29

A Modern Introduction to Probability and Statistics

Full Solutions

February 24, 2006

29.1 Full solutions

2.1 Using the relation $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$, we obtain $\mathrm{P}(A \cup B) = 2/3 + 1/6 - 1/9 = 13/18$.

2.2 The event "at least one of E and F occurs" is the event $E \cup F$. Using the second DeMorgan's law we obtain: $\mathrm{P}(E^c \cap F^c) = \mathrm{P}((E \cup F)^c) = 1 - \mathrm{P}(E \cup F) = 1 - 3/4 = 1/4$.

2.3 By additivity we have $\mathrm{P}(D) = \mathrm{P}(C^c \cap D) + \mathrm{P}(C \cap D)$. Hence $0.4 = \mathrm{P}(C^c \cap D) + 0.2$. We see that $\mathrm{P}(C^c \cap D) = 0.2$. (We did not need the knowledge $\mathrm{P}(C) = 0.3$!)

2.4 The event "*only* A occurs and not B or C" is the event $\{A \cap B^c \cap C^c\}$. We then have using DeMorgan's law and additivity

$$\mathrm{P}(A \cap B^c \cap C^c) = \mathrm{P}(A \cap (B \cup C)^c) = \mathrm{P}(A \cup B \cup C) - \mathrm{P}(B \cup C)\,.$$

The answer is yes , because of $\mathrm{P}(B \cup C) = \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C)$

2.5 The crux is that $B \subset A$ implies $\mathrm{P}(A \cap B) = \mathrm{P}(B)$. Using additivity we obtain $\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^c) = \mathrm{P}(B) + \mathrm{P}(A \setminus B)$. Hence $\mathrm{P}(A \setminus B) = \mathrm{P}(A) - \mathrm{P}(B)$.

2.6 a Using the relation $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$, we obtain $3/4 = 1/3 + 1/2 - \mathrm{P}(A \cap B)$, yielding $\mathrm{P}(A \cap B) = 4/12 + 6/12 - 9/12 = 1/12$.

2.6 b Using DeMorgan's laws we get $\mathrm{P}(A^c \cup B^c) = \mathrm{P}((A \cap B)^c) = 1 - \mathrm{P}(A \cap B) = 11/12$.

2.7 $\mathrm{P}((A \cup B) \cap (A \cap B)^c) = 0.7$.

2.8 From the rule for the probability of a union we obtain $\mathrm{P}(D_1 \cup D_2) \leq \mathrm{P}(D_1) + \mathrm{P}(D_2) = 2 \cdot 10^{-6}$. Since $D_1 \cap D_2$ is contained in both D_1 and D_2, we obtain $\mathrm{P}(D_1 \cap D_2) \leq \min\{\mathrm{P}(D_1)\,, \mathrm{P}(D_2)\} = 10^{-6}$. Equality may hold in both cases: for the union, take D_1 and D_2 disjoint, for the intersection, take D_1 and D_2 equal to each other.

2.9 a Simply by inspection we find that
$A = \{TTH, THT, HTT\}, B = \{TTH, THT, HTT, TTT\}$,
$C = \{HHH, HHT, HTH, HTT\}, D = \{TTT, TTH, THT, THH\}$.

2.9 b Here we find that $A^c = \{TTT, THH, HTH, HHT, HHH\}$,
$A \cup (C \cap D) = A \cup \emptyset = A, A \cap D^c = \{HTT\}$.

2.10 Cf. Exercise 2.7: the event "A or B occurs, but not both" equals $C = (A \cup B) \cap (A \cap B)^c$ Rewriting this using DeMorgan's laws (or paraphrasing "A or B occurs, but not both" as "A occurs but not B or B occurs but not A"), we can also write $C = (A \cap B^c) \cup (B \cap A^c)$.

2.11 Let the two outcomes be called 1 and 2. Then $\Omega = \{1, 2\}$, and $\mathrm{P}(1) = p, \mathrm{P}(2) = p^2$. We must have $\mathrm{P}(1) + \mathrm{P}(2) = \mathrm{P}(\Omega) = 1$, so $p + p^2 = 1$. This has two solutions: $p = (-1 + \sqrt{5})/2$ and $p = (-1 - \sqrt{5})/2$. Since we must have $0 \leq p \leq 1$ only one is allowed: $p = (-1 + \sqrt{5})/2$.

2.12 a This is the same situation as with the three envelopes on the doormat, but now with ten possibilities. Hence an outcome has probability $1/10!$ to occur.

2.12 b For the five envelopes labeled $1, 2, 3, 4, 5$ there are $5!$ possible orders, and for each of these there are $5!$ possible orders for the envelopes labeled $6, 7, 8, 9, 10$. Hence in total there are $5! \cdot 5!$ outcomes.

2.12 c There are $32{\cdot}5!{\cdot}5!$ outcomes in the event "dream draw." Hence the probability is $32 \cdot 5!5!/10! = 32 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5/(6 \cdot 7 \cdot 8 \cdot 9 \cdot 10) = 8/63 =$12.7 percent.

2.13 a The outcomes are pairs (x, y).

The outcome (a, a) has probability 0 to occur. The outcome (a, b) has probability $1/4 \times 1/3 = 1/12$ to occur.
The table becomes:

	a	b	c	d
a	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
b	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{1}{12}$
c	$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{12}$
d	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	0

2.13 b Let C be the event "c is one of the chosen possibilities". Then $C = \{(c, a), (c, b), (a, c), (b, c)\}$. Hence $\mathrm{P}(C) = 4/12 = 1/3$.

2.14 a Since door a is never opened, $\mathrm{P}((a, a)) = \mathrm{P}((b, a)) = \mathrm{P}((c, a)) = 0$. If the candidate chooses a (which happens with probability 1/3), then the quizmaster chooses without preference from doors b and c. This yields that $\mathrm{P}((a, b)) = \mathrm{P}((a, c)) = 1/6$. If the candidate chooses b (which happens with probability 1/3), then the quizmaster can only open door c. Hence $\mathrm{P}((b, c)) = 1/3$. Similarly, $\mathrm{P}((c, b)) = 1/3$. Clearly, $\mathrm{P}((b, b)) = \mathrm{P}((c, c)) = 0$.

2.14 b If the candidate chooses a then she or he wins; hence the corresponding event is $\{(a, a), (a, b), (a, c)\}$, and its probability is 1/3.

2.14 c To end with a the candidate should have chosen b or c. So the event is $\{(b, c), (c, b)\}$ and $\mathrm{P}(\{(b, c), (c, b)\}) = 2/3$.

2.15 The rule is:

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(A \cap B)-\mathrm{P}(A \cap C)-\mathrm{P}(B \cap C)+\mathrm{P}(A \cap B \cap C)\,.$$

That this is true can be shown by applying the sum rule twice (and using the set property $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$):

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

Here we did put $s := \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C)$ for typographical convenience.

2.16 Since $E \cap F \cap G = \emptyset$, the three sets $E \cap F$, $F \cap G$, and $E \cap G$ are disjoint. Since each has probability 1/3, they have probability 1 together. From these two facts one deduces $\mathrm{P}(E) = \mathrm{P}(E \cap F) + \mathrm{P}(E \cap G) = 2/3$ (make a diagram or use that $E = E \cap (E \cap F) \cup E \cap (F \cap G) \cup E \cap (E \cap G)$).

2.17 Since there are two queues we use pairs (i, j) of natural numbers to indicate the number of customers i in the first queue, and the number j in the second queue. Since we have no reasonable bound on the number of people that will queue, we take $\Omega = \{(i, j) : i = 0, 1, 2, \ldots, j = 0, 1, 2, \ldots\}$.

2.18 The probability r of no success at a certain day is equal to the probability that *both* experiments fail, hence $r = (1 - p)^2$. The probability of success for the first time on day n therefore equals $r^{n-1}(1 - r)$. (Cf. Section2.5.)

2.19 a We need at least two days to see two heads, hence $\Omega = \{2, 3, 4, \dots\}$.

2.19 b It takes 5 tosses if and only if the fifth toss is heads (which has probability p), and exactly one of the first 4 tosses is heads (which has probability $4p(1-p)^3$). Hence the probability asked for equals $4p^2(1-p)^3$.

3.1 Define the following events: B is the event "point B is reached on the second step," C is the event "the path to C is chosen on the first step," and similarly we define D and E. Note that the events C, D, and E are mutually exclusive and that one of them must occur. Furthermore, that we can only reach B by first going to C or D. For the computation we use the law of total probability, by conditioning on the result of the first step:

$$\begin{aligned} \mathrm{P}(B) &= \mathrm{P}(B \cap C) + \mathrm{P}(B \cap D) + \mathrm{P}(B \cap E) \\ &= \mathrm{P}(B \mid C)\,\mathrm{P}(C) + \mathrm{P}(B \mid D)\,\mathrm{P}(D) + \mathrm{P}(B \mid E)\,\mathrm{P}(E) \\ &= \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} + \frac{1}{3} \cdot 0 = \frac{7}{36}. \end{aligned}$$

3.2 a Event A has three outcomes, event B has 11 outcomes, and $A \cap B = \{(1,3),(3,1)\}$. Hence we find $\mathrm{P}(B) = 11/36$ and $\mathrm{P}(A \cap B) = 2/36$ so that

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{2/36}{11/36} = \frac{2}{11}.$$

3.2 b Because $\mathrm{P}(A) = 3/36 = 1/12$ and this is not equal to $2/11 = \mathrm{P}(A \mid B)$ the events A and B are *dependent*.

3.3 a There are 13 spades in the deck and each has probability $1/52$ of being chosen, hence $\mathrm{P}(S_1) = 13/52 = 1/4$. Given that the first card is a spade there are $13-1 = 12$ spades left in the deck with $52 - 1 = 51$ remaining cards, so $\mathrm{P}(S_2 \mid S_1) = 12/51$. If the first card is not a spade there are 13 spades left in the deck of 51, so $\mathrm{P}(S_2 \mid S_1^c) = 13/51$.

3.3 b We use the law of total probability (based on $\Omega = S_1 \cup S_1^c$):

$$\begin{aligned} \mathrm{P}(S_2) &= \mathrm{P}(S_2 \cap S_1) + \mathrm{P}(S_2 \cap S_1^c) = \mathrm{P}(S_2 \mid S_1)\,\mathrm{P}(S_1) + \mathrm{P}(S_2 \mid S_1^c)\,\mathrm{P}(S_1^c) \\ &= \frac{12}{51} \cdot \frac{1}{4} + \frac{13}{51} \cdot \frac{3}{4} = \frac{12 + 39}{51 \cdot 4} = \frac{1}{4}. \end{aligned}$$

3.4 We repeat the calculations from Section 3.3 based on $\mathrm{P}(B) = 1.3 \cdot 10^{-5}$:

$$\begin{aligned} \mathrm{P}(T \cap B) &= \mathrm{P}(T \mid B) \cdot \mathrm{P}(B) = 0.7 \cdot 0.000\,013 = 0.000\,0091 \\ \mathrm{P}(T \cap B^c) &= \mathrm{P}(T \mid B^c) \cdot \mathrm{P}(B^c) = 0.1 \cdot 0.999\,987 = 0.099\,9987 \end{aligned}$$

so $\mathrm{P}(T) = \mathrm{P}(T \cap B) + \mathrm{P}(T \cap B^c) = 0.000\,0091 + 0.099\,9987 = 0.100\,0078$ and

$$\mathrm{P}(B \mid T) = \frac{\mathrm{P}(T \cap B)}{\mathrm{P}(T)} = \frac{0.000\,0091}{0.100\,0078} = 0.000\,0910 = 9.1 \cdot 10^{-5}.$$

Further, we find

$$\mathrm{P}(T^c \cap B) = \mathrm{P}(T^c \mid B) \cdot \mathrm{P}(B) = 0.3 \cdot 0.000\,013 = 0.000\,0039$$

and combining this with $\mathrm{P}(T^c) = 1 - \mathrm{P}(T) = 0.899\,9922$:

$$\mathrm{P}(B \mid T^c) = \frac{\mathrm{P}(T^c \cap B)}{\mathrm{P}(T^c)} = \frac{0.000\,0039}{0.899\,9922} = 0.000\,0043 = 4.3 \cdot 10^{-6}.$$

3.5 Define the events R_1 and R_2 meaning: a red ball is drawn on the first and second draw, respectively. We are asked to compute $\mathrm{P}(R_1 \cap R_2)$. By conditioning on R_1 we find:

$$\mathrm{P}(R_1 \cap R_2) = \mathrm{P}(R_2 \,|\, R_1) \cdot \mathrm{P}(R_1) = \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8},$$

where the conditional probability $\mathrm{P}(R_2 \,|\, R_1)$ follows from the contents of the urn after R_1 has occurred: one white and three red balls.

3.6 a Let E denote the event "outcome is an even numbered month" and H the event "outcome is in the first half of the year." Then $\mathrm{P}(E) = 1/2$ and, because in the first half of the year there are three even and three odd numbered months, $\mathrm{P}(E \,|\, H) = 1/2$ as well; the events are independent.

3.6 b Let S denote the event "outcome is a summer month". Of the three summer months, June and August are even numbered, so $\mathrm{P}(E \,|\, S) = 2/3 \neq 1/2$. Therefore, E and S are dependent.

3.7 a The best approach to a problem like this one is to write out the conditional probability and then see if we can somehow combine this with $\mathrm{P}(A) = 1/3$ to solve the puzzle. Note that $\mathrm{P}(B \cap A^c) = \mathrm{P}(B \,|\, A^c)\,\mathrm{P}(A^c)$ and that $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B \cap A^c)$. So

$$\mathrm{P}(A \cup B) = \frac{1}{3} + \frac{1}{4} \cdot \left(1 - \frac{1}{3}\right) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

3.7 b From the conditional probability we find $\mathrm{P}(A^c \cap B^c) = \mathrm{P}(A^c \,|\, B^c)\,\mathrm{P}(B^c) = \frac{1}{2}\,(1 - \mathrm{P}(B))$. Recalling DeMorgan's law we know $\mathrm{P}(A^c \cap B^c) = \mathrm{P}((A \cup B)^c) = 1 - \mathrm{P}(A \cup B) = 1/3$. Combined this yields an equation for $\mathrm{P}(B)$: $\frac{1}{2}\,(1 - \mathrm{P}(B)) = 1/3$ from which we find $\mathrm{P}(B) = 1/3$.

3.8 a This asks for $\mathrm{P}(W)$. We use the law of total probability, decomposing $\Omega = F \cup F^c$. Note that $\mathrm{P}(W \,|\, F) = 0.99$.

$$\begin{aligned} \mathrm{P}(W) &= \mathrm{P}(W \cap F) + \mathrm{P}(W \cap F^c) = \mathrm{P}(W \,|\, F)\,\mathrm{P}(F) + \mathrm{P}(W \,|\, F^c)\,\mathrm{P}(F^c) \\ &= 0.99 \cdot 0.1 + 0.02 \cdot 0.9 = 0.099 + 0.018 = 0.117. \end{aligned}$$

3.8 b We need to determine $\mathrm{P}(F \,|\, W)$, and this can be done using Bayes' rule. Some of the necessary computations have already been done in **a**, we can copy $\mathrm{P}(W \cap F)$ and $\mathrm{P}(W)$ and get:

$$\mathrm{P}(F \,|\, W) = \frac{\mathrm{P}(F \cap W)}{\mathrm{P}(W)} = \frac{0.099}{0.117} = 0.846.$$

3.9 Deciphering the symbols we conclude that $\mathrm{P}(B \,|\, A)$ is to be determined. From the probabilities listed we find $\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B) = 3/4 + 2/5 - 4/5 = 7/20$, so that $\mathrm{P}(B \,|\, A) = \mathrm{P}(A \cap B)\,/\mathrm{P}(A) = (7/20)/(3/4) = 28/60 = 7/15$.

3.10 Let K denote the event "the student *knows* the answer" and C the event "the answer that is given is the correct one." We are to determine $\mathrm{P}(K \,|\, C)$. From the information given, we know that $\mathrm{P}(C \,|\, K) = 1$ and $\mathrm{P}(C \,|\, K^c) = 1/4$, and that $\mathrm{P}(K) = 0.6$. Therefore:

$$\mathrm{P}(C) = \mathrm{P}(C \,|\, K) \cdot \mathrm{P}(K) + \mathrm{P}(C \,|\, K^c) \cdot \mathrm{P}(K^c) = 1 \cdot 0.6 + \frac{1}{4} \cdot 0.4 = 0.6 + 0.1 = 0.7$$

and $\mathrm{P}(K \,|\, C) = 0.6/0.7 = 6/7$.

3.11 a The probability that a driver that is classified as exceeding the legal limit in fact does not exceed the limit.

3.11 b It is given that $\mathrm{P}(B) = 0.05$. We determine the answer via $\mathrm{P}(B^c \mid A) = \mathrm{P}(B^c \cap A)\,/\mathrm{P}(A)$. We find $\mathrm{P}(B^c \cap A) = \mathrm{P}(A \mid B^c) \cdot \mathrm{P}(B^c) = 0.95\,(1-p)$, $\mathrm{P}(B \cap A) = \mathrm{P}(A \mid B) \cdot \mathrm{P}(B) = 0.05\,p$, and by adding them $\mathrm{P}(A) = 0.95 - 0.9\,p$. So

$$\mathrm{P}(B^c \mid A) = \frac{0.95\,(1-p)}{0.95 - 0.9\,p} = \frac{95\,(1-p)}{95 - 90\,p} = 0.5 \quad \text{when } p = 0.95.$$

3.11 c From **b** we find

$$\mathrm{P}(B \mid A) = 1 - \mathrm{P}(B^c \mid A) = \frac{95 - 90\,p - 95\,(1-p)}{95 - 90\,p} = \frac{5p}{95 - 90\,p}.$$

Setting this equal to 0.9 and solving for p yields $p = 171/172 \approx 0.9942$.

3.12 We start by deriving some *unconditional* probabilities from what is given: $\mathrm{P}(B \cap C) = \mathrm{P}(B \mid C) \cdot \mathrm{P}(C) = 1/6$ and $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(B \cap C) \cdot \mathrm{P}(A \mid B \cap C) = 1/24$. Next, we should realize that $B \cap C$ is the union of the disjoint events $A \cap B \cap C$ and $A^c \cap B \cap C$, so that

$$\mathrm{P}(A^c \cap B \cap C) = \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B \cap C) = \frac{1}{6} - \frac{1}{24} = \frac{1}{8}.$$

3.13 a There are several ways to see that $\mathrm{P}(D_1) = 5/9$. Method 1: the first team we draw is irrelevant, for the second team there are always 5 "good" choices out of 9 remaining teams. Method 2: there are $\binom{10}{2} = 45$ possible outcomes when two teams are drawn; among these, there are $5 \cdot 5 = 25$ favorable outcomes (all weak-strong pairings), resulting in $\mathrm{P}(D_1) = 25/45 = 5/9$.

3.13 b Given D_1, there are 4 strong and 4 weak teams left. Using one of the methods from **a** on this reduced number of teams, we find $\mathrm{P}(D_2 \mid D_1) = 4/7$ and $\mathrm{P}(D_1 \cap D_2) = \mathrm{P}(D_2 \mid D_1) \cdot \mathrm{P}(D_1) = (4/7) \cdot (5/9) = 20/63$.

3.13 c Proceeding in the same fashion, we find $\mathrm{P}(D_3 \mid D_1 \cap D_2) = 3/5$ and $\mathrm{P}(D_1 \cap D_2 \cap D_3) = \mathrm{P}(D_3 \mid D_1 \cap D_2) \cdot \mathrm{P}(D_1 \cap D_2) = (3/5) \cdot (20/63) = 12/63$.

3.13 d Subsequently, we find $\mathrm{P}(D_4 \mid D_1 \cap \cdots \cap D_3) = 2/3$ and $\mathrm{P}(D_5 \mid D_1 \cap \cdots \cap D_4) = 1$. The final result can be written as

$$\mathrm{P}(D_1 \cap \cdots \cap D_5) = \frac{5}{9} \cdot \frac{4}{7} \cdot \frac{3}{5} \cdot \frac{2}{3} \cdot 1 = \frac{8}{63}.$$

The probability of a "dream draw" with n strong and n weak teams is $\mathrm{P}(D_1 \cap \cdots \cap D_n) = 2^n/\binom{2n}{n}$.

3.14 a If you chose the right door, switching will make you lose, so $\mathrm{P}(W \mid R) = 0$. If you chose the wrong door, switching will make you win for sure: $\mathrm{P}(W \mid R^c) = 1$.

3.14 b Using $\mathrm{P}(R) = 1/3$, we find:

$$\begin{aligned}\mathrm{P}(W) &= \mathrm{P}(W \cap R) + \mathrm{P}(W \cap R^c) = \mathrm{P}(W \mid R)\,\mathrm{P}(R) + \mathrm{P}(W \mid R^c)\,\mathrm{P}(R^c) \\ &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.\end{aligned}$$

3.15 This is a puzzle and there are many ways to solve it. First note that $\mathrm{P}(A) = 1/2$. We condition on $A \cup B$:

$$\begin{aligned}\mathrm{P}(B) &= \mathrm{P}(B \mid A \cup B) \cdot \mathrm{P}(A \cup B) \\ &= \frac{2}{3}\{\mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A)\,\mathrm{P}(B)\} \\ &= \frac{2}{3}\{\frac{1}{2} + \mathrm{P}(B) - \frac{1}{2}\mathrm{P}(B)\} \\ &= \frac{1}{3} + \frac{1}{3}\mathrm{P}(B)\,.\end{aligned}$$

Solving the resulting equation yields $\mathrm{P}(B) = \frac{\frac{1}{3}}{1-\frac{1}{3}} = \frac{1}{2}$.

3.16 a Using Bayes' rule we find:

$$\mathrm{P}(D \mid T) = \frac{\mathrm{P}(D \cap T)}{\mathrm{P}(T)} = \frac{0.98 \cdot 0.01}{0.98 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.1653.$$

3.16 b Denote the event "the second test indicates you have the disease" with the letter S.
Method 1 ("unconscious"): from the first test we know that the probability we have the disease is not 0.01 but 0.1653 and we reason that we shoud redo the calculation from **a** with $\mathrm{P}(D) = 0.1653$:

$$\mathrm{P}(D \mid S) = \frac{\mathrm{P}(D \cap S)}{\mathrm{P}(S)} = \frac{0.98 \cdot 0.1653}{0.98 \cdot 0.1653 + 0.05 \cdot 0.8347} \approx 0.7951.$$

This is the correct answer, but a more thorough consideration is warrented.
Method 2 ("conscientious"): we are in fact to determine

$$\mathrm{P}(D \mid S \cap T) = \frac{\mathrm{P}(D \cap S \cap T)}{\mathrm{P}(S \cap T)}$$

and we should wonder what "independent repetition of the test" exactly means. Clearly, both tests are dependent on whether you have the disease or not (and as a result, S and T are dependent), but *given that you have the disease* the outcomes are independent, and the same when you do not have the disease. Formally put: given D, the events S and T are independent; given D^c, the events S and T are independent. In formulae:

$$\begin{aligned}\mathrm{P}(S \cap T \mid D) &= \mathrm{P}(S \mid D) \cdot \mathrm{P}(T \mid D), \\ \mathrm{P}(S \cap T \mid D^c) &= \mathrm{P}(S \mid D^c) \cdot \mathrm{P}(T \mid D^c).\end{aligned}$$

So $\mathrm{P}(D \cap S \cap T) = \mathrm{P}(S \mid D) \cdot \mathrm{P}(T \mid D)\mathrm{P}(D) = (0.98)^2 \cdot 0.01 = 0.009604$ and $\mathrm{P}(D^c \cap S \cap T) = (0.05)^2 \cdot 0.99 = 0.002475$. Adding them yields $\mathrm{P}(S \cap T) = 0.012079$ and so $\mathrm{P}(D \mid S \cap T) = 0.009604/0.012079 \approx 0.7951$. Note that $\mathrm{P}(S \mid T) \approx 0.2037$ which is much larger than $\mathrm{P}(S) \approx 0.0593$ (and for a good test, it should be).

3.17 a I win the game after the next two rallies if I win both: $\mathrm{P}(W \cap G) = p^2$. Similarly for losing the game if you win both rallies: $\mathrm{P}(W^c \cap G) = (1-p)^2$. So $\mathrm{P}(W \mid G) = p^2/(p^2 + (1-p)^2)$.

3.17 b Note that $D = G^c$ and so $\mathrm{P}(D) = 1 - p^2 - (1-p)^2 = 2p(1-p)$. Using the law of total probability we find:

$$\mathrm{P}(W) = \mathrm{P}(W \cap G) + \mathrm{P}(W \cap D) = p^2 + \mathrm{P}(W \mid D)\mathrm{P}(D) = p^2 + 2p(1-p)\mathrm{P}(W \mid D).$$

Substituting $\mathrm{P}(W \mid D) = \mathrm{P}(W)$ (when it become deuce again, the initial situation repeats itself) and solving yields $\mathrm{P}(W) = p^2/(p^2 + (1-p)^2)$.

3.17 c If the game does not end after the next two rallies, each of us has again the same probability to win the game as we had initially. This implies, that the probability of winning the game is split between us in the same way as the probability of an "immediate win," i.e., in the next two rallies. So $\mathrm{P}(W) = \mathrm{P}(W \mid G)$.

3.18 a No: $A \cap B = \emptyset$ so $\mathrm{P}(A \mid B) = 0 \neq \mathrm{P}(A)$ which is positive.

3.18 b Since, by independence, $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B) > 0$ it must be that $A \cap B \neq \emptyset$.

3.18 c If $A \subset B$ then $\mathrm{P}(B \mid A) = 1$. However, $\mathrm{P}(B) < 1$, so $\mathrm{P}(B) \neq \mathrm{P}(B \mid A)$: A en B are dependent.

3.18 d Since $A \subset A \cap B$, $\mathrm{P}(A \cup B \mid A) = 1$, so for the required independence $\mathrm{P}(A \cup B) = 1$ should hold. On the other hand, $\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A) \cdot \mathrm{P}(B)$ and so $\mathrm{P}(A \cup B) = 1$ can be rewritten as $(1 - \mathrm{P}(A)) \cdot (1 - \mathrm{P}(B)) = 0$, which clearly contradicts the assumptions.

4.1 a In two independent throws of a die there are 36 possible outcomes, each occurring with probability 1/36. Since there are 25 ways to have no 6's, 10 ways to have one 6, and one way to have two 6's, we find that $p_Z(0) = 25/36$, $p_Z(1) = 10/36$, and $p_Z(2) = 1/36$. So the probability mass function p_Z of Z is given by the following table:

z	0	1	2
$p_Z(z)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

The distribution function F_Z is given by

$$F_Z(a) = \begin{cases} 0 & \text{for } a < 0 \\ \frac{25}{36} & \text{for } 0 \leq a < 1 \\ \frac{25}{36} + \frac{10}{36} = \frac{35}{36} & \text{for } 1 \leq a < 2 \\ \frac{25}{36} + \frac{10}{36} + \frac{1}{36} = 1 & \text{for } a \geq 2. \end{cases}$$

Z is the sum of two independent $Ber(1/6)$ distributed random variables, so Z has a $Bin(2, 1/6)$ distribution.

4.1 b If we denote the outcome of the two throws by (i, j), where i is the outcome of the first throw and j the outcome of the second, then $\{M = 2, Z = 0\} = \{\,(2,1),\,(1,2),\,(2,2)\,\}$, $\{S = 5, Z = 1\} = \emptyset$, $\{S = 8, Z = 1\} = \{\,(6,2),\,(2,6)\,\}$. Furthermore, $\mathrm{P}(M = 2, Z = 0) = 3/36$, $\mathrm{P}(S = 5, Z = 1) = 0$, and $\mathrm{P}(S = 8, Z = 1) = 2/36$.

4.1 c The events are dependent, because, e.g., $\mathrm{P}(M = 2, Z = 0) = \frac{3}{36}$ differs from $\mathrm{P}(M = 2) \cdot \mathrm{P}(Z = 0) = \frac{3}{36} \cdot \frac{25}{36}$.

4.2 a Since $\mathrm{P}(Y = 0) = \mathrm{P}(X = 0)$, $\mathrm{P}(Y = 1) = \mathrm{P}(X = -1) + \mathrm{P}(X = 1)$, $\mathrm{P}(Y = 4) = \mathrm{P}(X = 2)$, the following table gives the probability mass function of Y:

a	0	1	4
$p_Y(a)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$

4.2 b $F_X(1) = \mathrm{P}(X \le 1) = 1/2$; $F_Y(1) = 1/2$; $F_X(\frac{3}{4}) = \frac{3}{8}$; $F_Y(\frac{3}{4}) = \frac{1}{8}$; $F_X(\pi - 3) = F_X(0.1415...) = 3/8$; $F_Y(\pi - 3) = 1/8$.

4.3 For every random variable X and every real number a one has that $\{X = a\} = \{X \le a\} \setminus \{X < a\}$, so

$$\mathrm{P}(X = a) = \mathrm{P}(X \le a) - \mathrm{P}(X < a)\,.$$

Since F is right-continuous (see p. 49) we have that $F(a) = \lim_{\varepsilon\downarrow 0} F(a+\varepsilon) = \lim_{\varepsilon\downarrow 0} \mathrm{P}(X \le a+\varepsilon)$. Moreover, $\mathrm{P}(X < a) = \lim_{\varepsilon\downarrow 0} \mathrm{P}(X \le a-\varepsilon) = \lim_{\varepsilon\downarrow 0} F(a-\varepsilon)$, so we see that $\mathrm{P}(X = a) > 0$ precisely for those a for which $F(a)$ "makes a jump." In this exercise this is for $a = 0, \frac{1}{2}$, and $a = \frac{3}{4}$, and $p(0) = \mathrm{P}(X \le 0) - \mathrm{P}(X < 0) = 1/3$, etc. We find

a	0	1/2	3/4
$p(a)$	1/3	1/6	1/2.

4.4 a By conditioning one finds that each coin has probability $2p - p^2$ to give heads. The outcomes of the coins are independent, so the total number of heads has a binomial distribution with parameters n and $2p - p^2$.

4.4 b Since the total number of heads has a binomial distribution with parameters n and $2p - p^2$, we find for $k = 0, 1, \ldots, n$,

$$p_X(k) = \binom{n}{k} \left(2p - p^2\right)^k \left(p^2 - 2p + 1\right)^{n-k},$$

and $p_X(k) = 0$ otherwise.

4.5 In one throw of a die one cannot exceed 6, so $F(1) = \mathrm{P}(X \le 1) = 0$. Since $\mathrm{P}(X = 1) = 0$, we find that $F(2) = \mathrm{P}(X = 2)$, and from Table 4.1 it follows that $F(2) = 21/36$. No matter the outcomes, after seven throws we have that the sum certainly exceeds 6, so $F(7) = 1$.

4.6 a Let $(\omega_1, \omega_2, \omega_3)$ denote the outcome of the three draws, where ω_1 is the outcome of the first draw, ω_2 the outcome of the second draw, and ω_3 the outcome of the third one. Then the sample space Ω is given by

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2, \omega_3) \,:\, \omega_1,\, \omega_2,\, \omega_3 \in \{1, 2, 3\}\,\} \\ &= \{(1,1,1), (1,1,2), \ldots, (1,1,6), (1,2,1), \ldots, (3,3,3)\},\end{aligned}$$

and

$$\bar{X}(\omega_1, \omega_2, \omega_3) = \frac{\omega_1 + \omega_2 + \omega_3}{3} \quad \text{for } (\omega_1, \omega_2, \omega_3) \in \Omega.$$

The possible outcomes of $\bar{X}$ are $1, \frac{4}{3}, \frac{5}{3}, 2, \frac{7}{3}, \frac{8}{3}, 3$. Furthermore, $\{\bar{X} = 1\} = \{(1,1,1)\}$, so $\mathrm{P}\big(\bar{X} = 1\big) = \frac{1}{27}$, and

$$\{\bar{X} = \frac{4}{3}\} = \{(1,1,2), (1,2,1), (2,1,1)\}, \quad \text{so} \quad \mathrm{P}\left(\bar{X} = \frac{4}{3}\right) = \frac{3}{27}.$$

Since

$$\{\bar{X} = \frac{5}{3}\} = \{(1,1,3), (1,3,1), (3,1,1), (1,2,2), (2,1,2), (2,2,1)\},$$

we find that $\mathrm{P}\big(\bar{X} = \frac{5}{3}\big) = \frac{6}{27}$, etc. Continuing in this way we find that the probability mass function $p_{\bar{X}}$ of $\bar{X}$ is given by

a	1	$\frac{4}{3}$	$\frac{5}{3}$	2	$\frac{7}{3}$	$\frac{8}{3}$	3
$p_{\bar{X}}(a)$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{7}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{1}{27}$

4.6 b Setting, for $i = 1, 2, 3$,

$$Y_i = \begin{cases} 1 & \text{if} \quad X_i < 2 \\ 0 & \text{if} \quad X_i \geq 2, \end{cases}$$

and $Y = Y_1 + Y_2 + Y_3$, then Y is $Bin(3, \frac{1}{3})$. It follows that the probability that exactly two draws are smaller than 2 is equal to

$$\mathrm{P}(Y = 2) = \binom{3}{2} \left(\frac{1}{3}\right)^2 \left(1 - \frac{1}{3}\right) = \frac{6}{27}.$$

Another way to find this answer is to realize that the event that exactly two draws are equal to 1 is equal to

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

4.7 a Setting $X_i = 1$ if the ith lamp is defective, and $X_i = 0$ if not, for $i = 1, 2, \ldots, 1000$, we see that X_i is a $Ber(0.001)$ distributed random variable. Assuming that these X_i are independent, we find that X (as the sum of 1000 independent $Ber(0.001)$ random variables) has a $Bin(1000, 0.001)$ distribution.

4.7 b Since X has a $Bin(1000, 0.001)$ distribution, these probabilities are $\mathrm{P}(X = 0) = (\frac{999}{1000})^{1000} = 0.367695$, $\mathrm{P}(X = 1) = \frac{999^{999}}{1000^{1000}} = 0.36806$, and $\mathrm{P}(X > 2) = 1 - \mathrm{P}(X \leq 2) = 0.08021$.

4.8 a Assuming that O-rings fail independently of one-another, X can be seen as the sum of six independent random variables with a $Ber(0.8178)$ distribution. Consequently, X has a $Bin(6, 0.8178)$ distribution.

4.8 b Since X has a $Bin(6, 0.8178)$ distribution, we find that $\mathrm{P}(X \geq 1) = 1 - \mathrm{P}(X = 0) = 1 - (1 - 0.8178)^6 = 0.9999634$.

4.9 a With these assumptions, X has a $Bin(6, 0.0137)$ distribution, so we find that the probability that during a launch at least one O-ring fails is equal to $\mathrm{P}(X \geq 1) = 1 - \mathrm{P}(X = 0) = 1 - (1 - 0.0137)^6 = 0.079436$. But then the probability, that at least one O-ring fails for the first time during the 24th launch is equal to $\big(\mathrm{P}(X = 0)\big)^{23}\mathrm{P}(X \geq 1) = 0.01184$.

4.9 b The probability that no O-ring fails during a launch is $\mathrm{P}(X = 0) = (1 - 0.0137)^6 = 0.92056$, so no O-ring fails during 24 launches is equal to $\big(\mathrm{P}(X = 0)\big)^{24} = 0.13718$.

4.10 a Each R_i has a Bernoulli distribution, because it can only attain the values 0 and 1. The parameter is $p = \mathrm{P}(R_i = 1)$. It is not easy to determine $\mathrm{P}(R_i = 1)$, but it is fairly easy to determine $\mathrm{P}(R_i = 0)$. The event $\{R_i = 0\}$ occurs when none of the m people has chosen the ith floor. Since they make their choices independently of each other, and each floor is selected by each of these m people with probability 1/21, it follows that

$$\mathrm{P}(R_i = 0) = \left(\frac{20}{21}\right)^m.$$

Now use that $p = \mathrm{P}(R_i = 1) = 1 - \mathrm{P}(R_i = 0)$ to find the desired answer.

4.10 b If $\{R_1 = 0\}, \dots, \{R_{20} = 0\}$, we must have that $\{R_{21} = 1\}$, so we cannot conclude that the events $\{R_1 = a_1\}, \dots, \{R_{21} = a_{21}\}$, where a_i is 0 or 1, are independent. Consequently, we cannot use the argument from Section 4.3 to conclude that S_m is $Bin(21, p)$. In fact, S_m is not $Bin(21, p)$ distributed, as the following shows. The elevator will stop at least once, so $\mathrm{P}(S_m = 0) = 0$. However, if S_m would have a $Bin(21, p)$ distribution, then $\mathrm{P}(S_m = 0) = (1-p)^{21} > 0$, which is a contradiction.

4.10 c This exercise is a variation on finding the probability of no coincident birthdays from Section 3.2. For $m = 2$, $S_2 = 1$ occurs precisely if the two persons entering the elevator select the same floor. The first person selects any of the 21 floors, the second selects the same floor with probability $1/21$, so $\mathrm{P}(S_2 = 1) = 1/21$. For $m = 3$, $S_3 = 1$ occurs if the second and third persons entering the elevator both select the same floor as was selected by the first person, so $\mathrm{P}(S_3 = 1) = (1/21)^2 = 1/441$. Furthermore, $S_3 = 3$ occurs precisely when all three persons choose a different floor. Since there are $21 \cdot 20 \cdot 19$ ways to do this out of a total of 21^3 possible ways, we find that $\mathrm{P}(S_3 = 3) = 380/441$. Since S_3 can only attain the values $1, 2, 3$, it follows that $\mathrm{P}(S_3 = 2) = 1 - \mathrm{P}(S_3 = 1) - \mathrm{P}(S_3 = 3) = 60/441$.

4.11 Since the lotteries are different, the event to win something (or not) in one lottery is independent of the other. So the probability to win a prize the first time you play is $p_1 p_2 + p_1(1-p_2) + (1-p_1)p_2$. Clearly M has a geometric distribution, with parameter $p = p_1 p_2 + p_1(1-p_2) + (1-p_1)p_2$.

4.12 The "probability that your friend wins" is equal to

$$p + p(1-p)^2 + p(1-p)^4 + \cdots = p \cdot \frac{1}{1-(1-p)^2} = \frac{1}{2-p}.$$

The procedure is favorable to your friend if $1/(2-p) > 1/2$, and this is true if $p > 0$.

4.13 a Since we wait for the first time we draw the marked bolt in independent draws, each with a $Ber(p)$ distribution, where p is the probability to draw the bolt (so $p = 1/N$), we find, using a reasoning as in Section 4.4, that X has a $Geo(1/N)$ distribution.

4.13 b Clearly, $\mathrm{P}(Y = 1) = 1/N$. Let D_i be the event that the marked bolt was drawn (for the first time) in the ith draw. For $k = 2, \dots, N$ we have that

$$\begin{aligned}\mathrm{P}(Y = k) &= \mathrm{P}(D_1^c \cap \cdots \cap D_{k-1}^c \cap D_k) \\ &= \mathrm{P}(D_k \mid D_1^c \cap \cdots \cap D_{k-1}^c) \cdot \mathrm{P}(D_1^c \cap \cdots \cap D_{k-1}^c).\end{aligned}$$

Now $\mathrm{P}(D_k \mid D_1^c \cap \cdots \cap D_{k-1}^c) = \frac{1}{N-k+1}$,

$$\mathrm{P}(D_1^c \cap \cdots \cap D_{k-1}^c) = \mathrm{P}(D_{k-1}^c \mid D_1^c \cap \cdots \cap D_{k-2}^c) \cdot \mathrm{P}(D_1^c \cap \cdots \cap D_{k-2}^c),$$

and

$$\mathrm{P}(D_{k-1}^c \mid D_1^c \cap \cdots \cap D_{k-1}^c) = 1 - \mathrm{P}(D_{k-1} \mid D_1^c \cap \cdots \cap D_{k-1}^c) = 1 - \frac{1}{N-k+2}.$$

Continuing in this way, we find after k steps that

$$\mathrm{P}(Y = k) = \frac{1}{N-k+1} \cdot \frac{N-k+1}{N-k+2} \cdot \frac{N-k+2}{N-k+3} \cdots \frac{N-2}{N-1} \cdot \frac{N-1}{N} = \frac{1}{N}.$$

See also Section 9.3, where the distribution of Y is derived in a different way.

4.13 c For $k = 0, 1, \ldots, r$, the probability $\mathrm{P}(Z = k)$ is equal to the number of ways the event $\{Z = k\}$ can occur, divided by the number of ways $\binom{N}{r}$ we can select r objects from N objects, see also Section 4.3. Since one can select k marked bolts from m marked ones in $\binom{m}{k}$ ways, and $r-k$ nonmarked bolts from $N-m$ nonmarked ones in $\binom{N-m}{r-k}$ ways, it follows that

$$\mathrm{P}(Z = k) = \frac{\binom{m}{k}\binom{N-m}{r-k}}{\binom{N}{r}}, \quad \text{for } k = 0, 1, 2, \ldots, r.$$

4.14 a Denoting 'heads' by H and 'tails' by T, the event $\{X = 2\}$ occurs if we have thrown HH, i.e, if the outcome of both the first and the second throw was 'heads'. Since the probability of 'heads' is p, we find that $\mathrm{P}(X = 2) = p \cdot p = p^2$. Furthermore, $X = 3$ occurs if we either have thrown HTH, or THH, and since the probability of throwing 'tails' is $1-p$, we find that $\mathrm{P}(X = 3) = p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot p = 2p^2(1-p)$. Similarly, $X = 4$ can only occur if we throw $TTHH$, $THTH$, or $HTTH$, so $\mathrm{P}(X = 4) = 3p^2(1-p)^2$.

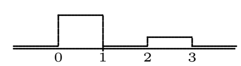
4.14 b If $X = n$, then the nth throw was heads (denoted by H), and all but one of the previous $n-1$ throws were tails (denoted by T). So the possible outcomes are

$$H\underbrace{TT\cdots T}_{n-2\text{ times}}H, \quad TH\underbrace{TT\cdots T}_{n-3\text{ times}}H, \quad TTH\underbrace{TT\cdots T}_{n-4\text{ times}}H, \quad \ldots, \quad \underbrace{TT\cdots T}_{n-2\text{ times}}HH$$

Notice there are exactly $\binom{n-1}{1} = n-1$ of such possible outcomes, each with probability $p^2(1-p)^{n-2}$, so $\mathrm{P}(X = n) = (n-1)p^2(1-p)^{n-2}$, for $n \geq 2$.

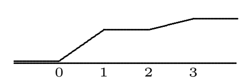
5.1 a

Sketch of probability density f:



5.1 b Since $f(x) = 0$ for $x < 0$, $F(b) = 0$ for $b < 0$. For $0 \leq b \leq 1$ we compute $F(b) = \int_{-\infty}^{b} f(x)\,\mathrm{d}x = \int_0^b 3/4\,\mathrm{d}x = 3b/4$. Since $f(x) = 0$ for $1 \leq x \leq 2$, $F(b) = F(1) = 3/4$ for $1 \leq b \leq 2$. For $2 \leq b \leq 3$ we compute $F(b) = \int_{-\infty}^{b} f(x)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x + \int_2^b f(x)\,\mathrm{d}x = F(1) + \int_2^b 1/4\,\mathrm{d}x = 3/4 + (1/4)(b-2) = b/4 + 1/4$. Since $f(x) = 0$ for $x > 3$, $F(b) = F(3) = 1$ for $b > 3$.

Sketch of distribution function F:



5.2 The event $\{1/2 < X \leq 3/4\}$ can be written as

$$\{1/2 < X \leq 3/4\} = \{X \leq 3/4\} \cap \{X \leq 1/2\}^c = \{X \leq 3/4\} \setminus \{X \leq 1/2\}.$$

Noting that $\{X \leq 1/2\} \subset \{X \leq 3/4\}$ we find

$$\mathrm{P}\left(\frac{1}{2} < X \leq \frac{3}{4}\right) = \mathrm{P}\left(X \leq \frac{3}{4}\right) - \mathrm{P}\left(X \leq \frac{1}{2}\right) = \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{5}{16}.$$

(cf. Exercise 2.5.)

