1

Events and their probabilities

1.2 Solutions. Events as sets

1. (a) Let $a \in (\bigcup A_i)^c$. Then $a \notin \bigcup A_i$, so that $a \in A_i^c$ for all i. Hence $(\bigcup A_i)^c \subseteq \bigcap A_i^c$. Conversely, if $a \in \bigcap A_i^c$, then $a \notin A_i$ for every i. Hence $a \notin \bigcup A_i$, and so $\bigcap A_i^c \subseteq (\bigcup A_i)^c$. The first De Morgan law follows.

(b) Applying part (a) to the family $\{A_i^c : i \in I\}$, we obtain that $\left(\bigcup_i A_i^c\right)^c = \bigcap_i (A_i^c)^c = \bigcap_i A_i$. Taking the complement of each side yields the second law.

2. Clearly
(i) $A \cap B = (A^c \cup B^c)^c$,
(ii) $A \setminus B = A \cap B^c = (A^c \cup B)^c$,
(iii) $A \triangle B = (A \setminus B) \cup (B \setminus A) = (A^c \cup B)^c \cup (A \cup B^c)^c$.
Now $\mathcal{F}$ is closed under the operations of countable unions and complements, and therefore each of these sets lies in $\mathcal{F}$.

3. Let us number the players $1, 2, \dots, 2^n$ in the order in which they appear in the initial table of draws. The set of victors in the first round is a point in the space $V_n = \{1, 2\} \times \{3, 4\} \times \cdots \times \{2^n - 1, 2^n\}$. Renumbering these victors in the same way as done for the initial draw, the set of second-round victors can be thought of as a point in the space V_{n-1}, and so on. The sample space of all possible outcomes of the tournament may therefore be taken to be $V_n \times V_{n-1} \times \cdots \times V_1$, a set containing $2^{2^{n-1}} 2^{2^{n-2}} \cdots 2^1 = 2^{2^n - 1}$ points.

Should we be interested in the ultimate winner only, we may take as sample space the set $\{1, 2, \dots, 2^n\}$ of all possible winners.

4. We must check that $\mathcal{G}$ satisfies the definition of a σ-field:
(a) $\varnothing \in \mathcal{F}$, and therefore $\varnothing = \varnothing \cap B \in \mathcal{G}$,
(b) if $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_i (A_i \cap B) = \left(\bigcup_i A_i\right) \cap B \in \mathcal{G}$,
(c) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ so that $B \setminus (A \cap B) = A^c \cap B \in \mathcal{G}$.
Note that $\mathcal{G}$ is a σ-field *of subsets of* B but not a σ-field of subsets of Ω, since $C \in \mathcal{G}$ does not imply that $C^c = \Omega \setminus C \in \mathcal{G}$.

5. (a), (b), and (d) are identically true; (c) is true if and only if $A \subseteq C$.

1.3 Solutions. Probability

1. (i) We have (using the fact that $\mathbb{P}$ is a non-decreasing set function) that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1 = \tfrac{1}{12}.$$

Also, since $A \cap B \subseteq A$ and $A \cap B \subseteq B$, $\mathbb{P}(A \cap B) \le \min\{\mathbb{P}(A), \mathbb{P}(B)\} = \frac{1}{3}$.

These bounds are attained in the following example. Pick a number at random from $\{1, 2, \dots, 12\}$. Taking $A = \{1, 2, \dots, 9\}$ and $B = \{9, 10, 11, 12\}$, we find that $A \cap B = \{9\}$, and so $\mathbb{P}(A) = \frac{3}{4}$, $\mathbb{P}(B) = \frac{1}{3}$, $\mathbb{P}(A \cap B) = \frac{1}{12}$. To attain the upper bound for $\mathbb{P}(A \cap B)$, take $A = \{1, 2, \dots, 9\}$ and $B = \{1, 2, 3, 4\}$.

(ii) Likewise we have in this case $\mathbb{P}(A \cup B) \le \min\{\mathbb{P}(A) + \mathbb{P}(B), 1\} = 1$, and $\mathbb{P}(A \cup B) \ge \max\{\mathbb{P}(A), \mathbb{P}(B)\} = \frac{3}{4}$. These bounds are attained in the examples above.

2. (i) We have (using the continuity property of $\mathbb{P}$) that

$$\mathbb{P}(\text{no head ever}) = \lim_{n\to\infty} \mathbb{P}(\text{no head in first } n \text{ tosses}) = \lim_{n\to\infty} 2^{-n} = 0,$$

so that $\mathbb{P}(\text{some head turns up}) = 1 - \mathbb{P}(\text{no head ever}) = 1$.

(ii) Given a fixed sequence s of heads and tails of length k, we consider the sequence of tosses arranged in disjoint groups of consecutive outcomes, each group being of length k. There is probability 2^{-k} that any given one of these is s, independently of the others. The event {one of the first n such groups is s} is a subset of the event {s occurs in the first nk tosses}. Hence (using the general properties of probability measures) we have that

$$\begin{aligned}
\mathbb{P}(s \text{ turns up eventually}) &= \lim_{n\to\infty} \mathbb{P}(s \text{ occurs in the first } nk \text{ tosses}) \\
&\ge \lim_{n\to\infty} \mathbb{P}(s \text{ occurs as one of the first } n \text{ groups}) \\
&= 1 - \lim_{n\to\infty} \mathbb{P}(\text{none of the first } n \text{ groups is } s) \\
&= 1 - \lim_{n\to\infty} (1 - 2^{-k})^n = 1.
\end{aligned}$$

3. Lay out the saucers in order, say as RRWWSS. The cups may be arranged in $6!$ ways, but since each pair of a given colour may be switched without changing the appearance, there are $6! \div (2!)^3 = 90$ distinct arrangements. By assumption these are equally likely. In how many such arrangements is no cup on a saucer of the same colour? The only acceptable arrangements in which cups of the same colour are paired off are WWSSRR and SSRRWW; by inspection, there are a further eight arrangements in which the first pair of cups is either SW or WS, the second pair is either RS or SR, and the third either RW or WR. Hence the required probability is $10/90 = \frac{1}{9}$.

4. We prove this by induction on n, considering first the case $n = 2$. Certainly $B = (A \cap B) \cup (B \setminus A)$ is a union of disjoint sets, so that $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$. Similarly $A \cup B = A \cup (B \setminus A)$, and so

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \big\{\mathbb{P}(B) - \mathbb{P}(A \cap B)\big\}.$$

Hence the result is true for $n = 2$. Let $m \ge 2$ and suppose that the result is true for $n \le m$. Then it is true for pairs of events, so that

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_1^{m+1} A_i\Big) &= \mathbb{P}\Big(\bigcup_1^{m} A_i\Big) + \mathbb{P}(A_{m+1}) - \mathbb{P}\Big\{\Big(\bigcup_1^{m} A_i\Big) \cap A_{m+1}\Big\} \\
&= \mathbb{P}\Big(\bigcup_1^{m} A_i\Big) + \mathbb{P}(A_{m+1}) - \mathbb{P}\Big\{\bigcup_1^{m} (A_i \cap A_{m+1})\Big\}.
\end{aligned}$$

Using the induction hypothesis, we may expand the two relevant terms on the right-hand side to obtain the result.

Let A_1, A_2, and A_3 be the respective events that you fail to obtain the ultimate, penultimate, and ante-penultimate Vice-Chancellors. Then the required probability is, by symmetry,

$$\begin{aligned}1 - \mathbb{P}\Big(\bigcup_1^3 A_i\Big) &= 1 - 3\mathbb{P}(A_1) + 3\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_2 \cap A_3)\\ &= 1 - 3(\tfrac{4}{5})^6 + 3(\tfrac{3}{5})^6 - (\tfrac{2}{5})^6.\end{aligned}$$

5. By the continuity of $\mathbb{P}$, Exercise (1.2.1), and Problem (1.8.11),

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{r=1}^{\infty} A_r\Big) &= \lim_{n\to\infty} \mathbb{P}\Big(\bigcap_{r=1}^{n} A_r\Big) = \lim_{n\to\infty}\Big[1 - \mathbb{P}\Big(\Big(\bigcap_{r=1}^{n} A_r\Big)^{\mathrm{c}}\Big)\Big]\\ &= 1 - \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{r=1}^{n} A_r^{\mathrm{c}}\Big) \ge 1 - \lim_{n\to\infty}\sum_{r=1}^{n}\mathbb{P}(A_r^{\mathrm{c}}) = 1.\end{aligned}$$

6. We have that $1 = \mathbb{P}\Big(\bigcup_1^n A_r\Big) = \sum_r \mathbb{P}(A_r) - \sum_{r<s}\mathbb{P}(A_r \cap A_s) = np - \frac{1}{2}n(n-1)q$. Hence $p \ge n^{-1}$, and $\frac{1}{2}n(n-1)q = np - 1 \le n - 1$.

7. Since at least one of the A_r occurs,

$$\begin{aligned}1 = \mathbb{P}\Big(\bigcup_1^n A_r\Big) &= \sum_r \mathbb{P}(A_r) - \sum_{r<s}\mathbb{P}(A_r \cap A_s) + \sum_{r<s<t}\mathbb{P}(A_r \cap A_s \cap A_t)\\ &= np - \binom{n}{2}q + \binom{n}{3}x.\end{aligned}$$

Since at least two of the events occur with probability $\frac{1}{2}$,

$$\tfrac{1}{2} = \mathbb{P}\Big(\bigcup_{r<s}(A_r \cap A_s)\Big) = \sum_{r<s}\mathbb{P}(A_r \cap A_s) - \tfrac{1}{2}\sum_{\substack{r<s\\ t<u\\ (r,s)\ne(t,u)}} \mathbb{P}(A_r \cap A_s \cap A_t \cap A_u) + \cdots.$$

By a careful consideration of the first three terms in the latter series, we find that

$$\frac{1}{2} = \binom{n}{2}q - 3\binom{n}{3}x + \binom{n}{3}x.$$

Hence $\frac{3}{2} = np - \binom{n}{3}x$, so that $p \ge 3/(2n)$. Also, $\binom{n}{2}q = 2np - \frac{5}{2}$, whence $q \le 4/n$.

1.4 Solutions. Conditional probability

1. By the definition of conditional probability,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}\frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \mathbb{P}(B \mid A)\frac{\mathbb{P}(A)}{\mathbb{P}(B)}$$

if $\mathbb{P}(A)\mathbb{P}(B) \neq 0$. Hence

$$\frac{\mathbb{P}(A \mid B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B \mid A)}{\mathbb{P}(B)},$$

whence the last part is immediate.

2. Set $A_0 = \Omega$ for notational convenience. Expand each term on the right-hand side to obtain

$$\prod_{r=1}^{n} \mathbb{P}\left(A_r \,\Big|\, \bigcap_{k=1}^{r-1} A_k\right) = \prod_{r=1}^{n} \frac{\mathbb{P}(\bigcap_1^r A_k)}{\mathbb{P}(\bigcap_1^{r-1} A_k)} = \mathbb{P}\left(\bigcap_1^n A_k\right).$$

3. Let M be the event that the first coin is double-headed, R the event that it is double-tailed, and N the event that it is normal. Let H_l^i be the event that the lower face is a head on the ith toss, T_u^i the event that the upper face is a tail on the ith toss, and so on. Then, using conditional probability *ad nauseam*, we find:

(i) $$\mathbb{P}(H_l^1) = \tfrac{2}{5}\mathbb{P}(H_l^1 \mid M) + \tfrac{1}{5}\mathbb{P}(H_l^1 \mid R) + \tfrac{2}{5}\mathbb{P}(H_l^1 \mid N) = \tfrac{2}{5} + 0 + \tfrac{2}{5}\cdot\tfrac{1}{2} = \tfrac{3}{5}.$$

(ii) $$\mathbb{P}(H_l^1 \mid H_u^1) = \frac{\mathbb{P}(H_l^1 \cap H_u^1)}{\mathbb{P}(H_u^1)} = \frac{\mathbb{P}(M)}{\mathbb{P}(H_l^1)} = \tfrac{2}{5}\big/\tfrac{3}{5} = \tfrac{2}{3}.$$

(iii) $$\begin{aligned}\mathbb{P}(H_l^2 \mid H_u^1) &= 1\cdot\mathbb{P}(M \mid H_u^1) + \tfrac{1}{2}\mathbb{P}(N \mid H_u^1) \\ &= \mathbb{P}(H_l^1 \mid H_u^1) + \tfrac{1}{2}(1 - \mathbb{P}(H_l^1 \mid H_u^1)) = \tfrac{2}{3} + \tfrac{1}{2}\cdot\tfrac{1}{3} = \tfrac{5}{6}.\end{aligned}$$

(iv)

$$\mathbb{P}(H_l^2 \mid H_u^1 \cap H_u^2) = \frac{\mathbb{P}(H_l^2 \cap H_u^1 \cap H_u^2)}{\mathbb{P}(H_u^1 \cap H_u^2)} = \frac{\mathbb{P}(M)}{1\cdot\mathbb{P}(M) + \frac{1}{4}\cdot\mathbb{P}(N)} = \frac{\frac{2}{5}}{\frac{2}{5} + \frac{1}{10}} = \frac{4}{5}.$$

(v) From (iv), the probability that he discards a double-headed coin is $\frac{4}{5}$, the probability that he discards a normal coin is $\frac{1}{5}$. (There is of course no chance of it being double-tailed.) Hence, by conditioning on the discard,

$$\mathbb{P}(H_u^3) = \tfrac{4}{5}\mathbb{P}(H_u^3 \mid M) + \tfrac{1}{5}\mathbb{P}(H_u^3 \mid N) = \tfrac{4}{5}\left(\tfrac{1}{4} + \tfrac{1}{2}\cdot\tfrac{1}{2}\right) + \tfrac{1}{5}\left(\tfrac{1}{2} + \tfrac{1}{2}\cdot\tfrac{1}{4}\right) = \tfrac{21}{40}.$$

4. The final calculation of $\frac{2}{3}$ refers not to a *single* draw of one ball from an urn containing three, but rather to a composite experiment comprising more than one stage (in this case, two stages). While it is true that {two black, one white} is the only fixed collection of balls for which a random choice is black with probability $\frac{2}{3}$, the composition of the urn is *not determined* prior to the final draw.

After all, if Carroll's argument were correct then it would apply also in the situation when the urn originally contains just one ball, either black or white. The final probability is now $\frac{3}{4}$, implying that the original ball was one half black and one half white! Carroll was himself aware of the fallacy in this argument.

5. (a) One cannot compute probabilities without knowing the rules governing the conditional probabilities. If the first door chosen conceals a goat, then the presenter has no choice in the door to be opened, since exactly one of the remaining doors conceals a goat. If the first door conceals the car, then a choice is necessary, and this is governed by the protocol of the presenter. Consider two 'extremal' protocols for this latter situation.

(i) The presenter opens a door chosen at random from the two available.

(ii) There is some ordering of the doors (left to right, perhaps) and the presenter opens the earlier door in this ordering which conceals a goat.

Analysis of the two situations yields $p = \frac{2}{3}$ under (i), and $p = \frac{1}{2}$ under (ii).

Let $\alpha \in [\frac{1}{2}, \frac{2}{3}]$, and suppose the presenter possesses a coin which falls with heads upwards with probability $\beta = 6\alpha - 3$. He flips the coin before the show, and adopts strategy (i) if and only if the coin shows heads. The probability in question is now $\frac{2}{3}\beta + \frac{1}{2}(1 - \beta) = \alpha$.

You never lose by swapping, but whether you gain depends on the presenter's protocol.

(b) Let D denote the first door chosen, and consider the following protocols:

(iii) If D conceals a goat, open it. Otherwise open one of the other two doors at random. In this case $p = 0$.

(iv) If D conceals the car, open it. Otherwise open the unique remaining door which conceals a goat. In this case $p = 1$.

As in part (a), a randomized algorithm provides the protocol necessary for the last part.

6. This is immediate by the definition of conditional probability.

7. Let C_i be the colour of the ith ball picked, and use the obvious notation.

(a) Since each urn contains the same number $n - 1$ of balls, the second ball picked is equally likely to be any of the $n(n - 1)$ available. One half of these balls are magenta, whence $\mathbb{P}(C_2 = \mathrm{M}) = \frac{1}{2}$.

(b) By conditioning on the choice of urn,

$$\mathbb{P}(C_2 = \mathrm{M} \mid C_1 = \mathrm{M}) = \frac{\mathbb{P}(C_1, C_2 = \mathrm{M})}{\mathbb{P}(C_1 = \mathrm{M})} = \sum_{r=1}^{n} \frac{(n-r)(n-r-1)}{n(n-1)(n-2)} \Big/ \frac{1}{2} = \frac{2}{3}.$$

1.5 Solutions. Independence

1. Clearly

$$\begin{aligned}\mathbb{P}(A^{\mathrm{c}} \cap B) &= \mathbb{P}(B \setminus \{A \cap B\}) = \mathbb{P}(B) - \mathbb{P}(A \cap B)\\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A^{\mathrm{c}})\mathbb{P}(B).\end{aligned}$$

For the final part, apply the first part to the pair B, A^{c}.

2. Suppose $i < j$ and $m < n$. If $j < m$, then A_{ij} and A_{mn} are determined by distinct independent rolls, and are therefore independent. For the case $j = m$ we have that

$$\begin{aligned}\mathbb{P}(A_{ij} \cap A_{jn}) &= \mathbb{P}(i\text{th}, j\text{th, and } n\text{th rolls show same number})\\ &= \sum_{r=1}^{6} \tfrac{1}{6}\mathbb{P}\big(j\text{th and } n\text{th rolls both show } r \,\big|\, i\text{th shows } r\big) = \tfrac{1}{36} = \mathbb{P}(A_{ij})\mathbb{P}(A_{jn}),\end{aligned}$$

as required. However, if $i \neq j \neq k$,

$$\mathbb{P}(A_{ij} \cap A_{jk} \cap A_{ik}) = \tfrac{1}{36} \neq \tfrac{1}{216} = \mathbb{P}(A_{ij})\mathbb{P}(A_{jk})\mathbb{P}(A_{ik}).$$

3. That (a) implies (b) is trivial. Suppose then that (b) holds. Consider the outcomes numbered $i_1, i_2, \dots, i_m$, and let $u_j \in \{\mathrm{H}, \mathrm{T}\}$ for $1 \leq j \leq m$. Let S_j be the set of all sequences of length $M = \max\{i_j : 1 \leq j \leq m\}$ showing u_j in the i_jth position. Clearly $|S_j| = 2^{M-1}$ and $\left|\bigcap_j S_j\right| = 2^{M-m}$. Therefore,

$$\mathbb{P}(S_j) = \frac{2^{M-1}}{2^M} = \frac{1}{2}, \qquad \mathbb{P}\Big(\bigcap_j S_j\Big) = \frac{2^{M-m}}{2^M} = \frac{1}{2^m},$$

so that $\mathbb{P}\left(\bigcap_j S_j\right) = \prod_j \mathbb{P}(S_j)$.

4. Suppose $|A| = a$, $|B| = b$, $|A \cap B| = c$, and A and B are independent. Then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, which is to say that $c/p = (a/p) \cdot (b/p)$, and hence $ab = pc$. If $ab \neq 0$ then $p \mid ab$ (i.e., p divides ab). However, p is prime, and hence either $p \mid a$ or $p \mid b$. Therefore, either $A = \Omega$ or $B = \Omega$ (or both).

5. (a) Flip two coins; let A be the event that the first shows H, let B be the event that the second shows H, and let C be the event that they show the same. Then A and B are independent, but not conditionally independent given C.

(b) Roll two dice; let A be the event that the smaller is 3, let B be the event that the larger is 6, and let C be the event that the smaller score is no more than 3, and the larger is 4 or more. Then A and B are conditionally independent given C, but not independent.

(c) The definitions are equivalent if $\mathbb{P}(C) = 1$.

6. $(\frac{9}{10})^7 < \frac{1}{2}$.

7. (a) $\mathbb{P}(A \cap B) = \frac{1}{8} = \frac{1}{4} \cdot \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)$, and $\mathbb{P}(B \cap C) = \frac{3}{8} = \frac{1}{2} \cdot \frac{3}{4} = \mathbb{P}(B)\mathbb{P}(C)$.

(b) $\mathbb{P}(A \cap C) = 0 \neq \mathbb{P}(A)\mathbb{P}(C)$.

(c) Only in the trivial cases when children are either almost surely boys or almost surely girls.

(d) No.

8. No. $\mathbb{P}(\text{all alike}) = \frac{1}{4}$.

9. $\mathbb{P}(\text{1st shows } r \text{ and sum is 7}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(\text{1st shows } r)\mathbb{P}(\text{sum is 7})$.

1.7 Solutions. Worked examples

1. Write EF for the event that there is an open road from E to F, and EF^c for the complement of this event; write $\mathrm{E} \leftrightarrow \mathrm{F}$ if there is an open route from E to F, and $\mathrm{E} \not\leftrightarrow \mathrm{F}$ if there is none. Now $\{\mathrm{A} \leftrightarrow \mathrm{C}\} = \mathrm{AB} \cap \mathrm{BC}$, so that

$$\mathbb{P}(\mathrm{AB} \mid \mathrm{A} \not\leftrightarrow \mathrm{C}) = \frac{\mathbb{P}(\mathrm{AB}, \mathrm{A} \not\leftrightarrow \mathrm{C})}{\mathbb{P}(\mathrm{A} \not\leftrightarrow \mathrm{C})} = \frac{\mathbb{P}(\mathrm{AB}, \mathrm{B} \not\leftrightarrow \mathrm{C})}{1 - \mathbb{P}(\mathrm{A} \leftrightarrow \mathrm{C})} = \frac{(1-p^2)p^2}{1-(1-p^2)^2}.$$

By a similar calculation (or otherwise) in the second case, one obtains the same answer:

$$\mathbb{P}(\mathrm{AB} \mid \mathrm{A} \not\leftrightarrow \mathrm{C}) = \frac{(1-p^2)p^3}{1-(1-p^2)^2p-(1-p)} = \frac{(1-p^2)p^2}{1-(1-p^2)^2}.$$

2. Let A be the event of exactly one ace, and KK be the event of exactly two kings. Then $\mathbb{P}(\mathrm{A} \mid \mathrm{KK}) = \mathbb{P}(\mathrm{A} \cap \mathrm{KK})/\mathbb{P}(\mathrm{KK})$. Now, by counting acceptable combinations,

$$\mathbb{P}(\mathrm{A} \cap \mathrm{KK}) = \binom{4}{1}\binom{4}{2}\binom{44}{10}\Big/\binom{52}{13}, \quad \mathbb{P}(\mathrm{KK}) = \binom{4}{2}\binom{48}{11}\Big/\binom{52}{13},$$

so the required probability is

$$\binom{4}{1}\binom{4}{2}\binom{44}{10}\Big/\binom{4}{2}\binom{48}{11} = \frac{7 \cdot 11 \cdot 37}{3 \cdot 46 \cdot 47} \simeq 0.44.$$

3. *First method*: Suppose that the coin is being tossed by a special machine which is not switched off when the walker is absorbed. If the machine ever produces N heads in succession, then either the game finishes at this point or it is already over. From Exercise (1.3.2), such a sequence of N heads must (with probability one) occur sooner or later.

Alternative method: Write down the difference equations for p_k, the probability the game finishes at 0 having started at k, and for $\widehat{p}_k$, the corresponding probability that the game finishes at N; actually these two difference equations are the same, but the respective boundary conditions are different. Solve these equations and add their solutions to obtain the total 1.

4. It is a tricky question. One of the present authors is in agreement, since if $\mathbb{P}(A \mid C) > \mathbb{P}(B \mid C)$ and $\mathbb{P}(A \mid C^c) > \mathbb{P}(B \mid C^c)$ then

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid C)\mathbb{P}(C) + \mathbb{P}(A \mid C^c)\mathbb{P}(C^c)\\ &> \mathbb{P}(B \mid C)\mathbb{P}(C) + \mathbb{P}(B \mid C^c)\mathbb{P}(C^c) = \mathbb{P}(B).\end{aligned}$$

The other author is more suspicious of the question, and points out that there is a difficulty arising from the use of the word 'you'. In Example (1.7.10), Simpson's paradox, whilst drug I is preferable to drug II for both males and females, it is drug II that wins overall.

5. Let L_k be the label of the kth card. Then, using symmetry,

$$\mathbb{P}\big(L_k = m \,\big|\, L_k > L_r \text{ for } 1 \le r < k\big) = \frac{\mathbb{P}(L_k = m)}{\mathbb{P}(L_k > L_r \text{ for } 1 \le r < k)} = \frac{1}{m}\Big/\frac{1}{k} = k/m.$$

1.8 Solutions to problems

1. (a) *Method I*: There are 36 equally likely outcomes, and just 10 of these contain exactly one six. The answer is therefore $\frac{10}{36} = \frac{5}{18}$.

Method II: Since the throws have independent outcomes,

$$\mathbb{P}(\text{first is 6, second is not 6}) = \mathbb{P}(\text{first is 6})\mathbb{P}(\text{second is not 6}) = \tfrac{1}{6}\cdot\tfrac{5}{6} = \tfrac{5}{36}.$$

There is an equal probability of the event {first is not 6, second is 6}.

(b) A die shows an odd number with probability $\frac{1}{2}$; by independence, $\mathbb{P}(\text{both odd}) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$.

(c) Write S for the sum, and $\{i, j\}$ for the event that the first is i and the second j. Then $\mathbb{P}(S = 4) = \mathbb{P}(1,3) + \mathbb{P}(2,2) + \mathbb{P}(3,1) = \frac{3}{36}$.

(d) Similarly

$$\begin{aligned}\mathbb{P}(S \text{ divisible by } 3) &= \mathbb{P}(S=3) + \mathbb{P}(S=6) + \mathbb{P}(S=9) + \mathbb{P}(S=12)\\ &= \{\mathbb{P}(1,2) + \mathbb{P}(2,1)\}\\ &\quad + \{\mathbb{P}(1,5) + \mathbb{P}(2,4) + \mathbb{P}(3,3) + \mathbb{P}(4,2) + \mathbb{P}(5,1)\}\\ &\quad + \{\mathbb{P}(3,6) + \mathbb{P}(4,5) + \mathbb{P}(5,4) + \mathbb{P}(6,3)\} + \mathbb{P}(6,6)\\ &= \tfrac{12}{36} = \tfrac{1}{3}.\end{aligned}$$

2. (a) By independence, $\mathbb{P}(n-1 \text{ tails, followed by a head}) = 2^{-n}$.

(b) If n is odd, $\mathbb{P}(\#\text{ heads} = \#\text{ tails}) = 0$; $\#A$ denotes the cardinality of the set A. If n is even, there are $\binom{n}{n/2}$ sequences of outcomes with $\frac{1}{2}n$ heads and $\frac{1}{2}n$ tails. Any given sequence of heads and tails has probability 2^{-n}; therefore $\mathbb{P}(\#\text{ heads} = \#\text{ tails}) = 2^{-n}\binom{n}{n/2}$.

(c) There are $\binom{n}{2}$ sequences containing 2 heads and $n-2$ tails. Each sequence has probability 2^{-n}, and therefore $\mathbb{P}(\text{exactly two heads}) = \binom{n}{2}2^{-n}$.
(d) Clearly

$$\mathbb{P}(\text{at least 2 heads}) = 1 - \mathbb{P}(\text{no heads}) - \mathbb{P}(\text{exactly one head}) = 1 - 2^{-n} - \binom{n}{1}2^{-n}.$$

3. (a) Recall De Morgan's Law (Exercise (1.2.1)): $\bigcap_i A_i = \left(\bigcup_i A_i^c\right)^c$, which lies in $\mathcal{F}$ since it is the complement of a countable union of complements of sets in $\mathcal{F}$.
(b) $\mathcal{H}$ is a σ-field because:
(i) $\varnothing \in \mathcal{F}$ and $\varnothing \in \mathcal{G}$; therefore $\varnothing \in \mathcal{H}$.
(ii) If $A_1, A_2, \dots$ is a sequence of sets belonging to both $\mathcal{F}$ and $\mathcal{G}$, then their union lies in both $\mathcal{F}$ and $\mathcal{G}$, which is to say that $\mathcal{H}$ is closed under the operation of taking countable unions.
(iii) Likewise A^c is in $\mathcal{H}$ if A is in both $\mathcal{F}$ and $\mathcal{G}$.
(c) We display an example. Let

$$\Omega = \{a, b, c\}, \quad \mathcal{F} = \big\{\{a\}, \{b, c\}, \varnothing, \Omega\big\}, \quad \mathcal{G} = \big\{\{a, b\}, \{c\}, \varnothing, \Omega\big\}.$$

Then $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$ is given by $\mathcal{H} = \big\{\{a\}, \{c\}, \{a, b\}, \{b, c\}, \varnothing, \Omega\big\}$. Note that $\{a\} \in \mathcal{H}$ and $\{c\} \in \mathcal{H}$, but the union $\{a, c\}$ is not in $\mathcal{H}$, which is therefore not a σ-field.

4. In each case $\mathcal{F}$ may be taken to be the set of all subsets of Ω, and the probability of any member of $\mathcal{F}$ is the sum of the probabilities of the elements therein.
(a) $\Omega = \{\mathrm{H}, \mathrm{T}\}^3$, the set of all triples of heads (H) and tails (T). With the usual assumption of independence, the probability of any given triple containing h heads and $t = 3-h$ tails is $p^h(1-p)^t$, where p is the probability of heads on each throw.
(b) In the obvious notation, $\Omega = \{\mathrm{U}, \mathrm{V}\}^2 = \{\mathrm{UU}, \mathrm{VV}, \mathrm{UV}, \mathrm{VU}\}$. Also $\mathbb{P}(\mathrm{UU}) = \mathbb{P}(\mathrm{VV}) = \frac{2}{4}\cdot\frac{1}{3}$ and $\mathbb{P}(\mathrm{UV}) = \mathbb{P}(\mathrm{VU}) = \frac{2}{4}\cdot\frac{2}{3}$.
(c) Ω is the set of finite sequences of tails followed by a head, $\{\mathrm{T}^n\mathrm{H} : n \geq 0\}$, together with the infinite sequence T^∞ of tails. Now, $\mathbb{P}(\mathrm{T}^n\mathrm{H}) = (1-p)^n p$, and $\mathbb{P}(\mathrm{T}^\infty) = \lim_{n\to\infty}(1-p)^n = 0$ if $p \neq 0$.

5. As usual, $\mathbb{P}(A \triangle B) = \mathbb{P}\big((A \cup B) \setminus \mathbb{P}(A \cap B)\big) = \mathbb{P}(A \cup B) - \mathbb{P}(A \cap B)$.

6. Clearly, by Exercise (1.4.2),

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}\big((A^c \cap B^c \cap C^c)^c\big) = 1 - \mathbb{P}(A^c \cap B^c \cap C^c) \\ &= 1 - \mathbb{P}(A^c \mid B^c \cap C^c)\mathbb{P}(B^c \mid C^c)\mathbb{P}(C^c).\end{aligned}$$

7. (a) If A is independent of itself, then $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, so that $\mathbb{P}(A) = 0$ or 1.
(b) If $\mathbb{P}(A) = 0$ then $0 = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all B. If $\mathbb{P}(A) = 1$ then $\mathbb{P}(A \cap B) = \mathbb{P}(B)$, so that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

8. $\Omega \cup \varnothing = \Omega$ and $\Omega \cap \varnothing = \varnothing$, and therefore $1 = \mathbb{P}(\Omega \cup \varnothing) = \mathbb{P}(\Omega) + \mathbb{P}(\varnothing) = 1 + \mathbb{P}(\varnothing)$, implying that $\mathbb{P}(\varnothing) = 0$.

9. (i) $\mathbb{Q}(\varnothing) = \mathbb{P}(\varnothing \mid B) = 0$. Also $\mathbb{Q}(\Omega) = \mathbb{P}(\Omega \mid B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.
(ii) Let $A_1, A_2, \dots$ be disjoint members of $\mathcal{F}$. Then $\{A_i \cap B : i \geq 1\}$ are disjoint members of $\mathcal{F}$, implying that

$$\mathbb{Q}\left(\bigcup_1^\infty A_i\right) = \mathbb{P}\left(\bigcup_1^\infty A_i \,\middle|\, B\right) = \frac{\mathbb{P}\big(\bigcup_1^\infty (A_i \cap B)\big)}{\mathbb{P}(B)} = \sum_1^\infty \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_1^\infty \mathbb{Q}(A_i).$$

Finally, since $\mathbb{Q}$ is a probability measure,

$$\mathbb{Q}(A \mid C) = \frac{\mathbb{Q}(A \cap C)}{\mathbb{Q}(C)} = \frac{\mathbb{P}(A \cap C \mid B)}{\mathbb{P}(C \mid B)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \mathbb{P}(A \mid B \cap C).$$

The order of the conditioning (C before B, or *vice versa*) is thus irrelevant.

10. As usual,

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_1^\infty (A \cap B_j)\right) = \sum_1^\infty \mathbb{P}(A \cap B_j) = \sum_1^\infty \mathbb{P}(A \mid B_j)\mathbb{P}(B_j).$$

11. The first inequality is trivially true if $n = 1$. Let $m \geq 1$ and assume that the inequality holds for $n \leq m$. Then

$$\begin{aligned}\mathbb{P}\left(\bigcup_1^{m+1} A_i\right) &= \mathbb{P}\left(\bigcup_1^m A_i\right) + \mathbb{P}(A_{m+1}) - \mathbb{P}\left(\bigcup_1^m (A_i \cap A_{m+1})\right)\\ &\leq \mathbb{P}\left(\bigcup_1^m A_i\right) + \mathbb{P}(A_{m+1}) \leq \sum_1^{m+1} \mathbb{P}(A_i),\end{aligned}$$

by the hypothesis. The result follows by induction. Secondly, by the first part,

$$\mathbb{P}\left(\bigcap_1^n A_i\right) = \mathbb{P}\left(\left(\bigcup_1^n A_i^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_1^n A_i^c\right) \geq 1 - \sum_1^n \mathbb{P}(A_i^c).$$

12. We have that

$$\begin{aligned}\mathbb{P}\left(\bigcap_1^n A_i\right) &= \mathbb{P}\left(\left(\bigcup_1^n A_i^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_1^n A_i^c\right)\\ &= 1 - \sum_i \mathbb{P}(A_i^c) + \sum_{i<j} \mathbb{P}(A_i^c \cap A_j^c) - \cdots + (-1)^n \mathbb{P}\left(\bigcap_1^n A_i^c\right) \quad \text{by Exercise (1.3.4)}\\ &= 1 - n + \sum_i \mathbb{P}(A_i) + \binom{n}{2} - \sum_{i<j} \mathbb{P}(A_i \cup A_j) - \binom{n}{3} + \cdots\\ &\qquad + (-1)^n \binom{n}{n} - (-1)^n \mathbb{P}\left(\bigcup_1^n A_i\right) \quad \text{using De Morgan's laws again}\\ &= (1-1)^n + \sum_i \mathbb{P}(A_i) - \cdots - (-1)^n \mathbb{P}\left(\bigcup_1^n A_i\right) \quad \text{by the binomial theorem.}\end{aligned}$$

13. Clearly,

$$\mathbb{P}(N_k) = \sum_{\substack{S \subseteq \{1,2,\ldots,n\}\\ |S|=k}} \mathbb{P}\left(\bigcap_{i \in S} A_i \bigcap_{j \notin S} A_j^c\right).$$

For any such given S, we write $A_S = \bigcap_{i \in S} A_i$. Then

$$\mathbb{P}\left(\bigcap_{i \in S} A_i \bigcap_{j \notin S} A_j^c\right) = \mathbb{P}(A_S) - \sum_{j \notin S} \mathbb{P}(A_{S \cup \{j\}}) + \sum_{\substack{j<k\\ j,k \notin S}} \mathbb{P}(A_{S \cup \{j,k\}}) - \cdots$$

by Exercise (1.3.4). Hence

$$\mathbb{P}(N_k) = \sum_{|S|=k} \mathbb{P}(A_S) - \sum_{|S|=k+1} \binom{k+1}{k}\mathbb{P}(A_S) + \cdots + (-1)^{n-k}\binom{n}{k}\mathbb{P}(A_1 \cap \cdots \cap A_n)$$

where a typical summation is over all subsets S of $\{1, 2, \ldots, n\}$ having the required cardinality.

Let A_i be the event that a copy of the ith bust is obtained. Then, by symmetry,

$$\mathbb{P}(N_3) = \binom{5}{3}\alpha_3 - \binom{5}{4}\binom{4}{3}\alpha_4 + \binom{5}{3}\alpha_5$$

where α_j is the probability that the j most recent Vice-Chancellors are obtained. Now α_3 is given in Exercise (1.3.4), and α_4 and α_5 may be calculated similarly.

14. Assuming the conditional probabilities are defined,

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\mathbb{P}\big(B \cap (\bigcup_1^n A_i)\big)} = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\sum_{i=1}^n \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

15. (a) We have that

$$\begin{aligned}\mathbb{P}(N=2 \mid S=4) &= \frac{\mathbb{P}(\{N=2\} \cap \{S=4\})}{\mathbb{P}(S=4)} = \frac{\mathbb{P}(S=4 \mid N=2)\mathbb{P}(N=2)}{\sum_i \mathbb{P}(S=4 \mid N=i)\mathbb{P}(N=i)} \\ &= \frac{\frac{1}{12}\cdot\frac{1}{4}}{\frac{1}{6}\cdot\frac{1}{2} + \frac{1}{12}\cdot\frac{1}{4} + \frac{3}{216}\cdot\frac{1}{8} + \frac{1}{6^4}\cdot\frac{1}{16}}.\end{aligned}$$

(b) Secondly,

$$\begin{aligned}\mathbb{P}(S=4 \mid N \text{ even}) &= \frac{\mathbb{P}(S=4 \mid N=2)\frac{1}{4} + \mathbb{P}(S=4 \mid N=4)\frac{1}{16}}{\mathbb{P}(N \text{ even})} \\ &= \frac{\frac{1}{12}\cdot\frac{1}{4} + \frac{1}{6^4}\cdot\frac{1}{16}}{\frac{1}{4} + \frac{1}{16} + \cdots} = \frac{4^2 3^3 + 1}{4^4 3^3}.\end{aligned}$$

(c) Writing D for the number shown by the first die,

$$\mathbb{P}(N=2 \mid S=4, D=1) = \frac{\mathbb{P}(N=2, S=4, D=1)}{\mathbb{P}(S=4, D=1)} = \frac{\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{4}}{\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{4} + \frac{1}{6}\cdot\frac{2}{36}\cdot\frac{1}{8} + \frac{1}{6^4}\cdot\frac{1}{16}}.$$

(d) Writing M for the maximum number shown, if $1 \le r \le 6$,

$$\mathbb{P}(M \le r) = \sum_{j=1}^{\infty} \mathbb{P}(M \le r \mid N=j)2^{-j} = \sum_{j=1}^{\infty}\left(\frac{r}{6}\right)^j \frac{1}{2^j} = \frac{r}{12}\left(1 - \frac{r}{12}\right)^{-1} = \frac{r}{12-r}.$$

Finally, $\mathbb{P}(M=r) = \mathbb{P}(M \le r) - \mathbb{P}(M \le r-1)$.

16. (a) $\omega \in B$ if and only if, for all n, $\omega \in \bigcup_{i=n}^{\infty} A_i$, which is to say that ω belongs to infinitely many of the A_n.

(b) $\omega \in C$ if and only if, for some n, $\omega \in \bigcap_{i=n}^{\infty} A_i$, which is to say that ω belongs to all but a finite number of the A_n.

(c) It suffices to note that B is a countable intersection of countable unions of events, and is therefore an event.
(d) We have that

$$C_n = \bigcap_{i=n}^{\infty} A_i \subseteq A_n \subseteq \bigcup_{i=n}^{\infty} A_i = B_n,$$

and therefore $\mathbb{P}(C_n) \le \mathbb{P}(A_n) \le \mathbb{P}(B_n)$. By the continuity of probability measures (1.3.5), if $C_n \to C$ then $\mathbb{P}(C_n) \to \mathbb{P}(C)$, and if $B_n \to B$ then $\mathbb{P}(B_n) \to \mathbb{P}(B)$. If $B = C = A$ then

$$\mathbb{P}(A) = \mathbb{P}(C) \le \lim_{n\to\infty} \mathbb{P}(A_n) \le \mathbb{P}(C) = \mathbb{P}(A).$$

17. If B_n and C_n are independent for all n then, using the fact that $C_n \subseteq B_n$,

$$\mathbb{P}(B_n)\mathbb{P}(C_n) = \mathbb{P}(B_n \cap C_n) = \mathbb{P}(C_n) \to \mathbb{P}(C) \qquad \text{as } n \to \infty,$$

and also $\mathbb{P}(B_n)\mathbb{P}(C_n) \to \mathbb{P}(B)\mathbb{P}(C)$ as $n \to \infty$, so that $\mathbb{P}(C) = \mathbb{P}(B)\mathbb{P}(C)$, whence either $\mathbb{P}(C) = 0$ or $\mathbb{P}(B) = 1$ or both. In any case $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$.

If $A_n \to A$ then $A = B = C$ so that $\mathbb{P}(A)$ equals 0 or 1.

18. It is standard (Lemma (1.3.5)) that $\mathbb{P}$ is continuous if it is countably additive. Suppose then that $\mathbb{P}$ is finitely additive and continuous. Let $A_1, A_2, \dots$ be disjoint events. Then $\bigcup_1^\infty A_i = \lim_{n\to\infty} \bigcup_1^n A_i$, so that, by continuity and finite-additivity,

$$\mathbb{P}\left(\bigcup_1^\infty A_i\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_1^n A_i\right) = \lim_{n\to\infty} \sum_1^n \mathbb{P}(A_i) = \sum_1^\infty \mathbb{P}(A_i).$$

19. The network of friendship is best represented as a square with diagonals, with the corners labelled A, B, C, and D. Draw a diagram. Each link of the network is absent with probability p. We write EF for the event that a typical link EF is present, and $\mathrm{EF^c}$ for its complement. We write $\mathrm{A} \leftrightarrow \mathrm{D}$ for the event that A is connected to D by present links.

(d) $$\begin{aligned}\mathbb{P}(\mathrm{A \leftrightarrow D \mid AD^c}) &= \mathbb{P}(\mathrm{A \leftrightarrow D \mid AD^c \cap BC^c})p + \mathbb{P}(\mathrm{A \leftrightarrow D \mid AD^c \cap BC})(1-p)\\ &= \{1-(1-(1-p)^2)^2\}p + (1-p^2)^2(1-p).\end{aligned}$$

(c) $$\begin{aligned}\mathbb{P}(\mathrm{A \leftrightarrow D \mid BC^c}) &= \mathbb{P}(\mathrm{A \leftrightarrow D \mid AD^c \cap BC^c})p + \mathbb{P}(\mathrm{A \leftrightarrow D \mid BC^c \cap AD})(1-p)\\ &= \{1-(1-(1-p)^2)^2\}p + (1-p).\end{aligned}$$

(b) $$\begin{aligned}\mathbb{P}(\mathrm{A \leftrightarrow D \mid AB^c}) &= \mathbb{P}(\mathrm{A \leftrightarrow D \mid AB^c \cap AD^c})p + \mathbb{P}(\mathrm{A \leftrightarrow D \mid AB^c \cap AD})(1-p)\\ &= (1-p)\{1-p(1-(1-p)^2)\}p + (1-p).\end{aligned}$$

(a) $$\begin{aligned}\mathbb{P}(\mathrm{A \leftrightarrow D}) &= \mathbb{P}(\mathrm{A \leftrightarrow D \mid AD^c})p + \mathbb{P}(\mathrm{A \leftrightarrow D \mid AD})(1-p)\\ &= \{1-(1-(1-p)^2)^2\}p^2 + (1-p^2)^2p(1-p) + (1-p).\end{aligned}$$

20. We condition on the result of the first toss. If this is a head, then we require an odd number of heads in the next $n-1$ tosses. Similarly, if the first toss is a tail, we require an even number of heads in the next $n-1$ tosses. Hence

$$p_n = p(1-p_{n-1}) + (1-p)p_{n-1} = (1-2p)p_{n-1} + p$$

with $p_0 = 1$. As an alternative to induction, we may seek a solution of the form $p_n = A + B\lambda^n$. Substitute this into the above equation to obtain

$$A + B\lambda^n = (1-2p)A + (1-2p)B\lambda^{n-1} + p$$

and $A + B = 1$. Hence $A = \frac{1}{2}$, $B = \frac{1}{2}$, $\lambda = 1 - 2p$.

21. Let $A = \{\text{run of } r \text{ heads precedes run of } s \text{ tails}\}$, $B = \{\text{first toss is a head}\}$, and $C = \{\text{first } s \text{ tosses are tails}\}$. Then

$$\mathbb{P}(A \mid B^c) = \mathbb{P}(A \mid B^c \cap C)\mathbb{P}(C \mid B^c) + \mathbb{P}(A \mid B^c \cap C^c)\mathbb{P}(C^c \mid B^c) = 0 + \mathbb{P}(A \mid B)(1 - q^{s-1}),$$

where $p = 1 - q$ is the probability of heads on any single toss. Similarly $\mathbb{P}(A \mid B) = p^{r-1} + \mathbb{P}(A \mid B^c)(1 - p^{r-1})$. We solve for $\mathbb{P}(A \mid B)$ and $\mathbb{P}(A \mid B^c)$, and use the fact that $\mathbb{P}(A) = \mathbb{P}(A \mid B)p + \mathbb{P}(A \mid B^c)q$, to obtain

$$\mathbb{P}(A) = \frac{p^{r-1}(1 - q^s)}{p^{r-1} + q^{s-1} - p^{r-1}q^{s-1}}.$$

22. (a) Since every cherry has the same chance to be this cherry, notwithstanding the fact that five are now in the pig, the probability that the cherry in question contains a stone is $\frac{5}{20} = \frac{1}{4}$.
(b) Think about it the other way round. *First* a random stone is removed, and *then* the pig chooses his fruit. This does not change the relevant probabilities. Let C be the event that the removed cherry contains a stone, and let P be the event that the pig gets at least one stone. Then $\mathbb{P}(P \mid C)$ is the probability that out of 19 cherries, 15 of which are stoned, the pig gets a stone. Therefore

$$\mathbb{P}(P \mid C) = 1 - \mathbb{P}\big(\text{pig chooses only stoned cherries} \mid C\big) = 1 - \tfrac{15}{19} \cdot \tfrac{14}{18} \cdot \tfrac{13}{17} \cdot \tfrac{12}{16} \cdot \tfrac{11}{15}.$$

23. Label the seats $1, 2, \dots, 2n$ clockwise. For the sake of definiteness, we dictate that seat 1 be occupied by a woman; this determines the sex of the occupant of every other seat. For $1 \le k \le 2n$, let A_k be the event that seats k, $k + 1$ are occupied by one of the couples (we identify seat $2n + 1$ with seat 1). The required probability is

$$\mathbb{P}\Big(\bigcap_1^{2n} A_i^c\Big) = 1 - \mathbb{P}\Big(\bigcup_1^{2n} A_i\Big) = 1 - \sum_i \mathbb{P}(A_i) + \sum_{i<j} \mathbb{P}(A_i \cap A_j) - \cdots.$$

Now, $\mathbb{P}(A_i) = n(n-1)!^2/n!^2$, since there are n couples who may occupy seats i and $i + 1$, $(n-1)!$ ways of distributing the remaining $n - 1$ women, and $(n-1)!$ ways of distributing the remaining $n - 1$ men. Similarly, if $1 \le i < j \le 2n$, then

$$\mathbb{P}(A_i \cap A_j) = \begin{cases} n(n-1)\dfrac{(n-2)!^2}{n!^2} & \text{if } |i - j| \ne 1 \\ 0 & \text{if } |i - j| = 1, \end{cases}$$

subject to $\mathbb{P}(A_1 \cap A_{2n}) = 0$. In general,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \frac{n!}{(n-k)!} \frac{(n-k)!^2}{n!^2} = \frac{(n-k)!}{n!}$$

if $i_1 < i_2 < \cdots < i_k$ and $i_{j+1} - i_j \ge 2$ for $1 \le j < k$, and $2n + i_1 - i_k \ge 2$; otherwise this probability is 0. Hence

$$\mathbb{P}\Big(\bigcap_1^{2n} A_i^c\Big) = \sum_{k=0}^{n} (-1)^k \frac{(n-k)!}{n!} S_{k,n}$$

where $S_{k,n}$ is the number of ways of choosing k non-overlapping pairs of adjacent seats.

Finally, we calculate $S_{k,n}$. Consider first the number $N_{k,m}$ of ways of picking k non-overlapping pairs of adjacent seats from a line (rather than a circle) of m seats labelled $1, 2, \dots, m$. There is a one–one correspondence between the set of such arrangements and the set of $(m - k)$–vectors containing

k 1's and $(m-2k)$ 0's. To see this, take such an arrangement of seats, and count 0 for an unchosen seat and 1 for a chosen pair of seats; the result is such a vector. Conversely take such a vector, read its elements in order, and construct the arrangement of seats in which each 0 corresponds to an unchosen seat and each 1 corresponds to a chosen pair. It follows that $N_{k,m} = \binom{m-k}{k}$.

Turning to $S_{k,n}$, either the pair $2n$, 1 is chosen or it is not. If it is chosen, we require another $k-1$ pairs out of a line of $2n-2$ seats. If it is not chosen, we require k pairs out of a line of $2n$ seats. Therefore

$$S_{k,n} = N_{k-1,2n-2} + N_{k,2n} = \binom{2n-k-1}{k-1} + \binom{2n-k}{k} = \binom{2n-k}{k}\frac{2n}{2n-k}.$$

24. Think about the experiment as laying down the $b+r$ balls from left to right in a random order. The number of possible orderings equals the number of ways of placing the blue balls, namely $\binom{b+r}{b}$. The number of ways of placing the balls so that the first k are blue, and the next red, is the number of ways of placing the red balls so that the first is in position $k+1$ and the remainder are amongst the $r+b-k-1$ places to the right, namely $\binom{r+b-k-1}{r-1}$. The required result follows.

The probability that the last ball is red is $r/(r+b)$, the same as the chance of being red for the ball in any other given position in the ordering.

25. We argue by induction on the total number of balls in the urn. Let p_{ac} be the probability that the last ball is azure, and suppose that $p_{ac} = \frac{1}{2}$ whenever $a, c \geq 1$, $a+c \leq k$. Let α and σ be such that $\alpha, \sigma \geq 1$, $\alpha + \sigma = k+1$. Let A_i be the event that i azure balls are drawn before the first carmine ball, and let C_j be the event that j carmine balls are drawn before the first azure ball. We have, by taking conditional probabilities and using the induction hypothesis, that

$$\begin{aligned} p_{\alpha\sigma} &= \sum_{i=1}^{\alpha} p_{\alpha-i,\sigma}\mathbb{P}(A_i) + \sum_{j=1}^{\sigma} p_{\alpha,\sigma-j}\mathbb{P}(C_j) \\ &= p_{0,\sigma}\mathbb{P}(A_\alpha) + p_{\alpha,0}\mathbb{P}(C_\sigma) + \tfrac{1}{2}\sum_{i=1}^{\alpha-1}\mathbb{P}(A_i) + \tfrac{1}{2}\sum_{j=1}^{\sigma-1}\mathbb{P}(C_j). \end{aligned}$$

Now $p_{0,\sigma} = 0$ and $p_{\alpha,0} = 1$. Also, by an easy calculation,

$$\mathbb{P}(A_\alpha) = \frac{\alpha}{\alpha+\sigma}\cdot\frac{\alpha-1}{\alpha+\sigma-1}\cdots\frac{1}{\sigma+1} = \frac{\alpha!\,\sigma!}{(\alpha+\sigma)!} = \mathbb{P}(C_\sigma).$$

It follows from the above two equations that

$$p_{\alpha\sigma} = \tfrac{1}{2}\left(\sum_{i=1}^{\alpha}\mathbb{P}(A_i) + \sum_{j=1}^{\sigma}\mathbb{P}(C_j)\right) + \tfrac{1}{2}\big(\mathbb{P}(C_\sigma) - \mathbb{P}(A_\alpha)\big) = \tfrac{1}{2}.$$

26. (a) If she says the ace of hearts is present, then this imparts no information about the other card, which is equally likely to be any of the three other possibilities.

(b) In the given protocol, interchange hearts and diamonds.

27. Writing A if A tells the truth, and A^c otherwise, etc., the only outcomes consistent with D telling the truth are ABCD, AB^cC^cD, A^cBC^cD, and A^cB^cCD, with a total probability of $\frac{13}{81}$. Likewise, the only outcomes consistent with D lying are $A^cB^cC^cD^c$, A^cBCD^c, AB^cCD^c, and ABC^cD^c, with a total probability of $\frac{28}{81}$. Writing S for the given statement, we have that

$$\mathbb{P}(D \mid S) = \frac{\mathbb{P}(D\cap S)}{\mathbb{P}(D\cap S) + \mathbb{P}(D^c\cap S)} = \frac{\frac{13}{81}}{\frac{13}{81}+\frac{28}{81}} = \frac{13}{41}.$$

Eddington himself thought the answer to be $\frac{25}{71}$; hence the 'controversy'. He argued that a truthful denial leaves things unresolved, so that if, for example, B truthfully denies that C contradicts D, then we cannot deduce that C supports D. He deduced that the only sequences which are inconsistent with the given statement are $\mathrm{AB^cCD}$ and $\mathrm{AB^cC^cD^c}$, and therefore

$$\mathbb{P}(\mathrm{D} \mid \mathrm{S}) = \frac{\frac{25}{81}}{\frac{25}{81} + \frac{46}{81}} = \frac{25}{71}.$$

Which side are *you* on?

28. Let B_r be the event that the rth vertex of a randomly selected cube is blue, and note that $\mathbb{P}(B_r) = \frac{1}{10}$. By Boole's inequality,

$$\mathbb{P}\left(\bigcup_{r=1}^{8} B_r\right) \le \sum_{r=1}^{8} \mathbb{P}(B_r) = \tfrac{8}{10} < 1,$$

so at least 20 per cent of such cubes have only red vertices.

29. (a) $\mathbb{P}(B \mid A) = \mathbb{P}(A \cap B)/\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B)/\mathbb{P}(A) > \mathbb{P}(B)$.
(b) $\mathbb{P}(A \mid B^c) = \mathbb{P}(A \cap B^c)/\mathbb{P}(B^c) = \{\mathbb{P}(A) - \mathbb{P}(A \cap B)\}/\mathbb{P}(B^c) < \mathbb{P}(A)$.
(c) No. Consider the case $A \cap C = \varnothing$.

30. The number of possible combinations of birthdays of m people is 365^m; the number of combinations of different birthdays is $365!/(365-m)!$. Use your calculator for the final part.

31. (a) $\displaystyle \binom{n-r+1}{r} \bigg/ \binom{n}{r}$.

(b) $\displaystyle (r-1)\binom{n-r+1}{r-1} \bigg/ \binom{n}{r}$.

(c) $\displaystyle \frac{1}{r!}$.

(d) $\displaystyle 1 \bigg/ \binom{n}{r}$.

(e) $\displaystyle \binom{r}{k}\binom{n-r}{r-k} \bigg/ \binom{n}{r}$.

32. In the obvious notation, $\mathbb{P}(w\mathrm{S}, x\mathrm{H}, y\mathrm{D}, z\mathrm{C}) = \binom{13}{w}\binom{13}{x}\binom{13}{y}\binom{13}{z} \big/ \binom{52}{13}$. Now use your calculator. Turning to the 'shape vector' (w, x, y, z) with $w \ge x \ge y \ge z$,

$$\mathbb{P}(w, x, y, z) = \begin{cases} 4\mathbb{P}(w\mathrm{S}, x\mathrm{H}, y\mathrm{D}, z\mathrm{C}) & \text{if } w \ne x = y = z, \\ 12\mathbb{P}(w\mathrm{S}, x\mathrm{H}, y\mathrm{D}, z\mathrm{C}) & \text{if } w = x \ne y \ne z, \end{cases}$$

on counting the disjoint ways of obtaining the shapes in question.

33. Use your calculator, and divide each of the following by $\binom{52}{5}$.

$$\binom{13}{1}\binom{4}{3}\binom{12}{3}\binom{4}{1}^3, \quad \binom{13}{2}\binom{4}{2}^2\binom{11}{1}\binom{4}{1}, \quad \binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2,$$

$$10\binom{4}{1}^5 - 10\binom{4}{1}, \quad \binom{4}{1}\binom{13}{5}, \quad \binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2},$$

$$\binom{13}{1}\binom{4}{4}\binom{12}{1}\binom{4}{1}, \quad 10\binom{4}{1}.$$

34. Divide each of the following by 6^5.

$$\frac{6!\,5!}{3!\,(2!)^2},\quad \frac{6!\,5!}{3!\,(2!)^3},\quad \frac{6!\,5!}{2!\,(3!)^2},$$
$$6!,\quad \frac{6!\,5!}{4!\,3!\,2!},\quad \frac{6!\,5!}{(4!)^2},$$
$$\frac{6!\,5!}{(5!)^2}.$$

35. Let S_r denote the event that you receive r similar answers, and T the event that they are correct. Denote the event that your interlocutor is a tourist by V. Then $T \cap V^c = \varnothing$, and

$$\mathbb{P}(T \mid S_r) = \frac{\mathbb{P}(T \cap V \cap S_r)}{\mathbb{P}(S_r)} = \frac{\mathbb{P}(T \cap S_r \mid V)\mathbb{P}(V)}{\mathbb{P}(S_r)}.$$

Hence:

(a) $\mathbb{P}(T \mid S_1) = \frac{3}{4} \times \frac{2}{3}/1 = \frac{1}{2}$.

(b) $\mathbb{P}(T \mid S_2) = (\frac{3}{4})^2 \cdot \frac{2}{3}/[\{(\frac{3}{4})^2 + (\frac{1}{4})^2\}\frac{2}{3} + \frac{1}{3}] = \frac{1}{2}$.

(c) $\mathbb{P}(T \mid S_3) = (\frac{3}{4})^3 \cdot \frac{2}{3}/[\{(\frac{3}{4})^3 + (\frac{1}{4})^3\}\frac{2}{3} + \frac{1}{3}] = \frac{9}{20}$.

(d) $\mathbb{P}(T \mid S_4) = (\frac{3}{4})^4 \cdot \frac{2}{3}/[\{(\frac{3}{4})^4 + (\frac{1}{4})^4\}\frac{2}{3} + \frac{1}{3}] = \frac{27}{70}$.

(e) If the last answer differs, then the speaker is surely a tourist, so the required probability is

$$\frac{(\frac{3}{4})^3 \cdot \frac{1}{4}}{(\frac{3}{4})^3 \times \frac{1}{4} + (\frac{1}{4})^3 . \frac{3}{4}} = \frac{9}{10}.$$

36. Let E (respectively W) denote the event that the answer East (respectively West) is given.

(a) Using conditional probability,

$$\mathbb{P}(\text{East correct} \mid \text{E}) = \frac{\epsilon\mathbb{P}(\text{E} \mid \text{East correct})}{\mathbb{P}(\text{E})} = \frac{\epsilon \cdot \frac{2}{3}.\frac{3}{4}}{\frac{1}{2}\epsilon + (\frac{2}{3}.\frac{1}{4} + \frac{1}{3})(1+\epsilon)} = \epsilon,$$
$$\mathbb{P}(\text{East correct} \mid \text{W}) = \frac{\epsilon(\frac{2}{3}.\frac{1}{4} + \frac{1}{3})}{\epsilon(\frac{1}{6} + \frac{1}{3}) + \frac{2}{3}.\frac{3}{4}(1-\epsilon)} = \epsilon.$$

(b) Likewise, one obtains for the answer EE,

$$\frac{\epsilon \cdot \frac{2}{3}(\frac{3}{4})^2}{\epsilon \cdot \frac{2}{3}(\frac{3}{4})^2 + (1-\epsilon)(\frac{2}{3}(\frac{1}{4})^2 + \frac{1}{3})} = \epsilon,$$

and for the answer WW,

$$\frac{\epsilon(\frac{2}{3}(\frac{1}{4})^2 + \frac{1}{3})}{\epsilon \cdot \frac{3}{8} + (1-\epsilon)\frac{3}{8}} = \epsilon.$$

(c) Similarly for EEE,

$$\epsilon(\tfrac{2}{3})(\tfrac{3}{4})^2(\{\epsilon(\tfrac{2}{3})(\tfrac{3}{4})^3 + (1-\epsilon)(\tfrac{2}{3}(\tfrac{1}{4})3 + \tfrac{1}{3})\} = \frac{9\epsilon}{11 - 2\epsilon}$$

and for WWW,

$$\frac{\epsilon\{(\frac{2}{3})(\frac{1}{4})^3 + \frac{1}{3}\}}{\epsilon[(\frac{2}{3})(\frac{1}{4})^3 + \frac{1}{3}] + (1-\epsilon)\frac{2}{3}(\frac{3}{4})^3} = \frac{11\epsilon}{9+2\epsilon}.$$

Then for $\epsilon = \frac{9}{20}$, the first is $\frac{81}{202}$; the second is $\frac{1}{2}$, as you would expect if you look at Problem (1.8.35).

37. Use induction. The inductive step employs Boole's inequality and the fact that

$$\mathbb{P}\left(\bigcup_{r=1}^{n+1} A_r\right) = \mathbb{P}(A_{n+1}) + \mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) - \mathbb{P}\left(\bigcup_{r=1}^{n} (A_r \cap A_{n+1})\right).$$

38. We propose to prove by induction that

$$\mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) \le \sum_{r=1}^{n} \mathbb{P}(A_r) - \sum_{2\le r\le n} \mathbb{P}(A_r \cap A_1).$$

There is nothing special about the choice of A_1 in this inequality, which will therefore hold with any suffix k playing the role of the suffix 1. Kounias's inequality is then implied.

The above inequality holds trivially when $n = 1$. Assume that it holds for some value of n (≥ 1). We have that

$$\begin{aligned}\mathbb{P}\left(\bigcup_{r=1}^{n+1} A_r\right) &= \mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(A_{n+1} \cap \bigcup_{r=1}^{n} A_r\right) \\ &\le \sum_{r=1}^{n} \mathbb{P}(A_r) - \sum_{2\le r\le n} \mathbb{P}(A_r \cap A_1) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(A_{n+1} \cap \bigcup_{r=1}^{n} A_r\right) \\ &\le \sum_{r=1}^{n+1} \mathbb{P}(A_r) - \sum_{2\le r\le n+1} \mathbb{P}(A_r \cap A_1)\end{aligned}$$

since $\mathbb{P}(A_{n+1} \cap A_1) \le \mathbb{P}\big(A_{n+1} \cap \bigcup_{r=1}^{n} A_r\big)$.

39. We take $n \ge 2$. We may assume without loss of generality that the seats are labelled $1, 2, \dots, n$, and that the passengers are labelled by their seat assignments. Write F for the event that the last passenger finds his assigned seat to be free. Let K (≥ 2) be the seat taken by passenger 1, so that $\mathbb{P}(F) = (n-1)^{-1}\sum_{k=2}^{n} \alpha_k$ where $\alpha_k = \mathbb{P}(F \mid K = k)$. Note that $\alpha_n = 0$. Passengers $2, 3, \dots, K-1$ occupy their correct seats. Passenger K either occupies seat 1, in which case all subsequent passengers take their correct seats, or he occupies some seat L satisfying $L > K$. In the latter case, passengers $K+1, K+2, \dots, L-1$ are correctly seated. We obtain thus that

$$\alpha_k = \frac{1}{n-k+1}(1 + \alpha_{k+1} + \alpha_{k+2} + \cdots + \alpha_n), \qquad 2 \le k < n.$$

Therefore $\alpha_k = \frac{1}{2}$ for $2 \le k < n$, by induction, and so $\mathbb{P}(F) = \frac{1}{2}(n-2)/(n-1)$.

2

Random variables and their distributions

2.1 Solutions. Random variables

1. (i) If $a > 0$, $x \in \mathbb{R}$, then $\{\omega : aX(\omega) \le x\} = \{\omega : X(\omega) \le x/a\} \in \mathcal{F}$ since X is a random variable. If $a < 0$,

$$\{\omega : aX(\omega) \le x\} = \{\omega : X(\omega) \ge x/a\} = \left\{ \bigcup_{n\ge 1} \left\{ \omega : X(\omega) \le \frac{x}{a} - \frac{1}{n} \right\} \right\}^{\mathrm{c}}$$

which lies in $\mathcal{F}$ since it is the complement of a countable union of members of $\mathcal{F}$. If $a = 0$,

$$\{\omega : aX(\omega) \le x\} = \begin{cases} \varnothing & \text{if } x < 0, \\ \Omega & \text{if } x \ge 0; \end{cases}$$

in either case, the event lies in $\mathcal{F}$.

(ii) For $\omega \in \Omega$, $X(\omega) - X(\omega) = 0$, so that $X - X$ is the zero random variable (that this *is* a random variable follows from part (i) with $a = 0$). Similarly $X(\omega) + X(\omega) = 2X(\omega)$.

2. Set $Y = aX + b$. We have that

$$\mathbb{P}(Y \le y) = \begin{cases} \mathbb{P}\big(X \le (y-b)/a\big) = F\big((y-b)/a\big) & \text{if } a > 0, \\ \mathbb{P}\big(X \ge (y-b)/a\big) = 1 - \lim_{x \uparrow (y-b)/a} F(x) & \text{if } a < 0. \end{cases}$$

Finally, if $a = 0$, then $Y = b$, so that $\mathbb{P}(Y \le y)$ equals 0 if $b > y$ and 1 if $b \le y$.

3. Assume that any specified sequence of heads and tails with length n has probability 2^{-n}. There are exactly $\binom{n}{k}$ such sequences with k heads.

If heads occurs with probability p then, assuming the independence of outcomes, the probability of any given sequence of k heads and $n-k$ tails is $p^k(1-p)^{n-k}$. The answer is therefore $\binom{n}{k} p^k (1-p)^{n-k}$.

4. Write $H = \lambda F + (1-\lambda)G$. Then $\lim_{x\to-\infty} H(x) = 0$, $\lim_{x\to\infty} H(x) = 1$, and clearly H is non-decreasing and right-continuous. Therefore H is a distribution function.

5. The function $g(F(x))$ is a distribution function whenever g is continuous and non-decreasing on $[0, 1]$, with $g(0) = 0$, $g(1) = 1$. This is easy to check in each special case.

2.2 Solutions. The law of averages

1. Let p be the potentially embarrassed fraction of the population, and suppose that each sampled individual would truthfully answer "yes" with probability p independently of all other individuals. In the modified procedure, the chance that someone says yes is $p + \frac{1}{2}(1-p) = \frac{1}{2}(1+p)$. If the proportion of yes's is now ϕ, then $2\phi - 1$ is a decent estimate of p.

The advantage of the given procedure is that it allows individuals to answer "yes" without their being identified with certainty as having the embarrassing property.

2. Clearly $H_n + T_n = n$, so that $(H_n - T_n)/n = (2H_n/n) - 1$. Therefore

$$\mathbb{P}\left(2p - 1 - \epsilon \le \frac{1}{n}(H_n - T_n) \le 2p - 1 + \epsilon\right) = \mathbb{P}\left(\left|\frac{1}{n}H_n - p\right| \le \frac{\epsilon}{2}\right) \to 1$$

as $n \to \infty$, by the law of large numbers (2.2.1).

3. Let $I_n(x)$ be the indicator function of the event $\{X_n \le x\}$. By the law of averages, $n^{-1}\sum_{r=1}^{n} I_r(x)$ converges in the sense of (2.2.1) and (2.2.6) to $\mathbb{P}(X_n \le x) = F(x)$.

2.3 Solutions. Discrete and continuous variables

1. With $\delta = \sup_m |a_m - a_{m-1}|$, we have that

$$|F(x) - G(x)| \le |F(a_m) - F(a_{m-1})| \le |F(x+\delta) - F(x-\delta)|$$

for $x \in [a_{m-1}, a_m)$. Hence $G(x)$ approaches $F(x)$ for any x at which F is continuous.

2. For y lying in the range of g, $\{Y \le y\} = \{X \le g^{-1}(y)\} \in \mathcal{F}$.

3. Certainly Y is a random variable, using the result of the previous exercise (2). Also

$$\mathbb{P}(Y \le y) = \mathbb{P}\big(F^{-1}(X) \le y\big) = \mathbb{P}\big(X \le F(y)\big) = F(y)$$

as required. If F is discontinuous then $F^{-1}(x)$ is not defined for all x, so that Y is not well defined. If F is non-decreasing and continuous, but not strictly increasing, then $F^{-1}(x)$ is not always defined uniquely. Such difficulties may be circumvented by defining $F^{-1}(x) = \inf\{y : F(y) \ge x\}$.

4. The function $\lambda f + (1-\lambda)g$ is non-negative and integrable over $\mathbb{R}$ to 1. Finally, fg is not necessarily a density, though it may be: e.g., if $f = g = 1$, $0 \le x \le 1$ then $f(x)g(x) = 1$, $0 \le x \le 1$.

5. (a) If $d > 1$, then $\int_1^\infty cx^{-d}\,dx = c/(d-1)$. Therefore f is a density function if $c = d - 1$, and $F(x) = 1 - x^{-(d-1)}$ when this holds. If $d \le 1$, then f has infinite integral and cannot therefore be a density function.

(b) By differentiating $F(x) = e^x/(1+e^x)$, we see that F is the distribution function, and $c = 1$.

2.4 Solutions. Worked examples

1. (a) If $y \ge 0$,

$$\mathbb{P}(X^2 \le y) = \mathbb{P}(X \le \sqrt{y}) - \mathbb{P}(X < -\sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}).$$

(b) We must assume that $X \ge 0$. If $y \ge 0$,

$$\mathbb{P}(\sqrt{X} \le y) = \mathbb{P}(0 \le X \le y^2) = F(y^2).$$

(c) If $-1 \le y \le 1$,

$$\begin{aligned}\mathbb{P}(\sin X \le y) &= \sum_{n=-\infty}^{\infty} \mathbb{P}\big((2n+1)\pi - \sin^{-1} y \le X \le (2n+2)\pi + \sin^{-1} y\big)\\ &= \sum_{n=-\infty}^{\infty} \Big\{F\big((2n+2)\pi + \sin^{-1} y\big) - F\big((2n+1)\pi - \sin^{-1} y\big)\Big\}.\end{aligned}$$

(d) $\mathbb{P}(G^{-1}(X) \le y) = \mathbb{P}(X \le G(y)) = F(G(y))$.
(e) If $0 \le y \le 1$, then $\mathbb{P}(F(X) \le y) = \mathbb{P}(X \le F^{-1}(y)) = F(F^{-1}(y)) = y$. There is a small difficulty if F is not *strictly* increasing, but this is overcome by defining $F^{-1}(y) = \sup\{x : F(x) = y\}$.
(f) $\mathbb{P}(G^{-1}(F(X)) \le y) = \mathbb{P}(F(X) \le G(y)) = G(y)$.

2. It is the case that, for $x \in \mathbb{R}$, $F_Y(x)$ and $F_Z(x)$ approach $F(x)$ as $a \to -\infty$, $b \to \infty$.

2.5 Solutions. Random vectors

1. Write $f_{xw} = \mathbb{P}(X = x, W = w)$. Then $f_{00} = f_{21} = \frac{1}{4}$, $f_{10} = \frac{1}{2}$, and $f_{xw} = 0$ for other pairs x, w.

2. (a) We have that

$$f_{X,Y}(x,y) = \begin{cases} p & \text{if } (x,y) = (1,0),\\ 1-p & \text{if } (x,y) = (0,1),\\ 0 & \text{otherwise.}\end{cases}$$

(b) Secondly,

$$f_{X,Z}(x,z) = \begin{cases} 1-p & \text{if } (x,z) = (0,0),\\ p & \text{if } (x,z) = (1,0),\\ 0 & \text{otherwise.}\end{cases}$$

3. Differentiating gives $f_{X,Y}(x,y) = e^{-x}/\{\pi(1+y^2)\}$, $x \ge 0$, $y \in \mathbb{R}$.

4. Let $A = \{X \le b,\ c < Y \le d\}$, $B = \{a < X \le b,\ Y \le d\}$. Clearly

$$\mathbb{P}(A) = F(b,d) - F(b,c), \quad \mathbb{P}(B) = F(b,d) - F(a,d), \quad \mathbb{P}(A \cup B) = F(b,d) - F(a,c);$$

now $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$, which gives the answer. Draw a map of $\mathbb{R}^2$ and plot the regions of values of (X, Y) involved.

5. The given expression equals

$$\mathbb{P}(X = x,\ Y \le y) - \mathbb{P}(X = x,\ Y \le y-1) = \mathbb{P}(X = x,\ Y = y).$$

Secondly, for $1 \le x \le y \le 6$,

$$f(x,y) = \begin{cases} \left(\dfrac{y-x+1}{6}\right)^r - 2\left(\dfrac{y-x}{6}\right)^r + \left(\dfrac{y-x-1}{6}\right)^r & \text{if } x < y,\\[2ex] \left(\dfrac{1}{6}\right)^r & \text{if } x = y.\end{cases}$$

6. No, because F is twice differentiable with $\partial^2 F/\partial x \partial y < 0$.

2.7 Solutions to problems

1. By the independence of the tosses,

$$\mathbb{P}(X > m) = \mathbb{P}(\text{first } m \text{ tosses are tails}) = (1-p)^m.$$

Hence

$$\mathbb{P}(X \le x) = \begin{cases} 1-(1-p)^{\lfloor x \rfloor} & \text{if } x \ge 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Remember that $\lfloor x \rfloor$ denotes the integer part of x.

2. (a) If X takes values $\{x_i : i \ge 1\}$ then $X = \sum_{i=1}^{\infty} x_i I_{A_i}$ where $A_i = \{X = x_i\}$.

(b) Partition the real line into intervals of the form $[k2^{-m}, (k+1)2^{-m})$, $-\infty < k < \infty$, and define $X_m = \sum_{k=-\infty}^{\infty} k2^{-m} I_{k,m}$ where $I_{k,m}$ is the indicator function of the event $\{k2^{-m} \le X < (k+1)2^{-m}\}$. Clearly X_m is a random variable, and $X_m(\omega) \uparrow X(\omega)$ as $m \to \infty$ for all ω.

(c) Suppose $\{X_m\}$ is a sequence of random variables such that $X_m(\omega) \uparrow X(\omega)$ for all ω. Then $\{X \le x\} = \bigcap_m \{X_m \le x\}$, which is a countable intersection of events and therefore lies in $\mathcal{F}$.

3. (a) We have that

$$\{X + Y \le x\} = \bigcap_{n=1}^{\infty} \bigcup_{r \in \mathbb{Q}^+} \left(\{X \le r\} \cap \{Y \le x - r + n^{-1}\}\right)$$

where $\mathbb{Q}^+$ is the set of positive rationals.

In the second case, if XY is a positive function, then $XY = \exp\{\log X + \log Y\}$; now use Exercise (2.3.2) and the above. For the general case, note first that $|Z|$ is a random variable whenever Z is a random variable, since $\{|Z| \le a\} = \{Z \le a\} \setminus \{Z < -a\}$ for $a \ge 0$. Now, if $a \ge 0$, then $\{XY \le a\} = \{XY < 0\} \cup \{|XY| \le a\}$ and

$$\{XY < 0\} = \big(\{X < 0\} \cap \{Y > 0\}\big) \cup \big(\{X > 0\} \cap \{Y < 0\}\big).$$

Similar relations are valid if $a < 0$.

Finally $\{\min\{X, Y\} > x\} = \{X > x\} \cap \{Y > y\}$, the intersection of events.

(b) It is enough to check that $\alpha X + \beta Y$ is a random variable whenever $\alpha, \beta \in \mathbb{R}$ and X, Y are random variables. This follows from the argument above.

If Ω is finite, we may take as basis the set $\{I_A : A \in \mathcal{F}\}$ of all indicator functions of events.

4. (a) $F(\frac{3}{2}) - F(\frac{1}{2}) = \frac{1}{2}$.

(b) $F(2) - F(1) = \frac{1}{2}$.

(c) $\mathbb{P}(X^2 \le X) = \mathbb{P}(X \le 1) = \frac{1}{2}$.

(d) $\mathbb{P}(X \le 2X^2) = \mathbb{P}(X \ge \frac{1}{2}) = \frac{3}{4}$.

(e) $\mathbb{P}(X + X^2 \le \frac{3}{4}) = \mathbb{P}(X \le \frac{1}{2}) = \frac{1}{4}$.

(f) $\mathbb{P}(\sqrt{X} \le z) = \mathbb{P}(X \le z^2) = \frac{1}{2}z^2$ if $0 \le z \le \sqrt{2}$.

5. $\mathbb{P}(X = -1) = 1 - p$, $\mathbb{P}(X = 0) = 0$, $\mathbb{P}(X \ge 1) = \frac{1}{2}p$.

6. There are 6 intervals of 5 minutes, preceding the arrival times of buses. Each such interval has probability $\frac{5}{60} = \frac{1}{12}$, so the answer is $6 \cdot \frac{1}{12} = \frac{1}{2}$.

7. Let T and B be the numbers of people on given typical flights of TWA and BA. From Exercise (2.1.3),

$$\mathbb{P}(T=k)=\binom{10}{k}\left(\frac{9}{10}\right)^k\left(\frac{1}{10}\right)^{10-k}, \quad \mathbb{P}(B=k)=\binom{20}{k}\left(\frac{9}{10}\right)^k\left(\frac{1}{10}\right)^{20-k}.$$

Now

$$\mathbb{P}(\text{TWA overbooked})=\mathbb{P}(T=10)=(\tfrac{9}{10})^{10},$$
$$\mathbb{P}(\text{BA overbooked})=\mathbb{P}(B\ge 19)=20(\tfrac{9}{10})^{19}(\tfrac{1}{10})+(\tfrac{9}{10})^{20},$$

of which the latter is the larger.

8. Assuming the coins are fair, the chance of getting at least five heads is $(\frac{1}{2})^6+6(\frac{1}{2})^6=\frac{7}{64}$.

9. (a) We have that

$$\mathbb{P}(X^+\le x)=\begin{cases}0 & \text{if } x<0,\\ F(x) & \text{if } x\ge 0.\end{cases}$$

(b) Secondly,

$$\mathbb{P}(X^-\le x)=\begin{cases}0 & \text{if } x<0,\\ 1-\lim_{y\uparrow -x}F(y) & \text{if } x\ge 0.\end{cases}$$

(c) $\mathbb{P}(|X|\le x)=\mathbb{P}(-x\le X\le x)$ if $x\ge 0$. Therefore

$$\mathbb{P}(|X|\le x)=\begin{cases}0 & \text{if } x<0,\\ F(x)-\lim_{y\uparrow -x}F(y) & \text{if } x\ge 0.\end{cases}$$

(d) $\mathbb{P}(-X\le x)=1-\lim_{y\uparrow -x}F(y)$.

10. By the continuity of probability measures (1.3.5),

$$\mathbb{P}(X=x_0)=\lim_{y\uparrow x_0}\mathbb{P}(y<X\le x_0)=F(x_0)-\lim_{y\uparrow x_0}F(y)=F(x_0)-F(x_0-),$$

using general properties of F. The result follows.

11. Define $m=\sup\{x:F(x)<\frac{1}{2}\}$. Then $F(y)<\frac{1}{2}$ for $y<m$, and $F(m)\ge\frac{1}{2}$ (if $F(m)<\frac{1}{2}$ then $F(m')<\frac{1}{2}$ for some $m'>m$, by the right-continuity of F, a contradiction). Hence m is a median, and is smallest with this property.

A similar argument may be used to show that $M=\sup\{x:F(x)\le\frac{1}{2}\}$ is a median, and is largest with this property. The set of medians is then the closed interval $[m,M]$.

12. Let the dice show X and Y. Write $S=X+Y$ and $f_i=\mathbb{P}(X=i)$, $g_i=\mathbb{P}(Y=i)$. Assume that $\mathbb{P}(S=2)=\mathbb{P}(S=7)=\mathbb{P}(S=12)=\frac{1}{11}$. Now

$$\mathbb{P}(S=2)=\mathbb{P}(X=1)\mathbb{P}(Y=1)=f_1g_1,$$
$$\mathbb{P}(S=12)=\mathbb{P}(X=6)\mathbb{P}(Y=6)=f_6g_6,$$
$$\mathbb{P}(S=7)\ge\mathbb{P}(X=1)\mathbb{P}(Y=6)+\mathbb{P}(X=6)\mathbb{P}(Y=1)=f_1g_6+f_6g_1.$$

It follows that $f_1g_1=f_6g_6$, and also

$$\frac{1}{11}=\mathbb{P}(S=7)\ge f_1g_1\left(\frac{g_6}{g_1}+\frac{f_6}{f_1}\right)=\frac{1}{11}\left(x+\frac{1}{x}\right)$$

where $x = g_6/g_1$. However $x + x^{-1} > 1$ for all $x > 0$, a contradiction.

13. (a) Clearly d_L satisfies (i). As for (ii), suppose that $d_L(F, G) = 0$. Then

$$F(x) \le \lim_{\epsilon \downarrow 0}\{G(x+\epsilon)+\epsilon\} = G(x)$$

and

$$F(y) \ge \lim_{\epsilon \downarrow 0}\{G(y-\epsilon)-\epsilon\} = G(y-).$$

Now $G(y-) \ge G(x)$ if $y > x$; taking the limit as $y \downarrow x$ we obtain

$$F(x) \ge \lim_{y \downarrow x} G(y-) \ge G(x),$$

implying that $F(x) = G(x)$ for all x.

Finally, if $F(x) \le G(x+\epsilon)+\epsilon$ and $G(x) \le H(x+\delta)+\delta$ for all x and some $\epsilon, \delta > 0$, then $F(x) \le H(x+\delta+\epsilon)+\epsilon+\delta$ for all x. A similar lower bound for $F(x)$ is valid, implying that $d_L(F, H) \le d_L(F, G) + d_L(G, H)$.

(b) Clearly d_{TV} satisfies (i), and $d_{TV}(X, Y) = 0$ if and only if $\mathbb{P}(X = Y) = 1$. By the usual triangle inequality,

$$\big|\mathbb{P}(X=k)-\mathbb{P}(Z=k)\big| \le \big|\mathbb{P}(X=k)-\mathbb{P}(Y=k)\big| + \big|\mathbb{P}(Y=k)-\mathbb{P}(Z=k)\big|,$$

and (iii) follows by summing over k.

We have that

$$\begin{aligned} 2\big|\mathbb{P}(X \in A)-\mathbb{P}(Y \in A)\big| &= \big|\big(\mathbb{P}(X \in A)-\mathbb{P}(Y \in A)\big) - \big(\mathbb{P}(X \in A^c)-\mathbb{P}(Y \in A^c)\big)\big| \\ &= \Big|\sum_k \big(\mathbb{P}(X=k)-\mathbb{P}(Y=k)\big) J_A(k)\Big| \end{aligned}$$

where $J_A(k)$ equals 1 if $k \in A$ and equals -1 if $k \in A^c$. Therefore,

$$2\big|\mathbb{P}(X \in A)-\mathbb{P}(Y \in A)\big| \le \sum_k \big|\mathbb{P}(X=k)-\mathbb{P}(Y=k)\big| \cdot |J_A(k)| \le d_{TV}(X, Y).$$

Equality holds if $A = \{k : \mathbb{P}(X=k) > \mathbb{P}(Y=k)\}$.

14. (a) Note that

$$\frac{\partial^2 F}{\partial x \partial y} = -e^{-x-y} < 0, \qquad x, y > 0,$$

so that F is not a joint distribution function.

(b) In this case

$$\frac{\partial^2 F}{\partial x \partial y} = \begin{cases} e^{-y} & \text{if } 0 \le x \le y, \\ 0 & \text{if } 0 \le y \le x, \end{cases}$$

and in addition

$$\int_0^\infty \int_0^\infty \frac{\partial^2 F}{\partial x \partial y}\, dx\, dy = 1.$$

Hence F is a joint distribution function, and easy substitutions reveal the marginals:

$$F_X(x) = \lim_{y \to \infty} F(x, y) = 1 - e^{-x}, \quad x \ge 0,$$

$$F_Y(y) = \lim_{x \to \infty} F(x, y) = 1 - e^{-y} - ye^{-y}, \quad y \ge 0.$$

15. Suppose that, for some $i \neq j$, we have $p_i < p_j$ and B_i is to the left of B_j. Write m for the position of B_i and r for the position of B_j, and consider the effect of interchanging B_i and B_j. For $k \leq m$ and $k > r$, $\mathbb{P}(T \geq k)$ is unchanged by the move. For $m < k \leq r$, $\mathbb{P}(T \geq k)$ is decreased by an amount $p_j - p_i$, since this is the increased probability that the search is successful at the mth position. Therefore the interchange of B_i and B_j is desirable.

It follows that the only ordering in which $\mathbb{P}(T \geq k)$ can be reduced for no k is that ordering in which the books appear in decreasing order of probability. In the event of ties, it is of no importance how the tied books are placed.

16. Intuitively, it may seem better to go *first* since the first person has greater choice. This conclusion is in fact false. Denote the coins by C_1, C_2, C_3 in order, and suppose you go *second*. If your opponent chooses C_1 then you choose C_3, because $\mathbb{P}(C_3 \text{ beats } C_1) = \frac{2}{5} + \frac{2}{5} \cdot \frac{3}{5} = \frac{16}{25} > \frac{1}{2}$. Likewise $\mathbb{P}(C_1 \text{ beats } C_2) = \mathbb{P}(C_2 \text{ beats } C_3) = \frac{3}{5} > \frac{1}{2}$. Whichever coin your opponent picks, you can arrange to have a better than evens chance of winning.

17. Various difficulties arise in sequential decision theory, even in simple problems such as this one. The following simple argument yields the optimal policy. Suppose that you have made a unsuccessful searches "ahead" and b unsuccessful searches "behind" (if any of these searches were successful, then there is no further problem). Let A be the event that the correct direction is ahead. Then

$$\begin{aligned}\mathbb{P}(A \mid \text{current knowledge}) &= \frac{\mathbb{P}(\text{current knowledge} \mid A)\mathbb{P}(A)}{\mathbb{P}(\text{current knowledge})} \\ &= \frac{(1-p)^a \alpha}{(1-p)^a \alpha + (1-p)^b(1-\alpha)},\end{aligned}$$

which exceeds $\frac{1}{2}$ if and only if $(1-p)^a \alpha > (1-p)^b(1-\alpha)$. The optimal policy is to compare $(1-p)^a \alpha$ with $(1-p)^b(1-\alpha)$. You search ahead if the former is larger and behind otherwise; in the event of a tie, do either.

18. (a) There are $\binom{64}{8}$ possible layouts, of which 8+8+2 are linear. The answer is $18/\binom{64}{8}$.
(b) Each row and column must contain exactly one pawn. There are 8 possible positions in the first row. Having chosen which of these is occupied, there are 7 positions in the second row which are admissible, 6 in the third, and so one. The answer is $8!/\binom{64}{8}$.

19. (a) The density function is $f(x) = F'(x) = 2xe^{-x^2}$, $x \geq 0$.
(b) The density function is $f(x) = F'(x) = x^2 e^{-1/x}$, $x > 0$.
(c) The density function is $f(x) = F'(x) = 2(e^x + e^{-x})^{-2}$, $x \in \mathbb{R}$.
(d) This is not a distribution function because $F'(1) < 0$.

20. We have that

$$\mathbb{P}(U = V) = \int_{\{(u,v):u=v\}} f_{U,V}(u, v)\, du\, dv = 0.$$

The random variables X, Y are continuous but not jointly continuous: there exists no integrable function $f : [0,1]^2 \to \mathbb{R}$ such that

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{u=0}^{x} \int_{v=0}^{y} f(u, v)\, du\, dv, \quad 0 \leq x, y \leq 1.$$

3

Discrete random variables

3.1 Solutions. Probability mass functions

1. (a) $C^{-1} = \sum_1^\infty 2^{-k} = 1$.
(b) $C^{-1} = \sum_1^\infty 2^{-k}/k = \log 2$.
(c) $C^{-1} = \sum_1^\infty k^{-2} = \pi^2/6$.
(d) $C^{-1} = \sum_1^\infty 2^k/k! = e^2 - 1$.

2. (i) $\frac{1}{2}$; $1 - (2\log 2)^{-1}$; $1 - 6\pi^{-2}$; $(e^2 - 3)/(e^2 - 1)$.
(ii) 1; 1; 1; 1 and 2.
(iii) It is the case that $\mathbb{P}(X \text{ even}) = \sum_{k=1}^\infty \mathbb{P}(X = 2k)$, and the answers are therefore
(a) $\frac{1}{3}$, (b) $1 - (\log 3)/(\log 4)$, (c) $\frac{1}{4}$. (d) We have that

$$\sum_{k=1}^{\infty} \frac{2^{2k}}{(2k)!} = \sum_{i=0}^{\infty} \frac{2^i + (-2)^i}{2(i!)} - 1 = \tfrac{1}{2}(e^2 + e^{-2}) - 1,$$

so the answer is $\frac{1}{2}(1 - e^{-2})$.

3. The number X of heads on the second round is the same as if we toss all the coins twice and count the number which show heads on both occasions. Each coin shows heads twice with probability p^2, so $\mathbb{P}(X = k) = \binom{n}{k} p^{2k}(1 - p^2)^{n-k}$.

4. Let D_k be the number of digits (to base 10) in the integer k. Then

$$\mathbb{P}(N = k) = \mathbb{P}(N = k \mid T = D_k)\mathbb{P}(T = D_k) = \frac{1}{|S_{D_k}|} 2^{-D_k}.$$

5. (a) The assertion follows for the binomial distribution because $k(n - k) \le (n - k + 1)(k + 1)$. The Poisson case is trivial.

(b) This follows from the fact that $k^8 \ge (k^2 - 1)^4$.

(c) The geometric mass function $f(k) = qp^k$, $k \ge 0$.

3.2 Solutions. Independence

1. We have that

$$\mathbb{P}(X = 1, Z = 1) = \mathbb{P}(X = 1, Y = 1) = \tfrac{1}{4} = \mathbb{P}(X = 1)\mathbb{P}(Z = 1).$$

This, together with three similar equations, shows that X and Z are independent. Likewise, Y and Z are independent. However

$$\mathbb{P}(X=1, Y=1, Z=-1) = 0 \neq \tfrac{1}{8} = \mathbb{P}(X=1)\mathbb{P}(Y=1)\mathbb{P}(Z=-1),$$

so that X, Y, and Z are not independent.

2. (a) If $x \geq 1$,

$$\begin{aligned}\mathbb{P}\big(\min\{X,Y\} \leq x\big) &= 1 - \mathbb{P}(X > x, Y > x) = 1 - \mathbb{P}(X > x)\mathbb{P}(Y > x)\\ &= 1 - 2^{-x} \cdot 2^{-x} = 1 - 4^{-x}.\end{aligned}$$

(b) $\mathbb{P}(Y > X) = \mathbb{P}(Y < X)$ by symmetry. Also

$$\mathbb{P}(Y > X) + \mathbb{P}(Y < X) + \mathbb{P}(Y = X) = 1.$$

Since

$$\mathbb{P}(Y = X) = \sum_x \mathbb{P}(Y = X = x) = \sum_x 2^{-x} \cdot 2^{-x} = \tfrac{1}{3},$$

we have that $\mathbb{P}(Y > X) = \frac{1}{3}$.

(c) $\frac{1}{3}$ by part (b).

(d)

$$\begin{aligned}\mathbb{P}(X \geq kY) &= \sum_{y=1}^{\infty} \mathbb{P}\big(X \geq kY, Y = y\big)\\ &= \sum_{y=1}^{\infty} \mathbb{P}\big(X \geq ky, Y = y\big) = \sum_{y=1}^{\infty} \mathbb{P}(X \geq ky)\mathbb{P}(Y = y)\\ &= \sum_{y=1}^{\infty}\sum_{x=0}^{\infty} 2^{-ky-x}2^{-y} = \frac{2}{2^{k+1}-1}.\end{aligned}$$

(e)

$$\begin{aligned}\mathbb{P}(X \text{ divides } Y) &= \sum_{k=1}^{\infty} \mathbb{P}(Y = kX) = \sum_{k=1}^{\infty}\sum_{x=1}^{\infty} \mathbb{P}(Y = kx, X = x)\\ &= \sum_{k=1}^{\infty}\sum_{x=1}^{\infty} 2^{-kx}2^{-x} = \sum_{k=1}^{\infty} \frac{1}{2^{k+1}-1}.\end{aligned}$$

(f) Let $r = m/n$ where m and n are coprime. Then

$$\mathbb{P}(X = rY) = \sum_{k=1}^{\infty} \mathbb{P}(X = km, Y = kn) = \sum_{k=1}^{\infty} 2^{-km}2^{-kn} = \frac{1}{2^{m+n}-1}.$$

3. (a) We have that

$$\begin{aligned}\mathbb{P}(X_1 < X_2 < X_3) &= \sum_{i<j<k} (1-p_1)(1-p_2)(1-p_3)p_1^{i-1}p_2^{j-1}p_3^{k-1}\\ &= \sum_{i<j} (1-p_1)(1-p_2)p_1^{i-1}p_2^{j-1}p_3^{j}\\ &= \sum_i \frac{(1-p_1)(1-p_2)p_1^{i-1}(p_2p_3)^i p_3}{1-p_2p_3}\\ &= \frac{(1-p_1)(1-p_2)p_2p_3^2}{(1-p_2p_3)(1-p_1p_2p_3)}.\end{aligned}$$

(b)
$$\mathbb{P}(X_1 \le X_2 \le X_3) = \sum_{i \le j \le k} (1-p_1)(1-p_2)(1-p_3)p_1^{i-1}p_2^{j-1}p_3^{k-1} = \frac{(1-p_1)(1-p_2)}{(1-p_2p_3)(1-p_1p_2p_3)}.$$

4. (a) Either substitute $p_1 = p_2 = p_3 = \frac{5}{6}$ in the result of Exercise (3b), or argue as follows, with the obvious notation. The event $\{A < B < C\}$ occurs only if one of the following occurs on the first round:

(i) A and B both rolled 6,

(ii) A rolled 6, B and C did not,

(iii) none rolled 6.

