Poisson approximation in this case.

Using normal approximation we get

$$\begin{aligned} P(X < 15) &= P\left(\frac{X - 100 \cdot \frac{1}{5}}{\sqrt{100 \cdot \frac{1}{5} \cdot \frac{4}{5}}} < \frac{15 - 100 \cdot \frac{1}{5}}{\sqrt{100 \cdot \frac{1}{5} \cdot \frac{4}{5}}}\right) \\ &= P\left(\frac{X - 100 \cdot \frac{1}{5}}{\sqrt{100 \cdot \frac{1}{5} \cdot \frac{4}{5}}} < -\frac{5}{4}\right) \\ &\approx \Phi(-1.25) = 1 - \Phi(1.25) \approx 1 - 0.8944 = 0.1056. \end{aligned}$$

With continuity correction we would get $\Phi(-1.375) = 1 - \Phi(1.375) \approx 0.08455$ (using linear interpolation to get $\Phi(1.375)$).

The actual value is 0.0804437 (calculated with a computer).

4.43. We first consider the probability $P(X \ge 48)$. Note that $X \sim \text{Binomial}(400, 0.1)$. Note also that the mean of X is 40 and the variance is $400 * 0.1 * 0.9 = 36$, which is large enough for a normal approximation to work. So, letting $Z \sim N(0,1)$ and using the correction for continuity, we have

$$\begin{aligned} P\{X \geq 48\} = P(X \geq 47.5) &= P\left(\frac{X-40}{6} \geq \frac{47.5-40}{6}\right) \\ &\approx P(Z \geq 1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056. \end{aligned}$$

Next we turn to approximating $P(Y \geq 2)$. Note that $Y \sim$ Binomial(400,0.0025), and since $400 \cdot 0.0025 = 1$ and $400 \cdot 0.0025^2 = 0.0025$ is small, it is clear that only a Poisson approximation is appropriate in this case. Letting $N \sim$ Poisson(1), we have

$$P(Y \geq 2) \approx P(N \geq 2) = 1 - P(N=0) - P(N=1) = 1 - e^{-1} - e^{-1} = 0.2642.$$

4.44. (a) Let X denote the number of defective watches in the box. Then $X \sim \text{Bin}(n,p)$ with $n = 400$ and $p = 1/2$. We are interested in the probability that at least 215 of the 400 watches are defective, this is the event $\{X \geq 215\}$. The exact probability is

$$P(X \geq 215) = \sum_{k=215}^{400} \binom{400}{k} \frac{1}{2^{400}}.$$

(b) We have $np(1-p) = 100 > 10$ and $np^2 = 100$. Thus it is more reasonable to use the normal approximation:

$$\begin{aligned} P(X \geq 215) &= P\left(\frac{X-400\cdot\frac{1}{2}}{\sqrt{400\cdot\frac{1}{2}\cdot\frac{1}{2}}} \geq \frac{215-400\cdot\frac{1}{2}}{\sqrt{400\cdot\frac{1}{2}\cdot\frac{1}{2}}}\right) \\ &= P\left(\frac{X-400\cdot\frac{1}{2}}{\sqrt{400\cdot\frac{1}{2}\cdot\frac{1}{2}}} \geq \frac{3}{2}\right) \\ &\approx 1 - \Phi(1.5) \approx 1 - 0.9332 = 0.0668. \end{aligned}$$

If we use continuity correction then we start with $P(X \geq 215) = P(X > 214.5)$ which leads to the approximation $1 - \Phi(1.45) \approx 0.0735$.

The actual probability is 0.07348 (calculated with a computer).

4.45. The probability of a four of a kind is $p = \frac{13\cdot 48}{\binom{52}{5}} = \frac{1}{4165}$. Denote by X the number of four of a kinds we see in 10,000 poker hands. Then $X \sim \text{Bin}(n,p)$ with $n = 10,000$. Since np^2 is tiny, we can approximate X with a Poisson(μ) random variable with $\mu = np$. Then

$$P(X = 0) \approx e^{-10,000\cdot\frac{1}{4165}} \approx 0.0907.$$

4.46. The probability that we get 5 tails when we flip a coin 5 times is $\frac{1}{2^5} = \frac{1}{32}$. Thus $X \sim \text{Bin}(n,p)$ with $n = 30$ and $p = \frac{1}{32}$. Since $np(1-p) = \frac{465}{512} < 1$, the normal approximation is not appropriate. On the other hand, $np^2 = \frac{15}{512} \approx 0.029$ is small, so the Poisson approximation should work. For this we approximate the distribution of X using a random variable $Y \sim \text{Poisson}(\lambda)$ with $\lambda = np = \frac{15}{16}$ to get

$$P(X = 2) \approx P(Y = 2) = \frac{\lambda^2}{2} e^{-\lambda} = \left(\frac{15}{16}\right)^2 e^{-\frac{15}{16}} \approx 0.1721.$$

The actual probability is 0.1746 (calculated with a computer).

4.47. (a) Let X be the number of times in a year that he needed more than 10 coin flips. Then $X \sim \text{Bin}(365, p)$ with

$$p = P(\text{more than 10 coin flips needed}) = P(\text{first 10 coin flips are tails}) = \frac{1}{2^{10}}$$

Since $np(1-p)$ is small (and np^2 is even smaller), we can use the Poisson approximation here with $\lambda = np = \frac{365}{2^{10}} = 0.356$. Then

$$P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2) \approx 1 - e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2}) \approx 0.00579.$$

(b) Denote the number of times that he needed exactly 3 coin flips by Y. This has a $\text{Bin}(365, r)$ distribution with success probability $r = \frac{1}{2^3} = \frac{1}{8}$. (The value of r is the probability that a Geom(1/2) random variable is equal to 3.) Since $nr(1-r) = 39.92 > 10$, we can use normal approximation. The expectation of Y is $E[Y] = nr = 45.625$.

$$\begin{aligned} P(X > 50) &= P\Big(\frac{X - 45.625}{\sqrt{39.92}} > \frac{50 - 45.625}{\sqrt{39.92}}\Big) = P\Big(\frac{X - 45.625}{\sqrt{39.92}} > 0.69\Big) \\ &\approx 1 - \Phi(0.69) = 1 - 0.7549 = 0.2451. \end{aligned}$$

4.48. Let $A = \{X \in [0,1]\}$ and $B = \{X \in [a,2]\}$. We need to find $a < 1$ so that $P(AB) = P(A)P(B)$.

If $a \leq 0$ then $AB = A$, and then $P(A)P(B) \neq P(AB)$. Thus we must have $0 < a < 1$ and hence $AB = \{X \in [a,1]\}$. The c.d.f. of X is $1 - e^{-2x}$ for $x \geq 0$ and 0 otherwise. From this we can compute

$$\begin{aligned} P(A) &= P(0 \leq X \leq 1) = 1 - e^{-2} \\ P(B) &= P(a \leq X \leq 2) = e^{-2a} - e^{-4} \\ P(AB) &= P(a \leq X \leq 1) = e^{-2a} - e^{-2}. \end{aligned}$$

Thus $P(AB) = P(A)P(B)$ is equivalent to

$$(1 - e^{-2})(e^{-2a} - e^{-4}) = e^{-2a} - e^{-2}.$$

Solving this we get $e^{-2a} = e^{-4} + 1 - e^{-2}$ and $a = -\frac{1}{2}\ln(1 - e^{-2} + e{-4}) \approx 0.0622$.

4.49. Let $T \sim \text{Exp}(1/10)$ be the lifetime of a particular stove. Let $r > 0$ and let X be the amount of money you earn on a particular extended warranty of length r. We see that

$$X = \begin{cases} C & \text{if } T > r \\ C - 800 & \text{if } T \leq r \end{cases}$$

We have $P(T > r) = e^{-(1/10)r}$, and so

$$\begin{aligned} E[X] &= CP(X = C) + (C - 800)P(X = C - 800) \\ &= CP(T > r) + (C - 800)P(T \leq r) \\ &= Ce^{-r/10} + (C - 800)(1 - e^{-r/10}). \end{aligned}$$

Thus, the pairs of numbers (C, r) will give an expected profit of zero are those satisfying:

$$0 = Ce^{-r/10} + (C - 800)(1 - e^{-r/10}).$$

4.50. By the memoryless property of the exponential distribution for any $x > 0$ we have

$$P(T > x + 7 | T > 7) = P(T > x).$$

Thus the conditional probability of waiting at least 3 more hours is $P(T > 3) = e^{-\frac{1}{3}\cdot 3} = e^{-1}$, and the conditional probability of waiting at least $x > 0$ more hours is $P(T > x) = e^{-\frac{1}{3}x}$.

4.51. We know from the condition that $0 \le T_1 \le t$, so $P(T_1 \le s \,|\, N_t = 1) = 0$ if $s < 0$ and $P(T_1 \le s \,|\, N_t = 1) = 1$ if $s > t$.

If $0 \le s \le t$ we have

$$P(T_1 \le s \,|\, N_t = 1) = \frac{P(T_1 \le s, N_t = 1)}{P(N_t = 1)}.$$

Since the arrival is a Poisson process with intensity λ, we have $P(N_1 = 1) = \lambda e^{-\lambda t}$. Also,

$$\begin{aligned} P(T_1 \le s, N_t = 1) &= P(N([0,s]) = 1, N([0,t]) = 1) = P(N([0,s]) = 1, N([s,t]) = 0) \\ &= P(N([0,s]) = 1)P(N([s,t]) = 0) = \lambda s e^{-\lambda s} \cdot e^{-\lambda(t-s)} \\ &= \lambda s e^{-\lambda t}. \end{aligned}$$

Then

$$P(T_1 \le s \,|\, N_t = 1) = \frac{P(T_1 \le s, N_t = 1)}{P(N_t = 1)} = \frac{\lambda s e^{-\lambda t}}{\lambda e^{-\lambda t}} = s.$$

Collecting all cases:

$$P(T_1 \le s \,|\, N_t = 1) = \begin{cases} 0, & s < 0 \\ s, & 0 \le s \le t \\ 1, & s > t. \end{cases}$$

This means that the conditional distribution is uniform on $[0, t]$.

4.52. (a) By definition $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$ for $r > 0$. Then $\Gamma(r+1) = \int_0^\infty x^r e^{-x} dx$. Using integration by parts with $(-e^{-x})' = e^{-x}$ we get

$$\begin{aligned} \Gamma(r+1) &= \int_0^\infty x^r e^{-x} dx \\ &= x^r(-e^{-x})\Big|_{x=0}^{x=\infty} - \int_0^\infty r x^{r-1}(-e^{-x}) dx \\ &= r \int_0^\infty x^{r-1} e^{-x} dx = r\Gamma(r). \end{aligned}$$

The two terms in $x^r(-e^{-x})\Big|_{x=0}^{x=\infty}$ disappear because $r > 0$ and $\lim_{x\to\infty} x^r e^{-x} = 0$.

(b) We use induction to prove the identity. For $n = 1$ the statement is true as

$$\Gamma(1) = \int_0^\infty e^{-x} dx = 1 = 0!.$$

Assume that the statement is true for some positive integer n: $\Gamma(n) = (n-1)!$, we need to show that it also holds for $n + 1$. But this is true because by part (a) we have

$$\Gamma(n+1) = n\Gamma(n) = n \cdot (n-1)! = n!,$$

where we used the induction hypothesis and the definition of $n!$.

4.53. We have

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x} dx.$$

We can modify the integrand so that we the probability density function of a Gamma$(r+1, \lambda)$ appears:

$$E[X] = \frac{\Gamma(r+1)}{\lambda \Gamma(r)} \int_0^{\infty} \frac{\lambda^{r+1} x^r}{\Gamma(r+1)} e^{-\lambda x} dx.$$

Since the probability density function of a Gamma$(r+1, \lambda)$ integrates to 1 this leads to

$$E[X] = \frac{\Gamma(r+1)}{\lambda \Gamma(r)} = \frac{r\Gamma(r)}{\lambda \Gamma(r)} = \frac{r}{\lambda}.$$

In the last step we used $\Gamma(r+1) = r\Gamma(r)$. We can use the same trick to compute the second moment:

$$\begin{aligned} E[X^2] &= \int_0^{\infty} x^2 \cdot \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x} dx = \frac{\Gamma(r+2)}{\lambda^2 \Gamma(r)} \int_0^{\infty} \frac{\lambda^{r+2} x^{r+1}}{\Gamma(r+2)} e^{-\lambda x} dx \\ &= \frac{\Gamma(r+2)}{\lambda^2 \Gamma(r)} = \frac{(r+1) r \Gamma(r)}{\lambda^2 \Gamma(r)} = \frac{(r+1) r}{\lambda^2}. \end{aligned}$$

Then the variance is

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}.$$

Solutions to Chapter 5

5.1. We have $M(t) = E[e^{tX}]$, and since X is discrete we have $E[e^{tX}] = \sum_k P(X = k)e^{tk}$. Using the given probability mass function we get

$$\begin{aligned} M(t) &= P(X=-6)e^{-6t} + P(X=-2)e^{-2t} + P(X=0) + P(X=3)e^{3t} \\ &= \tfrac{4}{9}e^{-6t} + \tfrac{1}{9}e^{-2t} + \tfrac{2}{9} + \tfrac{2}{9}e^{3t} \end{aligned}$$

5.2. (a) We have

$$M'(t) = -\tfrac{4}{3}e^{-4t} + \tfrac{5}{6}e^{5t}, \qquad M''(t) = \tfrac{16}{3}e^{-4t} + \tfrac{25}{6}e^{5t}.$$

Hence $E(X) = M'(0) = -\frac{4}{3} + \frac{5}{6} = \frac{1}{2}$, $E(X^2) = M''(0) = \frac{16}{3} + \frac{25}{6} = \frac{57}{6} = \frac{19}{2}$, and $\mathrm{Var}(X) = E(X^2) - (E[X])^2 = \frac{19}{2} - \frac{1}{4} = \frac{37}{4}$.

(b) From the moment generating function we see that X is discrete, the possible values are $-4, 0$ and 5. The corresponding probabilities can be read off from the coefficients of the appropriate exponential terms:

$$p(0) = \tfrac{1}{2}, \quad p(-4) = \tfrac{1}{3}, \quad p(5) = \tfrac{1}{6}.$$

From this we get

$$\begin{aligned} E(X) &= \tfrac{1}{3}\cdot(-4) + \tfrac{1}{6}\cdot 5 = \tfrac{1}{2}, \\ E(X^2) &= \tfrac{1}{3}\cdot 16 + \tfrac{1}{6}\cdot 25 = \tfrac{57}{6} = \tfrac{19}{2}, \\ \mathrm{Var}(X) &= E(X^2) - (E[X])^2 = \tfrac{19}{2} - \tfrac{1}{4} = \tfrac{37}{4} \end{aligned}$$

5.3. The probability density function of X is $f(x) = 1$ for $x \in [0,1]$ and 0 otherwise. The moment generating function can be computed as

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} f(x)e^{tx}dx = \int_0^1 e^{tx}dx.$$

If $t = 0$ then $M(t) = \int_0^1 dx = 1$. If $t \neq 0$ then

$$M(t) = \int_0^1 e^{tx}dx = \frac{e^t - 1}{t}.$$

5.4. (a) In Example 5.5 we have seen that the moment generating function of a $\mathcal{N}(\mu, \sigma^2)$ random variable is $e^{\frac{\sigma^2 t^2}{2}+\mu t}$. Thus if $\tilde{X} \sim \mathcal{N}(0, 12)$ then $M_{\tilde{X}}(t) = e^{6t^2}$ and $M_{\tilde{X}}(t) = M_X(t)$ for $|t| < 2$. But then by Fact 5.14 the distribution of X is the same as the distribution of $\tilde{X}$.
(b) In Example 5.6 we computed the moment generating function of an $\text{Exp}(\lambda)$ distribution, and it was $\frac{\lambda}{\lambda - t}$ for $t < \lambda$ and ∞ otherwise. Thus $M_Y(t)$ has the same moment generating function as an Exp(2) distribution in the interval $(-1/2, 1/2)$, hence by Fact 5.14 we have $Y \sim \text{Exp}(2)$.
(c) We cannot identify the distribution of Z, as there are many random variables with moment generating functions that are infinite for $t \geq 5$. For example, all $\text{Exp}(\lambda)$ distributions with $\lambda < 5$ have this property.
(d) We cannot identify the distribution of W, as there are many random variables where the moment generating function is equal to 2 at $t = 2$. Here are two examples: if $W_1 \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \frac{\ln 2}{2}$ then

$$M_{W_1}(2) = e^{\frac{\sigma^2 t^2}{2}} = e^{\frac{\frac{\ln 2}{2}(2^2)}{2}} = e^{\ln 2} = 2.$$

If $W_2 \sim \text{Poisson}(\lambda)$ with $\lambda = \frac{\ln 2}{e^2 - 1}$ then

$$M_{W_2}(2) = e^{\lambda(e^2-1)} = e^{\frac{\ln 2}{e^2-1}(e^2-1)} = e^{\ln 2} = 2.$$

5.5. We can recognize $M_X(t) = e^{3(e^t - 1)}$ as the moment generating function of a Poisson(3) random variable. Hence $P(X = 4) = e^{-3}\frac{3^4}{4!}$.

5.6. Then possible values of $Y = (X-1)^2$ are $1, 4$ and 9. The corresponding probabilities are

$$\begin{aligned}
P((X-1)^2 = 1) &= P(X = 0 \text{ or } X = 2) = P(X = 0) + P(X = 2) \\
&= \frac{1}{14} + \frac{3}{14} = \frac{2}{7} \\
P((X-1)^2 = 4) &= P(X = -1) = \frac{1}{7}, \\
P((X-1)^2 = 9) &= P(X = 4) = \frac{4}{7}.
\end{aligned}$$

5.7. The cumulative distribution function of X is $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise. Note that $X > 0$ with probability one, and $\ln(X)$ can take values from the whole $\mathbb{R}$.

We have

$$F_Y(y) = P(Y \leq y) = P(\ln(X) \leq y) = P(X \leq e^y) = 1 - e^{-\lambda e^y},$$

where we used $e^y > 0$. From this we get

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \left(1 - e^{-\lambda e^y}\right)' = \lambda e^{y - \lambda e^y}$$

for all $y \in \mathbb{R}$.

5.8. We first compute the cumulative distribution function of Y. Since $-1 \leq X \leq 2$, we have $0 \leq X^2 \leq 4$, thus $F_Y(y) = 1$ for $y \geq 4$ and $F_Y(y) = 0$ for $y < 0$.

For $0 \le y < 4$ we have

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Differentiating this we get the probability density function:

$$f_Y(y) = F_Y'(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}).$$

The probability density of X is $f_X(x) = \frac{1}{3}$ for $-1 \le x \le 2$ and zero otherwise. For $0 \le y \le 1$ then both $f_X(\sqrt{y})$ and $f_X(-\sqrt{y})$ is equal to $\frac{1}{3}$, and for $1 < y < 4$ we have $f_X(\sqrt{y}) = \frac{1}{3}$ and $f_X(-\sqrt{y}) = 0$.

From this we get

$$f_Y(y) = \begin{cases} \frac{1}{3\sqrt{y}} & \text{for } 0 \le y \le 1, \\ \frac{1}{6\sqrt{y}} & \text{for } 1 < y < 4, \\ 0 & \text{otherwise.} \end{cases}$$

5.9. (a) Using the probability mass function of the binomial distribution, and the binomial theorem:

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{tk} \\ &= \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1-p)^{n-k} \\ &= (e^t p + 1 - p)^n. \end{aligned}$$

(b) We have

$$\begin{aligned} E[X] &= M'(0) = npe^t \left(pe^t - p + 1\right)^{n-1} \Big|_{t=0} = np \\ E[X^2] &= M''(0) = (n-1)np^2 e^{2t} \left(pe^t - p + 1\right)^{n-2} + npe^t \left(pe^t - p + 1\right)^{n-1} \Big|_{t=0} \\ &= (n-1)np^2 + np. \end{aligned}$$

From these we get $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = (n-1)np^2 + np - n^2p^2 = np(1-p)$.

5.10. Using the Binomial Theorem we get

$$M(t) = \left(\frac{1}{5} + \frac{4}{5}e^t\right)^{30} = \sum_{k=0}^{30} \binom{30}{k} \left(\frac{4}{5}\right)^k e^{kt} \left(\frac{1}{5}\right)^{30-k}.$$

Since this is the sum of terms of the form $p_k e^{tk}$, we see that X is discrete. The possible values can be identified with the exponents: these are 0,1,2,…, 30. The coefficients are the corresponding probabilities:

$$P(X = k) = \binom{30}{k} \left(\frac{4}{5}\right)^k \left(\frac{1}{5}\right)^{30-k}, \qquad k = 0, 1, \ldots, 30.$$

We can recognize this as the probability mass function of a binomial distribution with $n = 30$ and $p = \frac{4}{5}$.

5.11. (a) The moment generating function is

$$M_X(t) = \int_{-\infty}^{\infty} f(x)e^{tX} = \int_0^{\infty} xe^{-(t-1)x}dx.$$

If $t-1 \leq 0$ then the integral is infinite. If $t-1>0$ then we can compute the integral by writing

$$\int_0^{\infty} xe^{-(t-1)x}dx = \frac{1}{t-1}\int_0^{\infty} x(t-1)e^{-(t-1)x}dx = \frac{1}{(t-1)^2}$$

where in the last step we recognized the integral to be the expectation of an Exp$(t-1)$ random variable. (One can also compute the integral by integrating by parts.)

Hence $M_X(t) = \frac{1}{(1-t)^2}$ for $t<1$, and $M_X(t)=\infty$ otherwise.

(b) Differentiating repeatedly:

$$M'(t) = \frac{2}{(1-t)^3}, \quad M''(t) = \frac{2\cdot 3}{(1-t)^4}, \quad M'''(t) = \frac{2\cdot 3\cdot 4}{(1-t)^5}.$$

Using mathematical induction one can show the general expression

$$M^{(n)} = \frac{2\cdot 3\cdots(n+1)}{(1-t)^{n+2}} = \frac{(n+1)!}{(1-t)^{n+2}},$$

from which we get

$$E[X^n] = M^{(n)}(0)(n+1)!.$$

5.12. We have

$$M(t) = \int_{-\infty}^{\infty} f(x)e^{tx}dx = \int_0^{\infty} \tfrac{1}{2}x^2e^{-x}e^{tx}dx = \int_0^{\infty} \tfrac{1}{2}x^2e^{(t-1)x}dx.$$

If $t \geq 1$ then $e^{(t-1)x} \geq 1$ for $x \geq 0$ and $M(t) \geq \int_0^{\infty}\frac{1}{2}x^2dx = \infty$. If $t<1$ then

$$\int_0^{\infty} \tfrac{1}{2}x^2e^{(t-1)x}dx = \frac{1}{2(1-t)}\int_0^{\infty} x^2(1-t)e^{-(1-t)x}dx = \frac{1}{2(1-t)}\cdot\frac{2}{(1-t)^2} = \frac{1}{(1-t)^3}.$$

The integral can be computed using integration by parts, or by recognizing it as the second moment of an Exp$(1-t)$ distributed random variable.

Thus we get

$$M(t) = \begin{cases} \frac{1}{(1-t)^3}, & \text{for } t<1 \\ \infty, & \text{otherwise.} \end{cases}$$

5.13. We can get $E[Y]$ by computing $M_Y'(0)$:

$$M_Y'(t) = -34\cdot\frac{1}{16}e^{-34t} - 5\cdot\frac{1}{8}e^{-5t} + 3\cdot\frac{1}{100}e^{3t} + 100\cdot\frac{121}{400}e^{100t}$$

and

$$E[Y] = M_Y'(0) = 27.53.$$

Since $M_Y(t)$ is of the form $\sum_k p_k e^{tk}$, we see that Y is discrete, the possible values are the numbers k for which $p_k \neq 0$ and p_k gives the probability $P(Y=k)$.

Hence the probability mass function of Y is

$$P(Y=0)=1/2, \quad P(Y=-34)=\frac{1}{16}, \quad P(Y=-5)=\frac{1}{8},$$
$$P(Y=3)=\frac{1}{100}, \quad P(Y=100)=\frac{121}{400}.$$

From this

$$E[Y]=0\cdot P(Y=0)+(-34)\cdot P(Y=-34)+(-5)\cdot P(Y=-5)$$
$$+3\cdot P(Y=3)+100\cdot P(Y=100)=27.53.$$

5.14. The probability mass function of X is

$$p_X(k)=\binom{4}{k}\frac{1}{2^4}, \quad k=0,1,\dots,4.$$

The possible values of X are $k=0,1,\dots,4$, which means that the possible values of Y are $0,1,4$. We have

$$P(Y=0)=P((X-2)^2=2)=P(X=2)=\binom{4}{2}\frac{1}{2^4}=\frac{3}{8}$$
$$P(Y=1)=P((X-2)^2=1)=P(X=3, \text{ or } X=1)=P(X=1)+P(X=3)$$
$$=\binom{4}{1}\frac{1}{2^4}+\binom{4}{3}\frac{1}{2^4}=\frac{1}{2}$$
$$P(Y=4)=P((X-2)^2=4)=P(X=4, \text{ or } X=0)=P(X=0)+P(X=4)$$
$$=\binom{4}{0}\frac{1}{2^4}+\binom{4}{4}\frac{1}{2^4}=\frac{1}{8}.$$

5.15. (a) We have

$$M_X(t)=\sum_k P(X=k)e^{tk}=\frac{1}{10}e^{-2t}+\frac{1}{5}e^{-t}+\frac{3}{10}+\frac{2}{5}e^{t}.$$

(b) The possible values of X are $\{-2,-1,0,1\}$, so the possible values of $Y=|X+1|$ are $\{0,1,2\}$. We get

$$P(Y=0)=P(X=-1)=\frac{3}{10}$$
$$P(Y=1)=P(X=-2)+P(X=0)=\frac{1}{10}+\frac{3}{10}=\frac{2}{5}$$
$$P(Y=2)=P(X=1)=\frac{2}{5}.$$

5.16. (a) We have $E[X^n]=\int_0^1 x^n\,dx=\frac{1}{n+1}$.
(b) In Exercise 5.3 we have seen that the moment generating function of X is given by the case defined function

$$M_X(t)=\begin{cases}1, & t=0\\ \frac{e^t-1}{t}, & t\neq 0.\end{cases}$$

We have $e^t = \sum_{k=0}^\infty \frac{t^k}{k!}$, hence $e^t - 1 = \sum_{k=1}^\infty \frac{t^k}{k!}$ and

$$M_X(t) = \frac{e^t-1}{t} = \frac{1}{t}\sum_{k=1}^\infty \frac{t^k}{k!} = \sum_{k=1}^\infty \frac{t^{k-1}}{k!} = \sum_{n=0}^\infty \frac{t^n}{(n+1)!}$$

for $t \neq 0$. In fact, this formula works for $t = 0$ as well, as the constant term of the series is equal to 1. Now we can read off the nth derivative at zero by taking the coefficient of t^n and multiplying by $n!$:

$$E[X^n] = M^{(n)}(0) = n! \cdot \frac{1}{(n+1)!} = \frac{1}{n+1}.$$

This agrees with the result we got for part (a).

5.17. (a) $M_X(0) = 1$. For $t \neq 0$ integrate by parts.

$$\begin{aligned}
M_X(t) &= E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \frac{1}{2}\int_0^2 x e^{tx}\,dx \\
&= \frac{1}{2}\left[\left(\frac{x}{t}e^{tx}\right)\Big|_{x=0}^{x=2} - \int_0^2 \frac{1}{t}e^{tx}\,dx\right] = \frac{1}{2}\left[\frac{2e^{2t}}{t} - \left(\frac{1}{t^2}e^{tx}\right)\Big|_{x=0}^{x=2}\right] \\
&= \frac{2te^{2t} - e^{2t} + 1}{2t^2}.
\end{aligned}$$

To summarize,

$$M_X(t) = \begin{cases} 1 & \text{for } t = 0, \\ \dfrac{2te^{2t} - e^{2t} + 1}{2t^2} & \text{for } t \neq 0. \end{cases}$$

(b) For $t \neq 0$ we insert the exponential series into $M_X(t)$ found in part (a) and then cancel terms:

$$\begin{aligned}
M_X(t) &= \frac{2te^{2t} - e^{2t} + 1}{2t^2} = \frac{1}{2t^2}\left(\sum_{k=0}^\infty \frac{(2t)^{k+1}}{k!} - \sum_{k=0}^\infty \frac{(2t)^k}{k!} + 1\right) \\
&= \frac{1}{2t^2}\sum_{k=2}^\infty (2t)^k \left(\frac{1}{(k-1)!} - \frac{1}{k!}\right) = \sum_{k=0}^\infty \frac{2^{k+1}}{k+2}\cdot\frac{t^k}{k!}
\end{aligned}$$

from which we read off $E(X^k) = M^{(k)}(0) = \frac{2^{k+1}}{k+2}$.

(c)

$$E(X^k) = \frac{1}{2}\int_0^2 x^{k+1}\,dx = \frac{2^{k+1}}{k+2}.$$

5.18. (a) Using the definition of a moment generating function we have

$$\begin{aligned}
M_X(t) &= E[e^{tX}] = \sum_{k=1}^\infty e^{tk} P(X=k) = \sum_{k=1}^\infty (e^t)^k (1-p)^{k-1} p \\
&= pe^t \sum_{k=1}^\infty (e^t(1-p))^{k-1} = pe^t \sum_{k=0}^\infty (e^t(1-p))^k
\end{aligned}$$

Note that the sum converges to a finite number if and only if $e^t(1-p)<1$, which holds if and only if $t<\ln\left(\frac{1}{1-p}\right)$. In this case we have

$$M_X(t)=pe^t\cdot\frac{1}{1-e^t(1-p)}.$$

Overall, we find:

$$M_X(t)=\begin{cases}\frac{pe^t}{1-e^t(1-p)} & t<\ln\left(\frac{1}{1-p}\right)\\ \infty & t\geq\ln\left(\frac{1}{1-p}\right)\end{cases}.$$

(b) For the mean,

$$\begin{aligned}E[X]=M_X'(0)&=\left.\frac{pe^t(1-e^t(1-p))-pe^t(-e^t(1-p))}{(1-e^t(1-p))^2}\right|_{t=0}\\&=\left.\frac{pe^t}{(1-e^t(1-p))^2}\right|_{t=0}\\&=\frac{p}{p^2}=\frac{1}{p}.\end{aligned}$$

For the variance we need the second moment,

$$\begin{aligned}E[X^2]=M_X''(0)&=\left.\frac{pe^t(1-e^t(1-p))^2-2pe^t(1-e^t(1-p))(-e^t(1-p))}{(1-e^t(1-p))^4}\right|_{t=0}\\&=\frac{p(1-(1-p))^2-2p(1-(1-p))(-(1-p))}{p^4}\\&=\frac{p^3-2p^3+2p^2}{p^4}=\frac{2}{p^2}-\frac{1}{p}.\end{aligned}$$

Finally the variance is

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=\frac{2}{p^2}-\frac{1}{p}-\frac{1}{p^2}=\frac{1}{p^2}-\frac{1}{p}.$$

5.19. (a) Since X is discrete, we get

$$M_X(t)=\sum_{k=0}^{\infty}P(X=k)e^{tk}=\frac{2}{5}+\frac{1}{5}\sum_{k=1}^{\infty}\left(\frac{3}{4}\right)^k e^{tk}=\frac{2}{5}+\frac{1}{5}\sum_{k=1}^{\infty}\left(\frac{3}{4}e^t\right)^k.$$

The geometric series is finite exactly if $\frac{3}{4}e^t<1$, which holds for $t\leq\ln(4/3)$. In that case

$$M_X(t)=\frac{2}{5}+\frac{1}{5}\sum_{k=1}^{\infty}\left(\frac{3}{4}e^t\right)^k=\frac{2}{5}+\frac{1}{5}\cdot\frac{\frac{3}{4}e^t}{1-\frac{3}{4}e^t}=\frac{8-3e^t}{20-15e^t}.$$

Hence

$$M_X(t)=\begin{cases}\frac{3e^t-8}{15e^t-20}, & t<\ln(4/3)\\ \infty & \text{else.}\end{cases}$$

(b) Differentiating $M_X(t)$ from part (a) we get

$$E[X] = M'(0) = \frac{15e^t\,(8-3e^t)}{(20-15e^t)^2} - \frac{3e^t}{20-15e^t}\Big|_{t=0} = \frac{12}{5}$$

$$E[X^2] = M''(0) = \frac{15e^t\,(8-3e^t)}{(20-15e^t)^2} + \frac{450e^{2t}\,(8-3e^t)}{(20-15e^t)^3} - \frac{3e^t}{20-15e^t} - \frac{90e^{2t}}{(20-15e^t)^2}\Big|_{t=0} = \frac{84}{5}.$$

From this we get

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 - \frac{84}{5} - (\frac{12}{5})^2 = \frac{276}{25}.$$

5.20. (a) From the definition we have

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{2}e^{-|x|}dx = \frac{1}{2}\int_0^{\infty} e^{-(1-t)x}dx + \frac{1}{2}\int_{-\infty}^{0} e^{(t+1)x}dx.$$

After the change of variables $x \to -x$ for the integral on $(-\infty, 0]$ we get

$$M_X(t) = \frac{1}{2}\int_0^{\infty} e^{-(1-t)x}dx + \frac{1}{2}\int_0^{\infty} e^{-(t+1)x}dx.$$

We have seen that the integral of $\int_0^\infty e^{-cx}dx$ is $\frac{1}{c}$ if $c > 0$ and ∞ otherwise. Thus $M_X(t)$ is finite if $1-t > 0$ and $1+t > 0$ (or $-1 < t < 1$) and ∞ otherwise. Moreover, if it is finite it is equal to

$$M_X(t) = \frac{1}{2}\cdot\frac{1}{1-t} + \frac{1}{2}\cdot\frac{1}{1+t} = \frac{1}{2(1-t^2)}.$$

Thus $M_X(t)$ is $\frac{1}{2(1-t^2)}$ for $|t| < 1$, and ∞ otherwise.
(b) We could try to differentiate $M_X(t)$ to get the moments, but it is simpler to take the Taylor expansion at $t = 0$. If $|t| < 1$ then $\frac{1}{1-t^2} = \sum_{k=0}^{\infty} t^{2k}$, hence

$$M_X(t) = \sum_{k=0}^{\infty}\frac{1}{2}t^{2k}.$$

The nth moment is the coefficient of t^n multiplied by $n!$. There are no odd exponent terms in the expansion, so all odd moments of X are zero. The term t^{2k} has a coefficient $\frac{1}{2}$, so the $(2k)$th moment is $\frac{(2k)!}{2}$.

5.21. We have

$$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = E[e^{bt+atX}] = e^{bt}E[e^{atX}] = e^{bt}M_X(at).$$

5.22. By the definition of the moment generating function and the properties of expectation we get

$$M_Y(t) = E[e^{tY}] = E[e^{(3X-2)t}] = E[e^{3tX}e^{-2t}] = e^{-2t}E[e^{3tX}].$$

Note that $E[e^{3tX}]$ is exactly the moment generating function $M_X(t)$ of X evaluated at $3t$. The moment generating function of $X \sim \mathrm{Exp}(\lambda)$ is $\frac{\lambda}{\lambda-t}$ for $t < \lambda$ and ∞ otherwise, thus $E[e^{3tX}] = \frac{\lambda}{\lambda-3t}$ for $t < \lambda/3$ and ∞ otherwise. This gives

$$M_Y(t) = \begin{cases} e^{-2t}\frac{\lambda}{\lambda-3t}, & \text{if } t < \lambda/3 \\ \infty, & \text{otherwise.} \end{cases}$$

5.23. We can notice that $M_Y(t)$ looks very similar to the moment generating function of a Poisson random variable. If $X \sim \text{Poisson}(2)$, then $M_X(t) = e^{2(e^t-1)}$, and $M_Y(t) = M_X(2t)$. From Exercise 5.21 we see that Y has the same moment generating function as $2X$, which means that they have the same distribution. Hence

$$P(Y = 4) = P(2X = 4) = P(X = 2) = e^{-2}\frac{2^2}{2!} = 2e^{-2}.$$

5.24. (a) Since $Y = e^X > 0$, we have $F_Y(t) = 0$ for $t \le 0$. For $t \le 0$,

$$F_Y(t) = P(Y \le t) = P(e^X \le t) = 0,$$

since $e^x > 0$ for all $x \in \mathbb{R}$. Next, for any $t > 0$

$$F_Y(t) = P(Y \le t) = P(e^X \le t) = P(X \le \ln t) = \Phi(\ln t).$$

Differentiating this gives the probability density function for $t > 0$:

$$f_Y(t) = \Phi'(\ln t)\frac{1}{t} = \frac{1}{t}\varphi(\ln t) = \frac{1}{\sqrt{2\pi t^2}}\exp\left(-\frac{(\ln(t))^2}{2}\right).$$

For $t \le 0$ the probability density function is 0.
(b) From the definition of Y we get that $E[Y^n] = E[(e^X)^n] = E[e^{nX}]$. Note that $E[e^{nX}] = M_X(n)$ is the moment generating function of X evaluated at n.

We computed the moment generating function for $X \sim \mathcal{N}(0, 1)$ and it is given by $M_X(t) = e^{t^2/2}$. Thus we have

$$E[Y^n] = e^{\frac{n^2}{2}}.$$

5.25. We start by expressing the cumulative distribution function $F_Y(y)$ of Y in terms of F_X. Since $Y = |X - 1| \ge 0$, we can concentrate on $y \ge 0$.

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(|X - 1| \le y) = P(-y \le X - 1 \le y) \\ &= P(1 - y \le X \le 1 + y) = F_X(1 + y) - F_X(1 - y). \end{aligned}$$

(In the last step we used $P(X = 1 - y) = 0$.) Differentiating the final expression:

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\left(F_X(1 + y) - F_X(1 - y)\right) = f_X(1 + y) + f_X(1 - y).$$

We have $f_X(x) = \frac{1}{5}$ if $-2 \le x \le 3$ and zero otherwise. Considering the various cases we get

$$f_Y(y) = \begin{cases} \frac{2}{5}, & 0 < y < 2 \\ \frac{1}{5}, & 2 \le y < 3 \\ 0 & \text{otherwise.} \end{cases}$$

5.26. The function $g(x) = x(x-3)$ is non-positive in $[0, 3]$ (as $0 \le x$ and $x - 3 \le 0$). It is a simple calculus exercise to show that the function $g(x))$ takes its minimum at $x = 3/2$ inside $[0, 3]$, and the minimum value is $-\frac{9}{4}$. Thus $Y = g(X)$ will take values from the interval $[-\frac{9}{4}, 0]$ and the probability density function $f_Y(y)$ is 0 for $y \notin [-\frac{9}{4}, 0]$.

We will determine the cumulative distribution function $F_Y(y)$ for $y \in [-\frac{9}{4}, 0]$. We have

$$F_Y(y) = P(Y \le y) = P(X(X - 3) \le y).$$

Next we solve the inequality $x(x-3) \le y$ for x. Since $x(x-3)$ is a parabola facing up, the solution will be an interval and the endpoints are exactly the solutions of $x(x-3) = y$. The solutions of this equation are

$$x_1 = \frac{3-\sqrt{9+4y}}{2}, \quad \text{and} \quad x_2 = \frac{3+\sqrt{9+4y}}{2},$$

thus for $-\frac{9}{4} \le y \le 0$ we get

$$\begin{aligned} F_Y(y) &= P(X(X-3) \le y) = P\left(\frac{3-\sqrt{9+4y}}{2} \le X \le \frac{3+\sqrt{9+4y}}{2}\right) \\ &= F_X\left(\tfrac{3+\sqrt{9+4y}}{2}\right) - F_X\left(\tfrac{3-\sqrt{9+4y}}{2}\right). \end{aligned}$$

Differentiating with respect to y gives

$$f_Y(y) = F_Y'(y) = \frac{1}{\sqrt{9+4y}} f_X\left(\tfrac{3+\sqrt{9+4y}}{2}\right) + \frac{1}{\sqrt{9+4y}} F_X\left(\tfrac{3-\sqrt{9+4y}}{2}\right).$$

Using the fact that $f_X(x) = \frac{2}{9}x$ for $0 \le x \le 3$ we obtain

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{9+4y}} \cdot \tfrac{2}{9}\left(\tfrac{3+\sqrt{9+4y}}{2}\right) + \frac{1}{\sqrt{9+4y}} \cdot \tfrac{2}{9} \cdot \left(\tfrac{3-\sqrt{9+4y}}{2}\right) \\ &= \frac{2}{9\sqrt{9+4y}}. \end{aligned}$$

Thus

$$f_Y(y) = \frac{2}{9\sqrt{9+4y}} \qquad \text{if } -\tfrac{9}{4} \le y \le 0$$

and 0 otherwise.
Finding the probability density via the Fact 5.27.
By Fact 5.27 we have

$$f_Y(y) = \sum_{x:g(x)=y, g'(x)\neq 0} f_X(x) \frac{1}{|g'(x)|}$$

with $g(x) = x(x-3)$. As we have seen before, if $0 \le x \le 3$ then $-\frac{9}{4} \le g(x) \le 0$. We also have $g'(x) = 2x - 3$. For $-\frac{9}{4} < y \le 0$ we have to possible x values with $g(x) = y$, these are the solutions x_1, x_2 found above. Then the formula gives

$$\begin{aligned} f_Y(y) &= f_X\left(\tfrac{3+\sqrt{9+4y}}{2}\right)\frac{1}{\sqrt{9+4y}} + f_X\left(\tfrac{3-\sqrt{9+4y}}{2}\right)\frac{1}{\sqrt{9+4y}} \\ &= \tfrac{2}{9}\left(\tfrac{3+\sqrt{9+4y}}{2}\right) \cdot \frac{1}{\sqrt{9+4y}} + \tfrac{2}{9} \cdot \left(\tfrac{3-\sqrt{9+4y}}{2}\right) \cdot \frac{1}{\sqrt{9+4y}} \\ &= \frac{2}{9\sqrt{9+4y}}. \end{aligned}$$

For y outside $[-\frac{9}{4}, 0]$ the probability density is 0 (and we can set it equal to zero for $y = -\frac{9}{4}$ as well).

5.27. We start by expressing the cumulative distribution function $F_Y(y)$ of Y in terms of F_X. Because $Y = e^X \ge 1$, we may assume $y \ge 1$.

$$F_Y(y) = P(Y \le y) = P(e^X \le y) = P(X \le \ln y) = F_X(\ln y).$$

Differentiating this we get

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X(\ln(y)) = f_X(\ln y)\frac{1}{y}.$$

The probability density function of X is $\lambda e^{-\lambda x}$ for $x \geq 0$ and zero otherwise. If $y > 1$ then $\ln y > 0$, hence in this case

$$f_Y(y) = \lambda e^{-\lambda \ln y} \frac{1}{y} = \lambda y^{-(\lambda+1)}.$$

For $y = 1$ we can set $f_Y(1) = 0$, so we get

$$f_Y(y) = \begin{cases} \lambda y^{-(\lambda+1)}, & y > 1 \\ 0 & \text{else.} \end{cases}$$

5.28. We have $f_X(x) = \frac{1}{3}$ for $-1 < x < 2$ and 0 otherwise. $Y = X^4$ takes values from $[0, 16]$, thus $f_Y(y) = 0$ outside this interval. For $0 < y \leq 16$ we have

$$F_Y(y) = P(Y \leq y) = P(X^4 \leq y) = P(-\sqrt[4]{y} \leq X \leq \sqrt[4]{y}) = F_X(\sqrt[4]{y}) - F_X(-\sqrt[4]{y}).$$

Differentiating this gives

$$f_Y(y) = F_Y'(y) = \frac{1}{4} y^{-3/4} f_X(\sqrt[4]{y}) + \frac{1}{4} y^{-3/4} f_X(-\sqrt[4]{y}).$$

Note that for $0 < y < 1$ both $\sqrt[4]{y}$ and $-\sqrt[4]{y}$ are in $(-1, 2)$, hence $f_X(\sqrt[4]{y})$ and $f_X(-\sqrt[4]{y})$ are both equal to $\frac{1}{3}$. This gives

$$f_Y(y) = 2 \cdot \frac{1}{4} y^{-3/4} \cdot \frac{1}{3} = \frac{1}{6} y^{-3/4}, \qquad \text{if } 0 < y < 1.$$

If $1 \leq y < 16$ then $\sqrt[4]{y} \in (-1, 2)$, but $-\sqrt[4]{y} \neq (-1, 2)$ which gives

$$f_Y(y) = \frac{1}{4} y^{-3/4} \cdot \frac{1}{3} = \frac{1}{12} y^{-3/4}, \qquad \text{if } 1 \leq y < 16.$$

Collecting everything

$$f_Y(y) = \begin{cases} \frac{1}{6} y^{-3/4}, & \text{if } 0 < y < 1 \\ \frac{1}{12} y^{-3/4}, & \text{if } 1 \leq y < 16 \\ 0, & \text{otherwise.} \end{cases}$$

5.29. $Y = |Z| \geq 0$. For $y \geq 0$ we get

$$F_Y(y) = P(Y \leq y) = P(|Z| \leq y) = P(-y \leq Z \leq y) = \Phi(y) - \Phi(-y) = 2\Phi(y) - 1.$$

Hence for $y \geq 0$ we have

$$f_Y(y) = F'(y) = (2\Phi(y) - 1)' = 2\phi(y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

and $f_Y(y) = 0$ otherwise.

5.30. We present two approaches for the solution.
Finding the probability density via the cumulative distribution function.
The probability density function of X is $f_X(x) = \frac{1}{3\pi}$ on $[-\pi, 2\pi]$ and 0 otherwise.

The $\sin(x)$ function takes values between -1 and 1, and it will take all these values on $[-\pi, 2\pi]$. Thus the set of possible values of Y are the interval $[-1, 1]$.

We will compute the cumulative distribution function of Y for $-1 < y < 1$. By definition,

$$F_Y(y) = P(Y \le y) = P(\sin(X) \le y).$$

In the next step we have to solve the inequality $\{\sin(X) \le y\}$ for X. Note that $\sin(x)$ is not one-to-one on $[-\pi, 2\pi]$. In order to solve the inequality, it helps to consider two cases: $0 \le y < 1$ and $-1 < y < 0$. If $0 \le y < 1$ then the solution of the inequality is

$$\{\pi - \arcsin(y) \le X \le 2\pi\} \cup \{-\pi \le X \le \arcsin(y)\}$$

and we get

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\sin(X) \le y) \\ &= P(-\pi \le X \le \arcsin(y)) + P(\pi - \arcsin(y) \le X \le 2\pi) \\ &= F_X(\arcsin(y)) + (1 - F_X(\pi - \arcsin(y)) \end{aligned}$$

Differentiating this (recall that $(\arcsin(x))' = \frac{1}{\sqrt{1-x^2}}$) we get

$$f_Y(y) = f_X(\arcsin(y))\frac{1}{\sqrt{1-y^2}} + f_X(\pi - \arcsin(y))\frac{1}{\sqrt{1-y^2}} = \frac{2}{3\pi\sqrt{1-y^2}}$$

(Note that $\arcsin(y)$ and $\pi - \arcsin(y)$ are both in $[-\pi, 2\pi]$.)

If $-1 < y < 0$ then the solution of the inequality is

$$\{-\pi - \arcsin(y) \le X \le \arcsin(y)\} \cup \{\pi - \arcsin(y) \le X \le 2\pi + \arcsin(y)\}$$

and we get

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\sin(X) \le y) \\ &= P(-\pi - \arcsin(y) \le X \le \arcsin(y)) + P(\pi - \arcsin(y) \le X \le 2\pi + \arcsin(y)) \\ &= F_X(\arcsin(y)) - F_X(-\pi - \arcsin(y)) + F_X(2\pi + \arcsin(y)) - F_X(\pi - \arcsin(y)) \end{aligned}$$

Differentiating this (and again using $(\arcsin(x))' = \frac{1}{\sqrt{1-x^2}}$) we get

$$\begin{aligned} f_Y(y) &= f_X(\arcsin(y))\frac{1}{\sqrt{1-y^2}} + f_X(-\pi - \arcsin(y))\frac{1}{\sqrt{1-y^2}} \\ &\quad + f_X(2\pi + \arcsin(y))\frac{1}{\sqrt{1-y^2}} + f_X(\pi - \arcsin(y))\frac{1}{\sqrt{1-y^2}} \\ &= \frac{4}{3\pi\sqrt{1-y^2}} \end{aligned}$$

This gives

$$f_Y(y) = \begin{cases} \frac{4}{3\pi\sqrt{1-y^2}}, & -1 < y < 0 \\ \frac{2}{3\pi\sqrt{1-y^2}}, & 0 \le y < 1 \\ 0, & |y| \ge 1 \end{cases}$$

Finding the probability density via the Fact 5.27.
By Fact 5.27 we have

$$f_Y(y) = \sum_{x: g(x)=y, g'(x)\neq 0} f_X(x)\frac{1}{|g'(x)|}$$

where $g(x) = \sin(x)$. Again, we only need to worry about the case $-1 \le y \le 1$, since Y can only take values from here. With a little bit of trigonometry you can check that the solutions of $\sin(x) = y$ for $|y| < 1$ are exactly the numbers

$$A_y = \{\arcsin(y) + 2\pi k, k \in \mathbb{Z}\} \cap \{\pi - \arcsin(y) + 2\pi k, k \in \mathbb{Z}\}.$$

Note that $g'(x) = \cos(x)$ and for any integer k

$$\frac{1}{|\cos(\arcsin(y) + 2\pi k)|} = \frac{1}{|\cos(\pi - \arcsin(y) + 2\pi k)|} = \frac{1}{\sqrt{1-y^2}}.$$

Since the density $f_X(x)$ is constant $\frac{1}{3\pi}$ on $[-\pi, 2\pi]$, we just need to check how many of the solutions from the set A_y are in this interval. It can be checked that there will be two solutions if $0 < y < 1$ and four solution for $-1 < y < 0$. (Sketching a graph of the sin function would help to visualize this.) Each one of these solutions will give a term $\frac{1}{3\pi\sqrt{1-y^2}}$ to the sum, so we get the case-defined function found wit the first approach.