5.3 a From $\mathrm{P}(X \leq 3/4) = \mathrm{P}(X \leq 1/4) + \mathrm{P}(1/4 < X \leq 3/4)$ we obtain

$$\mathrm{P}\left(\frac{1}{4} < X \leq \frac{3}{4}\right) = \mathrm{P}\left(X \leq \frac{3}{4}\right) - \mathrm{P}\left(X \leq \frac{1}{4}\right) = F\left(\frac{3}{4}\right) - F\left(\frac{1}{4}\right) = \frac{11}{16}.$$

5.3 b The probability density function $f(x)$ of X is obtained by differentiating the distribution function $F(x)$: $f(x) = \mathrm{d}/\mathrm{d}x\big(2x^2 - x^4\big) = 4x - 4x^3$ for $0 \leq x \leq 1$ (and $f(x) = 0$ elsewhere).

5.4 a Let T be the the time until the next arrival of a bus. Then T has $U(4,6)$ distribution. Hence $\mathrm{P}(X \leq 4.5) = \mathrm{P}(T \leq 4.5) = \int_4^{4.5} 1/2 \,\mathrm{d}x = 1/4$.

5.4 b Since Jensen leaves when the next bus arrives after more than 5 minutes, $\mathrm{P}(X = 5) = \mathrm{P}(T > 5) = \int_5^6 \frac{1}{2} \,\mathrm{d}x = 1/2$.

5.4 c Since $\mathrm{P}(X = 5) = 0.5 > 0$, X cannot be continuous. Since X can take any of the uncountable values in $[4,5]$, it can also not be discrete.

5.5 a A probability density f has to satisfy (I) $f(x) \geq 0$ for all x, and (II) $\int_{-\infty}^{\infty} f(x) \,\mathrm{d}x = 1$. Start with property (II):

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-3}^{-2} (cx+3) \,\mathrm{d}x + \int_{2}^{3} (3-cx) \,\mathrm{d}x = \left[\frac{c}{2}x^2 + 3x\right]_{-3}^{-2} + \left[3x - \frac{c}{2}x^2\right]_{2}^{3} = 6-5c.$$

So (II) leads to the conclusion that $c = 1$. Substituting this yields $f(x) = x + 3$ for $-3 \leq x \leq -2$, and $f(x) = 3 - x$ for $2 \leq x \leq 3$ (and $f(x) = 0$ elsewhere), so f also satisfies (I).

5.5 b Since $f(x) = 0$ for $x < -3$, $F(b) = 0$ for $b \leq -3$. For $-3 \leq b \leq -2$ we compute $F(b) = \int_{-\infty}^{b} f(x) \,\mathrm{d}x = \int_{-3}^{b} (x+3) \,\mathrm{d}x = (b+3)^2/2$. Similarly one finds $F(b) = 1 - (3-b)^2/2$ for $2 \leq b \leq 3$. For $-2 \leq x \leq 2$ $f(x) = 0$, and hence $F(b) = F(-2) = 1/2$ for $-2 \leq b \leq 2$. Finally $F(b) = 1$ for $b \geq 3$.

5.6 If X has an $Exp(0.2)$ distribution, then its distribution function is given by $F(b) = 1 - \mathrm{e}^{-\lambda b}$. Hence $\mathrm{P}(X > 5) = 1 - \mathrm{P}(X \leq 5) = 1 - F(5) = \mathrm{e}^{-5\lambda}$. Here $\lambda = 0.2$, so $\mathrm{P}(X > 5) = \mathrm{e}^{-1} = 0.367879\ldots$.

5.7 a $\mathrm{P}(\text{failure}) = \mathrm{P}(S < 0.55) = \int_{-\infty}^{0.55} f(x) \,\mathrm{d}x = \int_0^{0.55} 4x \,\mathrm{d}x + \int_{0.5}^{0.55} (4-4x) \,\mathrm{d}x = 0.595$.

5.7 b The 50-th percentile or median $q_{0.5}$ is obtained by solving $F(q_{0.5}) = 0.5$. For $b \leq 0.5$ we obtain $F(b) = \int_0^b 4x \,\mathrm{d}x = 2b^2$. Here $F(q_{0.5}) = q_{0.5}$ yields $2q_{0.5}^2 = 0.5$, or $q_{0.5} = 0.5$.

5.8 a The probability density $g(y) = 1/(2\sqrt{ry})$ has an asymptote in 0 and decreases to $1/2r$ in the point r. Outside $[0,r]$ the function is 0.

5.8 b The second darter is better: for each $0 < b < r$ one has $(b/r)^2 < \sqrt{b/r}$ so the second darter always has a larger probability to get closer to the center.

5.8 c Any function F that is 0 left from 0, increasing on $[0,r]$, takes the value 0.9 in $r/10$, and takes the value 1 in r and to the right of r is a correct answer to this question.

5.9 a The area of a triangle is 1/2 times base times height. So the largest area that can occur is 1/2 (when the y-coordinate of a point is 1).
The event $\{A \leq 1/4\}$ occurs if and only if the y-coordinate of the random point is at most 1/2, so the answer is $\{(x,y) : 2 \leq x \leq 3, 1 \leq y \leq 3/2\}$.

5.9 b Generalising **a** we see that $A \le a$ if and only if the y-coordinate of the random point is at most a. Hence $f(a) = \mathrm{P}(A \le a) = a$ for $0 \le a \le 1$. Furthermore $F(a) = 0$ for $a < 0$, and $F(a) = 1$ for $a \ge 1$.

5.9 c Differentiating F yields $f(x) = 2$ for $0 \le x \le 1/2$; $f(x) = 0$ elsewhere.

5.10 The residence time T has an $Exp(0.5)$ distribution, so its distribution function is $F(x) = 1 - \mathrm{e}^{-x/2}$. We are looking for $\mathrm{P}(T \le 2) = F(2) = 1 - \mathrm{e}^{-1} = 63.2\,\%$.

5.11 We have to solve $1 - \mathrm{e}^{-\lambda x} = 1/2$, or $\mathrm{e}^{-\lambda x} = 1/2$, which means that $-\lambda x = \ln(1/2)$ or $x = \ln(2)/\lambda$.

5.12 We have to solve $1 - x^{-\alpha} = 1/2$, or $x^{-\alpha} = 1/2$, which means that $x^{\alpha} = 2$ or $x = 2^{1/\alpha}$.

5.13 a This follows with a change of variable transformation $x \mapsto -x$ in the integral: $\Phi(-a) = \int_{-\infty}^{-a} \phi(x)\,\mathrm{d}x = \int_a^{\infty} \phi(-x)\,\mathrm{d}x = \int_a^{\infty} \phi(x)\,\mathrm{d}x = 1 - \Phi(a)$.

5.13 b This is straightforward: $\mathrm{P}(Z \le -2) = \Phi(-2) = 1 - \Phi(2) = 0.0228$.

5.14 The 10-th percentile $q_{0.1}$ is given by $\Phi(q_{0.1}) = 0.1$. To solve this we have to use Table **??**. Since the table only contains tail probabilities larger than 1/2, we have to use the symmetry of the normal distribution ($\phi(-x) = \phi(x)$): $\Phi(a) = 0.1$ if and only if $\Phi(-a) = 0.9$ if and only if $1 - \Phi(a) = 0.9$. This gives $-q_{0.1} = 1.28$, hence $q_{0.1} = -1.28$.

6.1 a If $0 \le U \le \frac{1}{6}$, put $X = 1$, if $\frac{1}{6} < U \le \frac{2}{6}$, put $X = 2$, etc., i.e., if $(i-1)/6 \le U < i/6$, then set $X = i$.

6.1 b The number $6U + 1$ is in the interval $[1, 7]$, so Y is one of the numbers 1, 2, 3, …, 7. For $k = 1, 2, \ldots, 6$ we see that $\mathrm{P}(Y = k) = \mathrm{P}(i \le 6U + 1 < i + 1) = \frac{1}{6}$; $\mathrm{P}(Y = 7) = \mathrm{P}(6U + 1 = 7) = \mathrm{P}(U = 1) = 0$.

6.2 a Substitute $u = 0.3782739$ in $x = 1 + 2\sqrt{u}$: $x = 1 + 2\sqrt{0.3782739} = 2.2300795$.

6.2 b The map $u \mapsto 1 + 2\sqrt{u}$ is increasing, so $u = 0.3$ will result in a smaller value, since $0.3 < 0.3782739$.

6.2 c Any number u smaller than 0.3782739 will result in a smaller realization than the one found in **a**. This happens with probability $\mathrm{P}(U < 0.3782739) = 0.3782739$.

6.3 The random variable Z attains values in the interval $[0, 1]$, so to show that its distribution is the $U(0, 1)$, it suffices to show that $F_Z(a) = a$, for $0 \le a \le 1$. We find:

$$F_Z(a) = \mathrm{P}(Z \le a) = \mathrm{P}(1 - U \le a) = \mathrm{P}(U \ge 1 - a) = 1 - F_U(1-a) = 1 - (1-a) = a.$$

6.4 Since $0 \le U \le 1$, $1 \le X \le 3$ follows, so $F_X(a) = 0$ for $a < 1$ and $F_X(a) = 1$ for $a > 3$. If we show that $F_X(a) = F(a)$ for $1 \le a \le 3$, then we have shown $F_X = F$:

$$\begin{aligned} F_X(a) &= \mathrm{P}(X \le a) = \mathrm{P}\left(1 + 2\sqrt{U} \le a\right) \\ &= \mathrm{P}\left(\sqrt{U} \le (a-1)/2\right) = \mathrm{P}\left(U \le ((a-1)/2)^2\right) \\ &= F_U\left(\frac{1}{4}(a-1)^2\right) = \frac{1}{4}(a-1)^2 = F(a). \end{aligned}$$

6.5 We see that

$$X \le a \quad \Leftrightarrow \quad -\ln U \le a \quad \Leftrightarrow \quad \ln U \ge -a \quad \Leftrightarrow \quad U \ge \mathrm{e}^{-a},$$

and so $\mathrm{P}(X \le a) = \mathrm{P}\big(U \ge \mathrm{e}^{-a}\big) = 1 - \mathrm{P}\big(U \le \mathrm{e}^{-a}\big) = 1 - \mathrm{e}^{-a}$, where we use $\mathrm{P}(U \le p) = p$ for $0 \le p \le 1$ applied to $p = \mathrm{e}^{-a}$ (remember that $a \ge 0$).

6.6 We know that $X = -\frac{1}{2}\ln U$ has an *Exp* (2) distribution when U is $U(0,1)$. Inverting this relationship, we find $U = \mathrm{e}^{-2X}$, which should be the way to get U from X. We check this, for $0 < a \le 1$:

$$\begin{aligned} \mathrm{P}(U \le a) &= \mathrm{P}\left(\mathrm{e}^{-2X} \le a\right) = \mathrm{P}(-2X \le \ln a) \\ &= \mathrm{P}\left(X \ge -\frac{1}{2}\ln a\right) = \mathrm{e}^{-2(-\frac{1}{2}\ln a)} \\ &= a = F_U(a). \end{aligned}$$

6.7 We need to obtain F^{inv}, and do this by solving $F(x) = u$, for $0 \le u \le 1$:

$$\begin{aligned} 1 - \mathrm{e}^{-5x^2} = u \quad &\Leftrightarrow \quad \mathrm{e}^{-5x^2} = 1 - u \quad \Leftrightarrow \quad -5x^2 = \ln(1-u) \\ &\Leftrightarrow \quad x^2 = -0.2\ln(1-u) \quad \Leftrightarrow \quad x = \sqrt{-0.2\ln(1-u)}. \end{aligned}$$

The solution is $Z = \sqrt{-0.2\ln U}$ (replacing $1 - U$ by U, see Exercise 6.3). Note that Z^2 has an *Exp* (5) distribution.

6.8 We need to solve $F(x) = u$ for $0 \le u \le 1$ to obtain $F^{\mathrm{inv}}(x)$:

$$1 - x^{-3} = u \quad \Leftrightarrow \quad 1 - u = x^{-3} \quad \Leftrightarrow \quad x = (1-u)^{-\frac{1}{3}}.$$

So, $X = (1-U)^{-\frac{1}{3}}$ has a *Par* (3) distribution, and the same holds for $X = U^{-\frac{1}{3}} = 1/\sqrt[3]{U}$.

6.9 a For six of the eight possible outcomes our algorithm terminates succesfully, so the probability of success is $6/8 = 3/4$.

6.9 b If the first toss (of the *three* coins) is unsuccessful, we try again, and repeat until the first success. This way, we generate a sequence of *Ber* (3/4) trials and stop after the first success. From Section 4.4 we know that the number N of trials needed has a *Geo* (3/4) distribution.

6.10 a Define random variables $B_i = 1$ if $U_i \le p$ and $B_i = 0$ if $U_i > p$. Then $\mathrm{P}(B_i = 1) = p$ and $\mathrm{P}(B_i = 0) = 1 - p$: each B_i has a *Ber* (p) distribution. If $B_1 = B_2 = \cdots = B_{k-1} = 0$ and $B_k = 1$, then $N = k$, i.e., N is the position in the sequence of Bernoulli random variables, where the first 1 occurs. This is a *Geo* (p) distribution. This can be verified by computing the probability mass function: for $k \ge 1$,

$$\begin{aligned} \mathrm{P}(N = k) &= \mathrm{P}(B_1 = B_2 = \cdots = B_{k-1} = 0, B_k = 1) \\ &= \mathrm{P}(B_1 = 0)\ \mathrm{P}(B_2 = 0) \cdots \mathrm{P}(B_{k-1} = 0)\ \mathrm{P}(B_k = 1) \\ &= (1-p)^{k-1}\, p. \end{aligned}$$

6.10 b If Y is (a real number!) greater than n, then rounding *upwards* means we obtain $n+1$ or higher, so $\{Y > n\} = \{Z \geq n+1\} = \{Z > n\}$. Therefore, $\mathrm{P}(Z > n) = \mathrm{P}(Y > n) = \mathrm{e}^{-\lambda n} = \left(\mathrm{e}^{-\lambda}\right)^n$. From $\lambda = -\ln(1-p)$ we see: $\mathrm{e}^{-\lambda} = 1-p$, so the last probability is $(1-p)^n$. From $\mathrm{P}(Z > n-1) = \mathrm{P}(Z = n) + \mathrm{P}(Z > n)$ we find: $\mathrm{P}(Z = n) = \mathrm{P}(Z > n-1) - \mathrm{P}(Z > n) = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1}\,p$. Z has a $Geo(p)$ distribution.

6.11 a There are many possibilities, one is: $Y_1 = g + 0.5 + Z_1$, where g and Z_1 are as defined earlier. In this case, the bribed jury member adds a bonus of 0.5 points. Of course, this is assuming that she was bribed to *support* the candidate. If she was bribed against her, $Y_1 = g - 0.5 + Z_1$ would be more appropriate.

6.11 b We call the resulting deviation R and it is the average of $Z_1, Z_2, \ldots, Z_7$ after we have removed two of them, chosen randomly. This can be done as follows. Let I and J be independent random variables, such that $\mathrm{P}(I = n) = 1/7$ for $n = 1, 2, \ldots, 7$ and $\mathrm{P}(J = m) = 1/6$ for $m = 1, 2, \ldots, 6$. Put $K = J$ if $J < I$; otherwise, put $K = J + 1$. Now, the pair $\{I, K\}$ is a random pair chosen from the set of indices $\{1, 2, \ldots, 7\}$. This can be verified, for example, the probability that 1 and 2 are selected equals $\mathrm{P}(I = 1, J = 1) + \mathrm{P}(I = 2, J = 1) = 1/21$, which is correct, because there are $\binom{7}{2} = 21$ pairs to choose from, each we the same probability.
Removing Z_I and Z_K from the jury list, we can compute R as the average of the remaining ones. We expect this rule to be *more* sensitive to bribery, because the bribed jury member is more likely to have assigned one of the extreme scores (highest or lowest). With either of the two other rules, this score has no influence at all, because it is not taken into account.

6.12 We need to generate stock prices for the next five years, or 60 months. So we need sixty $U(0,1)$ random variables $U_1, \ldots, U_{60}$. Let S_i denote the stock price in month i, and set $S_0 = 100$, the initial stock price. From the U_i we obtain the stock movement, as follows, for $i = 1, 2, \ldots$:

$$S_i = \begin{cases} 0.95\, S_{i-1} & \text{if } U_i < 0.25, \\ S_{i-1} & \text{if } 0.25 \leq U_i \leq 0.75, \\ 1.05\, S_{i-1} & \text{if } U_i > 0.75. \end{cases}$$

We have carried this out, using the realizations below:

1–10:	0.72	0.03	0.01	0.81	0.97	0.31	0.76	0.70	0.71	0.25
11–20:	0.88	0.25	0.89	0.95	0.82	0.52	0.37	0.40	0.82	0.04
21–30:	0.38	0.88	0.81	0.09	0.36	0.93	0.00	0.14	0.74	0.48
31–40:	0.34	0.34	0.37	0.30	0.74	0.03	0.16	0.92	0.25	0.20
41–50:	0.37	0.24	0.09	0.69	0.91	0.04	0.81	0.95	0.29	0.47
51–60:	0.19	0.76	0.98	0.31	0.70	0.36	0.56	0.22	0.78	0.41

We do not list all the stock prices, just the ones that matter for our investment strategy (you can verify this). We first wait until the price drops below € 95, which happens at $S_4 = 94.76$. Our money has been in the bank for four months, so we own € $1000 \cdot 1.005^4 =$ € 1020.15, for which we can buy $1020.15/94.76 = 10.77$ shares. Next we wait until the price hits € 110, this happens at $S_{15} = 114.61$. We sell the our shares for € $10.77 \cdot 114.61 =$ € 1233.85, and put the money in the bank. At $S_{42} = 92.19$ we buy stock again, for the € $1233.85 \cdot 1.005^{27} =$ € 1411.71 that has accrued in the bank. We can buy 15.31 shares. For the rest of the five year period nothing happens, the final price is $S_{60} = 100.63$, which puts the value of our portfolio at € 1540.65.

For a real simulation the above should be repeated, say, one thousand times. The one thousand net results then give us an impression of the probability distribution that corresponds to this model and strategy.

6.13 Set $p = \mathrm{P}(H)$. Toss the coin twice. If HT occurs, I win; if TH occurs, you win; if TT or HH occurs, we repeat. Since $\mathrm{P}(HT) = \mathrm{P}(TH) = p\,(1-p)$, we have the same probability of winning. The probability that the game ends is $2\,p\,(1-p)$ for each double toss. So, if $2N$ is the total number of tosses needed, where N has a $Geo(2\,p\,(1-p))$ distribution.

7.1 a Outcomes: 1, 2, 3, 4, 5, and 6. Each has probability 1/6.

7.1 b $\mathrm{E}[T] = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \cdots \frac{1}{6}\cdot 6 = \frac{1}{6}(1+2=\cdots+6) = \frac{7}{2}$.
For the variance, first compute $\mathrm{E}\left[T^2\right]$:

$$\mathrm{E}\left[T^2\right] = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 4 + \cdots \frac{1}{6}\cdot 36 = \frac{91}{6}.$$

Then: $\mathrm{Var}(T) = \mathrm{E}\left[T^2\right] - (\mathrm{E}[T])^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$.

7.2 a Here $\mathrm{E}[X] = (-1)\mathrm{P}(X=-1) + 0\cdot\mathrm{P}(X=0) + 1\cdot\mathrm{P}(X=1) = -\frac{1}{5} + 0 + \frac{2}{5} = \frac{1}{5}$.

7.2 b The discrete random variable Y can take the values 0 and 1. The probabilities are $\mathrm{P}(Y=0) = \mathrm{P}\big(X^2=0\big) = \mathrm{P}(X=0) = \frac{2}{5}$, and $\mathrm{P}(Y=1) = \mathrm{P}\big(X^2=1\big) = \mathrm{P}(X=-1) + \mathrm{P}(X=1) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$. Moreover, $\mathrm{E}[Y] = 0\cdot\frac{2}{5} + 1\cdot\frac{3}{5} = \frac{3}{5}$

7.2 c According to the change of variable formula: $\mathrm{E}\left[X^2\right] = (-1)^2\cdot\frac{1}{5} + 0^2\cdot\frac{2}{5} + 1^2\cdot\frac{2}{5} = \frac{3}{5}$.

7.2 d With the alternative expression for the variance: $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{3}{5} - (\frac{1}{5})^2 = \frac{14}{25}$.

7.3 Since $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$,we find $\mathrm{E}\left[X^2\right] = \mathrm{Var}(X) + (\mathrm{E}[X])^2 = 7$.

7.4 By the change of units rule: $\mathrm{E}[3-2X] = 3 - 2\mathrm{E}[X] = -1$, and $\mathrm{Var}(3-2X) = 4\mathrm{Var}(X) = 16$.

7.5 If X has a $Ber(p)$ distribution, then $\mathrm{E}[X] = 0\cdot(1-p) + 1\cdot p = p$ and $\mathrm{E}[X]^2 = 0\cdot(1-p) + 1.p = p$. So $\mathrm{Var}(X) = \mathrm{E}[X]^2 - (\mathrm{E}[X])^2 = p - p^2 = p(1-p)$.

7.6 Since f is increasing on the interval $[2,3]$ we know from the interpretation of expectation as center of gravity that the expectation should lie closer to 3 than to 2. The computation: $\mathrm{E}[Z] = \int_2^3 \frac{3}{19} z^3 \,\mathrm{d}z = \left[\frac{3}{76} z^4\right]_2^3 = 2\frac{43}{76}$.

7.7 We use the rule for expectation and variance under change of units. First, $\mathrm{E}[X] = \int_0^1 x\cdot(4x - 4x^3)\,\mathrm{d}x = \int_0^1 4x^2 - 4x^4\,\mathrm{d}x = \left[(4/3x^3 - 4/5x^5)\right]_0^1 = 4/3 - 4/5 = 8/15$. Then, changing units, $\mathrm{E}[2X+3] = 2\cdot 8/15 + 3 = 61/15$. For the variance, first compute $\mathrm{E}\left[X^2\right] = \int_0^1 x^2\cdot(4x - 4x^3)\,\mathrm{d}x = 1/3$. Hence $\mathrm{Var}(X) = 1/3 - (8/15)^2 = 11/225$, and changing units, $\mathrm{Var}(2X+3) = 44/225$.

7.8 Let f be the probability density of X Then $f(x) = \frac{d}{dx}F(x) = 2 - 2x$ for $0 \le x \le 1$, and $f(x) = 0$ elsewhere. So $\mathrm{E}[X] = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_0^1 (2x - 2x^2)\,\mathrm{d}x = [x^2 - \frac{2}{3}x^3]_0^1 = \frac{1}{3}$.

7.9 a If X has $U(\alpha,\beta)$ distribution, then X has probability density function $f(x) = 1/(\beta-\alpha)$ for $\alpha \le x \le \beta$, and 0 elsewhere. So

$$\mathrm{E}[X] = \int_\alpha^\beta x/(\beta-\alpha)\,\mathrm{d}x = 1/(\beta-\alpha)\cdot\left[\frac{1}{2}x^2\right]_\alpha^\beta = \frac{1}{2}\frac{(\beta^2-\alpha^2)}{\beta-\alpha} = \frac{1}{2}(\alpha+\beta).$$

7.9 b First compute $\mathrm{E}\left[X^2\right]$: $\mathrm{E}[X] = \int_\alpha^\beta \frac{x^2}{(\beta-\alpha)}\,\mathrm{d}x = \frac{1}{3}\frac{(\beta^3-\alpha^3)}{\beta-\alpha} = \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2)$. Then $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2) - \frac{1}{4}(\alpha+\beta)^2 = \frac{1}{12}(\beta-\alpha)^2$. A quicker way to do this is to note that X has the same variance as a $U(0,\beta-\alpha)$ distribution U. Hence $\mathrm{Var}(X) = \mathrm{Var}(U) = \frac{1}{12}(\beta-\alpha)^2$.

7.10 a If X has an $Exp(X)$ distribution, then X has probability density $f(X) = \lambda\mathrm{e}^{-\lambda x}$ for $x \geq 0$, and $f(x) = 0$ elsewhere. So, using the partial integration formula, $\mathrm{E}[X] = \int_0^\infty x\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \left[-x\mathrm{e}^{-\lambda x}\right]_0^\infty - \int_0^\infty -\mathrm{e}^{\lambda x}\,\mathrm{d}x = \left[-\frac{1}{\lambda}\mathrm{e}^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}$.

7.10 b First compute $\mathrm{E}\left[X^2\right]$, using partial integration, and using the result from part **a**:

$$\mathrm{E}\left[X^2\right] = \int_0^\infty x^2\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \left[-x^2\mathrm{e}^{-\lambda x}\right]_0^\infty - \int_0^\infty -2x\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda}\int_0^\infty x\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda^2}.$$

Then $\mathrm{Var}(X) = \mathrm{E}[X]^2 - (\mathrm{E}[X])^2 = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$.

7.11 a If X has a $Par(\alpha)$ distribution, then X has p.d $f(x) = \alpha x^{-\alpha-1}$ for $x \geq 1$, and $f(x) = 0$ elsewhere. So the $Par(2)$ distribution has probability density $f(x) = 2x^{-3}$, and then $\mathrm{E}[X] = \int_1^\infty x\cdot 2x^{-3}\,\mathrm{d}x = 2\int_1^\infty x^{-2}\,\mathrm{d}x = \left[-2x^{-1}\right]_1^\infty = 2$.

7.11 b Now X has probability density function $f(x) = \frac{1}{2}x^{-3/2}$. So now $\mathrm{E}[X] = \int_1^\infty \frac{1}{2}x^{-1/2}\,\mathrm{d}x = \left[\sqrt{x}\right]_1^\infty = \infty$.

7.11 c In the general case $\mathrm{E}[X] = \int_1^\infty x\frac{\alpha}{x^{\alpha+1}}\,\mathrm{d}x = \alpha\int_1^\infty x^{-\alpha}\,\mathrm{d}x = \alpha\left[\frac{x^{-\alpha+1}}{-\alpha+1}\right]_1^\infty = \frac{\alpha}{\alpha-1}$, provided $x^{-\alpha+1} \to 0$ as $x \to \infty$, which happens for all $\alpha > 1$.

7.12 From 7.11**c** we already know that $\mathrm{E}[X] = \alpha/(\alpha-1)$ if and only of $\alpha > 1$. Now we compute $\mathrm{E}\left[X^2\right]$: $\mathrm{E}\left[X^2\right] = \alpha\int_1^\infty x^{-\alpha+1}\,\mathrm{d}x = [\alpha\cdot\frac{x^{-\alpha+2}}{-\alpha+2}]_1^\infty = \frac{\alpha}{\alpha-2}$ provided $\alpha > 2$ (otherwise $\mathrm{E}\left[X^2\right] = \infty$). It follow that $\mathrm{Var}(X)$ exists only if $\alpha > 2$, and then $\mathrm{Var}(X) = \mathrm{E}[X]^2 - (\mathrm{E}[X])^2 = \frac{\alpha}{(\alpha-2)} - \frac{\alpha^2}{(\alpha-1)^2} = \frac{\alpha[\alpha^2-2\alpha+1-\alpha^2+2\alpha]}{(\alpha-1)^2(\alpha-2)} = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}$.

7.13 Since $\mathrm{Var}(X) = \mathrm{E}\left[\left(X - \mathrm{E}[X]\right)^2\right] \geq 0$, but also $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X]^2)$ we must have that $\mathrm{E}\left[X^2\right] \geq (\mathrm{E}[X])^2$.

7.14 The area of a triangle is half the length of the base time the height. Hence $A = \frac{1}{2}Y$, where Y is $U(0,1)$ distributed. It follows that $\mathrm{E}[A] = \frac{1}{2}\mathrm{E}[Y] = \frac{1}{4}$.

7.15 a We use the change-of-units rule for the expectation twice:

$$\begin{aligned}\mathrm{Var}(rX) &= \mathrm{E}\left[(rX - \mathrm{E}[rX])^2\right] = \mathrm{E}\left[(rX - r\mathrm{E}[X])^2\right] \\ &= \mathrm{E}\left[r^2(X - \mathrm{E}[X])^2\right] = r^2\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = r^2\mathrm{Var}(X)\,.\end{aligned}$$

7.15 b Now we use the change-of-units rule for the expectation once:

$$\begin{aligned}\mathrm{Var}(X+s) &= \mathrm{E}\left[((X+s) - \mathrm{E}[X+s])^2\right] \\ &= \mathrm{E}\left[((X+s) - \mathrm{E}[X]+s)^2\right] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{Var}(X)\,.\end{aligned}$$

7.15 c With first **b**, and then **a**: $\mathrm{Var}(rX+s) = \mathrm{Var}(rX) = r^2\mathrm{Var}(X)$.

7.16 We have to use partial integration: $\mathrm{E}[X] = \int_0^1 -4x^2\ln x\,\mathrm{d}x = \left[-\frac{4}{3}x^3\ln x\right]_0^1 + \int_0^1 \frac{4}{3}x^3\cdot\frac{1}{x}\,\mathrm{d}x = \frac{4}{3}\int_0^1 x^2\,\mathrm{d}x = \frac{4}{9}$.

7.17 a Since $a_i \geq 0$ and $p_i \geq 0$ it must follow that $a_1p_1 + \cdots + a_rp_r \geq 0$. So $0 = \mathrm{E}[U] = a_1p_1 + \cdots + a_rp_r \geq 0$. As we may assume that all $p_i > 0$, it follows that $a_1 = a_2 = \cdots = a_r = 0$.

7.17 b Let $m = \mathrm{E}[V] = p_1b_1 + \cdots + p_rb_r$. Then the random variable $U = (V - \mathrm{E}[V])^2$ takes the values $a_1 = (b_1 - m)^2, \ldots, a_r = (b_r - m)^2$. Since $\mathrm{E}[U] = \mathrm{Var}(V) = 0$, part **a** tells us that $0 = a_1 = (b_1 - m)^2, \ldots, 0 = a_r = (b_r - m)^2$. But this is only possible if $b_1 = m, \ldots, b_r = m$. Since $m = \mathrm{E}[V]$, this is the same as saying that $\mathrm{P}(V = \mathrm{E}[V]) = 1$.

8.1 The random variable Y can take the values $|80 - 100| = 20, |90 - 100| = 10, |100 - 100| = 0, |110 - 100| = 10$ and $|120 - 100| = 20$. We see that the values are $0, 10$ and 20, and the latter two occur in two ways: $\mathrm{P}(Y = 0) = \mathrm{P}(X = 100) = 0.2$, but $\mathrm{P}(Y = 10) = \mathrm{P}(X = 110) + \mathrm{P}(X = 90) = 0.4$; $\mathrm{P}(Y = 20) = \mathrm{P}(X = 120) + \mathrm{P}(X = 80) = 0.4$.

8.2 a First we determine the possible values that Y can take. Here these are $-1, 0$, and 1. Then we investigate which x-values lead to these y-values and sum the probabilities of the x-values to obtain the probability of the y-value. For instance,

$$\mathrm{P}(Y = 0) = \mathrm{P}(X = 2) + \mathrm{P}(X = 4) + \mathrm{P}(X = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Similarly, we obtain for the two other values

$$\mathrm{P}(Y = -1) = \mathrm{P}(X = 3) = \frac{1}{6}, \qquad \mathrm{P}(Y = 1) = \mathrm{P}(X = 1) + \mathrm{P}(X = 5) = \frac{1}{3}.$$

8.2 b The values taken by Z are -1, 0, and 1. Furthermore

$$\mathrm{P}(Z = 0) = \mathrm{P}(X = 1) + \mathrm{P}(X = 3) + \mathrm{P}(X = 5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

and similarly $\mathrm{P}(Z = -1) = 1/3$ and $\mathrm{P}(Z = 1) = 1/6$.

8.2 c Since for any α one has $\sin^2(\alpha) + \cos^2(\alpha) = 1$, W can only take the value 1, so $\mathrm{P}(W = 1) = 1$.

8.3 a Let F be the distribution function of U, and G the distribution function of V. Then we know that $F(x) = 0$ for $x < 1, F(x) = x$ for $0 \leq x \leq 1$, and $F(x) = 1$ for $x > 1$. Thus

$$G(y) = \mathrm{P}(V \leq y) = \mathrm{P}(2U + 7 \leq y) = \mathrm{P}\left(U \leq \frac{y-7}{2}\right) = \frac{y-7}{2},$$

provided $0 \leq (y - 7)/2 \leq 1$ which happens if and only if $0 \leq y - 7 \leq 2$ if and only $7 \leq y \leq 9$. Furthermore, $G(y) = 0$ if $y < 7$ and $G(y) = 1$ if $y > 9$. We recognize G as the distribution function of a $U(7, 9)$ random variable.

8.3 b Let F and G be as before. Now

$$G(y) = \mathrm{P}(V \leq y) = \mathrm{P}(rU + s \leq y) = \mathrm{P}\left(U \leq \frac{y-s}{r}\right) = \frac{y-s}{r},$$

provided $0 \leq (y - s)/r \leq 1$, which happens if and only if $0 \leq y - s \leq r$ (note that we use that $r > 0$), if and only if $s \leq y \leq s + r$. We see that G has a $U(s, s + r)$ distribution.

8.4 a Let F be the distribution function of X, and G that of Y. Then we know that $F(x) = 1 - \mathrm{e}^{-x/2}$ for $x \geq 0$, and we find that $G(y) = \mathrm{P}(Y \leq y) = \mathrm{P}\big(\frac{1}{2}X \leq y\big) = \mathrm{P}(X \leq 2y) = 1 - \mathrm{e}^{-y}$. We recognize G as the distribution function of an $Exp(1)$ distribution.

8.4 b Let F be the distribution function of X, and G that of Y. Then we know that $F(x) = 1 - \mathrm{e}^{-\lambda x}$ for $x \geq 0$, and we find that $G(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(\lambda X \leq y) = \mathrm{P}(X \leq y/\lambda) = 1 - \mathrm{e}^{-y}$. We recognize G as the distribution function of an $Exp(1)$ distribution.

8.5 a For $0 \leq b \leq 2$ we have that $F_X(b) = \int_0^b \frac{3}{4}x(2-x)\,\mathrm{d}x = \big[\frac{3}{4}x^2 - \frac{1}{4}x^3\big]_0^b = \frac{3}{4}b^2 - \frac{1}{4}b^3$. Furthermore $F_X(b) = 0$ for $b < 0$, and $F_X(b) = 1$ for $b > 2$.

8.5 b For $0 \leq y \leq \sqrt{2}$ we have $F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}\Big(\sqrt{X} \leq y\Big) = \mathrm{P}\big(X \leq y^2\big) = \frac{3}{4}y^4 - \frac{1}{4}y^6$.

8.5 c We simply differentiate F_Y : $f_Y(y) = \frac{d}{dy}F_Y(y) = 3y^3 - 3y^5/2$ for $0 \leq y \leq \sqrt{2}$, and $f_Y(y) = 0$ elsewhere.

8.6 a Compute the distribution function of Y : $F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}\big(\frac{1}{X} \leq y\big) = \mathrm{P}\Big(X \geq \frac{1}{y}\Big) = 1 - \mathrm{P}\Big(X < \frac{1}{y}\Big) = 1 - F_X(\frac{1}{y})$, where you use that $\mathrm{P}\Big(X < \frac{1}{y}\Big) = \mathrm{P}\Big(X \leq \frac{1}{y}\Big)$ since X has a continuous distribution. Differentiating we obtain: $f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}(1 - F_X(\frac{1}{y})) = \frac{1}{y^2}f_X(\frac{1}{y})$ for $y > 0$ ($f_Y(y) = 0$ for $y \leq 0$).

8.6 b Applying part **a** with $Z = 1/Y$ we obtain $f_Z(z) = \frac{1}{z^2}f_Y(\frac{1}{z})$. Then applying **a** again: $f_Z(z) = \frac{1}{z^2}f_Y\Big(\frac{1}{z}\Big) = \frac{1}{z^2}\frac{1}{(1/z)^2}f_X\Big(\frac{1}{(1/z)}\Big)) = f_X(z)$.
This is of course what should happen: $Z = 1/Y = 1/(1/X) = X$, so Z and X have the same probability density function.

8.7 Let X be any random variable that only takes positive values, and let $Y = \ln(X)$. Then for $y > 0$:

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(\ln(X) \leq y) = \mathrm{P}(X \leq \mathrm{e}^y) = F_X(\mathrm{e}^y).$$

If X has a $Par(\alpha)$ distribution, then $F_X(x) = 1 - x^{-\alpha}$ for $x \geq 1$. Hence $F_Y(y) = 1 - e^{-\alpha y}$ for $y \geq 0$ is the distribution function of Y. We recognize this as the distribution function of an $Exp(\alpha)$ distribution.

8.8 Let X be any random variable that only takes positive values, and let $W = X^{1/\alpha}/\lambda$. Then for $w > 0$:

$$F_W(w) = \mathrm{P}(W \leq w) = \mathrm{P}\Big(X^{1/\alpha}/\lambda \leq w\Big) = \mathrm{P}(X \leq (\lambda w)^\alpha) = F_X((\lambda w)^\alpha).$$

If X has an $Exp(1)$ distribution, then $F_X(x) = 1 - \mathrm{e}^{-x}$ for $x \geq 0$. Hence $F_W(w) = 1 - \mathrm{e}^{-(\lambda w)^\alpha}$ for $w \geq 0$.

8.9 If $Y = -X$, then $F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(-X \leq y) = \mathrm{P}(X \geq -y) = 1 - F_X(-y)$ for all Y (where you use that X has a continuous distribution). Differentiating we obtain $f_Y(y) = f_X(-y)$ for all y.

8.10 Because of symmetry: $\mathrm{P}(X \geq 3) = 0.500$. Furthermore: $\sigma^2 = 4$, so $\sigma = 2$. Then $Z = (X-3)/2$ is an $N(0,1)$ distributed random variable, so that $\mathrm{P}(X \leq 1) = \mathrm{P}((X-3)/2) \leq (1-3)/2 = \mathrm{P}(Z \leq -1) = \mathrm{P}(Z \geq 1) = 0.1587$.

8.11 Since $-g$ is a convex function, Jensen's inequality yields that $-g(\mathrm{E}[X]) \leq \mathrm{E}[-g(X)]$. Since $\mathrm{E}[-g(X)] = -\mathrm{E}[g(X)]$, the inequality follows by multiplying both sides by -1.

8.12 a The possible values Y can take are $\sqrt{0} = 0$, $\sqrt{1} = 1$, $\sqrt{100} = 10$, and $\sqrt{10\,000} = 100$. Hence the probability mass function is given by

y	0	1	10	100
$\mathrm{P}(Y = y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

8.12 b Compute the second derivative: $\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sqrt{x} = -\frac{1}{4}x^{-3/2} < 0$. Hence $g(x) = -\sqrt{x}$ is a convex function. Jensen's inequality yields that $\sqrt{\mathrm{E}[X]} \geq \mathrm{E}\left[\sqrt{X}\right]$.

8.12 c We obtain $\sqrt{\mathrm{E}[X]} = \sqrt{(0 + 1 + 100 + 10\,000)/4} = 50.25$, but

$$\mathrm{E}\left[\sqrt{X}\right] = \mathrm{E}[Y] = (0 + 1 + 10 + 100)/4 = 27.75.$$

8.13 On the interval $[\pi, 2\pi]$ the function $\sin w$ is a convex function, so by Jensen's inequality $\sin(\mathrm{E}[W]) \leq \mathrm{E}[\sin(W)]]$. Verified by computations : $\sin(\mathrm{E}[W]) = \sin(\frac{3}{2}\pi) = -1 \leq \mathrm{E}[\sin(W)] = \int_\pi^{2\pi} \sin(w)/\pi \,\mathrm{d}w = [-\cos(w)/\pi]_\pi^{2\pi} = -2/\pi$.

8.14 a An example is X with $\mathrm{P}(X = -1) = \mathrm{P}(X = 1) = \frac{1}{2}$. Then $\mathrm{E}[X] = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0$ and also $\mathrm{E}\left[X^3\right] = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot (1) = 0$.

8.14 b The function $g(x) = x^3$ is strictly convex on the interval $(0, \infty)$. Hence if $X > 0$ (and X is not constant) the inequality will hold.

8.15 a We know that $\mathrm{P}(Z \leq z) = [\mathrm{P}(X_1 \leq z)]^2$. So $F_Z(z) = z^2$ for $0 \leq z \leq 1$, and $f_Z(z) = 2z$ on this interval. Therefore $\mathrm{E}[Z] = \int_0^1 2z^2 \,\mathrm{d}z = \frac{2}{3}$. For V we have $F_V(v) = 1 - (1 - v)^2 = 2v - v^2$. Hence $f_V(v) = 2 - 2v$. Therefore $\mathrm{E}[V] = \int_0^1 (2v - 2v^2) \,\mathrm{d}v = 1 - 2/3 = 1/3$.

8.15 b For general n we have $F_Z(z) = z^n$ for $0 \leq z \leq 1$, and $f_Z(z) = nz^{n-1}$. Therefore $\mathrm{E}[Z] = \int_0^1 nz^n \,\mathrm{d}z = \frac{n}{n+1}$. For V we have $F_V(v) = 1 - (1 - v)^n$, and $f_V(v) = n(1 - v)^{n-1}$. Therefore $\mathrm{E}[V] = \int_0^1 nv(1 - v)^{n-1} \,\mathrm{d}v = \int_0^1 n(1 - u)u^{n-1} \,\mathrm{d}u = n\int_0^1 u^{n-1} \,\mathrm{d}u - n\int_0^1 u^n \,\mathrm{d}u = n.\frac{1}{n} - n.\frac{1}{n+1} = \frac{1}{n+1}$.

8.15 c Since $Y_i = 1 - X_i$ are also uniform we get $\mathrm{E}[Z] = \mathrm{E}[\max\{Y_1, \ldots, Y_n\}] = \mathrm{E}[\max\{1 - X_1, \ldots, 1 - X_n\}] = 1 - \mathrm{E}[\min\{X_1, \ldots, X_n\}] = 1 - \mathrm{E}[V]$.

8.16 a Suppose first that $a \leq b$ Then $\min\{a, b\} = a$, and also $|a - b| = b - a$. So $a + b - |a - b| = a + b - b + a = 2a$, and the formula holds. If $a > b$ then $\min\{a, b\} = b$, and $a + b - |a - b| = a + b - a + b = 2b$, and the formula holds also for this case.

8.16 b From part **a** and linearity of expectations we have

$$\begin{aligned}\mathrm{E}[\min\{X, Y\}] &= \mathrm{E}[(X + Y - |X - Y|)/2] = \frac{1}{2}\mathrm{E}[X] + \frac{1}{2}\mathrm{E}[Y] - \frac{1}{2}\mathrm{E}[X - Y|] \\ &= \mathrm{E}[X] - \frac{1}{2}\mathrm{E}[|X - Y|],\end{aligned}$$

since $\mathrm{E}[X] = \mathrm{E}[Y]$.

8.16 c From **b** we obtain, since $\mathrm{E}[|]\,X - Y| \geq 0$ that $\mathrm{E}[\min\{X, Y\}] \leq \mathrm{E}[X]$. Interchanging X and Y we also have $\mathrm{E}[\min\{X, Y\}] \leq \mathrm{E}[Y]$. Combining these two we have inequalities we obtain $\mathrm{E}[\min\{X, Y\}] \leq \min\{\mathrm{E}[X], \mathrm{E}[Y]\}$.

8.17 a For $n = 2$: if $x_1 \leq x_2$ then $\min\{x_1, x_2\} = x_1$ and $\max\{-x_1, -x_2\} = -x_1$ since $-x_1 \geq -x_2$; similarly for the case $x_1 > x_2$. In general, if x_i is the smallest of $x_1, \ldots, x_n$ then $-x_i$ will be the largest of $-x_1, \ldots, -x_n$.

8.17 b Using part **a** we have

$$\begin{aligned} F_V(a) &= \mathrm{P}(\min\{X_1, \ldots, X_n\} \leq a) = \mathrm{P}(-\max\{-X_1, \ldots, -X_n\} \leq a) \\ &= \mathrm{P}(\max\{-X_1, \ldots, -X_n\} \geq -a) = 1 - \mathrm{P}(\max\{-X_1, \ldots, -X_n\} \leq -a) \\ &= 1 - (\mathrm{P}(-X_1 \leq -a))^n = 1 - (1 - \mathrm{P}(X_1 \leq a))^n, \end{aligned}$$

using Exercise 8.9 in the last step.