Hence, using conditional probabilities,

$$\mathbb{P}(A < B < C) = \left(\tfrac{1}{6}\right)^2 + \tfrac{1}{6}\left(\tfrac{5}{6}\right)^2\mathbb{P}(B < C) + \left(\tfrac{5}{6}\right)^3\mathbb{P}(A < B < C),$$

In calculating $\mathbb{P}(B < C)$ we may ignore A's rolls, and an argument similar to the above tells us that

$$\mathbb{P}(B < C) = \left(\tfrac{5}{6}\right)^2\mathbb{P}(B < C) + \tfrac{1}{6}.$$

Hence $\mathbb{P}(B < C) = \frac{6}{11}$, yielding $\mathbb{P}(A < B < C) = \frac{216}{1001}$.

(b) One may argue as above. Alternatively, let N be the total number of rolls before the first 6 appears. The probability that A rolls the first 6 is

$$\mathbb{P}\big(N \in \{1, 4, 7, \dots\}\big) = \sum_{k=1,4,7,\dots} \left(\tfrac{5}{6}\right)^{k-1}\tfrac{1}{6} = \tfrac{36}{91}.$$

Once A has thrown the first 6, the game restarts with the players rolling in order BCABCA.... Hence the probability that B rolls the next 6 is $\frac{36}{91}$ also, and similarly for the probability that C throws the third 6. The answer is therefore $\left(\frac{36}{91}\right)^3$.

5. The vector $(-X_r : 1 \le r \le n)$ has the same joint distribution as $(X_r : 1 \le r \le n)$, and the claim follows.

Let $X + 2$ and $Y + 2$ have joint mass function f, where $f_{i,j}$ is the (i, j)th entry in the matrix

$$\begin{pmatrix} \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \end{pmatrix}, \qquad 1 \le i, j \le 3.$$

Then

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \mathbb{P}(Y = -1) = \mathbb{P}(Y = 1) = \tfrac{1}{3}, \quad \mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = \tfrac{1}{3},$$
$$\mathbb{P}(X + Y = -2) = \tfrac{1}{6} \ne \tfrac{1}{12} = \mathbb{P}(X + Y = 2).$$

3.3 Solutions. Expectation

1. (a) No!
(b) Let X have mass function: $f(-1) = \frac{1}{9}$, $f\left(\frac{1}{2}\right) = \frac{4}{9}$, $f(2) = \frac{4}{9}$. Then

$$\mathbb{E}(X) = -\tfrac{1}{9} + \tfrac{2}{9} + \tfrac{8}{9} = 1 = -\tfrac{1}{9} + \tfrac{8}{9} + \tfrac{2}{9} = \mathbb{E}(1/X).$$

2. (a) If you have already j distinct types of object, the probability that the next packet contains a different type is $(c-j)/c$, and the probability that it does not is j/c. Hence the number of days required has the geometric distribution with parameter $(c-j)/c$; this distribution has mean $c/(c-j)$.
(b) The time required to collect all the types is the sum of the successive times to collect each new type. The mean is therefore

$$\sum_{j=0}^{c-1} \frac{c}{c-j} = c\sum_{k=1}^{c} \frac{1}{k}.$$

3. (a) Let I_{ij} be the indicator function of the event that players i and j throw the same number. Then

$$\mathbb{E}(I_{ij}) = \mathbb{P}(I_{ij} = 1) = \sum_{i=1}^{6} \left(\tfrac{1}{6}\right)^2 = \tfrac{1}{6}, \qquad i \neq j.$$

The total score of the group is $S = \sum_{i<j} I_{ij}$, so

$$\mathbb{E}(S) = \sum_{i<j} \mathbb{E}(I_{ij}) = \frac{1}{6}\binom{n}{2}.$$

We claim that the family $\{I_{ij} : i < j\}$ is pairwise independent. The crucial calculation for this is as follows: if $i < j < k$ then

$$\mathbb{E}(I_{ij}I_{jk}) = \mathbb{P}(i, j, \text{and } k \text{ throw same number}) = \sum_{r=1}^{6} \left(\tfrac{1}{6}\right)^3 = \tfrac{1}{36} = \mathbb{E}(I_{ij})\mathbb{E}(I_{jk}).$$

Hence

$$\operatorname{var}(S) = \operatorname{var}\left(\sum_{i<j} I_{ij}\right) = \sum_{i<j} \operatorname{var}(I_{ij}) = \binom{n}{2} \operatorname{var}(I_{12})$$

by symmetry. But $\operatorname{var}(I_{12}) = \frac{1}{6}\left(1 - \frac{1}{6}\right)$.
(b) Let X_{ij} be the common score of players i and j, so that $X_{ij} = 0$ if their scores are different. This time the total score is $S = \sum_{i<j} X_{ij}$, and

$$\mathbb{E}(S) = \binom{n}{2}\mathbb{E}(X_{12}) = \binom{n}{2}\frac{1}{6}\cdot\frac{7}{2} = \frac{7}{12}\binom{n}{2}.$$

The X_{ij} are *not* pairwise independent, and you have to slog it out thus:

$$\begin{aligned}
\operatorname{var}(S) &= \mathbb{E}\left\{\left(\sum_{i<j} X_{ij}\right)^2\right\} - \mathbb{E}(S)^2 \\
&= \binom{n}{2}\mathbb{E}(X_{12}^2) + \binom{n}{3}\mathbb{E}(X_{12}X_{23}) + \left\{\binom{n}{2}^2 - \binom{n}{2} - \binom{n}{3}\right\}\mathbb{E}(X_{12})^2 - \left(\frac{7}{12}\right)^2\binom{n}{2}^2 \\
&= \frac{315}{144}\binom{n}{2} + \frac{35}{432}\binom{n}{3}.
\end{aligned}$$

4. The expected reward is $\sum_{k=1}^{\infty} 2^{-k} \cdot 2^k = \infty$. If your utility function is u, then your 'fair' entrance fee is $\sum_{k=1}^{\infty} 2^{-k} u(2^k)$. For example, if $u(k) = c(1 - k^{-\alpha})$ for $k \geq 1$, where $c, \alpha > 0$, then the fair fee is

$$c \sum_{k=1}^{\infty} 2^{-k}(1 - 2^{-\alpha k}) = c\left(1 - \frac{1}{2^{\alpha+1} - 1}\right).$$

This fee is certainly not 'fair' for the person offering the wager, unless possibly he is a noted philanthropist.

5. We have that $\mathbb{E}(X^\alpha) = \sum_{x=1}^{\infty} x^\alpha / \{x(x+1)\}$, which is finite if and only if $\alpha < 1$.

6. Clearly

$$\operatorname{var}(a + X) = \mathbb{E}\big(\{(a + X) - \mathbb{E}(a + X)\}^2\big) = \mathbb{E}(\{X - \mathbb{E}(X)\}^2) = \operatorname{var}(X).$$

7. For each r, bet $\{1 + \pi(r)\}^{-1}$ on horse r. If the rth horse wins, your payoff is $\{\pi(r) + 1\}\{1 + \pi(r)\}^{-1} = 1$, which is in excess of your total stake $\sum_k \{\pi(k) + 1\}^{-1}$.

8. We may assume that: (a) after any given roll of the die, your decision whether or not to stop depends only on the value V of the current roll; (b) if it is optimal to stop for $V = r$, then it is also optimal to stop when $V > r$.

Consider the strategy: stop the first time that the die shows r or greater. Let $S(r)$ be the expected score achieved by following this stategy. By elementary calculations,

$$S(6) = 6 \cdot \mathbb{P}(6 \text{ appears before } 1) + 1 \cdot \mathbb{P}(1 \text{ appears before } 6) = \tfrac{7}{2},$$

and similarly $S(5) = 4$, $S(4) = 4$, $S(3) = \frac{19}{5}$, $S(2) = \frac{7}{2}$. The optimal strategy is therefore to stop at the first throw showing 4, 5, or 6. Similar arguments may be used to show that 'stop at 5 or 6' is the rule to maximize the expected squared score.

9. Proceeding as in Exercise (8), we find the expected returns for the same strategies to be:

$$S(6) = \tfrac{7}{2} - 3c, \quad S(5) = 4 - 2c, \quad S(4) = 4 - \tfrac{3}{2}c, \quad S(3) = \tfrac{19}{5} - \tfrac{6}{5}c, \quad S(2) = \tfrac{7}{2} - c.$$

If $c = \frac{1}{3}$, it is best to stop when the score is at least 4; if $c = 1$, you should stop when the score is at least 3. The respective expected scores are $\frac{7}{2}$ and $\frac{13}{5}$.

3.4 Solutions. Indicators and matching

1. Let I_j be the indicator function of the event that the outcome of the $(j+1)$th toss is different from the outcome of the jth toss. The number R of distinct runs is given by $R = 1 + \sum_{j=1}^{n-1} I_j$. Hence

$$\mathbb{E}(R) = 1 + (n-1)\mathbb{E}(I_1) = 1 + (n-1)2pq,$$

where $q = 1 - p$. Now remark that I_j and I_k are independent if $|j - k| > 1$, so that

$$\begin{aligned}\mathbb{E}\{(R-1)^2\} = \mathbb{E}\left\{\left(\sum_{j=1}^{n-1} I_j\right)^2\right\} &= (n-1)\mathbb{E}(I_1) + 2(n-2)\mathbb{E}(I_1 I_2) \\ &\quad + \{(n-1)^2 - (n-1) - 2(n-2)\}\mathbb{E}(I_1)^2.\end{aligned}$$

Now $\mathbb{E}(I_1) = 2pq$ and $\mathbb{E}(I_1 I_2) = p^2 q + pq^2 = pq$, and therefore

$$\begin{aligned}\operatorname{var}(R) = \operatorname{var}(R-1) &= (n-1)\mathbb{E}(I_1) + 2(n-2)\mathbb{E}(I_1 I_2) - \big\{(n-1) + 2(n-2)\big\}\mathbb{E}(I_1)^2 \\ &= 2pq\big(2n - 3 - 2pq(3n-5)\big).\end{aligned}$$

2. The required total is $T = \sum_{i=1}^{k} X_i$, where X_i is the number shown on the ith ball. Hence $\mathbb{E}(T) = k\mathbb{E}(X_1) = \frac{1}{2}k(n+1)$. Now calculate, boringly,

$$\begin{aligned}\mathbb{E}\left\{\left(\sum_{i=1}^{k} X_i\right)^2\right\} &= k\mathbb{E}(X_1^2) + k(k-1)\mathbb{E}(X_1 X_2) \\ &= \frac{k}{n}\sum_{1}^{n} j^2 + \frac{k(k-1)}{n(n-1)} 2\sum_{i>j} ij \\ &= \frac{k}{n}\left\{\tfrac{1}{3}n(n+1)(n+2) - \tfrac{1}{2}n(n+1)\right\} \\ &\quad + \frac{k(k-1)}{n(n-1)}\sum_{j=1}^{n} j\{n(n+1) - j(j+1)\} \\ &= \tfrac{1}{6}k(n+1)(2n+1) + \tfrac{1}{12}k(k-1)(3n+2)(n+1).\end{aligned}$$

Hence

$$\operatorname{var}(T) = k(n+1)\left\{\tfrac{1}{6}(2n+1) + \tfrac{1}{12}(k-1)(3n+2) - \tfrac{1}{4}k(n+1)\right\} = \tfrac{1}{12}(n+1)k(n-k).$$

3. Each couple survives with probability

$$\binom{2n-2}{m}\bigg/\binom{2n}{m} = \left(1 - \frac{m}{2n}\right)\left(1 - \frac{m}{2n-1}\right),$$

so the required mean is

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

4. Any given red ball is in urn R after stage k if and only if it has been selected an even number of times. The probability of this is

$$\begin{aligned}\sum_{m \text{ even}} \binom{k}{m}\left(\frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{k-m} &= \frac{1}{2}\left\{\left[\left(1 - \frac{1}{n}\right) + \frac{1}{n}\right]^k + \left[\left(1 - \frac{1}{n}\right) - \frac{1}{n}\right]^k\right\} \\ &= \frac{1}{2}\left\{1 + \left(1 - \frac{2}{n}\right)^k\right\},\end{aligned}$$

and the mean number of such red balls is n times this probability.

5. Label the edges and vertices as in Figure 3.1. The structure function is

$$\begin{aligned}\zeta(X) = X_5 + (1 - X_5)\Big\{&(1 - X_1)X_4\big[X_3 + (1 - X_3)X_2 X_6\big] \\ &+ X_1\Big[X_2 + (1 - X_2)\Big(X_3\big\langle X_6 + X_4(1 - X_6)\big\rangle\Big)\Big]\Big\}.\end{aligned}$$

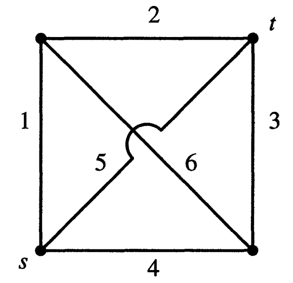


Figure 3.1. The network with source s and sink t.

For the reliability, see Problem (1.8.19a).

6. The structure function is $I_{\{S \ge k\}}$, the indicator function of $\{S \ge k\}$ where $S = \sum_{c=1}^{n} X_c$. The reliability is therefore $\sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$.

7. Independently colour each vertex livid or bronze with probability $\frac{1}{2}$ each, and let L be the random set of livid vertices. Then $\mathbb{E}N_L = \frac{1}{2}|E|$. There must exist one or more possible values of N_L which are at least as large as its mean.

8. Let I_r be the indicator function that the rth pair have opposite polarity, so that $X = 1 + \sum_{r=1}^{n-1} I_r$. We have that $\mathbb{P}(I_r = 1) = \frac{1}{2}$ and $\mathbb{P}(I_r = I_{r+1} = 1) = \frac{1}{4}$, whence $\mathbb{E}X = \frac{1}{2}(n+1)$ and $\operatorname{var} X = \frac{1}{4}(n-1)$.

9. (a) Let A_i be the event that the integer i remains in the ith position. Then

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{r=1}^{n} A_r\Big) &= \sum_r \mathbb{P}(A_r) - \sum_{r<s} \mathbb{P}(A_r \cap A_s) + \cdots + (-1)^{n-1}\mathbb{P}\Big(\bigcap_r A_r\Big) \\ &= n \cdot \frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \cdots + (-1)^{n-1}\frac{1}{n!}.\end{aligned}$$

Therefore the number M of matches satisfies

$$\mathbb{P}(M=0) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}.$$

Now

$$\begin{aligned}\mathbb{P}(M=r) &= \binom{n}{r}\mathbb{P}\big(r \text{ given numbers match, and the remaining } n-r \text{ are deranged}\big) \\ &= \frac{n!}{r!(n-r)!}\frac{(n-r)!}{n!}\left(\frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-r}\frac{1}{(n-r)!}\right).\end{aligned}$$

(b)

$$\begin{aligned}d_{n+1} &= \sum_{r=2}^{n+1} \#\{\text{derangements with 1 in the } r\text{th place}\} \\ &= n\big\{\#\{\text{derangements which swap 1 with 2}\} \\ &\qquad + \#\{\text{derangements in which 1 is in the 2nd place and 2 is not in the 1st place}\}\big\} \\ &= nd_{n-1} + nd_n,\end{aligned}$$

where $\#A$ denotes the cardinality of the set A. By rearrangement, $d_{n+1} - (n+1)d_n = -(d_n - nd_{n-1})$. Set $u_n = d_n - nd_{n-1}$ and note that $u_2 = 1$, to obtain $u_n = (-1)^n$, $n \geq 2$, and hence

$$d_n = \frac{n!}{2!} - \frac{n!}{3!} + \cdots + (-1)^n \frac{n!}{n!}.$$

Now divide by $n!$ to obtain the results above.

3.5 Solutions. Examples of discrete variables

1. There are $n!/(n_1!\, n_2! \cdots n_t!)$ sequences of outcomes in which the ith possible outcome occurs n_i times for each i. The probability of any such sequence is $p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}$, and the result follows.

2. The total number H of heads satisfies

$$\begin{aligned}\mathbb{P}(H = x) &= \sum_{n=x}^{\infty} \mathbb{P}(H = x \mid N = n)\mathbb{P}(N = n) = \sum_{n=x}^{\infty} \binom{n}{x} p^x (1-p)^{n-x} \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \frac{(\lambda p)^x e^{-\lambda p}}{x!} \sum_{n=x}^{\infty} \frac{\{\lambda(1-p)\}^{n-x} e^{-\lambda(1-p)}}{(n-x)!}.\end{aligned}$$

The last summation equals 1, since it is the sum of the values of the Poisson mass function with parameter $\lambda(1-p)$.

3. $dp_n/d\lambda = p_{n-1} - p_n$ where $p_{-1} = 0$. Hence $(d/d\lambda)\mathbb{P}(X \leq n) = p_n(\lambda)$.

4. The probability of a marked animal in the nth place is a/b. Conditional on this event, the chance of $n-1$ preceding places containing $m-1$ marked and $n-m$ unmarked animals is

$$\binom{a-1}{m-1}\binom{b-a}{n-m} \bigg/ \binom{b-1}{n-1},$$

as required. Now let X_j be the number of unmarked animals between the $j-1$th and jth marked animals, if all were caught. By symmetry, $\mathbb{E}X_j = (b-a)/(a+1)$, whence $\mathbb{E}X = m(\mathbb{E}X_1 + 1) = m(b+1)/(a+1)$.

3.6 Solutions. Dependence

1. Remembering Problem (2.7.3b), it suffices to show that $\text{var}(aX + bY) < \infty$ if $a, b \in \mathbb{R}$ and $\text{var}(X), \text{var}(Y) < \infty$. Now,

$$\begin{aligned}\text{var}(aX + bY) &= \mathbb{E}\left(\{aX + bY - \mathbb{E}(aX + bY)\}^2\right) \\ &= a^2 \text{var}(X) + 2ab \operatorname{cov}(X, Y) + b^2 \text{var}(Y) \\ &\leq a^2 \text{var}(X) + 2ab\sqrt{\text{var}(X)\,\text{var}(Y)} + b^2 \text{var}(Y) \\ &= \left(a\sqrt{\text{var}(X)} + b\sqrt{\text{var}(Y)}\right)^2\end{aligned}$$

where we have used the Cauchy–Schwarz inequality (3.6.9) applied to $X - \mathbb{E}(X)$, $Y - \mathbb{E}(Y)$.

2. Let N_i be the number of times the ith outcome occurs. Then N_i has the binomial distribution with parameters n and p_i.

3. For $x = 1, 2, \dots,$

$$\begin{aligned}\mathbb{P}(X = x) &= \sum_{y=1}^{\infty} \mathbb{P}(X = x, Y = y)\\ &= \sum_{y=1}^{\infty} \frac{C}{2}\left\{\frac{1}{(x+y-1)(x+y)} - \frac{1}{(x+y)(x+y+1)}\right\}\\ &= \frac{C}{2x(x+1)} = \frac{C}{2}\left(\frac{1}{x} - \frac{1}{x+1}\right),\end{aligned}$$

and hence $C = 2$. Clearly Y has the same mass function. Finally $\mathbb{E}(X) = \sum_{x=1}^{\infty}(x+1)^{-1} = \infty$, so the covariance does not exist.

4. $\operatorname{Max}\{u, v\} = \frac{1}{2}(u+v) + \frac{1}{2}|u - v|$, and therefore

$$\begin{aligned}\mathbb{E}\big(\max\{X^2, Y^2\}\big) &= \tfrac{1}{2}\mathbb{E}(X^2 + Y^2) + \tfrac{1}{2}\mathbb{E}\big|(X-Y)(X+Y)\big|\\ &\le 1 + \tfrac{1}{2}\sqrt{\mathbb{E}\big((X-Y)^2\big)\mathbb{E}\big((X+Y)^2\big)}\\ &= 1 + \tfrac{1}{2}\sqrt{(2-2\rho)(2+2\rho)} = 1 + \sqrt{1-\rho^2},\end{aligned}$$

where we have used the Cauchy–Schwarz inequality.

5. (a) $\log y \le y - 1$ with equality if and only if $y = 1$. Therefore,

$$\mathbb{E}\left(\log \frac{f_Y(X)}{f_X(X)}\right) \le \mathbb{E}\left[\frac{f_Y(X)}{f_X(X)} - 1\right] = 0,$$

with equality if and only if $f_Y = f_X$.
(b) This holds likewise, with equality if and only if $f(x, y) = f_X(x)f_Y(y)$ for all x, y, which is to say that X and Y are independent.

6. (a) $a + b + c = \mathbb{E}\{I_{\{X>Y\}} + I_{\{Y>Z\}} + I_{\{Z>X\}}\} = 2$, whence $\min\{a, b, c\} \le \frac{2}{3}$. Equality is attained, for example, if the vector (X, Y, Z) takes only three values with probabilities $f(2, 1, 3) = f(3, 2, 1) = f(1, 3, 2) = \frac{1}{3}$.
(b) $\mathbb{P}(X < Y) = \mathbb{P}(Y < X)$, etc.
(c) We have that $c = a = p$ and $b = 1 - p^2$. Also $\sup \min\{p, (1-p^2)\} = \frac{1}{2}(\sqrt{5} - 1)$.

7. We have for $1 \le x \le 9$ that

$$f_X(x) = \sum_{y=0}^{9} \log\left(1 + \frac{1}{10x+y}\right) = \log \prod_{y=0}^{9}\left(1 + \frac{1}{10x+y}\right) = \log\left(1 + \frac{1}{x}\right).$$

By calculation, $\mathbb{E}X \simeq 3.44$.

8. (i) $f_X(j) = c\displaystyle\sum_{k=0}^{\infty}\left\{\frac{j}{j!}a^j\frac{a^k}{k!} + \frac{ka^j a^k}{k!\,j!}\right\} = c\frac{e^a(j+a)a^j}{j!}$.

(ii) $1 = \displaystyle\sum_j f_X(j) = 2ace^{2a}$, whence $c = e^{-2a}/(2a)$.

(ii) $f_{X+Y}(r) = \displaystyle\sum_{j=0}^{r}\frac{cra^r}{j!\,(r-j)!} = \frac{cra^r 2^r}{r!}$, $r \ge 1$.

(iii) $\mathbb{E}(X+Y-1) = \sum_{r=1}^{\infty} \frac{cr(r-1)(2a)^r}{r!} = 2a$. Now $\mathbb{E}(X) = \mathbb{E}(Y)$, and therefore $\mathbb{E}(X) = a + \frac{1}{2}$.

3.7 Solutions. Conditional distributions and conditional expectation

1. (a) We have that

$$\begin{aligned}\mathbb{E}(aY + bZ \mid X = x) &= \sum_{y,z}(ay+bz)\mathbb{P}(Y=y, Z=z \mid X=x)\\ &= a\sum_{y,z} y\mathbb{P}(Y=y, Z=z \mid X=x) + b\sum_{y,z} z\mathbb{P}(Y=y, Z=z \mid X=x)\\ &= a\sum_{y} y\mathbb{P}(Y=y \mid X=x) + b\sum_{z} z\mathbb{P}(Z=z \mid X=x).\end{aligned}$$

Parts (b)–(e) are verified by similar trivial calculations. Turning to (f),

$$\begin{aligned}\mathbb{E}\{\mathbb{E}(Y \mid X, Z) \mid X = x\} &= \sum_z \left\{\sum_y y\mathbb{P}(Y=y \mid X=x, Z=z)\mathbb{P}(X=x, Z=z \mid X=x)\right\}\\ &= \sum_z \sum_y y \frac{\mathbb{P}(Y=y, X=x, Z=z)}{\mathbb{P}(X=x, Z=z)} \cdot \frac{\mathbb{P}(X=x, Z=z)}{\mathbb{P}(X=x)}\\ &= \sum_y y\mathbb{P}(Y=y \mid X=x) = \mathbb{E}(Y \mid X=x)\\ &= \mathbb{E}\Big\{\mathbb{E}(Y \mid X) \,\Big|\, X=x, Z=z\Big\}, \quad \text{by part (e).}\end{aligned}$$

2. If ϕ and ψ are two such functions then $\mathbb{E}\big((\phi(X)-\psi(X))g(X)\big) = 0$ for any suitable g. Setting $g(X) = I_{\{X=x\}}$ for any $x \in \mathbb{R}$ such that $\mathbb{P}(X=x) > 0$, we obtain $\phi(x) = \psi(x)$. Therefore $\mathbb{P}(\phi(X) = \psi(X)) = 1$.

3. We do not seriously expect you to want to do this one. However, if you insist, the method is to check in each case that both sides satisfy the appropriate definition, and then to appeal to uniqueness, deducing that the sides are almost surely equal (see Williams 1991, p. 88).

4. The natural definition is given by

$$\operatorname{var}(Y \mid X = x) = \mathbb{E}\big(\{Y - \mathbb{E}(Y \mid X = x)\}^2 \mid X = x\big).$$

Now,

$$\begin{aligned}\operatorname{var}(Y) = \mathbb{E}(\{Y - \mathbb{E}Y\}^2) &= \mathbb{E}\left[\mathbb{E}\big(\{Y - \mathbb{E}(Y \mid X) + \mathbb{E}(Y \mid X) - \mathbb{E}Y\}^2 \mid X\big)\right]\\ &= \mathbb{E}\big(\operatorname{var}(Y \mid X)\big) + \operatorname{var}\big(\mathbb{E}(Y \mid X)\big)\end{aligned}$$

since the mean of $\mathbb{E}(Y \mid X)$ is $\mathbb{E}Y$, and the cross product is, by Exercise (1e),

$$\begin{aligned}&2\mathbb{E}\Big[\mathbb{E}\big(\{Y - \mathbb{E}(Y \mid X)\}\{\mathbb{E}(Y \mid X) - \mathbb{E}Y\} \mid X\big)\Big]\\ &\qquad = 2\mathbb{E}\Big[\{\mathbb{E}(Y \mid X) - \mathbb{E}Y\}\mathbb{E}\{Y - \mathbb{E}(Y \mid X) \mid X\}\Big] = 0\end{aligned}$$

since $\mathbb{E}\{Y - \mathbb{E}(Y \mid X) \mid X\} = \mathbb{E}(Y \mid X) - \mathbb{E}(Y \mid X) = 0$.

5. We have that

$$\mathbb{E}(T - t \mid T > t) = \sum_{r=0}^{\infty} \mathbb{P}(T > t + r \mid T > t) = \sum_{r=0}^{\infty} \frac{\mathbb{P}(T > t + r)}{\mathbb{P}(T > t)}.$$

(a) $$\mathbb{E}(T - t \mid T > t) = \sum_{r=0}^{N-t} \frac{N - t - r}{N - t} = \tfrac{1}{2}(N - t + 1).$$

(b) $$\mathbb{E}(T - t \mid T > t) = \sum_{r=0}^{\infty} \frac{2^{-(t+r)}}{2^{-t}} = 2.$$

6. Clearly

$$\mathbb{E}(S \mid N = n) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \mu n,$$

and hence $\mathbb{E}(S \mid N) = \mu N$. It follows that $\mathbb{E}(S) = \mathbb{E}\{\mathbb{E}(S \mid N)\} = \mathbb{E}(\mu N)$.

7. A robot passed is in fact faulty with probability $\pi = \{\phi(1 - \delta)\}/(1 - \phi\delta)$. Thus the number of faulty passed robots, given Y, is $\text{bin}(n - Y, \pi)$, with mean $(n - Y)\{\phi(1 - \delta)\}/(1 - \phi\delta)$. Hence

$$\mathbb{E}(X \mid Y) = Y + \frac{(n - Y)\phi(1 - \delta)}{1 - \phi\delta}.$$

8. (a) Let m be the family size, ϕ_r the indicator that the rth child is female, and μ_r the indicator of a male. The numbers G, B of girls and boys satisfy

$$G = \sum_{r=1}^{m} \phi_r, \quad B = \sum_{r=1}^{m} \mu_r, \quad \mathbb{E}(G) = \tfrac{1}{2}m = \mathbb{E}(B).$$

(It will be shown later that the result remains true for random m under reasonable conditions.) We have not used the property of independence.

(b) With M the event that the selected child is male,

$$\mathbb{E}(G \mid M) = \mathbb{E}\left(\sum_{r=1}^{m-1} \phi_r\right) = \tfrac{1}{2}(m - 1) = \mathbb{E}(B).$$

The independence is necessary for this argument.

9. Conditional expectation is defined in terms of the conditional distribution, so the first step is not justified. Even if this step were accepted, the second equality is generally false.

10. By conditioning on X_{n-1},

$$\mathbb{E}X_n = \mathbb{E}\big[\mathbb{E}(X_n \mid X_{n-1})\big] = \mathbb{E}\big[p(X_{n-1} + 1) + (1 - p)(X_{n-1} + 1 + \widehat{X}_n)\big]$$

where $\widehat{X}_n$ has the same distribution as X_n. Hence $\mathbb{E}X_n = (1 + \mathbb{E}X_{n-1})/p$. Solve this subject to $\mathbb{E}X_1 = p^{-1}$.

3.8 Solutions. Sums of random variables

1. By the convolution theorem,

$$\mathbb{P}(X+Y=z)=\sum_k \mathbb{P}(X=k)\mathbb{P}(Y=z-k)$$
$$=\begin{cases}\dfrac{k+1}{(m+1)(n+1)} & \text{if } 0\le k\le m\wedge n,\\[2mm] \dfrac{(m\wedge n)+1}{(m+1)(n+1)} & \text{if } m\wedge n<k<m\vee n,\\[2mm] \dfrac{m+n+1-k}{(m+1)(n+1)} & \text{if } m\vee n\le k\le m+n,\end{cases}$$

where $m\wedge n=\min\{m,n\}$ and $m\vee n=\max\{m,n\}$.

2. If $z\ge 2$,

$$\mathbb{P}(X+Y=z)=\sum_{k=1}^{\infty}\mathbb{P}(X=k,Y=z-k)=\frac{C}{z(z+1)}.$$

Also, if $z\ge 0$,

$$\begin{aligned}\mathbb{P}(X-Y=z)&=\sum_{k=1}^{\infty}\mathbb{P}(X=k+z,Y=k)\\ &=C\sum_{k=1}^{\infty}\frac{1}{(2k+z-1)(2k+z)(2k+z+1)}\\ &=\tfrac12 C\sum_{k=1}^{\infty}\left\{\frac{1}{(2k+z-1)(2k+z)}-\frac{1}{(2k+z)(2k+z+1)}\right\}\\ &=\tfrac12 C\sum_{r=1}^{\infty}\frac{(-1)^{r+1}}{(r+z)(r+z+1)}.\end{aligned}$$

By symmetry, if $z\le 0$, $\mathbb{P}(X-Y=z)=\mathbb{P}(X-Y=-z)=\mathbb{P}(X-Y=|z|)$.

3. $$\sum_{r=1}^{z-1}\alpha(1-\alpha)^{r-1}\beta(1-\beta)^{z-r-1}=\frac{\alpha\beta\{(1-\beta)^{z-1}-(1-\alpha)^{z-1}\}}{\alpha-\beta}.$$

4. Repeatedly flip a coin that shows heads with probability p. Let X_r be the number of flips after the $r-1$th head up to, and including, the rth. Then X_r is geometric with parameter p. The number of flips Z to obtain n heads is negative binomial, and $Z=\sum_{r=1}^{n}X_r$ by construction.

5. Sam. Let X_n be the number of sixes shown by $6n$ dice, so that $X_{n+1}=X_n+Y$ where Y has the same distribution as X_1 and is independent of X_n. Then,

$$\begin{aligned}\mathbb{P}(X_{n+1}\ge n+1)&=\sum_{r=0}^{6}\mathbb{P}(X_n\ge n+1-r)\mathbb{P}(Y=r)\\ &=\mathbb{P}(X_n\ge n)+\sum_{r=0}^{6}\big[\mathbb{P}(X_n\ge n+1-r)-\mathbb{P}(X_n\ge n)\big]\mathbb{P}(Y=r).\end{aligned}$$

We set $g(k)=\mathbb{P}(X_n=k)$ and use the fact, easily proved, that $g(n)\ge g(n-1)\ge\cdots\ge g(n-5)$ to find that the last sum is no bigger than

$$g(n)\sum_{r=0}^{6}(r-1)\mathbb{P}(Y=r)=g(n)\big(\mathbb{E}(Y)-1\big).$$

The claim follows since $\mathbb{E}(Y) = 1$.

6. (i) $\text{LHS} = \sum_{n=0}^{\infty} ng(n)e^{-\lambda}\lambda^n/n! = \lambda \sum_{n=1}^{\infty} \frac{g(n)e^{-\lambda}}{(n-1)!}\lambda^{n-1} = \text{RHS}.$

(ii) Conditioning on N and X_N,

$$\begin{aligned}
\text{LHS} &= \mathbb{E}\big(\mathbb{E}(Sg(S) \mid N)\big) = \mathbb{E}\{N\mathbb{E}(X_N g(S) \mid N)\} \\
&= \sum_n \frac{e^{-\lambda}\lambda^n}{(n-1)!} \int x\mathbb{E}\left(g\left(\sum_{r=1}^{n-1} X_r + x\right)\right) dF(x) \\
&= \lambda \int x\mathbb{E}\big(g(S+x)\big)\, dF(x) = \text{RHS}.
\end{aligned}$$

3.9 Solutions. Simple random walk

1. (a) Consider an infinite sequence of tosses of a coin, any one of which turns up heads with probability p. With probability one there will appear a run of N heads sooner or later. If the coin tosses are 'driving' the random walk, then absorption occurs no later than this run, so that ultimate absorption is (almost surely) certain. Let S be the number of tosses before the first run of N heads. Certainly $\mathbb{P}(S > Nr) \le (1 - p^N)^r$, since Nr tosses may be divided into r blocks of N tosses, each of which is such a run with probability p^N. Hence $\mathbb{P}(S = s) \le (1 - p^N)^{\lfloor s/N \rfloor}$, and in particular $\mathbb{E}(S^k) < \infty$ for all $k \ge 1$. By the above argument, $\mathbb{E}(T^k) < \infty$ also.

2. If $S_0 = k$ then the first step X_1 satisfies

$$\mathbb{P}(X_1 = 1 \mid W) = \frac{\mathbb{P}(X_1 = 1)\mathbb{P}(W \mid X_1 = 1)}{\mathbb{P}(W)} = \frac{pp_{k+1}}{p_k}.$$

Let T be the duration of the walk. Then

$$\begin{aligned}
J_k &= \mathbb{E}(T \mid S_0 = k, W) \\
&= \mathbb{E}(T \mid S_0 = k, W, X_1 = 1)\mathbb{P}(X_1 = 1 \mid S_0 = k, W) \\
&\qquad + \mathbb{E}(T \mid S_0 = k, W, X_1 = -1)\mathbb{P}(X_1 = -1 \mid S_0 = k, W) \\
&= \big(1 + J_{k+1}\big)\frac{p_{k+1}p}{p_k} + \big(1 + J_{k-1}\big)\left(1 - \frac{p_{k+1}p}{p_k}\right) \\
&= 1 + \frac{pp_{k+1}J_{k+1}}{p_k} + \frac{\big(p_k - pp_{k+1}\big)J_{k-1}}{p_k},
\end{aligned}$$

as required.

Certainly $J_0 = 0$. If $p = \frac{1}{2}$ then $p_k = 1 - (k/N)$, so the difference equation becomes

$$(N - k - 1)J_{k+1} - 2(N - k)J_k + (N - k + 1)J_{k-1} = 2(k - N)$$

for $1 \le k \le N - 1$. Setting $u_k = (N - k)J_k$, we obtain

$$u_{k+1} - 2u_k + u_{k-1} = 2(k - N),$$

with general solution $u_k = A + Bk - \frac{1}{3}(N - k)^3$ for constants A and B. Now $u_0 = u_N = 0$, and therefore $A = \frac{1}{3}N^3$, $B = -\frac{1}{3}N^2$, implying that $J_k = \frac{1}{3}\{N^2 - (N - k)^2\}$, $0 \le k < N$.

3. The recurrence relation may be established as in Exercise (2). Set $u_k = (\rho^k - \rho^N)J_k$ and use the fact that $p_k = (\rho^k - \rho^N)/(1-\rho^N)$ where $\rho = q/p$, to obtain

$$pu_{k+1} - (1-r)u_k + qu_{k-1} = \rho^N - \rho^k.$$

The solution is

$$u_k = A + B\rho^k + \frac{k(\rho^k + \rho^N)}{p-q},$$

for constants A and B. The boundary conditions, $u_0 = u_N = 0$, yield the answer.

4. Conditioning in the obvious way on the result of the first toss, we obtain

$$p_{mn} = pp_{m-1,n} + (1-p)p_{m,n-1}, \qquad \text{if } m, n \geq 1.$$

The boundary conditions are $p_{m0} = 0$, $p_{0n} = 1$, if $m, n \geq 1$.

5. Let Y be the number of negative steps of the walk up to absorption. Then $\mathbb{E}(X+Y) = D_k$ and

$$X - Y = \begin{cases} N-k & \text{if the walk is absorbed at } N, \\ -k & \text{if the walk is absorbed at } 0. \end{cases}$$

Hence $\mathbb{E}(X-Y) = (N-k)(1-p_k) - kp_k$, and solving for $\mathbb{E}X$ gives the result.

3.10 Solutions. Random walk: counting sample paths

1. Conditioning on the first step X_1,

$$\begin{aligned} \mathbb{P}(T = 2n) &= \tfrac{1}{2}\mathbb{P}(T = 2n \mid X_1 = 1) + \tfrac{1}{2}\mathbb{P}(T = 2n \mid X_1 = -1) \\ &= \tfrac{1}{2}f_{-1}(2n-1) + \tfrac{1}{2}f_1(2n-1) \end{aligned}$$

where $f_b(m)$ is the probability that the first passage to b of a symmetric walk, starting from 0, takes place at time m. From the hitting time theorem (3.10.14),

$$f_1(2n-1) = f_{-1}(2n-1) = \frac{1}{2n-1}\mathbb{P}(S_{2n-1} = 1) = \frac{1}{2n-1}\binom{2n-1}{n}2^{-(2n-1)},$$

which therefore is the value of $\mathbb{P}(T = 2n)$.

For the last part, note first that $\sum_1^\infty \mathbb{P}(T = 2n) = 1$, which is to say that $\mathbb{P}(T < \infty) = 1$; either appeal to your favourite method in order to see this, or observe that $\mathbb{P}(T = 2n)$ is the coefficient of s^{2n} in the expansion of $F(s) = 1 - \sqrt{1-s^2}$. The required result is easily obtained by expanding the binomial coefficient using Stirling's formula.

2. By equation (3.10.13) of PRP, for $r \geq 0$,

$$\begin{aligned} \mathbb{P}(M_n = r) &= \mathbb{P}(M_n \geq r) - \mathbb{P}(M_n \geq r+1) \\ &= 2\mathbb{P}(S_n \geq r+1) + \mathbb{P}(S_n = r) - 2\mathbb{P}(S_n \geq r+2) - \mathbb{P}(S_n = r+1) \\ &= \mathbb{P}(S_n = r) + \mathbb{P}(S_n = r+1) \\ &= \max\{\mathbb{P}(S_n = r), \mathbb{P}(S_n = r+1)\} \end{aligned}$$

since only one of these two terms is non-zero.

3. By considering the random walk reversed, we see that the probability of a first visit to S_{2n} at time $2k$ is the same as the probability of a last visit to S_0 at time $2n - 2k$. The result is then immediate from the arc sine law (3.10.19) for the last visit to the origin.

3.11 Solutions to problems

1. (a) Clearly, for all $a, b \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}\big(g(X) = a, h(Y) = b\big) &= \sum_{\substack{x,y:\\ g(x)=a,h(y)=b}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{\substack{x,y:\\ g(x)=a,h(y)=b}} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \sum_{x:g(x)=a} \mathbb{P}(X = x) \sum_{y:h(y)=b} \mathbb{P}(Y = y) \\
&= \mathbb{P}(g(X) = a)\mathbb{P}(h(Y) = b).
\end{aligned}$$

(b) See the definition (3.2.1) of independence.

(c) The only remaining part which requires proof is that X and Y are independent if $f_{X,Y}(x, y) = g(x)h(y)$ for all $x, y \in \mathbb{R}$. Suppose then that this holds. Then

$$f_X(x) = \sum_y f_{X,Y}(x, y) = g(x)\sum_y h(y), \quad f_Y(y) = \sum_x f_{X,Y}(x, y) = h(y)\sum_x g(x).$$

Now

$$1 = \sum_x f_X(x) = \sum_x g(x)\sum_y h(y),$$

so that

$$f_X(x)f_Y(y) = g(x)h(y)\sum_x g(x)\sum_y h(y) = g(x)h(y) = f_{X,Y}(x, y).$$

2. If $\mathbb{E}(X^2) = \sum_x x^2\mathbb{P}(X = x) = 0$ then $\mathbb{P}(X = x) = 0$ for $x \neq 0$. Hence $\mathbb{P}(X = 0) = 1$. Therefore, if $\mathrm{var}(X) = 0$, it follows that $\mathbb{P}(X - \mathbb{E}X = 0) = 1$.

3. (a)

$$\mathbb{E}\big(g(X)\big) = \sum_y y\mathbb{P}\big(g(X) = y\big) = \sum_y \sum_{x:g(x)=y} y\mathbb{P}(X = x) = \sum_x g(x)\mathbb{P}(X = x)$$

as required.

(b)

$$\begin{aligned}
\mathbb{E}\big(g(X)h(Y)\big) &= \sum_{x,y} g(x)h(y)f_{X,Y}(x, y) && \text{by Lemma(3.6.6)} \\
&= \sum_{x,y} g(x)h(y)f_X(x)f_Y(y) && \text{by independence} \\
&= \sum_x g(x)f_X(x)\sum_y h(y)f_Y(y) = \mathbb{E}(g(X))\mathbb{E}(h(Y)).
\end{aligned}$$

4. (a) Clearly $f_X(i) = f_Y(i) = \frac{1}{3}$ for $i = 1, 2, 3$.

(b) $(X+Y)(\omega_1) = 3$, $(X+Y)(\omega_2) = 5$, $(X+Y)(\omega_3) = 4$, and therefore $f_{X+Y}(i) = \frac{1}{3}$ for $i = 3, 4, 5$.

(c) $(XY)(\omega_1) = 2$, $(XY)(\omega_2) = 6$, $(XY)(\omega_3) = 3$, and therefore $f_{XY}(i) = \frac{1}{3}$ for $i = 2, 3, 6$.

(d) Similarly $f_{X/Y}(i) = \frac{1}{3}$ for $i = \frac{1}{2}, \frac{2}{3}, 3$.

(e)
$$f_{Y|Z}(2 \mid 2) = \frac{\mathbb{P}(Y = 2, Z = 2)}{\mathbb{P}(Z = 2)} = \frac{\mathbb{P}(\omega_1)}{\mathbb{P}(\omega_1 \cup \omega_2)} = \frac{1}{2},$$

and similarly $f_{Y|Z}(3 \mid 2) = \frac{1}{2}$, $f_{Y|Z}(1 \mid 1) = 1$, and other values are 0.

(f) Likewise $f_{Z|Y}(2 \mid 2) = f_{Z|Y}(2 \mid 3) = f_{Z|Y}(1 \mid 1) = 1$.

5. (a) $\displaystyle\sum_{n=1}^{\infty} \frac{k}{n(n+1)} = k \sum_{n=1}^{\infty} \left\{ \frac{1}{n} - \frac{1}{n+1} \right\} = k$, and therefore $k = 1$.

(b) $\sum_{n=1}^{\infty} kn^{\alpha} = k\zeta(-\alpha)$ where ζ is the Riemann zeta function, and we require $\alpha < -1$ for the sum to converge. In this case $k = \zeta(-\alpha)^{-1}$.

6. (a) We have that

$$\begin{aligned}
\mathbb{P}(X + Y = n) &= \sum_{k=0}^{n} \mathbb{P}(X = n - k)\mathbb{P}(Y = k) = \sum_{k=0}^{n} \frac{e^{-\lambda}\lambda^{n-k}}{(n-k)!} \cdot \frac{e^{-\mu}\mu^k}{k!} \\
&= \frac{e^{-\lambda-\mu}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^{n-k}\mu^k = \frac{e^{-\lambda-\mu}(\lambda+\mu)^n}{n!}.
\end{aligned}$$

(b)
$$\begin{aligned}
\mathbb{P}(X = k \mid X + Y = n) &= \frac{\mathbb{P}(X = k, X + Y = n)}{\mathbb{P}(X + Y = n)} \\
&= \frac{\mathbb{P}(X = k)\mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)} = \binom{n}{k} \frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n}.
\end{aligned}$$

Hence the conditional distribution is $\text{bin}(n, \lambda/(\lambda + \mu))$.

7. (i) We have that

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

(ii) Many random variables of interest are 'waiting times', i.e., the time one must wait before the occurrence of some event A of interest. If such a time is geometric, the lack-of-memory property states that, given that A has not occurred by time n, the time to wait for A starting from n has the same distribution as it did to start with. With sufficient imagination this can be interpreted as a failure of memory by the process giving rise to A.

(iii) No. This is because, by the above, any such process satisfies $G(k + n) = G(k)G(n)$ where $G(n) = \mathbb{P}(X > n)$. Hence $G(k + 1) = G(1)^{k+1}$ and X is geometric.

8. Clearly,

$$\begin{aligned}
\mathbb{P}(X + Y = k) &= \sum_{j=0}^{k} \mathbb{P}(X = k - j, Y = j) \\
&= \sum_{j=0}^{k} \binom{m}{k-j} p^{k-j} q^{m-k+j} \binom{n}{j} p^j q^{n-j} \\
&= p^k q^{m+n-k} \sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j} = p^k q^{m+n-k} \binom{m+n}{k}
\end{aligned}$$

which is $\text{bin}(m+n,p)$.

9. Turning immediately to the second request, by the binomial theorem,

$$\tfrac{1}{2}(x+y)^n + \tfrac{1}{2}(y-x)^n = \tfrac{1}{2}\sum_k \binom{n}{k} y^{n-k}\{x^k + (-x)^k\} = \sum_{k \text{ even}} \binom{n}{k} x^k y^{n-k}$$

as required. Now,

$$\begin{aligned}\mathbb{P}(N \text{ even}) &= \sum_{k \text{ even}} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \tfrac{1}{2}\{(p+1-p)^n + (1-p-p)^n\} = \tfrac{1}{2}\{1+(1-2p)^n\}\end{aligned}$$

in agreement with Problem (1.8.20).

10. There are $\binom{b}{k}$ ways of choosing k blue balls, and $\binom{N-b}{n-k}$ ways of choosing $n-k$ red balls. The total number of ways of choosing n balls is $\binom{N}{n}$, and the claim follows. Finally,

$$\begin{aligned}\mathbb{P}(B=k) &= \binom{n}{k}\frac{b!}{(b-k)!}\cdot\frac{(N-b)!}{(N-b-n+k)!}\cdot\frac{(N-n)!}{N!} \\ &= \binom{n}{k}\left\{\frac{b}{N}\cdot\frac{b-1}{N}\cdots\frac{b-k+1}{N}\right\} \\ &\qquad \times\left\{\frac{N-b}{N}\cdots\frac{N-b-n+k+1}{N}\right\}\left\{\frac{N}{N}\cdots\frac{N-n+1}{N}\right\}^{-1} \\ &\to \binom{n}{k}p^k(1-p)^{n-k} \qquad \text{as } N\to\infty.\end{aligned}$$

11. Using the result of Problem (3.11.8),

$$\begin{aligned}\mathbb{P}(X=k \mid X+Y=N) &= \frac{\mathbb{P}(X=k)\mathbb{P}(Y=N-k)}{\mathbb{P}(X+Y=N)} \\ &= \frac{\binom{n}{k}p^k q^{n-k}\binom{n}{N-k}p^{N-k}q^{n-N+k}}{\binom{2n}{N}p^N q^{2n-N}} = \frac{\binom{n}{k}\binom{n}{N-k}}{\binom{2n}{N}}.\end{aligned}$$

12. (a) $\mathbb{E}(X) = c+d$, $\mathbb{E}(Y) = b+d$, and $\mathbb{E}(XY) = d$, so $\text{cov}(X,Y) = d-(c+d)(b+d)$, and X and Y are uncorrelated if and only if this equals 0.
(b) For independence, we require $f(i,j) = \mathbb{P}(X=i)\mathbb{P}(Y=j)$ for all i,j, which is to say that

$$a=(a+b)(a+c), \quad b=(a+b)(b+d), \quad c=(c+d)(a+c), \quad d=(b+d)(c+d).$$

Now $a+b+c+d=1$, and with a little work one sees that any one of these relations implies the other three. Therefore X and Y are independent if and only if $d=(b+d)(c+d)$, the same condition as for uncorrelatedness.

13. (a) We have that

$$\mathbb{E}(X) = \sum_{m=0}^{\infty} m\mathbb{P}(X=m) = \sum_{m=0}^{\infty}\sum_{n=0}^{m-1}\mathbb{P}(X=m) = \sum_{n=0}^{\infty}\sum_{m=n+1}^{\infty}\mathbb{P}(X=m) = \sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

(b) *First method.* Let N be the number of balls drawn. Then, by (a),

$$\begin{aligned}
\mathbb{E}(N) &= \sum_{n=0}^{r} \mathbb{P}(N > n) = \sum_{n=0}^{r} \mathbb{P}(\text{first } n \text{ balls are red}) \\
&= \sum_{n=0}^{r} \frac{r}{b+r}\frac{r-1}{b+r-1}\cdots\frac{r-n+1}{b+r-n+1} = \sum_{n=0}^{r} \frac{r!}{(b+r)!}\frac{(b+r-n)!}{(r-n)!} \\
&= \frac{r!\,b!}{(b+r)!}\sum_{n=0}^{r}\binom{n+b}{b} = \frac{b+r+1}{b+1},
\end{aligned}$$

where we have used the combinatorial identity $\sum_{n=0}^{r}\binom{n+b}{b} = \binom{r+b+1}{b+1}$. To see this, either use the simple identity $\binom{x}{r-1} + \binom{x}{r} = \binom{x+1}{r}$ repeatedly, or argue as follows. Changing the order of summation, we find that

$$\begin{aligned}
\sum_{r=0}^{\infty} x^r \sum_{n=0}^{r}\binom{n+b}{b} &= \frac{1}{1-x}\sum_{n=0}^{\infty} x^n \binom{n+b}{b} \\
&= (1-x)^{-(b+2)} = \sum_{r=0}^{\infty} x^r \binom{b+r+1}{b+1}
\end{aligned}$$

by the (negative) binomial theorem. Equating coefficients of x^r, we obtain the required identity.

Second method. Writing $m(b,r)$ for the mean in question, and conditioning on the colour of the first ball, we find that

$$m(b,r) = \frac{b}{b+r} + \{1 + m(b, r-1)\}\frac{r}{b+r}.$$

With appropriate boundary conditions and a little effort, one may obtain the result.

Third method. Withdraw all the balls, and let N_i be the number of red balls drawn between the ith and $(i+1)$th blue ball ($N_0 = N$, and N_b is defined analogously). Think of a possible 'colour sequence' as comprising r reds, split by b blues into $b+1$ red sequences. There is a one–one correspondence between the set of such sequences with $N_0 = i$, $N_m = j$ (for given i, j, m) and the set of such sequences with $N_0 = j$, $N_m = i$; just interchange the '0th' red run with the mth red run. In particular $\mathbb{E}(N_0) = \mathbb{E}(N_m)$ for all m. Now $N_0 + N_1 + \cdots + N_b = r$, so that $\mathbb{E}(N_m) = r/(b+1)$, whence the claim is immediate.

(c) We use the notation just introduced. In addition, let B_r be the number of blue balls remaining after the removal of the last red ball. The length of the last 'colour run' is $N_b + B_r$, only one of which is non-zero. The answer is therefore $r/(b+1) + b/(r+1)$, by the argument of the third solution to part (b).

14. (a) We have that $\mathbb{E}(X_k) = p_k$ and $\text{var}(X_k) = p_k(1-p_k)$, and the claims follow in the usual way, the first by the linearity of $\mathbb{E}$ and the second by the independence of the X_i; see Theorems (3.3.8) and (3.3.11).

(b) Let $s = \sum_k p_k$, and let Z be a random variable taking each of the values $p_1, p_2, \ldots, p_n$ with equal probability n^{-1}. Now $\mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \text{var}(Z) \geq 0$, so that

$$\sum_k \frac{1}{n}p_k^2 \geq \left(\sum_k \frac{1}{n}p_k\right)^2 = \frac{s^2}{n^2}$$

with equality if and only if Z is (almost surely) constant, which is to say that $p_1 = p_2 = \cdots = p_n$. Hence

$$\text{var}(Y) = \sum_k p_k - \sum_k p_k^2 \leq s - \frac{s^2}{n}$$

with equality if and only if $p_1 = p_2 = \cdots = p_n$.

Essentially the same route may be followed using a Lagrange multiplier.

(c) This conclusion is not contrary to informed intuition, but experience shows it to be contrary to much uninformed intuition.

15. A matrix $\mathbf{V}$ has zero determinant if and only if it is singular, that is to say if and only if there is a non-zero vector $\mathbf{x}$ such that $\mathbf{x}\mathbf{V}\mathbf{x}' = 0$. However,

$$\mathbf{x}\mathbf{V}(\mathbf{X})\mathbf{x}' = \mathbb{E}\left\{\left(\sum_k x_k(X_k - \mathbb{E}X_k)\right)^2\right\}.$$

Hence, by the result of Problem (3.11.2), $\sum_k x_k(X_k - \mathbb{E}X_k)$ is constant with probability one, and the result follows.

16. The random variables $X + Y$ and $|X - Y|$ are uncorrelated since

$$\begin{aligned}\operatorname{cov}\big(X + Y, |X - Y|\big) &= \mathbb{E}\big\{(X + Y)|X - Y|\big\} - \mathbb{E}(X + Y)\mathbb{E}(|X - Y|) \\ &= \tfrac{1}{4} + \tfrac{1}{4} - 1 \cdot \tfrac{1}{2} = 0.\end{aligned}$$

However,

$$\tfrac{1}{4} = \mathbb{P}(X + Y = 0, |X - Y| = 0) \neq \mathbb{P}(X + Y = 0)\mathbb{P}(|X - Y| = 0) = \tfrac{1}{4} \cdot \tfrac{1}{2} = \tfrac{1}{8},$$

so that $X + Y$ and $|X - Y|$ are dependent.

17. Let I_k be the indicator function of the event that there is a match in the kth place. Then $\mathbb{P}(I_k = 1) = n^{-1}$, and for $k \neq j$,

$$\mathbb{P}(I_k = 1, I_j = 1) = \mathbb{P}(I_j = 1 \mid I_k = 1)\mathbb{P}(I_k = 1) = \frac{1}{n(n-1)}.$$

Now $X = \sum_{k=1}^n I_k$, so that $\mathbb{E}(X) = \sum_{k=1}^n n^{-1} = 1$ and

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}\left(\sum_1^n I_k\right)^2 - 1 \\ &= \sum_1^n \mathbb{E}(I_k)^2 + \sum_{j\neq k} \mathbb{E}(I_j I_k) - 1 = 1 + 2\binom{n}{2}\frac{1}{n(n-1)} - 1 = 1.\end{aligned}$$

We have by the usual (mis)matching argument of Example (3.4.3) that

$$\mathbb{P}(X = r) = \frac{1}{r!}\sum_{i=0}^{n-r}\frac{(-1)^i}{i!}, \qquad 0 \le r \le n - 2,$$

which tends to $e^{-1}/r!$ as $n \to \infty$.

18. (a) Let $Y_1, Y_2, \dots, Y_n$ be Bernoulli with parameter p_2, and $Z_1, Z_2, \dots, Z_n$ Bernoulli with parameter p_1/p_2, and suppose the usual independence. Define $A_i = Y_i Z_i$, a Bernoulli random variable that has parameter $\mathbb{P}(A_i = 1) = \mathbb{P}(Y_i = 1)\mathbb{P}(Z_i = 1) = p_1$. Now $(A_1, A_2, \dots, A_n) \le (Y_1, Y_2, \dots, Y_n)$ so that $f(\mathbf{A}) \le f(\mathbf{Y})$. Hence $e(p_1) = \mathbb{E}(f(\mathbf{A})) \le \mathbb{E}(f(\mathbf{Y})) = e(p_2)$.

(b) Suppose first that $n = 1$, and let X and X' be independent Bernoulli variables with parameter p. We claim that

$$\{f(X) - f(X')\}\{g(X) - g(X')\} \ge 0;$$

to see this consider the three cases $X = X'$, $X < X'$, $X > X'$ separately, using the fact that f and g are increasing. Taking expectations, we obtain

$$\mathbb{E}\big(\{f(X) - f(X')\}\{g(X) - g(X')\}\big) \geq 0,$$

which may be expanded to find that

$$\begin{aligned} 0 &\leq \mathbb{E}\big(f(X)g(X)\big) - \mathbb{E}\big(f(X')g(X)\big) - \mathbb{E}\big(f(X)g(X')\big) + \mathbb{E}\big(f(X')g(X')\big) \\ &= 2\Big\{\mathbb{E}\big(f(X)g(X)\big) - \mathbb{E}(f(X))\mathbb{E}(g(X))\Big\} \end{aligned}$$

by the properties of X and X'.

Suppose that the result is valid for all n satisfying $n < k$ where $k \geq 2$. Now

$$(*) \qquad \mathbb{E}\big(f(\mathbf{X})g(\mathbf{X})\big) = \mathbb{E}\left\{\mathbb{E}\big(f(\mathbf{X})g(\mathbf{X}) \mid X_1, X_2, \ldots, X_{k-1}\big)\right\};$$

here, the conditional expectation given $X_1, X_2, \ldots, X_{k-1}$ is defined in very much the same way as in Definition (3.7.3), with broadly similar properties, in particular Theorem (3.7.4); see also Exercises (3.7.1, 3). If $X_1, X_2, \ldots, X_{k-1}$ are given, then $f(\mathbf{X})$ and $g(\mathbf{X})$ may be thought of as increasing functions of the single remaining variable X_k, and therefore

$$\mathbb{E}\big(f(\mathbf{X})g(\mathbf{X}) \mid X_1, X_2, \ldots, X_{k-1}\big) \geq \mathbb{E}\big(f(\mathbf{X}) \mid X_1, X_2, \ldots, X_{k-1}\big)\mathbb{E}\big(g(\mathbf{X}) \mid X_1, X_2, \ldots, X_{k-1}\big)$$

by the induction hypothesis. Furthermore

$$f'(\mathbf{X}) = \mathbb{E}\big(f(\mathbf{X}) \mid X_1, X_2, \ldots, X_{k-1}\big), \quad g'(\mathbf{X}) = \mathbb{E}\big(g(\mathbf{X}) \mid X_1, X_2, \ldots, X_{k-1}\big),$$

are increasing functions of the $k-1$ variables $X_1, X_2, \ldots, X_{k-1}$, implying by the induction hypothesis that $\mathbb{E}(f'(\mathbf{X})g'(\mathbf{X})) \geq \mathbb{E}(f'(\mathbf{X}))\mathbb{E}(g'(\mathbf{X}))$. We substitute this into $(*)$ to obtain

$$\mathbb{E}\big(f(\mathbf{X})g(\mathbf{X})\big) \geq \mathbb{E}(f'(\mathbf{X}))\mathbb{E}(g'(\mathbf{X})) = \mathbb{E}(f(\mathbf{X}))\mathbb{E}(g(\mathbf{X}))$$

by the definition of f' and g'.

19. Certainly $R(p) = \mathbb{E}(I_A) = \sum_\omega I_A(\omega)\mathbb{P}(\omega)$ and $\mathbb{P}(\omega) = p^{N(\omega)}q^{m-N(\omega)}$ where $p + q = 1$. Differentiating, we obtain

$$\begin{aligned} R'(p) &= \sum_\omega I_A(\omega)p^{N(\omega)}q^{m-N(\omega)}\left(\frac{N(\omega)}{p} - \frac{m - N(\omega)}{q}\right) \\ &= \frac{1}{pq}\sum_\omega I_A(\omega)\mathbb{P}(\omega)\big(N(\omega) - mp\big) \\ &= \frac{1}{pq}\mathbb{E}\big(I_A(N - mp)\big) = \frac{1}{pq}\{\mathbb{E}(I_A N) - \mathbb{E}(I_A)\mathbb{E}(N)\} = \frac{1}{pq}\operatorname{cov}(I_A, N). \end{aligned}$$

Applying the Cauchy–Schwarz inequality (3.6.9) to the latter covariance, we find that $R'(p) \leq (pq)^{-1}\sqrt{\operatorname{var}(I_A)\operatorname{var}(N)}$. However I_A is Bernoulli with parameter $R(p)$, so that $\operatorname{var}(I_A) = R(p)(1 - R(p))$, and finally N is bin(m, p) so that $\operatorname{var}(N) = mp(1 - p)$, whence the upper bound for $R'(p)$ follows.

As for the lower bound, use the general fact that $\operatorname{cov}(X+Y, Z) = \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z)$ to deduce that $\operatorname{cov}(I_A, N) = \operatorname{cov}(I_A, I_A) + \operatorname{cov}(I_A, N - I_A)$. Now I_A and $N - I_A$ are increasing functions of ω, in the sense of Problem (3.11.18); you should check this. Hence $\operatorname{cov}(I_A, N) \geq \operatorname{var}(I_A) + 0$ by the result of that problem. The lower bound for $R'(p)$ follows.

20. (a) Let each edge be *blue* with probability p_1 and *yellow* with probability p_2; assume these two events are independent of each other and of the colourings of all other edges. Call an edge *green* if it is both blue and yellow, so that each edge is *green* with probability $p_1 p_2$. If there is a working *green* connection from source to sink, then there is also a blue connection and a yellow connection. Thus

$$\begin{aligned}\mathbb{P}(\text{green connection}) &\le \mathbb{P}(\text{blue connection, and yellow connection})\\ &= \mathbb{P}(\text{blue connection})\mathbb{P}(\text{yellow connection})\end{aligned}$$

so that $R(p_1 p_2) \le R(p_1)R(p_2)$.

(b) This is somewhat harder, and may be proved by induction on the number n of edges of G. If $n = 1$ then a consideration of the two possible cases yields that either $R(p) = 1$ for all p, or $R(p) = p$ for all p. In either case the required inequality holds.

Suppose then that the inequality is valid whenever $n < k$ where $k \ge 2$, and consider the case when G has k edges. Let e be an edge of G and write $\omega(e)$ for the state of e; $\omega(e) = 1$ if e is working, and $\omega(e) = 0$ otherwise. Writing A for the event that there is a working connection from source to sink, we have that

$$\begin{aligned}R(p^\gamma) &= \mathbb{P}_{p^\gamma}(A \mid \omega(e) = 1)p^\gamma + \mathbb{P}_{p^\gamma}(A \mid \omega(e) = 0)(1 - p^\gamma)\\ &\le \mathbb{P}_p(A \mid \omega(e) = 1)^\gamma p^\gamma + \mathbb{P}_p(A \mid \omega(e) = 0)^\gamma (1 - p^\gamma)\end{aligned}$$

where $\mathbb{P}_\alpha$ is the appropriate probability measure when each edge is working with probability α. The inequality here is valid since, if $\omega(e)$ is given, then the network G is effectively reduced in size by one edge; the induction hypothesis is then utilized for the case $n = k - 1$. It is a minor chore to check that

$$x^\gamma p^\gamma + y^\gamma (1-p)^\gamma \le \{xp + y(1-p)\}^\gamma \qquad \text{if } x \ge y \ge 0;$$

to see this, check that equality holds when $x = y \ge 0$ and that the derivative of the left-hand side with respect to x is at most the corresponding derivative of the right-hand side when $x, y \ge 0$. Apply the latter inequality with $x = \mathbb{P}_p(A \mid \omega(e) = 1)$ and $y = \mathbb{P}_p(A \mid \omega(e) = 0)$ to obtain

$$R(p^\gamma) \le \big\{\mathbb{P}_p(A \mid \omega(e) = 1)p + \mathbb{P}_p(A \mid \omega(e) = 0)(1-p)\big\}^\gamma = R(p)^\gamma.$$

21. (a) The number X of such extraordinary individuals has the $\text{bin}(10^7, 10^{-7})$ distribution. Hence $\mathbb{E}X = 1$ and

$$\begin{aligned}\mathbb{P}(X > 1 \mid X \ge 1) &= \frac{\mathbb{P}(X > 1)}{\mathbb{P}(X > 0)} = \frac{1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1)}{1 - \mathbb{P}(X = 0)}\\ &= \frac{1 - (1 - 10^{-7})^{10^7} - 10^7 \cdot 10^{-7}(1 - 10^{-7})^{10^7 - 1}}{1 - (1 - 10^{-7})^{10^7}}\\ &\simeq \frac{1 - 2e^{-1}}{1 - e^{-1}} \simeq 0.4.\end{aligned}$$

(Shades of (3.5.4) here: X is approximately Poisson distributed with parameter 1.)

(b) Likewise

$$\mathbb{P}(X > 2 \mid X \ge 2) \simeq \frac{1 - 2e^{-1} - \frac{1}{2}e^{-1}}{1 - 2e^{-1}} \simeq 0.3.$$

(c) Provided $m \ll N = 10^7$,

$$\mathbb{P}(X = m) = \frac{N!}{m!\,(N-m)!}\left(\frac{1}{N}\right)^m \left(1 - \frac{1}{N}\right)^{N-m} \simeq \frac{e^{-1}}{m!},$$

the Poisson distribution. Assume that "reasonably confident that n is all" means that $\mathbb{P}(X > n \mid X \geq n) \leq r$ for some suitable small number r. Assuming the Poisson approximation, $\mathbb{P}(X > n) \leq r\mathbb{P}(X \geq n)$ if and only if

$$e^{-1} \sum_{k=n+1}^{\infty} \frac{1}{k!} \leq re^{-1} \sum_{k=n}^{\infty} \frac{1}{k!}.$$

For any given r, the smallest acceptable value of n may be determined numerically. If r is small, then very roughly $n \simeq 1/r$ will do (e.g., if $r = 0.05$ then $n \simeq 20$).

(d) No level p of improbability is sufficiently small for one to be sure that the person is specified uniquely. If $p = 10^{-7}\alpha$, then X is $\text{bin}(10^7, 10^{-7}\alpha)$, which is approximately Poisson with parameter α. Therefore, in this case,

$$\mathbb{P}(X > 1 \mid X \geq 1) \simeq \frac{1 - e^{-\alpha} - \alpha e^{-\alpha}}{1 - e^{-\alpha}} = \rho, \qquad \text{say.}$$

An acceptable value of ρ for a very petty offence might be $\rho \simeq 0.05$, in which case $\alpha \simeq 0.1$ and so $p = 10^{-8}$ might be an acceptable level of improbability. For a capital offence, one would normally require a much smaller value of ρ. We note that the rules of evidence do not allow an overt discussion along these lines in a court of law in the United Kingdom.

22. The number G of girls has the binomial distribution $\text{bin}(2n, p)$. Hence

$$\begin{aligned}\mathbb{P}(G \geq 2n - G) = \mathbb{P}(G \geq n) &= \sum_{k=n}^{2n} \binom{2n}{k} p^k q^{2n-k} \\ &\leq \binom{2n}{n} \sum_{k=n}^{\infty} p^k q^{2n-k} = \binom{2n}{n} p^n q^n \frac{q}{q-p},\end{aligned}$$

where we have used the fact that $\binom{2n}{k} \leq \binom{2n}{n}$ for all k.

With $p = 0.485$ and $n = 10^4$, we have using Stirling's formula (Exercise (3.10.1)) that

$$\begin{aligned}\binom{2n}{n} p^n q^n \frac{q}{q-p} &\leq \frac{1}{\sqrt{(n\pi)}} \{(1 - 0.03)(1 + 0.03)\}^n \frac{0.515}{0.03} \\ &= \frac{0.515}{3\sqrt{\pi}} \left(1 - \frac{9}{10^4}\right)^{10^4} \leq 1.23 \times 10^{-5}.\end{aligned}$$

It follows that the probability that boys outnumber girls for 82 successive years is at least $(1 - 1.23 \times 10^{-5})^{82} \geq 0.99899$.

23. Let M be the number of such visits. If $k \neq 0$, then $M \geq 1$ if and only if the particle hits 0 before it hits N, an event with probability $1 - kN^{-1}$ by equation (1.7.7). Having hit 0, the chance of another visit to 0 before hitting N is $1 - N^{-1}$, since the particle at 0 moves immediately to 1 whence there is probability $1 - N^{-1}$ of another visit to 0 before visiting N. Hence

$$\mathbb{P}(M \geq r \mid S_0 = k) = \left(1 - \frac{k}{N}\right)\left(1 - \frac{1}{N}\right)^{r-1}, \qquad r \geq 1,$$

so that

$$\begin{aligned}\mathbb{P}(M = j \mid S_0 = k) &= \mathbb{P}(M \geq j \mid S_0 = k) - \mathbb{P}(M \geq j + 1 \mid S_0 = 0) \\ &= \left(1 - \frac{k}{N}\right)\left(1 - \frac{1}{N}\right)^{j-1} \frac{1}{N}, \qquad j \geq 1.\end{aligned}$$

24. Either read the solution to Exercise (3.9.4), or the following two related solutions neither of which uses difference equations.

First method. Let T_k be the event that A wins and exactly k tails appear. Then $k < n$ so that $\mathbb{P}(\text{A wins}) = \sum_{k=0}^{n-1} \mathbb{P}(T_k)$. However $\mathbb{P}(T_k)$ is the probability that $m+k$ tosses yield m heads, k tails, and the last toss is heads. Hence

$$\mathbb{P}(T_k) = \binom{m+k-1}{m-1} p^m q^k,$$

whence the result follows.

Second method. Suppose the coin is tossed $m+n-1$ times. If the number of heads is m or more, then A must have won; conversely if the number of heads is $m-1$ or less, then the number of tails is n or more, so that B has won. The number of heads is $\text{bin}(m+n-1, p)$ so that

$$\mathbb{P}(\text{A wins}) = \sum_{k=m}^{m+n-1} \binom{m+n-1}{k} p^k q^{m+n-1-k}.$$

25. The chance of winning, having started from k, is

$$\frac{1-(q/p)^k}{1-(q/p)^N} \quad \text{which may be written as} \quad \frac{1-(q/p)^{\frac{1}{2}k}}{1-(q/p)^{\frac{1}{2}N}} \cdot \frac{1+(q/p)^{\frac{1}{2}k}}{1+(q/p)^{\frac{1}{2}N}};$$

see Example (3.9.6). If k and N are even, doubling the stake is equivalent to playing the original game with initial fortune $\frac{1}{2}k$ and the price of the Jaguar set at $\frac{1}{2}N$. The probability of winning is now

$$\frac{1-(q/p)^{\frac{1}{2}k}}{1-(q/p)^{\frac{1}{2}N}},$$

which is larger than before, since the final term in the above display is greater than 1 (when $p < \frac{1}{2}$).

If $p = \frac{1}{2}$, doubling the stake makes no difference to the chance of winning. If $p > \frac{1}{2}$, it is better to decrease the stake.

26. This is equivalent to taking the limit as $N \to \infty$ in the previous Problem (3.11.25). In the limit when $p \neq \frac{1}{2}$, the probability of ultimate bankruptcy is

$$\lim_{N\to\infty} \frac{(q/p)^k - (q/p)^N}{1-(q/p)^N} = \begin{cases} (q/p)^k & \text{if } p > \frac{1}{2}, \\ 1 & \text{if } p < \frac{1}{2}, \end{cases}$$

where $p+q=1$. If $p = \frac{1}{2}$, the corresponding limit is $\lim_{N\to\infty}(1-k/N) = 1$.

27. Using the technique of reversal, we have that

$$\begin{aligned} \mathbb{P}(R_n = R_{n-1}+1) &= \mathbb{P}(S_{n-1} \neq S_n, S_{n-2} \neq S_n, \dots, S_0 \neq S_n) \\ &= \mathbb{P}(X_n \neq 0, X_{n-1}+X_n \neq 0, \dots, X_1 + \cdots + X_n \neq 0) \\ &= \mathbb{P}(X_1 \neq 0, X_2 + X_1 \neq 0, \dots, X_n + \cdots + X_1 \neq 0) \\ &= \mathbb{P}(S_1 \neq 0, S_2 \neq 0, \dots, S_n \neq 0) = \mathbb{P}(S_1 S_2 \cdots S_n \neq 0). \end{aligned}$$

It follows that $\mathbb{E}(R_n) = \mathbb{E}(R_{n-1}) + \mathbb{P}(S_1 S_2 \cdots S_n \neq 0)$ for $n \geq 1$, whence

$$\frac{1}{n}\mathbb{E}(R_n) = \frac{1}{n}\left\{1 + \sum_{m=1}^{n} \mathbb{P}(S_1 S_2 \cdots S_m \neq 0)\right\} \to \mathbb{P}(S_k \neq 0 \text{ for all } k \geq 1)$$

since $\mathbb{P}(S_1 S_2 \cdots S_m \neq 0) \downarrow \mathbb{P}(S_k \neq 0 \text{ for all } k \geq 1)$ as $m \to \infty$.

There are various ways of showing that the last probability equals $|p - q|$, and here is one. Suppose $p > q$. If $X_1 = 1$, the probability of never subsequently hitting the origin equals $1 - (q/p)$, by the calculation in the solution to Problem (3.11.26) above. If $X_1 = -1$, the probability of staying away from the origin subsequently is 0. Hence the answer is $p(1 - (q/p)) + q \cdot 0 = p - q$.

If $q > p$, the same argument yields $q - p$, and if $p = q = \frac{1}{2}$ the answer is 0.

28. Consider first the event that M_{2n} is first attained at time $2k$. This event occurs if and only if: (i) the walk makes a first passage to S_{2k} (> 0) at time $2k$, and (ii) the walk thereafter does not exceed S_{2k}. These two events are independent. The chance of (i) is, by reversal and symmetry,

$$
\begin{aligned}
&\mathbb{P}(S_{2k-1} < S_{2k}, S_{2k-2} < S_{2k}, \dots, S_0 < S_{2k}) \\
&\quad = \mathbb{P}(X_{2k} > 0, X_{2k-1} + X_{2k} > 0, \dots, X_1 + \cdots + X_{2k} > 0) \\
&\quad = \mathbb{P}(X_1 > 0, X_1 + X_2 > 0, \dots, X_1 + \cdots + X_{2k} > 0) \\
&\quad = \mathbb{P}(S_i > 0 \text{ for } 1 \leq i \leq 2k) = \tfrac{1}{2}\mathbb{P}(S_i \neq 0 \text{ for } 1 \leq i \leq 2k) \\
&\quad = \tfrac{1}{2}\mathbb{P}(S_{2k} = 0) \qquad \text{by equation (3.10.23).}
\end{aligned}
$$

As for the second event, we may translate S_{2k} to the origin to obtain the probability of (ii):

$$
\mathbb{P}(S_{2k+1} \leq S_{2k}, \dots, S_{2n} \leq S_{2k}) = \mathbb{P}(M_{2n-2k} = 0) = \mathbb{P}(S_{2n-2k} = 0),
$$

where we have used the result of Exercise (3.10.2). The answer is therefore as given.

The probabilities of (i) and (ii) are unchanged in the case $i = 2k + 1$; the basic reason for this is that S_{2r} is even, and S_{2r+1} odd, for all r.

29. Let $u_k = \mathbb{P}(S_k = 0)$, $f_k = \mathbb{P}(S_k = 0, S_i \neq 0 \text{ for } 1 \leq i < k)$, and use conditional probability (or recall from equation (3.10.25)) to obtain

$$
(*) \qquad u_{2n} = \sum_{k=1}^{n} u_{2n-2k} f_{2k}.
$$

Now $N_1 = 2$, and therefore it suffices to prove that $\mathbb{E}(N_n) = \mathbb{E}(N_{n-1})$ for $n \geq 2$. Let N'_{n-1} be the number of points visited by the walk $S_1, S_2, \dots, S_n$ exactly once (we have removed S_0). Then

$$
N_n = \begin{cases} N'_{n-1} + 1 & \text{if } S_k \neq S_0 \text{ for } 1 \leq k \leq n, \\ N'_{n-1} - 1 & \text{if } S_k = S_0 \text{ for exactly one } k \text{ in } \{1, 2, \dots, n\}, \\ N'_{n-1} & \text{otherwise.} \end{cases}
$$

Hence, writing $\alpha_n = \mathbb{P}(S_k \neq 0 \text{ for } 1 \leq k \leq n)$,

$$
\begin{aligned}
\mathbb{E}(N_n) &= \mathbb{E}(N'_{n-1}) + \alpha_n - \mathbb{P}(S_k = S_0 \text{ exactly once}) \\
&= \mathbb{E}(N_{n-1}) + \alpha_n - \left\{ f_2 \alpha_{n-2} + f_4 \alpha_{n-4} + \cdots + f_{2\lfloor n/2 \rfloor} \right\}
\end{aligned}
$$

where $\lfloor x \rfloor$ is the integer part of x. Now $\alpha_{2m} = \alpha_{2m+1} = u_{2m}$ by equation (3.10.23). If $n = 2k$ is even, then

$$
\mathbb{E}(N_{2k}) - \mathbb{E}(N_{2k-1}) = u_{2k} - \{f_2 u_{2k-2} + \cdots + f_{2k}\} = 0 \qquad \text{by } (*).
$$

If $n = 2k + 1$ is odd, then

$$
\mathbb{E}(N_{2k+1}) - \mathbb{E}(N_{2k}) = u_{2k} - \left\{ f_2 u_{2k-2} + \cdots + f_{2k} \right\} = 0 \qquad \text{by } (*).
$$

In either case the claim is proved.

30. (a) Not much.

(b) The rhyme may be interpreted in any of several ways. Interpreting it as meaning that families stop at their first son, we may represent the sample space of a typical family as $\{B, GB, G^2B, \dots\}$, with $\mathbb{P}(G^nB) = 2^{-(n+1)}$. The mean number of girls is $\sum_{n=1}^{\infty} n\mathbb{P}(G^nB) = \sum_{n=1}^{\infty} n2^{-(n+1)} = 1$; there is exactly one boy.

The empirical sex ratio for large populations will be near to 1:1, by the law of large numbers. However the variance of the number of girls in a typical family is var(#girls) = 2, whilst var(#boys) = 0; $\#A$ denotes the cardinality of A. Considerable variation from 1:1 is therefore possible in smaller populations, but in either direction. In a large number of small populations, the number of large predominantly female families would be balanced by a large number of male singletons.

31. Any positive integer m has a unique factorization in the form $m = \prod_i p_i^{m(i)}$ for non-negative integers $m(1), m(2), \dots$. Hence,

$$\mathbb{P}(M = m) = \prod_i \mathbb{P}\big(N(i) = m(i)\big) = \prod_i \left(1 - \frac{1}{p_i^{\beta}}\right) \frac{1}{p_i^{\beta m(i)}} = C\left(\prod_i p_i^{-m(i)}\right)^{\beta} = \frac{C}{m^{\beta}}$$

where $C = \prod_i (1 - p_i^{-\beta})$. Now $\sum_m \mathbb{P}(M = m) = 1$, so that $C^{-1} = \sum_m m^{-\beta}$.

32. Number the plates $0, 1, 2, \dots, N$ where 0 is the starting plate, fix k satisfying $0 < k \le N$, and let A_k be the event that plate number k is the last to be visited. In order to calculate $\mathbb{P}(A_k)$, we cut the table open at k, and bend its outside edge into a line segment, along which the plate numbers read $k, k+1, \dots, N, 0, 1, \dots, k$ in order. It is convenient to relabel the plates as $-(N+1-k), -(N-k), \dots, -1, 0, 1, \dots, k$. Now A_k occurs if and only if a symmetric random walk, starting from 0, visits both $-(N-k)$ and $k-1$ *before* it visits either $-(N+1-k)$ or k. Suppose it visits $-(N-k)$ before it visits $k-1$. The (conditional) probability that it subsequently visits $k-1$ before visiting $-(N+1-k)$ is the same as the probability that a symmetric random walk, starting from 1, hits N before it hits 0, a probability of N^{-1} by (1.7.7). The same argument applies if the cake visits $k-1$ before it visits $-(N-k)$. Therefore $\mathbb{P}(A_k) = N^{-1}$.

33. With j denoting the jth best vertex, the walk has transition probabilities $p_{jk} = (j-1)^{-1}$ for $1 \le k < j$. By conditional expectation,

$$r_j = 1 + \frac{1}{j-1}\sum_{k=1}^{j-1} r_k, \qquad r_1 = 0.$$

Induction now supplies the result. Since $r_j \sim \log j$ for large j, the worst-case expectation is about $\log\binom{n}{m}$.

34. Let p_n denote the required probability. If (m_r, m_{r+1}) is first pair to make a dimer, then m_1 is ultimately uncombined with probability p_{r-1}. By conditioning on the first pair, we find that $p_n = (p_1 + p_2 + \dots + p_{n-2})/(n-1)$, giving $n(p_{n+1} - p_n) = -(p_n - p_{n-1})$. Therefore, $n!(p_{n+1} - p_n) = (-1)^{n-1}(p_2 - p_1) = (-1)^n$, and the claim follows by summing.

Finally,

$$\mathbb{E}U_n = \sum_{r=1}^{n} \mathbb{P}(m_r \text{ is uncombined}) = p_n + p_1 p_{n-1} + \dots + p_{n-1}p_1 + p_n,$$

since the rth molecule may be thought of as an end molecule of two sequences of length r and $n-r+1$. Now $p_n \to e^{-1}$ as $n \to \infty$, and it is an easy exercise of analysis to obtain that $n^{-1}\mathbb{E}U_n \to e^{-2}$.

35. First,

$$\lambda^k = \left(\sum_i p_i\right)^k = \sum_{r_1, r_2, \dots, r_k} p_{r_1} p_{r_2} \cdots p_{r_k} \ge k! \sum_{\{r_1, \dots, r_k\}} p_{r_1} p_{r_2} \cdots p_{r_k},$$

where the last summation is over all subsets $\{r_1, \dots, r_k\}$ of k distinct elements of $\{1, 2, \dots, n\}$. Secondly,

$$\begin{aligned}\lambda^k &\le k! \sum_{\{r_1, \dots, r_k\}} p_{r_1} p_{r_2} \cdots p_{r_k} + \binom{k}{2} \sum_i p_i^2 \sum_{r_1, \dots, r_{k-2}} p_{r_1} p_{r_2} \cdots p_{r_{k-2}} \\ &\le k! \sum_{r_1, \dots, r_k} p_{r_1} p_{r_2} \cdots p_{r_k} + \binom{k}{2} \max_i p_i \left(\sum_j p_j\right)^{k-1}.\end{aligned}$$

Hence

$$(*) \qquad \sum_{\{r_1, \dots, r_k\}} p_{r_1} p_{r_2} \cdots p_{r_k} = \frac{\lambda^k}{k!} \left\{1 + \mathrm{O}\left(\frac{k^2}{\lambda} \max_i p_i\right)\right\}.$$

By Taylor's theorem applied to the function $\log(1-x)$, there exist θ_r satisfying $0 < \theta_r < \{2(1-c)^2\}^{-1}$ such that

$$(**) \qquad \prod_{r=1}^n (1 - p_r) = \prod_r \exp\{-p_r - \theta_r p_r^2\} = \exp\left\{-\lambda - \lambda \mathrm{O}\left(\max_i p_i\right)\right\}.$$

Finally,

$$\mathbb{P}(X = k) = \left(\prod_r (1 - p_r)\right) \sum_{\{r_1, \dots, r_k\}} \frac{p_{r_1} \cdots p_{r_k}}{(1 - p_{r_1}) \cdots (1 - p_{r_k})}.$$

The claim follows from $(*)$ and $(**)$.

36. It is elementary that

$$\mathbb{E}(\overline{Y}) = \frac{1}{n} \sum_{r=1}^N \mathbb{E}(X_r) = \frac{1}{n} \sum_{r=1}^N x_r \cdot \frac{n}{N} = \mu.$$

We write $\overline{Y} - \mathbb{E}(\overline{Y})$ as the mixture of indicator variables thus:

$$\overline{Y} - \mathbb{E}(\overline{Y}) = \sum_{r=1}^N \frac{x_r}{n} \left(I_r - \frac{n}{N}\right).$$

It follows from the fact

$$\mathbb{E}(I_i I_j) = \frac{n}{N} \cdot \frac{n-1}{N-1}, \qquad i \ne j,$$

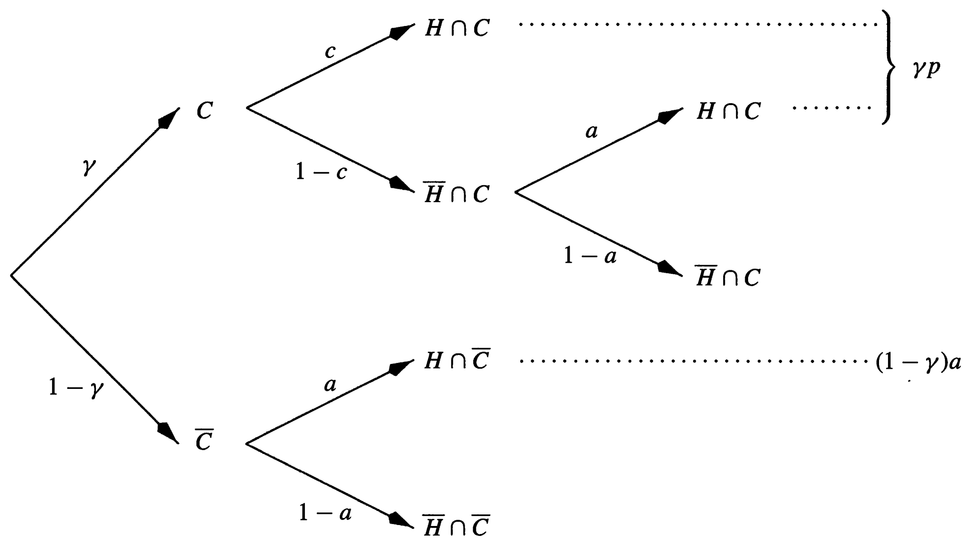


Figure 3.2. The tree of possibility and probability in Problem (3.11.37). The presence of the disease is denoted by C, and hospitalization by H; their negations are denoted by $\overline{C}$ and $\overline{H}$.

that

$$\begin{aligned}
\operatorname{var}(\overline{Y}) &= \sum_{r=1}^{N} \frac{x_r^2}{n^2} \mathbb{E}\left\{\left(I_r - \frac{n}{N}\right)^2\right\} + \sum_{i \neq j} \frac{x_i x_j}{n^2} \mathbb{E}\left\{\left(I_i - \frac{n}{N}\right)\left(I_j - \frac{n}{N}\right)\right\} \\
&= \sum_{r=1}^{N} \frac{x_r^2}{n^2} \frac{n}{N}\left(1 - \frac{n}{N}\right) + \sum_{i \neq j} \frac{x_i x_j}{n^2} \left\{\frac{n}{N}\frac{n-1}{N-1} - \frac{n^2}{N^2}\right\} \\
&= \sum_{r=1}^{N} x_r^2 \frac{N-n}{N^2 n} - \sum_{i \neq j} x_i x_j \frac{N-n}{n(N-1)N^2} \\
&= \frac{N-n}{Nn(N-1)}\left\{\sum_{r=1}^{N} x_r^2 - \frac{1}{N}\sum_{r=1}^{N} x_r^2 - \frac{1}{N}\sum_{i \neq j} x_i x_j\right\} \\
&= \frac{N-n}{n(N-1)}\frac{1}{N}\left\{\sum_{r=1}^{N} x_r^2 - N\overline{x}^2\right\} = \frac{N-n}{n(N-1)}\frac{1}{N}\sum_{r=1}^{N}(x_r - \overline{x})^2.
\end{aligned}$$

37. The tree in Figure 3.2 illustrates the possibilities and probabilities. If G contains n individuals, X is $\operatorname{bin}(n, \gamma p + (1-\gamma)a)$ and Y is $\operatorname{bin}(n, \gamma p)$. It is not difficult to see that $\operatorname{cov}(X, Y) = n\gamma p(1-\nu)$ where $\nu = \gamma p + (1-\gamma)a$. Also, $\operatorname{var}(Y) = n\gamma p(1-\gamma p)$ and $\operatorname{var}(X) = nk(1-\nu)$. The result follows from the definition of correlation.

38. (a) This is an extension of Exercise (3.5.2). With $\mathbb{P}_n$ denoting the probability measure conditional on $N = n$, we have that

$$\mathbb{P}_n(X_i = r_i \text{ for } 1 \le i \le k) = \frac{n!}{r_1!\, r_2! \cdots r_k!\, s!} f(1)^{r_1} f(2)^{r_2} \cdots f(k)^{r_k} (1 - F(k))^s,$$

where $s = n - \sum_{i=1}^{k} r_i$. Therefore,

$$\begin{aligned}\mathbb{P}(X_i = r_i \text{ for } 1 \le i \le k) &= \sum_{n=0}^{\infty} \mathbb{P}_n(X_i = r_i \text{ for } 1 \le i \le k)\mathbb{P}(N = n)\\ &= \prod_{i=1}^{k} \left\{ \frac{\nu^{r_i} f(i)^{r_i} e^{-\nu f(i)}}{r_i!} \right\} \sum_{s=0}^{\infty} \frac{\nu^s (1 - F(k))^s}{s!} e^{-\nu(1-F(k))}.\end{aligned}$$

The final sum is a Poisson sum, and equals 1.
(b) We use an argument relevant to Wald's equation. The event $\{T \le n-1\}$ depends only on the random variables $X_1, X_2, \dots, X_{n-1}$, and these are independent of X_n. It follows that X_n is independent of the event $\{T \ge n\} = \{T \le n-1\}^{\mathrm{c}}$. Hence,

$$\begin{aligned}\mathbb{E}(S) &= \sum_{i=1}^{\infty} \mathbb{E}\big(X_i I_{\{T \ge i\}}\big) = \sum_{i=1}^{\infty} \mathbb{E}(X_i)\mathbb{E}(I_{\{T\ge i\}}) = \sum_{i=1}^{\infty} \mathbb{E}(X_i)\mathbb{P}(T \ge i)\\ &= \sum_{i=1}^{\infty} \nu f(i) \sum_{t=i}^{\infty} \mathbb{P}(T = t) = \nu \sum_{t=1}^{\infty} \mathbb{P}(T = t) \sum_{i=1}^{t} f(i)\\ &= \nu \sum_{t=1}^{\infty} \mathbb{P}(T = t) F(t) = \mathbb{E}(F(T)).\end{aligned}$$

39. (a) Place an absorbing barrier at $a + 1$, and let p_a be the probability that the particle is absorbed at 0. By conditioning on the first step, we obtain that

$$p_n = \frac{1}{n+2}(p_0 + p_1 + p_2 + \cdots + p_{n+1}), \qquad 1 \le n \le a.$$

The boundary conditions are $p_0 = 1$, $p_{a+1} = 0$. It follows that $p_{n+1} - p_n = (n+1)(p_n - p_{n-1})$ for $2 \le n \le a$. We have also that $p_2 - p_1 = p_1 - 1$, and

$$p_{n+1} - p_n = \tfrac{1}{2}(n+1)!\,(p_2 - p_1) = \tfrac{1}{2}(n+1)!\,(p_1 - p_0).$$

Setting $n = a$ we obtain that $-p_a = \frac{1}{2}(a+1)!\,(p_1 - 1)$. By summing over $2 \le n < a$,

$$p_a - p_1 = (p_1 - p_0) + \tfrac{1}{2}(p_1 - p_0) \sum_{j=3}^{a} j!,$$

and we eliminate p_1 to conclude that

$$p_a = \frac{(a+1)!}{4 + 3! + 4! + \cdots + (a+1)!}.$$

It is now easy to see that, for given r, $p_r = p_r(a) \to 1$ as $a \to \infty$, so that ultimate absorption at 0 is (almost) certain, irrespective of the starting point.
(b) Let λ_r be the probability that the last step is from 1 to 0, having started at r. Then

$$(*) \qquad \lambda_1 = \tfrac{1}{3}(1 + \lambda_1 + \lambda_2),$$

$$(**) \qquad (r+2)\lambda_r = \lambda_1 + \lambda_2 + \cdots + \lambda_{r+1}, \qquad r \ge 2.$$

It follows that

$$\lambda_r - \lambda_{r-1} = \frac{1}{r+1}(\lambda_{r+1} - \lambda_r), \qquad r \geq 3,$$

whence

$$\lambda_3 - \lambda_2 = \frac{1}{4 \cdot 5 \cdots (r+1)}(\lambda_{r+1} - \lambda_r), \qquad r \geq 3.$$

Letting $r \to \infty$, we deduce that $\lambda_3 = \lambda_2$ so that $\lambda_r = \lambda_2$ for $r \geq 2$. From (**) with $r = 2$, $\lambda_2 = \frac{1}{2}\lambda_1$, and from (*) $\lambda_1 = \frac{2}{3}$.

(c) Let μ_r be the mean duration of the walk starting from r. As above, $\mu_0 = 0$, and

$$\mu_r = 1 + \frac{1}{r+2}(\mu_1 + \mu_2 + \cdots + \mu_{r+1}), \qquad r \geq 1,$$

whence $\mu_{r+1} - \mu_r = (r+1)(\mu_r - \mu_{r-1}) - 1$ for $r \geq 2$. Therefore, $v_{r+1} = (\mu_{r+1} - \mu_r)/(r+1)!$ satisfies $v_{r+1} - v_r = -1/(r+1)!$ for $r \geq 2$, and some further algebra yields the value of μ_1.

40. We label the vertices $1, 2, \ldots, n$, and we let π be a random permutation of this set. Let K be the set of vertices v with the property that $\pi(w) > \pi(v)$ for all neighbours w of v. It is not difficult to see that K is an independent set, whence $\alpha(G) \geq |K|$. Therefore, $\alpha(G) \geq \mathbb{E}|K| = \sum_v \mathbb{P}(v \in K)$. For any vertex v, a random permutation π is equally likely to assign any given ordering to the set comprising v and its neighbours. Also, $v \in K$ if and only if v is the earliest element in this ordering, whence $\mathbb{P}(v \in K) = 1/(d_v + 1)$. The result follows.