5.31. We have $Y = e^{\frac{U}{1-U}} \ge 1$. For $y \ge 1$:

$$F_Y(y) = P(Y \le y) = P(e^{\frac{U}{1-U}} \le y) = P(\frac{U}{1-U} \le \ln y) = P(U \le \frac{\ln y}{\ln y + 1}) = \frac{\ln y}{\ln y + 1},$$

where we used $U \sim \text{Unif}[0,1]$ and $0 < \frac{\ln y}{\ln y + 1} < 1$. For $y > 1$ we have

$$f_Y(y) = F_Y'(y) = \tfrac{1}{y(1+\ln(y))^2},$$

and $f_Y(y) = 0$ otherwise.

5.32. The set of possible values of X is $(0,1)$, hence the set of possible values for Y is the interval $[1,\infty)$. Thus, for $t < 1$, $f_Y(t) = 0$. For $t \ge 1$,

$$P(Y \le t) = P(\tfrac{1}{X} \le t) = P(X \ge \tfrac{1}{t}) = 1 - \tfrac{1}{t}.$$

Differentiating now shows that $f_Y(t) = \frac{1}{t^2}$ when $t \ge 1$.

5.33. The following function will work:

$$g(u) = \begin{cases} 1 & \text{if} \quad 0 < u < 1/7 \\ 4 & \text{if} \quad 1/7 \le u < 3/7 \\ 9 & \text{if} \quad 3/7 \le u \le 1. \end{cases}$$

5.34. We can see from the conditions that

$$P(1 < X < 3) = P(1 < X < 2) + P(X = 2) = P(2 < X < 3) = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1,$$

hence we will need to find a function g that maps $(0,1)$ to $(1,3)$. The conditions show that inside the intervals $(1,2)$ and $(2,3)$ the random variable X 'behaves' like a random variable with probability density function $\frac{1}{3}$ there, but it also takes the value 2 with probability $\frac{1}{3}$ (so it actually cannot have a probability density function). We get $P(g(U) = 2) = \frac{1}{3}$ if the function g is constant 2 on an interval of length $\frac{1}{3}$ inside $(0,1)$. To get the behavior in $(1,2)$ and $(2,3)$ we can have linear functions there with slope 3. This leads to the following construction:

$$g(x) = \begin{cases} 1+3x, & \text{if } 0 < x \le \frac{1}{3} \\ 2, & \text{if } \frac{1}{3} < x \le \frac{2}{3} \\ 2+3(x-\frac{2}{3}), & \text{if } \frac{2}{3} < x < 1. \end{cases}$$

We can define g any way we want it to outside $(1,3)$.

To check that this function works note that

$$P(g(U)=2) = P(\tfrac{1}{3} \le U \le \tfrac{2}{3}) = \frac{1}{3},$$

for $1 < a < 2$ we have

$$P(1 < g(U) < a) = P(1+3U < a) = P(U < \tfrac{1}{3}(a-1)) = \tfrac{1}{3}(a-1),$$

and for $2 < b < 3$ we have

$$P(b < g(U) < 3) = P(b < 2+3(U-\tfrac{2}{3})) = P(\tfrac{1}{3}(b-2)+\tfrac{2}{3} < U) = \tfrac{1}{3}-\tfrac{1}{3}(b-2) = \tfrac{1}{3}(3-b).$$

5.35. Note that $Y = \lfloor X \rfloor$ is an integer, and hence Y is discrete. Moreover, for an integer k we have $\lfloor X \rfloor = k$ if and only if $k \le X < 1$. Thus

$$P(\lfloor X \rfloor = k) = P(k \le X < k+1).$$

Since $X \sim \mathrm{Exp}(\lambda)$, we have $P(k \le X < k+1) = 0$ if $k \le -1$, and for $k \ge 0$:

$$P(k \le X < k+1) = \int_k^{k+1} \lambda e^{-\lambda y} dy = e^{-\lambda k} - e^{-\lambda(k+1)} = e^{-\lambda k}(1-e^{-\lambda}).$$

5.36. Note that $X \ge 0$ and thus the possible values of $\lfloor X \rfloor$ are $0,1,2,\dots$. To find the probability mass function, we have to compute $P(\lfloor X \rfloor = k)$ for all nonnegative integer k. Note that $\lfloor X \rfloor = k$ if and only if $k \le X < k+1$. Thus for $k \in \{0,1,\dots\}$ we have

$$\begin{aligned} P(\lfloor X \rfloor = k) = P(k \le X < k+1) &= \int_k^{k+1} \lambda e^{-\lambda t}\, dt \\ &= -e^{-\lambda t}\Big|_{t=k}^{t=k+1} = e^{-\lambda k} - e^{-\lambda(k+1)} \\ &= e^{-\lambda k}(1-e^{-\lambda}) = (e^{-\lambda})^k(1-e^{-\lambda}). \end{aligned}$$

Note that this implies the random variable $\lfloor X \rfloor + 1$ is geometric with a parameter of $e^{-\lambda}$.

5.37. Since $Y = \{X\}$, we have $0 \le Y < 1$. For $0 \le y < 1$ we have

$$F_Y(y) = P(Y \le y) = P(\{X\} \le y).$$

If $\{x\} \le y$ then $k \le x \le k+y$ for some integer k. Thus

$$P(\{X\} \le y) = \sum_k P(k \le X \le k+y) = \sum_k (F_X(k+y) - F_X(k)).$$

Since $X \sim \mathrm{Exp}(\lambda)$, we have $F_X(x) = 1 - e^{-\lambda x}$ for $x \ge 0$ and 0 otherwise. This gives

$$F_Y(y) = \sum_{k=0}^{\infty} \left(1 - e^{-\lambda(k+y)} - (1-e^{-\lambda k})\right) = \sum_{k=0}^{\infty} e^{-\lambda k}(1-e^{-\lambda y}) = \frac{1-e^{-\lambda y}}{1-e^{-\lambda}}.$$

Differentiating this gives

$$f_Y(y) = \begin{cases} \frac{\lambda e^{-\lambda y}}{1-e^{-\lambda}}, & 0 \le y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

5.38. The cumulative distribution function of X can be computed from the probability density:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \begin{cases} 1 - \frac{1}{x}, & x > 1, \\ 0, & x \le 1. \end{cases}$$

We will look for a strictly increasing continuous function g. The probability density function of X is positive on $(1, \infty)$, thus the function g must map $(1, \infty)$ to $(0, 1)$.

If $g(X)$ is uniform on $[0, 1]$ then for any $0 < y < 1$ we have $P(g(X) \le y) = y$. If g is strictly increasing and continuous then there is a well-defined inverse function g^{-1} and we have

$$y = P(g(X) \le y) = P(X \le g^{-1}(y)).$$

Since g maps $(1, \infty)$ to $(0, 1)$, g^{-1} maps $(0, 1)$ to $(-1, \infty)$, which means $g^{-1}(y) > 1$ and

$$y = P(X \le g^{-1}(y)) = 1 - \frac{1}{g^{-1}(y)}.$$

This gives $y = 1 - \frac{1}{g^{-1}(y)}$. By substituting $y = g(x)$ and we get $g(x) = 1 - \frac{1}{x}$ for $1 < x$. We can define g any way we want for $x \le 1$.

Solutions to Chapter 6

6.1. (a) We just need to compute the row sums to get $P(X = 1) = 0.3$, $P(X = 2) = 0.5$, and $P(X = 3) = 0.2$.
(b) The possible values for $Z = XY$ are $\{0, 1, 2, 3, 4, 6, 9\}$ and the probability mass function is

$$\begin{aligned}
&P(Z = 0) = P(Y = 0) = 0.35\\
&P(Z = 1) = P(X = 1, Y = 1) = 0.15\\
&P(Z = 2) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0.05\\
&P(Z = 3) = P(X = 1, Y = 3) + P(X = 3, Y = 1) = 0.05\\
&P(Z = 4) = P(X = 2, Y = 2) = 0.05\\
&P(Z = 6) = P(X = 2, Y = 3) + P(X = 3, Y = 2) = 0.2 + 0.1 = 0.3\\
&P(Z = 9) = P(X = 3, Y = 3) = 0.05.
\end{aligned}$$

(c) We can compute the expectation as follows:

$$\begin{aligned}
E[Xe^Y] &= \sum_{x=1}^{3}\sum_{y=0}^{3} xe^y\\
&= e^0 \cdot 0.1 + e^1 \cdot 0.15 + e^2 \cdot 0 + e^3 \cdot 0.05\\
&+ 2e^0 \cdot 0.2 + 2e^1 \cdot 0.05 + 2e^2 \cdot 0.05 + 2e^3 \cdot 0.2\\
&+ 3e^0 \cdot 0.05 + 3e^1 \cdot 0 + 3e^2 \cdot 0.1 + 3e^3 \cdot 0.05\\
&\approx 16.3365
\end{aligned}$$

6.2. (a) The marginal probability mass function of X is found by computing the row sums,

$$P(X = 1) = \frac{1}{3}, \quad P(X = 2) = \frac{1}{2}, \quad P(X = 3) = \frac{1}{6}.$$

Computing the column sums gives the probability mass function of Y,

$$P(Y=0)=\frac{1}{5},\quad P(Y=1)=\frac{1}{5},\quad P(Y=2)=\frac{1}{3},\quad P(Y=3)=\frac{4}{15}.$$

(b) First we find the combinations of X and Y where $X+Y^2\le 2$. These are $(1,0),(1,1)$, and $(2,0)$. So we have

$$\begin{aligned}P(X+Y^2\le 2)=&P(X=1,Y=0)+P(X=1,Y=1)+P(X=2,Y=0)\\ =&\frac{1}{15}+\frac{1}{15}+\frac{1}{10}=\frac{7}{30}.\end{aligned}$$

6.3. (a) Let (X_W,X_Y,X_P) denote the number of times the professor chooses white, yellow and purple chalks, respectively. Choosing the color of the chalk can be considered a trial with three possible outcomes (the three colors), and since the choices are independent the random vector (X_W,X_Y,X_P) has multinomial distribution with parameters $n=10$, $r=3$ and $p_W=0.5=1/2$, $p_Y=0.4=2/5$ and $p_P=0.1=1/10$. We can now compute the probability in question using the joint probability mass function of the multinomial:

$$P(X_W=5,X_Y=4,X_P=1)=\frac{10!}{5!4!1!}(\tfrac{1}{2})^5(\tfrac{2}{5})^4(\tfrac{1}{10})^1=\frac{63}{625}=0.1008.$$

(b) Using the same notations as in part (a) we need to compute $P(X_W=9)$. The marginal distribution of X_W is $\text{Bin}(10,1/2)$, since it counts the number of times in 10 trials we got a specific outcome (getting yellow chalk). Thus

$$P(X_W=9)=\binom{10}{9}(\tfrac{1}{2})^{10}=\frac{5}{512}\approx 0.009766.$$

6.4. (X,Y,Z,W) has a multinomial distribution with parameters $n=5$, $r=4$, $p_1=p_2=p_3=\frac{1}{8}$, and $p_4=\frac{5}{8}$. Hence, the joint probability mass function of (X,Y,Z,W) is

$$\begin{aligned}P(X=x,Y=y,Z=z,W=w)&=\frac{5!}{x!\,y!\,z!\,w!}\left(\frac{1}{8}\right)^x\left(\frac{1}{8}\right)^y\left(\frac{1}{8}\right)^z\left(\frac{5}{8}\right)^w\\ &=\frac{5!}{x!\,y!\,z!\,w!}\cdot\frac{5^w}{8^{x+y+z+w}},\end{aligned}$$

for those integers $x,y,z,w\ge 0$ satisfying $x+y+z+w=5$, and zero otherwise.

Let W be the number of times some sandwich other than salami, falafel, or veggie is chosen. Then (X,Y,Z,W) has a multinomial distribution with parameters $n=5$, $r=4$, $p_1=p_2=p_3=\frac{1}{8}$, and $p_4=\frac{5}{8}$.

6.5. (a)

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,dx\,dy&=\tfrac{12}{7}\int_0^1\left(\int_0^1(xy+y^2)\,dx\right)dy=\tfrac{12}{7}\int_0^1(\tfrac{1}{2}y+y^2)\,dy\\ &=\tfrac{12}{7}\left(\tfrac{1}{4}+\tfrac{1}{3}\right)=1.\end{aligned}$$

Since $f\ge 0$ by its definition and integrates to 1, it passes the test.

(b) Since $0\le X,Y\le 1$, the marginal density functions f_X and f_Y both vanish outside $[0,1]$.

For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \tfrac{12}{7}\int_0^1 (xy+y^2)\,dy = \tfrac{12}{7}(\tfrac{1}{2}x + \tfrac{1}{3})\,dy = \tfrac{6}{7}x + \tfrac{4}{7}.$$

For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \tfrac{12}{7}\int_0^1 (xy+y^2)\,dx = \tfrac{12}{7}(\tfrac{1}{2}y + y^2)\,dy = \tfrac{12}{7}y^2 + \tfrac{6}{7}y.$$

(c)

$$P(X<Y) = \iint_{x<y} f(x,y)\,dx\,dy = \tfrac{12}{7}\int_0^1 \left(\int_0^y (xy+y^2)\,dx\right) dy = \tfrac{12}{7}\int_0^1 \tfrac{3}{2}y^3\,dy$$
$$= \tfrac{12}{7}\cdot\tfrac{3}{8} = \tfrac{9}{14}.$$

(d)

$$E[X^2Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 y f(x,y)\,dx\,dy = \int_0^1\int_0^1 x^2y\,\tfrac{12}{7}(xy+y^2)\,dx\,dy$$
$$= \tfrac{12}{7}\int_0^1\int_0^1 (x^3y^2 + x^2y^3)\,dx\,dy = \tfrac{12}{7}\left(\tfrac{1}{4}\cdot\tfrac{1}{3} + \tfrac{1}{3}\cdot\tfrac{1}{4}\right) = \tfrac{2}{7}.$$

6.6. (a) The marginal of X is

$$f_X(x) = \int_0^{\infty} xe^{-x(1+y)}dy = xe^{-x}\int_0^{\infty} e^{-xy}dy = e^{-x},$$

for $x>0$ and zero otherwise. The marginal of Y is

$$f_Y(y) = \int_0^{\infty} xe^{-x(1+y)}dx = \frac{1}{(1+y)^2},$$

for $y>0$ and zero otherwise (use integration by parts).

(b) The expectation is

$$E[XY] = \int_0^{\infty}\int_0^{\infty} xy\cdot f(x,y)\,dx\,dy = \int_0^{\infty}\int_0^{\infty} x^2ye^{-x(1+y)}\,dy\,dx$$
$$= \int_0^{\infty} x^2e^{-x}\int_0^{\infty} ye^{-xy}\,dy\,dx = \int_0^{\infty} x^2e^{-x}\cdot\frac{1}{x^2}\,dx = \int_0^{\infty} e^{-x}dx = 1.$$

(c) The expectation is

$$E\left[\frac{X}{1+Y}\right] = \int_0^{\infty}\int_0^{\infty}\frac{x}{1+y}xe^{-x(1+y)}\,dx\,dy = \int_0^{\infty}\frac{1}{1+y}\int_0^{\infty} x^2e^{-x(1+y)}\,dx\,dy$$
$$= \int_0^{\infty}\frac{1}{1+y}\cdot\frac{2}{(1+y)^3}\,dy = 2\int_0^{\infty}\frac{1}{(1+y)^4}\,dy = \frac{2}{3}.$$

6.7. (a) The area of the triangle is 1/2, thus the joint density $f(x,y)$ is $\frac{1}{1/2} = 2$ inside the triangle and 0 outside. The triangle is the set $\{(x,y) : 0 \le x, 0 \le$

$y, x+y \le 1\}$, so we can also write

$$f(x,y) = \begin{cases} 2, & \text{if } 0 \le x, 0 \le y, x+y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

We can compute the marginal density of X by evaluating the integral $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$. If (x,y) is in the triangle then we must have $0 \le x \le 1$, so for values outside this interval $f_X(x) = 0$. If $0 \le x \le 1$ then $f(x,y) = 2$ for $0 \le y \le 1-x$ and thus in this case we have

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{1-x} 2dy = 2(2-x).$$

Thus

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

Similar computation shows that

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

(b) The expectation of X can be computed using the marginal density:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x2(1-x)dx = x^2 - \frac{2x^3}{3}\bigg|_{x=0}^{x=1} = \frac{1}{3}.$$

Similar computation gives $E[Y] = \frac{1}{3}$.
(c) To compute $E[XY]$ we need to integrate the function $xyf(x,y)$ on the whole plane, which in our case is the same as integrating $2xy$ on our triangle. We can write this double integral as two single variable integrals: for a given $0 \le x \le 1$ the possible y values are $0 \le y \le 1-x$ hence

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^{1-x} 2xy\,dy\,dx = \int_0^1 \left(xy^2\Big|_{y=0}^{y=1-x} \right) dx = \int_0^1 x(1-x)^2 dx \\ &= \frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}\Big|_{x=0}^{x=1} = \frac{1}{12}. \end{aligned}$$

6.8. (a) X and Y from Exercise 6.2 are not independent. For example, note that $P(X=3) > 0$ and $P(Y=2) > 0$, but $P(X=3, Y=2) = 0$.

(b) The marginals for X and Y from Exercise 6.5 are:
For $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \tfrac{12}{7} \int_0^1 (xy+y^2)\,dy = \tfrac{12}{7}(\tfrac{1}{2}x + \tfrac{1}{3})\,dy = \tfrac{6}{7}x + \tfrac{4}{7}.$$

For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \tfrac{12}{7} \int_0^1 (xy+y^2)\,dx = \tfrac{12}{7}(\tfrac{1}{2}y + y^2)\,dy = \tfrac{12}{7}y^2 + \tfrac{6}{7}y.$$

Thus, $f_X(x)f_Y(y) \neq f(x,y)$ and they are not independent. For example, $f_X(\frac{1}{4}) = \frac{11}{14}$ and $f_Y(\frac{1}{4}) = \frac{9}{28}$, so that $f_X(\frac{1}{4})f_Y(\frac{1}{4}) = \frac{99}{392}$. However, $f(\frac{1}{4},\frac{1}{4}) = \frac{3}{14}$.

(c) The marginal of X is

$$f_X(x) = \int_0^\infty xe^{-x(1+y)}dy = xe^{-x}\int_0^\infty e^{-xy}dy = e^{-x},$$

for $x > 0$ and zero otherwise. The marginal of Y is

$$f_Y(y) = \int_0^\infty xe^{-x(1+y)}dx = \frac{1}{(1+y)^2},$$

for $y > 0$ and zero otherwise. Hence, $f(x,y)$ is not the product of the marginals and X and Y are not independent.

(d) X and Y are not independent. For example, choose any point (x,y) contained in the square $\{(u,v) : 0 \leq u \leq 1,\ 0 \leq v \leq 1\}$, but *not* contained in the triangle with vertices $(0,0)$, $(1,0)$, $(0,1)$. Then $f_X(x) > 0$, $f_Y(y) > 0$, and so $f_X(x)f_Y(y) > 0$. However, $f(x,y) = 0$ (because the point is outside the triangle).

6.9. X is binomial with parameters 3 and 1/2, thus its probability mass function is $p_X(a) = \binom{3}{a}\frac{1}{8}$ for $a = 0,1,2,3$ and zero otherwise. The probability mass function of Y is $p_Y(b) = \frac{1}{6}$ for $b = 1,2,3,4,5,6$. Since X and Y are independent, the joint probability mass function is just the product of the individual probability mass functions which means that

$$p_{X,Y}(a,b) = p_X(a)p_Y(b) = \binom{3}{a}\tfrac{1}{48}, \quad \text{for } a \in \{0,1,2,3\} \text{ and } b \in \{1,2,3,4,5,6\}.$$

6.10. The marginals of X and Y are

$$f_X(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}, \qquad f_Y(y) = \begin{cases} 1, & y \in (0,1) \\ 0, & y \notin (0,1), \end{cases}$$

and because they are independent the joint density is their product

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \begin{cases} 1, & 0 < x < 1, \text{ and } 0 < y < 1 \\ 0, & \text{else.} \end{cases}$$

Therefore,

$$P(X < Y) = \iint_{x<y} f_{X,Y}(x,y)dxdy = \int_0^1\int_0^y 1dx\,dy = \int_0^1 y\,dy = \frac{1}{2}.$$

6.11. Because Y is uniform on $(1,2)$, the marginal density for $Y$is

$$f_Y(y) = \begin{cases} 1 & y \in (1,2) \\ 0 & \text{else} \end{cases}$$

By independence, the joint distribution of (X,Y) is therefore

$$f_{X,Y}(X,Y) = \begin{cases} 2x & 0 < x < 1, 1 < y < 2 \\ 0 & \text{else} \end{cases}$$

The required probability is

$$P(Y - X \geq \tfrac{3}{2}) = \int\int_{y-x \geq \frac{3}{2}} f_{XY}(x,y)\,dx\,dy$$
$$= \int_0^{\frac{1}{2}} \int_{x+\frac{3}{2}}^{2} 2x\,dy\,dx,$$

where you should draw a picture of the region to see why this is the case. Calculating the double integral yields:

$$P(Y - X \geq \tfrac{3}{2}) = \int_0^{\frac{1}{2}} \int_{x+\frac{3}{2}}^{2} 2x\,dy\,dx = \int_0^{1/2} 2x(\tfrac{1}{2} - x)\,dx = \tfrac{1}{24}.$$

6.12. $f_X(x) = 0$ if $x < 0$ and if $x > 0$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^{\infty} 2e^{-(x+2y)}\,dy = e^{-x}\int_0^{\infty} 2e^{-2y}\,dy = e^{-x}.$$

$f_Y(y) = 0$ if $y < 0$ and for $y > 0$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\infty} 2e^{-(x+2y)}\,dx = 2e^{-2y}\int_0^{\infty} e^{-\infty}\,dx = 2e^{-2y}.$$

Now note that $f(x,y)$ is the product of f_X and f_Y.

6.13. In Example 6.19 we computed the probability density functions f_X and f_Y, and these functions were positive on $(-r_0, r_0)$. If X and Y were independent then the joint density would be $f(x,y) = f_X(x)f_Y(y)$, a function that is positive on the square $(-r_0, r_0)^2$. But $f(x,y)$ is zero outside the disk D, which means that X and Y are not independent.

6.14. (a) $F(x,y) = \dfrac{\max\bigl(\min(a,x),0\bigr)\cdot\max\bigl(\min(b,y),0\bigr)}{ab}.$

(b) If (x,y) is not in the rectangle, then $F(x,y) = 0$ and $f(x,y) = 0$. When (x,y) is in the interior of the rectangle, (so that $0 < x < a$ and $0 < y < b$)

$$F(x,y) = \frac{\max\bigl(\min(a,x),0\bigr)\cdot\max\bigl(\min(b,y),0\bigr)}{ab} = \frac{\max(x,0)\cdot\max(y,0)}{ab} = \frac{xy}{ab}.$$

Hence,

$$\frac{\partial^2}{\partial x \partial y}F(x,y) = \frac{ab}{.}$$

6.15. We can express X and Y in terms of Z and W as $X = g(Z,W)$, $Y = h(Z,W)$ with $g(z,w) = z$ and $h(z,w) = \rho z + \sqrt{1-\rho^2}\,w$. Solving the equations

$$x = z, \quad y = \rho z + \sqrt{1-\rho^2}\,w$$

for z, w gives the inverse of the function $(g(z,w), h(z,w))$. The solution is

$$z = x, \quad w = \frac{y - \rho x}{\sqrt{1-\rho^2}},$$

thus the inverse of $(g(z,w), h(z,w))$ is the function $(q(x,y), r(x,y))$ with

$$q(x,y) = x, \qquad r(x,y) = \frac{y - \rho x}{\sqrt{1-\rho^2}}.$$

The Jacobian of $(q(x,y), r(x,y))$ with respect to x, y is

$$J(x,y) = \det\begin{bmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix} = \frac{1}{\sqrt{1-\rho^2}}.$$

Using Fact 6.41 we get the joint density of X and Y:

$$f_{X,Y}(x,y) = f_{Z,W}\left(x, \frac{y-\rho x}{\sqrt{1-\rho^2}}\right)\cdot\frac{1}{\sqrt{1-\rho^2}}.$$

Since Z and W are independent standard normals, we have $f_{Z,W}(z,w) = \frac{1}{2\pi}e^{-\frac{z^2+w^2}{2}}$. Thus

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{x^2+\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)^2}{2}}.$$

We can simplify the exponent of the exponential as follows:

$$-\frac{x^2+\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)^2}{2} = -\frac{x^2(1-\rho^2+\rho^2)+y^2-2\rho xy}{2(1-\rho^2)} = -\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}.$$

This shows that the joint probability density of X, Y is indeed the same as given in (6.28), and thus the pair (X, Y) has standard bivariate normal distribution with parameter ρ.

6.16. In terms of the polar coordinates (r, θ) the Cartesian coordinates (x, y) are expressed as

$$x = r\cos(\theta) \qquad \text{and} \qquad y = r\sin(\theta).$$

These equations give the coordinate functions of the inverse function $G^{-1}(r,\theta)$. The Jacobian is

$$J(r,\theta) = \det\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \det\begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

The joint density function of X, Y is $f_{X,Y}(x,y) = \frac{1}{\pi r_0^2}$ in D and 0 outside. Formula (6.32) gives

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(r\cos(\theta), r\sin(\theta))\,|J(r,\theta)| = \tfrac{1}{\pi r_0^2}r \quad \text{for} \quad (r,\theta)\in L.$$

This is exactly the joint density function obtained earlier in (6.26) of Example 6.37.

6.17. We can express (X, Y) as $(g(U,V), h(U,V))$ where $g(u,v) = uv$ and $h(u,v) = (1-u)v$. We can find the inverse of the function $(g(u,v), h(u,v))$ by solving the system of equations

$$x = uv, \qquad y = (1-u)v$$

for u and v. The solution is $u = \frac{x}{x+y}$, $v = x+y$, so the inverse of $(g(u,v), h(u,v))$ is the function $(q(x,y), r(x,y))$ with

$$q(x,y) = \frac{x}{x+y}, \qquad r(x,y) = x+y.$$

The Jacobian of $(q(x,y), r(x,y))$ with respect to x, y is

$$J(x,y) = \det\begin{bmatrix} \frac{y}{(x+y)^2} & -\frac{x}{(x+y)^2} \\ 1 & 1 \end{bmatrix} = \frac{y+x}{(x+y)^2} = \frac{1}{x+y}.$$

Using Fact 6.41 we get the joint density of X and Y:

$$f_{X,Y}(x,y) = f_{U,V}\left(\frac{x}{x+y}, x+y\right)\cdot\frac{1}{x+y}.$$

The joint density of (U,V) is given by

$$f_{U,V}(u,v) = f_U(u)f_V(v) = \lambda^2 v e^{-\lambda v}, \qquad \text{for} 0<u<1, 0<v$$

and zero otherwise. This gives

$$f_{X,Y}(x,y) = \lambda^2(x+y)e^{-\lambda(x+y)}\cdot\frac{1}{x+y} = \lambda^2 e^{-\lambda(x+y)}$$

for $0 < \frac{x}{x+y} < 1$ and $0 < x+y$, zero otherwise. This condition is equivalent to $0 < x$, $0 < y$, and the found joint density can be factorized as

$$f_{X,Y}(x,y) = \lambda e^{-\lambda x}\cdot\lambda e^{-\lambda y}.$$

This shows that X and Y are independent exponentials with parameter λ.

6.18. (a) The probability mass function can be visualized in tabular form

$X\backslash Y$	1	2	3	4
1	$\frac{1}{4}$	0	0	0
2	$\frac{1}{8}$	$\frac{1}{8}$	0	0
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	0
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

The terms are nonnegative and add to 1, which shows that $p_{X,Y}$ is a probability mass function.

(b) Adding the rows and columns gives the marginals. The marginal of X is

$$P(X=1)=\tfrac{1}{4},\quad P(X=2)=\tfrac{1}{4},\quad P(X=3)=\tfrac{1}{4},\quad P(X=4)=\tfrac{1}{4},$$

whereas the marginal of Y is

$$P(Y=1)=\tfrac{25}{48},\quad P(Y=2)=\tfrac{13}{48},\quad P(Y=3)=\tfrac{7}{48},\quad P(Y=4)=\tfrac{1}{16}.$$

(c)

$$\begin{aligned}P(X=Y+1) &= P(X=2,Y=1)+P(X=3,Y=2)+P(X=4,Y=3)\\ &= \tfrac{1}{8}+\tfrac{1}{12}+\tfrac{1}{16}=\tfrac{13}{48}.\end{aligned}$$

6.19. (a) By adding the probabilities in the respective rows we get $p_X(0) = \frac{1}{3}$, $p_X(1) = \frac{2}{3}$. By adding them in the appropriate columns we get the marginal probability mass function of Y: $p_Y(0) = \frac{1}{6}$, $p_Y(1) = \frac{1}{3}$, $p_Y(2) = \frac{1}{2}$.
(b) We have $p_{Z,W}(z,w) = p_Z(z)p_W(w)$ by the independence of Z and W. Using the probability mass functions from part (a) we get

$Z \backslash W$	0	1	2
0	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$

6.20. Note that the random variable X_1+X_2 counts the number of times that outcomes 1 or 2 occurred. This event has a probability of $\frac{1}{2}$. Hence, and similar to the argument made at the end of Example 6.10, $(X_1+X_2, X_3, X_4) \sim \text{Mult}(n, 3, \frac{1}{2}, \frac{1}{8}, \frac{3}{8})$. Therefore, for any pair of integers (k, ℓ) with $k+\ell \le n$

$$\begin{aligned} P(X_3 = k, X_4 = \ell) &= P(X_1 + X_2 = n - k - \ell, X_3 = k, X_4 = \ell) \\ &= \frac{n!}{(n-k-\ell)!\,k!\,\ell!} \left(\tfrac{1}{2}\right)^{n-k-\ell} \left(\tfrac{1}{8}\right)^{k} \left(\tfrac{3}{8}\right)^{\ell}. \end{aligned}$$

6.21. They are not independent. Both X_1 and X_2 can take the value n with positive probability. However, they cannot take it the same time, as $X_1 + X_2 \le n$. Thus

$$0 < P(X_1 = n)P(X_2 = n) \neq P(X_1 = n, X_2 = n) = 0$$

which shows that X_1 and X_2 are not independent.

6.22. The random variable $X_1 + X_2$ counts the number of times that outcomes 1 or 2 occurred. This event has a probability of $p_1 + p_2$. Therefore, $X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2)$.

6.23. Let X_g, X_r, X_y be the number of times we see a green ball, red ball, and yellow ball, respectively. Then, $(X_g, X_r, X_y) \sim \text{Mult}(4, 3, 1/3, 1/3, 1/3)$. We want the following probability,

$$\begin{aligned} P(X_g = 2, X_r = 1, X_y = 1) + P(X_g = 1, X_r = 2, X_y = 1) + P(X_g = 1, X_r = 1, X_y = 2) \\ = \tfrac{4!}{2!1!1!}\left(\tfrac{1}{3}\right)^2 \tfrac{1}{3}\tfrac{1}{3} + \tfrac{4!}{2!1!1!}\left(\tfrac{1}{3}\right)^2 \tfrac{1}{3}\tfrac{1}{3} + \tfrac{4!}{2!1!1!}\left(\tfrac{1}{3}\right)^2 \tfrac{1}{3}\tfrac{1}{3} \\ = \tfrac{4}{9}. \end{aligned}$$

6.24. The number of green balls chosen is binomially distributed with parameters $n = 3$ and $p = \frac{1}{4}$. Hence, the probability that exactly two balls are green and one is not green is

$$\binom{3}{2}\left(\frac{1}{4}\right)^2 \frac{3}{4} = \frac{9}{64}.$$

The same argument goes for seeing exactly two red balls, two yellow balls, or two white balls. Hence, the probability that exactly two balls are of the same color is

$$4 \cdot \frac{9}{64} = \frac{9}{16}.$$

6.25. (a) The possible values for X and Y are $0, 1, 2$. For each possible pair we compute the probability of the corresponding event, For example,

$$P(X = 0, Y = 0) = P\{(T, T, T)\} = 2^{-3}.$$

Similarly

$$\begin{aligned}
&P(X=0,Y=1)=P(\{(T,T,H)\})=2^{-3}\\
&P(X=0,Y=2)=0\\
&P(X=1,Y=0)=P(\{(H,T,T)\})=2^{-3}\\
&P(X=1,Y=1)=P(\{(H,T,H),(T,H,T)\})=2\times 2^{-3}=2^{-2}\\
&P(X=1,Y=2)=P(\{(T,H,H)\})=2^{-3}\\
&P(X=2,Y=1)=P(\{(H,H,T)\})=2^{-3}\\
&P(X=2,Y=2)=P(\{(H,H,H)\})=2^{-3}
\end{aligned}$$

and zero for every other value of X and Y.
(b) The discrete random variable XY can take values $\{0,1,2,4\}$. The probability mass function is found by considering the possible coin flip sequences for each value:

$$\begin{aligned}
P(XY=0)=&P(X=0,Y=0)+P(X=0,Y=1)+P(X=1,Y=0)=\tfrac{3}{8}\\
P(XY=1)=&P(X=1,Y=1)=\tfrac{1}{4}\\
P(XY=2)=&P(X=1,Y=2)+P(X=2,Y=1)=\tfrac{1}{4}\\
P(XY=4)=&P(X=2,Y=2)=\tfrac{1}{8}.
\end{aligned}$$

6.26. (a) By the setup of the experiment, X_A is uniformly distributed over $\{0,1,2\}$ whereas X_B is uniformly distributed over $\{1,2,\dots,6\}$. Moreover, X_A and X_B are independent. Hence, (X_A,X_B) is uniformly distributed over $\Omega=\{(k,\ell):0\le k\le 2, 1\le \ell\le 6\}$. That is, for $(k,\ell)\in\Omega$,

$$P((X_A,X_B)=(k,\ell))=\tfrac{1}{18}.$$

(b) The set of possible values of Y_1 is $\{0,1,2,3,4,5,6,8,10,12\}$ and the set of possible values of Y_2 is $\{1,2,3,4,5,6\}$. The joint distribution can be given in tabular form

$Y_1 \setminus Y_2$	1	2	3	4	5	6
0	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$
1	$\frac{1}{18}$	0	0	0	0	0
2	0	$\frac{2}{18}$	0	0	0	0
3	0	0	$\frac{1}{18}$	0	0	0
4	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0
5	0	0	0	0	$\frac{1}{18}$	0
6	0	0	$\frac{1}{18}$	0	0	$\frac{1}{18}$
8	0	0	0	$\frac{1}{18}$	0	0
10	0	0	0	0	$\frac{1}{18}$	0
12	0	0	0	0	0	$\frac{1}{18}$

For example,

$$P(Y_1=2,Y_2=2)=P(X_A=1,X_B=2)+P(X_A=2,X_B=1)=\tfrac{1}{18}+\tfrac{1}{18}.$$

(c) The marginals are found by summing along the rows and columns:

$$\begin{aligned} &P(Y_1 = 0) = \tfrac{6}{18}, \quad P(Y_1 = 1) = \tfrac{1}{18}, \quad P(Y_1 = 2) = \tfrac{2}{18} \\ &P(Y_1 = 3) = \tfrac{1}{18}, \quad P(Y_1 = 4) = \tfrac{2}{18}, \quad P(Y_1 = 5) = \tfrac{1}{18} \\ &P(Y_1 = 6) = \tfrac{2}{18}, \quad P(Y_1 = 8) = \tfrac{1}{18}, \quad P(Y_1 = 10) = \tfrac{1}{18} \\ &P(Y_1 = 12) = \tfrac{1}{18}, \end{aligned}$$

and

$$\begin{aligned} &P(Y_2 = 1) = \tfrac{2}{18}, \quad P(Y_2 = 2) = \tfrac{4}{18}, \quad P(Y_2 = 3) = \tfrac{3}{18} \\ &P(Y_2 = 4) = \tfrac{3}{18}, \quad P(Y_2 = 5) = \tfrac{3}{18}, \quad P(Y_2 = 6) = \tfrac{3}{18}. \end{aligned}$$

The random variables Y_1 and Y_2 are not independent. For example,

$$P(Y_1 = 2, Y_2 = 6) = 0 \quad \text{whereas} \quad P(Y_1 = 2) > 0 \text{ and } P(Y_2 = 6) > 0.$$

6.27. The possible values of Y are $-1, 1$, which is the same as X_2. Thus, we need to show four things:

$$\begin{aligned} P(X_2 = 1, Y = 1) &= P(X_2 = 1)P(Y = 1) \\ P(X_2 = -1, Y = 1) &= P(X_2 = -1)P(Y = 1) \\ P(X_2 = 1, Y = -1) &= P(X_2 = 1)P(Y = -1) \\ P(X_2 = -1, Y = -1) &= P(X_2 = -1)P(Y = -1). \end{aligned}$$

To check the first one

$$P(X_2 = 1, Y = 1) = P(X_2 = 1, X_2X_1 = 1) = P(X_2 = 1, X_1 = 1) = P(X_2 = 1)P(X_1 = 1) = \tfrac{p}{2}.$$

Also,

$$P(Y = 1) = P(X_1 = 1, X_2 = 1) + P(X_1 = -1, X_2 = -1) = \tfrac{p}{2} + \tfrac{1}{2} \cdot (1 - p) = \tfrac{1}{2},$$

and so,

$$P(X_2 = 1)P(Y = 1) = p\tfrac{1}{2} = P(X_2 = 1, Y = 1).$$

All the other terms are handled similarly, using $P(Y = 1) = P(Y = -1) = 1/2$ and $P(X_2 = a, Y = b) = P(X_1 = b/a, X_2 = a)$.

6.28. To help with notation we will use $q = 1 - p$. For the joint probability mass function we need to compute $P(V = k, W = \ell)$ for all $k \geq 1, \ell = 0, 1, 2$. We have

$$\begin{aligned} P(V = k, W = 0) &= P(\min(X, Y) = k, X < Y) = P(X = k, k < Y) \\ &= P(X = k)P(k < Y) = pq^{k-1} \cdot q^k = pq^{2k-1}, \end{aligned}$$

where we used the independence of X and Y in the third equality. We get $P(V = k, W = 2) = pq^{2k-1}$ in exactly the same way. Finally,

$$P(V = k, W = 1) = P(\min(X, Y) = k, X = Y) = P(X = k, Y = k) = p^2q^{2k-2}.$$

This gives us the joint probability mass function of V and W; for the independence we need to check if this is the product of the marginals.

By Example 6.31 we have $V \sim \text{Geom}(1 - q^2)$ so for any $k \in \{1, 2, \dots\}$ we get

$$P(V = k) = (1 - (1 - q^2))^{k-1}(1 - q^2) = q^{2k-2}(1 - q^2).$$

The probability mass function of W is also easy to compute. By symmetry we must have

$$P(W=0)=P(X<Y)=P(Y<X)=P(W=2).$$

Also, by the independence of X and Y,

$$\begin{aligned}P(W=1)=P(X=Y)&=\sum_{k=1}^{\infty}P(X=k,Y=k)=\sum_{k=1}^{\infty}P(X=k)P(Y=k)\\&=\sum_{k=1}^{\infty}pq^{k-1}\cdot pq^{k-1}=p^2\sum_{k=0}^{\infty}(q^2)^k=\frac{p^2}{1-q^2}\\&=\frac{p}{2-p}.\end{aligned}$$

Combining the above with the fact that $P(W=0)+P(W=1)+P(W=2)=1$ gives

$$P(W=0)=P(W=2)=\tfrac{1}{2}(1-P(W=1))=\frac{1-p}{2-p}.$$

Now we can check the independence of V and W. First note that

$$P(V=k)P(W=0)=q^{2k-2}(1-q^2)\tfrac{1-p}{2-p},\quad P(V=k,W=0)=pq^{2k-1},$$

and since $\frac{1-q^2}{2-p}=(1-q)(1+q)\frac{1}{1+q}=p$, we have

$$P(V=k)P(W=0)=P(V=k,W=0).$$

The same computation shows $P(V=k)P(W=2)=P(V=k,W=2)$. Finally,

$$P(V=k)P(W=1)=q^{2k-2}(1-q^2)\tfrac{p}{2-p},\quad P(V=k,W=1)=p^2q^{2k-2}$$

and using $\frac{1-q^2}{2-p}=p$ again we get

$$P(V=k)P(W=1)=P(V=k,W=1).$$

We showed that $P(V=k,W=\ell)=P(V=k)P(W=\ell)$ for all relevant k,ℓ, and this shows that V and W are independent.

6.29. Because of the independence, the joint probability mass function of X and Y is the product of the individual probability mass functions:

$$P(X=a,Y=b)=P(X=a)P(Y=b)=p(1-p)^{a-1}r(1-r)^{b-1},\qquad a,b\ge 1.$$

We can break up the $P(X < Y)$ as the sum of probabilities of events $\{X = a, Y = b\}$ with $b > a$:

$$\begin{aligned}
P(X < Y) &= \sum_{a<b} P(X = a)P(Y = b) = \sum_{a=1}^{\infty} \sum_{b=a+1}^{\infty} P(X = a)P(Y = b) \\
&= \sum_{a=1}^{\infty} P(X = a) \sum_{b=a+1}^{\infty} P(Y = b) = \sum_{a=1}^{\infty} P(X = a)P(Y > a) \\
&= \sum_{a=1}^{\infty} p(1-p)^{a-1}(1-r)^a = p(1-r)\sum_{a=1}^{\infty}(1-p)^{a-1}(1-r)^{a-1} \\
&= \frac{p(1-r)}{1-(1-p)(1-r)} = \frac{p - pr}{p + r - pr}.
\end{aligned}$$

6.30. Note the typo in the problem, it should say $P(X = Y+1)$, not $P(X+1 = Y)$.

For $k \geq 1$ and $\ell \geq 0$ the joint probability mass function of X and Y is

$$P(X = k, Y = \ell) = (1-p)^{k-1}p \cdot e^{-\lambda}\tfrac{\lambda^\ell}{\ell!}.$$

Breaking up $\{X = Y + 1\}$ into the disjoint union of smaller events $\{X = Y + 1\} = \cup_{k=0}^{\infty}\{X = k + 1, Y = k\}$. Thus

$$\begin{aligned}
P(X = Y + 1) &= \sum_{k=0}^{\infty} P(X = k + 1, Y = k) = \sum_{k=0}^{\infty}(1-p)^k p \cdot e^{-\lambda}\tfrac{\lambda^k}{k!} \\
&= pe^{-\lambda}\sum_{k=1}^{\infty}\frac{(\lambda(1-p))^k}{k!} \\
&= pe^{-\lambda}e^{\lambda(1-p)} = pe^{-p\lambda}.
\end{aligned}$$

For $P(X + 1 = Y)$ we need a couple of more steps to compute the answer. We start with $\{X + 1 = Y\} = \cup_{k=1}^{\infty}\{X = k, Y = k + 1\}$. Then

$$\begin{aligned}
P(X + 1 = Y) &= \sum_{k=1}^{\infty} P(X = k, Y = k + 1) = \sum_{k=1}^{\infty}(1-p)^{k-1}p \cdot e^{-\lambda}\tfrac{\lambda^{k+1}}{(k+1)!} \\
&= \tfrac{1}{(1-p)^2}pe^{-\lambda}\sum_{k=1}^{\infty}\frac{(\lambda(1-p))^{k+1}}{(k+1)!} = \tfrac{1}{(1-p)^2}pe^{-\lambda}\sum_{k=2}^{\infty}\frac{(\lambda(1-p))^k}{k!} \\
&= \tfrac{1}{(1-p)^2}pe^{-\lambda}\left(\sum_{k=0}^{\infty}\frac{(\lambda(1-p))^k}{k!} - 1 - \lambda(1-p)\right) \\
&= \tfrac{1}{(1-p)^2}pe^{-\lambda}\left(e^{\lambda(1-p)} - 1 - \lambda(1-p)\right) \\
&= \frac{p}{(1-p)^2}e^{-\lambda p} - \frac{p}{(1-p)^2}e^{-\lambda} - \frac{p\lambda}{(1-p)}e^{-\lambda}
\end{aligned}$$

6.31. We have $X_1 + X_2 + X_3 = 8$, and $0 \leq X_i$ for $i = 1, 2, 3$. Thus we have to find the probability $P(X_1 = a, X_2 = b, X_3 = c)$ for nonnegative integers a, b, c with $a + b + c = 8$. Imagining that all 45 balls are different (e.g. by numbering them) we get $\binom{45}{8}$ equally likely outcomes. Out of these $\binom{10}{a}\binom{15}{b}\binom{20}{c}$ outcomes produce a

red, b green and c yellow balls. Thus the joint probability mass function is

$$P(X_1 = a, X_2 = b, X_3 = c) = \frac{\binom{10}{a}\binom{15}{b}\binom{20}{c}}{\binom{45}{8}}$$

for $0 \le a$, $0 \le b$, $0 \le c$ and $a + b + c = 8$, and zero otherwise.

6.32. Note that N is geometrically distributed with $p = \frac{7}{9}$. Thus, for $n \ge 1$,

$$P(N = n) = (\tfrac{2}{9})^{n-1}\tfrac{7}{9}.$$

We turn to finding the joint probability mass function of N and Y. First, note that

$$\begin{aligned} P(Y = 1, N = n) &= P((n-1) \text{ white balls followed by a green ball}) \\ &= (\tfrac{2}{9})^{n-1}\tfrac{4}{9}. \end{aligned}$$

Similarly,

$$P(Y = 2, N = n) = (\tfrac{2}{9})^{n-1}\tfrac{3}{9}.$$

We can use the above to find the marginal of Y.

$$P(Y = 1) = \sum_{n=1}^{\infty} P(Y = 1, N = n) = \sum_{n=1}^{\infty} (\tfrac{2}{9})^{n-1}\tfrac{4}{9} = \tfrac{4}{9} \cdot \tfrac{1}{1-2/9} = \tfrac{4}{7}.$$

Similarly,

$$P(Y = 2) = \tfrac{3}{7}.$$

We see that Y and N are independent:

$$\begin{aligned} P(Y = 1)P(N = n) &= \tfrac{4}{7} \cdot (\tfrac{2}{9})^{n-1}\tfrac{7}{9} = (\tfrac{2}{9})^{n-1}\tfrac{4}{9} = P(Y = 1, N = n) \\ P(Y = 2)P(N = n) &= \tfrac{3}{7} \cdot (\tfrac{2}{9})^{n-1}\tfrac{7}{9} = (\tfrac{2}{9})^{n-1}\tfrac{3}{9} = P(Y = 2, N = n). \end{aligned}$$

The distribution of Y can be understood by noting that there are a total of 7 balls colored green or yellow, and the selection of one of the 4 green balls, conditioned on one of these 7 being chosen, is $\frac{4}{7}$.

6.33. Since $f(x, y)$ is positive only if $0 < y < 1$, we have $f_Y(y) = 0$ if $y \le 0$ or $y \ge 1$. For $0 < y < 1$, $f(x, y)$ is positive only if $y < x < 2 - y$, and so

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\,dx = \int_{y}^{2-y} f(x, y)\,dx = \int_{y}^{2-y} 3y(2 - x)\,dx \\ &= 6yx - \tfrac{3}{2}yx^2 \Big|_{x=y}^{x=2-y} = 6y - 6y^2. \end{aligned}$$

Thus

$$f_Y(y) = \begin{cases} 6y - 6y^2 & \text{if } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The joint density function is positive on the triangle

$$\{(x, y) : 0 < y < 1,\ y < x < 2 - y\}.$$

To calculate the probability that $X + Y \le 1$, we combine the restriction $x + y \le 1$ with the description of the triangle to find the region of integration. Some trial and error may be necessary to discover the easiest way to integrate.

$$\begin{aligned} P(X+Y \le 1) &= \iint\limits_{x+y\le 1} f(x,y)\,dx\,dy = \int_0^{1/2} \left(\int_y^{1-y} 3y(2-x)\,dx \right) dy \\ &= \int_0^{1/2} (\tfrac{9}{2}y - 9y^2)\,dy = \tfrac{3}{16}. \end{aligned}$$

6.34. (a) The area of D is $\frac{3}{2}$, and hence the joint p.d.f. is

$$f_{X,Y}(x,y) = \begin{cases} \frac{2}{3}, & (x,y) \in D \\ 0, & (x,y) \notin D. \end{cases}$$

The line segment from $(1,1)$ to $(2,0)$ that forms part of the boundary of D obeys the equation $y = 2 - x$. The marginal density functions are derived as follows. First for X.

For $x \le 0$ and $x \ge 2$, $\quad f_X(x) = 0$.

For $0 < x \le 1$, $\quad f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 \frac{2}{3}\,dy = \frac{2}{3}$.

For $1 < x < 2$, $\quad f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{2-x} \frac{2}{3}\,dy = \frac{4}{3} - \frac{2}{3}x$.

Let us check that this is a density function:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 \tfrac{2}{3}\,dx + \int_1^2 (\tfrac{4}{3} - \tfrac{2}{3}x)\,dx = 1,$$

so indeed it is.

Next the marginal density function of Y:

For $y \le 0$ and $y \ge 1$, $\quad f_Y(y) = 0$.

For $0 < y < 1$, $\quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{2-y} \frac{2}{3}\,dx = \frac{4}{3} - \frac{2}{3}y$.

(b)

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x\,f_X(x)\,dx = \int_0^1 \tfrac{2}{3}x\,dx + \int_1^2 (\tfrac{4}{3}x - \tfrac{2}{3}x^2)\,dx = \tfrac{7}{9}. \\ E[Y] &= \int_{-\infty}^{\infty} y\,f_Y(y)\,dy = \int_0^1 (\tfrac{4}{3}y - \tfrac{2}{3}y^2)\,dy = \tfrac{4}{9}. \end{aligned}$$

(c) X and Y are not independent. Their joint density is not a product of the marginal densities. Also, a picture of D shows that $P(X > \frac{3}{2}, Y > \frac{1}{2}) = 0$ because all points in D satisfy $x+y \le 2$. However, the marginal densities show that $P(X > \frac{3}{2}) \cdot P(Y > \frac{1}{2}) > 0$ so the probability of the intersection does not equal the product of the probabilities.