8.18 The distribution function of each of the X_i is $F(x) = 1 - \mathrm{e}^{-\lambda x}$. Then the distribution function of V is $1 - (1 - F(x))^n = 1 - \mathrm{e}^{-\lambda n x}$. This is the distribution function of an $Exp(n\lambda)$ random variable.

8.19 a This happens for all φ in the interval $[\pi/4, \pi/2]$, which corresponds to the upper right quarter of the circle.

8.19 b Since $\{Z \leq t\} = \{X \leq \arctan(t)\}$, we obtain

$$F_Z(t) = \mathrm{P}(Z \leq t) = \mathrm{P}(X \leq \arctan(t)) = \frac{1}{2} + \frac{1}{\pi} \arctan(t).$$

8.19 c Differentiating F_Z we obtain that the probability density function of Z is

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{1}{2} + \frac{1}{\pi} \arctan(z) \right) = \frac{1}{\pi(1 + z^2)} \quad \text{for } -\infty < z < \infty.$$

9.1 For a and b from 1 to 4 we have

$$\mathrm{P}(X = a) = \mathrm{P}(X = a, Y = 1) + \cdots + \mathrm{P}(X = a, Y = 4) = \frac{1}{4},$$

and

$$\mathrm{P}(Y = b) = \mathrm{P}(X = 1, Y = b) + \cdots + \mathrm{P}(X = 4, Y = b) = \frac{1}{4}.$$

9.2 a From $\mathrm{P}(X = 1, Y = 1) = 1/2$, $\mathrm{P}(X = 1) = 2/3$, and the fact that $\mathrm{P}(X = 1) = \mathrm{P}(X = 1, Y = 1) + \mathrm{P}(X = 1, Y = -1)$, it follows that $\mathrm{P}(X = 1, Y = -1) = 1/6$. Since $\mathrm{P}(Y = 1) = 1/2$ and $\mathrm{P}(X = 1, Y = 1) = 1/2$, we must have: $\mathrm{P}(X = 0, Y = 1)$ and $\mathrm{P}(X = 2, Y = 1)$ are both zero. From this and the fact that $\mathrm{P}(X = 0) = 1/6 = \mathrm{P}(X = 2)$ one finds that $\mathrm{P}(X = 0, Y = -1) = 1/6 = \mathrm{P}(X = 2, Y = -1)$.

9.2 b Since, e.g., $\mathrm{P}(X = 2, Y = 1) = 0$ is different from $\mathrm{P}(X = 2)\,\mathrm{P}(Y = 1) = \frac{1}{6} \cdot \frac{1}{2}$, one finds that X and Y are dependent.

9.3 a $\mathrm{P}(X = Y) = \mathrm{P}(X = 1, Y = 1) + \cdots + \mathrm{P}(X = 4, Y = 4) = \frac{1}{4}$.

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

9.3 c $\mathrm{P}(1 < X \leq 3, 1 < Y \leq 3) = \mathrm{P}(X = 2, Y = 2) + \mathrm{P}(X = 2, Y = 3) + \mathrm{P}(X = 3, Y = 2) + \mathrm{P}(X = 3, Y = 3) = \frac{1}{4}$.

9.3 d $\mathrm{P}((X, Y) \in \{1, 4\} \times \{1, 4\}) = \mathrm{P}(X = 1, Y = 1) + \mathrm{P}(X = 1, Y = 4) + \mathrm{P}(X = 4, Y = 1) + \mathrm{P}(X = 4, Y = 4) = \frac{1}{4}$.

9.4 Since $P(X = i, Y = j)$ is either 0, or is equal to $1/14$ for each i and j from 1 to 5, we know that all the entries of the first row and the second column of the table are equal to $1/14$. Since $P(X = 1) = 1/14$, this then determines the rest of the first column (apart of the first entry it contains only zeroes). Similarly, since $P(Y = 5) = 1/14$ one must have that—apart from the second entry—all the entries in the fifth row are zero. Continuing in this way we find

	a					
b	1	2	3	4	5	$P(Y = b)$
1	1/14	1/14	1/14	1/14	1/14	5/14
2	0	1/14	1/14	1/14	1/14	4/14
3	0	1/14	1/14	0	0	2/14
4	0	1/14	1/14	0	0	2/14
5	0	1/14	0	0	0	1/14
$P(X = a)$	1/14	5/14	4/14	2/14	2/14	1

9.5 a From the first row it follows that $\frac{1}{16} \leq \eta \leq \frac{1}{4}$. The second and third row do not add extra information, so we find that $\frac{1}{16} \leq \eta \leq \frac{1}{4}$.

9.5 b For any (allowed) value of η we have that $P(X = 1, Y = 4) = 0$. Since $\frac{1}{16} \leq P(X = 1) \leq \frac{1}{8}$, and $P(Y = 4) = \frac{3}{16}$, we find that $P(X = 1, Y = 4) \neq P(X = 1)\,P(Y = 4)$. Hence, there does not exist a value for η for which X and Y are independent.

9.6 a U attains the values 0, 1, and 2, while V attains the values 0 and 1. Then $P(U = 0, V = 0) = P(X = 0, Y = 0) = P(X = 0)\,P(Y = 0) = \frac{1}{4}$, due to the independence of X and Y. In a similar way the other joint probabilities of U and V are obtained, yielding the following table:

	u			
v	0	1	2	
0	1/4	0	1/4	1/2
1	0	1/2	0	1/2
	1/4	1/2	1/4	1

9.6 b Since $P(U = 0, V = 0) = \frac{1}{4} \neq \frac{1}{8} = P(U = 0)\,P(V = 0)$, we find that U and V are dependent.

9.7 a The joint probability distribution of X and Y is given by

	a			
b	1	2	3	$P(Y = b)$
1	0.22	0.15	0.06	0.43
2	0.11	0.24	0.22	0.57
$P(X = a)$	0.33	0.39	0.28	1

9.7 b Since $\mathrm{P}(X=1, Y=1)=\frac{1168}{5383} \neq \frac{1741}{5383} \frac{2298}{5383}=\mathrm{P}(X=1)\,\mathrm{P}(Y=1)$, we find that X and Y are dependent.

9.8 a Since X can attain the values 0 and 1 and Y the values 0 and 2, Z can attain the values 0, 1, 2, and 3 with probabilities: $\mathrm{P}(Z=0)=\mathrm{P}(X=0, Y=0)=1/4$, $\mathrm{P}(Z=1)=\mathrm{P}(X=1, Y=0)=1/4$, $\mathrm{P}(Z=2)=\mathrm{P}(X=0, Y=2)=1/4$, and $\mathrm{P}(Z=3)=\mathrm{P}(X=1, Y=2)=1/4$.

9.8 b Since $\tilde{X}=\tilde{Z}-\tilde{Y}$, $\tilde{X}$ can attain the values -2, -1, 0, 1, 2, and 3 with probabilities

$$
\begin{aligned}
\mathrm{P}\Big(\tilde{X}=-2\Big) &= \mathrm{P}\Big(\tilde{Z}=0, \tilde{Y}=2\Big)=1/8,\\
\mathrm{P}\Big(\tilde{X}=-1\Big) &= \mathrm{P}\Big(\tilde{Z}=1, \tilde{Y}=2\Big)=1/8,\\
\mathrm{P}\Big(\tilde{X}=0\Big) &= \mathrm{P}\Big(\tilde{Z}=0, \tilde{Y}=0\Big)+\mathrm{P}\Big(\tilde{Z}=2, \tilde{Y}=2\Big)=1/4,\\
\mathrm{P}\Big(\tilde{X}=1\Big) &= \mathrm{P}\Big(\tilde{Z}=1, \tilde{Y}=0\Big)+\mathrm{P}\Big(\tilde{Z}=3, \tilde{Y}=2\Big)=1/4,\\
\mathrm{P}\Big(\tilde{X}=2\Big) &= \mathrm{P}\Big(\tilde{Z}=2, \tilde{Y}=0\Big)=1/8,\\
\mathrm{P}\Big(\tilde{X}=3\Big) &= \mathrm{P}\Big(\tilde{Z}=3, \tilde{Y}=0\Big)=1/8.
\end{aligned}
$$

We have the following table:

z	-2	-1	0	1	2	3
$p_{\tilde{X}}(z)$	1/8	1/8	1/4	1/4	1/8	1/8

9.9 a One has that $F_X(x)=\lim_{y\to\infty} F(x,y)$. So for $x\leq 0$: $F_X(x)=0$, and for $x>0$: $F_X(x)=F(x,\infty)=1-\mathrm{e}^{-2x}$. Similarly, $F_Y(y)=0$ for $y\leq 0$, and for $y>0$: $F_Y(y)=F(\infty,y)=1-\mathrm{e}^{-y}$.

9.9 b For $x>0$ and $y>0$: $f(x,y)=\frac{\partial^2}{\partial x\,\partial y}F(x,y)=\frac{\partial}{\partial x}\big(\mathrm{e}^{-y}-\mathrm{e}^{-(2x+y)}\big)=2\mathrm{e}^{-(2x+y)}$.

9.9 c There are two ways to determine $f_X(x)$:

$$
f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y=\int_0^{\infty}\mathrm{e}^{-(2x+y)}\mathrm{d}y=2\mathrm{e}^{-2x}\quad\text{for } x>0
$$

and

$$
f_X(x)=\frac{\mathrm{d}}{\mathrm{d}x}F_X(x)=2\mathrm{e}^{-2x}\quad\text{for } x>0.
$$

Using either way one finds that $f_Y(y)=\mathrm{e}^{-y}$ for $y>0$.

9.9 d Since $F(x,y)=F_X(x)F_Y(y)$ for all x,y, we find that X and Y are independent.

9.10 a

$$
\begin{aligned}
\mathrm{P}\left(\frac{1}{4}\leq X\leq\frac{1}{2}, \frac{1}{3}\leq Y\leq\frac{2}{3}\right) &= \int_{\frac{1}{4}}^{\frac{1}{2}}\int_{\frac{1}{3}}^{\frac{2}{3}}\frac{12}{5}xy(1+y)\,\mathrm{d}x\,\mathrm{d}y\\
&= \frac{12}{5}\int_{\frac{1}{4}}^{\frac{1}{2}} x\Big(\int_{\frac{1}{3}}^{\frac{2}{3}} y(1+y)\,\mathrm{d}y\Big)\mathrm{d}x\\
&= \int_{\frac{1}{4}}^{\frac{1}{2}}\frac{82}{135}x\,\mathrm{d}x=\frac{41}{720}.
\end{aligned}
$$

9.10 b Since $f(x,y)=0$ for $x<0$ or $y<0$,

$$F(a,b)=\mathrm{P}(X\le a, Y\le b)=\int_{-\infty}^{a}\big(\int_{-\infty}^{b} f(x,y)\,\mathrm{d}y\big)\mathrm{d}x$$
$$=\int_{0}^{a}\big(\int_{0}^{b}\frac{12}{5}xy(1+y)\,\mathrm{d}y\big)\mathrm{d}x=\frac{3}{5}a^2b^2+\frac{2}{5}a^2b^3.$$

9.10 c Since $f(x,y)=0$ for $y>1$, we find for a between 0 and 1, and $b\ge 1$,

$$F(a,b)=\mathrm{P}(X\le a, Y\le b)=\mathrm{P}(X\le a, Y\le 1)=F(a,1)=a^2.$$

Hence, applying (9.1) one finds that $F_X(a)=a^2$, for a between 0 and 1.

9.10 d Another way to obtain f_X is by differentiating F_X.

9.10 e $f(x,y)=f_X(x)f_Y(y)$, so X and Y are independent.

9.11 To determine $\mathrm{P}(X<Y)$ we must integrate $f(x,y)$ over the region G of points (x,y) in $\mathbb{R}^2$ for which x is smaller than y:

$$\begin{aligned}\mathrm{P}(X<Y)&=\iint_{\{(x,y)\in\mathbb{R}^2;\,x<y\}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f(x,y)\,\mathrm{d}x\right)\mathrm{d}y=\int_0^1\left(\int_0^y\frac{12}{5}xy(1+y)\,\mathrm{d}x\right)\mathrm{d}y\\&=\frac{12}{5}\int_0^1 y(1+y)\left(\int_0^y x\,\mathrm{d}x\right)\mathrm{d}y=\frac{12}{10}\int_0^1 y^3(1+y)\,\mathrm{d}y=\frac{27}{50}.\end{aligned}$$

Here we used that $f(x,y)=0$ for (x,y) outside the unit square.

9.12 a Since the integral over $\mathbb{R}^2$ of the joint probability density function is equal to 1, we find from

$$\int_0^1\int_0^2(3u^2+8uv)\,\mathrm{d}u\,\mathrm{d}v=10$$

that $K=\frac{1}{10}$.

9.12 b

$$\begin{aligned}\mathrm{P}(2X\le Y)&=\int\int_{(x,y):2x<y}\frac{1}{10}(3x^2+8xy)\,\mathrm{d}x\,\mathrm{d}y\\&=\int_{x=0}^{1}\int_{y=2x}^{2}\frac{1}{10}(3x^2+8xy)\,\mathrm{d}x\,\mathrm{d}y\\&=0.45.\end{aligned}$$

9.13 a For $r\ge 0$, let D_r be the disc with origin $(0,0)$ and radius r. Since $1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\iint_{D_1}c\,\mathrm{d}x\,\mathrm{d}y=c\cdot\text{area of }D_1$, we find that $c=1/\pi$.

9.13 b Clearly $F_R(r)=0$ for $r<0$, and $F_R(r)=1$ for $r>1$. For r between 0 and 1,

$$F_R(r)=\mathrm{P}((X,Y)\in D_r)=\iint_{D_r}\frac{1}{\pi}\,\mathrm{d}x\,\mathrm{d}y=\frac{\pi r^2}{\pi}=r^2.$$

9.13 c For x between -1 and 1,

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,\mathrm{d}y = \frac{2}{\pi}\sqrt{1-x^2}.$$

For x outside the interval $-1, 1]$ we have that $f_X(x) = 0$.

9.14 a Let f be the joint probability density function of the pair (X, Y), and F their joint distribution function. Since $f(x,y) = \frac{\partial^2}{\partial x\,\partial y}F(x,y)$, we first determine F. Setting $G = (-\infty, a] \times (-\infty, b]$, for $-1 \le a, b \le 1$, we have that

$$F(a,b) = \mathrm{P}((X,Y) \in G) = \frac{(a+1)(b+1)}{4}.$$

Furthermore, if $a \le -1$ or $b \le -1$, we have that $F(a,b) = 0$, since $G \cap \square = \emptyset$. In a similar way, we find for $a > 1$ and $-1 \le b \le 1$ that $F(a,b) = (b+1)/2$, while for $-1 \le a \le 1$ and $b > 1$ we have that $F(a,b) = (a+1)/2$. Finally, if $a, b > 1$, then $F(a,b) = 1$. Taking derivatives, it then follows that $f(x,y) = 1/4$ for a and b between -1 and 1, and $f(x,y) = 0$ for all other values of x and y.

9.14 b Note that for x between -1 and 1, the marginal probability density function f_X of X is given by

$$f_X(x) = \int_{-1}^{1} f(x,y)\,\mathrm{d}y = \frac{1}{2},$$

and that $f_X(x) = 0$ for all other values of x. Similarly, $f_Y(y) = 1/2$ for y between -1 and 1, and $f_Y(y) = 0$ otherwise. But then we find that $f(x,y) = f_X(x)f_Y(y)$, for all possible xs and ys, and we find that X and Y are independent, $U(-1,1)$ distributed random variables.

9.15 a Setting $\square(a,b)$ as the set of points (x,y), for which $x \le a$ and $y \le b$, we have that

$$F(a,b) = \frac{\text{area}\,(\Delta \cap \square(a,b))}{\text{area of}\,\Delta}.$$

- If $a < 0$ or if $b < 0$ (or both), then area $(\Delta \cap \square(a,b)) = \emptyset$, so $F(a,b) = 0$,
- If $(a,b) \in \Delta$, then area $(\Delta \cap \square(a,b)) = a(b - \frac{1}{2}a)$, so $F(a,b) = a(2b - a)$,
- If $0 \le b \le 1$, and $a > b$, then area $(\Delta \cap \square(a,b)) = \frac{1}{2}b^2$, so $F(a,b) = b^2$,
- If $0 \le a \le 1$, and $b > 1$, then area $(\Delta \cap \square(a,b)) = a - \frac{1}{2}a^2$, so $F(a,b) = 2a - a^2$,
- If both $a > 1$ and $b > 1$, then area $(\Delta \cap \square(a,b)) = \frac{1}{2}$, so $F(a,b) = 1$.

9.15 b Since $f(x,y) = \frac{\partial^2}{\partial x\,\partial y}F(x,y)$, we find for $(x,y) \in \Delta$ that $f(x,y) = 2$. Furthermore, $f(x,y) = 0$ for (x,y) outside the triangle Δ.

9.15 c For x between 0 and 1,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_x^1 2\,\mathrm{d}y = 2(1-x).$$

For y between 0 and 1,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_0^y 2\,\mathrm{d}x = 2y.$$

9.16 Following the solution of Exercise 9.14 one finds that U and V are independent $U(0,1)$ distributed random variables. Using the results from Section 8.4 we find that $F_U(x) = 1 - (1-x)^2$ for x between 0 and 1, and that $F_V(y) = y^2$ for y between 0 and 1. Differentiating yields the desired result.

9.17 For $0 \le s \le t \le a$, it follows from the fact that U_1 and U_2 are independent uniformly distributed random variables over $[0, a]$, that

$$\mathrm{P}(U_1 \le t, U_2 \le t) = t^2/a^2,$$

and that

$$\mathrm{P}(s < U_1 \le t, s < U_2 \le t) = (t-s)^2/a^2.$$

But then the answer is an immediate consequence of the hint.
The statement can also be obtained as follows. Note that

$$\begin{aligned}\mathrm{P}(V \le s, Z \le t) = \mathrm{P}(U_2 \le s, s < U_1 \le t) + \mathrm{P}(U_1 \le s, s < U_2 \le t) \\ + \mathrm{P}(U_1 \le s, U_2 \le s)\,.\end{aligned}$$

Using independence and that fact that U_i has distribution function $F_{U_1}(u) = u/a$, we find for $0 \le s \le t \le a$:

$$\mathrm{P}(V \le s, Z \le t) = \frac{s(t-s)}{a^2} + \frac{s(t-s)}{a^2} + \frac{s^2}{a^2} = \frac{t^2 - (t-s)^2}{a^2}.$$

9.18 a By definition

$$\mathrm{E}[X_i] = \sum_{k=1}^{N} k p_{X_i}(k) = \frac{1}{N}(1 + 2 + \cdots + N) = \frac{1}{N} \cdot \frac{1}{2} N(N+1) = \frac{N+1}{2}.$$

9.18 b Using the identity, we find

$$\mathrm{E}\left[X_i^2\right] = \sum_{k=1}^{N} k^2 \mathrm{P}(X_i = k) = \frac{1}{N} \sum_{k=1}^{N} k^2 = \frac{(N+1)(2N+1)}{6}.$$

But then we have that

$$\mathrm{Var}(X_i) = \mathrm{E}\left[X_i^2\right] - \left(\mathrm{E}[X_i]\right)^2 = \frac{N^2 - 1}{12}.$$

9.19 a Clearly we must have that $a = \sqrt{50} = 5\sqrt{2}$, but then it follows that $2ab = 80$, implying that $b = 4\sqrt{2}$. Since $32 + c = 50$, we find that $c = 18$.

9.19 b Note that

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{e}^{-(5\sqrt{2}y - 4\sqrt{2}x)^2}\, \mathrm{d}y &= \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)}\, \mathrm{d}y \\ &= \sigma\sqrt{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)}\, \mathrm{d}y \\ &= \sigma\sqrt{2\pi},\end{aligned}$$

since

$$\frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)} \quad \text{for } -\infty < y < \infty$$

is the probability density function of an $N(\mu, \sigma^2)$ distributed random variable (and therefore integrates to 1).

9.19 c

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \frac{30}{\pi}\mathrm{e}^{-50x^2-50y^2+80xy}\,\mathrm{d}y \\
&= \frac{30}{\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-(5\sqrt{2}y-4\sqrt{2}x)^2}\mathrm{e}^{-18x^2}\,\mathrm{d}y \\
&= \frac{30}{\pi}\frac{\sqrt{2\pi}}{10}\mathrm{e}^{-18x^2} = \frac{1}{\frac{1}{6}\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{x}{1/6}\right)^2},
\end{aligned}
$$

for $-\infty < x < \infty$. So we see that X has a $N(0, \frac{1}{36})$ distribution.

9.20 a Since the needle hits the sheet of paper at an random position, the midpoint (X,Y) falls completely randomly between some lines. Consequently, the distance Z between (X,Y) and the line "under" (X,Y) has a $U(0,1)$ distribution. Also the orientation of the needle is completely random, so the angle between the needle and the positive x-axis can be anything between 0 and 180 degrees. But then H has a $U(0,\pi)$ distribution.

9.20 b From Figure 29.1 we see that—in case $Z \leq 1/2$—the needle hits the line under it when $Z \leq \frac{1}{2}\sin H$. In case $Z > 1/2$, we have a similar picture, but then the needle hits the line above it when $1 - Z \leq \frac{1}{2}\sin H$.

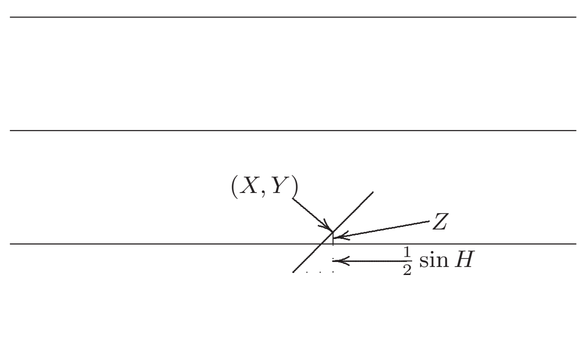


Fig. 29.1. Solution of Exercise 9.20, case $Z \leq 1/2$

9.20 c Since Z and H are independent, uniformly distributed random variables, the probability we are looking for is equal to the area of the two regions in $[0,\pi) \times [0,1]$ for which either

$$Z \leq \frac{1}{2}\sin H \quad \text{or} \quad 1 - Z \leq \frac{1}{2}\sin H,$$

divided (!) by the total area π. Note that these two regions both have the same are, and that this area is equal to

$$\int_0^{\pi} \frac{1}{2}\sin H\,\mathrm{d}H = \frac{1}{2}\cdot 2 = 1.$$

So the probability we are after is equal to

$$\frac{1}{\pi} + \frac{1}{\pi} = \frac{2}{\pi}.$$

10.1 a The joint probability distribution of X and Y is given by

	a			
b	1	2	3	$\mathrm{P}(Y = b)$
1	0.217	0.153	0.057	0.427
2	0.106	0.244	0.223	0.573
$\mathrm{P}(X = a)$	0.323	0.397	0.280	1

This means that

$$\begin{aligned}
\mathrm{E}[X] &= 1 \cdot 0.323 + 2 \cdot 0.397 + 3 \cdot 0.280 = 1.957 \\
\mathrm{E}[Y] &= 1 \cdot 0.427 + 2 \cdot 0.573 = 1.573 \\
\mathrm{E}[XY] &= 1 \cdot 1 \cdot 0.217 + \cdots + 2 \cdot 3 \cdot 0.223 = 3.220.
\end{aligned}$$

It follows that $\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 3.220 - 1.957 \cdot 0.245 = 0.142$.

10.1 b From the joint probability distribution of X and Y we compute

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= 1 \cdot 0.323 + 4 \cdot 0.397 + 9 \cdot 0.280 = 4.431 \\
\mathrm{Var}(X) &= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 0.601 \\
\mathrm{E}\left[Y^2\right] &= 1 \cdot 0.427 + 4 \cdot 0.573 = 2.719 \\
\mathrm{Var}(Y) &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 0.245.
\end{aligned}$$

Using the result from **a** we find

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{0.142}{\sqrt{0.601 \cdot 0.245}} = 0.369.$$

10.2 a From Exercise 9.2:

	x			
y	0	1	2	$p_Y(y)$
-1	1/6	1/6	1/6	1/2
1	0	1/2	0	1/2
$p_X(x)$	1/6	2/3	1/6	1

so that $\mathrm{E}[XY] = \frac{1}{6} \cdot 1 \cdot (-1) + \frac{1}{6} \cdot 2 \cdot (-1) + \frac{1}{2} \cdot 1 \cdot 1 = 0$.

10.2 b $\mathrm{E}[Y] = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0$, so that $\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y] = 0 - 0 = 0$.

10.2 c From the marginal distributions we find

$$\begin{aligned}
\mathrm{E}[X] &= 0 \cdot \frac{1}{6} + 1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{6} = 1 \\
\mathrm{E}\left[X^2\right] &= 0 \cdot \frac{1}{6} + 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{6} = 1 \\
\mathrm{Var}(X) &= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{4}{3} - 1^2 = \frac{1}{3} \\
\mathrm{E}\left[Y^2\right] &= (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1 \\
\mathrm{Var}(Y) &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = 1 - 0^2 = 1.
\end{aligned}$$

Since $\mathrm{Cov}(X,Y) = 0$ we have $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \frac{1}{3} + 1 = \frac{4}{3}$.

10.2 d Since $\mathrm{Cov}(X,-Y) = -\mathrm{Cov}(X,Y) = 0$ we have $\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(-Y) = \mathrm{Var}(X) + (-1)^2\mathrm{Var}(Y) = \frac{1}{3} + 1 = \frac{4}{3}$.

10.3 We have $\mathrm{E}[U] = 1$, $\mathrm{E}[V] = \frac{1}{2}$, $\mathrm{E}\left[U^2\right] = \frac{3}{2}$, $\mathrm{E}\left[V^2\right] = \frac{1}{2}$, $\mathrm{Var}(U) = \frac{1}{2}$, $\mathrm{Var}(V) = \frac{1}{4}$, and $\mathrm{E}[UV] = \frac{1}{2}$. Hence $\mathrm{Cov}(U,V) = 0$ and $\rho(U,V) = 0$.

10.4 Both X and Y have marginal probabilities $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, so that

$$\begin{aligned}
\mathrm{E}[X] &= 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 2.5 \\
\mathrm{E}[Y] &= 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 2.5 \\
\mathrm{E}[XY] &= 1 \cdot 1 \cdot \frac{16}{136} + \cdots + 4 \cdot 4 \cdot \frac{1}{136} = 6.25.
\end{aligned}$$

Hence $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 6.25 - 2.5 \cdot 2.5 = 0$.

10.5 a First find $\mathrm{P}(X=1, Y=0) = \frac{1}{3} - \frac{8}{72} - \frac{10}{72} = \frac{6}{72}$, and similarly $\mathrm{P}(X=0, Y=2) = \frac{4}{72}$ and $\mathrm{P}(Y=2) = \frac{1}{6}$. Then $\mathrm{P}(X=1) = \frac{1}{4}$ and $\mathrm{P}(X=2) = \frac{5}{12}$, and finally $\mathrm{P}(X=2, Y=2) = \frac{5}{71}$:

	a			
b	0	1	2	
0	8/72	6/72	10/72	1/3
1	12/72	9/72	15/72	1/2
2	4/72	3/72	5/72	1/6
	1/3	1/4	5/12	1

10.5 b With **a** we find:

$$\begin{aligned}
\mathrm{E}[X] &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{5}{12} = \frac{13}{12} \\
\mathrm{E}[Y] &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{6} = \frac{5}{6} \\
\mathrm{E}[XY] &= 1 \cdot 1 \cdot \frac{9}{72} + \cdots + 2 \cdot 2 \cdot \frac{5}{72} = \frac{65}{72}.
\end{aligned}$$

Hence $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \frac{65}{72} - \frac{13}{12} \cdot \frac{5}{6} = 0$.

10.5 c Yes, for all combinations (a,b) it holds that $\mathrm{P}(X=a, Y=b) = \mathrm{P}(X=a)\,\mathrm{P}(Y=b)$.

10.6 a When $c = 0$, the joint distribution becomes

b	a: -1	0	1	$P(Y = b)$
-1	2/45	9/45	4/45	1/3
0	7/45	5/45	3/45	1/3
1	6/45	1/45	8/45	1/3
$P(X = a)$	1/3	1/3	1/3	1

We find $E[X] = (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$, and similarly $E[Y] = 0$. By leaving out terms where either $X = 0$ or $Y = 0$, we find

$$E[XY] = (-1) \cdot (-1) \cdot \frac{2}{45} + (-1) \cdot 1 \cdot \frac{4}{45} + 1 \cdot (-1) \cdot \frac{6}{45} + 1 \cdot 1 \cdot \frac{8}{45} = 0,$$

which implies that $\mathrm{Cov}(X, Y) = E[XY] - E[X]\, E[Y] = 0$.

10.6 b Note that the variables X and Y in part **b** are equal to the ones from part **a**, shifted by c. If we write U and V for the variables from **a**, then $X = U + c$ and $Y = V + c$. According to the rule on the covariance under change of units, we then immediately find $\mathrm{Cov}(X, Y) = \mathrm{Cov}(U + c, V + c) = \mathrm{Cov}(U, V) = 0$.
Alternatively, one could also compute the covariance from $\mathrm{Cov}(X, Y) = E[XY] - E[X]\, E[Y]$. We find $E[X] = (c-1) \cdot \frac{1}{3} + c \cdot \frac{1}{3} + (c+1) \cdot \frac{1}{3} = c$, and similarly $E[Y] = c$. Since

$$\begin{aligned} E[XY] = {} & (c-1) \cdot (c-1) \cdot \frac{2}{45} + (c-1) \cdot c \cdot \frac{9}{45} + (c+1) \cdot (c+1) \cdot \frac{4}{45} \\ & + c \cdot (c-1) \cdot \frac{7}{45} + c \cdot c \cdot \frac{5}{45} + c \cdot (c+1) \cdot \frac{3}{45} \\ & + (c+1) \cdot (c-1) \cdot \frac{6}{45} + (c+1) \cdot c \cdot \frac{1}{45} + (c+1) \cdot (c+1) \cdot \frac{8}{45} = c^2, \end{aligned}$$

we find $\mathrm{Cov}(X, Y) = E[XY] - E[X]\, E[Y] = c^2 - c \cdot c = 0$.

10.6 c No, X and Y are not independent. For instance, $P(X = c, Y = c + 1) = 1/45$, which differs from $P(X = c)\; P(Y = c + 1) = 1/9$.

10.7 a $E[XY] = \frac{1}{8}$, $E[X] = E[Y] = \frac{1}{2}$, so that $\mathrm{Cov}(X, Y) = E[XY] - E[X]\, E[Y] = \frac{1}{8} - (\frac{1}{2})^2 = -\frac{1}{8}$.

10.7 b Since X has a $Ber(\frac{1}{2})$ distribution, $\mathrm{Var}(X) = \frac{1}{4}$. Similarly, $\mathrm{Var}(Y) = \frac{1}{4}$, so that

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-1/8}{1/4} = -\frac{1}{2}.$$

10.7 c For any ε between $-\frac{1}{4}$ and $\frac{1}{4}$, $\mathrm{Cov}(X, Y) = E[XY] - (\frac{1}{2})^2 = (\frac{1}{4} - \varepsilon) - (\frac{1}{2})^2 = -\varepsilon$, so that

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{-\varepsilon}{1/4} = -4\varepsilon.$$

Hence, $\rho(X, Y)$ is equal to -1, 0, or 1, for ε equal to 1/4, 0, or $-1/4$.

10.8 a $E\left[X^2\right] = \mathrm{Var}(X) + (E[X])^2 = 8$.

10.8 b $E\left[-2X^2 + Y\right] = -2E\left[X^2\right] + E[Y] = -2 \cdot 8 + 3 = -13$.

10.9 a If the aggregated blood sample tests negative, we do not have to perform additional tests, so that X_i takes on the value 1. If the aggregated blood sample tests positive, we have to perform 40 additional tests for the blood sample of each person in the group, so that X_i takes on the value 41. We first find that $\mathrm{P}(X_i = 1) = \mathrm{P}(\text{no infections in group of } 40) = (1 - 0.001)^{40} = 0.96$, and therefore $\mathrm{P}(X_i = 41) = 1 - \mathrm{P}(X_i = 1) = 0.04$.

10.9 b First compute $\mathrm{E}[X_i] = 1\cdot 0.96 + 41\cdot 0.04 = 2.6$. The expected total number of tests is $\mathrm{E}[X_1 + X_2 + \cdots + X_{25}] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_{25}] = 25\cdot 2.6 = 65$. With the original procedure of blood testing, the total number of tests is $25\cdot 40 = 1000$. On average the alternative procedure would only require 65 tests. Only with very small probability one would end up with doing more than 1000 tests, so the alternative procedure is better.

10.10 a We find

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_0^3 \frac{2}{225}\left(9x^3 + 7x^2\right)\mathrm{d}x = \frac{2}{225}\left[\frac{9}{4}x^4 + \frac{7}{3}x^3\right]_0^3 = \frac{109}{50},$$
$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y = \int_1^2 \frac{1}{25}(3y^3 + 12y^2)\,\mathrm{d}y = \frac{1}{25}\left[\frac{3}{4}y^4 + 4y^3\right]_1^2 = \frac{157}{100},$$

so that $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] = 15/4$.

10.10 b We find

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x = \int_0^3 \frac{2}{225}\left(9x^4 + 7x^3\right)\mathrm{d}x = \frac{2}{225}\left[\frac{9}{5}x^5 + \frac{7}{4}x^4\right]_0^3 = \frac{1287}{250},$$
$$\mathrm{E}\left[Y^2\right] = \int_{-\infty}^{\infty} y^2 f_Y(y)\,\mathrm{d}y = \int_1^2 \frac{1}{25}(3y^4 + 12y^3)\,\mathrm{d}y = \frac{1}{25}\left[\frac{3}{5}y^5 + 3y^4\right]_1^2 = \frac{318}{125},$$
$$\begin{aligned}\mathrm{E}[XY] &= \int_0^3\int_1^2 xyf(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^3\int_1^2 \frac{2}{75}\left(2x^3y^2 + x^2y^3\right)\mathrm{d}y\,\mathrm{d}x\\ &= \frac{4}{75}\int_0^3 x^3\left(\int_1^2 y^2\,\mathrm{d}y\right)\mathrm{d}x + \frac{2}{75}\int_0^3 x^2\left(\int_1^2 y^3\,\mathrm{d}y\right)\mathrm{d}x\\ &= \frac{4}{75}\frac{7}{3}\int_0^3 x^3\,\mathrm{d}x + \frac{2}{75}\frac{15}{4}\int_0^3 x^2\,\mathrm{d}x = \frac{171}{50},\end{aligned}$$

so that $\mathrm{E}\left[(X+Y)^2\right] = \mathrm{E}\left[X^2\right] + \mathrm{E}\left[Y^2\right] + 2\mathrm{E}[XY] = 3633/250$.

10.10 c We find

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{1287}{250} - \left(\frac{109}{50}\right)^2 = \frac{989}{2500},\\ \mathrm{Var}(Y) &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = \frac{318}{125} - \left(\frac{157}{100}\right)^2 = \frac{791}{10\,000},\\ \mathrm{Var}(X+Y) &= \mathrm{E}\left[(X+Y)^2\right] - (\mathrm{E}[X+Y])^2 = \frac{3633}{250} - \left(\frac{15}{4}\right)^2 = \frac{939}{2000}.\end{aligned}$$

Hence, $\mathrm{Var}(X) + \mathrm{Var}(Y) = 0.4747$, which differs from $\mathrm{Var}(X+Y) = 0.4695$.

10.11 $\mathrm{Cov}(T,S) = \mathrm{Cov}\left(\frac{9}{5}X + 32, \frac{9}{5}X + 32\right) = \left(\frac{9}{5}\right)^2\mathrm{Cov}(X,Y) = 9.72$ and $\rho(T,S) = \rho\left(\frac{9}{5}X + 32, \frac{9}{5}X + 32\right) = \rho(X,Y) = 0.8$.

10.12 Since H has a $U(25, 35)$ distribution, it follows immediately that $\mathrm{E}[H] = 30$. Furthermore

$$\mathrm{E}\left[R^2\right] = \int_{7.5}^{12.5} r^2 \frac{1}{5}\,\mathrm{d}r = \left[\frac{r^3}{15}\right]_{7.5}^{1} 2.5 = 102.0833.$$

Hence $\pi\mathrm{E}[H]\,\mathrm{E}\left[R^2\right] = \pi \cdot 30 \cdot 102.0833 = 9621.128$.

10.13 a You can apply Exercise 8.15 directly for the case $n = 2$ to conclude that $\mathrm{E}[X] = \frac{1}{3}$ and $\mathrm{E}[Y] = \frac{2}{3}$. In order to compute the variances we need the probability densities of X and Y. Using the rules about the distribution of the minimum and maximum of independent random variables in Section 8.4, we find

$$\begin{aligned} f_X(x) &= 2(1-x) \qquad \text{for } 0 \le x \le 1 \\ f_Y(y) &= 2y \qquad\qquad \text{for } 0 \le y \le 1. \end{aligned}$$

This means that $\mathrm{E}\left[X^2\right] = \int_0^1 2x^2(1-x)\,\mathrm{d}x = \frac{1}{6}$, so that $\mathrm{Var}(X) = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}$. Similarly, $\mathrm{E}\left[Y^2\right] = \int_0^1 2y^3\,\mathrm{d}y = \frac{1}{2}$, so that $\mathrm{Var}(Y) = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$.

10.13 b Since U and V are independent, each with variance $\frac{1}{12}$, it follows that $\mathrm{Var}(X+Y) = \mathrm{Var}(U+V) = \mathrm{Var}(U) + \mathrm{Var}(V) = \frac{1}{6}$.

10.13 c To compute $\mathrm{Cov}(X,Y)$ we need $\mathrm{E}[XY]$. This can be solved from

$$\begin{aligned} \frac{1}{6} &= \mathrm{Var}(X+Y) = \mathrm{E}\left[(X+Y)^2\right] - (\mathrm{E}[X+Y])^2 \\ &= \mathrm{E}\left[X^2\right] + 2\mathrm{E}[XY] + \mathrm{E}\left[Y^2\right] - (\mathrm{E}[X] + \mathrm{E}[Y])^2. \end{aligned}$$

Substitute $\mathrm{E}[X] = \frac{1}{3}$ and $\mathrm{E}\left[X^2\right] = \mathrm{Var}(X) + (\mathrm{E}[X])^2 = \frac{1}{18} + \frac{1}{9} = \frac{3}{18}$, and similarly $\mathrm{E}[Y] = \frac{2}{3}$ and $\mathrm{E}\left[Y^2\right] = \mathrm{Var}(Y) + (\mathrm{E}[Y])^2 = \frac{1}{18} + \frac{4}{9} = \frac{1}{2}$. This leads to $\mathrm{E}[XY] = \frac{1}{4}$.

10.14 a By using the alternative expression for the covariance and linearity of expectations, we find

$$\begin{aligned} &\mathrm{Cov}(X+s, Y+u) \\ &\quad = \mathrm{E}[(X+s)(Y+u)] - \mathrm{E}[X+s]\,\mathrm{E}[Y+u] \\ &\quad = \mathrm{E}[XY + sY + uX + su] - (\mathrm{E}[X] + s)(\mathrm{E}[Y] + u) \\ &\quad = (\mathrm{E}[XY] + s\mathrm{E}[Y] + u\mathrm{E}[X] + su) - (\mathrm{E}[X]\,\mathrm{E}[Y] + s\mathrm{E}[Y] + u\mathrm{E}[X] + su) \\ &\quad = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\ &\quad = \mathrm{Cov}(X,Y)\,. \end{aligned}$$

10.14 b By using the alternative expression for the covariance and the rule on expectations under change of units, we find

$$\begin{aligned} \mathrm{Cov}(rX, tY) &= \mathrm{E}[(rX)(tY)] - \mathrm{E}[rX]\,\mathrm{E}[tY] \\ &= \mathrm{E}[rtXY] - (r\mathrm{E}[X])(t\mathrm{E}[Y]) \\ &= rt\mathrm{E}[XY] - rt\mathrm{E}[X]\,\mathrm{E}[Y] \\ &= rt\,(\mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]) \\ &= rt\mathrm{Cov}(X,Y)\,. \end{aligned}$$

10.14 c First applying part **a** and then part **b** yields

$$\mathrm{Cov}(rX+s, tY+u) = \mathrm{Cov}(rX, tY) = rt\mathrm{Cov}(X,Y)\,.$$

10.15 a Left plot: looks like 500 realizations from a pair (X, Y) whose 2-dimensional distribution has contourlines that are circles. This means X and Y are uncorrelated. Middle plot: looks like 500 realizations from a pair (X, Y) whose 2-dimensional distribution has contourlines that are ellipsoids with the line $y = x$ as main axis. This means X and Y are positively correlated; Right plot: looks like 500 realizations from a pair (X, Y) whose 2-dimensional distribution has contourlines that are ellipsoids with the line $y = x$ as main axis. This means X and Y are negatively correlated.

10.15 b In the right picture the points are concentrated more closely than in the other pictures. Hence $|\rho(X, Y)|$ will be the largest for the right picture.

10.16 a $\mathrm{Cov}(X, X + Y) = \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) = \mathrm{Var}(X) + \mathrm{Cov}(X, Y)$

10.16 b Anything can happen, depending on the sign of $\mathrm{Cov}(X, Y)$ and its magnitude compared to $\mathrm{Var}(X)$.

10.16 c If X and Y are uncorrelated, $\mathrm{Cov}(X, Y) = \mathrm{Var}(X) \geq 0$, so apart from the special case where X is constant, X and $X + Y$ are positively correlated.

10.17 a Since all expectations are zero, it is sufficient to show

$$\mathrm{E}\left[(X + Y + Z)^2\right] = \mathrm{E}\left[X^2\right] + \mathrm{E}\left[Y^2\right] + \mathrm{E}\left[Y^2\right] + 2\mathrm{E}[XY] + 2\mathrm{E}[XZ] + 2\mathrm{E}[YZ].$$

This follows immediately from the hint with $n = 3$, and using linearity of expectations.

10.17 b Write $\tilde{X} = X - \mathrm{E}[X]$, and similarly $\tilde{Y}$ and $\tilde{Z}$. Then part **a** applies to $\tilde{X}$, $\tilde{Y}$, and $\tilde{Z}$:

$$\begin{aligned}\mathrm{Var}\left(\tilde{X} + \tilde{Y} + \tilde{Z}\right) = \mathrm{Var}\left(\tilde{X}\right) + \mathrm{Var}\left(\tilde{Y}\right) + \mathrm{Var}\left(\tilde{Z}\right) \\ + 2\mathrm{Cov}\left(\tilde{X}, \tilde{Y}\right) + 2\mathrm{Cov}\left(\tilde{X}, \tilde{Z}\right) + 2\mathrm{Cov}\left(\tilde{Y}, \tilde{Z}\right).\end{aligned}$$

According to the rules on pages 104 and 151 about the variance and covariance under a change of units, it follows that $\mathrm{Var}\left(\tilde{X}\right) = \mathrm{Var}(X)$, $\mathrm{Cov}\left(\tilde{X}, \tilde{Y}\right) = \mathrm{Cov}(X, Y)$, and similarly for all other variances and covariances.

10.17 c Similar to part **a** first consider the case with all random variables having expectation zero. As in part **a**, the hint in **a** together with linearity of expectations yields the desired equality. Then argue as in part **b** by introducing $\tilde{X}_i = X_i - \mathrm{E}[X_i]$.

10.17 d Use part **c** and note that there are n terms $\mathrm{Var}(X_i)$ and $n(n-1)$ terms $\mathrm{Cov}(X_i, X_j)$ with $i \neq j$.