4

Continuous random variables

4.1 Solutions. Probability density functions

1. (a) $\{x(1-x)\}^{-\frac{1}{2}}$ is the derivative of $\sin^{-1}(2x-1)$, and therefore $C = \pi^{-1}$.
(b) $C = 1$, since

$$\int_{-\infty}^{\infty} \exp(-x - e^{-x})\, dx = \lim_{K\to\infty} \big[\exp(-e^{-x})\big]_{-K}^{K} = 1.$$

(c) Substitute $v = (1+x^2)^{-1}$ to obtain

$$\int_{-\infty}^{\infty} \frac{dx}{(1+x^2)^m} = \int_0^1 v^{m-\frac{3}{2}}(1-v)^{-\frac{1}{2}}\, dv = B(\tfrac{1}{2}, m - \tfrac{1}{2})$$

where $B(\cdot,\cdot)$ is a beta function; see paragraph (4.4.8) and Exercise (4.4.2). Hence, if $m > \frac{1}{2}$,

$$C^{-1} = B(\tfrac{1}{2}, m - \tfrac{1}{2}) = \frac{\Gamma(\frac{1}{2})\Gamma(m - \frac{1}{2})}{\Gamma(m)}.$$

2. (i) The distribution function F_Y of Y is

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(aX \le y) = \mathbb{P}(X \le y/a) = F_X(y/a).$$

So, differentiating, $f_Y(y) = a^{-1} f_X(y/a)$.
(ii) Certainly

$$F_{-X}(x) = \mathbb{P}(-X \le x) = \mathbb{P}(X \ge -x) = 1 - \mathbb{P}(X \le -x)$$

since $\mathbb{P}(X = -x) = 0$. Hence $f_{-X}(x) = f_X(-x)$. If X and $-X$ have the same distribution function then $f_{-X}(x) = f_X(x)$, whence the claim follows. Conversely, if $f_X(-x) = f_X(x)$ for all x, then, by substituting $u = -x$,

$$\mathbb{P}(-X \le y) = \mathbb{P}(X \ge -y) = \int_{-y}^{\infty} f_X(x)\, dx = \int_{-\infty}^{y} f_X(-u)\, du = \int_{-\infty}^{y} f_X(u)\, du = \mathbb{P}(X \le y),$$

whence X and $-X$ have the same distribution function.

3. Since $\alpha \ge 0$, $f \ge 0$, and $g \ge 0$, it follows that $\alpha f + (1-\alpha)g \ge 0$. Also

$$\int_{\mathbb{R}} \{\alpha f + (1-\alpha)g\}\, dx = \alpha \int_{\mathbb{R}} f\, dx + (1-\alpha)\int_{\mathbb{R}} g\, dx = \alpha + 1 - \alpha = 1.$$

If X is a random variable with density f, and Y a random variable with density g, then $\alpha f+(1-\alpha)g$ is the density of a random variable Z which takes the value X with probability α and Y otherwise.

Some minor technicalities are necessary in order to find an appropriate probability space for such a Z. If X and Y are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, it is necessary to define the product space $(\Omega, \mathcal{F}, \mathbb{P}) \times (\Sigma, \mathcal{G}, \mathbb{Q})$ where $\Sigma = \{0, 1\}$, $\mathcal{G}$ is the set of all subsets of Σ, and $\mathbb{Q}(0) = \alpha$, $\mathbb{Q}(1) = 1 - \alpha$. For $\omega \times \sigma \in \Omega \times \Sigma$, we define

$$Z(\omega \times \sigma) = \begin{cases} X(\omega) & \text{if } \sigma = 0, \\ Y(\omega) & \text{if } \sigma = 1. \end{cases}$$

4. (a) By definition, $r(x) = \lim_{h \downarrow 0} \dfrac{1}{h} \dfrac{F(x+h) - F(x)}{1 - F(x)} = \dfrac{f(x)}{1 - F(x)}$.

(b) We have that

$$\frac{H(x)}{x} = \frac{d}{dx}\left\{\frac{1}{x}\int_0^x r(y)\,dy\right\} = \frac{r(x)}{x} - \frac{1}{x^2}\int_0^x r(y)\,dy = \frac{1}{x^2}\int_0^x [r(x) - r(y)]\,dy,$$

which is non-negative if r is non-increasing.

(c) $H(x)/x$ is non-decreasing if and only if, for $0 \le \alpha \le 1$,

$$\frac{1}{\alpha x} H(\alpha x) \le \frac{1}{x} H(x) \qquad \text{for all } x \ge 0,$$

which is to say that $-\alpha^{-1}\log[1 - F(\alpha x)] \le -\log[1 - F(x)]$. We exponentiate to obtain the claim.

(d) Likewise, if $H(x)/x$ is non-decreasing then $H(\alpha t) \le \alpha H(t)$ for $0 \le \alpha \le 1$ and $t \ge 0$, whence $H(\alpha t) + H(t - \alpha t) \le H(t)$ as required.

4.2 Solutions. Independence

1. Let N be the required number. Then $\mathbb{P}(N = n) = F(K)^{n-1}[1 - F(K)]$ for $n \ge 1$, the geometric distribution with mean $[1 - F(K)]^{-1}$.

2. (i) $\text{Max}\{X, Y\} \le v$ if and only if $X \le v$ and $Y \le v$. Hence, by independence,

$$\mathbb{P}\big(\max\{X, Y\} \le v\big) = \mathbb{P}(X \le v, Y \le v) = \mathbb{P}(X \le v)\mathbb{P}(Y \le v) = F(v)^2.$$

Differentiate to obtain the density function of $V = \max\{X, Y\}$.

(ii) Similarly $\min\{X, Y\} > u$ if and only if $X > u$ and $Y > u$. Hence

$$\mathbb{P}(U \le u) = 1 - \mathbb{P}(U > u) = 1 - \mathbb{P}(X > u)\mathbb{P}(Y > u) = 1 - [1 - F(u)]^2,$$

giving $f_U(u) = 2f(u)[1 - F(u)]$.

3. The 24 permutations of the order statistics are equally likely by symmetry, and thus have equal probability. Hence $\mathbb{P}(X_1 < X_2 < X_3 < X_4) = \frac{1}{24}$, and $\mathbb{P}(X_1 > X_2 < X_3 < X_4) = \frac{3}{24}$, by enumerating the possibilities.

4. $\mathbb{P}(Y(y) > k) = F(y)^k$ for $k \ge 1$. Hence $\mathbb{E}Y(y) = F(y)/[1 - F(y)] \to \infty$ as $y \to \infty$. Therefore,

$$\begin{aligned} \mathbb{P}\big(Y(y) > \mathbb{E}Y(y)\big) &= \{1 - [1 - F(y)]\}^{\lfloor F(y)/[1-F(y)]\rfloor} \\ &\sim \exp\left\{1 - [1 - F(y)]\left\lfloor \frac{F(y)}{1 - F(y)} \right\rfloor\right\} \to e^{-1} \quad \text{as } y \to \infty. \end{aligned}$$

4.3 Solutions. Expectation

1. (a) $\mathbb{E}(X^\alpha) = \int_0^\infty x^\alpha e^{-x}\,dx < \infty$ if and only if $\alpha > -1$.
(b) In this case

$$\mathbb{E}(|X|^\alpha) = \int_{-\infty}^{\infty} \frac{C|x|^\alpha}{(1+x^2)^m}\,dx < \infty$$

if and only if $-1 < \alpha < 2m-1$.

2. We have that

$$1 = \mathbb{E}\left(\frac{\sum_1^n X_i}{S_n}\right) = \sum_{i=1}^n \mathbb{E}(X_i/S_n).$$

By symmetry, $\mathbb{E}(X_i/S_n) = \mathbb{E}(X_1/S_n)$ for all i, and hence $1 = n\mathbb{E}(X_1/S_n)$. Therefore

$$\mathbb{E}(S_m/S_n) = \sum_{i=1}^m \mathbb{E}(X_i/S_n) = m\mathbb{E}(X_1/S_n) = m/n.$$

3. Either integrate by parts or use Fubini's theorem:

$$\begin{aligned} r\int_0^\infty x^{r-1}\mathbb{P}(X > x)\,dx &= r\int_0^\infty x^{r-1}\left\{\int_{y=x}^\infty f(y)\,dy\right\}dx \\ &= \int_{y=0}^\infty f(y)\left\{\int_{x=0}^y rx^{r-1}\,dx\right\}dy = \int_0^\infty y^r f(y)\,dy. \end{aligned}$$

An alternative proof is as follows. Let I_x be the indicator of the event that $X > x$, so that $\int_0^\infty I_x\,dx = X$. Taking expectations, and taking a minor liberty with the integral which may be made rigorous, we obtain $\mathbb{E}X = \int_0^\infty \mathbb{E}(I_x)\,dx$. A similar argument may be used for the more general case.

4. We may suppose without loss of generality that $\mu = 0$ and $\sigma = 1$. Assume further that $m > 1$. In this case, at least half the probability mass lies to the right of 1, whence $\mathbb{E}(XI_{\{X\ge m\}}) \ge \frac{1}{2}$. Now $0 = \mathbb{E}(X) = \mathbb{E}\{X[I_{\{X\ge m\}} + I_{\{X<m\}}]\}$, implying that $\mathbb{E}(XI_{\{X<m\}}) \le -\frac{1}{2}$. Likewise,

$$\mathbb{E}(X^2 I_{\{X\ge m\}}) \ge \tfrac{1}{2}, \quad \mathbb{E}(X^2 I_{\{X<m\}}) \le \tfrac{1}{2}.$$

By the definition of the median, and the fact that X is continuous,

$$\mathbb{E}(X \mid X < m) \le -1, \quad \mathbb{E}(X^2 \mid X < m) \le 1.$$

It follows that $\operatorname{var}(X \mid X < m) \le 0$, which implies in turn that, conditional on $\{X < m\}$, X is almost surely concentrated at a single value. This contradicts the continuity of X, and we deduce that $m \le 1$. The possibility $m < -1$ may be ruled out similarly, or by considering the random variable $-X$.

5. It is a standard to write $X = X^+ - X^-$ where $X^+ = \max\{X, 0\}$ and $X^- = -\min\{X, 0\}$. Now X^+ and X^- are non-negative, and so, by Lemma (4.3.4),

$$\begin{aligned} \mu = \mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) &= \int_0^\infty \mathbb{P}(X > x)\,dx - \int_0^\infty \mathbb{P}(X < -x)\,dx \\ &= \int_0^\infty [1 - F(x)]\,dx - \int_0^\infty F(-x)\,dx = \int_0^\infty [1-F(x)]\,dx - \int_{-\infty}^0 F(x)\,dx. \end{aligned}$$

It is a triviality that

$$\mu = \int_0^\mu F(x)\,dx + \int_0^\mu [1 - F(x)]\,dx$$

and the equation follows with $a = \mu$. It is easy to see that it cannot hold with any other value of a, since both sides are monotonic functions of a.

4.4 Solutions. Examples of continuous variables

1. (i) Integrating by parts,

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx = (t-1)\int_0^\infty x^{t-2}e^{-x}\,dx = (t-1)\Gamma(t-1).$$

If n is an integer, then it follows that $\Gamma(n) = (n-1)\Gamma(n-1) = \cdots = (n-1)!\,\Gamma(1)$ where $\Gamma(1) = 1$.
(ii) We have, using the substitution $u^2 = x$, that

$$\begin{aligned}\Gamma(\tfrac{1}{2})^2 &= \left\{\int_0^\infty x^{-\frac{1}{2}}e^{-x}\,dx\right\}^2 = \left\{\int_0^\infty 2e^{-u^2}\,du\right\}^2 \\ &= 4\int_0^\infty e^{-u^2}\,du\int_0^\infty e^{-v^2}\,dv = 4\int_{r=0}^\infty\int_{\theta=0}^{\pi/2} e^{-r^2}r\,dr\,d\theta = \pi\end{aligned}$$

as required. For integral n,

$$\Gamma(n+\tfrac{1}{2}) = (n-\tfrac{1}{2})\Gamma(n-\tfrac{1}{2}) = \cdots = (n-\tfrac{1}{2})(n-\tfrac{3}{2})\cdots\tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \frac{(2n)!}{4^n n!}\sqrt{\pi}.$$

2. By the definition of the gamma function,

$$\Gamma(a)\Gamma(b) = \int_0^\infty x^{a-1}e^{-x}\,dx\int_0^\infty y^{b-1}e^{-y}\,dy = \int_0^\infty\int_0^\infty e^{-(x+y)}x^{a-1}y^{b-1}\,dx\,dy.$$

Now set $u = x+y$, $v = x/(x+y)$, obtaining

$$\begin{aligned}&\int_{u=0}^\infty\int_{v=0}^1 e^{-u}u^{a+b-1}v^{a-1}(1-v)^{b-1}\,dv\,du \\ &\qquad = \int_0^\infty u^{a+b-1}e^{-u}\,du\int_0^1 v^{a-1}(1-v)^{b-1}\,dv = \Gamma(a+b)B(a,b).\end{aligned}$$

3. If g is strictly decreasing then $\mathbb{P}(g(X) \le y) = \mathbb{P}(X \ge g^{-1}(y)) = 1 - g^{-1}(y)$ so long as $0 \le g^{-1}(y) \le 1$. Therefore $\mathbb{P}(g(X) \le y) = 1 - e^{-y}$, $y \ge 0$, if and only if $g^{-1}(y) = e^{-y}$, which is to say that $g(x) = -\log x$ for $0 < x < 1$.

4. We have that

$$\mathbb{P}(X \le x) = \int_{-\infty}^x \frac{1}{\pi(1+u^2)}\,du = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x.$$

Also,

$$\mathbb{E}(|X|^\alpha) = \int_{-\infty}^\infty \frac{|x|^\alpha}{\pi(1+x^2)}\,dx$$

is finite if and only if $|\alpha| < 1$.

5. Writing Φ for the $N(0,1)$ distribution function, $\mathbb{P}(Y \le y) = \mathbb{P}(X \le \log y) = \Phi(\log y)$. Hence

$$f_Y(y) = \frac{1}{y}f_X(\log y) = \frac{1}{y\sqrt{2\pi}}e^{-\frac{1}{2}(\log y)^2}, \qquad 0 < y < \infty.$$

6. Integrating by parts,

$$\begin{aligned}\text{LHS} &= \int_{-\infty}^\infty g(x)\left\{(x-\mu)\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\right\}dx \\ &= -\left[g(x)\sigma\phi\left(\frac{x-\mu}{\sigma}\right)\right]_{-\infty}^\infty + \int_{-\infty}^\infty g'(x)\sigma\phi\left(\frac{x-\mu}{\sigma}\right)dx = \text{RHS}.\end{aligned}$$

7. (a) $r(x) = \alpha\beta x^{\beta-1}$.
(b) $r(x) = \lambda$.
(c) $r(x) = \dfrac{\lambda\alpha e^{-\lambda x} + \mu(1-\alpha)e^{-\mu x}}{\alpha e^{-\lambda x} + (1-\alpha)e^{-\mu x}}$, which approaches $\min\{\lambda, \mu\}$ as $x \to \infty$.

8. Clearly $\phi' = -x\phi$. Using this identity and integrating by parts repeatedly,

$$\begin{aligned}1 - \Phi(x) &= \int_x^\infty \phi(u)\,du = -\int_x^\infty \frac{\phi'(u)}{u}\,du = \frac{\phi(x)}{x} + \int_x^\infty \frac{\phi'(u)}{u^3}\,du \\ &= \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} - \int_x^\infty \frac{3\phi'(u)}{u^5}\,du = \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} + \frac{3\phi(x)}{x^5} - \int_x^\infty \frac{15\phi(u)}{u^6}\,du.\end{aligned}$$

4.5 Solutions. Dependence

1. (i) As the product of non-negative continuous functions, f is non- negative and continuous. Also

$$g(x) = \tfrac{1}{2}e^{-|x|}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi x^{-2}}} e^{-\frac{1}{2}x^2y^2}\,dy = \tfrac{1}{2}e^{-|x|}$$

if $x \neq 0$, since the integrand is the $N(0, x^{-2})$ density function. It is easily seen that $g(0) = 0$, so that g is discontinuous, while

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty} \tfrac{1}{2}e^{-|x|}\,dx = 1.$$

(ii) Clearly $f_Q \geq 0$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_Q(x, y)\,dx\,dy = \sum_{n=1}^{\infty} \left(\tfrac{1}{2}\right)^n \cdot 1 = 1.$$

Also f_Q is the uniform limit of continuous functions on any subset of $\mathbb{R}^2$ of the form $[-M, M] \times \mathbb{R}$; hence f_Q is continuous. Hence f_Q is a continuous density function. On the other hand

$$\int_{-\infty}^{\infty} f_Q(x, y)\,dy = \sum_{n=1}^{\infty} \left(\tfrac{1}{2}\right)^n g(x - q_n),$$

where g is discontinuous at 0.
(iii) Take Q to be the set of the rationals, in some order.

2. We may assume that the centre of the rod is uniformly positioned in a square of size $a \times b$, while the acute angle between the rod and a line of the first grid is uniform on $[0, \frac{1}{2}\pi]$. If the latter angle is θ then, with the aid of a diagram, one finds that there is no intersection if and only if the centre of the rod lies within a certain inner rectangle of size $(a - r\cos\theta) \times (b - r\sin\theta)$. Hence the probability of an intersection is

$$\frac{2}{\pi ab}\int_0^{\pi/2} \{ab - (a - r\cos\theta)(b - r\sin\theta)\}\,d\theta = \frac{2r}{\pi ab}(a + b - \tfrac{1}{2}r).$$

3. (i) Let I be the indicator of the event that the first needle intersects a line, and let J be the indicator that the second needle intersects a line. By the result of Exercise (4.5.2), $\mathbb{E}(I) = \mathbb{E}(J) = 2/\pi$; hence $Z = I + J$ satisfies $\mathbb{E}(\frac{1}{2}Z) = 2/\pi$.

(ii) We have that

$$\begin{aligned}\operatorname{var}(\tfrac{1}{2}Z) &= \tfrac{1}{4}\{\mathbb{E}(I^2)+\mathbb{E}(J^2)+2\mathbb{E}(IJ)\} - \mathbb{E}(\tfrac{1}{2}Z)^2 \\ &= \tfrac{1}{4}\{\mathbb{E}(I)+\mathbb{E}(J)+2\mathbb{E}(IJ)\} - \frac{4}{\pi^2} = \frac{1}{\pi} - \frac{4}{\pi^2} + \frac{1}{2}\mathbb{E}(IJ).\end{aligned}$$

In the notation of (4.5.8), if $0 < \theta < \frac{1}{2}\pi$, then two intersections occur if $z < \frac{1}{2}\min\{\sin\theta, \cos\theta\}$ or $1 - z < \frac{1}{2}\min\{\sin\theta, \cos\theta\}$. With a similar component when $\frac{1}{2}\pi \le \theta < \pi$, we find that

$$\begin{aligned}\mathbb{E}(IJ) = \mathbb{P}(\text{two intersections}) &= \frac{4}{\pi}\iint_{\substack{(z,\theta):\\ 0<z<\frac{1}{2}\min\{\sin\theta,\cos\theta\}\\ 0<\theta<\frac{1}{2}\pi}} dz\,d\theta \\ &= \frac{4}{\pi}\int_0^{\pi/2} \tfrac{1}{2}\min\{\sin\theta,\cos\theta\}\,d\theta = \frac{4}{\pi}\int_0^{\pi/4}\sin\theta\,d\theta = \frac{4}{\pi}\left(1 - \frac{1}{\sqrt{2}}\right),\end{aligned}$$

and hence

$$\operatorname{var}(\tfrac{1}{2}Z) = \frac{1}{\pi} - \frac{4}{\pi^2} + \frac{1}{\pi}(2-\sqrt{2}) = \frac{3-\sqrt{2}}{\pi} - \frac{4}{\pi^2}.$$

(iii) For Buffon's needle, the variance of the number of intersections is $(2/\pi) - (2/\pi)^2$ which exceeds $\operatorname{var}(\frac{1}{2}Z)$. You should therefore use Buffon's cross.

4. (i) $F_U(u) = 1-(1-u)(1-u)$ if $0 < u < 1$, and so $\mathbb{E}(U) = \int_0^1 2u(1-u)\,du = \frac{1}{3}$. (Alternatively, place three points independently at random on the circumference of a circle of circumference 1. Measure the distances X and Y from the first point to the other two, along the circumference clockwise. Clearly X and Y are independent and uniform on $[0,1]$. Hence by circular symmetry, $\mathbb{E}(U) = \mathbb{E}(V-U) = \mathbb{E}(1-V) = \frac{1}{3}$.)
(ii) Clearly $UV = XY$, so that $\mathbb{E}(UV) = \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{4}$. Hence

$$\operatorname{cov}(U,V) = \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V) = \tfrac{1}{4} - \tfrac{1}{3}(1-\tfrac{1}{3}) = \tfrac{1}{36},$$

since $\mathbb{E}(V) = 1 - \mathbb{E}(U)$ by 'symmetry'.

5. (i) If X and Y are independent then, by (4.5.6) and independence,

$$\begin{aligned}\mathbb{E}(g(X)h(Y)) &= \iint g(x)h(y)f_{X,Y}(x,y)\,dx\,dy \\ &= \int g(x)f_X(x)\,dx\int h(y)f_Y(y)\,dy = \mathbb{E}(g(X))\mathbb{E}(h(Y)).\end{aligned}$$

(ii) By independence

$$\mathbb{E}(e^{\frac{1}{2}(X+Y)}) = \mathbb{E}(e^{\frac{1}{2}X})^2 = \left\{\int_0^\infty e^{\frac{1}{2}x}e^{-x}\,dx\right\}^2 = 4.$$

6. If O is the centre of the circle, take the radius OA as origin of coordinates. That is, A $= (1,0)$, B $= (1,\Theta)$, C $= (1,\Phi)$, in polar coordinates, where we choose the labels in such a way that $0 \le \Theta \le \Phi$. The pair Θ, Φ has joint density function $f(\theta,\phi) = (2\pi^2)^{-1}$ for $0 < \theta < \phi < 2\pi$.

The three angles of ABC are $\frac{1}{2}\Theta$, $\frac{1}{2}(\Phi-\Theta)$, $\pi - \frac{1}{2}\Phi$. You should plot in the θ/ϕ plane the set of pairs (θ,ϕ) such that $0 < \theta < \phi < 2\pi$ and such that at least one of the three angles exceeds $x\pi$.

Then integrate f over this region to obtain the result. The shape of the region depends on whether or not $x < \frac{1}{2}$. The density function g of the largest angle is given by differentiation:

$$g(x) = \begin{cases} 6(3x-1) & \text{if } \frac{1}{3} \le x \le \frac{1}{2}, \\ 6(1-x) & \text{if } \frac{1}{2} \le x \le 1. \end{cases}$$

The expectation is found to be $\frac{11}{18}\pi$.

7. We have that $\mathbb{E}(\overline{X}) = \mu$, and therefore $\mathbb{E}(X_r - \overline{X}) = 0$. Furthermore,

$$\begin{aligned} \mathbb{E}\{\overline{X}(X_r - \overline{X})\} &= \frac{1}{n}\mathbb{E}\Big(\sum_s X_r X_s\Big) - \mathbb{E}(\overline{X}^2) = \frac{1}{n}\{\sigma^2 + n\mu^2\} - \big(\mathrm{var}(\overline{X}) + \mathbb{E}(\overline{X})^2\big) \\ &= \frac{1}{n}\{\sigma^2 + n\mu^2\} - \left(\frac{\sigma^2}{n} + \mu^2\right) = 0. \end{aligned}$$

8. The condition is that $\mathbb{E}(Y)\,\mathrm{var}(X) + \mathbb{E}(X)\,\mathrm{var}(Y) = 0$.

9. If X and Y are positive, then S positive entails T positive, which displays the dependence. Finally, $S^2 = X$ and $T^2 = Y$.

4.6 Solutions. Conditional distributions and conditional expectation

1. The point is picked according to the uniform distribution on the surface of the unit sphere, which is to say that, for any suitable subset C of the surface, the probability the point lies in C is the surface integral $\int_C (4\pi)^{-1}\,dS$. Changing to polar coordinates, $x = \cos\theta\cos\phi$, $y = \sin\theta\cos\phi$, $z = \sin\phi$, subject to $x^2 + y^2 + z^2 = 1$, this surface integral becomes $(4\pi)^{-1}\int_C |\cos\phi|\,d\theta\,d\phi$, whence the joint density function of Θ and Φ is

$$f(\theta, \phi) = \frac{1}{4\pi}|\cos\phi|, \qquad |\phi| \le \tfrac{1}{2}\pi,\ 0 \le \theta < 2\pi.$$

The marginals are then $f_\Theta(\theta) = (2\pi)^{-1}$, $f_\Phi(\phi) = \frac{1}{2}|\cos\phi|$, and the conditional density functions are

$$f_{\Theta|\Phi}(\theta \mid \phi) = \frac{1}{2\pi}, \qquad f_{\Phi|\Theta}(\phi \mid \theta) = \tfrac{1}{2}|\cos\phi|,$$

for appropriate θ and ϕ. Thus Θ and Φ are independent. The fact that the conditional density functions are different from each other is sometimes referred to as 'Borel's paradox'.

2. We have that

$$\psi(x) = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$$

and therefore

$$\begin{aligned} \mathbb{E}\big(\psi(X)g(X)\big) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)} g(x) f_X(x)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{y\,g(x)\} f_{X,Y}(x,y)\,dx\,dy = \mathbb{E}(Yg(X)). \end{aligned}$$

3. Take Y to be a random variable with mean ∞, say $f_Y(y) = y^{-2}$ for $1 \le y < \infty$, and let $X = Y$. Then $\mathbb{E}(Y \mid X) = X$ which is (almost surely) finite.

4. (a) We have that

$$f_X(x) = \int_x^\infty \lambda^2 e^{-\lambda y}\,dy = \lambda e^{-\lambda x}, \qquad 0 \le x < \infty,$$

so that $f_{Y|X}(y \mid x) = \lambda e^{\lambda(x-y)}$, for $0 \le x \le y < \infty$.
(b) Similarly,

$$f_X(x) = \int_0^\infty x e^{-x(y+1)}\,dy = e^{-x}, \qquad 0 \le x < \infty,$$

so that $f_{Y|X}(y \mid x) = xe^{-xy}$, for $0 \le y < \infty$.

5. We have that

$$\begin{aligned}\mathbb{P}(Y=k) &= \int_0^1 \mathbb{P}(Y=k \mid X=x) f_X(x)\,dx = \int_0^1 \binom{n}{k} x^k(1-x)^{n-k}\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\,dx \\ &= \binom{n}{k}\frac{B(a+k, n-k+b)}{B(a,b)}.\end{aligned}$$

In the special case $a = b = 1$, this yields

$$\mathbb{P}(Y=k) = \binom{n}{k}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{1}{n+1}, \qquad 0 \le k \le n,$$

whence Y is uniformly distributed.

We have in general that

$$\mathbb{E}(Y) = \int_0^1 \mathbb{E}(Y \mid X=x) f_X(x)\,dx = \frac{na}{a+b},$$

and, by a similar computation of $\mathbb{E}(Y^2)$,

$$\operatorname{var}(Y) = \frac{nab(a+b+n)}{(a+b)^2(a+b+1)}.$$

6. By conditioning on X_1,

$$G_n(x) = \mathbb{P}(N > n) = \int_0^x G_{n-1}(x-u)\,du = \int_0^x G_{n-1}(v)\,dv.$$

Now $G_0(v) = 1$ for all $v \in (0,1]$, and the result follows by induction. Now,

$$\mathbb{E}N = \sum_{n=0}^\infty \mathbb{P}(N > n) = e^x.$$

More generally,

$$G_N(s) = \sum_{n=1}^\infty s^n \mathbb{P}(N=n) = \sum_{n=1}^\infty s^n\left(\frac{x^{n-1}}{(n-1)!} - \frac{x^n}{n!}\right) = (s-1)e^{sx} + 1,$$

whence $\operatorname{var}(N) = G_N''(1) + G_N'(1) - G_N'(1)^2 = 2xe^x + e^x - e^{2x}$.

7. We may assume without loss of generality that $\mathbb{E}X = \mathbb{E}Y = 0$. By the Cauchy–Schwarz inequality,

$$\mathbb{E}(XY)^2 = \mathbb{E}\big(X\mathbb{E}(Y \mid X)\big)^2 \le \mathbb{E}(X^2)\mathbb{E}\big(\mathbb{E}(Y \mid X)^2\big).$$

Hence,

$$\mathbb{E}\big(\mathrm{var}(Y \mid X)\big) = \mathbb{E}(Y^2) - \mathbb{E}\big(\mathbb{E}(Y \mid X)^2\big) \le \mathbb{E}Y^2 - \frac{\mathbb{E}(XY)^2}{\mathbb{E}(X^2)} = (1-\rho^2)\,\mathrm{var}(Y).$$

8. One way is to evaluate

$$\int_0^\infty \int_x^\infty \int_y^\infty \lambda\mu\nu e^{-\lambda x - \mu y - \nu z}\,dx\,dy\,dz.$$

Another way is to observe that $\min\{Y, Z\}$ is exponentially distributed with parameter $\mu + \nu$, whence $\mathbb{P}(X < \min\{Y, Z\}) = \lambda/(\lambda + \mu + \nu)$. Similarly, $\mathbb{P}(Y < Z) = \mu/(\mu + \nu)$, and the product of these two terms is the required answer.

9. By integration, for $x, y > 0$,

$$f_Y(y) = \int_0^y f(x, y)\,dx = \tfrac{1}{6}cy^3e^{-y}, \quad f_X(x) = \int_x^\infty f(x, y)\,dy = cxe^{-x},$$

whence $c = 1$. It is simple to check the values of $f_{X|Y}(x \mid y) = f(x, y)/f_Y(y)$ and $f_{Y|X}(y \mid x)$, and then deduce by integration that $\mathbb{E}(X \mid Y = y) = \frac{1}{2}Y$ and $\mathbb{E}(Y \mid X = x) = x + 2$.

10. We have that $N > n$ if and only if X_0 is largest of $\{X_0, X_1, \dots, X_n\}$, an event having probability $1/(n+1)$. Therefore, $\mathbb{P}(N = n) = 1/\{n(n+1)\}$ for $n \ge 1$. Next, on the event $\{N = n\}$, X_n is the largest, whence

$$\mathbb{P}(X_N \le x) = \sum_{n=1}^\infty \frac{F(x)^{n+1}}{n(n+1)} = \sum_{n=1}^\infty \frac{F(x)^{n+1}}{n} - \sum_{n=1}^\infty \frac{F(x)^{n+1}}{n+1} + F(x),$$

as required. Finally,

$$\mathbb{P}(M = m) = \mathbb{P}(X_0 \ge X_1 \ge \cdots \ge X_{m-1}) - \mathbb{P}(X_0 \ge X_1 \ge \cdots \ge X_m) = \frac{1}{m!} - \frac{1}{(m+1)!}.$$

4.7 Solutions. Functions of random variables

1. We observe that, if $0 \le u \le 1$,

$$\begin{aligned}\mathbb{P}(XY \le u) &= \mathbb{P}(XY \le u, Y \le u) + \mathbb{P}(XY \le u, Y > u) = \mathbb{P}(Y \le u) + \mathbb{P}(X \le u/Y,\ Y > u)\\ &= u + \int_u^1 \frac{u}{y}\,dy = u(1 - \log u).\end{aligned}$$

By the independence of XY and Z,

$$\mathbb{P}(XY \le u, Z^2 \le v) = \mathbb{P}(XY \le u)\mathbb{P}(Z \le \sqrt{v}) = u\sqrt{v}(1 - \log u), \quad 0 < u, v < 1.$$

Differentiate to obtain the joint density function

$$g(u, v) = \frac{\log(1/u)}{2\sqrt{v}}, \qquad 0 \le u, v \le 1.$$

Hence

$$\mathbb{P}(XY \le Z^2) = \iint_{0 \le u \le v \le 1} \frac{\log(1/u)}{2\sqrt{v}}\, du\, dv = \tfrac{5}{9}.$$

Arguing more directly,

$$\mathbb{P}(XY \le Z^2) = \iiint_{\substack{0 \le x,y,z \le 1 \\ xy \le z^2}} dx\, dy\, dz = \tfrac{5}{9}.$$

2. The transformation $x = uv$, $y = u - uv$ has Jacobian

$$J = \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} = -u.$$

Hence $|J| = |u|$, and therefore $f_{U,V}(u, v) = ue^{-u}$, for $0 \le u < \infty$, $0 \le v \le 1$. Hence U and V are independent, and $f_V(v) = 1$ on $[0, 1]$ as required.

3. Arguing directly,

$$\mathbb{P}(\sin X \le y) = \mathbb{P}(X \le \sin^{-1} y) = \frac{2}{\pi} \sin^{-1} y, \qquad 0 \le y \le 1,$$

so that $f_Y(y) = 2/\big(\pi\sqrt{1 - y^2}\big)$, for $0 \le y \le 1$. Alternatively, make a one-dimensional change of variables.

4. (a) $\mathbb{P}(\sin^{-1} X \le y) = \mathbb{P}(X \le \sin y) = \sin y$, for $0 \le y \le \frac{1}{2}\pi$. Hence $f_Y(y) = \cos y$, for $0 \le y \le \frac{1}{2}\pi$.

(b) Similarly, $\mathbb{P}(\sin^{-1} X \le y) = \frac{1}{2}(1 + \sin y)$, for $-\frac{1}{2}\pi \le y \le \frac{1}{2}\pi$, so that $f_Y(y) = \frac{1}{2}\cos y$, for $-\frac{1}{2}\pi \le y \le \frac{1}{2}\pi$.

5. Consider the mapping $w = x$, $z = (y - \rho x)/\sqrt{1 - \rho^2}$ with inverse $x = w$, $y = \rho w + z\sqrt{1 - \rho^2}$ and Jacobian

$$J = \begin{vmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{vmatrix} = \sqrt{1 - \rho^2}.$$

The mapping is one–one, and therefore W $(= X)$ and Z satisfy

$$f_{W,Z}(w, z) = \frac{\sqrt{1 - \rho^2}}{2\pi\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}(1 - \rho^2)(w^2 + z^2)\right\} = \frac{1}{2\pi} e^{-\frac{1}{2}(w^2 + z^2)},$$

implying that W and Z are independent $N(0, 1)$ variables. Now

$$\{X > 0, Y > 0\} = \left\{W > 0, Z > -W\rho \Big/ \sqrt{1 - \rho^2}\right\},$$

and therefore, moving to polar coordinates,

$$\mathbb{P}(X > 0, Y > 0) = \int_{\theta=\alpha}^{\frac{1}{2}\pi} \int_{r=0}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r\, dr\, d\theta = \int_{\alpha}^{\frac{1}{2}\pi} \frac{1}{2\pi}\, d\theta$$

where $\alpha = -\tan^{-1}(\rho/\sqrt{1-\rho^2}) = -\sin^{-1}\rho$.

6. We confine ourselves to the more interesting case when $\rho \neq 1$. Writing $X = U$, $Y = \rho U + \sqrt{1-\rho^2}V$, we have that U and V are independent $N(0,1)$ variables. It is easy to check that $Y > X$ if and only if $(1-\rho)U < \sqrt{1-\rho^2}V$. Turning to polar coordinates,

$$\mathbb{E}(\max\{X,Y\}) = \int_0^\infty \frac{re^{-\frac{1}{2}r^2}}{2\pi}\left[\int_\psi^{\psi+\pi}\left\{\rho r\cos\theta + r\sqrt{1-\rho^2}\sin\theta\right\}d\theta + \int_{\psi-\pi}^{\psi} r\cos\theta\, d\theta\right]dr$$

where $\tan\psi = \sqrt{(1-\rho)/(1+\rho)}$. Some algebra yields the result. For the second part,

$$\mathbb{E}(\max\{X,Y\}^2) = \mathbb{E}(X^2 I_{\{X>Y\}}) + \mathbb{E}(Y^2 I_{\{Y>X\}}) = \mathbb{E}(X^2 I_{\{X<Y\}}) + \mathbb{E}(Y^2 I_{\{Y<X\}}),$$

by the symmetry of the marginals of X and Y. Adding, we obtain $2\mathbb{E}(\max\{X,Y\}^2) = \mathbb{E}(X^2) + \mathbb{E}(Y^2) = 2$.

7. We have that

$$\mathbb{P}(X < Y,\ Z > z) = \mathbb{P}(z < X < Y) = \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)z} = \mathbb{P}(X < Y)\mathbb{P}(Z > z).$$

(a) $\mathbb{P}(X = Z) = \mathbb{P}(X < Y) = \dfrac{\lambda}{\lambda+\mu}$.

(b) By conditioning on Y,

$$\mathbb{P}\big((X-Y)^+ = 0\big) = \mathbb{P}(X \le Y) = \frac{\lambda}{\lambda+\mu}, \quad \mathbb{P}\big((X-Y)^+ > w\big) = \frac{\mu}{\lambda+\mu}e^{-\lambda w} \quad \text{for } w > 0.$$

By conditioning on X,

$$\begin{aligned}\mathbb{P}(V > v) = \mathbb{P}(|X-Y| > v) &= \int_0^\infty \mathbb{P}(Y > v + x)f_X(x)\,dx + \int_v^\infty \mathbb{P}(Y < x - v)f_X(x)\,dx \\ &= \frac{\mu e^{-\lambda v} + \lambda e^{-\mu v}}{\lambda+\mu}, \qquad v > 0.\end{aligned}$$

(c) By conditioning on X, the required probability is found to be

$$\int_0^t \lambda e^{-\lambda x}\int_{t-x}^\infty \mu e^{-\mu y}\,dy\,dx = \frac{\lambda}{\mu-\lambda}\{e^{-\lambda t} - e^{-\mu t}\}.$$

8. Either make a change of variables and find the Jacobian, or argue directly. With the convention that $\sqrt{r^2-u^2} = 0$ when $r^2 - u^2 < 0$, we have that

$$\begin{aligned}F(r,x) &= \mathbb{P}(R \le r,\ X \le x) = \frac{2}{\pi}\int_{-r}^{x}\sqrt{r^2-u^2}\,du, \\ f(r,x) &= \frac{\partial^2 F}{\partial r\partial x} = \frac{2r}{\pi\sqrt{r^2-x^2}}, \qquad |x| < r < 1.\end{aligned}$$

9. As in the previous exercise,

$$\mathbb{P}(R \le r,\ Z \le z) = \frac{3}{4\pi}\int_{-r}^{z}\pi(r^2-w^2)\,dw.$$

Hence $f(r,z) = \frac{3}{2}r$ for $|z| < r < 1$. This question may be solved in spherical polars also.

10. The transformation $s = x + y$, $r = x/(x+y)$, has inverse $x = rs$, $y = (1-r)s$ and Jacobian $J = s$. Therefore,

$$f_R(r) = \int_0^\infty f_{R,S}(r,s)\,ds = \int_0^\infty f_{X,Y}\big(rs,(1-r)s\big)s\,ds$$
$$= \int_0^\infty \lambda e^{-\lambda rs}\mu e^{-\mu(1-r)s}s\,ds = \frac{\lambda\mu}{\{\lambda r + \mu(1-r)\}^2}, \qquad 0 \le r \le 1.$$

11. We have that

$$\mathbb{P}(Y \le y) = \mathbb{P}\left(X^2 \ge \frac{a}{y} - 1\right) = 2\mathbb{P}\left(X \le -\sqrt{\frac{a}{y} - 1}\right),$$

whence

$$f_Y(y) = 2f_X\big(-\sqrt{(a/y)-1}\big) = \frac{1}{\pi\sqrt{y(a-y)}}, \qquad 0 \le y \le a.$$

12. Using the result of Example (4.6.7), and integrating by parts, we obtain

$$\mathbb{P}(X > a, Y > b) = \int_a^\infty \phi(x)\left\{1 - \Phi\left(\frac{b-\rho x}{\sqrt{1-\rho^2}}\right)\right\}dx$$
$$= [1-\Phi(a)][1-\Phi(c)] + \int_a^\infty [1-\Phi(x)]\phi\left(\frac{b-\rho x}{\sqrt{1-\rho^2}}\right)\frac{\rho}{\sqrt{1-\rho^2}}\,dx.$$

Since $[1-\Phi(x)]/\phi(x)$ is decreasing, the last term on the right is no greater than

$$\frac{1-\Phi(a)}{\phi(a)}\int_a^\infty \phi(x)\phi\left(\frac{b-\rho x}{\sqrt{1-\rho^2}}\right)\frac{\rho}{\sqrt{1-\rho^2}}\,dx,$$

which yields the upper bound after an integration.

13. The random variable Y is symmetric and, for $a > 0$,

$$\mathbb{P}(Y > a) = \mathbb{P}(0 < X < a^{-1}) = \int_0^{a^{-1}} \frac{du}{\pi(1+u^2)} = \int_\infty^a \frac{-v^{-2}\,dv}{\pi(1+v^{-2})},$$

by the transformation $v = 1/u$. For another example, consider the density function

$$f(x) = \begin{cases} \frac{1}{2}x^{-2} & \text{if } x > 1,\\ \frac{1}{2} & \text{if } 0 \le x \le 1.\end{cases}$$

14. The transformation $w = x + y$, $z = x/(x+y)$ has inverse $x = wz$, $y = (1-z)w$, and Jacobian $J = w$, whence

$$f(w,z) = w \cdot \frac{\lambda(\lambda wz)^{\alpha-1}e^{-\lambda wz}}{\Gamma(\alpha)} \cdot \frac{\lambda(\lambda(1-z)w)^{\beta-1}e^{-\lambda(1-z)w}}{\Gamma(\beta)}$$
$$= \frac{\lambda(\lambda w)^{\alpha+\beta-1}e^{-\lambda w}}{\Gamma(\alpha+\beta)} \cdot \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha,\beta)}, \qquad w > 0,\ 0 < z < 1.$$

Hence W and Z are independent, and Z is beta distributed with parameters α and β.

4.8 Solutions. Sums of random variables

1. By the convolution formula (4.8.2), $Z = X + Y$ has density function

$$f_Z(z) = \int_0^z \lambda\mu e^{-\lambda x} e^{-\mu(z-x)}\,dx = \frac{\lambda\mu}{\mu-\lambda}\left(e^{-\lambda z} - e^{-\mu z}\right), \qquad z \ge 0,$$

if $\lambda \ne \mu$. What happens if $\lambda = \mu$? (Z has a gamma distribution in this case.)

2. Using the convolution formula (4.8.2), $W = \alpha X + \beta Y$ has density function

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{\pi\alpha\left(1+(x/\alpha)^2\right)} \cdot \frac{1}{\pi\beta\left(1+\{(w-x)/\beta\}^2\right)}\,dx,$$

which equals the limit of a complex integral:

$$\lim_{R\to\infty} \int_D \frac{\alpha\beta}{\pi^2} \cdot \frac{1}{z^2+\alpha^2} \cdot \frac{1}{(z-w)^2+\beta^2}\,dz$$

where D is the semicircle in the upper complex plane with diameter $[-R, R]$ on the real axis. Evaluating the residues at $z = i\alpha$ and $z = w + i\beta$ yields

$$\begin{aligned} f_W(w) &= \frac{\alpha\beta 2\pi i}{\pi^2}\left\{\frac{1}{2i\alpha} \cdot \frac{1}{(i\alpha-w)^2+\beta^2} + \frac{1}{2i\beta} \cdot \frac{1}{(w+i\beta)^2+\alpha^2}\right\} \\ &= \frac{1}{\pi(\alpha+\beta)} \cdot \frac{1}{1+\{w/(\alpha+\beta)\}^2}, \qquad -\infty < w < \infty \end{aligned}$$

after some manipulation. Hence W has a Cauchy distribution also.

3. Using the convolution formula (4.8.2),

$$f_Z(z) = \int_0^z \tfrac{1}{2} z e^{-z}\,dy = \tfrac{1}{2} z^2 e^{-z}, \qquad z \ge 0.$$

4. Let f_n be the density function of S_n. By convolution,

$$f_2(x) = \lambda_1 e^{-\lambda_1 x} \cdot \frac{\lambda_2}{\lambda_2-\lambda_1} + \lambda_2 e^{-\lambda_2 x} \cdot \frac{\lambda_1}{\lambda_1-\lambda_2} = \sum_{r=1}^{2} \lambda_r e^{-\lambda_r x} \prod_{\substack{s=1 \\ s\ne r}}^{n} \frac{\lambda_s}{\lambda_s-\lambda_r}.$$

This leads to the guess that

$$(*) \qquad f_n(x) = \sum_{r=1}^{n} \lambda_r e^{-\lambda_r x} \prod_{\substack{s=1 \\ s\ne r}}^{n} \frac{\lambda_s}{\lambda_s-\lambda_r}, \qquad n \ge 2,$$

which may be proved by induction as follows. Assume that $(*)$ holds for $n \le N$. Then

$$\begin{aligned} f_{N+1}(x) &= \int_0^x \sum_{r=1}^{N} \lambda_r e^{-\lambda_r(x-y)} \lambda_{N+1} e^{-\lambda_{N+1} y} \prod_{\substack{s=1 \\ s\ne r}}^{N} \frac{\lambda_s}{\lambda_s-\lambda_r}\,dy \\ &= \sum_{r=1}^{N} \lambda_r e^{-\lambda_r x} \prod_{\substack{s=1 \\ s\ne r}}^{N+1} \frac{\lambda_s}{\lambda_s-\lambda_r} + A e^{-\lambda_{N+1} x}, \end{aligned}$$

for some constant A. We integrate over x to find that

$$1 = \sum_{r=1}^{N} \prod_{\substack{s=1 \\ s \neq r}}^{N+1} \frac{\lambda_s}{\lambda_s - \lambda_r} + \frac{A}{\lambda_{N+1}},$$

and $(*)$ follows with $n = N + 1$ on solving for A.

5. The density function of $X + Y$ is, by convolution,

$$f_2(x) = \begin{cases} x & \text{if } 0 \le x \le 1, \\ 2 - x & \text{if } 1 \le x \le 2. \end{cases}$$

Therefore, for $1 \le x \le 2$,

$$f_3(x) = \int_0^1 f_2(x - y)\,dy = \int_{x-1}^{1} (x - y)\,dy + \int_0^{x-1} (2 - x + y)\,dy = \tfrac{3}{4} - (x - \tfrac{3}{2})^2.$$

Likewise,

$$f_3(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } 0 \le x \le 1, \\ \frac{1}{2}(3 - x)^2 & \text{if } 2 \le x \le 3. \end{cases}$$

A simple induction yields the last part.

6. The covariance satisfies $\text{cov}(U, V) = \mathbb{E}(X^2 - Y^2) = 0$, as required. If X and Y are symmetric random variables taking values ± 1, then

$$\mathbb{P}(U = 2,\ V = 2) = 0 \quad \text{but} \quad \mathbb{P}(U = 2)\mathbb{P}(V = 2) > 0.$$

If X and Y are independent $N(0, 1)$ variables, $f_{U,V}(u, v) = (4\pi)^{-1} e^{-\frac{1}{4}(u^2+v^2)}$, which factorizes as a function of u multiplied by a function of v.

7. From the representation $X = \sigma\rho U + \sigma\sqrt{1 - \rho^2}\,V$, $Y = \tau U$, where U and V are independent $N(0, 1)$, we learn that

$$\mathbb{E}(X \mid Y = y) = \mathbb{E}(\sigma\rho U \mid U = y/\tau) = \frac{\sigma\rho y}{\tau}.$$

Similarly,

$$\mathbb{E}(X^2 \mid Y = y) = \mathbb{E}\big((\sigma\rho U)^2 + \sigma^2(1 - \rho^2)V^2 \,\big|\, U = y/\tau\big) = \Big(\frac{\sigma\rho y}{\tau}\Big)^2 + \sigma^2(1 - \rho^2)$$

whence $\text{var}(X \mid Y) = \sigma^2(1 - \rho^2)$. For parts (c) and (d), simply calculate that $\text{cov}(X, X + Y) = \sigma^2 + \rho\sigma\tau$, $\text{var}(X + Y) = \sigma^2 + 2\rho\sigma\tau + \tau^2$, and

$$1 - \rho(X, X + Y)^2 = \frac{\tau^2(1 - \rho^2)}{\sigma^2 + 2\rho\sigma\tau + \tau^2}.$$

8. First recall that $\mathbb{P}(|X| \le y) = 2\Phi(y) - 1$. We shall use the fact that $U = (X + Y)/\sqrt{2}$, $V = (X - Y)/\sqrt{2}$ are independent and $N(0, 1)$ distributed. Let Δ be the triangle of $\mathbb{R}^2$ with vertices $(0, 0)$, $(0, Z)$, $(Z, 0)$. Then

$$\begin{aligned} \mathbb{P}(Z \le z \mid X > 0,\ Y > 0) &= 4\mathbb{P}\big((X, Y) \in \Delta\big) = \mathbb{P}\big(|U| \le z/\sqrt{2}, |V| \le z/\sqrt{2}\big) \quad \text{by symmetry} \\ &= 2\big\{2\Phi(z/\sqrt{2}) - 1\big\}^2, \end{aligned}$$

whence the conditional density function is

$$f(z) = 2\sqrt{2}\{2\Phi(z/\sqrt{2}) - 1\}\phi(z/\sqrt{2}).$$

Finally,

$$\begin{aligned}\mathbb{E}(Z \mid X > 0,\ Y > 0) &= 2\mathbb{E}(X \mid X > 0,\ Y > 0)\\ &= 2\mathbb{E}(X \mid X > 0) = 4\mathbb{E}(XI_{\{X>0\}}) = 4\int_0^\infty \frac{x}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx.\end{aligned}$$

4.9 Solutions. Multivariate normal distribution

1. Since $\mathbf{V}$ is symmetric, there exists a non-singular matrix $\mathbf{M}$ such that $\mathbf{M}' = \mathbf{M}^{-1}$ and $\mathbf{V} = \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^{-1}$, where $\boldsymbol{\Lambda}$ is the diagonal matrix with diagonal entries the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of $\mathbf{V}$. Let $\boldsymbol{\Lambda}^{\frac{1}{2}}$ be the diagonal matrix with diagonal entries $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$; $\boldsymbol{\Lambda}^{\frac{1}{2}}$ is well defined since $\mathbf{V}$ is non-negative definite. Writing $\mathbf{W} = \mathbf{M}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}'$, we have that $\mathbf{W} = \mathbf{W}'$ and also

$$\mathbf{W}^2 = (\mathbf{M}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{-1})(\mathbf{M}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{M}^{-1}) = \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^{-1} = \mathbf{V}$$

as required. Clearly $\mathbf{W}$ is non-singular if and only if $\boldsymbol{\Lambda}^{\frac{1}{2}}$ is non-singular. This happens if and only if $\lambda_i > 0$ for all i, which is to say that $\mathbf{V}$ is positive definite.

2. By Theorem (4.9.6), $\mathbf{Y}$ has the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix

$$\mathbf{W}^{-1}\mathbf{V}\mathbf{W}^{-1} = \mathbf{W}^{-1}(\mathbf{W}^2)\mathbf{W}^{-1} = \mathbf{I}.$$

3. Clearly $Y = (\mathbf{X} - \boldsymbol{\mu})\mathbf{a}' + \boldsymbol{\mu}\mathbf{a}'$ where $\mathbf{a} = (a_1, a_2, \dots, a_n)$. Using Theorem (4.9.6) as in the previous solution, $(\mathbf{X} - \boldsymbol{\mu})\mathbf{a}'$ is univariate normal with mean $\mathbf{0}$ and variance $\mathbf{a}\mathbf{V}\mathbf{a}'$. Hence Y is normal with mean $\boldsymbol{\mu}\mathbf{a}'$ and variance $\mathbf{a}\mathbf{V}\mathbf{a}'$.

4. Make the transformation $u = x + y$, $v = x - y$, with inverse $x = \frac{1}{2}(u + v)$, $y = \frac{1}{2}(u - v)$, so that $|J| = \frac{1}{2}$. The exponent of the bivariate normal density function is

$$-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2) = -\frac{1}{4(1-\rho^2)}\{u^2(1-\rho) + v^2(1+\rho)\},$$

and therefore $U = X + Y$, $V = X - Y$ have joint density

$$f(u, v) = \frac{1}{4\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{u^2}{4(1+\rho)} - \frac{v^2}{4(1-\rho)}\right\},$$

whence U and V are independent with respective distributions $N(0, 2(1+\rho))$ and $N(0, 2(1-\rho))$.

5. That Y is $N(0, 1)$ follows by showing that $\mathbb{P}(Y \le y) = \mathbb{P}(X \le y)$ for each of the cases $y \le -a$, $|y| < a$, $y \ge a$.

Secondly,

$$\rho(a) = \mathbb{E}(XY) = \int_{-a}^{a} x^2\phi(x)\,dx - \int_{-\infty}^{-a} x^2\phi(x)\,dx - \int_a^\infty x^2\phi(x)\,dx = 1 - 4\int_a^\infty x^2\phi(x)\,dx.$$

The answer to the final part is *no*; X and Y are $N(0, 1)$ variables, but the pair (X, Y) is *not* bivariate normal. One way of seeing this is as follows. There exists a root a of the equation $\rho(a) = 0$. With this value of a, *if* the pair X, Y is bivariate normal, then X and Y are independent. This conclusion is manifestly false: in particular, we have that $\mathbb{P}(X > a,\ Y > a) \neq \mathbb{P}(X > a)\mathbb{P}(Y > a)$.

6. Recall from Exercise (4.8.7) that for any pair of centred normal random variables

$$\mathbb{E}(X \mid Y) = \frac{\operatorname{cov}(X, Y)}{\operatorname{var} Y} Y, \quad \operatorname{var}(X \mid Y) = \{1 - \rho(X, Y)^2\} \operatorname{var} X.$$

The first claim follows immediately. Likewise,

$$\operatorname{var}(X_j \mid X_k) = \{1 - \rho(X_j, X_k)^2\} \operatorname{var} X_j = \left\{1 - \frac{\sum_r c_{jr}c_{kr}}{\sqrt{\sum_r c_{jr}^2 \sum_r c_{kr}^2}}\right\} \sum_r c_{jr}^2.$$

7. As in the above exercise, we calculate $a = \mathbb{E}(X_1 \mid \sum_1^n X_r)$ and $b = \operatorname{var}(X_1 \mid \sum_1^n X_r)$ using the facts that $\operatorname{var} X_1 = v_{11}$, $\operatorname{var}(\sum_1^n X_i) = \sum_{ij} v_{ij}$, and $\operatorname{cov}(X_1, \sum_1^n X_r) = \sum_r v_{1r}$.

8. Let $p = \mathbb{P}(X > 0,\ Y > 0,\ Z > 0) = \mathbb{P}(X < 0,\ Y < 0,\ Z < 0)$. Then

$$\begin{aligned}
1 - p &= \mathbb{P}(\{X > 0\} \cup \{Y > 0\} \cup \{Z > 0\}) \\
&= \mathbb{P}(X > 0) + \mathbb{P}(Y > 0) + \mathbb{P}(Z > 0) + p \\
&\quad - \mathbb{P}(X > 0,\ Y > 0) - \mathbb{P}(Y > 0,\ Z > 0) - \mathbb{P}(X > 0,\ Z > 0) \\
&= \frac{3}{2} + p - \left[\frac{3}{4} + \frac{1}{2\pi}\{\sin^{-1}\rho_1 + \sin^{-1}\rho_2 + \sin^{-1}\rho_3\}\right].
\end{aligned}$$

9. Let U, V, W be independent $N(0, 1)$ variables, and represent X, Y, Z as $X = U$, $Y = \rho_1 U + \sqrt{1 - \rho_1^2} V$,

$$Z = \rho_3 U + \frac{\rho_2 - \rho_1\rho_3}{\sqrt{1 - \rho_1^2}} V + \sqrt{\frac{1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3}{(1 - \rho_1^2)}} W.$$

We have that $U = X$, $V = (Y - \rho_1 X)/\sqrt{1 - \rho_1^2}$ and $\mathbb{E}(Z \mid X, Y)$ follows immediately, as does the conditional variance.

4.10 Solutions. Distributions arising from the normal distribution

1. *First method.* We have from (4.4.6) that the $\chi^2(m)$ density function is

$$f_m(x) = \frac{1}{\Gamma(m/2)} 2^{-m/2} x^{\frac{1}{2}m - 1} e^{-\frac{1}{2}x}, \qquad x \geq 0.$$

The density function of $Z = X_1 + X_2$ is, by the convolution formula,

$$\begin{aligned}
g(z) &= c \int_0^z x^{\frac{1}{2}m - 1} e^{-\frac{1}{2}x} (z - x)^{\frac{1}{2}n - 1} e^{-\frac{1}{2}(z - x)}\, dx \\
&= c z^{\frac{1}{2}(m+n) - 1} e^{-\frac{1}{2}z} \int_0^1 u^{\frac{1}{2}m - 1} (1 - u)^{\frac{1}{2}n - 1}\, du
\end{aligned}$$

by the substitution $u = x/z$, where c is a constant. Hence $g(z) = c' z^{\frac{1}{2}(m+n)-1} e^{-\frac{1}{2}z}$ for $z \ge 0$, for an appropriate constant c', as required.

Second method. If m and n are integral, the following argument is neat. Let $Z_1, Z_2, \dots, Z_{m+n}$ be independent $N(0,1)$ variables. Then X_1 has the same distribution as $Z_1^2 + Z_2^2 + \cdots + Z_m^2$, and X_2 the same distribution as $Z_{m+1}^2 + Z_{m+2}^2 + \cdots + Z_{m+n}^2$ (see Problem (4.14.12)). Hence $X_1 + X_2$ has the same distribution as $Z_1^2 + \cdots + Z_{m+n}^2$, i.e., the $\chi^2(m+n)$ distribution.

2. (i) The $t(r)$ distribution is symmetric with finite mean, and hence this mean is 0.

(ii) Here is one way. Let U and V be independent $\chi^2(r)$ and $\chi^2(s)$ variables (respectively). Then

$$\mathbb{E}\left(\frac{U/r}{V/s}\right) = \frac{s}{r}\mathbb{E}(U)\mathbb{E}(V^{-1})$$

by independence. Now U is $\Gamma(\frac{1}{2}, \frac{1}{2}r)$ and V is $\Gamma(\frac{1}{2}, \frac{1}{2}s)$, so that $\mathbb{E}(U) = r$ and

$$\mathbb{E}(V^{-1}) = \int_0^\infty \frac{1}{v}\frac{2^{-s/2}}{\Gamma(\frac{1}{2}s)} v^{\frac{1}{2}s-1} e^{-\frac{1}{2}v}\,dv = \frac{\Gamma(\frac{1}{2}s - 1)}{2\Gamma(\frac{1}{2}s)} \int_0^\infty \frac{2^{-\frac{1}{2}(s-2)}}{\Gamma(\frac{1}{2}s - 1)} v^{\frac{1}{2}s-2} e^{-\frac{1}{2}v}\,dv = \frac{1}{s-2}$$

if $s > 2$, since the integrand is a density function. Hence

$$\mathbb{E}\left(\frac{U/r}{V/s}\right) = \frac{s}{s-2} \qquad \text{if } s > 2.$$

(iii) If $s \le 2$ then $\mathbb{E}(V^{-1}) = \infty$.

3. Substitute $r = 1$ into the $t(r)$ density function.

4. *First method.* Find the density function of X/Y, using a change of variables. The answer is $F(2,2)$.

Second method. X and Y are independent $\chi^2(2)$ variables (just check the density functions), and hence X/Y is $F(2,2)$.

5. The vector $(\overline{X}, X_1 - \overline{X}, X_2 - \overline{X}, \dots, X_n - \overline{X})$ has, by Theorem (4.9.6), a multivariate normal distribution. We have as in Exercise (4.5.7) that $\operatorname{cov}(\overline{X}, X_r - \overline{X}) = 0$ for all r, which implies that $\overline{X}$ is independent of each X_r. Using the form of the multivariate normal density function, it follows that $\overline{X}$ is independent of the family $\{X_r - \overline{X} : 1 \le r \le n\}$, and hence of any function of these variables. Now $S^2 = (n-1)^{-1}\sum_r (X_r - \overline{X})^2$ is such a function.

6. The choice of fixed vector is immaterial, since the joint distribution of the X_j is spherically symmetric, and we therefore take this vector to be $(0, 0, \dots, 0, 1)$. We make the change of variables $U^2 = Q^2 + X_n^2$, $\tan\Psi = Q/X_n$, where $Q^2 = \sum_{r=1}^{n-1} X_r^2$ and $Q \ge 0$. Since Q has the $\chi^2(n-1)$ distribution, and is independent of X_n, the pair Q, X_n has joint density function

$$f(q,x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \cdot \frac{\frac{1}{2}(\frac{1}{2}q)^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}x}}{\Gamma(\frac{1}{2}(n-1))}, \qquad x \in \mathbb{R},\ q > 0.$$

The theory is now slightly easier than the practice. We solve for U, Ψ, find the Jacobian, and deduce the joint density function $f_{U,\Psi}(u,\psi)$ of U, Ψ. We now integrate over u, and choose the constant so that the total integral is 1.

4.11 Solutions. Sampling from a distribution

1. Uniform on the set $\{1, 2, \dots, n\}$.

2. The result holds trivially when $n = 2$, and more generally by induction on n.

3. We may assume without loss of generality that $\lambda = 1$ (since Z/λ is $\Gamma(\lambda, t)$ if Z is $\Gamma(1, t)$). Let U, V be independent random variables which are uniformly distributed on $[0, 1]$. We set $X = -t \log V$ and note that X has the exponential distribution with parameter $1/t$. It is easy to check that

$$\frac{1}{\Gamma(t)} x^{t-1} e^{-x} \le c f_X(x) \qquad \text{for } x > 0,$$

where $c = t^t e^{-t+1}/\Gamma(t)$. Also, conditional on the event A that

$$U \le \frac{X^{t-1} e^{-t}}{\Gamma(t)} t e^{-X/t},$$

X has the required gamma distribution. This observation may be used as a basis for sampling using the rejection method. We note that $A = \{\log U \le (n-1)\big(\log(X/n) - (X/n) + 1\big)\}$. We have that $\mathbb{P}(A) = 1/c$, and therefore there is a mean number c of attempts before a sample of size 1 is obtained.

4. Use your answer to Exercise (4.11.3) to sample X from $\Gamma(1, \alpha)$ and Y from $\Gamma(1, \beta)$. By Exercise (4.7.14), $Z = X/(X + Y)$ has the required distribution.

5. (a) This is the beta distribution with parameters 2, 2. Use the result of Exercise (4).

(b) The required $\Gamma(1, 2)$ variables may be more easily obtained and used by forming $X = -\log(U_1 U_2)$ and $Y - \log(U_3 U_4)$ where $\{U_i : 1 \le i \le 4\}$ are independent and uniform on $[0, 1]$.

(c) Let U_1, U_2, U_3 be as in (b) above, and let Z be the second order statistic $U_{(2)}$. That is, Z is the middle of the three values taken by the U_i; see Problem (4.14.21). The random variable Z has the required distribution.

(d) As a slight variant, take $Z = \max\{U_1, U_2\}$ conditional on the event $\{Z \le U_3\}$.

(e) Finally, let $X = \sqrt{U_1}/(\sqrt{U_1} + \sqrt{U_2})$, $Y = \sqrt{U_1} + \sqrt{U_2}$. The distribution of X, conditional on the event $\{Y \le 1\}$, is as required.

6. We use induction. The result is obvious when $n = 2$. Let $n \ge 3$ and let $\mathbf{p} = (p_1, p_2, \dots, p_n)$ be a probability vector. Since $\mathbf{p}$ sums to 1, its minimum entry $p_{(1)}$ and maximum entry $p_{(n)}$ must satisfy

$$p_{(1)} \le \frac{1}{n} < \frac{1}{n-1}, \quad p_{(1)} + p_{(n)} \ge p_{(1)} + \frac{1 - p_{(1)}}{n-1} = \frac{1 + (n-2)p_{(1)}}{n-1} \ge \frac{1}{n-1}.$$

We relabel the entries of the vector $\mathbf{p}$ such that $p_1 = p_{(1)}$ and $p_2 = p_{(n)}$, and set $\mathbf{v}_1 = \big((n-1)p_1, 1 - (n-1)p_1, 0, \dots, 0\big)$. Then

$$\mathbf{p} = \frac{1}{n-1}\mathbf{v}_1 + \frac{n-2}{n-1}\mathbf{p}_{n-1} \quad \text{where} \quad \mathbf{p}_{n-1} = \frac{n-1}{n-2}\left(0, p_1 + p_2 - \frac{1}{n-1}, p_3, \dots, p_n\right),$$

is a probability vector with at most $n - 1$ non-zero entries. The induction step is complete.

It is a consequence that sampling from a discrete distribution may be achieved by sampling from a collection of Bernoulli random variables.

7. It is an elementary exercise to show that $\mathbb{P}(R^2 \le 1) = \frac{1}{4}\pi$, and that, conditional on this event, the vector (T_1, T_2) is uniformly distributed on the unit disk. Assume henceforth that $R^2 \le 1$, and write (R, Θ) for the point (T_1, T_2) expressed in polar coordinates. We have that R and Θ are independent with joint density function $f_{R,\Theta}(r, \theta) = r/\pi$, $0 \le r \le 1$, $0 \le \theta < 2\pi$. Let (Q, Ψ) be the polar coordinates of (X, Y), and note that $\Psi = \Theta$ and $e^{-\frac{1}{2}Q^2} = R^2$. The random variables Q and Ψ are independent, and, by a change of variables, Q has density function $f_Q(q) = qe^{-\frac{1}{2}q^2}$, $q > 0$. We recognize the distribution of (Q, Ψ) as that of the polar coordinates of (X, Y) where X and Y are independent $N(0, 1)$ variables. [Alternatively, the last step may be achieved by a two-dimensional change of variables.]

8. We have that

$$\mathbb{P}(X = k) = \mathbb{P}\left(\left\lfloor \frac{\log U}{\log q} \right\rfloor = k - 1\right) = \mathbb{P}(q^k < U \le q^{k-1}) = q^{k-1}(1-q), \quad k \ge 1.$$

9. The polar coordinates (R, Θ) of (X, Y) have joint density function

$$f_{R,\Theta}(r, \theta) = \frac{2r}{\pi}, \qquad 0 \le r \le 1, \ -\tfrac{1}{2}\pi \le \theta \le \tfrac{1}{2}\pi.$$

Make a change of variables to find that $Y/X = \tan\Theta$ has the Cauchy distribution.

10. By the definition of Z,

$$\begin{aligned}\mathbb{P}(Z = m) &= h(m) \prod_{r=0}^{m-1} \big(1 - h(r)\big) \\ &= \mathbb{P}(X > 0)\mathbb{P}(X > 1 \mid X > 0) \cdots \mathbb{P}(X = m \mid X > m-1) = \mathbb{P}(X = m).\end{aligned}$$

11. Suppose g is increasing, so that $h(\mathbf{x}) = -g(\mathbf{1} - \mathbf{x})$ is increasing also. By the FKG inequality of Problem (3.11.18b), $\kappa = \operatorname{cov}(g(\mathbf{U}), -g(\mathbf{1} - \mathbf{U})) \ge 0$, yielding the result.

Estimating I by the average $(2n)^{-1}\sum_{r=1}^{2n} g(\mathbf{U}_r)$ of $2n$ random vectors $\mathbf{U}_r$ requires a sample of size $2n$ and yields an estimate having some variance $2n\sigma^2$. If we estimate I by the average $(2n)^{-1}\{\sum_{r=1}^{n} g(\mathbf{U}_r) + g(\mathbf{1} - \mathbf{U}_r)\}$, we require a sample of size only n, and we obtain an estimate with the smaller variance $2n(\sigma^2 - \kappa)$.

12. (a) By the law of the unconscious statistician,

$$\mathbb{E}\left[\frac{g(Y)f_X(Y)}{f_Y(Y)}\right] = \int \frac{g(y)f_X(y)}{f_Y(y)} f_Y(y)\,dy = I.$$

(b) This is immediate from the fact that the variance of a sum of independent variables is the sum of their variances; see Theorem (3.3.11b).

(c) This is an application of the strong law of large numbers, Theorem (7.5.1).

13. (a) If U is uniform on $[0, 1]$, then $X = \sin(\frac{1}{2}\pi U)$ has the required distribution. This is an example of the inverse transform method.

(b) If U is uniform on $[0, 1]$, then $1 - U^2$ has density function $g(x) = \{2\sqrt{1-x}\}^{-1}$, $0 \le x \le 1$. Now $g(x) \ge (\pi/4)f(x)$, which fact may be used as a basis for the rejection method.

4.12 Solutions. Coupling and Poisson approximation

1. Suppose that $\mathbb{E}(u(X)) \ge \mathbb{E}(u(Y))$ for all increasing functions u. Let $c \in \mathbb{R}$ and set $u = I_c$ where

$$I_c(x) = \begin{cases} 1 & \text{if } x > c, \\ 0 & \text{if } x \le c, \end{cases}$$

to find that $\mathbb{P}(X > c) = \mathbb{E}(I_c(X)) \ge \mathbb{E}(I_c(Y)) = \mathbb{P}(Y > c)$.

Conversely, suppose that $X \ge_{\text{st}} Y$. We may assume by Theorem (4.12.3) that X and Y are defined on the same sample space, and that $\mathbb{P}(X \ge Y) = 1$. Let u be an increasing function. Then $\mathbb{P}(u(X) \ge u(Y)) \ge \mathbb{P}(X \ge Y) = 1$, whence $\mathbb{E}(u(X) - u(Y)) \ge 0$ whenever this expectation exists.

2. Let $\alpha = \mu/\lambda$, and let $\{I_r : r \ge 1\}$ be independent Bernoulli random variables with parameter α. Then $Z = \sum_{r=1}^{X} I_r$ has the Poisson distribution with parameter $\lambda\alpha = \mu$, and $Z \le X$.

3. Use the argument in the solution to Problem (2.7.13).

4. For any $A \subseteq \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(X \ne Y) \ge \mathbb{P}(X \in A,\ Y \in A^{\text{c}}) &= \mathbb{P}(X \in A) - \mathbb{P}(X \in A,\ Y \in A)\\ &\ge \mathbb{P}(X \in A) - \mathbb{P}(Y \in A),\end{aligned}$$

and similarly with X and Y interchanged. Hence,

$$\mathbb{P}(X \ne Y) \ge \sup_{A \subseteq \mathbb{R}} \big|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\big| = \tfrac{1}{2} d_{\text{TV}}(X, Y).$$

5. For any positive x and y, we have that $(y - x)^+ + x \wedge y = y$, where $x \wedge y = \min\{x, y\}$. It follows that

$$\sum_k \{f_X(k) - f_Y(k)\}^+ = \sum_k \{f_Y(k) - f_X(k)\}^+ = 1 - \sum_k f_X(k) \wedge f_Y(k),$$

and by the definition of $d_{\text{TV}}(X, Y)$ that the common value in this display equals $\frac{1}{2} d_{\text{TV}}(X, Y) = \delta$. Let U be a Bernoulli variable with parameter $1 - \delta$, and let V, W, Z be independent integer-valued variables with

$$\begin{aligned}\mathbb{P}(V = k) &= \{f_X(k) - f_Y(k)\}^+/\delta,\\ \mathbb{P}(W = k) &= \{f_Y(k) - f_X(k)\}^+/\delta,\\ \mathbb{P}(Z = k) &= f_X(k) \wedge f_Y(k)/(1 - \delta).\end{aligned}$$

Then $X' = UZ + (1 - U)V$ and $Y' = UZ + (1 - U)W$ have the required marginals, and $\mathbb{P}(X' = Y') = \mathbb{P}(U = 1) = 1 - \delta$. See also Problem (7.11.16d).

6. Evidently $d_{\text{TV}}(X, Y) = |p - q|$, and we may assume without loss of generality that $p \ge q$. We have from Exercise (4.12.4) that $\mathbb{P}(X = Y) \le 1 - (p - q)$. Let U and Z be independent Bernoulli variables with respective parameters $1 - p + q$ and $q/(1 - p + q)$. The pair $X' = U(Z - 1) + 1$, $Y' = UZ$ has the same marginal distributions as the pair X, Y, and $\mathbb{P}(X' = Y') = \mathbb{P}(U = 1) = 1 - p + q$ as required.

To achieve the minimum, we set $X'' = 1 - X'$ and $Y'' = Y'$, so that $\mathbb{P}(X'' = Y'') = 1 - \mathbb{P}(X' = Y') = p - q$.

4.13 Solutions. Geometrical probability

1. The angular coordinates Ψ and Σ of A and B have joint density $f(\psi, \sigma) = (2\pi)^{-2}$. We make the change of variables from $(p, \theta) \mapsto (\psi, \sigma)$ by $p = \cos\{\frac{1}{2}(\sigma - \psi)\}$, $\theta = \frac{1}{2}(\pi + \sigma + \psi)$, with inverse

$$\psi = \theta - \tfrac{1}{2}\pi - \cos^{-1} p, \quad \sigma = \theta - \tfrac{1}{2}\pi + \cos^{-1} p,$$

and Jacobian $|J| = 2/\sqrt{1 - p^2}$.

2. Let A be the left shaded region and B the right shaded region in the figure. Writing λ for the random line, by Example (4.13.2),

$$\begin{aligned}\mathbb{P}(\lambda \text{ meets both } S_1 \text{ and } S_2) &= \mathbb{P}(\lambda \text{ meets both } A \text{ and } B)\\ &= \mathbb{P}(\lambda \text{ meets } A) + \mathbb{P}(\lambda \text{ meets } B) - \mathbb{P}(\lambda \text{ meets either } A \text{ or } B)\\ &\propto b(A) + b(B) - b(H) = b(X) - b(H),\end{aligned}$$

whence $\mathbb{P}(\lambda \text{ meets } S_2 \mid \lambda \text{ meets } S_1) = [b(X) - b(H)]/b(S_1)$.

The case when $S_2 \subseteq S_1$ is treated in Example (4.13.2). When $S_1 \cap S_2 \neq \varnothing$ and $S_1 \triangle S_2 \neq \varnothing$, the argument above shows the answer to be $[b(S_1) + b(S_2) - b(H)]/b(S_1)$.

3. With $|I|$ the length of the intercept I of λ_1 with S_2, we have that $\mathbb{P}(\lambda_2 \text{ meets } I) = 2|I|/b(S_1)$, by the Buffon needle calculation (4.13.2). The required probability is

$$\frac{1}{2}\int_0^{2\pi}\int_{-\infty}^{\infty} \frac{2|I|}{b(S_1)} \cdot \frac{dp\,d\theta}{b(S_1)} = \int_0^{2\pi} \frac{|S_2|}{b(S_1)^2}\,d\theta = \frac{2\pi|S_2|}{b(S_1)^2}.$$

4. If the two points are denoted $\mathrm{P} = (X_1, Y_1)$, and $\mathrm{Q} = (X_2, Y_2)$, then

$$\mathbb{E}(Z^2) = \mathbb{E}(|\mathrm{PQ}|^2) = 2\mathbb{E}\big((X_1 - X_2)^2\big) = 4\operatorname{var}(X_1) = \frac{8}{\pi a^2}\int_{-a}^{a} x^2\sqrt{a^2 - x^2}\,dx = a^2.$$

We use Crofton's method in order to calculate $\mathbb{E}(Z)$. Consider a disc D of radius x surrounded by an annulus A of width h. We set $\lambda(x) = \mathbb{E}(Z \mid \mathrm{P}, \mathrm{Q} \in D)$, and find that

$$\lambda(x+h) = \lambda(x)\left(1 - \frac{4h}{x} - \mathrm{o}(h)\right) + 2\mathbb{E}(Z \mid \mathrm{P} \in D,\ \mathrm{Q} \in A)\left(\frac{2h}{x} + \mathrm{o}(h)\right).$$

Now

$$\mathbb{E}(Z \mid \mathrm{P} \in D,\ \mathrm{Q} \in A) = \frac{2}{\pi x^2}\int_0^{\frac{1}{2}\pi}\int_0^{2x\cos\theta} r^2\,dr\,d\theta + \mathrm{o}(1) = \frac{32x}{9\pi},$$

whence

$$\frac{d\lambda}{dx} = -\frac{4\lambda}{x} + \frac{128}{9\pi},$$

which is easily integrated subject to $\lambda(0) = 0$ to give the result.

5. (i) We may assume without loss of generality that the sphere has radius 1. The length $X = |\mathrm{AO}|$ has density function $f(x) = 3x^2$ for $0 \le x \le 1$. The triangle includes an obtuse angle if B lies either in the hemisphere opposite to A, or in the sphere with centre $\frac{1}{2}X$ and radius $\frac{1}{2}X$, or in the segment cut off by the plane through A perpendicular to AO. Hence,

$$\begin{aligned}\mathbb{P}(\text{obtuse}) &= \tfrac{1}{2} + \mathbb{E}\big((\tfrac{1}{2}X)^3\big) + (\tfrac{4}{3}\pi)^{-1}\mathbb{E}\left(\int_X^1 \pi(1-y^2)\,dy\right)\\ &= \tfrac{1}{2} + \tfrac{1}{16} + (\tfrac{4}{3})^{-1}\mathbb{E}(\tfrac{2}{3} - X + \tfrac{1}{3}X^3) = \tfrac{5}{8}.\end{aligned}$$

(ii) In the case of the circle, X has density function $2x$ for $0 \le x \le 1$, and similar calculations yield

$$\mathbb{P}(\text{obtuse}) = \frac{1}{2} + \frac{1}{8} + \frac{1}{\pi}\mathbb{E}\big(\cos^{-1}X - X\sqrt{1 - X^2}\big) = \frac{3}{4}.$$

6. Choose the x-axis along AB. With $\mathrm{P} = (X, Y)$ and $\mathrm{G} = (\gamma_1, \gamma_2)$,

$$\mathbb{E}|\mathrm{ABP}| = \tfrac{1}{2}|\mathrm{AB}|\,\mathbb{E}(Y) = \tfrac{1}{2}|\mathrm{AB}|\gamma_2 = |\mathrm{ABG}|.$$

7. We use Exercise (4.13.6). First fix P, and then Q, to find that

$$\mathbb{E}|\mathrm{APQ}| = \mathbb{E}\big[\mathbb{E}(|\mathrm{APQ}| \mid \mathrm{P})\big] = \mathbb{E}|\mathrm{APG_2}| = |\mathrm{AG_1G_2}|.$$

With $b = |\mathrm{AB}|$ and h the height of the triangle ABC on the base AB, we have that $|\mathrm{G_1G_2}| = \frac{1}{3}b$ and the height of the triangle $\mathrm{AG_1G_2}$ is $\frac{2}{3}h$. Hence,

$$\mathbb{E}|\mathrm{APQ}| = \tfrac{1}{2} \cdot \tfrac{1}{3}b \cdot \tfrac{2}{3}h = \tfrac{2}{9}|\mathrm{ABC}|.$$

8. Let the scale factor for the random triangle be X, where $X \in (0, 1)$. For a triangle with scale factor x, any given vertex can lie anywhere in a certain triangle having area $(1-x)^2|\mathrm{ABC}|$. Picking one at random from all possible such triangles amounts to supposing that X has density function $f(x) = 3(1-x)^2$, $0 \le x \le 1$. Hence the mean area is

$$|\mathrm{ABC}|\,\mathbb{E}(X^2) = |\mathrm{ABC}| \int_0^1 3x^2(1-x)^2\,dx = \tfrac{1}{10}|\mathrm{ABC}|.$$

9. We have by conditioning that, for $0 \le z \le a$,

$$\begin{aligned} F(z, a + da) &= F(z, a)\left(\frac{a}{a + da}\right)^2 + \mathbb{P}(X \ge a - z) \cdot \frac{2a\,da}{(a + da)^2} + \mathrm{o}(da) \\ &= F(z, a)\left(1 - \frac{2\,da}{a}\right) + \frac{z}{a} \cdot \frac{2da}{a} + \mathrm{o}(da), \end{aligned}$$

and the equation follows by taking the limit as $da \downarrow 0$. The boundary condition may be taken to be $F(a, a) = 1$, and we deduce that

$$F(z, a) = \frac{2z}{a} - \left(\frac{z}{a}\right)^2, \qquad 0 \le z \le a.$$

Likewise, by use of conditional expectation,

$$m_r(a + da) = m_r(a)\left(1 - \frac{2\,da}{a}\right) + \mathbb{E}\big((a - X)^r\big) \cdot \frac{2\,da}{a} + \mathrm{o}(da).$$

Now, $\mathbb{E}((a-X)^r) = a^r/(r+1)$, yielding the required equation. The boundary condition is $m_r(0) = 0$, and therefore

$$m_r(a) = \frac{2a^r}{(r+1)(r+2)}.$$

10. If n great circles meet each other, not more than two at any given point, then there are $2\binom{n}{2}$ intersections. It follows that there are $4\binom{n}{2}$ segments between vertices, and Euler's formula gives the number of regions as $n(n-1) + 2$. We may think of the plane as obtained by taking the limit as $R \to \infty$ and 'stretching out' the sphere. Each segment is a side of two polygons, so the average number of sides satisfies

$$\frac{4n(n-1)}{2 + n(n-1)} \to 4 \qquad \text{as } n \to \infty.$$

11. By making an affine transformation, we may without loss of generality assume the triangle has vertices $\mathrm{A} = (0, 1)$, $\mathrm{B} = (0, 0)$, $\mathrm{C} = (1, 0)$. With $\mathrm{P} = (X, Y)$, we have that

$$\mathrm{L} = \left(\frac{X}{1-Y}, 0\right), \quad \mathrm{M} = \left(\frac{X}{X+Y}, \frac{Y}{X+Y}\right), \quad \mathrm{N} = \left(0, \frac{Y}{1-X}\right).$$

Hence,

$$\mathbb{E}|\mathrm{BLN}| = 2\int_{\mathrm{ABC}} \frac{xy}{2(1-x)(1-y)}\,dx\,dy = \int_0^1 \left(-x - \frac{x}{1-x}\log x\right)dx = \frac{\pi^2}{6} - \frac{3}{2},$$

and likewise $\mathbb{E}|\mathrm{CLM}| = \mathbb{E}|\mathrm{ANM}| = \frac{1}{6}\pi^2 - \frac{3}{2}$. It follows that $\mathbb{E}|\mathrm{LMN}| = \frac{1}{2}(10-\pi^2) = (10-\pi^2)|\mathrm{ABC}|$.

12. Let the points be P, Q, R, S. By Example (4.13.6),

$$\mathbb{P}\big(\text{one lies inside the triangle formed by the other three}\big) = 4\mathbb{P}(\mathrm{S} \in \mathrm{PQR}) = 4 \cdot \tfrac{1}{12}.$$

13. We use Crofton's method. Let $m(a)$ be the mean area, and condition on whether points do or do not fall in the annulus with internal and external radii a, $a+h$. Then

$$m(a+h) = m(a)\left(\frac{a}{a+h}\right)^6 + \left[\frac{6h}{a} + \mathrm{o}(h)\right]\widehat{m}(a),$$

where $\widehat{m}(a)$ is the mean area of a triangle having one vertex P on the boundary of the circle. Using polar coordinates with P as origin,

$$\begin{aligned}\pi^2 a^4 \widehat{m}(a) &= \frac{1}{2}\int_0^\pi\int_0^\pi\int_0^{2a\cos\theta}\int_0^{2a\cos\psi} r_1^2 r_2^2\,dr_1\,dr_2\,\sin|\theta-\psi|\,d\theta\,d\psi\\ &= \frac{32a^6}{9}\int_0^\pi\int_0^\pi \sin^3\theta\,\sin^3\psi\,\sin|\theta-\psi|\,d\theta\,d\psi = \frac{35a^6\pi}{36}.\end{aligned}$$

Letting $h \downarrow 0$ above, we obtain

$$\frac{dm}{da} = -\frac{6m}{a} + \frac{6}{a}\cdot\frac{35a^2}{36\pi},$$

whence $m(a) = (35a^2)/(48\pi)$.

14. Let a be the radius of C, and let R be the distance of A from the centre. Conditional on R, the required probability is $(a-R)^2/a^2$, whence the answer is $\mathbb{E}((a-R)^2/a^2) = \int_0^1 (1-r)^2 2r\,dr = \frac{1}{6}$.

15. Let a be the radius of C, and let R be the distance of A from the centre. As in Exercise (4.13.14), the answer is $\mathbb{E}((a-R)^3/a^3) = \int_0^1 (1-r)^3 3r^2\,dr = \frac{1}{20}$.

4.14 Solutions to problems

1. (a) We have that

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy = \iint_{\mathbb{R}^2} e^{-r^2} r\,dr\,d\theta = \pi.$$

Secondly, $f \ge 0$, and it is easily seen that $\int_{-\infty}^{\infty} f(x)\,dx = 1$ using the substitution $y = (x-\mu)/(\sigma\sqrt{2})$.

(b) The mean is $\int_{-\infty}^{\infty} x(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}\,dx$, which equals 0 since $xe^{-\frac{1}{2}x^2}$ is an odd integrable function. The variance is $\int_{-\infty}^{\infty} x^2(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}\,dx$, easily integrated by parts to obtain 1.

(c) Note that

$$\frac{d}{dy}\left\{y^{-1}e^{-\frac{1}{2}y^2}\right\} = -(1+y^{-2})e^{-\frac{1}{2}y^2},$$
$$\frac{d}{dy}\left\{(y^{-1}-y^{-3})e^{-\frac{1}{2}y^2}\right\} = -(1-3y^{-4})e^{-\frac{1}{2}y^2},$$

and also $1-3y^{-4} < 1 < 1+y^{-2}$. Multiply throughout these inequalities by $e^{-\frac{1}{2}y^2}/\sqrt{2\pi}$, and integrate over $[x,\infty)$, to obtain the required inequalities. More extensive inequalities may be found in Exercise (4.4.8).
(d) The required probability is $\alpha(x) = [1-\Phi(x+a/x)]/[1-\Phi(x)]$. By (c),

$$\alpha(x) = (1+\mathrm{o}(1))\frac{e^{-\frac{1}{2}(x+a/x)^2}}{e^{-\frac{1}{2}x^2}} \to e^{-a} \qquad \text{as } x\to\infty.$$

2. Clearly $f \geq 0$ if and only if $0 \leq \alpha < \beta \leq 1$. Also

$$C^{-1} = \int_\alpha^\beta (x-x^2)\,dx = \tfrac{1}{2}(\beta^2-\alpha^2) - \tfrac{1}{3}(\beta^3-\alpha^3).$$

3. The A_i partition the sample space, and $i-1 \leq X(\omega) < i$ if $\omega \in A_i$. Taking expectations and using the fact that $\mathbb{E}(I_i) = \mathbb{P}(A_i)$, we find that $S \leq \mathbb{E}(X) \leq 1+S$ where

$$S = \sum_{i=2}^{\infty}(i-1)\mathbb{P}(A_i) = \sum_{i=2}^{\infty}\sum_{j=1}^{i-1} 1\cdot\mathbb{P}(A_i) = \sum_{j=1}^{\infty}\sum_{i=j+1}^{\infty}\mathbb{P}(A_i) = \sum_{j=1}^{\infty}\mathbb{P}(X \geq j).$$

4. (a) (i) Let $F^{-1}(y) = \sup\{x : F(x) = y\}$, so that

$$\mathbb{P}(F(X) \leq y) = \mathbb{P}(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y, \qquad 0 \leq y \leq 1.$$

(ii) $\mathbb{P}(-\log F(X) \leq y) = \mathbb{P}(F(X) \geq e^{-y}) = 1-e^{-y}$ if $y \geq 0$.
(b) Draw a picture. With $D = \mathrm{PR}$,

$$\mathbb{P}(D \leq d) = \mathbb{P}(\tan\widehat{\mathrm{PQR}} \leq d) = \mathbb{P}(\widehat{\mathrm{PQR}} \leq \tan^{-1} d) = \frac{1}{\pi}\left(\frac{\pi}{2} + \tan^{-1} d\right).$$

Differentiate to obtain the result.

5. Clearly

$$\mathbb{P}(X > s+x \mid X > s) = \frac{\mathbb{P}(X > s+x)}{\mathbb{P}(X > s)} = \frac{e^{-\lambda(s+x)}}{e^{-\lambda s}} = e^{-\lambda x}$$

if $x, s \geq 0$, where λ is the parameter of the distribution.

Suppose that the non-negative random variable X has the lack-of-memory property. Then $G(x) = \mathbb{P}(X > x)$ is monotone and satisfies $G(0) = 1$ and $G(s+x) = G(s)G(x)$ for $s, x \geq 0$. Hence $G(s) = e^{-\lambda s}$ for some λ; certainly $\lambda > 0$ since $G(s) \leq 1$ for all s.

Let us prove the hint. Suppose that g is monotone with $g(0) = 1$ and $g(s+t) = g(s)g(t)$ for $s, t \geq 0$. For an integer m, $g(m) = g(1)g(m-1) = \cdots = g(1)^m$. For rational $x = m/n$, $g(x)^n = g(m) = g(1)^m$ so that $g(x) = g(1)^x$; all such powers are interpreted as $\exp\{x\log g(1)\}$. Finally, if x is irrational, and g is monotone non-increasing (say), then $g(u) \leq g(x) \leq g(v)$ for all rationals u and v satisfying $v \le x \le u$. Hence $g(1)^u \le g(x) \le g(1)^v$. Take the limits as $u \downarrow x$ and $v \uparrow x$ through the rationals to obtain $g(x) = e^{\mu x}$ where $\mu = \log g(1)$.

6. If X and Y are independent, we may take $g = f_X$ and $h = f_Y$. Suppose conversely that $f(x, y) = g(x)h(y)$. Then

$$f_X(x) = g(x)\int_{-\infty}^{\infty} h(y)\,dy, \quad f_Y(y) = h(y)\int_{-\infty}^{\infty} g(x)\,dx$$

and

$$1 = \int_{-\infty}^{\infty} f_Y(y)\,dy = \int_{-\infty}^{\infty} g(x)\,dx \int_{-\infty}^{\infty} h(y)\,dy.$$

Hence $f_X(x)f_Y(y) = g(x)h(y) = f(x, y)$, so that X and Y are independent.

7. They are not independent since $\mathbb{P}(Y < 1,\ X > 1) = 0$ whereas $\mathbb{P}(Y < 1) > 0$ and $\mathbb{P}(X > 1) > 0$. As for the marginals,

$$f_X(x) = \int_x^{\infty} 2e^{-x-y}\,dy = 2e^{-2x}, \qquad f_Y(y) = \int_0^y 2e^{-x-y}\,dx = 2e^{-y}(1 - e^{-y}),$$

for $x, y \ge 0$. Finally,

$$\mathbb{E}(XY) = \int_{x=0}^{\infty}\int_{y=x}^{\infty} xy2e^{-x-y}\,dx\,dy = 1$$

and $\mathbb{E}(X) = \frac{1}{2}$, $\mathbb{E}(Y) = \frac{3}{2}$, implying that $\operatorname{cov}(X, Y) = \frac{1}{4}$.

8. As in Example (4.13.1), the desired property holds if and only if the length X of the chord satisfies $X \le \sqrt{3}$. Writing R for the distance from P to the centre O, and Θ for the acute angle between the chord and the line OP, we have that $X = 2\sqrt{1 - R^2 \sin^2\Theta}$, and therefore $\mathbb{P}(X \le \sqrt{3}) = \mathbb{P}(R\sin\Theta \ge \frac{1}{2})$. The answer is therefore

$$\mathbb{P}\left(R \ge \frac{1}{2\sin\Theta}\right) = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} \mathbb{P}\left(R \ge \frac{1}{2\sin\theta}\right) d\theta,$$

which equals $\frac{2}{3} - \sqrt{3}/(2\pi)$ in case (a) and $\frac{2}{3} + \pi^{-1}\log\tan(\pi/12)$ in case (b).

9. Evidently,

$$\mathbb{E}(U) = \mathbb{P}(Y \le g(X)) = \iint_{\substack{0\le x,y\le 1\\ y\le g(x)}} dx\,dy = \int_0^1 g(x)\,dx,$$

$$\mathbb{E}(V) = \mathbb{E}(g(X)) = \int_0^1 g(x)\,dx,$$

$$\mathbb{E}(W) = \tfrac{1}{2}\int_0^1 \{g(x) + g(1-x)\}\,dx = \int_0^1 g(x)\,dx.$$

Secondly,

$$\mathbb{E}(U^2) = \mathbb{E}(U) = J, \qquad \mathbb{E}(V^2) = \int_0^1 g(x)^2\,dx \le J \quad \text{since } g \le 1,$$

$$\begin{aligned}\mathbb{E}(W^2) &= \tfrac{1}{4}\left\{2\int_0^1 g(x)^2\,dx + 2\int_0^1 g(x)g(1-x)\,dx\right\}\\ &= \mathbb{E}(V^2) - \tfrac{1}{2}\int_0^1 g(x)\{g(x) - g(1-x)\}\,dx\\ &= \mathbb{E}(V^2) - \tfrac{1}{4}\int_0^1 \{g(x) - g(1-x)\}^2\,dx \le \mathbb{E}(V^2).\end{aligned}$$

Hence $\operatorname{var}(W) \le \operatorname{var}(V) \le \operatorname{var}(U)$.

10. Clearly the claim is true for $n = 1$, since the $\Gamma(\lambda, 1)$ distribution *is* the exponential distribution. Suppose it is true for $n \le k$ where $k \ge 1$, and consider the case $n = k + 1$. Writing f_n for the density function of S_n, we have by the convolution formula (4.8.2) that

$$f_{k+1}(x) = \int_0^x f_k(y) f_1(x-y)\,dy = \int_0^x \frac{\lambda^k}{\Gamma(k)} y^{k-1} e^{-\lambda y} \lambda e^{-\lambda(x-y)}\,dy = \frac{\lambda^{k+1} e^{-\lambda x}}{\Gamma(k)} \int_0^x y^{k-1}\,dy,$$

which is easily seen to be the $\Gamma(\lambda, k+1)$ density function.

11. (a) Let $Z_1, Z_2, \dots, Z_{m+n}$ be independent exponential variables with parameter λ. Then, by Problem (4.14.10), $X' = Z_1 + \cdots + Z_m$ is $\Gamma(\lambda, m)$, $Y' = Z_{m+1} + \cdots + Z_{m+n}$ is $\Gamma(\lambda, n)$, and $X' + Y'$ is $\Gamma(\lambda, m+n)$. The pair (X, Y) has the same joint distribution as the pair (X', Y'), and therefore $X + Y$ has the same distribution as $X' + Y'$, i.e., $\Gamma(\lambda, m+n)$.