6.35. (a) Since f_{XY} is non-negative, we just need to prove that the integral of f_{XY} is 1:

$$\int f_{XY}(x,y)dxdy = \int_{0\le x\le y\le 2} \frac{1}{4}(x+y)dx\,dy = \frac{1}{4}\int_0^2 \Big(\int_0^y (x+y)dx\Big)dy$$
$$= \frac{1}{4}\int_0^2 \big(\frac{3}{2}y^2\big)dy = 1.$$

(b) We calculate the probability using the joint density function:

$$P\{Y<2X\} = \int_{0\le x\le y\le 2, y<2x} \frac{1}{4}(x+y)dxdy = \int_0^2 \Big(\int_{\frac{y}{2}}^y \frac{1}{4}(x+y)dx\Big)dy$$
$$= \frac{1}{4}\int_0^2 \big(\frac{3}{2}y^2 - \frac{5}{8}y^2\big)dy = \frac{7}{32}\int_0^2 y^2 dy = \frac{7}{32}\cdot\frac{8}{3} = \frac{7}{12}$$

(c) According to the definition, when $0\le y\le 2$:

$$f_Y(y) = \int f_{XY}(x,y)dx = \int_0^y \frac{1}{4}(x+y)dx = \frac{1}{4}\big(\frac{3}{2}y^2 - 0\big) = \frac{3}{8}y^2$$

Otherwise, the density function $f_{XY}(x,y) = 0$. Thus:

$$f_Y(y) = \begin{cases} \frac{3}{8}y^2 & y\in[0,2] \\ 0 & \text{else} \end{cases}$$

6.36. (a) We need to find c so that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$. For this we need to compute

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}-\frac{(x-y)^2}{2}}\,dx\,dy$$

We can decide whether we should integrate with respect to x or y first, and choosing y gives a slightly easier path.

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}-\frac{(x-y)^2}{2}}dy = e^{-\frac{x^2}{2}}\int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{2}}dy$$
$$= \sqrt{2\pi}e^{-\frac{x^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-y)^2}{2}}dy = \sqrt{2\pi}e^{-\frac{x^2}{2}}.$$

In the last step we could recognize the integral of the pdf of a $\mathcal{N}(x,1)$ distributed random variable. From this we get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}-\frac{(x-y)^2}{2}}dydx = \int_0^{\infty}\sqrt{2\pi}e^{-\frac{x^2}{2}}dx$$
$$= 2\pi\int_0^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = 2\pi.$$

In the last step we integrated the pdf of the standard normal. Hence, $c = \frac{1}{2\pi}$.

(b) We have basically computed f_X (without the constant c) in part (a) already.

$$f_X(x) = \int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-\frac{x^2}{2}-\frac{(x-y)^2}{2}}dy$$
$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-y)^2}{2}}dy = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Now we compute f_Y:

$$f_Y(y) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2}{2} - \frac{(x-y)^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} - \frac{(x-y)^2}{2}} dx.$$

We can complete the square in the exponent of the exponential:

$$-\frac{x^2}{2} - \frac{(x-y)^2}{2} = -\left(x^2 - xy - \tfrac{1}{2}y^2\right) = -(x - y/2)^2 - y^2/4,$$

and we can now compute the integral:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} - \frac{(x-y)^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-y/2)^2 - y^2/4} dx \\ &= \frac{1}{\sqrt{4\pi}} e^{-y^2/4} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x-y/2)^2} dx = \frac{1}{\sqrt{4\pi}} e^{-y^2/4}. \end{aligned}$$

In the last step we used the fact that $\frac{1}{\sqrt{\pi}} e^{-(x-y/2)^2}$ is the pdf of a $\mathcal{N}(y/2, 1)$ distributed random variable. It follows that $Y \sim \mathcal{N}(0, 2)$.

Thus $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 2)$.

(c) X and Y are not independent since there joint density function is not the same as the product of the marginal densities.

6.37. We want to find $f_X(x)$ for which $P(c < X < d) = \int_c^d f_X(x)dx$ for all $c < d$. Because the x-coordinate of any point in D is in (a, b), we can assume that $a < c < d < b$. In this case

$$A = \{c < X < d\} = \{(x, y) : c < x < d,\ 0 < y < h(x)\}.$$

Because we chose (X, Y) uniformly from D, we get $P(A) = \frac{\text{area}(A)}{\text{area}(D)}$. We can compute the areas by integration:

$$P(c < X < d) = P(A) = \frac{\int_c^d \int_0^{h(x)} dy dx}{\int_a^b \int_0^{h(x)} dy dx} = \frac{\int_c^d h(x) dx}{\int_a^b h(x) dx}.$$

We can rewrite the last expression as

$$P(c < X < d) = \int_c^d \frac{h(x)}{\int_a^b h(s) ds} dx$$

which shows that

$$f_X(x) = \begin{cases} \frac{h(x)}{\int_a^b h(s)ds}, & \text{if } a < x < b \\ 0, & \text{otherwise.} \end{cases}$$

6.38. The marginal of Y is

$$f_Y(y) = \int_0^{\infty} x e^{-x(1+y)} dx = \frac{1}{(1+y)^2},$$

for $y > 0$ and zero otherwise (use integration by parts). Hence,

$$E[Y] = \int_0^{\infty} \frac{y}{(1+y)^2} dy = \infty.$$

6.39. $F(p,q)$ is the probability corresponding to the quarter plane $\{(x,y) : x < p, y < q\}$. (Because X, Y are jointly continuous it does not matter whether we write $<$ or $\leq$.) Our goal is to get the probability of (X,Y) being in the rectangle $\{(x,y) : a < x < b, c < y < d\}$ using quarter planes probabilities. We start with the probability $F(b,d)$, this is the probability corresponding to the quarter plane with corner (b,d). If we subtract $F(a,d) + F(b,c)$ from this then we remove the probabilities of the quarter planes corresponding to (a,d) and (b,c), and we have exactly the rectangle $(a,b) \times (c,d)$ left. However, the probability corresponding to the quarter plane with corner (a,c) was subtracted twice (instead of once), so we have to add it back. This gives

$$P(a < X < b, c < Y < d) = F(b,d) - F(b,c) - F(a,d) + F(a,b).$$

6.40. First note that the relevant set of values is $s \in [0,2]$ since $0 \leq X + Y \leq 2$. The joint density function is positive on the triangle

$$\{(x,y) : 0 < y < 1,\ y < x < 2 - y\}.$$

To calculate the probability that $X + Y \leq s$, for $0 \leq s \leq 2$, we combine the restriction $x + y \leq s$ with the description of the triangle to find the region of integration. (A picture could help.)

$$\begin{aligned} P(X+Y \leq s) &= \iint\limits_{x+y\leq s} f(x,y)\,dx\,dy = \int_0^{s/2} \left(\int_y^{s-y} 3y(2-x)\,dx \right) dy \\ &= \int_0^{s/2} \left(-\tfrac{3}{2}\, s^2 y + 3\, s y^2 + 6\, sy - 12\, y^2 \right) dy \\ &= \frac{(3\,s - 12)\, s^3}{24} + \frac{\left(-\tfrac{3}{2}\, s^2 + 6\, s\right) s^2}{8}. \end{aligned}$$

Differentiating to give the density yields

$$f(s) = \frac{3}{4}s^2 - \frac{1}{4}s^3 \text{ for } 0 < s < 2, \text{ and zero elsewhere.}$$

6.41. Let A be the intersection of the ball with radius r centered at the origin and D. Because $r < h$, this is just the 'top' half of the ball. We need to compute $P((X,Y,Z) \in A)$, and because (X,Y,Z) is chosen uniformly from D this is just the ratio of volumes of D and A. The volume of D is $r^2 h \pi$ while the volume of A is $\frac{2}{3} r^3 \pi$, so the probability in question is $\frac{\frac{2}{3} r^3 \pi}{r^2 h \pi} = \frac{2r}{3h}$.

6.42. Drawing a picture is key to understanding the solution as there are multiple cases requiring the computation of the areas of relevant regions.

Note that $0 \leq X \leq 2$ and $0 \leq Z = X + Y \leq 5$. This means that for $x < 0$ or $z < 0$ we have

$$F_{X,Z}(x,z) = P(X \leq x, Z \leq z) = 0.$$

If x and z are both nonnegative then we can compute $P(X \leq x, Z \leq z) = P(X \leq x, X + Y \leq z)$ by integrating the joint density of X, Y on the region $A_{x,z} = \{(s,t) : s \leq x, s + t \leq z\}$. This is just the area of the intersection of $A_{x,z}$ and D divided by the area of D (which is 6). The rest of the solution boils down to identifying the region $A_{x,z} \cap D$ in various cases and finding the corresponding area.

If $0 \leq x \leq 2$ and z is nonnegative then we need to consider four cases:

- If $0 \le z \le x$ then $A_{x,z} \cap D$ is the triangle with vertices $(0,0)$, $(z,0)$, $(0,z)$, with area $\frac{z^2}{2}$.
- If $x < z \le 3$ then $A_{x,z} \cap D$ is a trapezoid with vertices $(0,0)$, $(x,0)$, $(0,z)$ and $(x, z-x)$. Its area is $\frac{x(2z-x)}{2}$.
- If $3 < z \le 3+x$ then $A_{x,z} \cap D$ is a pentagon with vertices $(0,0)$, $(x,0)$, $(x, z-x)$, $(z-3, 3)$ and $(0,3)$. Its area is $3x - \frac{(3+x-z)^2}{2}$
- If $3+x < z$ then $A_{x,z} \cap D$ is the rectangle with vertices $(0,0)$, $(x,0)$, $(x,3)$ and $(0,3)$, with area $3x$.

We get the corresponding probabilities by dividing the area of $A_{x,z} \cap D$ with 6. Thus for $0 \le x \le 2$ we have

$$F_{X,Z}(x,z) = \begin{cases} 0, & \text{if } z < 0 \\ \frac{z^2}{12}, & \text{if } 0 \le z \le x \\ \frac{x(2z-x)}{12}, & \text{if } x < z \le 3 \\ \frac{x}{2} - \frac{(3+x-z)^2}{12}, & \text{if } 3 < z \le 3+x \\ \frac{x}{2}, & \text{if } 3+x < z. \end{cases}$$

For $2 < x$ we get $P(X \le x, Z \le z) = P(X \le 2, Z \le z) = F_{X,Z}(2,z)$. Using the previous results, in this case we get

$$F(x,z) = \begin{cases} 0, & \text{if } z < 0 \\ \frac{z^2}{12}, & \text{if } 0 \le z \le 2 \\ \frac{(z-x)}{3}, & \text{if } 2 < z \le 3 \\ 1 - \frac{(5-z)^2}{12}, & \text{if } 3 < z \le 5 \\ 1, & \text{if } 5 < z. \end{cases}$$

6.43. Following the reasoning of Example 6.40,

$$f_{T,V}(u,v) = f_{X,Y}(u,v) + f_{X,Y}(v,u).$$

Substituting in the definition of $f_{X,Y}$ gives the answer

$$f_{T,V}(u,v) = \begin{cases} 2u^2v + \sqrt{v} + 2v^2u + \sqrt{u} & \text{if } 0 < u < v < 1 \\ 0 & \text{else.} \end{cases}$$

6.44. Drawing a picture of the cone would help with this problem. The joint density of the uniform distribution in the teepee is

$$f_{X,Y,Z}(x,y,z) = \begin{cases} \frac{1}{\text{vol(Cone)}} & \text{if } (x,y,z) \in \text{Cone} \\ 0 & \text{else .} \end{cases}$$

The volume of the cone is $\pi r^2 h/3$. Thus the joint density is,

$$f_{X,Y,Z}(x,y,z) = \begin{cases} \frac{3}{\pi r^2 h} & \text{if } (x,y,z) \in \text{Cone} \\ 0 & \text{else .} \end{cases}$$

To find the joint density of (X, Y) we must integrate out the Z variable. To do so, we switch to cylindrical variables. Let $(\tilde{R}, \Theta, Z)$ be the distance from the center of the teepee, angle, and height where the fly dies. The height that we must integrate depends where we are on the floor. That is, if we are in the middle of the teepee $\tilde{R} = 0$, we must integrate Z from $z = 0$ to $z = h$. If we are near the edge of the teepee, we only integrate a small amount, for example $z = 0$ to $z = \epsilon$. For an arbitrary radius $\tilde{R}'$, the height we must integrate to is $h' = (1 - \frac{\tilde{R}}{r})h$.

Then the integral we must compute is

$$f_{\tilde{R},\Theta}(r, \theta) = \int_0^{(1-\frac{\tilde{r}}{r})h} \frac{3}{\pi r^2 h}\, dz = \frac{3(1 - \frac{\tilde{r}}{r})}{\pi r^2}.$$

We can check that this integrates to one. Recall that we are integrating with respect to cylindrical coordinates and thus

$$\begin{aligned}\int\int_{\text{circle}} f_{X,Y}(x, y)\, dx\, dy &= \int_0^{2\pi} \int_0^{r} \frac{3(1 - \frac{\tilde{r}}{r})}{\pi r^2}\, \tilde{r}\, d\tilde{r}\, d\theta \\ &= \int_0^{2\pi} \frac{3(\frac{r^2}{2} - \frac{r^3}{3})}{\pi r^2}\, d\theta = \frac{3r^2 \frac{1}{6}}{\pi r^2}(2\pi) = 1.\end{aligned}$$

Thus, switching back to rectangular coordinates,

$$f_{X,Y}(x, y) = f_{R,\Theta}(\sqrt{x^2 + y^2}, \theta) = \frac{3(1 - \frac{\sqrt{x^2+y^2}}{r})}{\pi r^2}$$

for $x^2 + y^2 \le r^2$.

For the marginal in Z, consider the height to be z. Then we must integrate over the circle with radius $r' = r(1 - \frac{z}{h})$. Thus, in cylindrical coordinates,

$$f_Z(z) = \int_0^{2\pi} \int_0^{r(1-z/h)} \frac{3}{\pi r^2 h}\, \tilde{r}\, d\tilde{r}\, d\theta$$

which yields,

$$f_Z(z) = \int_0^{2\pi} \frac{3r^2(1 - z/h)^2}{2\pi r^2 h} d\theta = \frac{3}{h}\left(1 - \frac{z}{h}\right)^2.$$

6.45. We first note that

$$\begin{aligned}F_V(v) &= P(V \le v) = P(\max(X, Y) \le v) = P(X \le v, Y \le v) \\ &= P(X \le v)P(Y \le v) = F_X(v)F_Y(v).\end{aligned}$$

Differentiating this we get the p.d.f. of V:

$$f_V(v) = \frac{d}{dv} F_V(v) = \big(F_X(v)F_Y(v)\big)' = f_X(v)F_Y(v) + F_X(v)f_Y(v).$$

For the minimum we use

$$P(T > z) = P(\min(X, Y) > z) = P(X > z, Y > z) = P(X > z)P(Y > z),$$

then

$$\begin{aligned}F_T(z) &= P(T \le z) = 1 - P(T > z) = 1 - P(X > z)P(Y > z) \\ &= 1 - (1 - F_X(z))(1 - F_Y(z)),\end{aligned}$$

and

$$\begin{aligned} f_T(z) &= \left[1 - (1 - F_X(z))(1 - F_Y(z))\right]' \\ &= f_X(z)(1 - F_Y(z)) + f_Y(z)(1 - F_X(z)). \end{aligned}$$

We computed the probabilities of the events $\{\max(X, Y) \le v\}$ and $\{\min(X, Y) > z\}$ because these events can be written as intersections to take advantage of independence.

6.46. We know from (6.31) and the independence of X and Y that

$$f_{T,V}(t, v) = f_X(t)f_Y(v) + f_X(v)f_Y(t),$$

if $t < v$ and zero otherwise. The marginal of $T = \min(X, Y)$ is found by integrating the v variable:

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_{T,V}(t, v)dv = \int_t^{\infty} f_X(t)f_Y(v) + f_X(v)f_Y(t)\,dv \\ &= f_X(t)(1 - F_Y(t)) + f_Y(t)(1 - F_X(t)). \end{aligned}$$

Turning to $V = \max(X, Y)$, we integrate away the t variable:

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_{T,V}(t, v)dt = \int_{-\infty}^{v} f_X(t)f_Y(v) + f_X(v)f_Y(t)\,dt \\ &= f_Y(v)F_X(v) + f_X(v)F_Y(v). \end{aligned}$$

6.47. (a) We will write F_X for F to avoid confusion. We need

$$F_Z(z) = P(\min(X_1, \dots, X_n) \le z).$$

We would like to write this in terms of the intersections of independent events, so we consider the complement:

$$1 - P(\min(X_1, \dots, X_n) \le z) = P(\min(X_1, \dots, X_n) > z).$$

The minimum of a group of numbers is larger than z if and only if every number is larger than z:

$$\begin{aligned} P(\min(X_1, \dots, X_n) > z) &= P(X_1 > z, \dots, X_n > z) = P(X_1 > z) \cdots P(X_n > z) \\ &= (1 - P(X_1 \le z)) \cdots (1 - P(X_n \le z)) = (1 - F_X(z))^n. \end{aligned}$$

Thus

$$F_Z(z) = 1 - (1 - F_X(z))^n$$

For the cumulative distribution of the maximum we need

$$F_W(w) = P(\max(X_1, X_2, \dots, X_n) \le w).$$

The maximum of some numbers is at most w if and only if every number is at most w:

$$\begin{aligned} P(\max(X_1, X_2, \dots, X_n) \le w) &= P(X_1 \le w, \dots, X_n \le w) \\ &= P(X_1 \le w) \cdots P(X_n \le w) = F_X(w)^n. \end{aligned}$$

(b) We can find the density functions by differentiation (using the chain rule):

$$\begin{aligned} f_Z(z) &= \frac{d}{dz}F_Z(z) = \frac{d}{dz}\left(1-(1-F_X(z))^n\right) = nf_X(x)(1-F_X(x))^{n-1}, \\ f_W(w) &= \frac{d}{dw}F_W(w) = \frac{d}{dw}F_X(w)^n = nf_X(x)F_X(x)^{n-1}. \end{aligned}$$

6.48. Let $t > 0$. We will show that $P(Y > t) = e^{-(\lambda_1+\cdots+\lambda_n)t}$. Using the independence of the random variables we have

$$\begin{aligned} P(Y > t) &= P(\min(X_1, X_2, \dots, X_n) > t) = P(X_1 > t, X_2 > t, \dots, X_n > t) \\ &= \prod_{i=1}^{n} P(X_i > t) = \prod_{i=1}^{n} e^{-\lambda_i t} \\ &= e^{-(\lambda_1+\cdots+\lambda_n)t}. \end{aligned}$$

Hence, Y is exponentially distributed with parameter $\lambda_1 + \cdots + \lambda_n$.

6.49. In the setting of Fact 6.41, let $G(x,y) = (\min(x,y), \max(x,y))$ and $L = \{(t,v) : t < v\}$. When $x \neq y$ this function G is two-to-one. Hence we define two separate regions $K_1 = \{(x,y) : x < y\}$ and $K_2 = \{(x,y) : x > y\}$, so that G is one-to-one and onto L from both K_1 and K_2. The inverse functions are as follows: from L onto K_1 it is $(q_1(t,v), r_1(t,v)) = (t,v)$ and from L onto K_2 it is $(q_2(t,v), r_2(t,v)) = (v,t)$. Their Jacobians are

$$J_1(t,v) = \det\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1 \quad \text{and} \quad J_2(t,v) = \det\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1.$$

Let again w be an arbitrary function whose expectation we wish to compute.

$$\begin{aligned} E[w(U,V)] &= \int_{-\infty}^{\infty}\int_{\infty}^{\infty} w\big(\min(x,y), \max(x,y)\big)\, f_{X,Y}(x,y)\,dx\,dy \\ &= \iint_{x<y} w(x,y) f_{X,Y}(x,y)\,dx\,dy + \iint_{y>x} w(y,x) f_{X,Y}(x,y)\,dx\,dy \\ &= \iint_L w(t,v)\, f_{X,Y}(q_1(t,v), r_1(t,v))\ |J_1(t,v)|\ dt\,dv \\ &\qquad + \iint_L w(t,v)\, f_{X,Y}(q_2(t,v), r_2(t,v))\ |J_2(t,v)|\ dt\,dv \\ &= \iint_{t<v} w(t,v)\left(f_{X,Y}(t,v) + f_{X,Y}(v,t)\right) dt\,dv. \end{aligned}$$

Since the diagonal $\{(x,y) : x = y\}$ has zero area it was legitimate to drop it from the first double integral. From the last line we can read off the joint density function $f_{T,V}(t,v) = f_{X,Y}(t,v) + f_{X,Y}(v,t)$ for $t < v$.

6.50. (a) Since $X \sim \text{Gamma}(r,\lambda)$ and $Y \sim \text{Gamma}(s,\lambda)$ are independent, we have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{x^{r-1}\lambda^r}{\Gamma(r)}e^{-\lambda x}\frac{y^{s-1}\lambda^s}{\Gamma(s)}e^{-\lambda y}$$

for $x > 0, y > 0$, and $f_{X,Y}(x,y) = 0$ otherwise.

In the setting of Fact 6.41, for $x, y \in (0,\infty)$ we are using the change of variables

$$u = g(x,y) = \frac{x}{x+y} \in (0,1), \quad v = h(x,y) = x + y \in (0,\infty).$$

The inverse functions are

$$q(u,v) = uv \in (0,\infty), \quad r(u,v) = v(1-u) \in (0,\infty).$$

The relevant Jacobian is

$$J(u,v) = \begin{vmatrix} \frac{\partial q}{\partial u}(u,v) & \frac{\partial q}{\partial v}(u,v) \\ \frac{\partial r}{\partial u}(u,v) & \frac{\partial r}{\partial v}(u,v) \end{vmatrix} = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v.$$

From this we get

$$\begin{aligned} f_{B,G}(u,v) &= f_X(uv) f_Y(v(1-u))v \\ &= \frac{\lambda^r (uv)^{r-1}}{\Gamma(r)} e^{-\lambda uv} \frac{\lambda^s (v(1-u))^{s-1}}{\Gamma(s)} e^{-\lambda(v(1-u))} v \\ &= \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} u^{r-1}(1-u)^{s-1} \cdot \frac{1}{\Gamma(r+s)} \lambda^{r+s} v^{(r+s)-1} e^{-\lambda v}. \end{aligned}$$

for $u \in (0,1), v \in (0,\infty)$, and 0 otherwise. We can recognize that this is exactly the product of a Beta(r,s) probability density (in u) and a Gamma$(r+s,\lambda)$ probability density (in v), hence $B \sim$ Beta(r,s), $G \sim$ Gamma$(r+s,\lambda)$, and they are independent.

(b) The transformation described is the inverse of that found in part (a). Therefore, X and Y are independent with $X \sim$ Gamma(r,λ) and $Y \sim$ Gamma(s,λ).

For the detailed solution note that

$$f_{B,G}(b,g) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} b^{r-1}(1-b)^{s-1} \cdot \frac{1}{\Gamma(r+s)} \lambda^{r+s} g^{(r+s)-1} e^{-\lambda g}$$

for $b \in (0,1), g \in (0,\infty)$ and it is zero otherwise.

We use the change of variables

$$x = b \cdot g, \qquad y = (1-b) \cdot g.$$

The inverse function is

$$b = \frac{x}{x+y}, \qquad g = x + y.$$

The Jacobian is

$$J(x,y) = \begin{vmatrix} \frac{y}{(x+y)^2} & -\frac{x}{(x+y)^2} \\ 1 & 1 \end{vmatrix} = \frac{1}{x+y}.$$

From this we get

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} (\tfrac{x}{x+y})^{r-1}(1 - \tfrac{x}{x+y})^{s-1} \cdot \frac{1}{\Gamma(r+s)} \lambda^{r+s} (x+y)^{(r+s)-1} e^{-\lambda(x+y)} \frac{1}{x+y} \\ &= \frac{x^{r-1}\lambda^r}{\Gamma(r)} e^{-\lambda x} \frac{y^{s-1}\lambda^s}{\Gamma(s)} e^{-\lambda y} \end{aligned}$$

for $x > 0, y > 0$ (and zero otherwise). This shows that indeed X and Y are independent with $X \sim$ Gamma(r,λ) and $Y \sim$ Gamma(s,λ).

6.51. (a) Apply the two-variable expectation formula to the function $h(x,y) = g(x)$. Then

$$E[g(X)] = E[h(X,Y)] = \sum_{k,\ell} h(k,\ell)P(X=k,Y=\ell) = \sum_{k,\ell} g(k)P(X=k,Y=\ell)$$
$$= \sum_k g(k) \sum_\ell P(X=k,Y=\ell) = \sum_k g(k)P(X=k).$$

(b) Similarly with integrals:

$$E[g(X)] = E[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\, f_{X,Y}(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} g(x)\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\right)dx = \int_{-\infty}^{\infty} g(x)\, f_X(x)\,dx.$$

6.52. For any $t_1,\dots,t_r \in \mathbb{R}$ we have

$$E\big[e^{t_1X_1+\cdots+t_rX_r}\big] = \sum_{k_1+k_2+\cdots+k_r=n} e^{t_1k_1+\cdots+t_rk_r}\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r}$$
$$= \sum_{k_1+k_2+\cdots+k_r=n} \frac{n!}{k_1!\cdots k_r!}\big(p_1e^{t_1}\big)^{k_1}\cdots\big(p_re^{t_r}\big)^{k_r}$$
$$= (p_1e^{t_1}+\cdots+p_re^{t_r})^n,$$

where the final step follows from the multinomial theorem.

6.53.

$$p_{X_1,\dots,X_m}(k_1,\dots,k_m) = P(X_1=k_1,\dots,X_m=k_m)$$
$$= \sum_{\ell_{m+1},\dots,\ell_n} P(X_1=k_1,\dots,X_m=k_m,X_{m+1}=\ell_{m+1},\dots,X_n=\ell_n)$$
$$= \sum_{\ell_{m+1},\dots,\ell_n} p_{X_1,\dots,X_m,\dots,X_n}(k_1,\dots,k_m,\ell_{m+1},\dots,\ell_n).$$

6.54. Let $X_1,\dots,X_n$ be jointly continuous random variables with joint density function f. Then for any $1\le m\le n$ the joint density function $f_{X_1,\dots,X_m}$ of random variables $X_1,\dots,X_m$ is

$$f_{X_1,\dots,X_m}(x_1,\dots,x_m) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\dots,x_m,y_{m+1},\dots,y_n)\,dy_{m+1}\dots\,dy_n.$$

Proof. One way to prove this is with the infinitesimal method. For $\varepsilon>0$ we have

$$P(X_1\in(x_1,x_1+\varepsilon),\dots,X_m\in(x_m,x_m+\varepsilon))$$
$$= \int_{x_1}^{x_1+\varepsilon}\cdots\int_{x_m}^{x_m+\varepsilon}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(y_1,\dots,y_n)\,dy_1\dots dy_n$$
$$\approx \left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\dots,x_m,y_{m+1},\dots,y_n)\,dy_{m+1}\dots dy_n\right)\varepsilon^m.$$

The result is shown by an application of Fact 6.39. $\square$

Another possible proof would be to express the joint cumulative distribution function of $X_1, \ldots, X_m$ as a multiple integral, and to read off the joint probability density function from that.

6.55. Consider the table for the joint probability mass function:

X_B \ X_D	0	1
0	0	a
1	b	$1-a-b$

We set $P(X_B = X_D = 0) = 0$ to make sure that a call comes. a and b are unknowns that have to satisfy $a \geq 0$, $b \geq 0$ and $a + b \leq 1$, in order for the table to represent a legitimate joint probability mass function.

(a) The given marginal p.m.f.s force the following solution:

X_B \ X_D	0	1
0	0	0.7
1	0.2	0.1

(b) There is still a solution when $P(X_D = 1) = 0.7$ but no longer when $P(X_D = 1) = 0.6$.

6.56. Pick an x for which $P(X = x) > 0$. Then,

$$0 < P(X = x) = \sum_y P(X = x, Y = y) = \sum_y a(x)b(y) = a(x) \sum_y b(y).$$

Hence, $\sum_y b(y) \neq 0$ and

$$a(x) = \frac{P(X = x)}{\sum_y b(y)}.$$

Similarly, for a y for which $P(Y = y) > 0$ we have

$$b(y) = \frac{P(Y = y)}{\sum_x a(x)}.$$

Combining the above we have

$$P(X = x, Y = y) = a(x)b(y) = \frac{P(X = x)P(Y = y)}{\sum_{\tilde{y}} b(\tilde{y}) \sum_{\tilde{x}} a(\tilde{x})}.$$

However, the denominator is equal to 1:

$$1 = \sum_{x,y} P(X = x, Y = y) = \sum_{x,y} a(x)b(y) = \sum_x a(x) \sum_y b(y),$$

and so the result is shown.

6.57. We can assume that $n \geq 2$. (If $n = 1$ then $Z = W = X_1$ and there is no joint density.)

Since all X_i are in $[0,1]$, this will be true for Z and W as well. We also know that the maximum is at least as large as the minimum: $P(Z \leq W) = 1$. We start by computing the probability $P(z < Z \leq W \leq w)$ for $0 \leq z < w \leq 1$. The maximum and minimum are between z and w if and only if all the numbers are between z and w. Thus

$$\begin{aligned} P(z < Z \leq W \leq w) &= P(z < X_1 \leq w, \ldots, z < X_n \leq w) \\ &= P(z < X_1 \leq w) \cdots P(z < X_n \leq w) \\ &= (w - z)^n. \end{aligned}$$

We would like to find the joint cumulative distribution function $F_{Z,W}(z,w) = P(Z \leq z, W \leq w)$. Because $0 \leq Z \leq W \leq 1$, it is enough to focus on $0 \leq z \leq w \leq 1$. Note that

$$P(z < Z \leq W \leq w) = P(W \leq w) - P(Z \leq z, W \leq w)$$

hence for $0 \leq z \leq w \leq 1$ we have

$$F_{Z,W}(z,w) = P(W \leq w) - (w - z)^n.$$

(This also holds for $w = z$, because then $P(Z \leq w, W \leq w) = P(W \leq w)$.) Taking the mixed partial derivatives gives the joint density (note that the $P(W \leq w)$ disappears when we differentiate with respect to z):

$$\begin{aligned} f_{Z,W}(z,w) &= \frac{\partial^2}{\partial z \partial w} F_{Z,W}(z,w) = \frac{\partial^2}{\partial z \partial w} (P(W \leq w) - (w - z)^n) \\ &= n(n-1)(w - z)^{n-2}. \end{aligned}$$

Thus $f_{Z,W}(z,w) = n(n-1)(w - z)^{n-2}$ if $0 \leq z < w \leq 1$ and zero otherwise.

Solutions to Chapter 7

7.1. We have

$$P(Z=3) = P(X+Y=3) = \sum_k P(X=k)P(Y=3-k).$$

Since X is Poisson, $P(X=k)=0$ for $k<0$. The random variable Y is geometric, hence $P(Y=3-k)=0$ if $3-k\le 0$. Thus $P(X=k)P(Y=3-k)$ is nonzero for $k=0,1$ and 2 and we get

$$\begin{aligned} P(Z=3) &= P(X=0)P(Y=3)+P(X=1)P(Y=2)+P(X=2)P(Y=1)\\ &= e^{-2}\tfrac{2}{3}\cdot(\tfrac{1}{3})^2 + 2e^{-2}\tfrac{2}{3}\cdot\tfrac{1}{3} + \tfrac{2^2}{2!}e^{-2}\tfrac{2}{3} = \frac{50}{27}e^{-2}. \end{aligned}$$

7.2. The possible values for both X and Y are 0 and 1, hence $X+Y$ can take the values $0,1$ and 2. If $X+Y=0$ then we must have $X=0$ and $Y=0$ and by independence we get

$$P(X+Y=0) = P(X=0,Y=0) = P(X=0)P(Y=0) = (1-p)(1-r).$$

Similarly, if $X+Y=2$ then we must have $X=1$ and $Y=1$:

$$P(X+Y=2) = P(X=1,Y=1) = P(X+1)P(Y=1) = pr.$$

We can now compute $P(X+Y)=1$ by considering the complement:

$$P(X+Y=1) = 1-P(X+Y=0)-P(X+Y=2) = 1-(1-p)(1-r)-pr = p+r-2pr.$$

We have computed the probability mass function of $X+Y$ which identifies its distribution.

7.3. Let X_1 and X_2 be the change in price tomorrow and the day after tomorrow. We know that X_1 and X_2 are independent, they have probability mass functions given by the table. We need to compute $P(X_1+X_2=2)$, which is given by

$$P(X_1+X_2=2) = \sum_k P(X_1=k)P(X_2=2-k).$$

Going through the possible values of k for which $P(X_1 = k) > 0$, and keeping only the terms for which $P(X_2 = 2 - k)$ is also positive:

$$\begin{aligned}P(X_1 + X_2 = 2) &= P(X_1 = -1)P(X_2 = 3) + P(X_1 = 0)P(X_2 = 2)\\&\quad + P(X_1 = 1)P(X_2 = 1) + P(X_1 = 2)P(X_2 = 0)\\&\quad + P(X_1 = 3)P(X_2 = -1)\\&= \tfrac{1}{64} + \tfrac{1}{64} + \tfrac{1}{16} + \tfrac{1}{64} + \tfrac{1}{64} = \tfrac{1}{8}\end{aligned}$$

7.4. We have

$$f_X(x) = \begin{cases}\lambda e^{-\lambda x}, & \text{if } x > 0\\ 0, & \text{otherwise,}\end{cases} \qquad f_Y(y) = \begin{cases}\mu e^{-\mu y}, & \text{if } y > 0\\ 0, & \text{otherwise.}\end{cases}$$

Since X and Y are both positive, $X+Y > 0$ with probability one, and $f_{X+Y}(z) = 0$ for $z \le 0$. For $z > 0$, using the convolution formula

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx = \int_0^z \lambda e^{-\lambda x}\mu e^{-\mu(z-x)}dx.$$

In the second step we used that $f_X(x)f_Y(z-x) \neq 0$ if and only if $x > 0$ and $z - x > 0$ which means that $0 < x < z$.

Returning to the integral

$$\begin{aligned}f_{X+Y}(z) &= \int_0^z \lambda e^{-\lambda x}\mu e^{-\mu(z-x)}dx = \lambda\mu e^{-\mu z}\int_0^z e^{(\mu-\lambda)x}dx\\&= \lambda\mu e^{-\mu z}\frac{e^{(\mu-\lambda)x}}{\mu-\lambda}\bigg|_{x=0}^{x=z} = \lambda\mu e^{-\mu z}\frac{e^{(\mu-\lambda)z}-1}{\mu-\lambda} = \lambda\mu\frac{e^{-\lambda z}-e^{-\mu z}}{\mu-\lambda}.\end{aligned}$$

Note that we used $\lambda \neq \mu$ when we integrated $e^{(\mu-\lambda)x}$.

Hence the probability density function of $X + Y$ is

$$f_{X+Y}(z) = \begin{cases}\lambda\mu\frac{e^{-\lambda z}-e^{-\mu z}}{\mu-\lambda}, & \text{if } z > 0\\ 0, & \text{otherwise.}\end{cases}$$

7.5. (a) By Fact 7.9 the distribution of W is normal, with

$$\mu_W = 2\mu_x - 4\mu_Y + \mu_Z = -7, \qquad \sigma_W^2 = \sigma_X^2 + 16\sigma_Y^2 + \sigma_Z^2 = 25.$$

Thus $W \sim \mathcal{N}(-7, 25)$.

(b) Using part (a) we know that $\frac{W+7}{\sqrt{25}}$ is a standard normal. Thus

$$P(W > -2) = P\left(\frac{W+7}{5} > \frac{-2+7}{5}\right) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587.$$

7.6. By exchangeability

$$\begin{aligned}&P(\text{3rd card is a king, 5th card is the ace of spades})\\&\qquad = P(\text{1st card is the ace of spades, 2nd card is king}).\end{aligned}$$

The second probability can now be computed by counting favorable outcomes within the first two picks:

$$P(\text{1st card is the ace of spades, 2nd card is king}) = \frac{1 \cdot 4}{\binom{52}{2}} = \frac{2}{663}.$$

7.7. By exchangeability

$$P(X_3 \text{ is the second largest}) = P(X_i \text{ is the second largest})$$

for any $i = 1, 2, 4$. Because the X_i are jointly continuous the probability that any two are equal is zero. Thus

$$1 = \sum_{i=1}^{4} P(X_i \text{ is the second largest}) = 4P(X_3 \text{ is the second largest})$$

and $P(X_3 \text{ is the second largest}) = \frac{1}{4}$.

7.8. Let X_k denote the color of the kth pick. Since the random variables $X_1, \dots, X_{10}$ are exchangeable, we have

$$\begin{aligned} P(X_3 = \texttt{green} \,|\, \texttt{X}_5 = \texttt{yellow}) &= \frac{P(X_3 = \texttt{green}, \texttt{X}_5 = \texttt{yellow})}{P(X_5 = \texttt{yellow})} \\ &= \frac{P(X_2 = \texttt{green}, \texttt{X}_1 = \texttt{yellow})}{P(X_1 = \texttt{yellow})} \\ &= P(X_2 = \texttt{green} \,|\, \texttt{X}_1 = \texttt{yellow}) = \frac{2}{7}. \end{aligned}$$

The fact that $P(X_3 = \texttt{green} \,|\, \texttt{X}_5 = \texttt{yellow}) = \frac{6}{21} = \frac{2}{7}$ follows by counting favorable outcomes, or noting that given that the first pick is yellow there are 6 out of the 21 balls left are green.

7.9. (a) The waiting time W_5 between the 4th and 5th call has Exp(6) distribution (with hours as units). Thus

$$P(W_5 < 10 \text{ min}) = P(W_5 < \tfrac{1}{6}) = 1 - e^{\frac{1}{6} \cdot 6} = 1 - e^{-1}.$$

(b) The waiting time between the 9th and 7th call is $W_8 + W_9$ where W_i is the waiting time between the $(i-1)$th and ith calls. These are independent exponentials with parameter 6. The sum of two independent Exp(6) distributed random variables has Gamma$(2, 6)$ distribution (see Example 7.29 and the discussion before that). Thus

$$P(W_8 + W_9 \le 15 \text{ min}) = P(W_8 + W_9 \le \tfrac{1}{4}) = \int_0^{\frac{1}{4}} 6^2 t e^{-6t} dt = 1 - \frac{5}{2} e^{-3/2}.$$

The final computation comes from the pdf of the gamma random variable and integration by parts. Alternatively, you can use the explicit cdf of the Gamma$(2, \lambda)$ distribution that we derived in Example 4.36.

7.10. By the memoryless property of the exponential distribution the waiting time until the first bulb replacement has distribution Exp$(\frac{1}{6})$ (where we use months as units). The waiting time from the first bulb replacement until the second one has the same Exp$(\frac{1}{6})$ distribution, and we can assume that it is independent of the first wait time. The same holds for the waiting time between the kth and $(k+1)$st bulb replacements. This means that the replacement times form a Poisson process with intensity $\frac{1}{6}$. Denoting the number of points in $[0, t)$ for the process by $N([0, t])$

we need to compute $P(N([0,3]) = 3)$. But $N([0,3])$ has Poisson distribution with parameter $3 \cdot \frac{1}{6} = \frac{1}{2}$, hence

$$\begin{aligned}
&P(\text{exactly 3 bulbs are replaced before the end of March}) \\
&\qquad = P(N([0,3]) = 3) = \frac{(1/2)^3}{3!} e^{-\frac{1}{2}} = \frac{e^{-\frac{1}{2}}}{48}
\end{aligned}$$

7.11. (a) Let X be the number of trials you perform and let Y be the number of trials I perform. Then, using that X and Y are independent Geom(p) and Geom(r) distributed random variables

$$\begin{aligned}
P(X = Y) &= \sum_{k=1}^{\infty} P(X = Y = k) = \sum_{k=1}^{\infty} P(X = k)P(Y = k) \\
&= \sum_{k=1}^{\infty} p(1-p)^{k-1} r(1-r)^{k-1} = pr \sum_{k=0}^{\infty} [(1-p)(1-r)]^k \\
&= pr \frac{1}{1-(1-p)(1-r)} = \frac{pr}{r+p-rp}.
\end{aligned}$$

(b) We have $Z = X + Y$. Thus, the range of Z is $\{2, 3, \dots\}$ and the probability mass function can be computed as

$$\begin{aligned}
P(Z = n) &= \sum_{i=1}^{n-1} P(X = i)P(Y = n-i) = \sum_{i=1}^{n-1} p(1-p)^{i-1} r(1-r)^{n-i-1} \\
&= pr \sum_{i=1}^{n-1} (1-p)^{i-1}(1-r)^{n-i-1} = pr \sum_{i=0}^{n-2} (1-p)^i (1-r)^{n-(i+1)-1} \\
&= pr(1-r)^{n-2} \sum_{i=0}^{n-2} \left[\frac{1-p}{1-r}\right]^i = pr(1-r)^{n-2} \frac{1-[(1-p)/(1-r)]^{n-1}}{1-(1-p)/(1-r)} \\
&= pr(1-r)^{n-1} \frac{1-[(1-p)/(1-r)]^{n-1}}{(1-r)-(1-p)} = pr \frac{(1-r)^{n-1}-(1-p)^{n-1}}{p-r}.
\end{aligned}$$

7.12. The probability mass function of Z is $p_Z(0) = 1-p$, $p_Z(1) = p$. The probability mass function of W is

$$p_W(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n.$$

The possible values of $Z + W$ are $0, 1, \dots, n+1$. Using the convolution formula we get

$$p_{Z+W}(k) = \sum_{\ell} p_Z(\ell) p_W(k-\ell).$$

We only need to evaluate this for $k = 0, 1, \dots, n+1$. Since $p_Z(\ell)$ is nonzero only for $\ell = 0$ and $\ell = 1$:

$$\begin{aligned} p_{Z+W}(k) &= p_Z(0)p_W(k) + p_Z(1)p_W(k-1) \\ &= (1-p)\cdot\binom{n}{k}p^k(1-p)^{n-k} + p\cdot\binom{n}{k-1}p^{k-1}(1-p)^{n-k+1} \\ &= \left(\binom{n}{k} + \binom{n}{k-1}\right)p^k(1-p)^{n+1-k}. \end{aligned}$$

In the last formula we used the convention that $\binom{n}{a} = 0$ if $a < 0$ or $a > n$. The final formula looks very similar to the probability mass function of a $\text{Bin}(n+1, p)$ distribution. In fact, it is exactly the same, as by Exercise C.11 we have $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$. Thus $Z + W \sim \text{Bin}(n+1, p)$.

Once we find (or conjecture) the answer, we can find a simpler argument. We can represent a $\text{Bin}(n, p)$ distributed random variable as the number of successes among n independent trials with success probability p. Now imagine that we have $n+1$ independent trials with success probability p. Denote the number of successes among the first n trials by $\tilde{W}$ and denote the outcome of the last trial by $\tilde{Z}$. Then $\tilde{Z} \sim \text{Ber}(p)$, $\tilde{W} \sim \text{Bin}(n, p)$ and these are independent (since the last trial is independent of the first n). But $\tilde{Z} + \tilde{W}$ counts the number of successes among the $n+1$ trials, so its distribution is $\text{Bin}(n+1, p)$. This shows that the sum of a $\text{Ber}(p)$ and and independent $\text{Bin}(n, p)$ distributed random variable is distributed as $\text{Bin}(n+1, p)$.

7.13. We could use the convolution formula, but it is easier to use the way we introduced the negative binomial distribution. (See the discussion before Definition 7.6.) If $Z_1, Z_2, \dots$ are independent $\text{Geom}(p)$ random variables, then adding n of them gives a $\text{Negbin}(n, p)$ distributed random variable. In particular, $Z_1 + \cdots + Z_k \sim \text{Negbin}(k, p)$ and $Z_{k+1} + \cdots + Z_m \sim \text{Negbin}(m, p)$ and these are independent. Thus $X + Y$ has the same distribution as $Z_1 + \cdots + Z_{m+n}$ which has $\text{Negbin}(k+m, p)$ distribution. Thus $X + Y$ has possible values $k+m, k+m+1, \dots$ and pmf

$$P(X+Y=n) = \binom{n-1}{k+m-1}p^{k+m}(1-p)^{n-k-m} \qquad \text{for } n \geq k+m.$$

7.14. Using the same notation as in Example 7.7 we get that

$$P(X=k) = \binom{k-1}{3}p^4(1-p)^{k-4}, \qquad k = 4, 5, 6, 7.$$

Evaluating $P(X = 6)$ for the various values of p gives the following numerical values:

p	0.40	0.35	0.30
P(Brewers win in 6)	0.09216	0.06340	0.03969

We also get

$$P(\text{Brewers win}) = \sum_{k=4}^{7} P(X=k) = \sum_{k=4}^{7}\binom{k-1}{3}p^4(1-p)^{k-4}.$$

Evaluating this sum for the various values of p gives the following numerical values:

p	0.40	0.35	0.30
$P(\text{Brewers win})$	0.2898	0.1998	0.1260

7.15. We have the following probability mass functions for X and Y:

$$p_X(k) = \frac{1}{n}, \quad \text{for } 1 \le k \le n, \quad \text{and} \quad p_Y(k) = \frac{1}{m}, \quad \text{for } 1 \le k \le m.$$

Both functions can be extended to all integers by setting them equal to zero outside the given domain. The domain of $X+Y$ is the set $\{2, 3, \dots, n+m\}$. The pmf can be computed using the convolution formula:

$$p_{X+Y}(a) = \sum_k p_X(k) p_Y(a-k).$$

The value of $p_X(k)p_Y(a-k)$ is either zero or $\frac{1}{mn}$, so we just have to compute the number of nonzero terms in the sum for a given $2 \le a \le n+m$. In order for $p_X(k)p_Y(a-k)$ to be nonzero we need $1 \le k \le n$ and $1 \le a-k \le m$. The second inequality gives $a-m \le k \le a-1$. Solving the system of inequalities by considering the 'worse' of the upper and lower bounds we get

$$\max(1, a-m) \le k \le \min(n, a-1).$$

There are $\min(n, a-1) - \max(1, a-m) + 1$ integer solutions to this inequality, so

$$p_{X+Y}(a) = \frac{1}{mn} \left(\min(n, a-1) - \max(1, a-m) + 1 \right), \text{ for } 2 \le a \le n+m.$$

By considering the cases $2 \le a \le n$, $n+1 \le a \le m+1$ and $m+2 \le a \le m+n$ separately, we can simplify the answer to get the following function:

$$p_{X+Y}(a) = \begin{cases} \frac{a-1}{mn} & 2 \le a \le n, \\ \frac{1}{m} & n+1 \le a \le m+1, \\ \frac{m+n+1-a}{mn} & m+2 \le a \le m+n. \end{cases}$$

7.16. The probability mass function of X is

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \dots$$

while the probability mass function of Y is $p_Y(0) = 1-p$, $p_Y(1) = p$. Using the convolution formula we get

$$p_{X+Y}(n) = \sum_k p_X(k) p_Y(n-k).$$

The possible values of $X+Y$ are $0, 1, 2, \dots$, so we only need to deal with $n \ge 0$. We only have $p_Y(n-k) \ne 0$ if $n-k=0$ or $n-k=1$ so we get

$$p_{X+Y}(n) = p_X(n)p_Y(0) + p_X(n-1)p_Y(1).$$

If $n=0$ then $p_X(n-1) = 0$, so

$$p_{X+Y}(0) = p_X(0)p_Y(0) = (1-p)e^{-\lambda}.$$

For $n > 0$ we get

$$p_{X+Y}(n) = p_X(n)p_Y(0) + p_X(n-1)p_Y(1) = (1-p)\frac{\lambda^n}{n!}e^{-\lambda} + p\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda}$$
$$= \lambda^{n-1}\frac{(\lambda(1-p)+np)}{n!}e^{-\lambda}.$$

Thus the probability mass function of $X + Y$ is

$$p_{X+Y}(n) = \begin{cases} (1-p)e^{-\lambda}, & \text{if } n = 0, \\ \lambda^{n-1}\frac{(\lambda(1-p)+np)}{n!}e^{-\lambda}, & \text{if } n \geq 1. \end{cases}$$

7.17. Let X be the the number of trials needed until we reach k successes, then $X \sim \text{Negbin}(k, p)$. The event that the number of successes reaches k before the number of failures reaches ℓ is the same as $\{X < k + \ell\}$. Moreover this event is the same as having at least k successes within the first $k + \ell - 1$ trials. Thus

$$P(X < k+\ell) = \sum_{j=0}^{\ell-1}\binom{k+j}{k-1}p^k(1-p)^j = \sum_{a=k}^{k+\ell-1}\binom{k+\ell-1}{a}p^a(1-p)^{k+\ell-1-a}.$$

7.18. Both X and Y have probability densities that are zero for negative values, this will hold for $X + Y$ as well. Using the convolution formula, for $z \geq 0$ we get

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_0^z f_X(x)f_Y(z-x)dx$$
$$= \int_0^z 2e^{-2x}4(z-x)e^{-2(z-x)}dx = \int_0^z 8(z-x)e^{-2z}dx$$
$$= 8e^{-2z}\int_0^z (z-x)dx = 4z^2e^{-2z}.$$

Thus

$$f_{X+Y}(z) = \begin{cases} 4z^2e^{-2z}, & \text{if } z \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

7.19. (a) We need to compute

$$P(Y \geq X \geq 2) = \iint_{y \geq x \geq 2} f_X(x)f_Y(y)\,dx\,dy = \int_2^\infty \int_x^\infty e^{-x-y}dydx$$
$$= \int_2^\infty e^{-2x}dx = \tfrac{1}{2}e^{-4}.$$

(b) The density of f_{-Y} is given by $f_{-Y}(y) = f_Y(-y)$. Then from the convolution formula we get

$$f_{X-Y}(z) = \int_{-\infty}^{\infty} f_X(t)f_{-Y}(z-t)dt = \int_{-\infty}^{\infty} f_X(t)f_{-Y}(z-t)dt = \int_{-\infty}^{\infty} f_X(t)f_Y(t-z)dt.$$

Note that $f_X(t)f_Y(t-z) > 0$ if $t > 0$ and $t - z > 0$, which is the same as $t > \max(z, 0)$. Thus

$$f_{X-Y}(z) = \int_{\max(z,0)}^{\infty} f_X(t)f_Y(t-z)dt = \int_{\max(z,0)}^{\infty} e^{-2t+z}dt = \frac{1}{2}e^{-2\max(z,0)+z}.$$

If $z \ge 0$ then this gives $\frac{1}{2}e^{-2\max(z,0)+z} = \frac{1}{2}e^{-z}$. If $z < 0$ then $\frac{1}{2}e^{-2\max(z,0)+z} = \frac{1}{2}e^{z}$. We can summarize these two cases with the formula $f_{X-Y}(z) = \frac{1}{2}e^{-|z|}$.