10.18 First note that $X_1 + X_2 + \cdots + X_N$ is the sum of all numbers, which is a nonrandom constant. Therefore, $\mathrm{Var}(X_1 + X_2 + \cdots + X_N) = 0$. In Section 9.3 we argued that, although we draw without replacement, each X_i has the same distribution. By the same reasoning, we find that each pair (X_i, X_j), with $i \neq j$, has the same joint distribution, so that $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_1, X_2)$ for all pairs with $i \neq j$. Direct application of Exercise 10.17 with $\sigma^2 = (N-1)(N+1)$ and $\gamma = \mathrm{Cov}(X_1, X_2)$ gives

$$0 = \mathrm{Var}(X_1 + X_2 + \cdots + X_N) = N \cdot \frac{(N-1)(N+1)}{12} + N(N-1)\mathrm{Cov}(X_1, X_2).$$

Solving this identity gives $\mathrm{Cov}(X_1, X_2) = -(N+1)/12$.

10.19 By definition and linearity of expectations:

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \mathrm{E}\left[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right]\\ &= \mathrm{E}\left[XY - X\mathrm{E}[Y] - \mathrm{E}[X]Y + \mathrm{E}[X]\mathrm{E}[Y]\right]\\ &= \mathrm{E}[XY] - \mathrm{E}[X\mathrm{E}[Y]] - \mathrm{E}[\mathrm{E}[X]Y] + \mathrm{E}[X]\mathrm{E}[Y]\\ &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] - \mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}[X]\mathrm{E}[Y]\\ &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].\end{aligned}$$

10.20 To compute the correlation, we need:

$$\begin{aligned}\operatorname{Cov}\left(U,U^2\right) &= \mathrm{E}\left[U^3\right] - \mathrm{E}[U]\,\mathrm{E}\left[U^2\right],\\ \operatorname{Var}(U) &= \mathrm{E}\left[U^2\right] - (\mathrm{E}[U])^2,\\ \operatorname{Var}\left(U^2\right) &= \mathrm{E}\left[U^4\right] - \left(\mathrm{E}\left[U^2\right]\right)^2.\end{aligned}$$

Hence we determine

$$\mathrm{E}\left[U^k\right] = \int_0^a u^k\,\frac{1}{a}\,du = \frac{a^k}{k+1},$$

which yields

$$\begin{aligned}\operatorname{Cov}\left(U,U^2\right) &= \frac{a^3}{4} - \frac{a}{2}\frac{a^2}{3} = \left(\frac{1}{4} - \frac{1}{2}\frac{1}{3}\right)a^3 = \frac{1}{12}a^3,\\ \operatorname{Var}(U) &= \frac{a^2}{3} - \left(\frac{a}{2}\right)^2 = \left(\frac{1}{3} - \frac{1}{4}\right)a^2 = \frac{1}{12}a^2,\\ \operatorname{Var}\left(U^2\right) &= \frac{a^4}{5} - \left(\frac{a^2}{3}\right)^2 = \left(\frac{1}{5} - \frac{1}{9}\right)a^4 = \frac{4}{45}a^4.\end{aligned}$$

Hence

$$\rho\left(U,U^2\right) = \frac{1}{12}\sqrt{135} = \frac{1}{4}\sqrt{15} = 0.968.$$

Note that the answer does not depend on a.

11.1 a In the addition rule, k is between 2 and 12. We must always have $1 \le k-\ell \le 6$ and $1 \le \ell \le 6$, or equivalently, $k-6 \le \ell \le k-1$ and $1 \le \ell \le 6$. For $k = 2,\ldots,6$, this means

$$p_Z(k) = \mathrm{P}(X+Y=k) = \sum_{\ell=1}^{k-1} p_X(k-\ell)p_Y(\ell) = \sum_{\ell=1}^{k-1} \frac{1}{6}\cdot\frac{1}{6} = \frac{k-1}{36}$$

and for $k = 7,\ldots,12$,

$$p_Z(k) = \mathrm{P}(X+Y=k) = \sum_{\ell=k-6}^{6} p_X(k-\ell)p_Y(\ell) = \sum_{\ell=k-6}^{6} \frac{1}{6}\cdot\frac{1}{6} = \frac{13-k}{36}.$$

11.1 b In the addition rule, k is between 2 and $2N$. We must always have $1 \le k-\ell \le N$ and $1 \le \ell \le N$, or equivalently, $k-N \le \ell \le k-1$ and $1 \le \ell \le N$. For $k = 2,\ldots,N$, this means

$$p_Z(k) = \mathrm{P}(X+Y=k) = \sum_{\ell=1}^{k-1} p_X(k-\ell)p_Y(\ell) = \sum_{\ell=1}^{k-1} \frac{1}{N}\cdot\frac{1}{N} = \frac{k-1}{N^2}$$

and for $k = N+1, \ldots, 2N$,

$$p_Z(k) = \mathrm{P}(X+Y=k) = \sum_{\ell=k-N}^{N} p_X(k-\ell)p_Y(\ell) = \sum_{\ell=k-N}^{N} \frac{1}{N}\cdot\frac{1}{N} = \frac{2N-k+1}{N^2}.$$

11.2 a By using the rule on addition of two independent discrete random variables, we have

$$\mathrm{P}(X+Y=k) = p_Z(k) = \sum_{\ell=0}^{\infty} p_X(k-\ell)p_Y(\ell).$$

Because $p_X(a)=0$ for $a \le -1$, all terms with $\ell \ge k+1$ vanish, so that

$$\mathrm{P}(X+Y=k) = \sum_{\ell=0}^{k} \frac{1^{k-\ell}}{(k-\ell)!}\mathrm{e}^{-1}\cdot\frac{1^\ell}{\ell!}\mathrm{e}^{-1} = \frac{\mathrm{e}^{-2}}{k!}\sum_{\ell=0}^{k}\binom{k}{\ell} = \frac{2^k}{k!}\mathrm{e}^{-2},$$

also using $\sum_{\ell=0}^{k}\binom{k}{\ell} = 2^k$ in the last equality.

11.2 b Similar to part **a**, by using the rule on addition of two independent discrete random variables and leaving out terms for which $p_X(a)=0$, we have

$$\mathrm{P}(X+Y=k) = \sum_{\ell=0}^{k} \frac{\lambda^{k-\ell}}{(k-\ell)!}\mathrm{e}^{-\lambda}\cdot\frac{\mu^\ell}{\ell!}\mathrm{e}^{-\mu} = \frac{(\lambda+\mu)^k}{k!}\mathrm{e}^{-(\lambda+\mu)}\sum_{\ell=0}^{k}\binom{k}{\ell}\frac{\lambda^{k-\ell}\mu^\ell}{(\lambda+\mu)^k}.$$

Next, write

$$\frac{\lambda^{k-\ell}\mu^\ell}{(\lambda+\mu)^k} = \left(\frac{\mu}{\lambda+\mu}\right)^\ell\left(\frac{\lambda}{\lambda+\mu}\right)^{k-\ell} = \left(\frac{\mu}{\lambda+\mu}\right)^\ell\left(1-\frac{\mu}{\lambda+\mu}\right)^{k-\ell} = p^\ell(1-p)^{k-\ell}$$

with $p = \mu/(\lambda+\mu)$. This means that

$$\mathrm{P}(X+Y=k) = \frac{(\lambda+\mu)^k}{k!}\mathrm{e}^{-(\lambda+\mu)}\sum_{\ell=0}^{k}\binom{k}{\ell}p^\ell(1-p)^{k-\ell} = \frac{(\lambda+\mu)^k}{k!}\mathrm{e}^{-(\lambda+\mu)},$$

using that $\sum_{\ell=0}^{k}\binom{k}{\ell}p^\ell(1-p)^{k-\ell} = 1$.

11.3

$$\begin{aligned} p_Z(k) &= \sum_{\ell=0}^{k}\binom{n}{k-\ell}\binom{m}{\ell}(\tfrac{1}{2})^{k-\ell}(\tfrac{1}{4})^{\ell}(\tfrac{1}{2})^{n-(k-\ell)}(\tfrac{3}{4})^{m-\ell} \\ &= \sum_{\ell=0}^{k}\binom{n}{k-\ell}\binom{m}{\ell}(\tfrac{1}{2})^{n}(\tfrac{1}{4})^{\ell}(\tfrac{3}{4})^{m-\ell}. \end{aligned}$$

This cannot be simplified and is not equal to a binomial probability of the type $\binom{n+m}{k}r^k(1-r)^{n+m-k}$ for some $0<r<1$.

11.4 a From the fact that X has an $N(2,5)$ distribution, it follows that $\mathrm{E}[X]=2$ and $\mathrm{Var}(X)=5$. Similarly, $\mathrm{E}[Y]=5$ and $\mathrm{Var}(Y)=9$. Hence by linearity of expectations,

$$\mathrm{E}[Z] = \mathrm{E}[3X-2Y+1] = 3\mathrm{E}[X]-2\mathrm{E}[Y]+1 = 3\cdot 2 - 2\cdot 5 + 1 = -3.$$

By the rules for the variance and covariance,

$$\mathrm{Var}(Z) = 9\mathrm{Var}(X)+4\mathrm{Var}(Y)-12\mathrm{Cov}(X,Y) = 9\cdot 5+4\cdot 9-12\cdot 0 = 81,$$

using that $\mathrm{Cov}(X,Y)=0$, due to independence of X and Y.

11.4 b The random variables $3X$ and $-2Y+1$ are independent and, according to the rule for the normal distribution under a change of units (page 112), it follows that they both have a normal distribution. Next, the sum rule for independent normal random variables then yields that $Z=(3X)+(-2Y+1)$ also has a normal distribution. Its parameters are the expectation and variance of Z. From **a** it follows that Z has an $N(-3,81)$ distribution.

11.4 c From **b** we know that Z has an $N(-3,81)$ distribution, so that $(Z+3)/9$ has a standard normal distribution. Therefore

$$\mathrm{P}(Z \leq 6) = \mathrm{P}\left(\frac{Z+3}{9} \leq \frac{6+3}{9}\right) = \Phi(1),$$

where Φ is the standard normal distribution function. From Table **??** we find that $\Phi(1) = 1 - 0.1587 = 0.8413$.

11.5 According to the addition rule, the probability density of $Z=X+Y$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,\mathrm{d}y = \int_0^1 f_X(z-y) f_Y(y)\,\mathrm{d}y,$$

where we use that $f_Y(y)=0$ for $y \notin [0,1]$. When $0 \leq y \leq 1$, the following holds. For $z<0$, also $z-y<0$ so that $f_X(z-y)=0$, and for $z>2$, $z-y>1$ so that $f_X(z-y)=0$. For $0 \leq z < 1$:

$$f_Z(z) = \int_0^1 f_X(z-y) f_Y(y)\,\mathrm{d}y = \int_0^z 1\,\mathrm{d}y = z,$$

whereas for $1 \leq z \leq 2$:

$$f_Z(z) = \int_0^1 f_X(z-y) f_Y(y)\,\mathrm{d}y = \int_{z-1}^1 1\,\mathrm{d}y = 2-z.$$

11.6 According to the addition rule

$$\begin{aligned} f_Z(z) &= \int_0^z f_X(z-y) f_Y(y)\,\mathrm{d}y = \int_0^z \frac{1}{4}(z-y)\mathrm{e}^{-(z-y)/2}\frac{1}{4}y\mathrm{e}^{-y/2}\,\mathrm{d}y. \\ &= \frac{1}{16}\mathrm{e}^{-z/2}\int_0^z (z-y)y\,\mathrm{d}y = \frac{1}{16}\mathrm{e}^{-z/2}\left[\frac{1}{2}zy^2 - \frac{1}{3}y^3\right]_0^z = \frac{z^3}{96}\mathrm{e}^{-z/2}. \end{aligned}$$

11.7 Each X_i has the same distribution as the sum of k independent $Exp(\lambda)$ distributed random variables. Since the X_i are independent, $X_1 + X_2 + \cdots + X_n$ has the same distribution as the sum of nk independent $Exp(\lambda)$ distributed random variables, which is a $Gam(nk,\lambda)$ distribution.

11.8 a $Y = rX + s$ has probability density function

$$f_Y(y) = \frac{1}{r} f_X\left(\frac{y-s}{r}\right) = \frac{1}{r\pi\left(1+\left(\frac{y-s}{r}\right)^2\right)} = \frac{r}{\pi\left(r^2+(y-s)^2\right)}.$$

We see that $Y = rX + s$ has a $Cau(s,r)$ distribution.

11.8 b If $Y = (X-\beta)/\alpha$ then Y has a $Cau(0,1)$ distribution.

11.9 a According to the product rule on page 172,

$$f_Z(z) = \int_1^z f_Y\left(\frac{z}{x}\right) f_X(x)\frac{1}{x}\,\mathrm{d}x = \int_1^z \frac{1}{\left(\frac{z}{x}\right)^2}\frac{3}{x^4}\frac{1}{x}\,\mathrm{d}x$$
$$= \frac{3}{z^2}\int_1^z \frac{1}{x^3}\,\mathrm{d}x = \frac{3}{z^2}\left[-\frac{1}{2}x^{-2}\right]_1^z = \frac{3}{2}\frac{1}{z^2}\left(1-\frac{1}{z^2}\right)$$
$$= \frac{3}{2}\left(\frac{1}{z^2}-\frac{1}{z^4}\right).$$

11.9 b According to the product rule,

$$f_Z(z) = \int_1^z f_Y\left(\frac{z}{x}\right) f_X(x)\frac{1}{x}\,\mathrm{d}x = \int_1^z \frac{\beta}{\left(\frac{z}{x}\right)^{\beta+1}}\frac{\alpha}{x^{\alpha+1}}\frac{1}{x}\,\mathrm{d}x$$
$$= \frac{\alpha\beta}{z^{\beta+1}}\int_1^z x^{\beta-\alpha-1}\,\mathrm{d}x = \frac{\alpha\beta}{z^{\beta+1}}\left[\frac{x^{\beta-\alpha}}{\beta-\alpha}\right]_1^z = \frac{\alpha\beta}{\alpha-\beta}\frac{1}{z^{\beta+1}}\left(1-z^{\beta-\alpha}\right)$$
$$= \frac{\alpha\beta}{\beta-\alpha}\left(\frac{1}{z^{\beta+1}}-\frac{1}{z^{\alpha+1}}\right).$$

11.10 a In the quotient rule for $Z = X/Y$ for $0 < z < \infty$ fixed, we must have $1 \le zx < \infty$ and $1 \le x < \infty$. Hence for $0 < z < 1$,

$$f_Z(z) = \int_{1/z}^\infty f_X(zx)f_Y(y)x\,\mathrm{d}x = \int_{1/z}^\infty \frac{2}{(zx)^3}\frac{2}{x^3}x\,\mathrm{d}x$$
$$= \frac{4}{z^3}\int_{1/z}^\infty x^{-5}\,\mathrm{d}x = \frac{4}{z^3}\left[\frac{-x^{-4}}{4}\right]_{1/z}^\infty = \frac{1}{z^3}\left(0-(-z^4)\right) = z.$$

For $z \ge 1$, we find

$$f_Z(z) = \int_1^\infty f_X(zx)f_Y(y)x\,\mathrm{d}x = \int_1^\infty \frac{2}{(zx)^3}\frac{2}{x^3}x\,\mathrm{d}x$$
$$= \frac{4}{z^3}\int_1^\infty x^{-5}\,\mathrm{d}x = \frac{4}{z^3}\left[\frac{-x^{-4}}{4}\right]_1^\infty = \frac{1}{z^3}.$$

11.10 b In the quotient rule for $Z = X/Y$ for $0 < z < \infty$ fixed, we must have $1 \le zx < \infty$ and $1 \le x < \infty$. Hence for $0 < z < 1$,

$$f_Z(z) = \int_{1/z}^\infty f_X(zx)f_Y(y)x\,\mathrm{d}x = \int_{1/z}^\infty \frac{\alpha}{(zx)^{\alpha+1}}\frac{\beta}{x^{\beta+1}}x\,\mathrm{d}x$$
$$= \frac{\alpha\beta}{z^{\alpha+1}}\int_{1/z}^\infty x^{-\alpha-\beta-1}\,\mathrm{d}x = \frac{\alpha\beta}{z^{\alpha+1}}\left[\frac{-x^{-\alpha-\beta}}{\alpha+\beta}\right]_{1/z}^\infty$$
$$= \frac{\alpha\beta}{\alpha+\beta}\frac{1}{z^{\alpha+1}}\left(0-(-z^{\alpha+\beta})\right) = \frac{\alpha\beta}{\alpha+\beta}z^{\beta-1}.$$

For $z \ge 1$, we find

$$f_Z(z) = \int_{1/z}^\infty f_X(zx)f_Y(y)x\,\mathrm{d}x = \int_1^\infty \frac{\alpha}{(zx)^{\alpha+1}}\frac{\beta}{x^{\beta+1}}\,\mathrm{d}x$$
$$= \frac{\alpha\beta}{z^{\alpha+1}}\int_1^\infty x^{-\alpha-\beta-1}\,\mathrm{d}x = \frac{\alpha\beta}{z^{\alpha+1}}\left[\frac{-x^{-\alpha-\beta}}{\alpha+\beta}\right]_1^\infty = \frac{\alpha\beta}{\alpha+\beta}\frac{1}{z^{\alpha+1}}.$$

11.11 a Put $T_3 = X_1 + X_2 + X_3$. Then T_3 can be interpreted as the time of the third success in a series of independent experiments with probability p of success. Then $T_3 = k$ means that the first $(k-1)$ experiments contained exactly 2 successes, and the k-th experiment was also a success. Since the number of successes in $(k-1)$ experiments has a $Bin(k-1, p)$ distribution, it follows that

$$\mathrm{P}(T_3 = k) = \binom{k-1}{2} p^2 (1-p)^{(k-1-2)} \cdot p = \frac{1}{2}(k-1)(k-2)p^3(1-p)^{k-3}.$$

11.11 b This follows from a smart calculation:

$$\begin{aligned}
&\sum_{k=3}^{\infty} p_Z(k) = 1 \\
\Leftrightarrow &\sum_{k=3}^{\infty} \frac{1}{2}(k-1)(k-2)p^3(1-p)^{k-3} = 1 \\
\Leftrightarrow &\frac{1}{2}p^2 \sum_{k=3}^{\infty}(k-1)(k-2)p(1-p)^{k-3} = 1 \quad (\text{now put } k-2=m) \\
\Leftrightarrow &\frac{1}{2}p^2 \sum_{m=1}^{\infty} m(m+1)p(1-p)^{m-1} = 1 \\
\Leftrightarrow &\frac{1}{2}p^2 \sum_{m=1}^{\infty} (m^2+m)\mathrm{P}(X_1 = m) = 1 \\
\Leftrightarrow &\frac{1}{2}p^2 \left(\mathrm{E}\left[X_1^2\right] + \mathrm{E}\left[X_1\right]\right) = 1 \\
\Leftrightarrow &\, p^2 \left(\mathrm{E}\left[X_1^2\right] + \mathrm{E}\left[X_1\right]\right) = 2.
\end{aligned}$$

11.11 c The first part follows directly from **b**:

$$\mathrm{E}\left[X_1^2\right] = \frac{2}{p^2} - \mathrm{E}\left[X_1\right] = \frac{2}{p^2} - \frac{1}{p} = \frac{2-p}{p^2}.$$

For the second part:

$$\mathrm{Var}(X_1) = \mathrm{E}\left[X_1^2\right] - \left(\mathrm{E}\left[X_1\right]\right)^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

11.12 $\Gamma(1) = \int_0^\infty \mathrm{e}^{-t}\,\mathrm{d}t = 1$ and

$$\Gamma(x+1) = \int_0^\infty t^x \mathrm{e}^{-t}\,\mathrm{d}t = \left[-t^x \mathrm{e}^{-t}\right]_{t=0}^{t=\infty} + \int_0^\infty x t^{x-1} \mathrm{e}^{-t}\,\mathrm{d}t.$$

Since $x > 0$, the first term on the right hand side is zero and the second term is equal to $x\Gamma(x)$.

11.13 a

$$\begin{aligned}
\mathrm{P}(Z_n \le a) &= \int_0^a \frac{\lambda^n z^{n-1} \mathrm{e}^{-\lambda z}}{(n-1)!}\,\mathrm{d}z \\
&= \left[-\frac{\lambda^{n-1} z^{n-1} \mathrm{e}^{-\lambda z}}{(n-1)!}\right]_{z=0}^{z=a} + \int_0^a \frac{\lambda^{n-1} z^{n-2} \mathrm{e}^{-\lambda z}}{(n-2)!}\,\mathrm{d}z \\
&= -\frac{(\lambda a)^{n-1}}{(n-1)!}\mathrm{e}^{-\lambda a} + \int_0^a \frac{\lambda^{n-1} z^{n-2} \mathrm{e}^{-\lambda z}}{(n-2)!}\,\mathrm{d}z \\
&= -\frac{(\lambda a)^{n-1}}{(n-1)!}\mathrm{e}^{-\lambda a} + \mathrm{P}(Z_{n-1} \le a)\,.
\end{aligned}$$

11.13 b Use part **a** recursively:

$$\begin{aligned}
\mathrm{P}(Z_n \le a) &= -\frac{(\lambda a)^{n-1}}{(n-1)!}\,\mathrm{e}^{-\lambda a} + \mathrm{P}(Z_{n-1} \le a) \\
&= -\frac{(\lambda a)^{n-1}}{(n-1)!}\mathrm{e}^{-\lambda a} - \frac{(\lambda a)^{n-2}}{(n-2)!}\mathrm{e}^{-\lambda a} + \mathrm{P}(Z_{n-2} \le a) \\
&\;\vdots \\
&= -\sum_{i=1}^{n-1} \frac{(\lambda a)^i}{i!}\mathrm{e}^{-\lambda a} + \mathrm{P}(Z_1 \le a)\,.
\end{aligned}$$

11.13 c Since $\mathrm{P}(Z_1 \le a) = 1 - \mathrm{e}^{-\lambda a}$, we find

$$\begin{aligned}
\mathrm{P}(Z_n \le a) &= \mathrm{P}(Z_1 \le a) - \sum_{i=1}^{n-1} \frac{(\lambda a)^i}{i!}\mathrm{e}^{-\lambda a} \\
&= 1 - \mathrm{e}^{-\lambda a} - \sum_{i=1}^{n-1} \frac{(\lambda a)^i}{i!}\mathrm{e}^{-\lambda a} = 1 - \sum_{i=0}^{n-1} \frac{(\lambda a)^i}{i!}\mathrm{e}^{-\lambda a}.
\end{aligned}$$

12.1 e This is certainly open to discussion. Bankruptcies: no (they come in clusters, don't they?). Eggs: no (I suppose after one egg it takes the chicken some time to produce another). Examples 3 and 4 are the best candidates. Example 5 could be modeled by the Poisson process if the crossing is not a dangerous one; otherwise authorities might take measures and destroy the homogeneity.

12.2 Let X be the number of customers on a day. Given is that $\mathrm{P}(X = 0) = 10^{-5}$. Since X is Poisson distributed, $\mathrm{P}(X = 0) = \mathrm{e}^{-\lambda}$. So $\mathrm{e}^{-\lambda} = 10^{-5}$, which implies $-\lambda = -5\ln(10)$, and hence $\lambda = 11.5$. Then also $\mathrm{E}[X] = \lambda = 11.5$.

12.3 When N has a $Pois(4)$ distribution, $\mathrm{P}(N = 4) = 4^4\mathrm{e}^{-4}/4! = 0.195$.

12.4 When X has a $Pois(2)$ distribution, $\mathrm{P}(X \le 1) = \mathrm{P}(X = 0) + \mathrm{P}(X = 1) = \mathrm{e}^{-2} + 2\mathrm{e}^{-2} = 3\mathrm{e}^{-2} = 0.406$.

12.5 a Model the errors in the bytes of the hard disk as a Poisson process, with intensity λ per byte. Given is that $\lambda \cdot 2^{20} = 1$, or $\lambda = 2^{-20}$. The expected number of errors in $512 = 2^9$ bytes is $\lambda 2^9 = 2^{-11} = 0.00049$.

12.5 b Let Y be the number of errors on the hard disk. Then Y has a Poisson distribution with parameter $\mu = 39054015 \times 0.00049 = 19069.34$. Then $\mathrm{P}(Y \ge 1) = 1 - \mathrm{P}(Y = 0) = 1 - \mathrm{e}^{-19069.34} = 1.00000\cdots$. For all practical purposes this happens with probability 1.

12.6 The expected numbers of flaws in 1 meter is $100/40 = 2.5$, and hence the number of flaws X has a $Pois(2.5)$ distribution. The answer is $\mathrm{P}(X = 2) = \frac{1}{2!}(2.5)^2 \mathrm{e}^{-2.5} = 0.256$.

12.7 a It is reasonable to estimate λ with (nr. of cars)/(total time in sec.) $= 0.192$.

12.7 b $19/120 = 0.1583$, and if $\lambda = 0.192$ then $\mathrm{P}(N(10) = 0) = \mathrm{e}^{-0.192 \cdot 10} = 0.147$.

12.7 c $\mathrm{P}(N(10) = 10)$ with λ from **a** seems a reasonable approximation of this probability. It equals $\mathrm{e}^{-1.92} \cdot (0.192 \cdot 10)^{10}/10! = 2.71 \cdot 10^{-5}$.

12.8 a We have

$$\begin{aligned} \mathrm{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1)\mathrm{e}^{-\lambda}\lambda^k/k! = \sum_{k=2}^{\infty} \mathrm{e}^{-\lambda}\lambda^k/(k-2)! \\ &= \lambda^2 \sum_{k=2}^{\infty} \mathrm{e}^{-\lambda}\lambda^{k-2}/(k-2)! = \lambda^2 \sum_{j=0}^{\infty} \mathrm{e}^{-\lambda}\lambda^j/j! = \lambda^2. \end{aligned}$$

12.8 b Since $\mathrm{Var}(X) = \mathrm{E}[X(X-1)] + \mathrm{E}[X] - (\mathrm{E}[X])^2$, we have $\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

12.9 In a Poisson process with intensity 1, the number of points in a interval of length t has a Poisson distribution with parameter t. So the number Y_1 of points in $[0, \mu_1]$ is $Pois(\mu_1)$ distributed, and the number Y_2 in $[\mu_1 + \mu_1 + \mu_2]$ is $Pois(\mu_2)$ distributed. But the sum $Y_1 + Y_2$ of these is equal to the number of points in $[0, \mu_1 + \mu_2]$, and so is $Pois(\mu_1 + \mu_2)$ distributed.

12.10 We have to consider the numbers $p_k = \mathrm{P}(X = k) = \frac{\mu^k}{k!}\mathrm{e}^{-\mu}$. To compare them, divide: $p_{k+1}/p_k = \mu/(k+1)$, and note that $p_k < (\text{ or } >) p_{k+1}$ is equivalent to $p_{k+1}/p_k > (\text{ or } <) 1$.
So it follows immediately that
— for $\mu < 1$ the probabilities $\mathrm{P}(X = k)$ are decreasing, and
— for $\mu > 1$ the probabilities are increasing as long as $\frac{\mu}{k+1} > 1$,
and decreasing from the moment where this fraction has become less than 1.
Lastly if $\mu = 1$, then $p_0 = p_1 > p_2 > p_3 > \ldots$.

12.11 Following the hint, we obtain:

$$\begin{aligned} \mathrm{P}(N([0,s] = k, N([0,2s]) = n) &= \mathrm{P}(N([0,s]) = k, N((s,2s]) = n-k) \\ &= \mathrm{P}(N([0,s]) = k) \cdot \mathrm{P}(N((s,2s]) = n-k) \\ &= (\lambda s)^k \mathrm{e}^{-\lambda s}/(k!) \cdot (\lambda s)^{n-k}\mathrm{e}^{-\lambda s}/((n-k)!) \\ &= (\lambda s)^n \mathrm{e}^{-\lambda 2s}/(k!(n-k)!). \end{aligned}$$

So

$$\begin{aligned} \mathrm{P}(N([0,s]) = k \mid N([0,2s]) = n) &= \frac{\mathrm{P}(N([0,s]) = k, N([0,2s]) = n)}{\mathrm{P}(N([0,2s]) = n)} \\ &= n!/(k!(n-k)!) \cdot (\lambda s)^n/(2\lambda s)^n \\ &= n!/(k!(n-k)!) \cdot (1/2)^n. \end{aligned}$$

This holds for $k = 0, \ldots, n$, so we find the $Bin(n, \frac{1}{2})$ distribution.

12.12 a The event $\{X_2 \leq t\}$ is a disjoint union of the events $\{X_1 \leq s, X_2 \leq t\}$ and $\{X_1 > s, X_2 \leq t\}$.

12.12 b If $X_2 \le t$ and $N_a = 2$, then the first two points of the Poisson process must lie in $[0,t]$, and *no* points lie in $(t,a]$, and conversely. Therefore

$$\begin{aligned} \mathrm{P}(X_2 \le t, N_a = 2) &= \mathrm{P}(N([0,t]) = 2, N((t,a]) = 0) \\ &= \tfrac{\lambda^2 t^2}{2!}\mathrm{e}^{-\lambda t} \cdot \mathrm{e}^{-\lambda(a-t)} \\ &= \tfrac{1}{2}\lambda^2 t^2 \mathrm{e}^{-\lambda a}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathrm{P}(X_1 > s, X_2 \le t, N_a = 2) &= \mathrm{P}(N([0,s)) = 0, N([s,t]) = 2, N((t,a]) = 0) \\ &= \mathrm{e}^{-\lambda s} \cdot \tfrac{\lambda^2 (t-s)^2}{2!}\mathrm{e}^{-\lambda(t-s)} \cdot \mathrm{e}^{-\lambda(a-t)} \\ &= \tfrac{1}{2}\lambda^2 (t-s)^2 \mathrm{e}^{-\lambda a}. \end{aligned}$$

The result now follows by using part **a.**

12.12 c This follows immediately from part **b** since

$$\mathrm{P}(X_1 \le s, X_2 \le t \,|\, N_a = 2) = \frac{\mathrm{P}(X_1 \le s, X_2 \le t, N_a = 2)}{\mathrm{P}(N_a = 2)},$$

where the denominator equals $\frac{1}{2}\lambda^2 a^2\, \mathrm{e}^{-\lambda a}$.

12.13 a Given that $N_t = n + m$, the number of daisies X_t has a binomial distribution with parameters $n + m$ and $1/4$. So $\mathrm{P}(X_t = n, Y_t = m | N_t = n + m) = \mathrm{P}(X_t = n | N_t = n + m) = \binom{n+m}{n}\left(\frac{1}{4}\right)^n\left(\frac{3}{4}\right)^m$.

12.13 b Using part **a** we find that $\mathrm{P}(X_t = n, Y_t = m) = \mathrm{P}(X_t = n, Y_t = m | N_t = n + m)\,\mathrm{P}(N_t = n + m) = \binom{n+m}{n}\left(\frac{1}{4}\right)^n\left(\frac{3}{4}\right)^m \frac{(\lambda t)^{n+m}\mathrm{e}^{-\lambda t}}{(n+m)!} = \frac{1}{n!}\frac{1}{m!}\left(\frac{1}{4}\right)^n\left(\frac{3}{4}\right)^m (\lambda t)^{n+m}\mathrm{e}^{-\lambda t}$.

12.13 c Using **b** we find that $\mathrm{P}(X_t = n) = \sum_{m=0}^{\infty} \mathrm{P}(X_t = n, Y_t = m) = \frac{1}{n!}(\frac{1}{4})^n (\lambda t)^n \mathrm{e}^{-(\lambda/4)t} \sum_{m=0}^{\infty} \frac{1}{m!}(3\lambda/4)^m \mathrm{e}^{-(3\lambda/4)t} = \frac{1}{n!}((\lambda/4)t)^n \mathrm{e}^{-(\lambda/4)t}$, since the sum over m just adds all the probabilities $\mathrm{P}(Z = m)$ where Z has a *Pois*$(3\lambda/4)$ distribution.

12.14 a This follows since $1 - 1/n \to 1$ and $1/n \to 0$.

12.14 b This is an easy computation: $\mathrm{E}[X_n] = (1 - \frac{1}{n}) \cdot 0 + (\frac{1}{n}) \cdot 7n = 7$ for all n.

13.1 For $U(-1,1)$, $\mu = 0$ and $\sigma = 1/\sqrt{3}$, and we obtain the following table:

k	1	2	3	4
$\mathrm{P}(\lvert Y - \mu\rvert < k\sigma)$	0.577	1	1	1

For $U(-a,a)$, $\mu = 0$ and $\sigma = a/\sqrt{3}$, and we obtain the following table:

k	1	2	3	4
$\mathrm{P}(\lvert Y - \mu\rvert < k\sigma)$	0.577	1	1	1

For $N(0,1)$, $\mu = 0$ and $\sigma = 1$, and we obtain the following table:

k	1	2	3	4
$\mathrm{P}(\lvert Y - \mu\rvert < k\sigma)$	0.6826	0.9544	0.9974	1

For $N(\mu, \sigma^2)$, $\mathrm{P}(|Y - \mu| < k\sigma) = \mathrm{P}(|Y - \mu|/\sigma < k)$, and we obtain the same result as for $N(0, 1)$.
For $Par(3)$, $\mu = 3/2$, $\sigma = \sqrt{3/4}$, and

$$\mathrm{P}(|Y - \mu| < k\sigma) = \int_1^{3/2+k\sqrt{3}/2} 3x^{-4}\,\mathrm{d}x = 1 - (3/2 + k\sqrt{3}/2)^{-3},$$

and we obtain the following table:

k	1	2	3	4
$\mathrm{P}(\lvert Y - \mu\rvert < k\sigma)$	0.925	0.970	0.985	0.992

For $Geo(3)$, $\mu = 2$, $\sigma = 2$, and $\mathrm{P}(|Y - \mu| < k\sigma) = \mathrm{P}\big(|Y - 2| < k\sqrt{2}\big) = \mathrm{P}\big(2 - k\sqrt{2} < Y < 2 + k\sqrt{2}\big) = \mathrm{P}\big(Y < 2 + k\sqrt{2}\big)$, and we obtain the following table:

k	1	2	3	4
$\mathrm{P}(\lvert Y - \mu\rvert < k\sigma)$	0.875	0.9375	0.9844	0.9922

13.2 a From the formulas for the $U(a, b)$ distribution, substituting $a = -1/2$ and $b = 1/2$, we derive that $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1/12$.

13.2 b We write $S = X_1 + X_2 + \cdots + X_{100}$, for which we find $\mathrm{E}[S] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_{100}] = 0$ and, by independence, $\mathrm{Var}(S) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_{100}) = 100 \cdot \frac{1}{12} = 100/12$. We find from Chebyshev's inequality:

$$\mathrm{P}(|S| > 10) = \mathrm{P}(|S - 0| > 10) \leq \frac{\mathrm{Var}(S)}{10^2} = \frac{1}{12}.$$

13.3 We can apply the law of large numbers to the sequence (Y_i), with $Y_i = |X_i|$ for $i = 1, \ldots$. Since $\mathrm{E}[Y_i] = \mathrm{E}[|X_i|] = 2\int_0^{0.5} x\,\mathrm{d}x = 2 \cdot \frac{1}{2} \cdot (0.5)^2 = 0.25$, and since the Y_i have finite variance, it follows that $\frac{1}{n}\sum_{i=1}^n |X_i| \to 0.25$ as $n \to \infty$.

13.4 a Because X_i has a $Ber(p)$ distribution, $\mathrm{E}[X_i] = p$ and $\mathrm{Var}(X_i) = p\,(1 - p)$, and so $\mathrm{E}\big[\bar{X}_n\big] = p$ and $\mathrm{Var}\big(\bar{X}_n\big) = \mathrm{Var}(X_i)\,/n = p\,(1 - p)/n$. By Chebyshev's inequality:

$$\mathrm{P}\big(|\bar{X}_n - p| \geq 0.2\big) \leq \frac{p\,(1 - p)/n}{(0.2)^2} = \frac{25p(1 - p)}{n}.$$

The right-hand side should be at most 0.1 (note that we switched to the complement). If $p = 1/2$ we therefore require $25/(4n) \leq 0.1$, or $n \geq 25/(4 \cdot 0.1) = 62.5$, i.e., $n \geq 63$. Now, suppose $p \neq 1/2$, using $n = 63$ and $p(1 - p) \leq 1/4$ we conclude that $25p(1 - p)/n \leq 25 \cdot (1/4)/63 = 0.0992 < 0.1$, so (because of the inequality) the computed value satisfies for other values of p as well.

13.4 b For arbitrary $a > 0$ we conclude from Chebyshev's inequality:

$$\mathrm{P}\big(|\bar{X}_n - p| \geq a\big) \leq \frac{p\,(1 - p)/n}{a^2} = \frac{p(1 - p)}{na^2} \leq \frac{1}{4na^2},$$

where we used $p\,(1 - p) \leq 1/4$ again. The question now becomes: when $a = 0.1$, for what n is $1/(4na^2) \leq 0.1$? We find: $n \geq 1/(4 \cdot 0.1 \cdot (0.1)^2) = 250$, so $n = 250$ is large enough.

13.4 c From part **a** we know that an error of size 0.2 or occur with a probability of at most $25/4n$, regardless of the values of p. So, we need $25/(4n) \leq 0.05$, i.e., $n \geq 25/(4 \cdot 0.05) = 125$.

13.4 d We compute $\mathrm{P}(\bar{X}_n \leq 0.5)$ for the case that $p = 0.6$. Then $\mathrm{E}[\bar{X}_n] = 0.6$ and $\mathrm{Var}(\bar{X}_n) = 0.6 \cdot 0.4/n$. Chebyshev's inequality cannot be used directly, we need an intermediate step: the probability that $\bar{X}_n \leq 0.5$ is contained in the event "the prediction is off by at least 0.1, in either direction." So

$$\mathrm{P}(\bar{X}_n \leq 0.5) \leq \mathrm{P}(|\bar{X}_n - 0.6| \geq 0.1) \leq \frac{0.6 \cdot 0.4/n}{(0.1)^2} = \frac{24}{n}$$

For $n \geq 240$ this probability is 0.1 or smaller.

13.5 To get $\mathrm{P}(|\bar{M}_n - c| \leq 0.5) \geq 0.9$, we must have $\mathrm{P}(|\bar{U}_n| \leq 0.5) \geq 0.9$. Now $\mathrm{P}(|\bar{U}_n| \geq 0.5) \leq \frac{\mathrm{Var}(U_i)/n}{(0.5)^2} = \frac{3/n}{0.25} = \frac{12}{n}$. If we want this below 0.1, we should take $n \geq 120$.

13.6 The probability distribution for an individual game has $\mathrm{P}(X = -1) = \frac{18}{37}$, and $\mathrm{P}(X = 1) = \frac{19}{37}$. From this we compute $\mathrm{E}[X] = (-1) \cdot \frac{18}{37} + 1 \cdot \frac{19}{37} = \frac{1}{37}$. If the game is played $365 \cdot 1000$ times, and we suppose that the roulette machine is "fair", the total gain will, according to the law of large numbers, be close to $365 \cdot 1000 \cdot \frac{1}{37} = 9865$ (Euro). (Using Chebyshev's inequality one can, for example, find a lower bound for the probability that the total gain will be between $9865 - 250$ and $9865 + 250$.)

13.7 Following the hint we define $Y_i = 1$ when $X_i \in (-\infty, a]$, and 0 otherwise. In this way we have written $F_n(a) = \frac{1}{n}\sum_{i=1}^{n} Y_i$. The expectation of Y_i equals $\mathrm{E}[Y_i] = \mathrm{P}(X_i \in (-\infty, a]) = F(a)$. Moreover, $\mathrm{Var}(Y_i) = F(a)(1 - F(a))$ is finite. Since $F_n(a) = \bar{Y}_n$, the law of large numbers tells us that

$$\lim_{n\to\infty} \mathrm{P}(|F_n(a) - F(a)| > \varepsilon) = 0.$$

13.8 a We have $p = \mathrm{E}[Y] = \int_{a-h}^{a+h} x\mathrm{e}^{-x}\,\mathrm{d}x \approx 2h \cdot 2\mathrm{e}^{-2}$. (Of course the integral can be computed exactly, but this approximation is excellent: for $h = 0.25$ we get $p = 0.13534$, while the exact result is $p = 0.13533$.) Now $\mathrm{Var}(Y) = p(1-p) = 4h\mathrm{e}^{-2}(1 - 4h\mathrm{e}^{-2})$, and

$$\mathrm{Var}(\bar{Y}_n/2h) = \frac{1}{4h^2 n} 4h\mathrm{e}^{-2}(1 - 4h\mathrm{e}^{-2}).$$

13.8 b What is required is that

$$\mathrm{P}(|\bar{Y}_n/2h - f(a)| \geq 0.2f(a)) \leq 0.2.$$

Chebyshev's inequality yields that

$$\mathrm{P}(|\bar{Y}_n/2h - f(a)| \geq 0.2f(a)) \leq \frac{\mathrm{Var}(\bar{Y}_n/2h)}{0.04f(a)^2},$$

so we want $\mathrm{Var}(\bar{Y}_n/2h) \leq 0.008 \cdot (2\mathrm{e}^{-2})^2$. Using part **a** this leads to

$$\frac{1}{n} \leq \frac{0.008 \cdot \mathrm{e}^{-2}}{1 - \mathrm{e}^{-2}} \quad \Longrightarrow \quad n \geq \frac{\mathrm{e}^2 - 1}{0.008} = 798.6 \quad \Longrightarrow \quad n = 799.$$

13.9 a The statement looks like the law of large numbers, and indeed, if we look more closely, we see that T_n is the average of an i.i.d. sequence: define $Y_i = X_i^2$, then $T_n = \bar{Y}_n$. The law of large numbers now states: if $\bar{Y}_n$ is the average of n independent random variables with expectation μ and variance σ^2, then for any $\varepsilon > 0$: $\lim_{n\to\infty} \mathrm{P}\big(|\bar{Y}_n - \mu| > \varepsilon\big) = 0$. So, if $a = \mu$ and the variance σ^2 is finite, then it is true.

13.9 b We compute expectation and variance of Y_i: $\mathrm{E}[Y_i] = \mathrm{E}\left[X_i^2\right] = \int_{-1}^{1} \frac{1}{2}x^2 \,\mathrm{d}x = 1/3$. And: $\mathrm{E}\left[Y_i^2\right] = \mathrm{E}\left[X_i^4\right] = \int_{-1}^{1} \frac{1}{2}x^4 \,\mathrm{d}x = 1/5$, so $\mathrm{Var}(Y_i) = 1/5 - (1/3)^2 = 4/45$. The variance is finite, so indeed, the law of large numbers applies, and the statement is true if $a = \mathrm{E}\left[X_i^2\right] = 1/3$.

13.10 a For $0 \leq \varepsilon \leq 1$:

$$\mathrm{P}(|M_n - 1| > \varepsilon) = \mathrm{P}(1 - \varepsilon < M_n < 1 + \varepsilon) = \mathrm{P}(1 - \varepsilon < M_n) = (1 - \varepsilon)^n.$$

13.10 b For any $0 < \varepsilon$: $\lim_{n\to\infty}(1 - \varepsilon)^n = 0$. The conclusion is that M_n converges to 1, as n goes to infinity. This cannot be obtained by a straightforward application of Chebyshev's inequality or the law of large numbers.

13.11 a We have $\mathrm{P}(|X - 1| > 8) = \mathrm{P}(X = 10) = 1/10$. On the other hand we get from Chebyshev's inequality $\mathrm{P}(|X - 1| > 8) \leq (t - 1)/64 = 9/64$.

13.11 b For $a = 5$ we have
EXACT: $\mathrm{P}(|X - 1| > 5) = \mathrm{P}(X = 10) = 1/10$, CHEB: $\mathrm{P}(|X - 1| > 5) \leq 9/25$,
so the Chebyshev gap equals 9/25-1/10=0.26. For $a = 10$ we have
EXACT: $\mathrm{P}(|X - 1| > 10) = 0$, CHEB: $\mathrm{P}(|X - 1| > 10) \leq 9/100$,
so the Chebyshev gap equals 9/100=0.09.

13.11 c Choosing $a = t$ as in the previous question, we have $\mathrm{P}(|X - 1| > t) = 0$, and $\mathrm{P}(|X - 1| > t) \leq (t - 1)/t^2 < 1/t$. So to answer the question we can simply choose $t = 100, t = 1000$ and $t = 10\,000$.