(b) Using the transformation $u = x + y$, $v = x/(x+y)$, with inverse $x = uv$, $y = u(1-v)$, and Jacobian

$$J = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u,$$

we find that $U = X + Y$, $V = X/(X+Y)$ have joint density function

$$\begin{aligned} f_{U,V}(u, v) = f_{X,Y}\big(uv, u(1-v)\big)|u| &= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} (uv)^{m-1} \{u(1-v)\}^{n-1} e^{-\lambda u} u \\ &= \left\{ \frac{\lambda^{m+n}}{\Gamma(m+n)} u^{m+n-1} e^{-\lambda u} \right\} \left\{ \frac{v^{m-1}(1-v)^{n-1}}{B(m,n)} \right\} \end{aligned}$$

for $u \ge 0$, $0 \le v \le 1$. Hence U and V are independent, U being $\Gamma(\lambda, m+n)$, and V having the beta distribution with parameters m and n.

(c) Integrating by parts,

$$\begin{aligned} \mathbb{P}(X > t) &= \int_t^\infty \frac{\lambda^m}{(m-1)!} x^{m-1} e^{-\lambda x}\,dx \\ &= \left[-\frac{\lambda^{m-1}}{(m-1)!} x^{m-1} e^{-\lambda x} \right]_t^\infty + \int_t^\infty \frac{\lambda^{m-1}}{(m-2)!} x^{m-2} e^{-\lambda x}\,dx \\ &= e^{-\lambda t} \frac{(\lambda t)^{m-1}}{(m-1)!} + \mathbb{P}(X' > t) \end{aligned}$$

where X' is $\Gamma(\lambda, m-1)$. Hence, by induction,

$$\mathbb{P}(X > t) = \sum_{k=0}^{m-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \mathbb{P}(Z < m).$$

(d) This may be achieved by the usual change of variables technique. Alternatively, reflect that, using the notation and result of part (b), the invertible mapping $u = x + y$, $v = x/(x+y)$ maps a pair X, Y of independent ($\Gamma(\lambda, m)$ and $\Gamma(\lambda, n)$) variables to a pair U, V of independent ($\Gamma(\lambda, m+n)$ and $B(m, n)$) variables. Now $UV = X$, so that (figuratively)

$$\text{“}\Gamma(\lambda, m+n) \times B(m, n) = \Gamma(\lambda, m)\text{”}.$$

Replace n by $n - m$ to obtain the required conclusion.

12. (a) $Z = X_1^2$ satisfies

$$f_Z(z) = \frac{d}{dz}\mathbb{P}(X_1^2 \le z) = \frac{d}{dz}\left\{2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\,du\right\} = \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{1}{2}z}, \qquad z \ge 0,$$

the $\Gamma(\frac{1}{2}, \frac{1}{2})$ or $\chi^2(1)$ density function.

(b) If $z \ge 0$, $Z = X_1^2 + X_2^2$ satisfies

$$\begin{aligned}\mathbb{P}(Z \le z) = \mathbb{P}(X_1^2 + X_2^2 \le z) &= \iint_{x^2+y^2\le z} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy \\ &= \int_{r=0}^{\sqrt{z}} \int_{\theta=0}^{2\pi} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r\,dr\,d\theta = 1 - e^{-\frac{1}{2}z},\end{aligned}$$

the $\chi^2(2)$ distribution function.

(c) One way is to work in n-dimensional polar coordinates! Alternatively use induction. It suffices to show that if X and Y are independent, X being $\chi^2(n)$ and Y being $\chi^2(2)$ where $n \ge 1$, then $Z = X + Y$ is $\chi^2(n+2)$. However, by the convolution formula (4.8.2),

$$f_Z(z) = \int_0^z \left\{\frac{2^{-n/2}}{\Gamma(n/2)} x^{\frac{1}{2}n-1} e^{-\frac{1}{2}x}\right\}\left\{\tfrac{1}{2} e^{-\frac{1}{2}(z-x)}\right\} dx = c e^{-\frac{1}{2}z} z^{\frac{1}{2}n}, \qquad z \ge 0,$$

for some constant c. This is the $\chi^2(n+2)$ density function as required.

13. Concentrate on where x occurs in $f_{X|Y}(x \mid y)$; any multiplicative constant can be sorted out later:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = c_1(y)\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2x\mu_1}{\sigma_1^2} - \frac{2\rho x(y-\mu_2)}{\sigma_1\sigma_2}\right)\right\}$$

by Example (4.5.9), where $c_1(y)$ depends on y only. Hence

$$f_{X|Y}(x \mid y) = c_2(y)\exp\left\{-\frac{[x - \mu_1 - \rho\sigma_1(y-\mu_2)/\sigma_2]^2}{2(1-\rho^2)\sigma_1^2}\right\}, \qquad x \in \mathbb{R},$$

for some $c_2(y)$. This is the normal density function with mean $\mu_1 + \rho\sigma_1(y - \mu_2)/\sigma_2$ and variance $\sigma_1^2(1-\rho^2)$. See also Exercise (4.8.7).

14. Set $u = y/x$, $v = x$, with inverse $x = v$, $y = uv$, and $|J| = |v|$. Hence the pair $U = Y/X$, $V = X$ has joint density $f_{U,V}(u,v) = f_{X,Y}(v, uv)|v|$ for $-\infty < u, v < \infty$. Therefore $f_U(u) = \int_{-\infty}^{\infty} f(v, uv)|v|\,dv$.

15. By the result of Problem (4.14.14), $U = Y/X$ has density function

$$f_U(u) = \int_{-\infty}^{\infty} f(y) f(uy)|y|\,dy,$$

and therefore it suffices to show that U has the Cauchy distribution if and only if $Z = \tan^{-1} U$ is uniform on $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$. Clearly

$$\mathbb{P}(Z \le \theta) = \mathbb{P}(U \le \tan\theta), \quad -\tfrac{1}{2}\pi < \theta < \tfrac{1}{2}\pi,$$

whence $f_Z(\theta) = f_U(\tan\theta)\sec^2\theta$. Therefore $f_Z(\theta) = \pi^{-1}$ (for $|\theta| < \frac{1}{2}\pi$) if and only if

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad -\infty < u < \infty.$$

When f is the $N(0,1)$ density,

$$\int_{-\infty}^{\infty} f(x)f(xy)|x|\,dx = 2\int_0^{\infty} \frac{1}{2\pi}e^{-\frac{1}{2}x^2(1+y^2)}x\,dx,$$

which is easily integrated directly to obtain the Cauchy density. In the second case, we have the following integral:

$$\int_{-\infty}^{\infty} \frac{a^2|x|}{(1+x^4)(1+x^4y^4)}\,dx.$$

In this case, make the substitution $z = x^2$ and expand as partial fractions.

16. The transformation $x = r\cos\theta$, $y = r\sin\theta$ has Jacobian $J = r$, so that

$$f_{R,\Theta}(r,\theta) = \frac{1}{2\pi}re^{-\frac{1}{2}r^2}, \qquad r > 0, \quad 0 \le \theta < 2\pi.$$

Therefore R and Θ are independent, Θ being uniform on $[0, 2\pi)$, and R^2 having distribution function

$$\mathbb{P}(R^2 \le a) = \int_0^{\sqrt{a}} re^{-\frac{1}{2}r^2}\,dr = 1 - e^{-\frac{1}{2}a};$$

this is the exponential distribution with parameter $\frac{1}{2}$ (otherwise known as $\Gamma(\frac{1}{2}, 1)$ or $\chi^2(2)$). The density function of R is $f_R(r) = re^{-\frac{1}{2}r^2}$ for $r > 0$.

Now, by symmetry,

$$\mathbb{E}\left(\frac{X^2}{R^2}\right) = \frac{1}{2}\mathbb{E}\left(\frac{X^2+Y^2}{R^2}\right) = \frac{1}{2}.$$

In the first octant, i.e., in $\{(x,y) : 0 \le y \le x\}$, we have $\min\{x,y\} = y$, $\max\{x,y\} = x$. The joint density $f_{X,Y}$ is invariant under rotations, and hence the expectation in question is

$$8\int_{0\le y\le x} \frac{y}{x} f_{X,Y}(x,y)\,dx\,dy = 8\int_{\theta=0}^{\pi/4}\int_{r=0}^{\infty} \frac{\tan\theta}{2\pi}re^{-\frac{1}{2}r^2}\,dr\,d\theta = \frac{2}{\pi}\log 2.$$

17. (i) Using independence,

$$\mathbb{P}(U \le u) = 1 - \mathbb{P}(X > u, Y > u) = 1 - \big(1 - F_X(u)\big)\big(1 - F_Y(u)\big).$$

Similarly

$$\mathbb{P}(V \le v) = \mathbb{P}(X \le v, Y \le v) = F_X(v)F_Y(v).$$

(ii) (a) By (i), $\mathbb{P}(U \le u) = 1 - e^{-2u}$ for $u \ge 0$.

(b) Also, $Z = X + \frac{1}{2}Y$ satisfies

$$\begin{aligned}\mathbb{P}(Z > v) &= \int_0^{\infty} \mathbb{P}\big(Y > 2(v-x)\big)f_X(x)\,dx = \int_0^{v} e^{-2(v-x)}e^{-x}\,dx + \int_v^{\infty} e^{-x}\,dx \\ &= e^{-2v}(e^v - 1) + e^{-v} = 1 - (1 - e^{-v})^2 = \mathbb{P}(V > v).\end{aligned}$$

Therefore $\mathbb{E}(V) = \mathbb{E}(X) + \frac{1}{2}\mathbb{E}(Y) = \frac{3}{2}$, and $\operatorname{var}(V) = \operatorname{var}(X) + \frac{1}{4}\operatorname{var}(Y) = \frac{5}{4}$ by independence.

18. (a) We have that

$$\mathbb{P}(X \le Y) = \int_0^\infty \mathbb{P}(X \le y)\mu e^{-\mu y}\,dy = \int_0^\infty (1 - e^{-\lambda y})\mu e^{-\mu y}\,dy = \frac{\lambda}{\mu + \lambda}.$$

(b) Clearly, for $w > 0$,

$$\mathbb{P}(U \le u, W > w) = \mathbb{P}(U \le u, W > w, X \le Y) + \mathbb{P}(U \le u, W > w, X > Y).$$

Now

$$\begin{aligned}\mathbb{P}(U \le u, W > w, X \le Y) = \mathbb{P}(X \le u, Y > X + w) &= \int_0^u \lambda e^{-\lambda x} e^{-\mu(x+w)}\,dx \\ &= \frac{\lambda}{\lambda + \mu} e^{-\mu w}(1 - e^{-(\lambda+\mu)u})\end{aligned}$$

and similarly

$$\mathbb{P}(U \le u, W > w, X > Y) = \frac{\mu}{\lambda + \mu} e^{-\lambda w}(1 - e^{-(\lambda+\mu)u}).$$

Hence, for $0 \le u \le u + w < \infty$,

$$\mathbb{P}(U \le u, W > w) = (1 - e^{-(\lambda+\mu)u})\left(\frac{\lambda}{\lambda+\mu} e^{-\mu w} + \frac{\mu}{\lambda+\mu} e^{-\lambda w}\right),$$

an expression which factorizes into the product of a function of u with a function of w. Hence U and W are independent.

19. $U = X + Y$, $V = X$ have joint density function $f_Y(u - v)f_X(v)$, $0 \le v \le u$. Hence

$$f_{V|U}(v \mid u) = \frac{f_{U,V}(u, v)}{f_U(u)} = \frac{f_Y(u - v)f_X(v)}{\int_0^u f_Y(u - y)f_X(y)\,dy}.$$

(a) We are given that $f_{V|U}(v \mid u) = u^{-1}$ for $0 \le v \le u$; then

$$f_Y(u - v)f_X(v) = \frac{1}{u}\int_0^u f_Y(u - y)f_X(y)\,dy$$

is a function of u alone, implying that

$$\begin{aligned}f_Y(u - v)f_X(v) &= f_Y(u)f_X(0) &&\text{by setting } v = 0 \\ &= f_Y(0)f_X(u) &&\text{by setting } v = u.\end{aligned}$$

In particular $f_Y(u)$ and $f_X(u)$ differ only by a multiplicative constant; they are both density functions, implying that this constant is 1, and $f_X = f_Y$. Substituting this throughout the above display, we find that $g(x) = f_X(x)/f_X(0)$ satisfies $g(0) = 1$, g is continuous, and $g(u - v)g(v) = g(u)$ for $0 \le v \le u$. From the hint, $g(x) = e^{-\lambda x}$ for $x \ge 0$ for some $\lambda > 0$ (remember that g is integrable).

(b) Arguing similarly, we find that

$$f_Y(u - v)f_X(v) = \frac{c}{u^{\alpha+\beta-1}} v^{\alpha-1}(u - v)^{\beta-1}\int_0^u f_Y(u - y)f_X(y)\,dy$$

for $0 \le v \le u$ and some constant c. Setting $f_X(v) = \chi(v)v^{\alpha-1}$, $f_Y(y) = \eta(y)y^{\beta-1}$, we obtain $\eta(u-v)\chi(v) = h(u)$ for $0 \le v \le u$, and for some function h. Arguing as before, we find that η and χ are proportional to negative exponential functions, so that X and Y have gamma distributions.

20. We are given that U is uniform on [0, 1], so that $0 \le X, Y \le \frac{1}{2}$ almost surely. For $0 < \epsilon < \frac{1}{4}$,

$$\epsilon = \mathbb{P}(X + Y < \epsilon) \le \mathbb{P}(X < \epsilon, Y < \epsilon) = \mathbb{P}(X < \epsilon)^2,$$

and similarly

$$\epsilon = \mathbb{P}(X + Y > 1 - \epsilon) \le \mathbb{P}(X > \tfrac{1}{2} - \epsilon, Y > \tfrac{1}{2} - \epsilon) = \mathbb{P}(X > \tfrac{1}{2} - \epsilon)^2,$$

implying that $\mathbb{P}(X < \epsilon) \ge \sqrt{\epsilon}$ and $\mathbb{P}(X > \frac{1}{2} - \epsilon) \ge \sqrt{\epsilon}$. Now

$$\begin{aligned} 2\epsilon &= \mathbb{P}\big(\tfrac{1}{2} - \epsilon < X + Y < \tfrac{1}{2} + \epsilon\big) \ge \mathbb{P}\big(X > \tfrac{1}{2} - \epsilon, Y < \epsilon\big) + \mathbb{P}\big(X < \epsilon, Y > \tfrac{1}{2} - \epsilon\big) \\ &= 2\mathbb{P}(X > \tfrac{1}{2} - \epsilon)\mathbb{P}(X < \epsilon) \ge 2(\sqrt{\epsilon})^2. \end{aligned}$$

Therefore all the above inequalities are in fact equalities, implying that $\mathbb{P}(X < \epsilon) = \mathbb{P}(X > \frac{1}{2} - \epsilon) = \sqrt{\epsilon}$ if $0 < \epsilon < \frac{1}{4}$. Hence a contradiction:

$$\tfrac{1}{8} = \mathbb{P}(X + Y < \tfrac{1}{8}) = \mathbb{P}(X, Y < \tfrac{1}{8}) - \mathbb{P}(X, Y < \tfrac{1}{8},\ X + Y \ge \tfrac{1}{8}) < \mathbb{P}(X < \tfrac{1}{8}, Y < \tfrac{1}{8}) = \tfrac{1}{8}.$$

21. Evidently

$$\mathbb{P}(X_{(1)} \le y_1, \dots, X_{(n)} \le y_n) = \sum_{\pi} \mathbb{P}\left(X_{\pi_1} \le y_1, \dots, X_{\pi_n} \le y_n, X_{\pi_1} < \cdots < X_{\pi_n}\right)$$

where the sum is over all permutations $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ of $(1, 2, \dots, n)$. By symmetry, each term in the sum is equal, whence the sum equals

$$n!\,\mathbb{P}(X_1 \le y_1, \dots, X_n \le y_n, X_1 < X_2 < \cdots < X_n).$$

The integral form is then immediate. The joint density function is, by its definition, the integrand.

22. (a) In the notation of Problem (4.14.21), the joint density function of $X_{(2)}, \dots, X_{(n)}$ is

$$\begin{aligned} g_2(y_2, \dots, y_n) &= \int_{-\infty}^{y_2} g(y_1, \dots, y_n)\,dy_1 \\ &= n!\,L(y_2, \dots, y_n)F(y_2)f(y_2)f(y_3)\cdots f(y_n) \end{aligned}$$

where F is the common distribution function of the X_i. Similarly $X_{(3)}, \dots, X_{(n)}$ have joint density

$$g_3(y_3, \dots, y_n) = \tfrac{1}{2}n!\,L(y_3, \dots, y_n)F(y_3)^2 f(y_3)\cdots f(y_n),$$

and by iteration, $X_{(k)}, \dots, X_{(n)}$ have joint density

$$g_k(y_k, \dots, y_n) = \frac{n!}{(k-1)!}L(y_k, \dots, y_n)F(y_k)^{k-1}f(y_k)\cdots f(y_n).$$

We now integrate over $y_n, y_{n-1}, \dots, y_{k+1}$ in turn, arriving at

$$f_{X_{(k)}}(y_k) = \frac{n!}{(k-1)!\,(n-k)!}F(y_k)^{k-1}\{1 - F(y_k)\}^{n-k}f(y_k).$$

(b) It is neater to argue directly. Fix x, and let I_r be the indicator function of the event $\{X_r \le x\}$, and let $S = I_1 + I_2 + \cdots + I_n$. Then S is distributed as $\text{bin}(n, F(x))$, and

$$\mathbb{P}(X_{(k)} \le x) = \mathbb{P}(S \ge k) = \sum_{l=k}^{n} \binom{n}{l} F(x)^l (1 - F(x))^{n-l}.$$

Differentiate to obtain, with $F = F(x)$,

$$\begin{aligned} f_{X_{(k)}}(x) &= \sum_{l=k}^{n} \binom{n}{l} f(x) \left\{ l F^{l-1}(1-F)^{n-l} - (n-l)F^l(1-F)^{n-l-1} \right\} \\ &= k\binom{n}{k} f(x) F^{k-1}(1-F)^{n-k} \end{aligned}$$

by successive cancellation of the terms in the series.

23. Using the result of Problem (4.14.21), the joint density function is $g(\mathbf{y}) = n!\, L(\mathbf{y}) T^{-n}$ for $0 \le y_i \le T$, $1 \le i \le n$, where $\mathbf{y} = (y_1, y_2, \ldots, y_n)$.

24. (a) We make use of Problems (4.14.22)–(4.14.23). The density function of $X_{(k)}$ is $f_k(x) = k\binom{n}{k} x^{k-1}(1-x)^{n-k}$ for $0 \le x \le 1$, so that the density function of $nX_{(k)}$ is

$$\frac{1}{n} f_k(x/n) = \frac{k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} x^{k-1} \left(1 - \frac{x}{n}\right)^{n-k} \to \frac{1}{(k-1)!} x^{k-1} e^{-x}$$

as $n \to \infty$. The limit is the $\Gamma(1, k)$ density function.

(b) For an increasing sequence $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ in $[0, 1]$, we define the sequence $u_n = -n \log x_{(n)}$, $u_k = -k \log(x_{(k)}/x_{(k+1)})$ for $1 \le k < n$. This mapping has inverse

$$x_{(n)} = e^{-u_n/n}, \quad x_{(k)} = x_{(k+1)} e^{-u_k/k} = \exp\left\{ -\sum_{i=k}^{n} i^{-1} u_i \right\},$$

with Jacobian $J = (-1)^n e^{-u_1 - u_2 - \cdots - u_n}/n!$. Applying the mapping to the sequence $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ we obtain a family $U_1, U_2, \ldots, U_n$ of random variables with joint density $g(u_1, u_2, \ldots, u_n) = e^{-u_1 - u_2 - \cdots - u_n}$ for $u_i \ge 0$, $1 \le i \le n$, yielding that the U_i are independent and exponentially distributed, with parameter 1. Finally $\log X_{(k)} = -\sum_{i=k}^{n} i^{-1} U_i$.

(c) In the notation of part (b), $Z_k = \exp(-U_k)$ for $1 \le k \le n$, a collection of independent variables. Finally, U_k is exponential with parameter 1, and therefore

$$\mathbb{P}(Z_k \le z) = \mathbb{P}(U_k \ge -\log z) = e^{\log z} = z, \qquad 0 < z \le 1.$$

25. (i) (X_1, X_2, X_3) is uniformly distributed over the unit cube of $\mathbb{R}^3$, and the answer is therefore the volume of that set of points (x_1, x_2, x_3) of the cube which allow a triangle to be formed. A triangle is impossible if $x_1 \ge x_2 + x_3$, or $x_2 \ge x_1 + x_3$, or $x_3 \ge x_1 + x_2$. This defines three regions of the cube which overlap only at the origin. Each of these regions is a tetrahedron; for example, the region $x_3 \ge x_1 + x_2$ is an isosceles tetrahedron with vertices $(0, 0, 0)$, $(1, 0, 1)$, $(0, 1, 1)$, $(0, 0, 1)$, with volume $\frac{1}{6}$. Hence the required probability is $1 - 3 \cdot \frac{1}{6} = \frac{1}{2}$.

(ii) The rods of length $x_1, x_2, \ldots, x_n$ fail to form a polygon if either $x_1 \ge x_2 + x_3 + \cdots + x_n$ or any of the other $n - 1$ corresponding inequalities hold. We therefore require the volume of the n-dimensional hypercube with n corners removed. The inequality $x_1 \ge x_2 + x_3 + \cdots + x_n$ corresponds to the convex hull of the points $(0, 0, \ldots, 0)$, $(1, 0, \ldots, 0)$, $(1, 1, 0, \ldots, 0)$, $(1, 0, 1, 0, \ldots, 0)$, $\ldots$, $(1, 0, \ldots, 0, 1)$.

Mapping $x_1 \mapsto 1 - x_1$, we see that this has the same volume V_n as has the convex hull of the origin $\mathbf{0}$ together with the n unit vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$. Clearly $V_2 = \frac{1}{2}$, and we claim that $V_n = 1/n!$. Suppose this holds for $n < k$, and consider the case $n = k$. Then

$$V_k = \int_0^1 dx_1\, V_{k-1}(\mathbf{0}, x_1\mathbf{e}_2, \dots, x_1\mathbf{e}_k)$$

where $V_{k-1}(\mathbf{0}, x_1\mathbf{e}_2, \dots, x_1\mathbf{e}_k)$ is the $(k-1)$-dimensional volume of the convex hull of $\mathbf{0}, x_1\mathbf{e}_2, \dots, x_1\mathbf{e}_k$. Now

$$V_{k-1}(\mathbf{0}, x_1\mathbf{e}_2, \dots, x_1\mathbf{e}_k) = x_1^{k-1} V_{k-1} = \frac{x_1^{k-1}}{(k-1)!},$$

so that

$$V_k = \int_0^1 \frac{x_1^{k-1}}{(k-1)!}\, dx_1 = \frac{1}{k!}.$$

The required probability is therefore $1 - n/(n!) = 1 - \{(n-1)!\}^{-1}$.

26. (i) The lengths of the pieces are $U = \min\{X_1, X_2\}$, $V = |X_1 - X_2|$, $W = 1 - U - V$, and we require that $U < V + W$, etc, as in the solution to Problem (4.14.25). In terms of the X_i we require

$$\begin{aligned} \text{either}: &\quad X_1 < \tfrac{1}{2}, \quad |X_1 - X_2| < \tfrac{1}{2}, \quad 1 - X_2 < \tfrac{1}{2}, \\ \text{or}: &\quad X_2 < \tfrac{1}{2}, \quad |X_1 - X_2| < \tfrac{1}{2}, \quad 1 - X_1 < \tfrac{1}{2}. \end{aligned}$$

Plot the corresponding region of $\mathbb{R}^2$. One then sees that the area of the region is $\frac{1}{4}$, which is therefore the probability in question.

(ii) The pieces may form a polygon if no piece is as long as the sum of the lengths of the others. Since the total length is 1, this requires that each piece has length less than $\frac{1}{2}$. Neglecting certain null events, this fails to occur if and only if the disjoint union of events $A_0 \cup A_1 \cup \dots \cup A_n$ occurs, where

$$A_0 = \{\text{no break in } (0, \tfrac{1}{2}]\}, \quad A_k = \{\text{no break in } (X_k, X_k + \tfrac{1}{2}]\} \text{ for } 1 \le k \le n;$$

remember that there is a permanent break at 1. Now $\mathbb{P}(A_0) = (\frac{1}{2})^n$, and for $k \ge 1$,

$$\mathbb{P}(A_k) = \int_0^1 \mathbb{P}(A_k \mid X_k = x)\, dx = \int_0^{\frac{1}{2}} (\tfrac{1}{2})^{n-1}\, dx = (\tfrac{1}{2})^n;$$

Hence $\mathbb{P}(A_0 \cup A_1 \cup \dots \cup A_n) = (n+1)2^{-n}$ whence the required probability is $1 - (n+1)2^{-n}$.

27. (a) The function $g(t) = (t^p/p) + (t^{-q}/q)$, for $t > 0$, has a unique minimum at $t = 1$, and hence $g(t) \ge g(1) = 1$ for $t > 0$. Substitute $t = x^{1/q} y^{-1/p}$ where

$$x = \frac{|X|}{\{\mathbb{E}|X^p|\}^{1/p}}, \quad y = \frac{|Y|}{\{\mathbb{E}|Y^q|\}^{1/q}},$$

(we may as well suppose that $\mathbb{P}(XY = 0) \ne 1$) to find that

$$\frac{|X|^p}{p\mathbb{E}|X^p|} + \frac{|Y|^q}{q\mathbb{E}|Y^q|} \ge \frac{|XY|}{\{\mathbb{E}|X^p|\}^{1/p}\{\mathbb{E}|Y^q|\}^{1/q}}.$$

Hölder's inequality follows by taking expectations.

(b) We have, with $Z = |X + Y|$,

$$\begin{aligned}\mathbb{E}(Z^p) &= \mathbb{E}(Z \cdot Z^{p-1}) \le \mathbb{E}(|X|Z^{p-1}) + \mathbb{E}(|Y|Z^{p-1}) \\ &\le \{\mathbb{E}|X^p|\}^{1/p}\{\mathbb{E}(Z^p)\}^{1/q} + \{\mathbb{E}|Y^p|\}^{1/p}\{\mathbb{E}(Z^p)\}^{1/q}\end{aligned}$$

by Hölder's inequality, where $p^{-1} + q^{-1} = 1$. Divide by $\{\mathbb{E}(Z^p)\}^{1/q}$ to get the result.

28. Apply the Cauchy–Schwarz inequality to $|Z|^{\frac{1}{2}(b-a)}$ and $|Z|^{\frac{1}{2}(b+a)}$, where $0 \le a \le b$, to obtain $\{\mathbb{E}|Z^b|\}^2 \le \mathbb{E}|Z^{b-a}|\,\mathbb{E}|Z^{b+a}|$. Now take logarithms: $2g(b) \le g(b-a) + g(b+a)$ for $0 \le a \le b$. Also $g(p) \to g(0) = 1$ as $p \downarrow 0$ (by dominated convergence). These two properties of g imply that g is convex on intervals of the form $[0, M)$ if it is finite on this interval. The reference to dominated convergence may be avoided by using Hölder instead of Cauchy–Schwarz.

By convexity, $g(x)/x$ is non-decreasing in x, and therefore $g(r)/r \ge g(s)/s$ if $0 < s \le r$.

29. Assume that X, Y, Z are jointly continuously distributed with joint density function f. Then

$$\mathbb{E}(X \mid Y = y, Z = z) = \int x f_{X|Y,Z}(x \mid y, z)\,dx = \int x \frac{f(x, y, z)}{f_{Y,Z}(y, z)}\,dx.$$

Hence

$$\begin{aligned}\mathbb{E}\{\mathbb{E}(X \mid Y, Z) \mid Y = y\} &= \int \mathbb{E}(X \mid Y = y, Z = z) f_{Z|Y}(z \mid y)\,dz \\ &= \iint x \frac{f(x, y, z)}{f_{Y,Z}(y, z)} \frac{f_{Y,Z}(y, z)}{f_Y(y)}\,dx\,dz \\ &= \iint x \frac{f(x, y, z)}{f_Y(y)}\,dx\,dz = \mathbb{E}(X \mid Y = y).\end{aligned}$$

Alternatively, use the general properties of conditional expectation as laid down in Section 4.6.

30. The first car to arrive in a car park of length $x + 1$ effectively divides it into two disjoint parks of length Y and $x - Y$, where Y is the position of the car's leftmost point. Now Y is uniform on $[0, x]$, and the formula follows by conditioning on Y. Laplace transforms are the key to exploring the asymptotic behaviour of $m(x)/x$ as $x \to \infty$.

31. (a) If the needle were of length d, the answer would be $2/\pi$ as before. Think about the new needle as being obtained from a needle of length d by dividing it into two parts, an 'active' part of length L, and a 'passive' part of length $d - L$, and then counting only intersections involving the active part. The chance of an 'active intersection' is now $(2/\pi)(L/d) = 2L/(\pi d)$.

(b) As in part (a), the angle between the line and the needle is independent of the distance between the line and the needle's centre, each having the same distribution as before. The answer is therefore unchanged.

(c) The following argument lacks a little rigour, which may be supplied as a consequence of the statement that S has finite length. For $\epsilon > 0$, let $x_1, x_2, \dots, x_n$ be points on S, taken in order along S, such that x_0 and x_n are the endpoints of S, and $|x_{i+1} - x_i| < \epsilon$ for $0 \le i < n$; $|x - y|$ denotes the Euclidean distance from x to y. Let J_i be the straight line segment joining x_i to x_{i+1}, and let I_i be the indicator function of $\{J_i \cap \lambda \ne \varnothing\}$. If ϵ is sufficiently small, the total number of intersections between $J_0 \cup J_1 \cup \cdots \cup J_{n-1}$ and S has mean

$$\sum_{i=0}^{n-1} \mathbb{E}(I_i) = \frac{2}{\pi d} \sum_{i=0}^{n-1} |x_{i+1} - x_i|$$

by part (b). In the limit as $\epsilon \downarrow 0$, we have that $\sum_i \mathbb{E}(I_i)$ approaches the required mean, while

$$\frac{2}{\pi d}\sum_{i=0}^{n-1}|x_{i+1}-x_i| \to \frac{2L(S)}{\pi d}.$$

32. (i) Fix Cartesian axes within the gut in question. Taking one end of the needle as the origin, the other end is uniformly distributed on the unit sphere of $\mathbb{R}^3$. With the X-ray plate parallel to the (x, y)-plane, the projected length V of the needle satisfies $V \ge v$ if and only if $|Z| \le \sqrt{1-v^2}$, where Z is the (random) z-coordinate of the 'loose' end of the needle. Hence, for $0 \le v \le 1$,

$$\mathbb{P}(V \ge v) = \mathbb{P}\big(-\sqrt{1-v^2} \le Z \le \sqrt{1-v^2}\big) = \frac{4\pi\sqrt{1-v^2}}{4\pi} = \sqrt{1-v^2},$$

since $4\pi\sqrt{1-v^2}$ is the surface area of that part of the unit sphere satisfying $|z| \le \sqrt{1-v^2}$ (use Archimedes's theorem of the circumscribing cylinder, or calculus). Therefore V has density function $f_V(v) = v/\sqrt{1-v^2}$ for $0 \le v \le 1$.

(ii) Draw a picture, if you can. The lengths of the projections are determined by the angle Θ between the plane of the cross and the X-ray plate, together with the angle Ψ of rotation of the cross about an axis normal to its arms. Assume that Θ and Ψ are independent and uniform on $[0, \frac{1}{2}\pi]$. If the axis system has been chosen with enough care, we find that the lengths A, B of the projections of the arms are given by

$$A = \sqrt{\cos^2\Psi + \sin^2\Psi\cos^2\Theta}, \qquad B = \sqrt{\sin^2\Psi + \cos^2\Psi\cos^2\Theta},$$

with inverse

$$\Theta = \cos^{-1}\sqrt{A^2+B^2-1}, \qquad \Psi = \tan^{-1}\sqrt{\frac{1-A^2}{1-B^2}}.$$

Some slog is now required to calculate the Jacobian J of this mapping, and the answer will be $f_{A,B}(a, b) = 4|J|\pi^{-2}$ for $0 < a, b < 1$, $a^2 + b^2 > 1$.

33. The order statistics of the X_i have joint density function

$$h(x_1, x_2, \dots, x_n) = \lambda^n n! \exp\Big(-\sum_{i=1}^n x_i\Big)$$

on the set I of increasing sequences of positive reals. Define the one–one mapping from I onto $(0, \infty)^n$ by

$$y_1 = nx_1, \qquad y_r = (n+1-r)(x_r - x_{r-1}) \quad \text{for } 1 < r \le n,$$

with inverse $x_r = \sum_{k=1}^r y_k/(n-k+1)$ for $r \ge 1$. The Jacobian is $(n!)^{-1}$, whence the joint density function of $Y_1, Y_2, \dots, Y_n$ is

$$\frac{1}{n!}\lambda^n n! \exp\Big(-\sum_{i=1}^n x_i(\mathbf{y})\Big) = \lambda^n \exp\Big(-\sum_{k=1}^n y_k\Big).$$

34. Recall Problem (4.14.4). First, $Z_i = F(X_i)$, $1 \le i \le n$, is a sequence of independent variables with the uniform distribution on $[0, 1]$. Secondly, a variable U has the latter distribution if and only if $-\log U$ has the exponential distribution with parameter 1.

It follows that $L_i = -\log F(X_i)$, $1 \le i \le n$, is a sequence of independent variables with the exponential distribution. The order statistics $L_{(1)}, \dots, L_{(n)}$ are *in order* $-\log F(X_{(n)}), \dots, -\log F(X_{(1)})$, since the function $-\log F(\cdot)$ is non-increasing. Applying the result of Problem (4.14.33), $E_1 = -n \log F(X_{(n)})$ and

$$E_r = -(n+1-r)\big\{\log F(X_{(n+1-r)}) - \log F(X_{(n+2-r)})\big\}, \quad 1 < r \le n,$$

are independent with the exponential distribution. Therefore $\exp(-E_r)$, $1 \le r \le n$, are independent with the uniform distribution.

35. One may be said to be in state j if the first $j-1$ prizes have been rejected and the jth prize has just been viewed. There are two possible decisions at this stage: accept the jth prize if it is the best so far (there is no point in accepting it if it is not), or reject it and continue. The mean return of the first decision equals the probability j/n that the jth prize is the best so far, and the mean return of the second is the maximal probability $f(j)$ that one may obtain the best prize having rejected the first j. Thus the maximal mean return $V(j)$ in state j satisfies

$$V(j) = \max\{j/n, f(j)\}.$$

Now j/n increases with j, and $f(j)$ decreases with j (since a possible stategy is to reject the $(j+1)$th prize also). Therefore there exists J such that $j/n \le f(j)$ if and only if $j \le J$. This confirms the optimal stategy as having the following form: reject the first J prizes out of hand, and accept the subsequent prize which is the best of those viewed so far. If there is no such prize, we pick the last prize presented.

Let Π_J be the probability of achieving the best prize by following the above strategy. Let A_k be the event that you pick the kth prize, and B the event that the prize picked is the best. Then,

$$\Pi_J = \sum_{k=J+1}^{n} \mathbb{P}(B \mid A_k)\mathbb{P}(A_k) = \sum_{k=J+1}^{n} \left(\frac{k}{n}\right)\left(\frac{J}{k-1} \cdot \frac{1}{k}\right) = \frac{J}{n} \sum_{k=J+1}^{n} \frac{1}{k-1},$$

and we choose the integer J which maximizes this expresion.

When n is large, we have the asymptotic relation $\Pi_J \simeq (J/n)\log(n/J)$. The maximum of the function $h_n(x) = (x/n)\log(n/x)$ occurs at $x = n/e$, and we deduce that $J \simeq n/e$. [A version of this problem was posed by Cayley in 1875. Our solution is due to Lindley (1961).]

36. The joint density function of (X, Y, Z) is

$$f(x, y, z) = \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\{-\tfrac{1}{2}(r^2 - 2\lambda x - 2\mu y - 2\nu z + \lambda^2 + \mu^2 + \nu^2)\}$$

where $r^2 = x^2 + y^2 + z^2$. The conditional density of X, Y, Z given $R = r$ is therefore proportional to $\exp\{\lambda x + \mu y + \nu z\}$. Now choosing spherical polar coordinates with axis in the direction (λ, μ, ν), we obtain a density function proportional to $\exp(a\cos\theta)\sin\theta$, where $a = r\sqrt{\lambda^2 + \mu^2 + \nu^2}$. The constant is chosen in such a way that the total probability is unity.

37. (a) $\phi'(x) = -x\phi(x)$, so $H_1(x) = x$. Differentiate the equation for H_n to obtain $H_{n+1}(x) = xH_n(x) - H_n'(x)$, and use induction to deduce that H_n is a polynomial of degree n as required. Integrating by parts gives, when $m \le n$,

$$\begin{aligned}\int_{-\infty}^{\infty} H_m(x)H_n(x)\phi(x)\,dx &= (-1)^n \int_{-\infty}^{\infty} H_m(x)\phi^{(n)}(x)\,dx \\ &= (-1)^{n-1} \int_{-\infty}^{\infty} H_m'(x)\phi^{(n-1)}(x)\,dx \\ &= \cdots = (-1)^{n-m} \int_{-\infty}^{\infty} H_m^{(m)}(x)\phi^{(n-m)}(x)\,dx,\end{aligned}$$

and the claim follows by the fact that $H_m^{(m)}(x) = m!$.
(b) By Taylor's theorem and the first part,

$$\phi(x)\sum_{n=0}^{\infty}\frac{t^n}{n!}H_n(x) = \sum_{n=0}^{\infty}\frac{(-t)^n}{n!}\phi^{(n)}(x) = \phi(x-t),$$

whence

$$\sum_{n=0}^{\infty}\frac{t^n}{n!}H_n(x) = \exp\{-\tfrac{1}{2}(x-t)^2 + \tfrac{1}{2}x^2\} = \exp(xt - \tfrac{1}{2}t^2).$$

38. The polynomials of Problem (4.14.37) are orthogonal, and there are unique expansions (subject to mild conditions) of the form $u(x) = \sum_{r=0}^{\infty} a_r H_r(x)$ and $v(x) = \sum_{r=0}^{\infty} b_r H_r(x)$. Without loss of generality, we may assume that $\mathbb{E}(U) = \mathbb{E}(V) = 0$, whence, by Problem (4.14.37a), $a_0 = b_0 = 0$. By (4.14.37a) again,

$$\operatorname{var}(U) = \mathbb{E}(u(X)^2) = \sum_{r=1}^{\infty} a_r^2 r!, \quad \operatorname{var}(V) = \sum_{r=1}^{\infty} b_r^2 r!.$$

By (4.14.37b),

$$\mathbb{E}\left(\sum_{m=0}^{\infty}\frac{H_m(X)s^m}{m!}\sum_{n=0}^{\infty}\frac{H_n(Y)t^n}{n!}\right) = \mathbb{E}\big(\exp\{sX - \tfrac{1}{2}s^2 + tY - \tfrac{1}{2}t^2\}\big) = e^{st\rho}.$$

By considering the coefficient of $s^m t^n$,

$$\mathbb{E}\big(H_m(X)H_n(Y)\big) = \begin{cases} \rho^n n! & \text{if } m = n,\\ 0 & \text{if } m \neq n,\end{cases}$$

and so

$$\operatorname{cov}(U, V) = \mathbb{E}\left(\sum_{m=1}^{\infty} a_m H_m(X)\sum_{n=1}^{\infty} b_n H_n(Y)\right) = \sum_{n=1}^{\infty} a_n b_n \rho^n n!,$$

$$|\rho(U, V)| = \frac{|\rho|\left|\sum_{n=1}^{\infty} a_n b_n \rho^{n-1} n!\right|}{\sqrt{\sum_{n=1}^{\infty} a_n^2 n! \sum_{n=1}^{\infty} b_n^2 n!}} \le |\rho|\frac{\sum_{n=1}^{\infty} |a_n b_n| n!}{\sqrt{\sum_{n=1}^{\infty} a_n^2 n! \sum_{n=1}^{\infty} b_n^2 n!}} \le |\rho|,$$

where we have used the Cauchy–Schwarz inequality at the last stage.

39. (a) Let $Y_r = X_{(r)} - X_{(r-1)}$ with the convention that $X_{(0)} = 0$ and $X_{(n+1)} = 1$. By Problem (4.14.21) and a change of variables, we may see that $Y_1, Y_2, \dots, Y_{n+1}$ have the distribution of a point chosen uniformly at random in the simplex of non-negative vectors $\mathbf{y} = (y_1, y_2, \dots, y_{n+1})$ with sum 1. [This may also be derived using a Poisson process representation and Theorem (6.12.7).] Consequently, the Y_j are identically distributed, and their joint distribution is invariant under permutations of the indices of the Y_j. Now $\sum_{r=1}^{n+1} Y_r = 1$ and, by taking expectations, $(n+1)\mathbb{E}(Y_1) = 1$, whence $\mathbb{E}(X_{(r)}) = r\mathbb{E}(Y_1) = r/(n+1)$.
(b) We have that

$$\mathbb{E}(Y_1^2) = \int_0^1 x^2 n(1-x)^{n-1}\,dx = \frac{2}{(n+1)(n+2)},$$

$$1 = \mathbb{E}\left[\left(\sum_{r=1}^{n+1} Y_r\right)^2\right] = (n+1)\mathbb{E}(Y_1^2) + n(n+1)\mathbb{E}(Y_1 Y_2),$$

implying that

$$\mathbb{E}(Y_1Y_2) = \frac{1}{(n+1)(n+2)},$$

and also

$$\mathbb{E}(X_{(r)}X_{(s)}) = r\mathbb{E}(Y_1^2) + r(s-1)\mathbb{E}(Y_1Y_2) = \frac{r(s+1)}{(n+1)(n+2)}.$$

The required covariance follows.

40. (a) By paragraph (4.4.6), X^2 is $\Gamma(\frac{1}{2}, \frac{1}{2})$ and $Y^2 + Z^2$ is $\Gamma(\frac{1}{2}, 1)$. Now use the results of Exercise (4.7.14).

(b) Since the distribution of X^2/R^2 is independent of the value of $R^2 = X^2 + Y^2 + Z^2$, it is valid also if the three points are picked independently and uniformly within the sphere.

41. (a) Immediate, because the $N(0, 1)$ distribution is symmetric.

(b) We have that

$$\begin{aligned}\int_{-\infty}^{\infty} 2\phi(x)\Phi(\lambda x)\,dx &= \int_{-\infty}^{\infty} \{\phi(x)\Phi(\lambda x) + \phi(x)[1 - \Phi(-\lambda x)]\}\,dx \\ &= \int_{-\infty}^{\infty} \phi(x)\,dx + \int_{-\infty}^{\infty} \phi(x)[\Phi(\lambda x) - \Phi(-\lambda x)]\,dx = 1,\end{aligned}$$

because the second integrand is odd. Using the symmetry of ϕ, the density function of $|Y|$ is

$$\phi(x) + \phi(x)\big\{\Phi(\lambda x) - \Phi(-\lambda x)\big\} + \phi(-x) + \phi(-x)\big\{\Phi(-\lambda x) - \Phi(\lambda x)\big\} = 2\phi(x).$$

(c) Finally, make the change of variables $W = |Y|$, $Z = (X + \lambda|Y|)/\sqrt{1+\lambda^2}$, with inverse $|y| = w$, $x = z\sqrt{1+\lambda^2} - \lambda w$, and Jacobian $\sqrt{1+\lambda^2}$. Then

$$\begin{aligned}f_{W,Z}(w,z) &= \sqrt{1+\lambda^2}\, f_{X,|Y|}\big(z\sqrt{1+\lambda^2} - \lambda w, w\big) \\ &= \sqrt{1+\lambda^2}\cdot\phi\big(z\sqrt{1+\lambda^2} - \lambda w\big)\cdot 2\phi(w), \qquad w > 0,\ x \in \mathbb{R}.\end{aligned}$$

The result follows by integrating over w and using the fact that

$$\int_0^{\infty} \phi\big(z\sqrt{1+\lambda^2} - \lambda w\big)\phi(w)\,dw = \frac{\phi(z)\Phi(\lambda z)}{\sqrt{1+\lambda^2}}.$$

42. The required probability equals

$$\begin{aligned}\mathbb{P}\Big(\{X_3 - \tfrac{1}{2}(X_1+X_2)\}^2 + \{Y_3 - \tfrac{1}{2}(Y_1+Y_2)\}^2 \le \tfrac{1}{4}(X_1-X_2)^2 + \tfrac{1}{4}(Y_1-Y_2)^2\Big) \\ = \mathbb{P}(U_1^2 + U_2^2 \le V_1^2 + V_2^2)\end{aligned}$$

where U_1, U_2 are $N(0, \frac{3}{2})$, V_1, V_2 are $N(0, \frac{1}{2})$, and U_1, U_2, V_1, V_2 are independent. The answer is therefore

$$\begin{aligned}p &= \mathbb{P}\big(\tfrac{3}{2}(N_1^2 + N_2^2) \le \tfrac{1}{2}(N_3^2 + N_4^2)\big) \quad \text{where the } N_i \text{ are independent } N(0,1) \\ &= \mathbb{P}(K_1 \le \tfrac{1}{3}K_2) \qquad \text{where the } K_i \text{ are independent chi-squared } \chi^2(2) \\ &= \mathbb{P}\left(\frac{K_1}{K_1+K_2} \le \frac{1}{4}\right) = \mathbb{P}(B \le \tfrac{1}{4}) = \tfrac{1}{4}\end{aligned}$$

where we have used the result of Exercise (4.7.14), and B is a beta-distributed random variable with parameters 1, 1.

43. The argument of Problem (4.14.42) leads to the expression

$$\begin{aligned}\mathbb{P}(U_1^2 + U_2^2 + U_3^2 \le V_1^2 + V_2^2 + V_3^2) &= \mathbb{P}(K_1 \le \tfrac{1}{3}K_2) \quad \text{where the } K_i \text{ are } \chi^2(3)\\ &= \mathbb{P}(B \le \tfrac{1}{4}) = \frac{1}{3} - \frac{\sqrt{3}}{4\pi},\end{aligned}$$

where B is beta-distributed with parameters $\frac{3}{2}, \frac{3}{2}$.

44. (a) Simply expand thus: $\mathbb{E}[(X-\mu)^3] = \mathbb{E}[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3]$ where $\mu = \mathbb{E}(X)$.
(b) $\mathrm{var}(S_n) = n\sigma^2$ and $\mathbb{E}[(S_n - n\mu)^3] = n\mathbb{E}[(X_1-\mu)^3]$ plus terms which equal zero because $\mathbb{E}(X_1 - \mu) = 0$.
(c) If Y is Bernoulli with parameter p, then $\mathrm{skw}(Y) = (1-2p)/\sqrt{pq}$, and the claim follows by (b).
(d) $m_1 = \lambda$, $m_2 = \lambda + \lambda^2$, $m_3 = \lambda^3 + 3\lambda^2 + \lambda$, and the claim follows by (a).
(e) Since λX is $\Gamma(1, t)$, we may as well assume that $\lambda = 1$. It is immediate that $\mathbb{E}(X^n) = \Gamma(t+n)/\Gamma(t)$, whence

$$\mathrm{skw}(X) = \frac{t(t+1)(t+2) - 3t \cdot t(t+1) + 2t^3}{t^{3/2}} = \frac{2}{\sqrt{t}}.$$

45. We find as above that $\mathrm{kur}(X) = (m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4)/\sigma^4$ where $m_k = \mathbb{E}(X^k)$.
(a) $m_4 = 3\sigma^4$ for the $N(0, \sigma^2)$ distribution, whence $\mathrm{kur}(X) = 3\sigma^4/\sigma^4$.
(b) $m_r = r!/\lambda^r$, and the result follows.
(c) In this case, $m_4 = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$, $m_3 = \lambda^3 + 3\lambda^2 + \lambda$, $m_2 = \lambda^2 + \lambda$, and $m_1 = \lambda$.
(d) $(\mathrm{var}\, S_n)^2 = n^2\sigma^4$ and $\mathbb{E}[(S_n - nm_1)^4] = n\mathbb{E}[(X_1 - m_1)^4] + 3n(n-1)\sigma^4$.

46. We have as $n \to \infty$ that

$$\mathbb{P}(X_{(n)} \le x + \log n) = \{1 - e^{-(x+\log n)}\}^n = \left(1 - \frac{e^{-x}}{n}\right)^n \to e^{-e^{-x}}.$$

By the lack-of-memory property,

$$\mathbb{E}(X_{(1)}) = \frac{1}{n}, \quad \mathbb{E}(X_{(2)}) = \frac{1}{n} + \frac{1}{n-1}, \quad \dots,$$

whence, by Lemma (4.3.4),

$$\int_0^\infty \{1 - e^{-e^{-x}}\}\,dx = \lim_{n\to\infty} \mathbb{E}(X_{(n)} - \log n) = \lim_{n\to\infty}\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1 - \log n\right) = \gamma.$$

47. By the argument presented in Section 4.11, conditional on acceptance, X has density function f_S. You might use this method when f_S is itself particularly tiresome or expensive to calculate. If $a(x)$ and $b(x)$ are easy to calculate and are close to f_S, much computational effort may be saved.

48. $M = \max\{U_1, U_2, \dots, U_Y\}$ satisfies

$$\mathbb{P}(M \le t) = \mathbb{E}(t^Y) = \frac{e^t - 1}{e - 1}.$$

Thus,

$$\begin{aligned}\mathbb{P}(Z \ge z) &= \mathbb{P}(X \ge \lfloor z \rfloor + 2) + \mathbb{P}\big(X = \lfloor z \rfloor + 1,\ Y \le \lfloor z \rfloor + 1 - z\big)\\ &= \frac{(e-1)e^{-\lfloor z \rfloor - 2}}{1-e^{-1}} + (e-1)e^{-\lfloor z \rfloor - 1} \cdot \frac{e^{\lfloor z \rfloor + 1 - z} - 1}{e-1} = e^{-z}.\end{aligned}$$

49. (a) Y has density function e^{-y} for $y > 0$, and X has density function $f_X(x) = \alpha e^{-\alpha x}$ for $x > 0$. Now $Y \ge \frac{1}{2}(X-\alpha)^2$ if and only if

$$V\alpha e^{-\alpha X} \frac{e^{\frac{1}{2}\alpha^2}}{\alpha}\sqrt{\frac{2}{\pi}} \le \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}X^2},$$

which is to say that $aVf_X(X) \le f(X)$, where $a = \alpha^{-1} e^{\frac{1}{2}\alpha^2}\sqrt{2/\pi}$. Recalling the argument of Example (4.11.5), we conclude that, conditional on this event occurring, X has density function f.

(b) The number of rejections is geometrically distributed with mean a^{-1}, so the optimal value of α is that which minimizes $\alpha e^{-\frac{1}{2}\alpha^2}\sqrt{\pi/2}$, that is, $\alpha = 1$.

(c) Setting

$$Z = \left\{\begin{array}{ll} +X & \text{with probability } \frac{1}{2} \\ -X & \text{with probability } \frac{1}{2} \end{array}\right\} \quad \text{conditional on } Y > \tfrac{1}{2}(X-\alpha)^2,$$

we obtain a random variable Z with the $N(0,1)$ distribution.

50. (a) $\mathbb{E}(X) = \displaystyle\int_0^1 \sqrt{1-u^2}\,du = \pi/4$.

(b) $\mathbb{E}(Y) = \displaystyle\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} \sin\theta\,d\theta = 2/\pi$.

51. You are asked to calculate the mean distance of a randomly chosen pebble from the nearest collection point. Running through the cases, where we suppose the circle has radius a and we write P for the position of the pebble,

(i)
$$\mathbb{E}|\mathrm{OP}| = \frac{1}{\pi a^2}\int_0^{2\pi}\int_0^a r^2\,dr\,d\theta = \frac{2a}{3}.$$

(ii)
$$\mathbb{E}|\mathrm{AP}| = \frac{2}{\pi a^2}\int_0^{\frac{1}{2}\pi}\int_0^{2a\cos\theta} r^2\,dr\,d\theta = \frac{32a}{9\pi}.$$

(iii)
$$\begin{aligned}\mathbb{E}(|\mathrm{AP}| \wedge |\mathrm{BP}|) &= \frac{4}{\pi a^2}\left[\int_0^{\frac{1}{4}\pi}\int_0^{a\sec\theta} r^2\,dr\,d\theta + \int_{\frac{1}{4}\pi}^{\frac{1}{2}\pi}\int_0^{2a\cos\theta} r^2\,dr\,d\theta\right]\\ &= \frac{4a}{3\pi}\left\{\frac{16}{3} - \frac{17}{6}\sqrt{2} + \frac{1}{2}\log(1+\sqrt{2})\right\} \simeq \frac{2a}{3} \times 1.13.\end{aligned}$$

(iv)
$$\mathbb{E}(|\mathrm{AP}| \wedge |\mathrm{BP}| \wedge |\mathrm{CP}|) = \frac{6}{\pi a^2}\left\{\int_0^{\frac{1}{3}\pi}\int_0^{x} r^2\,dr\,d\theta + \int_{\frac{1}{3}\pi}^{\frac{1}{2}\pi}\int_0^{2a\cos\theta} r^2\,dr\,d\theta\right\}$$

$$\text{where } x = a\sin(\tfrac{1}{3}\pi)\operatorname{cosec}(\tfrac{2}{3}\pi - \theta)$$

$$\begin{aligned}&= \frac{2a}{\pi}\left\{\int_0^{\frac{1}{3}\pi}\frac{1}{8}3\sqrt{3}\operatorname{cosec}^3\left(\frac{\pi}{3}+\theta\right)d\theta + \int_{\frac{1}{3}\pi}^{\frac{1}{2}\pi} 8\cos^3\theta\,d\theta\right\}\\ &= \frac{2a}{\pi}\left\{\frac{16}{3} - \frac{11}{4}\sqrt{3} + \frac{3\sqrt{3}}{16}\log\frac{3}{2}\right\} \simeq \frac{2a}{3} \times 0.67.\end{aligned}$$

52. By Problem (4.14.4), the displacement of R relative to P is the sum of two independent Cauchy random variables. By Exercise (4.8.2), this sum has also a Cauchy distribution, and inverting the transformation shows that Θ is uniformly distributed.

53. We may assume without loss of generality that R has length 1. Note that Δ occurs if and only if the sum of any two parts exceeds the length of the third part.

(a) If the breaks are at X, Y, where $0 < X < Y < 1$, then Δ occurs if and only if $2Y > 1$, and $2(Y - X) < 1$ and $2X < 1$. These inequalities are satisfied with probability $\frac{1}{4}$.

(b) The length X of the shorter piece has density function $f_X(x) = 2$ for $0 \le x \le \frac{1}{2}$. The other pieces are of length $(1 - X)Y$ and $(1 - X)(1 - Y)$, where Y is uniform on $(0, 1)$. The event Δ occurs if and only if $2Y < 2XY + 1$ and $X + Y - XY > \frac{1}{2}$, and this has probability

$$2\int_0^{\frac{1}{2}} \left\{ \frac{1}{2(1-x)} - \frac{1-2x}{2(1-x)} \right\} dx = \log(4/e).$$

(c) The three lengths are X, $\frac{1}{2}(1 - X)$, $\frac{1}{2}(1 - X)$, where X is uniform on $(0, 1)$. The event Δ occurs if and only if $X < \frac{1}{2}$.

(d) This triangle is obtuse if and only if

$$\frac{\frac{1}{2}X}{\frac{1}{2}(1-X)} > \frac{1}{\sqrt{2}},$$

which is to say that $X > \sqrt{2} - 1$. Hence,

$$\mathbb{P}(\text{obtuse} \mid \Delta) = \frac{\mathbb{P}(\sqrt{2} - 1 < X < \frac{1}{2})}{\mathbb{P}(X < \frac{1}{2})} = 3 - 2\sqrt{2}.$$

54. The shorter piece has density function $f_X(x) = 2$ for $0 \le x \le \frac{1}{2}$. Hence,

$$\mathbb{P}(R \le r) = \mathbb{P}\left(\frac{X}{1-X} \le r\right) = \frac{2r}{1+r},$$

with density function $f_R(r) = 2/(1 - r)^2$ for $0 \le r \le 1$. Therefore,

$$\mathbb{E}(R) = \int_0^1 \mathbb{P}(R > r)\,dr = \int_0^1 \frac{1-r}{1+r}\,dr = 2\log 2 - 1,$$
$$\mathbb{E}(R^2) = \int_0^1 2r\mathbb{P}(R > r)\,dr = \int_0^1 \frac{2r(1-r)}{1+r}dr = 3 - 4\log 2,$$

and $\mathrm{var}(R) = 2 - (2\log 2)^2$.

55. With an obvious notation,

$$\mathbb{E}(R^2) = \mathbb{E}[(X_1 - X_2)^2] + \mathbb{E}[(Y_1 - Y_2)^2] = 4\mathbb{E}(X_1^2) - 4\{\mathbb{E}(X_1)\}^2 = 4 \cdot \tfrac{1}{3}a^2 - 4(\tfrac{1}{2}a)^2 = \tfrac{1}{3}a^2.$$

By a natural re-scaling, we may assume that $a = 1$. Now, $X_1 - X_2$ and $Y_1 - Y_2$ have the same triangular density symmetric on $(-1, 1)$, whence $(X_1 - X_2)^2$ and $(Y_1 - Y_2)^2$ have distribution function $F(z) = 2\sqrt{z} - z$ and density function $f_Z(z) = z^{-\frac{1}{2}} - 1$, for $0 \le z \le 1$. Therefore R^2 has the density f given by

$$f(r) = \begin{cases} \displaystyle\int_0^r \left(\frac{1}{\sqrt{z}} - 1\right)\left(\frac{1}{\sqrt{r-z}} - 1\right) dz & \text{if } 0 \le r \le 1, \\ \displaystyle\int_{r-1}^1 \left(\frac{1}{\sqrt{z}} - 1\right)\left(\frac{1}{\sqrt{r-z}} - 1\right) dz & \text{if } 1 \le r \le 2. \end{cases}$$

The claim follows since

$$\int_a^b \frac{1}{\sqrt{z}} \frac{1}{\sqrt{r-z}} dz = 2\left(\sin^{-1}\sqrt{\frac{b}{r}} - \sin^{-1}\sqrt{\frac{a}{r}}\right) \quad \text{for } 0 \le a \le b \le 1.$$

56. We use an argument similar to that used for Buffon's needle. Dropping the paper at random amounts to dropping the lattice at random on the paper. The mean number of points of the lattice in a small element of area dA is dA. By the additivity of expectations, the mean number of points on the paper is A. There must therefore exist a position for the paper in which it covers at least $\lceil A \rceil$ points.

57. Consider a small element of surface dS. Positioning the rock at random amounts to shining light at this element from a randomly chosen direction. On averaging over all possible directions, we see that the mean area of the shadow cast by dS is proportional to the area of dS. We now integrate over the surface of the rock, and use the additivity of expectation, to deduce that the area A of the random shadow satisfies $\mathbb{E}(A) = cS$ for some constant c which is independent of the shape of the rock. By considering the special case of the sphere, we find $c = \frac{1}{4}$. It follows that at least one orientation of the rock gives a shadow of area at least $\frac{1}{4}S$.

58. (a) We have from Problem (4.14.11b) that $Y_r = X_r/(X_1 + \cdots + X_r)$ is independent of $X_1 + \cdots + X_r$, and therefore of the variables $X_{r+1}, X_{r+2}, \cdots, X_{k+1}, X_1 + \cdots + X_{k+1}$. Therefore Y_r is independent of $\{Y_{r+s} : s \ge 1\}$, and the claim follows.

(b) Let $S = X_1 + \cdots + X_{k+1}$. The inverse transformation $x_1 = z_1 s$, $x_2 = z_2 s, \ldots, x_k = z_k s$, $x_{k+1} = s - z_1 s - z_2 s - \cdots - z_k s$ has Jacobian

$$J = \begin{vmatrix} s & 0 & 0 & \cdots & z_1 \\ 0 & s & 0 & \cdots & z_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & z_k \\ -s & -s & -s & \cdots & 1 - z_1 - \cdots - z_k \end{vmatrix} = s^k.$$

The joint density function of $X_1, X_2, \ldots, X_k, S$ is therefore (with $\sigma = \sum_{r=1}^{k+1} \beta_r$),

$$\left\{\prod_{r=1}^k \frac{\lambda^{\beta_r}(z_r s)^{\beta_r - 1} e^{-\lambda z_r s}}{\Gamma(\beta_r)}\right\} \cdot \frac{\lambda^{\beta_{k+1}}\{s(1 - z_1 - \cdots - z_k)\}^{\beta_{k+1}-1} e^{-\lambda s(1-z_1-\cdots-z_k)}}{\Gamma(\beta_{k+1})}$$
$$= f(\lambda, \boldsymbol{\beta}, s)\left(\prod_{r=1}^k z_r^{\beta_r - 1}\right)(1 - z_1 - \cdots - z_k)^{\beta_{k+1}-1},$$

where f is a function of the given variables. The result follows by integrating over s.

59. Let $\mathbf{C} = (c_{rs})$ be an orthogonal $n \times n$ matrix with $c_{ni} = 1/\sqrt{n}$ for $1 \le i \le n$. Let $Y_{ir} = \sum_{s=1}^n X_{is} c_{rs}$, and note that the vectors $\mathbf{Y}_r = (Y_{1r}, Y_{2r}, \ldots, Y_{nr})$, $1 \le r \le n$, are multivariate normal. Clearly $\mathbb{E}Y_{ir} = 0$, and

$$\mathbb{E}(Y_{ir}Y_{js}) = \sum_{t,u} c_{rt}c_{su}\mathbb{E}(X_{it}X_{ju}) = \sum_{t,u} c_{rt}c_{su}\delta_{tu}v_{ij} = \sum_t c_{rt}c_{st}v_{ij} = \delta_{rs}v_{ij},$$

where δ_{tu} is the Kronecker delta, since $\mathbf{C}$ is orthogonal. It follows that the set of vectors $\mathbf{Y}_r$ has the same joint distribution as the set of $\mathbf{X}_r$. Since $\mathbf{C}$ is orthogonal, $X_{ir} = \sum_{s=1}^{n} c_{sr} Y_{is}$, and therefore

$$S_{ij} = \sum_{r,s,t} c_{sr} c_{tr} Y_{is} Y_{jt} - \frac{1}{n} \sum_r X_{ir} \sum_r X_{jr} = \sum_{s,t} \delta_{st} Y_{is} Y_{jt} - \frac{1}{\sqrt{n}} \sum_r X_{ir} \frac{1}{\sqrt{n}} \sum_r X_{jr}$$

$$= \sum_s Y_{is} Y_{js} - Y_{in} Y_{jn} = \sum_{s=1}^{n-1} Y_{is} Y_{js}.$$

This has the same distribution as T_{ij} because the $\mathbf{Y}_r$ and the $\mathbf{X}_r$ are identically distributed.

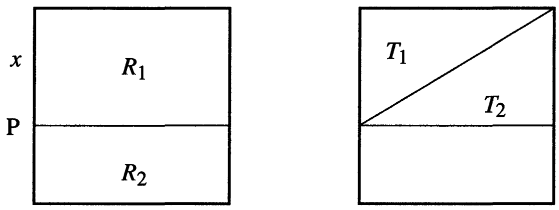
60. We sketch this. Let $\mathbb{E}|\mathrm{PQR}| = m(a)$, and use Crofton's method. A point randomly dropped in $S(a + da)$ lies in $S(a)$ with probability

$$\left(\frac{a}{a + da} \right)^2 = 1 - \frac{2da}{a} + \mathrm{o}(da).$$

Hence

$$\frac{dm}{da} = -\frac{6m}{a} + \frac{6m_b}{a},$$

where $m_b(a)$ is the conditional mean of $|\mathrm{PQR}|$ given that P is constrained to lie on the boundary of $S(a)$. Let $b(x)$ be the conditional mean of $|\mathrm{PQR}|$ given that P lies a distance x down one vertical edge.



By conditioning on whether Q and R lie above or beneath P we find, in an obvious notation, that

$$b(x) = \left(\frac{x}{a} \right)^2 m_{R_1} + \left(\frac{a - x}{a} \right)^2 m_{R_2} + \frac{2x(a - x)}{a^2} m_{R_1,R_2}.$$

By Exercise (4.13.6) (see also Exercise (4.13.7)), $m_{R_1,R_2} = \frac{1}{2}(\frac{1}{2}a)(\frac{1}{2}a) = \frac{1}{8}a^2$. In order to find m_{R_1}, we condition on whether Q and R lie in the triangles T_1 or T_2, and use an obvious notation.

Recalling Example (4.13.6), we have that $m_{T_1} = m_{T_2} = \frac{4}{27} \cdot \frac{1}{2}ax$. Next, arguing as we did in that example,

$$m_{T_1,T_2} = \tfrac{1}{2} \cdot \tfrac{4}{9} \{ ax - \tfrac{1}{4}ax - \tfrac{1}{4}ax - \tfrac{1}{8}ax \}.$$

Hence, by conditional expectation,

$$m_{R_1} = \tfrac{1}{4} \cdot \tfrac{4}{27} \cdot \tfrac{1}{2}ax + \tfrac{1}{4} \cdot \tfrac{4}{27} \cdot \tfrac{1}{2}ax + \tfrac{1}{2} \cdot \tfrac{4}{9} \cdot \tfrac{3}{8}ax = \tfrac{13}{108}ax.$$

We replace x by $a - x$ to find m_{R_2}, whence in total

$$b(x) = \left(\frac{x}{a} \right)^2 \frac{13ax}{108} + \left(\frac{a - x}{a} \right)^2 \frac{13a(a - x)}{108} + \frac{2x(a - x)}{a^2} \cdot \frac{a^2}{8} = \frac{13}{108}a^2 - \frac{12ax}{108} + \frac{12x^2}{108}.$$

Since the height of P is uniformly distributed on $[0, a]$, we have that

$$m_b(a) = \frac{1}{a}\int_0^a b(x)\,dx = \frac{11a^2}{108}.$$

We substitute this into the differential equation to obtain the solution $m(a) = \frac{11}{144}a^2$.

Turning to the last part, by making an affine transformation, we may without loss of generality take the parallelogram to be a square. The points form a convex quadrilateral when no point lies inside the triangle formed by the other three, and the required probability is therefore $1 - 4m(a)/a^2 = 1 - \frac{44}{144} = \frac{25}{36}$.

61. Choose four points P, Q, R, S uniformly at random inside C, and let T be the event that their convex hull is a triangle. By considering which of the four points lies in the interior of the convex hull of the other three, we see that $\mathbb{P}(T) = 4\mathbb{P}(\text{S} \in \text{PQR}) = 4\mathbb{E}|\text{PQR}|/|C|$. Having chosen P, Q, R, the four points form a triangle if and only if S lies in either the triangle PQR or the shaded region A. Thus, $\mathbb{P}(T) = \{|A| + \mathbb{E}|\text{PQR}|\}/|C|$, and the claim follows on solving for $\mathbb{P}(T)$.

62. Since $\mathbf{X}$ has zero means and covariance matrix $\mathbf{I}$, we have that $\mathbb{E}(\mathbf{Z}) = \boldsymbol{\mu} + \mathbb{E}(\mathbf{X})\mathbf{L} = \boldsymbol{\mu}$, and the covariance matrix of $\mathbf{Z}$ is $\mathbb{E}(\mathbf{L}'\mathbf{X}'\mathbf{X}\mathbf{L}) = \mathbf{L}'\mathbf{I}\mathbf{L} = \mathbf{V}$.

63. Let $\mathbf{D} = (d_{ij}) = \mathbf{AB} - \mathbf{C}$. The claim is trivial if $\mathbf{D} = \mathbf{0}$, and so we assume the converse. Choose i, k such that $d_{ik} \neq 0$, and write $y_i = \sum_{j=1}^n d_{ij}x_j = S + d_{ik}x_k$. Now $\mathbb{P}(y_i = 0) = \mathbb{E}\big(\mathbb{P}(x_k = -S/d_{ik} \mid S)\big)$. For any given S, there is probability at least $\frac{1}{2}$ that $x_k \neq -S/d_{ik}$, and the second claim follows.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ be independent random vectors with the given distribution. If $\mathbf{D} \neq \mathbf{0}$, the probability that $\mathbf{D}\mathbf{x}_s = \mathbf{0}$ for $1 \le s \le m$ is at most $(\frac{1}{2})^m$, which may be made as small as required by choosing m sufficiently large.

5

Generating functions and their applications

5.1 Solutions. Generating functions

1. (a) If $|s| < (1-p)^{-1}$,

$$G(s) = \sum_{m=0}^{\infty} s^m \binom{n+m-1}{m} p^n (1-p)^m = \left\{ \frac{p}{1-s(1-p)} \right\}^n.$$

Therefore the mean is $G'(1) = n(1-p)/p$. The variance is $G''(1)+G'(1)-G'(1)^2 = n(1-p)/p^2$.
(b) If $|s| < 1$,

$$G(s) = \sum_{m=1}^{\infty} s^m \left(\frac{1}{m} - \frac{1}{m+1} \right) = 1 + \left(\frac{1-s}{s} \right) \log(1-s).$$

Therefore $G'(1) = \infty$, and there exist no moments of order 1 or greater.
(c) If $p < |s| < p^{-1}$,

$$G(s) = \sum_{m=-\infty}^{\infty} s^m \left(\frac{1-p}{1+p} \right) p^{|m|} = \frac{1-p}{1+p} \left\{ 1 + \frac{sp}{1-sp} + \frac{p/s}{1-(p/s)} \right\}.$$

The mean is $G'(1) = 0$, and the variance is $G''(1) = 2p(1-p)^{-2}$.

2. (i) Either hack it out, or use indicator functions I_A thus:

$$T(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X > n) = \mathbb{E}\left(\sum_{n=0}^{\infty} s^n I_{\{n<X\}} \right) = \mathbb{E}\left(\sum_{n=0}^{X-1} s^n \right) = \mathbb{E}\left(\frac{1-s^X}{1-s} \right) = \frac{1-G(s)}{1-s}.$$

(ii) It follows that

$$T(1) = \lim_{s\uparrow 1} \left\{ \frac{1-G(s)}{1-s} \right\} = \lim_{s\uparrow 1} \frac{G'(s)}{1} = G'(1) = \mathbb{E}(X)$$

by L'Hôpital's rule. Also,

$$\begin{aligned} T'(1) &= \lim_{s\uparrow 1} \left\{ \frac{-(1-s)G'(s) + 1 - G(s)}{(1-s)^2} \right\} \\ &= \tfrac{1}{2} G''(1) = \tfrac{1}{2} \{ \mathrm{var}(X) - G'(1) + G'(1)^2 \} \end{aligned}$$

whence the claim is immediate.

3. (i) We have that $G_{X,Y}(s,t) = \mathbb{E}(s^X t^Y)$, whence $G_{X,Y}(s,1) = G_X(s)$ and $G_{X,Y}(1,t) = G_Y(t)$.
(ii) If $\mathbb{E}|XY| < \infty$ then

$$\mathbb{E}(XY) = \mathbb{E}\left(XYs^{X-1}t^{Y-1}\right)\Big|_{s=t=1} = \frac{\partial^2}{\partial s\,\partial t} G_{X,Y}(s,t)\bigg|_{s=t=1}.$$

4. We write $G(s,t)$ for the joint generating function.

(a)
$$\begin{aligned} G(s,t) &= \sum_{j=0}^{\infty}\sum_{k=0}^{j} s^j t^k (1-\alpha)(\beta-\alpha)\alpha^j \beta^{k-j-1} \\ &= \sum_{j=0}^{\infty}\left(\frac{\alpha s}{\beta}\right)^j \frac{(1-\alpha)(\beta-\alpha)}{\beta}\cdot\frac{1-(\beta t)^{j+1}}{1-\beta t} && \text{if } \beta|t| < 1 \\ &= \frac{(1-\alpha)(\beta-\alpha)}{(1-\beta t)\beta}\left\{\frac{1}{1-(\alpha s/\beta)} - \frac{\beta t}{1-\alpha s t}\right\} && \text{if } \frac{\alpha}{\beta}|s| < 1 \\ &= \frac{(1-\alpha)(\beta-\alpha)}{(1-\alpha s t)(\beta-\alpha s)} \end{aligned}$$

(the condition $\alpha|st| < 1$ is implied by the other two conditions on s and t). The marginal generating functions are

$$G(s,1) = \frac{(1-\alpha)(\beta-\alpha)}{(1-\alpha s)(\beta-\alpha s)}, \quad G(1,t) = \frac{1-\alpha}{1-\alpha t},$$

and the covariance is easily calculated by the conclusion of Exercise (5.1.3) as $\alpha(1-\alpha)^{-2}$.
(b) Arguing similarly, we obtain $G(s,t) = (e-1)/\{e(1-te^{s-2})\}$ if $|t|e^{s-2} < 1$, with marginals

$$G(s,1) = \frac{1-e^{-1}}{1-e^{s-2}}, \quad G(1,t) = \frac{1-e^{-1}}{1-te^{-1}},$$

and covariance $e(e-1)^{-2}$.
(c) Once again,

$$G(s,t) = \frac{\log\{1-tp(1-p+sp)\}}{\log(1-p)} \qquad \text{if } |tp(1-p+sp)| < 1.$$

The marginal generating functions are

$$G(s,1) = \frac{\log\{1-p+p^2(1-s)\}}{\log(1-p)}, \quad G(1,t) = \frac{\log(1-tp)}{\log(1-p)},$$

and the covariance is

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

5. (i) We have that

$$\mathbb{E}(x^H y^T) = \sum_{k=0}^{n} x^k y^{n-k}\binom{n}{k} p^k (1-p)^{n-k} = (px+qy)^n$$

where $p + q = 1$.
(ii) More generally, if each toss results in one of t possible outcomes, the ith of which has probability p_i, then the corresponding quantity is a function of t variables, $x_1, x_2, \dots, x_t$, and is found to be $(p_1x_1 + p_2x_2 + \cdots + p_tx_t)^n$.

6. We have that

$$\mathbb{E}(s^X) = \mathbb{E}\{\mathbb{E}(s^X \mid U)\} = \int_0^1 \{1 + u(s-1)\}^n \, du = \frac{1}{n+1} \cdot \frac{1 - s^{n+1}}{1 - s},$$

the probability generating function of the uniform distribution. See also Exercise (4.6.5).

7. We have that

$$\begin{aligned} G_{X,Y,Z}(x, y, z) = G(x, y, z, 1) &= \tfrac{1}{8}(xyz + xy + yz + zx + x + y + z + 1) \\ &= \tfrac{1}{2}(x+1)\tfrac{1}{2}(y+1)\tfrac{1}{2}(z+1) = G_X(x)G_Y(y)G_Z(z), \end{aligned}$$

whence X, Y, Z are independent. The same conclusion holds for any other set of exactly three random variables. However, $G(x, y, z, w) \neq G_X(x)G_Y(y)G_Z(z)G_W(w)$.

8. (a) We have by differentiating that $\mathbb{E}(X^{2n}) = 0$, whence $\mathbb{P}(X = 0) = 1$. This is not a moment generating function.
(b) This is a moment generating function if and only if $\sum_r p_r = 1$, in which case it is that of a random variable X with $\mathbb{P}(X = a_r) = p_r$.

9. The coefficients of s^n in both combinations of G_1, G_2 are non-negative and sum to 1. They are therefore probability generating functions, as is $G(\alpha s)/G(\alpha)$ for the same reasons.

5.2 Solutions. Some applications

1. Let $G(s) = \mathbb{E}(s^X)$ and $G_S(s) = \sum_{j=0}^n s^j S_j$. By the result of Exercise (5.1.2),

$$T(s) = \sum_{m=0}^{\infty} s^m \mathbb{P}(X \ge m) = 1 + s\sum_{k=0}^{\infty} s^k \mathbb{P}(X > k) = 1 + \frac{s(1 - G(s))}{1 - s} = \frac{1 - sG(s)}{1 - s}.$$

Now,

$$G_S(s) = \sum_{m=0}^{n} s^m \mathbb{E}\binom{X}{m} = \mathbb{E}\left\{\sum_{m=0}^{n} s^m \binom{X}{m}\right\} = \mathbb{E}\{(1+s)^X\} = G(1+s)$$

so that

$$\frac{T(s) - T(0)}{s} = \frac{G_S(s-1) - G_S(0)}{s - 1}$$

where we have used the fact that $T(0) = G_S(0) = 1$. Therefore

$$\sum_{i=1}^{n} s^{i-1}\mathbb{P}(X \ge i) = \sum_{j=1}^{n} (s-1)^{j-1} S_j.$$

Equating coefficients of s^{i-1}, we obtain as required that

$$\mathbb{P}(X \ge i) = \sum_{j=i}^{n} S_j \binom{j-1}{i-1}(-1)^{j-i}, \qquad 1 \le i \le n.$$

Similarly,

$$\frac{G_S(s) - G_S(0)}{s} = \frac{T(1+s) - T(0)}{1+s}$$

whence the second formula follows.

2. Let A_i be the event that the ith person is chosen by nobody, and let X be the number of events $A_1, A_2, \dots, A_n$ which occur. Clearly

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j}) = \left(\frac{n-j}{n-1}\right)^j \left(\frac{n-j-1}{n-1}\right)^{n-j}$$

if $i_1 \neq i_2 \neq \cdots \neq i_j$, since this event requires each of $i_1, \dots, i_j$ to choose from a set of $n - j$ people, and each of the others to choose from a set of size $n - j - 1$. Using Waring's Theorem (Problem (1.8.13) or equation (5.2.14)),

$$\mathbb{P}(X = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} S_j$$

where

$$S_j = \binom{n}{j} \left(\frac{n-j}{n-1}\right)^j \left(\frac{n-j-1}{n-1}\right)^{n-j}.$$

Using the result of Exercise (5.2.1),

$$\mathbb{P}(X \geq k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j-1}{k-1} S_j, \qquad 1 \leq k \leq n,$$

while $\mathbb{P}(X \geq 0) = 1$.

3. (a)

$$\mathbb{E}(x^{X+Y}) = \mathbb{E}\{\mathbb{E}(x^{X+Y} \mid Y)\} = \mathbb{E}\{x^Y e^{Y(x-1)}\} = \mathbb{E}\{(xe^{x-1})^Y\} = \exp\{\mu(xe^{x-1} - 1)\}.$$

(b) The probability generating function of X_1 is

$$G(s) = \sum_{k=1}^{\infty} \frac{\{s(1-p)\}^k}{k \log(1/p)} = \frac{\log\{1 - s(1-p)\}}{\log p}.$$

Using the 'compounding theorem' (5.1.25),

$$G_Y(s) = G_N(G(s)) = e^{\mu(G(s)-1)} = \left(\frac{p}{1 - s(1-p)}\right)^{-\mu/\log p}.$$

4. Clearly,

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \mathbb{E}\left(\int_0^1 t^X \, dt\right) = \int_0^1 \mathbb{E}(t^X) \, dt = \int_0^1 (q + pt)^n \, dt = \frac{1 - q^{n+1}}{p(n+1)}$$

where $q = 1 - p$. In the limit,

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \frac{1 - (1 - \lambda/n)^{n+1}}{\lambda(n+1)/n} + o(1) \to \frac{1 - e^{-\lambda}}{\lambda},$$

the corresponding moment of the Poisson distribution with parameter λ.

5. Conditioning on the outcome of the first toss, we obtain $h_n = qh_{n-1} + p(1 - h_{n-1})$ for $n \geq 1$, where $q = 1 - p$ and $h_0 = 1$. Multiply throughout by s^n and sum to find that $H(s) = \sum_{n=0}^{\infty} s^n h_n$ satisfies $H(s) - 1 = (q - p)sH(s) + ps/(1 - s)$, and so

$$H(s) = \frac{1 - qs}{(1 - s)\{1 - (q - p)s\}} = \frac{1}{2}\left\{\frac{1}{1 - s} + \frac{1}{1 - (q - p)s}\right\}.$$

6. By considering the event that HTH does not appear in n tosses and then appears in the next three, we find that $\mathbb{P}(X > n)p^2q = \mathbb{P}(X = n + 1)pq + \mathbb{P}(X = n + 3)$. We multiply by s^{n+3} and sum over n to obtain

$$\frac{1 - \mathbb{E}(s^X)}{1 - s} p^2qs^3 = pqs^2\mathbb{E}(s^X) + \mathbb{E}(s^X),$$

which may be solved as required. Let Z be the time at which THT first appears, so $Y = \min\{X, Z\}$. By a similar argument,

$$\mathbb{P}(Y > n)p^2q = \mathbb{P}(X = Y = n + 1)pq + \mathbb{P}(X = Y = n + 3) + \mathbb{P}(Z = Y = n + 2)p,$$
$$\mathbb{P}(Y > n)q^2p = \mathbb{P}(Z = Y = n + 1)pq + \mathbb{P}(Z = Y = n + 3) + \mathbb{P}(X = Y = n + 2)q.$$

We multipy by s^{n+1}, sum over n, and use the fact that $\mathbb{P}(Y = n) = \mathbb{P}(X = Y = n) + \mathbb{P}(Z = Y = n)$.

7. Suppose there are $n + 1$ matching letter/envelope pairs, numbered accordingly. Imagine the envelopes lined up in order, and the letters dropped at random onto these envelopes. Assume that exactly $j + 1$ letters land on their correct envelopes. The removal of any one of these $j + 1$ letters, together with the corresponding envelope, results after re-ordering in a sequence of length n in which exactly j letters are correctly placed. It is not difficult to see that, for each resulting sequence of length n, there are exactly $j + 1$ originating sequences of length $n + 1$. The first result follows. We multiply by s^j and sum over j to obtain the second. It is evident that $G_1(s) = s$. Either use induction, or integrate repeatedly, to find that $G_n(s) = \sum_{r=0}^{n}(s - 1)^r/r!$.

8. We have for $|s| < \mu + 1$ that

$$\mathbb{E}(s^X) = \mathbb{E}\{\mathbb{E}(s^X \mid \Lambda)\} = \mathbb{E}(e^{\Lambda(s-1)}) = \frac{\mu}{\mu - (s - 1)} = \frac{\mu}{\mu + 1}\sum_{k=0}^{\infty}\left(\frac{s}{\mu + 1}\right)^k.$$

9. Since the waiting times for new objects are geometric and independent,

$$\mathbb{E}(s^T) = s\left(\frac{3s}{4 - s}\right)\left(\frac{s}{2 - s}\right)\left(\frac{s}{4 - 3s}\right).$$

Using partial fractions, the coefficient of s^k is $\frac{3}{32}\{\frac{1}{2}(\frac{1}{4})^{k-4} - 4(\frac{1}{2})^{k-4} + \frac{9}{2}(\frac{3}{4})^{k-4}\}$, for $k \geq 4$.

5.3 Solutions. Random walk

1. Let A_k be the event that the walk ever reaches the point k. Then $A_k \supseteq A_{k+1}$ if $k \geq 0$, so that

$$\mathbb{P}(M \geq r) = \mathbb{P}(A_r) = \mathbb{P}(A_0)\prod_{k=0}^{r-1}\mathbb{P}(A_{k+1} \mid A_k) = (p/q)^r, \qquad r \geq 0,$$

since $\mathbb{P}(A_{k+1} \mid A_k) = \mathbb{P}(A_1 \mid A_0) = p/q$ for $k \geq 0$ by Corollary (5.3.6).

2. (a) We have by Theorem (5.3.1c) that

$$\sum_{k=1}^{\infty} s^{2k} 2k f_0(2k) = sF_0'(s) = \frac{s^2}{\sqrt{1-s^2}} = s^2 \mathbb{P}_0(s) = \sum_{k=1}^{\infty} s^{2k} p_0(2k-2),$$

and the claim follows by equating the coefficients of s^{2k}.
(b) It is the case that $\alpha_n = \mathbb{P}(S_1 S_2 \cdots S_{2n} \neq 0)$ satisfies

$$\alpha_n = \sum_{\substack{k=2n+2 \\ k \text{ even}}}^{\infty} f_0(k),$$

with the convention that $\alpha_0 = 1$. We have used the fact that ultimate return to 0 occurs with probability 1. This sequence has generating function given by

$$\begin{aligned}\sum_{n=0}^{\infty} s^{2n} \sum_{\substack{k=2n+2 \\ k \text{ even}}}^{\infty} f_0(k) &= \sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} f_0(k) \sum_{n=0}^{\frac{1}{2}k-1} s^{2n} \\ &= \frac{1 - F_0(s)}{1 - s^2} = \frac{1}{\sqrt{1-s^2}} \quad \text{by Theorem (5.3.1c)} \\ &= \mathbb{P}_0(s) = \sum_{n=0}^{\infty} s^{2n} \mathbb{P}(S_{2n} = 0).\end{aligned}$$

Now equate the coefficients of s^{2n}. (Alternatively, use Exercise (5.1.2) to obtain the generating function of the α_n directly.)

3. Draw a diagram of the square with the letters ABCD in clockwise order. Clearly $p_{\mathrm{AA}}(m) = 0$ if m is odd. The walk is at A after $2n$ steps if and only if the numbers of leftward and rightward steps are equal *and* the numbers of upward and downward steps are equal. The number of ways of choosing $2k$ horizontal steps out of $2n$ is $\binom{2n}{2k}$. Hence

$$p_{\mathrm{AA}}(2n) = \sum_{k=0}^{n} \binom{2n}{2k} \alpha^{2k} \beta^{2n-2k} = \tfrac{1}{2}\{(\alpha+\beta)^{2n} + (\alpha-\beta)^{2n}\} = \tfrac{1}{2}\{1 + (\alpha-\beta)^{2n}\}$$

with generating function

$$G_{\mathrm{A}}(s) = \sum_{n=0}^{\infty} s^{2n} p_{\mathrm{AA}}(2n) = \frac{1}{2}\left\{\frac{1}{1-s^2} + \frac{1}{1 - \{s(\alpha-\beta)\}^2}\right\}.$$

Writing $F_{\mathrm{A}}(s)$ for the probability generating function of the time T of first return, we use the argument which leads to Theorem (5.3.1a) to find that $G_{\mathrm{A}}(s) = 1 + F_{\mathrm{A}}(s)G_{\mathrm{A}}(s)$, and therefore $F_{\mathrm{A}}(s) = 1 - G_{\mathrm{A}}(s)^{-1}$.

4. Write (X_n, Y_n) for the position of the particle at time n. It is an elementary calculation to show that the relations $U_n = X_n + Y_n$, $V_n = X_n - Y_n$ define independent simple symmetric random walks U and V. Now $T = \min\{n : U_n = m\}$, and therefore $G_T(s) = \{s^{-1}(1 - \sqrt{1-s^2})\}^m$ for $|s| \leq 1$ by Theorem (5.3.5).

Now $X - Y = V_T$, so that

$$G_{X-Y}(s) = \mathbb{E}\{\mathbb{E}(s^{V_T} \mid T)\} = \mathbb{E}\left\{\left(\frac{s+s^{-1}}{2}\right)^T\right\} = G_T\left(\tfrac{1}{2}(s+s^{-1})\right)$$

where we have used the independence of U and V. This converges if and only if $|\frac{1}{2}(s+s^{-1})| \le 1$, which is to say that $s = \pm 1$. Note that $G_T(s)$ converges in a non-trivial region of the complex plane.

5. Let T be the time of the first return of the walk S to its starting point 0. During the time-interval $(0, T)$, the walk is equally likely to be to the left or to the right of 0, and therefore

$$L_{2n} = \begin{cases} TR + L' & \text{if } T \le 2n, \\ 2nR & \text{if } T > 2n, \end{cases}$$

where R is Bernoulli with parameter $\frac{1}{2}$, L' has the distribution of L_{2n-T}, and R and L' are independent. It follows that $G_{2n}(s) = \mathbb{E}(s^{L_{2n}})$ satisfies

$$G_{2n}(s) = \sum_{k=1}^{n} \tfrac{1}{2}(1+s^{2k})G_{2n-2k}(s)f(2k) + \sum_{k>n} \tfrac{1}{2}(1+s^{2n})f(2k)$$

where $f(2k) = \mathbb{P}(T = 2k)$. (Remember that L_{2n} and T are even numbers.) Let $H(s,t) = \sum_{n=0}^{\infty} t^{2n}G_{2n}(s)$. Multiply through the above equation by t^{2n} and sum over n, to find that

$$H(s,t) = \tfrac{1}{2}H(s,t)\{F(t) + F(st)\} + \tfrac{1}{2}\{J(t) + J(st)\}$$

where $F(x) = \sum_{k=0}^{\infty} x^{2k} f(2k)$ and

$$J(x) = \sum_{n=0}^{\infty} x^{2n} \sum_{k>n} f(2k) = \frac{1}{\sqrt{1-x^2}}, \qquad |x| < 1,$$

by the calculation in the solution to Exercise (5.3.2). Using the fact that $F(x) = 1 - \sqrt{1-x^2}$, we deduce that $H(s,t) = 1/\sqrt{(1-t^2)(1-s^2t^2)}$. The coefficient of $s^{2k}t^{2n}$ is

$$\begin{aligned} \mathbb{P}(L_{2n} = 2k) &= \binom{-\frac{1}{2}}{n-k}(-1)^{n-k} \cdot \binom{-\frac{1}{2}}{k}(-1)^k \\ &= \binom{2k}{k}\binom{2n-2k}{n-k}\left(\frac{1}{2}\right)^{2n} = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0). \end{aligned}$$

6. We show that all three terms have the same generating function, using various results established for simple symmetric random walk. First, in the usual notation,

$$\sum_{m=0}^{\infty} 4m\mathbb{P}(S_{2m} = 0)s^{2m} = 2sP_0'(s) = \frac{2s^2}{(1-s^2)^{3/2}}.$$

Secondly, by Exercise (5.3.2a, b),

$$\mathbb{E}(T \wedge 2m) = 2m\mathbb{P}(T > 2m) + \sum_{k=1}^{m} 2kf_0(2k) = 2m\mathbb{P}(S_{2m} = 0) + \sum_{k=1}^{m} \mathbb{P}(S_{2k-2} = 0).$$

Hence,

$$\sum_{m=0}^{\infty} s^{2m}\mathbb{E}(T \wedge 2m) = \frac{s^2 P_0(s)}{1-s^2} + s P_0'(s) = \frac{2s^2}{(1-s^2)^{3/2}}.$$

Finally, using the hitting time theorem (3.10.14), (5.3.7), and some algebra at the last stage,

$$\begin{aligned}\sum_{m=0}^{\infty} s^{2m} 2\mathbb{E}|S_{2m}| &= 4\sum_{m=0}^{\infty} s^{2m}\sum_{k=1}^{m} 2k\mathbb{P}(S_{2m}=2k) = 4\sum_{m=0}^{\infty} s^{2m}\sum_{k=1}^{m} 2mf_{2k}(2m)\\ &= 4s\frac{d}{ds}\sum_{m=0}^{\infty} s^{2m}\sum_{k=1}^{m} f_{2k}(2m) = 4s\frac{d}{ds}\frac{F_1(s)^2}{1-F_1(s)^2} = \frac{2s^2}{(1-s^2)^{3/2}}.\end{aligned}$$

7. Let I_n be the indicator of the event $\{S_n = 0\}$, so that $S_{n+1} = S_n + X_{n+1} + I_n$. In equilibrium, $\mathbb{E}(S_0) = \mathbb{E}(S_0) + \mathbb{E}(X_1) + \mathbb{E}(I_0)$, which implies that $\mathbb{P}(S_0 = 0) = \mathbb{E}(I_0) = -\mathbb{E}(X_1)$ and entails $\mathbb{E}(X_1) \le 0$. Furthermore, it is impossible that $\mathbb{P}(S_0 = 0) = 0$ since this entails $\mathbb{P}(S_0 = a) = 0$ for all $a < \infty$. Hence $\mathbb{E}(X_1) < 0$ if S is in equilibrium. Next, in equilibrium,

$$\mathbb{E}(z^{S_0}) = \mathbb{E}(z^{S_{n+1}}) = \mathbb{E}\big(z^{S_n+X_{n+1}+I_n}\{(1-I_n)+I_n\}\big).$$

Now,

$$\begin{aligned}\mathbb{E}\{z^{S_n+X_{n+1}+I_n}(1-I_n)\} &= \mathbb{E}(z^{S_n} \mid S_n > 0)\mathbb{E}(z^{X_1})\mathbb{P}(S_n > 0)\\ \mathbb{E}(z^{S_n+X_{n+1}+I_n} I_n) &= z\mathbb{E}(z^{X_1})\mathbb{P}(S_n = 0).\end{aligned}$$

Hence

$$\mathbb{E}(z^{S_0}) = \mathbb{E}(z^{X_1})\big[\{\mathbb{E}(z^{S_0}) - \mathbb{P}(S_0=0)\} + z\mathbb{P}(S_0=0)\big]$$

which yields the appropriate choice for $\mathbb{E}(z^{S_0})$.

8. The hitting time theorem (3.10.14), (5.3.7), states that $\mathbb{P}(T_{0b} = n) = (|b|/n)\mathbb{P}(S_n = b)$, whence

$$\mathbb{E}(T_{0b} \mid T_{0b} < \infty) = \frac{b}{\mathbb{P}(T_{0b} < \infty)}\sum_n \mathbb{P}(S_n = b).$$

The walk is transient if and only if $p \ne \frac{1}{2}$, and therefore $\mathbb{E}(T_{0b} \mid T_{0b} < \infty) < \infty$ if and only if $p \ne \frac{1}{2}$. Suppose henceforth that $p \ne \frac{1}{2}$.

The required conditional mean may be found by conditioning on the first step, or alternatively as follows. Assume first that $p < q$, so that $\mathbb{P}(T_{0b} < \infty) = (p/q)^b$ by Corollary (5.3.6). Then $\sum_n \mathbb{P}(S_n = b)$ is the mean of the number N of visits of the walk to b. Now

$$\mathbb{P}(N = r) = \left(\frac{p}{q}\right)^b \rho^{r-1}(1-\rho), \qquad r \ge 1,$$

where $\rho = \mathbb{P}(S_n = 0 \text{ for some } n \ge 1) = 1 - |p-q|$. Therefore $\mathbb{E}(N) = (p/q)^b/|p-q|$ and

$$\mathbb{E}(T_{0b} \mid T_{0b} < \infty) = \frac{b}{(p/q)^b}\cdot\frac{(p/q)^b}{|p-q|}.$$

We have when $p > q$ that $\mathbb{P}(T_{0b} < \infty) = 1$, and $\mathbb{E}(T_{01}) = (p-q)^{-1}$. The result follows from the fact that $\mathbb{E}(T_{0b}) = b\mathbb{E}(T_{01})$.