7.20. (a) Since X and Y are independent, we have $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ where

$$f_X(x) = \begin{cases} 2x, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} 1, & \text{if } 1 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

To compute $P(Y - X \ge \frac{3}{2})$ we need to integrate $f_{X,Y}(x,y)$ on the set $\{(x,y) : y - x \ge \frac{3}{2}\}$. Since $f_{X,Y}(x,y)$ is positive only if $0 < x < 1$ and $1 < y < 2$, it is enough to consider the intersection

$$\{(x,y) : y - x \ge \tfrac{3}{2}\} \cap \{(x,y) : 0 < x < 1, 1 < y < 2\}.$$

By sketching this region (or solving the inequalities) we get the region is the same as $\{(x,y) : 0 < x < 1/2, 3/2 + x < y < 2\}$.Thus we get

$$\begin{aligned} P(Y - X \ge \tfrac{3}{2}) &= \iint_{y-x\ge 3/2} f_{X,Y}(x,y)dxdy = \int_0^{1/2}\int_{3/2+x}^{2} 2xdydx \\ &= \int_0^{1/2} (1/2 - x)2xdx = \frac{1}{24}. \end{aligned}$$

(b) Note that X takes values in $(0,1)$, Y takes values in $(1,2)$ so $X+Y$ will take values in $(1,3)$. For a given $z \in (1,3)$ the convolution formula gives

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_0^1 f_X(x)f_Y(z-x)dx,$$

where we used the fact that $f_X(x) = 0$ outside $(0,1)$. For a given $1 < z < 3$ the function $f_Y(z-x)$ is nonzero if and only if $1 < z - x < 2$, which is equivalent to $z - 2 < x < z - 1$. Since we must have $0 < x < 1$ for $f_X(x)$ to be nonzero, this means that $f_X(x)f_Y(z-x)$ is nonzero only if $\max(0, z-2) < x < \min(1, z-1)$. Thus

$$\begin{aligned} f_{X+Y}(z) &= \int_0^1 f_X(x)f_Y(z-x)dx = \int_{\max(0,z-2)}^{\min(1,z-1)} 2xdx \\ &= \min(1, z-1)^2 - \max(0, z-2)^2. \end{aligned}$$

Considering the $1 < z \le 2$ and $2 < z < 3$ cases separately:

$$f_{X+Y}(z) = \begin{cases} (z-1)^2, & \text{if } 1 < z \le 2, \\ 1 - (z-2)^2, & \text{if } 2 < z < 3, \\ 0, & \text{otherwise.} \end{cases}$$

7.21. (a) By Fact 7.9 the distribution of W is normal, with

$$\mu_W = 3\mu_x + 4\mu_Y = 10, \qquad \sigma_W^2 = 9\sigma_X^2 + 16\sigma_Y^2 = 59.$$

Thus $W \sim \mathcal{N}(9, 57)$.

(b) Using part (a) we know that $\frac{W-10}{\sqrt{57}}$ is a standard normal. Thus

$$P(W > 15) = P\left(\frac{W-10}{\sqrt{57}} > \frac{15-10}{\sqrt{57}}\right) = 1 - \Phi(\tfrac{5}{\sqrt{57}}) \approx 1 - \Phi(0.66) \approx 0.2578.$$

7.22. Using Fact 3.61 we have $2X \sim \mathcal{N}(2\mu, 4\sigma^2)$. From Fact 7.9 by the independence of X and Y we get $X + Y \sim \mathcal{N}(2\mu, 2\sigma^2)$. Since $\sigma^2 > 0$, the two distributions can never be the same.

7.23. By Fact 7.9 $X - Y \sim \mathcal{N}(0, 2)$ and thus $\frac{X-Y}{\sqrt{2}} \sim \mathcal{N}(0, 1)$. From this we get

$$P(X > Y + 2) = P(\tfrac{X-Y}{\sqrt{2}} > \sqrt{2}) = 1 - \Phi(\sqrt{2}) \approx 1 - \Phi(1.41) \approx 0.0793.$$

7.24. Suppose that the variances of X, Y and Z are σ_X^2, σ_Y^2 and σ_Z^2. Using Fact 7.9 we have that $X + 2Y - 3Z \sim \mathcal{N}(0, \sigma_X^2 + 4\sigma_Y^2 + 9\sigma_Z^2)$, and $\frac{X+2Y-3Z}{\sqrt{\sigma_X^2+4\sigma_Y^2+9\sigma_Z^2}} \sim \mathcal{N}(0, 1)$.

This gives

$$P(X + 2Y - 3Z > 0) = P\left(\frac{X + 2Y - 3Z}{\sqrt{\sigma_X^2 + 4\sigma_Y^2 + 9\sigma_Z^2}} > 0\right) = 1 - \Phi(0) = \frac{1}{2}.$$

7.25. We have $f_X(x) = 1$ for $0 < x < 1$ and zero otherwise. For Y we have $f_Y(y) = \frac{1}{2}$ for $8 < y < 10$ and zero otherwise. Note that $8 < X + Y < 11$.

The density of $X + Y$ is given by

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z - t) dt.$$

The product $f_X(t) f_Y(z-t)$ is $\frac{1}{2}$ if $0 < t < 1$ and $8 < z-t < 10$, and zero otherwise. The second inequality is equivalent to $z - 10 < t < z - 8$. The the solution of the inequality system is $\max(0, z - 10) < t < \min(1, z - 8)$. Hence, for $8 < z < 11$ we have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z - t) dt = \frac{1}{2}(\min(1, z - 8) - \max(0, z - 10)).$$

Evaluating the formula on $(8, 9)$, $[9, 10)$ and $[10, 11)$ we get the following case defined function:

$$f_{X+Y}(z) = \begin{cases} \frac{z-8}{2} & 8 < z < 9 \\ \frac{1}{2} & 9 \le z < 10 \\ \frac{11-z}{2} & 10 \le z < 11, \\ 0 & \text{otherwise} \end{cases}$$

7.26. The probability density functions of X and Y are

$$f_X(x) = \begin{cases} \frac{1}{2}, & \text{if } 1 < x < 3 \\ 0, & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} 1, & \text{if } 9 < y < 10 \\ 0, & \text{otherwise} \end{cases}$$

Since $1 \le X \le 3$ and $9 \le Y \le 10$ we must have $10 \le X + Y \le 13$. For a $z \in [10, 13]$ the convolution formula gives

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx = \int_1^3 f_X(x) f_Y(z - x) dx.$$

We must have $9 \le z - x \le 10$ for $f_Y(z - x)$ to be nonzero, and this means $z - 10 \le x \le z - 9$. Combining this with the inequality $1 \le x \le 3$ we get that $f_X(x) f_Y(z - x)$ is nonzero if

$$\max(1, z - 10) \le x \le \min(3, z - 9).$$

Thus

$$f_{X+Y}(z) = \int_1^3 f_X(x) f_Y(z-x)dx = \int_{\max(1,z-10)}^{\min(3,z-9)} \frac{1}{2}dx$$
$$= \frac{1}{2}\left(\min(3, z-9) - \max(1, z-10)\right).$$

Evaluating these expressions for $10 \le z < 11$, $11 \le z < 12$ and $12 \le z < 13$ we get the following case defined function:

$$f_{X+Y}(z) = \begin{cases} \frac{1}{2}(z-10) & \text{if } 10 \le z < 11 \\ \frac{1}{2} & \text{if } 11 \le z < 12 \\ \frac{1}{2}(13-z) & \text{if } 12 \le z < 13 \\ 0 & \text{otherwise.} \end{cases}.$$

7.27. Using the convolution formula:

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f(s) f_Y(t-s)ds.$$

We have $f_Y(t-s) = 1$ for $0 \le t-s \le 1$ and zero otherwise. The inequality $0 \le t-s \le 1$ is equivalent to $t-1 \le s \le t$. Thus

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f(s) f_Y(t-s)ds = \int_{t-1}^{t} f(s)ds.$$

7.28. Because X_1, X_2, X_3 are jointly continuous, the probability that any two of them are equal is 0. This means that $P(X_1, X_2, X_3 \text{ are all different}) = 1$. By the exchangeability of X_1, X_2, X_3 we have

$$P(X_1 < X_2 < X_3) = P(X_2 < X_1 < X_3) = P(X_1 < X_3 < X_2)$$
$$= P(X_3 < X_2 < X_1) = P(X_2 < X_3 < X_1) = P(X_3 < X_1 < X_2),$$

where we listed all six possible orderings of X_1, X_2, X_3. Since the sum of the six probabilities is $P(X_1, X_2, X_3 \text{ are all different})$, we get that $P(X_1 < X_2 < X_3) = \frac{1}{6}$.

7.29. By exchangeability, each X_i, $1 \le i \le 100$ has the same probability to be the 50th largest. Since the X_i are jointly continuous, the probability of any two being equal is 0. Hence

$$1 = \sum_{i=1}^{100} P(X_i \text{ is the 50th largest number}) = 100 P(X_{20} \text{ is the 50th largest number})$$

and the probability in question must be $\frac{1}{100}$.

7.30. (a) By exchangeability

$$P(\text{2nd card is A, 4th card is K}) = P(\text{1st card is A, 2nd card is K}) = \frac{4 \cdot 4}{52 \cdot 51} = \frac{4}{663},$$

where the final probability comes from counting the favorable outcomes for the first two picks.

(b) Again, by exchangeability and counting the favorable outcomes within the first two picks:

$$P(\text{1st card is } \spadesuit, \text{ 5th card is } \spadesuit) = P(\text{1st card is } \spadesuit, \text{ 2nd card is } \spadesuit) = \frac{\binom{13}{2}}{\binom{52}{2}} = \frac{1}{17}.$$

(c) Using the same arguments:

$$\begin{aligned}
P(\text{2nd card is K}|\text{last two cards are aces}) &= \frac{P(\text{2nd card is K, last two cards are aces})}{P(\text{last two cards are aces})} \\
&= \frac{P(\text{3rd card is K, first two cards are aces})}{P(\text{first two cards are aces})} \\
&= P(\text{3rd card is K}|\text{first two cards are aces}) \\
&= \frac{4}{50} = \frac{2}{25}.
\end{aligned}$$

The final probability comes either from counting favorable outcomes for the first three picks, or by noting that if we choose two aces for the first two picks then we always have 50 cards left with 4 of them being kings.

7.31. By exchangeability the probability that the 3rd, 10th and 23rd picks are of different colors is the same as the probability of the first three picks being of different color. For this event the order of the first three picks does not matter, so we can assume that we choose the three balls without order, and we just need the probability that these are of different colors. Thus the probability is

$$P(\text{we choose one of each color}) = \frac{20 \cdot 10 \cdot 15}{\binom{45}{3}} = \frac{100}{473}.$$

7.32. Denote by X_k the numerical value of the kth pick. By exchangeability of $X_1, \dots, X_{23}$ we get

$$P(X_9 \le 5, X_{14} \le 5, X_{21} \le 5) = P(X_1 \le 5, X_2 \le 5, X_3 \le 5).$$

The probability that the first three picks are from $\{1, 2, 3, 4, 5\}$ is $\frac{\binom{5}{3}}{\binom{23}{3}} = \frac{10}{1771}$.

7.33. Denote the color of the kth chip by X_k. By exchangeability

$$P(X_5 = \texttt{black}|X_3 = X_{10} = \texttt{red}) = P(X_3 = \texttt{black}|X_1 = X_2 = \texttt{red}) = \frac{4}{22} = \frac{2}{11},$$

where the last step follows from the fact that if the first two choices were red then there are 4 out of the remaining 22 chops are black.

7.34. By Fact 7.17 we have to show that the joint probability mass function of $X_1, \dots, X_4$ is a symmetric function.

We will compute $P(X_1 = a_1, X_2 = a_2, X_3 = a_3, X_4 = a_4)$ for all choices of $a_1, a_2, a_3, a_4 \in \{0, 1\}$. For a given choice of $a_1, a_2, a_3, a_4 \in \{0, 1\}$ we know which aces were chosen and which were not. We can compute $P(X_1 = a_1, X_2 = a_2, X_3 = a_3, X_4 = a_4)$ by counting the favorable outcomes among the $\binom{52}{5}$ choices of unordered samples of 5. Since we know which aces are in the sample, and which are not, we just have to count the number of ways we can choose the remaining non-aces. This is given by $\binom{48}{5-k}$, where $k = a_1 + a_2 + a_3 + a_4$ is the number of aces among the 5 cards. (48 is the total number of non-ace cards, $5-k$ is the number of non-ace cards among the 5.)

Thus

$$P(X_1=a_1, X_2=a_2, X_3=a_3, X_4=a_4)=\frac{\binom{48}{5-(a_1+\cdots+a_4)}}{\binom{52}{5}}$$

if $a_1,a_2,a_3,a_4\in\{0,1\}$. But this is a symmetric function of a_1,a_2,a_3,a_4 (as the sum does not change when we permute these numbers), which shows that the random variables X_1,X_2,X_3,X_4 are indeed exchangeable.

7.35. By exchangeability, it is enough to compute the probability that the values of first three picks are increasing. By using exchangeability again, any of the possible $3!=6$ order for the first three picks are equally likely. Hence the probability in question is $\frac{1}{6}$.

7.36. (a) The waiting times between replacements are independent exponentials with parameter 1/2 (with years as the time units). This means that the replacements form a Poisson process with parameter 1/2. Then the number of replacements within the next year is Poisson distributed with parameter 1/2, and hence

$$\begin{aligned}&P(\text{have to replace a light bulb during the year})\\&\qquad=1-P(\text{no replacements within the year})=1-e^{-1/2}.\end{aligned}$$

(b) The number of points in two non-overlapping intervals are independent for a Poisson process. Thus the conditional probability is the same as the unconditional one, and using the same approach as in part (b) we get

$$P(\text{two replacements in the year})=\frac{(1/2)^2}{2!}e^{-1/2}=\frac{e^{-1/2}}{8}.$$

7.37. The joint probability mass function of $g(X_1),g(X_2),g(X_3)$ can be expressed in terms of the joint probability mass function $p(x_1,x_2,x_3)$ of X_1,X_2,X_3:

$$P(g(X_1)=a_1,g(X_2)=a_2,g(X_3)=a_3)=\sum_{\substack{b_1:g(b_1)=a_1\\ b_2:g(b_2)=a_2\\ b_3:g(b_3)=a_3}}p(x_1,x_2,x_3).$$

Similarly, for any permutation (k_1,k_2,k_3) of $(1,2,3)$ we can write

$$P(g(X_{k_1})=a_1,g(X_{k_2})=a_2,g(X_{k_3})=a_3)=\sum_{\substack{b_1:g(b_1)=a_1\\ b_2:g(b_2)=a_2\\ b_3:g(b_3)=a_3}}P(X_{k_1}=a_1,X_{k_2}=a_3,X_{k_3}=a_3).$$

Since X_1,X_2,X_3 are exchangeable, we have

$$P(X_{k_1}=a_1,X_{k_2}=a_3,X_{k_3}=a_3)=P(X_1=a_1,X_2=a_3,X_3=a_3)=p(x_1,x_2,x_3)$$

which means that

$$P(g(X_{k_1})=a_1,g(X_{k_2})=a_2,g(X_{k_3})=a_3)=P(g(X_1)=a_1,g(X_2)=a_2,g(X_3)=a_3).$$

This proves that $g(X_1),g(X_2),g(X_3)$ are exchangeable.

Solutions to Chapter 8

8.1. From the information given and properties of the random variables we deduce

$$EX = \frac{1}{p}, \quad E(X^2) = \frac{2-p}{p^2}, \quad EY = nr, \quad E(Y^2) = n(n-1)r^2 + nr.$$

(a) By linearity of expectation, $E[X+Y] = EX + EY = \frac{1}{p} + nr$.

(b) We cannot calculate $E[XY]$ without knowing something about the joint distribution of (X,Y). But no such information is given.

(c) By linearity of expectation, $E[X^2+Y^2] = E[X^2]+E[Y^2] = \frac{2-p}{p^2}+n(n-1)r^2+nr$.

(d) $E[\,(X+Y)^2\,] = E[X^2+2XY+Y^2\,] = E[X^2]+2E[XY]+E[Y^2\,]$. Again we would need $E[XY]$ which we cannot calculate.

8.2. Let X_k be the number showing on the k-sided die. We need $E[X_4+X_6+X_{12}]$. By linearity of expectation

$$E[X_4+X_6+X_{12}] = E[X_4]+E[X_6]+E[X_{12}].$$

We can compute the expectation of X_k by taking the average of the numbers $1,2,\dots,k$:

$$E[X_k] = \sum_{j=1}^{k} j \cdot \frac{1}{k} = \frac{k(k+1)}{2k} = \frac{k+1}{2}.$$

This gives

$$E[X_4+X_6+X_{12}] = \frac{4+1}{2} + \frac{6+1}{2} + \frac{12+1}{2} = \frac{25}{2}.$$

8.3. Introduce indicator variables X_B, X_C, X_D so that $X = X_B + X_C + X_D$, by defining $X_B = 1$ if Ben calls and zero otherwise, and similarly for X_C and X_D. Then $E[X] = E[X_B+X_C+X_D] = E[X_B]+E[X_C]+E[X_D] = 0.3+0.4+0.7 = 1.4$.

8.4. Let I_k be the indicator of the event that the number 4 is showing on the k-sided die. Then $Z = I_4 + I_6 + I_{12}$. For each $k \geq 4$ we have

$$E[I_k] = P(\text{the number 4 is showing on the } k\text{-sided die}) = \frac{1}{k}.$$

Hence, by linearity of expectation

$$E[Z] = E[I_4] + E[I_6] + E[I_{12}] = \frac{1}{4} + \frac{1}{6} + \frac{1}{12} = \frac{1}{2}.$$

8.5. We have $E[X] = \frac{1}{p} = 3$ and $E[Y] = \lambda = 4$ from the given distributions. The perimeter of the rectangle is given by $2(X + Y + 1)$ and the area is $X(Y + 1)$. The expectation of the perimeter is

$$E[2(X + Y + 1)] = E[2X + 2Y + 2] = 2E[X] + 2E[Y] + 2 = 2 \cdot 3 + 2 \cdot 4 + 2 = 16,$$

where we used the linearity of expectation.

The expectation of the area is

$$E[X(Y+1)] = E[XY + X] = E[XY] + E[X] = E[X]E[Y] + E[X] = 3 \cdot 4 + 3 = 15.$$

We used the linearity of expectation, and also that because of the independence of X and Y we have $E[XY] = E[X]E[Y]$.

8.6. The answer to parts (a) and (c) do not change. However, we can now compute $E[XY]$ and $E[(X + Y)^2]$ using the additional information that X and Y are independent. Using the facts from the solution of Exercise 8.1 about the first and second moments of X and Y, and the independence of these random variables we get

$$E[XY] = E[X]E[Y] = \frac{1}{p} \cdot nr = \frac{nr}{p},$$

and

$$\begin{aligned} E[(X+Y)^2] &= E[X^2 + 2XY + Y^2] = E[X^2] + 2E[XY] + E[Y^2] \\ &= \frac{2-p}{p^2} + \frac{2nr}{p} + n(n-1)r^2 + nr. \end{aligned}$$

8.7. The mean of X is given by the solution of Exercise 8.3. As in the solution of Exercise 8.3, introduce indicators so that $X = X_B + X_C + X_D$. Using the assumed independence,

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(X_B + X_C + X_D) = \mathrm{Var}(X_B) + \mathrm{Var}(X_C) + \mathrm{Var}(X_D) \\ &= 0.3 \cdot 0.7 + 0.4 \cdot 0.6 + 0.7 \cdot 0.3 = 0.66. \end{aligned}$$

8.8. Let X be the arrival time of the plumber and T the time needed to complete the project. Then $X \sim \mathrm{Unif}[1, 7]$ and $T \sim \mathrm{Exp}(2)$ (with hours as units), and these are independent. The parameter of the exponential comes from the fact that an $\mathrm{Exp}(\lambda)$ distributed random variable has expectation $1/\lambda$.

We need to compute $E[X + T]$ and $\mathrm{Var}(X + T)$. Using the distributions of X and T we get

$$E[X] = \frac{1+7}{2} = 4, \quad \mathrm{Var}(X) = \frac{6^2}{12} = 3, \quad E[T] = \frac{1}{2}, \qquad \mathrm{Var}(T) = \frac{1}{2^2} = \frac{1}{4}.$$

By linearity we get

$$E[X+T] = E[X] + E[T] = 4 + \frac{1}{2} = \frac{9}{2}.$$

From the independence

$$\mathrm{Var}(X+T) = \mathrm{Var}(X) + \mathrm{Var}(T) = 3 + \frac{1}{4} = \frac{13}{4}.$$

8.9. (a) We have

$$E[3X - 2Y + 7] = 3E[X] - 2E[Y] + 7 = 3 \cdot 3 - 2 \cdot 5 + 7 = 6,$$

where we used the linearity of expectation.
(b) Using the independence of X and Y:

$$\mathrm{Var}(3X - 2Y + 7) = 9 \cdot \mathrm{Var}(X) + 4 \cdot \mathrm{Var}(Y) = 92 + 43 = 30.$$

(c) From the definition of the variance

$$\mathrm{Var}(XY) = E[(XY)^2] - E[XY]^2.$$

By independence we have $E[XY] = E[X]E[Y]$ and $E[(XY)^2] = E[X^2]E[Y^2]$, thus

$$\begin{aligned}\mathrm{Var}(XY) &= E[X^2]E[Y^2] - E[X]^2E[Y]^2 \\ &= E[X^2]E[Y^2] - 925 = E[X^2]E[Y^2] - 225,\end{aligned}$$

To compute the second moments we use the variance:

$$2 = \mathrm{Var}(X) = E[X^2] - E[X]^2 = E[X^2] - 9$$

hence $E[X^2] = 9+2 = 11$. Similarly, $E[Y^2] = E[Y]^2 + \mathrm{Var}(Y) = 25+3 = 28$. Thus

$$\mathrm{Var}(XY) = 11 \cdot 28 - 225 = 83.$$

8.10. The moment generating function of X_1 is given by

$$M_{X_1}(t) = E[e^{tX}] = \sum_k e^{tk} P(X_1 = k) = \frac{1}{2} + \frac{1}{3}e^t + \frac{1}{6}e^{2t}.$$

The moment generating function of X_2 is the same. Since X_1 and X_2 are independent, we can compute the moemnt generating function of $S = X_1 + X_2$ as follows:

$$M_S(t) = M_{X_1}(t)M_{X_2}(t) = \left(\frac{1}{2} + \frac{1}{3}e^t + \frac{1}{6}e^{2t}\right)^2.$$

Expanding the square we get

$$M_S(t) = \frac{1}{4} + \frac{1}{3}e^t + \frac{5}{18}e^{2t} + \frac{1}{9}e^{3t} + \frac{1}{36}e^{4t}.$$

We can read off the probability mass function of S from this by identifying the coefficients of the exponential terms:

$P(S=0) = \frac{1}{4}, \quad P(S=1) = \frac{1}{3}, \quad P(S=2) = \frac{5}{18}, \quad P(S=3) = \frac{1}{9}, \quad P(S=4) = \frac{1}{36}.$

8.11. Introduce indicator variables X_B, X_C, X_D so that $X = X_B + X_C + X_D$, by defining $X_B = 1$ if Ben calls and zero otherwise, and similarly for X_C and X_D. These are independent Bernoulli random variables with parameters 0.3, 0.4 and 0.7, respectively. By the independence, the moment generating function of $X = X_B + X_C + X_D$ can be written as

$$M_X(t) = M_{X_A}(t)M_{X_B}(t)M_{X_C}(t).$$

The generating function of a parameter p Bernoulli random variable is $pe^t + 1 - p$, which means that

$$M_X(t) = (0.3e^t+0.7)(0.4e^t+0.6)(0.7e^t+0.3) = 0.126+0.432e^t+0.358e^{2t}+0.084e^{3t}.$$

8.12. (a) We need to compute

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} f_Z(z)dz = \int_0^{\infty} e^{tz}\lambda^2 ze^{-\lambda z}dz = \lambda^2 \int_0^{\infty} ze^{-(\lambda-t)z}dz.$$

If $\lambda - t \le 0$ then this integral is at least as large as $\lambda^2 \int_0^\infty zdz$ which is infinite. If $\lambda - t > 0$ then we can compute the integral using integration by parts, or by noting that $\int_0^\infty z(\lambda-t)e^{-(\lambda-t)z}dz = \frac{1}{\lambda-t}$ as the integral is the expectation of an $\text{Exp}(\lambda - t)$ distributed random variable. This gives

$$M_Z(t) = \begin{cases} \frac{\lambda^2}{(\lambda-t)^2}, & \text{if } t < \lambda \\ \infty, & \text{if } t \ge \lambda. \end{cases}$$

(b) We have seen in Example 5.6 that

$$M_X(t) = M_Y(t) = \begin{cases} \frac{\lambda}{\lambda-t}, & \text{if } t < \lambda \\ \infty, & \text{if } t \ge \lambda. \end{cases}$$

Since X and Y are independent, we have $M_{X+Y}(t) = M_X(t)M_Y(t)$. Comparing with part (a) we see that $X+Y$ has the same moment generating function as Z, which means that they must have a the same distribution. (Since the moment generating function is finite in a neighborhood of 0.)

8.13. We first find a random variable that has the moment generating function $\frac{1}{2}e^{-t} + \frac{2}{5} + \frac{1}{10}e^{t/2}$. Reading off the coefficients of the e^{-t}, $e^{t/2}$ and also considering the constant term we get that if X has probability mass function

$$p(-1) = \tfrac{1}{2}, \quad p(0) = \tfrac{2}{5}, \quad p(\tfrac{1}{2}) = \tfrac{1}{10}.$$

then $M_X(t) = \frac{1}{2}e^{-t}+\frac{2}{5}+\frac{1}{10}e^{t/2}$. Now take independent random variables $X_1, \dots, X_{36}$ with the same distribution as X. By independence, the sum $X_1 + \cdots + X_{36}$ has a moment generating function which is the product of the individual moment generating functions, which is exactly $\left(\frac{1}{2}e^{-t} + \frac{2}{5} + \frac{1}{10}e^{t/2}\right)^{36} = M_Z(t)$. Hence Z has the same distribution as $X_1 + \cdots + X_{36}$.

8.14. We need to compute $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$, $E[XY]$. All of these can be computed using the joint probability mass function given in the table. For example,

$$\begin{aligned} E[X] &= 1 \cdot (\tfrac{1}{15} + \tfrac{1}{15} + \tfrac{2}{15} + \tfrac{1}{15}) + 2 \cdot (\tfrac{1}{10} + \tfrac{1}{10} + \tfrac{1}{5} + \tfrac{1}{10}) + 3 \cdot (\tfrac{1}{30} + \tfrac{1}{30} + 0 + \tfrac{1}{10}) \\ &= \frac{11}{6} \end{aligned}$$

and

$$\begin{aligned}E[XY] &= 1\cdot 0\cdot \tfrac{1}{15} 1\cdot 1\cdot \tfrac{1}{15} + 1\cdot 2\cdot \tfrac{2}{15} + 1\cdot 3\cdot \tfrac{1}{15} + 2\cdot 0\cdot \tfrac{1}{10} + 2\cdot 1\cdot \tfrac{1}{10}\\ &\quad + 2\cdot 2\cdot \tfrac{1}{5} + 2\cdot 3\cdot \tfrac{1}{10} + 3\cdot 0\cdot \tfrac{1}{30} + 3\cdot 1\cdot \tfrac{1}{30} + 3\cdot 3\cdot \tfrac{1}{10}\\ &= \frac{47}{15}.\end{aligned}$$

Similarly,

$$E[Y] = \frac{5}{3}, \quad E[X^2] = \frac{23}{6}, \quad E[Y^2] = \frac{59}{15}.$$

Then

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{47}{15} - \frac{11}{6}\cdot\frac{5}{3} = \frac{7}{90}.$$

For the correlation we first compute the variances:

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] - (E[X])^2 = \frac{23}{6} - \left(\frac{11}{6}\right)^2 = \frac{17}{36}\\ \mathrm{Var}(Y) &= E[Y^2] - (E[Y])^2 = \frac{59}{15} - \left(\frac{5}{3}\right)^2 = \frac{52}{45}.\end{aligned}$$

From this we have

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{7}{2\sqrt{1105}} \approx 0.1053$$

8.15. We first compute the joint probability density of (X,Y). The quadrilateral D is composed of a unit square and a triangle which is half of the unit square, thus the area of D is $\frac{3}{2}$. Thus the joint density function is

$$f_{X,Y}(x,y) = \frac{2}{3}1_{\{(x,y)\in D\}}.$$

To calculate the covariance we need to calculate

$$E[XY], \quad E[X], \quad E[Y].$$

We have

$$\begin{aligned}E[XY] &= \int_0^1\int_0^{2-y} \frac{2}{3}xy\,dx\,dy = \int_0^1 \frac{2}{6}y(2-y)^2\,dy\\ &= \frac{2}{6}\left(\frac{4}{2}y^2 - \frac{4}{3}y^3 + \frac{1}{4}y^4\right)\Big|_0^1 = \frac{2}{6}\cdot\frac{11}{12} = \frac{11}{36}.\end{aligned}$$

$$\begin{aligned}E[X] &= \int_0^1\int_0^{2-y} \frac{2}{3}x\,dx\,dy = \int_0^1 \frac{2}{6}(2-y)^2\,dy\\ &= \frac{2}{6}\left(4y - \frac{4}{2}y^2 + \frac{1}{3}y^3\right)\Big|_0^1 = \frac{2}{6}\cdot\frac{7}{3} = \frac{7}{9}.\end{aligned}$$

$$\begin{aligned}E[Y] &= \int_0^1\int_0^{2-y} \frac{2}{3}y\,dx\,dy = \int_0^1 \frac{2}{3}(2-y)y\,dy\\ &= \frac{2}{3}\left(\frac{2}{2}y^2 - \frac{1}{3}y^3\right)\Big|_0^1 = \frac{2}{3}\cdot\frac{2}{3} = \frac{4}{9}.\end{aligned}$$

By the definition of covariance, we get

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{11}{36} - \frac{7}{9}\cdot\frac{4}{9} = -\frac{13}{324}.$$

The fact that X and Y are negatively correlated could have been guessed from the shape of D: as Y gets smaller, the value of X tend to get larger on average.

8.16. We have

$$\mathrm{Cov}(X, 2X+Y-3) = 2\,\mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) = 2\,\mathrm{Var}(X) + \mathrm{Cov}(X,Y).$$

The variance of X can be computed as follows:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = 3 - 1^2 = 2.$$

The covariance can be calculated as

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = -4 - 1\cdot 2 = -6.$$

Thus

$$\mathrm{Cov}(X, 2X+Y-3) = 2\,\mathrm{Var}(X) + \mathrm{Cov}(X,Y) = 2\cdot 2 - 6 = -2.$$

8.17. We need $E[X^2]$ and $E[X]$. By linearity:

$$E[X] = E[I_A + I_B] = E[I_A] + E[I_B] = P(A) + P(B) = 0.7.$$

Similarly:

$$\begin{aligned} E[X^2] &= E[(I_A+I_B)^2] = E[I_A^2 + I_B^2 + 2I_AI_B], \\ &= E[I_A^2] + E[I_B^2] + 2E[I_AI_B]. \end{aligned}$$

We have $I_A^2 = I_A$, $I_B^2 = I_B$ and $I_AI_B = I_{AB}$, hence

$$\begin{aligned} E[X^2] &= E[I_A^2] + E[I_B^2] + 2E[I_AI_B] \\ &= P(A) + P(B) + 2P(AB) = 0.9. \end{aligned}$$

Then

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = 0.9 - 0.7^2 = 0.41.$$

8.18. By the discussion in Section 8.6 if X, Y are independent standard normals and A is a 2×2 matrix then the coordinates of the random vector $A[X,Y]^T$ are distributed as a bivariate normal with expectation vector $[0,0]^T$ and covariance matrix AA^T. Choosing $A = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ we get $A[X,Y]^t = [U,V]^T$. Since $AA^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ we get that the variance of U and V are both 1, and the covariance of U and V is 0. Hence U and V are indeed independent standard normals.

Here is another solution using the Jacobian technique of Section 6.4. We have $U = g(X,Y)$, $V = h(X,Y)$ with

$$g(x,y) = \frac{1}{\sqrt{2}}(x-y), \qquad h(x,y) = \frac{1}{\sqrt{2}}(x+y).$$

Then the inverse of these functions is given by

$$q(u,v) = \frac{1}{\sqrt{2}}(u+v), \qquad r(u,v) = \frac{1}{\sqrt{2}}(v-u),$$

and the Jacobian is

$$J(u,v) = \det \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = 1.$$

Now using Fact 6.41 we get that the joint density of U, V is given by

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}\Big(\frac{u+v}{\sqrt{2}}, \frac{v-u}{\sqrt{2}}\Big) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(\frac{u+v}{\sqrt{2}}\right)^2 - \frac{1}{2}\left(\frac{v-u}{\sqrt{2}}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}. \end{aligned}$$

The final result shows that U and V are independent standard normals.

8.19. This is the same problem as Exercise 6.15.

8.20. By linearity, $E[X_3 + X_{10} + X_{22}] = E[X_3] + E[X_{10}] + E[X_{22}]$. The random variables $X_1, \dots, X_{30}$ are exchangeable, thus $E[X_k] = E[X_1]$ for all $1 \le k \le 30$. This gives

$$E[X_3 + X_{10} + X_{22}] = 3E[X_1].$$

The value of the first pick is equally likely to be any of the first 30 positive integers, hence

$$E[X_1] = \sum_{k=1}^{30} k \frac{1}{30} = \frac{30 \cdot 31}{2 \cdot 30} = \frac{31}{2},$$

and

$$E[X_3 + X_{10} + X_{22}] = 3E[X_1] = \frac{93}{2}.$$

8.21. Label the coins from 1 to 10, for example so that coins 1-5 are the dimes, coins 6-8 are the quarters, and coins 9-10 are the pennies. Let a_k be the value of coin k and let I_k be the indicator variable that is 1 if coin k is chosen, for $k = 1, \dots, 10$. Then

$$X = \sum_{k=1}^{10} a_k I_k = 10(I_1 + \cdots + I_5) + 25(I_6 + I_7 + I_8) + I_9 + I_{10}.$$

The probability that any particular coin is chosen is

$$E(I_k) = P(\text{coin } k \text{ chosen}) = \frac{\binom{9}{2}}{\binom{10}{3}} = \tfrac{3}{10}.$$

Hence

$$EX = \sum_{k=1}^{10} a_k E(I_k) = 10 \cdot 5 \cdot \tfrac{3}{10} + 25 \cdot 3 \cdot \tfrac{3}{10} + 2 \cdot \tfrac{3}{10} = 38.1 \text{ (cents)}.$$

8.22. There are several ways to approach this problem. One possibility that gives the answer without doing complicated computations is as follows. For each $1 \le j \le 89$ let I_j be the indicator of the event that both j and $j+1$ are chosen among the five numbers. Then $X = \sum_{j=1}^{89} I_j$, since if j and $j+1$ are both chosen then they will be next to each other in the ordered sample. By linearity

$$E[X] = E[\sum_{j=1}^{89} I_j] = \sum_{j=1}^{89} E[I_j].$$

We can compute $E[I_j]$ directly by counting favorable outcomes:

$$E[I_j] = P(\text{both } j \text{ and } j+1 \text{ are chosen}) = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{2}{801}.$$

Thus

$$E[X] = 89 \cdot \frac{2}{801} = \frac{20}{89}.$$

Note that we could have expressed X differently as a sum of indicators, e.g. by considering the indicator that the jth and $(j+1)$st number among the chosen numbers have a difference of 1. However, this would lead to indicators that are not exchangeable, and the corresponding probabilities would be hard to compute.

8.23. (a) Let Y_i denote the color of the i^{th} pick (i.e. $Y_i \in \{\texttt{red, green}\}$). Then $Y_1, \dots, Y_{50}$ are exchangeable so

$$\begin{aligned}P(Y_{28} \neq Y_{29}) &= P(Y_{28} = \texttt{red}, Y_{29} = \texttt{green}) + P(Y_{29} = \texttt{red}, Y_{28} = \texttt{green}) \\ &= 2P(Y_1 = \texttt{red}, Y_2 = \texttt{green}) = 2\frac{20 \cdot 30}{50 \cdot 49} = \frac{24}{49}\end{aligned}$$

(b) Let I_j be the indicator that $Y_j \neq Y_{j+1}$ for $j = 1, \dots, 49$. Then $X = I_1 + \cdots + I_{49}$ and by linearity

$$E[X] = \sum_{i=1}^{49} E[I_i] = \sum_{i=1}^{49} P(Y_i \neq Y_{i+1})$$

By the exchangeability of the Y_i random variables and part (a) we get

$$E[X] = \sum_{i=1}^{49} P(Y_i \neq Y_{i+1}) = 49P(Y_1 \neq Y_2) = 49\frac{24}{49} = 24.$$

Another (bit more complicated) solution for part (b):
Introduce labels for the 20 red balls (from 1 to 20). Let $J_i, 1 \leq i \leq 20$ be the indicator that the i^{th} red ball has a green ball right after it, and K_i be the indicator that the i^{th} red ball has a green ball right before it. Then

$$X = \sum_{i=1}^{20} (J_i + K_i),$$

and by the linearity of expectation and exchangeability we have

$$E[X] = \sum_{i=1}^{20} E[J_i] + \sum_{i=1}^{20} E[K_i] = 20E[J_1] + 20E[K_1]$$

Using exchangeability again:

$$\begin{aligned}P(J_1 = 1) &= \sum_{i=1}^{49} P(\text{red ball \# 1 is picked at position } i \text{ and a green ball is picked at } i+1) \\ &= 49P(\text{red ball \# 1 is picked at position 1 and a green ball is picked at 2}) \\ &= 49\frac{1 \cdot 30}{50 \cdot 49} = \frac{3}{5}.\end{aligned}$$

Same way we get $P(K_1 = 1) = \frac{3}{5}$. Putting everything together:

$$E[X] = 20E[J_1] + 20E[K_1] = 2 \cdot 20 \cdot \frac{3}{5} = 24.$$

8.24. Let I_j be the indicator of the event that Jane's jth pick has the same color as Sam's jth pick. Imagine that we write down the picked colors as they appear, all 80 of them. Then I_j depends on the color of the $(2j-1)$st and $2j$th pick, and since the colors are exchangeable, the I_j random variables will be exchangeable as well. We have $N = \sum_{j=1}^{40} I_j$, and by linearity of expectation and exchangeability we get

$$E[N] = E[\sum_{j=1}^{40} I_j] = \sum_{i=1}^{40} E[I_j] = 40E[I_1].$$

But

$$E[I_1] = P(\text{first two colors are the same color}) = \frac{\binom{30}{2} + \binom{50}{2}}{\binom{80}{2}} = \frac{83}{158},$$

by counting favorable outcomes within the first two pick. This gives

$$E[N] = 40 \cdot \frac{83}{158} = \frac{1660}{79} \approx 21.0127.$$

8.25. (a) Let Y_i denote the number of the ith pick. Then $(Y_1, Y_2, \dots, Y_{10})$ is exchangeable, and hence

$$P(Y_5 > Y_4) = P(Y_1 > Y_2) = P(Y_2 > Y_1) = 1/2.$$

In the last step we used that the numbers are different and this $P(Y_1 > Y_2) + P(Y_2 > Y_1) = 1$.

(b) Let I_j be the indicator of the event that the number on the jth ball is larger than the number on the $(j-1)$st. (For $j = 2, 3, \dots, 10$.) Then

$$X = I_2 + I_3 + \cdots + I_{10}$$

and

$$E[X] = E[I_2 + I_3 + \cdots + I_{10}] = \sum_{j=2}^{10} P(j\text{th number is larger than the } (j-1)\text{st}).$$

Using part (a) we get that

$$P(j\text{th number is larger than the } (j-1)\text{st}) = 1/2$$

for all $2 \le j \le 10$, which means that $E[X] = \sum_{j=2}^{10} \frac{1}{2} = \frac{9}{2}$.

8.26. (a) Let I_j be the indicator that the jth ball is green and the $(j+1)$st ball is yellow. Then $X_n = \sum_{j=1}^{n-1} I_j$. By linearity

$$E[X_n] = E[\sum_{j=1}^{n-1} I_j] = \sum_{j=1}^{n-1} E[I_j].$$

Because we draw with replacement, the colors of different picks are independent:

$$\begin{aligned} E[I_j] &= P(j\text{th ball is green and the } (j+1)\text{st ball is yellow}) \\ &= P(j\text{th ball is green})P((j+1)\text{st ball is yellow}) = \frac{4}{9} \cdot \frac{3}{9} = \frac{4}{27}. \end{aligned}$$

This gives

$$E[X_n] = \sum_{j=1}^{n-1} \frac{4}{27} = \frac{4(n-1)}{27}.$$

(b) We will see a different (maybe more straightforward) technique in Chapter 10, but here we will give a solution using the indicator method. Let J_k denote the indicator that the kth ball is green and there are no white balls among the first $k-1$. Then $Y = \sum_{k=1}^{\infty} J_k$. (In the sum a term is equal to 1 if the corresponding ball is green and came before the first white.) Using linearity

$$\begin{aligned} E[Y] &= E[\sum_{k=1}^{\infty} J_k] = \sum_{k=1}^{\infty} E[J_k] \\ &= \sum_{k=1}^{\infty} P(k\text{th ball is green, no white balls among the first } k-1). \end{aligned}$$

(We can exchange the expectation and the infinite sum here as each term is nonnegative.) Using independence we can compute the probability in question for each k:

$$\begin{aligned} &P(k\text{th ball is green, no white balls among the first } k-1) \\ &\qquad = P(k\text{th ball is green})P(\text{first } k-1 \text{ balls are all green or yellow}) \\ &\qquad = \tfrac{4}{9} \cdot \left(\tfrac{7}{9}\right)^{k-1}. \end{aligned}$$

This gives

$$E[Y] = \sum_{k=1}^{\infty} \tfrac{4}{9} \cdot \left(\tfrac{7}{9}\right)^{k-1} = \tfrac{4}{9} \cdot \tfrac{1}{1-\frac{7}{9}} = 2.$$

Here is an intuitive explanation for the result that we got. The yellow draws are irrelevant in this problem: the only thing that matters is the position of the first white, and the number of green choices before that. Imagine that we remove the yellow balls from the urn, and we repeat the same experiment (sampling with replacement), stopping at the first white ball. Then the number of picks is a geometric random variable with parameter $\frac{2}{6} = \frac{1}{3}$. The expectation of this geometric random variable is 3. Moreover, the number of total picks is equal to the number of green balls chosen before the first white plus the 1 (the first white). This explains why the expectation of Y is $3 - 1 = 2$.

8.27. For $1 \le i < j \le n$ let $I_{i,j}$ be the indicator of the event that $a_i = a_j$. We need to compute the expected value of the random variable $X = \sum_{i<j} I_{i,j}$. By linearity $E[X] = \sum_{i<j} E[I_{i,j}]$. Using the exchangeability of the sample $(a_1, \dots, a_n)$ we get for all $i < j$ that $E[I_{i,j}] = E[I_{1,2}] = P(a_1 = a_2)$. Counting favorable outcomes (or by conditioning on the first pick) we get $P(a_1 = a_2) = \frac{1}{n}$. This gives

$$E[X] = \sum_{i<j} E[I_{i,j}] = \binom{n}{2} P(a_1 = a_2) = \binom{n}{2} \cdot \frac{1}{n} = \frac{n-1}{2}.$$

8.28. Imagine that we take the sample with order and for each $1 \le k \le 10$ let I_k be the indicator that we got a yellow marble for the kth pick, and J_k be the indicator that we got a green pick. Then $X = \sum_{k=1}^{10} I_k$, $Y = \sum_{k=1}^{10} J_k$ and $X - Y = \sum_{k=1}^{10}(I_k - J_k)$. Using the linearity of expectation we get

$$E[X-Y] = E[\sum_{k=1}^{10}(I_k - J_k)] = \sum_{k=1}^{10}(E[I_k] - E[J_k]).$$

Using the exchangeability of $I_1, \dots, I_{10}$, and $J_1, \dots, J_{10}$:

$$E[X-Y] = \sum_{k=1}^{10}(E[I_k] - E[J_k]) = 10E[I_1] - 10E[J_1].$$

By counting favorable outcomes:

$$E[I_1] = P(\text{first pick is yellow}) = \frac{25}{95} = \frac{5}{19}$$
$$E[J_1] = P(\text{first pick is green}) = \frac{30}{95} = \frac{6}{19}.$$

which leads to

$$E[X-Y] = 10 \cdot \frac{5}{19} - 10 \cdot \frac{6}{19} = -\frac{10}{19}.$$

8.29. Let I_j be the indicator that the cards flipped at $j, j+1$ and $j+2$ are all number cards. (Here $1 \le j \le 50$.) Then $X = \sum_{j=1}^{50} I_j$ and $E[X] = \sum_{j=1}^{50} E[I_j]$. By exchangeability we have

$$E[X] = \sum_{j=1}^{50} E[I_j] = 50E[I_1] = 50P(\text{the first three cards flipped are number cards}).$$

Counting favorable outcomes (noting that there are $4 \cdot 9 = 36$ number cards in the deck) gives

$$P(\text{the first three cards flipped are number cards}) = \frac{\binom{36}{3}}{\binom{52}{3}} = \frac{21}{65}$$

and

$$E[X] = 50 \cdot \frac{21}{65} = \frac{210}{13}.$$

8.30. Let X_k be the number of the kth chosen ball and let I_k be the indicator of the event that $X_k > X_{k-1}$. Then

$$N = I_2 + I_3 + \cdots + I_{20},$$

and using linearity and exchangeability

$$E[N] = E[\sum_{k=2}^{20} I_k] = \sum_{k=2}^{20} E[I_k] = 19E[I_2].$$

We also have

$$E[I_2] = P(X_1 < X_2) = P(\text{first number is smaller than the second}).$$

One could compute the probability $P(X_1 < X_2)$ by counting favorable outcomes for the first two picks. Another way is to notice that

$$1 = P(X_1 < X_2) + P(X_1 > X_2) + P(X_1 = X_2) = 2P(X_1 < X_2) + P(X_1 = X_2),$$

where we used exchangeability again. By conditioning on the first outcome we see that $P(X_1 = X_2) = \frac{1}{19}$, which gives

$$2P(X_1 < X_2) = \frac{1 - P(X_1 = X_2)}{2} = \frac{9}{19}$$

and $E[N] = 19P(X_1 < X_2) = 9$.

8.31. Write the uniformly chosen number with exactly 4 digits (by putting zeros at the beginning if needed), and denote the four digits by X_1, X_2, X_3, X_4. (Thus for 128 we have $X_1 = 0$, $X_2 = 1$, $X_3 = 2$, $X_4 = 8$.) Then each digit will be uniform on the set $\{0, \dots, 9\}$ (you can check this by counting), hence $E[X_i] = \frac{0+1+2+\cdots+0}{2} = \frac{9}{2}$. We have $X = X_1 + X_2 + X_3 + X_4$ and hence

$$EX = E[X_1 + X_2 + X_3 + X_4] = 4EX_1 = 4 \cdot 9/2 = 18.$$

8.32. (a) We have

$$I_{A\cup B} = I_{(A^c \cap B^c)^c} = 1 - I_{A^c B^c} = 1 - I_{A^c} I_{B^c} = 1 - (1 - I_A)(1 - I_B).$$

Expanding the last expression gives

$$I_{A\cup B} = 1 - (1 - I_A)(1 - I_B) = 1 - (1 - I_A - I_B + I_A I_B) = I_A + I_B - I_A I_B.$$

The identity now follows by noting that $I_A I_B = I_{AB}$.

Another approach would be to note that $AB \cup AB^c \cup A^c B \cup A^c B^c$ gives a partition of Ω, so any $\omega \in \Omega$ will be a member of exactly one of AB, AB^c, $A^c B$ or $A^c B^c$. For each of these four cases we can evaluate $I_{A\cup B}, I_A, I_B, I_{AB}$ and check that the two sides of the equation are equal.

(b) This is immediate after taking expectation in the identity proved in part (a). We have

$$E[I_{A\cup B}] = P(A \cup B)$$

and using linearity

$$E[I_A + I_B - I_{A\cap B}] = E[I_A] + E[I_B] - E[I_{A\cap B}] = P(A) + P(B) - P(AB).$$

Since the two expectations agree by part (a), we get $P(A \cup B) = P(A) + P(B) - P(AB)$.