13.11 d As we have seen that the Chebyshev gap can be made arbitrarily small (taking arbitrarily large t in part **c**) we can not find a closer bound for this family of probability distributions.

13.12 a Since $\mathrm{E}\left[\bar{X}_n\right] = \frac{1}{n}\sum_{i=1}^{n} \mu_i$, and $\mathrm{Var}\big(\bar{X}_n\big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i)$, this inequality follows directly from Chebyshev's.

13.12 b This follows directly from

$$\mathrm{Var}\big(\bar{X}_n\big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \leq \frac{1}{n}M \to 0 \quad \text{as} \quad n \to \infty.$$

14.1 Since $\mu = 2$ and $\sigma = 2$, we find that

$$\begin{aligned}
\mathrm{P}(X_1 + X_2 + \cdots + X_{144} > 144) &= \mathrm{P}\big(\bar{X}_{144} > 1\big) \\
&= \mathrm{P}\left(\sqrt{144}\frac{\bar{X}_{144} - \mu}{\sigma} > \sqrt{144}\frac{1 - \mu}{\sigma}\right) \\
&= \mathrm{P}\left(Z_{144} > 12\frac{1 - 2}{2}\right) \\
&\approx 1 - \Phi(-6) = 1.
\end{aligned}$$

14.2 Since $\mathrm{E}[X_i] = 1/4$ and $\mathrm{Var}(X_i) = 3/80$, using the central limit theorem we find

$$\begin{aligned}\mathrm{P}(X_1 + X_2 + \cdots + X_{625} < 170) &= \mathrm{P}\left(\sqrt{625}\frac{\bar{X}_{625} - 1/4}{\sqrt{3/80}} < \sqrt{625}\frac{170/625 - 1/4}{\sqrt{3/80}}\right) \\ &\approx \Phi(2.8402) = 1 - 0.0023 = 0.9977.\end{aligned}$$

14.3 First note that $\mathrm{P}\big(|\bar{X}_n - p| < 0.2\big) = 1 - \mathrm{P}\big(\bar{X}_n - p \geq 0.2\big) - \mathrm{P}\big(\bar{X}_n - p \leq -0.2\big)$. Because $\mu = p$ and $\sigma^2 = p(1-p)$, we find, using the central limit theorem:

$$\begin{aligned}\mathrm{P}\big(\bar{X}_n - p \geq 0.2\big) &= \mathrm{P}\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \\ &= \mathrm{P}\left(Z_n \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \approx \mathrm{P}\left(Z \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right),\end{aligned}$$

where Z has an $N(0,1)$ distribution. Similarly,

$$\mathrm{P}\big(\bar{X}_n - p \leq -0.2\big) \approx \mathrm{P}\left(Z \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right),$$

so we are looking for the smallest positive integer n such that

$$1 - 2\mathrm{P}\left(Z \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \geq 0.9,$$

i.e., the smallest positive integer n such that

$$\mathrm{P}\left(Z \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \leq 0.05.$$

From Table **??** it follows that

$$\sqrt{n}\frac{0.2}{\sqrt{p(1-p)}} \geq 1.645.$$

Since $p(1-p) \leq 1/4$ for all p between 0 and 1, we see that n should be at least 17.

14.4 Using the central limit theorem, with $\mu = 2$, and $\sigma^2 = 4$,

$$\begin{aligned}\mathrm{P}(T_1 + T_2 + \cdots + T_{30} \leq 60) &= \mathrm{P}\left(\sqrt{30}\frac{\bar{T}_{30} - 2}{2} \leq \sqrt{30}\frac{2-2}{2}\right) \\ &\approx \Phi(0) = \frac{1}{2}.\end{aligned}$$

14.5 In Section 4.3 we have seen that X has the same probability distribution as $X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent $Ber(p)$ distributed random variables. Recall that $\mathrm{E}[X_i] = p$, and $\mathrm{Var}(X_i) = p(1-p)$. But then we have for any real number a that

$$\mathrm{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \leq a\right) = \mathrm{P}\left(\frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} \leq a\right) = \mathrm{P}(Z_n \leq a);$$

see also (14.1). It follows from the central limit theorem that

$$\mathrm{P}\left(\frac{X-np}{\sqrt{np(1-p)}} \leq a\right) \approx \Phi(a),$$

i.e., the random variable $\frac{X-np}{\sqrt{np(1-p)}}$ has a distribution that is approximately standard normal.

14.6 a Using the result of Exercise 14.5 yields that

$$\begin{aligned}\mathrm{P}(X \leq 25) &= \mathrm{P}\left(\frac{X-25}{\sqrt{75/4}} \leq \frac{25-25}{\sqrt{75/4}}\right)\\ &= \mathrm{P}\left(\frac{X-25}{\sqrt{75/4}} \leq 0\right)\\ &\approx \mathrm{P}(Z \leq 0)\\ &= \frac{1}{2},\end{aligned}$$

where Z has a standard normal distribution. In the same way,

$$\begin{aligned}\mathrm{P}(X < 26) &= \mathrm{P}\left(\frac{X-25}{\sqrt{75/4}} \leq \frac{26-25}{\sqrt{75/4}}\right)\\ &= \mathrm{P}\left(\frac{X-25}{\sqrt{75/4}} \leq \frac{2}{5\sqrt{3}}\right)\\ &\approx \mathrm{P}(Z \leq 0.2309)\\ &= 0.591.\end{aligned}$$

14.6 b $\mathrm{P}(X \leq 2) \approx \mathrm{P}(Z \leq -5.31) \approx 0$ (the table ends with $z = 3.49$).

14.7 Setting $Y = \frac{n^{1/4}}{\sigma}\bar{X}_n$, one easily shows that $\mathrm{E}[Y] = \frac{n^{1/4}}{\sigma}\mu$, and $\mathrm{Var}(Y) = 1/\sqrt{n}$. Since

$$\mathrm{P}\left(\left|n^{\frac{1}{4}}\frac{\bar{X}_n-\mu}{\sigma}\right| \geq a\right) = \mathrm{P}(|Y - \mathrm{E}[Y]| \geq a),$$

it follows from Chebyshev's inequality that

$$\mathrm{P}\left(\left|n^{\frac{1}{4}}\frac{\bar{X}_n-\mu}{\sigma}\right| \geq a\right) \leq \frac{1}{a^2\sqrt{n}}.$$

As a consequence, if n goes to infinity, we see that most of the probability mass of the random variable $n^{\frac{1}{4}}\frac{\bar{X}_n-\mu}{\sigma}$ is concentrated in the interval between $-a$ and a, for every $a > 0$. Since we can choose a arbitrarily small, this explains the spike in Figure 14.1.

14.8 a $1 = \mathrm{Var}(X_i) = \mathrm{E}\left[X_i^2\right] - (\mathrm{E}[X_i])^2 = \mathrm{E}\left[X_i^2\right] - 0$, so $\mathrm{E}\left[X_i^2\right] = 1$.

14.8 b (integration by parts: $\mathrm{E}\left[X_i^4\right] = \int x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx = \frac{-1}{\sqrt{2\pi}}\left(\left[x^3 e^{-\frac{1}{2}x^2}\right] - \int 3x^2 e^{-\frac{1}{2}x^2}\,dx\right) = 3\int x e^{-\frac{1}{2}x^2}\,dx = 3\mathrm{E}\left[X_i^2\right] = 3$)
$\mathrm{Var}\left(X_i^2\right) = \mathrm{E}\left[X_i^4\right] - \left(\mathrm{E}\left[X_i^2\right]\right)^2 = 3 - 1 = 2$.

14.8 c $\mathrm{P}(Y_{100} > 110) \approx \mathrm{P}\left(Z > \frac{1}{\sqrt{2}}\right) \approx 0.242.$

14.9 a The probability that for a chain of at least 50 meters more than 1002 links are needed is the same as the probability that a chain of 1002 chains is shorter than 50 meters. Assuming that the random variables $X_1, X_2, \ldots, X_{1002}$ are independent, and using the central limit theorem, we have that

$$\mathrm{P}(X_1 + X_2 + \cdots + X_{1002} < 5000) \approx \mathrm{P}\left(Z < \sqrt{1002} \cdot \frac{\frac{5000}{1002} - 5}{\sqrt{0.04}}\right) = 0.0571,$$

where Z has an $N(0,1)$ distribution. So about 6% of the customers will receive a free chain.

14.9 b We now have that

$$\mathrm{P}(X_1 + X_2 + \cdots + X_{1002} < 5000) \approx \mathrm{P}(Z < 0.0032)\,,$$

which is slightly larger than 1/2. So about half of the customers will receive a free chain. Clearly something has to be done: a seemingly minor change of expected value has major consequences!

14.10 a Note that

$$\begin{aligned}\mathrm{P}\left(|\bar{M}_n - c| \le 0.5\right) &= \mathrm{P}\left(\frac{-\sqrt{3}}{6}\sqrt{n}\frac{\bar{M}_n - c}{\sigma} < \frac{\sqrt{3}}{6}\sqrt{n}\right)\\ &\approx \mathrm{P}\left(\frac{-\sqrt{3}}{6}\sqrt{n} < Z < \frac{\sqrt{3}}{6}\sqrt{n}\right)\\ &= 1 - 2\mathrm{P}\left(Z \ge \frac{\sqrt{3}}{6}\sqrt{n}\right),\end{aligned}$$

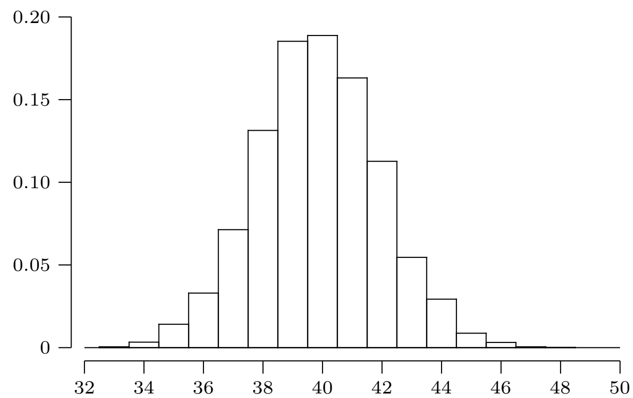
where Z is $N(0,1)$. Now choose n such that $1 - 2\mathrm{P}\left(Z \ge \frac{\sqrt{3}}{6}\sqrt{n}\right) = 0.9$, i.e., let n be such that $\mathrm{P}\left(Z \ge \frac{\sqrt{3}}{6}\sqrt{n}\right) = 0.05$. Then $\frac{\sqrt{3}}{6}\sqrt{n} = 2.75$ yielding that $n = 90.75$. Since n is an integer, we find $n = 91$.

14.10 b In case the U_i are normally distributed, the random variable $(\bar{M}_n - c)/\sigma$ has an $N(0,1)$ distribution, and the calculations in the answer of **a** are 'exact' (cf. Section 5.5). In all other cases this random variable has a distribution which is by approximation equal to an $N(0,1)$ distribution (see also the discussion in Section 14.2 on the size of n).

15.1 a For bin (32.5, 33.5], the height equals the number of x_j in B_i divided by $n|B_i|$: $3/(5732 \cdot 1) = 0.00052$. Similar computations for the other bins give

bin	height	bin	height
(32.5, 33.5]	0.00052	(40.5, 41.5]	0.16312
(33.5, 34.5]	0.00331	(41.5, 42.5]	0.11270
(34.5, 35.5]	0.01413	(42.5, 43.5]	0.05461
(35.5, 36.5]	0.03297	(43.5, 44.5]	0.02931
(36.5, 37.5]	0.07135	(44.5, 45.5]	0.00872
(37.5, 38.5]	0.13137	(45.5, 46.5]	0.00314
(38.5, 39.5]	0.18526	(46.5, 47.5]	0.00052
(39.5, 40.5]	0.18876	(47.5, 48.5]	0.00017

15.1 b Symmetric, since the histogram has one mode with more or less the same decay in both directions.



15.2 a For bin (50, 55], the height equals the number of x_j in B_i divided by $n|B_i|$: $1/(23 \cdot 5) = 0.0087$. The heights of the other bins can be computed similarly.

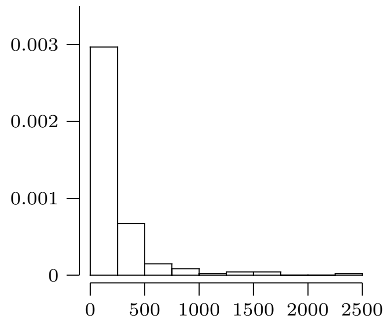
bin	height
(50,55]	0.0087
(55,60]	0.0174
(60,65]	0.0087
(65,70]	0.0870
(70,75]	0.0348
(75,80]	0.0348
(80,85]	0.0087

15.2 b No, 31 degrees is extremely low compared to the temperature experienced on earlier takeoffs.

15.3 a For bin (0, 250], the height equals the number of x_j in B_i divided by $n|B_i|$: $141/(190 \cdot 250) = 0.00297$. The heights of the other bins can be computed similarly.

Bin	Height
(0,250]	0.00297
(250,500]	0.00067
(500,750]	0.00015
(750,1000]	0.00008
(1000,1250]	0.00002
(1250,1500]	0.00004
(1500,1750]	0.00004
(1750,2000]	0
(2250,2500]	0
(2250,2500]	0.00002

15.3 b Skewed, since the histogram has one mode and only decays to the right.



15.4 a For bin (0, 500], the height equals the number of x_j in B_i divided by $n|B_i|$: $86/(135 \cdot 500) = 0.0012741$. The heights of the other bins can be computed similarly.

Bin	Height
[0,500]	0.0012741
(500,1000]	0.0003556
(1000,1500]	0.0001778
(1500,2000]	0.0000741
(2000,2500]	0.0000148
(2500,3000]	0.0000148
(3000,3500]	0.0000296
(3500,4000]	0
(4000,4500]	0.0000148
(4500,5000]	0
(5000,5500]	0.0000148
(5500,6000]	0.0000148
(6000,6500]	0.0000148

15.4 b Since all elements (not rounded) are strictly positive, the value of F_n at zero is zero. At 500 the value of F_n is the number of $x_i \leq 500$ divided by n: $86/135 = 0.6370$. The values at the other endpoints can be computed similarly:

t	$F_n(t)$	t	$F_n(t)$
0	0	3500	0.9704
500	0.6370	4000	0.9704
1000	0.8148	4500	0.9778
1500	0.9037	5000	0.9778
2000	0.9407	5500	0.9852
2500	0.9481	6000	0.9926
3000	0.9556	6500	1

15.4 c The area under the histogram on bin (1000, 1500] is $500 \cdot 0.000178 = 0.0889$, and $F_n(1500) - F_n(1000) = 0.9037 - 0.8148 = 0.0889$.

15.5 Since the increase of F_n on bin (0, 1] is equal to the area under the histogram on the same bin, the height on this bin must be $0.2250/(1-0) = 0.2250$. Similarly, the height on bin (1, 3] is (0.445-0.225)/(3-1)=0.1100, and so on:

Bin	Height
$(0,1]$	0.2250
$(1,3]$	0.1100
$(3,5]$	0.0850
$(5,8]$	0.0400
$(8,11]$	0.0230
$(11,14]$	0.0350
$(14,18]$	0.0225

15.6 Because $(2-0)\cdot 0.245 + (4-2)\cdot 0.130 + (7-4)\cdot 0.050 + (11-7)\cdot 0.020 + (15-11)\cdot 0.005 = 1$, there are no data points outside the listed bins. Hence

$$\begin{aligned} F_n(7) &= \frac{\text{number of } x_i \leq 7}{n} \\ &= \frac{\text{number of } x_i \text{ in bins } (0,2],\ (2,4] \text{ and } (4,7]}{n} \\ &= \frac{n\cdot(2-0)\cdot 0.245 + n\cdot(4-2)\cdot 0.130 + n\cdot(7-4)\cdot 0.050}{n} \\ &= 0.490 + 0.260 + 0.150 = 0.9. \end{aligned}$$

15.7 Note that the increase of F_n on each bin is equal to the area under the histogram on the same bin. Since all bins have the same width, it follows from the results of Exercise 15.2 that the increase of F_n is the largest ($5\cdot 0.0.870 = 0.435$) on $(65,70]$.

15.8 K does not satisfy (K1), since it is negative between 0 and 1 and does not integrate to one.

15.9 a The scatterplot indicates that larger durations correspond to longer waiting times.

15.9 b By judgement by the eye, the authors predicted a waiting time of about 80.

15.9 c If you project the points on the vertical axes you get the dataset of waiting times. Since the two groups of points in the scatterplot are separated in North East direction (both vertical as well as horizontal), you will see two modes.

15.10 Each steep part of F_n corresponds to a mode of the dataset. One counts four of such parts.

15.11 The height of the histogram on a bin $(a,b]$ is

$$\begin{aligned} \frac{\text{number of } x_i \text{ in } (a,b]}{n(b-a)} &= \frac{(\text{number of } x_i \leq b) - (\text{number of } x_i \leq a)}{n(b-a)} \\ &= \frac{F_n(b) - F_n(a)}{b-a}. \end{aligned}$$

15.12 a By inserting the expression for $f_{n,h}(t)$, we get

$$\begin{aligned}\int_{-\infty}^{\infty} t \cdot f_{n,h}(t)\,\mathrm{d}t &= \int_{-\infty}^{\infty} t \cdot \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{t}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t.\end{aligned}$$

For each i fixed we find with change of integration variables $u = (t - x_i)/h$,

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{t}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t &= \int_{-\infty}^{\infty}(x_i + hu)K(u)\,\mathrm{d}u \\ &= x_i\int_{-\infty}^{\infty}K(u)\,\mathrm{d}u + h\int_{-\infty}^{\infty}uK(u)\,\mathrm{d}u = x_i,\end{aligned}$$

using that K integrates to one and that $\int_{-\infty}^{\infty} uK(u)\,\mathrm{d}u = 0$, because K is symmetric. Hence

$$\int_{-\infty}^{\infty} t \cdot f_{n,h}(t)\,\mathrm{d}t = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{t}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t = \frac{1}{n}\sum_{i=1}^{n}x_i.$$

15.12 b By means of a similar reasoning

$$\begin{aligned}\int_{-\infty}^{\infty} t^2 \cdot f_{n,h}(t)\,\mathrm{d}t &= \int_{-\infty}^{\infty} t^2 \cdot \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{t^2}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t.\end{aligned}$$

For each i:

$$\begin{aligned}&\int_{-\infty}^{\infty}\frac{t^2}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty}(x_i + hu)^2K(u)\,\mathrm{d}u = \int_{-\infty}^{\infty}(x_i^2 + 2x_ihu + h^2u^2)K(u)\,\mathrm{d}u \\ &= x_i^2\int_{-\infty}^{\infty}K(u)\,\mathrm{d}u + 2x_ih\int_{-\infty}^{\infty}uK(u)\,\mathrm{d}u + h^2\int_{-\infty}^{\infty}u^2K(u)\,\mathrm{d}u \\ &= x_i^2 + h^2\int_{-\infty}^{\infty}u^2K(u)\,\mathrm{d}u,\end{aligned}$$

again using that K integrates to one and that K is symmetric.

16.1 a The dataset has 135 elements, so the sample median is the 68th element in order of magnitude. From the table in Exercise 15.4 we see that this is 290.

16.1 b The dataset has $n = 135$ elements. The lower quartile is the 25th empirical percentile. We have $k = \lfloor 0.25 \cdot (135 + 1)\rfloor = 34$, so that $\alpha = 0$, and $q_n(0.25) = x_{(34)} = 81$, according to the table in Exercise 15.4. Similarly, the upper quartile is $q_n(0.75) = x_{(102)} = 843$, and the IQR is $843 - 81 = 762$.

16.1 c We have $k = \lfloor 0.37 \cdot (135 + 1)\rfloor = \lfloor 50.32\rfloor = 50$, so that $\alpha = 0.32$, and

$$q_n(0.37) = x_{(50)} + 0.32 \cdot (x_{(51)} - x_{(50)}) = 143 + 0.32 \cdot (148 - 143) = 144.6.$$

16.2 To compute the lower quartile, we use that $n = 272$, so that $k = \lfloor 0.25 \cdot 273 \rfloor = \lfloor 68.25 \rfloor = 68$ and $\alpha = 0.25$, and hence from Table 15.2:

$$q_n(0.25) = x_{(68)} + 0.25 \cdot (x_{(69)} - x_{(68)}) = 129 + 0.25 \cdot (130 - 129) = 129.25.$$

Similarly, $q_n(0.5) = 240$ and $q_n(0.75) = 267.75$. Hence, $|q_n(0.75) - q_n(0.5)| = 27.75$ and $|q_n(0.25) - q_n(0.5)| = 110.75$. The kernel density estimate (see Figure 15.3) has a 'peak' just right of 240, so there is relatively a lot of probability mass just right of the sample median, namely about 25% between 240 and 110.75.

16.3 a Because $n = 24$, the sample median is the average of the 12th and 13th element. Since these are both equal to 70, the sample median is also 70. The lower quartile is the pth empirical quantile for $p = 1/4$. We get $k = \lfloor p(n+1) \rfloor = \lfloor 0.25 \cdot (24+1) \rfloor = \lfloor 6.25 \rfloor = 6$, so that

$$q_n(0.25) = x_{(6)} + 0.25 \cdot (x_{(7)} - x_{(6)}) = 66 + 0.25 \cdot (67 - 66) = 66.25.$$

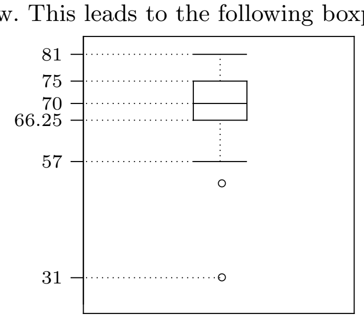
Similarly, the upper quartile is the pth empirical quantile for $p = 3/4$:

$$q_n(0.75) = x_{(18)} + 0.75 \cdot (x_{(19)} - x_{(18)}) = 75 + 0.75 \cdot (75 - 75) = 75.$$

16.3 b In part **a** we found the sample median and the two quartiles. From this we compute the IQR: $q_n(0.75) - q_n(0.25) = 75 - 66.25 = 8.75$. This means that

$$\begin{aligned} q_n(0.25) - 1.5 \cdot \mathrm{IQR} &= 66.25 - 1.5 \cdot 8.75 = 53.125, \\ q_n(0.75) + 1.5 \cdot \mathrm{IQR} &= 75 + 1.5 \cdot 8.75 = 88.125. \end{aligned}$$

Hence, the last element below 88.125 is 88, and the first element above 53.125 is 57. Therefore, the upper whisker runs until 88 and the lower whisker until 57, with two elements 53 and 31 below. This leads to the following boxplot:



16.3 c The values 53 and 31 are outliers. Value 31 is far away from the bulk of the data and appears to be an *extreme* outlier.

16.4 a Check that $\bar{y}_n = 700/99$ and $\bar{x}_n = 492/11$, so that

$$\frac{5}{9}(\bar{x}_n - 32) = \frac{5}{9}\left(\frac{492}{11} - 32\right) = \frac{700}{99}.$$

16.4 b Since $n = 11$, for both datasets the sample median is 6th element in order of magnitude. Check that $\mathrm{Med}(y_1, \ldots, y_n) = 50/9$ and $\mathrm{Med}(x_1, \ldots, x_n) = 42$, so that

$$\frac{5}{9}(\mathrm{Med}(x_1, \ldots, x_n) - 32) = \frac{5}{9}(42 - 32) = \frac{50}{9},$$

which is equal to $\mathrm{Med}(y_1, \ldots, y_n)$.

16.4 c For any real number a, we have

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + b = a\bar{x}_n + b.$$

For $a \geq 0$, the order of magnitude of $y_i = ax_i + b$ is the same as that of the x_i's, and therefore

$$\text{Med}(y_1, \ldots, y_n) = a\text{Med}(x_1, \ldots, x_n) + b.$$

When $a < 0$, the order of $y_i = ax_i + b$ is reverse to that of the x_i's, but the position of the middle number in order of magnitude (or average of the two middle numbers) remains the same, so that the above rule still holds.

16.5 a Check that the sample variance of the y_i's is $(s_C)^2 = 132550/99^2$, so that

$$\left(\frac{5}{9}\right)^2 (s_F)^2 = \left(\frac{5}{9}\right)^2 \frac{482}{11} = \frac{132550}{99^2},$$

which is equal to $(s_C)^2$, and hence $s_C = \frac{5}{9}s_F$.

16.5 b For the x_i's we have $\text{Med}(x_1, \ldots, x_n) = 42$. This leads to the following table

x_i	43	43	41	41	41	42	43	58	58	41	41
$\lvert x_i - 42\rvert$	1	1	1	1	1	0	1	16	16	1	1
$\lvert x_i - 42\rvert$ ordered	0	1	1	1	1	1	1	1	1	16	16

so that $\text{MAD}_F = \text{Med}|x_i - 42| = 1$. Similarly, we have $\text{Med}(y_1, \ldots, y_n) = \frac{50}{9}$, and

y_i	$\frac{55}{9}$	$\frac{55}{9}$	5	5	5	$\frac{50}{9}$	$\frac{55}{9}$	$\frac{130}{9}$	$\frac{130}{9}$	5	5
$\lvert y_i - \frac{50}{9}\rvert$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	0	$\frac{5}{9}$	$\frac{80}{9}$	$\frac{80}{9}$	$\frac{5}{9}$	$\frac{5}{9}$
$\lvert y_i - \frac{50}{9}\rvert$ ordered	0	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{5}{9}$	$\frac{80}{9}$	$\frac{80}{9}$

so that $\text{MAD}_C = \text{Med}|y_i - \frac{50}{9}| = \frac{5}{9}$. Therefore, $\text{MAD}_C = \frac{5}{9}\text{MAD}_F$.

16.5 c Since, $\bar{y}_n = a\bar{x}_n + b$, for any real number a we have for the sample variance of the y_i's

$$\begin{aligned} s_Y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n}((ax_i + b) - (a\bar{x}_n + b))^2 \\ &= a^2\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = a^2 s_X^2 \end{aligned},$$

so that $s_Y = |a|\, s_X$. Because $\text{Med}_Y = a\text{Med}_X + b$, it follows that

$$\begin{aligned} \text{MAD}_Y &= \text{Med}|y_i - \text{Med}_Y| = \text{Med}|(ax_i + b) - (a\text{Med}_X + b)| \\ &= |a|\text{Med}|x_i - \text{Med}_X| = |a|\text{MAD}_X. \end{aligned}$$

16.6 a Yes, we find $\bar{x} = (1+5+9)/3 = 15/3 = 5$, $\bar{y} = (2+4+6+8)/4 = 20/4 = 5$, so that $(\bar{x} + \bar{y})/2 = 5$. The average for the combined dataset is also equal to 5: $(15 + 20)/7 = 5$.

16.6 b The mean of $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m$ equals

$$\frac{x_1 + \cdots + x_n + y_1 + \cdots + y_m}{n+m} = \frac{n\bar{x}_n + m\bar{y}_m}{n+m} = \frac{n}{n+m}\,\bar{x}_n + \frac{m}{n+m}\,\bar{y}_m.$$

In general, this is not equal to $(\bar{x}_n+\bar{y}_m)/2$. For instance, replace 1 in the first dataset by 4. Then $\bar{x}_n = 6$ and $\bar{y}_m = 5$, so that $(\bar{x}_n + \bar{y}_m)/2 = 5\frac{1}{2}$. However, the average of the combined dataset is $38/7 = 5\frac{2}{7}$.

16.6 c Yes, $m = n$ implies $n/(n+m) = m/(n+m) = 1/2$. From the expressions found in part **b** we see that the sample mean of the combined dataset equals $(\bar{x}_n + \bar{y}_m)/2$.

16.7 a We have $\mathrm{Med}_x = 5$ and $\mathrm{Med}_y = (4+6)/2 = 5$, whereas for the combined dataset

$$1 \quad 2 \quad 4 \quad 5 \quad 6 \quad 8 \quad 9$$

the sample median is the fourth number in order of magnitude: 5.

16.7 b This will not be true in general: take $1, 2, 3$ and $5, 7$ with sample medians 2 and 6. The combined dataset has sample median 3, whereas $(2+6)/2 = 4$.

16.7 c This will not be true in general: take $1, 9$ and $2, 4$ with sample medians 5 and 3. The combined dataset has sample median 3, whereas $(5+3)/2 = 4$.

16.8 The ordered combined dataset is 1, 2, 4, 5, 6, 8, 9, so that the sample median equals 5. The absolute deviations from 5 are: 4, 3, 1, 0, 1, 3, 4, and if we put them in order: 0, 1, 1, 3, 3, 4, 4. The MAD is the sample median of the absolute deviations, which is 3.

16.9 a One can easily check that $\bar{x}_n = 10/3$. The average of the y_i's:

$$\frac{1}{3}\left(-\frac{1}{6} + 1 + \frac{1}{15}\right) = \frac{1}{3} \cdot \frac{54}{60} = \frac{3}{10},$$

so that $\bar{y}_7 = 1/\bar{x}_7$.

16.9 b No, for instance take 1, 2, and 3. Then $\bar{x}_n = 2$, and $\bar{y}_n = (1 + \frac{1}{2} + \frac{1}{3})/3 = 11/18$.

16.9 c First note that the transformation $y = 1/x$ puts the y_i's in reverse order compared to the x_i's. If n is odd, then the middle element in order of magnitude remains the middle element under the transformation $y = 1/x$. Therefore, if n is odd, the answer is yes. If n is even, the answer is no, for example take 1 and 2. Then $\mathrm{Med}_X = \frac{3}{2}$ and $\mathrm{Med}_Y = \mathrm{Med}(1, \frac{3}{2}) = (1 + \frac{3}{2})/2 = \frac{5}{4}$.

16.10 a The sum of the elements is $16.8 + y$ which goes to infinity as $y \to \infty$, and therefore the sample mean also goes to infinity, because we only divide the sum by the sample size n. When $y \to \infty$, the ordered data will be

$$3.0 \quad 4.2 \quad 4.6 \quad 5.0 \quad y$$

Therefore, no matter how large y gets, the sample median, which is the middle number, remains 4.6.

16.10 b In order to let the middle number go to infinity, we must replace at least three numbers. For instance, replace 3.2, 4.2, and 5.0 by some real number y that goes to infinity. In that case, the ordered data will be

$$3.0 \quad 4.6 \quad y \quad y \quad y$$

of which the middle number y goes to infinity.

16.10 c Without loss of generality suppose that the dataset is already ordered. By the same reasoning as in part **a**, one can argue that we only have to replace one element in order to let the sample mean go to infinity.
When $n = 2k+1$ is odd, the middle number is the $(k+1)$st number. By the same reasoning as in part **b** one can argue that in order to let that go to infinity, we must replace the middle number, as well as k other elements by y, which means we have to replace $k+1 = (n-1)/2 = \lfloor (n+1)/2 \rfloor$ elements.
When $n = 2k$ is even, the middle number is the average of the kth and $(k+1)$st number. In order to let that go to infinity, it suffices to replace k elements, that include either the kth or the $(k+1)$st number, by y, This means we have to replace $k = n/2 = \lfloor (n+1)/2 \rfloor$ elements.

16.11 a. From Exercise 16.10 we already know that the sample mean goes to infinity. This implies that $(4.6 - \bar{x}_n)^2$ also goes to infinity and therefore also the sample variance, as well as the sample standard deviation.
From Exercise 16.10 we know that the sample median remains 4.6. Hence, the ordered absolute deviations are

$$0 \quad 0.4 \quad 0.4 \quad 1.6 \quad |y - 4.6|$$

Therefore, no matter how large y gets, the MAD, which is the middle number of the ordered absolute deviations, remains 0.4.
b. In order to let the middle number the ordered absolute deviations go to infinity, we must at least replace three numbers. For instance, replace 3.2 and 4.2 by y and 5.0 by $-y$, where y is some real number that goes to infinity. In that case, the ordered data will be

$$-y \quad 3.0 \quad 4.6 \quad y \quad y$$

of which the middle number is 4.6. The ordered absolute deviations are

$$0 \quad 0.4 \quad |y - 4.6| \quad |y - 4.6| \quad |4.6 + y|$$

The MAD is the middle number, which goes to infinity.
c. Without loss of generality suppose that the dataset is already ordered. By the same reasoning as in part **a** we only need to replace one element in order to let the sample standard deviation go to infinity.
When $n = 2k+1$ is odd, the middle number is the $(k+1)$st number. By the same reasoning as in part **b** one can argue that in order to let the MAD go to infinity, we must replace the middle number by y, as well as k other numbers. This means we have to replace $k+1 = (n-1)/2 = \lfloor (n+1)/2 \rfloor$ elements, for instance one by $-y$ and the rest by y. In that case, the sample median remains finite and the majority of the absolute deviations from the median tends to infinity.
When $n = 2k$ is even, the middle number is the average of the kth and $(k+1)$st number. In order to let the MAD go to infinity, it suffices to replace k elements, that include either the kth or the $(k+1)$st number, by y. This means we have to replace $k = n/2 = \lfloor (n+1)/2 \rfloor$ elements, for instance one by $-y$ and the rest by y. In that case, the sample median remains finite and half of the absolute deviations from the median tends to infinity.

16.12 The sample mean is

$$\frac{1}{N}(1 + 2 + \cdots + N) = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}$$

When $N = 2k+1$ is odd, the sample median is the $(k+1)$st number, which is $k+1 = (N-1)/2+1 = (N+1)/2$. When $N = 2k$ is even, the sample median is the average of the kth and $(k+1)$st number:

$$\frac{(N/2 + (N/2+1))}{2} = \frac{N+1}{2}.$$

16.13 First note that the sample mean is zero, and $n = 2N+1$. Therefore the sample variance is

$$\begin{aligned} s_n^2 &= \frac{1}{n-1}\left((-N)^2 + \cdots + (-1)^2 + 0^2 + 1^2 + \cdots N^2\right) \\ &= \frac{1}{2N} \cdot 2 \cdot \left(1^2 + \cdots N^2\right) = \frac{1}{N} \cdot \frac{N(N+1)(2N+1)}{6} = \frac{(N+1)(2N+1)}{6}, \end{aligned}$$

so that $s_n = \sqrt{\frac{1}{6}(N+1)(2N+1)}$. Since $n = 2N+1$ is always odd, the sample median is the middle number, which is zero. Therefore we have the following

x_i	$-N$	$\cdots$	-1	0	1	$\cdots$	N	
$\lvert x_i\rvert$	N	$\cdots$	1	0	1	$\cdots$	N	
$\lvert x_i\rvert$ ordered	0	1	1	$\cdots$	$\cdots$	N	N	

Since $n = 2N+1$, which is odd, MAD = Med$|x_i|$ is the $(N+1)$st number in the bottom row: $(N+1)/2$.

16.14 When $n = 2i+1$ is odd, then in the formula $q_n(\frac{1}{2}) = x_{(k)} + \alpha[x_{(k+1)} - x_{(k)}]$ for the 50th empirical percentile, we find $k = \lfloor \frac{1}{2}(n+1) \rfloor = \lfloor i+1 \rfloor = i+1$ and $\alpha = 0$, so that $q_n(\frac{1}{2}) = x_{(i+1)}$, which is the middle number in order of magnitude.
When $n = 2i$ is even, we find $k = \lfloor \frac{1}{2}(n+1) \rfloor = \lfloor i + \frac{1}{2} \rfloor = i$ and $\alpha = \frac{1}{2}$, so that

$$q_n(\tfrac{1}{2}) = x_{(i)} + \tfrac{1}{2}\left(x_{(i+1)} - x_{(i)}\right) = \tfrac{x_{(i+1)} + x_{(i)}}{2},$$

which is the average of the two middle numbers in order of magnitude.

16.15 First write

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2\right) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}_n^2.$$

Next, by inserting

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}_n \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\bar{x}_n^2 = \frac{1}{n}\cdot n \cdot \bar{x}_n^2 = \bar{x}_n^2,$$

we find

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n^2 + \bar{x}_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}_n^2.$$

16.16 According to Exercise 15.12

$$\int_{-\infty}^{\infty} t f_{n,h}(t)\,\mathrm{d}t = \bar{x}_n$$

and

$$\int_{-\infty}^{\infty} t^2 f_{n,h}(t)\,\mathrm{d}t = \frac{1}{n}\sum_{i=1}^{n} x_i^2 + h^2 \int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u.$$

Therefore by the rule for the variance, $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$, the variance corresponding to $f_{n,h}$ is

$$\begin{aligned}
\int_{-\infty}^{\infty} t^2 f_{n,h}(t)\,\mathrm{d}t - \left(\int_{-\infty}^{\infty} t f_{n,h}(t)\,\mathrm{d}t\right)^2 &= \int_{-\infty}^{\infty} t^2 f_{n,h}(t)\,\mathrm{d}t - (\bar{x}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x}_n)^2 + h^2 \int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + h^2 \int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u,
\end{aligned}$$

according to Exercise 16.15.

16.17 For $p = i/(n+1)$ we find that

$$k = \lfloor p(n+1) \rfloor = i \quad \text{and} \quad \alpha = p(n+1) - k = 0,$$

so that

$$q_n(p) = x_{(k)} + \alpha\left(x_{(k+1)} - x_{(k)}\right) = x_{(i)}.$$

17.1 Before one starts looking at the figures, it is useful to recall a couple features of the two distributions involved:

1. most of the probability mass of the $N(0,1)$ distribution lies between ± 3, and between ± 9 for the $N(0,9)$ distribution.
2. the height at zero of the exponential density is λ and the median of the $Exp(\lambda)$ is $\ln(2)/\lambda$ (see Exercise 5.11).

We then have the following:

1. $N(3,1)$, since the mode is about 3.
2. $N(0,1)$, since F_n is 0.5 at about zero and all elements are between about ± 2.
3. $N(0,1)$, since the mode is about zero and all elements are between about -4 and 3.
4. $N(3,1)$, since the mode is about 3.
5. $Exp(1/3)$, since the sample median is about 2.
6. $Exp(1)$, since the sample median is less than 1.
7. $N(0,1)$, since the mode is about zero and all elements are between about ± 4.
8. $N(0,9)$, since the mode is about zero, and all elements are between about -6 and 9.
9. $Exp(1)$, since the height of the histogram at zero is about 0.7. Moreover, almost all elements are less than 6, whereas the probability of exceeding 6 for the $Exp(1/3)$ distribution is 0.135.
10. $N(3,1)$, since F_n is 0.5 at about 3.
11. $N(0,9)$, since the mode is about zero and all elements are between about ± 12.
12. $Exp(1/3)$, since the height of the histogram at zero is about 0.24. Moreover, there are several elements beyond 10, which has probability 0.000045 for the $Exp(1)$ distribution.
13. $N(0,9)$, since the mode is about zero and all elements are between about -5 and 8.

14. $Exp(1/3)$, since the height of the kernel density estimate at zero is about 0.35.
15. $Exp(1)$, since the sample median is about 1.

17.2 We continue as in Exercise 17.1:

1. $Exp(1/3)$, since the sample median is about 2.
2. $N(0,9)$, since the sample median is about zero and all elements are between about -6 and 8.
3. $Exp(1/3)$, since the sample median is about 2.
4. $N(0,1)$, since the sample median is about zero and all elements are between about -2 and 3.
5. $N(3,1)$, since the sample median is about 3.
6. $Exp(1)$, since the sample median is less than one.
7. $N(0,9)$, since the sample median is about zero and all elements are between about -9 and 7.
8. $N(0,9)$, since the sample median is about zero and all elements are between about -7 and 9.
9. $N(3,1)$, since the sample median is about 3.
10. $Exp(1)$, since the sample median is less than one.
11. $N(3,1)$, since the sample median is about 3.
12. $Exp(1)$, since the sample median is less than one.
13. $N(0,1)$, since the sample median is about zero and all elements are between about ± 3.
14. $N(0,1)$, since the sample median is about zero and all elements are between about -3 and 4.
15. $Exp(1/3)$, since the sample median is about 2.

17.3 a The model distribution corresponds to the number of women in a queue. A queue has 10 positions. The occurrence of a woman in any position is independent of the occurrence of a woman in other positions. At each position a woman occurs with probability p. Counting the occurrence of a woman as a "success," the number of women in a queue corresponds to the number of successes in 10 independent experiments with probability p of success and is therefore modeled by a $Bin(10,p)$ distribution.

17.3 b We have 100 queues and the number of women x_i in the ith queue is a realization of a $Bin(10,p)$ random variable. Hence, according to Table 17.2, the average number of women $\bar{x}_{100}$ resembles the expectation $10p$ of the $Bin(10,p)$ distribution. We find $\bar{x}_{100} = 435/100 = 4.35$, so an estimate for p is $4.35/10 = 0.435$.

17.4 a Recall that the parameter μ is the expectation of the $Pois(\mu)$ distribution. Hence, according to Table 17.2 the sample mean seems a reasonable estimate. Since, the dataset contains 229 zero's, 211 ones, etc.,

$$\bar{x}_n = \frac{211 \cdot 1 + 93 \cdot 2 + 35 \cdot 3 + 7 \cdot 4 + 1 \cdot 7}{576} = \frac{537}{576} = 0.9323.$$

17.4 b From part **a** the parameter $\mu = 0.9323$. Then $\mathrm{P}(X=0) = \mathrm{e}^{-0.9323} = 0.393$, and the other Poisson probabilities can be computed similarly. The following table compares the relative frequencies to the Poisson probabilities.

hits	0	1	2	3	4	5	6	7
rel.freq.	0.397	0.367	0.162	0.06	0.012	0.0000	0.0000	0.0017
prob.	0.393	0.367	0.171	0.053	0.012	0.0023	0.0004	0.0000

17.5 a One possibility is to use the fact that the geometric distribution has expectation $1/p$, so that the sample mean will be close to $1/p$. From the average number of cycles needed, 331/93, we then estimate p by $93/331 = 0.2809$.
Another possibility is to use that fact that p is equal to the probability of getting pregnant in the first cycle. We can estimate this by the relative frequency of women that got pregnant in the first cycle: $29/93 = 0.3118$.

17.5 b From the average number of cycles needed, 1285/474, we could estimate p by $474/1285 = 0.3689$. Another possibility is to estimate p by the relative frequency of women that got pregnant in the first cycle: $198/474 = 0.4177$.

17.5 c

	p	$p + p(1-p) + p(1-p)^2$
smokers	0.2809	0.6281
	0.3118	0.6741
nonsmokers	0.3689	0.7486
	0.4177	0.8026

17.6 a The parameters μ and σ are the expectation and standard deviation of the $N(\mu, \sigma^2)$ distribution. Hence, according to Table 17.2 we can estimate μ by the sample mean:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2283772}{5732} = 39.84,$$

and σ by the sample standard deviation s_n. To compute this, use Exercise 16.15 to get

$$\begin{aligned} s_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right] \\ &= \frac{n}{n-1}\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^2 - (\bar{x}_n)^2\right] = \frac{5732}{5731}\left[\frac{9124064}{5732} - (39.8)^2\right] = 4.35. \end{aligned}$$

Hence, take $s_n = \sqrt{4.35} = 2.09$ as an estimate for σ.

17.6 b Since we model the chest circumference by means of a random variable X with a $N(\mu, \sigma^2)$ distribution, the required probability is equal to

$$\begin{aligned} \mathrm{P}(38.5 < X < 42.5) &= \mathrm{P}(X < 42.5) - \mathrm{P}(X < 38.5) \\ &= \mathrm{P}\left(Z < \frac{42.5 - \mu}{\sigma}\right) - \mathrm{P}\left(Z < \frac{38.5 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{42.5 - \mu}{\sigma}\right) - \Phi\left(\frac{38.5 - \mu}{\sigma}\right) \end{aligned}$$

where $Z = (X - \mu)/\sigma$ has a standard normal distribution. We can estimate this by plugging in the estimates 39.84 and 2.09 for μ and σ. This gives

$$\Phi(1.27) - \Phi(-0.64) = 0.8980 - 0.2611 = 0.6369.$$

Another possibility is to estimate $\mathrm{P}(38.5 < X < 42.5)$ by the relative frequency of chest circumferences between 38.5 and 42.5. Using the information in Exercise 15.1 we find $3725/5732 = 0.6499$.