5.4 Solutions. Branching processes

1. Clearly $\mathbb{E}(Z_n \mid Z_m) = Z_m \mu^{n-m}$ since, given Z_m, Z_n is the sum of the numbers of $(n-m)$th generation descendants of Z_m progenitors. Hence $\mathbb{E}(Z_m Z_n \mid Z_m) = Z_m^2 \mu^{n-m}$ and $\mathbb{E}(Z_m Z_n) = \mathbb{E}\{\mathbb{E}(Z_m Z_n \mid Z_m)\} = \mathbb{E}(Z_m^2)\mu^{n-m}$. Hence

$$\operatorname{cov}(Z_m, Z_n) = \mu^{n-m}\mathbb{E}(Z_m^2) - \mathbb{E}(Z_m)\mathbb{E}(Z_n) = \mu^{n-m} \operatorname{var}(Z_m),$$

and, by Lemma (5.4.2),

$$\rho(Z_m, Z_n) = \mu^{n-m}\sqrt{\frac{\operatorname{var} Z_m}{\operatorname{var} Z_n}} = \begin{cases} \sqrt{\mu^{n-m}(1-\mu^m)/(1-\mu^n)} & \text{if } \mu \neq 1, \\ \sqrt{m/n} & \text{if } \mu = 1. \end{cases}$$

2. Suppose $0 \le r \le n$, and that everything is known about the process up to time r. Conditional on this information, and using a symmetry argument, a randomly chosen individual in the nth generation has probability $1/Z_r$ of having as rth generation ancestor any given member of the rth generation. The chance that two individuals from the nth generation, chosen randomly and independently of each other, have the same rth generation ancestor is therefore $1/Z_r$. Therefore

$$\mathbb{P}(L < r) = \mathbb{E}\{\mathbb{P}(L < r \mid Z_r)\} = \mathbb{E}(1 - Z_r^{-1})$$

and so

$$\mathbb{P}(L = r) = \mathbb{P}(L < r+1) - \mathbb{P}(L < r) = \mathbb{E}(Z_r^{-1}) - \mathbb{E}(Z_{r+1}^{-1}), \qquad 0 \le r < n.$$

If $0 < \mathbb{P}(Z_1 = 0) < 1$, then almost the same argument proves that $\mathbb{P}(L = r \mid Z_n > 0) = \eta_r - \eta_{r+1}$ for $0 \le r < n$, where $\eta_r = \mathbb{E}(Z_r^{-1} \mid Z_n > 0)$.

3. The number Z_n of nth generation decendants satisfies

$$\mathbb{P}(Z_n = 0) = G_n(0) = \begin{cases} \dfrac{n}{n+1} & \text{if } p = q, \\ \dfrac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{if } p \neq q, \end{cases}$$

whence, for $n \ge 1$,

$$\mathbb{P}(T = n) = \mathbb{P}(Z_n = 0) - \mathbb{P}(Z_{n-1} = 0) = \begin{cases} \dfrac{1}{n(n+1)} & \text{if } p = q, \\ \dfrac{p^{n-1}q^n(p-q)^2}{(p^n - q^n)(p^{n+1} - q^{n+1})} & \text{if } p \neq q. \end{cases}$$

It follows that $\mathbb{E}(T) < \infty$ if and only if $p < q$.

4. (a) As usual,

$$G_2(s) = G(G(s)) = 1 - \alpha\{\alpha(1-s)^\beta\}^\beta = 1 - \alpha^{1+\beta}(1-s)^{\beta^2}.$$

This suggests that $G_n(s) = 1 - \alpha^{1+\beta+\cdots+\beta^{n-1}}(1-s)^{\beta^n}$ for $n \ge 1$; this formula may be proved easily by induction, using the fact that $G_n(s) = G(G_{n-1}(s))$.
(b) As in the above part (a),

$$G_2(s) = f^{-1}(P(f(f^{-1}(P(f(s)))))) = f^{-1}(P(P(f(s)))) = f^{-1}(P_2(f(s)))$$

where $P_2(s) = P(P(s))$. Similarly $G_n(s) = f^{-1}(P_n(f(s)))$ for $n \ge 1$, where $P_n(s) = P(P_{n-1}(s))$.
(c) With $P(s) = \alpha s/\{1-(1-\alpha)s\}$ where $\alpha = \gamma^{-1}$, it is an easy exercise to prove, by induction, that $P_n(s) = \alpha^n s/\{1-(1-\alpha^n)s\}$ for $n \ge 1$, implying that

$$G_n(s) = P_n(s^m)^{1/m} = \left\{\frac{\alpha^n s^m}{1-(1-\alpha^n)s^m}\right\}^{1/m}.$$

5. Let Z_n be the number of members of the nth generation. The $(n+1)$th generation has size $C_{n+1} + I_{n+1}$ where C_{n+1} is the number of natural offspring of the previous generation, and I_{n+1} is the number of immigrants. Therefore by the independence,

$$\mathbb{E}(s^{Z_{n+1}} \mid Z_n) = \mathbb{E}(s^{C_{n+1}} \mid Z_n)H(s) = G(s)^{Z_n}H(s),$$

whence

$$G_{n+1}(s) = \mathbb{E}(s^{Z_{n+1}}) = \mathbb{E}\{G(s)^{Z_n}\}H(s) = G_n(G(s))H(s).$$

6. By Example (5.4.3),

$$\mathbb{E}(s^{Z_n}) = \frac{n-(n-1)s}{n+1-ns} = \frac{n-1}{n} + \frac{1}{n^2(1+n^{-1}-s)}, \qquad n \ge 0.$$

Differentiate and set $s = 0$ to find that

$$\mathbb{E}(V_1) = \sum_{n=0}^{\infty} \mathbb{P}(Z_n = 1) = \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} = \tfrac{1}{6}\pi^2.$$

Similarly,

$$\mathbb{E}(V_2) = \sum_{n=0}^{\infty} \frac{n}{(n+1)^3} = \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} - \sum_{n=0}^{\infty} \frac{1}{(n+1)^3} = \tfrac{1}{6}\pi^2 - \sum_{n=0}^{\infty} \frac{1}{(n+1)^3},$$

$$\mathbb{E}(V_3) = \sum_{n=0}^{\infty} \frac{n^2}{(n+1)^4} = \sum_{n=0}^{\infty} \frac{(n+1)^2 - 2(n+1) + 1}{(n+1)^4} = \tfrac{1}{6}\pi^2 + \tfrac{1}{90}\pi^4 - 2\sum_{n=0}^{\infty} \frac{1}{(n+1)^3}.$$

The conclusion is obtained by eliminating $\sum_n (n+1)^{-3}$.

5.5 Solutions. Age-dependent branching processes

1. (i) The usual renewal argument shows as in Theorem (5.5.1) that

$$G_t(s) = \int_0^t G(G_{t-u}(s)) f_T(u)\,du + \int_t^{\infty} s f_T(u)\,du.$$

Differentiate with respect to t, to obtain

$$\frac{\partial}{\partial t} G_t(s) = G(G_0(s)) f_T(t) + \int_0^t \frac{\partial}{\partial t}\{G(G_{t-u}(s))\} f_T(u)\,du - s f_T(t).$$

Now $G_0(s) = s$, and

$$\frac{\partial}{\partial t}\{G(G_{t-u}(s))\} = -\frac{\partial}{\partial u}\{G(G_{t-u}(s))\},$$

so that, using the fact that $f_T(u) = \lambda e^{-\lambda u}$ if $u \ge 0$,

$$\int_0^t \frac{\partial}{\partial t}\{G(G_{t-u}(s))\} f_T(u)\,du = -\Big[G(G_{t-u}(s))f_T(u)\Big]_0^t - \lambda\int_0^t G(G_{t-u}(s))f_T(u)\,du,$$

having integrated by parts. Hence

$$\begin{aligned}\frac{\partial}{\partial t}G_t(s) &= G(s)\lambda e^{-\lambda t} + \Big\{-G(s)\lambda e^{-\lambda t} + \lambda G(G_t(s))\Big\} - \lambda\left\{G_t(s) - \int_t^\infty sf_T(u)\,du\right\} - s\lambda e^{-\lambda t}\\ &= \lambda\{G(G_t(s)) - G_t(s)\}.\end{aligned}$$

(ii) Substitute $G(s) = s^2$ into the last equation to obtain

$$\frac{\partial G_t}{\partial t} = \lambda(G_t^2 - G_t)$$

with boundary condition $G_0(s) = s$. Integrate to obtain $\lambda t + c(s) = \log\{1 - G_t^{-1}\}$ for some function $c(s)$. Using the boundary condition at $t = 0$, we find that $c(s) = \log\{1 - G_0^{-1}\} = \log\{1 - s^{-1}\}$, and hence $G_t(s) = se^{-\lambda t}/\{1 - s(1 - e^{-\lambda t})\}$. Expand in powers of s to find that $Z(t)$ has the geometric distribution $\mathbb{P}(Z(t) = k) = (1 - e^{-\lambda t})^{k-1}e^{-\lambda t}$ for $k \ge 1$.

2. The equation becomes

$$\frac{\partial G_t}{\partial t} = \tfrac{1}{2}(1 + G_t^2) - G_t$$

with boundary condition $G_0(s) = s$. This differential equation is easily solved with the result

$$G_t(s) = \frac{2s + t(1-s)}{2 + t(1-s)} = \frac{4/t}{2 + t(1-s)} - \frac{2-t}{t}.$$

We pick out the coefficient of s^n to obtain

$$\mathbb{P}(Z(t) = n) = \frac{4}{t(2+t)}\left(\frac{t}{2+t}\right)^n, \quad n \ge 1,$$

and therefore

$$\mathbb{P}(Z(t) \ge k) = \sum_{n=k}^\infty \frac{4}{t(2+t)}\left(\frac{t}{2+t}\right)^n = \frac{2}{t}\left(\frac{t}{2+t}\right)^k, \qquad k \ge 1.$$

It follows that, for $x > 0$ and in the limit as $t \to \infty$,

$$\mathbb{P}(Z(t) \ge xt \mid Z(t) > 0) = \frac{\mathbb{P}(Z(t) \ge xt)}{\mathbb{P}(Z(t) \ge 1)} = \left(\frac{t}{2+t}\right)^{\lfloor xt\rfloor - 1} = \left(1 + \frac{2}{t}\right)^{1-\lfloor xt\rfloor} \to e^{-2x}.$$

5.6 Solutions. Expectation revisited

1. Set $a = \mathbb{E}(X)$ to find that $u(X) \ge u(\mathbb{E}X) + \lambda(X - \mathbb{E}X)$ for some fixed λ. Take expectations to obtain the result.

2. Certainly $Z_n = \sum_{i=1}^{n} X_i$ and $Z = \sum_{i=1}^{\infty} |X_i|$ are such that $|Z_n| \le Z$, and the result follows by dominated convergence.

3. Apply Fatou's lemma to the sequence $\{-X_n : n \ge 1\}$ to find that

$$\mathbb{E}\Big(\limsup_{n\to\infty} X_n\Big) = -\mathbb{E}\Big(\liminf_{n\to\infty} -X_n\Big) \ge -\liminf_{n\to\infty} \mathbb{E}(-X_n) = \limsup_{n\to\infty} \mathbb{E}(X_n).$$

4. Suppose that $\mathbb{E}|X^r| < \infty$ where $r > 0$. We have that, if $x > 0$,

$$x^r \mathbb{P}(|X| \ge x) \le \int_{[x,\infty)} u^r \, dF(u) \to 0 \quad \text{as } x \to \infty,$$

where F is the distribution function of $|X|$.

Conversely suppose that $x^r \mathbb{P}(|X| \ge x) \to 0$ where $r \ge 0$, and let $0 \le s < r$. Now $\mathbb{E}|X^s| = \lim_{M\to\infty} \int_0^M u^s \, dF(u)$ and, by integration by parts,

$$\int_0^M u^s \, dF(u) = \Big[-u^s\big(1 - F(u)\big)\Big]_0^M + \int_0^M s u^{s-1}\big(1 - F(u)\big)\, du.$$

The first term on the right-hand side is negative. The integrand in the second term satisfies $su^{s-1}\mathbb{P}(|X| > u) \le su^{s-1} \cdot u^{-r}$ for all large u. Therefore the integral is bounded uniformly in M, as required.

5. Suppose first that, for all $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$, such that $\mathbb{E}(|X| I_A) < \epsilon$ for all A satisfying $\mathbb{P}(A) < \delta$. Fix $\epsilon > 0$, and find x (> 0) such that $\mathbb{P}(|X| > x) < \delta(\epsilon)$. Then, for $y > x$,

$$\int_{-y}^{y} |u| \, dF_X(u) \le \int_{-x}^{x} |u| \, dF_X(u) + \mathbb{E}\left(|X| I_{\{|X|>x\}}\right) \le \int_{-x}^{x} |u| \, dF_X(u) + \epsilon.$$

Hence $\int_{-y}^{y} |u| \, dF_X(u)$ converges as $y \to \infty$, whence $\mathbb{E}|X| < \infty$.

Conversely suppose that $\mathbb{E}|X| < \infty$. It follows that $\mathbb{E}\left(|X| I_{\{|X|>y\}}\right) \to 0$ as $y \to \infty$. Let $\epsilon > 0$, and find y such that $\mathbb{E}\left(|X| I_{\{|X|>y\}}\right) < \frac{1}{2}\epsilon$. For any event A, $I_A \le I_{A\cap B^c} + I_B$ where $B = \{|X| > y\}$. Hence

$$\mathbb{E}(|X| I_A) \le \mathbb{E}\left(|X| I_{A\cap B^c}\right) + \mathbb{E}\left(|X| I_B\right) \le y\mathbb{P}(A) + \tfrac{1}{2}\epsilon.$$

Writing $\delta = \epsilon/(2y)$, we have that $\mathbb{E}(|X| I_A) < \epsilon$ if $\mathbb{P}(A) < \delta$.

5.7 Solutions. Characteristic functions

1. Let X have the Cauchy distribution, with characteristic function $\phi(s) = e^{-|s|}$. Setting $Y = X$, we have that $\phi_{X+Y}(t) = \phi(2t) = e^{-2|t|} = \phi_X(t)\phi_Y(t)$. However, X and Y are certainly dependent.

2. (i) It is the case that $\mathrm{Re}\{\phi(t)\} = \mathbb{E}(\cos tX)$, so that, in the obvious notation,

$$\begin{aligned}\mathrm{Re}\{1 - \phi(2t)\} &= \int_{-\infty}^{\infty} \{1 - \cos(2tx)\}\, dF(x) = 2\int_{-\infty}^{\infty} \{1 - \cos(tx)\}\{1 + \cos(tx)\}\, dF(x) \\ &\le 4\int_{-\infty}^{\infty} \{1 - \cos(tx)\}\, dF(x) = 4\,\mathrm{Re}\{1 - \phi(t)\}.\end{aligned}$$

(ii) Note first that, if X and Y are independent with common characteristic function ϕ, then $X - Y$ has characteristic function

$$\psi(t) = \mathbb{E}(e^{itX})\mathbb{E}(e^{-itY}) = \phi(t)\phi(-t) = \phi(t)\overline{\phi(t)} = |\phi(t)|^2.$$

Apply the result of part (i) to the function ψ to obtain that $1 - |\phi(2t)|^2 \le 4(1 - |\phi(t)|^2)$. However $|\phi(t)| \le 1$, so that

$$1 - |\phi(2t)| \le 1 - |\phi(2t)|^2 \le 4(1 - |\phi(t)|^2) \le 8(1 - |\phi(t)|).$$

3. (a) With $m_k = \mathbb{E}(X^k)$, we have that

$$\mathbb{E}(e^{\theta X}) = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} m_k \theta^k = 1 + S(\theta),$$

say, and therefore, for sufficiently small values of θ,

$$K_X(\theta) = \sum_{r=1}^{\infty} \frac{(-1)^{r+1}}{r} S(\theta)^r.$$

Expand $S(\theta)^r$ in powers of θ, and equate the coefficients of $\theta, \theta^2, \theta^3$, in turn, to find that $k_1(X) = m_1$, $k_2(X) = m_2 - m_1^2$, $k_3(X) = m_3 - 3m_1m_2 + 2m_1^3$.

(b) If X and Y are independent, $K_{X+Y}(\theta) = \log\{\mathbb{E}(e^{\theta X})\mathbb{E}(e^{\theta Y})\} = K_X(\theta) + K_Y(\theta)$, whence the claim is immediate.

4. The $N(0, 1)$ variable X has moment generating function $\mathbb{E}(e^{\theta X}) = e^{\frac{1}{2}\theta^2}$, so that $K_X(\theta) = \frac{1}{2}\theta^2$.

5. (a) Suppose X takes values in $L(a, b)$. Then

$$|\phi_X(2\pi/b)| = \left|\sum_x e^{2\pi i x/b}\mathbb{P}(X = x)\right| = |e^{2\pi i a/b}|\left|\sum_m e^{2\pi i m}\mathbb{P}(X = a + bm)\right| = 1$$

since only numbers of the form $x = a + bm$ make non-zero contributions to the sum.

Suppose in addition that X has span b, and that $|\phi_X(T)| = 1$ for some $T \in (0, 2\pi/b)$. Then $\phi_X(T) = e^{ic}$ for some $c \in \mathbb{R}$. Now

$$\mathbb{E}\big(\cos(TX - c)\big) = \tfrac{1}{2}\mathbb{E}\big(e^{iTX-ic} + e^{-iTX+ic}\big) = 1,$$

using the fact that $\mathbb{E}(e^{-iTX}) = \overline{\phi_X(T)} = e^{-ic}$. However $\cos x \le 1$ for all x, with equality if and only if x is a multiple of 2π. It follows that $TX - c$ is a multiple of 2π, with probability 1, and hence that X takes values in the set $L(c/T, 2\pi/T)$. However $2\pi/T > b$, which contradicts the maximality of the span b. We deduce that no such T exists.

(b) This follows by the argument above.

6. This is a form of the 'Riemann–Lebesgue lemma'. It is a standard result of analysis that, for $\epsilon > 0$, there exists a step function g_ϵ such that $\int_{-\infty}^{\infty} |f(x) - g_\epsilon(x)|\,dx < \epsilon$. Let $\phi_\epsilon(t) = \int_{-\infty}^{\infty} e^{itx} g_\epsilon(x)\,dx$. Then

$$|\phi_X(t) - \phi_\epsilon(t)| = \left|\int_{-\infty}^{\infty} e^{itx}\big(f(x) - g_\epsilon(x)\big)\,dx\right| \le \int_{-\infty}^{\infty} |f(x) - g_\epsilon(x)|\,dx < \epsilon.$$

If we can prove that, for each ϵ, $|\phi_\epsilon(t)| \to 0$ as $t \to \pm\infty$, then it will follow that $|\phi_X(t)| < 2\epsilon$ for all large t, and the claim then follows.

Now $g_\epsilon(x)$ is a finite linear combination of functions of the form $cI_A(x)$ for reals c and intervals A, that is $g_\epsilon(x) = \sum_{k=1}^K c_k I_{A_k}(x)$; elementary integration yields

$$\phi_\epsilon(t) = \sum_{k=1}^K c_k \frac{e^{itb_k} - e^{ita_k}}{it}$$

where a_k and b_k are the endpoints of A_k. Therefore

$$|\phi_\epsilon(t)| \le \frac{2}{t}\sum_{k=1}^K c_k \to 0, \qquad \text{as } t \to \pm\infty.$$

7. If X is $N(\mu, 1)$, then the moment generating function of X^2 is

$$M_{X^2}(s) = \mathbb{E}(e^{sX^2}) = \int_{-\infty}^{\infty} e^{sx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}\, dx = \frac{1}{\sqrt{1-2s}} \exp\left(\frac{\mu^2 s}{1-2s}\right),$$

if $s < \frac{1}{2}$, by completing the square in the exponent. It follows that

$$M_Y(s) = \prod_{j=1}^n \left\{\frac{1}{\sqrt{1-2s}} \exp\left(\frac{\mu_j^2 s}{1-2s}\right)\right\} = \frac{1}{(1-2s)^{n/2}} \exp\left(\frac{s\theta}{1-2s}\right).$$

It is tempting to substitute $s = it$ to obtain the answer. This procedure may be justified in this case using the theory of analytic continuation.

8. (a) $T^2 = X^2/(Y/n)$, where X^2 is $\chi^2(1; \mu^2)$ by Exercise (5.7.7), and Y is $\chi^2(n)$. Hence T^2 is $F(1, n; \mu^2)$.
(b) F has the same distribution function as

$$Z = \frac{(A^2 + B)/m}{V/n}$$

where A, B, V are independent, A being $N(\sqrt{\theta}, 1)$, B being $\chi^2(m-1)$, and V being $\chi^2(n)$. Therefore

$$\begin{aligned}\mathbb{E}(Z) &= \frac{1}{m}\left\{\mathbb{E}(A^2)\mathbb{E}\left(\frac{n}{V}\right) + (m-1)\mathbb{E}\left(\frac{B/(m-1)}{V/n}\right)\right\} \\ &= \frac{1}{m}\left\{(1+\theta)\frac{n}{n-2} + (m-1)\frac{n}{n-2}\right\} = \frac{n(m+\theta)}{m(n-2)},\end{aligned}$$

where we have used the fact (see Exercise (4.10.2)) that the $F(r, s)$ distribution has mean $s/(s-2)$ if $s > 2$.

9. Let $\widetilde{X}$ be independent of X with the same distribution. Then $|\phi|^2$ is the characteristic function of $X - \widetilde{X}$ and, by the inversion theorem,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} |\phi(t)|^2 e^{-itx}\, dt = f_{X-\widetilde{X}}(x) = \int_{-\infty}^{\infty} f(y) f(x+y)\, dy.$$

Now set $x = 0$. We require that the density function of $X - \widetilde{X}$ be differentiable at 0.

10. By definition,

$$e^{-ity}\phi_X(y) = \int_{-\infty}^{\infty} e^{iy(x-t)} f_X(x)\,dx.$$

Now multiply by $f_Y(y)$, integrate over $y \in \mathbb{R}$, and change the order of integration with an appeal to *Fubini's theorem*.

11. (a) We adopt the usual convention that integrals of the form $\int_u^v g(y)\,dF(y)$ include any atom of the distribution function F at the upper endpoint v but not at the lower endpoint u. It is a consequence that F_τ is right-continuous, and it is immediate that F_τ increases from 0 to 1. Therefore F_τ is a distribution function. The corresponding moment generating function is

$$M_\tau(t) = \int_{-\infty}^{\infty} e^{tx}\,dF_\tau(x) = \frac{1}{M(t)}\int_{-\infty}^{\infty} e^{tx+\tau x}\,dF(x) = \frac{M(t+\tau)}{M(t)}.$$

(b) The required moment generating function is

$$\frac{M_{X+Y}(t+\tau)}{M_{X+Y}(t)} = \frac{M_X(t+\tau)M_Y(t+\tau)}{M_X(t)M_Y(t)},$$

the product of the moment generating functions of the individual tilted distributions.

5.8 Solutions. Examples of characteristic functions

1. (i) We have that $\overline{\phi}(t) = \overline{\mathbb{E}(e^{itX})} = \mathbb{E}(e^{-itX}) = \phi_{-X}(t)$.

(ii) If X_1 and X_2 are independent random variables with common characteristic function ϕ, then $\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t) = \phi(t)^2$.

(iii) Similarly, $\phi_{X_1-X_2}(t) = \phi_{X_1}(t)\phi_{-X_2}(t) = \phi(t)\overline{\phi(t)} = |\phi(t)|^2$.

(iv) Let X have characteristic function ϕ, and let Z be equal to X with probability $\frac{1}{2}$ and to $-X$ otherwise. The characteristic function of Z is given by

$$\phi_Z(t) = \tfrac{1}{2}\big(\mathbb{E}(e^{itX}) + \mathbb{E}(e^{-itX})\big) = \tfrac{1}{2}\big(\phi(t) + \overline{\phi(t)}\big) = \mathrm{Re}(\phi)(t),$$

where we have used the argument of part (i) above.

(v) If X is Bernoulli with parameter $\frac{1}{3}$, then its characteristic function is $\phi(t) = \frac{2}{3} + \frac{1}{3}e^{it}$. Suppose Y is a random variable with characteristic function $\psi(t) = |\phi(t)|$. Then $\psi(t)^2 = \phi(t)\phi(-t)$. Written in terms of random variables this asserts that $Y_1 + Y_2$ has the same distribution as $X_1 - X_2$, where the Y_i are independent with characteristic function ψ, and the X_i are independent with characteristic function ϕ. Now $X_j \in \{0, 1\}$, so that $X_1 - X_2 \in \{-1, 0, 1\}$, and therefore $Y_j \in \{-\frac{1}{2}, \frac{1}{2}\}$. Write $\alpha = \mathbb{P}(Y_j = \frac{1}{2})$. Then

$$\mathbb{P}(Y_1 + Y_2 = 1) = \alpha^2 = \mathbb{P}(X_1 - X_2 = 1) = \tfrac{2}{9},$$
$$\mathbb{P}(Y_1 + Y_2 = -1) = (1-\alpha)^2 = \mathbb{P}(X_1 - X_2 = -1) = \tfrac{2}{9},$$

implying that $\alpha^2 = (1-\alpha)^2$ so that $\alpha = \frac{1}{2}$, contradicting the fact that $\alpha^2 = \frac{2}{9}$. We deduce that no such variable Y exists.

2. For $t \ge 0$,

$$\mathbb{P}(X \ge x) = \mathbb{P}(e^{tX} \ge e^{tx}) \le e^{-tx}\mathbb{E}(e^{tX}).$$

Now minimize over $t \ge 0$.

3. The moment generating function of Z is

$$M_Z(t) = \mathbb{E}\left\{\mathbb{E}(e^{tXY} \mid Y)\right\} = \mathbb{E}\{M_X(tY)\} = \mathbb{E}\left\{\left(\frac{\lambda}{\lambda - tY}\right)^m\right\}$$
$$= \int_0^1 \left(\frac{\lambda}{\lambda - ty}\right)^m \frac{y^{n-1}(1-y)^{m-n-1}}{B(n, m-n)}\,dy.$$

Substitute $v = 1/y$ and integrate by parts to obtain that

$$I_{mn} = \int_1^\infty \frac{(v-1)^{m-n-1}}{(\lambda v - t)^m}\,dv$$

satisfies

$$I_{mn} = \left[-\frac{1}{\lambda(m-1)}\frac{(v-1)^{m-n-1}}{(\lambda v - t)^{m-1}}\right]_1^\infty + \frac{m-n-1}{\lambda(m-1)} I_{m-1,n} = c(m,n,\lambda) I_{m-1,n}$$

for some $c(m, n, \lambda)$. We iterate this to obtain

$$I_{mn} = c' I_{n+1,n} = c' \int_1^\infty \frac{dv}{(\lambda v - t)^{n+1}} = \frac{c'}{n\lambda} \cdot \frac{1}{(\lambda - t)^n}$$

for some c' depending on m, n, λ. Therefore $M_Z(t) = c''(\lambda - t)^{-n}$ for some c'' depending on m, n, λ. However $M_Z(0) = 1$, and hence $c'' = \lambda^n$, giving that Z is $\Gamma(\lambda, n)$. Throughout these calculations we have assumed that t is sufficiently small and positive. Alternatively, we could have set $t = is$ and used characteristic functions. See also Problem (4.14.12).

4. We have that

$$\mathbb{E}(e^{itX^2}) = \int_{-\infty}^{\infty} e^{itx^2} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left[x - \mu(1-2\sigma^2 it)^{-1}\right]^2}{2\sigma^2(1-2\sigma^2 it)^{-1}}\right) \exp\left(\frac{it\mu^2}{1-2\sigma^2 it}\right) dx$$
$$= \frac{1}{\sqrt{1-2\sigma^2 it}} \exp\left(\frac{it\mu^2}{1-2\sigma^2 it}\right).$$

The integral is evaluated by using Cauchy's theorem when integrating around a sector in the complex plane. It is highly suggestive to observe that the integrand differs only by a multiplicative constant from a hypothetical normal density function with (complex) mean $\mu(1-2\sigma^2 it)^{-1}$ and (complex) variance $\sigma^2(1-2\sigma^2 it)^{-1}$.

5. (a) Use the result of Exercise (5.8.4) with $\mu = 0$ and $\sigma^2 = 1$: $\phi_{X_1^2}(t) = (1-2it)^{-\frac{1}{2}}$, the characteristic function of the $\chi^2(1)$ distribution.

(b) From (a), the sum S has characteristic function $\phi_S(t) = (1-2it)^{-\frac{1}{2}n}$, the characteristic function of the $\chi^2(n)$ distribution.

(c) We have that

$$\mathbb{E}(e^{itX_1/X_2}) = \mathbb{E}\{\mathbb{E}(e^{itX_1/X_2} \mid X_2)\} = \mathbb{E}\left(\phi_{X_1}(t/X_2)\right) = \mathbb{E}(\exp\{-\tfrac{1}{2}t^2/X_2^2\}).$$

Now

$$\mathbb{E}\big(\exp\{-\tfrac{1}{2}t^2/X_2^2\}\big) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2x^2} - \frac{x^2}{2}\right) dx.$$

There are various ways of evaluating this integral. Using the result of Problem (5.12.18c), we find that the answer is $e^{-|t|}$, whence X_1/X_2 has the Cauchy distribution.

(d) We have that

$$\begin{aligned}\mathbb{E}(e^{itX_1X_2}) &= \mathbb{E}\{\mathbb{E}(e^{itX_1X_2} \mid X_2)\} = \mathbb{E}\big(\phi_{X_1}(tX_2)\big) = \mathbb{E}\big(e^{-\frac{1}{2}t^2X_2^2}\big)\\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}x^2(1+t^2)\}\, dx = \frac{1}{\sqrt{1+t^2}},\end{aligned}$$

on observing that the integrand differs from the $N(0, (1+t^2)^{-\frac{1}{2}})$ density function only by a multiplicative constant. Now, examination of a standard work of reference, such as Abramowitz and Stegun (1965, Section 9.6.21), reveals that

$$\int_0^{\infty} \frac{\cos(xt)}{\sqrt{1+t^2}}\, dt = K_0(x),$$

where $K_0(x)$ is the second kind of modified Bessel function. Hence the required density, by the inversion theorem, is $f(x) = K_0(|x|)/\pi$. Note that, for small x, $K_0(x) \sim -\log x$, and for large positive x, $K_0(x) \sim e^{-x}\sqrt{\pi x/2}$.

As a matter of interest, note that we may also invert the more general characteristic function $\phi(t) = (1-it)^{-\alpha}(1+it)^{-\beta}$. Setting $1 - it = -z/x$ in the integral gives

$$\begin{aligned}f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx}}{(1-it)^{\alpha}(1+it)^{\beta}}\, dt = \frac{e^{-x}x^{\alpha-1}}{2^{\beta}2\pi i}\int_{-x-ix\infty}^{-x+ix\infty} \frac{e^{-z}\, dz}{(-z)^{\alpha}\,(1+z/(2x))^{\beta}}\\ &= \frac{e^{x}(2x)^{\frac{1}{2}(\beta-\alpha)}}{\Gamma(\alpha)} W_{\frac{1}{2}(\alpha-\beta),\frac{1}{2}(1-\alpha-\beta)}(2x)\end{aligned}$$

where W is a confluent hypergeometric function. When $\alpha = \beta$ this becomes

$$f(x) = \frac{(x/2)^{\alpha-\frac{1}{2}}}{\Gamma(\alpha)\sqrt{\pi}} K_{\alpha-\frac{1}{2}}(x)$$

where K is a Bessel function of the second kind.

(e) Using (d), we find that the required characteristic function is $\phi_{X_1X_2}(t)\phi_{X_3X_4}(t) = (1+t^2)^{-1}$. In order to invert this, either use the inversion theorem for the Cauchy distribution to find the required density to be $f(x) = \frac{1}{2}e^{-|x|}$ for $-\infty < x < \infty$, or alternatively express $(1+t^2)^{-1}$ as partial fractions, $(1+t^2)^{-1} = \frac{1}{2}\{(1-it)^{-1} + (1+it)^{-1}\}$, and recall that $(1-it)^{-1}$ is the characteristic function of an exponential distribution.

6. The joint characteristic function of $\mathbf{X} = (X_1, X_2, \dots, X_n)$ satisfies $\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}\mathbf{X}'}) = \mathbb{E}(e^{iY})$ where $\mathbf{t} = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n$ and $Y = \mathbf{t}\mathbf{X}' = t_1X_1 + \cdots + t_nX_n$. Now Y is normal with mean and variance

$$\mathbb{E}(Y) = \sum_{j=1}^{n} t_j\mathbb{E}(X_j) = \mathbf{t}\boldsymbol{\mu}', \qquad \operatorname{var}(Y) = \sum_{j,k=1}^{n} t_jt_k\operatorname{cov}(X_j, X_k) = \mathbf{t}\mathbf{V}\mathbf{t}',$$

where $\boldsymbol{\mu}$ is the mean vector of $\mathbf{X}$, and $\mathbf{V}$ is the covariance matrix of $\mathbf{X}$. Therefore $\phi_{\mathbf{X}}(\mathbf{t}) = \phi_Y(1) = \exp(i\mathbf{t}\boldsymbol{\mu}' - \frac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}')$ by (5.8.5).

Let $\mathbf{Z} = \mathbf{X} - \boldsymbol{\mu}$. It is easy to check that the vector $\mathbf{Z}$ has joint characteristic function $\phi_{\mathbf{Z}}(\mathbf{t}) = e^{-\frac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}'}$, which we recognize by (5.8.6) as being that of the $N(\mathbf{0}, \mathbf{V})$ distribution.

7. We have that $\mathbb{E}(Z) = 0$, $\mathbb{E}(Z^2) = 1$, and $\mathbb{E}(e^{tZ}) = \mathbb{E}\{\mathbb{E}(e^{tZ} \mid U, V)\} = \mathbb{E}(e^{\frac{1}{2}t^2}) = e^{\frac{1}{2}t^2}$. If X and Y have the bivariate normal distribution with correlation ρ, then the random variable $Z = (UX + VY)/\sqrt{U^2 + 2\rho UV + V^2}$ is $N(0, 1)$.

8. By definition, $\mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))$. By integrating by parts,

$$\int_0^\infty \cos(tx)\lambda e^{-\lambda x}\,dx = \frac{\lambda^2}{\lambda^2 + t^2}, \qquad \int_0^\infty \sin(tx)\lambda e^{-\lambda x}\,dx = \frac{\lambda t}{\lambda^2 + t^2},$$

and

$$\frac{\lambda^2 + i\lambda t}{\lambda^2 + t^2} = \frac{\lambda}{\lambda - it}.$$

9. (a) We have that $e^{-|x|} = e^{-x}I_{\{x\ge 0\}} + e^{x}I_{\{x<0\}}$, whence the required characteristic function is

$$\phi(t) = \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2}.$$

(b) By a similar argument applied to the $\Gamma(1, 2)$ distribution, we have in this case that

$$\phi(t) = \frac{1}{2}\left(\frac{1}{(1-it)^2} + \frac{1}{(1+it)^2}\right) = \frac{1-t^2}{(1+t^2)^2}.$$

10. Suppose X has moment generating function $M(t)$. The proposed equation gives

$$M(t) = \int_0^1 M(ut)^2\,du = \frac{1}{t}\int_0^t M(v)^2\,dv.$$

Differentiate to obtain $tM' + M = M^2$, with solution $M(t) = \lambda/(\lambda + t)$. Thus the exponential distribution has the stated property.

11. We have that

$$\phi_{X,Y}(s,t) = \mathbb{E}(e^{isX+itY}) = \phi_{sX+tY}(1).$$

Now $sX + tY$ is $N(0, s^2\sigma^2 + 2st\sigma\tau\rho + \tau^2)$ where $\sigma^2 = \mathrm{var}(X)$, $\tau^2 = \mathrm{var}(Y)$, $\rho = \mathrm{corr}(X, Y)$, and therefore

$$\phi_{X,Y}(s,t) = \exp\{-\tfrac{1}{2}(s^2\sigma^2 + 2st\sigma\tau\rho + t^2\tau^2)\}.$$

The fact that $\phi_{X,Y}$ may be expressed in terms of the characteristic function of a single normal variable is sometimes referred to as the *Cramér–Wold device*.

5.9 Solutions. Inversion and continuity theorems

1. Clearly, for $0 \le y \le 1$, $\mathbb{P}(X_n \le ny) = n^{-1}\lfloor ny\rfloor \to y$ as $n \to \infty$.

2. (a) The derivative of F_n is $f_n(x) = 1 - \cos(2n\pi x)$, for $0 \le x \le 1$. It is easy to see that f_n is non-negative and $\int_0^1 f_n(x)\,dx = 1$. Therefore F_n is a distribution function with density function f_n.

(b) As $n \to \infty$,

$$\left|\frac{\sin(2n\pi x)}{2n\pi}\right| \le \frac{1}{2n\pi} \to 0,$$

and so $F_n(x) \to x$ for $0 \le x \le 1$. On the other hand, $\cos(2n\pi x)$ does not converge unless $x \in \{0, 1\}$, and therefore $f_n(x)$ does not converge on $(0, 1)$.

3. We may express N as the sum $N = T_1 + T_2 + \cdots + T_k$ of independent variables each having the geometric distribution $\mathbb{P}(T_j = r) = pq^{r-1}$ for $r \ge 1$, where $p + q = 1$. Therefore

$$\phi_N(t) = \phi_{T_1}(t)^k = \left\{\frac{pe^{it}}{1 - qe^{it}}\right\}^k,$$

implying that $Z = 2Np$ has characteristic function

$$\phi_Z(t) = \phi_N(2pt) = \left\{\frac{pe^{2pit}}{1 - (1-p)e^{2pit}}\right\}^k = \left\{\frac{p(1 + 2pit + o(p))}{p(1 - 2it + o(1))}\right\}^k \to (1 - 2it)^{-k}$$

as $p \downarrow 0$, the characteristic function of the $\Gamma(\frac{1}{2}, k)$ distribution. The result follows by the continuity theorem (5.9.5).

4. All you need to know is the fact, easily proved, that $\psi_m(t) = e^{itm}$ satisfies

$$\int_{-\pi}^{\pi} \psi_j(t)\psi_k(t)\,dt = \begin{cases} 2\pi & \text{if } j + k = 0, \\ 0 & \text{if } j + k \ne 0, \end{cases}$$

for integers j and k.

Now, $\phi(t) = \sum_{j=-\infty}^{\infty} e^{itj}\mathbb{P}(X = j)$, so that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi(t)\,dt = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \mathbb{P}(X = j)\int_{-\pi}^{\pi} \psi_j(t)\psi_{-k}(t)\,dt = \frac{1}{2\pi}\cdot\mathbb{P}(X = k)\cdot 2\pi.$$

If X is arithmetic with span λ, then X/λ is integer valued, whence

$$\mathbb{P}(X = k\lambda) = \frac{\lambda}{2\pi}\int_{-\pi/\lambda}^{\pi/\lambda} e^{-itk\lambda}\phi_X(t)\,dt.$$

5. Let X be uniformly distributed on $[-a, a]$, Y be uniformly distributed on $[-b, b]$, and let X and Y be independent. Then X has characteristic function $\sin(at)/(at)$, and Y has characteristic function $\sin(bt)/(bt)$. We apply the inversion theorem (5.9.1) to the characteristic function of $X + Y$ to find that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_{X+Y}(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\sin(at)\sin(bt)}{abt^2}\,dt = f_{X+Y}(0) = \frac{a \wedge b}{2ab}.$$

6. It is elementary that

$$\int_0^{\infty} \exp\{f_n(x)\}\,dx = \int_0^{\infty} x^n e^{-x}\,dx = \Gamma(n + 1) = n!.$$

In addition, $a = n$, $f_n''(a) = -n^{-1}$, and

$$\int_0^{\infty} \exp\{f_n(a) + \tfrac{1}{2}(x - a)^2 f_n''(a)\}\,dx = n^n e^{-n}\int_0^{\infty} \exp\left\{-\frac{(x - n)^2}{2n}\right\}dx \sim n^n e^{-n}\sqrt{2\pi n},$$

and Stirling's formula follows.

7. The vector $\mathbf{X}$ has joint characteristic function $\phi(\mathbf{t}) = \exp(-\frac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}')$. By the multidimensional version of the inversion theorem (5.9.1), the joint density function of $\mathbf{X}$ is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \exp\left(-i\mathbf{t}\mathbf{x}' - \tfrac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}'\right) d\mathbf{t}.$$

Therefore, if $i \neq j$,

$$\frac{\partial f}{\partial v_{ij}} = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} t_i t_j \exp\left(-i\mathbf{t}\mathbf{x}' - \tfrac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}'\right) d\mathbf{t} = \frac{\partial^2 f}{\partial x_i \partial x_j},$$

and similarly when $i = j$. When $i \neq j$,

$$\begin{aligned}\frac{\partial}{\partial v_{ij}}\mathbb{P}\big(\max_k X_k \leq u\big) &= \int_Q \frac{\partial f}{\partial v_{ij}}\, d\mathbf{x} \quad \text{where } Q = \big\{\mathbf{x} : x_k \leq u \text{ for } k = 1, 2, \dots, n\big\} \\ &= \int_Q \frac{\partial^2 f}{\partial x_i \partial x_j}\, d\mathbf{x} = \int' f\Big|_{x_i=x_j=-\infty}^{x_i=x_j=u} d\mathbf{x}' \geq 0,\end{aligned}$$

where $\int' \cdot\, d\mathbf{x}'$ is an integral over the variables x_k for $k \neq i, j$.

Therefore, $\mathbb{P}(\max_k X_k \leq u)$ increases in every parameter v_{ij}, and is therefore greater than its value when $v_{ij} = 0$ for $i \neq j$, namely $\prod_k \mathbb{P}(X_k \leq u)$.

8. By a two-dimensional version of the inversion theorem (5.9.1) applied to $\mathbb{E}(e^{i\mathbf{t}\mathbf{X}'})$, $\mathbf{t} = (t_1, t_2)$,

$$\begin{aligned}\frac{\partial}{\partial\rho}\mathbb{P}(X_1 > 0,\, X_2 > 0) &= \frac{\partial}{\partial\rho}\int_0^\infty\int_0^\infty \left\{\frac{1}{4\pi^2}\iint_{\mathbb{R}^2} \exp\left(-i\mathbf{t}\mathbf{x}' - \tfrac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}'\right) d\mathbf{t}\right\} d\mathbf{x} \\ &= \frac{\partial}{\partial\rho}\frac{1}{4\pi^2}\iint_{\mathbb{R}^2} \frac{\exp(-\frac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}')}{(it_1)(it_2)}\, d\mathbf{t} \\ &= \frac{1}{4\pi^2}\iint_{\mathbb{R}^2} \exp(-\tfrac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}')\, d\mathbf{t} = \frac{2\pi\sqrt{|\mathbf{V}^{-1}|}}{4\pi^2} = \frac{1}{2\pi\sqrt{1-\rho^2}}.\end{aligned}$$

We integrate with respect to ρ to find that, in agreement with Exercise (4.7.5),

$$\mathbb{P}(X_1 > 0,\, X_2 > 0) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho.$$

5.10 Solutions. Two limit theorems

1. (a) Let $\{X_i : i \geq 1\}$ be a collection of independent Bernoulli random variables with parameter $\frac{1}{2}$. Then $S_n = \sum_1^n X_i$ is binomially distributed as $\text{bin}(n, \frac{1}{2})$. Hence, by the central limit theorem,

$$2^{-n}\sum_{\substack{k:\\ |k-\frac{1}{2}n|\leq\frac{1}{2}x\sqrt{n}}}\binom{n}{k} = \mathbb{P}\left(\frac{|S_n - \frac{1}{2}n|}{\frac{1}{2}\sqrt{n}} \leq x\right) \to \Phi(x) - \Phi(-x) = \int_{-x}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}\, dy,$$

where Φ is the $N(0, 1)$ distribution function.

(b) Let $\{X_i : i \geq 1\}$ be a collection of independent Poisson random variables, each with parameter 1. Then $S_n = \sum_1^n X_i$ is Poisson with parameter n, and by the central limit theorem

$$e^{-n} \sum_{\substack{k:\\ |k-n| \leq x\sqrt{n}}} \frac{n^k}{k!} = \mathbb{P}\left(\frac{|S_n - n|}{\sqrt{n}} \leq x\right) \to \Phi(x) - \Phi(-x), \quad \text{as above.}$$

2. A superficially plausible argument asserts that, if all babies look the same, then the number X of correct answers in n trials is a random variable with the bin$(n, \frac{1}{2})$ distribution. Then, for large n,

$$\mathbb{P}\left(\frac{X - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} > 3\right) \simeq 1 - \Phi(3) \simeq \frac{1}{1000}$$

by the central limit theorem. For the given values of n and X,

$$\frac{X - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{910 - 750}{5\sqrt{15}} \simeq 8.$$

Now we might say that the event $\{X - \frac{1}{2}n > \frac{3}{2}\sqrt{n}\}$ is sufficiently unlikely that its occurrence casts doubt on the original supposition that babies look the same.

A statistician would level a good many objections at drawing such a clear cut decision from such murky data, but this is beyond our scope to elaborate.

3. Clearly

$$\phi_Y(t) = \mathbb{E}\{\mathbb{E}(e^{itY} \mid X)\} = \mathbb{E}\{\exp\big(X(e^{it} - 1)\big)\} = \left(\frac{1}{1 - (e^{it} - 1)}\right)^s = \left(\frac{1}{2 - e^{it}}\right)^s.$$

It follows that

$$\mathbb{E}(Y) = \frac{1}{i}\phi_Y'(0) = s, \qquad \mathbb{E}(Y^2) = -\phi_Y''(0) = s^2 + 2s,$$

whence $\text{var}(Y) = 2s$. Therefore the characteristic function of the normalized variable $Z = (Y - \mathbb{E}Y)/\sqrt{\text{var}(Y)}$ is

$$\phi_Z(t) = e^{-it\sqrt{s/2}}\phi_Y\big(t/\sqrt{2s}\big).$$

Now,

$$\begin{aligned}\log\{\phi_Y\big(t/\sqrt{2s}\big)\} &= -s\log\big(2 - e^{it/\sqrt{2s}}\big) = s\big(e^{it/\sqrt{2s}} - 1\big) + \tfrac{1}{2}s\big(e^{it/\sqrt{2s}} - 1\big)^2 + \mathrm{o}(1)\\ &= it\sqrt{\tfrac{1}{2}s} - \tfrac{1}{4}t^2 - \tfrac{1}{4}t^2 + \mathrm{o}(1),\end{aligned}$$

where the o(1) terms are as $s \to \infty$. Hence $\log\{\phi_Z(t)\} \to -\frac{1}{2}t^2$ as $s \to \infty$, and the result follows by the continuity theorem (5.9.5).

Let $P_1, P_2, \ldots$ be an infinite sequence of independent Poisson variables with parameter 1. Then $S_n = P_1 + P_2 + \cdots + P_n$ is Poisson with parameter n. Now Y has the Poisson distribution with parameter X, and so Y is distributed as S_X. Also, X has the same distribution as the sum of s independent exponential variables, implying that $X \to \infty$ as $s \to \infty$, with probability 1. This suggests by the central limit theorem that S_X (and hence Y also) is approximately normal in the limit as $s \to \infty$. We have neglected the facts that s and X are not generally integer valued.

4. Since X_1 is non-arithmetic, there exist integers $n_1, n_2, \ldots, n_k$ with greatest common divisor 1 and such that $\mathbb{P}(X_1 = n_i) > 0$ for $1 \leq i \leq k$. There exists N such that, for all $n \geq N$, there exist non-negative integers $\alpha_1, \alpha_2, \ldots, \alpha_k$ such that $n = \alpha_1 n_1 + \cdots + \alpha_k n_k$. If x is a non-negative integer, write

$N = \beta_1 n_1 + \cdots + \beta_k n_k$, $N + x = \gamma_1 n_1 + \cdots + \gamma_k n_k$ for non-negative integers $\beta_1, \ldots, \beta_k, \gamma_1 \ldots, \gamma_k$. Now $S_n = X_1 + \cdots + X_n$ is such that

$$\mathbb{P}(S_B = N) \ge \mathbb{P}\big(X_j = n_i \text{ for } B_{i-1} < j \le B_i,\ 1 \le i \le k\big) = \prod_{i=1}^{k} \mathbb{P}(X_1 = n_i)^{\beta_i} > 0$$

where $B_0 = 0$, $B_i = \beta_1 + \beta_2 + \cdots + \beta_i$, $B = B_k$. Similarly $\mathbb{P}(S_G = N + x) > 0$ where $G = \gamma_1 + \gamma_2 + \cdots + \gamma_k$. Therefore

$$\mathbb{P}(S_G - S_{G,B+G} = x) \ge \mathbb{P}(S_G = N + x)\mathbb{P}(S_B = N) > 0$$

where $S_{G,B+G} = \sum_{i=G+1}^{B+G} X_i$. Also, $\mathbb{P}(S_B - S_{B,B+G} = -x) > 0$ as required.

5. Let $X_1, X_2, \ldots$ be independent integer-valued random variables with mean 0, variance 1, span 1, and common characteristic function ϕ. We are required to prove that $\sqrt{n}\mathbb{P}(U_n = x) \to e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$ as $n \to \infty$ where

$$U_n = \frac{1}{\sqrt{n}} S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$$

and x is any number of the form $k/\sqrt{n}$ for integral k. The case of general μ and σ^2 is easily derived from this.

By the result of Exercise (5.9.4), for any such x,

$$\mathbb{P}(U_n = x) = \frac{1}{2\pi\sqrt{n}} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} e^{-itx} \phi_{U_n}(t)\,dt,$$

since U_n is arithmetic. Arguing as in the proof of the local limit theorem (6),

$$2\pi \left|\sqrt{n}\mathbb{P}(U_n = x) - f(x)\right| \le I_n + J_n$$

where f is the $N(0, 1)$ density function, and

$$I_n = \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \left|e^{-itx}\big(\phi_{U_n}(t) - e^{-\frac{1}{2}t^2}\big)\right| dt, \qquad J_n = \int_{|t|>\pi\sqrt{n}} |e^{-itx} e^{-\frac{1}{2}t^2}|\,dt.$$

Now $J_n = 2\sqrt{2\pi}\big(1 - \Phi(\pi\sqrt{n})\big) \to 0$ as $n \to \infty$, where Φ is the $N(0, 1)$ distribution function. As for I_n, pick $\delta \in (0, \pi)$. Then

$$I_n \le \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left|\phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2}\right| dt + \int_{\delta\sqrt{n}<|t|<\pi\sqrt{n}} \left\{\left|\phi(t/\sqrt{n})^n\right| + e^{-\frac{1}{2}t^2}\right\} dt.$$

The final term involving $e^{-\frac{1}{2}t^2}$ is dealt with as was J_n. By Exercise (5.7.5a), there exists $\lambda \in (0, 1)$ such that $|\phi(t)| < \lambda$ if $\delta \le |t| \le \pi$. This implies that

$$\int_{\delta\sqrt{n}<|t|<\pi\sqrt{n}} \left|\phi\big(t/\sqrt{n}\big)^n\right| dt \le (\pi - \delta)\lambda^n\sqrt{n} \to 0,$$

and it remains only to show that

$$\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left|\phi\big(t/\sqrt{n}\big)^n - e^{-\frac{1}{2}t^2}\right| dt \to 0 \quad \text{as } n \to \infty. \quad \bullet$$

The proof of this is considerably simpler if we make the extra (though unnecessary) assumption that $m_3 = \mathbb{E}|X_1^3| < \infty$, and we assume this henceforth. It is a consequence of Taylor's theorem (see Theorem (5.7.4)) that $\phi(t) = 1 - \frac{1}{2}t^2 - \frac{1}{6}it^3 m_3 + \mathrm{o}(t^3)$ as $t \to 0$. It follows that $\phi(t) = e^{-\frac{1}{2}t^2 + t^3\theta(t)}$ for some finite $\theta(t)$. Now $|e^x - 1| \le |x|e^{|x|}$, and therefore

$$\begin{aligned}\left|\phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2}\right| &= e^{-\frac{1}{2}t^2}\left|\exp\left(t^3 n^{-\frac{1}{2}}\theta(tn^{-\frac{1}{2}})\right) - 1\right| \\ &\le \frac{|t^3\theta(tn^{-\frac{1}{2}})|}{\sqrt{n}} \exp\left(\frac{|t^3\theta(tn^{-\frac{1}{2}})|}{\sqrt{n}} - \frac{1}{2}t^2\right).\end{aligned}$$

Let $K_\delta = \sup\{|\theta(u)| : |u| \le \delta\}$, noting that $K_\delta < \infty$, and pick δ sufficiently small that $0 < \delta < \pi$ and $\delta K_\delta < \frac{1}{4}$. For $|t| < \delta\sqrt{n}$,

$$\left|\phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2}\right| \le K_\delta \frac{|t|^3}{\sqrt{n}} \exp\left(t^2\delta K_\delta - \tfrac{1}{2}t^2\right) \le K_\delta \frac{|t|^3}{\sqrt{n}} e^{-\frac{1}{4}t^2},$$

and therefore

$$\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left|\phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2}\right| dt \le \frac{K_\delta}{\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^3 e^{-\frac{1}{4}t^2}\, dt \to 0 \quad \text{as } n \to \infty$$

as required.

6. The second moment of the X_i is

$$2\int_0^{e^{-1}} \frac{x^2}{2x(\log x)^2}\, dx = \int_{-\infty}^{-1} \frac{e^{2u}}{u^2}\, du$$

(substitute $x = e^u$), a finite integral. Therefore the X's have finite mean and variance. The density function is symmetric about 0, and so the mean is 0.

By the convolution formula, if $0 < x < e^{-1}$,

$$f_2(x) = \int_{-e^{-1}}^{e^{-1}} f(y)f(x-y)\, dy \ge \int_0^x f(y)f(x-y)\, dy \ge f(x)\int_0^x f(y)\, dy,$$

since $f(x-y)$, viewed as a function of y, is increasing on $[0, x]$. Hence

$$f_2(x) \ge \frac{f(x)}{2\log|x|} = \frac{1}{4|x|(\log|x|)^3}$$

for $0 < x < e^{-1}$. Continuing this procedure, we obtain

$$f_n(x) \ge \frac{k_n}{|x|(\log|x|)^{n+1}}, \quad 0 < x < e^{-1},$$

for some positive constant k_n. Therefore $f_n(x) \to \infty$ as $x \to 0$, and in particular the density function of $(X_1 + \cdots + X_n)/\sqrt{n}$ does not converge to the appropriate normal density at the origin.

7. We have for $s > 0$ that

$$\begin{aligned}\phi(is) &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \exp\left(-(2x)^{-1} - xs\right)x^{-3/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \exp\left(-\tfrac{1}{2}y^2 - sy^{-2}\right)2\, dy \quad \text{by substituting } x = y^{-2} \\ &= \exp(-\sqrt{2s}),\end{aligned}$$

by the result of Problem (5.12.18c), or by consulting a table of integrals. The required conclusion follows by analytic continuation in the upper half-plane. See Moran 1968, p. 271.

8. (a) The sum $S_n = \sum_{r=1}^n X_r$ has characteristic function $\mathbb{E}(e^{itS_n}) = \phi(t)^n = \phi(tn^2)$, whence $U_n = S_n/n$ has characteristic function $\phi(tn) = \mathbb{E}(e^{itnX_1})$. Therefore,

$$\mathbb{P}(S_n < c) = \mathbb{P}(nX_1 < c) = \mathbb{P}\left(X_1 < \frac{c}{n}\right) \to 0 \qquad \text{as } n \to \infty.$$

(b) $\mathbb{E}(e^{itT_n}) = \phi(t) = \mathbb{E}(e^{itX_1})$.

9. (a) Yes, because X_n is the sum of independent identically distributed random variables with non-zero variance.

(b) It cannot in general obey what we have called the central limit theorem, because $\text{var}(X_n) = (n^2 - n)\,\text{var}(\Theta) + n\mathbb{E}(\Theta)(1 - \mathbb{E}(\Theta))$ and $n\,\text{var}(X_1) = n\mathbb{E}(\Theta)(1 - \mathbb{E}(\Theta))$ are different whenever $\text{var}(\Theta) \neq 0$. Indeed the right 'normalization' involves dividing by n rather than $\sqrt{n}$. It may be shown when $\text{var}(\Theta) \neq 0$ that the distribution of X_n/n converges to that of the random variable Θ.

5.11 Solutions. Large deviations

1. We may write $S_n = \sum_1^n X_i$ where the X_i have moment generating function $M(t) = \frac{1}{2}(e^t + e^{-t})$. Applying the large deviation theorem (5.11.4), we obtain that, for $0 < a < 1$, $\mathbb{P}(S_n > an)^{1/n} \to \inf_{t>0}\{g(t)\}$ where $g(t) = e^{-at}M(t)$. Now g has a minimum when $e^t = \sqrt{(1+a)/(1-a)}$, where it takes the value $1/\sqrt{(1+a)^{1+a}(1-a)^{1-a}}$ as required. If $a \geq 1$, then $\mathbb{P}(S_n > an) = 0$ for all n.

2. (i) Let Y_n have the binomial distribution with parameters n and $\frac{1}{2}$. Then $2Y_n - n$ has the same distribution as the random variable S_n in Exercise (5.11.1). Therefore, if $0 < a < 1$,

$$\mathbb{P}(Y_n - \tfrac{1}{2}n > \tfrac{1}{2}an)^{1/n} = \mathbb{P}(S_n > an)^{1/n} \to \frac{1}{\sqrt{(1+a)^{1+a}(1-a)^{1-a}}},$$

and similarly for $\mathbb{P}(Y_n - \frac{1}{2}n < -\frac{1}{2}an)$, by symmetry. Hence

$$T_n^{1/n} = \left\{2^n \mathbb{P}\big(|Y_n - \tfrac{1}{2}n| > \tfrac{1}{2}an\big)\right\}^{1/n} \to \frac{4}{\sqrt{(1+a)^{1+a}(1-a)^{1-a}}}.$$

(ii) This time let $S_n = X_1 + \cdots + X_n$, the sum of independent Poisson variables with parameter 1. Then $T_n = e^n\mathbb{P}(S_n > n(1+a))$. The moment generating function of $X_1 - 1$ is $M(t) = \exp(e^t - 1 - t)$, and the large deviation theorem gives that $T_n^{1/n} \to e\inf_{t>0}\{g(t)\}$ where $g(t) = e^{-at}M(t)$. Now $g'(t) = (e^t - a - 1)\exp(e^t - at - t - 1)$ whence g has a minimum at $t = \log(a+1)$. Therefore $T_n^{1/n} \to eg(\log(1+a)) = \{e/(a+1)\}^{a+1}$.

3. Suppose that $M(t) = \mathbb{E}(e^{tX})$ is finite on the interval $[-\delta, \delta]$. Now, for $a > 0$, $M(\delta) \geq e^{\delta a}\mathbb{P}(X > a)$, so that $\mathbb{P}(X > a) \leq M(\delta)e^{-\delta a}$. Similarly, $\mathbb{P}(X < -a) \leq M(-\delta)e^{-\delta a}$.

Suppose conversely that such λ, μ exist. Then

$$M(t) \leq \mathbb{E}(e^{|tX|}) = \int_{[0,\infty)} e^{|t|x}\,dF(x)$$

where F is the distribution function of $|X|$. Integrate by parts to obtain

$$M(t) \leq 1 + \left[-e^{|t|x}[1 - F(x)]\right]_0^\infty + \int_0^\infty |t|e^{|t|x}[1 - F(x)]\,dx$$

(the term '1' takes care of possible atoms at 0). However $1 - F(x) \le \mu e^{-\lambda x}$, so that $M(t) < \infty$ if $|t|$ is sufficiently small.

4. The characteristic function of S_n/n is $\{e^{-|t/n|}\}^n = e^{-|t|}$, and hence S_n/n is Cauchy. Hence

$$\mathbb{P}(S_n > an) = \int_a^\infty \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi}\left(\frac{\pi}{2} - \tan^{-1} a\right).$$

5.12 Solutions to problems

1. The probability generating function of the sum is

$$\left\{\frac{1}{6}\sum_{i=1}^{6} s^i\right\}^{10} = \left(\frac{1}{6}s\right)^{10}\left\{\frac{1-s^6}{1-s}\right\}^{10} = \left(\frac{1}{6}s\right)^{10}(1 - 10s^6 + \cdots)(1 + 10s + \cdots).$$

The coefficient of s^{27} is

$$\left(\frac{1}{6}\right)^{10}\left\{\binom{10}{2}\binom{14}{5} - \binom{10}{1}\binom{20}{11} + \binom{26}{17}\right\}.$$

2. (a) The initial sequences T, HT, HHT, HHH induce a partition of the sample space. By conditioning on this initial sequence, we obtain $f(k) = qf(k-1) + pqf(k-2) + p^2qf(k-3)$ for $k > 3$, where $p + q = 1$. Also $f(1) = f(2) = 0$, $f(3) = p^3$. In principle, this difference equation may be solved in the usual way (see Appendix I). An alternative is to use generating functions. Set $G(s) = \sum_{k=1}^{\infty} s^k f(k)$, multiply throughout the difference equation by s^k and sum, to find that $G(s) = p^3s^3/\{1 - qs - pqs^2 - p^2qs^3\}$. To find the coefficient of s^k, factorize the denominator, expand in partial fractions, and use the binomial series.

Another equation for $f(k)$ is obtained by observing that $X = k$ if and only if $X > k-4$ and the last four tosses were THHH. Hence

$$f(k) = qp^3\left(1 - \sum_{i=1}^{k-4} f(i)\right), \qquad k > 3.$$

Applying the first argument to the mean, we find that $\mu = \mathbb{E}(X)$ satisfies $\mu = q(1+\mu) + pq(2+\mu) + p^2q(3+\mu) + 3p^3$ and hence $\mu = (1 + p + p^2)/p^3$.

As for HTH, consider the event that HTH does not occur in n tosses, and in addition the next three tosses give HTH. The number Y until the first occurrence of HTH satisfies

$$\mathbb{P}(Y > n)p^2q = \mathbb{P}(Y = n+1)pq + \mathbb{P}(Y = n+3), \quad n \ge 2.$$

Sum over n to obtain $\mathbb{E}(Y) = (pq+1)/(p^2q)$.

(b) $G_N(s) = (q + ps)^n$, in the obvious notation.

(i) $\mathbb{P}(2 \text{ divides } N) = \frac{1}{2}\{G_N(1) + G_N(-1)\}$, since only the coefficients of the *even* powers of s contribute to this probability.

(ii) Let ω be a complex cube root of unity. Then the coefficient of $\mathbb{P}(X = k)$ in $\frac{1}{3}\{G_N(1) + G_N(\omega) + G_N(\omega^2)\}$ is

$$\begin{aligned} &\tfrac{1}{3}\{1 + \omega^3 + \omega^6\} = 1, && \text{if } k = 3r,\\ &\tfrac{1}{3}\{1 + \omega + \omega^2\} = 0, && \text{if } k = 3r+1,\\ &\tfrac{1}{3}\{1 + \omega^2 + \omega^4\} = 0, && \text{if } k = 3r+2,\end{aligned}$$

for integers r. Hence $\frac{1}{3}\{G_N(1)+G_N(\omega)+G_N(\omega^2)\} = \sum_{r=0}^{\lfloor \frac{1}{3}n \rfloor} \mathbb{P}(N=3r)$, the probability that N is a multiple of 3. Generalize this conclusion.

3. We have that $T=k$ if no run of n heads appears in the first $k-n-1$ throws, then there is a tail, and then a run of n heads. Therefore $\mathbb{P}(T=k)=\mathbb{P}(T>k-n-1)qp^n$ for $k \ge n+1$ where $p+q=1$. Finally $\mathbb{P}(T=n)=p^n$. Multiply by s^k and sum to obtain a formula for the probability generating function G of T:

$$\begin{aligned} G(s)-p^n s^n &= qp^n \sum_{k=n+1}^{\infty} s^k \sum_{j>k-n-1} \mathbb{P}(T=j) = qp^n \sum_{j=1}^{\infty} \mathbb{P}(T=j) \sum_{k=n+1}^{n+j} s^k \\ &= \frac{qp^n s^{n+1}}{1-s} \sum_{j=1}^{\infty} \mathbb{P}(T=j)(1-s^j) = \frac{qp^n s^{n+1}}{1-s}(1-G(s)). \end{aligned}$$

Therefore

$$G(s) = \frac{p^n s^n - p^{n+1}s^{n+1}}{1-s+qp^n s^{n+1}}.$$

4. The required generating function is

$$G(s) = \sum_{k=r}^{\infty} s^k \binom{k-1}{r-1} p^r (1-p)^{k-r} = \left(\frac{ps}{1-qs}\right)^r$$

where $p+q=1$. The mean is $G'(1)=r/p$ and the variance is $G''(1)+G'(1)-\{G'(1)\}^2 = rq/p^2$.

5. It is standard (5.3.3) that $p_0(2n)=\binom{2n}{n}(pq)^n$. Using Stirling's formula,

$$p_0(2n) \sim \frac{(2n)^{2n+\frac{1}{2}} e^{-2n}\sqrt{2\pi}}{\{n^{n+\frac{1}{2}} e^{-n}\sqrt{2\pi}\}^2}(pq)^n = \frac{(4pq)^n}{\sqrt{\pi n}}.$$

The generating function $F_0(s)$ for the first return time is given by $F_0(s) = 1-P_0(s)^{-1}$ where $P_0(s)=\sum_n s^{2n} p_0(2n)$. Therefore the probability of ultimate return is $F_0(1)=1-\lambda^{-1}$ where, by Abel's theorem,

$$\lambda = \sum_n p_0(2n) \begin{cases} =\infty & \text{if } p=q=\frac{1}{2}, \\ <\infty & \text{if } p\ne q. \end{cases}$$

Hence $F_0(1)=1$ if and only if $p=\frac{1}{2}$.

6. (a) $R_n = X_n^2+Y_n^2$ satisfies

$$\begin{aligned} \mathbb{E}(R_{n+1}-R_n) &= \mathbb{E}\{(X_{n+1}^2-X_n^2)+(Y_{n+1}^2-Y_n^2)\} \\ &= 2\mathbb{E}(X_{n+1}^2-X_n^2) = 2\mathbb{E}\{\mathbb{E}(X_{n+1}^2-X_n^2 \mid X_n)\} \\ &= 2\mathbb{E}\{\tfrac{1}{4}[(X_n+1)^2-X_n^2]+\tfrac{1}{4}[(X_n-1)^2-X_n^2]\} = 1. \end{aligned}$$

Hence $R_n = n+R_0 = n$.

(b) The quick way is to argue as in the solution to Exercise (5.3.4). Let $U_n = X_n+Y_n$, $V_n=X_n-Y_n$. Then U and V are simple symmetric random walks, and furthermore they are independent. Therefore

$$p_0(2n) = \mathbb{P}(U_{2n}=0, V_{2n}=0) = \mathbb{P}(U_{2n}=0)\mathbb{P}(V_{2n}=0) = \left\{\left(\frac{1}{2}\right)^{2n}\binom{2n}{n}\right\}^2,$$

by (5.3.3). Using Stirling's formula, $p_0(2n) \sim (n\pi)^{-1}$, and therefore $\sum_n p_0(2n) = \infty$, implying that the chance of eventual return is 1.

A longer method is as follows. The walk is at the origin at time 0 if and only if it has taken equal numbers of leftward and rightward steps, and also equal numbers of upward and downward steps. Therefore

$$p_0(2n) = \left(\frac{1}{4}\right)^{2n} \sum_{m=0}^{n} \frac{(2n)!}{(m!)^2\{(n-m)!\}^2} = \left(\frac{1}{2}\right)^{4n} \binom{2n}{n}^2.$$

7. Let e_{ij} be the probability the walk ever reaches j having started from i. Clearly $e_{a0} = e_{a,a-1}e_{a-1,a-2}\cdots e_{10}$, since a passage to 0 from a requires a passage to $a-1$, then a passage to $a-2$, and so on. By homogeneity, $e_{a0} = (e_{10})^a$.

By conditioning on the value of the first step, we find that $e_{10} = pe_{30} + qe_{00} = pe_{10}^3 + q$. The cubic equation $x = px^3 + q$ has roots $x = 1, c, d$, where

$$c = \frac{-p - \sqrt{p^2 + 4pq}}{2p}, \quad d = \frac{-p + \sqrt{p^2 + 4pq}}{2p}.$$

Now $|c| > 1$, and $|d| \geq 1$ if and only if $p^2 + 4pq \geq 9p^2$ which is to say that $p \leq \frac{1}{3}$. It follows that $e_{10} = 1$ if $p \leq \frac{1}{3}$, so that $e_{a0} = 1$ if $p \leq \frac{1}{3}$.

When $p > \frac{1}{3}$, we have that $d < 1$, and it is actually the case that $e_{10} = d$, and hence

$$e_{a0} = \left(\frac{-p + \sqrt{p^2 + 4pq}}{2p}\right)^a \quad \text{if } p > \tfrac{1}{3}.$$

In order to prove this, it suffices to prove that $e_{a0} < 1$ for all large a; this is a minor but necessary chore. Write $T_n = S_n - S_0 = \sum_{i=1}^n X_i$, where X_i is the value of the ith step. Then

$$\begin{aligned} e_{a0} &= \mathbb{P}(T_n \leq -a \text{ for some } n \geq 1) = \mathbb{P}(n\mu - T_n \geq n\mu + a \text{ for some } n \geq 1) \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}(n\mu - T_n \geq n\mu + a) \end{aligned}$$

where $\mu = \mathbb{E}(X_1) = 2p - q > 0$. As in the theory of large deviations, for $t > 0$,

$$\mathbb{P}(n\mu - T_n \geq n\mu + a) \leq e^{-t(n\mu+a)}\{\mathbb{E}(e^{t(\mu-X)})\}^n$$

where X is a typical step. Now $\mathbb{E}(e^{t(\mu-X)}) = 1 + \mathrm{o}(t)$ as $t \downarrow 0$, and therefore we may pick $t > 0$ such that $\theta(t) = e^{-t\mu}\mathbb{E}(e^{t(\mu-X)}) < 1$. It follows that $e_{a0} \leq \sum_{n=1}^{\infty} e^{-ta}\theta(t)^n$ which is less than 1 for all large a, as required.

8. We have that

$$\mathbb{E}(s^X t^Y \mid X + Y = n) = \sum_{k=0}^{n} s^k t^{n-k} \binom{n}{k} p^k q^{n-k} = (ps + qt)^n,$$

where $p + q = 1$. Hence $G_{X,Y}(s,t) = G(ps + qt)$ where G is the probability generating function of $X + Y$. Now X and Y are independent, so that

$$G(ps + qt) = G_X(s)G_Y(t) = G_{X,Y}(s,1)G_{X,Y}(1,t) = G(ps + q)G(p + qt).$$

Write $f(u) = G(1+u)$, $x = s-1$, $y = t-1$, to obtain $f(px+qy) = f(px)f(qy)$, a functional equation valid at least when $-2 < x, y \le 0$. Now f is continuous within its disc of convergence, and also $f(0) = 1$; the usual argument (see Problem (4.14.5)) implies that $f(x) = e^{\lambda x}$ for some λ, and therefore $G(s) = f(s-1) = e^{\lambda(s-1)}$. Therefore $X+Y$ has the Poisson distribution with parameter λ. Furthermore, $G_X(s) = G(ps+q) = e^{\lambda p(s-1)}$, whence X has the Poisson distribution with parameter λp. Similarly Y has the Poisson distribution with parameter λq.

9. In the usual notation, $G_{n+1}(s) = G_n(G(s))$. It follows that $G''_{n+1}(1) = G''_n(1)G'(1)^2 + G'_n(1)G''(1)$ so that, after some work, $\operatorname{var}(Z_{n+1}) = \mu^2 \operatorname{var}(Z_n) + \mu^n\sigma^2$. Iterate to obtain

$$\operatorname{var}(Z_{n+1}) = \sigma^2(\mu^n + \mu^{n+1} + \cdots + \mu^{2n}) = \frac{\sigma^2\mu^n(1-\mu^{n+1})}{1-\mu}, \qquad n \ge 0,$$

for the case $\mu \ne 1$. If $\mu = 1$, then $\operatorname{var}(Z_{n+1}) = \sigma^2(n+1)$.

10. (a) Since the coin is unbiased, we may assume that each player, having won a round, continues to back the same face (heads or tails) until losing. The duration D of the game equals k if and only if k is the first time at which there has been either a run of $r-1$ heads or a run of $r-1$ tails; the probability of this may be evaluated in a routine way. Alternatively, argue as follows. We record S (for 'same') each time a coin shows the same face as its predecessor, and we record C (for 'change') otherwise; start with a C. It is easy to see that each symbol in the resulting sequence is independent of earlier symbols and is equally likely to be S or C. Now $D = k$ if and only if the first run of $r-2$ S's is completed at time k. It is immediate from the result of Problem (5.12.3) that

$$G_D(s) = \frac{(\frac{1}{2}s)^{r-2}(1-\frac{1}{2}s)}{1-s+(\frac{1}{2}s)^{r-1}}.$$

(b) The probability that A_k wins is

$$\pi_k = \sum_{n=1}^{\infty} \mathbb{P}\big(D = n(r-1)+k-1\big).$$

Let ω be a complex $(r-1)$th root of unity, and set

$$W_k(s) = \frac{1}{r-1}\left\{G_D(s) + \frac{1}{\omega^{k-1}}G_D(\omega s) + \frac{1}{\omega^{2(k-1)}}G_D(\omega^2 s) + \cdots + \frac{1}{\omega^{(r-2)(k-1)}}G_D(\omega^{r-2}s)\right\}.$$

It may be seen (as for Problem (5.12.2)) that the coefficient of s^i in $W_k(s)$ is $\mathbb{P}(D = i)$ if i is of the form $n(r-1)+(k-1)$ for some n, and is 0 otherwise. Therefore $\mathbb{P}(A_k \text{ wins}) = W_k(1)$.

(c) The pool contains £D when it is won. The required mean is therefore

$$\mathbb{E}(D \mid A_k \text{ wins}) = \frac{\mathbb{E}\big(DI_{\{A_k \text{ wins}\}}\big)}{\mathbb{P}(A_k \text{ wins})} = \frac{W'_k(1)}{W_k(1)}.$$

(d) Using the result of Exercise (5.1.2), the generating function of the sequence $\mathbb{P}(D > k)$, $k \ge 0$, is $T(s) = (1-G_D(s))/(1-s)$. The required probability is the coefficient of s^n in $T(s)$.

11. Let T_n be the total number of people in the first n generations. By considering the size Z_1 of the first generation, we see that

$$T_n = 1 + \sum_{i=1}^{Z_1} T_{n-1}(i)$$

where $T_{n-1}(1), T_{n-1}(2), \dots$ are independent random variables, each being distributed as T_{n-1}. Using the compounding formula (5.1.25), $H_n(s) = sG(H_{n-1}(s))$.

12. We have that

$$\begin{aligned}
\mathbb{P}(Z_n > N \mid Z_m = 0) &= \frac{\mathbb{P}(Z_n > N, Z_m = 0)}{\mathbb{P}(Z_m = 0)} \\
&= \sum_{r=1}^{\infty} \frac{\mathbb{P}(Z_m = 0 \mid Z_n = N + r)\mathbb{P}(Z_n = N + r)}{\mathbb{P}(Z_m = 0)} \\
&= \sum_{r=1}^{\infty} \frac{\mathbb{P}(Z_{m-n} = 0)^{N+r}\mathbb{P}(Z_n = N + r)}{\mathbb{P}(Z_m = 0)} \\
&\le \frac{\mathbb{P}(Z_m = 0)^{N+1}}{\mathbb{P}(Z_m = 0)} \sum_{r=1}^{\infty} \mathbb{P}(Z_n = N + r) \le \mathbb{P}(Z_m = 0)^N = G_m(0)^N.
\end{aligned}$$

13. (a) We have that $G_W(s) = G_N(G(s)) = e^{\lambda(G(s)-1)}$. Also, $G_W(s)^{1/n} = e^{\lambda((G(s)-1)/n}$, the same probability generating function as G_W but with λ replaced by λ/n.

(b) We can suppose that $H(0) < 1$, since if $H(0) = 1$ then $H(s) = 1$ for all s, and we may take $\lambda = 0$ and $G(s) = 1$. We may suppose also that $H(0) > 0$. To see this, suppose instead that $H(0) = 0$ so that $H(s) = s^r \sum_{j=0}^{\infty} s^j h_{j+r}$ for some sequence (h_k) and some $r \ge 1$ such that $h_r > 0$. Find a positive integer n such that r/n is non-integral; then $H(s)^{1/n}$ is not a power series, which contradicts the assumption that H is infinitely divisible.

Thus we take $0 < H(0) < 1$, and so $0 < 1 - H(s) < 1$ for $0 \le s < 1$. Therefore

$$\log H(s) = \log\big(1 - \{1 - H(s)\}\big) = \lambda\big(-1 + A(s)\big)$$

where $\lambda = -\log H(0)$ and $A(s)$ is a power series with $A(0) = 0$, $A(1) = 1$. Writing $A(s) = \sum_{j=1}^{\infty} a_j s^j$, we have that

$$\left. \frac{d^j}{ds^j}\{H(s)e^{\lambda}\}^{1/n} \right|_{s=0} = \frac{\lambda}{n} j!\, a_j + o(n^{-1})$$

as $n \to \infty$. Now $H(s)^{1/n}$ is a probability generating function, so that each such expression is non-negative. Therefore $a_j \ge 0$ for all j, implying that $A(s)$ is a probability generating function, as required.

14. It is clear from the definition of infinite divisibility that a distribution has this property if and only if, for each n, there exists a characteristic function ψ_n such that $\phi(t) = \psi_n(t)^n$ for all t.

(a) The characteristic functions in question are

$$\begin{aligned}
N(\mu, \sigma^2): &\quad \phi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2} \\
\text{Poisson } (\lambda): &\quad \phi(t) = e^{\lambda(e^{it}-1)} \\
\Gamma(\lambda, \mu): &\quad \phi(t) = \left(\frac{\lambda}{\lambda - it}\right)^{\mu}.
\end{aligned}$$

In these respective cases, the 'nth root' ψ_n of ϕ is the characteristic function of the $N(\mu/n, \sigma^2/n)$, Poisson (λ/n), and $\Gamma(\lambda, \mu/n)$ distributions.

(b) Suppose that ϕ is the characteristic function of an infinitely divisible distribution, and let ψ_n be a characteristic function such that $\phi(t) = \psi_n(t)^n$. Now $|\phi(t)| \le 1$ for all t, so that

$$|\psi_n(t)| = |\phi(t)|^{1/n} \to \begin{cases} 1 & \text{if } |\phi(t)| \neq 0, \\ 0 & \text{if } |\phi(t)| = 0. \end{cases}$$

For any value of t such that $\phi(t) \neq 0$, it is the case that $\psi_n(t) \to 1$ as $n \to \infty$. To see this, suppose instead that there exists θ satisfying $0 < \theta < 2\pi$ such that $\psi_n(t) \to e^{i\theta}$ along some subsequence. Then $\psi_n(t)^n$ does not converge along this subsequence, a contradiction. It follows that

$$(*) \qquad \psi(t) = \lim_{n\to\infty} \psi_n(t) = \begin{cases} 1 & \text{if } \phi(t) \neq 0, \\ 0 & \text{if } \phi(t) = 0. \end{cases}$$

Now ϕ is a characteristic function, so that $\phi(t) \neq 0$ on some neighbourhood of the origin. Hence $\psi(t) = 1$ on some neighbourhood of the origin, so that ψ is continuous at the origin. Applying the continuity theorem (5.9.5), we deduce that ψ is itself a characteristic function. In particular, ψ is continuous, and hence $\psi(t) = 1$ for all t, by $(*)$. We deduce that $\phi(t) \neq 0$ for all t.

15. We have that

$$\mathbb{P}(N = n \mid S = N) = \frac{\mathbb{P}(S = n \mid N = n)\mathbb{P}(N = n)}{\sum_k \mathbb{P}(S = k \mid N = k)\mathbb{P}(N = k)} = \frac{p^n \mathbb{P}(N = n)}{\sum_{k=1}^{\infty} p^k \mathbb{P}(N = k)}.$$

Hence $\mathbb{E}(x^N \mid S = N) = G(px)/G(p)$.

If N is Poisson with parameter λ, then

$$\mathbb{E}(x^N \mid S = N) = \frac{e^{\lambda(px-1)}}{e^{\lambda(p-1)}} = e^{\lambda p(x-1)} = G(x)^p.$$

Conversely, suppose that $\mathbb{E}(x^N \mid S = N) = G(x)^p$. Then $G(px) = G(p)G(x)^p$, valid for $|x| \le 1$, $0 < p < 1$. Therefore $f(x) = \log G(x)$ satisfies $f(px) = f(p) + pf(x)$, and in addition f has a power series expansion which is convergent at least for $0 < x \le 1$. Substituting this expansion into the above functional equation for f, and equating coefficients of $p^i x^j$, we obtain that $f(x) = -\lambda(1 - x)$ for some $\lambda \ge 0$. It follows that N has a Poisson distribution.

16. Certainly

$$G_X(s) = G_{X,Y}(s, 1) = \left(\frac{1 - (p_1 + p_2)}{1 - p_2 - p_1 s}\right)^n, \quad G_Y(t) = G_{X,Y}(1, t) = \left(\frac{1 - (p_1 + p_2)}{1 - p_1 - p_2 t}\right)^n,$$
$$G_{X+Y}(s) = G_{X,Y}(s, s) = \left(\frac{1 - (p_1 + p_2)}{1 - (p_1 + p_2)s}\right)^n,$$

giving that X, Y, and $X + Y$ have distributions similar to the negative binomial distribution. More specifically,

$$\mathbb{P}(X = k) = \binom{n + k - 1}{k}\alpha^k(1 - \alpha)^n, \quad \mathbb{P}(Y = k) = \binom{n + k - 1}{k}\beta^k(1 - \beta)^n,$$
$$\mathbb{P}(X + Y = k) = \binom{n + k - 1}{k}\gamma^k(1 - \gamma)^n,$$

for $k \ge 0$, where $\alpha = p_1/(1 - p_2)$, $\beta = p_2/(1 - p_1)$, $\gamma = p_1 + p_2$.

Now

$$\mathbb{E}(s^X \mid Y = y) = \frac{\mathbb{E}(s^X I_{\{Y=y\}})}{\mathbb{P}(Y = y)} = \frac{A}{B}$$

where A is the coefficient of t^y in $G_{X,Y}(s,t)$ and B is the coefficient of t^y in $G_Y(t)$. Therefore

$$\begin{aligned}\mathbb{E}(s^X \mid Y = y) &= \left(\frac{1-p_1-p_2}{1-p_1 s}\right)^n \left(\frac{p_2}{1-p_1 s}\right)^y \Big/ \left\{\left(\frac{1-p_1-p_2}{1-p_1}\right)^n \left(\frac{p_2}{1-p_1}\right)^y\right\} \\ &= \left(\frac{1-p_1}{1-p_1 s}\right)^{n+y}.\end{aligned}$$

17. As in the previous solution,

$$G_X(s) = e^{(\alpha+\gamma)(s-1)}, \quad G_Y(s) = e^{(\beta+\gamma)(t-1)}, \quad G_{X+Y}(s) = e^{(\alpha+\beta)(s-1)} e^{\gamma(s^2-1)}.$$

18. (a) Substitute $u = y/a$ to obtain

$$I(a,b) = \int_0^\infty \exp(-y^2 - a^2 b^2 y^{-2}) a^{-1}\, dy = a^{-1} I(1, ab).$$

(b) Differentiating through the integral sign,

$$\begin{aligned}\frac{\partial I}{\partial b} &= \int_0^\infty \left\{-\frac{2b}{u^2}\exp(-a^2u^2 - b^2u^{-2})\right\} du \\ &= -\int_0^\infty 2\exp(-a^2b^2y^{-2} - y^2)\, dy = -2I(1, ab),\end{aligned}$$

by the substitution $u = b/y$.

(c) Hence $\partial I/\partial b = -2aI$, whence $I = c(a)e^{-2ab}$ where

$$c(a) = I(a, 0) = \int_0^\infty e^{-a^2u^2}\, du = \frac{\sqrt{\pi}}{2a}.$$

(d) We have that

$$\mathbb{E}(e^{-tX}) = \int_0^\infty e^{-tx} \frac{d}{\sqrt{x}} e^{-c/x - gx}\, dx = 2dI\left(\sqrt{g+t}, \sqrt{c}\right)$$

by the substitution $x = y^2$.

(e) Similarly

$$\mathbb{E}(e^{-tX}) = \int_0^\infty e^{-tx} \frac{1}{\sqrt{2\pi x^3}} e^{-1/(2x)}\, dx = \sqrt{\frac{2}{\pi}}\, I\left(\frac{1}{\sqrt{2}}, \sqrt{t}\right)$$

by substituting $x = y^{-2}$.

19. (a) We have that

$$\begin{aligned}\mathbb{E}(e^{itU}) &= \mathbb{E}\{\mathbb{E}(e^{itX/Y} \mid Y)\} = \mathbb{E}\{\phi_X(t/Y)\} = \mathbb{E}\{e^{-\frac{1}{2}t^2/Y^2}\} \\ &= \int_{-\infty}^\infty e^{-\frac{1}{2}t^2/y^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy = \sqrt{\frac{2}{\pi}}\, I\left(\frac{1}{\sqrt{2}}, \frac{|t|}{\sqrt{2}}\right) = e^{-|t|}\end{aligned}$$

in the notation of Problem (5.12.18). Hence U has the Cauchy distribution.
(b) Similarly

$$\mathbb{E}(e^{-tV}) = \int_{-\infty}^{\infty} e^{-tx^{-2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = \sqrt{\frac{2}{\pi}} I\left(\frac{1}{\sqrt{2}}, \sqrt{t}\right) = e^{-\sqrt{2t}}$$

for $t > 0$. Using the result of Problem (5.12.18e), V has density function

$$f(x) = \frac{1}{\sqrt{2\pi x^3}} e^{-1/(2x)}, \quad x > 0.$$

(c) We have that $W^{-2} = X^{-2} + Y^{-2} + Z^{-2}$. Therefore, using (b),

$$\mathbb{E}(e^{-tW^{-2}}) = e^{-3\sqrt{2t}} = e^{-\sqrt{18t}} = \mathbb{E}(e^{-9Vt})$$

for $t > 0$. It follows that W^{-2} has the same distribution as $9V = 9X^{-2}$, and so W^2 has the same distribution as $\frac{1}{9}X^2$. Therefore, using the fact that both X and W are symmetric random variables, W has the same distribution as $\frac{1}{3}X$, that is $N(0, \frac{1}{9})$.

20. It follows from the inversion theorem that

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{2\pi} \lim_{N\to\infty} \int_{-N}^{N} \frac{1 - e^{-ith}}{it} e^{-itx} \phi(t)\, dt.$$

Since $|\phi|$ is integrable, we may use the dominated convergence theorem to take the limit as $h \downarrow 0$ within the integral:

$$f(x) = \frac{1}{2\pi} \lim_{N\to\infty} \int_{-N}^{N} e^{-itx} \phi(t)\, dt.$$

The condition that ϕ be absolutely integrable is stronger than necessary; note that the characteristic function of the exponential distribution fails this condition, in reflection of the fact that its density function has a discontinuity at the origin.

21. Let G_n denote the probability generating function of Z_n. The (conditional) characteristic function of Z_n/μ^n is

$$\mathbb{E}\left(e^{itZ_n/\mu^n} \mid Z_n > 0\right) = \frac{G_n(e^{it/\mu^n}) - G_n(0)}{1 - G_n(0)}.$$

It is a standard exercise (or see Example (5.4.3)) that

$$G_n(s) = \frac{\mu^n - 1 - \mu s(\mu^{n-1} - 1)}{\mu^{n+1} - 1 - \mu s(\mu^n - 1)},$$

whence by an elementary calculation

$$\mathbb{E}\left(e^{itZ_n/\mu^n} \mid Z_n > 0\right) \to \frac{\mu - 1}{\mu - 1 - \mu i t} \quad \text{as } n \to \infty,$$

the characteristic function of the exponential distribution with parameter $1 - \mu^{-1}$.

22. The imaginary part of $\phi_X(t)$ satisfies

$$\tfrac{1}{2}\{\phi_X(t) - \overline{\phi_X(t)}\} = \tfrac{1}{2}\{\phi_X(t) - \phi_X(-t)\} = \tfrac{1}{2}\{\mathbb{E}(e^{itX}) - \mathbb{E}(e^{-itX})\} = 0$$

for all t, if and only if X and $-X$ have the same characteristic function, or equivalently the same distribution.

23. (a) $U = X + Y$ and $V = X - Y$ are independent, so that $\phi_{U+V} = \phi_U \phi_V$, which is to say that $\phi_{2X} = \phi_{X+Y}\phi_{X-Y}$, or

$$\phi(2t) = \{\phi(t)^2\}\{\phi(t)\phi(-t)\} = \phi(t)^3\phi(-t).$$

Write $\psi(t) = \phi(t)/\phi(-t)$. Then

$$\psi(2t) = \frac{\phi(2t)}{\phi(-2t)} = \frac{\phi(t)^3\phi(-t)}{\phi(-t)^3\phi(t)} = \psi(t)^2.$$

Therefore

$$\psi(t) = \psi(\tfrac{1}{2}t)^2 = \psi(\tfrac{1}{4}t)^4 = \cdots = \psi(t/2^n)^{2^n} \qquad \text{for } n \geq 0.$$

However, as $h \to 0$,

$$\psi(h) = \frac{\phi(h)}{\phi(-h)} = \frac{1 - \frac{1}{2}h^2 + o(h^2)}{1 - \frac{1}{2}h^2 + o(h^2)} = 1 + o(h^2),$$

so that $\psi(t) = \{1 + o(t^2/2^{2n})\}^{2^n} \to 1$ as $n \to \infty$, whence $\psi(t) = 1$ for all t, giving that $\phi(-t) = \phi(t)$. It follows that

$$\begin{aligned}\phi(t) &= \phi(\tfrac{1}{2}t)^3\phi(-\tfrac{1}{2}t) = \phi(\tfrac{1}{2}t)^4 = \phi(t/2^n)^{2^{2n}} && \text{for } n \geq 1\\ &= \left\{1 - \frac{1}{2}\cdot\frac{t^2}{2^{2n}} + o(t^2/2^{2n})\right\}^{2^{2n}} \to e^{-\frac{1}{2}t^2} && \text{as } n \to \infty,\end{aligned}$$

so that X and Y are $N(0, 1)$.

(b) With $U = X + Y$ and $V = X - Y$, we have that $\psi(s, t) = \mathbb{E}(e^{isU+itV})$ satisfies

$$(*) \qquad \psi(s, t) = \mathbb{E}(e^{i(s+t)X+i(s-t)Y}) = \phi(s+t)\phi(s-t).$$

Using what is given,

$$\left.\frac{\partial^2\psi}{\partial t^2}\right|_{t=0} = -\mathbb{E}(V^2e^{isU}) = -\mathbb{E}\{e^{isU}\mathbb{E}(V^2 \mid U)\} = -\mathbb{E}(2e^{isU}) = -2\phi(s)^2.$$

However, by $(*)$,

$$\left.\frac{\partial^2\psi}{\partial t^2}\right|_{t=0} = 2\{\phi''(s)\phi(s) - \phi'(s)^2\},$$

yielding the required differential equation, which may be written as

$$\frac{d}{ds}(\phi'/\phi) = -1.$$

Hence $\log\phi(s) = a + bs - \frac{1}{2}s^2$ for constants a, b, whence $\phi(s) = e^{-\frac{1}{2}s^2}$.

24. (a) Using characteristic functions, $\phi_Z(t) = \phi_X(t/n)^n = e^{-|t|}$.

(b) $\mathbb{E}|X_i| = \infty$.

25. (a) See the solution to Problem (5.12.24).
(b) This is much longer. Having established the hint, the rest follows thus:

$$\begin{aligned} f_{X+Y}(y) &= \int_{-\infty}^{\infty} f(x)f(y-x)\,dx \\ &= \frac{1}{\pi(4+y^2)}\int_{-\infty}^{\infty}\big\{f(x)+f(y-x)\big\}\,dx + Jg(y) = \frac{2}{\pi(4+y^2)} + Jg(y) \end{aligned}$$

where

$$\begin{aligned} J &= \int_{-\infty}^{\infty}\big\{xf(x)+(y-x)f(y-x)\big\}\,dx \\ &= \lim_{M,N\to\infty}\left[\frac{1}{2\pi}\big\{\log(1+x^2)-\log\big(1+(y-x)^2\big)\big\}\right]_{-M}^{N} = 0. \end{aligned}$$

Finally,

$$f_Z(z) = 2f_{X+Y}(2z) = \frac{1}{\pi(1+z^2)}.$$

26. (a) $X_1 + X_2 + \cdots + X_n$.
(b) $X_1 - X_1'$, where X_1 and X_1' are independent and identically distributed.
(c) X_N, where N is a random variable with $\mathbb{P}(N = j) = p_j$ for $1 \le j \le n$, independent of $X_1, X_2, \ldots, X_n$.
(d) $\sum_{j=1}^{M} Z_j$ where $Z_1, Z_2, \ldots$ are independent and distributed as X_1, and M is independent of the Z_j with $\mathbb{P}(M = m) = (\frac{1}{2})^{m+1}$ for $m \ge 0$.
(e) YX_1, where Y is independent of X_1 with the exponential distribution parameter 1.

27. (a) We require

$$\phi(t) = \int_{-\infty}^{\infty} \frac{2e^{itx}}{e^{\pi x}+e^{-\pi x}}\,dx.$$

First method. Consider the contour integral

$$I_K = \int_C \frac{2e^{itz}}{e^{\pi z}+e^{-\pi z}}\,dz$$

where C is a rectangular contour with vertices at $\pm K$, $\pm K + i$. The integrand has a simple pole at $z = \frac{1}{2}i$, with residue $e^{-\frac{1}{2}t}/(i\pi)$. Hence, by Cauchy's theorem,

$$I_K \to \frac{2e^{-\frac{1}{2}t}}{1+e^{-t}} = \frac{1}{\cosh(\frac{1}{2}t)} \qquad \text{as } K \to \infty.$$

Second method. Expand the denominator to obtain

$$\frac{1}{\cosh(\pi x)} = \sum_{k=0}^{\infty}(-1)^k \exp\big\{-(2k+1)\pi|x|\big\}.$$

Multiply by e^{itx} and integrate term by term.