(c) Let A, B, C be events on the same sample space. Then

$$\begin{aligned} I_{A\cup B\cup C} &= I(A^c B^c C^c)^c = 1 - I_{A^c B^c C^c} \\ &= 1 - I_{A^c} I_{B^c} I_{C^c} = 1 - (1 - I_A)(1 - I_B)(1 - I_C). \end{aligned}$$

Expanding the product

$$\begin{aligned} I_{A\cup B\cup C} &= 1 - (1 - I_A)(1 - I_B)(1 - I_C) \\ &= 1 - (1 - I_A - I_B - I_C + I_A I_B + I_A I_C + I_B I_C - I_A I_B I_C) \\ &= I_A + I_B + I_C - I_A I_B - I_A I_C - I_B I_C + I_A I_B I_C. \end{aligned}$$

Using $I_D I_E = I_{DE}$ repeatedly:

$$\begin{aligned} I_{A\cup B\cup C} &= I_A + I_B + I_C - I_A I_B - I_A I_C - I_B I_C + I_A I_B I_C \\ &= I_A + I_B + I_C - I_{AB} - I_{AC} - I_{BC} + I_{ABC}. \end{aligned}$$

Taking expectations of both sides now gives

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

8.33. (a) For each $1 \le a \le 10$ let I_a be the indicator of the event that the ath player won exactly 2 matches. Then we need

$$E[\sum_{k=1}^{10} I_k] = \sum_{k=1}^{10} P(\text{the } a\text{th player won exactly 2 matches}).$$

By exchangeability the probability is the same for each a. Since the outcomes of the matches are independent and a player plays 9 matches, we have

$$P(\text{the first player won exactly 2 matches}) = \binom{9}{2} 2^{-9}.$$

Thus the expectation is $10 \cdot \binom{9}{2} 2^{-9} = \frac{45}{64}$.
(b) For each $1 \le a < b < c \le 10$ let $J_{a,b,c}$ be the indicator that the players numbered a, b and c form a 3-cycle. We need $E[\sum_{a<b<c} J_{a,b,c}] = \sum_{a<b<c} E[J_{a,b,c}]$. There are $\binom{10}{3}$ such triples, and the expectation is the same for each one, so it is enough to find

$$E[J_{1,2,3}] = P(\text{Players 1, 2 and 3 form a 3-cycle}).$$

Players 1, 2 and 3 form a 3-cycle if 1 beats 2, 2 beats 3, 3 beats 1 (this has probability 1/8) or if 1 beats 3, 3 beats 2 and 2 beats 1 (this also has probability 1/8). Thus $E[J_{1,2,3}] = 1/8 + 1/8 = \frac{1}{4}$, and the expectation in question is $\binom{10}{3}\frac{1}{4} = 30$.
(c) We use the indicator method again. For each possible sequence of different players $a_1, a_2, \dots, a_k$ we set up an indicator that this sequence is a k-path. The number of such indicators is $\binom{10}{k} \cdot k! = \frac{10!}{(10-k)!}$ (we choose the k players, then their order). The probability that a given indicator is 1 is the probability that a_1 beats a_2, a_2 beats a_3, ..., a_{k-1} beats a_k which is $2^{-(k-1)}$. Thus the expectation is $\frac{10!}{(10-k)!}\left(\frac{1}{2}\right)^{k-1}$.

8.34. We show the proof for $n = 2$, the general case can be done similarly. Assume that the joint probability density function of X_1, X_2 is $f(x_1, x_2)$. Then

$$E[g_1(X_1) + g_2(X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (g_1(x_1) + g_2(x_2)) f(x_1, x_2) dx_1 dx_2.$$

Using the linearity of the integral we can write this as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1) f(x_1, x_2) dx_1 dx_2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(x_2) f(x_1, x_2) dx_1 dx_2.$$

Integrating out x_2 in the first integral gives

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1) f(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} g_1(x_1) \left(\int_{-\infty}^{\infty} f(x_1, x_2) dx_2\right) dx_1.$$

Note that $\int_{-\infty}^{\infty} f(x_1, x_2) dx_2$ is equal to $f_{X_1}(x_1)$, the marginal probability density of X_1. Hence

$$\int_{-\infty}^{\infty} g_1(x_1) \left(\int_{-\infty}^{\infty} f(x_1, x_2) dx_2\right) dx_1 = \int_{-\infty}^{\infty} g_1(x_1) f_{X_1}(x_1) dx_1 = E[g_1(X_1)].$$

Similar computation shows that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(x_2) f(x_1, x_2) dx_1 dx_2 = E[g_2(X_2)].$$

Thus $E[g_1(X_1) + g_2(X_2)] = E[g_1(X_1)] + E[g_2(X_2)]$.

8.35. (a) We may assume that the choices we made each day are independent. Let J_k be the indicator for the event that the sweater k is worn at least once in the 5 days. Then $X = J_1 + J_2 + J_3 + J_4$. By linearity and exchangeability

$$\begin{aligned} E[X] &= E[J_1 + J_2 + J_3 + J_4] = \sum_{k=1}^{4} E[J_k] = 4E[J_1] \\ &= 4P(\text{the first sweater was worn at least once}). \end{aligned}$$

Considering the complement of the event in the last line:

$$\begin{aligned} P(\text{the first sweater was worn at least once}) &= 1 - P(\text{the first sweater was not worn at all}) \\ &= 1 - \left(\frac{3}{4}\right)^5, \end{aligned}$$

where we used the independence assumption. This gives

$$E[X] = 4\left(1 - \left(\frac{3}{4}\right)^5\right) = \frac{781}{256}.$$

(b) We use the notation introduced in part (a). For the variance of X we need $E[X^2]$. Using linearity and exchangeability:

$$\begin{aligned} E[X^2] &= E[(J_1 + J_2 + J_3 + J_4)^2] = E[\sum_{k=1}^{4} J_k^2 + 2\sum_{k<\ell} J_k J_\ell] \\ &= 4E[J_1^2] + 2\binom{4}{2}E[J_1 J_2] = 4E[J_1^2] + 12E[J_1 J_2] \end{aligned}$$

Since J_1 is one or zero, we have $J_1^2 = J_1$ and by part (a)

$$4E[J_1^2] = 4E[J_1] = E[X] = \frac{781}{256}.$$

We also have

$$E[J_1 J_2] = P(\text{both the first and second sweater were worn at least once}).$$

Let A_k denote the event that the kth sweater was not worn at all during the week. Then

$$\begin{aligned} &P(\text{both the first and second sweater were worn at least once}) = P(A_1^c A_2^c) \\ &\quad = 1 - P((A_1^c A_2^c)^c) = 1 - P(A_1 \cup A_2) \\ &\quad = 1 - (P(A_1) + P(A_2) - P(A_1 A_2)). \end{aligned}$$

From part (a) we get $P(A_1) = P(A_2) = (\frac{3}{4})^5$, and similarly

$$P(A_1 A_2) = P(\text{neither the first nor the second sweater was worn}) = (\tfrac{2}{4})^5.$$

Thus

$$E[J_1 J_2] = 1 - P(A_1) - P(A_2) + P(A_1 A_2) = 1 - 2(\tfrac{3}{4})^5 + (\tfrac{2}{4})^5$$

and

$$E[X^2] = \frac{781}{256} + 12\left(1 - 2(\tfrac{3}{4})^5 + (\tfrac{2}{4})^5\right) = \frac{2491}{256}.$$

Finally,

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \tfrac{2491}{256} - (\tfrac{781}{256})^2 \approx 0.4232.$$

8.36. (a) Let I_k be the indicator of the event that the number k appears at least once among the four die rolls. Then $X = I_1 + \cdots + I_6$ and we get

$$E[X] = E[I_1 + \cdots + I_6] = E[I_1] + \cdots + E[I_6] = 6E[I_1],$$

where the last step comes from exchangeability. We have

$$E[I_1] = P(\text{the number 1 shows up}) = 1 - P(\text{none of the rolls are equal to 1}) = 1 - \left(\tfrac{5}{6}\right)^4$$

which gives

$$E[X] = 6\left(1 - \left(\tfrac{5}{6}\right)^4\right).$$

(b) We need to compute the second moment of X. Using the notation of part (a):

$$\begin{aligned} E[X^2] = E[(I_1 + \cdots + I_6)^2] &= E[\sum_{k=1}^{6} I_k^2 + 2 \sum_{j<k\le 6} I_j I_k] \\ &= \sum_{k=1}^{6} E[I_k^2] + 2 \sum_{j<k\le 6} E[I_j I_k]. \end{aligned}$$

Since I_k is either 0 or 1, we have $I_k^2 = I_k$. Using exchangeability

$$\begin{aligned} E[X^2] &= \sum_{k=1}^{6} E[I_k^2] + 2 \sum_{j<k\le 6} E[I_j I_k] \\ &= \sum_{k=1}^{6} E[I_k] + 2 \sum_{j<k\le 6} E[I_j I_k] \\ &= 6E[I_1] + 30E[I_1 I_2]. \end{aligned}$$

We computed $6E[I_1]$ in part (a), it is exactly $E[X] = 6\left(1 - \left(\tfrac{5}{6}\right)^4\right)$. To compute $E[I_1 I_2]$ we first note that $I_1 I_2$ is the indicator of the event that both the numbers 1 and 2 show up at least once. Taking complements and using inclusion-exclusion:

$$\begin{aligned} E[I_1 I_2] &= P(\text{both 1 and 2 show up at least once}) \\ &= 1 - P(\text{none of the rolls are equal to 1 or none of the rolls are equal to 2}) \\ &= 1 - \big(P(\text{the number 1 shows up}) + P(\text{the number 2 shows up}) \\ &\qquad - P(\text{neither 1 nor 2 shows up})\big) \\ &= 1 - \left(\left(\tfrac{5}{6}\right)^4 + \left(\tfrac{5}{6}\right)^4 - \left(\tfrac{2}{3}\right)^4\right) = 1 + \left(\tfrac{2}{3}\right)^4 - 2\cdot\left(\tfrac{5}{6}\right)^4 \end{aligned}$$

Collecting everything:

$$E[X^2] = 6\left(1 - \left(\tfrac{5}{6}\right)^4\right) + 30\left(1 + \left(\tfrac{2}{3}\right)^4 - 2\cdot\left(\tfrac{5}{6}\right)^4\right)$$

and

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - E[X]^2 \\ &= 6\left(1 - \left(\tfrac{5}{6}\right)^4\right) + 30\left(1 + \left(\tfrac{2}{3}\right)^4 - 2\cdot\left(\tfrac{5}{6}\right)^4\right) - 36\left(1 - \left(\tfrac{5}{6}\right)^4\right)^2 \\ &\approx 0.447. \end{aligned}$$

8.37. (a) Let J_k be the indicator for the event that the toy k is in at least one of the 4 boxes. Then $X = J_1 + J_2 + \cdots + J_{10}$. By linearity and exchangeability

$$\begin{aligned} E[X] &= E[\sum_{k=1}^{10} J_k] = \sum_{k=1}^{10} E[J_k] = 10E[J_1] \\ &= 10P(\text{the first toy was in one of the boxes}). \end{aligned}$$

Let A_k be the event that the kth toy was not in any of the four boxes. Then

$$E[X] = 10P(A_1^c) = 10(1 - P(A_1)).$$

We may assume that the toys in the boxes are chosen independently of each other, and hence

$$P(A_1) = P(\text{first box does not contain the first toy})^4 = \left(\frac{\binom{9}{2}}{\binom{10}{2}}\right)^4 = (\tfrac{4}{5})^4$$

and

$$E[X] = 10\left(1 - (\tfrac{4}{5})^4\right) = \frac{738}{125}.$$

(b) We need $E[X^2]$ which can be expressed using the introduced indicators as

$$\begin{aligned} E[X^2] &= E[(\sum_{k=1}^{10} J_k)^2] = E[\sum_{k=1}^{10} J_k^2 + 2\sum_{j<k} J_j J_k] \\ &= \sum_{k=1}^{10} E[J_k^2] + 2\sum_{j<k} E[J_j J_k] \\ &= 10E[J_1^2] + 2\binom{10}{2} E[J_1 J_2] \\ &= 10E[J_1] + 90E[J_1 J_2]. \end{aligned}$$

We used linearity, exchangeability and $J_1 = J_1^2$. Note that $10E[J_1] = E[X] = \frac{738}{125}$ by part (a). Recalling the definition of A_k from part (a) we get

$$E[J_1 J_2] = P(A_1^c A_2^c).$$

By taking complements,

$$P(A_1^c A_2^c) = 1 - P((A_1^c A_2^c)^c) = 1 - P(A_1 \cup A_2) = 1 - (P(A_1) + P(A_2) - P(A_1 A_2)).$$

As we have seen in part (a):

$$P(A_1) = P(A_2) = \left(\frac{\binom{9}{2}}{\binom{10}{2}}\right)^4 = (\tfrac{4}{5})^4,$$

and a similar computation gives

$$P(A_1 A_2) = \left(\frac{\binom{8}{2}}{\binom{10}{2}}\right)^4 = (\tfrac{28}{45})^4.$$

This gives

$$E[J_1 J_2] = 1 - 2(\tfrac{4}{5})^4 + (\tfrac{28}{45})^4$$

and

$$E[X^2] = \frac{738}{125} + 90\left(1 - 2(\tfrac{4}{5})^4 + (\tfrac{28}{45})^4\right),$$

which leads to

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\ &= \frac{738}{125} + 90\left(1 - 2(\tfrac{4}{5})^4 + (\tfrac{28}{45})^4\right) - (\tfrac{738}{125})^2 \\ &\approx 0.8092.\end{aligned}$$

8.38. Consider the coupon collector's problem with $n = 6$ (see Example 8.17). Then we have one of 6 possible toys in each box of cereal, each with probability 1/6, independently of the others. Thus we can imagine that the toy in a given box is chosen as the result of a die roll. Then finding all 6 toys means that we see all 6 numbers as outcomes among the die rolls. Hence the answer to our question is given by the solution of the coupon collector's problems with $n = 6$, by Example 8.17 the mean is $6(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}) = 14.7$ and the variance is

$$6^2(1 + \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{16} + \tfrac{1}{25}) - 6(1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5}) = 38.99.$$

8.39. Let $J_i = 1$ if a boy is chosen with the ith selection, and zero otherwise. Note that $E[J_i] = P\{X_i = 1\} = 17/40$. Then $X = \sum_{i=1}^{15} J_i$ and using linearity and exchangeability

$$E[X] = \sum_{i=1}^{15} P\{J_i = 1\} = 15 \times \frac{17}{40} = \frac{51}{8}.$$

Using the formula for the variance of the sum (together with exchangeability) gives

$$\begin{aligned}\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{15} J_i\right) &= \sum_{i=1}^{15} \mathrm{Var}(J_i) + 2\sum_{i<k} \mathrm{Cov}(J_i, J_k) \\ &= 15\mathrm{Var}(J_1) + 15 \cdot 14\,\mathrm{Cov}(J_1, J_2),\end{aligned}$$

Finding the variance of J_1 is easy since J_1 is a Bernoulli random variable:

$$\mathrm{Var}(J_1) = P(X_1 = 1)(1 - P(X_1) = 1) = \frac{17}{40} \cdot \frac{23}{40}.$$

To find the covariance, we have

$$\mathrm{Cov}(J_1, J_2) = E[J_1 J_2] - E[J_1]E[J_2] = E[J_1 J_2] - (\tfrac{17}{40})^2.$$

To find $E[J_1 J_2]$ note that $J_1 J_2 = 1$ only if a boy is called upon twice to start, and zero otherwise. Thus, by counting favorable outcomes we get

$$E[J_1 J_2] = \frac{\binom{17}{2}}{\binom{40}{2}} = \frac{34}{195}.$$

Collecting everything:

$$\mathrm{Var}(X) = 15 \cdot \frac{17}{40} \cdot \frac{23}{40} + 15 \cdot 14 \cdot \left(\tfrac{34}{195} - (\tfrac{17}{40})^2\right) = \frac{1955}{832}.$$

8.40. (a) We use the method of indicators. Let J_k be the indicator for the event that the number k is drawn in at least one of the 4 weeks. Then $X = J_1 + J_2 +$

$\cdots + J_{90}$. Then by the linearity of expectation and exchangeability we get

$$E[X] = E\left[\sum_{k=1}^{90} J_k\right] = \sum_{k=1}^{90} E[J_k]$$
$$= 90E[J_1].$$

We have

$$E[J_1] = P(1 \text{ is drawn at least one of the 4 weeks})$$
$$= 1 - P(1 \text{ is not drawn in any of the 4 weeks}))$$
$$= 1 - \left(\frac{89 \cdot 88 \cdot 87 \cdot 86 \cdot 85}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}\right)^4 = 1 - \left(\frac{85}{90}\right)^4.$$

From this

$$E[X] = 90E[J_1] = 90\left(1 - \left(\frac{85}{90}\right)^4\right) \approx 18.394.$$

(b) We first compute the second moment of X. Using the notation from part (b) we have

$$E[X^2] = E\left[\left(\sum_{k=1}^{90} J_k\right)^2\right] = E\left[\sum_{k=1}^{90} J_k^2 + 2\sum_{1\le k<\ell\le 90} J_k J_\ell\right]$$
$$= \sum_{k=1}^{90} E[J_k^2] + 2\sum_{1\le k<\ell\le 90} E[J_k J_\ell]$$
$$= 90E[J_1^2] + 2\cdot\binom{90}{2}E[J_1J_2],$$

where we used exchangeability again in the last step. Since J_1 is either zero or one, we have $J_1^2 = J_1$. Thus the term $90E[J_1^2]$ is the same as $90E[J_1]$ which is equal to $E[X]$. The second term can be computed as follows:

$$E[J_1J_2] = P(\text{both 1 and 2 are drawn at least once within the 4 weeks})$$
$$= 1 - P(\text{at least one of 1 and 2 is not drawn within of the 4 weeks}))$$
$$= 1 - \big(P(1 \text{ is not drawn in any of the 4 weeks})$$
$$+ P(2 \text{ is not drawn in any of the 4 weeks})$$
$$+ P(\text{neither 1 nor 2 is drawn in any of the 4 weeks})\big),$$

where we used inclusion-exclusion in the last step. We have

$$P(1 \text{ is not drawn in any of the 4 weeks})$$
$$= P(2 \text{ is not drawn in any of the 4 weeks}) = \left(\frac{85}{90}\right)^4,$$

and

$$P(\text{neither 1 nor 2 is drawn in any of the 4 weeks}) = \left(\frac{88 \cdot 87 \cdot 86 \cdot 85 \cdot 84}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}\right)^4 = \left(\frac{85 \cdot 84}{90 \cdot 89}\right)^4.$$

Putting everything together:

$$E[X^2] = 90\left(1 - \left(\frac{85}{90}\right)^4\right) + 90 \cdot 89\left(1 - 2 \cdot \left(\frac{85}{90}\right)^4 + \left(\frac{85 \cdot 84}{90 \cdot 89}\right)^4\right) \approx 339.59.$$

Now we can compute the variance:

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 \approx 339.59 - (18.394)^2 \approx 1.25.$$

8.41. We have

$$E[\bar{X}_n^3] = E\left[\left(\frac{X_1 + \cdots + X_n}{n}\right)^3\right] = \frac{1}{n^3}E\left[(X_1 + \cdots + X_n)^3\right].$$

By expanding the cube of the sum and using linearity and exchangeability

$$\begin{aligned} E[\bar{X}_n^3] &= \frac{1}{n^3}E\left[\sum_{k=1}^n X_k^3 + 6\sum_{i<j<k} X_i X_j X_k + 3\sum_{j\neq k} X_j^2 X_k\right] \\ &= \frac{1}{n^3}\left(\sum_{k=1}^n E[X_k^3] + 6\sum_{i<j<k} E[X_i X_j X_k] + 3\sum_{j\neq k} E[X_j^2 X_k]\right) \\ &= \frac{1}{n^3} \cdot n\,E[X_1^3] + 6\binom{n}{3}E[X_1X_2X_3] + 3n(n-1)E[X_1^2 X_2]. \end{aligned}$$

By independence

$$E[X_1X_2X_3] = E[X_1]E[X_2]E[X_3] = 0, \quad \text{and} \quad E[X_1^2X_2] = E[X_1^2]E[X_2] = 0,$$

hence

$$E[\bar{X}_n^3] = \frac{1}{n^3} \cdot n\,E[X_1^3] = \frac{b}{n^2}.$$

8.42. We have

$$E[\bar{X}_n^4] = E\left[\left(\frac{X_1 + \cdots + X_n}{n}\right)^4\right] = \frac{1}{n^3}E\left[(X_1 + \cdots + X_n)^4\right].$$

By expanding the fourth power of the sum and using linearity and exchangeability

$$\begin{aligned}
E[\bar{X}_n^4] &= \frac{1}{n^4} E\Big[\sum_{k=1}^n X_k^4 + 24 \sum_{i<j<k<\ell} X_i X_j X_k X_\ell \\
&\qquad + 12 \sum_{\substack{k<\ell \\ j\neq k, j\neq \ell}} X_j^2 X_k X_\ell + 6 \sum_{j<k} X_j^2 X_k^2 + 4 \sum_{j\neq k} X_j^3 X_k \Big] \\
&= \frac{1}{n^4} \sum_{k=1}^n E[X_k^4] + 24 \sum_{i<j<k<\ell} E[X_i X_j X_k X_\ell] \\
&\qquad + 12 \sum_{\substack{k<\ell \\ j\neq k, j\neq \ell}} E[X_j^2 X_k X_\ell] + 6 \sum_{j<k} E[X_j^2 X_k^2] + 4 \sum_{j\neq k} E[X_j^3 X_k] \\
&= \frac{1}{n^3} E[X_1^4] + 24 \binom{n}{4} E[X_1 X_2 X_3 X_4] \\
&\qquad + 12 \cdot \cdot \binom{n}{3} E[X_1^2 X_2 X_3] + 6 \binom{n}{2} E[X_1^2 X_2^2] + 4n(n-1) E[X_1^3 X_2].
\end{aligned}$$

By independence

$$E[X_1X_2X_3X_4] = E[X_1]E[X_2]E[X_3]E[X_4] = 0, \quad E[X_1^2X_2X_3] = E[X_1^2]E[X_2]E[X_3] = 0,$$
$$E[X_1^3X_2] = E[X_1^3]E[X_2] = 0, \qquad E[X_1^2X_2^2] = E[X_1^2]E[X_2^2] = E[X_1^2]^2.$$

Hence

$$E[\bar{X}_n^4] = \frac{1}{n^3} E[X_1^4] + \frac{3n(n-1)}{n^4} E[X_1^2]^2 = \frac{c}{n^3} + \frac{3(n-1)a^2}{n^3}.$$

8.43. (a) Note that $E[Z_i^2] = E[Z_i^2] - E[Z_i]^2 = \mathrm{Var}(Z_i) = 1$, because $E[Z_i] = 0$. Therefore by linearity we have

$$E[Y] = \sum_{i=1}^n E[Z_i^2] = nE[Z_1^2] = n.$$

For the variance, by independence, using independence

$$\mathrm{Var}(Y) = \sum_{i=1}^n \mathrm{Var}(Z_i^2) = n\mathrm{Var}(Z_1^2).$$

We have

$$\mathrm{Var}(Z_1^2) = E[Z_1^4] - E[Z_1^2]^2.$$

The fourth moment of a standard normal random variable in Exercise 3.69: $E[Z_1^4] = 3$. Thus,

$$\mathrm{Var}(Y) = n\mathrm{Var}(Z_1^2) = n(3-1) = 2n.$$

(b) The moment generating function of Y is

$$M_Y(t) = E[e^{tY}] = E[e^{t(Z_1^2+Z_2^2+\cdots+Z_n^2)}].$$

By the independence of Z_i we can write the right hand side as a product of the individual moment generating functions, and using the fact that the Z_i are i.i.d. we get

$$M_Y(t) = M_{Z_1^2}(t)^n.$$

We compute the moment generating function of Z_1^2 by computing the expectation $E[e^{tZ_1^2}]$. We have

$$E[e^{tZ_1^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz^2} e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{(2t-1)z^2}{2}}\, dz.$$

This integral convergences only for $t < 1/2$ (otherwise we integrate a function that is always at least 1). Moreover, we can write this using the integral of the probability density function of an $\mathcal{N}(0, \frac{1}{2t-1})$ random variable:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{z^2}{2\frac{1}{2t-1}}}\, dz = \frac{1}{\sqrt{2t-1}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\frac{1}{2t-1}}} e^{\frac{(2t-1)z^2}{2}}\, dz = \frac{1}{\sqrt{2t-1}}.$$

Therefore,

$$M_Y(t) = \begin{cases} (1-2t)^{-n/2} & \text{for } t < 1/2 \\ \infty & \text{for } t \geq 1/2. \end{cases}$$

Using the moment generating function we calculate the mean to be

$$E[Y] = M_Y'(0) = n.$$

For the variance, we first calculate the second moment,

$$E[Y^2] = M_Y''(0) = n(n-2) = n(n-2).$$

From this the variance is

$$\mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = n^2 - 2n - n^2 = 2n.$$

8.44. (a) From the definition

$$M_X(t) = E[e^{tX}] = \sum_{k=1}^{3} p_X(k) e^{tk} = \frac{1}{4}e^t + \frac{1}{4}e^{2t} + \frac{1}{2}e^{3t}$$

and similarly,

$$M_Y(t) = \frac{1}{7}e^{2t} + \frac{2}{7}e^{3t} + \frac{4}{7}e^{4t}.$$

(b) Since X and Y are independent, we have $M_{X+Y}(t) = M_X(t)M_Y(t)$. Using the result of part (a) we get

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \left(\tfrac{1}{4}e^t + \tfrac{1}{4}e^{2t} + \tfrac{1}{2}e^{3t}\right)\left(\tfrac{1}{7}e^{2t} + \tfrac{2}{7}e^{3t} + \tfrac{4}{7}e^{4t}\right).$$

Expanding the product gives

$$M_{X+Y}(t) = \frac{e^{3t}}{28} + \frac{3e^{4t}}{28} + \frac{2e^{5t}}{7} + \frac{2e^{6t}}{7} + \frac{2e^{7t}}{7}.$$

We can identify the possible values of $X+Y$ by looking at the exponents. The probability mass function at k is just the coefficient of e^{kt}. This gives

$$p_{X+Y}(3) = \frac{1}{28},\ p_{X+Y}(4) = \frac{3}{28},\ p_{X+Y}(5) = \frac{2}{7},\ p_{X+Y}(6) = \frac{2}{7},\ p_{X+Y}(7) = \frac{2}{7}.$$

8.45. Using the joint probability mass function we can compute

$$\begin{aligned}
E[XY] &= 1 \cdot 1 \cdot p_{X,Y}(1,1) + 1 \cdot 2 \cdot p_{X,Y}(1,2) + 2 \cdot 0 \cdot p_{X,Y}(2,0) \\
&\quad + 2 \cdot 1 \cdot p_{X,Y}(2,1) + 3 \cdot 1 \cdot p_{X,Y}(3,1) + 3 \cdot 2 \cdot p_{X,Y}(3,2) = \frac{16}{9}, \\
E[X] &= 1 \cdot p_{X,Y}(1,1) + 1 \cdot p_{X,Y}(1,2) + 2 \cdot p_{X,Y}(2,0) \\
&\quad + 2 \cdot p_{X,Y}(2,1) + 3 \cdot p_{X,Y}(3,1) + 3 \cdot p_{X,Y}(3,2) = 2, \\
E[Y] &= 1 \cdot p_{X,Y}(1,1) + 2 \cdot p_{X,Y}(1,2) + 0 \cdot p_{X,Y}(2,0) \\
&\quad + 1 \cdot p_{X,Y}(2,1) + 1 \cdot p_{X,Y}(3,1) + 2 \cdot p_{X,Y}(3,2) = \frac{8}{9}.
\end{aligned}$$

Then $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{16}{9} - 2 \cdot \frac{8}{9} = 0$, which means that $\mathrm{Corr}(X,Y) = 0$ as well.

8.46. The first five and last five draws together will give all the draws, thus $X+Y = 6$ and $Y = 6 - X$. Then

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X, 6-X) = -\mathrm{Cov}(X,X) = -\mathrm{Var}(X).$$

The number of red balls in the first five draws has a hypergeometric distribution with $N_A = 6$, $N_B = 4$, $N = 10$, $n = 5$. In Example we computed the variance of such a random variable to get

$$\mathrm{Var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{N_A}{N} \cdot \frac{N_B}{N} = \frac{10-5}{10-1} \cdot 5 \cdot \frac{6}{10} \cdot \frac{4}{10} = \frac{2}{3}.$$

This leads to $\mathrm{Cov}(X,Y) = -\mathrm{Var}(X) = -\frac{2}{3}$.

8.47. The mean of X is given by the solution of Exercise 8.3. As in the solution of Exercise 8.3, introduce indicators so that $X = X_B + X_C + X_D$. Assumption (i) of the problem implies that $\mathrm{Cov}(X_B, X_D) = \mathrm{Cov}(X_C, X_D) = 0$. Assumption (ii) of the problem implies that

$$\begin{aligned}
\mathrm{Cov}(X_B, X_C) &= E[X_B X_C] - E[X_B]E[X_C] \\
&= P(X_B = 1, X_C = 1) - P(X_B = 1)P(X_C = 1) \\
&= P(X_C = 1 | X_B = 1)P(X_B = 1) - P(X_B = 1)P(X_C = 1) \\
&= 0.8 \cdot 0.3 - 0.3 \cdot 0.4 = 0.12.
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(X_B + X_C + X_D) = \mathrm{Var}(X_B) + \mathrm{Var}(X_C) + \mathrm{Var}(X_D) \\
&\quad + 2[\mathrm{Cov}(X_B, X_C) + \mathrm{Cov}(X_B, X_D) + \mathrm{Cov}(X_C, X_D)] \\
&= 0.3 \cdot 0.7 + 0.4 \cdot 0.6 + 0.7 \cdot 0.3 + 2 \cdot 0.12 = 0.9
\end{aligned}$$

8.48. The joint probability mass function of the random variables (X, Y) can be represented by the following table.

$X \backslash Y$	0	1	2
1	$\frac{9}{100}$	0	0
2	$\frac{81}{100}$	$\frac{9}{100}$	0
3	0	0	$\frac{1}{100}$

Hence, the marginal distribution are:

$$p_X(1) = \tfrac{9}{100}, \quad p_X(2) = \tfrac{90}{100}, \quad p_X(3) = \tfrac{1}{100}$$
$$p_Y(0) = \tfrac{90}{100}, \quad p_Y(1) = \tfrac{9}{100}, \quad p_Y(2) = \tfrac{1}{100}.$$

From these we can compute the following expectations:

$$E[X] = \tfrac{48}{25}, \quad E[Y] = \tfrac{11}{100}, \quad E[XY] = \tfrac{6}{25},$$

and so

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = \tfrac{6}{25} - \tfrac{48}{25} \cdot \tfrac{11}{100} = \tfrac{18}{625}.$$

8.49. We need $E[X], E[Y], E[XY]$. The joint density of X, Y is $f(x, y) = \mathbf{1}((x, y) \in D))$ (the area is 1) and the bounding lines of D are $y = 1$, $y = x$, $y = -x$. We get

$$E[X] = \iint_{(x,y)\in D} x f(x, y) dx dy = \int_0^1 \int_{-y}^{y} x dx dy = \int_0^1 (y^2/2 - (-y)^2/2) dy = 0,$$
$$E[Y] = \iint_{(x,y)\in D} y f(x, y) dx dy = \int_0^1 \int_{-y}^{y} y dx dy = \int_0^1 2y^2 dy = \frac{2}{3},$$
$$E[XY] = \iint_{(x,y)\in D} xy f(x, y) dy dx = \int_0^1 \int_{-y}^{y} xy dx dy = \int_0^1 (y^3/2 - y(-y)^2/2) dy = 0.$$

This gives

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.$$

Solution without computation:
By symmetry we see that (X, Y) has the same distribution as $(-X, Y)$. This implies $E[X] = E[-X] = -E[X]$ yielding $E[X] = 0$. It also implies $E[XY] = E[-XY] = -E[XY]$ which gives $E[XY] = 0$. This immediately shows that

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.$$

8.50. Note that if (x, y) is on the union of the line segments AB and AC then either x or y is equal to zero. This means that $XY = 0$, and $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = -E[X]E[Y]$.

To compute $E[X]$ and $E[Y]$ is a little bit tricky, since X and Y are neither continuous, nor discrete. However, we can write both of them as a function of a continuous random variable. Imagine that we rotate AC 90 degrees about $(0, 0)$ so that it C is rotated into $(-1,0)$. Let Z be a uniformly chosen point on the line segment connecting $(-1,0)$ and $(1,0)$. We can get (X,Y) as the following function of Z:

$$g(z) = \begin{cases} (z,0), & \text{if } z \geq 0 \\ (0,-z), & \text{if } z < 0. \end{cases}$$

In other words: we 'fold out' the union of AB and AC so that it becomes the line segment connecting $(-1,0)$ and $(1,0)$, choose a point Z on it uniformly, and then 'fold' it back into the original $AB \cup AC$.

The density function of Z is $\frac{1}{2}$ on $(-1,1)$, and zero otherwise and $X = h(Z) = \max(z,0)$. Thus

$$E[X] = \int_{-1}^{1} \frac{1}{2} \max(z,0) dz = \int_{0}^{1} \frac{z}{2} dz = \frac{1}{4}.$$

Similarly,

$$E[Y] = \int_{-1}^{1} \frac{1}{2} \max(-z,0) dz = -\int_{-1}^{0} \frac{z}{2} dz = \frac{1}{4}.$$

This gives $\mathrm{Cov}(X,Y) = -E[X]E[Y] = -\frac{1}{16}$.

8.51. We start by computing the second moment:

$$\begin{aligned} E[(X+2Y+Z)^2] &= E[X^2 + 4Y^2 + Z^2 + 4XY + 2XZ + 4YZ] \\ &= E[X^2] + 4E[Y^2] + E[Z^2] + 4E[XY] + 2E[XZ] + 4E[YZ] \\ &= 2 + 4 \cdot 12 + 12 + 4 \cdot 2 + 2 \cdot 4 + 4 \cdot 9 \\ &= 114. \end{aligned}$$

Then the variance is given by

$$\mathrm{Var}(X+2Y+Z) = E[(X+2Y+Z)^2] - (E[X+2Y+Z])^2 = 114 - (1+2\cdot 3+3)^2 = 114 - 100 = 14$$

One could also compute all the variances and pairwise covariances first and use

$$\mathrm{Var}(X+2Y+Z) = \mathrm{Var}(X) + 4\,\mathrm{Var}(Y) + \mathrm{Var}(Z) + 4\,\mathrm{Cov}(X,Y) + 2\,\mathrm{Cov}(X,Z) + 4\,\mathrm{Cov}(Y,Z).$$

8.52. For the correlation we need $\mathrm{Cov}(X,Y)$, $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$. Both X and Y have $\mathrm{Bin}(20, \frac{1}{2})$ distribution, thus

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = 20 \cdot \frac{1}{2} \cdot \frac{1}{2} = 5.$$

Denote by Z_i the number of heads among the coin flips $10(i-1)+1, 10(i-1)+2, \ldots, 10i$. Then Z_1, Z_2, Z_3 are independent, they all have $\mathrm{Bin}(10, \frac{1}{2})$ distribution, and we have $X = Z_1 + Z_2$ and $Y = Z_2 + Z_3$. Using the properties of the covariance and the independence of Z_1, Z_2, Z_3:

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathrm{Cov}(Z_1 + Z_2, Z_2 + Z_3) \\ &= \mathrm{Cov}(Z_1, Z_2) + \mathrm{Cov}(Z_2, Z_2) + \mathrm{Cov}(Z_1, Z_3) + \mathrm{Cov}(Z_2, Z_3) \\ &= \mathrm{Cov}(Z_2, Z_2) = \mathrm{Var}(Z_2) = 10 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{2}. \end{aligned}$$

Now we can compute the correlation:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\frac{5}{2}}{\sqrt{5 \cdot 5}} = \frac{1}{2}.$$

Here is another way to compute the covariance. Let I_j be the indicator of the event that the jth flip is heads. These are independent Ber(1/2) distributed random variables. We have $X = \sum_{k=1}^{20} I_k$ and $Y = \sum_{k=21}^{30} I_k$, and using the properties of covariance and the independence we get

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{Cov}\Big(\sum_{k=1}^{20} I_k, \sum_{j=11}^{30} I_j\Big) \\
&= \sum_{k=1}^{20}\sum_{j=11}^{30} \mathrm{Cov}(I_k, I_j) \\
&= \sum_{k=11}^{20} \mathrm{Cov}(I_k, I_k) = \sum_{k=11}^{20} \mathrm{Var}(I_k) = 10 \cdot \frac{1}{2} \cdot \frac{1}{2}.
\end{aligned}$$

8.53. (a) We have $\mathrm{Cov}(3X+2, 2Y-3) = 3 \cdot 2\, \mathrm{Cov}(X,Y)$. Also:

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = -1 - 1 \cdot 2 = -3.$$

Thus $\mathrm{Cov}(3X+2, 2Y-3) = 3 \cdot 2 \cdot (-3) = -18$.
(b) We have

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = 3 - 1^2 = 2, \qquad \mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = 13 - 2^2 = 9.$$

Using that $\mathrm{Cov}(X,Y) = -3$ we get

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-3}{\sqrt{2 \cdot 9}} = -\frac{1}{\sqrt{2}}.$$

8.54. (a) We have

$$\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 = 5 - 2^2 = 1 \\
\mathrm{Var}(Y) &= E[Y^2] - (E[Y])^2 = 10 - 1^2 = 9 \\
\mathrm{Cov}(X,Y) &= E[XY] - E[X]E[Y] = 1 - 2 \cdot 1 = -1.
\end{aligned}$$

Then

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-1}{\sqrt{1 \cdot 9}} = -\frac{1}{3}.$$

(b) We have

$$\mathrm{Cov}(X, X+cY) = \mathrm{Var}(X) + c\,\mathrm{Cov}(X,Y) = 1 - c(-1) = 1 + c.$$

Thus for $c = -1$ the random variables X and $X + cY$ are uncorrelated.

8.55. Note that $I_{A^c} = 1 - I_A$ and $I_{B^c} = 1 - I_B$. Then from Theorem 8.36 we have

$$\mathrm{Corr}(I_{A^c}, I_{B^c}) = \mathrm{Corr}(1-I_A, 1-I_B) = (-1)\cdot\mathrm{Corr}(I_A, 1-I_B) = (-1)\cdot(-1)\cdot\mathrm{Corr}(I_A, I_B).$$

8.56. From the properties of variance and covariance:

$$\begin{aligned}
\mathrm{Var}(aX+c) &= a^2\,\mathrm{Var}(X) \\
\mathrm{Var}(bY+d) &= b^2\,\mathrm{Var}(Y) \\
\mathrm{Cov}(aX+c, bY+d) &= ab\,\mathrm{Cov}(X,Y).
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Corr}(aX+c, bY+d) &= \frac{\mathrm{Cov}(aX+c, bY+d)}{\sqrt{\mathrm{Var}(aX+c)\,\mathrm{Var}(bY+d)}} \\
&= \frac{ab\,\mathrm{Cov}(X,Y)}{\sqrt{a^2b^2\,\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \\
&= \frac{ab}{|a|\cdot|b|}\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{ab}{|a|\cdot|b|}\,\mathrm{Corr}(X,Y).
\end{aligned}$$

The coefficient $\frac{ab}{|a|\cdot|b|}$ is 1 if $ab > 0$ and -1 if $ab < 0$.

8.57. Assume that there are random variables satisfying the listed conditions. Then

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = 3 - 1^2 = 2, \quad \mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = 5 - 2^2 = 1$$

and

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = -1 - 1\cdot 2 = -3.$$

From this the correlation is

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-3}{\sqrt{2\cdot 1}} = -\frac{3}{\sqrt{2}}.$$

But $-\frac{3}{\sqrt{2}} < -1$, and we know that the correlation must be in $[-1,1]$. The found contradiction shows that we cannot find such random variables.

8.58. By the discussion in Section 8.5 if Z and W are independent standard normals then with

$$X = \sigma_X Z + \mu_X, \qquad Y = \sigma_Y \rho Z + \sigma_Y\sqrt{1-\rho^2}\,W + \mu_Y$$

the random variables (X,Y) have bivariate normal distribution with marginals $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and correlation $\mathrm{Corr}(X,Y) = \rho$. Then we have

$$\begin{aligned}
U &= 2X + Y = (2\sigma_X + \sigma_Y\rho)Z + \sigma_Y\sqrt{1-\rho^2}\,W + 2\mu_X + \mu_Y \\
V &= X - Y = (\sigma_X - \sigma_Y\rho)Z - \sigma_Y\sqrt{1-\rho^2}\,W + \mu_X - \mu_Y.
\end{aligned}$$

We can turn this system of equations into a single vector valued equation:

$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} 2\sigma_X + \sigma_Y\rho & \sigma_Y\sqrt{1-\rho^2} \\ \sigma_X - \sigma_Y\rho & -\sigma_Y\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} Z \\ W \end{bmatrix} + \begin{bmatrix} 2\mu_X + \mu_Y \\ \mu_X - \mu_Y \end{bmatrix}$$

In Section 8.6 it was shown that if Z, W are independent standard normals, A is a 2×2 matrix and μ is an $\mathbb{R}^2$ valued vector then $A[Z,W]^T + \mu$ is a bivariate normal with mean vector μ and covariance matrix AA^T. Thus (U,V) is a bivariate normal and we just have to identify the individual means, variances and the correlation of U and V.

Using the properties of mean, variance and covariance together gives

$$\begin{aligned}
E[U] &= E[2X+Y] = 2\mu_X + \mu_Y \\
E[V] &= E[X-Y] = \mu_X - \mu_Y \\
\operatorname{Var}(U) &= \operatorname{Var}(2X+Y) = 4\operatorname{Var}(X) + \operatorname{Var}(Y) + 4\operatorname{Cov}(X,Y) = 4\sigma_X^2 + \sigma_Y^2 + 4\sigma_X\sigma_Y\rho \\
\operatorname{Var}(V) &= \operatorname{Var}(X-Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X,Y) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho \\
\operatorname{Cov}(U,V) &= \operatorname{Cov}(2X+Y, X-Y) = 2\operatorname{Var}(X) + \operatorname{Cov}(X,Y) - 2\operatorname{Cov}(X,Y) - \operatorname{Var}(Y,Y) \\
&= 2\sigma_X^2 - \sigma_Y^2 - 2\sigma_X\sigma_Y\rho.
\end{aligned}$$

We also used the fact that $\operatorname{Cov}(X,Y) = \operatorname{Corr}(X,Y)\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}$.

Finally,

$$\operatorname{Corr}(U,V) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(U)\operatorname{Var}(V)}} = \frac{2\sigma_X^2 - \sigma_Y^2 - 2\sigma_X\sigma_Y\rho}{\sqrt{(4\sigma_X^2 + \sigma_Y^2 + 4\sigma_X\sigma_Y\rho)(\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho)}}.$$

Thus (U,V) has bivariate normal distribution with the parameters identified above. Remark: the joint density of U,V can also be identified by considering the joint probability density of (X,Y) from (8.32) and using the Jacobian technique of Section 6.4 to derive the joint density function of $(U,V) = (2X+Y, X-Y)$.

8.59. We can express X and Y in terms of Z and W as $X = g(Z,W)$, $Y = h(Z,W)$ with $g(z,w) = \sigma_X z + \mu_X$ and $h(z,w) = \sigma_Y\rho z + \sigma_Y\sqrt{1-\rho^2}\,w + \mu_Y$. Solving the equations

$$x = \sigma_X z + \mu_X, \quad y = \sigma_Y\rho z + \sigma_Y\sqrt{1-\rho^2}\,w + \mu_Y$$

for z,w gives the inverse of the function $(g(z,w), h(z,w))$. The solution is

$$z = \frac{x-\mu_X}{\sigma_X}, \quad w = \frac{(y-\mu_Y)\sigma_X - (x-\mu_X)\rho\sigma_Y}{\sqrt{1-\rho^2}\sigma_X\sigma_Y},$$

thus the inverse of $(g(z,w), h(z,w))$ is the function $(q(x,y), r(x,y))$ with

$$q(x,y) = \frac{x-\mu_X}{\sigma_X}, \qquad r(x,y) = \frac{(y-\mu_Y)\sigma_X - (x-\mu_X)\rho\sigma_Y}{\sqrt{1-\rho^2}\sigma_X\sigma_Y}.$$

The Jacobian of $(q(x,y), r(x,y))$ with respect to x,y is

$$J(x,y) = \det\begin{bmatrix} 1/\sigma_X & 0 \\ -\frac{\rho}{\sigma_X\sqrt{1-\rho^2}} & \frac{1}{\sigma_Y\sqrt{1-\rho^2}} \end{bmatrix} = \frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

Using Fact 6.41 we get the joint density of X and Y:

$$f_{X,Y}(x,y) = f_{Z,W}\left(\frac{x-\mu_X}{\sigma_X}, \frac{(y-\mu_Y)\sigma_X - (x-\mu_X)\rho\sigma_Y}{\sqrt{1-\rho^2}\sigma_X\sigma_Y}\right)\cdot\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

Since Z and W are independent standard normals, we have $f_{Z,W}(z,w) = \frac{1}{2\pi}e^{-\frac{z^2+w^2}{2}}$. Thus

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{1}{2}\left(\frac{(y-\mu_Y)\sigma_X - (x-\mu_X)\rho\sigma_Y}{\sqrt{1-\rho^2}\sigma_X\sigma_Y}\right)^2\right]$$

Rearranging the terms in the exponent shows that the found joint density is the same as the one given in (8.32). This shows that the distribution of (X, Y) is bivariate normal with parameters $\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho$.

8.60. The number of ways in which toys can be chosen so that new toys appear at times $1, 1+a_1, 1+a_1+a_2, \dots, 1+a_1+\dots+a_{n-1}$ is

$$n\cdot 1^{a_1-1}\cdot(n-1)\cdot 2^{a_2-1}\cdot(n-2)\cdot 3^{a_3-1}\cdot(n-3)\cdots 2\cdot(n-1)^{a_{n-1}-1}\cdot 1 = n\cdot\prod_{k=1}^{n-1}(n-k)\cdot k^{a_k-1}.$$

The total number of sequences of $1+a_1+\dots+a_{n-1}$ toys is $n^{1+a_1+\dots+a_{n-1}}$. The probability is

$$\begin{aligned} P(W_1 = a_1, \dots, W_{n-1} = a_{n-1}) &= \frac{n\cdot\prod_{k=1}^{n-1}(n-k)\cdot k^{a_k-1}}{n^{1+a_1+\dots+a_{n-1}}} = \prod_{k=1}^{n-1}\frac{n-k}{n}\left(\frac{k}{n}\right)^{a_k-1} \\ &= \prod_{k=1}^{n-1} P(W_k = a_k). \end{aligned}$$

where in the last step we used the fact that $W_1, W_2, \dots, W_{k-1}$ are independent with $W_j \sim \text{Geom}(\frac{n-j}{n})$.

8.61. (a) Since $f(x) = \frac{1}{x}$ is a decreasing function, by the bounds shown in Figure D.1 we get

$$\sum_{k=2}^{n}\frac{1}{k} \le \int_1^n \frac{1}{x}dx \le \sum_{k=1}^{n-1}\frac{1}{k}.$$

Since $\int_1^n \frac{1}{x}dx = \ln n$ this gives

$$\ln n \ge \sum_{k=2}^{n}\frac{1}{k} = \sum_{k=1}^{n}\frac{1}{k} - 1$$

and

$$\ln n \le \sum_{k=1}^{n-1}\frac{1}{k} \le \sum_{k=1}^{n}\frac{1}{k}$$

which together give $0 \le \sum_{k=1}^{n}\frac{1}{k} - \ln n \le 1$.

(c) In Example 8.17 we have shown that $E[T_n] = n\sum_{k=1}^{n}\frac{1}{n}$. Using the bounds in part (a) we have

$$n\ln n \le nE[T_n] \le n(\ln n + 1)$$

from which $\lim_{n\to\infty}\frac{E(T_n)}{n\ln n} = 1$ follows.

We have also shown

$$\text{Var}(T_n) = n^2\sum_{j=1}^{n-1}\frac{1}{j^2} - n\sum_{j=1}^{n-1}\frac{1}{j},$$

and hence

$$\frac{\text{Var}(T_n)}{n^2} = \sum_{j=1}^{n-1}\frac{1}{j^2} - \frac{1}{n}\sum_{j=1}^{n-1}\frac{1}{j}.$$

Since $\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}$ we have $\lim_{n\to\infty} \sum_{j=1}^{n-1} \frac{1}{j^2} = \frac{\pi^2}{6}$. We also have $0 \le \sum_{j=1}^{n-1} \frac{1}{j} \le \ln n$ by part (a), and we know that $\lim_{n\to\infty} \frac{\ln n}{n} = 0$, thus $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} \frac{1}{j} = 0$. But this means that $\lim_{n\to\infty} \frac{\mathrm{Var}(T_n)}{n^2} = \frac{\pi^2}{6}$.

Solutions to Chapter 9

9.1. (a) The expected value of Y is $E[Y] = \frac{1}{p} = 6$. Since Y is nonnegative, we can use Markov's inequality to get the bound $P(Y \geq 16) \leq \frac{E[Y]}{16} = \frac{6}{16} = \frac{3}{8}$.
(b) The variance of Y is $\text{Var}(Y) = \frac{q}{p^2} = \frac{\frac{5}{6}}{\frac{1}{36}} = 30$. Using Chebyshev's inequality we get

$$P(Y \geq 16) = P(Y - E[Y] \geq 10) \leq P(|Y - E[Y]| \geq 10) \leq \frac{\text{Var}(Y)}{10^2} = \frac{30}{100} = \frac{3}{10}.$$

(c) The exact value of $P(Y \geq 16)$ can be computed for example by treating Y as the number trials needed for the first success in a sequence of independent trials with success probability p. Then

$$P(Y \geq 16) = P(\text{first 15 trials all failed}) = q^{15} = (5/6)^15 \approx 0.0649.$$

We can see that the estimates in (a) and (b) are valid, although they are not very close to the truth.