17.7 a If we model the series of disasters by a Poisson process, then as a property of the Poisson process, the interdisaster times should follow an exponential distribution (see Section 12.3). This is indeed confirmed by the histogram and empirical distribution of the observed interdisaster times; they resemble the probability density and distribution function of an exponential distribution.

17.7 b The average length of a time interval is $40\,549/190 = 213.4$ days. Following Table 17.2 this should resemble the expectation of the $Exp(\lambda)$ distribution, which is $1/\lambda$. Hence, as an estimate for λ we could take $190/40\,549 = 0.00469$.

17.8 a The distribution function of Y is given by

$$F_Y(a) = \mathrm{P}(Y \le a) = \mathrm{P}\left(X \le y^{1/\alpha}\right) = 1 - \mathrm{e}^{\lambda^\alpha y}.$$

This is the distribution function of the $Exp(\lambda^\alpha)$ distribution, which has expectation $1/\lambda^\alpha$. Therefore, $\mathrm{E}[X^\alpha] = \mathrm{E}[Y] = 1/\lambda^\alpha$.

17.8 b Take α-powers of the data: $y_i = x_i^\alpha$. Then, according to part **a**, the y_i's are a realization of a sample from an $Exp(\lambda^\alpha)$ distribution, with expectation $1/\lambda^\alpha$. Hence, $1/\lambda^\alpha$ can be estimated by the average of these numbers:

$$\frac{1}{\lambda^\alpha} \approx \frac{1}{n}\sum_{i=1}^{n} x_i^\alpha.$$

Next, solve for λ to get an estimate for λ:

$$\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^\alpha\right)^{-1/\alpha}.$$

When we plug in $\alpha = 2.102$ and apply this formula to the dataset, we get $\lambda = (10654.85)^{-1/2.102} = 0.0121$.

17.9 a A (perfect) cylindrical cone with diameter d (at the base) and height h has volume $\pi d^2 h/12$, or about $0.26 d^2 h$. The effective wood of a tree is the trunk without the branches. Since the trunk is similar to a cylindrical cone, one can expect a linear relation between the effective wood and $d^2 h$.

17.9 b We find

$$\begin{aligned}
\bar{z}_n &= \frac{\sum y_i/x_i}{n} = \frac{9.369}{31} = 0.3022 \\
\bar{y}/\bar{x} &= \frac{(\sum y_i)/n}{(\sum x_i)/n} = \frac{26.486/31}{87.456/31} = 0.3028 \\
\text{least squares} &= \frac{\sum x_i y_i}{\sum x_i^2} = \frac{95.498}{314.644} = 0.3035.
\end{aligned}$$

17.10 a

$$\begin{aligned}
G(y) &= \mathrm{P}(Y \le y) \\
&= \mathrm{P}(|X - m| \le y) \\
&= \mathrm{P}(m - y \le X \le m + y) \\
&= \mathrm{P}(X \le m + y) - \mathrm{P}(X \le m - y) \\
&= F(m + y) - F(m - y).
\end{aligned}$$

17.10 b If f is symmetric around its median, then

$$f(m-y) = f(m+y)$$

for all y. This implies that the area under f on $(-\infty, m-y]$ is equal to the area under f on $[m+y, \infty)$, i.e. $F(m-y) = 1-F(m+y)$. By **a** this means $G(y) = 2F(m+y)-1$. To derive $G^{\text{inv}}(\frac{1}{2})$ put $\frac{1}{2} = G(y)$. Then, since $m = F^{\text{inv}}(\frac{1}{2})$

$$\begin{aligned}\frac{1}{2} = 2F(m+y) - 1 \Leftrightarrow \frac{3}{4} &= F(m+y) \\ \Leftrightarrow y &= F^{\text{inv}}(\tfrac{3}{4}) - m = F^{\text{inv}}(\tfrac{3}{4}) - F^{\text{inv}}(\tfrac{1}{2}).\end{aligned}$$

17.10 c First determine an expression for $F^{\text{inv}}(x)$, by putting $x = F(u)$. Then

$$F^{\text{inv}}(x) = \mu + \sigma\Phi^{\text{inv}}(x).$$

With **b** it follows that the MAD of a $N(\mu, \sigma^2)$ is equal to

$$F^{\text{inv}}(\tfrac{3}{4}) - F^{\text{inv}}(\tfrac{1}{2}) = \sigma\Phi^{\text{inv}}(\tfrac{3}{4}) - \sigma\Phi^{\text{inv}}(\tfrac{1}{2}) = \sigma\Phi^{\text{inv}}(\tfrac{3}{4}),$$

using that $\Phi^{\text{inv}}(\frac{1}{2}) = 0$.
For the $N(5, 4)$ distribution the MAD is $2\Phi^{\text{inv}}(\frac{3}{4}) = 2 \cdot 0.6745 = 1.3490$.

17.11 a Using Exercise 17.10 **a** it follows that

$$\begin{aligned}G(y) &= F(m+y) - F(m-y) \\ &= \left(1 - \mathrm{e}^{-\lambda(m-y)}\right) - \left(1 - \mathrm{e}^{-\lambda(m+y)}\right) \\ &= \mathrm{e}^{-\lambda m}\left(\mathrm{e}^{\lambda y} - \mathrm{e}^{-\lambda y}\right) \\ &= \frac{1}{2}\left(\mathrm{e}^{\lambda y} - \mathrm{e}^{-\lambda y}\right),\end{aligned}$$

using that m is the median of the $Exp(\lambda)$ distribution, and satisfies $1 - \mathrm{e}^{-\lambda m} = \frac{1}{2}$.

17.11 b Combining **a** with Exercise 17.10 **b** the MAD of the $Exp(\lambda)$ distribution is a solution of $G(y) = \frac{1}{2}$, so that $\frac{1}{2}\left(\mathrm{e}^{\lambda y} - \mathrm{e}^{-\lambda y}\right) = \frac{1}{2}$. Multiplying this equation with $\mathrm{e}^{\lambda y}$, yields that the MAD must satisfy $\mathrm{e}^{2\lambda y} - \mathrm{e}^{\lambda y} - 1 = 0$.

17.11 c Put $x = \mathrm{e}^{\lambda y}$ and solve $x^2 - x - 1 = 0$ for x. This gives

$$x = \frac{1 \pm \sqrt{5}}{2}.$$

Since the MAD must be positive, it can be found from the relation

$$\mathrm{e}^{\lambda y} = \frac{1 + \sqrt{5}}{2}.$$

If follows that the MAD of the $Exp(\lambda)$ distribution is equal to

$$\frac{\ln(1+\sqrt{5}) - \ln 2}{\lambda} = \frac{\ln(1+\sqrt{5})}{\lambda} - F^{\text{inv}}(\tfrac{1}{2}).$$

18.1 If we view a bootstrap dataset as a vector $(x_1^*, \ldots, x_n^*)$, we have five possibilities 1, 2, 3, 4, and 6 for each of the six positions. Therefore there are 5^6 different vectors possible. They are not equally likely. For instance, $(1, 1, 1, 1, 1, 1)$ has probability $(\frac{1}{3})^6$, whereas $(2, 2, 2, 2, 2, 2)$ has probability $(\frac{1}{6})^6$ to occur.

18.2 a Because the bootstrap sample can only be 1 if and only if all elements in the bootstrap random sample are 1, we get:

$$\begin{aligned}\mathrm{P}\big(\bar{X}_n^* = 1\big) &= \mathrm{P}(X_1^* = 1, \dots, X_4^* = 1)\\ &= \mathrm{P}(X_1^* = 1) \cdots \mathrm{P}(X_4^* = 1) = \left(\frac{1}{4}\right)^4 = 0.0039.\end{aligned}$$

18.2 b Because the maximum is less than 4, is equivalent to all numbers being less than 4, we get

$$\begin{aligned}\mathrm{P}(\max X_i^* = 6) &= 1 - \mathrm{P}(\max X_i^* \le 4) = 1 - \mathrm{P}(X_1^* \le 4, \dots, X_4^* \le 4)\\ &= 1 - \mathrm{P}(X_1^* \le 4) \cdots \mathrm{P}(X_4^* \le 4) = 1 - \left(\frac{3}{4}\right)^4 = 0.6836\end{aligned}$$

18.3 a Note that generating from the empirical distribution function is the same as choosing one of the elements of the original dataset with equal probability. Hence, an element in the bootstrap dataset equals 0.35 with probability 0.1. The number of ways to have exactly three out of ten elements equal to 0.35 is $\binom{10}{3}$, and each has probability $(0.1)^3(0.9)^7$. Therefore, the probability that the bootstrap dataset has exactly three elements equal to 0.35 is equal to $\binom{10}{3}(0.1)^3(0.9)^7 = 0.0574$.

18.3 b Having at most two elements less than or equal to 0.38 means that 0, 1, or 2 elements are less than or equal to 0.38. Five elements of the original dataset are smaller than or equal to 0.38, so that an element in the bootstrap dataset is less than or equal to 0.38 with probability 0.5. Hence, the probability that the bootstrap dataset has at most two elements less than or equal to 0.38 is equal to $(0.5)^{10} + \binom{10}{1}(0.5)^{10} + \binom{10}{2}(0.5)^{10} = 0.0547$.

18.3 c Five elements of the dataset are smaller than or equal to 0.38 and two are greater than 0.42. Therefore, obtaining a bootstrap dataset with two elements less than or equal to 0.38, and the other elements greater than 0.42 has probability $(0.5)^2\,(0.2)^8$. The number of such bootstrap datasets is $\binom{10}{2}$. So the answer is $\binom{10}{2}\,(0.5)^2\,(0.2)^8 = 0.000029$.

18.4 a There are 9 elements strictly less than 0.46, so that $\mathrm{P}(M_{10}^* < 0.46) = (\frac{9}{10})^{10} = 0.3487$.

18.4 b There are $n-1$ elements strictly less than m_n, so that $\mathrm{P}(M_n^* < m_n) = (\frac{n-1}{n})^n = (1 - 1/n)^n$.

18.5 Since each X_i^* is either 0, 3, or 6, it is not so difficult to see that $X_1^*+X_2^*+X_3^*$ can only take the values 0, 3, 6, 9, 12, 15, and 18. The sample mean of the dataset is $\bar{x}_n = 3$. Therefore $\bar{X}_3^* - \bar{x}_n = (X_1^* + X_2^* + X_3^*)/3 - 3$ can only take the values -3, -2, -1, 0, 1, 2, and 3. The value $\bar{X}_3^* - \bar{x}_n = -3$ corresponds to $(X_1^*, X_2^*, X_3^*) = (0,0,0)$. Because there are $3^3 = 27$ possibilities for (X_1^*, X_2^*, X_3^*), the probability $\mathrm{P}\big(\bar{X}_3^* - \bar{x}_n = -3\big) = 1/27$. Similarly, the value $\bar{X}_3^* - \bar{x}_n = -2$ corresponds to (X_1^*, X_2^*, X_3^*) being equal to $(3,0,0)$, $(0,3,0)$, or $(0,0,3)$, so that $\mathrm{P}\big(\bar{X}_3^* - \bar{x}_n = -2\big) = 3/27$. The other probabilities can be computed in the same way, which leads to

a	-3	-2	-1	0	1	2	3
$\mathrm{P}\big(\bar{X}_n^* - \bar{x}_n = a\big)$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{7}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{1}{27}$

18.6 a We have

$$F_\lambda(m) = \tfrac{1}{2} \Leftrightarrow 1 - \mathrm{e}^{-\lambda m} = \tfrac{1}{2} \Leftrightarrow -\lambda m = \ln(\tfrac{1}{2}) = -\ln(2) \Leftrightarrow m = \frac{\ln 2}{\lambda}.$$

18.6 b Since we know that the dataset is a realization of a sample from an $Exp(\lambda)$ distribution, we are dealing with a parametric bootstrap. Therefore we must generate the bootstrap datasets from the distribution function of the $Exp(\lambda)$ distribution with the parameter λ estimated by $1/\bar{x}_n = 0.2$. Because the bootstrap simulation is for $\mathrm{Med}(X_1, X_2, \ldots, X_n) - m_\lambda$, in each iteration we must compute $\mathrm{Med}(x_1^*, x_2^*, \ldots, x_n^*) - m^*$, where m^* denotes the median of the estimated exponential distribution: $m^* = \ln(2)/(1/\bar{x}_n) = 5\ln(2)$. This leads to the following parametric bootstrap simulation procedure:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from $\hat{F}(x) = 1 - \mathrm{e}^{-0.2x}$.
2. Compute the centered sample median for the bootstrap dataset:

$$\mathrm{Med}(x_1^*, x_2^*, \ldots, x_n^*) - 5\ln(2),$$

   where $\mathrm{Med}(x_1^*, x_2^*, \ldots, x_n^*)$ is the sample median of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 many times.

18.7 For the parametric bootstrap, we must estimate the parameter θ by $\hat{\theta} = (n+1)m_n/n$, and generate bootstrap samples from the $U(0, \hat{\theta})$ distribution. This distribution has expectation $\mu_{\hat{\theta}} = \hat{\theta}/2 = (n+1)m_n/(2n)$. Hence, for each bootstrap sample $x_1^*, x_2^*, \ldots, x_n^*$ compute $\bar{x}_n^* - \mu_{\hat{\theta}} = \bar{x}_n^* - (n+1)m_n/(2n)$.
Note that this is different from the *empirical* bootstrap simulation, where one would estimate μ by $\bar{x}_n$ and compute $\bar{x}_n^* - \bar{x}_n$.

18.8 a Since we know nothing about the distribution of the interfailure times, we estimate F by the empirical distribution function F_n of the software data and we estimate the expectation μ of F by the expectation $\mu^* = \bar{x}_n = 656.8815$ of F_n. The bootstrapped centered sample mean is the random variable $\bar{X}_n^* - 656.8815$. The corresponding empirical bootstrap simulation is described as follows:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from F_n, i.e., draw with replacement 135 numbers from the software data.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 656.8815$$

   where $\bar{x}_n^*$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times.

18.8 b Because the interfailure times are now assumed to have an $Exp(\lambda)$ distribution, we must estimate λ by $\hat{\lambda} = 1/\bar{x}_n = 0.0015$ and estimate F by the distribution function of the $Exp(0.0015)$ distribution. Estimate the expectation $\mu = 1/\lambda$ of the $Exp(\lambda)$ distribution by $\mu^* = 1/\hat{\lambda} = \bar{x}_n = 656.8815$. Also now, the bootstrapped centered sample mean is the random variable $\bar{X}_n^* - 656.8815$. The corresponding parametric bootstrap simulation is described as follows:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the $Exp(0.0015)$ distribution.

2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 656.8815,$$

where $\bar{x}_n^*$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times. We see that in this simulation the bootstrapped centered sample mean is the *same* in both cases: $\bar{X}_n^* - \bar{x}_n$, but the corresponding simulation procedures differ in step 1.

18.8 c Estimate λ by $\hat{\lambda} = \ln 2/m_n = 0.0024$ and estimate F by the distribution function of the *Exp*(0.0024) distribution. Estimate the expectation $\mu = 1/\lambda$ of the *Exp*(λ) distribution by $\mu^* = 1/\hat{\lambda} = 418.3816$. The corresponding parametric bootstrap simulation is described as follows:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the *Exp*(0.0024) distribution.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 418.3816,$$

where $\bar{x}_n^*$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times. We see that in this parametric bootstrap simulation the bootstrapped centered sample mean is *different* from the one in the empirical bootstrap simulation: $\bar{X}_n^* - (\ln 2)/m_n$ instead of $\bar{X}_n^* - \bar{x}_n$.

18.9 Estimate μ by $\bar{x}_n = 39.85$, σ by $s_n = 2.09$, and estimate F by the distribution function of the $N(39.85, 4.37)$ distribution.

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the $N(39.85, 4.37)$ distribution.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 39.85$$

where $\bar{x}_n$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times. Estimate $\mathrm{P}\big(|\bar{X}_n - \mu| > 1\big)$ by the relative frequency of bootstrapped centered sample means that are greater than 1 in absolute value:

$$\frac{\text{number of } \bar{x}_n^* \text{ with } |\bar{x}_n^* - 39.85| \text{ greater than } 1}{1000}.$$

18.10 a Perform the empirical bootstrap simulation as in part **a** of Exercise 18.8.

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from F_n, i.e., draw with replacement 135 numbers from the software data.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 656.8815$$

where $\bar{x}_n^*$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times.
Estimate $\mathrm{P}\big(|\bar{X}_n - \mu| > 10\big)$ by the relative frequency of bootstrapped centered sample means that are greater than 1 in absolute value:

$$\frac{\text{number of } \bar{x}_n^* \text{ with } |\bar{x}_n^* - 656.8815| \text{ greater than } 10}{1000}.$$

18.10 b Perform the parametric bootstrap simulation as in part **b** of Exercise 18.8

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the $Exp(0.0015)$ distribution.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 656.8815,$$

where $\bar{x}_n^*$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times.
Estimate $\mathrm{P}\big(|\bar{X}_n - \mu| > 10\big)$ as in part **a**.

18.11 Estimate μ by $\bar{x}_n = 39.85$, σ by $s_n = 2.09$, and estimate F by the distribution function of the $N(39.85, 4.37)$ distribution.

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the $N(39.85, 4.37)$ distribution.
2. Compute the sample mean $\bar{x}_n^*$, sample standard deviation s_n^* and empirical distribution function F_n^* of $x_1^*, x_2^*, \ldots, x_n^*$. Use these to compute the bootstrapped KS distance

$$t_{\mathrm{ks}}^* = \sup_{a\in\mathbb{R}} |F_n^*(a) - F_{\bar{x}_n^*, s_n^*}(a)|$$

Repeat steps 1 and 2 a large number of times. To investigate to which degree the value 0.0987 agrees with the assumed normality of the dataset, one may compute the percentage of t_{ks}^* values that are greater then 0.0987. The closer this percentage is to one, the better the normal distribution fits the data.

18.12 a Note that $\mathrm{P}(T_n \leq 0) = \mathrm{P}(M_n \geq \theta) = 0$, because all X_i are less than θ with probability one. To compute $\mathrm{P}(T_n^* \leq 0)$, note that, since the bootstrap random sample is from the original dataset, we always have that $M_n^* \leq m_n$. Hence

$$\mathrm{P}(T_n^* \leq 0) = \mathrm{P}(M_n^* \geq m_n) = \mathrm{P}(M_n^* = m_n) = 1 - \mathrm{P}(M_n^* < m_n)\,.$$

Furthermore, from Exercise 18.4 we know that $\mathrm{P}(M_n^* < m_n) = \big(1 - \frac{1}{n}\big)^n$.

18.12 b Note that

$$G_n(0) = \mathrm{P}(T_n \leq 0) = \mathrm{P}(M_n \leq 0) = \mathrm{P}(X_1 \leq 0) \cdots \mathrm{P}(X_n \leq 0)\,,$$

which is zero because the X_i have a $U(0, \theta)$ distribution. With part **a** it follows that

$$\sup_{t\in\mathbb{R}} |G_n^*(t) - G_n(t)| \geq |G_n^*(0) - G_n(0)| = \mathrm{P}(T_n^* \leq 0) = 1 - \left(1 - \frac{1}{n}\right)^n.$$

18.12 c From the inequality $\mathrm{e}^{-x} \geq 1 - x$, it follows that

$$1 - \left(1 - \frac{1}{n}\right)^n \geq 1 - \left(\mathrm{e}^{-1/n}\right)^n = 1 - \mathrm{e}^{-1}.$$

18.13 a The density of a $U(0, \theta)$ distribution is given by $f_\theta(x) = 1/\theta$, for $0 \leq x \leq \theta$. Hence for $0 \leq a \leq \theta$,

$$F_\theta(a) = \int_0^a \frac{1}{\theta}\,\mathrm{d}x = \frac{a}{\theta}.$$

18.13 b We have

$$P(T_n \le t) = P(1 - M_n/\theta \le t) = P(M_n \ge \theta(1-t)) = 1 - P(M_n \le \theta(1-t)).$$

Using the rule on page 115 about the distribution function of the maximum, it follows that for $0 \le t \le 1$,

$$G_n(t) = 1 - P(M_n \le \theta(1-t)) = 1 - F_\theta(\theta(1-t))^n = 1 - (1-t)^n.$$

18.13 c By the same argument as before

$$G_n^*(t) = P(T_n^* \le t) = 1 - P\left(M_n^* \le \hat{\theta}(1-t)\right).$$

Since M_n^* is the maximum of a random sample $X_1^*, X_2^*, \ldots, X_n^*$ from a $U(0, \hat{\theta})$ distribution, again the rule on page 115 about the distribution function of the maximum yields that for $0 \le t \le 1$,

$$G_n^*(t) = 1 - P\left(M_n^* \le \hat{\theta}(1-t)\right) = 1 - F_{\hat{\theta}}(\hat{\theta}(1-t))^n = 1 - (1-t)^n.$$

19.1 a We must show that $E[T] = \theta^2$. From the formulas for the expectation and variance of uniform random variables we deduce that $E[X_i] = 0$ and $\mathrm{Var}(X_i) = (2\theta)^2/12 = \theta^2/3$. Hence $E\left[X_i^2\right] = \mathrm{Var}(X_i) + (E[X_i])^2 = \theta^2/3$. Therefore, by linearity of expectations

$$\begin{aligned} E[T] &= \frac{3}{n}(E\left[X_1^2\right] + E\left[X_2^2\right] + \cdots + E\left[X_n^2\right]) \\ &= \frac{3}{n}\left(\frac{\theta^2}{3} + \cdots + \frac{\theta^2}{3}\right) = \frac{3}{n} \cdot n \cdot \frac{\theta^2}{3} = \theta^2. \end{aligned}$$

Since $E[T] = \theta^2$, the random variable T is an unbiased estimator for θ^2.

19.1 b The function $g(x) = -\sqrt{x}$ is a strictly convex function, because $g''(x) = (x^{-3/2})/4 > 0$. Therefore, by Jensen's inequality, $-\sqrt{E[T]} < -E\left[\sqrt{T}\right]$. Since, from part **a** we know that $E[T] = \theta^2$, this means that $E\left[\sqrt{T}\right] < \theta$. In other words, $\sqrt{T}$ is a biased estimator for θ, with negative bias.

19.2 a We must check whether $E[S] = \mu$. By linearity of expectations, we have

$$\begin{aligned} E[S] &= E\left[\frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3\right] = \frac{1}{2}E[X_1] + \frac{1}{3}E[X_2] + \frac{1}{6}E[X_3] \\ &= \frac{1}{2}\mu + \frac{1}{3}\mu + \frac{1}{6}\mu = \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{6}\right)\mu = \mu. \end{aligned}$$

So indeed, S is an unbiased estimator for μ.

19.2 b We must check under what conditions $E[T] = \mu$ for all μ. By linearity of expectations, we have

$$\begin{aligned} E[T] &= E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n] \\ &= a_1\mu + a_2\mu + \cdots + a_n\mu = (a_1 + a_2 + \cdots + a_n)\,\mu. \end{aligned}$$

This is equal to μ if and only if $a_1 + a_2 + \cdots + a_n = 1$. Therefore, T is an unbiased estimator for μ if and only if $a_1 + a_2 + \cdots + a_n = 1$.

19.3 We must check under what conditions $\mathrm{E}[T] = \mu$ for all μ. By linearity of expectations, we have

$$\begin{aligned}\mathrm{E}[T] &= \mathrm{E}[a(X_1 + X_2 + \cdots + X_n) + b] = a\left(\mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]\right) + b \\ &= a\left(\mu + \mu + \cdots + \mu\right) + b = an\mu + b.\end{aligned}$$

For T to be an unbiased estimator for μ, one must have $an\mu + b = \mu$ for *all* μ, with a and b being constants *not depending* on μ. This can only be the case, when $b = 0$ and $a = 1/n$.

19.4 a The function $g(x) = 1/x$ is convex, so that according to Jensen's inequality:

$$\mathrm{E}[T] = \mathrm{E}\left[\frac{1}{\bar{X}_n}\right] > \frac{1}{\mathrm{E}\left[\bar{X}_n\right]}.$$

Furthermore, $\mathrm{E}\left[\bar{X}_n\right] = \mathrm{E}[X_1] = 1/p$, so that $\mathrm{E}[T] > 1/(1/p) = p$. We conclude that T is a biased estimator for p with positive bias.

19.4 b For each i label '$X_i \leq 3$' as a success, then the total number Y of all $X_i \leq 3$ has a $Bin(n, \theta)$ distribution where $\theta = p + (1-p)p + (1-p)^2 p$ represents the probability that a woman becomes pregnant within three or fewer cycles. This implies that $\mathrm{E}[Y] = n\theta$ and therefore $\mathrm{E}[S] = \mathrm{E}[Y/n] = \mathrm{E}[Y]/n = \theta$. This means that S is an unbiased estimator for θ.

19.5 We must find out for which c, $\mathrm{E}[T] = \mu$, where $\mu = 1/\lambda$ is the expectation of the $Exp(\lambda)$ distribution. Because M_n has an $Exp(n\lambda)$ distribution with expectation $1/(n\lambda)$, we have

$$\mathrm{E}[T] = \mathrm{E}[cM_n] = c\mathrm{E}[M_n] = c \cdot \frac{1}{n\lambda}.$$

Hence, with the choice $c = n$, we get $\mathrm{E}[T] = 1/\lambda = \mu$, which means that T is an unbiased estimator for μ.

19.6 a We must show that $\mathrm{E}[T] = 1/\lambda$. By linearity of expectations

$$\begin{aligned}\mathrm{E}[T] &= \frac{n}{n-1}\left(\mathrm{E}\left[\bar{X}_n\right] - \mathrm{E}[M_n]\right) = \frac{n}{n-1}\left[\left(\delta + \frac{1}{\lambda}\right) - \left(\delta + \frac{1}{n\lambda}\right)\right] \\ &= \frac{n}{n-1}\left[\frac{1}{\lambda} - \frac{1}{n\lambda}\right] = \frac{n}{n-1}\left(1 - \frac{1}{n}\right) \cdot \frac{1}{\lambda} = \frac{1}{\lambda}.\end{aligned}$$

19.6 b Find a linear combination of $\bar{X}_n$ and M_n of which the expectation is δ. From the expressions for $\mathrm{E}\left[\bar{X}_n\right]$ and $\mathrm{E}[M_n]$ we see that we can eliminate λ by subtracting $\mathrm{E}\left[\bar{X}_n\right]$ from $n\mathrm{E}[M_n]$. Therefore, first consider $nM_n - \bar{X}_n$, which has expectation

$$\mathrm{E}\left[nM_n - \bar{X}_n\right] = n\mathrm{E}[M_n] - \mathrm{E}\left[\bar{X}_n\right] = n\left(\delta + \frac{1}{n\lambda}\right) - \left(\delta + \frac{1}{\lambda}\right) = (n-1)\delta.$$

This means that

$$T = \frac{nM_n - \bar{X}_n}{n-1}$$

has expectation δ: $\mathrm{E}[T] = \mathrm{E}\left[nM_n - \bar{X}_n\right]/(n-1) = \delta$, so that T is an unbiased estimator for δ.

19.6 c Plug in $\bar{x}_n = 8563.5$, $m_n = 2398$ and $n = 20$ in the estimator of part **b**:

$$t = \frac{20 \cdot 2398 - 8563.5}{19} = 2073.5.$$

19.7 a Note that by linearity of expectations

$$\mathrm{E}[T_1] = \frac{4}{n}\mathrm{E}[N_1] - 2.$$

Because N_1 has a $Bin(n, p_1)$ distribution, with $p_1 = (\theta+2)/4$, it follows that $\mathrm{E}[N_1] = np_1 = n(\theta+2)/4$, so that $\mathrm{E}[T_1] = \theta$. The argument for T_2 is similar.

19.7 b Plug in $n = 3839$, $n_1 = 1997$, and $n_2 = 32$ in the estimators T_1 and T_2:

$$\begin{aligned} t_1 &= \frac{4}{3839} \cdot 1997 - 2 = 0.0808, \\ t_2 &= \frac{4}{3839} \cdot 32 = 0.0333. \end{aligned}$$

19.8 From the model assumptions it follows that $\mathrm{E}[Y_i] = \beta x_i$ for each i. Using linearity of expectations, this implies that

$$\begin{aligned} \mathrm{E}[B_1] &= \frac{1}{n}\left(\frac{\mathrm{E}[Y_1]}{x_1} + \cdots + \frac{\mathrm{E}[Y_n]}{x_n}\right) = \frac{1}{n}\left(\frac{\beta x_1}{x_1} + \cdots + \frac{\beta x_n}{x_n}\right) = \beta, \\ \mathrm{E}[B_2] &= \frac{\mathrm{E}[Y_1] + \cdots + \mathrm{E}[Y_n]}{x_1 + \cdots + x_n} = \frac{\beta x_1 + \cdots + \beta x_n}{x_1 + \cdots + x_n} = \beta, \\ \mathrm{E}[B_3] &= \frac{x_1\mathrm{E}[Y_1] + \cdots + x_n\mathrm{E}[Y_n]}{x_1^2 + \cdots + x_n^2} = \frac{\beta x_1^2 + \cdots + \beta x_n^2}{x_1^2 + \cdots + x_n^2} = \beta. \end{aligned}$$

19.9 Write $T = \mathrm{e}^{-Z/n}$, where $Z = X_1 + X_2 + \cdots + X_n$ has a $Pois(n\mu)$ distribution, with probabilities $\mathrm{P}(Z = k) = \mathrm{e}^{-n\mu}(n\mu)^k/k!$. Therefore

$$\begin{aligned} \mathrm{E}[T] = \mathrm{E}\left[\mathrm{e}^{-Z/n}\right] &= \sum_{k=0}^{\infty} \mathrm{e}^{-k/n} \cdot \mathrm{P}(Z = k) \\ &= \mathrm{e}^{-n\mu} \sum_{k=0}^{\infty} \mathrm{e}^{-k/n} \cdot \frac{(n\mu)^k}{k!} = \mathrm{e}^{-n\mu} \sum_{k=0}^{\infty} \frac{\left(n\mu\mathrm{e}^{-1/n}\right)^k}{k!} \end{aligned}$$

Use that $\sum_{k=0}^{\infty} x^k/k! = \mathrm{e}^x$, with $x = n\mu\mathrm{e}^{-1/n}$, and conclude that

$$\mathrm{E}[T] = \mathrm{e}^{-n\mu} \cdot \mathrm{e}^{n\mu\mathrm{e}^{-1/n}} = \mathrm{e}^{-n\mu(1-\mathrm{e}^{-1/n})}.$$

20.1 We have $\mathrm{Var}(\bar{X}_n) = \mathrm{MSE}(X_n) = \sigma^2/n$, which is decreasing in n, so the larger the sample, the more efficient $\bar{X}_n$ is. In particular $\mathrm{Var}(\bar{X}_n)/\mathrm{Var}(\bar{X}_{2n}) = (\sigma^2/n)/(\sigma^2/2n) = 2$ shows that $\bar{X}_{2n}$ is twice as efficient as $\bar{X}_n$.

20.2 a Compute the mean squared errors of S and T: $\mathrm{MSE}(S) = \mathrm{Var}(S) + [\mathrm{bias}(S)]^2 = 40 + 0 = 40$; $\mathrm{MSE}(T) = \mathrm{Var}(T) + [\mathrm{bias}(T)]^2 = 4 + 9 = 13$. We prefer T, because it has a smaller MSE.

20.2 b Compute the mean squared errors of S and T: $\mathrm{MSE}(S) = 40$, as in **a**; $\mathrm{MSE}(T) = \mathrm{Var}(T) + [\mathrm{bias}(T)]^2 = 4 + a^2$. So, if $a < 6$: prefer T. If $a \geq 6$: prefer S. The preferences are based on the MSE criterion.

20.3 $\mathrm{Var}(T_1) = 1/(n\lambda^2), \mathrm{Var}(T_2) = 1/\lambda^2$; hence we prefer T_1, because of its smaller variance.

20.4 a Since L has the same distribution as $N+1-M$, we find

$$\mathrm{E}[T_3] = \mathrm{E}[3L-1] = \mathrm{E}[3N+3-3M-1] = 3N+2-\mathrm{E}[2T_2+2] = 3N-2N = N.$$

Here we use that $\mathrm{E}[2T_2] = 2N$, since T_2 is unbiased.

20.4 b We have

$$\mathrm{Var}(T_3) = \mathrm{Var}(3L-1) = 9\mathrm{Var}(L) = 9\mathrm{Var}(M) = \frac{1}{2}(N+1)(N-2).$$

20.4 c We compute:

$$\mathrm{Var}(T_3)/\mathrm{Var}(T_2) = 9\mathrm{Var}(M)/\mathrm{Var}\big(\tfrac{3}{2}M-1\big) = 9\mathrm{Var}(M)/\tfrac{9}{4}\mathrm{Var}(M) = 4.$$

So using the maximum is 4 times as efficient as using the minimum!

20.5 From the variance of the sum rule: $\mathrm{Var}((U+V)/2) = \frac{1}{4}(\mathrm{Var}(U) + \mathrm{Var}(V) + 2\mathrm{Cov}(U,V))$. Using that $\mathrm{Var}(V)$=$\mathrm{Var}(U)$, we get from this that that the relative efficiency of U with respect to W is equal to

$$\frac{\mathrm{Var}(W)}{\mathrm{Var}(U)} = \frac{\mathrm{Var}((U+V)/2)}{\mathrm{Var}(U)} = \frac{1}{4}\Big(1+1+2\frac{\mathrm{Cov}(U,V)}{\sqrt{\mathrm{Var}(V)}\sqrt{\mathrm{Var}(U)}}\Big) = \frac{1}{2} + \frac{1}{2}\rho(U,V)\,.$$

Since the correlation coefficient is always between -1 and 1, it follows that the relative efficiency of U with respect to W is always between 0 and 1. Hence it ia always better (in the MSE sense) to use W.

20.6 a By linearity of expectations, $\mathrm{E}[U_1] = \mathrm{E}[T_1] + \frac{1}{3}(\pi - \mathrm{E}[T_1] - \mathrm{E}[T_2] - \mathrm{E}[T_3])$, which equals $\alpha_1 + \frac{1}{3}(\pi - \alpha_1 - \alpha_2 - \alpha_3) = \alpha_1$.
To compute the variances, rewrite U_1 as $U_1 = \frac{2}{3}T_1 - \frac{1}{3}T_2 - \frac{1}{3}T_3 + \frac{1}{3}\pi$. Then, by independence, $\mathrm{Var}(U_1) = \frac{4}{9}\mathrm{Var}(T_1) + \frac{1}{9}\mathrm{Var}(T_2) + \frac{1}{9}\mathrm{Var}(T_3) = \frac{2}{3}\sigma^2$.

20.6 b We have $\mathrm{Var}(T_1)/\mathrm{Var}(U_1) = 3/2$, so U_1 is 50% more efficient than T_1.

20.6 c There are at least two ways to obtain an efficient estimator for $\alpha_1 = \alpha_2$. The first is via the insight that T_1 and T_2 both estimate α_1, so $(T_1+T_2)/2$ is an efficient estimator for α_1 (c.f. Exercise 20.1). Then we can improve the efficiency in the same way as in part **a**. This yields the estimator

$$V_1 = \tfrac{1}{2}(T_1+T_2) + \tfrac{1}{3}(\pi - T_1 - T_2 - T_3) = \tfrac{1}{6}T_1 + \tfrac{1}{6}T_2 - \tfrac{1}{3}T_3 + \tfrac{1}{3}\pi.$$

The second way is to find the linear estimator $V_1 = uT_1 + vT_2 + wT_3 + t$ with the smallest MSE, optimising over u, v, w and t. This will result in the same estimator as obtained with the first way.

20.7 We compare the MSE's, which by unbiasedness are equal to the variances. Both N_1 and N_2 are binomally distributed, so $\mathrm{Var}(N_1) = np_1(1-p_1) = n(\theta+2)(1-(\theta+2)/4)/4$, and $\mathrm{Var}(N_2) = np_2(1-p_2) = n\theta(1-\theta)/4$. It follows that

$$\mathrm{Var}(T_1) = \frac{16}{n^2}\mathrm{Var}(N_1) = \frac{1}{n}(4-\theta^2); \mathrm{Var}(T_2) = \frac{16}{n^2}\mathrm{Var}(N_2) = \frac{1}{n}\theta(4-\theta).$$

Since $(4-\theta^2) > \theta(4-\theta)$ for all θ, we prefer T_2.

20.8 a This follows directly from linearity of expectations:

$$\mathrm{E}[T] = \mathrm{E}\left[r\bar{X}_n + (1-r)\bar{Y}_m\right] = r\mathrm{E}\left[\bar{X}_n\right] + (1-r)\mathrm{E}\left[\bar{Y}_m\right] = r\mu + (1-r)\mu = \mu.$$

20.8 b Using that $\bar{X}_n$ and $\bar{Y}_m$ are independent, we find $\mathrm{MSE}(T) = \mathrm{Var}(T) = r^2\mathrm{Var}(\bar{X}_n) + (1-r)^2\mathrm{Var}(\bar{Y}_m) = r^2 \cdot \sigma^2/n + (1-r)^2 \cdot \sigma^2/m$.
To find the minimum of this parabola we differentiate with respect to r and equate the result to 0: $2r/n - 2(1-r)/m = 0$. This gives the minimum value: $2rm - 2n(1-r) = 0$ or $r = n/(n+m)$.

20.9 a Since $\mathrm{E}[T_1] = p$, T_1 is unbiased. The estimator T_2 takes only the values 0 and 1, the latter with probability p^n. So $\mathrm{E}[T_2] = p^n$, and T_2 is biased for all $n > 1$.

20.9 b Since T_1 is unbiased, $\mathrm{MSE}(T_1) = \mathrm{Var}(T_1) = np(1-p)/n^2 = \frac{1}{n}p(1-p)$. Now for T_2: this random variable has a $Ber(p^n)$ distribution, hence
$\mathrm{MSE}(T_2) = \mathrm{E}\left[(T_2 - \theta)^2\right] = \mathrm{E}\left[(T_2 - p)^2\right] = p^2 \cdot \mathrm{P}(T_2 = 0) + (1-p)^2 \cdot \mathrm{P}(T_2 = 1) = p^2(1-p^n) + (1-p)^2 p^n = p^n - 2p^{n+1} + p^2$.

20.9 c For $n = 2$: $\mathrm{MSE}(T_2) = 4p\,\mathrm{MSE}(T_1)$, so for $p < \frac{1}{4}$ T_2 is more efficient than T_1, but otherwise T_1 is more efficient.

20.10 a Recall that the variance of an $Exp(\lambda)$ distribution equals $1/\lambda^2$, hence the mean squared error of T equals

$$\begin{aligned}\mathrm{MSE}(T) &= \mathrm{Var}(T) + \left(\mathrm{E}[T] - \lambda^{-1}\right)^2 \\ &= c^2 \cdot n \cdot \lambda^{-2} + \left(c \cdot n \cdot \lambda^{-1} - \lambda^{-1}\right)^2 \\ &= \lambda^{-2}\left[c^2 \cdot n + (c \cdot n - 1)^2\right] \\ &= \lambda^{-2}\left[c^2 \cdot (n^2 + n) - 2cn + 1\right].\end{aligned}$$

20.10 b This is a parabola in c, taking its smallest value when $2c(n^2+n) = 2n$, which happens for $c = 1/(n+1)$. So the estimator with the smallest MSE is

$$U = \frac{1}{n+1} \cdot (X_1 + X_2 + \cdots + X_n).$$

Substituting $c = 1/n$, and $c = 1/(n+1)$ in the formula for the MSE, we find that

$$\mathrm{MSE}(\bar{X}_n) = \frac{1}{n}, \quad \text{and} \quad \mathrm{MSE}(U) = \frac{1}{n+1}.$$

So U performs better in terms of MSE, but not much.

20.11 We have $\mathrm{MSE}(T_1) = \mathrm{Var}(T_1)$, which equals

$$\mathrm{Var}(T_1) = \mathrm{Var}\left(\sum_{i=1}^n x_i Y_i \Big/ \sum_{i=1}^n x_i^2\right) = \sum_{i=1}^n x_i^2 \mathrm{Var}(Y_i) \Big/ \left(\sum_{i=1}^n x_i^2\right)^2 = \sigma^2 \Big/ \left(\sum_{i=1}^n x_i^2\right).$$

Also, $\mathrm{MSE}(T_2) = \mathrm{Var}(T_2)$, which equals

$$\mathrm{Var}(T_2) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i}\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left(\frac{Y_i}{x_i}\right) = \frac{\sigma^2}{n^2}\sum_{i=1}^n \left(1/x_i^2\right).$$

Finally, $\mathrm{MSE}(T_3) = \mathrm{Var}(T_3)$, which equals

$$\mathrm{Var}(T_3) = \mathrm{Var}\left(\sum_{i=1}^{n} Y_i \Big/ \sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \mathrm{Var}(Y_i) \Big/ \left(\sum_{i=1}^{n} x_i\right)^2 = n\sigma^2 \Big/ \left(\sum_{i=1}^{n} x_i\right)^2.$$

To show that $\mathrm{Var}(T_3) < \mathrm{Var}(T_2)$: introduce a random variable X by $\mathrm{P}(X = x_i) = p_i$ for $i = 1 \dots n$, and apply Jensen with the function $g(x) = 1/x^2$, which is strictly convex on $(0, \infty)$ (note that all x_i in the cherry tree example are positive).

20.12 a Following the hint:

$$\begin{aligned}\mathrm{P}(M_n \le k) &= \mathrm{P}(X_1 \le k, X_2 \le k, \dots, X_n \le k) \\ &= \frac{k}{N} \cdot \frac{k-1}{N-1}, \cdots \cdot \frac{k-n+1}{N-n+1} \\ &= \frac{k!}{(k-n)!} \frac{(N-n)!}{N!}.\end{aligned}$$

20.12 b To have $M_n = n$ we should have drawn all numbers $1, 2 \dots, n$, and conversely, so

$$\mathrm{P}(M_n = n) = \mathrm{P}(M_n \le n) = \frac{n!(N-n)!}{N!}.$$

20.12 c This also follows directly from part **a**:

$$\begin{aligned}\mathrm{P}(M_n = k) &= \mathrm{P}(M_n \le k) - \mathrm{P}(M_n \le k-1) \\ &= \frac{k!}{(k-n)!} \frac{(N-n)!}{N!} - \frac{(k-1)!}{(k-1-n)!} \frac{(N-n)!}{N!} \\ &= \frac{(k-1)![k-(k-n)]}{(k-1-n)!} \frac{(N-n)!}{N!} \\ &= n \cdot \frac{(k-1)!}{(k-n)!} \frac{(N-n)!}{N!}.\end{aligned}$$

21.1 Setting $X_i = j$ if *red* appears in the ith experiment for the first time on the jth throw, we have that X_1, X_2, and X_3 are independent $Geo(p)$ distributed random variables, where p is the probability that *red* appears when throwing the selected die. The likelihood function is

$$\begin{aligned}L(p) &= \mathrm{P}(X_1 = 3, X_2 = 5, X_3 = 4) = (1-p)^2 p \cdot (1-p)^4 p \cdot (1-p)^3 p \\ &= p^3 (1-p)^9,\end{aligned}$$

so for D_1 one has that $L(p) = L(\frac{5}{6}) = \left(\frac{5}{6}\right)^3 \left(1 - \frac{5}{6}\right)^9$, whereas for D_2 one has that $L(p) = L(\frac{1}{6}) = \left(\frac{1}{6}\right)^3 \left(1 - \frac{1}{6}\right)^9 = 5^6 \cdot L(\frac{5}{6})$. It is very likely that we picked D_2.

21.2 As in the solution of Exercise 21.1, the likelihood is given by $L(p) = p^3(1-p)^9$, where p is between 0 and 1. So the loglikelihood is given by $\ell(p) = 3 \ln p + 9 \ln(1-p)$, and differentiating the loglikelihood with respect to p gives:

$$\frac{\mathrm{d}}{\mathrm{d}p}\big(\ell(p)\big) = \frac{3}{p} - \frac{9}{1-p}.$$

We find that $\frac{\mathrm{d}}{\mathrm{d}p}\big(\ell(p)\big) = 0$ if and only if $p = 1/4$, and since $\ell(p)$ (and also $L(p)$!) attains its maximum for this value of p, we find that the maximum likelihood estimate of p is $\hat{p} = 1/4$.