(b) Define $\phi(t) = 1 - |t|$ for $|t| \le 1$, and $\phi(t) = 0$ otherwise. Then

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,dt &= \frac{1}{2\pi}\int_{-1}^{1} e^{-itx}(1-|t|)\,dt \\ &= \frac{1}{\pi}\int_0^1 (1-t)\cos(tx)\,dt = \frac{1}{\pi x^2}(1-\cos x).\end{aligned}$$

Using the inversion theorem, ϕ is the required characteristic function.

(c) In this case,

$$\int_{-\infty}^{\infty} e^{itx}e^{-x-e^{-x}}\,dx = \int_0^{\infty} y^{-it}e^{-y}\,dy = \Gamma(1-it)$$

where Γ is the gamma function.

(d) Similarly,

$$\begin{aligned}\int_{-\infty}^{\infty} \tfrac{1}{2}e^{itx}e^{-|x|}\,dx &= \tfrac{1}{2}\left\{\int_0^{\infty} e^{itx}e^{-x}\,dx + \int_0^{\infty} e^{-itx}e^{-x}\,dx\right\} \\ &= \frac{1}{2}\left\{\frac{1}{1-it} + \frac{1}{1+it}\right\} = \frac{1}{1+t^2}.\end{aligned}$$

(e) We have that $\mathbb{E}(X) = -i\phi'(0) = -\Gamma'(1)$. Now, Euler's product for the gamma function states that

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!\,n^z}{z(z+1)\cdots(z+n)}$$

where the convergence is uniform on a neighbourhood of the point $z = 1$. By differentiation,

$$\Gamma'(1) = \lim_{n\to\infty}\left\{\frac{n}{n+1}\left(\log n - 1 - \frac{1}{2} - \cdots - \frac{1}{n+1}\right)\right\} = -\gamma.$$

28. (a) See Problem (5.12.27b).

(b) Suppose ϕ is the characteristic function of X. Since $\phi'(0) = \phi''(0) = \phi'''(0) = 0$, we have that $\mathbb{E}(X) = \mathrm{var}(X) = 0$, so that $\mathbb{P}(X = 0) = 1$, and hence $\phi(t) = 1$, a contradiction. Hence ϕ is not a characteristic function.

(c) As for (b).

(d) We have that $\cos t = \frac{1}{2}(e^{it} + e^{-it})$, whence ϕ is the characteristic function of a random variable taking values ± 1 each with probability $\frac{1}{2}$.

(e) By the same working as in the solution to Problem (5.12.27b), ϕ is the characteristic function of the density function

$$f(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1, \\ 0 & \text{otherwise.}\end{cases}$$

29. We have that

$$\begin{aligned}|1-\phi(t)| &\le \mathbb{E}|1-e^{itX}| = \mathbb{E}\sqrt{(1-e^{itX})(1-e^{-itX})} \\ &= \mathbb{E}\sqrt{2\{1-\cos(tX)\}} \le \mathbb{E}|tX|\end{aligned}$$

since $2(1-\cos x) \le x^2$ for all x.

30. This is a consequence of Taylor's theorem for functions of two variables:

$$\phi(s,t) = \sum_{\substack{m\le M \\ n\le N}} \frac{s^m t^n}{m!\,n!}\phi_{mn}(0,0) + R_{MN}(s,t)$$

where ϕ_{mn} is the derivative of ϕ in question, and R_{MN} is the remainder. However, subject to appropriate conditions,

$$\phi(s,t) = \sum_{\substack{m \le M \\ n \le N}} \frac{(is)^m (it)^n}{m!\,n!} \mathbb{E}(X^m Y^n) + o(s^M t^N)$$

whence the claim follows.

31. (a) We have that

$$\frac{x^2}{3} \le \frac{x^2}{2!} - \frac{x^4}{4!} \le 1 - \cos x$$

if $|x| \le 1$, and hence

$$\begin{aligned}\int_{[-t^{-1},t^{-1}]} (tx)^2\, dF(x) &\le \int_{[-t^{-1},t^{-1}]} 3\{1 - \cos(tx)\}\, dF(x) \\ &\le 3 \int_{-\infty}^{\infty} \{1 - \cos(tx)\}\, dF(x) = 3\{1 - \operatorname{Re}\phi(t)\}.\end{aligned}$$

(b) Using Fubini's theorem,

$$\begin{aligned}\frac{1}{t}\int_0^t \{1 - \operatorname{Re}\phi(v)\}\, dv &= \int_{x=-\infty}^{\infty} \frac{1}{t}\int_{v=0}^{t} \{1 - \cos(vx)\}\, dv\, dF(x) \\ &= \int_{-\infty}^{\infty} \left(1 - \frac{\sin(tx)}{tx}\right) dF(x) \\ &\ge \int_{\substack{x:\\ |tx| \ge 1}} \left(1 - \frac{\sin(tx)}{tx}\right) dF(x)\end{aligned}$$

since $1 - (tx)^{-1}\sin(tx) \ge 0$ if $|tx| < 1$. Also, $\sin(tx) \le (tx)\sin 1$ for $|tx| \ge 1$, whence the last integral is at least

$$\int_{\substack{x:\\ |tx| \ge 1}} (1 - \sin 1)\, dF(x) \ge \tfrac{1}{7}\mathbb{P}(|X| \ge t^{-1}).$$

32. It is easily seen that, if $y > 0$ and n is large,

$$\mathbb{P}\big(n(1 - M_n) > y\big) = \mathbb{P}\left(M_n < 1 - \frac{y}{n}\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i < 1 - \frac{y}{n}\right) = \left(1 - \frac{y}{n}\right)^n \to e^{-y}.$$

33. (a) The characteristic function of Y_λ is

$$\psi_\lambda(t) = \mathbb{E}\big\{\exp\big(it(X - \lambda)/\sqrt{\lambda}\big)\big\} = \exp\big\{\lambda(e^{it/\sqrt{\lambda}} - 1) - it\sqrt{\lambda}\big\} = \exp\big\{-\tfrac{1}{2}t^2 + o(1)\big\}$$

as $\lambda \to \infty$. Now use the continuity theorem.

(b) In this case,

$$\psi_\lambda(t) = e^{-it\sqrt{\lambda}}\left(1 - \frac{it}{\sqrt{\lambda}}\right)^{-\lambda},$$

so that, as $\lambda \to \infty$,

$$\log \psi_\lambda(t) = -it\sqrt{\lambda} - \lambda \log\left(1 - \frac{it}{\sqrt{\lambda}}\right) = -it\sqrt{\lambda} + \lambda\left(\frac{it}{\sqrt{\lambda}} - \frac{t^2}{2\lambda} + o(\lambda^{-1})\right) \to -\tfrac{1}{2}t^2.$$

(c) Let Z_n be Poisson with parameter n. By part (a),

$$\mathbb{P}\left(\frac{Z_n - n}{\sqrt{n}} \le 0\right) \to \Phi(0) = \tfrac{1}{2}$$

where Φ is the $N(0,1)$ distribution function. The left hand side equals $\mathbb{P}(Z_n \le n) = \sum_{k=0}^{n} e^{-n} n^k / k!$.

34. If you are in possession of $r-1$ different types, the waiting time for the acquisition of the next new type is geometric with probability generating function

$$G_r(s) = \frac{(n-r+1)s}{n-(r-1)s}.$$

Therefore the characteristic function of $U_n = (T_n - n\log n)/n$ is

$$\psi_n(t) = e^{-it\log n} \prod_{r=1}^{n} G_r(e^{it/n}) = n^{-it} \prod_{r=1}^{n} \left\{ \frac{(n-r+1)e^{it/n}}{n-(r-1)e^{it/n}} \right\} = \frac{n^{-it} n!}{\prod_{r=0}^{n-1}(ne^{-it/n} - r)}.$$

The denominator satisfies

$$\prod_{r=0}^{n-1}(ne^{-it/n} - r) = (1 + \mathrm{o}(1)) \prod_{r=0}^{n-1}(n - it - r)$$

as $n \to \infty$, by expanding the exponential function, and hence

$$\lim_{n\to\infty} \psi_n(t) = \lim_{n\to\infty} \frac{n^{-it} n!}{\prod_{r=0}^{n-1}(n - it - r)} = \Gamma(1 - it),$$

where we have used Euler's product for the gamma function:

$$\frac{n!\, n^z}{\prod_{r=0}^{n}(z + r)} \to \Gamma(z) \quad \text{as } n \to \infty$$

the convergence being uniform on any region of the complex plane containing no singularity of Γ. The claim now follows by the result of Problem (5.12.27c).

35. Let X_n be uniform on $[-n, n]$, with characteristic function

$$\phi_n(t) = \int_{-n}^{n} \frac{1}{2n} e^{itx}\, dx = \begin{cases} \dfrac{\sin(nt)}{nt} & \text{if } t \ne 0, \\ 1 & \text{if } t = 0. \end{cases}$$

It follows that, as $n \to \infty$, $\phi_n(t) \to \delta_{0t}$, the Kronecker delta. The limit function is discontinuous at $t = 0$ and is therefore not itself a characteristic function.

36. Let $G_i(s)$ be the probability generating function of the number shown by the ith die, and suppose that

$$G_1(s)G_2(s) = \sum_{k=2}^{12} \tfrac{1}{11} s^k = \frac{s^2(1 - s^{11})}{11(1 - s)},$$

so that $1 - s^{11} = 11(1 - s)H_1(s)H_2(s)$ where $H_i(s) = s^{-1}G_i(s)$ is a real polynomial of degree 5. However

$$1 - s^{11} = (1 - s) \prod_{k=1}^{5} (\omega_k - s)(\overline{\omega}_k - s)$$

where $\omega_1, \overline{\omega}_1, \dots, \omega_5, \overline{\omega}_5$ are the ten complex eleventh roots of unity. The ω_k come in conjugate pairs, and therefore no five of the ten terms in $\prod_{k=1}^{5}(\omega_k - s)(\overline{\omega}_k - s)$ have a product which is a real polynomial. This is a contradiction.

37. (a) Let H and T be the numbers of heads and tails. The joint probability generating function of H and T is

$$G_{H,T}(s,t) = \mathbb{E}(s^H t^T) = \mathbb{E}(s^H t^{N-H}) = \mathbb{E}\{\mathbb{E}((s/t)^H t^N \mid N)\} = \mathbb{E}\left\{t^N \left(q + \frac{ps}{t}\right)^N\right\}$$

where $p = 1 - q$ is the probability of heads on each throw. Hence

$$G_{H,T}(s,t) = G_N(qt + ps) = \exp\{\lambda(qt + ps - 1)\}.$$

It follows that

$$G_H(s) = G_{H,T}(s,1) = e^{\lambda p(s-1)}, \qquad G_T(t) = G_{H,T}(1,t) = e^{\lambda q(s-1)},$$

so that $G_{H,T}(s,t) = G_H(s)G_T(t)$, whence H and T are independent.

(b) Suppose conversely that H and T are independent, and write G for the probability generating function of N. From the above calculation, $G_{H,T}(s,t) = G(qt + ps)$, whence $G_H(s) = G(q + ps)$ and $G_T(t) = G(qt + p)$, so that $G(qt + ps) = G(q + ps)G(qt + p)$ for all appropriate s, t. Write $f(x) = G(1 - x)$ to obtain $f(x + y) = f(x)f(y)$, valid at least for all $0 \le x, y \le \min\{p, q\}$. The only continuous solutions to this functional equation which satisfy $f(0) = 1$ are of the form $f(x) = e^{\mu x}$ for some μ, whence it is immediate that $G(x) = e^{\lambda(x-1)}$ where $\lambda = -\mu$.

38. The number of such paths π containing exactly n nodes is 2^{n-1}, and each such π satisfies $\mathbb{P}(B(\pi) \ge k) = \mathbb{P}(S_n \ge k)$ where $S_n = Y_1 + Y_2 + \cdots + Y_n$ is the sum of n independent Bernoulli variables having parameter p $(= 1 - q)$. Therefore $\mathbb{E}\{X_n(k)\} = 2^{n-1}\mathbb{P}(S_n \ge k)$. We set $k = n\beta$, and need to estimate $\mathbb{P}(S_n \ge n\beta)$. It is a consequence of the large deviation theorem (5.11.4) that, if $p \le \beta < 1$,

$$\mathbb{P}(S_n \ge n\beta)^{1/n} \to \inf_{t>0}\left\{e^{-t\beta}M(t)\right\}$$

where $M(t) = \mathbb{E}(e^{tY_1}) = (q + pe^t)$. With the aid of a little calculus, we find that

$$\mathbb{P}(S_n \ge n\beta)^{1/n} \to \left(\frac{p}{\beta}\right)^{\beta}\left(\frac{1-p}{1-\beta}\right)^{1-\beta}, \qquad p \le \beta < 1.$$

Hence

$$\mathbb{E}\{X_n(\beta n)\} \to \begin{cases} 0 & \text{if } \gamma(\beta) < 1, \\ \infty & \text{if } \gamma(\beta) > 1, \end{cases}$$

where

$$\gamma(\beta) = 2\left(\frac{p}{\beta}\right)^{\beta}\left(\frac{1-p}{1-\beta}\right)^{1-\beta}$$

is a decreasing function of β. If $p < \frac{1}{2}$, there is a unique $\beta_c \in [p, 1)$ such that $\gamma(\beta_c) = 1$; if $p \ge \frac{1}{2}$ then $\gamma(\beta) > 1$ for all $\beta \in [p, 1)$ so that we may take $\beta_c = 1$.

Turning to the final part,

$$\mathbb{P}\big(X_n(\beta n) \ge 1\big) \le \mathbb{E}\{X_n(\beta n)\} \to 0 \qquad \text{if } \beta > \beta_c.$$

As for the other case, we shall make use of the inequality

$$(*) \qquad \mathbb{P}(N \ne 0) \ge \frac{\mathbb{E}(N)^2}{\mathbb{E}(N^2)}$$

for any N taking values in the non-negative integers. This is easily proved: certainly

$$\operatorname{var}(N \mid N \neq 0) = \mathbb{E}(N^2 \mid N \neq 0) - \mathbb{E}(N \mid N \neq 0)^2 \geq 0,$$

whence

$$\frac{\mathbb{E}(N^2)}{\mathbb{P}(N \neq 0)} \geq \frac{\mathbb{E}(N)^2}{\mathbb{P}(N \neq 0)^2}.$$

We have that $\mathbb{E}\{X_n(\beta n)^2\} = \sum_{\pi,\rho} \mathbb{E}(I_\pi I_\rho)$ where the sum is over all such paths π, ρ, and I_π is the indicator function of the event $\{B(\pi) \geq \beta n\}$. Hence

$$\mathbb{E}\{X_n(\beta n)^2\} = \sum_\pi \mathbb{E}(I_\pi) + \sum_{\pi \neq \rho} \mathbb{E}(I_\pi I_\rho) = \mathbb{E}\{X_n(\beta n)\} + 2^{n-1} \sum_{\rho \neq L} \mathbb{E}(I_L I_\rho)$$

where L is the path which always takes the left fork (there are 2^{n-1} choices for π, and by symmetry each provides the same contribution to the sum). We divide up the last sum according to the number of nodes in common to ρ and L, obtaining $\sum_{m=1}^{n-1} 2^{n-m-1}\mathbb{E}(I_L I_M)$ where M is a path having exactly m nodes in common with L. Now

$$\mathbb{E}(I_L I_M) = \mathbb{E}(I_M \mid I_L = 1)\mathbb{E}(I_L) \leq \mathbb{P}(T_{n-m} \geq \beta n - m)\mathbb{E}(I_L)$$

where T_{n-m} has the $\operatorname{bin}(n-m, p)$ distribution (the 'most value' to I_M of the event $\{I_L = 1\}$ is obtained when all m nodes in $L \cap M$ are black). However

$$\mathbb{E}(I_M) = \mathbb{P}(T_n \geq \beta n) \geq p^m \mathbb{P}(T_{n-m} \geq \beta n - m)$$

so that $\mathbb{E}(I_L I_M) \leq p^{-m}\mathbb{E}(I_L)\mathbb{E}(I_M)$. It follows that $N = X_n(\beta n)$ satisfies

$$\mathbb{E}(N^2) \leq \mathbb{E}(N) + 2^{n-1} \sum_{m=1}^{n-1} 2^{n-m-1} \cdot \frac{1}{p^m}\mathbb{E}(I_L)\mathbb{E}(I_M) = \mathbb{E}(N) + \tfrac{1}{2}\mathbb{E}(N)^2 \sum_{m=1}^{n-1} \left(\frac{1}{2p}\right)^m$$

whence, by $(*)$,

$$\mathbb{P}(N \neq 0) \geq \frac{1}{\mathbb{E}(N)^{-1} + \frac{1}{2}\sum_{m=1}^{n-1}(2p)^{-m}}.$$

If $\beta < \beta_c$ then $\mathbb{E}(N) \to \infty$ as $n \to \infty$. It is immediately evident that $\mathbb{P}(N \neq 0) \to 1$ if $p \leq \frac{1}{2}$. Suppose finally that $p > \frac{1}{2}$ and $\beta < \beta_c$. By the above inequality,

$$(**) \qquad \mathbb{P}\big(X_n(\beta n) > 0\big) \geq c(\beta) \qquad \text{for all } n$$

where $c(\beta)$ is some positive constant. Find $\epsilon > 0$ such that $\beta + \epsilon < \beta_c$. Fix a positive integer m, and let $\mathcal{P}_m$ be a collection of 2^m disjoint paths each of length $n - m$ starting from depth m in the tree. Now

$$\mathbb{P}\big(X_n(\beta n) = 0\big) \leq \mathbb{P}\big(B(\nu) < \beta n \text{ for all } \nu \in \mathcal{P}_m\big) = \mathbb{P}\big(B(\nu) < \beta n\big)^{2^m}$$

where $\nu \in \mathcal{P}_m$. However

$$\mathbb{P}\big(B(\nu) < \beta n\big) \leq \mathbb{P}\big(B(\nu) < (\beta + \epsilon)(n - m)\big)$$

if $\beta n < (\beta + \epsilon)(n - m)$, which is to say that $n \geq (\beta + \epsilon)m/\epsilon$. Hence, for all large n,

$$\mathbb{P}\big(X_n(\beta n) = 0\big) \leq \{1 - c(\beta + \epsilon)\}^{2^m}$$

by $(**)$; we let $n \to \infty$ and $m \to \infty$ in that order, to obtain $\mathbb{P}(X_n(\beta n) = 0) \to 0$ as $n \to \infty$.

39. (a) The characteristic function of X_n satisfies

$$\mathbb{E}(e^{itX_n}) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^{it}\right)^n = \left(1 + \frac{\lambda}{n}[e^{it} - 1]\right)^n \to \exp\left(\lambda[e^{it} - 1]\right),$$

the characteristic function of the Poisson distribution.

(b) Similarly,

$$\mathbb{E}(e^{itY_n/n}) = \frac{pe^{it/n}}{1 - (1-p)e^{it/n}} \to \frac{\lambda}{\lambda - it}$$

as $n \to \infty$, the limit being the characteristic function of the exponential distribution.

40. If you cannot follow the hints, take a look at one or more of the following: Moran 1968 (p. 389), Breiman 1968 (p. 186), Loève 1977 (p. 287), Laha and Rohatgi 1979 (p. 288).

41. With $Y_k = kX_k$, we have that $\mathbb{E}(Y_k) = 0$, $\operatorname{var}(Y_k) = k^2$, $\mathbb{E}|Y_k^3| = k^3$. Note that $S_n = Y_1 + Y_2 + \cdots + Y_n$ is such that

$$\frac{1}{\{\operatorname{var}(S_n)\}^{3/2}} \sum_{k=1}^{n} \mathbb{E}|Y_k^3| \sim c\frac{n^4}{n^{9/2}} \to 0$$

as $n \to \infty$, where c is a positive constant. Applying the central limit theorem ((5.10.5) or Problem (5.12.40)), we find that

$$\frac{S_n}{\sqrt{\operatorname{var} S_n}} \xrightarrow{\mathrm{D}} N(0,1), \qquad \text{as } n \to \infty,$$

where $\operatorname{var} S_n = \sum_{k=1}^{n} k^2 \sim \frac{1}{3}n^3$ as $n \to \infty$.

42. We may suppose that $\mu = 0$ and $\sigma = 1$; if this is not so, then replace X_i by $Y_i = (X_i - \mu)/\sigma$. Let $\mathbf{t} = (t_0, t_1, t_2, \dots, t_n) \in \mathbb{R}^{n+1}$, and set $\bar{t} = n^{-1}\sum_{j=1}^{n} t_j$. The joint characteristic function of the $n+1$ variables $\overline{X}, Z_1, Z_2, \dots, Z_n$ is

$$\begin{aligned}\phi(\mathbf{t}) &= \mathbb{E}\left\{\exp\left(it_0\overline{X} + \sum_{j=1}^{n} it_jZ_j\right)\right\} = \mathbb{E}\left\{\prod_{j=1}^{n} \exp\left(i\left[\frac{t_0}{n} + t_j - \bar{t}\right]X_j\right)\right\} \\ &= \prod_{j=1}^{n} \exp\left(-\frac{1}{2}\left[\frac{t_0}{n} + t_j - \bar{t}\right]^2\right)\end{aligned}$$

by independence. Hence

$$\phi(\mathbf{t}) = \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left[\frac{t_0}{n} + (t_j - \bar{t})\right]^2\right) = \exp\left\{-\frac{t_0^2}{2n} - \frac{1}{2}\sum_{j=1}^{n}(t_j - \bar{t})^2\right\}$$

where we have used the fact that $\sum_{j=1}^{n}(t_j - \bar{t}) = 0$. Therefore

$$\phi(\mathbf{t}) = \mathbb{E}(e^{it_0\overline{X}})\mathbb{E}\left(\exp\left\{i\sum_{1}^{n}(t_j - \bar{t})X_j\right\}\right) = \mathbb{E}(e^{it_0\overline{X}})\mathbb{E}\left(\exp\left\{i\sum_{1}^{n}t_jZ_j\right\}\right),$$

whence $\overline{X}$ is independent of the collection $Z_1, Z_2, \dots, Z_n$. It follows that $\overline{X}$ is independent of $S^2 = (n-1)^{-1}\sum_{j=1}^{n} Z_j^2$. Compare with Exercise (4.10.5).

43. (i) Clearly, $\mathbb{P}(Y \le y) = \mathbb{P}(X \le \log y) = \Phi(\log y)$ for $y > 0$, where Φ is the $N(0, 1)$ distribution function. The density function of Y follows by differentiating.
(ii) We have that $f_a(x) \ge 0$ if $|a| \le 1$, and

$$\int_0^\infty a \sin(2\pi \log x) \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\log x)^2}\, dx = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} a \sin(2\pi y) e^{-\frac{1}{2}y^2}\, dy = 0$$

since sine is an odd function. Therefore $\int_{-\infty}^\infty f_a(x)\, dx = 1$, so that each such f_a is a density function.

For any positive integer k, the kth moment of f_a is $\int_{-\infty}^\infty x^k f(x)\, dx + I_a(k)$ where

$$I_a(k) = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} a \sin(2\pi y) e^{ky - \frac{1}{2}y^2}\, dy = 0$$

since the integrand is an odd function of $y - k$. It follows that each f_a has the same moments as f.

44. Here is one way of proving this. Let $X_1, X_2, \dots$ be the steps of the walk, and let S_n be the position of the walk after the nth step. Suppose $\mu = \mathbb{E}(X_1)$ satisfies $\mu < 0$, and let $e_m = \mathbb{P}(S_n = 0 \text{ for some } n \ge 1 \mid S_0 = -m)$ where $m > 0$. Then $e_m \le \sum_{n=1}^\infty \mathbb{P}(T_n > m)$ where $T_n = X_1 + X_2 + \dots + X_n = S_n - S_0$. Now, for $t > 0$,

$$\mathbb{P}(T_n > m) = \mathbb{P}(T_n - n\mu > m - n\mu) \le e^{-t(m-n\mu)}\mathbb{E}(e^{t(T_n - n\mu)}) = e^{-tm}\left\{e^{t\mu}M(t)\right\}^n$$

where $M(t) = \mathbb{E}(e^{t(X_1 - \mu)})$. Now $M(t) = 1 + \mathrm{O}(t^2)$ as $t \to 0$, and therefore there exists t (> 0) such that $\theta(t) = e^{t\mu}M(t) < 1$ (remember that $\mu < 0$). With this choice of t, $e_m \le \sum_{n=1}^\infty e^{-tm}\theta(t)^n \to 0$ as $m \to \infty$, whence there exists K such that $e_m < \frac{1}{2}$ for $m \ge K$.

Finally, there exist $\delta, \epsilon > 0$ such that $\mathbb{P}(X_1 < -\delta) > \epsilon$, implying that $\mathbb{P}(S_N < -K \mid S_0 = 0) > \epsilon^N$ where $N = \lceil K/\delta \rceil$, and therefore

$$\mathbb{P}(S_n \ne 0 \text{ for all } n \ge 1 \mid S_0 = 0) \ge (1 - e_K)\epsilon^N \ge \tfrac{1}{2}\epsilon^N;$$

therefore the walk is transient. This proof may be shortened by using the Borel–Cantelli lemma.

45. Obviously,

$$L = \begin{cases} a & \text{if } X_1 > a, \\ X_1 + \widetilde{L} & \text{if } X_1 \le a, \end{cases}$$

where $\widetilde{L}$ has the same distribution as L. Therefore,

$$\mathbb{E}(s^L) = s^a \mathbb{P}(X_1 > a) + \sum_{r=1}^a s^r \mathbb{E}(s^L)\mathbb{P}(X_1 = r).$$

46. We have that

$$W_n = \begin{cases} W_{n-1} + 1 & \text{with probability } p, \\ W_{n-1} + 1 + \widetilde{W_n} & \text{with probability } q, \end{cases}$$

where $\widetilde{W_n}$ is independent of W_{n-1} and has the same distribution as W_n. Hence $G_n(s) = psG_{n-1}(s) + qsG_{n-1}(s)G_n(s)$. Now $G_0(s) = 1$, and the recurrence relation may be solved by induction. (Alternatively use Problem (5.12.45) with appropriate X_i.)

47. Let W_r be the number of flips until you first see r consecutive heads, so that $\mathbb{P}(L_n < r) = \mathbb{P}(W_r > n)$. Hence,

$$1 + \sum_{n=1}^\infty s^n \mathbb{P}(L_n < r) = \sum_{n=0}^\infty s^n \mathbb{P}(W > n) = \frac{1 - \mathbb{E}(s^{W_r})}{1 - s},$$

where $\mathbb{E}(s^{W_r}) = G_r(s)$ is given in Problem (5.12.46).

48. We have that
$$X_{n+1} = \begin{cases} \frac{1}{2}X_n & \text{with probability } \frac{1}{2}, \\ \frac{1}{2}X_n + Y_n & \text{with probability } \frac{1}{2}. \end{cases}$$
Hence the characteristic functions satisfy
$$\begin{aligned}\phi_{n+1}(t) &= \mathbb{E}(e^{itX_{n+1}}) = \tfrac{1}{2}\phi_n(\tfrac{1}{2}t) + \tfrac{1}{2}\phi_n(\tfrac{1}{2}t)\frac{\lambda}{\lambda - it} \\ &= \phi_n(\tfrac{1}{2}t)\frac{\lambda - \frac{1}{2}it}{\lambda - it} = \phi_{n-1}(\tfrac{1}{4}t)\frac{\lambda - \frac{1}{4}it}{\lambda - it} = \cdots = \phi_1(t2^{-n})\frac{\lambda - it2^{-n}}{\lambda - it} \to \frac{\lambda}{\lambda - it}\end{aligned}$$
as $n \to \infty$. The limiting distribution is exponential with parameter λ.

49. We have that
$$\int_0^1 G(s)\,ds = \mathbb{E}\left(\int_0^1 s^X\,ds\right) = \mathbb{E}\left(\left.\frac{s^{X+1}}{X+1}\right|_0^1\right) = \mathbb{E}\left(\frac{1}{X+1}\right).$$
(a) $(1-e^{-\lambda})/\lambda$, (b) $-(p/q^2)(q+\log p)$, (c) $(1-q^{n+1})/[(n+1)p]$, (d) $-[1+(p/q)\log p]/\log p$.
(e) Not if $\mathbb{P}(X+1>0)=1$, by Jensen's inequality (see Exercise (5.6.1)) and the strict concavity of the function $f(x) = 1/x$. If $X+1$ is permitted to be negative, consider the case when $\mathbb{P}(X+1 = -1) = \mathbb{P}(X+1=1) = \frac{1}{2}$.

50. By compounding, as in Theorem (5.1.25), the sum has characteristic function
$$G_N(\phi_X(t)) = \frac{p\phi_X(t)}{1 - q\phi_X(t)} = \frac{\lambda p}{\lambda p - it},$$
whence the sum is exponentially distributed with parameter λp.

51. Consider the function $G(x) = \{\mathbb{E}(X^2)\}^{-1}\int_{-\infty}^x y^2\,dF(y)$. This function is right-continuous and increases from 0 to 1, and is therefore a distribution function. Its characteristic function is
$$\int_{-\infty}^{\infty} \frac{e^{itx}}{\mathbb{E}(X^2)}x^2\,dF(x) = -\frac{1}{\mathbb{E}(X^2)}\frac{d^2}{dt^2}\phi(t).$$

52. By integration, $f_X(x) = f_Y(y) = \frac{1}{2}$, $|x|<1$, $|y|<1$. Since $f(x,y) \neq f_X(x)f_Y(y)$, X and Y are not independent. Now,
$$f_{X+Y}(z) = \int_{-1}^1 f(x, z-x)\,dx = \begin{cases} \frac{1}{4}(z+2) & \text{if } -2<z<0, \\ \frac{1}{4}(2-z) & \text{if } 0<z<2, \end{cases}$$
the 'triangular' density function on $(-2,2)$. This is the density function of the sum of two independent random variables uniform on $(-1,1)$.

6

Markov chains

6.1 Solutions. Markov processes

1. The sequence $X_1, X_2, \dots$ of independent random variables satisfies

$$\mathbb{P}(X_{n+1} = j \mid X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = j),$$

whence the sequence is a Markov chain. The chain is homogeneous if the X_i are identically distributed.

2. (a) With Y_n the outcome of the nth throw, $X_{n+1} = \max\{X_n, Y_{n+1}\}$, so that

$$p_{ij} = \begin{cases} 0 & \text{if } j < i \\ \frac{1}{6}i & \text{if } j = i \\ \frac{1}{6} & \text{if } j > i, \end{cases}$$

for $1 \le i, j \le 6$. Similarly,

$$p_{ij}(n) = \begin{cases} 0 & \text{if } j < i \\ (\frac{1}{6}i)^n & \text{if } j = i. \end{cases}$$

If $j > i$, then $p_{ij}(n) = \mathbb{P}(Z_n = j)$, where $Z_n = \max\{Y_1, Y_2, \dots, Y_n\}$, and an elementary calculation yields

$$p_{ij}(n) = \left(\frac{j}{6}\right)^n - \left(\frac{j-1}{6}\right)^n, \qquad i < j \le 6.$$

(b) $N_{n+1} - N_n$ is independent of $N_1, N_2, \dots, N_n$, so that N is Markovian with

$$p_{ij} = \begin{cases} \frac{1}{6} & \text{if } j = i + 1, \\ \frac{5}{6} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

(c) The evolution of C is given by

$$C_{r+1} = \begin{cases} 0 & \text{if the die shows 6,} \\ C_r + 1 & \text{otherwise,} \end{cases}$$

whence C is Markovian with

$$p_{ij} = \begin{cases} \frac{1}{6} & j = 0, \\ \frac{5}{6} & j = i + 1, \\ 0 & \text{otherwise.} \end{cases}$$

(d) This time,

$$B_{r+1} = \begin{cases} B_r - 1 & \text{if } B_r > 0, \\ Y_r & \text{if } B_r = 0, \end{cases}$$

where Y_r is a geometrically distributed random variable with parameter $\frac{1}{6}$, independent of the sequence $B_0, B_2, \cdots, B_r$. Hence B is Markovian with

$$p_{ij} = \begin{cases} 1 & \text{if } j = i - 1 \geq 0, \\ (\frac{5}{6})^{j-1}\frac{1}{6} & \text{if } i = 0,\ j \geq 1. \end{cases}$$

3. (i) If $X_n = i$, then $X_{n+1} \in \{i-1, i+1\}$. Now, for $i \geq 1$,

$(*)$

$$\begin{aligned} \mathbb{P}(X_{n+1} = i+1 \mid X_n = i, B) &= \mathbb{P}(X_{n+1} = i+1 \mid S_n = i, B)\mathbb{P}(S_n = i \mid X_n = i, B) \\ &\quad + \mathbb{P}(X_{n+1} = i+1 \mid S_n = -i, B)\mathbb{P}(S_n = -i \mid X_n = i, B) \end{aligned}$$

where $B = \{X_r = i_r \text{ for } 0 \leq r < n\}$ and $i_0, i_1, \ldots, i_{n-1}$ are integers. Clearly

$$\mathbb{P}(X_{n+1} = i+1 \mid S_n = i, B) = p, \quad \mathbb{P}(X_{n+1} = i+1 \mid S_n = -i, B) = q,$$

where $p\ (= 1 - q)$ is the chance of a rightward step. Let l be the time of the last visit to 0 prior to the time n, $l = \max\{r : i_r = 0\}$. During the time-interval $(l, n]$, the path lies entirely in either the positive integers or the negative integers. If the former, it is required to follow the route prescribed by the event $B \cap \{S_n = i\}$, and if the latter by the event $B \cap \{S_n = -i\}$. The absolute probabilities of these two routes are

$$\pi_1 = p^{\frac{1}{2}(n-l+i)} q^{\frac{1}{2}(n-l-i)}, \quad \pi_2 = p^{\frac{1}{2}(n-l-i)} q^{\frac{1}{2}(n-l+i)},$$

whence

$$\mathbb{P}(S_n = i \mid X_n = i, B) = \frac{\pi_1}{\pi_1 + \pi_2} = \frac{p^i}{p^i + q^i} = 1 - \mathbb{P}(S_n = -i \mid X_n = i, B).$$

Substitute into $(*)$ to obtain

$$\mathbb{P}(X_{n+1} = i+1 \mid X_n = i, B) = \frac{p^{i+1} + q^{i+1}}{p^i + q^i} = 1 - \mathbb{P}(X_{n+1} = i-1 \mid X_n = i, B).$$

Finally $\mathbb{P}(X_{n+1} = 1 \mid X_n = 0, B) = 1$.

(ii) If $Y_n > 0$, then $Y_n - Y_{n+1}$ equals the $(n+1)$th step, a random variable which is independent of the past history of the process. If $Y_n = 0$ then $S_n = M_n$, so that Y_{n+1} takes the values 0 and 1 with respective probabilities p and q, independently of the past history. Therefore Y is a Markov chain with transition probabilities

$$\text{for } i > 0, \quad p_{ij} = \begin{cases} p & \text{if } j = i-1 \\ q & \text{if } j = i+1, \end{cases} \qquad p_{0j} = \begin{cases} p & \text{if } j = 0 \\ q & \text{if } j = 1. \end{cases}$$

The sequence Y is a random walk with a retaining barrier at 0.

4. For any sequence $i_0, i_1, \ldots$ of states,

$$\begin{aligned} \mathbb{P}(Y_{k+1} = i_{k+1} \mid Y_r = i_r \text{ for } 0 \leq r \leq k) &= \frac{\mathbb{P}(X_{n_s} = i_s \text{ for } 0 \leq s \leq k+1)}{\mathbb{P}(X_{n_s} = i_s \text{ for } 0 \leq s \leq k)} \\ &= \frac{\prod_{s=0}^{k} p_{i_s, i_{s+1}}(n_{s+1} - n_s)}{\prod_{s=0}^{k-1} p_{i_s, i_{s+1}}(n_{s+1} - n_s)} \\ &= p_{i_k, i_{k+1}}(n_{k+1} - n_k) = \mathbb{P}(Y_{k+1} = i_{k+1} \mid Y_k = i_k), \end{aligned}$$

where $p_{ij}(n)$ denotes the appropriate n-step transition probability of X.
(a) With the usual notation, the transition matrix of Y is

$$\pi_{ij} = \begin{cases} p^2 & \text{if } j = i+2, \\ 2pq & \text{if } j = i, \\ q^2 & \text{if } j = i-2. \end{cases}$$

(b) With the usual notation, the transition probability π_{ij} is the coefficient of s^j in $G(G(s))^i$.

5. Writing $\mathbf{X} = (X_1, X_2, \dots, X_n)$, we have that

$$\mathbb{P}\big(F \mid I(\mathbf{X}) = 1, X_n = i\big) = \frac{\mathbb{P}\big(F, I(\mathbf{X}) = 1, X_n = i\big)}{\mathbb{P}\big(I(\mathbf{X}) = 1, X_n = i\big)}$$

where F is any event defined in terms of $X_n, X_{n+1}, \dots$. Let A be the set of all sequences $\mathbf{x} = (x_1, x_2, \dots, x_{n-1}, i)$ of states such that $I(\mathbf{x}) = 1$. Then

$$\mathbb{P}\big(F, I(\mathbf{X}) = 1, X_n = i\big) = \sum_{\mathbf{x}\in A} \mathbb{P}(F, \mathbf{X} = \mathbf{x}) = \mathbb{P}(F \mid X_n = i) \sum_{\mathbf{x}\in A} \mathbb{P}(\mathbf{X} = \mathbf{x})$$

by the Markov property. Divide through by the final summation to obtain $\mathbb{P}\big(F \mid I(\mathbf{X}) = 1, X_n = i\big) = \mathbb{P}(F \mid X_n = i)$.

6. Let $H_n = \{X_k = x_k \text{ for } 0 \le k < n,\ X_n = i\}$. The required probability may be written as

$$\frac{\mathbb{P}(X_{T+m} = j, H_T)}{\mathbb{P}(H_T)} = \frac{\sum_n \mathbb{P}(X_{T+m} = j, H_T, T = n)}{\mathbb{P}(H_T)}.$$

Now $\mathbb{P}(X_{T+m} = j \mid H_T, T = n) = \mathbb{P}(X_{n+m} = j \mid H_n, T = n)$. Let I be the indicator function of the event $H_n \cap \{T = n\}$, an event which depends only upon the values of $X_1, X_2, \dots, X_n$. Using the result of Exercise (6.1.5),

$$\mathbb{P}(X_{n+m} = j \mid H_n, T = n) = \mathbb{P}(X_{n+m} = j \mid X_n = i) = p_{ij}(m).$$

Hence

$$\mathbb{P}(X_{T+m} = j \mid H_T) = \frac{p_{ij}(m) \sum_n \mathbb{P}(H_n, T = n)}{\mathbb{P}(H_T)} = p_{ij}(m).$$

7. Clearly

$$\mathbb{P}(Y_{n+1} = j \mid Y_r = i_r \text{ for } 0 \le r \le n) = \mathbb{P}(X_{n+1} = b \mid X_r = a_r \text{ for } 0 \le r \le n)$$

where $b = h^{-1}(j)$, $a_r = h^{-1}(i_r)$; the claim follows by the Markov property of X.

It is easy to find an example in which h is not one-one, for which X is a Markov chain but Y is not. The first part of Exercise (6.1.3) describes such a case if $S_0 \ne 0$.

8. Not necessarily! Take as example the chains S and Y of Exercise (6.1.3). The sum is $S_n + Y_n = M_n$, which is not a Markov chain.

9. All of them. (a) Using the Markov property of X,

$$\mathbb{P}(X_{m+r} = k \mid X_m = i_m, \dots, X_{m+r-1} = i_{m+r-1}) = \mathbb{P}(X_{m+r} = k \mid X_{m+r-1} = i_{m+r-1}).$$

(b) Let $\{\text{even}\} = \{X_{2r} = i_{2r} \text{ for } 0 \le r \le m\}$ and $\{\text{odd}\} = \{X_{2r+1} = i_{2r+1} \text{ for } 0 \le r \le m-1\}$. Then,

$$\begin{aligned}\mathbb{P}(X_{2m+2} = k \mid \text{even}) &= \sum{}' \frac{\mathbb{P}(X_{2m+2} = k, X_{2m+1} = i_{2m+1}, \text{ even, odd})}{\mathbb{P}(\text{even})} \\ &= \sum{}' \frac{\mathbb{P}(X_{2m+2} = k, X_{2m+1} = i_{2m+1} \mid X_{2m} = i_{2m})\mathbb{P}(\text{even, odd})}{\mathbb{P}(\text{even})} \\ &= \mathbb{P}(X_{2m+2} = k \mid X_{2m} = i_{2m}),\end{aligned}$$

where the sum is taken over all possible values of i_s for odd s.
(c) With $Y_n = (X_n, X_{n+1})$,

$$\begin{aligned}\mathbb{P}\big(Y_{n+1} = (k,l) \,\big|\, Y_0 = (i_0, i_1), \dots, Y_n = (i_n, k)\big) &= \mathbb{P}\big(Y_{n+1} = (k,l) \,\big|\, X_{n+1} = k\big) \\ &= \mathbb{P}\big(Y_{n+1} = (k,l) \,\big|\, Y_n = (i_n, k)\big),\end{aligned}$$

by the Markov property of X.

10. We have by Lemma (6.1.8) that, with $\mu_j^{(i)} = \mathbb{P}(X_i = j)$,

$$\text{LHS} = \frac{\mu_{x_1}^{(1)} p_{x_1 x_2} \cdots p_{x_{r-1},k} p_{k,x_{r+1}} \cdots p_{x_{n-1}x_n}}{\mu_{x_1}^{(1)} \cdots p_{x_{r-1}x_{r+1}}(2) \cdots p_{x_{n-1}x_n}} = \frac{\mu_{x_1}^{(1)} p_{x_{r-1},k} p_{k,x_{r+1}}}{\mu_{x_1}^{(1)} p_{x_{r-1}x_{r+1}}(2)} = \text{RHS}.$$

11. (a) Since $S_{n+1} = S_n + X_{n+1}$, a sum of independent random variables, S is a Markov chain.
(b) We have that

$$\mathbb{P}(Y_{n+1} = k \mid Y_i = x_i + x_{i-1} \text{ for } 1 \le i \le n) = \mathbb{P}(Y_{n+1} = k \mid X_n = x_n)$$

by the Markov property of X. However, conditioning on X_n is not generally equivalent to conditioning on $Y_n = X_n + X_{n-1}$, so Y does not generally constitute a Markov chain.
(c) $Z_n = nX_1 + (n-1)X_2 + \cdots + X_n$, so Z_{n+1} is the sum of X_{n+1} and a certain linear combination of $Z_1, Z_2, \dots, Z_n$, and so cannot be Markovian.
(d) Since $S_{n+1} = S_n + X_{n+1}$, $Z_{n+1} = Z_n + S_n + X_{n+1}$, and X_{n+1} is independent of $X_1, \dots, X_n$, this is a Markov chain.

12. With $\mathbf{1}$ a row vector of 1's, a matrix $\mathbf{P}$ is stochastic (respectively, doubly stochastic, sub-stochastic) if $\mathbf{P}\mathbf{1}' = \mathbf{1}$ (respectively, $\mathbf{1P} = \mathbf{1}, \mathbf{P1}' \le \mathbf{1}$, with inequalities interpreted coordinatewise). By recursion, $\mathbf{P}$ satisfies any of these equations if and only if $\mathbf{P}^n$ satisfies the same equation.

6.2 Solutions. Classification of states

1. Let A_k be the event that the last visit to i, prior to n, took place at time k. Suppose that $X_0 = i$, so that $A_0, A_1, \dots, A_{n-1}$ form a partition of the sample space. It follows, by conditioning on the A_i, that

$$p_{ij}(n) = \sum_{k=0}^{n-1} p_{ii}(k) l_{ij}(n-k)$$

for $i \ne j$. Multiply by s^n and sum over n (≥ 1) to obtain $P_{ij}(s) = P_{ii}(s)L_{ij}(s)$ for $i \ne j$. Now $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$ if $i \ne j$, so that $F_{ij}(s) = L_{ij}(s)$ whenever $P_{ii}(s) = P_{jj}(s)$.

As examples of chains for which $P_{ii}(s)$ does not depend on i, consider a simple random walk on the integers, or a symmetric random walk on a complete graph.

2. Let i ($\neq s$) be a state of the chain, and define $n_i = \min\{n : p_{is}(n) > 0\}$. If $X_0 = i$ and $X_{n_i} = s$ then, with probability one, X makes no visit to i during the intervening period $[1, n_i - 1]$; this follows from the minimality of n_i. Now s is absorbing, and hence

$$\mathbb{P}(\text{no return to } i \mid X_0 = i) \geq \mathbb{P}(X_{n_i} = s \mid X_0 = i) > 0.$$

3. Let I_k be the indicator function of the event $\{X_k = i\}$, so that $N = \sum_{k=0}^{\infty} I_k$ is the number of visits to i. Then

$$\mathbb{E}(N) = \sum_{k=0}^{\infty} \mathbb{E}(I_k) = \sum_{k=0}^{\infty} p_{ii}(k)$$

which diverges if and only if i is persistent. There is another argument which we shall encounter in some detail when solving Problem (6.15.5).

4. We write $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot \mid X_0 = i)$. One way is as follows, another is via the calculation of Problem (6.15.5). Note that $\mathbb{P}_i(V_j \geq 1) = \mathbb{P}_i(T_j < \infty)$.

(a) We have that

$$\mathbb{P}_i(V_i \geq 2) = \mathbb{P}_i(V_i \geq 2 \mid V_i \geq 1)\mathbb{P}_i(V_i \geq 1) = \mathbb{P}_i(V_i \geq 1)^2$$

by the strong Markov property (Exercise (6.1.6)) applied at the stopping time T_i. By iteration, $\mathbb{P}_i(V_i \geq n) = \mathbb{P}_i(V_i \geq 1)^n$, and allowing $n \to \infty$ gives the result.

(b) Suppose $i \neq j$. For $m \geq 1$,

$$\mathbb{P}_i(V_j \geq m) = \mathbb{P}_i(V_j \geq m \mid T_j < \infty)\mathbb{P}_i(T_j < \infty) = \mathbb{P}_j(V_j \geq m - 1)\mathbb{P}_i(T_j < \infty)$$

by the strong Markov property. Now let $m \to \infty$, and use the result of (a).

5. Let $\theta = \mathbb{P}(T_j < T_i \mid X_0 = i) = \mathbb{P}(T_i < T_j \mid X_0 = j)$, and let N be the number of visits to j before visiting i. Then

$$\mathbb{P}(N \geq 1 \mid X_0 = i) = \mathbb{P}(T_j < T_i \mid X_0 = i) = \theta.$$

Likewise, $\mathbb{P}(N \geq k \mid X_0 = i) = \theta(1 - \theta)^{k-1}$ for $k \geq 1$, whence

$$\mathbb{E}(N \mid X_0 = i) = \sum_{k=1}^{\infty} \theta(1 - \theta)^{k-1} = 1.$$

6.3 Solutions. Classification of chains

1. If $r = 1$, then state i is absorbing for $i \geq 1$; also, 0 is transient unless $a_0 = 1$.

Assume $r < 1$ and let $J = \sup\{j : a_j > 0\}$. The states $0, 1, \ldots, J$ form an irreducible persistent class; they are aperiodic if $r > 0$. All other states are transient. For $0 \leq i \leq J$, the recurrence time T_i of i satisfies $\mathbb{P}(T_i = 1) = r$. If $T_i > 1$ then T_i may be expressed as the sum of

$$\begin{aligned} T_i^{(1)} &:= \text{time to reach 0, given that the first step is leftwards,} \\ T_i^{(2)} &:= \text{time spent in excursions from 0 not reaching } i, \\ T_i^{(3)} &:= \text{time taken to reach } i \text{ in final excursion.} \end{aligned}$$

It is easy to see that $\mathbb{E}(T_i^{(1)}) = 1 + (i-1)/(1-r)$ if $i \geq 1$, since the waiting time at each intermediate point has mean $(1-r)^{-1}$. The number N of such 'small' excursions has mass function $\mathbb{P}(N = n) = \alpha_i(1-\alpha_i)^n$, $n \geq 0$, where $\alpha_i = \sum_{j=i}^{\infty} a_j$; hence $\mathbb{E}(N) = (1-\alpha_i)/\alpha_i$. Each such small excursion has mean duration

$$\sum_{j=0}^{i-1}\left(\frac{j}{1-r}+1\right)\frac{a_j}{1-\alpha_i} = 1 + \sum_{j=0}^{i-1}\frac{ja_j}{(1-\alpha_i)(1-r)}$$

and therefore

$$\mathbb{E}(T_i^{(2)}) = \frac{1}{\alpha_i}\left\{(1-\alpha_i) + \sum_{j=0}^{i-1}\frac{ja_j}{1-r}\right\}.$$

By a similar argument,

$$\mathbb{E}(T_i^{(3)}) = \frac{1}{\alpha_i}\sum_{j=i}^{\infty}\left(1+\frac{j-i}{1-r}\right)a_j.$$

Combining this information, we obtain that

$$\mathbb{E}(T_i) = r + (1-r)\mathbb{E}(T_i^{(1)} + T_i^{(2)} + T_i^{(3)}) = \frac{1}{\alpha_i}\left(1 - r + \sum_{j=0}^{\infty} ja_j\right), \quad i \geq 1,$$

and a similar argument yields $\mathbb{E}(T_0) = 1 + \sum_j ja_j/(1-r)$. The apparent simplicity of these formulae suggests the possibility of an easier derivation; see Exercise (6.4.2). Clearly $\mathbb{E}(T_i) < \infty$ for $i \leq J$ whenever $\sum_j ja_j < \infty$, a condition which certainly holds if $J < \infty$.

2. Assume that $0 < p < 1$. The mean jump-size is $3p - 1$, whence the chain is persistent if and only if $p = \frac{1}{3}$; see Theorem (5.10.17).

3. (a) All states are absorbing if $p = 0$. Assume henceforth that $p \neq 0$. Diagonalize $\mathbf{P}$ to obtain $\mathbf{P} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^{-1}$ where

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{pmatrix},$$

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-2p & 0 \\ 0 & 0 & 1-4p \end{pmatrix}.$$

Therefore

$$\mathbf{P}^n = \mathbf{B}\mathbf{\Lambda}^n\mathbf{B}^{-1} = \mathbf{B}\begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-2p)^n & 0 \\ 0 & 0 & (1-4p)^n \end{pmatrix}\mathbf{B}^{-1}$$

whence $p_{ij}(n)$ is easily found.

In particular,

$$p_{11}(n) = \tfrac{1}{4} + \tfrac{1}{2}(1-2p)^n + \tfrac{1}{4}(1-4p)^n, \quad p_{22}(n) = \tfrac{1}{2} + \tfrac{1}{2}(1-4p)^n,$$

and $p_{33}(n) = p_{11}(n)$ by symmetry.

Now $F_{ii}(s) = 1 - P_{ii}(s)^{-1}$, where

$$P_{11}(s) = P_{33}(s) = \frac{1}{4(1-s)} + \frac{1}{2\{1-s(1-2p)\}} + \frac{1}{4\{1-s(1-4p)\}},$$
$$P_{22}(s) = \frac{1}{2(1-s)} + \frac{1}{2\{1-s(1-4p)\}}.$$

After a little work one obtains the mean recurrence times $\mu_i = F'_{ii}(1)$: $\mu_1 = \mu_3 = 4$, $\mu_2 = 2$.
(b) The chain has period 2 (if $p \neq 0$), and all states are non-null and persistent. By symmetry, the mean recurrence times μ_i are equal. One way of calculating their common value (we shall encounter an easier way in Section 6.4) is to observe that the sequence of visits to any given state j is a renewal process (see Example (5.2.15)). Suppose for simplicity that $p \neq 0$. The times between successive visits to j must be even, and therefore we work on a new time-scale in which one new unit equals two old units. Using the renewal theorem (5.2.24), we obtain

$$p_{ij}(2n) \to \frac{2}{\mu_j} \text{ if } |j-i| \text{ is even}, \qquad p_{ij}(2n+1) \to \frac{2}{\mu_j} \text{ if } |j-i| \text{ is odd;}$$

note that the mean recurrence time of j in the new time-scale is $\frac{1}{2}\mu_j$. Now $\sum_j p_{ij}(m) = 1$ for all m, and so, letting $m = 2n \to \infty$, we find that $4/\mu = 1$ where μ is a typical mean recurrence time.

There is insufficient space here to calculate $p_{ij}(n)$. One way is to diagonalize the transition matrix. Another is to write down a family of difference equations of the form $p_{12}(n) = p \cdot p_{22}(n-1) + (1-p) \cdot p_{42}(n-1)$, and solve them.

4. (a) By symmetry, all states have the same mean-recurrence time. Using the renewal-process argument of the last solution, the common value equals 8, being the number of vertices of the cube. Hence $\mu_v = 8$.

Alternatively, let s be a neighbour of v, and let t be a neighbour of s other than v. In the obvious notation, by symmetry,

$$\mu_v = 1 + \tfrac{3}{4}\mu_{sv}, \qquad \mu_{sv} = 1 + \tfrac{1}{4}\mu_{sv} + \tfrac{1}{2}\mu_{tv},$$
$$\mu_{tv} = 1 + \tfrac{1}{2}\mu_{sv} + \tfrac{1}{4}\mu_{tv} + \tfrac{1}{4}\mu_{wv}, \qquad \mu_{wv} = 1 + \tfrac{1}{4}\mu_{wv} + \tfrac{3}{4}\mu_{tv},$$

a system of equations which may be solved to obtain $\mu_v = 8$.
(b) Using the above equations, $\mu_{wv} = \frac{40}{3}$, whence $\mu_{vw} = \frac{40}{3}$ by symmetry.
(c) The required number X satisfies $\mathbb{P}(X = n) = \theta^{n-1}(1-\theta)^2$ for $n \geq 1$, where θ is the probability that the first return of the walk to its starting point precedes its first visit to the diametrically opposed vertex. Therefore

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} n\theta^{n-1}(1-\theta)^2 = 1.$$

5. (a) Let $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot \mid X_0 = i)$. Since i is persistent,

$$\begin{aligned} 1 = \mathbb{P}_i(V_i = \infty) &= \mathbb{P}_i(V_j = 0,\ V_i = \infty) + \mathbb{P}_i(V_j > 0,\ V_i = \infty) \\ &\leq \mathbb{P}_i(V_j = 0) + \mathbb{P}_i(T_j < \infty,\ V_i = \infty) \\ &\leq 1 - \mathbb{P}_i(T_j < \infty) + \mathbb{P}_i(T_j < \infty)\mathbb{P}_j(V_i = \infty), \end{aligned}$$

by the strong Markov property. Since $i \to j$, we have that $\mathbb{P}_j(V_i = \infty) \geq 1$, which implies $\eta_{ji} = 1$. Also, $\mathbb{P}_i(T_j < \infty) = 1$, and hence $j \to i$ and j is persistent. This implies $\eta_{ij} = 1$.
(b) This is an immediate consequence of Exercise (6.2.4b).

6. Let $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot \mid X_0 = i)$. It is trivial that $\eta_j = 1$ for $j \in A$. For $j \notin A$, condition on the first step and use the Markov property to obtain

$$\eta_j = \sum_{k \in S} p_{jk} \mathbb{P}(T_A < \infty \mid X_1 = k) = \sum_k p_{jk} \eta_k.$$

If $\mathbf{x} = (x_j : j \in S)$ is any non-negative solution of these equations, then $x_j = 1 \geq \eta_j$ for $j \in A$. For $j \notin A$,

$$\begin{aligned} x_j &= \sum_{k \in S} p_{jk} x_k = \sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk} x_k = \mathbb{P}_j(T_A = 1) + \sum_{k \notin A} p_{jk} x_k \\ &= \mathbb{P}_j(T_A = 1) + \sum_{k \notin A} p_{jk} \left\{ \sum_{i \in A} p_{ki} + \sum_{i \notin A} p_{ki} x_i \right\} = \mathbb{P}_j(T_A \leq 2) + \sum_{k \notin A} p_{jk} \sum_{i \notin A} p_{ki} x_i. \end{aligned}$$

We obtain by iteration that, for $j \notin A$,

$$x_j = \mathbb{P}_j(T_A \leq n) + \sum p_{jk_1} p_{k_1 k_2} \cdots p_{k_{n-1}, k_n} x_{k_n} \geq \mathbb{P}(T_A \leq n),$$

where the sum is over all $k_1, k_2, \ldots, k_n \notin A$. We let $n \to \infty$ to find that $x_j \geq \mathbb{P}_j(T_A < \infty) = \eta_j$.

7. The first part follows as in Exercise (6.3.6). Suppose $\mathbf{x} = (x_j : j \in S)$ is a non-negative solution to the equations. As above, for $j \notin A$,

$$\begin{aligned} x_j &= 1 + \sum_k p_{jk} x_k = \mathbb{P}_j(T_A \geq 1) + \sum_{k \notin A} p_{jk} \left(1 + \sum_{i \notin A} p_{ki} x_i \right) \\ &= \mathbb{P}_j(T_A \geq 1) + \mathbb{P}_j(T_A \geq 2) + \cdots + \mathbb{P}_j(T_A \geq n) + \sum p_{jk_1} p_{k_1 k_2} \cdots p_{k_{k-1} k_n} x_{k_n} \\ &\geq \sum_{m=1}^{n} \mathbb{P}(T_A \geq m), \end{aligned}$$

where the penultimate sum is over all paths of length n that do not visit A. We let $n \to \infty$ to obtain that $x_j \geq \mathbb{E}_j(T_A) = \rho_j$.

8. Yes, because the S_r and T_r are stopping times whenever they are finite. Whether or not the exit times are stopping times depends on their exact definition. The times $U_r = \min\{k > U_{r-1} : X_{U_r} \in A,\ X_{U_r+1} \notin A\}$ are not stopping times, but the times $U_r + 1$ are stopping times.

9. (a) Using the aperiodicity of j, there exist integers $r_1, r_2, \ldots, r_s$ having highest common factor 1 and such that $p_{jj}(r_k) > 0$ for $1 \leq k \leq s$. There exists a positive integer M such that, if $r \geq M$, then $r = \sum_{k=1}^{s} a_k r_k$ for some sequence $a_1, a_2, \ldots, a_s$ of non-negative integers. Now, by the Chapman–Kolmogorov equations,

$$p_{jj}(r) \geq \prod_{k=1}^{s} p_{jj}(r_k)^{a_k} > 0,$$

so that $p_{jj}(r) > 0$ for all $r \geq M$.

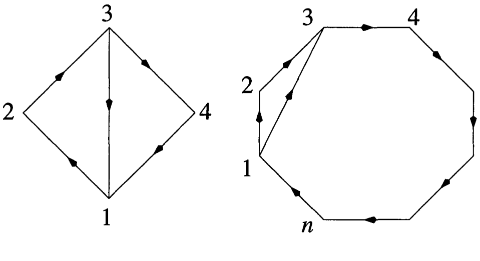
Finally, find m such that $p_{ij}(m) > 0$. Then

$$p_{ij}(r+m) \geq p_{ij}(m) p_{jj}(r) > 0 \qquad \text{if } r \geq M.$$

(b) Since there are only finitely many pairs i, j, the maximum $R(\mathbf{P}) = \max\{N(i,j) : i, j \in S\}$ is finite. Now $R(\mathbf{P})$ depends only on the positions of the non-negative entries in the transition matrix $\mathbf{P}$.

There are only finitely many subsets of entries of $\mathbf{P}$, and so there exists $f(n)$ such that $R(\mathbf{P}) \le f(n)$ for all relevant $n \times n$ transition matrices $\mathbf{P}$.

(c) Consider the two chains with diagrams in the figure beneath. In the case on the left, we have that $p_{11}(5) = 0$, and in the case on the right, we may apply the postage stamp lemma with $a = n$ and $b = n - 1$.



10. Let X_n be the number of green balls after n steps. Let e_j be the probability that X_n is ever zero when $X_0 = j$. By conditioning on the first removal,

$$e_j = \frac{j+2}{2(j+1)} e_{j+1} + \frac{j}{2(j+1)} e_{j-1}, \quad j \ge 1,$$

with $e_0 = 1$. Solving recursively gives

$$(*) \qquad e_j = 1 - (1 - e_1)\left\{1 + \frac{q_1}{p_1} + \cdots + \frac{q_1 q_2 \cdots q_{j-1}}{p_1 p_2 \cdots p_{j-1}}\right\},$$

where

$$p_j = \frac{j+2}{2(j+1)}, \quad q_j = \frac{j}{2(j+1)}.$$

It is easy to see that

$$\sum_{r=0}^{j-1} \frac{q_1 q_2 \cdots q_{j-1}}{p_1 p_2 \cdots p_{j-1}} = 2 - \frac{2}{j+1} \to 2 \quad \text{as } j \to \infty.$$

By the result of Exercise (6.3.6), we seek the minimal non-negative solution (e_j) to $(*)$, which is attained when $2(1 - e_1) = 1$, that is, $e_1 = \frac{1}{2}$. Hence

$$e_j = 1 - \frac{1}{2} \sum_{r=0}^{j-1} \frac{q_1 q_2 \cdots q_{j-1}}{p_1 p_2 \cdots p_{j-1}} = \frac{1}{j+1}.$$

For the second part, let d_j be the expected time until $j - 1$ green balls remain, having started with j green balls and $j + 2$ red. We condition as above to obtain

$$d_j = 1 + \frac{j}{2(j+1)}\{d_{j+1} + d_j\}.$$

We set $e_j = d_j - (2j + 1)$ to find that $(j + 2)e_j = j e_{j+1}$, whence $e_j = \frac{1}{2} j(j + 1)e_1$. The expected time to remove all the green balls is

$$\sum_{j=1}^{n} d_j = \sum_{j=1}^{n} \{e_j + 2(j-1)\} = n(n+2) + e_1 \sum_{j=1}^{n} \tfrac{1}{2} j(j+1).$$

The minimal non-negative solution is found by setting $e_1 = 0$, and the conclusion follows by Exercise (6.3.7).

6.4 Solutions. Stationary distributions and the limit theorem

1. Let Y_n be the number of new errors introduced at the nth stage, and let G be the common probability generating function of the Y_n. Now $X_{n+1} = S_n + Y_{n+1}$ where S_n has the binomial distribution with parameters X_n and q $(= 1 - p)$. Thus the probability generating function G_n of X_n satisfies

$$\begin{aligned} G_{n+1}(s) &= G(s)\mathbb{E}(s^{S_n}) = G(s)\mathbb{E}\{\mathbb{E}(s^{S_n} \mid X_n)\} = G(s)\mathbb{E}\{(p+qs)^{X_n}\} \\ &= G(s)G_n(p+qs) = G(s)G_n\big(1 - q(1-s)\big). \end{aligned}$$

Therefore, for $s < 1$,

$$\begin{aligned} G_n(s) &= G(s)G\big(1-q(1-s)\big)G_{n-2}(1-q^2(1-s)) = \cdots \\ &= G_0\big(1-q^n(1-s)\big)\prod_{r=0}^{n-1} G\big(1-q^r(1-s)\big) \to \prod_{r=0}^{\infty} G\big(1-q^r(1-s)\big) \end{aligned}$$

as $n \to \infty$, assuming $q < 1$. This infinite product is therefore the probability generating function of the stationary distribution whenever this exists. If $G(s) = e^{\lambda(s-1)}$, then

$$\prod_{r=0}^{\infty} G\big(1-q^r(1-s)\big) = \exp\left\{\lambda(s-1)\sum_{r=0}^{\infty} q^r\right\} = e^{\lambda(s-1)/p},$$

so that the stationary distribution is Poisson with parameter λ/p.

2. (6.3.1): Assume for simplicity that $\sup\{j : a_j > 0\} = \infty$. The chain is irreducible if $r < 1$. Look for a stationary distribution $\boldsymbol{\pi}$ with probability generating function G. We have that

$$\pi_0 = a_0\pi_0 + (1-r)\pi_1, \quad \pi_i = a_i\pi_0 + r\pi_i + (1-r)\pi_{i+1} \text{ for } i \geq 1.$$

Hence

$$sG(s) = \pi_0 sA(s) + rs(G(s) - \pi_0) + (1-r)(G(s) - \pi_0)$$

where $A(s) = \sum_{j=0}^{\infty} a_j s^j$, and therefore

$$G(s) = \pi_0\left(\frac{sA(s) - (1 - r + sr)}{(1-r)(s-1)}\right).$$

Taking the limit as $s \uparrow 1$, we obtain by L'Hôpital's rule that

$$G(1) = \pi_0\left(\frac{A'(1) + 1 - r}{1-r}\right).$$

There exists a stationary distribution if and only if $r < 1$ and $A'(1) < \infty$, in which case

$$G(s) = \frac{sA(s) - (1 - r + sr)}{(s-1)(A'(1) + 1 - r)}.$$

Hence the chain is non-null persistent if and only if $r < 1$ and $A'(1) < \infty$. The mean recurrence time μ_i is found by expanding G and setting $\mu_i = 1/\pi_i$.

(6.3.2): Assume that $0 < p < 1$, and suppose first that $p \neq \frac{1}{3}$. Look for a solution $\{y_j : j \neq 0\}$ of the equations

$$(*) \qquad y_i = \sum_{j \neq 0} p_{ij} y_j, \qquad i \neq 0,$$

as in (6.4.10). Away from the origin, this equation is $y_i = qy_{i-1} + py_{i+2}$ where $p + q = 1$, with auxiliary equation $p\theta^3 - \theta + q = 0$. Now $p\theta^3 - \theta + q = p(\theta - 1)(\theta - \alpha)(\theta - \beta)$ where

$$\alpha = \frac{-p - \sqrt{p^2 + 4pq}}{2p} < -1, \quad \beta = \frac{-p + \sqrt{p^2 + 4pq}}{2p} > 0.$$

Note that $0 < \beta < 1$ if $p > \frac{1}{3}$, while $\beta > 1$ if $p < \frac{1}{3}$.

For $p > \frac{1}{3}$, set

$$y_i = \begin{cases} A + B\beta^i & \text{if } i \geq 1, \\ C + D\alpha^i & \text{if } i \leq -1, \end{cases}$$

the constants A, B, C, D being chosen in such a manner as to 'patch over' the omission of 0 in the equations ($*$):

$$(**) \qquad y_{-2} = qy_{-3}, \quad y_{-1} = qy_{-2} + py_1, \quad y_1 = py_3.$$

The result is a bounded non-zero solution $\{y_j\}$ to ($*$), and it follows that the chain is transient.

For $p < \frac{1}{3}$, follow the same route with

$$y_i = \begin{cases} 0 & \text{if } i \geq 1, \\ C + D\alpha^i + \mathbb{E}\beta^i & \text{if } i \leq -1, \end{cases}$$

the constants being chosen such that $y_{-2} = qy_{-3}$, $y_{-1} = qy_{-2}$.

Finally suppose that $p = \frac{1}{3}$, so that $\alpha = -2$ and $\beta = 1$. The general solution to ($*$) is

$$y_i = \begin{cases} A + Bi + C\alpha^i & \text{if } i \geq 1, \\ D + Ei + F\alpha^i & \text{if } i \leq -1, \end{cases}$$

subject to ($**$). Any bounded solution has $B = E = C = 0$, and ($**$) implies that $A = D = F = 0$. Therefore the only bounded solution to ($*$) is the zero solution, whence the chain is persistent. The equation $\mathbf{x} = \mathbf{xP}$ is satisfied by the vector $\mathbf{x}$ of 1's; by an appeal to (6.4.6), the walk is null.

(6.3.3): (a) Solve the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ to find a stationary distribution $\boldsymbol{\pi} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ when $p \neq 0$. Hence the chain is non-null and persistent, with $\mu_1 = \pi_1^{-1} = 4$, and similarly $\mu_2 = 2$, $\mu_3 = 4$.

(b) Similarly, $\boldsymbol{\pi} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ is a stationary distribution, and $\mu_i = \pi_i^{-1} = 4$.

(6.3.4): (a) The stationary distribution may be found to be $\pi_i = \frac{1}{8}$ for all i, so that $\mu_v = 8$.

3. The quantities $X_1, X_2, \ldots, X_n$ depend only on the initial contents of the reservoir and the rainfalls $Y_0, Y_1, \ldots, Y_{n-1}$. The contents on day $n + 1$ depend only on the value X_n of the previous contents and the rainfall Y_n. Since Y_n is independent of all earlier rainfalls, the process X is a Markov chain. Its state space is $S = \{0, 1, 2, \ldots, K - 1\}$ and it has transition matrix

$$\mathbb{P} = \begin{pmatrix} g_0 + g_1 & g_2 & g_3 & \cdots & g_{K-1} & G_K \\ g_0 & g_1 & g_2 & \cdots & g_{K-2} & G_{K-1} \\ 0 & g_0 & g_1 & \cdots & g_{K-3} & G_{K-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g_0 & G_1 \end{pmatrix}$$

where $g_i = \mathbb{P}(Y_1 = i)$ and $G_i = \sum_{j=i}^{\infty} g_j$. The equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ is as follows:

$$\begin{aligned} \pi_0 &= \pi_0(g_0 + g_1) + \pi_1 g_0, \\ \pi_r &= \pi_0 g_{r+1} + \pi_1 g_r + \cdots + \pi_{r+1} g_0, \qquad 0 < r < K - 1, \\ \pi_{K-1} &= \pi_0 G_K + \pi_1 G_{K-1} + \cdots + \pi_{K-1} G_1. \end{aligned}$$

The final equation is a consequence of the previous ones, since $\sum_{i=0}^{K-1} \pi_i = 1$. Suppose then that $\nu = (\nu_1, \nu_2, \dots)$ is an infinite vector satisfying

$$\nu_0 = \nu_0(g_0 + g_1) + \nu_1 g_0, \quad \nu_r = \nu_0 g_{r+1} + \nu_1 g_r + \cdots + \nu_{r+1} g_0 \text{ for } r > 0.$$

Multiply through the equation for ν_r by s^{r+1}, and sum over r to find (after a little work) that

$$N(s) = \sum_{i=0}^{\infty} \nu_i s^i, \qquad G(s) = \sum_{i=0}^{\infty} g_i s^i$$

satisfy $sN(s) = N(s)G(s) + \nu_0 g_0(s-1)$, and hence

$$\frac{1}{\nu_0} N(s) = \frac{g_0(s-1)}{s - G(s)}.$$

The probability generating function of the π_i is therefore a constant multiplied by the coefficients of $s^0, s^1, \dots, s^{K-1}$ in $g_0(s-1)/(s-G(s))$, the constant being chosen in such a way that $\sum_{i=0}^{K-1} \pi_i = 1$.

When $G(s) = p(1-qs)^{-1}$, then $g_0 = p$ and

$$\frac{g_0(s-1)}{s-G(s)} = \frac{p(1-qs)}{p-qs} = p + \frac{q}{1-(qs/p)}.$$

The coefficient of s^0 is 1, and of s^i is q^{i+1}/p^i if $i \ge 1$. The stationary distribution is therefore given by $\pi_i = q\pi_0(q/p)^i$ for $i \ge 1$, where

$$\pi_0 = \frac{1}{1 + \sum_1^{K-1} q(q/p)^i} = \frac{p-q}{p-q+q^2\big(1-(q/p)^{K-1}\big)}$$

if $p \ne q$, and $\pi_0 = 2/(K+1)$ if $p = q = \frac{1}{2}$.

4. The transition matrices

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{P}_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

have respective stationary distributions $\boldsymbol{\pi}_1 = (p, 1-p)$ and $\boldsymbol{\pi}_2 = \big(\frac{1}{2}p, \frac{1}{2}p, \frac{1}{2}(1-p), \frac{1}{2}(1-p)\big)$ for any $0 \le p \le 1$.

5. (a) Set $i = 1$, and find an increasing sequence $n_1(1), n_1(2), \dots$ along which $x_1(n)$ converges. Now set $i = 2$, and find a subsequence of $(n_1(j) : j \ge 1)$ along which $x_2(n)$ converges; denote this subsequence by $n_2(1), n_2(2), \dots$. Continue inductively to obtain, for each i, a sequence $\mathbf{n}_i = (n_i(j) : j \ge 1)$, noting that:

(i) $\mathbf{n}_{i+1}$ is a subsequence of $\mathbf{n}_i$, and

(ii) $\lim_{r\to\infty} x_i(n_i(r))$ exists for all i.

Finally, define $m_k = n_k(k)$. For each $i \ge 1$, the sequence $m_i, m_{i+1}, \dots$ is a subsequence of $\mathbf{n}_i$, and therefore $\lim_{r\to\infty} x_i(m_r)$ exists.

(b) Let S be the state space of the irreducible Markov chain X. There are countably many pairs i, j of states, and part (a) may be applied to show that there exists a sequence $(n_r : r \ge 1)$ and a family $(\alpha_{ij} : i, j \in S)$, not all zero, such that $p_{ij}(n_r) \to \alpha_{ij}$ as $r \to \infty$.

Now X is persistent, since otherwise $p_{ij}(n) \to 0$ for all i, j. The coupling argument in the proof of the ergodic theorem (6.4.17) is valid, so that $p_{aj}(n) - p_{bj}(n) \to 0$ as $n \to \infty$, implying that $\alpha_{aj} = \alpha_{bj}$ for all a, b, j.

6. Just check that $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ and $\sum_v \pi_v = 1$.

7. Let X_n be the Markov chain which takes the value r if the walk is at any of the 2^r nodes at level r. Then X_n executes a simple random walk with retaining barrier having $p = 1 - q = \frac{2}{3}$, and it is thus transient by Example (6.4.15).

8. Assume that X_n includes particles present just after the entry of the fresh batch Y_n. We may write

$$X_{n+1} = \sum_{i=1}^{X_n} B_{i,n} + Y_n$$

where the $B_{i,n}$ are independent Bernoulli variables with parameter $1 - p$. Therefore X is a Markov chain. It follows also that

$$G_{n+1}(s) = \mathbb{E}(s^{X_{n+1}}) = G_n(p + qs)e^{\lambda(s-1)}.$$

In equilibrium, $G_{n+1} = G_n = G$, where $G(s) = G(p + qs)e^{\lambda(s-1)}$. There is a unique stationary distribution, and it is easy to see that $G(s) = e^{\lambda(s-1)/p}$ must therefore be the solution. The answer is the Poisson distribution with parameter λ/p.

9. The Markov chain X has a uniform transition distribution $p_{jk} = 1/(j + 2)$, $0 \le k \le j + 1$. Therefore,

$$\begin{aligned}\mathbb{E}(X_n) &= \mathbb{E}\big(\mathbb{E}(X_n \mid X_{n-1})\big) = \tfrac{1}{2}\big(1 + \mathbb{E}(X_{n-1})\big) = \cdots \\ &= 1 - (\tfrac{1}{2})^n + (\tfrac{1}{2})^n X_0.\end{aligned}$$

The equilibrium probability generating function satisfies

$$G(s) = \mathbb{E}(s^{X_n}) = \mathbb{E}\big(\mathbb{E}(s^{X_n} \mid X_{n-1})\big) = \mathbb{E}\left\{\frac{1 - s^{X_n+2}}{(1-s)(X_n+2)}\right\},$$

whence

$$\frac{d}{ds}\big\{(1-s)G(s)\big\} = -sG(s),$$

subject to $G(1) = 1$. The solution is $G(s) = e^{s-1}$, which is the probability generating function of the Poisson distribution with parameter 1.

10. This is the claim of Theorem (6.4.13). Without loss of generality we may take $s = 0$ and the y_j to be non-negative (since if the y_j solve the equations, then so do $y_j + c$ for any constant c). Let $\mathbf{T}$ be the matrix obtained from $\mathbf{P}$ by deleting the row and column labelled 0, and write $\mathbf{T}^n = (t_{ij}(n) : i, j \neq 0)$. Then $\mathbf{T}^n$ includes all the n-step probabilities of paths that never visit zero.

We claim first that, for all i, j it is the case that $t_{ij}(n) \to 0$ as $n \to \infty$. The quantity $t_{ij}(n)$ may be thought of as the n-step transition probability from i to j in an altered chain in which s has been made absorbing. Since the original chain is assumed irreducible, all states communicate with s, and therefore all states other than s are transient in the altered chain, implying by the summability of $t_{ij}(n)$ (Corollary (6.2.4)) that $t_{ij}(n) \to 0$ as required.

Iterating the inequality $\mathbf{y} \ge \mathbf{T}\mathbf{y}$ yields $\mathbf{y} \ge \mathbf{T}^n\mathbf{y}$, which is to say that

$$y_i \ge \sum_{j=1}^{\infty} t_{ij}(n) y_j \ge \min_{s \ge 1}\{y_{r+s}\} \sum_{j=r+1}^{\infty} t_{ij}(n), \qquad i \ge 1.$$

Let $A_n = \{X_k \neq 0 \text{ for } k \leq n\}$. For $i \geq 1$,

$$\mathbb{P}(A_\infty \mid X_0 = i) = \lim_{n\to\infty} \mathbb{P}(A_n \mid X_0 = i) = \sum_{j=1}^{\infty} t_{ij}(n)$$
$$\leq \lim_{n\to\infty} \left\{ \sum_{j=1}^{r} t_{ij}(n) + \frac{y_i}{\min_{s\geq 1}\{y_{r+s}\}} \right\}.$$

Let $\epsilon > 0$, and pick R such that

$$\frac{y_i}{\min_{s\geq 1}\{y_{R+s}\}} < \epsilon.$$

Take $r = R$ and let $n \to \infty$, implying that $\mathbb{P}(A_\infty \mid X_0 = i) = 0$. It follows that 0 is persistent, and by irreducibility that all states are persistent.

11. By Exercise (6.4.6), the stationary distribution is $\pi_A = \pi_B = \pi_D = \pi_E = \frac{1}{6}$, $\pi_C = \frac{1}{3}$.
(a) By Theorem (6.4.3), the answer is $\mu_A = 1/\pi_A = 6$.
(b) By the argument around Lemma (6.4.5), the answer is $\rho_D(A) = \pi_D \mu_A = \pi_D/\pi_A = 1$.
(c) Using the same argument, the answer is $\rho_C(A) = \pi_C/\pi_A = 2$.
(d) Let $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot \mid X_0 = i)$, let T_j be the time of the first passage to state j, and let $\nu_i = \mathbb{P}_i(T_A < T_E)$. By conditioning on the first step,

$$\nu_B = \tfrac{1}{2} + \tfrac{1}{2}\nu_C, \quad \nu_C = \tfrac{1}{4} + \tfrac{1}{4}\nu_B + \tfrac{1}{4}\nu_D, \quad \nu_A = \tfrac{1}{2}\nu_B + \tfrac{1}{2}\nu_C, \quad \nu_D = \tfrac{1}{2}\nu_C,$$

with solution $\nu_A = \frac{5}{8}$, $\nu_B = \frac{3}{4}$, $\nu_C = \frac{1}{2}$, $\nu_D = \frac{1}{4}$.

A typical conditional transition probability $\tau_{ij} = \mathbb{P}_i(X_1 = j \mid T_A < T_E)$ is calculated as follows:

$$\tau_{AB} = \mathbb{P}_A(X_1 = B \mid T_A < T_E) = \frac{\mathbb{P}_A(X_1 = B)\mathbb{P}_B(T_A < T_E)}{\mathbb{P}_A(T_A < T_E)} = \frac{\nu_B}{2\nu_A} = \frac{3}{5},$$

and similarly,

$$\tau_{AC} = \tfrac{2}{5}, \ \tau_{BA} = \tfrac{2}{3}, \ \tau_{BC} = \tfrac{1}{3}, \ \tau_{CA} = \tfrac{1}{2}, \ \tau_{CB} = \tfrac{3}{8}, \ \tau_{CD} = \tfrac{1}{8}, \ \tau_{DC} = 1.$$

We now compute the conditional expectations $\widetilde{\mu}_i = \mathbb{E}_i(T_A \mid T_A < T_E)$ by conditioning on the first step of the conditioned process. This yields equations of the form $\widetilde{\mu}_A = 1 + \frac{3}{5}\widetilde{\mu}_B + \frac{2}{5}\widetilde{\mu}_C$, whose solution gives $\widetilde{\mu}_A = \frac{14}{5}$.
(e) Either use the stationary distribution of the conditional transition matrix $\boldsymbol{\tau}$, or condition on the first step as follows. With N the number of visits to D, and $\eta_i = \mathbb{E}_i(N \mid T_A < T_E)$, we obtain

$$\eta_A = \tfrac{3}{5}\eta_B + \tfrac{2}{5}\eta_C, \quad \eta_B = 0 + \tfrac{1}{3}\eta_C, \quad \eta_C = 0 + \tfrac{3}{8}\eta_B + \tfrac{1}{8}(1 + \eta_D), \quad \eta_D = \eta_C,$$

whence in particular $\eta_A = \frac{1}{10}$.

12. By Exercise (6.4.6), the stationary distribution has $\pi_A = \frac{1}{14}$, $\pi_B = \frac{1}{7}$. Using the argument around Lemma (6.4.5), the answer is $\rho_B(A) = \pi_B \mu_A = \pi_B/\pi_A = 2$.

6.5 Solutions. Reversibility

1. Look for a solution to the equations $\pi_i p_{ij} = \pi_j p_{ji}$. The only non-trivial cases of interest are those with $j = i + 1$, and therefore $\lambda_i \pi_i = \mu_{i+1}\pi_{i+1}$ for $0 \le i < b$, with solution

$$\pi_i = \pi_0 \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}, \qquad 0 \le i \le b,$$

an empty product being interpreted as 1. The constant π_0 is chosen in order that the π_i sum to 1, and the chain is therefore time-reversible.

2. Let $\boldsymbol{\pi}$ be the stationary distribution of X, and suppose X is reversible. We have that $\pi_i p_{ij} = p_{ji}\pi_j$ for all i, j, and furthermore $\pi_i > 0$ for all i. Hence

$$\pi_i p_{ij} p_{jk} p_{ki} = p_{ji}\pi_j p_{jk} p_{ki} = p_{ji} p_{kj} \pi_k p_{ki} = p_{ji} p_{kj} p_{ik} \pi_i$$

as required when $n = 3$. A similar calculation is valid when $n > 3$.

Suppose conversely that the given display holds for all finite sequences of states. Sum over all values of the subsequence $j_2, \dots, j_{n-1}$ to deduce that $p_{ij}(n-1) p_{ji} = p_{ij} p_{ji}(n-1)$, where $i = j_1$, $j = j_n$. Take the limit as $n \to \infty$ to obtain $\pi_j p_{ji} = p_{ij}\pi_i$ as required for time-reversibility.

3. With $\boldsymbol{\pi}$ the stationary distribution of X, look for a stationary distribution $\boldsymbol{\nu}$ of Y of the form

$$\nu_i = \begin{cases} c\beta\pi_i & \text{if } i \notin C, \\ c\pi_i & \text{if } i \in C. \end{cases}$$

There are four cases.
(a) $i \in C, j \notin C$: $\nu_i q_{ij} = c\pi_i \beta p_{ij} = c\beta\pi_j p_{ji} = \nu_j q_{ji}$,
(b) $i, j \in C$: $\nu_i q_{ij} = c\pi_i p_{ij} = c\pi_j p_{ji} = \nu_j q_{ji}$,
(c) $i \notin C, j \in C$: $\nu_i q_{ij} = c\beta\pi_i p_{ij} = c\beta\pi_j p_{ji} = \nu_j q_{ji}$,
(d) $i, j \notin C$: $\nu_i q_{ij} = c\beta\pi_i p_{ij} = c\beta\pi_j p_{ji} = \nu_j q_{ji}$.
Hence the modified chain is reversible in equilibrium with stationary distribution $\boldsymbol{\nu}$, when

$$c\left\{\beta \sum_{i \notin C} \pi_i + \sum_{i \in C} \pi_i\right\} = 1.$$

In the limit as $\beta \downarrow 0$, the chain Y never leaves the set C once it has arrived in it.

4. Only if the period is 2, because of the detailed balance equations.

5. With $Y_n = X_n - \frac{1}{2}m$,

$$\begin{aligned} \mathbb{E}(Y_n) &= \mathbb{E}(Y_{n-1}) + \mathbb{E}(X_n - X_{n-1}) \\ &= \mathbb{E}(Y_{n-1}) + \mathbb{E}\left\{\left(1 - \frac{X_{n-1}}{m}\right) - \frac{X_{n-1}}{m}\right\} = \mathbb{E}(Y_{n-1}) - \frac{2}{m}\mathbb{E}(Y_{n-1}). \end{aligned}$$

Now iterate.

6. (a) The distribution $\pi_1 = \beta/(\alpha+\beta)$, $\pi_2 = \alpha/(\alpha+\beta)$ satisfies the detailed balance equations, so this chain is reversible.
(b) By symmetry, the stationary distribution is $\boldsymbol{\pi} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, which satisfies the detailed balance equations if and only if $p = \frac{1}{2}$.
(c) This chain is reversible if and only if $p = \frac{1}{2}$.

7. A simple random walk which moves rightwards with probability p has a stationary measure $\pi_n = A(p/q)^n$, in the sense that $\boldsymbol{\pi}$ is a vector satisfying $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. It is not necessarily the case that this $\boldsymbol{\pi}$ has finite sum. It may then be checked that the recipe given in the solution to Exercise (6.5.3) yields $\pi(i, j) = \rho_1^i \rho_2^j / \sum_{(r,s)\in C} \rho_1^r \rho_2^s$ as stationary distribution for the given process, where C is the relevant region of the plane, and $\rho_i = p_i/q_i$ and p_i $(= 1 - q_i)$ is the chance that the ith walk moves rightwards on any given step.

8. Since the chain is irreducible with a finite state space, we have that $\pi_i > 0$ for all i. Assume the chain is reversible. The balance equations $\pi_i p_{ij} = \pi_j p_{ji}$ give $p_{ij} = \pi_j p_{ji}/\pi_i$. Let $\mathbf{D}$ be the matrix with entries $1/\pi_i$ on the diagonal, and $\mathbf{S}$ the matrix $(\pi_j p_{ji})$, and check that $\mathbf{P} = \mathbf{DS}$.

Conversely, if $\mathbf{P} = \mathbf{DS}$, then $d_i^{-1} p_{ij} = d_j^{-1} p_{ji}$, whence $\pi_i = d_i^{-1} / \sum_k d_k^{-1}$ satisfies the detailed balance equations.

Note that

$$p_{ij} = \frac{1}{\sqrt{\pi_i}} \sqrt{\frac{\pi_i}{\pi_j}}\, p_{ij} \sqrt{\pi_j}.$$

If the chain is reversible in equilibrium, the matrix $\mathbf{M} = \big(\sqrt{\pi_i/\pi_j}\, p_{ij}\big)$ is symmetric, and therefore $\mathbf{M}$, and, by the above, $\mathbf{P}$, has real eigenvalues. An example of the failure of the converse is the transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} \\ 1 & 0 & 0 \end{pmatrix},$$

which has real eigenvalues 1, and $-\frac{1}{2}$ (twice), and stationary distribution $\boldsymbol{\pi} = (\frac{4}{9}, \frac{4}{9}, \frac{1}{9})$. However, $\pi_1 p_{13} = 0 \neq \frac{1}{9} = \pi_3 p_{31}$, so that such a chain is not reversible.

9. Simply check the detailed balance equations $\pi_i p_{ij} = \pi_j p_{ji}$.

6.6 Solutions. Chains with finitely many states

1. Let $\mathbf{P} = (p_{ij} : 1 \le i, j \le n)$ be a stochastic matrix and let C be the subset of $\mathbb{R}^n$ containing all vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ satisfying $x_i \ge 0$ for all i and $\sum_{i=1}^n x_i = 1$; for $\mathbf{x} \in C$, let $\|\mathbf{x}\| = \max_j \{x_j\}$. Define the linear mapping $T : C \to \mathbb{R}^n$ by $T(\mathbf{x}) = \mathbf{xP}$. Let us check that T is a continuous function from C into C. First,

$$\|T(\mathbf{x})\| = \max_j \Big\{ \sum_i x_i p_{ij} \Big\} \le \alpha \|\mathbf{x}\|$$

where

$$\alpha = \max_j \Big\{ \sum_i p_{ij} \Big\};$$

hence $\|T(\mathbf{x}) - T(\mathbf{y})\| \le \alpha \|\mathbf{x} - \mathbf{y}\|$. Secondly, $T(\mathbf{x})_j \ge 0$ for all j, and

$$\sum_j T(\mathbf{x})_j = \sum_j \sum_i x_i p_{ij} = \sum_i x_i \sum_j p_{ij} = 1.$$

Applying the given theorem, there exists a point $\boldsymbol{\pi}$ in C such that $T(\boldsymbol{\pi}) = \boldsymbol{\pi}$, which is to say that $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.

2. Let $\mathbf{P}$ be a stochastic $m \times m$ matrix and let $\mathbf{T}$ be the $m \times (m + 1)$ matrix with (i, j)th entry

$$t_{ij} = \begin{cases} p_{ij} - \delta_{ij} & \text{if } j \le m, \\ 1 & \text{if } j = m + 1, \end{cases}$$

where δ_{ij} is the Kronecker delta. Let $\mathbf{v} = (0, 0, \dots, 0, 1) \in \mathbb{R}^{m+1}$. If statement (ii) of the question is valid, there exists $\mathbf{y} = (y_1, y_2, \dots, y_{m+1})$ such that

$$y_{m+1} < 0, \qquad \sum_{j=1}^{m} (p_{ij} - \delta_{ij}) y_j + y_{m+1} \geq 0 \text{ for } 1 \leq i \leq m;$$

this implies that

$$\sum_{j=1}^{m} p_{ij} y_j \geq y_i - y_{m+1} > y_i \qquad \text{for all } i,$$

and hence the impossibility that $\sum_{j=1}^{m} p_{ij} y_j > \max_i\{y_i\}$. It follows that statement (i) holds, which is to say that there exists a non-negative vector $\mathbf{x} = (x_1, x_2, \dots, x_m)$ such that $\mathbf{x}(\mathbf{P} - \mathbf{I}) = \mathbf{0}$ and $\sum_{i=1}^{m} x_i = 1$; such an $\mathbf{x}$ is the required eigenvector.

3. Thinking of x_{n+1} as the amount you may be sure of winning, you seek a betting scheme $\mathbf{x}$ such that x_{n+1} is maximized subject to the inequalities

$$x_{n+1} \leq \sum_{i=1}^{n} x_i t_{ij} \quad \text{for } 1 \leq j \leq m.$$

Writing $a_{ij} = -t_{ij}$ for $1 \leq i \leq n$ and $a_{n+1,j} = 1$, we obtain the linear program:

$$\text{maximize} \quad x_{n+1} \quad \text{subject to} \quad \sum_{i=1}^{n+1} x_i a_{ij} \leq 0 \quad \text{for } 1 \leq j \leq m.$$

The dual linear program is:

$$\text{minimize} \quad 0 \quad \text{subject to} \quad \sum_{j=1}^{m} a_{ij} y_j = 0 \quad \text{for } 1 \leq i \leq n,$$
$$\sum_{j=1}^{m} a_{n+1,j} y_j = 1, \quad y_j \geq 0 \quad \text{for } 1 \leq j \leq m.$$

Re-expressing the a_{ij} in terms of the t_{ij} as above, the dual program takes the form:

$$\text{minimize} \quad 0 \quad \text{subject to} \quad \sum_{j=1}^{m} t_{ij} p_j = 0 \quad \text{for } 1 \leq i \leq n,$$
$$\sum_{j=1}^{m} p_j = 1, \quad p_j \geq 0 \quad \text{for } 1 \leq j \leq m.$$

The vector $\mathbf{x} = \mathbf{0}$ is a feasible solution of the primal program. The dual program has a feasible solution if and only if statement (a) holds. Therefore, if (a) holds, the dual program has minimal value 0, whence by the duality theorem of linear programming, the maximal value of the primal program is 0, in contradiction of statement (b). On the other hand, if (a) does not hold, the dual has no feasible solution, and therefore the primal program has no optimal solution. That is, the objective function of the primal is unbounded, and therefore (b) holds. [This was proved by De Finetti in 1937.]

4. Use induction, the claim being evidently true when $n = 1$. Suppose it is true for $n = m$. Certainly $\mathbf{P}^{m+1}$ is of the correct form, and the equation $\mathbf{P}^{m+1}\mathbf{x}' = \mathbf{P}(\mathbf{P}^m\mathbf{x}')$ with $\mathbf{x} = (1, \omega, \omega^2)$ yields in its first row

$$a_{1,m+1} + a_{2,m+1}\omega + a_{3,m+1}\omega^2 = (1 - p + p\omega)^m (\mathbf{P}\mathbf{x}')_1 = (1 - p + p\omega)^{m+1}$$

as required.

5. The first part follows from the fact that $\boldsymbol{\pi}\mathbf{1}' = 1$ if and only if $\boldsymbol{\pi}\mathbf{U} = \mathbf{1}$. The second part follows from the fact that $\pi_i > 0$ for all i if $\mathbf{P}$ is finite and irreducible, since this implies the invertibility of $\mathbf{I} - \mathbf{P} + \mathbf{U}$.

6. The chessboard corresponds to a graph with $8 \times 8 = 64$ vertices, pairs of which are connected by edges when the corresponding move is legitimate for the piece in question. By Exercises (6.4.6), (6.5.9), we need only check that the graph is connected, and to calculate the degree of a corner vertex.

(a) For the king there are 4 vertices of degree 3, 24 of degree 5, 36 of degree 8. Hence, the number of edges is 210 and the degree of a corner is 3. Therefore $\mu(\text{king}) = 420/3 = 140$.

(b) $\mu(\text{queen}) = (28 \times 21 + 20 \times 23 + 12 \times 25 + 4 \times 27)/21 = 208/3$.

(c) We restrict ourselves to the set of 32 vertices accessible from a given corner. Then $\mu(\text{bishop}) = (14 \times 7 + 10 \times 9 + 6 \times 11 + 2 \times 13)/7 = 40$.

(d) $\mu(\text{knight}) = (4 \times 2 + 8 \times 3 + 20 \times 4 + 16 \times 6 + 16 \times 8)/2 = 168$.

(e) $\mu(\text{rook}) = 64 \times 14/14 = 64$.