9.2. (a) We have $E[X] = \frac{1}{\lambda} = 2$ and $X \geq 0$. By Markov's inequality

$$P(X > 6) \leq \frac{E[X]}{6} = \frac{1}{3}.$$

(b) We have $E[X] = \frac{1}{\lambda} = 2$, $\text{Var}[X] = \frac{1}{\lambda^2} = 4$. By Chebyshev's inequality

$$P(X > 6) = P(X - E[X] > 4) \leq P(|X - E[X]| > 4) \leq \frac{\text{Var}(X)}{4^2} = \frac{4}{4^2} = \frac{1}{4}.$$

9.3. Let X_i be the price change between day $i - 1$ and day i (with day 0 being today). Then $C_n - C_0 = X_1 + X_2 + \cdots + X_n$. The expectation of X_i (for each i) is given by $E[X_i] = E[X_1] = 0.45 \cdot 1 + 0.5 \cdot (-2) + 0.05 \cdot (10) = -0.05$. We can also check that the variance is finite. We have

$$P(C_n > C_0) = P(C_n - C_0 > 0) = P(\sum_{i=1}^{n} X_i > 0) = P(\tfrac{1}{n}\sum_{i=1}^{n} X_i > 0)$$
$$= P(\tfrac{1}{n}\sum_{i=1}^{n} X_i - E[X_1] > 0.05).$$

By the law of large numbers (Theorem 9.9) we have

$$P(\tfrac{1}{n}\sum_{i=1}^{n} X_i - E[X_1] > 0.05) \le P(|\tfrac{1}{n}\sum_{i=1}^{n} X_i - E[X_1]| > 0.05) \to 0$$

as $n \to \infty$. Thus $\lim_{n\to\infty} P(C_n > C_0) = 0$.

9.4. In each round Ben wins \$1 with probability $\frac{18}{37}$ and loses \$1 with probability $\frac{19}{37}$. Let X_k be Ben's net winnings in the kth round, we may assume that $X_1, X_2, \dots$ are independent. We have $\mu = E[X_k] = \frac{18}{37} - \frac{19}{37} = -\frac{1}{37}$. If we denote by S_k the total net winnings within the first k rounds then $S_k = X_1 + \cdots + X_k$. By the law of large numbers $\frac{S_n}{n}$ will be close to $\mu = -\frac{1}{37}$ with high probability. More precisely, for any $\varepsilon > 0$ we the probability $P\big(\left|\frac{S_n}{n} + \frac{1}{37}\right| < \varepsilon\big)$ converges to 1 as $n \to \infty$.

This means that for large n with high probability Ben will lose many after n rounds.

9.5. (a) Using Markov's inequality:

$$P(X \ge 15) \le \frac{E[X]}{15} = \frac{10}{15} = \frac{2}{3}.$$

(b) Using Chebyshev's inequality:

$$P(X \ge 15) = P(X - 10 \ge 5) \le \frac{Var(X)}{5^2} = \frac{3}{25}$$

(c) Let $S = \sum_{i=1}^{300} Y_i$. Use the general version of the Central Limit Theorem to estimate $P(S > 3030)$, by first standardizing the sum, then replacing the standardized sum with a standard normal:

$$\begin{aligned} P(S > 3030) &= P\left(\frac{S - 300\cdot 10}{\sqrt{3\cdot 300}} > \frac{3030 - 300\cdot 10}{\sqrt{3\cdot 300}}\right) \\ &= P\left(\frac{S - 300\cdot 10}{\sqrt{3\cdot 300}} > 1\right) \\ &\approx 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587 \end{aligned}$$

9.6. Let X_k denote the time needed in seconds to it the kth hot dog, and denote by S_n the sum $X_1 + \cdots + X_n$. Since 15 minutes is 900 seconds, we need to estimate the probability $P(S_{64} < 900)$. By the CLT the standardized random variable $\frac{S_{64} - 64\cdot 15}{\sqrt{64\cdot 4^2}}$ is close to a standard normal. Thus

$$\begin{aligned} P(S_{64} < 900) &= P\left(\frac{S_{64} - 64\cdot 15}{\sqrt{64\cdot 5^2}} < \frac{900 - 64\cdot 15}{\sqrt{64\cdot 4^2}}\right) \\ &\approx \Phi\left(\frac{900 - 64\cdot 15}{\sqrt{64\cdot 4^2}}\right) = \Phi(-1.875) = 1 - \Phi(1.875) \\ &\approx 0.0304, \end{aligned}$$

where we used linear interpolation to approximate $\Phi(1.875)$ using the table in the Appendix.

9.7. Let X_i be the size of the claim made by the ith policyholder. Let m be the premium they charge. We desire a premium m for which

$$P\left(\sum_{i=1}^{2,500} X_i \leq 2,500 \cdot m\right) \geq 0.999.$$

We first use Chebyshev's inequality to estimate the probability of the complement. Recall that $\mu = E[X_i] = 1000$ and $\sigma = \sqrt{\mathrm{Var}(X_i)} = 900$. Using the notation $S = \sum_{i=1}^{2,500} X_i$ we have

$$E[S] = 2500\mu, \qquad \mathrm{Var}(S) = 2500\sigma^2.$$

By Chebyshev's inequality (assuming $m > \mu$)

$$\begin{aligned} P(S \geq 2,500 \cdot m) &= P(S - 2500\mu \geq 2,500 \cdot (m-\mu)) \\ &\leq \frac{\mathrm{Var}(S)}{2500^2 \cdot (m-\mu)^2} = \frac{2500 \cdot 900^2}{2500^2 \cdot (m-\mu)^2} = \frac{324}{(m-1000)^2}. \end{aligned}$$

We need this probability to be at most $1 - 0.999 = 0.001$, which leads to $\frac{324}{(m-1000)^2} \leq 0.001$ and

$$m \geq 1000 + \frac{18}{\sqrt{0.001}} \approx 1569.21.$$

Note that we assumed $m > \mu$ which was natural: for $m \leq \mu$ we can use Chebyshev's inequality that the probability in question cannot be at least 0.999.

Now let us see how we can estimate $P\left(\sum_{i=1}^{2,500} X_i \leq 2,500 \cdot m\right)$ using the central limit theorem. We have

$$\begin{aligned} P(S \leq 2500 \cdot m) &= P\left(\frac{S - 2,500 \cdot 1,000}{\sqrt{2,500} \cdot 900} \leq \frac{2,500 \cdot m - 2,500 \cdot 1,000}{\sqrt{2,500} \cdot 900}\right) \\ &\approx \Phi\left(\frac{2500(m-1,000)}{\sqrt{2,500} \cdot 900}\right) = \Phi\left(\frac{m-1,000}{18}\right) \end{aligned}$$

We would like this probability to be at most 0.999. Using the table in Appendix E we get that $\Phi\left(\frac{m-1,000}{18}\right) \geq 0.999$ if $\frac{m-1,000}{18} \geq 3.1$ which leads to $m \geq 1055.8$.

9.8. (a) This is just the area of the quarter of the unit disk, multiplied by 4.

(b) We have

$$\int_0^1 \int_0^1 4 \cdot I(x^2 + y^2 \leq 1)\, dx\, dy = E[g(U_1, U_2)]$$

where U_1, U_2 are independent Unif$[0,1]$ random variables and $g(x,y) = 4 \cdot I(x^2 + y^2 \leq 1)$.

(c) We need to generate $n = 10^6$ independent samples of the random variable $g(U_1, U_2)$. If $\bar{\mu}$ is the sample mean and s_n^2 is the sample variance then the appropriate confidence interval is $(\bar{\mu} - \frac{1.96 \cdot s_n}{\sqrt{n}}, \bar{\mu} + \frac{1.96 \cdot s_n}{\sqrt{n}})$.

9.9. (a) Using Markov's inequality we have

$$P\{X \geq 7,000\} \leq \frac{E[X]}{7,000} = \frac{5}{7}.$$

(b) Using Chebyshev's inequality we have

$$P\{X \geq 7,000\} = P(X - 5,000 \geq 2,000) \leq \frac{4,500}{2000^2} = \frac{9}{8000} = 0.001125.$$

(c) We want n so that

$$P\left(\left|\frac{S_n}{n} - 5,000\right| \geq 50\right) \leq 0.05.$$

Using Chebyshev's inequality we have that

$$P\left(\left|\frac{S_n}{n} - 5,000\right| \geq 50\right) \leq \frac{\mathrm{Var}(S_n/n)}{50^2} = \frac{n\mathrm{Var}(X_1)}{n^2 50^2} = \frac{4,500}{n \cdot 50^2} = \frac{9}{n \cdot 5}.$$

Hence, it is sufficient to choose an n so that

$$\frac{9}{n \cdot 5} \leq 0.05 = \frac{1}{20} \implies n \geq \frac{9 \cdot 20}{5} = 9 \cdot 4 = 36.$$

9.10. We have

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j\leq n} \mathrm{Cov}(X_i, X_j).$$

Since we have $\mathrm{Var}(X_i) = 4500$, this gives $\mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i,X_j)}{4500}$. Hence

$$\mathrm{Cov}(X_i, X_j) = \begin{cases} 0.5 \cdot 4500, & \text{if } j = i+1, \\ 0, & \text{if } j - i \geq 2. \end{cases}$$

There are $n-1$ pairs of the form $i, i+1$ in the sum above, which gives

$$\mathrm{Var}(X_1 + \cdots + X_n) = 4500n + 4500(n-1) = 9000n - 4500.$$

Using the outline given in Exercise 9.9(c) we get

$$P\left(\left|\frac{S_n}{n} - 5,000\right| \geq 50\right) \leq \frac{\mathrm{Var}(S_n/n)}{50^2} = \frac{9000n - 4500}{n^2 2500}.$$

We need $\frac{9000n-4500}{n^2 2500} < 0.05$ which leads to $n \geq 72$.

9.11. (a) We have

$$M_X'(t) = \tfrac{3}{2} \cdot 2(1-2t)^{-5/2} = 3(1-2t)^{-5/2}.$$

Thus,

$$M_X'(0) = E[X] = 3.$$

We may now use Markov's inequality to conclude that

$$P(X > 8) \leq \frac{E[X]}{8} = \frac{3}{8} = 0.375.$$

(b) In order to use Chebyshev's inequality, we must find the variance of X. So, differentiating again yields

$$M_X''(t) = 15(1-2t)^{-7/2},$$

and so,

$$M''(0) = E[X^2] = 15 \implies \mathrm{Var}(X) = 15 - 9 = 6.$$

Thus, Chebyshev's inequality yields

$$P(X > 8) = P(X - 3 > 5) \leq \frac{\mathrm{Var}(X)}{5^2} = \frac{6}{25} = 0.24.$$

9.12. (a) We have $E[X] = 2$ and $E[Y] = 1/2$ which gives $E[X+Y] = 5/2$. Since $X + Y \geq 0$, we may use Markov's inequality to get

$$P(X + Y > 10) \leq \frac{E[X+Y]}{10} = \frac{5}{20} = \frac{1}{4}.$$

(b) We have $\mathrm{Var}(X) = 2$ and $\mathrm{Var}(Y) = \frac{1}{12}$, and by independence $\mathrm{Var}(X+Y) = \frac{25}{12}$. Using Chebyshev's inequality:

$$\begin{aligned} P(X+Y > 10) &= P(X + Y - \tfrac{5}{2} > 10 - \tfrac{5}{2}) \leq P(|X+Y-\tfrac{5}{2}| > \tfrac{15}{2}) \\ &\leq \frac{\mathrm{Var}(X+Y)}{(\frac{15}{2})^2} = \frac{\frac{25}{12}}{(\frac{15}{2})^2} = \frac{1}{27}. \end{aligned}$$

9.13. We have

$$\begin{aligned} E[X] &= \frac{10}{3}, \qquad \mathrm{Var}(X) = 10 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{20}{9} \\ E[Y] &= \frac{1}{3}, \qquad \mathrm{Var}(Y) = \frac{1}{9}. \end{aligned}$$

From this we get

$$E[X - Y] = \frac{10}{3} - \frac{1}{3} = 3, \quad \mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \frac{20}{9} + \frac{1}{9} = \frac{7}{3}.$$

Now we can apply Chebyshev's inequality:

$$P(X - Y < -1) = P(X - Y - 3 < -4) \leq P(|X - Y - 3| > 4) \leq \frac{\mathrm{Var}(X-Y)}{4^2} = \frac{7}{48}.$$

9.14. To get a meaningful bounds we consider only $t > 2$.

Markov's inequality gives the bound

$$P(X > t) \leq \frac{E[X]}{t} = \frac{2}{t}.$$

Chebyshev's inequality (for $t > 2$) yields

$$P(X > t) = P(X - E[X] > t - 2) \leq P(|X - E[X]| > t - 2) \leq \frac{\mathrm{Var}(X)}{(t-2)^2} = \frac{9}{(t-2)^2}.$$

Solving the inequality $\frac{2}{t} < \frac{9}{(t-2)^2}$ gives $1/2 < t < 8$, and since $t > 2$, this leads to $2 < t < 8$.

9.15. Let X_i and Y_i the number of customers coming to Omar's and Cheryl's truck on the ith day, respectively. We need to estimate $P(\sum_{k=1}^n X_i \geq \sum_{k=1}^n Y_i)$ as n gets larger. This is the same as the probability

$$P(\sum_{k=1}^n (X_i - Y_i) \geq 0) = P\left(\frac{1}{n}\sum_{k=1}^n (X_i - Y_i) \geq 0\right)$$

The random variables $Z_i = X_i - Y_i$ are independent, have mean $E[Z_i] = E[X_i] - E[Y_i] = 10$ and a finite variance. By the law of large numbers the average of these random variables will converge to 10, in particular

$$P\left(\frac{1}{n}\sum_{k=1}^{n}(X_i - Y_i) < 0\right) = P\left(\frac{1}{n}\sum_{k=1}^{n}(Z_i - E[Z_i]) < -10\right)$$

will converge to 0 by Theorem 9.9. But this means that the probability of the complement will converge to 1, in other words $P(\sum_{k=1}^{n} X_i \geq \sum_{k=1}^{n} Y_i)$ converges to 1 as n gets larger and larger.

9.16. Let U_i be the waiting time for number 5 on morning i, and V_i the waiting time for number 8 on morning i. From the problem, $U_i \sim \text{Exp}(\frac{1}{10})$ and $V_i \sim \text{Exp}(\frac{1}{20})$. The actual waiting time on morning i is $X_i = \min(U_i, V_i)$. Let Y_i be the Bernoulli variable that records 1 if I take the number 5 on morning i. Then from properties of exponential variables (from Examples 6.33 and 6.34)

$$X_i \sim \text{Exp}(\tfrac{3}{20}), \quad E(X_i) = \tfrac{20}{3}, \quad E(Y_i) = P(Y_i = 1) = P(U_i < V_i) = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{1}{20}} = \tfrac{2}{3}.$$

Since $S_n = \sum_{i=1}^{n} X_i$ and $T_n = \sum_{i=1}^{n} Y_i$, we can answer the questions by the LLN.

(a)

$$\begin{aligned}\lim_{n\to\infty} P(S_n \leq 7n) &= \lim_{n\to\infty} P(S_n - nE(X_1) \leq \tfrac{1}{3}n) \\ &\geq \lim_{n\to\infty} P(\left|\tfrac{S_n}{n} - E(X_1)\right| \leq \tfrac{1}{3}) = 1.\end{aligned}$$

(b)

$$\begin{aligned}\lim_{n\to\infty} P(T_n \geq 0.6n) &= \lim_{n\to\infty} P(T_n - nE(Y_1) \geq -\tfrac{1}{15}n) \\ &\geq \lim_{n\to\infty} P(\left|\tfrac{T_n}{n} - E(Y_1)\right| \leq \tfrac{1}{15}) = 1.\end{aligned}$$

9.17. (a) Using Markov's inequality we have

$$P(X > 120) \leq \frac{E[X]}{120} = \frac{100}{120} = \frac{5}{6}.$$

(b) Using Chebyshev's inequality we have

$$P(X > 120) = P(X - 100 > 20) \leq \frac{\text{Var}(X)}{20^2} = \frac{100}{400} = \frac{1}{4}.$$

(c) We have that $X = \sum_{i=1}^{100} X_i$ where the X_i are i.i.d. Poisson random variables with a parameter of one (hence, they all have mean 1 and variance 1). Thus,

$$\begin{aligned}P(X > 120) &= P\left(\sum_{i=1}^{100} X_i > 120\right) = P\left(\sum_{i=1}^{100}(X_i - 1) > 20\right) \\ &= P\left(\frac{\sum_{i=1}^{100}(X_i - 1)}{\sqrt{100}} > 2\right) \approx P(Z > 2),\end{aligned}$$

where Z is a standard normal random variable and we have applied the CLT in the last line. Hence,

$$P(X > 120) \approx 1 - \Phi(2) = 1 - 0.9772 = 0.0228.$$

9.18. (a) From Example 8.13 we have $E[X] = 100 \cdot \frac{1}{\frac{1}{3}} = 300$. Hence by Markov's inequality we get

$$P(X > 500) \le \frac{E[X]}{500} = \frac{300}{500} = \frac{3}{5}.$$

(b) Again, from Example 8.13 we have $\mathrm{Var}[X] = 100 \cdot \frac{1-\frac{1}{3}}{(\frac{1}{3})^2} = 600$. Then from Chebyshev's inequality:

$$\begin{aligned} P(X > 500) &= P(X - E[X] > 500 - 300) \\ &\le \frac{\mathrm{Var}(X)}{200^2} = \frac{600}{200^2} = \frac{3}{200} = 0.015. \end{aligned}$$

(c) By the CLT the distribution of the standardized version of X is close to that of a standard normal. The standardized version is $\frac{X-300}{\sqrt{600}}$, hence

$$P(X > 500) = P\left(\tfrac{X-300}{\sqrt{600}} > \tfrac{500-300}{\sqrt{600}}\right) \approx 1 - \Phi(\tfrac{20}{\sqrt{6}}) \approx 1 - \Phi(8.16) < 0.0002.$$

(In fact $1 - \Phi(8.16)$ is way smaller than 0.0002, it is approximately $2.2 \cdot 10^{-16}$.)

(d) We need more than 500 trials for the 100th success exactly if there are at most 99 successes within the first 500 trials. Thus denoting by S the number of successes within the first 500 trials we have $P(X > 500) = P(S \le 99)$. Since $S \sim \mathrm{Bin}(500, \frac{1}{3})$, we may use normal approximation to get

$$P(S \le 99) = P\left(\frac{S - \frac{500}{3}}{\sqrt{500 \cdot \frac{2}{9}}} \le \frac{99 - \frac{500}{3}}{\sqrt{500 \cdot \frac{2}{9}}}\right) \approx \Phi\left(\frac{99 - \frac{500}{3}}{\sqrt{500 \cdot \frac{2}{9}}}\right) \approx \Phi(-6.42) < 0.002.$$

(Again, the real value of $\Phi(-6.42)$ is a lot smaller than 0.0002, it is approximately $6.8 \cdot 10^{-11}$.)

9.19. Let X_i be the amount of time it takes the child to spin around on his ith revolution. Then the total time it will take to spin around 100 times is

$$S_{100} = X_1 + \cdots + X_{100}.$$

We assume that the X_i are independent with mean 1/2 and standard deviation 1/3. Then $E[S_{100}] = 50$ and $\mathrm{Var}(S_{100}) = \frac{100}{3^2}$. Using Chebyshev's inequality:

$$P(X_1 + \cdots + X_{100} > 55) = P(X_1 + \cdots + X_{100} - 50 > 5) \le \frac{\mathrm{Var}(S_{100})}{5^2} = \frac{100}{9 \cdot 25} = \frac{4}{9}.$$

If we use the CLT then

$$\begin{aligned} P(X_1 + \cdots + X_{100} > 55) &= P\left(\frac{X_1 + \cdots + X_{100} - 50}{\sqrt{100} \cdot (1/3)} > \frac{55 - 50}{\sqrt{100} \cdot (1/3)}\right) \\ &\approx P(Z > \frac{5}{10 \cdot (1/3)} \\ &= P(Z > 1.5) = 1 - P(Z \le 1.5) \\ &= 1 - 0.9332 = 0.0668. \end{aligned}$$

9.20. (a) We can use the law of large numbers:

$$\lim_{n\to\infty} P(S_n \geq 0.01n) = \lim_{n\to\infty} P(S_n - nE[X_1] \geq 0.01n) \leq \lim_{n\to\infty} P(|\tfrac{S_n}{n} - E[X_1]| \geq 0.01) = 0.$$

Hence the limit is 0.

(b) Here the central limit theorem will be helpful:

$$\lim_{n\to\infty} P(S_n \geq 0) = \lim_{n\to\infty} P(\tfrac{S_n - nE[X_1]}{\sqrt{n\operatorname{Var}(X_1)}} \geq 0) = 1 - \Phi(0) = \frac{1}{2}.$$

The limit is $\frac{1}{2}$.

(c) We can use the law of large numbers:

$$\lim_{n\to\infty} P(S_n \geq -0.01n) = \lim_{n\to\infty} P(S_n - nE[X_1] \geq -0.01n) \geq \lim_{n\to\infty} P(|\tfrac{S_n}{n} - E[X_1]| \leq 0.01) = 1.$$

Hence the limit is 1.

9.21. Let $Z_i = X_i - Y_i$. Then

$$E[Z_i] = E[X_i] - E[Y_i] = 2 - 2 = 0, \quad \operatorname{Var}(Z_i) = \operatorname{Var}(X_i - Y_i) = \operatorname{Var}(X_i) + \operatorname{Var}(Y_i) = 3 + 2 = 5.$$

We have

$$P\left(\sum_{i=1}^{500} X_i > \sum_{i=1}^{500} Y_i + 50\right) = P\left(\sum_{i=1}^{500} Z_i > 50\right).$$

Applying the central limit theorem we get

$$\begin{aligned} P\left(\sum_{i=1}^{500} Z_i > 50\right) &= P\left(\frac{\sum_{i=1}^{500} Z_i}{\sqrt{500\cdot 5}} > \frac{50}{\sqrt{500\cdot 5}}\right) \\ &\approx 1 - \Phi\left(\frac{50}{\sqrt{500\cdot 5}}\right) = 1 - \Phi(1) \\ &\approx 1 - 0.8413 = 0.1587. \end{aligned}$$

9.22. If we can generate a Unif$[0,1]$ distributed random variable, then by Example 5.19 we can also generate an Exp(1) random variable by plugging it into $\ln(1-x)$. Then we can produce a sample of $n = 10^5$ independent copies of the Y random variable given in the exercise. If $\bar{\mu}$ is the sample mean and s_n^2 is the sample variance from this sample then the 95% confidence interval for the integral is $(\bar{\mu} - \frac{1.96\cdot s_n}{\sqrt{n}}, \bar{\mu} + \frac{1.96\cdot s_n}{\sqrt{n}})$.

Solutions to Chapter 10

10.1. (a) By summing the probabilities in the appropriate columns we get the marginal probability mass function of Y:

$$p_Y(0) = \tfrac{1}{3}, \quad p_Y(1) = \tfrac{4}{9}, \quad p_Y(2) = \tfrac{2}{9}.$$

We can now compute the conditional probability mass function $p_{X|Y}(x|y)$ for $y = 0, 1, 2$ using the formula $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$. We get

$$\begin{aligned}
&p_{X|Y}(2|0) = 1, \\
&p_{X|Y}(1|1) = \tfrac{1}{4}, \qquad p_{X|Y}(2|1) = \tfrac{1}{2}, \qquad p_{X|Y}(3|1) = \tfrac{1}{4}, \\
&p_{X|Y}(2|2) = \tfrac{1}{2}, \qquad p_{X|Y}(3|2) = \tfrac{1}{2}
\end{aligned}$$

(b) The conditional expectations can be computed using the conditional probability mass functions:

$$\begin{aligned}
E[X|Y = 0] &= 2p_{X|Y}(2|0) = 2 \\
E[X|Y = 1] &= 1p_{X|Y}(1|1) + 2p_{X|Y}(2|1) + 3p_{X|Y}(3|1) = \tfrac{1}{4} + 2 \cdot \tfrac{1}{2} + 3 \cdot \tfrac{1}{4} = 2 \\
E[X|Y = 2] &= 2p_{X|Y}(2|2) + 3p_{X|Y}(3|2) = 2 \cdot \tfrac{1}{2} + 3 \cdot \tfrac{1}{2} = \tfrac{5}{2}.
\end{aligned}$$

10.2. (i) Given $X = 1$, Y is uniformly distributed. This implies $p_{X,Y}(1,1) = \frac{1}{8}$.

(ii) $p_{X|Y}(0|0) = \frac{2}{3}$. This implies that

$$\tfrac{2}{3} = \frac{p_{X,Y}(0,0)}{p_Y(0)} = \frac{p_{X,Y}(0,0)}{p_{X,Y}(0,0) + p_{X,Y}(1,0)} = \frac{p_{X,Y}(0,0)}{p_{X,Y}(0,0) + \frac{1}{8}}$$

which implies $p_{X,Y}(0,0) = \frac{2}{8}$.

(iii) $E(Y|X=0)=\frac{4}{5}$. This implies

$$\begin{aligned}\tfrac{4}{5} &= 0\cdot p_{Y|X}(0|0)+1\cdot p_{Y|X}(1|0)+2\cdot p_{Y|X}(2|0)\\ &= \frac{p_{X,Y}(0,1)+2p_{X,Y}(0,2)}{p_X(0)}=\frac{p_{X,Y}(0,1)+2p_{X,Y}(0,2)}{p_{X,Y}(0,0)+p_{X,Y}(0,1)+p_{X,Y}(0,2)}\\ &= \frac{p_{X,Y}(0,1)+2(\frac{3}{8}-p_{X,Y}(0,1))}{\frac{2}{8}+p_{X,Y}(0,1)+\frac{3}{8}-p_{X,Y}(0,1)}.\end{aligned}$$

With the previously known values of the table, the fact that probabilities sum to 1 gives $\frac{5}{8}+p_{X,Y}(0,1)+p_{X,Y}(0,2)=1$ and we can replace $p_{X,Y}(0,2)$ with $\frac{3}{8}-p_{X,Y}(0,1)$. From the equation above we deduce $p_{X,Y}(0,1)=\frac{2}{8}$ and then $p_{X,Y}(0,2)=\frac{1}{8}$.

The final table is

$X \backslash Y$	0	1	2
0	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

10.3. Given $Y=y$, the random variable X is binomial with parameters y and $1/2$. Hence, for x between 0 and 6, we have

$$p_X(x)=\sum_{y=1}^{6}p_{X|Y}(x|y)p_Y(y)=\sum_{y=1}^{6}\binom{y}{x}\frac{1}{2^y}\cdot\frac{1}{6},$$

where $\binom{y}{x}=0$ if $y<x$ (as usual).

For the expectation, we have

$$E[X]=\sum_{y=1}^{6}E[X|Y=y]p_Y(y)=\sum_{y=1}^{6}\frac{y}{2}\cdot\frac{1}{6}=\frac{7}{4}.$$

10.4. (a) Directly from the description of the problem we get that

$$p_{X|N}(k|n)=\binom{n}{k}(\tfrac{1}{2})^n \qquad \text{for } 0\le k\le n\le 100.$$

(b) From knowing the mean of the binomial, $E[X|N=n]=n/2$ for $0\le n\le 100$.

(c)

$$E[X]=\sum_{n=0}^{100}E[X|N=n]\,p_N(n)=\tfrac{1}{2}\sum_{n=0}^{100}n\,p_N(n)=\tfrac{1}{2}E[N]=\tfrac{1}{2}\cdot 100\cdot\tfrac{1}{4}=\tfrac{25}{2}.$$

Above we used the fact that N is binomial.

10.5. (a) The conditional probability density function is given by the formula:

$$f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)},$$

if $f_Y(y) > 0$. Since the joint density is only nonzero for $0 < y < 1$, the Y variable will have a density which is only nonzero in $0 < y < 1$. In that case we have

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(w,y)dw = \int_0^1 \frac{12}{5}w(2-w-y)dw \\
&= \frac{12}{5}(w^2 - \frac{1}{3}w^3 - \frac{1}{2}yw^2)\big|_0^1 = \frac{12}{5}(1-\frac{1}{3}-\frac{1}{2}y) = \frac{8}{5}-\frac{6}{5}y
\end{aligned}
$$

Thus, for $0 < y < 1$ we have

$$f_{X|Y}(x|y) = \frac{\frac{12}{5}x(2-x-y)}{\frac{8}{5}-\frac{6}{5}y} = \frac{6x(2-x-y)}{4-3y}.$$

(b) We have

$$
\begin{aligned}
P(X > \frac{1}{2}|Y = \frac{3}{4}) &= \int_{\frac{1}{2}}^1 f_{X|Y}(x|y=\frac{3}{4})dx = \int_{\frac{1}{2}}^1 \frac{6x(2-x-\frac{3}{4})}{4-\frac{9}{4}}dx \\
&= \frac{24}{7}\int_{\frac{1}{2}}^1 x(\frac{5}{4}-x)dx = \frac{24}{7}(\frac{5}{8}x^2 - \frac{1}{3}x^3)\big|_{\frac{1}{2}}^1 = \frac{24}{7}(\frac{5}{8}-\frac{1}{3}-\frac{5}{32}+\frac{1}{24}) \\
&= \frac{24}{7}(\frac{7}{24}-\frac{11}{96}) = \frac{24}{7}\frac{17}{96} = \frac{17}{28}.
\end{aligned}
$$

$$
\begin{aligned}
E[X|Y=\frac{3}{4}] &= \int_0^1 x\frac{6x(\frac{5}{4}-x)}{\frac{7}{4}}dx = \frac{24}{7}\int_0^1 x^2(\frac{5}{4}-x)dx = \frac{24}{7}(\frac{5}{12}x^3 - \frac{1}{4}x^4)\big|_0^1 \\
&= \frac{24}{7}\frac{1}{6} = \frac{4}{7}.
\end{aligned}
$$

10.6. (a) Begin by finding the marginal density function of Y. For $0 < y < 2$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \tfrac{1}{4}\int_0^y (x+y)\,dx = \tfrac{3}{8}y^2.$$

Then for $0 < x < y < 2$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{3}\cdot\frac{x+y}{y^2}.$$

(b) For $y = 1$ the conditional density function of X is

$$f_{X|Y}(x|1) = \tfrac{2}{3}(x+1) \quad \text{for } 0 < x < 1 \text{ and zero otherwise.}$$

We compute the conditional probabilities with the conditional density function:

$$P(X < \tfrac{1}{2}\,|\,Y=1) = \int_{-\infty}^{1/2} f_{X|Y}(x|1)dx = \tfrac{2}{3}\int_0^{1/2}(x+1)\,dx = \tfrac{5}{12}$$

and

$$P(X < \tfrac{3}{2}\,|\,Y=1) = \int_{-\infty}^{3/2} f_{X|Y}(x|1)dx = \tfrac{2}{3}\int_0^{1}(x+1)\,dx = 1.$$

Note that integrating all the way to $3/2$ would be wrong in the last integral above because conditioning on $Y = 1$ restricts X to $0 < X < 1$.

(c) The conditional expectation: for $0 < y < 2$,

$$E[X^2|Y=y] = \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y)\,dy = \tfrac{2}{3}\int_0^y x^2 \cdot \frac{x+y}{y^2}\,dx = \tfrac{7}{18}y^2.$$

For $0 < x < 2$, the marginal density function of X can be obtained either from

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \tfrac{1}{4}\int_x^2 (x+y)\,dy = \tfrac{1}{2} + \tfrac{1}{2}x - \tfrac{3}{8}x^2,$$

or equivalently from

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\,dy = \tfrac{1}{4}\int_x^2 (x+y)\,dy = \tfrac{1}{2} + \tfrac{1}{2}x - \tfrac{3}{8}x^2.$$

With the marginal density function we calculate $E[X^2]$:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^2 x^2\left(\tfrac{1}{2} + \tfrac{1}{2}x - \tfrac{3}{8}x^2\right)dx = \tfrac{14}{15}.$$

We can get the same answer by averaging the conditional expectation:

$$\begin{aligned}\int_{-\infty}^{\infty} E[X^2|Y=y]\,f_Y(y)\,dy &= \tfrac{7}{18}\int_{-\infty}^{\infty} y^2 f_Y(y)dy = \tfrac{7}{18}E[Y^2] \\ &= \tfrac{7}{18}\int_0^2 y^2 \cdot \tfrac{3}{8}y^2\,dy = \tfrac{14}{15}.\end{aligned}$$

10.7. (a) Directly by multiplying, $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = 6x$ for $0 < x < y < 1$.

(b)

$$f_X(x) = \int_x^1 \frac{2x}{y^2}\cdot 3y^2\,dy = 6x(1-x), \qquad 0 < x < 1.$$

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

Thus given $X = x$, Y is uniform on the interval $(x, 1)$. Valid for $0 < x < 1$.

10.8. (a) From the description of the problem,

$$p_{Y|X}(m|\ell) = \binom{\ell}{m}(\tfrac{4}{9})^m(\tfrac{5}{9})^{\ell-m} \qquad \text{for } 0 \le m \le \ell.$$

From knowing the mean of a binomial, $E[Y|X=\ell] = \frac{4}{9}\ell$. Thus $E[Y|X] = \frac{4}{9}X$.

(b) $X \sim \text{Geom}(\frac{1}{6})$, and so $E(X) = 6$. For the mean of Y,

$$E[Y] = E[E(Y|X)] = \tfrac{4}{9}E[X] = \tfrac{4}{9}\cdot 6 = \tfrac{8}{3}.$$

10.9. (a) We have

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\infty} \frac{1}{y}e^{-x/y}e^{-y}dx = e^{-y}$$

if $0 < y$ and zero otherwise. We can evaluate the last integral without computation if we recognize that $\frac{1}{y}e^{-x/y}$ is the probability density function of an $\text{Exp}(1/y)$ distribution and hence its integral on $[0, \infty)$ is equal to 1.

From the found probability density $f_Y(y)$ we see that $Y \sim \text{Exp}(1)$ and hence $E[Y] = 1$. We also get

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{y} e^{-x/y} \quad \text{if } 0 < x, 0 < y,$$

and zero otherwise.
(b) The conditional probability density function $f_{X|Y}(x|y)$ found in part (a) shows that given $Y = y > 0$ the conditional distribution of X is $\text{Exp}(1/y)$. Hence $E[X|Y = y] = \frac{1}{\frac{1}{y}} = y$ and $E[X|Y] = Y$.
(c) We can compute $E[X]$ by conditioning on Y and then averaging the conditional expectation:

$$E[X] = E[E[X|Y]] = E[Y] = 1,$$

where in the last step we used part (a).

10.10. (a)

$$p_{X|N}(k\,|\,n) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } 0 \le k \le n.$$

From knowing the expectation of a binomial, $E(X\,|\,N = n) = np$ and then $E(X\,|\,N) = pN$.

(b) $E[X] = E[E(X|N)] = pE[N] = p\lambda$.

(c) We use formula (10.36) to compute the expectation of the product:

$$E[NX] = E[E(NX|N)] = E[N\,E(X|N)] = E[N \cdot pN] = pE[N^2] = p(\lambda^2 + \lambda).$$

In the last step we used $E[N] = \text{Var}[N] = \lambda$ and $E[N^2] = (E[N])^2 + \text{Var}[N]$. The calculation above can be done without formula (10.36) also, by manipulating the sums involved:

$$\begin{aligned} E[XN] &= \sum_{k,n} kn\, p_{X,N}(k,n) = \sum_{k,n} kn\, p_{X|N}(k\,|\,n)\, p_N(n) \\ &= \sum_n n\, p_N(n) \sum_k k\, p_{X|N}(k\,|\,n) = \sum_n n\, p_N(n)\, E(X\,|\,N = n) \\ &= p \sum_n n^2\, p_N(n) = pE[N^2] = p(\lambda^2 + \lambda). \end{aligned}$$

Now for the covariance:

$$\text{Cov}(N, X) = E[NX] - EN \cdot EX = p(\lambda^2 + \lambda) - \lambda \cdot p\lambda = p\lambda.$$

10.11. The expected value of a $\text{Poisson}(y)$ random variable is y, and the second moment is $y + y^2$. Thus

$$E[X|Y = y] = y, \qquad E[X^2|Y = y] = y^2 + y,$$

and $E[X|Y] = Y$, $E[X^2|Y] = Y^2 + Y$. Now taking expectations and using the the moments of the exponential distribution gives

$$E[X] = E[E[X|Y]] = E[Y] = \frac{1}{\lambda}$$

and

$$E[E[X^2|Y]] = E[Y^2 + Y] = \frac{2}{\lambda^2} + \frac{1}{\lambda}.$$

This gives

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} + \frac{1}{\lambda} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} + \frac{1}{\lambda}$$

10.12. (a) This question is for Wald's identity.

$$E[S_N] = E[N] \cdot E[X_1] = p^{-1} \cdot \lambda^{-1} = \frac{1}{p\lambda}.$$

(b) We derive the moment generating function of S_N by conditioning on N. Let $t \in \mathbb{R}$. First the conditional moment generating function. As in equation (10.35) and in the proof of Wald's identity, conditioning on $N = n$ turns S_N into S_n. Then we use independence and identical distribution of the terms X_i.

$$\begin{aligned} E[e^{tS_N}|N = n] &= E[e^{tS_n}] = E\Big[\prod_{i=1}^{n} e^{tX_i}\Big] = \prod_{i=1}^{n} E[e^{tX_i}] \\ &= \begin{cases} \infty & \text{if } t \ge \lambda, \\ \left(\dfrac{\lambda}{\lambda - t}\right)^n & \text{if } t < \lambda. \end{cases} \end{aligned}$$

Above we took the moment generating function of the exponential distribution from Example 5.6.

Next, for $t < \lambda$, we take expectations over the conditioning variable N:

$$\begin{aligned} E[e^{tS_N}] &= E[E(e^{tS_N}|N)] = \sum_{n=1}^{\infty} E[e^{tS_N}|N = n]\, p_N(n) \\ &= \sum_{n=1}^{\infty} \left(\frac{\lambda}{\lambda - t}\right)^n (1-p)^{n-1} p = \frac{p\lambda}{\lambda - t} \sum_{n=1}^{\infty} \left(\frac{\lambda(1-p)}{\lambda - t}\right)^{n-1} \\ &= \frac{\frac{p\lambda}{\lambda - t}}{1 - \frac{\lambda(1-p)}{\lambda - t}} = \frac{p\lambda}{p\lambda - t}. \end{aligned}$$

With $t < \lambda$ the geometric series above converges if and only if

$$\frac{\lambda(1-p)}{\lambda - t} < 1 \quad \text{if and only if} \quad t < p\lambda.$$

The outcome of the calculation is

$$E[e^{tS_N}] = \begin{cases} \infty & \text{if } t \ge p\lambda, \\ \dfrac{p\lambda}{p\lambda - t} & \text{if } t < p\lambda. \end{cases}$$

Comparison with Example 5.6 shows that $S_N \sim \mathrm{Exp}(p\lambda)$.

This problem can be solved without calculation by appeal to the properties of the Poisson process in Section 7.3 and Example 10.14. Namely, start with a Poisson process of rate λ of customers that arrive at my store. By Fact 7.26 the interarrival times of the customers are i.i.d. $\mathrm{Exp}(\lambda)$ random variables that we call X_1, X_2, X_3, etc. Suppose each customer independently buys something with probability p. Then the first customer who buys something is the Nth customer for a $\mathrm{Geom}(p)$ random variable N. This customer's arrival time is S_N.

On the other hand, according to the thinning property of Example 10.14, the process of arrival times of buying customers is a Poisson process of rate $p\lambda$. Hence again by Fact 7.26 the time of arrival of the first buying customer has $\text{Exp}(p\lambda)$ distribution. Thus we conclude that $S_N \sim \text{Exp}(p\lambda)$. From this, $E[S_N] = 1/(p\lambda)$.

10.13. The price should be the expected value of X. The expectation of a Poisson(λ) distributed random variable is λ, hence we have $E[X|U = u] = u$ and $E[X|U] = U$. Taking expectations again:

$$E[X] = E[E[X|U]] = E[U] = 5$$

since $U \sim \text{Unif}[0, 10]$.

10.14. Given the vector $(t_1, \dots, t_n)$ of zeroes and ones, let m be the number of ones among $t_1, \dots, t_n$. Permutation does not alter the number of ones in the vector and so m is also the number of ones among $t_{k_1}, \dots, t_{k_n}$. Consequently

$$\begin{aligned} &P(X_1 = t_1, X_2 = t_2, \dots, X_n = t_n) \\ &\qquad = \int_0^1 P(X_1 = t_1, X_2 = t_2, \dots, X_n = t_n \,|\, \xi = p)\, dp \\ &\qquad = \int_0^1 p^m (1-p)^{n-m}\, dp \end{aligned}$$

and similarly

$$\begin{aligned} &P(X_1 = t_{k_1}, X_2 = t_{k_2}, \dots, X_n = t_{k_n}) \\ &\qquad = \int_0^1 P(X_1 = t_{k_1}, X_2 = t_{k_2}, \dots, X_n = t_{k_n} \,|\, \xi = p)\, dp \\ &\qquad = \int_0^1 p^m (1-p)^{n-m}\, dp. \end{aligned}$$

The two probabilities agree.

10.15. (a) This is very similar to Example 10.13 and can be solved similarly. Let N be the number of claims in one day. We know that $N \sim \text{Poisson}(12)$. Let N_A be the number of claims from A policies in one day, and N_B be the number of claims from B policies in one day. We assume that each claim comes independently from policy A or policy B. Hence, given $N = n$, N_A is distributed as a binomial random variable with parameters n and $1/4$. Therefore, for any nonnegative k,

$$\begin{aligned} P(N_A = k) &= \sum_{n=0}^{\infty} P(N_A = k|N = n) P(N = n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} e^{-12} \frac{12^n}{n!} \\ &= \frac{1}{k!} \left(\frac{1}{4}\right)^k 12^k e^{-12} \sum_{n=k}^{\infty} \frac{1}{(n-k)!} \left(\frac{3}{4} \cdot 12\right)^{n-k} \\ &= \frac{1}{k!} 3^k e^{-12} \sum_{j=0}^{\infty} \frac{9^j}{j!} = \frac{1}{k!} 3^k e^{-12} e^9 = e^{-3} \frac{3^k}{k!}. \end{aligned}$$

Hence, $N_A \sim \text{Poisson}(3)$, and we can use this to calculate $P(N_A \geq 5)$:

$$P(N_A \geq 5) = 1 - \sum_{k=0}^{4} P(N_A = k) = 1 - \sum_{k=0}^{4} e^{-3}\frac{3^k}{k!} \approx 0.1847.$$

(b) As in part (a), we can show that $N_B \sim \text{Poisson}(9)$, which gives

$$P(N_B \geq 5) = 1 - \sum_{k=0}^{4} P(N_B = k) = 1 - \sum_{k=0}^{4} e^{-9}\frac{9^k}{k!} \approx 0.9450.$$

(c) Since $N \sim \text{Poisson}(12)$, we have

$$P(N \geq 10) = 1 - \sum_{k=0}^{9} P(N = k) = 1 - \sum_{k=0}^{9} e^{-12}\frac{12^k}{k!} \approx 0.7576.$$

10.16. There are several ways to approach this problem. We begin with an approach of direct calculation. The total number of claims is $N \sim \text{Poisson}(12)$. Consider any particular claim. Let A be the event that this claim is from policy A, B the event that this claim is from policy B, and C the event that this claim is greater than \$100,000. By the law of total probability

$$P(C) = P(C|A)P(A) + P(C|B)P(B) = \tfrac{4}{5}\cdot\tfrac{1}{4} + \tfrac{1}{5}\cdot\tfrac{3}{4} = \tfrac{7}{20}.$$

Let X denote the number of claims that are greater than \$100,000. We must assume that each claim is greater than \$100,000 independently of the other claims. It follows then that given $N = n$, X is conditionally $\text{Bin}(n, \frac{7}{20})$. We can deduce the p.m.f. of X. For $k \geq 0$,

$$\begin{aligned}
P(X = k) &= \sum_{n=k}^{\infty} P(X = k|N = n)P(N = n) = \sum_{n=k}^{\infty} \binom{n}{k} (\tfrac{7}{20})^k (\tfrac{13}{20})^{n-k} e^{-12}\frac{12^n}{n!} \\
&= (\tfrac{7}{20})^k e^{-12}\frac{12^k}{k!} \sum_{n=k}^{\infty} \frac{(\frac{13}{20})^{n-k} 12^{n-k}}{(n-k)!} = (\tfrac{21}{5})^k e^{-12}\frac{1}{k!}\sum_{j=0}^{\infty}\frac{(\frac{39}{5})^j}{j!} \\
&= (\tfrac{21}{5})^k e^{-12}\frac{1}{k!} e^{\frac{39}{5}} = e^{-\frac{21}{5}}\frac{(\frac{21}{5})^k}{k!}.
\end{aligned}$$

We found that $X \sim \text{Poisson}(\frac{21}{5})$. From this we answer the questions.

(a) $E[X] = \frac{21}{5}$.

(b) $P(X \leq 2) = e^{-\frac{21}{5}}(1 + \frac{21}{5} + \frac{1}{2}(\frac{21}{5})^2) = e^{-\frac{21}{5}}\frac{701}{50} \approx 0.21$.

We can arrive at the distribution of X also without calculation, and then solve the problem as above. From the solution to Exercise 10.15, $N_A \sim \text{Poisson}(3)$ and $N_B \sim \text{Poisson}(9)$. These two variables are independent by the same kind of calculation that was done in Example 10.13. Let X_A be the number of claims from policy A that are greater than \$100,000 and let X_B be the number of claims from policy B that are greater than \$100,000. The situation is exactly as in Problem 10.15 and in Example 10.13, and we conclude that X_A and X_B are independent with distributions $N_A \sim \text{Poisson}(\frac{12}{5})$ and $N_B \sim \text{Poisson}(\frac{9}{5})$. Consequently $X = X_A + X_B \sim \text{Poisson}(\frac{21}{5})$.

10.17. (a) Let B be the event that the coin lands on heads. Then the conditional distribution of X given B is binomial with parameters 3 and $\frac{1}{6}$, while the conditional distribution of X given B^c is $\text{Bin}(5, \frac{1}{6})$. From this we can write down the conditional probability mass functions, and using (10.5) the unconditional one:

$$\begin{aligned} P(X = k) &= P(X = k|B)P(B) + P(X = k|B^c)P(B^c) \\ &= \binom{3}{k}\left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{3-k} \cdot \frac{1}{2} + \binom{5}{k}\left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{5-k} \cdot \frac{1}{2}. \end{aligned}$$

The set of possible values of X are $\{0, 1, \dots, 5\}$, and the formula makes sense for all k if we define $\binom{a}{b}$ as 0 if $b > a$.

(b) We could use the probability mass function to compute the expectation of X, but it is much easier to use the conditional expectations. Because the conditional distributions are binomial, the conditional expectation of X given B is $E[X|B] = 3 \cdot \frac{1}{6} = \frac{1}{2}$ and the conditional expectation of X given B^c is $E[X|B^c] = 5 \cdot \frac{1}{6} = \frac{5}{6}$. Thus,

$$E[X] = E[X|B]P(B) + E[X|B^c]P(B^c) = \tfrac{1}{2} \cdot \tfrac{1}{2} + \tfrac{5}{6} \cdot \tfrac{1}{2} = \tfrac{2}{3}.$$

10.18. Let N be the number of trials needed for seeing the first outcome s, and Y the number of outcomes t in the first $N - 1$ trials.

(a) For the equally likely outcomes case $P(N = n) = (\frac{r-1}{r})^{n-1}\frac{1}{r}$ for $n \geq 1$. The joint distribution is, for $0 \leq m < n$,

$$\begin{aligned} P(Y = m, N = n) &= P(m \text{ outcomes } t \text{ and no outcomes } s \\ &\qquad\qquad \text{in the first } n-1 \text{ trials, outcome } s \text{ in trial } n) \\ &= \binom{n-1}{m}\left(\tfrac{1}{r}\right)^m \left(\tfrac{r-2}{r}\right)^{n-1-m} \cdot \tfrac{1}{r}. \end{aligned}$$

The conditional probability mass function of Y given $N = n$ is therefore

$$\begin{aligned} p_{Y|N}(m\,|\,n) &= \frac{P(Y = m, N = n)}{P(N = n)} = \frac{\binom{n-1}{m}\left(\frac{1}{r}\right)^m \left(\frac{r-2}{r}\right)^{n-1-m} \cdot \frac{1}{r}}{\left(\frac{r-1}{r}\right)^{n-1}\frac{1}{r}} \\ &= \binom{n-1}{m}\left(\tfrac{1}{r-1}\right)^m \left(\tfrac{r-2}{r-1}\right)^{n-1-m}, \qquad 0 \leq m \leq n-1. \end{aligned}$$

Thus given $N = n$, the conditional distribution of Y is $\text{Bin}(n-1, \frac{1}{r-1})$. From knowing the mean of a binomial,

$$E[Y\,|\,N = n] = \tfrac{n-1}{r-1}.$$

Hence $E(Y\,|\,N) = \frac{N-1}{r-1}$ and then

$$E[Y] = E[E[Y\,|\,N]] = E[\tfrac{N-1}{r-1}] = \tfrac{1}{r-1}(E[N] - 1) = \tfrac{1}{r-1}(r-1) = 1.$$

(b) In this case $P(N=n)=(1-p_s)^{n-1}p_s$ for $n \geq 1$. The joint distribution is, for $0 \leq m < n$,

$$\begin{aligned}P(Y=m, N=n) &= P(m \text{ outcomes } t \text{ and no outcomes } s \\ &\qquad\qquad \text{in the first } n-1 \text{ trials, outcome } s \text{ in trial } n) \\ &= \binom{n-1}{m} p_t^m (1-p_s-p_t)^{n-1-m} p_s.\end{aligned}$$

The conditional probability mass function of Y given $N=n$ is therefore

$$\begin{aligned}p_{Y|N}(m \,|\, n) &= \frac{P(Y=m, N=n)}{P(N=n)} = \frac{\binom{n-1}{m} p_t^m (1-p_s-p_t)^{n-1-m} p_s}{(1-p_s)^{n-1} p_s} \\ &= \binom{n-1}{m} \left(\tfrac{p_t}{1-p_s}\right)^m \left(1-\tfrac{p_t}{1-p_s}\right)^{n-1-m}, \qquad 0 \leq m \leq n-1.\end{aligned}$$

Thus given $N=n$, the conditional distribution of Y is $\text{Bin}(n-1, \frac{p_t}{1-p_s})$. From knowing the mean of a binomial,

$$E[Y|\, N=n] = \frac{p_t(n-1)}{1-p_s}.$$

Hence $E(Y \,|\, N) = \frac{p_t(N-1)}{1-p_s}$ and then

$$\begin{aligned}E[Y] = E[E[Y \,|\, N]] &= E\left[\frac{p_t(N-1)}{1-p_s}\right] = \frac{p_t(E[N]-1)}{1-p_s} \\ &= \frac{p_t(p_s^{-1}-1)}{1-p_s} = \frac{p_t}{p_s}.\end{aligned}$$

10.19. (a) We know that $X_1 \sim \text{Bin}(n, p_1)$ and $(X_1, X_2, X_3) \sim \text{Mult}(n, 3, p_1, p_2, p_3)$. Using the probability mass function of X_1 and the joint probability mass function of (X_1, X_2, X_3) we get that if $k+\ell+m=n$ and $0 \leq k, \ell, m$ then

$$\begin{aligned}P(X_2=k, X_3=\ell \,|\, X_1=m) &= \frac{P(X_2=k, X_3=\ell \,|\, X_1=m)}{P(X_1=m)} \\ &= \frac{\binom{n}{k,l,m} p_1^m p_2^k p_3^\ell}{\binom{n}{m} p_1^m (p_2+p_3)^{n-m}} = \frac{\frac{n!}{k!\ell!m!}}{\frac{n!}{(n-m)!m!}} \frac{p_1^m p_2^k p_3^\ell}{p_1^m (p_2+p_3)^{n-m}} \\ &= \frac{(n-m)!}{k!\ell!} \frac{p_2^k}{(p_2+p_3)^k} \frac{p_3^\ell}{(p_2+p_3)^\ell} = \binom{k+\ell}{k} \left(\tfrac{p_2}{p_2+p_3}\right)^k \left(1-\tfrac{p_2}{p_2+p_3}\right)^\ell.\end{aligned}$$

(b) The conditional probability mass function found in (a) is binomial with parameters $k+\ell=n-m$ and $\frac{p_2}{p_2+p_3}$. Thus conditioned upon $X_1=m$, the distribution of X_2 is $\text{Bin}(n-m, \frac{p_2}{p_2+p_3})$.