21.3 a The likelihood $L(\mu)$ is given by

$$\begin{aligned} L(\mu) &= C\mathrm{P}(X=0)^{229}\,\mathrm{P}(X=1)^{211}\,\mathrm{P}(X=2)^{93}\,\mathrm{P}(X=3)^{35}\,\mathrm{P}(X=4)^{7}\,\mathrm{P}(X=7) \\ &= \tilde{C}\mu^{537}\mathrm{e}^{-576\mu}. \end{aligned}$$

But then the loglikelihood $\ell(\mu)$ satisfies

$$\ell(\mu) = \ln\tilde{C} + 537\ln\mu - 576\mu.$$

But then we have, that $\ell'(\mu) = 537/\mu - 576$, from which it follows that $\ell'(\mu) = 0$ if and only if $\mu = 537/576 = 0.93229$.

21.3 b Now the likelihood $L(\mu)$ is given by

$$\begin{aligned} L(\mu) &= C\mathrm{P}(X=0,1)^{440}\,\mathrm{P}(X=2)^{93}\,\mathrm{P}(X=3)^{35}\,\mathrm{P}(X=4)^{7}\,\mathrm{P}(X=7) \\ &= \tilde{C}(1+\mu)^{440}\mu^{316}\mathrm{e}^{-576\mu}. \end{aligned}$$

But then the loglikelihood $\ell(\mu)$ satisfies

$$\ell(\mu) = \ln\tilde{C} + 440\ln(1+\mu) + 316\ln\mu - 576\mu.$$

But then we have, that

$$\ell'(\mu) = \frac{440}{1+\mu} + 316/\mu - 576,$$

from which it follows that $\ell'(\mu) = 0$ if and only if $576\mu^2 - 180\mu + 260 = 0$. We find that $\mu = 0.9351086$.

21.4 a The likelihood $L(\mu)$ is given by

$$\begin{aligned} L(\mu) &= \mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathrm{P}(X_1 = x_1)\cdots\mathrm{P}(X_n = x_n) \\ &= \frac{\mu^{x_1}}{x_1!}\cdot\mathrm{e}^{-\mu}\cdots\frac{\mu^{x_n}}{x_n!}\cdot\mathrm{e}^{-\mu} = \frac{\mathrm{e}^{-n\mu}}{x_1!\cdots x_n!}\mu^{x_1+x_2+\cdots+x_n}. \end{aligned}$$

21.4 b We find that the loglikelihood $\ell(\mu)$ is given by

$$\ell(\mu) = \left(\sum_{i=1}^{n} x_i\right)\ln(\mu) - \ln\left(x_1!\cdots x_n!\right) - n\mu.$$

Hence

$$\frac{\mathrm{d}\ell}{\mathrm{d}\mu} = \frac{\sum x_i}{\mu} - n,$$

and we find—after checking that we indeed have a maximum!—that $\bar{x}_n$ is the maximum likelihood estimate for μ.

21.4 c In **b** we have seen that $\bar{x}_n$ is the maximum likelihood estimate for μ. Due to the invariance principle from Section 21.4 we thus find that $\mathrm{e}^{-\bar{x}_n}$ is the maximum likelihood estimate for $\mathrm{e}^{-\mu}$.

21.5 a By definition, the likelihood $L(\mu)$ is given by

$$\begin{aligned} L(\mu) &= f_\mu(x_1)\cdots f_\mu(x_n) \\ &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x_1-\mu)^2}\cdots\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x_n-\mu)^2} \\ &= (2\pi)^{-n/2}\cdot\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}. \end{aligned}$$

But then the loglikelihood $\ell(\mu)$ satisfies

$$\ell(\mu) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Differentiating $\ell(\mu)$ with respect to μ yields

$$\ell'(\mu) = \sum_{i=1}^{n}(x_i - \mu) = -n\mu + \sum_{i=1}^{n} x_i.$$

So $\ell'(\mu) = 0$ if and only if $\mu = \bar{x}_n$. Since $\ell(\mu)$ attains a maximum at this value of μ (check this!), we find that $\bar{x}_n$ is the maximum likelihood estimate for μ.

21.5 b Again by the definition of the likelihood, we find that $L(\sigma)$ is given by

$$\begin{aligned} L(\sigma) &= f_\sigma(x_1)\cdots f_\sigma(x_n) \\ &= \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x_1^2/\sigma^2}\cdots\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x_n^2/\sigma^2} \\ &= \sigma^{-n}(2\pi)^{-n/2}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2}. \end{aligned}$$

But then the loglikelihood $\ell(\sigma)$ satisfies

$$\ell(\sigma) = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2.$$

Differentiating $\ell(\sigma)$ with respect to σ yields

$$\ell'(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}x_i^2.$$

So $\ell'(\sigma) = 0$ if and only if $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}$. Since $\ell(\sigma)$ attains a maximum at this value of σ (check this!), we find that $\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}$ is the maximum likelihood estimate for σ.

21.6 a In this case the likelihood function $L(\delta) = 0$ for $\delta > x_{(1)}$, and

$$L(\delta) = \mathrm{e}^{n\delta - \sum x_i} \quad \text{for } \delta \leq x_{(1)}.$$

21.6 b It follows from the graph of $L(\delta)$ in **a** that $x_{(1)}$ is the maximum likelihood estimate for δ.

21.7 By definition, the likelihood $L(\theta)$ is given by

$$\begin{aligned} L(\theta) &= f_\theta(x_1)\cdots f_\theta(x_n) \\ &= \frac{x_1}{\theta^2}\mathrm{e}^{-\frac{1}{2}x_1^2/\theta^2}\cdots\frac{x_n}{\theta^2}\mathrm{e}^{-\frac{1}{2}x_n^2/\theta^2} \\ &= \theta^{-2n}\left(\prod_{i=1}^{n}x_i\right)\mathrm{e}^{-\frac{1}{2\theta^2}\sum_{i=1}^{n}x_i^2}. \end{aligned}$$

But then the loglikelihood $\ell(\theta)$ is equal to

$$\ell(\theta) = -2n \ln \theta + \ln \left(\prod_{i=1}^{n} x_i \right) - \frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2.$$

Differentiating $\ell(\theta)$ with respect to θ yields

$$\ell'(\theta) = \frac{-2n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^{n} x_i^2,$$

and we find that $\ell'()\theta) = 0$ if and only if

$$\theta = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} x_i^2}.$$

Since $\ell(\theta)$ attains a maximum at this value of θ (check this!), we find that the maximum likelihood estimate for θ is given by $\sqrt{\frac{1}{2n} \sum_{i=1}^{n} x_i^2}$.

21.8 a The likelihood $L(\theta)$ is given by

$$\begin{aligned} L(\theta) &= C \cdot \left(\frac{1}{4}(2+\theta)\right)^{1997} \cdot \left(\frac{1}{4}\theta\right)^{32} \cdot \left(\frac{1}{4}(1-\theta)\right)^{906} \cdot \left(\frac{1}{4}(1-\theta)\right)^{904} \\ &= \frac{C}{4^{3839}} \cdot (2+\theta)^{1997} \cdot \theta^{32} \cdot (1-\theta)^{1810}, \end{aligned}$$

where C is the number of ways we can assign 1997 starchy-greens, 32 sugary-whites, 906 starchy-whites, and 904 sugary-greens to 3839 plants. Hence the loglikelihood $\ell(\theta)$ is given by

$$\ell(\theta) = \ln(C) - 3839 \ln(4) + 1997 \ln(2+\theta) + 32 \ln(\theta) + 1810 \ln(1-\theta).$$

21.8 b A short calculation shows that

$$\frac{\mathrm{d}\ell(\theta)}{\mathrm{d}\theta} = 0 \qquad \Leftrightarrow \qquad 3810\theta^2 - 1655\theta - 64 = 0,$$

so the maximum likelihood estimate of θ is (after checking that $L(\theta)$ indeed attains a maximum for this value of θ):

$$\frac{-1655 + \sqrt{3714385}}{7620} = 0.0357.$$

21.8 c In this general case the likelihood $L(\theta)$ is given by

$$\begin{aligned} L(\theta) &= C \cdot \left(\frac{1}{4}(2+\theta)\right)^{n_1} \cdot \left(\frac{1}{4}\theta\right)^{n_2} \cdot \left(\frac{1}{4}(1-\theta)\right)^{n_3} \cdot \left(\frac{1}{4}(1-\theta)\right)^{n_4} \cdot \\ &= \frac{C}{4^n} \cdot (2+\theta)^{n_1} \cdot \theta^{n_2} \cdot (1-\theta)^{n_3+n_4}, \end{aligned}$$

where C is the number of ways we can assign n_1 starchy-greens, n_2 sugary-whites, n_3 starchy-whites, and n_4 sugary-greens to n plants. Hence the loglikelihood $\ell(\theta)$ is given by

$$\ell(\theta) = \ln(C) - n \ln(4) + n_1 \ln(2+\theta) + n_2 \ln(\theta) + (n_3 + n_4) \ln(1-\theta).$$

A short calculation shows that

$$\frac{\mathrm{d}\ell(\theta)}{\mathrm{d}\theta} = 0 \qquad \Leftrightarrow \qquad n\theta^2 - (n_1 - n_2 - 2n_3 - 2n_4)\theta - 2n_2 = 0,$$

so the maximum likelihood estimate of θ is (after checking that $L(\theta)$ indeed attains a maximum for this value of θ):

$$\frac{n_1 - n_2 - 2n_3 - 2n_4 + \sqrt{(n_1 - n_2 - 2n_3 - 2n_4)^2 + 8nn_2}}{2n}.$$

21.9 The probability density of this distribution is given by $f_{\alpha,\beta}(x) = 0$ if x is not between α and β, and

$$f_{\alpha,\beta}(x) = \frac{1}{\beta - \alpha} \qquad \text{for} \quad \alpha \le x \le \beta.$$

Since the x_i must be in the interval between α and β, the likelihood (which is a function of two variables!) is given by

$$L(\alpha, \beta) = \left(\frac{1}{\beta - \alpha}\right)^n \qquad \text{for} \quad \alpha \le x_{(1)} \quad \text{and} \quad \beta \ge x_{(n)},$$

and $L(\alpha, \beta) = 0$ for all other values of α and β. So outside the 'rectangle' $(-\infty, x_{(1)}] \times [x_{(n)}, \infty)$ the likelihood is zero, and clearly on this 'rectangle' it attains its maximum in $(x_{(1)}, x_{(n)})$. The maximum likelihood estimates of α and β are therefore $\hat{\alpha} = x_{(1)}$ and $\hat{\beta} = x_{(n)}$.

21.10 The likelihood is

$$L(\alpha) = \frac{\alpha}{x_1^{\alpha+1}} \frac{\alpha}{x_2^{\alpha+1}} \cdots \frac{\alpha}{x_n^{\alpha+1}} = \alpha^n \left(\prod_{i=1}^n x_i\right)^{-(\alpha+1)},$$

so the loglikelihood is

$$\ell(\alpha) = n \ln \alpha - (\alpha + 1) \ln \left(\prod_{i=1}^n x_i\right).$$

Differentiating $\ell(\alpha)$ to α yields as maximum likelihood $\hat{\alpha} = n / \ln \left(\prod_{i=1}^n x_i\right)$.

21.11 a Since the dataset is a realization of a random sample from a $Geo(1/N)$ distribution, the likelihood is $L(N) = \mathrm{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$, where each X_i has a $Geo(1/N)$ distribution. So

$$\begin{aligned} L(N) &= \left(1 - \frac{1}{N}\right)^{x_1 - 1} \frac{1}{N} \left(1 - \frac{1}{N}\right)^{x_2 - 1} \frac{1}{N} \cdots \left(1 - \frac{1}{N}\right)^{x_n - 1} \frac{1}{N} \\ &= \left(1 - \frac{1}{N}\right)^{\left(-n + \sum_{i=1}^n x_i\right)} \left(\frac{1}{N}\right)^n. \end{aligned}$$

But then the loglikelihood is equal to

$$\ell(N) = -n \ln N + \left(-n + \sum_{i=1}^n x_i\right) \ln \left(1 - \frac{1}{N}\right).$$

Differentiating to N yields

$$\frac{\mathrm{d}}{\mathrm{d}N}\big(\ell(N)\big) = \frac{-n}{N} + \left(-n + \sum_{i=1}^{n} x_i\right)\frac{1}{N(N-1)},$$

Now $\frac{\mathrm{d}}{\mathrm{d}N}\big(\ell(N)\big) = 0$ if and only if $N = \bar{x}_n$. Because $\ell(N)$ attains its maximum at $\bar{x}_n$, we find that the maximum likelihood estimate of N is $\hat{N} = \bar{x}_n$.

21.11 b Since $\mathrm{P}(Y = k) = 1/N$ for $k = 1, 2, \ldots, N$, the likelihood is given by

$$L(N) = \left(\frac{1}{N}\right)^n \quad \text{for } N \geq y_{(n)},$$

and $L(N) = 0$ for $N < y_{(n)}$. So $L(N)$ attains its maximum at $y_{(n)}$; the maximum likelihood estimate of N is $\hat{N} = y_{(n)}$.

21.12 Since $L(N) = \mathrm{P}(Z = k)$, it follows from Exercise 4.13**c** that

$$L(N) = \frac{\binom{m}{k}\binom{N-m}{r-k}}{\binom{N}{r}}.$$

In order to see that $L(N)$ increases for $N < \frac{mr}{k}$, and decreases for $N > \frac{mr}{k}$, consider the ratio $L(N)/L(N-1)$. After some elementary calculations one finds that

$$\frac{L(N)}{L(N-1)} = \frac{(N-m)(n-r)}{N(N-m-r+k)}.$$

So $L(N)$ is increasing if $\frac{L(N)}{L(N-1)} = \frac{(N-m)(N-r)}{N(N-m-r+k)} > 1$, and another elementary calculation shows that this is when $N < \frac{mr}{k}$.

21.13 Let $N(t_1, t_2)$ be the number of customers arriving in the show between time t_1 and time t_2. Then it follows from the assumption that customers arrive at the shop according to a Poisson process with rate λ, that

$$\mathrm{P}(N(t_1, t_2) = k) = \frac{(\lambda(t_2 - t_1))^k}{k!}\,\mathrm{e}^{-\lambda(t_2 - t_1)}, \qquad \text{for} \quad k = 0, 1, 2, \ldots.$$

The likelihood $L(\lambda)$ is given by

$$\begin{aligned} L(\lambda) = {} & \mathrm{P}(N(12.00, 12.15) = 2, N(12.15, 12.45) = 0, N(12.45, 13.00) = 1) \\ & + \mathrm{P}(N(12.00, 12.15) = 1, N(12.15, 12.45) = 1, N(12.45, 13.00) = 0)\,. \end{aligned}$$

Since $N(12.00, 12.15)$, $N(12.15, 12.45)$, and $N(12.45, 13.00)$ are independent random variables, we find that

$$\begin{aligned} L(\lambda) &= \frac{(\frac{1}{4}\lambda)^2}{2!}\mathrm{e}^{-\frac{1}{4}\lambda} \cdot \frac{(\frac{1}{2}\lambda)^0}{0!}\mathrm{e}^{-\frac{1}{2}\lambda} \cdot \frac{(\frac{1}{4}\lambda)^1}{1!}\mathrm{e}^{-\frac{1}{4}\lambda} \\ &\quad + \frac{(\frac{1}{4}\lambda)^1}{1!}\mathrm{e}^{-\frac{1}{4}\lambda} \cdot \frac{(\frac{1}{2}\lambda)^1}{1!}\mathrm{e}^{-\frac{1}{2}\lambda} \cdot \frac{(\frac{1}{4}\lambda)^0}{0!}\mathrm{e}^{-\frac{1}{4}\lambda} \\ &= \left(\frac{1}{128}\lambda^3 + \frac{1}{8}\lambda^2\right)\mathrm{e}^{-\lambda}. \end{aligned}$$

Now $L'(\lambda) = 0$ if and only if $\lambda^3 + 13\lambda^2 - 32\lambda = 0$. Since $\lambda > 0$, we find that $\lambda = \frac{-13+\sqrt{297}}{2} = 2.1168439$.

21.14 For $i = 1, 2, \ldots, n$, r_i is the realization of a continuous random variable R_i. Since the shots at the disc do not influence each other, the R_i are independent, and all have the same distribution function $F_\theta(x) = x^2/\theta^2$ if x is between 0 and θ. But then $f_\theta(x) = 2x/\theta^2$ for x between 0 and θ (and $f(x) = 0$ otherwise). Since the disc is hit each of the n shots, the likelihood is:

$$L(\theta) = f_\theta(r_1) f_\theta(r_2) \cdots f_\theta(r_n) = \frac{2r_1}{\theta^2}\frac{2r_2}{\theta^2} \cdots \frac{2r_n}{\theta^2} = \frac{2^n \prod_{i=1}^n r_i}{\theta^{2n}},$$

for $\theta \geq r_{(n)}$, and $L(\theta) = 0$ otherwise. But then we at once see that $L(\theta)$ attains its maximum at $\theta = r_{(n)}$, i.e., the maximum likelihood estimate for θ is $\hat{\theta} = r_{(n)}$.

21.15 At temperature t, the probability of failure of an O-ring is given by $p(t)$, so the probability of a failure of k O-rings at this temperature, for $k = 0, 1, \ldots, 6$, is given by

$$\binom{6}{k}(p(t))^k (1 - p(t))^{6-k}.$$

Setting

$$C = \binom{6}{0}^{16}\binom{6}{1}^{5}\binom{6}{2}^{2},$$

we find that the likelihood $L(a, b)$ is given by

$$\begin{aligned} L(a,b) &= C \cdot (p(53))^2 (1 - p(53))^4 \cdots (p(53))^0 (1 - p(53))^6 \\ &= C \cdot \prod_{i=1}^{23} (p(t_i))^{n_i} (1 - p(t_i))^{6-n_i}, \end{aligned}$$

where t_i is the temperature and n_i is the number of failing O-rings during the ith launch, for $i = 1, 2 \ldots, 23$. But then the loglikelihood $\ell(a, b)$ is given by

$$\ell(a,b) = \ln C + \sum_{i=1}^{23} n_i \ln p(t_i) + \sum_{i=1}^{23} (6 - n_i) \ln(1 - p(t_i)).$$

21.16 Since s_n is the realization of a $Bin(n, p)$ distributed random variable S_n, we find that the likelihood $L(p)$ is given by

$$L(p) = \binom{n}{s_n} p^{s_n} (1 - p)^{n - s_n},$$

from which we see that the loglikelihood $\ell(p)$ satisfies

$$\ell(p) = \ln \binom{n}{s_n} + s_n \ln p + (n - s_n) \ln(1 - p).$$

But then differentiating $\ell(p)$ with respect to p yields that

$$\ell'(p) = \frac{s_n}{p} - \frac{n - s_n}{1 - p},$$

and we find that $\ell'(p) = 0$ if and only if $p = s_n/n$. Since $\ell(p)$ attains a maximum at this value of p (check this!), we have obtained that s_n/n is the maximum likelihood estimate (and S_n/n the maximum likelihood estimator) for p. Due to the invariance principle from Section 21.4 we find that $2n/s_n$ is the maximum likelihood estimate for π, and that

$$T = \frac{2n}{S_n}$$

is the maximum likelihood estimator for π.

22.1 a Since $\sum x_i y_i = 12.4$, $\sum x_i = 9$, $\sum y_i = 4.8$, $\sum x_i^2 = 35$, and $n = 3$, we find (c.f. (22.1) and (22.2)), that

$$\hat{\beta} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} = \frac{3 \cdot 12.4 - 9 \cdot 4.8}{3 \cdot 35 - 9^2} = -\frac{1}{4},$$

and $\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n = 2.35$.

22.1 b Since $r_i = y_i - \hat{\alpha} - \hat{\beta}x_i$, for $i = 1, \ldots, n$, we find $r_1 = 2 - 2.35 + 0.25 = -0.1$, $r_2 = 1.8 - 2.35 + 0.75 = 0.2$, $r_3 = 1 - 2.35 + 1.25 = -0.1$, and $r_1 + r_2 + r_3 = -0.1 + 0.2 - 0.1 = 0$.

22.1 c See Figure 29.2.

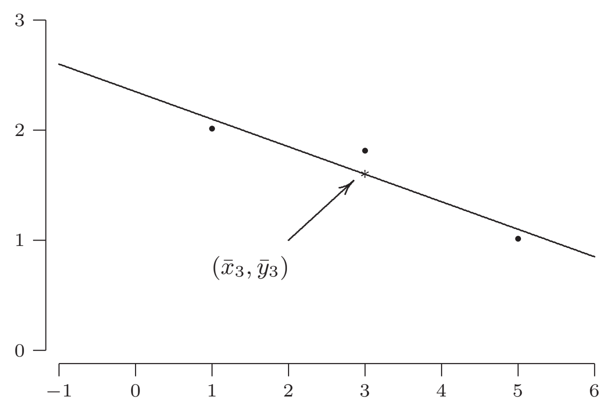


Fig. 29.2. Solution of Exercise 22.1 **c**.

22.2 As in the previous exercise, we have that $\sum x_i y_i = 12.4$, $\sum x_i = 9$, $\sum y_i = 4.8$, $\sum x_i^2 = 35$. However, now $n = 4$, and we find that

$$\hat{\beta} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} = \frac{4 \cdot 12.4 - 9 \cdot 4.8}{4 \cdot 35 - 9^2} = 0.10847.$$

Since we now have that $\bar{x}_n = 9/4$, and $\bar{y}_n = 1.2$, we find that $\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n = 0.9559$.

22.3 a Ignoring the subscripts to the sums, we have

$$\sum x_i = 10, \ \sum x_i^2 = 21.84, \ \sum y_i = 20, \qquad \text{and} \qquad \sum x_i y_i = 41.61$$

From (22.1) we find that the least squares estimate $\hat{\beta}$ of β is given by

$$\hat{\beta} = \frac{5 \cdot 41.61 - 10 \cdot 20}{5 \cdot 21.84 - 10^2} = 0.875,$$

while from (22.2) it follows that the least squares estimate $\hat{\alpha}$ of α is given by

$$\hat{\alpha} = \frac{20}{5} - 0.875 \cdot \frac{10}{5} = 2.25$$

22.3 b See Figure 29.3.

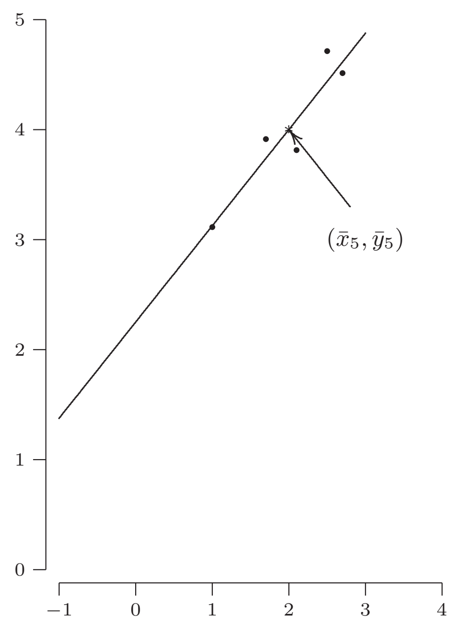


Fig. 29.3. Solution of Exercise 22.3 **b**.

22.4 Since the least squares estimate $\hat{\beta}$ of β is given by (22.1), we find that

$$\hat{\beta} = \frac{100 \cdot 5189 - 231.7 \cdot 321}{100 \cdot 2400.8 - 231.7^2} = 2.385.$$

From (22.2) we find that

$$\hat{\alpha} = 3.21 - 2.317 cdot \hat{\beta} = -2.316.$$

22.5 With the assumption that $\alpha = 0$, the method of least squares tells us now to minimize

$$S(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2 .$$

Now

$$\frac{\mathrm{d}S(\beta)}{\mathrm{d}\beta} = -2 \sum_{i=1}^{n} (y_i - \beta x_i) x_i = -2 \left(\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2 \right) ,$$

so

$$\frac{\mathrm{d}S(\beta)}{\mathrm{d}\beta} = 0 \quad \Leftrightarrow \quad \beta = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Because $S(\beta)$ has a minimum for this last value of β, we see that the least squares estimator $\hat{\beta}$ of β is given by

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

22.6 In Exercise 22.5 we have seen that

$$\hat{\beta} = \frac{\displaystyle\sum_{i=1}^n x_i Y_i}{\displaystyle\sum_{i=1}^n x_i^2},$$

so for the timber dataset we find that $\hat{\beta} = 34.13$. We use the estimated regression line $y = 34.13x$ to predict the Janka hardness. For density $x = 65$ we find as a prediction for the Janka hardness $y = 2218.45$.

22.7 The sum of squares function $S(\alpha, \beta)$ is now given by

$$S(\alpha, \beta) = \sum_{i=1}^n \left(y_i - \mathrm{e}^{\alpha + \beta x_i}\right)^2.$$

In order to find the values of α and β for which $S(\alpha, \beta)$ attains a maximum, one could differentiate $S(\alpha, \beta)$ to α and β. However, this does not yield workable expressions such as those in (22.1) and (22.2). In order to find α and β for a given bivariate dataset $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, an iterative method is best suited.

22.8 In the model with intercept, α and β are chosen in such a way, that $S(\alpha, \beta) = \sum_{i=1}^n r_i^2$ is minimal. In the model with no intercept we chose β in such a way, that $S(0, \beta) = \sum_{i=1}^n r_i^2$ is minimal. Clearly, in this last model the residual sum of squares $\sum_{i=1}^n r_i^2$ is greater than, or equal to the residual sum of squares in the model with intercept.

22.9 With the assumption that $\beta = 0$, the method of least squares tells us now to minimize

$$S(\alpha) = \sum_{i=1}^n (y_i - \alpha)^2.$$

Now

$$\frac{\mathrm{d}S(\alpha)}{\mathrm{d}\alpha} = -2\sum_{i=1}^n (y_i - \alpha) = -2(n\bar{y}_n - n\alpha) = -2n(\bar{y}_n - \alpha).$$

So $\frac{\mathrm{d}S(\alpha)}{\mathrm{d}\alpha} = 0$ if and only if $\alpha = \bar{y}_n$. Since $S(\alpha)$ has a minimum for this last value of α, we see that $\bar{Y}_n$ is the least squares estimator $\hat{\alpha}$ of α.

22.10 a One has that $n = 3$, $\sum x_i = 3$, $\sum y_i = 4$, $\sum x_i^2 = 5$, and $\sum x_i y_i = 2$, so $\hat{\beta} = \frac{3\cdot 2 - 3\cdot 4}{3\cdot 5 - 3^2} = -1$. Since $\bar{x}_3 = 1$ and $\bar{y}_3 = 4/3$, we find that $\hat{\alpha} = 7/3$. Since $r_1 = -1/3$, $r_2 = 2/3$, and $r_3 = -1/3$, we find that $A(\hat{\alpha}, \hat{\beta}) = |r_1| + |r_2| + |r_3| = 4/3$. Note that $S(\hat{\alpha}, \hat{\beta}) = r_1^2 + r_2^2 + r_3^2 = 1/3$.

22.10 b $A(\alpha,-1) = |2-\alpha+0| + |2-\alpha+1| + |0-\alpha+2|$. For $2 \le \alpha < 3$ we find that $A(\alpha,-1) = \alpha - 1$. For $\alpha \ge 3$ we find that $A(\alpha,-1) = 3\alpha - 7$, while for $\alpha < 2$ we find that $A(\alpha,-1) = 7 - 3\alpha$. We find that $A(\alpha,-1) < (\hat{\alpha},\hat{\beta})$ when α is between 17/9 and 7/3. Note that $A(\alpha,-1)$ is minimal (and equal to 1) for $\alpha - 2$.

22.10 c Since $A(2,-1) = 1$, we must have that $1 < \alpha < 3$ (since otherwise $A(\alpha,\beta) \ge |2-\alpha+0| > 1$). Clearly we must have that $\beta < 0$ (since otherwise $A(\alpha,\beta) \ge |0-\alpha-2\beta| > 1$). Considering the various cases for α yields that $A(\alpha,\beta)$ attains its minimum at $\alpha = 2$ and $\beta = -1$.

22.11 a In the present set-up one has to minimize

$$S(\beta,\gamma) = \sum_{i=1}^{n} \left(y_i - (\beta x_i + \gamma x_i^2)\right)^2 .$$

Differentiating $S(\beta,\gamma)$ to β and γ yields

$$\frac{\partial S}{\partial \beta} = -2 \sum_{i=1}^{n} x_i \left(y_i - \beta x_i - \gamma x_i^2\right)$$

and

$$\frac{\partial S}{\partial \gamma} = -2 \sum_{i=1}^{n} x_i^2 \left(y_i - \beta x_i - \gamma x_i^2\right).$$

We find that

$$\frac{\partial S}{\partial \beta} = 0 \quad \Leftrightarrow \quad \beta \sum_{i=1}^{n} x_i^2 + \gamma \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

and

$$\frac{\partial S}{\partial \gamma} = 0 \quad \Leftrightarrow \quad \beta \sum_{i=1}^{n} x_i^3 + \gamma \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i.$$

22.11 b Using Cramer's rule from linear algebra now yields that

$$\beta = \frac{\begin{vmatrix} \sum x_i y_i & \sum x_i^3 \\ \sum x_i^2 y_i & \sum x_i^4 \end{vmatrix}}{\begin{vmatrix} \sum x_i^2 & \sum x_i^3 \\ \sum x_i^3 & \sum x_i^4 \end{vmatrix}} = \frac{(\sum x_i Y_i)(\sum x_i^4) - (\sum x_i^3)(\sum x_i^2 Y_i)}{(\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2}$$

and

$$\gamma = \frac{\begin{vmatrix} \sum x_i^2 & \sum x_i y_i \\ \sum x_i^3 & \sum x_i^2 y_i \end{vmatrix}}{\begin{vmatrix} \sum x_i^2 & \sum x_i^3 \\ \sum x_i^3 & \sum x_i^4 \end{vmatrix}} = \frac{(\sum x_i^2)(\sum x_i^2 Y_i) - (\sum x_i^3)(\sum x_i Y_i)}{(\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2}.$$

Since $S(\beta,\gamma)$ is a 'vase', the above stationary point (β,γ) is a global minimum for $S(\beta,\gamma)$. This finishes the exercise.

22.12 a Since the denominator of $\hat{\beta}$ is a number, *not* a random variable, one has that

$$\mathrm{E}\left[\hat{\beta}\right] = \frac{\mathrm{E}\left[n(\sum x_i Y_i) - (\sum x_i)(\sum Y_i)\right]}{x \sum x_i^2 - (\sum x_i)^2}.$$

Furthermore, the numerator of this last fraction can be written as

$$\mathrm{E}\left[n\sum x_iY_i\right]-\mathrm{E}\left[(\sum x_i)(\sum Y_i)\right],$$

which is equal to

$$n\sum(x_i\mathrm{E}\left[Y_i\right])-(\sum x_i)\sum\mathrm{E}\left[Y_i\right].$$

22.12 b Substituting $\mathrm{E}\left[Y_i\right]=\alpha+\beta x_i$ in the last expression, we find that

$$\mathrm{E}\left[\hat{\beta}\right]=\frac{n\sum(x_i(\alpha+\beta x_i))-(\sum x_i)\left[\sum(\alpha+\beta x_i)\right]}{x\sum x_i^2-(\sum x_i)^2}.$$

22.12 c The numerator of the previous expression for $\mathrm{E}\left[\hat{\beta}\right]$ can be simplified to

$$\frac{n\alpha\sum x_i+n\beta\sum x_i^2-n\alpha\sum x_i-\beta(\sum x_i)(\sum x_i)}{n\sum x_i^2-(\sum x_i)^2},$$

which is equal to

$$\frac{\beta(n\sum x_i^2-(\sum x_i)^2)}{n\sum x_i^2-(\sum x_i)^2}.$$

22.12 d From **c** it now follows that $\mathrm{E}\left[\hat{\beta}\right]=\beta$, i.e., $\hat{\beta}$ is an unbiased estimator for β.

23.1 This is the case: normal data with variance *known*. So we should use the formula from Section 23.2 (the case variance known):

$$\left(\bar{x}_n-1.96\frac{\sigma}{\sqrt{n}},\bar{x}_n+1.96\frac{\sigma}{\sqrt{n}}\right),$$

where $\bar{x}_n=743$, $\sigma=5$ and $n=16$. Because $z_{\alpha/2}=z_{0.025}=1.96$, the 95% confidence interval is:

$$\left(743-1.96\cdot\frac{5}{\sqrt{16}},743+1.96\cdot\frac{5}{\sqrt{16}}\right)=(740.55,\ 745.45).$$

23.2 This is the case: normal data with variance *unknown*. So we should use the formula from Section 23.2 (the case variance unknown):

$$\left(\bar{x}_n-t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n+t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right),$$

where $\bar{x}_n=3.54$, $s_n=0.13$ and $n=34$. Because $t_{n-1,\alpha/2}=t_{33,0.01}\approx t_{30,0.01}=2.457$, the 98% confidence interval is:

$$\left(3.54-2.457\cdot\frac{0.13}{\sqrt{34}},3.54+2.457\cdot\frac{0.13}{\sqrt{34}}\right)=(3.485,\ 3.595).$$

One can redo the same calculation using $t_{33,0.01}=2.445$ (obtained from a software package), and find $(3.4855, 3.5945)$.

23.3 This is the case: normal data with variance *unknown*. So we should use the formula from Section 23.2 (the case variance unknown):

$$\left(\bar{x}_n-t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n+t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right),$$

where $\bar{x}_n=93.5$, $s_n=0.75$ and $n=10$. Because $t_{n-1,\alpha/2}=t_{9,0.025}=2.262$, the 95% confidence interval is:

$$\left(93.5-2.262\cdot\frac{0.75}{\sqrt{10}},93.5+2.262\cdot\frac{0.75}{\sqrt{10}}\right)=(92.96,\ 94.036).$$

23.4 This is the case: normal data with variance *unknown*. So we should use the formula from Section 23.2 (the case variance unknown):

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right),$$

where $\bar{x}_n = 195.3$, $s_n = 16.7$ and $n = 18$. Because $t_{n-1,\alpha/2} = t_{17,0.025} = 2.110$, the 95% confidence interval is:

$$\left(195.3 - 2.110 \cdot \frac{16.7}{\sqrt{18}}, 195.3 + 2.110 \cdot \frac{16.7}{\sqrt{18}}\right) = (186.99,\ 203.61).$$

23.5 a The standard confidence interval for the mean of a normal sample with unknown variance applies, with $n = 23$, $\bar{x} = 0.82$ and $s = 1.78$, so:

$$\left(\bar{x} - t_{22,0.025} \cdot \frac{s}{\sqrt{23}},\ \bar{x} + t_{22,0.025} \cdot \frac{s}{\sqrt{23}}\right).$$

The critical values come from the $t(22)$ distribution: $t_{22,0.025} = 2.074$. The actual interval becomes:

$$\left(0.82 - 2.074 \cdot \frac{1.78}{\sqrt{23}},\ 0.82 + 2.074 \cdot \frac{1.78}{\sqrt{23}}\right) = (0.050,\ 1.590).$$

23.5 b Generate one thousand samples of size 23, by drawing with replacement from the 23 numbers

$$1.06, \quad 1.04, \quad 2.62, \quad \ldots, \quad 2.01.$$

For each sample $x_1^*, x_2^*, \ldots, x_{23}^*$ compute: $t^* = \bar{x}_{23}^* - 0.82/(s_{23}^*/\sqrt{23})$, where $s_{23}^* = \sqrt{\frac{1}{22}\sum(x_i^* - \bar{x}_{23}^*)^2}$.

23.5 c We need to estimate the critical value c_l^* such that $\mathrm{P}(T^* \le c_l^*) \approx 0.025$. We take $c_l^* = -2.101$, the 25th of the ordered values, an estimate for the $25/1000 = 0.025$ quantile. Similarly, c_l^* is estimated by the 976th, which is 2.088.
The bootstrap confidence interval uses the c^* values instead of the t-distribution values $\pm t_{n-1,\alpha/2}$, but beware: c_l^* is from the *left tail* and appears on the *right-hand side* of the interval and c_u^* on the left-hand side:

$$\left(\bar{x}_n - c_u^* \frac{s_n}{\sqrt{n}},\ \bar{x}_n - c_l^* \frac{s_n}{\sqrt{n}}\right).$$

Substituting $c_l^* = -2.101$ and $c_u^* = 2.088$, the confidence interval becomes:

$$\left(0.82 - 2.088 \cdot \frac{1.78}{\sqrt{23}},\ 0.82 + 2.101 \cdot \frac{1.78}{\sqrt{23}}\right) = (0.045,\ 1.600).$$

23.6 a Because events described by inequalities do not change when we multiply the inqualities by a positive constant or add or subtract a constant, the following equalities hold: $\mathrm{P}\left(\tilde{L}_n < \theta < \tilde{U}_n\right) = \mathrm{P}(3L_n + 7 < 3\mu + 7 < 3U_n + 7) = \mathrm{P}(3L_n < 3\mu < 3U_n) = \mathrm{P}(L_n < \mu < U_n)$, and this equals 0.95, as is given.

23.6 b The confidence interval for θ is obtained as the realization of $(\tilde{L}_n, \tilde{U}_n)$, that is: $(\tilde{l}_n, \tilde{u}_n) = (3l_n + 7, 3u_n + 7)$. This is obtained by transforming the confidence interval for μ (using the transformation that is applied to μ to get θ).

23.6 c We start with $\mathrm{P}(L_n < \mu < U_n) = 0.95$ and try to get $1-\mu$ in the middle: $\mathrm{P}(L_n < \mu < U_n) = \mathrm{P}(-L_n > -\mu > -U_n) = \mathrm{P}(1-L_n > 1-\mu > 1-U_n) = \mathrm{P}(1-U_n < 1-\mu < 1-L_n)$, where we see that the minus sign causes an interchange: $\tilde{L}_n = 1-U_n$ and $\tilde{U}_n = 1-L_n$. The confidence interval: $(1-5, 1-(-2)) = (-4, 3)$.

23.6 d If we knew that L_n and U_n were always positive, then we could conclude: $\mathrm{P}(L_n < \mu < U_n) = \mathrm{P}\big(L_n^2 < \mu^2 < U_n^2\big)$ and we could just square the numbers in the confidence interval for μ to get the one for θ. Without the positivity assumption, the sharpest conclusion you can draw from $L_n < \mu < U_n$ is that μ^2 is smaller than the maximum of L_n^2 and U_n^2. So, $0.95 = \mathrm{P}(L_n < \mu < U_n) \le \mathrm{P}\big(0 \le \mu^2 < \max\{L_n^2, U_n^2\}\big)$ and the confidence interval $[0, \max\{l_n^2, u_n^2\}) = [0, 25)$ has a confidence of *at least* 95%. This kind of problem may occur when the transformation is not one-to-one (both -1 and 1 are mapped to 1 by squaring).

23.7 We know that $(l_n, u_n) = (2, 3)$, where l_n and u_n are the realizations of L_n and U_n, that have the property $\mathrm{P}(L_n < \mu < U_n) = 0.95$. This is equivalent with

$$\mathrm{P}\Big(\mathrm{e}^{-U_n} < \mathrm{e}^{-\mu} < \mathrm{e}^{-L_n}\Big) = 0.95,$$

so that $\big(\mathrm{e}^{-3}, \mathrm{e}^{-2}\big) = (0.050, 0.135)$ is a 95% confidence interval for $\mathrm{P}(X=0) = \mathrm{e}^{-\mu}$.

23.8 Define random variables

$$\begin{aligned} X_i &= \text{weight of } i\text{th bottle together with filling amount} \\ W_i &= \text{weight of } i\text{th bottle alone} \\ Y_i &= \text{weight of } i\text{th filling amount} \end{aligned}$$

so that $X_i = W_i + Y_i$. It is given that W_i has a $N(250, 15^2)$ distribution and according to Exercise 23.1, Y_i has a $N(\mu_y, 5^2)$. Since they are independent $X_i = W_i + Y_i$ has a normal distribution with expectation $\mu = 250 + \mu_y$ and variance $15^2 + 5^2 = 250$: $X_i \sim N(250 + \mu_y, 250)$. Our data consist of the weights of 16 filled bottles of wine, $x_1, \ldots, x_{16}$. On the basis of these we can construct a confidence interval for μ. Since we are in the case: normal data with known variance, this is

$$\left(\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

where $\bar{x}_n = 998$, $\sigma = \sqrt{250}$ and $n = 16$. Because $z_{\alpha/2} = z_{0.025} = 1.96$, the 95% confidence interval is:

$$\left(998 - 1.96 \cdot \frac{\sqrt{250}}{\sqrt{16}}, 998 + 1.96 \cdot \frac{\sqrt{250}}{\sqrt{16}}\right) = (990.25,\ 1005.75).$$

Since, this is a 95% confidence interval for $\mu = 250 + \mu_y$, the 95% confidence interval for μ_y, is given by $(990.25 - 250,\ 1005.75 - 50) = (740.25,\ 755.75)$.

23.9 a Since we do not assume a particular parametric model, we are dealing with an empirical bootstrap simulation. Generate bootstrap samples $x_1^*, x_2^*, \ldots, x_n^*$ size $n = 2608$, from the empirical distribution function of the dataset, which is equivalent to generate from the discrete distribution with probability mass function:

a	0	1	2	3	4
$p(a)$	57/2608	203/2608	383/2608	525/2608	532/2608
a	5	6	7	8	9
$p(a)$	408/2608	273/2608	139/2608	45/2608	27/2608
a	10	11	12	13	14
$p(a)$	10/2608	4/2608	0	1/2608	1/2608

Then determine:

$$t^* = \frac{\bar{x}_n^* - 3.8715}{s_n^*/\sqrt{2608}}.$$

Repeat this one thousand times. Next estimate the values c_l^* and c_u^* such that

$$\mathrm{P}(T^* \leq c_l^*) = 0.025 \quad \text{and} \quad \mathrm{P}(T^* \geq c_u^*) = 0.025.$$

The bootstrap confidence interval is then given by:

$$\left(\bar{x}_n - c_u^* \frac{s_n}{\sqrt{n}}, \bar{x}_n - c_l^* \frac{s_n}{\sqrt{n}}\right).$$

23.9 b For a 95% confidence interval we need the empirical 0.025-quantile and the 0.975-th quantile, which we estimate by the 25th and 976th order statistic: $c_l \approx 1.862$ and $c_u \approx -1.888$. This results in

$$\left(3.8715 - 1.862\frac{1.9225}{\sqrt{2608}},\ 3.8715 - (-1.888)\frac{1.9225}{\sqrt{2608}}\right) = (3.801,\ 3.943).$$

23.9 c The (estimated) critical values that we would obtain from the table are -2.228 and 2.234, instead of -1.888 and 1.862 we used in part **b**. Hence, the resulting interval would be larger.

23.10 a This interval has been obtained from

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right),$$

so that the sample mean is in the middle and must be equal to $(1.6 + 7.8)/2 = 4.7$.

23.10 b From the formula we see that half the width of the 95% confidence interval is

$$\frac{7.8 - 1.6}{2} = 3.1 = t_{15,0.025}\frac{s_n}{\sqrt{n}} = 2.131\frac{s_n}{\sqrt{n}}.$$

Similarly, half the width of the 99% confidence interval is

$$t_{15,0.005}\frac{s_n}{\sqrt{n}} = 2.947\frac{s_n}{\sqrt{n}} = \frac{2.947}{2.131}\cdot 2.131\frac{s_n}{\sqrt{n}} = \frac{2.947}{2.131}\cdot 3.1 = 4.287.$$

Hence the 99% confidence interval is

$$(4.7 - 4.287,\ 4.7 + 4.287) = (0.413,\ 8.987).$$

23.11 a For the 98% confidence interval the same formula is used as for the 95% interval, replacing the critical values by larger ones. This is the case, no matter whether the critical values are from the normal or t-distribution, or from a bootstrap experiment. Therefore, the 98% interval *contains* the 95%, and so must also contain the number 0.