7. They are walking on a product space of 8×16 vertices. Of these, 6×16 have degree 6×3 and 16×2 have degree 6×5. Hence

$$\mu(C) = (6 \times 16 \times 6 \times 3 + 16 \times 2 \times 6 \times 5)/18 = 448/3.$$

8. $|\mathbf{P} - \lambda I| = (\lambda - 1)(\lambda + \frac{1}{2})(\lambda + \frac{1}{6})$. Tedious computation yields the eigenvectors, and thus

$$\mathbf{P}^n = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \frac{1}{12}(-\tfrac{1}{2})^n \begin{pmatrix} 8 & -4 & -4 \\ 2 & -1 & -1 \\ -10 & 5 & 5 \end{pmatrix} + \frac{1}{4}(-\tfrac{1}{6})^n \begin{pmatrix} 0 & 0 & 0 \\ -2 & 3 & -1 \\ 2 & -3 & 1 \end{pmatrix}.$$

6.7 Solutions. Branching processes revisited

1. We have (using Theorem (5.4.3), or the fact that $G_{n+1}(s) = G(G_n(s))$) that the probability generating function of Z_n is

$$G_n(s) = \frac{n - (n-1)s}{n + 1 - ns},$$

so that

$$\mathbb{P}(Z_n = k) = \left(\frac{n}{n+1}\right)^{k+1} - \left(\frac{n-1}{n+1}\right)\left(\frac{n}{n+1}\right)^{k-1} = \frac{n^{k-1}}{(n+1)^{k+1}}$$

for $k \geq 1$. Therefore, for $y > 0$, as $n \to \infty$,

$$\mathbb{P}(Z_n \leq 2yn \mid Z_n > 0) = \frac{1}{1 - G_n(0)} \sum_{k=1}^{\lfloor 2yn \rfloor} \frac{n^{k-1}}{(n+1)^{k+1}} = 1 - \left(1 + \frac{1}{n}\right)^{-\lfloor 2yn \rfloor} \to 1 - e^{-2y}.$$

2. Using the independence of different lines of descent,

$$\mathbb{E}(s^{Z_n} \mid \text{extinction}) = \sum_{k=0}^{\infty} \frac{s^k \mathbb{P}(Z_n = k, \text{extinction})}{\mathbb{P}(\text{extinction})} = \sum_{k=0}^{\infty} \frac{s^k \mathbb{P}(Z_n = k)\eta^k}{\eta} = \frac{1}{\eta} G_n(s\eta),$$

where G_n is the probability generating function of Z_n.

3. We have that $\eta = G(\eta)$. In this case $G(s) = q(1-ps)^{-1}$, and therefore $\eta = q/p$. Hence

$$\frac{1}{\eta}G_n(s\eta) = \frac{p}{q}\cdot\frac{q\{p^n - q^n - p(sq/p)(p^{n-1}-q^{n-1})\}}{p^{n+1}-q^{n+1}-p(sq/p)(p^n-q^n)} = \frac{p\{q^n - p^n - qs(q^{n-1}-p^{n-1})\}}{q^{n+1}-p^{n+1}-qs(q^n-p^n)},$$

which is $G_n(s)$ with p and q interchanged.

4. (a) Using the fact that $\mathrm{var}(X \mid X > 0) \ge 0$,

$$\mathbb{E}(X^2) = \mathbb{E}(X^2 \mid X > 0)\mathbb{P}(X > 0) \ge \mathbb{E}(X \mid X > 0)^2\mathbb{P}(X > 0) = \mathbb{E}(X)\mathbb{E}(X \mid X > 0).$$

(b) Hence

$$\mathbb{E}(Z_n/\mu^n \mid Z_n > 0) \le \frac{\mathbb{E}(Z_n^2)}{\mu^n\mathbb{E}(Z_n)} = \mathbb{E}(W_n^2)$$

where $W_n = Z_n/\mathbb{E}(Z_n)$. By an easy calculation (see Lemma (5.4.2)),

$$\mathbb{E}(W_n^2) = \frac{\sigma^2(1-\mu^{-n})}{\mu^2-\mu} + 1 \le \frac{\sigma^2}{\mu^2-\mu} + 1 = \frac{2p}{p-q}$$

where $\sigma^2 = \mathrm{var}(Z_1) = p/q^2$.

(c) Doing the calculation exactly,

$$\mathbb{E}(Z_n/\mu^n \mid Z_n > 0) = \frac{\mathbb{E}(Z_n/\mu^n)}{\mathbb{P}(Z_n > 0)} = \frac{1}{1-G_n(0)} \to \frac{1}{1-\eta}$$

where $\eta = \mathbb{P}(\text{ultimate extinction}) = q/p$.

6.8 Solutions. Birth processes and the Poisson process

1. Let F and W be the incoming Poisson processes, and let $N(t) = F(t)+W(t)$. Certainly $N(0) = 0$ and N is non-decreasing. Arrivals of flies during $[0, s]$ are independent of arrivals during $(s, t]$, if $s < t$; similarly for wasps. Therefore the aggregated arrival process during $[0, s]$ is independent of the aggregated process during $(s, t]$. Now

$$\mathbb{P}\big(N(t+h) = n+1 \,\big|\, N(t) = n\big) = \mathbb{P}(A \mathbin{\triangle} B)$$

where

$$A = \big\{\text{one fly arrives during } (t, t+h]\big\}, \quad B = \big\{\text{one wasp arrives during } (t, t+h]\big\}.$$

We have that

$$\mathbb{P}(A \mathbin{\triangle} B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \lambda h + \mu h - (\lambda h)(\mu h) + \mathrm{o}(h) = (\lambda+\mu)h + \mathrm{o}(h).$$

Finally

$$\mathbb{P}\big(N(t+h) > n+1 \,\big|\, N(t) = n\big) \le \mathbb{P}(A \cap B) + \mathbb{P}(C \cup D),$$

where $C = \{\text{two or more flies arrive in } (t, t+h]\}$ and $D = \{\text{two or more wasps arrive in } (t, t+h]\}$. This probability is no greater than $(\lambda h)(\mu h) + o(h) = o(h)$.

2. Let I be the incoming Poisson process, and let G be the process of arrivals of green insects. Matters of independence are dealt with as above. Finally,

$$\mathbb{P}\big(G(t+h) = n+1 \,\big|\, G(t) = n\big) = p\mathbb{P}\big(I(t+h) = n+1 \,\big|\, I(t) = n\big) + o(h) = p\lambda h + o(h),$$
$$\mathbb{P}\big(G(t+h) > n+1 \,\big|\, G(t) = n\big) \le \mathbb{P}\big(I(t+h) > n+1 \,\big|\, I(t) = n\big) = o(h).$$

3. Conditioning on T_1 and using the time-homogeneity of the process,

$$\mathbb{P}\big(E(t) > x \,\big|\, T_1 = u\big) = \begin{cases} \mathbb{P}\big(E(t-u) > x\big) & \text{if } u \le t, \\ 0 & \text{if } t < u \le t+x, \\ 1 & \text{if } u > t+x, \end{cases}$$

(draw a diagram to help you see this). Therefore

$$\begin{aligned} \mathbb{P}\big(E(t) > x\big) &= \int_0^\infty \mathbb{P}\big(E(t) > x \mid T_1 = u\big)\lambda e^{-\lambda u}\,du \\ &= \int_0^t \mathbb{P}\big(E(t-u) > x\big)\lambda e^{-\lambda u}\,du + \int_{t+x}^\infty \lambda e^{-\lambda u}\,du. \end{aligned}$$

You may solve the integral equation using Laplace transforms. Alternately you may guess the answer and then check that it works. The answer is $\mathbb{P}(E(t) \le x) = 1 - e^{-\lambda x}$, the exponential distribution. Actually this answer is obvious since $E(t) > x$ if and only if there is no arrival in $[t, t+x]$, an event having probability $e^{-\lambda x}$.

4. The forward equation is

$$p'_{ij}(t) = \lambda(j-1)p_{i,j-1}(t) - \lambda j p_{ij}(t), \qquad i \le j,$$

with boundary conditions $p_{ij}(0) = \delta_{ij}$, the Kronecker delta. We write $G_i(s,t) = \sum_j s^j p_{ij}(t)$, the probability generating function of $B(t)$ conditional on $B(0) = i$. Multiply through the differential equation by s^j and sum over j:

$$\frac{\partial G_i}{\partial t} = \lambda s^2 \frac{\partial G_i}{\partial s} - \lambda s \frac{\partial G_i}{\partial s},$$

a partial differential equation with boundary condition $G_i(s,0) = s^i$. This may be solved in the usual way to obtain $G_i(s,t) = g(e^{\lambda t}(1 - s^{-1}))$ for some function g. Using the boundary condition, we find that $g(1 - s^{-1}) = s^i$ and so $g(u) = (1-u)^{-i}$, yielding

$$G_i(s,t) = \frac{1}{\{1 - e^{\lambda t}(1 - s^{-1})\}^i} = \frac{(se^{-\lambda t})^i}{\{1 - s(1 - e^{-\lambda t})\}^i}.$$

The coefficient of s^j is, by the binomial series,

$$(*) \qquad p_{ij}(t) = e^{-i\lambda t}\binom{j-1}{i-1}(1 - e^{-\lambda t})^{j-i}, \qquad j \ge i,$$

as required.

Alternatively use induction. Set $j = i$ to obtain $p'_{ii}(t) = -\lambda i p_{ii}(t)$ (remember $p_{i,i-1}(t) = 0$), and therefore $p_{ii}(t) = e^{-\lambda i t}$. Rewrite the differential equation as

$$\frac{d}{dt}\left(p_{ij}(t)e^{\lambda j t}\right) = \lambda(j-1)p_{i,j-1}(t)e^{\lambda j t}.$$

Set $j = i+1$ and solve to obtain $p_{i,i+1}(t) = ie^{-\lambda i t}\left(1 - e^{-\lambda t}\right)$. Hence $(*)$ holds, by induction.

The mean is

$$\mathbb{E}(B(t)) = \frac{\partial}{\partial s}G_I(s,t)\bigg|_{s=1} = Ie^{\lambda t},$$

by an easy calculation. Similarly $\mathrm{var}(B(t)) = A + \mathbb{E}(B(t)) - \mathbb{E}(B(t))^2$ where

$$A = \frac{\partial^2}{\partial s^2}G_I(s,t)\bigg|_{s=1}.$$

Alternatively, note that $B(t)$ has the negative binomial distribution with parameters $e^{-\lambda t}$ and I.

5. The forward equations are

$$p'_n(t) = \lambda_{n-1}p_{n-1}(t) - \lambda_n p_n(t), \qquad n \geq 0,$$

where $\lambda_i = i\lambda + \nu$. The process is honest, and therefore $m(t) = \sum_n np_n(t)$ satisfies

$$\begin{aligned} m'(t) &= \sum_{n=1}^{\infty} n\big[(n-1)\lambda + \nu\big]p_{n-1}(t) - \sum_{n=0}^{\infty} n(n\lambda + \nu)p_n(t) \\ &= \sum_{n=0}^{\infty}\big\{\lambda[(n+1)n - n^2] + \nu[(n+1) - n]\big\}p_n(t) \\ &= \sum_{n=0}^{\infty}(\lambda n + \nu)p_n(t) = \lambda m(t) + \nu. \end{aligned}$$

Solve subject to $m(0) = 0$ to obtain $m(t) = \nu(e^{\lambda t} - 1)/\lambda$.

6. Using the fact that the time to the nth arrival is the sum of exponential interarrival times (or using equation (6.8.15)), we have that

$$\widehat{p}_n(\theta) = \int_0^{\infty} e^{-\theta t}p_n(t)\,dt$$

is given by

$$\widehat{p}_n(\theta) = \frac{1}{\lambda_n}\prod_{i=0}^{n}\frac{\lambda_i}{\lambda_i + \theta}$$

which may be expressed, using partial fractions, as

$$\widehat{p}_n(\theta) = \frac{1}{\lambda_n}\sum_{i=0}^{n}\frac{a_i\lambda_i}{\lambda_i + \theta}$$

where

$$a_i = \prod_{\substack{j=0 \\ j\neq i}}^{n}\frac{\lambda_j}{\lambda_j - \lambda_i}$$

so long as $\lambda_i \neq \lambda_j$ whenever $i \neq j$. The Laplace transform $\widehat{p}_n$ may now be inverted as

$$p_n(t) = \frac{1}{\lambda_n} \sum_{i=0}^{n} a_i \lambda_i e^{-\lambda_i t}.$$

See also Exercise (4.8.4).

7. Let T_n be the time of the nth arrival, and let $T = \lim_{n\to\infty} T_n = \sup\{t : N(t) < \infty\}$. Now, as in Exercise (6.8.6),

$$\lambda_n \widehat{p}_n(\theta) = \prod_{i=0}^{n} \frac{\lambda_i}{\lambda_i + \theta} = \mathbb{E}(e^{-\theta T_n})$$

since $T_n = X_0 + X_1 + \cdots + X_n$ where X_k is the $(k+1)$th interarrival time, a random variable which is exponentially distributed with parameter λ_k. Using the continuity theorem, $\mathbb{E}(e^{-\theta T_n}) \to \mathbb{E}(e^{-\theta T})$ as $n \to \infty$, whence $\lambda_n \widehat{p}_n(\theta) \to \mathbb{E}(e^{-\theta T})$ as $n \to \infty$, which may be inverted to obtain $\lambda_n p_n(t) \to f(t)$ as $n \to \infty$ where f is the density function of T. Now

$$\mathbb{E}\big(N(t) \,\big|\, N(t) < \infty\big) = \frac{\sum_{n=0}^{\infty} n p_n(t)}{\sum_{n=0}^{\infty} p_n(t)}$$

which converges or diverges according to whether or not $\sum_n n p_n(t)$ converges. However $p_n(t) \sim \lambda_n^{-1} f(t)$ as $n \to \infty$, so that $\sum_n n p_n(t) < \infty$ if and only if $\sum_n n\lambda_n^{-1} < \infty$.

When $\lambda_n = (n + \frac{1}{2})^2$, we have that

$$\mathbb{E}(e^{-\theta T}) = \prod_{n=0}^{\infty} \left\{1 + \frac{\theta}{(n+\frac{1}{2})^2}\right\}^{-1} = \operatorname{sech}\big(\pi\sqrt{\theta}\big).$$

Inverting the Laplace transform (or consulting a table of such transforms) we find that

$$f(t) = -\frac{1}{\pi^2} \frac{\partial}{\partial \nu} \theta_1\big(\tfrac{1}{2}\nu \,\big|\, t/\pi^2\big)\bigg|_{\nu=0}$$

where θ_1 is the first Jacobi theta function.

6.9 Solutions. Continuous-time Markov chains

1. (a) We have that

$$p_{11}' = -\mu p_{11} + \lambda p_{12}, \quad p_{22}' = -\lambda p_{22} + \mu p_{21},$$

where $p_{12} = 1 - p_{11}$, $p_{21} = 1 - p_{22}$. Solve these subject to $p_{ij}(t) = \delta_{ij}$, the Kronecker delta, to obtain that the matrix $\mathbf{P}_t = (p_{ij}(t))$ is given by

$$\mathbf{P}_t = \frac{1}{\lambda+\mu} \begin{pmatrix} \lambda + \mu e^{-(\lambda+\mu)t} & \mu - \mu e^{-(\lambda+\mu)t} \\ \lambda - \lambda e^{-(\lambda+\mu)t} & \mu + \lambda e^{-(\lambda+\mu)t} \end{pmatrix}.$$

(b) There are many ways of calculating $\mathbf{G}^n$; let us use generating functions. Note first that $\mathbf{G}^0 = \mathbf{I}$, the identity matrix. Write

$$\mathbf{G}^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}, \qquad n \geq 0,$$

and use the equation $\mathbf{G}^{n+1} = \mathbf{G} \cdot \mathbf{G}^n$ to find that

$$a_{n+1} = -\mu a_n + \mu c_n, \quad c_{n+1} = \lambda a_n - \lambda c_n.$$

Hence $a_{n+1} = -(\mu/\lambda)c_{n+1}$ for $n \geq 0$, and the first difference equation becomes $a_{n+1} = -(\lambda+\mu)a_n$, $n \geq 1$, which, subject to $a_1 = -\mu$, has solution $a_n = (-1)^n \mu(\lambda+\mu)^{n-1}$, $n \geq 1$. Therefore $c_n = (-1)^{n+1}\lambda(\lambda+\mu)^{n-1}$ for $n \geq 1$, and one may see similarly that $b_n = -a_n$, $d_n = -c_n$ for $n \geq 1$. Using the facts that $a_0 = d_0 = 1$ and $b_0 = c_0 = 0$, we deduce that $\sum_{n=0}^{\infty}(t^n/n!)\mathbf{G}^n = \mathbf{P}_t$ where $\mathbf{P}_t$ is given in part (a).

(c) With $\boldsymbol{\pi} = (\pi_1, \pi_2)$, we have that $-\mu\pi_1 + \lambda\pi_2 = 0$ and $\mu\pi_1 - \lambda\pi_2 = 0$, whence $\pi_1 = (\lambda/\mu)\pi_2$. In addition, $\pi_1 + \pi_2 = 1$ if $\pi_1 = \lambda/(\lambda+\mu) = 1 - \pi_2$.

2. (a) The required probability is

$$\frac{\mathbb{P}\big(X(t) = 2, X(3t) = 1 \mid X(0) = 1\big)}{\mathbb{P}\big(X(3t) = 1 \mid X(0) = 1\big)} = \frac{p_{12}(t)p_{21}(2t)}{p_{11}(3t)}$$

using the Markov property and the homogeneity of the process.

(b) Likewise, the required probability is

$$\frac{p_{12}(t)p_{21}(2t)p_{11}(t)}{p_{11}(3t)p_{11}(t)},$$

the same as in part (a).

3. The interarrival times and runtimes are independent and exponentially distributed. It is the lack-of-memory property which guarantees that X has the Markov property.

The state space is $S = \{0, 1, 2, \dots\}$ and the generator is

$$\mathbf{G} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ 0 & \mu & -(\lambda+\mu) & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Solutions of the equation $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$ satisfy

$$-\lambda\pi_0 + \mu\pi_1 = 0, \quad \lambda\pi_{j-1} - (\lambda+\mu)\pi_j + \mu\pi_{j+1} = 0 \text{ for } j \geq 1,$$

with solution $\pi_i = \pi_0(\lambda/\mu)^i$. We have in addition that $\sum_i \pi_i = 1$ if $\lambda < \mu$ and $\pi_0 = \{1 - (\lambda/\mu)\}^{-1}$.

4. One may use the strong Markov property. Alternatively, by the Markov property,

$$\mathbb{P}(Y_{n+1} = j \mid Y_n = i, T_n = t, B) = \mathbb{P}(Y_{n+1} = j \mid Y_n = i, T_n = t)$$

for any event B defined in terms of $\{X(s) : s \leq T_n\}$. Hence

$$\begin{aligned} \mathbb{P}(Y_{n+1} = j \mid Y_n = i, B) &= \int_0^{\infty} \mathbb{P}(Y_{n+1} = j \mid Y_n = i, T_n = t) f_{T_n}(t)\, dt \\ &= \mathbb{P}(Y_{n+1} = j \mid Y_n = i), \end{aligned}$$

so that Y is a Markov chain. Now $q_{ij} = \mathbb{P}(Y_{n+1} = j \mid Y_n = i)$ is given by

$$q_{ij} = \int_0^{\infty} p_{ij}(t)\lambda e^{-\lambda t}\, dt,$$

by conditioning on the $(n+1)$th interarrival time of N; here, as usual, $p_{ij}(t)$ is a transition probability of X. Now

$$\sum_i \pi_i q_{ij} = \int_0^\infty \Big(\sum_i \pi_i p_{ij}(t)\Big)\lambda e^{-\lambda t}\,dt = \int_0^\infty \pi_j \lambda e^{-\lambda t}\,dt = \pi_j.$$

5. The jump chain $Z = \{Z_n : n \ge 0\}$ has transition probabilities $h_{ij} = g_{ij}/g_i$, $i \ne j$. The chance that Z ever reaches A from j is also η_j, and $\eta_j = \sum_k h_{jk}\eta_k$ for $j \notin A$, by Exercise (6.3.6). Hence $-g_j\eta_j = \sum_k g_{jk}\eta_k$, as required.

6. Let $T_1 = \inf\{t : X(t) \ne X(0)\}$, and more generally let T_m be the time of the mth change in value of X. For $j \notin A$,

$$\mu_j = \mathbb{E}_j(T_1) + \sum_{k \ne j} h_{jk}\mu_k,$$

where $\mathbb{E}_j$ denotes expectation conditional on $X_0 = j$. Now $\mathbb{E}_j(T_1) = g_j^{-1}$, and the given equations follow. Suppose next that $(a_k : k \in S)$ is another non-negative solution of these equations. With $U_i = T_{i+1} - T_i$ and $R = \min\{n \ge 1 : Z_n \in A\}$, we have for $j \notin A$ that

$$\begin{aligned} a_j &= \frac{1}{g_j} + \sum_{k \notin A} h_{jk}a_k = \frac{1}{g_j} + \sum_{k\notin A} h_{jk}\bigg\{\frac{1}{g_k} + \sum_{m \notin A} h_{km}a_m\bigg\} \\ &= \mathbb{E}_j(U_0) + \mathbb{E}_j(U_1 I_{\{R>1\}}) + \mathbb{E}_j(U_2 I_{\{R>2\}}) + \cdots + \mathbb{E}_j(U_n I_{\{R>n\}}) + \Sigma, \end{aligned}$$

where Σ is a sum of non-negative terms. It follows that

$$\begin{aligned} a_j &\ge \mathbb{E}_j(U_0) + \mathbb{E}_j(U_1 I_{\{R>1\}}) + \cdots + \mathbb{E}_j(U_n I_{\{R>n\}}) \\ &= \mathbb{E}_j\bigg(\sum_{r=0}^{n} U_r I_{\{R>r\}}\bigg) = \mathbb{E}_j\big(\min\{T_n, H_A\}\big) \to \mathbb{E}_j(H_A) \end{aligned}$$

as $n \to \infty$, by monotone convergence. Therefore, $\boldsymbol{\mu}$ is the minimal non-negative solution.

7. First note that i is persistent if and only if it is also a persistent state in the jump chain Z. The integrand being positive, we can write

$$\int_0^\infty p_{ii}(t)\,dt = \mathbb{E}\bigg[\int_0^\infty I_{\{X(t)=i\}}\,dt \,\bigg|\, X(0) = i\bigg] = \mathbb{E}\bigg[\sum_{n=0}^{\infty}(T_{n+1} - T_n)I_{\{Y_n=i\}} \,\bigg|\, Y_0 = i\bigg]$$

where $\{T_n : n \ge 1\}$ are the times of the jumps of X. The right side equals

$$\sum_{n=0}^{\infty} \mathbb{E}(T_1 \mid X(0) = i)h_{ii}(n) = \frac{1}{g_i}\sum_{n=0}^{\infty} h_{ii}(n)$$

where $\mathbf{H} = (h_{ij})$ is the transition matrix of Z. The sum diverges if and only if i is persistent for Z.

8. Since the imbedded jump walk is persistent, so is X. The probability of visiting m during an excursion is $\alpha = (2m)^{-1}$, since such a visit requires an initial step to the right, followed by a visit to m before 0, cf. Example (3.9.6). Having arrived at m, the chance of returning to m before visiting 0 is $1 - \alpha$, by the same argument with 0 and m interchanged. In this way one sees that the number N of visits to m during an excursion from 0 has distribution given by $\mathbb{P}(N \ge k) = \alpha(1-\alpha)^{k-1}$, $k \ge 1$. The 'total jump rate' from any state is λ, whence T may be expressed as $\sum_{i=0}^{N} V_i$ where the V_i are exponential with parameter λ. Therefore,

$$\mathbb{E}(e^{\theta T}) = G_N\left(\frac{\lambda}{\lambda - \theta}\right) = (1-\alpha) + \alpha\frac{\alpha\lambda}{\alpha\lambda - \theta}.$$

The distribution of T is a mixture of an atom at 0 and the exponential distribution with parameter $\alpha\lambda$.

9. The number N of sojourns in i has a geometric distribution $\mathbb{P}(N = k) = f^{k-1}(1 - f)$, $k \ge 1$, for some $f < 1$. The length of each of these sojourns has the exponential distribution with some parameter g_i. By the independence of these lengths, the total time T in state i has moment generating function

$$\mathbb{E}(e^{\theta T}) = \sum_{k=1}^{\infty} f^{k-1}(1-f)\left(\frac{g_i}{g_i - \theta}\right)^k = \frac{g_i(1-f)}{g_i(1-f) - \theta}.$$

The distribution of T is exponential with parameter $g_i(1 - f)$.

10. The jump chain is the simple random walk with probabilities $\lambda/(\lambda + \mu)$ and $\mu/(\lambda + \mu)$, and with $p_{01} = 1$. By Corollary (5.3.6), the chance of ever hitting 0 having started at 1 is μ/λ, whence the probability of returning to 0 having started there is $f = \mu/\lambda$. By the result of Exercise (6.9.9),

$$\mathbb{E}(e^{\theta V_0}) = \frac{\lambda - \mu}{\lambda - \mu - \theta},$$

as required. Having started at 0, the walk visits the state $r \ge 1$ with probability 1. The probability of returning to r having started there is

$$f_r = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu}\cdot\frac{\mu}{\lambda} = \frac{2\mu}{\lambda + \mu},$$

and each sojourn is exponentially distributed with parameter $g_r = \lambda + \mu$. Now $g_r(1 - f_r) = \lambda - \mu$, whence, as above,

$$\mathbb{E}(e^{\theta V_r}) = \frac{\lambda - \mu}{\lambda - \mu - \theta}.$$

The probability of ever reaching 0 from $X(0)$ is $(\mu/\lambda)^{X(0)}$, and the time spent there subsequently is exponential with parameter $\lambda - \mu$. Therefore, the mean total time spent at 0 is

$$\mathbb{E}\left(\frac{(\mu/\lambda)^{X(0)}}{\lambda - \mu}\right) = \frac{G(\mu/\lambda)}{\lambda - \mu}.$$

11. (a) The imbedded chain has transition probabilities

$$h_{kj} = \begin{cases} g_{kj}/g_k & \text{if } k \ne j, \\ 0 & \text{if } k = j, \end{cases}$$

where $g_k = -g_{kk}$. Therefore, for any state j,

$$\sum_k \widehat{\pi}_k h_{kj} = \frac{\sum_{k \ne j} \pi_k g_k (g_{kj}/g_k)}{\sum_i \pi_i g_i} = \frac{\pi_j g_j}{\sum_i \pi_i g_i} = \widehat{\pi}_j,$$

where we have used the fact that $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$. Also $\widehat{\pi}_k \ge 0$ and $\sum_k \widehat{\pi}_k = 1$, and therefore $\widehat{\boldsymbol{\pi}}$ is a stationary distribution of Y.

Clearly $\widehat{\pi}_k = \pi_k$ for all k if and only if $g_k = \sum_i \pi_i g_i$ for all k, which is to say that $g_i = g_k$ for all pairs i, k. This requires that the 'holding times' have the same distribution.

(b) Let T_n be the time of the nth change of value of X, with $T_0 = 0$, and let $U_n = T_{n+1} - T_n$. Fix a state k, and let $H = \min\{n \ge 1 : Z_n = k\}$. Let $\gamma_i(k)$ be the mean time spent in state i between two consecutive visits to k, and let $\widehat{\gamma}_i(k)$ be the mean number of visits to i by the jump chain in between two visits to k (so that, in particular, $\gamma_k(k) = g_k^{-1}$ and $\widehat{\gamma}_k(k) = 1$). With $\mathbb{E}_j$ and $\mathbb{P}_j$ denoting expectation and probability conditional on $X(0) = j$, we have that

$$\begin{aligned}\gamma_i(k) &= \mathbb{E}_k\left(\sum_{n=0}^{\infty} U_n I_{\{Z_n=i,\ H>n\}}\right) = \sum_{n=0}^{\infty} \mathbb{E}_k(U_n \mid I_{\{Z_n=i\}})\mathbb{P}_k(Z_n = i,\ H > n)\\ &= \sum_{n=0}^{\infty} \frac{1}{g_i}\mathbb{P}_k(Z_n = i,\ H > n) = \frac{1}{g_i}\widehat{\gamma}_i(k).\end{aligned}$$

The vector $\widehat{\gamma}(k) = (\widehat{\gamma}_i(k) : i \in S)$ satisfies $\widehat{\gamma}(k)\mathbf{H} = \widehat{\gamma}(k)$, by Lemma (6.4.5), where $\mathbf{H}$ is the transition matrix of the jump chain Z. That is to say,

$$\sum_{i:i\neq j} \gamma_i(k)\frac{g_{ij}}{g_i} = \gamma_j(k) \qquad \text{for } j \in S,$$

whence $\sum_i \gamma_i(k)g_{ij} = 0$ for all $j \in S$. If $\mu_k = \sum_i \gamma_i(k) < \infty$, the vector $(\gamma_i(k)/\mu_k)$ is a stationary distribution for X, whence $\pi_i = \gamma_i(k)/\mu_k$ for all i. Setting $i = k$ we deduce that $\pi_k = 1/(g_k\mu_k)$.

Finally, if $\sum_i \pi_i g_i < \infty$, then

$$\widehat{\mu}_k = \frac{1}{\widehat{\pi}_k} = \frac{\sum_i \pi_i g_i}{\pi_k g_k} = \mu_k \sum_i \pi_i g_i.$$

12. Define the generator $\mathbf{G}$ by $g_{ii} = -\nu_i$, $g_{ij} = \nu_i h_{ij}$, so that the imbedded chain has transition matrix $\mathbf{H}$. A root of the equation $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$ satisfies

$$0 = \sum_i \pi_i g_{ij} = -\pi_j\nu_j + \sum_{i:i\neq j} (\pi_i\nu_i)h_{ij}$$

whence the vector $\boldsymbol{\zeta} = (\pi_j\nu_j : j \in S)$ satisfies $\boldsymbol{\zeta} = \boldsymbol{\zeta}\mathbf{H}$. Therefore $\boldsymbol{\zeta} = \alpha\boldsymbol{\nu}$, which is to say that $\pi_j\nu_j = \alpha\nu_j$, for some constant α. Now $\nu_j > 0$ for all j, so that $\pi_j = \alpha$, which implies that $\sum_j \pi_j \neq 1$. Therefore the continuous-time chain X with generator $\mathbf{G}$ has no stationary distribution.

6.11 Solutions. Birth–death processes and imbedding

1. The jump chain is a walk $\{Z_n\}$ on the set $S = \{0, 1, 2, \dots\}$ satisfying, for $i \geq 1$,

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i) = \begin{cases} p_i & \text{if } j = i+1,\\ 1 - p_i & \text{if } j = i-1,\end{cases}$$

where $p_i = \lambda_i/(\lambda_i + \mu_i)$. Also $\mathbb{P}(Z_{n+1} = 1 \mid Z_n = 0) = 1$.

2. The transition matrix $\mathbf{H} = (h_{ij})$ of Z is given by

$$h_{ij} = \begin{cases} \dfrac{i\mu}{\lambda + i\mu} & \text{if } j = i-1,\\[2mm] \dfrac{\lambda}{\lambda + i\mu} & \text{if } j = i+1.\end{cases}$$

To find the stationary distribution of Y, either solve the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{Q}$, or look for a solution of the detailed balance equations $\pi_i h_{i,i+1} = \pi_{i+1}h_{i+1,i}$. Following the latter route, we have that

$$\pi_i = \pi_0 \frac{h_{01}h_{12}\cdots h_{i-1,i}}{h_{i,i-1}\cdots h_{21}h_{10}}, \qquad i \geq 1,$$

whence $\pi_i = \pi_0 \rho^i (1 + i/\rho)/i!$ for $i \geq 1$. Choosing π_0 accordingly, we obtain the result.

It is a standard calculation that X has stationary distribution $\boldsymbol{\nu}$ given by $\nu_i = \rho^i e^{-\rho}/i!$ for $i \geq 0$. The difference between $\boldsymbol{\pi}$ and $\boldsymbol{\nu}$ arises from the fact that the holding-times of X have distributions which depend on the current state.

3. We have, by conditioning on $X(h)$, that

$$\begin{aligned}\eta(t+h) &= \mathbb{E}\{\mathbb{P}(X(t+h) = 0 \mid X(h))\}\\ &= \mu h \cdot 1 + (1 - \lambda h - \mu h)\eta(t) + \lambda h \xi(t) + \mathrm{o}(h)\end{aligned}$$

where $\xi(t) = \mathbb{P}(X(t) = 0 \mid X(0) = 2)$. The process X may be thought of as a collection of particles each of which dies at rate μ and divides at rate λ, different particles enjoying a certain independence; this is a consequence of the linearity of λ_n and μ_n. Hence $\xi(t) = \eta(t)^2$, since each of the initial pair is required to have no descendants at time t. Therefore

$$\eta'(t) = \mu - (\lambda + \mu)\eta(t) + \lambda\eta(t)^2$$

subject to $\eta(0) = 0$.

Rewrite the equation as

$$\frac{\eta'}{(1-\eta)(\mu - \lambda\eta)} = 1$$

and solve using partial fractions to obtain

$$\eta(t) = \begin{cases} \dfrac{\lambda t}{\lambda t + 1} & \text{if } \lambda = \mu, \\ \dfrac{\mu(1 - e^{t(\mu-\lambda)})}{\lambda - \mu e^{t(\mu-\lambda)}} & \text{if } \lambda \neq \mu. \end{cases}$$

Finally, if $0 < t < u$,

$$\mathbb{P}(X(t) = 0 \mid X(u) = 0) = \mathbb{P}(X(u) = 0 \mid X(t) = 0)\frac{\mathbb{P}(X(t) = 0)}{\mathbb{P}(X(u) = 0)} = \frac{\eta(t)}{\eta(u)}.$$

4. The random variable $X(t)$ has generating function

$$G(s,t) = \frac{\mu(1-s) - (\mu - \lambda s)e^{-t(\lambda-\mu)}}{\lambda(1-s) - (\mu - \lambda s)e^{-t(\lambda-\mu)}}$$

as usual. The generating function of $X(t)$, conditional on $\{X(t) > 0\}$, is therefore

$$\sum_{n=1}^{\infty} s^n \frac{\mathbb{P}(X(t) = n)}{\mathbb{P}(X(t) > 0)} = \frac{G(s,t) - G(0,t)}{1 - G(0,t)}.$$

Substitute for G and take the limit as $t \to \infty$ to obtain as limit

$$H(s) = \frac{(\mu - \lambda)s}{\mu - \lambda s} = \sum_{n=1}^{\infty} s^n p_n$$

where, with $\rho = \lambda/\mu$, we have that $p_n = \rho^{n-1}(1 - \rho)$ for $n \geq 1$.

5. Extinction is certain if $\lambda < \mu$, and in this case, by Theorem (6.11.10),

$$\begin{aligned}\mathbb{E}(T) &= \int_0^\infty \mathbb{P}(T > t)\,dt = \int_0^\infty \left\{1 - \mathbb{E}(s^{X(t)})\big|_{s=0}\right\} dt \\ &= \int_0^\infty \frac{(\mu-\lambda)e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}\,dt = \frac{1}{\lambda}\log\left(\frac{\mu}{\mu-\lambda}\right).\end{aligned}$$

If $\lambda > \mu$ then $\mathbb{P}(T < \infty) = \mu/\lambda$, so

$$\mathbb{E}(T \mid T < \infty) = \int_0^\infty \left\{1 - \frac{\lambda}{\mu}\mathbb{E}(s^{X(t)})\big|_{s=0}\right\} dt = \int_0^\infty \frac{(\lambda-\mu)e^{(\mu-\lambda)t}}{\lambda - \mu e^{(\mu-\lambda)t}}\,dt = \frac{1}{\mu}\log\left(\frac{\lambda}{\lambda-\mu}\right).$$

In the case $\lambda = \mu$, $\mathbb{P}(T < \infty) = 1$ and $\mathbb{E}(T) = \infty$.

6. By considering the imbedded random walk, we find that the probability of ever returning to 1 is $\max\{\lambda, \mu\}/(\lambda + \mu)$, so that the number of visits is geometric with parameter $\min\{\lambda, \mu\}/(\lambda + \mu)$. Each visit has an exponentially distributed duration with parameter $\lambda + \mu$, and a short calculation using moment generating functions shows that $V_1(\infty)$ is exponential with parameter $\min\{\lambda, \mu\}$.

Next, by a change of variables, Theorem (6.11.10), and some calculation,

$$\begin{aligned}\sum_r s^r \mathbb{E}(V_r(t)) &= \mathbb{E}\left(\sum_r \int_0^t s^r I_{\{X(u)=r\}}\,du\right) = \mathbb{E}\left(\int_0^t s^{X(u)}\,du\right) \\ &= \int_0^t \mathbb{E}(s^{X(u)})\,du = \frac{\mu t}{\lambda} - \frac{1}{\lambda}\log\left\{\frac{\lambda(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}}{\lambda - \mu}\right\} \\ &= -\frac{1}{\lambda}\log\left\{1 - \frac{\lambda s(e^{\rho t} - 1)}{\mu e^{\rho t} - \lambda}\right\} + \text{terms not involving } s,\end{aligned}$$

where $\rho = \mu - \lambda$. We take the limit as $t \to \infty$ and we pick out the coefficient of s^r.

7. If $\lambda = \mu$ then, by Theorem (6.11.10),

$$\mathbb{E}(s^{X(t)}) = \frac{\lambda t(1-s) + s}{\lambda t(1-s) + 1} = 1 - \frac{1-s}{\lambda t(1-s) + 1},$$

and

$$\begin{aligned}\int_0^t \mathbb{E}(s^{X(u)})\,du &= t - \frac{1}{\lambda}\log\{\lambda t(1-s) + 1\} \\ &= -\frac{1}{\lambda}\log\left\{1 - \frac{\lambda t s}{1 + \lambda t}\right\} + \text{terms not involving } s.\end{aligned}$$

Letting $t \to \infty$ and picking out the coefficient of s^r gives $\mathbb{E}(V_r(\infty)) = (r\lambda)^{-1}$. An alternative method utilizes the imbedded simple random walk and the exponentiality of the sojourn times.

6.12 Solutions. Special processes

1. The jump chain is simple random walk with step probabilities $\lambda/(\lambda + \mu)$ and $\mu/(\lambda + \mu)$. The expected time μ_{10} to pass from 1 to 0 satisfies

$$\mu_{10} = 1 + \frac{\lambda}{\lambda + \mu}(\mu_{21} + \mu_{10}) = 1 + \frac{2\lambda}{\lambda + \mu}\mu_{10},$$

whence $\mu_{10} = (\mu + \lambda)/(\mu - \lambda)$. Since each sojourn is exponentially distributed with parameter $\mu + \lambda$, the result follows by an easy calculation. See also Theorem (11.3.17).

2. We apply the method of Theorem (6.12.11) with

$$G_N(s, u) = \frac{se^{-\lambda u}}{1 - s + se^{-\lambda u}},$$

the probability generating function of the population size at time u in a simple birth process. In the absence of disasters, the probability generating function of the ensuing population size at time v is

$$H(s, v) = \exp\left(\nu \int_0^v [G_N(s, u) - 1]\, du\right) = \left\{s + (1 - s)e^{\lambda v}\right\}^{-\nu/\lambda}.$$

The individuals alive at time t arose subsequent to the most recent disaster at time $t - D$, where D has density function $\delta e^{-\delta x}$, $x > 0$. Therefore,

$$\mathbb{E}(s^{X(t)}) = \mathbb{E}(H(s, D)) = \int_0^t \frac{\delta e^{-\delta x} e^{-\nu t}\, dx}{(1 - s + se^{-\lambda t})^{\nu/\delta}} + e^{-\delta t} \frac{e^{-\nu t}}{(1 - s + se^{-\lambda t})^{\nu/\delta}}.$$

3. The mean number of descendants after time t of a single progenitor at time 0 is $e^{(\lambda - \mu)t}$. The expected number due to the arrival of a single individual at a uniformly distributed time in the interval on $[0, x]$ is therefore

$$\frac{1}{x} \int_0^x e^{(\lambda - \mu)u}\, du = \frac{e^{(\lambda - \mu)x} - 1}{(\lambda - \mu)x}.$$

The aggregate effect at time x of N earlier arrivals is the same, by Theorem (6.12.7), as that of N arrivals at independent times which are uniformly distributed on $[0, x]$. Since $\mathbb{E}(N) = \nu x$, the mean population size at time x is $\nu[e^{(\lambda - \mu)x} - 1]/(\lambda - \mu)$. The most recent disaster occurred at time $t - D$, where D has density function $\delta e^{-\delta x}$, $x > 0$, and it follows that

$$\mathbb{E}(X(t)) = \int_0^t \delta e^{-\delta x} \frac{\nu}{\lambda - \mu} [e^{(\lambda - \mu)x} - 1]\, dx + \frac{\nu}{\lambda - \mu} e^{-\delta x} [e^{(\lambda - \mu)t} - 1].$$

This is bounded as $t \to \infty$ if and only if $\delta > \lambda - \mu$.

4. Let N be the number of clients who arrive during the interval $[0, t]$. Conditional on the event $\{N = n\}$, the arrival times have, by Theorem (6.12.7), the same joint distribution as n independent variables chosen uniformly from $[0, t]$. The probability that an arrival at a uniform time in $[0, t]$ is still in service at time t is $\beta = \int_0^t [1 - G(t - x)] t^{-1}\, dx$, whence, conditional on $\{N = n\}$, the total number M still in service is $\text{bin}(n, \beta)$. Therefore,

$$\mathbb{E}(e^{\theta M}) = \mathbb{E}\big(\mathbb{E}(e^{\theta M} \mid N)\big) = \mathbb{E}\big((\beta e^\theta + 1 - \beta)^N\big) = G_N(\beta e^\theta + 1 - \beta) = e^{\lambda \beta t (e^\theta - 1)},$$

whence M has the Poisson distribution with parameter $\lambda \beta t = \lambda \int_0^t [1 - G(x)]\, dx$. Note that this parameter approaches $\lambda \mathbb{E}(S)$ as $t \to \infty$.

6.13 Solutions. Spatial Poisson processes

1. It is easy to check from the axioms that the combined process $N(t) = B(t) + G(t)$ is a Poisson process with intensity $\beta + \gamma$.
(a) The time S (respectively, T) until the arrival of the first brown (respectively, grizzly) bear is exponentially distributed with parameter β (respectively, γ), and these times are independent. Now,

$$\mathbb{P}(S < T) = \int_0^\infty \beta e^{-\beta s} e^{-\gamma s}\, ds = \frac{\beta}{\beta + \gamma}.$$

(b) Using (a), and the lack-of-memory of the process, the required probability is

$$\left(\frac{\gamma}{\beta + \gamma}\right)^n \frac{\beta}{\beta + \gamma}.$$

(c) Using Theorem (6.12.7),

$$\mathbb{E}\big(\min\{S, T\} \mid B(1) = 1\big) = \mathbb{E}\left\{\frac{1}{G(1) + 2}\right\} = \frac{\gamma - 1 + e^{-\gamma}}{\gamma^2}.$$

2. Let B_r be the ball with centre $\mathbf{0}$ and radius r, and let $N_r = |\Pi \cap B_r|$. We have by Theorem (6.13.11) that $S_r = \sum_{\mathbf{x}\in\Pi\cap B_r} g(\mathbf{x})$ satisfies

$$\mathbb{E}(S_r \mid N_r) = N_r \int_{B_r} g(\mathbf{u}) \frac{\lambda(\mathbf{u})}{\Lambda(B_r)}\, d\mathbf{u},$$

where $\Lambda(B) = \int_{\mathbf{y}\in B} \lambda(\mathbf{y})\, d\mathbf{y}$. Therefore, $\mathbb{E}(S_r) = \int_{B_r} g(\mathbf{u})\lambda(\mathbf{u})\, d\mathbf{u}$, implying by monotone convergence that $\mathbb{E}(S) = \int_{\mathbb{R}^d} g(\mathbf{u})\lambda(\mathbf{u})\, d\mathbf{u}$. Similarly,

$$\begin{aligned}
\mathbb{E}(S_r^2 \mid N_r) &= \mathbb{E}\left(\left[\sum_{\mathbf{x}\in\Pi\cap B_r} g(\mathbf{x})\right]^2\right) \\
&= \mathbb{E}\left(\sum_{\mathbf{x}\in\Pi\cap B_r} g(\mathbf{x})^2)\right) + \mathbb{E}\left(\sum_{\substack{\mathbf{x}\neq\mathbf{y}\\ \mathbf{x},\mathbf{y}\in\Pi\cap B_r}} g(\mathbf{x})g(\mathbf{y})\right) \\
&= N_r \int_{B_r} g(\mathbf{u})^2 \frac{\lambda(\mathbf{u})}{\Lambda(B_r)}\, d\mathbf{u} + N_r(N_r - 1) \iint_{\mathbf{u},\mathbf{v}\in B_r} g(\mathbf{u})g(\mathbf{v}) \frac{\lambda(\mathbf{u})\lambda(\mathbf{v})}{\Lambda(B_r)^2}\, d\mathbf{u}\, d\mathbf{v},
\end{aligned}$$

whence

$$\mathbb{E}(S_r^2) = \int_{B_r} g(\mathbf{u})^2\lambda(\mathbf{u})\, d\mathbf{u} + \left(\int_{B_r} g(\mathbf{u})\lambda(\mathbf{u})\, d\mathbf{u}\right)^2.$$

By monotone convergence,

$$\mathbb{E}(S^2) = \int_{B_r} g(\mathbf{u})^2\lambda(\mathbf{u})\, d\mathbf{u} + \mathbb{E}(S)^2,$$

and the formula for the variance follows.

3. If $B_1, B_2, \dots, B_n$ are disjoint regions of the disc, then the numbers of projected points therein are Poisson-distributed and independent, since they originate from disjoint regions of the sphere. By elementary coordinate geometry, the intensity function in plane polar coordinates is $2\lambda/\sqrt{1-r^2}$, $0 \le r \le 1$, $0 \le \theta < 2\pi$.

4. The same argument is valid with resulting intensity function $2\lambda\sqrt{1-r^2}$.

5. The Mercator projection represents the spherical coordinates (θ, ϕ) as Cartesian coordinates in the range $0 \le \phi < 2\pi$, $0 \le \theta \le \pi$. (Recall that θ is the angle made with the axis through the north pole.) Therefore a uniform intensity on the globe corresponds to an intensity function $\lambda \sin\theta$ on the map. Likewise, a uniform intensity on the map corresponds to an intensity $\lambda/\sin\theta$ on the globe.

6. Let the X_r have characteristic function ϕ. Conditional on the value of $N(t)$, the corresponding arrival times have the same distribution as $N(t)$ independent variables with the uniform distribution, whence

$$\begin{aligned}\mathbb{E}(e^{i\theta S(t)}) &= \mathbb{E}\{\mathbb{E}(e^{i\theta S(t)} \mid N(t))\} = \mathbb{E}\{\mathbb{E}(e^{i\theta X e^{-\alpha U}})^{N(t)}\}\\ &= \exp\{\lambda t(\mathbb{E}(e^{i\theta X e^{-\alpha U}}) - 1)\} = \exp\left\{\lambda \int_0^t \{\phi(\theta e^{-\alpha u}) - 1\}\, du\right\},\end{aligned}$$

where U is uniformly distributed on $[0, t]$. By differentiation,

$$\begin{aligned}\mathbb{E}(S(t)) &= -i\phi'_{S(t)}(0) = \frac{\lambda}{\alpha}\mathbb{E}(X)(1 - e^{-\alpha t}),\\ \mathbb{E}(S(t)^2) &= -\phi''_{S(t)}(0) = \mathbb{E}(S(t))^2 + \frac{\lambda\mathbb{E}(X^2)}{2\alpha}(1 - e^{-2\alpha t}).\end{aligned}$$

Now, for $s < t$, $S(t) = S(s)e^{-\alpha(t-s)} + \widehat{S}(t-s)$ where $\widehat{S}(t-s)$ is independent of $S(s)$ with the same distribution as $S(t-s)$. Hence, for $s < t$,

$$\operatorname{cov}\big(S(s), S(t)\big) = \operatorname{var}(S(s))e^{-\alpha(t-s)} = \frac{\lambda\mathbb{E}(X^2)}{2\alpha}(1 - e^{-2\alpha s})e^{-\alpha(t-s)} \to \frac{\lambda\mathbb{E}(X^2)}{2\alpha}e^{-\alpha v}$$

as $s \to \infty$ with $v = t - s$ fixed. Therefore, $\rho(S(s), S(s+v)) \to e^{-\alpha v}$ as $s \to \infty$.

7. The first two arrival times T_1, T_2 satisfy

$$\mathbb{P}(T_1 \le x,\ T_2 - T_1 > y) = \int_0^x \lambda(u)e^{-\Lambda(u)}e^{-(\Lambda(u+y)-\Lambda(u))}\, du = \int_0^x \lambda(u)e^{-\Lambda(u+y)}\, du.$$

Differentiate with respect to x and y to obtain the joint density function $\lambda(x)\lambda(x+y)e^{-\Lambda(x+y)}$, $x, y \ge 0$. Since this does not generally factorize as the product of a function of x and a function of y, T_1 and T_2 are dependent in general.

8. Let X_i be the time of the first arrival in the process N_i. Then

$$\begin{aligned}\mathbb{P}(I = 1,\ T \ge t) &= \mathbb{P}\big(t \le X_1 < \inf\{X_2, X_3, \dots\}\big)\\ &= \int_t^\infty \mathbb{P}\big(\inf\{X_2, X_3, \dots\} > x\big)\lambda_1 e^{-\lambda_1 x}\, dx = \frac{\lambda_1}{\lambda}e^{-\lambda t}.\end{aligned}$$

6.14 Solutions. Markov chain Monte Carlo

1. If $\mathbf{P}$ is reversible then

$$\text{RHS} = \sum_i \Big(\sum_j p_{ij} x_j\Big) y_i \pi_i = \sum_{i,j} \pi_i p_{ij} x_j y_i = \sum_{i,j} \pi_j p_{ji} y_i x_j = \sum_j \pi_j x_j \Big(\sum_i p_{ji} y_i\Big) = \text{LHS}.$$

Suppose conversely that $\langle \mathbf{x}, \mathbf{Py}\rangle = \langle \mathbf{Px}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in l^2(\pi)$. Choose $\mathbf{x}, \mathbf{y}$ to be unit vectors with 1 in the ith and jth place respectively, to obtain the detailed balance equations $\pi_i p_{ij} = \pi_j p_{ji}$.

2. Just check that $0 \le b_{ij} \le 1$ and that the $p_{ij} = g_{ij} b_{ij}$ satisfy the detailed balance equations (6.14.3).

3. It is immediate that $p_{jk} = |A_{jk}|$, the Lebesgue measure of A_{jk}. This is a method for simulating a Markov chain with a given transition matrix.

4. (a) Note first from equation (4.12.7) that $d(\mathbf{U}) = \frac{1}{2}\sup_{i\ne j} d_{\mathrm{TV}}(u_{i\cdot}, u_{j\cdot})$, where $u_{i\cdot}$ is the mass function u_{it}, $t \in T$. The required inequality may be hacked out, but instead we will use the maximal coupling of Exercises (4.12.4, 5); see also Problem (7.11.16). Thus requires a little notation. For $i, j \in S$, $i \ne j$, we find a pair (X_i, X_j) of random variables taking values in T according to the marginal mass functions $u_{i\cdot}$, $u_{j\cdot}$, and such that $\mathbb{P}(X_i \ne X_j) = \frac{1}{2} d_{\mathrm{TV}}(u_{i\cdot}, u_{j\cdot})$. The existence of such a pair was proved in Exercise (4.12.5). Note that the value of X_i depends on j, but this fact has been suppressed from the notation for ease of reading. Having found (X_i, X_j), we find a pair $(Y(X_i), Y(X_j))$ taking values in U according to the marginal mass functions $v_{X_i\cdot}$, $v_{X_j\cdot}$, and such that $\mathbb{P}(Y(X_i) \ne Y(X_j) \mid X_i, X_j) = \frac{1}{2} d_{\mathrm{TV}}(v_{X_i\cdot}, v_{X_j\cdot})$. Now, taking a further liberty with the notation,

$$\begin{aligned}
\mathbb{P}\big(Y(X_i) \ne Y(X_j)\big) &= \sum_{\substack{r,s\in S\\ r\ne s}} \mathbb{P}(X_i = r,\ X_j = s)\mathbb{P}\big(Y(r) \ne Y(s)\big)\\
&= \sum_{\substack{r,s\in S\\ r\ne s}} \mathbb{P}(X_i = r,\ X_j = s)\tfrac{1}{2} d_{\mathrm{TV}}(v_{r\cdot}, v_{s\cdot})\\
&\le \Big\{\tfrac{1}{2}\sup_{r\ne s} d_{\mathrm{TV}}(v_{r\cdot}, v_{s\cdot})\Big\}\mathbb{P}(X_i \ne X_j),
\end{aligned}$$

whence

$$d(\mathbf{UV}) = \sup_{i\ne j} \mathbb{P}\big(Y(X_i) \ne Y(X_j)\big) \le \Big\{\tfrac{1}{2}\sup_{r\ne s} d_{\mathrm{TV}}(v_{r\cdot}, v_{s\cdot})\Big\}\Big\{\sup_{i,j} \mathbb{P}(X_i \ne X_j)\Big\}$$

and the claim follows.

(b) Write $S = \{1, 2, \dots, m\}$, and take

$$\mathbf{U} = \begin{pmatrix} \mathbb{P}(X_0 = 1) & \mathbb{P}(X_0 = 2) & \cdots & \mathbb{P}(X_0 = m)\\ \mathbb{P}(Y_0 = 1) & \mathbb{P}(Y_0 = 2) & \cdots & \mathbb{P}(Y_0 = m) \end{pmatrix}.$$

The claim follows by repeated application of the result of part (a).

It may be shown with the aid of a little matrix theory that the second largest eigenvalue of a finite stochastic matrix $\mathbf{P}$ is no larger in modulus that $d(\mathbf{P})$; cf. the equation prior to Theorem (6.14.9).

6.15 Solutions to problems

1. (a) The state 4 is absorbing. The state 3 communicates with 4, and is therefore transient. The set $\{1, 2\}$ is finite, closed, and aperiodic, and hence ergodic. We have that $f_{34}(n) = (\frac{1}{4})^{n-1}\frac{1}{2}$, so that $f_{34} = \sum_n f_{34}(n) = \frac{2}{3}$.
(b) The chain is irreducible with period 2. All states are non-null persistent. Solve the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ to find the stationary distribution $\boldsymbol{\pi} = \left(\frac{3}{8}, \frac{3}{16}, \frac{5}{16}, \frac{1}{8}\right)$ whence the mean recurrence times are $\frac{8}{3}, \frac{16}{3}, \frac{16}{5}, 8$, in order.

2. (a) Let $\mathbf{P}$ be the transition matrix, assumed to be doubly stochastic. Then

$$\sum_i p_{ij}(n) = \sum_i \sum_k p_{ik}(n-1)p_{kj} = \sum_k \Big(\sum_i p_{ik}(n-1)\Big) p_{kj}$$

whence, by induction, the n-step transition matrix $\mathbf{P}^n$ is doubly stochastic for all $n \geq 1$.

If j is not non-null persistent, then $p_{ij}(n) \to 0$ as $n \to \infty$, for all i, implying that $\sum_i p_{ij}(n) \to 0$, a contradiction. Therefore all states are non-null persistent.

If in addition the chain is irreducible and aperiodic then $p_{ij}(n) \to \pi_j$, where $\boldsymbol{\pi}$ is the unique stationary distribution. However, it is easy to check that $\boldsymbol{\pi} = (N^{-1}, N^{-1}, \dots, N^{-1})$ is a stationary distribution if $\mathbf{P}$ is doubly stochastic.
(b) Suppose the chain is persistent. In this case there exists a positive root of the equation $\mathbf{x} = \mathbf{xP}$, this root being unique up to a multiplicative constant (see Theorem (6.4.6) and the forthcoming Problem (7)). Since the transition matrix is doubly stochastic, we may take $\mathbf{x} = \mathbf{1}$, the vector of 1's. By the above uniqueness of $\mathbf{x}$, there can exist no stationary distribution, and therefore the chain is null. We deduce that the chain cannot be non-null persistent.

3. By the Chapman–Kolmogorov equations,

$$p_{ii}(m+r+n) \geq p_{ij}(m)p_{jj}(r)p_{ji}(n), \qquad m, r, n \geq 0.$$

Choose two states i and j, and pick m and n such that $\alpha = p_{ij}(m)p_{ji}(n) > 0$. Then

$$p_{ii}(m+r+n) \geq \alpha p_{jj}(r).$$

Set $r = 0$ to find that $p_{ii}(m+n) > 0$, and so $d(i) \mid (m+n)$. If $d(i) \nmid r$ then $p_{ii}(m+r+n) = 0$, so that $p_{jj}(r) = 0$; therefore $d(i) \mid d(j)$. Similarly $d(j) \mid d(i)$, giving that $d(i) = d(j)$.

4. (a) See the solution to Exercise (6.3.9a).
(b) Let $i, j, r, s \in S$, and choose $N(i, r)$ and $N(j, s)$ according to part (a). Then

$$\mathbb{P}\big(Z_n = (r, s) \,\big|\, Z_0 = (i, j)\big) = p_{ir}(n)p_{js}(n) > 0$$

if $n \geq \max\{N(i, r), N(j, s)\}$, so that the chain is irreducible and aperiodic.
(c) Suppose $S = \{1, 2\}$ and

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In this case $\big\{\{1, 1\}, \{2, 2\}\big\}$ and $\big\{\{1, 2\}, \{2, 1\}\big\}$ are closed sets of states for the bivariate chain.

5. Clearly $\mathbb{P}(N = 0) = 1 - f_{ij}$, while, by conditioning on the time of the nth visit to j, we have that $\mathbb{P}(N \geq n + 1 \mid N \geq n) = f_{jj}$ for $n \geq 1$, whence the answer is immediate. Now $\mathbb{P}(N = \infty) = 1 - \sum_{n=0}^{\infty} \mathbb{P}(N = n)$ which equals 1 if and only if $f_{ij} = f_{jj} = 1$.

6. Fix $i \neq j$ and let $m = \min\{n : p_{ij}(n) > 0\}$. If $X_0 = i$ and $X_m = j$ then there can be no intermediate visit to i (with probability one), since such a visit would contradict the minimality of m.

Suppose $X_0 = i$, and note that $(1 - f_{ji})p_{ij}(m) \le 1 - f_{ii}$, since if the chain visits j at time m and subsequently does not return to i, then no return to i takes place at all. However $f_{ii} = 1$ if i is persistent, so that $f_{ji} = 1$.

7. (a) We may take $S = \{0, 1, 2, \dots\}$. Note that $q_{ij}(n) \ge 0$, and

$$\sum_j q_{ij}(n) = 1, \quad q_{ij}(n+1) = \sum_{l=0}^{\infty} q_{il}(1)q_{lj}(n),$$

whence $\mathbf{Q} = (q_{ij}(1))$ is the transition matrix of a Markov chain, and $\mathbf{Q}^n = (q_{ij}(n))$. This chain is persistent since

$$\sum_n q_{ii}(n) = \sum_n p_{ii}(n) = \infty \qquad \text{for all } i,$$

and irreducible since i communicates with j in the new chain whenever j communicates with i in the original chain.

That

$$(*) \qquad g_{ij}(n) = \frac{x_j}{x_i} l_{ji}(n), \qquad i \ne j,\ n \ge 1,$$

is evident when $n = 1$ since both sides are $q_{ij}(1)$. Suppose it is true for $n = m$ where $m \ge 1$. Now

$$l_{ji}(m+1) = \sum_{k:k \ne j} l_{jk}(m)p_{ki}, \qquad i \ne j,$$

so that

$$\frac{x_j}{x_i} l_{ji}(m+1) = \sum_{k:k \ne j} g_{kj}(m)q_{ik}(1), \qquad i \ne j,$$

which equals $g_{ij}(m+1)$ as required.

(b) Sum $(*)$ over n to obtain that

$$(**) \qquad 1 = \frac{x_j}{x_i}\rho_i(j), \qquad i \ne j,$$

where $\rho_i(j)$ is the mean number of visits to i between two visits to j; we have used the fact that $\sum_n g_{ij}(n) = 1$, since the chain is persistent (see Problem (6.15.6)). It follows that $x_i = x_0\rho_i(0)$ for all i, and therefore $\mathbf{x}$ is unique up to a multiplicative constant.

(c) The claim is trivial when $i = j$, and we assume therefore that $i \ne j$. Let $N_i(j)$ be the number of visits to i before reaching j for the first time, and write $\mathbb{P}_k$ and $\mathbb{E}_k$ for probability and expectation conditional on $X_0 = k$. Clearly, $\mathbb{P}_j(N_i(j) \ge r) = h_{ji}(1 - h_{ij})^{r-1}$ for $r \ge 1$, whence

$$\rho_i(j) = \mathbb{E}_j(N_i(j)) = \sum_{r=1}^{\infty} \mathbb{P}_j(N_i(j) \ge r) = \frac{h_{ji}}{h_{ij}}.$$

The claim follows by $(**)$.

8. (a) If such a Markov chain exists, then

$$u_n = \sum_{i=1}^{n} f_i u_{n-i}, \qquad n \ge 1,$$

where f_i is the probability that the first return of X to its persistent starting point s takes place at time i. Certainly $u_0 = 1$.

Conversely, suppose u is a renewal sequence with respect to the collection $(f_m : m \geq 1)$. Let X be a Markov chain on the state space $S = \{0, 1, 2, \dots\}$ with transition matrix

$$p_{ij} = \begin{cases} \mathbb{P}(T \geq i+2 \mid T \geq i+1) & \text{if } j = i+1, \\ 1 - \mathbb{P}(T \geq i+2 \mid T \geq i+1) & \text{if } j = 0, \end{cases}$$

where T is a random variable having mass function $f_m = \mathbb{P}(T = m)$. With $X_0 = 0$, the chance that the first return to 0 takes place at time n is

$$\begin{aligned} \mathbb{P}\left(X_n = 0, \prod_1^{n-1} X_i \neq 0 \,\middle|\, X_0 = 0 \right) &= p_{01} p_{12} \cdots p_{n-2,n-1} p_{n-1,0} \\ &= \left(1 - \frac{G(n+1)}{G(n)}\right) \prod_{i=1}^{n-1} \frac{G(i+1)}{G(i)} \\ &= G(n) - G(n+1) = f_n \end{aligned}$$

where $G(m) = \mathbb{P}(T \geq m) = \sum_{n=m}^{\infty} f_n$. Now $v_n = \mathbb{P}(X_n = 0 \mid X_0 = 0)$ satisfies

$$v_0 = 1, \qquad v_n = \sum_{i=1}^{n} f_i v_{n-i} \qquad \text{for } n \geq 1,$$

whence $v_n = u_n$ for all n.

(b) Let X and Y be the two Markov chains which are associated (respectively) with u and v in the above sense. We shall assume that X and Y are independent. The product $(u_n v_n : n \geq 1)$ is now the renewal sequence associated with the bivariate Markov chain (X_n, Y_n).

9. Of the first $2n$ steps, let there be i rightwards, j upwards, and k inwards. Now $\mathbf{X}_{2n} = \mathbf{0}$ if and only if there are also i leftwards, j downwards, and k outwards. The number of such possible combinations is $(2n)!/\{(i!\,j!\,k!)^2\}$, and each such combination has probability $(\frac{1}{6})^{2(i+j+k)} = (\frac{1}{6})^{2n}$. The first equality follows, and the second is immediate.

Now

$$(*) \qquad \mathbb{P}(\mathbf{X}_{2n} = \mathbf{0}) \leq \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} M \sum_{i+j+k=n} \frac{n!}{3^n i!\,j!\,k!}$$

where

$$M = \max\left\{ \frac{n!}{3^n i!\,j!\,k!} : i, j, k \geq 0,\ i + j + k = n \right\}.$$

It is not difficult to see that the maximum M is attained when i, j, and k are all closest to $\frac{1}{3}n$, so that

$$M \leq \frac{n!}{3^n (\lfloor \frac{1}{3}n \rfloor !)^3}.$$

Furthermore the summation in $(*)$ equals 1, since the summand is the probability that, in allocating n balls randomly to three urns, the urns contain respectively i, j, and k balls. It follows that

$$\mathbb{P}(\mathbf{X}_{2n} = \mathbf{0}) \leq \frac{(2n)!}{12^n n!\,(\lfloor \frac{1}{3}n \rfloor !)^3}$$

which, by an application of Stirling's formula, is no bigger than $Cn^{-\frac{3}{2}}$ for some constant C. Hence $\sum_n \mathbb{P}(\mathbf{X}_{2n} = \mathbf{0}) < \infty$, so that the origin is transient.

10. No. The line of ancestors of any cell-state is a random walk in three dimensions. The difference between two such lines of ancestors is also a type of random walk, which in three dimensions is transient.

11. There are one or two absorbing states according as whether one or both of α and β equal zero. If $\alpha\beta \neq 0$, the chain is irreducible and persistent. It is periodic if and only if $\alpha = \beta = 1$, in which case it has period 2.

If $0 < \alpha\beta < 1$ then

$$\boldsymbol{\pi} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

is the stationary distribution. There are various ways of calculating $\mathbf{P}^n$; see Exercise (6.3.3) for example. In this case the answer is given by

$$(\alpha+\beta)\mathbf{P}^n = (1-\alpha-\beta)^n \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} + \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix};$$

proof by induction. Hence

$$(\alpha+\beta)\mathbf{P}^n \to \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \qquad \text{as } n \to \infty.$$

The chain is reversible in equilibrium if and only if $\pi_1 p_{12} = \pi_2 p_{21}$, which is to say that $\alpha\beta = \beta\alpha$!

12. The transition matrix is given by

$$p_{ij} = \begin{cases} \left(\dfrac{N-i}{N}\right)^2 & \text{if } j = i+1, \\ 1 - \left(\dfrac{i}{N}\right)^2 - \left(\dfrac{N-i}{N}\right)^2 & \text{if } j = i, \\ \left(\dfrac{i}{N}\right)^2 & \text{if } j = i-1, \end{cases}$$

for $0 \le i \le N$. This process is a birth–death process in discrete time, and by Exercise (6.5.1) is reversible in equilibrium. Its stationary distribution satisfies the detailed balance equation $\pi_i p_{i,i+1} = \pi_{i+1} p_{i+1,i}$ for $0 \le i < N$, whence $\pi_i = \pi_0 \binom{N}{i}^2$ for $0 \le i \le N$, where

$$\frac{1}{\pi_0} = \sum_{i=0}^{N} \binom{N}{i}^2 = \binom{2N}{N}.$$

13. (a) The chain X is irreducible; all states are therefore of the same type. The state 0 is aperiodic, and so therefore is every other state. Suppose that $X_0 = 0$, and let T be the time of the first return to 0. Then $\mathbb{P}(T > n) = a_0 a_1 \cdots a_{n-1} = b_n$ for $n \ge 1$, so that 0 is persistent if and only if $b_n \to 0$ as $n \to \infty$.

(b) The mean of T is

$$\mathbb{E}(T) = \sum_{n=0}^{\infty} \mathbb{P}(T > n) = \sum_{n=0}^{\infty} b_n.$$

The stationary distribution $\boldsymbol{\pi}$ satisfies

$$\pi_0 = \sum_{k=0}^{\infty} \pi_k(1 - a_k), \qquad \pi_n = \pi_{n-1}a_{n-1} \text{ for } n \geq 1.$$

Hence $\pi_n = \pi_0 b_n$ and $\pi_0^{-1} = \sum_{n=0}^{\infty} b_n$ if this sum is finite.

(c) Suppose a_i has the stated form for $i \geq I$. Then

$$b_n = b_I \prod_{i=I}^{n-1}(1 - Ai^{-\beta}), \qquad n \geq I.$$

Hence $b_n \to 0$ if and only if $\sum_i Ai^{-\beta} = \infty$, which is to say that $\beta \leq 1$. The chain is therefore persistent if and only if $\beta \leq 1$.

(d) We have that $1 - x \leq e^{-x}$ for $x \geq 0$, and therefore

$$\sum_{n=I}^{\infty} b_n \leq b_I \sum_{n=I}^{\infty} \exp\left\{-A \sum_{i=I}^{n-1} i^{-\beta}\right\} \leq b_I \sum_{n=I}^{\infty} \exp\left\{-An \cdot n^{-\beta}\right\} < \infty \quad \text{if } \beta < 1.$$

(e) If $\beta = 1$ and $A > 1$, there is a constant c_I such that

$$\sum_{n=I}^{\infty} b_n \leq b_I \sum_{n=I}^{\infty} \exp\left\{-A \sum_{i=I}^{n-1} \frac{1}{i}\right\} \leq c_I \sum_{n=I}^{\infty} \exp\{-A \log n\} = c_I \sum_{n=I}^{\infty} n^{-A} < \infty,$$

giving that the chain is non-null.

(f) If $\beta = 1$ and $A \leq 1$,

$$b_n = b_I \prod_{i=I}^{n-1}\left(1 - \frac{A}{i}\right) \geq b_I \prod_{i=I}^{n-1}\left(\frac{i-1}{i}\right) = b_I \left(\frac{I-1}{n-1}\right).$$

Therefore $\sum_n b_n = \infty$, and the chain is null.

14. Using the Chapman–Kolmogorov equations,

$$\begin{aligned}|p_{ij}(t+h) - p_{ij}(t)| &= \left|\sum_k \big(p_{ik}(h) - \delta_{ik}\big)p_{kj}(t)\right| \leq \big(1 - p_{ii}(h)\big)p_{ij}(t) + \sum_{k \neq i} p_{ik}(h) \\ &\leq \big(1 - p_{ii}(h)\big) + \big(1 - p_{ii}(h)\big) \to 0\end{aligned}$$

as $h \downarrow 0$, if the semigroup is standard.

Now $\log x$ is continuous for $0 < x \leq 1$, and therefore g is continuous. Certainly $g(0) = 0$. In addition $p_{ii}(s+t) \geq p_{ii}(s)p_{ii}(t)$ for $s, t \geq 0$, whence $g(s+t) \leq g(s) + g(t)$, $s, t \geq 0$.

For the last part

$$\frac{1}{t}\big(p_{ii}(t) - 1\big) = \frac{g(t)}{t} \cdot \frac{p_{ii}(t) - 1}{-\log\{1 - (1 - p_{ii}(t))\}} \to -\lambda$$

as $t \downarrow 0$, since $x/\log(1-x) \to -1$ as $x \downarrow 0$.

15. Let i and j be distinct states, and suppose that $p_{ij}(t) > 0$ for some t. Now

$$p_{ij}(t) = \sum_{n=0}^{\infty} \frac{1}{n!} t^n (G^n)_{ij},$$

implying that $(G^n)_{ij} > 0$ for some n, which is to say that

$$(*) \qquad g_{i,k_1} g_{k_1,k_2} \cdots g_{k_n,j} \neq 0$$

for some sequence $k_1, k_2, \dots, k_n$ of states.

Suppose conversely that $(*)$ holds for some sequence of states, and choose the minimal such value of n. Then $i, k_1, k_2, \dots, k_n, j$ are distinct states, since otherwise n is not minimal. It follows that $(G^n)_{ij} > 0$, while $(G^m)_{ij} = 0$ for $0 \le m < n$. Therefore

$$p_{ij}(t) = t^n \sum_{k=n}^{\infty} \frac{1}{k!} t^k (G^k)_{ij}$$

is strictly positive for all sufficiently small positive values of t. Therefore i communicates with j.

16. (a) Suppose X is reversible, and let i and j be distinct states. Now

$$\mathbb{P}\big(X(0) = i,\ X(t) = j\big) = \mathbb{P}\big(X(t) = i,\ X(0) = j\big),$$

which is to say that $\pi_i p_{ij}(t) = \pi_j p_{ji}(t)$. Divide by t and take the limit as $t \downarrow 0$ to obtain that $\pi_i g_{ij} = \pi_j g_{ji}$.

Suppose now that the chain is uniform, and $X(0)$ has distribution $\boldsymbol{\pi}$. If $t > 0$, then

$$\mathbb{P}(X(t) = j) = \sum_i \pi_i p_{ij}(t) = \pi_j,$$

so that $X(t)$ has distribution $\boldsymbol{\pi}$ also. Now let $t < 0$, and suppose that $X(t)$ has distribution $\boldsymbol{\mu}$. The distribution of $X(s)$ for $s \ge 0$ is $\boldsymbol{\mu}\mathbf{P}_{s-t} = \boldsymbol{\pi}$, a polynomial identity in the variable $s - t$, valid for all $s \ge 0$. Such an identity must be valid for all s, and particularly for $s = t$, implying that $\boldsymbol{\mu} = \boldsymbol{\pi}$.

Suppose in addition that $\pi_i g_{ij} = \pi_j g_{ji}$ for all i, j. For any sequence $k_1, k_2, \dots, k_n$ of states,

$$\pi_i g_{i,k_1} g_{k_1,k_2} \cdots g_{k_n,j} = g_{k_1,i} \pi_{k_1} g_{k_1,k_2} \cdots g_{k_n,j} = \cdots = g_{k_1,i} g_{k_2,k_1} \cdots g_{j,k_n} \pi_j.$$

Sum this expression over all sequences $k_1, k_2, \dots, k_n$ of length n, to obtain

$$\pi_i (G^{n+1})_{ij} = \pi_j (G^{n+1})_{ji}, \qquad n \ge 0.$$

It follows, by the fact that $\mathbf{P}_t = e^{t\mathbf{G}}$, that

$$\pi_i p_{ij}(t) = \pi_i \sum_{m=0}^{\infty} \frac{1}{m!} t^m (G^m)_{ij} = \pi_j \sum_{m=0}^{\infty} \frac{1}{m!} t^m (G^m)_{ji} = \pi_j p_{ji}(t)$$

for all i, j, t. For $t_1 < t_2 < \cdots < t_n$,

$$\begin{aligned}
\mathbb{P}\big(&X(t_1) = i_1, X(t_2) = i_2, \dots, X(t_n) = i_n\big) \\
&= \pi_{i_1} p_{i_1,i_2}(t_2 - t_1) \cdots p_{i_{n-1},i_n}(t_n - t_{n-1}) \\
&= p_{i_2,i_1}(t_2 - t_1) \pi_{i_2} p_{i_2,i_3}(t_3 - t_2) \cdots p_{i_{n-1},i_n}(t_n - t_{n-1}) = \cdots \\
&= p_{i_2,i_1}(t_2 - t_1) \cdots p_{i_n,i_{n-1}}(t_n - t_{n-1}) \pi_{i_n} \\
&= \mathbb{P}\big(Y(t_1) = i_1, Y(t_2) = i_2, \dots, Y(t_n) = i_n\big),
\end{aligned}$$

giving that the chain is reversible.
(b) Let $S = \{1, 2\}$ and

$$\mathbf{G} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

where $\alpha\beta > 0$. The chain is uniform with stationary distribution

$$\boldsymbol{\pi} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right),$$

and therefore $\pi_1 g_{12} = \pi_2 g_{21}$.
(c) Let X be a birth–death process with birth rates λ_i and death rates μ_i. The stationary distribution $\boldsymbol{\pi}$ satisfies

$$\pi_1\mu_1 - \pi_0\mu_0 = 0, \quad \pi_{k+1}\mu_{k+1} - \pi_k\lambda_k = \pi_k\lambda_k - \pi_{k-1}\lambda_{k-1} \text{ for } k \geq 1.$$

Therefore $\pi_{k+1}\mu_{k+1} = \pi_k\lambda_k$ for $k \geq 0$, the conditions for reversibility.

17. Consider the continuous-time chain with generator

$$\mathbf{G} = \begin{pmatrix} -\beta & \beta \\ \gamma & -\gamma \end{pmatrix}.$$

It is a standard calculation (Exercise (6.9.1)) that the associated semigroup satisfies

$$(\beta+\gamma)\mathbf{P}_t = \begin{pmatrix} \gamma + \beta h(t) & \beta(1-h(t)) \\ \gamma(1-h(t)) & \beta + \gamma h(t) \end{pmatrix}$$

where $h(t) = e^{-t(\beta+\gamma)}$. Now $\mathbf{P}_1 = \mathbf{P}$ if and only if $\gamma + \beta h(1) = \beta + \gamma h(1) = \alpha(\beta+\gamma)$, which is to say that $\beta = \gamma = -\frac{1}{2}\log(2\alpha - 1)$, a solution which requires that $\alpha > \frac{1}{2}$.

18. The forward equations for $p_n(t) = \mathbb{P}(X(t) = n)$ are

$$p_0'(t) = \mu p_1 - \lambda p_0,$$
$$p_n'(t) = \lambda p_{n-1} - (\lambda + n\mu)p_n + \mu(n+1)p_{n+1}, \qquad n \geq 1.$$

In the usual way,

$$\frac{\partial G}{\partial t} = (s-1)\left(\lambda G - \mu \frac{\partial G}{\partial s}\right)$$

with boundary condition $G(s, 0) = s^I$. The characteristics are given by

$$dt = \frac{ds}{\mu(s-1)} = \frac{dG}{\lambda(s-1)G},$$

and therefore $G = e^{\rho(s-1)} f\big((s-1)e^{-\mu t}\big)$, for some function f, determined by the boundary condition to satisfy $e^{\rho(s-1)} f(s-1) = s^I$. The claim follows.

As $t \to \infty$, $G(s, t) \to e^{\rho(s-1)}$, the generating function of the Poisson distribution, parameter ρ.

19. (a) The forward equations are

$$\frac{\partial}{\partial t} p_{ii}(s, t) = -\lambda(t) p_{ii}(s, t),$$
$$\frac{\partial}{\partial t} p_{ij}(s, t) = -\lambda(t) p_{ij}(s, t) + \lambda(t) p_{i,j-1}(t), \qquad i < j.$$

Assume $N(s) = i$ and $s < t$. In the usual way,

$$G(s,t;x) = \sum_{j=i}^{\infty} x^j \mathbb{P}\big(N(t) = j \mid N(s) = i\big)$$

satisfies

$$\frac{\partial G}{\partial t} = \lambda(t)(x-1)G.$$

We integrate subject to the boundary condition to obtain

$$G(s,t;x) = x^i \exp\left\{(x-1)\int_s^t \lambda(u)\,du\right\},$$

whence $p_{ij}(t)$ is found to be the probability that $A = j - i$ where A has the Poisson distribution with parameter $\int_s^t \lambda(u)\,du$.

The backward equations are

$$\frac{\partial}{\partial s} p_{ij}(s,t) = \lambda(s)p_{i+1,j}(s,t) - \lambda(s)p_{ij}(s,t);$$

using the fact that $p_{i+1,j}(t) = p_{i,j-1}(t)$, we are led to

$$-\frac{\partial G}{\partial s} = \lambda(s)(x-1)G.$$

The solution is the same as above.

(b) We have that

$$\mathbb{P}(T > t) = p_{00}(t) = \exp\left\{-\int_0^t \lambda(u)\,du\right\},$$

so that

$$f_T(t) = \lambda(t)\exp\left\{-\int_0^t \lambda(u)\,du\right\}, \qquad t \ge 0.$$

In the case $\lambda(t) = c/(1+t)$,

$$\mathbb{E}(T) = \int_0^\infty \mathbb{P}(T > t)\,dt = \int_0^\infty \frac{du}{(1+u)^c}$$

which is finite if and only if $c > 1$.

20. Let $s > 0$. Each offer has probability $1 - F(s)$ of exceeding s, and therefore the first offer exceeding s is the Mth offer overall, where $\mathbb{P}(M = m) = F(s)^{m-1}[1 - F(s)]$, $m \ge 1$. Conditional on $\{M = m\}$, the value of X_M is independent of the values of $X_1, X_2, \dots, X_{M-1}$, with

$$\mathbb{P}(X_M > u \mid M = m) = \frac{1 - F(u)}{1 - F(s)}, \qquad 0 < s \le u,$$

and $X_1, X_2, \dots, X_{M-1}$ have shared (conditional) distribution function

$$G(u \mid s) = \frac{F(u)}{F(s)}, \qquad 0 \le u \le s.$$

For any event B defined in terms of $X_1, X_2, \dots, X_{M-1}$, we have that

$$\begin{aligned}\mathbb{P}(X_M > u, B) &= \sum_{m=1}^{\infty} \mathbb{P}(X_M > u, B \mid M = m)\mathbb{P}(M = m)\\ &= \sum_{m=1}^{\infty} \mathbb{P}(X_M > u \mid M = m)\mathbb{P}(B \mid M = m)\mathbb{P}(M = m)\\ &= \mathbb{P}(X_M > u) \sum_{m=1}^{\infty} \mathbb{P}(B \mid M = m)\mathbb{P}(M = m)\\ &= \mathbb{P}(X_M > u)\mathbb{P}(B), \qquad 0 < s \le u,\end{aligned}$$

where we have used the fact that $\mathbb{P}(X_M > u \mid M = m)$ is independent of m. It follows that the first record value exceeding s is independent of all record values not exceeding s. By a similar argument (or an iteration of the above) all record values exceeding s are independent of all record values not exceeding s.

The chance of a record value in $(s, s+h]$ is

$$\mathbb{P}(s < X_M \le s+h) = \frac{F(s+h) - F(s)}{1 - F(s)} = \frac{f(s)h}{1 - F(s)} + o(h).$$

A very similar argument works for the runners-up. Let $X_{M_1}, X_{M_2}, \dots$ be the values, in order, of offers exceeding s. It may be seen that this sequence is independent of the sequence of offers not exceeding s, whence it follows that the sequence of runners-up is a non-homogeneous Poisson process. There is a runner-up in $(s, s+h]$ if (neglecting terms of order $o(h)$) the first offer exceeding s is larger than $s+h$, and the second is in $(s, s+h]$. The probability of this is

$$\left(\frac{1 - F(s+h)}{1 - F(s)}\right)\left(\frac{F(s+h) - F(s)}{1 - F(s)}\right) + o(h) = \frac{f(s)h}{1 - F(s)} + o(h).$$

21. Let $F_t(x) = \mathbb{P}(N^*(t) \le x)$, and let A be the event that N has a arrival during $(t, t+h)$. Then

$$F_{t+h}(x) = \lambda h \mathbb{P}\big(N^*(t+h) \le x \bigm| A\big) + (1 - \lambda h)F_t(x) + o(h)$$

where

$$\mathbb{P}\big(N^*(t+h) \le x \bigm| A\big) = \int_{-\infty}^{\infty} F_t(x-y) f(y)\, dy.$$

Hence

$$\frac{\partial}{\partial t} F_t(x) = -\lambda F_t(x) + \lambda \int_{-\infty}^{\infty} F_t(x-y) f(y)\, dy.$$

Take Fourier transforms to find that $\phi_t(\theta) = \mathbb{E}(e^{i\theta N^*(t)})$ satisfies

$$\frac{\partial \phi_t}{\partial t} = -\lambda \phi_t + \lambda \phi_t \phi,$$

an equation which may be solved subject to $\phi_0(\theta) = 1$ to obtain $\phi_t(\theta) = e^{\lambda t(\phi(\theta) - 1)}$.

Alternatively, using conditional expectation,

$$\phi_t(\theta) = \mathbb{E}\big\{\mathbb{E}\big(e^{i\theta N^*(t)} \bigm| N(t)\big)\big\} = \mathbb{E}\big\{\phi(\theta)^{N(t)}\big\}$$

where $N(t)$ is Poisson with parameter λt.

22. We have that

$$\mathbb{E}(s^{N(t)}) = \mathbb{E}\{\mathbb{E}(s^{N(t)} \mid \Lambda)\} = \tfrac{1}{2}\{e^{\lambda_1 t(s-1)} + e^{\lambda_2 t(s-1)}\},$$

whence $\mathbb{E}(N(t)) = \frac{1}{2}(\lambda_1 + \lambda_2)t$ and $\mathrm{var}(N(t)) = \frac{1}{2}(\lambda_1 + \lambda_2)t + \frac{1}{4}(\lambda_1 - \lambda_2)^2 t^2$.

23. Conditional on $\{X(t) = i\}$, the next arrival in the birth process takes place at rate λi.

24. The forward equations for $p_n(t) = \mathbb{P}(X(t) = n)$ are

$$p_n'(t) = \frac{1 + \mu(n-1)}{1 + \mu t} p_{n-1}(t) - \frac{1 + \mu n}{1 + \mu t} p_n(t), \qquad n \geq 0,$$

with the convention that $p_{-1}(t) = 0$. Multiply by s^n and sum to deduce that

$$(1 + \mu t)\frac{\partial G}{\partial t} = sG + \mu s^2 \frac{\partial G}{\partial s} - G - \mu s \frac{\partial G}{\partial s}$$

as required.

Differentiate with respect to s and take the limit as $s \uparrow 1$. If $\mathbb{E}(X(t)^2) < \infty$, then

$$m(t) = \mathbb{E}(X(t)) = \left.\frac{\partial G}{\partial s}\right|_{s=1}$$

satisfies $(1 + \mu t)m'(t) = 1 + \mu m(t)$ subject to $m(0) = I$. Solving this in the usual way, we obtain $m(t) = I + (1 + \mu I)t$.

Differentiate again to find that

$$n(t) = \mathbb{E}\big(X(t)(X(t) - 1)\big) = \left.\frac{\partial^2 G}{\partial s^2}\right|_{s=1}$$

satisfies $(1 + \mu t)n'(t) = 2\big(m(t) + \mu m(t) + \mu n(t)\big)$ subject to $n(0) = I(I - 1)$. The solution is

$$n(t) = I(I - 1) + 2I(1 + \mu I)t + (1 + \mu I)(1 + \mu + \mu I)t^2.$$

The variance of $X(t)$ is $n(t) + m(t) - m(t)^2$.

25. (a) Condition on the value of the first step:

$$\eta_j = \frac{\lambda_j}{\lambda_j + \mu_j} \cdot \eta_{j+1} + \frac{\mu_j}{\lambda_j + \mu_j} \cdot \eta_{j-1}, \qquad j \geq 1,$$

as required. Set $x_i = \eta_{i+1} - \eta_i$ to obtain $\lambda_j x_j = \mu_j x_{j-1}$ for $j \geq 1$, so that

$$x_j = x_0 \prod_{i=1}^{j} \frac{\mu_i}{\lambda_i}, \qquad j \geq 1.$$

It follows that

$$\eta_{j+1} = \eta_0 + \sum_{k=0}^{j} x_k = 1 + (\eta_1 - 1)\sum_{k=0}^{j} e_k.$$

The η_j are probabilities, and lie in $[0, 1]$. If $\sum_1^\infty e_k = \infty$ then we must have $\eta_1 = 1$, which implies that $\eta_j = 1$ for all j.

(b) By conditioning on the first step, the probability η_j, of visiting 0 having started from j, satisfies

$$\eta_j = \frac{(j+1)^2\eta_{j+1} + j^2\eta_{j-1}}{j^2 + (j+1)^2}.$$

Hence, $(j+1)^2(\eta_{j+1} - \eta_j) = j^2(\eta_j - \eta_{j-1})$, giving $(j+1)^2(\eta_{j+1} - \eta_j) = \eta_1 - \eta_0$. Therefore,

$$1 - \eta_{j+1} = (1-\eta_1)\sum_{k=0}^{j}\frac{1}{(k+1)^2} \to (1-\eta_1)\tfrac{1}{6}\pi^2 \qquad \text{as } j \to \infty.$$

By Exercise (6.3.6), we seek the minimal non-negative solution, which is achieved when $\eta_1 = 1 - (6/\pi^2)$.

26. We may suppose that $X(0) = 0$. Let $T_n = \inf\{t : X(t) = n\}$. Suppose $T_n = T$, and let $Y = T_{n+1} - T$. Condition on all possible occurrences during the interval $(T, T+h)$ to find that

$$\mathbb{E}(Y) = (\lambda_n h)h + \mu_n h(h + \mathbb{E}(Y')) + (1 - \lambda_n h - \mu_n h)(h + \mathbb{E}(Y)) + o(h),$$

where Y' is the mean time which elapses before reaching $n+1$ from $n-1$. Set $m_n = \mathbb{E}(T_{n+1} - T_n)$ to obtain that

$$m_n = \mu_n h(m_{n-1} + m_n) + m_n + h\{1 - (\lambda_n + \mu_n)m_n\} + o(h).$$

Divide by h and take the limit as $h \downarrow 0$ to find that $\lambda_n m_n = 1 + \mu_n m_{n-1}$, $n \geq 1$. Therefore

$$m_n = \frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_n}m_{n-1} = \cdots = \frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_n\lambda_{n-1}} + \cdots + \frac{\mu_n\mu_{n-1}\cdots\mu_1}{\lambda_n\lambda_{n-1}\cdots\lambda_0},$$

since $m_0 = \lambda_0^{-1}$. The process is dishonest if $\sum_{n=0}^{\infty} m_n < \infty$, since in this case $T_\infty = \lim T_n$ has finite mean, so that $\mathbb{P}(T_\infty < \infty) = 1$.

On the other hand, the process grows no faster than a birth process with birth rates λ_i, which is honest if $\sum_{n=0}^{\infty} 1/\lambda_n = \infty$. Can you find a better condition?

27. We know that, conditional on $X(0) = I$, $X(t)$ has generating function

$$G(s,t) = \left(\frac{\lambda t(1-s) + s}{\lambda t(1-s) + 1}\right)^I,$$

so that

$$\mathbb{P}(T \leq x \mid X(0) = I) = \mathbb{P}(X(x) = 0 \mid X(0) = I) = G(0,x) = \left(\frac{\lambda x}{\lambda x + 1}\right)^I.$$

It follows that, in the limit as $x \to \infty$,

$$\mathbb{P}(T \leq x) = \sum_{I=0}^{\infty}\left(\frac{\lambda x}{\lambda x + 1}\right)^I \mathbb{P}(X(0) = I) = G_{X(0)}\left(\frac{\lambda x}{\lambda x + 1}\right) \to 1.$$

For the final part, the required probability is $\{xI/(xI+1)\}^I = \{1 + (xI)^{-1}\}^{-I}$, which tends to $e^{-1/x}$ as $I \to \infty$.

28. Let Y be an immigration–death process *without* disasters, with $Y(0) = 0$. We have from Problem (6.15.18) that $Y(t)$ has generating function $G(s,t) = \exp\{\rho(s-1)(1-e^{-\mu t})\}$ where $\rho = \lambda/\mu$. As seen earlier, and as easily verified by taking the limit as $t \to \infty$, Y has a stationary distribution.

From the process Y we may generate the process X in the following way. At the epoch of each disaster, we paint every member of the population grey. At any given time, the unpainted individuals constitute X, and the aggregate population constitutes Y. When constructed in this way, it is the case that $Y(t) \le X(t)$, so that Y is a Markov chain which is dominated by a chain having a stationary distribution. It follows that X has a stationary distribution $\boldsymbol{\pi}$ (the state 0 is persistent for X, and therefore persistent for Y also).

Suppose X is in equilibrium. The times of disasters form a Poisson process with intensity δ. At any given time t, the elapsed time T since the last disaster is exponentially distributed with parameter δ. At the time of this disaster, the value of $X(t)$ is reduced to 0 whatever its previous value.

It follows by averaging over the value of T that the generating function $H(s) = \sum_{n=0}^{\infty} s^n \pi_n$ of $X(t)$ is given by

$$H(s) = \int_0^\infty \delta e^{-\delta u} G(s,u)\,du = \frac{\delta}{\mu} e^{\rho(s-1)} \int_0^1 x^{(\delta/\mu)-1} e^{-\rho(s-1)x}\,dx$$

by the substitution $x = e^{-\mu u}$. The mean of $X(t)$ is

$$H'(1) = \int_0^\infty \delta e^{-\delta u} \mathbb{E}(Y(u))\,du = \int_0^\infty \delta e^{-\delta u} \rho(1 - e^{-\mu u})\,du = \frac{\rho\mu}{\delta+\mu} = \frac{\lambda}{\delta+\mu}.$$

29. Let $G(|B|, s)$ be the generating function of $X(B)$. If $B \cap C = \varnothing$, then $X(B \cup C) = X(B) + X(C)$, so that $G(\alpha+\beta, s) = G(\alpha, s)G(\beta, s)$ for $|s| \le 1, \alpha, \beta \ge 0$. The only solutions to this equation which are monotone in α are of the form $G(\alpha, s) = e^{\alpha\lambda(s)}$ for $|s| \le 1$, and for some function $\lambda(s)$. Now any interval may be divided into n equal sub-intervals, and therefore $G(\alpha, s)$ is the generating function of an infinitely divisible distribution. Using the result of Problem (5.12.13b), $\lambda(s)$ may be written in the form $\lambda(s) = (A(s) - 1)\lambda$ for some λ and some probability generating function $A(s) = \sum_0^\infty a_i s^i$. We now use (iii): if $|B| = \alpha$,

$$\frac{\mathbb{P}(X(B) \ge 1)}{\mathbb{P}(X(B) = 1)} = \frac{1 - e^{\alpha\lambda(a_0 - 1)}}{\alpha\lambda a_1 e^{\alpha\lambda(a_0-1)}} \to 1$$

as $\alpha \downarrow 0$. Therefore $a_0 + a_1 = 1$, and hence $A(s) = a_0 + (1 - a_0)s$, and $X(B)$ has a Poisson distribution with parameter proportional to $|B|$.

30. (a) Let $M(r, s)$ be the number of points of the resulting process on $\mathbb{R}_+$ lying in the interval $(r, s]$. Since disjoint intervals correspond to disjoint annuli of the plane, the process M has independent increments in the sense that $M(r_1, s_1), M(r_2, s_2), \dots, M(r_n, s_n)$ are independent whenever $r_1 < s_1 < r_2 < \cdots < r_n < s_n$. Furthermore, for $r < s$ and $k \ge 0$,

$$\mathbb{P}\big(M(r,s) = k\big) = \mathbb{P}\big(N \text{ has } k \text{ points in the corresponding annulus}\big) = \frac{\{\lambda\pi(s-r)\}^k e^{-\lambda\pi(s-r)}}{k!}.$$

(b) We have similarly that

$$\mathbb{P}(R_{(k)} \le x) = \mathbb{P}(N \text{ has least } k \text{ points in circle of radius } x) = \sum_{r=k}^{\infty} \frac{(\lambda\pi x^2)^r e^{-\lambda\pi x^2}}{r!},$$

and the claim follows by differentiating, and utilizing the successive cancellation.

31. The number $X(S)$ of points within the sphere with volume S and centre at the origin has the Poisson distribution with parameter λS. Hence $\mathbb{P}(X(S) = 0) = e^{-\lambda S}$, implying that the volume V of the largest such empty ball has the exponential distribution with parameter λ.