10.20. (a) Let $n \geq 1$ and $0 \leq k \leq n$ so that $P(S_n=k) > 0$ and conditioning on the event $\{S_n=k\}$ is sensible. By the definition of conditional probability,

$$\begin{aligned}&P(X_1=a_1, X_2=a_2, \ldots, X_n=a_n \,|\, S_n=k) \\ &\qquad = \frac{P(X_1=a_1, X_2=a_2, \ldots, X_n=a_n, S_n=k)}{P(S_n=k)}.\end{aligned}$$

Unless the vector $(a_1, \ldots, a_n)$ has exactly k ones, the numerator above equals zero. Hence assume that $(a_1, \ldots, a_n)$ has exactly k ones. Then the condition

$S_n = k$ is superfluous in the numerator and can be dropped. The ratio above equals

$$\frac{P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n)}{P(S_n = k)} = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}.$$

Summarize this as a formula: for $0 \le k \le n$,

$$P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n \,|\, S_n = k) = \begin{cases} \dfrac{1}{\binom{n}{k}} & \text{if } \sum_{i=1}^n a_i = k \\ 0 & \text{otherwise.} \end{cases}$$

(b) The equation above shows that the conditional probability $P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n \,|\, S_n = k)$ depends only on the number of ones in the vector $(a_1, \ldots, a_n)$. A permutation of $(a_1, \ldots, a_n)$ does not change the number of ones. Hence for any permutation $(a_{\ell_1}, \ldots, a_{\ell_n})$ of $(a_1, \ldots, a_n)$,

$$\begin{aligned} &P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n \,|\, S_n = k) \\ &\quad = P(X_1 = a_{\ell_1}, X_2 = a_{\ell_2}, \ldots, X_n = a_{\ell_n} \,|\, S_n = k). \end{aligned}$$

This shows that, given $S_n = k$, $X_1, \ldots, X_n$ are exchangeable.

We show that independence fails for any $n \ge 2$ and $0 < k < n$. First deduce for a fixed index $j \in \{1, \ldots, n\}$ that

$$\begin{aligned} P(X_j = 1 \,|\, S_n = k) &= \frac{P(X_j = 1, S_n = k)}{P(S_n = k)} \\ &= \frac{P(X_j = 1, \text{exactly } k-1 \text{ successes among } X_i \text{ for } i \ne j)}{P(S_n = k)} \\ &= \frac{p \cdot \binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{k}{n}. \end{aligned}$$

Thus

$$P(X_1 = 1 \,|\, S_n = k) \cdot P(X_2 = 0 \,|\, S_n = k) = \frac{k(n-k)}{n^2}.$$

To complete the proof that independence fails we show that the product above does not agree with $P(X_1 = 1, X_2 = 0 \,|\, S_n = k)$, as long as $0 < k < n$.

$$\begin{aligned} P(X_1 = 1, X_2 = 0 \,|\, S_n = k) &= \frac{P(X_1 = 1, X_2 = 0, S_n = k)}{P(S_n = k)} \\ &= \frac{P(X_1 = 1, X_2 = 0, \text{exactly } k-1 \text{ successes among } X_i \text{ for } i \ge 3)}{P(S_n = k)} \\ &= \frac{p(1-p) \cdot \binom{n-2}{k-1}p^{k-1}(1-p)^{n-k-1}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{k(n-k)}{n(n-1)}. \end{aligned}$$

The condition $0 < k < n$ guarantees that the numerators of $\frac{k(n-k)}{n^2}$ and $\frac{k(n-k)}{n(n-1)}$ agree and do not vanish. Hence the disagreement of the denominators forces $\frac{k(n-k)}{n^2} \ne \frac{k(n-k)}{n(n-1)}$.

10.21. (a) We have for $1 \le m < n$

$$P(S_m = \ell | S_n = k) = \frac{P(S_m = \ell, S_n = k)}{P(S_n = k)} = \frac{P(S_m = \ell, S_n - S_m = k - \ell)}{P(S_n = k)}.$$

We know that $S_n \sim \text{Bin}(n,p)$ and $S_k \sim \text{Bin}(k,p)$ as these random variables count the number of successes within the first n and k trials. The random variable $S_n - S_k$ counts the number of successes within the trials $k+1, k+2, \dots, n$, so its distribution is $\text{Bin}(n-k,p)$. Moreover, $S_n - S_k$ is independent of S_k, since S_k depends on the outcome of the first k, and $S_n - S_k$ depends on the next $n-k$ trials. Thus

$$\begin{aligned} P(S_m = \ell | S_n = k) &= \frac{P(S_m = \ell, S_n - S_m = k - \ell)}{P(S_n = k)} = \frac{P(S_m = \ell)P(S_n - S_m = k - \ell)}{P(S_n = k)} \\ &= \frac{\binom{m}{\ell}p^\ell(1-p)^{m-\ell}\binom{n-m}{k-\ell}p^{k-\ell}(1-p)^{(n-m)-(k-\ell)}}{\binom{n}{k}p^k(1-p)^{n-k}} \\ &= \frac{\binom{m}{\ell}\binom{n-m}{k-\ell}}{\binom{n}{k}}. \end{aligned}$$

This means that the conditional distribution of S_m given $S_n = k$ is hypergeometric with parameters n, k, m. Intuitively, the conditional distribution of S_m given $S_n = k$ is identical to the distribution of the number of successes that occur by sampling m times without replacement from a set containing k successes and $n-k$ failures.

(b) From Example 8.7 we know that the expectation of a $\text{Hypgeom}(n,k,m)$ distributed random variable is $\frac{mk}{n}$. Hence $E[S_m|S_n = k] = \frac{mk}{n}$ and $E[S_m|S_n] = S_n \frac{m}{n}$.

10.22. (a) Start by observing that either $X = 1$ and $Y \geq 2$ (when the first trial is a success) or $X \geq 2$ and $Y = 1$ (when the first trial is a failure). Thus when $Y = 1$ we have, for $m \geq 2$,

$$\begin{aligned} p_{X|Y}(m|1) &= \frac{p_{X,Y}(m,1)}{p_Y(1)} = \frac{P(\text{first } m-1 \text{ trials fail, } m\text{th trial succeeds})}{P(\text{first trial fails})} \\ &= \frac{(1-p)^{m-1}p}{1-p} = (1-p)^{m-2}p. \end{aligned}$$

In the other case when $Y = \ell \geq 2$ we must have $X = 1$, and the calculation also verifies this:

$$\begin{aligned} p_{X|Y}(1|\ell) &= \frac{p_{X,Y}(1,\ell)}{p_Y(\ell)} = \frac{P(\text{first } \ell-1 \text{ trials succeed, } \ell\text{th trial fails})}{P(\text{first trial succeeds})} \\ &= \frac{p^{\ell-1}(1-p)}{p^{\ell-1}(1-p)} = 1. \end{aligned}$$

We can summarize the answer in the following pair of formulas that capture all the possible values of both X and Y:

$$p_{X|Y}(m|1) = \begin{cases} 0, & m = 1 \\ (1-p)^{m-2}p, & m \geq 2, \end{cases}$$

and for $\ell \geq 2$,

$$p_{X|Y}(m|\ell) = \begin{cases} 1, & m = 1 \\ 0, & m \geq 2. \end{cases}$$

(b) We reason as in Example 10.6. Let B be the event that the first trial is a success. Then

$$\begin{aligned}E[\max(X,Y)] &= pE[\max(X,Y)\,|\,B] + (1-p)E[\max(X,Y)\,|\,B^c] \\ &= pE[Y\,|\,B] + (1-p)E[X\,|\,B^c] = pE[Y+1] + (1-p)E[X+1] \\ &= p\left(\frac{1}{p-1}+1\right) + (1-p)\left(\frac{1}{p}+1\right) = \frac{1-p+p^2}{p(1-p)}.\end{aligned}$$

10.23. (a) The distribution of Y is negative binomial with parameters 3 and $1/6$ and the probability mass function is

$$P(Y=y) = \binom{y-1}{2}\frac{1}{6^3}\left(\frac{5}{6}\right)^{y-2}, \qquad y = 3,4,\dots$$

To find the conditional probability $P(X=x|Y=y) = \frac{P(X=x,Y=y)}{P(Y=y)}$ we just need to compute the joint probability mass function of X, Y. Note that $X+2 \le Y$ (since we need at least two more rolls to get the third six after the first six). For $1 \le x$, $x+2 \le y$ the event $\{X=x, Y=y\}$ is exactly the same as getting no sixes within the first $x-1$ rolls, six on the xth roll, exactly one six from $x+1$ to $y-1$ and a six on the yth roll. These can be written as intersection of independent events, thus

$$\begin{aligned}P(X=x,Y=y) &= P(\text{no sixes within the first } x-1 \text{ rolls})P(x\text{th roll is a six}) \\ &\quad \cdot P(\text{exactly one six from } x+1 \text{ to } y-1)P(y\text{th roll is a six}) \\ &= \left(\frac{5}{6}\right)^{x-1}\cdot\frac{1}{6}\cdot\left((y-x-2)\left(\frac{5}{6}\right)^{y-x-2}\cdot\frac{1}{6}\right)\cdot\frac{1}{6} \\ &= (y-x-2)\left(\frac{5}{6}\right)^{y-3}\cdot\frac{1}{6^3}.\end{aligned}$$

This leads to

$$\begin{aligned}P(X=x|Y=y) &= \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{(y-x-2)\left(\frac{5}{6}\right)^{y-3}\cdot\frac{1}{6^3}}{\binom{y-1}{2}\frac{1}{6^3}\left(\frac{5}{6}\right)^{y-2}} \\ &= \frac{y-x-2}{\frac{(y-1)(y-2)}{2}} = \frac{2(y-x-1)}{(y-1)(y-2)},\end{aligned}$$

if $1 \le x$, $x+2 \le y$ and zero otherwise.

(b) For a given $y \ge 3$ the possible values of X are $1, 2, \dots, y-2$. Using the result of part(a) we get

$$E[X|Y=y] = \sum_{x=1}^{y-2} x\frac{2(y-x-1)}{(y-1)(y-2)}.$$

To evaluate the sum $\sum_{x=1}^{y-2} 2x(y-x-1)$ we separate it in parts and then use the identities (D.6) and (D.7):

$$\begin{aligned}\sum_{x=1}^{y-2} 2x(y-x-1) &= 2(y-1)\sum_{x=1}^{y-2} x - 2\sum_{x=1}^{y-2} x^2 \\ &= 2(y-1)\frac{(y-2)(y-1)}{2} - 2\frac{(y-2)(y-1)(2(y-2)+1)}{6} \\ &= \frac{(y-2)(y-1)y}{3}.\end{aligned}$$

This gives

$$E[X|Y=y] = \sum_{x=1}^{y-2} x\frac{2(y-x-1)}{(y-1)(y-2)} = \frac{(y-2)(y-1)y}{3(y-2)(y-1)} = \frac{y}{3},$$

and $E[X|Y] = \frac{Y}{3}$.

10.24. (a) Given $\{Y=y\}$ the distribution of X is $\text{Bin}(y, \frac{1}{6})$. Thus

$$p_{X|Y}(x\,|\,y) = \binom{y}{x}\left(\tfrac{1}{6}\right)^x\left(\tfrac{5}{6}\right)^{y-x}, \qquad 0 \le x \le y \le 10.$$

Since $Y \sim \text{Bin}(10, \frac{1}{2})$ we have $p_Y(y) = \binom{10}{y}(\frac{1}{2})^{10}$ and then

$$p_{X,Y}(x,y) = p_{X|Y}(x\,|\,y)p_Y(y) = \binom{y}{x}\left(\tfrac{1}{6}\right)^x\left(\tfrac{5}{6}\right)^{y-x}\binom{10}{y}\left(\tfrac{1}{2}\right)^{10}, \qquad 0 \le x \le y \le 10.$$

The unconditional probability mass function of X can be computed as

$$\begin{aligned}p_X(x) &= \sum_y p_{X|Y}(x\,|\,y)p_Y(y) = \sum_y p_{X,Y}(x,y) = \sum_{y=x}^{10}\binom{y}{x}\left(\tfrac{1}{6}\right)^x\left(\tfrac{5}{6}\right)^{y-x}\binom{10}{y}\tfrac{1}{2^{10}} \\ &= \sum_{y=x}^{10}\frac{10!}{x!(y-x)!(10-y)!}\left(\tfrac{1}{6}\right)^x\left(\tfrac{5}{6}\right)^{y-x}2^{-10} \\ &= \frac{10!}{x!(10-x)!}\left(\tfrac{1}{6}\right)^x\left(\tfrac{1}{2}\right)^{10}\sum_{k=0}^{10-x}\frac{(10-x)!}{k!(10-x-k)!}\left(\tfrac{5}{6}\right)^k \\ &= \binom{10}{x}\left(\tfrac{1}{6}\right)^x\left(\tfrac{1}{2}\right)^{10}\left(\tfrac{11}{6}\right)^{10-x} = \binom{10}{x}\left(\tfrac{1}{12}\right)^x\left(\tfrac{11}{12}\right)^{10-x}.\end{aligned}$$

The conditional expectation $E[X|Y=y]$ for a fixed y is just the expected value of $\text{Bin}(y, \frac{1}{6})$ which is $\frac{y}{6}$. This means that $E(X|Y) = \frac{Y}{6}$ and

$$E[X] = E[E(X|Y)] = E[\tfrac{Y}{6}] = \tfrac{5}{6},$$

since $Y \sim \text{Bin}(10, \frac{1}{2})$.

(b) A closer inspection of the joint probability mass function shows that $(X, Y - X, 10 - Y)$ has a multinomial distribution with parameters $(10, \frac{1}{12}, \frac{5}{12}, \frac{1}{2})$:

$$\begin{aligned} P(X = x, Y - X =&y - x, 10 - Y = 10 - y) = P(X = x, Y = y) \\ &= \binom{y}{x} \left(\tfrac{1}{6}\right)^x \left(\tfrac{5}{6}\right)^{(y-x)} \binom{10}{y} \tfrac{1}{2^{10}} \\ &= \tfrac{10!}{x!(y-x)!(10-y)!} \left(\tfrac{1}{12}\right)^x \left(\tfrac{5}{12}\right)^{y-x} \left(\tfrac{1}{2}\right)^{10-y}. \end{aligned}$$

This implies again that X is just a $\text{Bin}(10, \frac{1}{12})$ random variable.

To see the joint distribution without computation, imagine that after we flip the 10 coins, we roll 10 dice, but only count the sixes if the corresponding coin showed heads. This is the same experiment because the number of 'counted' sixes has the same distribution as X. This is the number of successes for 10 identical experiments where success for the kth experiment means that the kth coin shows heads and the kth die shows six. The probability of success is $\frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$. Moreover, $(X, Y - X, 10 - Y)$ gives the number of outcomes where we have heads and a six, heads and not a six, and tails. This explains why the the joint distribution is multinomial with probabilities $(\frac{1}{12}, \frac{5}{12}, \frac{1}{2})$.

10.25. (a) The conditional distribution of Y given $X = x$ is a negative binomial with parameters x, $1/2$: so we have

$$P(Y = y|X = x) = \binom{y-1}{x-1} \frac{1}{2^y}, \qquad 1 \le x \le y.$$

(b) We have $P(X = x) = (5/6)^{x-1}(1/6)$ and $X \le Y$ so

$$\begin{aligned} P(Y = y) &= \sum_x P(Y = y|X = x) P(X = x) \\ &= \sum_{x=1}^{y} \binom{y-1}{x-1} \frac{1}{2^y} (\tfrac{5}{6})^{x-1} (\frac{1}{6}) = \frac{1}{6} \cdot \frac{1}{2^y} \sum_{i=0}^{y-1} \binom{y-1}{i} (\tfrac{5}{6})^i \\ &= \frac{1}{6} \frac{1}{2^y} (1 + \tfrac{5}{6})^{y-1} = \frac{1}{12} \left(\frac{11}{12}\right)^{y-1}. \end{aligned}$$

We can recognize this as the probability mass function of the geometric distribution with parameter $\frac{1}{12}$.

(c) We have for $1 \le x \le y$:

$$\begin{aligned} P(X = x|Y = y) &= \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{\binom{y-1}{x-1} \frac{1}{2^y} (5/6)^{x-1} (1/6)}{\frac{1}{12} (\frac{11}{12})^{y-1}} \\ &= \binom{y-1}{x-1} (\tfrac{5}{11})^{x-1} (\tfrac{6}{11})^{y-x}. \end{aligned}$$

Thus the conditional distribution of $X - 1$ given $Y = y$ is $\text{Bin}(y - 1, \frac{5}{11})$.

10.26. Let B be the event that the first trial is a success. Recall that $E[N] = p^{-1}$.

$$\begin{aligned} E(N^2) &= E[N^2|B]P(B) + E[N^2|B^c]P(B^c) = 1 \cdot p + E[(N+1)^2] \cdot (1-p) \\ &= p + (1-p)(E[N^2] + 2E[N] + 1) = p + (1-p)(2p^{-1} + 1) + (1-p)E[N^2] \\ &= \frac{2-p}{p} + (1-p)E[N^2]. \end{aligned}$$

From the equation above we solve

$$E[N^2] = \frac{2-p}{p^2}.$$

From this,

$$\mathrm{Var}(N) = E[N^2] - (E[N])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

10.27. Utilize again the temporary notation $E[X|Y] = v(Y)$ from Definition 10.23 and identity (10.11):

$$E\big[E[X|Y]\big] = E[v(Y)] = \sum_y v(y)p_Y(y) = \sum_y E[X|Y=y]p_Y(y) = E(X).$$

10.28. We reason as in Example 10.13. First deduction of the joint p.m.f. Let $k_1, k_2, \dots, k_r \in \{0, 1, 2, \dots\}$ and set $k = k_1 + k_2 + \cdots + k_r$. In the first equality below we can add the condition $X = k$ into the probability because the event $\{X_1 = k_1, X_2 = k_2, \dots, X_r = k_r\}$ is a subset of the event $\{X = k\}$.

$$\begin{aligned} &P(X_1 = k_1, X_2 = k_2, \dots, X_r = k_r) \\ &\quad = P(X_1 = k_1, X_2 = k_2, \dots, X_r = k_r, X = k) \\ &\quad = P(X = k)\, P(X_1 = k_1, X_2 = k_2, \dots, X_r = k_r \,|\, X = k) \\ \text{(A)} \qquad &\quad = \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{k!}{k_1!\, k_2! \cdots k_r!}\, p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \\ &\quad = \frac{e^{-p_1\lambda}(p_1\lambda)^{k_1}}{k_1!} \cdot \frac{e^{-p_2\lambda}(p_2\lambda)^{k_2}}{k_2!} \cdots \frac{e^{-p_r\lambda}(p_r\lambda)^{k_r}}{k_r!}. \end{aligned}$$

In the passage from line 3 to line 4 we used the conditional joint probability mass function of $(X_1, X_2, \dots, X_r)$, given that $X = k$, namely

$$P(X_1 = k_1, X_2 = k_2, \dots, X_r = k_r \,|\, X = k) = \frac{k!}{k_1!\, k_2! \cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r},$$

which came from the description of the problem. In the last equality of (A) we cancelled $k!$ and then used both $k = k_1 + k_2 + \cdots + k_r$ and $p_1 + p_2 + \cdots + p_r = 1$.

From the joint p.m.f. we deduce the marginal p.m.f.s by summing away the other variables. Let $1 \le j \le r$ and $\ell \ge 0$. In the second equality below substitute in the last line from (A). Then observe that each sum over the entire Poisson p.m.f.

evaluates to 1.

$$\begin{aligned}P(X_j = \ell) = \sum_{\substack{k_1,\dots,k_{j-1},\\ k_{j+1},\dots,k_r\geq 0}} & P\big(X_1 = k_1, \dots, X_{j-1} = k_{j-1}, \\ & \qquad X_j = \ell,\ X_{j+1} = k_{j+1}, \dots, X_r = k_r\big)\end{aligned}$$

$$\begin{aligned}&= \left(\sum_{k_1=0}^{\infty} \frac{e^{-p_1\lambda}(p_1\lambda)^{k_1}}{k_1!}\right)\cdots\left(\sum_{k_{j-1}=0}^{\infty} \frac{e^{-p_{j-1}\lambda}(p_{j-1}\lambda)^{k_{j-1}}}{k_{j-1}!}\right)\frac{e^{-p_j\lambda}(p_j\lambda)^{\ell}}{\ell!} \\ &\qquad \cdot \left(\sum_{k_{j+1}=0}^{\infty} \frac{e^{-p_{j+1}\lambda}(p_{j+1}\lambda)^{k_{j+1}}}{k_{j+1}!}\right)\cdots\left(\sum_{k_r=0}^{\infty} \frac{e^{-p_r\lambda}(p_r\lambda)^{k_r}}{k_r!}\right) \\ &= \frac{e^{-p_j\lambda}(p_j\lambda)^{\ell}}{\ell!}.\end{aligned}$$

This gives us $X_j \sim \text{Poisson}(p_j\lambda)$ for each j. Together with the earlier calculation (A) we now know that $X_1, X_2, \dots, X_r$ are independent with Poisson marginals $X_j \sim \text{Poisson}(p_j\lambda)$.

10.29. For $0 \leq \ell \leq n$,

$$\begin{aligned}p_L(\ell) &= \sum_{m=\ell}^{n} p_{L|M}(\ell|m)\, p_M(m) \\ &= \sum_{m=\ell}^{n} \frac{m!}{\ell!(m-\ell)!} r^{\ell}(1-r)^{m-\ell} \cdot \frac{n!}{m!(n-m)!} p^m(1-p)^{n-m} \\ &= \frac{n!}{\ell!(n-\ell)!}(pr)^{\ell} \sum_{m=\ell}^{n} \frac{(n-\ell)!}{(m-\ell)!(n-m)!}(1-r)^{m-\ell}p^{m-\ell}(1-p)^{n-m} \\ &= \frac{n!}{\ell!(n-\ell)!}(pr)^{\ell} \sum_{j=0}^{n-\ell} \frac{(n-\ell)!}{j!(n-\ell-j)!}((1-r)p)^j(1-p)^{n-\ell-j} \\ &= \frac{n!}{\ell!(n-\ell)!}(pr)^{\ell}((1-r)p+1-p)^{n-\ell} = \binom{n}{\ell}(pr)^{\ell}(1-pr)^{n-\ell}.\end{aligned}$$

In other words, $L \sim \text{Bin}(n, pr)$.

Here is a way to get the distribution of L without calculation. Imagine that we allow everybody to write the second test (even those applicants who fail the first one). For a given applicant the probability of passing both tests is pr by independence. Since L is the number of applicants passing both tests out of the n applicants, we immediately get $L \sim \text{Bin}(n, pr)$.

10.30. First deduction of the joint p.m.f. Let $k, \ell \in \{0, 1, 2, \dots\}$.

$$\begin{aligned}P(X_1 = k, X_2 = \ell) &= P(X_1 = k,\ X_2 = \ell,\ X = k+\ell) \\ &= P(X = k+\ell)\, P(X_1 = k,\ X_2 = \ell \,|\, X = k+\ell) \\ &= (1-p)^{k+\ell}p \cdot \frac{(k+\ell)!}{k!\,\ell!}\alpha^k(1-\alpha)^{\ell}.\end{aligned}$$

To find the marginal p.m.f. we manipulate the series into a form where we can apply identity (10.52). Let $k \geq 0$.

$$
\begin{aligned}
P(X_1 = k) &= \sum_{\ell=0}^{\infty} P(X_1 = k, X_2 = \ell) = \sum_{\ell=0}^{\infty} (1-p)^{k+\ell} p \cdot \frac{(k+\ell)!}{k!\,\ell!} \alpha^k (1-\alpha)^\ell \\
&= (\alpha(1-p))^k p \sum_{\ell=0}^{\infty} \frac{(k+1)(k+2)\cdots(k+\ell)}{\ell!} \big((1-p)(1-\alpha)\big)^\ell \\
&= (\alpha(1-p))^k p \sum_{\ell=0}^{\infty} \frac{(-k-1)(-k-2)\cdots(-k-1-\ell+1)}{\ell!} \big(-(1-p)(1-\alpha)\big)^\ell \\
&= (\alpha(1-p))^k p \sum_{\ell=0}^{\infty} \binom{-k-1}{\ell} \big(-(1-p)(1-\alpha)\big)^\ell \\
&= (\alpha(1-p))^k \cdot p \cdot \big(1-(1-p)(1-\alpha)\big)^{-k-1} \\
&= \left(\frac{\alpha(1-p)}{p+\alpha(1-p)}\right)^k \cdot \frac{p}{p+\alpha(1-p)}.
\end{aligned}
$$

Same reasoning (or simply replacing α with $1-\alpha$) gives for $\ell \geq 0$

$$
P(X_2 = \ell) = \left(\frac{(1-\alpha)(1-p)}{p+(1-\alpha)(1-p)}\right)^\ell \cdot \frac{p}{p+(1-\alpha)(1-p)}.
$$

Thus marginally X_1 and X_2 are shifted geometric random variables. However, the conditional p.m.f. of X_2, given that $X_1 = k$, is of a different form and furthermore depends on k:

$$
\begin{aligned}
p_{Y|X}(\ell|k) &= \frac{p_{X,Y}(k,\ell)}{p_X(k)} = \frac{(1-p)^{k+\ell} p \cdot \frac{(k+\ell)!}{k!\,\ell!} \alpha^k (1-\alpha)^\ell}{\left(\frac{\alpha(1-p)}{p+\alpha(1-p)}\right)^k \cdot \frac{p}{p+\alpha(1-p)}} \\
&= (p+\alpha(1-p))^{k+1} \frac{(k+1)(k+2)\cdots(k+\ell)}{\ell!} \big((1-p)(1-\alpha)\big)^\ell.
\end{aligned}
$$

We conclude in particular that X_1 and X_2 are not independent.

10.31. We have

$$
p_{X|I_B}(x\,|\,1) = P(X = x\,|\,I_B = 1) = P(X = x\,|\,B) = p_{X|B}(x),
$$

and

$$
p_{X|I_B}(x\,|\,0) = P(X = x\,|\,I_B = 0) = P(X = x\,|\,B^c) = p_{X|B^c}(x).
$$

10.32. From Exercise 6.34 we record the joint and marginal density functions:

$$
f_{X,Y}(x,y) = \begin{cases} \frac{2}{3} & (x,y) \in D, \\ 0 & (x,y) \notin D, \end{cases}
$$

$$
f_X(x) = \begin{cases} 0 & x \leq 0 \text{ or } x \geq 2, \\ \frac{2}{3} & 0 < x \leq 1, \\ \frac{4}{3} - \frac{2}{3}x & 1 < x < 2, \end{cases} \qquad f_Y(y) = \begin{cases} 0 & y \leq 0 \text{ or } y \geq 1, \\ \frac{4}{3} - \frac{2}{3}y & 0 < y < 1. \end{cases}
$$

From these we deduce the conditional densities. Note that the line segment from $(1,1)$ to $(2,0)$ that forms part of the boundary of D obeys the equation

$y = 2 - x$ and consequently all points of D (excluding boundary points) satisfy $x > 0$, $0 < y < 1$, and $x + y < 2$.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{2}{3}}{\frac{4}{3} - \frac{2}{3}y} = \frac{1}{2-y} \quad \text{for } 0 < x < 2-y \text{ and } 0 < y < 1.$$

This shows that given $Y = y \in (0,1)$, X is uniform on the interval $(0, 2-y)$. Since the mean of a uniform random variable is the midpoint of the interval,

$$E[X|Y = y] = 1 - \tfrac{y}{2} \quad \text{for } 0 < y < 1.$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{\frac{2}{3}}{\frac{2}{3}} = 1 & \text{for } 0 < y < 1 \text{ and } 0 < x \le 1, \\ \frac{\frac{2}{3}}{\frac{4}{3} - \frac{2}{3}x} = \frac{1}{2-x} & \text{for } 0 < y < 2-x \text{ and } 1 < x < 2. \end{cases}$$

Thus given $X = x \in (0,1]$, Y is uniform on the interval $(0,1)$, while given $X = x \in (1,2)$, Y is uniform on the interval $(0, 2-x)$. Hence

$$E[Y|X = x] = \begin{cases} \frac{1}{2} & 0 < x \le 1, \\ 1 - \frac{x}{2} & 1 < x < 2. \end{cases}$$

We combine the answers in the formulas for the conditional expectations as random variables:

$$E[X|Y] = 1 - \tfrac{1}{2}Y \quad \text{and} \quad E[Y|X] = \begin{cases} \frac{1}{2} & \text{if } X \le 1, \\ 1 - \frac{1}{2}X & \text{if } X > 1. \end{cases}$$

(Note that not all bounds are needed explicitly in the cases above because with probability one we have $0 < Y < 1$ and $0 < X < 2$.)

Last, we calculate the expectations of the conditional expectations.

$$\begin{aligned} E[X] &= E[E(X|Y)] = E[1 - \tfrac{1}{2}Y] = 1 - \tfrac{1}{2}E[Y] = 1 - \tfrac{1}{2}\int_0^1 y(\tfrac{4}{3} - \tfrac{2}{3}y)\,dy \\ &= 1 - \tfrac{1}{2} \cdot \tfrac{4}{9} = \tfrac{7}{9}. \\ E[Y] &= E[E(Y|X)] = \int_{-\infty}^{\infty} E[Y|X = x]\, f_X(x)\,dx \\ &= \int_0^1 \tfrac{1}{2} \cdot \tfrac{2}{3}\,dx + \int_1^2 (1 - \tfrac{1}{2}x)(\tfrac{4}{3} - \tfrac{2}{3}x)\,dx = \tfrac{1}{3} + \tfrac{1}{9} = \tfrac{4}{9}. \end{aligned}$$

10.33. (a) By formula (10.15),

$$P(X \le \tfrac{1}{2} \,|\, Y = y) = \int_{-\infty}^{1/2} f_{X|Y}(x \,|\, y)\,dx.$$

To find the correct limits of integration, look at (10.20) and check where the integrand $f_{X|Y}(x \,|\, y)$ is nonzero on the integration interval $(-\infty, \frac{1}{2}]$. There are three cases, depending on whether the right endpoint $\frac{1}{2}$ is to the left of, in the middle of, or to the right of the interval $[1-y, 2-2y]$. We get these three cases.

(i) $y < \frac{1}{2}$: $P(X \le \frac{1}{2} \,|\, Y = y) = 0$.

(ii) $\frac{1}{2} \le y < \frac{3}{4}$: $P(X \le \frac{1}{2} \,|\, Y = y) = \displaystyle\int_{1-y}^{1/2} \frac{1}{1-y}\, dx = \frac{y - \frac{1}{2}}{1-y}$.

(iii) $y \ge \frac{3}{4}$: $P(X \le \frac{1}{2} \,|\, Y = y) = \displaystyle\int_{1-y}^{2-2y} \frac{1}{1-y}\, dx = 1$.

(b) From Figure 6.4 or from the formula for f_X in Example 6.20 we deduce $P(X \le \frac{1}{2}) = \frac{1}{8}$. Then integrate the conditional probability from part (a) to find

$$\begin{aligned}
&\int_{-\infty}^{\infty} P(X \le \tfrac{1}{2} \,|\, Y = y)\, f_Y(y)\, dy \\
&\qquad = \int_{1/2}^{3/4} \frac{y - \frac{1}{2}}{1-y}(2-2y)\, dy + \int_{3/4}^{1} (2-2y)\, dy = \tfrac{1}{8}.
\end{aligned}$$

10.34. The discrete case, utilizing $p_{X|Y}(x|y)p_Y(y) = p_{X,Y}(x,y)$:

$$\begin{aligned}
E[Y \cdot E(X|Y)] &= \sum_y y\, E(X|Y = y)\, p_Y(y) = \sum_y y \sum_x x\, p_{X|Y}(x|y)\, p_Y(y) \\
&= \sum_{x,y} xy\, p_{X|Y}(x|y)\, p_Y(y) = \sum_{x,y} xy\, p_{X,Y}(x,y) = E[XY].
\end{aligned}$$

The jointly continuous case, utilizing $f_{X|Y}(x|y)f_Y(y) = f_{X,Y}(x,y)$:

$$\begin{aligned}
E[Y \cdot E(X|Y)] &= \int_{-\infty}^{\infty} y\, E(X|Y = y)\, f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx \right) f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_{X|Y}(x|y)\, f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f_{X,Y}(x,y)\, dx\, dy = E[XY].
\end{aligned}$$

10.35. (a) We first find the joint density of (X, S). Using the same idea as in Example 10.22, we write an expression for the joint cumulative distribution function $F_{X,S}(x,s)$.

$$\begin{aligned}
F_{X,S}(x,s) &= P(X \le x, S \le s) = P(X \le x, X + Y \le s) \\
&= \iint_{u \le x,\, u+v \le s} f_{X,Y}(u,v)\, du\, dv = \iint_{u \le x,\, v \le s-u} \varphi(u)\varphi(v)\, du\, dv \\
&= \int_{-\infty}^{x} \int_{-\infty}^{s-u} \varphi(u)\varphi(v)\, du\, dv = \int_{-\infty}^{x} \varphi(u)\Phi(s-u) du.
\end{aligned}$$

We can get the joint density of (X, S) by taking the mixed partial derivative, and we will do that by taking the x-derivative first:

$$\begin{aligned}
f_{X,S}(x,s) &= \frac{\partial}{\partial s}\frac{\partial}{\partial x} F_{X,S}(x,s) = \frac{\partial}{\partial s}\left(\frac{\partial}{\partial x} \int_{-\infty}^{x} \varphi(u)\Phi(s-u) du \right) \\
&= \frac{\partial}{\partial s}(\varphi(x)\Phi(s-x)) = \varphi(x)\varphi(s-x) = \frac{1}{2\pi} e^{-\frac{x^2+(s-x)^2}{2}}.
\end{aligned}$$

Since S is the sum of two independent standard normals, we have $S \sim \mathcal{N}(0,2)$ and $f_S(s) = \frac{1}{2\sqrt{\pi}} e^{-\frac{s^2}{4}}$. Then

$$f_{X|S}(x \mid s) = \frac{f_{X,S}(x,s)}{f_S(s)} = \frac{\frac{1}{2\pi} e^{-\frac{x^2+(s-x)^2}{2}}}{\frac{1}{2\sqrt{\pi}} e^{-\frac{s^2}{4}}} = \frac{1}{\sqrt{\pi}} e^{-(\frac{s^2}{4}-sx+x^2)} = \frac{1}{\sqrt{\pi}} e^{-(x-\frac{s}{2})^2}.$$

We can recognize the final result as the probability density function of the $\mathcal{N}(\frac{s}{2}, \frac{1}{2})$ distribution.

(b) Since the conditional distribution of X given $S = s$ is $\mathcal{N}(\frac{s}{2}, \frac{1}{2})$, we get

$$E[X|S=s] = \frac{s}{2}, \quad \text{and} \quad E[X^2|S=s] = \frac{1}{2} + (\tfrac{s}{2})^2,$$

from which $E[X|S] = \frac{S}{2}$, $E[X^2|S] = \frac{1}{2} + \frac{S^2}{4}$.

Taking expectations again:

$$E[E[X|S]] = E[S/2] = 0, \qquad E[E[X^2|S]] = E[\tfrac{1}{2} + \tfrac{S^2}{4}] = \tfrac{1}{2} + \tfrac{2}{4} = 1,$$

where we used $S \sim \mathcal{N}(0,2)$. The final answers agree with the fact that X is standard normal.

10.36. To find the joint density function of (X, S), we change variables in an integral that calculates the expectation of a function $g(X, S)$.

$$\begin{aligned} E[g(X,S)] = E[g(X,X+Y)] &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, x+y) e^{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{(y-\mu)^2}{2\sigma^2}} \, dy \, dx \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,s) e^{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{(s-x-\mu)^2}{2\sigma^2}} \, ds \, dx. \end{aligned}$$

From this we read off

$$f_{X,S}(x,s) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{(s-x-\mu)^2}{2\sigma^2}} \qquad \text{for} \quad x, y \in \mathbb{R}.$$

From the properties of sums of normals we know that $S \sim \mathcal{N}(2\mu, 2\sigma^2)$ and hence

$$f_S(s) = \frac{1}{\sqrt{4\pi\sigma^2}} e^{-\frac{(s-2\mu)^2}{4\sigma^2}}.$$

From these ingredients we write down the conditional density function of X, given that $S = s$:

$$f_{X|S}(x|s) = \frac{f_{X,S}(x,s)}{f_S(s)} = \frac{\sqrt{4\pi\sigma^2}}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{(s-x-\mu)^2}{2\sigma^2} + \frac{(s-2\mu)^2}{4\sigma^2}}.$$

After some algebra and cancellation in the exponent, this turns into

$$f_{X|S}(x|s) = \frac{1}{\sqrt{2\pi\sigma^2/2}} \exp\left\{ -\frac{(x-\frac{s}{2})^2}{2\sigma^2/2} \right\}.$$

The conclusion is that given $S = s$, $X \sim \mathcal{N}(s/2, \sigma^2/2)$. Knowledge of the normal expectation gives $E(X|S=s) = s/2$, from which $E[X|S] = \frac{1}{2}S$.

10.37. Let A be the event $\{Z > 0\}$. Random variable Y has the same distribution as Z conditioned on the event A. Hence the density function $f_Y(y)$ is the same as the conditional probability density function $f_{Z|A}(y)$. This conditional density will be 0 for $y \le 0$, so we can focus on $y > 0$. The conditional density will satisfy

$$P(a \le Z \le b | Z > 0) = \int_a^b f_{Y|A}(y) dy$$

for any $0 < a < b$. But if $0 < a < b$ then

$$\begin{aligned} P(a \le Z \le b | Z > 0) &= \frac{P(a \le Z \le b, Z > 0)}{P(Z > 0)} = \frac{P(a \le Z \le b)}{P(Z > 0)} \\ &= \frac{\int_a^b \varphi(y) dy}{1/2} = \int_a^b 2\varphi(y) dy. \end{aligned}$$

Thus $f_Y(y) = f_{Z|A}(y) = 2\varphi(y)$ for $y > 0$ and 0 otherwise.

10.38. (a) The problem statement gives us these density functions for $x, y > 0$:

$$f_Y(y) = e^{-y} \quad \text{and} \quad f_{X|Y}(x|y) = ye^{-yx}.$$

Then the joint density function is given by

$$f_{X,Y}(x, y) = f_{X|Y}(x|y) f_Y(y) = ye^{-y(x+1)} \qquad \text{for } x > 0,\ y > 0.$$

(b) Once we observe $X = x$, the distribution of Y should be conditioned on $X = x$. First find the marginal density function of X for $x > 0$.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^{\infty} ye^{-y(x+1)} dy = \frac{1}{(1 + x)^2}.$$

Then, again for $x > 0$ and $y > 0$,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = y(1 + x)^2 e^{-y(x+1)}.$$

The conclusion is that, given $X = x$, $Y \sim \text{Gamma}(2, x + 1)$. The gamma distribution was defined in Definition 4.37.

10.39. From the problem we get that the conditional distribution of Y given $X = x$ is uniform on $[x, 1]$. From this we get that $f_{Y|X}(y|x)$ is defined for every $0 \le x < 1$ and is equal to

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } x \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

By averaging out x we can get the unconditional probability density function of Y, for any $0 \le y \le 1$ we have

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{Y|X}(y|x) f_X(x) dx \\ &= \int_0^y \frac{1}{1 - x} \cdot 20x^3(1 - x) dx \\ &= 20 \int_0^y x^3 dx = 20 \frac{x^4}{4} \Big|_0^y = 5y^4 \end{aligned}$$

If $y < 0$ or $y > 1$ then we have $f_Y(y) = 0$, thus

$$f_Y(y) = \begin{cases} 5y^4 & \text{if } 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

10.40. The conditional density function of Y given $X = x$ is

$$f_{Y|X}(y|x) = \begin{cases} x, & 0 < y < 1/x \\ 0, & y \le 0 \text{ or } y \ge 1/x. \end{cases}$$

(a) Conditional on $X = x$, $Y < 1/x$. Hence $P(Y > 2|X = x) = 0$ if $1/x \le 2$ which is equivalent to $x \ge 1/2$. For $0 < x < 1/2$ we have

$$P(Y > 2|X = x) = \int_2^{1/x} x\,dy = x\Big(\frac{1}{x} - 2\Big) = 1 - 2x.$$

In summary,

$$P(Y > 2|X = x) = \begin{cases} 0, & \text{if } x \ge 1/2 \\ 1 - 2x, & \text{if } 0 < x < 1/2. \end{cases}$$

(b) Since the expectation of a uniform random variable is the midpoint of the interval, $E[Y|X = x] = \frac{1}{2x}$ and from this $E[Y|X] = 1/(2X)$. Finally,

$$E[Y] = E[E[Y|X]] = E\big[\frac{1}{2X}\big] = \int_0^\infty \frac{1}{2x}\cdot xe^{-x}\,dx = \frac{1}{2}\int_0^\infty e^{-x}\,dx = \frac{1}{2}.$$

10.41. Let X be the length of the stick after two stick-breaking steps. From Example 10.26 we have $f_X(x) = -\ln x$ for $0 < x < 1$ and zero elsewhere, and from the problem description $f_{Z|X}(z|x) = \frac{1}{x}$ for $0 < z < x < 1$. Thus for $0 < z < 1$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{Z|X}(z|x)\, f_X(x)\,dx = -\int_z^1 \frac{\ln x}{x}\,dx = -\tfrac{1}{2}\int_z^1 \frac{d}{dx}(\ln x)^2\,dx \\ &= -\tfrac{1}{2}((\ln 1)^2 - (\ln z)^2) = \tfrac{1}{2}(\ln z)^2. \end{aligned}$$

As already computed in Example 10.26, $E(Z|X) = \frac{1}{2}X$ and $E(Z^2|X) = \frac{1}{3}X^2$. Next compute

$$E(Z) = E[E(Z|X)] = \tfrac{1}{2}E(X) = \tfrac{1}{8}$$

and

$$E(Z^2) = E[E(Z^2|X)] = \tfrac{1}{3}E(X^2) = \tfrac{1}{27}.$$

Finally, $\mathrm{Var}(Z) = E(Z^2) - (E[Z])^2 = \frac{1}{27} - \frac{1}{64} = \frac{37}{1728} \approx 0.021$.

10.42. We introduce several random variables to get to X. First let $U \sim \mathrm{Unif}(0,1)$ and then $Y = \min(U, 1-U)$. Then Y is the length of the shorter piece after the first stick breaking. Let us deduce the density function $f_Y(y)$ by differentiating the c.d.f. of Y. Y cannot be larger than $1/2$, and hence we can restrict to $0 < y \le 1/2$. Exclusion of one point makes no difference to the density function so we can restrict to $0 < y < 1/2$. This is convenient because for $0 < y < 1/2$ the events $\{U \le y\}$ and $\{U \ge 1-y\}$ are disjoint. This makes the addition of probabilities in the next calculation legitimate.

$$F_Y(y) = P(Y \le y) = P(U \le y) + P(U \ge 1-y) = y + 1 - (1-y) = 2y.$$

From this $f_Y(y) = F_Y'(y) = 2$ for $0 < y < 1/2$.

Next, given $Y = y$, let $V \sim \mathrm{Unif}(0,y)$ and then $X = \min(V, Y-V)$. Now X is the length of the shorter piece after the second stick breaking. We apply the same strategy to find the conditional density function $f_{X|Y}(x|y)$, namely, we differentiate the conditional c.d.f. Since $X \le Y/2$, when conditioning on $Y = y$ we discard the value $y/2$ and restrict to $0 < x < y/2$:

$$\begin{aligned} P(X \le x|Y = y) &= P(V \le x|Y = y) + P(V \ge y - x|Y = y) \\ &= \frac{x}{y} + \frac{y-(y-x)}{y} = \frac{2x}{y}. \end{aligned}$$

From this,

$$f_{X|Y}(x|y) = \frac{d}{dx} P(X \le x|Y = y) = \frac{2}{y} \qquad \text{for } 0 < x < y/2 \text{ and } 0 < y < 1/2.$$

From these ingredients we find the density function $f_X(x)$. Concerning the range, the inequalities $0 < x < y/2$ and $0 < y < 1/2$ combine to give $0 < x < 1/4$. For such x,

$$\text{(A)} \qquad f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy = \int_{2x}^{1/2} \frac{2}{y} \cdot 2\, dy = -4 \ln 4x.$$

Alternative. Instead of the two separate calculations above for finding f_Y and $f_{X|Y}$, we can do a single calculation for a stick of general length. Let Z be the length of the shorter piece when a stick of length ℓ is broken at a uniformly random position. Let $U \sim \text{Unif}(0, \ell)$. Then as above, for $0 < z < \ell/2$,

$$F_Z(z) = P(Z \le z) = P(U \le z) + P(U \ge \ell - z) = \frac{z}{\ell} + \frac{\ell - (\ell - z)}{\ell} = \frac{2z}{\ell}$$

from which $f_Z(z) = F_Z'(z) = 2/\ell$ for $0 < z < \ell/2$. We apply this first with $\ell = 1$ to get $f_Y(y) = 2$ for $0 < y < 1/2$ and then with $\ell = y$ to get $f_{X|Y}(x|y) = 2/y$ for $0 < x < y/2$. The solution is then completed with (A) as above.

10.43. (a) Since $0 < Y < 2$ we can assume that $0 < y < 2$. The area of the triangle is 2, thus the joint density $f_{X,Y}(x, y)$ is $\frac{1}{2}$ inside the triangle, and 0 outside. Note that the points (x, y) in the triangle are the points satisfying $0 \le x, 0 \le y$ and $x + y \le 2$. For $0 < y < 2$ we have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_{y}^{2} \tfrac{1}{2} dx = \tfrac{2-y}{2}$$

and $f_Y(y) = 0$ otherwise. Thus

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \\ &= \begin{cases} \frac{\frac{1}{2}}{\frac{2-y}{2}} = \frac{1}{2-y} & \text{if } x < 2 - y \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

This shows that the conditional distribution of X given $Y = y$ is Uniform$[y, 2]$.
(b) From part (a) we have $E[X|Y = y] = \frac{y+2}{2}$ and $E[X|Y] = \frac{Y+2}{2}$.

10.44. The calculation below begins with the averaging principle. Conditioning on $Y = y$ permits us to replace Y with y inside the probability, and then the conditioning can be dropped because X and Y are independent. Manipulation of the integrals then gives us the convolution formula.

$$\begin{aligned}
P(X+Y\le z) &= \int_{-\infty}^{\infty} P(X+Y\le z\,|\,Y=y)\,f_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} P(X\le z-y\,|\,Y=y)\,f_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} P(X\le z-y)\,f_Y(y)\,dy = \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{z-y} f_X(w)\,dw\Big) f_Y(y)\,dy \\
&= \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{z} f_X(x-y)\,dx\Big) f_Y(y)\,dy = \int_{-\infty}^{z}\Big(\int_{-\infty}^{\infty} f_X(x-y)\,f_Y(y)\,dy\Big)\,dx.
\end{aligned}$$

10.45. (a) We have the joint density $f_{X,Y}(a,y)$ given in (8.32). The distribution of Y is $\mathcal{N}(\mu_Y,\sigma_Y^2)$ and thus the marginal density is $f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$. Then $f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$. To help with the notation let us introduce $\tilde{x}=\frac{x-\mu_X}{\sigma_X}$ and $\tilde{y}=\frac{y-\mu_Y}{\sigma_Y}$. Then

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}(\tilde{x}^2+\tilde{y}^2-2\rho\tilde{x}\tilde{y})},\quad f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{\tilde{y}^2}{2}}$$

and

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}(\tilde{x}^2+\tilde{y}^2-2\rho\tilde{x}\tilde{y})}}{\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{\tilde{y}^2}{2}}} = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X}e^{-\frac{\tilde{x}^2-2\tilde{x}\tilde{y}\rho+\tilde{y}^2\rho^2}{2(1-\rho^2)}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X}e^{-\frac{(\tilde{x}-y\tilde{\rho})^2}{2(1-\rho^2)}}
\end{aligned}$$

Substituting back $\tilde{x}=\frac{x-\mu_X}{\sigma_X}$ and $\tilde{y}=\frac{y-\mu_Y}{\sigma_Y}$ we see that the conditional distribution of X given $Y=y$ is normal distribution with mean $\frac{\sigma_X}{\sigma_X}\rho(y-\mu_Y)+\mu_X$ and variance $\sigma_X^2(1-\rho^2)$.