23.11 b From a new bootstrap experiment we would obtain new and, most probably, different values c_u^* and c_l^*. It therefore could be, if the number 0 is close to the edge of the first bootstrap confidence interval, that it is just outside the new interval.

23.11 c The new dataset will resemble the old one in many ways, but things like the sample mean would most likely differ from the old one, and so there is no guarantee that the number 0 will again be in the confidence interval.

23.12 a This follows immediately from the change of units rule for normal random variables on page 112 and the fact that if the Z_i's are independent, so are the $\mu + \sigma Z_i$'s.

23.12 b We have

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}(\mu + \sigma Z_i) = \mu + \sigma \cdot \frac{1}{n}\sum_{i=1}^{n} Z_i = \mu + \sigma \bar{Z}.$$

From this we also find that

$$X_i - \bar{X} = (\mu + \sigma Z_i) - (\mu + \sigma \bar{Z}) = \sigma(Z_i - \bar{Z})$$

It follows that

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\sigma^2(Z_i - \bar{Z}_i)^2 = \sigma^2 S_Z^2,$$

so that $S_X = \sigma S_Z$.

23.12 c The equality follows immediately from part **b**, by inserting $\bar{X} - \mu = \sigma\bar{Z}$, and $S_X = \sigma S_Z$. The right hand side is the studentized mean for a random sample from a $N(0, 1)$ distribution, and therefore its distribution does not depend on μ and σ. The left hand side is the studentized mean for a random sample from a $N(\mu, \sigma^2)$ distribution. Since the two are equal, also their distributions are, which means that the distribution of the studentized mean for a random sample from a $N(\mu, \sigma^2)$ distribution (the left hand side) also does not depend on μ and σ.

24.1 From Section 24.1, using $z_{0.05} = 1.645$ we find the equation:

$$\left(\frac{70}{100} - p\right)^2 - \frac{(1.645)^2}{100}\, p(1-p) < 0.$$

This reduces to

$$1.0271\, p^2 - 1.4271\, p + 0.49 < 0.$$

The zeroes of this parabola are

$$p_{1,2} = \frac{-(-1.4271) \pm \sqrt{(-1.4271)^2 - 4 \cdot 1.0271 \cdot 0.49}}{2 \cdot 1.0271} = 0.6947 \pm 0.0746$$

and the 90% confidence interval is $(0.6202, 0.7693)$.

24.2 a Since $n = 6$ is too small to expect a good approximation from the central limit theorem, we cannot apply the Wilson method.

24.2 b Solve

$$\left(\frac{140}{250} - p\right)^2 - \frac{(1.96)^2}{250}\, p(1-p) = 0$$

or

$$1.0154\, p^2 - 1.1354\, p + 0.3136 = 0.$$

That is

$$p_{1,2} = \frac{-(-1.1354) \pm \sqrt{(-1.1354)^2 - 4 \cdot 1.0154 \cdot 0.3136}}{2 \cdot 1.0154} = 0.5591 \pm 0.0611$$

and the 90% confidence interval is $(0.4980, 0.6202)$.

24.3 The width of the confidence interval is $2 \cdot 2.576 \cdot \sigma/\sqrt{n}$, where $\sigma = 5$. So, we require (see Section 24.4):

$$n \geq \left(\frac{2 \cdot 2.576 \cdot 5}{1}\right)^2 = (25.76)^2 = 663.6,$$

that is, at least a sample size of 664.

24.4 a The width of the confidence interval will be about $2t_{n-1,0.05}\, s/\sqrt{n}$. For s we substitute our current estimate of σ, 0.75, and we use $t_{n-1,0.05} = z_{0.05} = 1.645$, for the moment assuming that n will be large. This results in

$$n \geq \left(\frac{2 \cdot 1.645 \cdot 0.75}{0.1}\right)^2 = 608.9,$$

so we use $n = 609$ (which is indeed large, so it is appropriate to use the critical value from the normal distribution).

24.4 b In our computation, we used $s = 0.75$. From the new dataset of size 609 we are going to compute s_{609} and use that in the computation. If $s_{609} > 0.75$ then the confidence interval will be too wide.

24.5 a From

$$\mathrm{P}\left(-z_{\alpha/2} < \frac{X - np}{\sqrt{n/4}} z_{\alpha/2}\right) = 1 - \alpha$$

we deduce

$$\mathrm{P}\left(\frac{X}{n} - \frac{z_{\alpha/2}}{2\sqrt{n}} < p < \frac{X}{n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right) = 1 - \alpha.$$

So, the approximate 95% confidence interval is

$$\left(\frac{x}{n} - \frac{z_{0.05}}{2\sqrt{n}}, \frac{x}{n} + \frac{z_{0.05}}{2\sqrt{n}}\right).$$

The width is $2z_{0.05}/(2\sqrt{n}) = z_{0.05}/\sqrt{n}$ and so n should satisfy $1.96/\sqrt{n} \leq 0.01$ or $n \geq (196)^2 = 38416$.

24.5 b The confidence interval is

$$\frac{19477}{38416} \pm \frac{1.96}{2 \cdot \sqrt{38416}} = 0.5070 \pm 0.005,$$

so of the intended width.

24.6 a The environmentalists are interested in a *lower* confidence bound, because they would like to make a statement like "We are 97.5% confidence that the concentration exceeds 1.68 ppm [and that is much too high.]" We have normal data, with σ unknown so we use $s_{16} = \sqrt{1.12} = 1.058$ as an estimate and use the critical value corresponding to 2.5% from the $t(15)$ distribution: $t_{15,0.025} = 2.131$. The lower confidence bound is $2.24 - 2.131 \cdot 1.058/\sqrt{16} = 2.24 - 0.56 = 1.68$, the interval: $(1.68, \infty)$.

24.6 b For similar reasons, the plant management constructs an *upper* confidence bound ("We are 97.5% confident pollution does not exceed 2.80 [and this is acceptable.]"). The computation is the same except for a minus sign: $2.24 + 2.131 \cdot 1.058/\sqrt{16} = 2.24 + 0.56 = 2.80$, so the interval is $[0, 2.80)$. Note that the computed upper and lower bounds are in fact the endpoints of the 95% two-sided confidence interval.

24.7 a From the normal approximation we know

$$\mathrm{P}\left(-z_{0.025} < \frac{\bar{X}_n - \mu}{\sqrt{\mu}/\sqrt{n}} < z_{0.025}\right) \approx 0.95$$

or

$$\mathrm{P}\left(\left(\frac{\bar{X}_n - \mu}{\sqrt{\mu}/\sqrt{n}}\right)^2 < z_{0.025}^2\right) \approx 0.95,$$

i.e.,

$$\mathrm{P}\left(\left(\bar{X}_n - \mu\right)^2 < (1.96)^2 \, \frac{\mu}{\sqrt{n}}\right) \approx 0.95.$$

Just as with the derivation of the Wilson method (Section 24.1) we now conclude that the 95% confidence interval contains those μ for which

$$(\bar{x}_n - \mu)^2 \leq (1.96)^2 \frac{\mu}{n}.$$

24.7 b We need to solve $(\bar{x}_n - \mu)^2 - (1.96)^2\mu/n = 0$, where $\bar{x}_n = 3.8715$ and $n = 2608$, resulting in

$$\mu^2 - \left(2 \cdot 3.8715 + \frac{(1.96)^2}{2608}\right)\mu + (3.8715)^2 = 0,$$

or

$$\mu^2 - 7.7446\mu + 14.9889 = 0.$$

From the roots we find the confidence interval $(3.7967, 3.9478)$.

24.7 c The confidence interval $(3.7967, 3.9478)$is almost the same as the one in Exercise 23.9 **b**. This is not surprising: n is very large, so the normal approximation should be very good.

24.8 a We solve

$$\left(\frac{15}{23} - p\right)^2 - \frac{(1.96)^2}{23}\, p(1-p) = 0$$

or

$$1.1670\, p^2 - 1.4714\, p + 0.4253 = 0,$$

from which we find the confidence interval $(0.4489, 0.8119)$.

24.8 b Replacing the 1.96 in the equation above by $z_{0.05} = 1.645$, we find $(0.4808, 0.7915)$. The appropriate one-sided confidence interval will provide us with a *lower* bound on the outer lane winning probability p, so it is $(0.4808, 1]$.

24.9 a From Section 8.4 we know: $\mathrm{P}(M \leq a) = [F_X(a)]^{12}$, so $\mathrm{P}(M/\theta \leq t) = \mathrm{P}(M \leq \theta t) = [F_X(\theta t)]^{12}$. Since X_i has a $U(0,\theta)$ distribution, $F_X(\theta t) = t$, for $0 \leq t \leq 1$. Substituting this shows the result.

24.9 b For c_l we need to solve $(c_l)^{12} = \alpha/2$, or $c_l = (\alpha/2)^{1/12} = (0.05)^{1/12} = 0.7791$. For c_u we need to solve $(c_u)^{12} = 1-\alpha/2$, or $c_u = (1-\alpha/2)^{1/12} = (0.95)^{1/12} = 0.9958$.

24.9 c From **b** we know that $\mathrm{P}(c_l < M/\theta < c_u) = \mathrm{P}(0.7790 < M/\theta < 0.9958) = 0.90$. Rewriting this equation, we get: $\mathrm{P}(0.7790\,\theta < M < 0.9958\,\theta) = 0.90$ and $\mathrm{P}(M/0.9958 < \theta < M/0.7790) = 0.90$. This means that $(m/0.9958,\ m/0.7790) = (3.013,\ 3.851)$ is a 90% confidence interval for θ.

24.9 d From **b** we derive the general formula:

$$\mathrm{P}\left((\alpha/2)^{1/n} < \frac{M}{\theta} < (1-\alpha/2)^{1/n}\right) = 1-\alpha.$$

The left hand inequality can be rewritten as $\theta < M/(\alpha/2)^{1/n}$ and the right hand one as $M/(1-\alpha/2)^{1/n} < \theta$. So, the statement above can be rewritten as:

$$\mathrm{P}\left(\frac{M}{(1-\alpha/2)^{1/n}} < \theta < \frac{M}{(\alpha/2)^{1/n}}\right) = 1-\alpha,$$

so that the general formula for the confidence interval becomes:

$$\left(\frac{m}{(1-\alpha/2)^{1/n}}, \frac{m}{(\alpha/2)^{1/n}}\right).$$

24.10 a From Section 11.2 we know that S_n, being the sum of n independent $Exp(\lambda)$ random variables, has a $Gam(n,\lambda)$ distribution. From Exercise 8.4 we know: λX_i has an $Exp(1)$ distribution. Combining these facts, it follows that $\lambda S_n = \lambda X_1 + \cdots + \lambda X_n$ has a $Gam(n,1)$ distribution.

24.10 b From the quantiles we see

$$\begin{aligned} 0.9 &= \mathrm{P}(q_{0.05} \leq \lambda S_{20} \leq q_{0.95}) \\ &= \mathrm{P}(13.25 \leq \lambda S_{20} \leq 27.88) \\ &= \mathrm{P}(13.25/S_{20} \leq \lambda \leq 27.88/S_{20}). \end{aligned}$$

Noting that the realization of S_{20} is $x_1 + \cdots + x_{20} = 20\,\bar{x}_{20}$, we conclude that

$$\left(\frac{13.25}{20\,\bar{x}_{20}}, \frac{27.88}{20\,\bar{x}_{20}}\right) = \left(\frac{0.6625}{\bar{x}_{20}}, \frac{1.394}{\bar{x}_{20}}\right)$$

is a 95% confidence interval for λ.

25.1 The alternative hypothesis should reflect the belief that arrival delays of trains exhibit *more* variation during rush hours than during quiet hours. Therefore take $H_1 : \sigma_1 > \sigma_2$.

25.2 The alternative hypothesis should reflect the belief that the number of babies born in Cleveland, Ohio, in the month of September in 1977 is *higher* than 1472. Therefore, take $H_1 : \mu > 1472$.

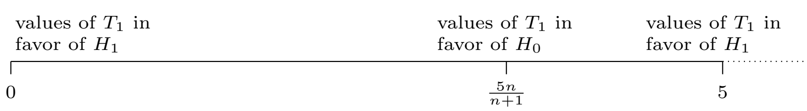
25.3 When the regression line runs through the origin, then $\alpha = 0$. One possible testing problem is to test $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$. If $H_0 : \alpha = 0$ is rejected in favor of $H_1 : \alpha \neq 0$, the parameter α should be left in the model.

25.4 a Denote the observed numbers of cycles for the smokers by $X_1, X_2, \ldots, X_{n_1}$ and similarly $Y_1, Y_2, \ldots, Y_{n_2}$ for the nonsmokers. A test statistic should compare estimators for p_1 and p_2. Since the geometric distributions have expectations $1/p_1$ and $1/p_2$, we could compare the estimator $1/\bar{X}_{n_1}$ for p_1 with the estimator $1/\bar{Y}_{n_2}$ for p_2, or simply compare $\bar{X}_{n_1}$ with $\bar{Y}_{n_2}$. For instance, take test statistic $T = \bar{X}_{n_1} - \bar{Y}_{n_2}$. Values of T close to zero are in favor of H_0, and values far away from zero are in favor of H_1. Another possibility is $T = \bar{X}_{n_1}/\bar{Y}_{n_2}$.

25.4 b In this case, the maximum likelihood estimators $\hat{p}_1$ and $\hat{p}_2$ give better indications about p_1 and p_2. They can be compared in the same way as the estimators in **a**.

25.4 c The probability of getting pregnant during a cycle is p_1 for the smoking women and p_2 for the nonsmokers. The alternative hypothesis should express the belief that smoking women are *less likely* to get pregnant than nonsmoking women. Therefore take $H_1 : p_1 < p_2$.

25.5 a When the maximum is greater than 5, at least one X_i is greater than 5 so that θ must be greater than 5, and we know for sure that the null hypothesis is false. Therefore the set of relevant values of $T_1 = \max\{X_1, X_2, \ldots, X_n\}$ is the interval $[0, 5]$. Similar to Exercise 8.15, one can argue that $\mathrm{E}[T_1] = n\theta/(n+1)$. Hence values of T_1 close to $5n/(n+1)$ are in favor of H_0. Values of T_1 in the neighborhood of 0 indicate that $\theta < 5$, and values of T_1 very close to 5 indicate that $\theta > 5$. *Both* these regions are in favor of H_1.



25.5 b When the distance between $2\bar{X}_n$ and 5 is greater than 5, then $2\bar{X}_n$ must be greater than 10, which means that at least one X_i is greater than 5. In that case we know for sure that the null hypothesis is false. Therefore the set of relevant values of $T_2 = |2\bar{X}_n - 5|$ is the interval $[0, 5]$. Since $\bar{X}_n$ will be close to $\theta/2$, values of T_2 close to zero are in favor of H_0. Values of T_2 far away from zero either correspond to $2\bar{X}_n$ far below 5, which indicates $\theta < 5$, or correspond to $2\bar{X}_n$ far above 5, which indicates $\theta > 5$. Hence values of T_2 far away from zero are in favor of H_1.

25.6 a The p-value $\mathrm{P}(T \geq 2.34) = 0.23$ is larger than 0.05, so do not reject.

25.6 b The p-value $\mathrm{P}(T \geq 2.34) = 1 - \mathrm{P}(T \leq 2.34) = 0.77$ is larger than 0.05, so do not reject.

25.6 c The p-value $\mathrm{P}(T \geq 0.03) = 0.968$ is larger than 0.05, so do not reject.

25.6 d The p-value $P(T \geq 1.07) = 1 - P(T \leq 1.07) = 0.019$ is less than 0.05, so reject.

25.6 e The p-value $P(T \geq 1.07) \geq P(T \geq 2.34) = 0.99$, which is larger than 0.05, so do not reject.

25.6 f The p-value $P(T \geq 2.34) \leq P(T \geq 1.07) = 0.0.019$, which is smaller than 0.05, so reject.

25.6 g The p-value $P(T \geq 2.34) \leq P(T \geq 1.07) = 0.200$. Therefore, the p-value is smaller than 0.200, but that does not give enough information to decide about the null hypothesis.

25.7 a Since the parameter μ is the expectation of T, values of T much larger than 1472 suggest that $\mu > 1472$. Because we test $H_0 : \mu = 1472$ against $H_1 : \mu > 1472$, the more values of T are to the right, the stronger evidence they provide in favor of H_1.

25.7 b According to part **a**, values to the right of $t = 1718$ bare stronger evidence in favor of H_1. Therefore, the p-value is $P(T \geq 1718)$, where T has a Poisson distribution with $\mu = 1472$. Because the distribution of T can be approximated by a normal distribution with mean 1472 and variance 1472, we can approximate this probability as follows

$$P(T \geq 1718) = P\left(\frac{T - 1472}{\sqrt{1472}} \geq \frac{1718 - 1472}{\sqrt{1472}}\right) \approx P(Z \geq 6.412)$$

where Z has an $N(0, 1)$ distribution. From Table **??** we see that the latter probability is almost zero (to be precise, $7.28 \cdot 10^{-11}$, which was obtained using a statistical software package).

25.8 The values of F_n and Φ lie between 0 and 1, so that the maximal distance between the two graphs also lies between 0 and 1. In fact, since F_n has jumps of size $1/n$, the minimal value of T must be half the size of a jump: $1/(2n)$. This would correspond to the situation, where at each observation, the graph of Φ precisely runs through the middle of the two heights of F_n. When the graph of F_n lies far to the right (or to the left) of that of Φ, the maximum distance between the two graphs can be arbitrary close to 1.
When the dataset is a realization from a distribution different from the standard normal, the corresponding distribution function F will differ from Φ. Since $F_n \approx F$ (recall Table 17.2), the graph of F_n will show large differences with that of Φ resulting in a relatively large value of T. On the other hand, when the dataset is indeed a realization from the standard normal, then $F_n \approx \Phi$, resulting in a relative small value of T. We conclude that only large values of T close to 1 are evidence against the null hypothesis.

25.9 Only values of T_{ks} close to 1 are evidence against the null hypothesis. Therefore the p-value is $P(T_{\text{ks}} \geq 0.176)$. On the basis of the bootstrap results, this probability is approximated by the relative frequency of T_{ks}-values greater than or equal to 0.176, which is zero.

25.10 a The alternative hypothesis should express the belief that the gross calorific *exceeds* 23.75 MJ/kg. Therefore take $H_1 : \mu > 23.75$.

25.10 b The p-value is the probability $\mathrm{P}\big(\bar{X}_n \geq 23.788\big)$ under the null hypothesis. We can compute this probability by using that under the null hypothesis $\bar{X}_n$ has a $N(23.75, (0.1)^2/23)$ distribution:

$$\mathrm{P}\big(\bar{X}_n \geq 23.788\big) = \mathrm{P}\left(\frac{\bar{X}_n - 23.75}{0.1/\sqrt{23}} \geq \frac{23.788 - 23.75}{0.1/\sqrt{23}}\right) = \mathrm{P}(Z \geq 1.82)\,,$$

where Z has an $N(0,1)$ distribution. From Table **??** we find $\mathrm{P}(Z \geq 1.82) = 0.0344$.

25.11 A type I error occurs when $\mu = 0$ and $|t| \geq 2$. When $\mu = 0$, then T has an $N(0,1)$ distribution. Hence, by symmetry of the $N(0,1)$ distribution and Table **??**, we find that the probability of committing a type I error is

$$\mathrm{P}(|T| \geq 2) = \mathrm{P}(T \leq -2) + \mathrm{P}(T \geq 2) = 2 \cdot \mathrm{P}(T \geq 2) = 2 \cdot 0.0228 = 0.0456.$$

26.1 A type I error is to falsely reject the null hypothesis, i.e., to falsely conclude "suspect is guilty". This happened in 9 out of 140 cases. Hence, the probability of a type I error is $9/140 = 0.064$.
A type II error is to falsely accept the null hypothesis, i.e., to falsely conclude "suspect is innocent". This happened in 15 out of 140 cases. Hence, the probability of a type II error is $15/140 = 0.107$.

26.2 According to Exercise 25.11, we do not reject if $|T| < 2$. Therefore the probability of a type II error is

$$\mathrm{P}(|T| < 2) = \mathrm{P}(-2 < T < 2) = \mathrm{P}(T < 2) - \mathrm{P}(T < -2)\,,$$

where T has a $N(1,1)$ distribution. Using that $T-1$ has a $N(0,1)$ distribution, we find that

$$\begin{aligned}\mathrm{P}(T < 2) - \mathrm{P}(T < -2) &= \mathrm{P}(T - 1 < 1) - \mathrm{P}(T - 1 < -3)\\ &= 0.8413 - 0.0013 = 0.84.\end{aligned}$$

26.3 a A type I error is to falsely reject the null hypothesis $H_0 : \theta = 2$. We reject when $X \leq 0.1$ or $X \geq 1.9$. Because under the null hypothesis, X has a $U(0,2)$ distribution, the probability of committing a type I error is

$$\mathrm{P}(X \leq 0.1 \,|\, \theta = 2) + \mathrm{P}(X \geq 1.9 \,|\, \theta = 2) = 0.05 + 0.05 = 0.1.$$

26.3 b In this case a type II error is to falsely accept the null hypothesis when $\theta = 2.5$ We accept when $0.1 < X < 1.9$. Because under $\theta = 2.5$, X has a $U(0, 2.5)$ distribution, the probability of committing a type I error is

$$\mathrm{P}(0.1 < X < 1.9 \,|\, \theta = 2.5) = \frac{1.9 - 0.1}{2.5} = 0.72.$$

26.4 a Since T has a $Bin(144, p)$ distribution, values of T close to $144/8 = 18$ are in favor of the null hypothesis. Values of T far above 18 indicate that $p > 1/8$, whereas values far below 18 indicate that $p < 1/8$. This means we reject only for values of T far above 18. Hence we are only dealing with a right critical value.

26.4 b Denote the right critical value by c. Then we must solve $\mathrm{P}(T \geq c) = 0.01$. Using the normal approximation for the binomial distribution,

$$\mathrm{P}(T \geq c) \approx \mathrm{P}\left(Z \geq \frac{c - np}{\sqrt{np(1-p)}}\right) = \mathrm{P}\left(Z \geq \frac{c - 18}{\sqrt{18 \cdot \frac{7}{8}}}\right),$$

where Z has a $N(0,1)$ distribution. We find c by solving

$$\frac{c - 18}{\sqrt{18 \cdot \frac{7}{8}}} = z_{0.01} = 2.326,$$

which gives $c = 27.2$. Because T can only take integer values, the right critical value is taken to be 28, and the critical region is $\{28, 29, \cdots, 144\}$. Since the observed number $t = 29$ falls in the critical region, we reject $H_0 : p = 1/8$ in favor of $H_1 : p > 1/8$.

26.5 a The p-value is $\mathrm{P}(X \geq 15)$ under the null hypothesis $H_0 : p = 1/2$. Using Table 26.3 we find $\mathrm{P}(X \geq 15) = 1 - \mathrm{P}(X \leq 14) = 1 - 0.8950 = 0.1050$.

26.5 b Only values close to 23 are in favor of $H_1 : p > 1/2$, so the critical region is of the form $K = \{c, c+1, \ldots, 23\}$. The critical value c is the smallest value, such that $\mathrm{P}(X \geq c) \leq 0.05$ under $H_0 : p = 1/2$, or equivalently, $1 - \mathrm{P}(X \leq c - 1) \leq 0.05$, which means $\mathrm{P}(X \leq c - 1) \geq 0.95$. From Table 26.3 we conclude that $c - 1 = 15$, so that $K = \{16, 17, \ldots, 23\}$.

26.5 c A type I error occurs if $p = 1/2$ and $X \geq 16$. The probability that this happens is $\mathrm{P}(X \geq 16 \mid p = 1/2) = 1 - \mathrm{P}(X \leq 15 \mid p = 1/2) = 1 - 0.9534 = 0.0466$, where we have used Table 26.3 once more.

26.5 d In this case, a type II error occurs if $p = 0.6$ and $X \leq 15$. To approximate $\mathrm{P}(X \leq 15 \mid p = 0.6)$, we use the same reasoning as in Section 14.2, but now with $n = 23$ and $p = 0.6$. Write X as the sum of independent Bernoulli random variables: $X = R_1 + \cdots + R_n$, and apply the central limit theorem with $\mu = p = 0.6$ and $\sigma^2 = p(1-p) = 0.24$. Then

$$\begin{aligned}\mathrm{P}(X \leq 15) &= \mathrm{P}(R_1 + \cdots + R_n \leq 15) \\ &= \mathrm{P}\left(\frac{R_1 + \cdots + R_n - n\mu}{\sigma\sqrt{n}} \leq \frac{15 - n\mu}{\sigma\sqrt{n}}\right) \\ &= \mathrm{P}\left(Z_{23} \geq \frac{15 - 13.8}{\sqrt{0.24}\sqrt{23}}\right) \approx \Phi(0.51) = 0.6950.\end{aligned}$$

26.6 a Because T has a $Pois(\mu)$ distribution (see Exercise 25.7), we always have $\mathrm{E}[T] = \mu$. Therefore values of T around 1472 are in favor of the null hypothesis, values of T far to the left of 1472 are in favor of $\mu < 1472$, and values of T far to the right 1472 are in favor of $\mu > 1472$. Therefore, only values far to the right of 1472 are in favor of $H_1 : \mu > 1472$, so that we only have a right critical value.

26.6 b Since, according to part **a**, we only have a right critical value c, we must solve $\mathrm{P}(T \geq c) = 0.05$. Using the normal approximation

$$\mathrm{P}(T \geq c) = \mathrm{P}\left(\frac{T - \mu}{\sqrt{\mu}} \geq \frac{c - \mu}{\sqrt{\mu}}\right) \approx \mathrm{P}\left(Z \geq \frac{c - 1472}{\sqrt{1472}}\right),$$

where Z has a $N(0,1)$ distribution. We find c by solving

$$\frac{c - 1472}{\sqrt{1472}} = z_{0.05} = 1.645,$$

which gives $c = 1535.1$. Because T can only take integer values we take right critical value 1536, and critical region $\{1536, 1537, \infty\}$. The observed number 1718 falls into the critical region, so that we reject $H_0 : \mu = 1472$ in favor of $H_1 : \mu > 1472$.

26.7 We must solve $\mathrm{P}(X_1 + X_2 \leq c) = 0.05$ under the null hypothesis. Under the null hypothesis, X_1 and X_2 are independent random variables with a $U(0,1)$ distribution. According to Exercise 11.5, the random variable $T = X_1 + X_2$ has density $f(t) = t$, for $0 \leq t \leq 1$, and $f(t) = 2 - t$, for $1 \leq t \leq 2$. Since this density integrates to 0.5 for $0 \leq t \leq 1$, we can find c by solving

$$\int_0^c t \,\mathrm{d}t = \mathrm{P}(T \leq c) = 0.05,$$

or equivalently $\frac{1}{2}c^2 = 0.05$, which gives left critical value $c = 0.316$. The corresponding critical region for $T = X_1 + X_2$ is $[0,\ 0.316]$.

26.8 a Test statistic $T = \bar{X}_n$ takes values in $(0, \infty)$. Recall that the $Exp(\lambda)$ distribution has expectation $1/\lambda$, and that according to the law of large numbers $\bar{X}_n$ will be close to $1/\lambda$. Hence, values of $\bar{X}_n$ close to 1 are in favor of $H_0 : \lambda = 1$, and *only* values of $\bar{X}_n$ close to zero are in favor $H_1 : \lambda > 1$. Large values of $\bar{X}_n$ also provide evidence against $H_0 : \lambda = 1$, but even stronger evidence against $H_1 : \lambda > 1$. We conclude that $T = \bar{X}_n$ has critical region $K = (0, c_l]$. This is an example in which the alternative hypothesis and the test statistic deviate from the null hypothesis in *opposite* directions.
Test statistic $T' = \mathrm{e}^{-\bar{X}_n}$ takes values in $(0,1)$. Values of $\bar{X}_n$ close to zero correspond to values of T' close to 1, and large values of $\bar{X}_n$ correspond to values of T' close to 0. Hence, *only* values of T' close to 1 are in favor $H_1 : \lambda > 1$. We conclude that T' has critical region $K' = [c_u, 1)$. Here the alternative hypothesis and the test statistic deviate from the null hypothesis in the *same* direction.

26.8 b Again, values of $\bar{X}_n$ close to 1 are in favor of $H_0 : \lambda = 1$. Values of $\bar{X}_n$ close to zero suggest $\lambda > 1$, whereas large values of $\bar{X}_n$ suggest $\lambda < 1$. Hence, both small and large values of $\bar{X}_n$ are in favor of $H_1 : \lambda \neq 1$. We conclude that $T = \bar{X}_n$ has critical region $K = (0, c_l] \cup [c_u, \infty)$.
Small and large values of $\bar{X}_n$ correspond to values of T' close to 1 and 0. Hence, values of T' both close to 0 and close 1 are in favor of $H_1 : \lambda \neq 1$. We conclude that T' has critical region $K' = (0, c'_l] \cup [c'_u, 1)$. Both test statistics deviate from the null hypothesis in the same directions as the alternative hypothesis.

26.9 a Test statistic $T = (\bar{X}_n)^2$ takes values in $[0, \infty)$. Since μ is the expectation of the $N(\mu, 1)$ distribution, according to the law of large numbers, $\bar{X}_n$ is close to μ. Hence, values of $\bar{X}_n$ close to zero are in favor of $H_0 : \mu = 0$. Large negative values of $\bar{X}_n$ suggest $\mu < 0$, and large positive values of $\bar{X}_n$ suggest $\mu > 0$. Therefore, both large negative and large positive values of $\bar{X}_n$ are in favor of $H_1 : \mu \neq 0$. These values correspond to large positive values of T, so T has critical region $K = [c_u, \infty)$. This is an example in which the test statistic deviates from the null hypothesis in *one* direction, whereas the alternative hypothesis deviates in *two* directions.
Test statistic T' takes values in $(-\infty, 0) \cup (0, \infty)$. Large negative values and large positive values of $\bar{X}_n$ correspond to values of T' close to zero. Therefore, T' has critical region $K' = [c'_l, 0) \cup (0, c'_u]$. This is an example in which the test statistic deviates from the null hypothesis for small values, whereas the alternative hypothesis deviates for large values.

26.9 b Only large positive values of $\bar{X}_n$ are in favor of $\mu > 0$, which correspond to large values of T. Hence, T has critical region $K = [c_u, \infty)$. This is an example where the test statistic has the *same type* of critical region with a one-sided or two-sided alternative. Of course, the critical value c_u in part **b** is different from the one in part **a**.
Large positive values of $\bar{X}_n$ correspond to small positive values of T'. Hence, T' has critical region $K' = (0, c'_u]$. This is another example where the test statistic deviates from the null hypothesis for small values, whereas the alternative hypothesis deviates for large values.

27.1 a The value of the t-test statistic is

$$t = \frac{\bar{x}_n - 10}{s_n/\sqrt{n}} = \frac{11 - 10}{2/\sqrt{4}} = 2.$$

The right critical value is $t_{n-1,\alpha/2} = t_{15,0.025} = 2.131$. The observed value $t = 2$ is smaller than this, so we do not reject $H_0 : \mu = 10$ in favor of $H_1 : \mu \neq 10$.

27.1 b The right critical value is now $t_{n-1,\alpha} = t_{15,0.05} = 1.753$. The observed value $t = 2$ is larger than this, so we reject $H_0 : \mu = 10$ in favor of $H_1 : \mu > 10$.

27.2 a The belief that the pouring temperature is at the right target setting is put to the test. The alternative hypothesis should represent the belief that the pouring temperature *differs* from the target setting. Hence, test $H_0 : \mu = 2550$ against $H_1 : \mu \neq 2550$.

27.2 b The value of the t-test statistic is

$$t = \frac{\bar{x}_n - 2550}{s_n/\sqrt{n}} = \frac{2558.7 - 2550}{\sqrt{517.34}/\sqrt{10}} = 1.21.$$

Because $H_1 : \mu \neq 2550$, both small and large values of T are in favor of H_1. Therefore, the right critical value is $t_{n-1,\alpha/2} = t_{9,0.025} = 3.169$. The observed value $t = 1.21$ is smaller than this, so we do not reject $H_0 : \mu = 2550$ in favor of $H_1 : \mu \neq 2550$.

27.3 a The alternative hypothesis should represent the belief that the load at failure *exceeds* 10 MPa. Therefore, take $H_1 : \mu > 10$.

27.3 b The value of the t-test statistic is

$$t = \frac{\bar{x}_n - 10}{s_n/\sqrt{n}} = \frac{13.71 - 10}{3.55/\sqrt{22}} = 4.902.$$

Because $H_1 : \mu > 10$, only large values of T are in favor of H_1. Therefore, the right critical value is $t_{n-1,\alpha} = t_{21,0.05} = 1.721$. The observed value $t = 4.902$ is larger than this, so we reject $H_0 : \mu = 10$ in favor of $H_1 : \mu > 10$.

27.4 The value of the t-test statistic is

$$t = \frac{\bar{x}_n - 31}{s_n/\sqrt{n}} = \frac{31.012 - 31}{0.1294/\sqrt{22}} = 0.435.$$

Because $H_1 : \mu > 31$, only large values of T are in favor of H_1. Therefore, the right critical value is $t_{n-1,\alpha} = t_{21,0.01} = 2.518$. The observed value $t = 0.435$ is smaller than this, so we do not reject $H_0 : \mu = 31$ in favor of $H_1 : \mu > 31$.

27.5 a The interest is whether the inbreeding coefficient exceeds 0. Let μ represent this coefficient for the species of wasps. The value 0 is the a priori specified value of the parameter, so test null hypothesis $H_0 : \mu = 0$. The alternative hypothesis should express the belief that the inbreeding coefficient *exceeds* 0. Hence, we take alternative hypothesis $H_1 : \mu > 0$. The value of the test statistic is

$$t = \frac{0.044}{0.884/\sqrt{197}} = 0.70.$$

27.5 b Because $n = 197$ is large, we approximate the distribution of T under the null hypothesis by an $N(0, 1)$ distribution. The value $t = 0.70$ lies to the right of zero, so the p-value is the right tail probability $\mathrm{P}(T \geq 0.70)$. By means of the normal approximation we find from Table **??** that the right tail probability

$$\mathrm{P}(T \geq 0.70) \approx 1 - \Phi(0.70) = 0.2420.$$

This means that the value of the test statistic is not very far in the (right) tail of the distribution and is therefore not to be considered exceptionally large. We do not reject the null hypothesis.

27.6 The belief that the intercept is zero is put to the test. The alternative hypothesis should represent the belief that the intercept *differs* from zero. Therefore, test $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$. The value of the t-test statistic is

$$t_a = \frac{\hat{\alpha}}{s_a} = \frac{5.388}{1.874} = 2.875.$$

Because $H_1 : \alpha \neq 0$, both small and large values of T_a are in favor of H_1. Therefore, the right critical value is $t_{n-1,\alpha/2} = t_{5,0.05} = 2.015$. The observed value $t = 2.875$ is larger than this, so we reject $H_0 : \alpha = 0$ in favor of $H_1 : \alpha \neq 0$.

27.7 a The data are modeled by a simple linear regression model: $Y_i = \alpha + \beta x_i$, where Y_i is the gas consumption and x_i is the average outside temperature in the ith week. Higher gas consumption as a consequence of smaller temperatures corresponds to $\beta < 0$. It is natural to consider the value 0 as the a priori specified value of the parameter (it corresponds to no change of gas consumption). Therefore, we take null hypothesis $H_0 : \beta = 0$. The alternative hypothesis should express the belief that the gas consumption increases as a consequence of smaller temperatures. Hence, we take alternative hypothesis $H_1 : \beta < 0$. The value of the test statistic is

$$t_b = \frac{\hat{\beta}}{s_b} = \frac{-0.3932}{0.0196} = -20.06.$$

The test statistic T_b has a t-distribution with $n - 2 = 24$ degrees of freedom. The value -20.06 is smaller than the left critical value $t_{24,0.05} = -1.711$, so we reject.

27.7 b For the data after insulation, the value of the test statistic is

$$t_b = \frac{-0.2779}{0.0252} = -11.03,$$

and T_b has a $t(28)$ distribution. The value -11.03 is smaller than the left critical value $t_{28,0.05} = -1.701$, so we reject.

28.1 a $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$; The value of the test statistic is $t_p = -2.130$. The critical values are $\pm t_{142,0.025}$. These are not in Table **??**, but $1.96 < t_{142,0.025} < 2.009$, so that we reject the null hypothesis.

28.1 b The value of the test statistic is the *same* (see Exercise 28.4). The normal approximation yields that the critical values are ± 1.96. The observed value $t_d = -2.130$ is smaller than the left critical value -1.96, so that we reject the null hypothesis.

28.1 c The observed value $t_d = -2.130$ is smaller than the left critical value -2.004, so that we reject the null hypothesis. The salaries differ significantly.

28.2 First consider testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. In view of the observed sample variances, there is no reason to assume equal variances, so compute

$$s_d^2 = \frac{s_x^2}{n} + \frac{s_y^2}{m} = \frac{7.77}{775} + \frac{25.33}{261} = 0.1071,$$

and

$$t_d = \frac{\bar{x}_n - \bar{y}_m}{s_d} = \frac{39.08 - 37.59}{\sqrt{0.1071}} = 4.553.$$

With sample sizes $n_1 = 775$ and $n_2 = 261$ we can use the normal approximation. The p-value is $\mathrm{P}(T_d \geq 4.553) \approx 0$, so that we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_1 : \mu_1 \neq \mu_2$. The other testing problems are handled in the same way. For testing $H_0 : \mu_1 = \mu_3$ against $H_1 : \mu_1 \neq \mu_3$, we find

$$s_d^2 = \frac{7.77}{775} + \frac{4.95}{633} = 0.0178 \quad \text{and} \quad t_d = \frac{39.08 - 39.60}{\sqrt{0.0178}} = -3.898$$

with p-value $\mathrm{P}(T_d \leq -3.898) \approx 0$, so that we reject $H_0 : \mu_1 = \mu_3$ in favor of $H_1 : \mu_1 \neq \mu_3$.
For testing $H_0 : \mu_2 = \mu_3$ against $H_1 : \mu_2 \neq \mu_3$, we find

$$s_d^2 = \frac{25.33}{261} + \frac{4.95}{633} = 0.1049 \quad \text{and} \quad t_d = \frac{37.59 - 39.60}{\sqrt{0.1049}} = -6.206$$

with p-value $\mathrm{P}(T_d \leq -6.206) \approx 0$, so that we reject $H_0 : \mu_2 = \mu_3$ in favor of $H_1 : \mu_2 \neq \mu_3$.

28.3 a The value of the test statistic is

$$t_p = \frac{22.43 - 11.01}{4.58} = 2.492.$$

Under the assumption of normal data with equal variances, we must compare this with the right critical value $t_{43,0.025}$. This is not Table **??**, but $2.009 < t_{43,0.025} < 2.021$. Hence, $t_p > t_{43,0.025}$, so that we reject the null hypothesis.

28.3 b The value $t_p = 2.492$ is greater than the right critical value 1.959, so that we reject on the basis of the bootstrap simulation.

28.3 c The value of the test statistic is

$$t_d = \frac{22.43 - 11.01}{4.64} = 2.463.$$

Without the assumption of equal variances, and using the normal approximation, we must compare this with the right critical value $z_{0.025} = 1.96$. Since $t_p > 1.96$, we reject the null hypothesis.

28.3 d Because we test at level 0.05, we reject if the right tail probability corresponding to $t_d = 2.463$ is smaller than 0.025. Since this is the case, we reject on the basis of the bootstrap simulation.

28.4 When $n = m$, then

$$\begin{aligned} S_p^2 &= \frac{(n-1)S_X^2 + (n-1)S_Y^2}{n+n-2}\left(\frac{1}{n}+\frac{1}{n}\right) \\ &= \frac{(n-1)(S_X^2+S_Y^2)}{2(n-1)}\frac{2}{n} = \frac{S_X^2}{n} + \frac{S_Y^2}{n} = S_d^2. \end{aligned}$$

28.5 a When $aS_X^2 + bS_Y^2$ is unbiased for σ^2, we should have $\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = \sigma^2$. Using that S_X^2 and S_Y^2 are both unbiased for σ^2, i.e., $\mathrm{E}\left[S_X^2\right] = \sigma^2$ and $\mathrm{E}\left[S_Y^2\right] = \sigma^2$, we get

$$\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = a\mathrm{E}\left[S_X^2\right] + b\mathrm{E}\left[S_Y^2\right] = (a+b)\sigma^2.$$

Hence, $\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = \sigma^2$ for all $\sigma > 0$ if and only if $a + b = 1$.

28.5 b By independence of S_X^2 and S_Y^2 write

$$\begin{aligned} \mathrm{Var}\big(aS_X^2 + (1-a)S_Y^2\big) &= a^2\mathrm{Var}\big(S_X^2\big) + (1-a)^2\mathrm{Var}\big(S_Y^2\big) \\ &= \left(\frac{a^2}{n-1} + \frac{(1-a)^2}{m-1}\right)2\sigma^4. \end{aligned}$$

To find the value of a that minimizes this, differentiate with respect to a and put the derivative equal to zero. This leads to

$$\frac{2a}{n-1} - \frac{2(1-a)}{m-1} = 0.$$

Solving for a yields $a = (n-1)/(n+m-2)$. Note that the second derivative of $\mathrm{Var}\big(aS_X^2 + (1-a)S_Y^2\big)$ is positive so that this is indeed a minimum.

28.6 a By independence of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ we have

$$\mathrm{Var}\big(\bar{X}_n - \bar{Y}_m\big) = \mathrm{Var}\big(\bar{X}_n\big) + \mathrm{Var}\big(\bar{Y}_m\big) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

28.6 b

$$\begin{aligned} \mathrm{E}\left[S_p^2\right] &= \frac{(n-1)\mathrm{E}\left[S_X^2\right] + (m-1)\mathrm{E}\left[S_Y^2\right]}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right) \\ &= \frac{(n-1)\sigma_X^2 + (m-1)\sigma_Y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right). \end{aligned}$$

In principle this may differ from $\mathrm{Var}\big(\bar{X}_n - \bar{Y}_m\big) = \sigma_X^2/n + \sigma_Y^2/m$.

28.6 c First note that

$$\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = a\mathrm{E}\left[S_X^2\right] + b\mathrm{E}\left[S_Y^2\right] = a\sigma_X^2 + b\sigma_Y^2.$$

For unbiasedness this must equal $\sigma_X^2/n + \sigma_Y^2/m$ for all $\sigma_X > 0$ and $\sigma_Y > 0$, with a, b not depending on σ_X and σ_Y. This is only possible for $a = 1/n$ and $b = 1/m$.

28.6 d From part **a** together with $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, we have

$$\begin{aligned}\mathrm{E}\left[S_p^2\right] &= \frac{(n-1)\sigma_X^2 + (m-1)\sigma_Y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)\\ &= \sigma^2 \cdot \frac{(n-1)+(m-1)}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right) = \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right).\end{aligned}$$

28.6 e No, not in general, see part **b**. If $n = m$, then according to (the computations in) Exercise 28.4, $S_p^2 = S_d^2$. Since, according to part **c**, S_d^2 is always an unbiased estimator for $\mathrm{Var}(\bar{X}_n - \bar{Y}_m)$ it follows that S_p^2 is also an unbiased estimator for $\mathrm{Var}(\bar{X}_n - \bar{Y}_m)$. One may also check this as follows:

$$\begin{aligned}\mathrm{E}\left[S_p^2\right] &= \frac{(n-1)\sigma_X^2 + (m-1)\sigma_Y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)\\ &= \frac{(n-1)\sigma_X^2 + (n-1)\sigma_Y^2}{n+n-2}\left(\frac{1}{n} + \frac{1}{n}\right)\\ &= (\sigma_X^2 + \sigma_Y^2) \cdot \frac{n-1}{n+n-2}\left(\frac{1}{n} + \frac{1}{n}\right) = (\sigma_X^2 + \sigma_Y^2)\frac{1}{n} = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}.\end{aligned}$$