(b) The conditional expectation of X given $Y=y$ is the mean of the normal distribution we found: $\frac{\sigma_X}{\sigma_X}\rho(y-\mu_Y)+\mu_X$. Thus

$$E[X|Y]=\frac{\sigma_X}{\sigma_X}\rho(Y-\mu_Y)+\mu_X.$$

Note that this is just a linear function of Y.

10.46. The definitions of conditional p.m.f.s and density functions use a ratio of a joint probability or density function over a marginal. Following the same joint/marginal pattern, a sensible suggestion would be

$$f_X(x\,|\,Y\in B)=\frac{1}{P(Y\in B)}\int_B f(x,y)\,dy.$$

A conditional probability of X should come by integrating the conditional density, and so we would expect

$$P(X\in A\,|\,Y\in B)=\int_A f_X(x\,|\,Y\in B)\,dx.$$

We can check that the formula given above for $f_X(x \,|\, Y \in B)$ satisfies this identity. By the definition of conditional probability,

$$\begin{aligned} P(X \in A \,|\, Y \in B) &= \frac{P(X \in A, Y \in B)}{P(Y \in B)} = \frac{1}{P(Y \in B)} \iint_{A \times B} f(x,y)\,dx\,dy \\ &= \frac{1}{P(Y \in B)} \int_A \left(\int_B f(x,y)\,dy \right) dx = \int_A f_X(x \,|\, Y \in B)\,dx. \end{aligned}$$

10.47.

$$\begin{aligned} &E[\,g(X) \,|\, Y = y] = \sum_m m\, P(g(X) = m \,|\, Y = y) = \sum_m m \sum_{k: g(k) = m} P(X = k \,|\, Y = y) \\ &= \sum_m \sum_{k: g(k)=m} m\, P(X = k \,|\, Y = y) = \sum_m \sum_{k: g(k) = m} g(k) P(X = k \,|\, Y = y) \\ &= \sum_k g(k) P(X = k \,|\, Y = y). \end{aligned}$$

10.48.

$$\begin{aligned} E[X + Z \,|\, Y = y] &= \sum_m m\, P(X + Z = m \,|\, Y = y) \\ &= \sum_m m \sum_{k,\ell:\, k+\ell=m} P(X = k, Z = \ell \,|\, Y = y) \\ &= \sum_{k,\ell,m:\, k+\ell=m} m\, P(X = k, Z = \ell \,|\, Y = y) \\ &= \sum_{k,\ell} (k + \ell) P(X = k, Z = \ell \,|\, Y = y) \\ &= \sum_{k,\ell} k\, P(X = k, Z = \ell \,|\, Y = y) + \sum_{k,\ell} \ell\, P(X = k, Z = \ell \,|\, Y = y) \\ &= \sum_k k \sum_\ell P(X = k, Z = \ell \,|\, Y = y) + \sum_\ell \ell \sum_k P(X = k, Z = \ell \,|\, Y = y) \\ &= \sum_k k\, P(X = k \,|\, Y = y) + \sum_\ell \ell\, P(Z = \ell \,|\, Y = y) \\ &= E[X \,|\, Y = y] + E[Z \,|\, Y = y]. \end{aligned}$$

10.49. (a) If it takes me more than one time unit to complete the job I'm simply paid 1 dollar, so for $t \geq 1$, $p_{X|T}(1|t) = 1$. For $0 < t < 1$ we get either 1 or 2 dollars with probability $1/2 - 1/2$, so the conditional probability mass function is

$$p_{X|T}(1|t) = \tfrac{1}{2} \quad \text{and} \quad p_{X|T}(2|t) = \tfrac{1}{2}.$$

(b) From part (a) we get that

$$E[X|T = t] = \begin{cases} 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}, & \text{if } 0 < t < 1 \\ 1 \cdot 1 = 1 & \text{if } 1 \leq t. \end{cases}$$

We can compute $E[X]$ by averaging $E[X|T = t]$ using the probability density $f_T(t)$ of T. Since $T \sim \text{Exp}(\lambda)$, we have $f_T(t) = \lambda e^{-\lambda t}$ for $t > 0$ and 0 otherwise. Thus

$$\begin{aligned} E[X] &= \int_0^\infty E[X|T=t] f_T(t) dy = \int_0^1 \frac{3}{2}\lambda e^{-\lambda t} dt + \int_1^\infty \lambda e^{-\lambda t} dt \\ &= \frac{3}{2}(1-e^{-\lambda}) + e^{-\lambda} = \frac{3}{2} - \frac{1}{2}e^{-\lambda}. \end{aligned}$$

10.50. For $0 \le k < n$ we have

$$P(S_n = k) = \int_0^1 P(S_n = k \,|\, \xi = p) f_\xi(p)\, dp = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\, dp.$$

We use integration by parts on the right-hand side to show that $P(S_n = k) = P(S_n = k+1)$.

$$\begin{aligned} P(S_n = k) &= \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\, dp \\ &= \binom{n}{k} \left[\frac{p^{k+1}}{k+1}(1-p)^{n-k} \Big|_{p=0}^{p=1} + \frac{n-k}{k+1} \int_0^1 p^{k+1}(1-p)^{n-k-1}\, dp \right] \\ &= \binom{n}{k} \frac{n-k}{k+1} \int_0^1 p^{k+1}(1-p)^{n-k-1}\, dp \\ &= \binom{n}{k+1} \int_0^1 p^{k+1}(1-p)^{n-k-1}\, dp \; = \; P(S_n = k+1). \end{aligned}$$

10.51. (a) By independence we have

$$\begin{aligned} P(Z \in [-1,1], X = 3) &= P(Z \in [-1,1]) P(X = 3) \\ &= (\Phi(1) - \Phi(-1)) \binom{n}{3} p^3 (1-p)^{n-3} = (2\Phi(1) - 1) p^3 (1-p)^{n-3}. \end{aligned}$$

(b) We have $P(Y < 1|X = 3) = \frac{P(Y<1, X=3)}{P(X=3)}$ and

$$\begin{aligned} P(Y < 1, X = 3) &= P(X + Z < 1, X = 3) = P(3 + Z < 1, X = 3) \\ &= P(Z < -2, X = 3) = P(Z < -2) P(X = 3). \end{aligned}$$

Thus

$$\begin{aligned} P(Y < 1|X = 3) &= \frac{P(Y < 1, X = 3)}{P(X = 3)} = \frac{P(Z < -2)P(X = 3)}{P(X = 3)} \\ &= P(Z < -2) = \Phi(-2). \end{aligned}$$

(c) We can condition on X to get

$$P(Y < x) = \sum_{k=0}^n P(Y < x | X = k) \binom{n}{k} p^k (1-p)^{n-k}.$$

Using the same argument as in part (b) we get

$$\begin{aligned} P(Y < x|X = k) &= \frac{P(Z + X < x, X = k)}{P(X = k)} = \frac{P(Z < x - k)P(X = k)}{P(X = k)} \\ &= P(Z < x - k) = \Phi(x - k). \end{aligned}$$

Thus

$$P(Y < x) = \sum_{k=0}^{n} \Phi(x-k) \binom{n}{k} p^k (1-p)^{n-k}.$$

10.52. (a)

$$p_{Y|X}(y|k) = \frac{p_{X,Y}(k,y)}{p_X(k)} = \frac{p_{X|Y}(k|y)\, p_Y(y)}{p_{X|Y}(k|0)\, p_Y(0) + p_{X|Y}(k|1)\, p_Y(1)}$$

$$= \begin{cases} \dfrac{\frac{1}{2}\cdot e^{-2}\frac{2^k}{k!}}{\frac{1}{2}\cdot e^{-2}\frac{2^k}{k!} + \frac{1}{2}\cdot e^{-3}\frac{3^k}{k!}} = \dfrac{2^k e^{-2}}{2^k e^{-2} + 3^k e^{-3}}\,, & y = 0 \\ \dfrac{\frac{1}{2}\cdot e^{-3}\frac{3^k}{k!}}{\frac{1}{2}\cdot e^{-2}\frac{2^k}{k!} + \frac{1}{2}\cdot e^{-3}\frac{3^k}{k!}} = \dfrac{3^k e^{-3}}{2^k e^{-2} + 3^k e^{-3}}\,, & y = 1. \end{cases}$$

(b)

$$\lim_{k\to\infty} p_{Y|X}(1|k) = \lim_{k\to\infty} \frac{3^k e^{-3}}{2^k e^{-2} + 3^k e^{-3}} = \lim_{k\to\infty} \frac{1}{(\frac{2}{3})^k e + 1} = 1.$$

Since $Y = 1$ makes X typically larger than $Y = 0$ does, a very large X makes $Y = 1$ overwhelmingly likelier than $Y = 0$.

10.53. To see that X_2 and X_3 are not independent, observe the following. Both X_2 and X_3 can take the value $(0,1)$ with positive probability, but

$$P(X_2 = (0,1), X_3 = (0,1)) = 0 \neq P(X_2 = (0,1))P(X_3 = (0,1)) > 0.$$

Now we show that $X_2, X_3, X_4, \dots$ is a Markov chain. Suppose that we have a sequence $x_2, x_3, \dots, x_n$ from the set $\{(0,0),(0,1),(1,0),(1,1)\}$ so that $P(X_2 = x_2, X_3 = x_3, \dots . X_n = x_n) > 0$. Denote the two coordinates of x_i by a_i and b_i. Then we must have $b_k = a_{k+1}$ for $k = 2, 3, \dots, n-1$ and

$$P(X_2 = x_2, X_3 = x_3, \dots . X_n = x_n) = P(Y_1 = a_1, Y_2 = a_2, \dots, Y_{n-1} = a_{n-1}, Y_n = b_n).$$

Let

$$x_{n+1} = (a_{n+1}, b_{n+1}) \in \{(0,0),(0,1),(1,0),(1,1)\}.$$

Then

$$\begin{aligned} P(X_{n+1} = x_{n+1}|X_n = x_n) &= P(X_{n+1} = (a_{n+1}, b_{n+1})|X_n = (a_n, b_n))) \\ &= \frac{P(X_{n+1} = (a_{n+1}, b_{n+1}), X_n = (a_n, b_n))}{P(X_n = (a_n, b_n))} \\ &= \frac{P(Y_n = a_{n+1}, Y_{n+1} = b_{n+1}, Y_{n-1} = a_n, Y_n = b_n)}{P(Y_{n-1} = a_n, Y_n = b_n)} \\ &= \begin{cases} P(Y_{n+1} = b_{n+1}), & \text{if } a_{n+1} = b_n \\ 0, & \text{if } a_{n+1} \neq b_n. \end{cases} \end{aligned}$$

Now consider the conditional distribution of X_{n+1} with respect to the full past:

$$\begin{aligned}
&P(X_{n+1} = x_{n+1} \,|\, X_2 = x_2, \ldots, X_n = x_n) \\
&\quad = \frac{P(X_2 = x_2, \ldots, X_n = x_n, X_{n+1} = x_{n+1})}{P(X_2 = x_2, \ldots, X_n = x_n)} \\
&\quad = \frac{P(Y_1 = a_1, Y_2 = a_2, \ldots, Y_{n-1} = a_{n-1}, Y_n = b_n, Y_n = a_{n+1}, Y_{n+1} = b_{n+1})}{P(Y_1 = a_1, Y_2 = a_2, \ldots, Y_{n-1} = a_{n-1}, Y_n = b_n)}.
\end{aligned}$$

This ratio is zero if $b_n \neq a_{n+1}$, and if $b_n = a_{n+1}$ then it becomes $P(Y_{n+1} = b_{n+1})$ by the independence of the Y_k. Thus

$$P(X_{n+1} = x_{n+1} | X_n = x_n) = P(X_{n+1} = x_{n+1} \,|\, X_2 = x_2, \ldots, X_n = x_n)$$

which shows that the process is a Markov chain.

Solutions to the Appendix

Appendix B.

B.1.

(a) We want to collect the elements which are either (in A and in B, but not in C), or (in A and in C, but not in B), or (in B and in C, but not in A).

The elements described by the first parentheses are given by the set ABC^c (or equivalently $A \cap B \cap C^c$). The set in the second parentheses is ACB^c while the third is BCA^c. By taking the union of these sets we have exactly the elements of D:

$$D = ABC^c \cup ACB^c \cup BCA^c.$$

(b) This is similar to part (a), but now we should also include the elements that are in all three sets. These are exactly the elements of $ABC = A \cap B \cap C$, so by taking the union of this set with the answer of (a) we get the required result.

$$D = ABC^c \cup BCA^c \cup ACB^c \cup ABC.$$

Alternately, we can write simply

$$D = AB \cup AC \cup BC = (A \cap B) \cup (A \cap C) \cup (B \cap C).$$

In this last expression there can be overlap between the members of the union but it is still a legitimate way to express the set D.

B.2. (a) $A \cap B \cap C$

(b) $A \cap (B \cup C)^c$ which can also be written as $A \cap B^c \cap C^c$.

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

(d) $A \cap B \cap C^c$

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

B.3.

(a) $B \setminus A = \{15, 25, 35, 45, 51, 53, 55, 57, 59, 65, 75, 85, 95\}$.

(b) $A \cap B \cap C^c = \{50, 52, 54, 56, 58\} \cap C^c = \{50, 52, 56, 58\}$.

(c) Observe that a two-digit number $10a + b$ is a multiple of 3 if and only if $a + b$ is a multiple of 3: $10a + b = 3k \iff a + b = 3(k - 3a)$. Thus $C \cap D = \varnothing$ because the sum of the digits cannot be both 10 and a multiple of 3. Consequently $((A \setminus D) \cup B) \cap (C \cap D) = \varnothing$.

B.4. We have $\omega \in \left(\bigcap_i A_i \right)^c$ if and only if $\omega \notin \left(\bigcap_i A_i \right)$. An element ω is not in the intersection of the sets A_i if and only if there is at least one i with $\omega \notin A_i$, which is the same as $\omega \in A_i^c$. But $\omega \in A_i^c$ for one of the i if and only if $\omega \in \bigcup_i A_i^c$. This proves the identity.

B.5. (a) The elements in $A \triangle B$ are either elements of A, but not B or elements of B, but not A. Thus we have $A \triangle B = AB^c \cup A^c B$.

(b) First note that for any two sets $E, F \subset \Omega$ we have

$$\Omega = EF \cup E^c F \cup EF^c \cup E^c F^c$$

where the four sets on the right are disjoint. From this and part (a) it follows that

$$(E \triangle F)^c = (EF^c \cup E^c F)^c = EF \cup E^c F^c.$$

This gives

$$\begin{aligned} A \triangle (B \triangle C) &= A(B \triangle C)^c \cup A^c (B \triangle C) \\ &= A(BC \cup B^c C^c) \cup A^c (BC^c \cup B^c C) \\ &= ABC \cup AB^c C^c \cup A^c BC^c \cup A^c B^c C. \end{aligned}$$

and

$$\begin{aligned} (A \triangle B) \triangle C &= (A \triangle B) C^c \cup (A \triangle B)^c C \\ &= (AB^c \cup A^c B) C^c \cup (AB \cup A^c B^c) C \\ &= AB^c C \cup A^c BC^c \cup ABC \cup A^c B^c C \end{aligned}$$

which shows that the two sets are the same.

B.6. (a) We have $\omega \in E = A \cap B$ if and only if $\omega \in A$ and $\omega \in B$. Similarly, $\omega \in E = A \cap B^c$ if and only if $\omega \in A$ and $\omega \in B^c$. This shows that we cannot have $\omega \in E$ and $\omega \in F$ the same time: this would imply $\omega \in B$ and $\omega \in B^c$ the same time, which cannot happen. Thus the intersection of E and F must be the empty set.

(b) We first show that if $\omega \in A$ then either $\omega \in E$ or $\omega \in F$, this shows that $\omega \in E \cup F$. We either have $\omega \in B$ or $\omega \in B^c$. If $\omega \in B$ then ω is an element of both A and B, and hence an element of $E = A \cap B$. If $\omega \in B^c$ then ω is an element of A and B^c, and hence $F = A \cap B^c$. This proves that if $\omega \in A$ then $\omega \in E \cup F$.

On the other hand, if $\omega \in E \cup F$ then we must have either $\omega \in E = A \cap B$ or $\omega \in F = A \cap B^c$. In both cases $\omega \in A$. Thus $\omega \in E \cup F$ implies $\omega \in A$.

This proves that the elements of A are exactly the elements of $E \cup F$, and thus $A = E \cup F$.

B.7. (a) Yes. One possibility is $D = CB^c$.

(b) Note that whenever 2 appears in one of the sets (A or B) then 6 is there as well, and vice versa. This means that we cannot separate these two elements with the set operations, whatever set expression we come up with, the result will either have both 2 and 6 or neither. Thus we cannot get $\{2, 4\}$ as the result.

Appendix C.

C.1. We can construct all allowed license plates using the following procedure: we choose one of the 26 letters to be the first letter, then one of the remaining 25 letters to be the 2nd, and then one of the remaining 24 letters to be the third letter. Similarly, we choose one of the 10 digits to be the first digit, then choose the second and third digits (with 9 and 8 possible choices). By the multiplication principle this gives us $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 = 11,232,000$ different license plates.

C.2. There are 26 choices for each of the three letters. Further, there are 10 choices for each of the digits. Thus, there are a total of $26^3 \cdot 10^3$ ways to construct license plates when any combination is allowed. However, there are $26^3 \cdot 1^3$ ways to construct license plates with three zeros (we have 26 choices for each of the three letters, and exactly one choice for each number). Subtracting those off gives a solution of

$$26^3(10^3 - 1) = 17{,}558{,}424.$$

Another way to get the same answer is as follows: we have 26^3 choices for the three letters and 999 choices for the three digits (10^3 minus the three zero case) which gives again $26^3 \cdot 999 = 17{,}558{,}424$.

C.3. There are 25 license plates that differ from $UWU144$ only at the first position (as there are 25 other letters we can choose there), the same is true for the second and third positions. There are 9 license plates that differ from $UWU144$ only at the fourth position (there are 9 other possible digits), and the same is true for the 5th and 6th positions. This gives $3 \cdot 25 + 3 \cdot 9 = 102$ possibilities.

C.4. We can arrange the 6 letters in $6! = 120$ different orders, so the answer is 120.

C.5. Imagine that we differentiate between the two Ps: there is a P_1 and a P_2. Then we could order the five letters $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ different ways. Each ordering of the letters gives a word, but we counted each word twice (as the two Ps can be in two different orders). Thus we can construct $\frac{120}{2} = 60$ different words.

C.6. (a) This is the choice of a subset of size 5 from a set of size 90, hence we have $\binom{90}{5} = 43,949,268$ outcomes.

If you want to first choose the numbers in order, then first you produce an ordered list of 5 numbers: $90 \cdot 89 \cdot 88 \cdot 87 \cdot 86$ outcomes. But now each set of 5 numbers is counted $5!$ times (in each of its orderings). Thus the answer is again

$$\frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}{5!} = \binom{90}{5} = 43,949,268.$$

(b) If 1 is forced into the set, then we choose the remaining 4 winning numbers from the 89 numbers $\{2, 3, \dots, 90\}$. We can do that $\binom{89}{4} = 2,441,626$ different ways, this is the number of outcomes with 1 appearing among the five numbers.

(c) These outcomes can be produced by first picking 2 numbers from the set $\{1, 2, \dots, 49\}$ and 3 numbers from $\{61, 62, \dots, 90\}$. By the multiplication principle of counting there are $\binom{49}{2}\binom{30}{3} = 4,774,560$ ways we can do that, so that is the number of outcomes. *Note:* It does not matter in what order the steps are performed, or you can imagine them performed simultaneously.

(d) Here are two possible ways of solving this problem:

(i) First choose a set of 5 distinct second digits from the set $\{0, 1, 2, \ldots, 9\}$: $\binom{10}{5}$ choices. The for each last digit in turn, choose a first digit. There are always 9 choices: if the last digit is 0, then the choices for the first digit are $\{1, 2, \ldots, 9\}$, while if the last digit is in the range $1 - 9$ then the choices for the first digit are $\{0, 1, \ldots, 8\}$. By the multiplication principle of counting there are $\binom{10}{5}9^5 = 14,880,348$ outcomes.

(ii) Here is another presentation of the same idea: divide the 90 numbers into subsets according to last digit:

$$A_0 = \{10, 20, 30, \ldots, 90\}, \quad A_1 = \{1, 11, 21, \ldots, 81\},$$
$$A_2 = \{2, 12, 22, \ldots, 82\}, \ldots, \quad A_9 = \{9, 19, 29, \ldots, 89\}.$$

The rule is that at most 1 number comes from each A_k. Hence first choose 5 subsets $A_{k_1}, A_{k_2}, \ldots, A_{k_5}$ out of the ten possible: $\binom{10}{5}$ choices. Then choose one number from the 9 in each set A_{k_j}: 9^5 total possibilities. By the multiplication principle $\binom{10}{5}9^5$ outcomes.

C.7. Denote the four players by A, B, C and D. Note that if we choose the partner of A (which we can do three possible ways) then this will determine the other team as well. Thus there are 3 ways to set up the doubles match.

C.8. (a) Once we choose the opponent of team A, the whole tournament is set up. Thus there are 3 ways to set up the tournament.

(b) In the tournament there are three games, each have two possible outcomes. Thus for a given set up we have $2^3 = 8$ outcomes, and since there are 3 ways to set up the tournament this gives $8 \cdot 3 = 24$ possible outcomes for the tournament.

C.9. (a) In order to produce all pairs we can first choose the rank of the pair (2, 3, ..., J, Q, K or A), which gives 13 choices. Then we choose the two cards from the 4 possibilities for that rank (for example, if the rank is K then we choose 2 cards from ♡ K, ♣ K, ♢ K, ♠ K), which gives $\binom{4}{2}$ choices. By the multiplication principle we have altogether $13 \cdot \binom{4}{2} = 78$ choices.

(b) To produce two cards with the same suit we first choose the suit (4 choices) and then choose the two cards from the 13 possibilities with the given suit ($\binom{13}{2} = 78$ choices). By the multiplication principle the result is $4 \cdot \binom{13}{2} = 312$.

(c) To produce a suited connector, first choose the suit (4 choices) then one of the 13 neighboring pairs. This gives $4 \cdot 13 = 52$ choices.

C.10. (a) We can construct a hand with two pairs the following way. First we choose the ranks of the repeated ranks, we can do that $\binom{13}{2}$ different ways. For the lower ranked pair we can choose the two suits $\binom{4}{2}$ ways, and the for the larger ranked pair we again have $\binom{4}{2}$ choices for the suits. The fifth card must have a different rank than the two pairs we have already chosen, there are $52 - 2 \cdot 4 = 44$ choices for that. This gives $\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 44 = 123552$ choices.

(b) We can choose the rank of the three cards of the same rank 13 ways, and the three suits $\binom{4}{3} = 4$ ways. The other two cards have different ranks, we can choose those ranks $\binom{12}{2}$ different ways. For each of these two ranks we can choose the suit four ways, which gives 4^2 choices. This gives $13 \cdot 4 \cdot \binom{12}{2} \cdot 4^2 = 54912$ possible three of a kinds.

(c) We can choose the rank of the starting card 10 ways (A, 2, ..., 10) if we want five cards in sequential order, this identifies the ranks of the other cards. For each of the 5 ranks we can choose the suit 4 ways. But for each sequence we have four cases where all five cards are of the same suit, we have to remove these from the 4^5 possibilities. This gives $10 \cdot (4^5 - 4) = 10200$ choices for a straight.

(d) The suit of the five cards can be chosen 4 ways. There are $\binom{13}{5}$ ways to choose five cards, but we have to remove the cases when these are in sequential order. We can choose the rank of the starting card 10 ways (A, 2, ..., 10) if we want five cards in sequential order. This gives $4 \cdot (\binom{13}{5} - 10) = 5108$ choices for a flush.

(e) We can construct a full house the following way. First choose the rank that appears three times (13 choices), and then the rank appearing twice (there are 12 remaining choices). Then choose the three suits for the rank appearing three times ($\binom{4}{3} = 4$ choices) and the suits for the other two cards ($\binom{4}{2} = 6$ choices). In each step the number of choices does not depend on the previous decisions, so we can multiply these together to get the number of ways we can get a full house: $13 \cdot 12 \cdot 4 \cdot 6 = 3744$.

(f) We can choose the rank of the 4 times repeated card 13 ways, and the fifth card 48 ways (since we have 48 other cards), this gives $13 \cdot 48 = 624$ poker hands with four of a kind.

(g) We can choose the value of the starting card 10 ways (A, 2, ..., 10), and the suit 4 ways, which gives $10 \cdot 4 = 40$ poker hands with straight flush. (Often the case when the starting card is a 10 is called a royal flush. There are 4 such hands.)

C.11. From the definition:

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

Here is another way to prove the identity. Assume that in a class there are n students, and one of them is called Dana. There are $\binom{n}{k}$ ways to choose a team of k students from the class. When we choose the team there are two possibilities: Dana is either on the team or not. There are $\binom{n-1}{k}$ ways to choose the team if we cannot include Dana. There are $\binom{n-1}{k-1}$ ways to choose the team if we have to include Dana. These two numbers must add up to the total number of ways we can select the team, which gives the identity.

C.12. (a) We have to divide up the remaining 48 (non-ace) cards into four groups so that the first group has 9 cards, and the second, third and fourth groups have 13 cards. This can be done by $\binom{48}{9,13,13,13} = \frac{48!}{9!(13!)^3}$ different ways.

(b) To describe such a configuration we just have to assign a different suit for each player. This can be done $4! = 24$ different ways.

(c) We can construct such a configuration by first choosing the 13 cards of Player 4 (there are 39 non-$\heartsuit$ cards, so we can do that $\binom{39}{13}$ different ways), then choosing the 13 cards of Player 3 (there are 26 non-$\heartsuit$ cards remaining, so we can do that $\binom{26}{13}$ different ways), and then choosing the 13 cards of Player 2 out of the remaining 26 cards (out of which 13 are $\heartsuit$), we can do that $\binom{26}{13}$ different ways. (Player 1 gets the remaining 13 cards.) Since the number of choices in each step do not depend on the outcomes of the previous choices, the total number of configurations is the product $\binom{39}{13} \cdot \binom{26}{13}\binom{26}{13} = \frac{39!26!}{(13!)^5}$.

C.13. Label the sides of the square with north, west, south and east. For any coloring we can always rotate the square in a unique way so that the red side is the north side. We can choose the colors of the other two sides (W, S, E) $3 \cdot 2 \cdot 1 = 6$ different ways, which means that there are 6 different colorings.

C.14. We will use one color twice and the other colors once. Let us first count the number of ways we can color the sides so there are two red sides. Label the sides of the square with north, west, south, east. We can rotate any coloring uniquely so the (only) blue side is the north side. The yellow side can be chosen now three different ways (from the other three positions), and once we have that, the positions of the red sides are determined. Thus there are three ways we can color the sides of the square so that there are 2 red, 1 blue and 1 yellow side and colorings that can be rotated to each other are treated the same. Similarly, we have three colorings with 2 blue, 1 red and 1 yellow side, and three colorings with 2 yellow, 1 red and 1 blue side. This gives 9 possible colorings.

C.15. Imagine that we place the colored cube on the table so that one of the faces is facing us. There are 6 different colorings of the cube where the red and blue faces are on the opposite sides. Indeed: for such a coloring we can always rotate the cube uniquely so that it rests on the red face and the yellow face is facing us (with blue on the top). Now we can choose the colors of the other three faces $3 \cdot 2 \cdot 1$ different ways, which gives us 6 such colorings.
If the red and the blue faces are next to each other then we can always rotate the cube uniquely so it rests on the red face and the blue face is facing us. The remaining four faces can be colored $4 \cdot 3 \cdot 2 \cdot 1$ different ways, thus we have 24 such colorings.

This gives $24 + 6 = 30$ colorings all together.

C.16. Number the bead positions clockwise with $0, 1, \dots, 17$. We can choose the positions of the 7 green beads out of the 18 possibilities $\binom{18}{7}$ different ways. However this way we over counted the number of necklaces, as we counted the rotated versions of each necklace separately. We will show that each necklace was counted exactly 18 times. A given necklace can be rotated 18 different ways (with the first position going into one of the eighteen possible positions), we just have to check that two different rotations cannot give the same set of positions for the green beads. We prove this by contradiction. Assume that we have seven different positions $g_1, \ldots, g_7 \in \{0, 1, \ldots, 17\}$ so that if we rotate them by $0 < d < 18$ then we get the same set of positions. It can be shown that this can only happen if each two neighboring position are separated by the same number of steps. But 7 does not divide 16, so this is impossible. Thus all 18 rotations of a necklace were counted separately, which means that the number of necklaces is $\frac{1}{18}\binom{18}{7} = 1768$.

C.17. Suppose that in a class there are n girls and n boys. There are $\binom{2n}{n}$ different ways we can choose a team of n students out of this class of $2n$. For any $0 \le k \le n$ there are $\binom{n}{k} \cdot \binom{n}{n-k}$ ways to choose the team so that there are exactly k girls and $n-k$ boys chosen. For $0 \le k \le n$ we have $\binom{n}{n-k} = \binom{n}{k}$ and thus $\binom{n}{k} \cdot \binom{n}{n-k} = \binom{n}{k}^2$.

By considering the possible values of the number of girls in the team we now get the identity

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

C.18. If $x = -1$ then the inequality is $0 \ge 1 - n$ which certainly holds.

Now assume $x > -1$. For $n = 1$ both sides are equal to $1+x$, so the inequality is true. Assume now that the inequality holds for some positive integer n, we need to show that it holds for $n+1$ as well. By our induction assumption $(1+x)^n \ge 1+nx$, and because $x > -1$, we have $1 + x > 0$. Hence we can multiply both sides of the previous inequality with $1 + x$ to get

$$(1+x)^{n+1} \ge (1+nx)(1+x) = 1 + (n+1)x + nx^2.$$

Since $nx^2 \ge 0$ we get $(1+x)^{n+1} \ge 1 + (n+1)x$ which proves the induction step, and finishes the proof.

C.19. Let $a_n = 11^n - 6$. We have $a_1 = 5$, which is divisible by 5. Now assume that for some positive integer n the number a_n is divisible by 5. We have

$$a_{n+1} = 11^{n+1} - 6 = 11(a_n + 6) - 6 = 11a_n + 60.$$

If $\frac{a_n}{5}$ is an integer then $\frac{a_{n+1}}{5} = 11\frac{a_n}{5} + 12$ is also an integer. This shows the induction step, which finishes the proof.

C.20. By checking the first couple of values of n we see that

$$2^1 < 4 \cdot 1, \qquad 2^2 < 4 \cdot 2, \qquad 2^3 < 4 \cdot 3, \qquad 2^4 = 4 \cdot 4.$$

We will show that for all $n \ge 4$ we have $2^n \ge 4n$. This certainly holds for $n = 4$. Now assume that it holds for some integer $n \ge 4$, we will show that it also holds for $n + 1$. Multiplying both sides of the inequality $2^n \ge 4n$ (which we assumed to be true) by 2 we get

$$2^{n+1} \ge 8n.$$

But $8n = 4(n+1) + 4(n-1) > 4(n+1)$ if $n \ge 4$. Thus $2^{n+1} \ge 4(n+1)$, which finishes the proof.

Appendix D.

D.1. We can separate the terms into two sums:

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

Note that in the first sum we add n times the constant term n, so the sum is equal to n^2. The second sum is just twice the sum (D.6), so its value is $n(n+1)$. Thus

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

D.2. For any fixed $i \geq 1$ we have $\sum_{j=1}^{\infty} a_{i,j} = a_{i,i} + a_{i,i+1} = 1 - 1 = 0$. Thus $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{i,j} = 0$.

If we fix $j \geq 1$ then

$$\sum_{i=1}^{\infty} a_{i,j} = \begin{cases} a_{1,1} = 1, & \text{if } j = 1, \\ a_{j-1,j} + a_{j,j} = -1 + 1 = 0, & \text{if } j > 1. \end{cases}$$

Thus $\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{i,j} = 1$. This shows that for this particular choice of numbers $a_{i,j}$ we have

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{i,j} \neq \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{i,j} = 1.$$

D.3. (a) Evaluating the sum on the inside first using (D.6):

$$\sum_{k=1}^{n}\sum_{\ell=1}^{k} \ell = \sum_{k=1}^{n} \frac{k(k+1)}{2} = \sum_{k=1}^{n}\left(\frac{1}{2}k^2 + \frac{1}{2}k\right).$$

Separating the sum in two parts and then using (D.6) and (D.7):

$$\begin{aligned} \sum_{k=1}^{n}\left(\frac{1}{2}k^2 + \frac{1}{2}k\right) &= \frac{1}{2}\sum_{k=1}^{n} k^2 + \frac{1}{2}\sum_{k=1}^{n} k \\ &= \frac{1}{2}\cdot\frac{n(n+1)(2n+1)}{6} + \frac{1}{2}\cdot\frac{n(n+1)}{2} \\ &= \frac{(n(n+1)}{12}\cdot(2n+1+3) = \frac{n^3}{6} + \frac{n^2}{2} + \frac{n}{3}. \end{aligned}$$

(b) Since the sum on the inside has k terms that are all equal to k we get

$$\sum_{k=1}^{n}\sum_{\ell=1}^{k} k = \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n.$$

(c) Separating the sum into three parts:

$$\sum_{k=1}^{n}\sum_{\ell=1}^{k}(7+2k+\ell) = \sum_{k=1}^{n}\sum_{\ell=1}^{k} 7 + 2\sum_{k=1}^{n}\sum_{\ell=1}^{k} k + \sum_{k=1}^{n}\sum_{\ell=1}^{k} \ell.$$

The second and third sums can be evaluated using parts (a) and (b). The first sum is

$$\sum_{k=1}^{n}\sum_{\ell=1}^{k} 7 = \sum_{k=1}^{n} 7k = \frac{7n(n+1)}{2} = \frac{7}{2}n^2 + \frac{7}{2}n.$$

Thus we get

$$\begin{aligned}\sum_{k=1}^{n}\sum_{\ell=1}^{k}(7+2k+\ell) &= \frac{7}{2}n^2 + \frac{7}{2}n + 2\cdot\left(\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n\right) + \frac{n^3}{6} + \frac{n^2}{2} + \frac{n}{3}\\ &= \frac{5}{6}n^3 + 5n^2 + \frac{25}{6}n.\end{aligned}$$

D.4. $\sum_{j=i}^{n} j$ is the sum of the arithmetic progression $i, i+1, \ldots, n$ which has $n-i+1$ elements, so its value is $(n-i+1)\frac{n+i}{2}$. Thus

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=i}^{n} j &= \sum_{i=1}^{n}(n-i+1)\frac{n+i}{2} = \sum_{i=1}^{n}\frac{1}{2}\left(-i^2 + i + n^2 + n\right)\\ &= -\frac{1}{2}\sum_{i=1}^{n} i^{+}\frac{1}{2}\sum_{i=1}^{n} i + \frac{1}{2}\sum_{i=1}^{n}(n^2+n).\end{aligned}$$

The terms in the last sum do not depend on i, so

$$\frac{1}{2}\sum_{i=1}^{n}(n^2+n) = \frac{1}{2}(n^2+n)n = \frac{n^2(n+1)}{2}.$$

The first and second sums can be computed using the identities (D.6) and (D.7):

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^{n} i^2 &= \frac{n(n+1)(2n+1)}{12}\\ \frac{1}{2}\sum_{i=1}^{n} i &= \frac{n(n+1)}{4}.\end{aligned}$$

Collecting all the terms:

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=i}^{n} j &= -\frac{n(n+1)(2n+1)}{12} + \frac{n(n+1)}{2} + \frac{n^2(n+1)}{2}\\ &= \frac{n(n+1)}{12}(-(2n+1) + 3 + 6n) = \frac{n(n+1)(2n+1)}{6}.\end{aligned}$$

Here is a quicker solution using the exchange of sums. In the double sum we have $1 \le i \le j \le n$. If we switch the order of the summation, then i will go from 1 to j, and then j will go from 1 to n:

$$\sum_{i=1}^{n}\sum_{j=i}^{n} j = \sum_{j=1}^{n}\sum_{i=1}^{j} j.$$

(The switching of the order of the summation is justified because we have a finite sum.) The inside sum is easy to evaluate because the summand does not depend on i: $\sum_{i=1}^{j} j = j \cdot j = j^2$. Then

$$\sum_{j=1}^{n}\sum_{i=1}^{j} j = \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6},$$

by (D.7).

D.5. (a) From (D.1) we have

$$\sum_{j=i}^{\infty} x^j = x^i + x^{i+1} + x^{i+2} + \cdots = x^i \sum_{n=0}^{\infty} x^n = \frac{x^i}{1-x}.$$

Thus

$$\begin{aligned}\sum_{i=1}^{\infty}\sum_{j=i}^{\infty} x^j &= \sum_{i=1}^{\infty} \frac{x^i}{1-x} = \frac{x}{1-x}(1 + x + x^2 + \dots) \\ &= \frac{x}{1-x}\sum_{n=0}^{\infty} x^n = \frac{x}{1-x}\cdot\frac{1}{1-x} = \frac{x}{(1-x)^2}.\end{aligned}$$

(b) Using the hint we can write

$$\sum_{k=1}^{\infty} kx^k = \sum_{k=1}^{\infty}\sum_{j=1}^{k} x^k.$$

In the sum we have all k, j with $1 \le j \le k$. Thus if we switch the order of summation then we first have k going from j to ∞ and then j going from 1 to ∞:

$$\sum_{k=1}^{\infty}\sum_{j=1}^{k} x^k = \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} x^k.$$

This is exactly the sum that we computed in part (a), which shows that the answer is again $\frac{x}{(1-x)^2}$. The fact that we can switch the order of the summation follows from the fact that the double sum in (a) is finite even if we put absolute values around each term.

D.6. We use induction. For $n = 1$ the two sides are equal: $1^2 = \frac{1\cdot 2\cdot(2\cdot 1+1)}{6}$. Assume that the identity holds for $n \ge 1$, we will show that it also holds for $n+1$. By the induction hypothesis

$$\begin{aligned}1^2 + 2^2 + \cdots + n^2 + (n+1)^2 &= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \\ &= \frac{n+1}{6}\left(n(2n+1) + 6(n+1)\right) = \frac{n+1}{6}\left(2n^2 + 7n + 6\right) \\ &= \frac{(n+1)(2n^2+7n+6)}{6} = \frac{(n+1)(n+2)(2n+3)}{6}.\end{aligned}$$

The last formula is exactly the right side of (D.7) for $n+1$ in place of n, which proves the induction step and the statement.

D.7. We prove the identity by induction. The identity holds for $n = 1$. Assume that it holds for $n \ge 1$, we will show that it also holds for $n+1$. By the induction hypothesis

$$\begin{aligned}1^3+2^3+\dots+n^3+(n+1)^3 &= \frac{n^2(n+1)^2}{4}+(n+1)^3\\ &= (n+1)^2\left(\frac{n^2}{4}+n+1\right) = (n+1)^2\frac{n^2+4n+4}{4}\\ &= \frac{(n+1)^2(n+2)^2}{4}.\end{aligned}$$

This is exactly (D.8) stated for $n+1$, which completes the proof.

D.8. First note that both sums have finitely many terms, because $\binom{n}{k}=0$ if $k>n$.

If we move every term to the left side then we get

$$\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\binom{n}{3}+\binom{n}{4}-+\dots$$

We would like to show that this expression is zero. Note that the alternating signs can be expressed using powers of -1, hence the expression above is equal to $\sum_{k=0}^{n}(-1)^k\binom{n}{k}=\sum_{k=0}^{n}(-1)^k\cdot 1^{n-k}\binom{n}{k}$. But this is exactly equal to $(-1+1)^n=0^n=0$ by the binomial theorem. Hence $\sum_{k=0}^{n}(-1)^k\binom{n}{k}=0$ and

$$\binom{n}{0}+\binom{n}{2}+\binom{n}{4}+\dots=\binom{n}{1}+\binom{n}{3}+\binom{n}{5}+\dots$$

Using the binomial theorem for $(1+1)^n$ we get $\sum_{k=0}^{n}\binom{n}{k}=2^n$. Introducing

$$\begin{aligned}a_n &= \binom{n}{0}+\binom{n}{2}+\binom{n}{4}+\dots\\ b_n &= \binom{n}{1}+\binom{n}{3}+\binom{n}{5}+\dots,\end{aligned}$$

we have just shown that $a_n=b_n$ and $a_n+b_n=2^n$. This yields $a_n=b_n=2^{n-1}$. But a_n is exactly the number of even subsets of a set of size n (as it counts the number of subsets with $0,2,4\dots$ elements), thus the number of even subsets is 2^{n-1}. Similarly, the number of odd subsets is also 2^{n-1}.

D.9. We would like to show (D.10) for all x,y and $n\ge 1$. For $n=1$ the two sides are equal. Assume that the statement holds for n, we will prove that it also holds for $n+1$. By the induction hypothesis

$$\begin{aligned}(x+y)^{n+1} &= (x+y)\cdot(x+y)^n=(x+y)\sum_{k=0}^{n}\binom{n}{k}x^ky^{n-k}\\ &= \sum_{k=0}^{n}\binom{n}{k}x^ky^{n-k}(x+y)=\sum_{k=0}^{n}\binom{n}{k}x^{k+1}y^{n-k}+\sum_{k=0}^{n}\binom{n}{k}x^ky^{n-k+1}.\end{aligned}$$

Shifting the index in the first sum gives

$$\begin{aligned}\sum_{k=0}^{n}\binom{n}{k}x^{k+1}y^{n-k}+\sum_{k=0}^{n}\binom{n}{k}x^ky^{n-k+1} &= \sum_{k=1}^{n+1}\binom{n}{k-1}x^ky^{n+1-k}+\sum_{k=0}^{n}\binom{n}{k}x^ky^{n-k+1}\\ &= x^{n+1}+y^{n+1}+\sum_{k=1}^{n}\left(\binom{n}{k-1}+\binom{n}{k}\right)x^ky^{n+1-k}\end{aligned}$$

where in the last step we separated the last and first term of the two sums. Using Exercise C.11 we get that $\left(\binom{n}{k-1}+\binom{n}{k}\right)=\binom{n+1}{k}$ which gives

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

which is exactly what we wanted to prove.

D.10. For $r=2$ the statement is the binomial theorem, which we have proved in Fact D.2. Assume that for a certain $r\geq 2$ the statement is true, we will prove that it holds for $r+1$ as well.

We start by noting that

$$(x_1+x_2+\cdots+x_{r+1})^n=(x_1+x_2+\cdots+(x_r+x_{r+1}))^n.$$

We can use our induction assumption for the r numbers $x_1,x_2,\dots,x_{r-1},x_r+x_{r+1}$ to get

$$\begin{aligned}&(x_1+x_2+\cdots+(x_r+x_{r+1}))^n\\&\qquad=\sum_{\substack{k_1\geq 0,\,k_2\geq 0,\dots,\,k_r\geq 0\\ k_1+k_2+\cdots+k_r=n}}\binom{n}{k_1,k_2,\dots,k_r}x_1^{k_1}x_2^{k_2}\cdots x_{r-1}^{k_{r-1}}(x_r+x_{r+1})^{k_r}\end{aligned}$$

Using the binomial theorem for $(x_r+x_{r+1})^{k_r}$ gives

$$\begin{aligned}&(x_1+x_2+\cdots+(x_r+x_{r+1}))^n\\&\qquad=\sum_{\substack{k_1\geq 0,\,k_2\geq 0,\dots,\,k_r\geq 0\\ k_1+k_2+\cdots+k_r=n}}\sum_{j=0}^{k_r}\binom{n}{k_1,k_2,\dots,k_r}\binom{k_r}{j}x_1^{k_1}x_2^{k_2}\cdots x_{r-1}^{k_{r-1}}x_r^j x_{r+1})^{k_r-j}.\end{aligned}$$

Introducing the new notation $a=j, b=k_r-j$ we can rewrite the double sum as follows

$$\begin{aligned}&\sum_{\substack{k_1\geq 0,\,k_2\geq 0,\dots,\,k_r\geq 0\\ k_1+k_2+\cdots+k_r=n}}\sum_{j=0}^{k_r}\binom{n}{k_1,k_2,\dots,k_r}\binom{k_r}{j}x_1^{k_1}x_2^{k_2}\cdots x_{r-1}^{k_{r-1}}x_r^j x_{r+1})^{k_r-j}\\&=\sum_{\substack{k_1\geq 0,\,k_2\geq 0,\dots,\,k_{r-1}\geq 0,a\geq 0,b\geq 0\\ k_1+k_2+\cdots+k_{r-1}+a+b=n}}\binom{n}{k_1,k_2,\dots,k_{r-1},a+b}\binom{a+b}{a}x_1^{k_1}x_2^{k_2}\cdots x_{r-1}^{k_{r-1}}x_r^a x_{r+1})^b.\end{aligned}$$

Now note that

$$\begin{aligned}\binom{n}{k_1,k_2,\dots,k_{r-1},a+b}\binom{a+b}{a}&=\frac{n!}{k_1!k_2!\cdots k_{r-1}!(a+b)!}\cdot\frac{(a+b)!}{a!b!}\\&=\binom{n}{k_1,k_2,\dots,k_{r-1},a,b}.\end{aligned}$$

This means that

$$(x_1 + x_2 + \cdots + (x_r + x_{r+1}))^n$$

$$= \sum_{\substack{k_1 \geq 0,\, k_2 \geq 0, \ldots,\, k_{r-1} \geq 0, a \geq 0, b \geq 0 \\ k_1 + k_2 + \cdots + k_{r-1} + a + b = n}} \binom{n}{k_1, k_2, \ldots, k_{r-1}, a, b} x_1^{k_1} x_2^{k_2} \cdots x_{r-1}^{k_{r-1}} x_r^a x_{r+1})^b$$

which is exactly the statement we have to prove for $r+1$. This proves the induction step and the theorem.

D.11. This can be done similarly to Exercise D.9. We outline the proof for $r = 3$, the general case is similar (with more indices). We need to show that

$$(x_1 + x_2 + x_3)^n = \sum_{\substack{k_1 \geq 0,\, k_2 \geq 0, k_3 \geq 0 \\ k_1 + k_2 + k_3 = n}} \binom{n}{k_1, k_2, k_3} x_1^{k_1} x_2^{k_2} x_3^{k_3}.$$

For $n = 1$ the two sides are equal: the only possible triples (k_1, k_2, k_3) are $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ and these give the terms x_1, x_2 and x_3. Now assume that the equation holds for some n, we would like to show it for $n+1$. Take the equation for n and multiply both sides with $x_1+x_2+x_3$. Then on one side we get $(x_1+x_2+x_3)^{n+1}$, while the other side is

$$\sum_{\substack{k_1 \geq 0,\, k_2 \geq 0, k_3 \geq 0 \\ k_1 + k_2 + k_3 = n}} \binom{n}{k_1, k_2, k_3} \left(x_1^{k_1+1} x_2^{k_2} x_3^{k_3} + x_1^{k_1} x_2^{k_2+1} x_3^{k_3} + x_1^{k_1} x_2^{k_2} x_3^{k_3+1} \right).$$

The coefficient of $x_1^{a_1} x_2^{a_2} x_3^{a_3}$ for a given $0 \leq a_1, 0 \leq a_2, 0 \leq a_3$with $a_1+a_2+a_3 = n+1$ is equal to

$$\binom{n}{a_1 - 1, a_2, a_3} + \binom{n}{a_1, a_2 - 1, a_3} + \binom{n}{a_1, a_2, a_3 - 1}$$

which can be shown to be equal to $\binom{n+1}{a_1, a_2, a_3}$. (This is a generalization of Exercise D.9 and can be shown the same way.) But this means that

$$(x_1 + x_2 + x_3)^{n+1} = \sum_{\substack{k_1 \geq 0,\, k_2 \geq 0, k_3 \geq 0 \\ k_1 + k_2 + k_3 = n}} \binom{n}{k_1, k_2, k_3} \left(x_1^{k_1+1} x_2^{k_2} x_3^{k_3} + x_1^{k_1} x_2^{k_2+1} x_3^{k_3} + x_1^{k_1} x_2^{k_2} x_3^{k_3+1} \right)$$

$$= \sum_{\substack{a_1 \geq 0,\, a_2 \geq 0, a_3 \geq 0 \\ a_1 + a_2 + a_3 = n+1}} \binom{n+1}{a_1, a_2, a_3} x_1^{a_1} x_2^{a_2} x_3^{a_3},$$

which is exactly what we needed for the induction step.

D.12. Imagine that we expand all the parentheses in the product

$$(x_1 + \cdots + x_r)^n = (x_1 + \cdots + x_r)(x_1 + \cdots + x_r) \cdots (x_1 + \cdots + x_r).$$

Then each term in the resulting expansion will be of the form of $x_1^{k_1} \cdots x_r^{k_r}$ with $k_i \geq 0$ and $k_1 + \cdots + k_r = n$. This is because from each of the $(x_1 + \cdots + x_r)$ term we will pick exactly one of the x_i, and we have n factors in the end. Now we have to determine the coefficient of a the term $x_1^{k_1} \cdots x_r^{k_r}$ in the expansion for a given choice of $k_1, \ldots, k_r$ with $k_i \geq 0$ and $k_1 + \cdots + k_r = n$. In order to get such a term from the expansion we need to choose k_1 times x_1, k_2 times x_2 and so on. But the number of ways we can do that is exactly the multinomial coefficient $\binom{n}{k_1,k_2,\ldots,k_r}$. This proves the identity (D.11).