It follows that $\mathbb{P}(R > r) = \mathbb{P}(V > cr^n) = e^{-\lambda c r^n}$ for $r \ge 0$, where c is the volume of the unit ball in n dimensions. Therefore

$$f_R(r) = \lambda n c r^{n-1} e^{-\lambda c r^n}, \qquad r \ge 0.$$

Finally, $\mathbb{E}(R) = \int_0^\infty e^{-\lambda c r^n}\,dr$, and we set $v = \lambda c r^n$.

32. The time between the kth and $(k+1)$th infection has mean λ_k^{-1}, whence

$$\mathbb{E}(T) = \sum_{k=1}^{N} \frac{1}{\lambda_k}.$$

Now

$$\begin{aligned}\sum_{k=1}^{N} \frac{1}{k(N+1-k)} &= \frac{1}{N+1}\left\{\sum_{k=1}^{N}\frac{1}{k} + \sum_{k=1}^{N}\frac{1}{N+1-k}\right\}\\ &= \frac{2}{N+1}\sum_{k=1}^{N}\frac{1}{k} = \frac{2}{N+1}\{\log N + \gamma + \mathrm{O}(N^{-1})\}.\end{aligned}$$

It may be shown with more work (as in the solution to Problem (5.12.34)) that the moment generating function of $\lambda(N+1)T - 2\log N$ converges as $N \to \infty$, the limit being $\{\Gamma(1-\theta)\}^2$.

33. (a) The forward equations for $p_n(t) = \mathbb{P}(V(t) = n + \frac{1}{2})$ are

$$p_n'(t) = (n+1)p_{n+1}(t) - (2n+1)p_n(t) + np_{n-1}(t), \qquad n \ge 0,$$

with the convention that $p_{-1}(t) = 0$. It follows as usual that

$$\frac{\partial G}{\partial t} = \frac{\partial G}{\partial s} - \left(2s\frac{\partial G}{\partial s} + G\right) + \left(s^2\frac{\partial G}{\partial s} + sG\right)$$

as required. The general solution is

$$G(s,t) = \frac{1}{1-s} f\left(t + \frac{1}{1-s}\right)$$

for some function f. The boundary condition is $G(s,0) = 1$, and the solution is as given.

(b) Clearly

$$m_n(T) = \mathbb{E}\left(\int_0^T I_{nt}\,dt\right) = \int_0^T \mathbb{E}(I_{nt})\,dt$$

by Fubini's theorem, where I_{nt} is the indicator function of the event that $V(t) = n + \frac{1}{2}$.

As for the second part,

$$\sum_{n=0}^{\infty} s^n m_n(T) = \int_0^T G(s,t)\,dt = \frac{\log[1 + (1-s)T]}{1-s},$$

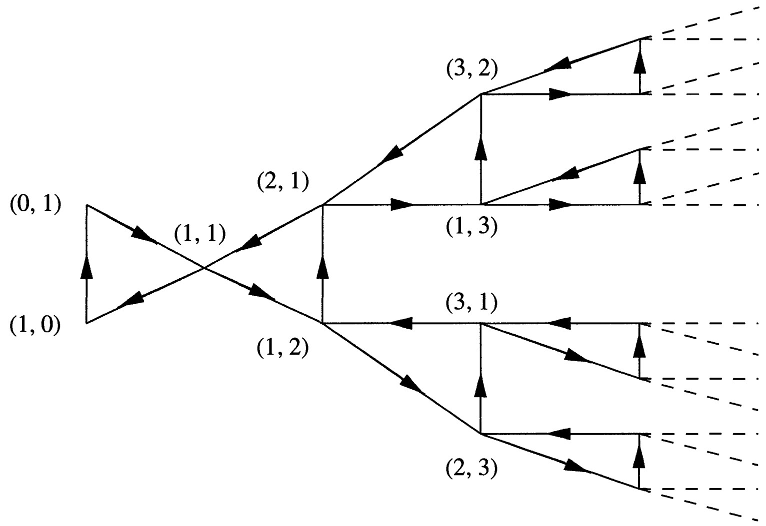
so that, in the limit as $T \to \infty$,

$$\sum_{n=0}^{\infty} s^n\big(m_n(T) - \log T\big) = \frac{1}{1-s}\log\left(\frac{1+(1-s)T}{T}\right) \to \frac{\log(1-s)}{1-s} = -\sum_{n=1}^{\infty} s^n a_n$$

if $|s| < 1$, where $a_n = \sum_{i=1}^{n} i^{-1}$, as required.

(c) The mean velocity at time t is

$$\frac{1}{2} + \left.\frac{\partial G}{\partial s}\right|_{s=1} = t + \frac{1}{2}.$$



34. It is clear that Y is a Markov chain, and its possible transitions are illustrated in the above diagram. Let x and y be the probabilities of ever reaching $(1,1)$ from $(1,2)$ and $(2,1)$, respectively. By conditioning on the first step and using the translational symmetry, we see that $x = \frac{1}{2}y + \frac{1}{2}x^2$ and $y = \frac{1}{2} + \frac{1}{2}xy$. Hence $x^3 - 4x^2 + 4x - 1 = 0$, an equation with roots $x = 1, \frac{1}{2}(3 \pm \sqrt{5})$. Since x is a probability, it must be that either $x = 1$ or $x = \frac{1}{2}(3 - \sqrt{5})$, with the corresponding values of $y = 1$ and $y = \frac{1}{2}(\sqrt{5} - 1)$. Starting from any state to the right of $(1,1)$ in the above diagram, we see by recursion that the chance of ever visiting $(1,1)$ is of the form $x^\alpha y^\beta$ for some non-negative integers α, β. The minimal non-negative solution is therefore achieved when $x = \frac{1}{2}(3 - \sqrt{5})$ and $y = \frac{1}{2}(\sqrt{5} - 1)$. Since $x < 1$, the chain is transient.

35. We write A, 1, 2, 3, 4, 5 for the vertices of the hexagon in clockwise order. Let $T_i = \min\{n \geq 1 : X_n = i\}$ and $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot \mid X_0 = i)$.

(a) By symmetry, the probabilities $p_i = \mathbb{P}_i(T_A < T_C)$ satisfy

$$p_A = \tfrac{2}{3}p_1, \quad p_1 = \tfrac{1}{3} + \tfrac{1}{3}p_2, \quad p_2 = \tfrac{1}{3}p_1 + \tfrac{1}{3}p_3, \quad p_3 = \tfrac{2}{3}p_2,$$

whence $p_A = \frac{7}{27}$.

(b) By Exercise (6.4.6), the stationary distribution is $\pi_C = \frac{1}{4}$, $\pi_i = \frac{1}{8}$ for $i \neq C$, whence $\mu_A = \pi_A^{-1} = 8$.

(c) By the argument leading to Lemma (6.4.5), this equals $\mu_A \pi_C = 2$.

(d) We condition on the event $E = \{T_A < T_C\}$ as in the solution to Exercise (6.2.7). The probabilities $b_i = \mathbb{P}_i(E)$ satisfy

$$b_1 = \tfrac{1}{3} + \tfrac{1}{3}b_2, \quad b_2 = \tfrac{1}{3}b_1 + \tfrac{1}{3}b_3, \quad b_3 = \tfrac{2}{3}b_2,$$

yielding $b_1 = \frac{7}{18}$, $b_2 = \frac{1}{6}$, $b_3 = \frac{1}{9}$. The transition probabilities conditional on E are now found by equations of the form

$$\tau_{12} = \frac{\mathbb{P}_2(E)p_{12}}{\mathbb{P}_1(E)} = \frac{b_2}{3b_1} = \frac{1}{7},$$

and similarly $\tau_{21} = \frac{7}{9}$, $\tau_{23} = \frac{2}{9}$, $\tau_{32} = \frac{1}{2}$, $\tau_{1\mathrm{A}} = \frac{6}{7}$. Hence, with the obvious notation,

$$\mu_{2\mathrm{A}} = 1 + \tfrac{7}{9}\mu_{1\mathrm{A}} + \tfrac{2}{9}\mu_{3\mathrm{A}}, \quad \mu_{3\mathrm{A}} = 1 + \mu_{2\mathrm{A}}, \quad \mu_{1\mathrm{A}} = 1 + \tfrac{1}{7}\mu_{2\mathrm{A}},$$

giving $\mu_{1\mathrm{A}} = \frac{10}{7}$, and the required answer is $1 + \mu_{1\mathrm{A}} = 1 + \frac{10}{7} = \frac{17}{7}$.

36. (a) We have that

$$p_{i,i+1} = \frac{\beta(m-i)^2}{m^2}, \quad p_{i+1,i} = \frac{\alpha(i+1)^2}{m^2}.$$

Look for a solution to the detailed balance equations

$$\pi_i \frac{\beta(m-i)^2}{m^2} = \pi_{i+1}\frac{\alpha(i+1)^2}{m^2}$$

to find the stationary distribution

$$\pi_i = \binom{m}{i}^2 (\beta/\alpha)^i \pi_0, \quad \text{where} \quad \pi_0 = \left\{ \sum_{i=0}^{m} \binom{m}{i}^2 (\beta/\alpha)^i \right\}^{-1}.$$

(b) In this case,

$$p_{i,i+1} = \frac{\beta(m-i)}{m}, \quad p_{i+1,i} = \frac{\alpha(i+1)}{m}.$$

Look for a solution to the detailed balance equations

$$\pi_i \frac{\beta(m-i)}{m} = \pi_{i+1}\frac{\alpha(i+1)}{m},$$

yielding the stationary distribution

$$\pi_i = (\beta/\alpha)^i \binom{m}{i} \pi_0, \quad \text{where} \quad \pi_0 = \left\{ \sum_{i=0}^{m} \binom{m}{i} (\beta/\alpha)^i \right\}^{-1} = \left(\frac{\alpha}{\alpha+\beta}\right)^m.$$

37. We have that

$$\begin{aligned}
d(s+t) &= \sum_k \pi_k c\left(\frac{a_k(s+t)}{\pi_k}\right) \\
&= \sum_k \pi_k c\left(\sum_j \frac{a_j(s)p_{jk}(t)}{\pi_j}\frac{\pi_j}{\pi_k}\right) && \text{by the Chapman–Kolmogorov equations} \\
&\geq \sum_k \pi_k \sum_j \frac{\pi_j p_{jk}(t)}{\pi_k} c\left(\frac{a_j(s)}{\pi_j}\right) && \text{by the concavity of } c \\
&= \sum_j \pi_j c\left(\frac{a_j(s)}{\pi_j}\right) = d(s),
\end{aligned}$$

where we have used the fact that $\sum_j \pi_j p_{jk}(t) = \pi_k$. Now $a_j(s) \to \pi_j$ as $s \to \infty$, and therefore $d(t) \to c(1)$.

38. By the Chapman–Kolmogorov equations and the reversibility,

$$\begin{aligned} u_0(2t) &= \sum_j \mathbb{P}\big(X(2t) = 0 \mid X(t) = j\big)\mathbb{P}\big(X(t) = j \mid X(0) = 0\big) \\ &= \sum_j \frac{\pi_0}{\pi_j}\mathbb{P}\big(X(2t) = j \mid X(t) = 0\big)u_j(t) = \pi_0 \sum_j \pi_j \left(\frac{u_j(t)}{\pi_j}\right)^2. \end{aligned}$$

The function $c(x) = -x^2$ is concave, and the claim follows by the result of the previous problem.

39. This may be done in a variety of ways, by breaking up the distribution of a typical displacement and using the superposition theorem (6.13.5), by the colouring theorem (6.13.14), or by Rényi's theorem (6.13.17) as follows. Let B be a closed bounded region of $\mathbb{R}^d$. We colour a point of Π at $\mathbf{x} \in \mathbb{R}^d$ *black* with probability $\mathbb{P}(\mathbf{x} + X \in B)$, where X is a typical displacement. By the colouring theorem, the number of black points has a Poisson distribution with parameter

$$\begin{aligned} \int_{\mathbb{R}^d} \lambda \mathbb{P}(\mathbf{x} + X \in B)\, d\mathbf{x} &= \lambda \int_{\mathbf{y} \in B} d\mathbf{y} \int_{\mathbf{x} \in \mathbb{R}^d} \mathbb{P}(X \in d\mathbf{y} - \mathbf{x}) \\ &= \lambda \int_{\mathbf{y} \in B} d\mathbf{y} \int_{\mathbf{v} \in \mathbb{R}^d} \mathbb{P}(X \in d\mathbf{v}) = \lambda |B|, \end{aligned}$$

by the change of variables $\mathbf{v} = \mathbf{y} - \mathbf{x}$. Therefore the probability that no displaced point lies in B is $e^{-\lambda|B|}$, and the claim follows by Rényi's theorem.

40. Conditional on the number $N(s)$ of points originally in the interval $(0, s)$, the positions of these points are jointly distributed as uniform random variables, so the mean number of these points which lie in $(-\infty, a)$ after the perturbation satisfies

$$\lambda s \int_0^s \frac{1}{s}\mathbb{P}(X + u \le a)\, du \to \lambda \int_0^\infty F_X(a - u)\, du = \mathbb{E}(\mathrm{R_L}) \qquad \text{as } s \to \infty,$$

where X is a typical displacement. Likewise, $\mathbb{E}(\mathrm{L_R}) = \lambda \int_0^\infty [1 - F_X(a + u)]\, du$. Equality is valid if and only if

$$\int_a^\infty [1 - F_X(v)]\, dv = \int_{-\infty}^a F_X(v)\, dv,$$

which is equivalent to $a = \mathbb{E}(X)$, by Exercise (4.3.5).

The last part follows immediately on setting $X_r = V_r t$, where V_r is the velocity of the rth car.

41. Conditional on the number $N(t)$ of arrivals by time t, the arrival times of these ants are distributed as independent random variables with the uniform distribution. Let U be a typical arrival time, so that U is uniformly distributed on $(0, t)$. The arriving ant is in the pantry at time t with probability $\pi = \mathbb{P}(U + X > t)$, or in the sink with probability $\rho = \mathbb{P}(U + X < t < U + X + Y)$, or departed with probability $1 - \rho - \pi$. Thus,

$$\begin{aligned} \mathbb{E}(x^{A(t)}y^{B(t)}) &= \mathbb{E}\{\mathbb{E}(x^{A(t)}y^{B(t)} \mid N(t))\} \\ &= \mathbb{E}\{(\pi x + \rho y + 1 - \pi - \rho)^{N(t)}\} = e^{\lambda\pi(x-1)}e^{\lambda\rho(y-1)}. \end{aligned}$$

Thus $A(t)$ and $B(t)$ are independent Poisson-distributed random variables. If the ants arrive in pairs and then separate,

$$\mathbb{E}\big(x^{A(t)}y^{B(t)} \mid N(t)\big) = \{\pi^2x^2 + 2\pi\rho xy + \rho^2y^2 + 2\gamma\pi x + 2\gamma\rho y + \gamma^2\}^{N(t)}$$

where $\gamma = 1 - \pi - \rho$. Hence,

$$\mathbb{E}(x^{A(t)}y^{B(t)}) = \exp\{\lambda\{(\pi x + \rho y + \gamma)^2 - 1\}\},$$

whence $A(t)$ and $B(t)$ are *not* independent in this case.

42. The sequence $\{X_r\}$ generates a Poisson process $N(t) = \max\{n : S_n \le t\}$. The statement that $S_n = t$ is equivalent to saying that there are $n-1$ arrivals in $(0, t)$, and in addition an arrival at t. By Theorem (6.12.7) or Theorem (6.13.11), the first $n-1$ arrival times have the required distribution.

Part (b) follows similarly, on noting that $f_{\mathbf{U}}(\mathbf{u})$ depends on $\mathbf{u} = (u_1, u_2, \dots, u_n)$ only through the constraints on the u_r.

43. Let Y be a Markov chain independent of X, having the same transition matrix and such that Y_0 has the stationary distribution $\boldsymbol{\pi}$. Let $T = \min\{n \ge 1 : X_n = Y_n\}$ and suppose $X_0 = i$. As in the proof of Theorem (6.4.17),

$$|p_{ij}(n) - \pi_j| = \left|\sum_k \pi_k\big(p_{ij}(n) - p_{kj}(n)\big)\right| \le \sum_k \pi_k \mathbb{P}(T > n) = \mathbb{P}(T > n).$$

Now,

$$\mathbb{P}(T > r+1 \mid T > r) \le 1 - \epsilon^2 \qquad \text{for } r \ge 0,$$

where $\epsilon = \min_{ij}\{p_{ij}\} > 0$. The claim follows with $\lambda = 1 - \epsilon^2$.

44. Let $I_k(n)$ be the indicator function of a visit to k at time n, so that $\mathbb{E}(I_k(n)) = \mathbb{P}(X_n = k) = a_k(n)$, say. By Problem (6.15.43), $|a_k(n) - \pi_k| \le \lambda^n$. Now,

$$\begin{aligned}\mathbb{E}\left(\left|\frac{1}{n}V_i(n) - \pi_i\right|^2\right) &= \frac{1}{n^2}\mathbb{E}\left(\left[\sum_{r=0}^{n-1}\{I_i(r) - \pi_i\}\right]^2\right)\\ &= \frac{1}{n^2}\sum_r\sum_m \mathbb{E}\{(I_i(r) - \pi_i)(I_i(m) - \pi_i)\}.\end{aligned}$$

Let $s = \min\{m, r\}$ and $t = |m - r|$. The last summation equals

$$\begin{aligned}&\frac{1}{n^2}\sum_r\sum_m\{a_i(s)p_{ii}(t) - a_i(r)\pi_i - a_i(m)\pi_i + \pi_i^2\}\\ &\quad = \frac{1}{n^2}\sum_r\sum_m\Big\{(a_i(s) - \pi_i)(p_{ii}(t) - \pi_i) + \pi_i(p_{ii}(t) - \pi_i)\\ &\qquad\qquad + \pi_i(a_i(s) - \pi_i) - \pi_i(a_i(r) - \pi_i) - \pi_i(a_i(m) - \pi_i)\Big\}\\ &\quad \le \frac{1}{n^2}\sum_r\sum_m(\lambda^{s+t} + \lambda^t + \lambda^s + \lambda^r + \lambda^m)\\ &\quad \le \frac{An}{n^2} \to 0 \qquad \text{as } n \to \infty,\end{aligned}$$

where $0 < A < \infty$. For the last part, use the fact that $\sum_{r=0}^{n-1} f(X_r) = \sum_{i \in S} f(i)V_i(n)$. The result is obtained by Minkowski's inequality (Problem (4.14.27b)) and the first part.

45. We have by the Markov property that $f(X_{n+1} \mid X_n, X_{n-1}, \dots, X_0) = f(X_{n+1} \mid X_n)$, whence

$$\mathbb{E}\big(\log f(X_{n+1} \mid X_n, X_{n-1}, \dots, X_0) \,\big|\, X_n, \dots, X_0\big) = \mathbb{E}\big(\log f(X_{n+1} \mid X_n) \,\big|\, X_n\big).$$

Taking the expectation of each side gives the result. Furthermore,

$$H(X_{n+1} \mid X_n) = -\sum_{i,j}(p_{ij}\log p_{ij})\mathbb{P}(X_n = i).$$

Now X has a unique stationary distribution $\boldsymbol{\pi}$, so that $\mathbb{P}(X_n = i) \to \pi_i$ as $n \to \infty$. The state space is finite, and the claim follows.

46. Let $T = \inf\{t : X_t = Y_t\}$. Since X and Y are persistent, and since each process moves by distance 1 at continuously distributed times, it is the case that $\mathbb{P}(T < \infty) = 1$. We define

$$Z_t = \begin{cases} X_t & \text{if } t < T, \\ Y_t & \text{if } t \geq T, \end{cases}$$

noting that the processes X and Z have the same distributions.

(a) By the above remarks,

$$\begin{aligned} |\mathbb{P}(X_t = k) - \mathbb{P}(Y_t = k)| &= |\mathbb{P}(Z_t = k) - \mathbb{P}(Y_t = k)| \\ &\leq \big|\mathbb{P}(Z_t = k,\ T \leq t) + \mathbb{P}(Z_t = k,\ T > t) - \mathbb{P}(Y_t = k,\ T \leq t) - \mathbb{P}(Y_t = k,\ T > t)\big| \\ &\leq \mathbb{P}(X_t = k,\ T > t) + \mathbb{P}(Y_t = k,\ T > t). \end{aligned}$$

We sum over $k \in A$, and let $t \to \infty$.

(b) We have in this case that $Z_t \leq Y_t$ for all t. The claim follows from the fact that X and Z are processes with the same distributions.

47. We reformulate the problem in the following way. Suppose there are two containers, W and N, containing n particles in all. During the time interval $(t, t + dt)$, any particle in W moves to N with probability $\mu dt + \mathrm{o}(dt)$, and any particle in N moves to W with probability $\lambda dt + \mathrm{o}(dt)$. The particles move independently of one another. The number $Z(t)$ of particles in W has the same rules of evolution as the process X in the original problem. Now, $Z(t)$ may be expressed as the sum of two independent random variables U and V, where U is $\text{bin}(r, \theta_t)$, V is $\text{bin}(n - r, \psi_t)$, and θ_t is the probability that a particle starting in W is in W at time t, ψ_t is the probability that a particle starting in N at 0 is in W at t. By considering the two-state Markov chain of Exercise (6.9.1),

$$\theta_t = \frac{\lambda + \mu e^{-(\lambda+\mu)t}}{\lambda + \mu}, \qquad \psi_t = \frac{\lambda - \lambda e^{-(\lambda+\mu)t}}{\lambda + \mu},$$

and therefore

$$\mathbb{E}(s^{X(t)}) = \mathbb{E}(s^U)\mathbb{E}(s^V) = (s\theta_t + 1 - s)^r(s\psi_t + 1 - s)^{n-r}.$$

Also, $\mathbb{E}(X(t)) = r\theta_t + (n - r)\psi_t$ and $\text{var}(X(t)) = r\theta_t(1 - \theta_t) + (n - r)\psi_t(1 - \psi_t)$. In the limit as $n \to \infty$, the distribution of $X(t)$ approaches the $\text{bin}(n, \lambda/(\lambda + \mu))$ distribution.

48. Solving the equations

$$\pi_0 = q_1\pi_1 + p_2\pi_2, \quad \pi_1 = q_2\pi_2 + p_0\pi_0, \quad \sum_i \pi_i = 1,$$

gives the first claim. We have that $\gamma = \sum_i (p_i - q_i)\pi_i$, and the formula for γ follows.

Considering the three walks in order, we have that:

A. $\pi_i = \frac{1}{3}$ for each i, and $\gamma_{\text{A}} = -2a < 0$.

B. Substitution in the formula for γ_{B} gives the numerator as $3\{-\frac{49}{40}a + \mathrm{o}(a)\}$, which is negative for small a whereas the denominator is positive.

C. The transition probabilities are the averages of those for A and B, namely, $p_0 = \frac{1}{2}(\frac{1}{10} - a) + \frac{1}{2}(\frac{1}{2} - a) = \frac{3}{10} - a$, and so on. The numerator in the formula for γ_C equals $\frac{9}{160} + o(1)$, which is positive for small a.

49. Call a car *green* if it satisfies the given condition. The chance that a green car arrives on the scene during the time interval $(u, u+h)$ is $\lambda h \mathbb{P}(V < x/(t-u))$ for $u < t$. Therefore, the arrival process of green cars is an inhomogeneous Poisson process with rate function

$$\lambda(u) = \begin{cases} \lambda \mathbb{P}\big(V < x/(t-u)\big) & \text{if } u < t, \\ 0 & \text{if } u \geq t. \end{cases}$$

Hence the required number has the Poisson distribution with mean

$$\begin{aligned} \lambda \int_0^t \mathbb{P}\left(V < \frac{x}{t-u}\right) du &= \lambda \int_0^t \mathbb{P}\left(V < \frac{x}{u}\right) du \\ &= \lambda \int_0^t \mathbb{E}(I_{\{Vu<x\}})\, du = \lambda \mathbb{E}\big(V^{-1} \min\{x, Vt\}\big). \end{aligned}$$

50. The answer is the probability of exactly one arrival in the interval (s, t), which equals $g(s) = \lambda(t-s)e^{-\lambda(t-s)}$. By differentiation, g has its maximum at $\bar{s} = \max\{0, t - \lambda^{-1}\}$, and $g(\bar{s}) = e^{-1}$ when $t \geq \lambda^{-1}$.

51. We measure money in millions and time in hours. The number of available houses has the Poisson distribution with parameter 30λ, whence the number A of affordable houses has the Poisson distribution with parameter $\frac{1}{6} \cdot 30\lambda = 5\lambda$ (cf. Exercise (3.5.2)). Since each viewing time T has moment generating function $\mathbb{E}(e^{\theta T}) = (e^{2\theta} - e^{\theta})/\theta$, the answer is

$$G_A\big(\mathbb{E}(e^{\theta T})\big) = \exp\big\{5\lambda(e^{2\theta} - e^{\theta} - \theta)/\theta\big\}.$$

7

Convergence of random variables

7.1 Solutions. Introduction

1. (a) $\mathbb{E}|(cX)^r| = |c|^r \cdot \{\|X\|_r\}^r$.
(b) This is Minkowski's inequality.
(c) Let $\epsilon > 0$. Certainly $|X| \geq I_\epsilon$ where I_ϵ is the indicator function of the event $\{|X| > \epsilon\}$. Hence $\mathbb{E}|X^r| \geq \mathbb{E}|I_\epsilon^r| = \mathbb{P}(|X| > \epsilon)$, implying that $\mathbb{P}(|X| > \epsilon) = 0$ for all $\epsilon > 0$. The converse is trivial.

2. (a) $\mathbb{E}(\{aX + bY\}Z) = a\mathbb{E}(XZ) + b\mathbb{E}(YZ)$.
(b) $\mathbb{E}(\{X+Y\}^2) + \mathbb{E}(\{X-Y\}^2) = 2\mathbb{E}(X^2) + 2\mathbb{E}(Y^2)$.
(c) Clearly

$$\mathbb{E}\left(\left\{\sum_{i=1}^n X_i\right\}^2\right) = \sum_{i=1}^n \mathbb{E}(X_i^2) + 2\sum_{i<j} \mathbb{E}(X_i X_j).$$

3. Let $f(u) = \frac{2}{3}\epsilon$, $g(u) = 0$, $h(u) = -\frac{2}{3}\epsilon$, for all u. Then $d_\epsilon(f,g) + d_\epsilon(g,h) = 0$ whereas $d_\epsilon(f,h) = 1$.

4. Either argue directly, or as follows. With any distribution function F, we may associate a graph $\widetilde{F}$ obtained by adding to the graph of F vertical line segments connecting the two endpoints at each discontinuity of F. By drawing a picture, you may see that $\sqrt{2}\,d(F,G)$ equals the maximum distance between $\widetilde{F}$ and $\widetilde{G}$ measured along lines of slope -1. It is now clear that $d(F,G) = 0$ if and only if $F = G$, and that $d(F,G) = d(G,F)$. Finally, by the triangle inequality for real numbers, we have that $d(F,H) \leq d(F,G) + d(G,H)$.

5. Take X to be any random variable satisfying $\mathbb{E}(X^2) = \infty$, and define $X_n = X$ for all n.

7.2 Solutions. Modes of convergence

1. (a) By Minkowski's inequality,

$$\{\mathbb{E}|X^r|\}^{1/r} \leq \{\mathbb{E}(|X_n - X|^r)\}^{1/r} + \{\mathbb{E}|X_n^r|\}^{1/r};$$

let $n \to \infty$ to obtain $\liminf_{n\to\infty} \mathbb{E}|X_n^r| \geq \mathbb{E}|X^r|$. By another application of Minkowski's inequality,

$$\{\mathbb{E}|X_n^r|\}^{1/r} \leq \{\mathbb{E}(|X_n - X|^r)\}^{1/r} + \{\mathbb{E}|X^r|\}^{1/r},$$

whence $\limsup_{n\to\infty} \mathbb{E}|X_n^r| \leq \mathbb{E}|X^r|$.

(b) We have that

$$|\mathbb{E}(X_n) - \mathbb{E}(X)| = |\mathbb{E}(X_n - X)| \le \mathbb{E}|X_n - X| \to 0$$

as $n \to \infty$. The converse is clearly false. If each X_n takes the values ± 1, each with probability $\frac{1}{2}$, then $\mathbb{E}(X_n) = 0$, but $\mathbb{E}|X_n - 0| = 1$.

(c) By part (a), $\mathbb{E}(X_n^2) \to \mathbb{E}(X^2)$. Now $X_n \xrightarrow{1} X$ by Theorem (7.2.3), and therefore $\mathbb{E}(X_n) \to \mathbb{E}(X)$ by part (b). Therefore $\text{var}(X_n) = \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2 \to \text{var}(X)$.

2. Assume that $X_n \xrightarrow{\text{P}} X$. Since $|X_n| \le Z$ for all n, it is the case that $|X| \le Z$ a.s. Therefore $Z_n = |X_n - X|$ satisfies $Z_n \le 2Z$ a.s. In addition, if $\epsilon > 0$,

$$\mathbb{E}|Z_n| = \mathbb{E}\left(Z_n I_{\{Z_n \le \epsilon\}}\right) + \mathbb{E}\left(Z_n I_{\{Z_n > \epsilon\}}\right) \le \epsilon + 2\mathbb{E}\left(Z I_{\{Z_n > \epsilon\}}\right).$$

As $n \to \infty$, $\mathbb{P}(|Z_n| > \epsilon) \to 0$, and therefore the last term tends to 0; to see this, use the fact that $\mathbb{E}(Z) < \infty$, together with the result of Exercise (5.6.5). Now let $\epsilon \downarrow 0$ to obtain that $\mathbb{E}|Z_n| \to 0$ as $n \to \infty$.

3. We have that $X - n^{-1} \le X_n \le X$, so that $\mathbb{E}(X_n) \to \mathbb{E}(X)$, and similarly $\mathbb{E}(Y_n) \to \mathbb{E}(Y)$. By the independence of X_n and Y_n,

$$\mathbb{E}(X_n Y_n) = \mathbb{E}(X_n)\mathbb{E}(Y_n) \to \mathbb{E}(X)\mathbb{E}(Y).$$

Finally, $(X - n^{-1})(Y - n^{-1}) \le X_n Y_n \le XY$, and

$$\mathbb{E}\left\{\left(X - \frac{1}{n}\right)\left(Y - \frac{1}{n}\right)\right\} = \mathbb{E}(XY) - \frac{\mathbb{E}(X) + \mathbb{E}(Y)}{n} + \frac{1}{n^2} \to \mathbb{E}(XY)$$

as $n \to \infty$, so that $\mathbb{E}(X_n Y_n) \to \mathbb{E}(XY)$.

4. Let $F_1, F_2, \dots$ be distribution functions. As in Section 5.9, we write $F_n \to F$ if $F_n(x) \to F(x)$ for all x at which F is continuous. We are required to prove that $F_n \to F$ if and only if $d(F_n, F) \to 0$.

Suppose that $d(F_n, F) \to 0$. Then, for $\epsilon > 0$, there exists N such that

$$F(x - \epsilon) - \epsilon \le F_n(x) \le F(x + \epsilon) + \epsilon \qquad \text{for all } x.$$

Take the limits as $n \to \infty$ and $\epsilon \to 0$ in that order, to find that $F_n(x) \to F(x)$ whenever F is continuous at x.

Suppose that $F_n \to F$. Let $\epsilon > 0$, and find real numbers $a = x_1 < x_2 < \cdots < x_n = b$, each being points of continuity of F, such that

(i) $F_i(a) < \epsilon$ for all i, $F(b) > 1 - \epsilon$,

(ii) $|x_{i+1} - x_i| < \epsilon$ for $1 \le i < n$.

In order to pick a such that $F_i(a) < \epsilon$ for all i, first choose a' such that $F(a') < \frac{1}{2}\epsilon$ and F is continuous at a', then find M such that $|F_m(a') - F(a')| < \frac{1}{2}\epsilon$ for $m \ge M$, and lastly find a continuity point a of F such that $a \le a'$ and $F_m(a) < \epsilon$ for $1 \le m < M$.

There are finitely many points x_i, and therefore there exists N such that $|F_m(x_i) - F(x_i)| < \epsilon$ for all i and $m \ge N$. Now, if $m \ge N$ and $x_i \le x < x_{i+1}$,

$$F_m(x) \le F_m(x_{i+1}) < F(x_{i+1}) + \epsilon \le F(x + \epsilon) + \epsilon,$$

and similarly

$$F_m(x) \ge F_m(x_i) > F(x_i) - \epsilon \ge F(x - \epsilon) - \epsilon.$$

Similar inequalities hold if $x \le a$ or $x \ge b$, and it follows that $d(F_m, F) < \epsilon$ if $m \ge N$. Therefore $d(F_m, F) \to 0$ as $m \to \infty$.

5. (a) Suppose $c > 0$ and pick δ such that $0 < \delta < c$. Find N such that $\mathbb{P}(|Y_n - c| > \delta) < \delta$ for $n \ge N$. Now, for $x \ge 0$,

$$\mathbb{P}(X_n Y_n \le x) \le \mathbb{P}\big(X_n Y_n \le x, |Y_n - c| \le \delta\big) + \mathbb{P}\big(|Y_n - c| > \delta\big) \le \mathbb{P}\left(X_n \le \frac{x}{c-\delta}\right) + \delta,$$

and similarly

$$\mathbb{P}(X_n Y_n > x) \le \mathbb{P}\big(X_n Y_n > x, |Y_n - c| \le \delta\big) + \delta \le \mathbb{P}\left(X_n > \frac{x}{c+\delta}\right) + \delta.$$

Taking the limits as $n \to \infty$ and $\delta \downarrow 0$, we find that $\mathbb{P}(X_n Y_n \le x) \to \mathbb{P}(X \le x/c)$ if x/c is a point of continuity of the distribution function of X. A similar argument holds if $x < 0$, and we conclude that $X_n Y_n \xrightarrow{\text{D}} cX$ if $c > 0$. No extra difficulty arises if $c < 0$, and the case $c = 0$ is similar.

For the second part, it suffices to prove that $Y_n^{-1} \xrightarrow{\text{P}} c^{-1}$ if $Y_n \xrightarrow{\text{P}} c$ $(\ne 0)$. This is immediate from the fact that $|Y_n^{-1} - c^{-1}| < \epsilon/\{|c|(|c| - \epsilon)\}$ if $|Y_n - c| < \epsilon$ $(< |c|)$.
(b) Let $\epsilon > 0$. There exists N such that

$$\mathbb{P}(|X_n| > \epsilon) < \epsilon, \quad \mathbb{P}\big(|Y_n - Y| > \epsilon\big) < \epsilon, \quad \text{if } n \ge N,$$

and in addition $\mathbb{P}(|Y| > N) < \epsilon$. By an elementary argument, g is uniformly continuous at points of the form $(0, y)$ for $|y| \le N$. Therefore there exists δ (> 0) such that

$$|g(x', y') - g(0, y)| < \epsilon \qquad \text{if } |x'| \le \delta,\ |y' - y| \le \delta.$$

If $|X_n| \le \delta$, $|Y_n - Y| \le \delta$, and $|Y| \le N$, then $|g(X_n, Y_n) - g(0, Y)| < \epsilon$, so that

$$\mathbb{P}\big(|g(X_n, Y_n) - g(0, Y)| \ge \epsilon\big) \le \mathbb{P}(|X_n| > \delta) + \mathbb{P}\big(|Y_n - Y| > \delta\big) + \mathbb{P}(|Y| > N) \le 3\epsilon,$$

for $n \ge N$. Therefore $g(X_n, Y_n) \xrightarrow{\text{P}} g(0, Y)$ as $n \to \infty$.

6. The subset A of the sample space Ω may be expressed thus:

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \{|X_{n+m} - X_n| < k^{-1}\},$$

a countable sequence of intersections and unions of events.

For the last part, define

$$X(\omega) = \begin{cases} \lim_{n\to\infty} X_n(\omega) & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The function X is $\mathcal{F}$-measurable since $A \in \mathcal{F}$.

7. (a) If $X_n(\omega) \to X(\omega)$ then $c_n X_n(\omega) \to cX(\omega)$.
(b) We have by Minkowski's inequality that, as $n \to \infty$,

$$\mathbb{E}\big(|c_n X_n - cX|^r\big) \le |c_n|^r \mathbb{E}\big(|X_n - X|^r\big) + |c_n - c|^r \mathbb{E}|X^r| \to 0.$$

(c) If $c = 0$, the claim is nearly obvious. Otherwise $c \ne 0$, and we may assume that $c > 0$. For $0 < \epsilon < c$, there exists N such that $|c_n - c| < \epsilon$ whenever $n \ge N$. By the triangle inequality, $|c_n X_n - cX| \le |c_n(X_n - X)| + |(c_n - c)X|$, so that, for $n \ge N$,

$$\begin{aligned}
\mathbb{P}\big(|c_n X_n - cX| > \epsilon\big) &\le \mathbb{P}\big(c_n|X_n - X| > \tfrac{1}{2}\epsilon\big) + \mathbb{P}\big(|c_n - c| \cdot |X| > \tfrac{1}{2}\epsilon\big) \\
&\le \mathbb{P}\left(|X_n - X| > \frac{\epsilon}{2(c+\epsilon)}\right) + \mathbb{P}\left(|X| > \frac{\epsilon}{2|c_n - c|}\right) \\
&\to 0 \qquad \text{as } n \to \infty.
\end{aligned}$$

(d) A neat way is to use the Skorokhod representation (7.2.14). If $X_n \xrightarrow{\text{D}} X$, find random variables Y_n, Y with the same distributions such that $Y_n \xrightarrow{\text{a.s.}} Y$. Then $c_nY_n \xrightarrow{\text{a.s.}} cY$, so that $c_nY_n \xrightarrow{\text{D}} cY$, implying the same conclusion for the X's.

8. If X is not a.s. constant, there exist real numbers c and ϵ such that $0 < \epsilon < \frac{1}{2}$ and $\mathbb{P}(X < c) > 2\epsilon$, $\mathbb{P}(X > c + \epsilon) > 2\epsilon$. Since $X_n \xrightarrow{\text{P}} X$, there exists N such that

$$\mathbb{P}(X_n < c) > \epsilon, \quad \mathbb{P}(X_n > c + \epsilon) > \epsilon, \quad \text{if } n \ge N.$$

Also, by the triangle inequality, $|X_r - X_s| \le |X_r - X| + |X_s - X|$; therefore there exists M such that $\mathbb{P}(|X_r - X_s| > \epsilon) < \epsilon^3$ for $r, s \ge M$. Assume now that the X_n are independent. Then, for $r, s \ge \max\{M, N\}$, $r \ne s$,

$$\epsilon^3 > \mathbb{P}(|X_r - X_s| > \epsilon) \ge \mathbb{P}(X_r < c, X_s > c + \epsilon) = \mathbb{P}(X_r < c)\mathbb{P}(X_s > c + \epsilon) > \epsilon^2,$$

a contradiction.

9. Either use the fact (Exercise (4.12.3)) that convergence in total variation implies convergence in distribution, together with Theorem (7.2.19), or argue directly thus. Since $|u(\cdot)| \le K < \infty$,

$$\left|\mathbb{E}(u(X_n)) - \mathbb{E}(u(X))\right| = \left|\sum_k u(k)\{f_n(k) - f(k)\}\right| \le K \sum_k |f_n(k) - f(k)| \to 0.$$

10. The partial sum $S_n = \sum_{r=1}^n X_r$ is Poisson-distributed with parameter $\sigma_n = \sum_{r=1}^n \lambda_r$. For fixed x, the event $\{S_n \le x\}$ is decreasing in n, whence by Lemma (1.3.5), if $\sigma_n \to \sigma < \infty$ and x is a non-negative integer,

$$\mathbb{P}\left(\sum_{r=1}^{\infty} X_r \le x\right) = \lim_{n\to\infty} \mathbb{P}(S_n \le x) = \sum_{j=0}^{x} \frac{e^{-\sigma}\sigma^j}{j!}.$$

Hence if $\sigma < \infty$, $\sum_{r=1}^{\infty} X_r$ converges to a Poisson random variable. On the other hand, if $\sigma_n \to \infty$, then $e^{-\sigma_n}\sum_{j=0}^{x} \sigma_n^j/j! \to 0$, giving that $\mathbb{P}\left(\sum_{r=1}^{\infty} X_r > x\right) = 1$ for all x, and therefore the sum diverges with probability 1, as required.

7.3 Solutions. Some ancillary results

1. (a) If $|X_n - X_m| > \epsilon$ then either $|X_n - X| > \frac{1}{2}\epsilon$ or $|X_m - X| > \frac{1}{2}\epsilon$, so that

$$\mathbb{P}(|X_n - X_m| > \epsilon) \le \mathbb{P}(|X_n - X| > \tfrac{1}{2}\epsilon) + \mathbb{P}(|X_m - X| > \tfrac{1}{2}\epsilon) \to 0$$

as $n, m \to \infty$, for $\epsilon > 0$.

Conversely, suppose that $\{X_n\}$ is Cauchy convergent in probability. For each positive integer k, there exists n_k such that

$$\mathbb{P}(|X_n - X_m| \ge 2^{-k}) < 2^{-k} \qquad \text{for } n, m \ge n_k.$$

The sequence (n_k) may not be increasing, and we work instead with the sequence defined by $N_1 = n_1$, $N_{k+1} = \max\{N_k + 1, n_{k+1}\}$. We have that

$$\sum_k \mathbb{P}(|X_{N_{k+1}} - X_{N_k}| \ge 2^{-k}) < \infty,$$

whence, by the first Borel–Cantelli lemma, a.s. only finitely many of the events $\{|X_{N_{k+1}} - X_{N_k}| \geq 2^{-k}\}$ occur. Therefore, the expression

$$X = X_{N_1} + \sum_{k=1}^{\infty}(X_{N_{k+1}} - X_{N_k})$$

converges absolutely on an event C having probability one. Define $X(\omega)$ accordingly for $\omega \in C$, and $X(\omega) = 0$ for $\omega \notin C$. We have, by the definition of X, that $X_{N_k} \xrightarrow{\text{a.s.}} X$ as $k \to \infty$. Finally, we 'fill in the gaps'. As before, for $\epsilon > 0$,

$$\mathbb{P}\big(|X_n - X| > \epsilon\big) \leq \mathbb{P}\big(|X_n - X_{N_k}| > \tfrac{1}{2}\epsilon\big) + \mathbb{P}\big(|X_{N_k} - X| > \tfrac{1}{2}\epsilon\big) \to 0$$

as $n, k \to \infty$, where we are using the assumption that $\{X_n\}$ is Cauchy convergent in probability.

(b) Since $X_n \xrightarrow{\text{P}} X$, the sequence $\{X_n\}$ is Cauchy convergent in probability. Hence

$$\mathbb{P}\big(|Y_n - Y_m| > \epsilon\big) = \mathbb{P}\big(|X_n - X_m| > \epsilon\big) \to 0 \quad \text{as } n, m \to \infty,$$

for $\epsilon > 0$. Therefore $\{Y_n\}$ is Cauchy convergent also, and the sequence converges in probability to some limit Y. Finally, $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{D}} X$, so that X and Y have the same distribution.

2. Since $A_n \subseteq \bigcup_{m=n}^{\infty} A_m$, we have that

$$\limsup_{n\to\infty} \mathbb{P}(A_n) \leq \lim_{n\to\infty} \mathbb{P}\bigg(\bigcup_{m=n}^{\infty} A_m\bigg) = \mathbb{P}\bigg(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} A_m\bigg) = \mathbb{P}(A_n \text{ i.o.}),$$

where we have used the continuity of $\mathbb{P}$. Alternatively, apply Fatou's lemma to the sequence $I_{A_n^c}$ of indicator functions.

3. (a) Suppose $X_{2n} = 1$, $X_{2n+1} = -1$, for $n \geq 1$. Then $\{S_n = 0 \text{ i.o.}\}$ occurs if $X_1 = -1$, and not if $X_1 = 1$. The event is therefore not in the tail σ-field of the X's.

(b) Here is a way. As usual, $\mathbb{P}(S_{2n} = 0) = \binom{2n}{n}\{p(1-p)\}^n$, so that

$$\sum_n \mathbb{P}(S_{2n} = 0) < \infty \quad \text{if } p \neq \tfrac{1}{2},$$

implying by the first Borel–Cantelli lemma that $\mathbb{P}(S_n = 0 \text{ i.o.}) = 0$.

(c) Changing the values of any finite collection of the steps has no effect on $I = \liminf T_n$ and $J = \limsup T_n$, since such changes are extinguished in the limit by the denominator '$\sqrt{n}$'. Hence I and J are tail functions, and are measurable with respect to the tail σ-field. In particular, $\{I \leq -x\} \cap \{J \geq x\}$ lies in the σ-field.

Take $x = 1$, say. Then, $\mathbb{P}(I \leq -1) = \mathbb{P}(J \geq 1)$ by symmetry; using Exercise (7.3.2) and the central limit theorem,

$$\mathbb{P}(J \geq 1) \geq \mathbb{P}(S_n \geq \sqrt{n} \text{ i.o.}) \geq \limsup_{n\to\infty} \mathbb{P}(S_n \geq \sqrt{n}) = 1 - \Phi(1) > 0,$$

where Φ is the $N(0, 1)$ distribution function. Since $\{J \geq 1\}$ is a tail event of an independent sequence, it has probability either 0 or 1, and therefore $\mathbb{P}(I \leq -1) = \mathbb{P}(J \geq 1) = 1$, and also $\mathbb{P}(I \leq -1, J \geq 1) = 1$. That is, on an event having probability one, each visit of the walk to the left of $-\sqrt{n}$ is followed by a visit of the walk to the right of $\sqrt{n}$, and vice versa. It follows that the walk visits 0 infinitely often, with probability one.

4. Let A be exchangeable. Since A is defined in terms of the X_i, it follows by a standard result of measure theory that, for each n, there exists an event $A_n \in \sigma(X_1, X_2, \dots, X_n)$, such that $\mathbb{P}(A \triangle A_n) \to 0$ as $n \to \infty$. We may express A_n and A in the form

$$A_n = \{\mathbf{X}_n \in B_n\}, \quad A = \{\mathbf{X} \in B\},$$

where $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$, and B_n and B are appropriate subsets of $\mathbb{R}^n$ and $\mathbb{R}^\infty$. Let

$$A'_n = \{\mathbf{X}'_n \in B_n\}, \quad A' = \{\mathbf{X}' \in B\},$$

where $\mathbf{X}'_n = (X_{n+1}, X_{n+2}, \dots, X_{2n})$ and $\mathbf{X}' = (X_{n+1}, X_{n+2}, \dots, X_{2n}, X_1, X_2, \dots, X_n, X_{2n+1}, X_{2n+2}, \dots)$.

Now $\mathbb{P}(A_n \cap A'_n) = \mathbb{P}(A_n)\mathbb{P}(A'_n)$, by independence. Also, $\mathbb{P}(A_n) = \mathbb{P}(A'_n)$, and therefore

$$(*) \qquad \mathbb{P}(A_n \cap A'_n) = \mathbb{P}(A_n)^2 \to \mathbb{P}(A)^2 \quad \text{as } n \to \infty.$$

By the exchangeability of A, we have that $\mathbb{P}(A \triangle A'_n) = \mathbb{P}(A' \triangle A'_n)$, which in turn equals $\mathbb{P}(A \triangle A_n)$, using the fact that the X_i are independent and identically distributed. Therefore,

$$|\mathbb{P}(A_n \cap A'_n) - \mathbb{P}(A)| \le \mathbb{P}(A \triangle A_n) + \mathbb{P}(A \triangle A'_n) \to 0 \quad \text{as } n \to \infty.$$

Combining this with $(*)$, we obtain that $\mathbb{P}(A) = \mathbb{P}(A)^2$, and hence $\mathbb{P}(A)$ equals 0 or 1.

5. The value of S_n does not depend on the order of the first n steps, but only on their sum. If $S_n = 0$ i.o., then $S'_n = 0$ i.o. for all walks $\{S'_n\}$ obtained from $\{S_n\}$ by permutations of finitely many steps.

6. Since f is continuous on a closed interval, it is bounded: $|f(y)| \le c$ for all $y \in [0, 1]$ for some c. Furthermore f is uniformly continuous on $[0, 1]$, which is to say that, if $\epsilon > 0$, there exists δ (> 0), such that $|f(y) - f(z)| < \epsilon$ if $|y - z| \le \delta$. With this choice of ϵ, δ, we have that $|\mathbb{E}(ZI_{A^c})| < \epsilon$, and

$$|\mathbb{E}(ZI_A)| \le 2c\mathbb{P}(A) \le 2c \cdot \frac{x(1-x)}{n\delta^2}$$

by Chebyshov's inequality. Therefore

$$|\mathbb{E}(Z)| < \epsilon + \frac{2c}{n\delta^2},$$

which is less than 2ϵ for values of n exceeding $2c/(\epsilon\delta^2)$.

7. If $\{X_n\}$ converges completely to X then, by the first Borel–Cantelli lemma, $|X_n - X| > \epsilon$ only finitely often with probability one, for all $\epsilon > 0$. This implies that $X_n \xrightarrow{\text{a.s.}} X$; see Theorem (7.2.4c).

Suppose conversely that $\{X_n\}$ is a sequence of independent variables which converges almost surely to X. By Exercise (7.2.8), X is almost surely constant, and we may therefore suppose that $X_n \xrightarrow{\text{a.s.}} c$ where $c \in \mathbb{R}$. It follows that, for $\epsilon > 0$, only finitely many of the (independent) events $\{|X_n - c| > \epsilon\}$ occur, with probability one. Using the second Borel–Cantelli lemma,

$$\sum_n \mathbb{P}\big(|X_n - c| > \epsilon\big) < \infty.$$

8. Of the various ways of doing this, here is one. We have that

$$\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} X_i X_j = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 - \frac{1}{n(n-1)}\sum_{i=1}^n X_i^2.$$

Now $n^{-1}\sum_1^n X_i \xrightarrow{\text{D}} \mu$, by the law of large numbers (5.10.2); hence $n^{-1}\sum_1^n X_i \xrightarrow{\text{P}} \mu$ (use Theorem (7.2.4a)). It follows that $(n^{-1}\sum_1^n X_i)^2 \xrightarrow{\text{P}} \mu^2$; to see this, either argue directly or use Problem (7.11.3). Now use Exercise (7.2.7) to find that

$$\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \xrightarrow{\text{P}} \mu^2.$$

Arguing similarly,

$$\frac{1}{n(n-1)}\sum_{i=1}^n X_i^2 \xrightarrow{\text{P}} 0,$$

and the result follows by the fact (Theorem (7.3.9)) that the sum of these two expressions converges in probability to the sum of their limits.

9. Evidently,

$$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1+\epsilon\right) = \frac{1}{n^{1+\epsilon}}, \qquad \text{for } |\epsilon| < 1.$$

By the Borel–Cantelli lemmas, the events $A_n = \{X_n/\log n \geq 1+\epsilon\}$ occur a.s. infinitely often for $-1 < \epsilon \leq 0$, and a.s. only finitely often for $\epsilon > 0$.

10. (a) Mills's ratio (Exercise (4.4.8) or Problem (4.14.1c)) informs us that $1-\Phi(x) \sim x^{-1}\phi(x)$ as $x \to \infty$. Therefore,

$$\mathbb{P}\big(|X_n| \geq \sqrt{2\log n}(1+\epsilon)\big) \sim \frac{1}{\sqrt{2\pi}\,\log n(1+\epsilon)n^{(1+\epsilon)^2}}.$$

The sum over n of these terms converges if and only if $\epsilon > 0$, and the Borel–Cantelli lemmas imply the claim.

(b) This is an easy implication of the Borel–Cantelli lemmas.

11. Let X be uniformly distributed on the interval $[-1, 1]$, and define $X_n = I_{\{X \leq (-1)^n/n\}}$. The distribution of X_n approaches the Bernoulli distribution which takes the values ± 1 with equal probability $\frac{1}{2}$. The median of X_n is 1 if n is even and -1 if n is odd.

12. (i) We have that

$$\sum_{r=1}^{\infty}\mathbb{P}\left(\frac{X_r}{r} \geq x\right) = \sum_{r=1}^{\infty}\mathbb{P}\left(\frac{X_r}{x} \geq r\right) = \frac{\mathbb{E}(X_r)}{x} = \infty$$

for $x > 0$. The result follows by the second Borel–Cantelli lemma.

(ii) (a) The stationary distribution $\boldsymbol{\pi}$ is found in the usual way to satisfy

$$\pi_k = \frac{k-1}{k+1}\pi_{k-1} = \cdots = \frac{2}{k(k+1)}\pi_1, \qquad k \geq 2.$$

Hence $\pi_k = \{k(k+1)\}^{-1}$ for $k \geq 1$, a distribution with mean $\sum_{k=1}^{\infty}(k+1)^{-1} = \infty$.

(b) By construction, $\mathbb{P}(X_n \leq X_0 + n) = 1$ for all n, whence

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{X_n}{n} \leq 1\right) = 1.$$

It may in fact be shown that $\mathbb{P}\big(\limsup_{n\to\infty} X_n/n = 0\big) = 1$.

13. We divide the numerator and denominator by $\sqrt{n}\sigma$. By the central limit theorem, the former converges in distribution to the $N(0, 1)$ distribution. We expand the new denominator, squared, as

$$\frac{1}{n\sigma^2}\sum_{r=1}^{n}(X_r-\mu)^2 - \frac{2}{n\sigma^2}(\overline{X}-\mu)\sum_{r=1}^{n}(X_r-\mu) + \frac{1}{\sigma^2}(\overline{X}-\mu)^2.$$

By the weak law of large numbers (Theorem (5.10.2), combined with Theorem (7.2.3)), the first term converges in probability to 1, and the other terms to 0. Their sum converges to 1, by Theorem (7.3.9), and the result follows by Slutsky's theorem, Exercise (7.2.5).

7.4 Solutions. Laws of large numbers

1. Let $S_n = X_1 + X_2 + \cdots + X_n$. Then

$$\mathbb{E}(S_n^2) = \sum_{i=2}^{n}\frac{i}{\log i} \le \frac{n^2}{\log n}$$

and therefore $S_n/n \xrightarrow{\text{m.s.}} 0$. On the other hand, $\sum_i \mathbb{P}(|X_i| \ge i) = 1$, so that $|X_i| \ge i$ i.o., with probability one, by the second Borel–Cantelli lemma. For such a value of i, we have that $|S_i - S_{i-1}| \ge i$, implying that S_n/n does not converge, with probability one.

2. Let the X_n satisfy

$$\mathbb{P}(X_n = -n) = 1 - \frac{1}{n^2}, \qquad \mathbb{P}(X_n = n^3 - n) = \frac{1}{n^2},$$

whence they have zero mean. However,

$$\sum_n \mathbb{P}\left(\frac{X_n}{n} \ne -1\right) = \sum_n \frac{1}{n^2} < \infty,$$

implying by the first Borel–Cantelli lemma that $\mathbb{P}(X_n/n \to -1) = 1$. It is an elementary result of real analysis that $n^{-1}\sum_{r=1}^{n} x_n \to -1$ if $x_n \to -1$, and the claim follows.

3. The random variable $N(S)$ has mean and variance $\lambda|S| = cr^d$, where c is a constant depending only on d. By Chebyshov's inequality,

$$\mathbb{P}\left(\left|\frac{N(S)}{|S|} - \lambda\right| \ge \epsilon\right) \le \frac{\lambda}{\epsilon^2|S|} = \left(\frac{\lambda}{\epsilon}\right)^2 \frac{1}{cr^d}.$$

By the first Borel–Cantelli lemma, $\big||S_k|^{-1}N(S_k) - \lambda\big| \ge \epsilon$ for only finitely many integers k, a.s., where S_k is the sphere of radius k. It follows that $N(S_k)/|S_k| \xrightarrow{\text{a.s.}} \lambda$ as $k \to \infty$. The same conclusion holds as $k \to \infty$ through the reals, since $N(S)$ is non-decreasing in the radius of S.

7.5 Solutions. The strong law

1. Let I_{ij} be the indicator function of the event that X_j lies in the ith interval. Then

$$\log R_m = \sum_{i=1}^{n} Z_m(i) \log p_i = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} \log p_i = \sum_{j=1}^{m} Y_j$$

where, for $1 \le j \le m$, $Y_j = \sum_{i=1}^{n} I_{ij} \log p_i$ is the sum of independent identically distributed variables with mean

$$\mathbb{E}(Y_j) = \sum_{i=1}^{n} p_i \log p_i = -h.$$

By the strong law, $m^{-1} \log R_m \xrightarrow{\text{a.s.}} -h$.

2. The following two observations are clear:

(a) $N(t) < n$ if and only if $T_n > t$,

(b) $T_{N(t)} \le t < T_{N(t)+1}$.

If $\mathbb{E}(X_1) < \infty$, then $\mathbb{E}(T_n) < \infty$, so that $\mathbb{P}(T_n > t) \to 0$ as $t \to \infty$. Therefore, by (a),

$$\mathbb{P}(N(t) < n) = \mathbb{P}(T_n > t) \to 0 \quad \text{as } t \to \infty,$$

implying that $N(t) \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$.

Secondly, by (b),

$$\frac{T_{N(t)}}{N(t)} \le \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)+1} \cdot (1 + N(t)^{-1}).$$

Take the limit as $t \to \infty$, using the fact that $T_n/n \xrightarrow{\text{a.s.}} \mathbb{E}(X_1)$ by the strong law, to deduce that $t/N(t) \xrightarrow{\text{a.s.}} \mathbb{E}(X_1)$.

3. By the strong law, $S_n/n \xrightarrow{\text{a.s.}} \mathbb{E}(X_1) \ne 0$. In particular, with probability 1, $S_n = 0$ only finitely often.

7.6 Solution. The law of the iterated logarithm

1. The sum S_n is approximately $N(0, n)$, so that

$$\mathbb{P}\big(S_n > \sqrt{\alpha n \log n}\big) = 1 - \Phi\big(\sqrt{\alpha \log n}\big) < \frac{n^{-\frac{1}{2}\alpha}}{\sqrt{\alpha \log n}}$$

for all large n, by the tail estimate of Exercise (4.4.8) or Problem (4.14.1c) for the normal distribution. This is summable if $\alpha > 2$, and the claim follows by an application of the first Borel–Cantelli lemma.

7.7 Solutions. Martingales

1. Suppose $i < j$. Then

$$\begin{aligned}\mathbb{E}(X_j X_i) &= \mathbb{E}\Big\{\mathbb{E}\big[(S_j - S_{j-1})X_i \mid S_0, S_1, \dots, S_{j-1}\big]\Big\} \\ &= \mathbb{E}\Big\{X_i\big[\mathbb{E}(S_j \mid S_0, S_1, \dots, S_{j-1}) - S_{j-1}\big]\Big\} = 0\end{aligned}$$

by the martingale property.

2. Clearly $\mathbb{E}|S_n| < \infty$ for all n. Also, for $n \geq 0$,

$$\mathbb{E}(S_{n+1} \mid Z_0, Z_1, \dots, Z_n) = \frac{1}{\mu^{n+1}}\left\{\mathbb{E}(Z_{n+1} \mid Z_0, \dots, Z_n) - m\left(\frac{1-\mu^{n+1}}{1-\mu}\right)\right\}$$
$$= \frac{1}{\mu^{n+1}}\left\{m + \mu Z_n - m\left(\frac{1-\mu^{n+1}}{1-\mu}\right)\right\} = S_n.$$

3. Certainly $\mathbb{E}|S_n| < \infty$ for all n. Secondly, for $n \geq 1$,

$$\mathbb{E}(S_{n+1} \mid X_0, X_1, \dots, X_n) = \alpha\mathbb{E}(X_{n+1} \mid X_0, \dots, X_n) + X_n$$
$$= (\alpha a + 1)X_n + \alpha b X_{n-1},$$

which equals S_n if $\alpha = (1-a)^{-1}$.

4. The gambler stakes $Z_i = f_{i-1}(X_1, \dots, X_{i-1})$ on the ith play, at a return of X_i per unit. Therefore $S_i = S_{i-1} + X_i Z_i$ for $i \geq 2$, with $S_1 = X_1 Y$. Secondly,

$$\mathbb{E}(S_{n+1} - S_n \mid X_1, \dots, X_n) = Z_{n+1}\mathbb{E}(X_{n+1} \mid X_1, \dots, X_n) = 0,$$

where we have used the fact that Z_{n+1} depends only on $X_1, X_2, \dots, X_n$.

7.8 Solutions. Martingale convergence theorem

1. It is easily checked that S_n defines a martingale with respect to itself, and the claim follows from the Doob–Kolmogorov inequality, using the fact that

$$\mathbb{E}(S_n^2) = \sum_{j=1}^{n} \operatorname{var}(X_j).$$

2. It would be easy but somewhat perverse to use the martingale convergence theorem, and so we give a direct proof based on Kolmogorov's inequality of Exercise (7.8.1). Applying this inequality to the sequence $Z_m, Z_{m+1}, \dots$ where $Z_i = (X_i - \mathbb{E}X_i)/i$, we obtain that $S_n = Z_1 + Z_2 + \dots + Z_n$ satisfies, for $\epsilon > 0$,

$$\mathbb{P}\left(\max_{m \leq n \leq r} |S_n - S_m| > \epsilon\right) \leq \frac{1}{\epsilon^2}\sum_{n=m+1}^{r} \operatorname{var}(Z_n).$$

We take the limit as $r \to \infty$, using the continuity of $\mathbb{P}$, to obtain

$$\mathbb{P}\left(\sup_{n \geq m} |S_n - S_m| > \epsilon\right) \leq \frac{1}{\epsilon^2}\sum_{n=m+1}^{\infty} \frac{1}{n^2}\operatorname{var}(X_n).$$

Now let $m \to \infty$ to obtain (after a small step)

$$\mathbb{P}\left(\lim_{m\to\infty}\sup_{n \geq m} |S_n - S_m| \leq \epsilon\right) = 1 \qquad \text{for all } \epsilon > 0.$$

Any real sequence (x_n) satisfying

$$\lim_{m\to\infty}\sup_{n \geq m} |x_n - x_m| \leq \epsilon \qquad \text{for all } \epsilon > 0,$$

is Cauchy convergent, and hence convergent. It follows that S_n converges a.s. to some limit Y.

The last part is an immediate consequence, using Kronecker's lemma.

3. By the martingale convergence theorem, $S = \lim_{n\to\infty} S_n$ exists a.s., and $S_n \xrightarrow{\text{m.s.}} S$. Using Exercise (7.2.1c), $\text{var}(S_n) \to \text{var}(S)$, and therefore $\text{var}(S) = 0$.

7.9 Solutions. Prediction and conditional expectation

1. (a) Clearly the best predictors are $\mathbb{E}(X \mid Y) = Y^2$, $\mathbb{E}(Y \mid X) = 0$.
(b) We have, after expansion, that

$$\mathbb{E}\{(X - aY - b)^2\} = \text{var}(Y^2) + a^2\mathbb{E}(Y^2) + \{b - \mathbb{E}(Y^2)\}^2,$$

since $\mathbb{E}(Y) = \mathbb{E}(Y^3) = 0$. This is a minimum when $b = \mathbb{E}(Y^2) = \frac{1}{3}$, and $a = 0$. The best linear predictor of X given Y is therefore $\frac{1}{3}$.

Note that $\mathbb{E}(Y \mid X) = 0$ *is* a linear function of X; it is therefore the best linear predictor of Y given X.

2. By the result of Problem (4.14.13), $\mathbb{E}(Y \mid X) = \mu_2 + \rho\sigma_2(X - \mu_1)/\sigma_1$, in the natural notation.

3. Write

$$g(\mathbf{a}) = \sum_{i=1}^{n} a_i X_i = \mathbf{a}\mathbf{X}',$$

and

$$v(\mathbf{a}) = \mathbb{E}\{(Y - g(\mathbf{a}))^2\} = \mathbb{E}(Y^2) - 2\mathbf{a}\mathbb{E}(Y\mathbf{X}') + \mathbf{a}\mathbf{V}\mathbf{a}'.$$

Let $\widehat{\mathbf{a}}$ be a vector satisfying $\mathbf{V}\widehat{\mathbf{a}}' = \mathbb{E}(Y\mathbf{X}')$. Then

$$\begin{aligned} v(\mathbf{a}) - v(\widehat{\mathbf{a}}) &= \mathbf{a}\mathbf{V}\mathbf{a}' - 2\mathbf{a}\mathbb{E}(Y\mathbf{X}') + 2\widehat{\mathbf{a}}\mathbb{E}(Y\mathbf{X}') - \widehat{\mathbf{a}}\mathbf{V}\widehat{\mathbf{a}}' \\ &= \mathbf{a}\mathbf{V}\mathbf{a}' - 2\mathbf{a}\mathbf{V}\widehat{\mathbf{a}}' + \widehat{\mathbf{a}}\mathbf{V}\widehat{\mathbf{a}}' = (\mathbf{a} - \widehat{\mathbf{a}})\mathbf{V}(\mathbf{a} - \widehat{\mathbf{a}})' \geq 0, \end{aligned}$$

since $\mathbf{V}$ is non-negative definite. Hence $v(\mathbf{a})$ is a minimum when $\mathbf{a} = \widehat{\mathbf{a}}$, and the answer is $g(\widehat{\mathbf{a}})$. If $\mathbf{V}$ is non-singular, $\widehat{\mathbf{a}} = \mathbb{E}(Y\mathbf{X})\mathbf{V}^{-1}$.

4. Recall that $Z = \mathbb{E}(Y \mid \mathcal{G})$ is the ('almost') unique $\mathcal{G}$-measurable random variable with finite mean and satisfying $\mathbb{E}\{(Y - Z)I_G\} = 0$ for all $G \in \mathcal{G}$.
(i) $\Omega \in \mathcal{G}$, and hence $\mathbb{E}\{\mathbb{E}(Y \mid \mathcal{G})I_\Omega\} = \mathbb{E}(ZI_\Omega) = \mathbb{E}(YI_\Omega)$.
(ii) $U = \alpha\mathbb{E}(Y \mid \mathcal{G}) + \beta\mathbb{E}(Z \mid \mathcal{G})$ satisfies

$$\begin{aligned} \mathbb{E}(UI_G) &= \alpha\mathbb{E}\{\mathbb{E}(Y \mid \mathcal{G})I_G\} + \beta\mathbb{E}\{\mathbb{E}(Z \mid \mathcal{G})I_G\} \\ &= \alpha\mathbb{E}(YI_G) + \beta\mathbb{E}(ZI_G) = \mathbb{E}\{(\alpha Y + \beta Z)I_G\}, \qquad G \in \mathcal{G}. \end{aligned}$$

Also, U is $\mathcal{G}$-measurable.
(iii) Suppose there exists m (> 0) such that $G = \{\mathbb{E}(Y \mid \mathcal{G}) < -m\}$ has strictly positive probability. Then $G \in \mathcal{G}$, and so $\mathbb{E}(YI_G) = \mathbb{E}\{\mathbb{E}(Y \mid \mathcal{G})I_G\}$. However $YI_G \geq 0$, whereas $\mathbb{E}(Y \mid \mathcal{G})I_G < -m$. We obtain a contradiction on taking expectations.
(iv) Just check the definition of conditional expectation.
(v) If Y is independent of $\mathcal{G}$, then $\mathbb{E}(YI_G) = \mathbb{E}(Y)\mathbb{P}(G)$ for $G \in \mathcal{G}$. Hence $\mathbb{E}\{(Y - \mathbb{E}(Y))I_G\} = 0$ for $G \in \mathcal{G}$, as required.

(vi) If g is convex then, for all $a \in \mathbb{R}$, there exists $\lambda(a)$ such that

$$g(y) \geq g(a) + (y - a)\lambda(a);$$

furthermore λ may be chosen to be a measurable function of a. Set $a = \mathbb{E}(Y \mid \mathcal{G})$ and $y = Y$, to obtain

$$g(Y) \geq g\{\mathbb{E}(Y \mid \mathcal{G})\} + \{Y - \mathbb{E}(Y \mid \mathcal{G})\}\lambda\{\mathbb{E}(Y \mid \mathcal{G})\}.$$

Take expectations conditional on $\mathcal{G}$, and use the fact that $\mathbb{E}(Y \mid \mathcal{G})$ is $\mathcal{G}$-measurable.

(vii) We have that

$$\big|\mathbb{E}(Y_n \mid \mathcal{G}) - \mathbb{E}(Y \mid \mathcal{G})\big| \leq \mathbb{E}\{|Y_n - Y| \mid \mathcal{G}\} \leq V_n$$

where $V_n = \mathbb{E}\{\sup_{m \geq n} |Y_m - Y| \mid \mathcal{G}\}$. Now $\{V_n : n \geq 1\}$ is non-increasing and bounded below. Hence $V = \lim_{n\to\infty} V_n$ exists and satisfies $V \geq 0$. Also

$$\mathbb{E}(V) \leq \mathbb{E}(V_n) = \mathbb{E}\left\{\sup_{m \geq n} |Y_m - Y|\right\},$$

which tends to 0 as $m \to \infty$, by the dominated convergence theorem. Therefore $\mathbb{E}(V) = 0$, and hence $\mathbb{P}(V = 0) = 1$. The claim follows.

5. $\mathbb{E}(Y \mid X) = X$.

6. (a) Let $\{X_n : n \geq 1\}$ be a sequence of members of H which is Cauchy convergent in mean square, that is, $\mathbb{E}\{|X_n - X_m|^2\} \to 0$ as $m, n \to \infty$. By Chebyshov's inequality, $\{X_n : n \geq 1\}$ is Cauchy convergent in probability, and therefore converges in probability to some limit X (see Exercise (7.3.1)). It follows that there exists a subsequence $\{X_{n_k} : k \geq 1\}$ which converges to X almost surely. Since each X_{n_k} is $\mathcal{G}$-measurable, we may assume that X is $\mathcal{G}$-measurable. Now, as $n \to \infty$,

$$\mathbb{E}\{|X_n - X|^2\} = \mathbb{E}\left\{\liminf_{k\to\infty} |X_n - X_{n_k}|^2\right\} \leq \liminf_{k\to\infty} \mathbb{E}\{|X_n - X_{n_k}|^2\} \to 0,$$

where we have used Fatou's lemma and Cauchy convergence in mean square. Therefore $X_n \xrightarrow{\text{m.s.}} X$. That $\mathbb{E}(X^2) < \infty$ is a consequence of Exercise (7.2.1a).

(b) That (i) implies (ii) is obvious, since $I_G \in H$. Suppose that (ii) holds. Any Z ($\in H$) may be written as the limit, as $n \to \infty$, of random variables of the form

$$Z_n = \sum_{i=1}^{m(n)} a_i(n) I_{G_i(n)}$$

for reals $a_i(n)$ and events $G_i(n)$ in $\mathcal{G}$; furthermore we may assume that $|Z_n| \leq |Z|$. It is easy to see that $\mathbb{E}\{(Y - M)Z_n\} = 0$ for all n. By dominated convergence, $\mathbb{E}\{(Y - M)Z_n\} \to \mathbb{E}\{(Y - M)Z\}$, and the claim follows.

7.10 Solutions. Uniform integrability

1. It is easily checked by considering whether $|x| \leq a$ or $|y| \leq a$ that, for $a > 0$,

$$|x + y| I_{\{|x+y| \geq 2a\}} \leq 2\left(|x| I_{\{|x| \geq a\}} + |y| I_{\{|y| \geq a\}}\right).$$

Now substitute $x = X_n$ and $y = Y_n$, and take expectations.

2. (a) Let $\epsilon > 0$. There exists N such that $\mathbb{E}(|X_n - X|^r) < \epsilon$ if $n > N$. Now $\mathbb{E}|X^r| < \infty$, by Exercise (7.2.1a), and therefore there exists δ (> 0) such that

$$\mathbb{E}(|X|^r I_A) < \epsilon, \quad \mathbb{E}(|X_n|^r I_A) < \epsilon \quad \text{for } 1 \le n \le N,$$

for all events A such that $\mathbb{P}(A) < \delta$. By Minkowski's inequality,

$$\left\{\mathbb{E}(|X_n|^r I_A)\right\}^{1/r} \le \left\{\mathbb{E}(|X_n - X|^r I_A)\right\}^{1/r} + \left\{\mathbb{E}(|X|^r I_A)\right\}^{1/r} \le 2\epsilon^{1/r} \qquad \text{if } n > N$$

if $\mathbb{P}(A) < \delta$. Therefore $\{|X_n|^r : n \ge 1\}$ is uniformly integrable.

If r is an integer then $\{X_n^r : n \ge 1\}$ is uniformly integrable also. Also $X_n^r \xrightarrow{\text{P}} X^r$ since $X_n \xrightarrow{\text{P}} X$ (use the result of Problem (7.11.3)). Therefore $\mathbb{E}(X_n^r) \to \mathbb{E}(X^r)$ as required.

(b) Suppose now that the collection $\{|X_n|^r : n \ge 1\}$ is uniformly integrable and $X_n \xrightarrow{\text{P}} X$. We show first that $\mathbb{E}|X^r| < \infty$, as follows. There exists a subsequence $\{X_{n_k} : k \ge 1\}$ which converges to X almost surely. By Fatou's lemma,

$$\mathbb{E}|X^r| = \mathbb{E}\left(\liminf_{k\to\infty} |X_{n_k}|^r\right) \le \liminf_{k\to\infty} \mathbb{E}|X_{n_k}^r| \le \sup_n \mathbb{E}|X_n^r| < \infty.$$

If $\epsilon > 0$, there exists δ (> 0) such that

$$\mathbb{E}(|X^r| I_A) < \epsilon, \quad \mathbb{E}(|X_n^r| I_A) < \epsilon \qquad \text{for all } n,$$

whenever A is such that $\mathbb{P}(A) < \delta$. There exists N such that $B_n(\epsilon) = \{|X_n - X| > \epsilon\}$ satisfies $\mathbb{P}(B_n(\epsilon)) < \delta$ for $n > N$. Consequently

$$\mathbb{E}(|X_n - X|^r) \le \epsilon^r + \mathbb{E}\left(|X_n - X|^r I_{B_n(\epsilon)}\right), \qquad n > N,$$

of which the final term satisfies

$$\left\{\mathbb{E}\left(|X_n - X|^r I_{B_n(\epsilon)}\right)\right\}^{1/r} \le \left\{\mathbb{E}\left(|X_n^r| I_{B_n(\epsilon)}\right)\right\}^{1/r} + \left\{\mathbb{E}\left(|X^r| I_{B_n(\epsilon)}\right)\right\}^{1/r} \le 2\epsilon^{1/r}.$$

Therefore, $X_n \xrightarrow{r} X$.

3. Fix $\epsilon > 0$, and find a real number a such that $g(x) > x/\epsilon$ if $x > a$. If $b \ge a$,

$$\mathbb{E}\left(|X_n| I_{\{|X_n| > b\}}\right) < \epsilon \mathbb{E}\{g(|X_n|)\} \le \epsilon \sup_n \mathbb{E}\{g(|X_n|)\},$$

whence the left side approaches 0, uniformly in n, as $b \to \infty$.

4. Here is a quick way. Extinction is (almost) certain for such a branching process, so that $Z_n \xrightarrow{\text{a.s.}} 0$, and hence $Z_n \xrightarrow{\text{P}} 0$. If $\{Z_n : n \ge 0\}$ were uniformly integrable, it would follow that $\mathbb{E}(Z_n) \to 0$ as $n \to \infty$; however $\mathbb{E}(Z_n) = 1$ for all n.

5. We may suppose that X_n, Y_n, and Z_n have finite means, for all n. We have that $0 \le Y_n - X_n \le Z_n - X_n$ where, by Theorem (7.3.9c), $Z_n - X_n \xrightarrow{\text{P}} Z - X$. Also

$$\mathbb{E}|Z_n - X_n| = \mathbb{E}(Z_n - X_n) \to \mathbb{E}(Z - X) = \mathbb{E}|Z - X|,$$

so that $\{Z_n - X_n : n \ge 1\}$ is uniformly integrable, by Theorem (7.10.3). It follows that $\{Y_n - X_n : n \ge 1\}$ is uniformly integrable. Also $Y_n - X_n \xrightarrow{\text{P}} Y - X$, and therefore by Theorem (7.10.3), $\mathbb{E}|Y_n - X_n| \to \mathbb{E}|Y - X|$, which is to say that $\mathbb{E}(Y_n) - \mathbb{E}(X_n) \to \mathbb{E}(Y) - \mathbb{E}(X)$; hence $\mathbb{E}(Y_n) \to \mathbb{E}(Y)$.

It is not necessary to use uniform integrability; try doing it using the 'more primitive' Fatou's lemma.

6. For any event A, $\mathbb{E}(|X_n| I_A) \le \mathbb{E}(Z I_A)$ where $Z = \sup_n |X_n|$. The uniform integrability follows by the assumption that $\mathbb{E}(Z) < \infty$.

7.11 Solutions to problems

1. $\mathbb{E}|X_n^r| = \infty$ for $r \geq 1$, so there is no convergence in any mean. However, if $\epsilon > 0$,

$$\mathbb{P}(|X_n| > \epsilon) = 1 - \frac{2}{\pi}\tan^{-1}(n\epsilon) \to 0 \qquad \text{as } n \to \infty,$$

so that $X_n \xrightarrow{\text{P}} 0$.

You have insufficient information to decide whether or not X_n converges almost surely:
(a) Let X be Cauchy, and let $X_n = X/n$. Then X_n has the required density function, and $X_n \xrightarrow{\text{a.s.}} 0$.
(b) Let the X_n be independent with the specified density functions. For $\epsilon > 0$,

$$\mathbb{P}(|X_n| > \epsilon) = \frac{2}{\pi}\sin^{-1}\left(\frac{1}{\sqrt{1+n^2\epsilon^2}}\right) \sim \frac{2}{\pi n\epsilon},$$

so that $\sum_n \mathbb{P}(|X_n| > \epsilon) = \infty$. By the second Borel–Cantelli lemma, $|X_n| > \epsilon$ i.o. with probability one, implying that X_n does not converge a.s. to 0.

2. (i) Assume all the random variables are defined on the same probability space; otherwise it is meaningless to add them together.
(a) Clearly $X_n(\omega) + Y_n(\omega) \to X(\omega) + Y(\omega)$ whenever $X_n(\omega) \to X(\omega)$ and $Y_n(\omega) \to Y(\omega)$. Therefore

$$\big\{X_n + Y_n \nrightarrow X + Y\big\} \subseteq \{X_n \nrightarrow X\} \cup \{Y_n \nrightarrow Y\},$$

a union of events having zero probability.
(b) Use Minkowski's inequality to obtain that

$$\big\{\mathbb{E}\left(|X_n + Y_n - X - Y|^r\right)\big\}^{1/r} \leq \big\{\mathbb{E}(|X_n - X|^r)\big\}^{1/r} + \big\{\mathbb{E}(|Y_n - Y|^r)\big\}^{1/r}.$$

(c) If $\epsilon > 0$, we have that

$$\big\{|X_n + Y_n - X - Y| > \epsilon\big\} \subseteq \big\{|X_n - X| > \tfrac{1}{2}\epsilon\big\} \cup \big\{|Y_n - Y| > \tfrac{1}{2}\epsilon\big\},$$

and the probability of the right side tends to 0 as $n \to \infty$.
(d) If $X_n \xrightarrow{\text{D}} X$ and the X_n are symmetric, then $-X_n \xrightarrow{\text{D}} X$. However $X_n + (-X_n) \xrightarrow{\text{D}} 0$, which generally differs from $2X$ in distribution.
(ii) (e) Almost-sure convergence follows as in (a) above.
(f) The corresponding statement for convergence in rth mean is false in general. Find a random variable Z such that $\mathbb{E}|Z^r| < \infty$ but $\mathbb{E}|Z^{2r}| = \infty$, and define $X_n = Y_n = Z$ for all n.
(g) Suppose $X_n \xrightarrow{\text{P}} X$ and $Y_n \xrightarrow{\text{P}} Y$. Let $\epsilon > 0$. Then

$$\begin{aligned}\mathbb{P}\big(|X_nY_n - XY| > \epsilon\big) &= \mathbb{P}\Big(\big|(X_n - X)(Y_n - Y) + (X_n - X)Y + X(Y_n - Y)\big| > \epsilon\Big)\\ &\leq \mathbb{P}\big(|X_n - X|\cdot|Y_n - Y| > \tfrac{1}{3}\epsilon\big) + \mathbb{P}\big(|X_n - X|\cdot|Y| > \tfrac{1}{3}\epsilon\big)\\ &\qquad + \mathbb{P}\big(|X|\cdot|Y_n - Y| > \tfrac{1}{3}\epsilon\big).\end{aligned}$$

Now, for $\delta > 0$,

$$\mathbb{P}\big(|X_n - X|\cdot|Y| > \tfrac{1}{3}\epsilon\big) \leq \mathbb{P}\big(|X_n - X| > \epsilon/(3\delta)\big) + \mathbb{P}\big(|Y| > \delta\big),$$

which tends to 0 in the limit as $n \to \infty$ and $\delta \to \infty$ in that order. Together with two similar facts, we obtain that $X_n Y_n \xrightarrow{\mathrm{P}} XY$.

(h) The example of (d) above indicates that the corresponding statement is false for convergence in distribution.

3. Let $\epsilon > 0$. We may pick M such that $\mathbb{P}(|X| \geq M) \leq \epsilon$. The continuous function g is uniformly continuous on the bounded interval $[-M, M]$. There exists $\delta > 0$ such that

$$|g(x) - g(y)| \leq \epsilon \quad \text{if } |x - y| \leq \delta \text{ and } |x| \leq M.$$

If $|g(X_n) - g(X)| > \epsilon$, then either $|X_n - X| > \delta$ or $|X| \geq M$. Therefore

$$\mathbb{P}\big(|g(X_n) - g(X)| > \epsilon\big) \leq \mathbb{P}\big(|X_n - X| > \delta\big) + \mathbb{P}\big(|X| \geq M\big) \to \mathbb{P}\big(|X| \geq M\big) \leq \epsilon,$$

in the limit as $n \to \infty$. It follows that $g(X_n) \xrightarrow{\mathrm{P}} g(X)$.

4. Clearly

$$\begin{aligned}\mathbb{E}(e^{itX_n}) &= \prod_{j=1}^{n} \mathbb{E}(e^{itY_j/10^j}) = \prod_{j=1}^{n} \left\{ \frac{1}{10} \cdot \frac{1 - e^{it/10^{j-1}}}{1 - e^{it/10^j}} \right\} \\ &= \frac{1 - e^{it}}{10^n(1 - e^{it/10^n})} \to \frac{1 - e^{it}}{it}\end{aligned}$$

as $n \to \infty$. The limit is the characteristic function of the uniform distribution on $[0, 1]$.

Now $X_n \leq X_{n+1} \leq 1$ for all n, so that $Y(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists for all ω. Therefore $X_n \xrightarrow{\text{a.s.}} Y$; hence $X_n \xrightarrow{\mathrm{D}} Y$, whence Y has the uniform distribution.

5. (a) Suppose $s < t$. Then

$$\mathbb{E}\big(N(s)N(t)\big) = \mathbb{E}(N(s)^2) + \mathbb{E}\{N(s)(N(t) - N(s))\} = \mathbb{E}(N(s)^2) + \mathbb{E}(N(s))\mathbb{E}\big(N(t) - N(s)\big),$$

since N has independent increments. Therefore

$$\begin{aligned}\operatorname{cov}(N(s), N(t)) &= \mathbb{E}\big(N(s)N(t)\big) - \mathbb{E}(N(s))\mathbb{E}(N(t)) \\ &= (\lambda s)^2 + \lambda s + \lambda s\{\lambda(t - s)\} - (\lambda s)(\lambda t) = \lambda s.\end{aligned}$$

In general, $\operatorname{cov}(N(s), N(t)) = \lambda \min\{s, t\}$.

(b) $N(t + h) - N(t)$ has the same distribution as $N(h)$, if $h > 0$. Hence

$$\mathbb{E}\left(\{N(t + h) - N(t)\}^2\right) = \mathbb{E}(N(h)^2) = (\lambda h)^2 + \lambda h$$

which tends to 0 as $h \to 0$.

(c) By Markov's inequality,

$$\mathbb{P}\big(|N(t + h) - N(t)| > \epsilon\big) \leq \frac{1}{\epsilon^2}\mathbb{E}\left(\{N(t + h) - N(t)\}^2\right),$$

which tends to 0 as $h \to 0$, if $\epsilon > 0$.

(d) Let $\epsilon > 0$. For $0 < h < \epsilon^{-1}$,

$$\mathbb{P}\left(\left|\frac{N(t + h) - N(t)}{h}\right| > \epsilon\right) = \mathbb{P}\big(N(t + h) - N(t) \geq 1\big) = \lambda h + \mathrm{o}(h),$$

which tends to 0 as $h \to 0$.

On the other hand,

$$\mathbb{E}\left(\left\{\frac{N(t+h)-N(t)}{h}\right\}^2\right) = \frac{1}{h^2}\{(\lambda h)^2 + \lambda h\}$$

which tends to ∞ as $h \downarrow 0$.

6. By Markov's inequality, $S_n = \sum_{i=1}^n X_i$ satisfies

$$\mathbb{P}(|S_n| > n\epsilon) \le \frac{\mathbb{E}(S_n^4)}{(n\epsilon)^4}$$

for $\epsilon > 0$. Using the properties of the X's,

$$\begin{aligned}\mathbb{E}(S_n^4) &= n\mathbb{E}(X_1^4) + 4\binom{n}{2}\mathbb{E}(X_1^3X_2) + \binom{4}{2}\binom{n}{2}\mathbb{E}(X_1^2X_2^2) \\ &\quad + 3\binom{4}{2}\binom{n}{3}\mathbb{E}(X_1^2X_2X_3) + 4!\binom{n}{4}\mathbb{E}(X_1X_2X_3X_4) \\ &= n\mathbb{E}(X_1^4) + \binom{4}{2}\binom{n}{2}\mathbb{E}(X_1^2X_2^2),\end{aligned}$$

since $\mathbb{E}(X_i) = 0$ for all i. Therefore there exists a constant C such that

$$\sum_n \mathbb{P}(|n^{-1}S_n| > \epsilon) \le \sum_n \frac{C}{n^2} < \infty,$$

implying (via the first Borel–Cantelli lemma) that $n^{-1}S_n \xrightarrow{\text{a.s.}} 0$.

7. We have by Markov's inequality that

$$\sum_n \mathbb{P}(|X_n - X| > \epsilon) \le \sum_n \frac{\mathbb{E}\{|X_n - X|^r\}}{\epsilon^r} < \infty$$

for $\epsilon > 0$, so that $X_n \xrightarrow{\text{a.s.}} X$ (via the first Borel–Cantelli lemma).

8. Either use the Skorokhod representation or characteristic functions. Following the latter route, the characteristic function of $aX_n + b$ is

$$\mathbb{E}(e^{it(aX_n+b)}) = e^{itb}\phi_n(at) \to e^{itb}\phi_X(at) = \mathbb{E}(e^{it(aX+b)})$$

where ϕ_n is the characteristic function of X_n. The result follows by the continuity theorem.

9. For any positive reals c, t,

$$\mathbb{P}(X \ge t) = \mathbb{P}(X + c \ge t + c) \le \frac{\mathbb{E}\{(X+c)^2\}}{(t+c)^2}.$$

Set $c = \sigma^2/t$ to obtain the required inequality.

10. Note that $g(u) = u/(1+u)$ is an increasing function on $[0, \infty)$. Therefore, for $\epsilon > 0$,

$$\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}\left(\frac{|X_n|}{1+|X_n|} > \frac{\epsilon}{1+\epsilon}\right) \le \frac{1+\epsilon}{\epsilon} \cdot \mathbb{E}\left(\frac{|X_n|}{1+|X_n|}\right)$$

by Markov's inequality. If this expectation tends to 0 then $X_n \xrightarrow{\mathrm{P}} 0$.

Suppose conversely that $X_n \xrightarrow{\mathrm{P}} 0$. Then

$$\mathbb{E}\left(\frac{|X_n|}{1+|X_n|}\right) \le \frac{\epsilon}{1+\epsilon} \cdot \mathbb{P}(|X_n| \le \epsilon) + 1 \cdot \mathbb{P}(|X_n| > \epsilon) \to \frac{\epsilon}{1+\epsilon}$$

as $n \to \infty$, for $\epsilon > 0$. However ϵ is arbitrary, and hence the expectation has limit 0.

11. (i) The argument of the solution to Exercise (7.9.6a) shows that $\{X_n\}$ converges in mean square if it is mean-square Cauchy convergent. Conversely, suppose that $X_n \xrightarrow{\text{m.s.}} X$. By Minkowski's inequality,

$$\left\{\mathbb{E}\left((X_m - X_n)^2\right)\right\}^{1/2} \le \left\{\mathbb{E}\left((X_m - X)^2\right)\right\}^{1/2} + \left\{\mathbb{E}\left((X_n - X)^2\right)\right\}^{1/2} \to 0$$

as $m, n \to \infty$, so that $\{X_n\}$ is mean-square Cauchy convergent.

(ii) The corresponding result is valid for convergence almost surely, in rth mean, and in probability. For a.s. convergence, it is self-evident by the properties of Cauchy-convergent sequences of real numbers. For convergence in probability, see Exercise (7.3.1). For convergence in rth mean ($r \ge 1$), just adapt the argument of (i) above.

12. If $\operatorname{var}(X_i) \le M$ for all i, the variance of $n^{-1}\sum_{i=1}^n X_i$ is

$$\frac{1}{n^2}\sum_{i=1}^n \operatorname{var}(X_i) \le \frac{M}{n} \to 0 \quad \text{as } n \to \infty.$$

13. (a) We have that

$$\mathbb{P}(M_n \le a_n x) = F(a_n x)^n \to H(x) \quad \text{as } n \to \infty.$$

If $x \le 0$ then $F(a_n x)^n \to 0$, so that $H(x) = 0$. Suppose that $x > 0$. Then

$$-\log H(x) = -\lim_{n\to\infty}\left\{n \log\left[1 - (1 - F(a_n x))\right]\right\} = \lim_{n\to\infty}\left\{n(1 - F(a_n x))\right\}$$

since $-y^{-1}\log(1-y) \to 1$ as $y \downarrow 0$. Setting $x = 1$, we obtain $n(1 - F(a_n)) \to -\log H(1)$, and the second limit follows.

(b) This is immediate from the fact that it is valid for all sequences $\{a_n\}$.

(c) We have that

$$\frac{1 - F(te^{x+y})}{1 - F(t)} = \frac{1 - F(te^{x+y})}{1 - F(te^x)} \cdot \frac{1 - F(te^x)}{1 - F(t)} \to \frac{\log H(e^y)}{\log H(1)} \cdot \frac{\log H(e^x)}{\log H(1)}$$

as $t \to \infty$. Therefore $g(x + y) = g(x)g(y)$. Now g is non-increasing with $g(0) = 1$. Therefore $g(x) = e^{-\beta x}$ for some β, and hence $H(u) = \exp(-\alpha u^{-\beta})$ for $u > 0$, where $\alpha = -\log H(1)$.

14. Either use the result of Problem (7.11.13) or do the calculations directly thus. We have that

$$\mathbb{P}\left(M_n \le xn/\pi\right) = \left\{\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{xn}{\pi}\right)\right\}^n = \left\{1 - \frac{1}{\pi}\tan^{-1}\left(\frac{\pi}{xn}\right)\right\}^n$$

if $x > 0$, by elementary trigonometry. Now $\tan^{-1} y = y + \mathrm{o}(y)$ as $y \to 0$, and therefore

$$\mathbb{P}\left(M_n \le xn/\pi\right) = \left(1 - \frac{1}{xn} + \mathrm{o}(n^{-1})\right)^n \to e^{-1/x} \qquad \text{as } n \to \infty.$$

15. The characteristic function of the average satisfies

$$\phi(t/n)^n = \left(1 + \frac{i\mu t}{n} + \mathrm{o}(n^{-1})\right)^n \to e^{i\mu t} \qquad \text{as } n \to \infty.$$

By the continuity theorem, the average converges in distribution to the *constant* μ, and hence in probability also.

16. (a) With $u_n = u(x_n)$, we have that

$$\left|\mathbb{E}(u(X)) - \mathbb{E}(u(Y))\right| \le \sum_n |u_n| \cdot |f_n - g_n| \le \sum_n |f_n - g_n|$$

if $\|u\|_\infty \le 1$. There is equality if u_n equals the sign of $f_n - g_n$. The second equality holds as in Problem (2.7.13) and Exercise (4.12.3).

(b) Similarly, if $\|u\|_\infty \le 1$,

$$\left|\mathbb{E}(u(X)) - \mathbb{E}(u(Y))\right| \le \int_{-\infty}^{\infty} |u(x)| \cdot |f(x) - g(x)|\,dx \le \int_{-\infty}^{\infty} |f(x) - g(x)|\,dx$$

with equality if $u(x)$ is the sign of $f(x) - g(x)$. Secondly, we have that

$$\left|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\right| = \tfrac{1}{2}\left|\mathbb{E}(u(X)) - \mathbb{E}(u(Y))\right| \le \tfrac{1}{2} d_{\mathrm{TV}}(X, Y),$$

where

$$u(x) = \begin{cases} 1 & \text{if } x \in A, \\ -1 & \text{if } x \notin A. \end{cases}$$

Equality holds when $A = \{x \in \mathbb{R} : f(x) \ge g(x)\}$.

(c) Suppose $d_{\mathrm{TV}}(X_n, X) \to 0$. Fix $a \in \mathbb{R}$, and let u be the indicator function of the interval $(-\infty, a]$. Then $|\mathbb{E}(u(X_n)) - \mathbb{E}(u(X))| = |\mathbb{P}(X_n \le a) - \mathbb{P}(X \le a)|$, and the claim follows.

On the other hand, if $X_n = n^{-1}$ with probability one, then $X_n \xrightarrow{\mathrm{D}} 0$. However, by part (a), $d_{\mathrm{TV}}(X_n, 0) = 2$ for all n.

(d) This is tricky without a knowledge of Radon–Nikodým derivatives, and we therefore restrict ourselves to the case when X and Y are discrete. (The continuous case is analogous.) As in the solution to Exercise (4.12.4), $\mathbb{P}(X \ne Y) \ge \frac{1}{2} d_{\mathrm{TV}}(X, Y)$. That equality is possible was proved for Exercise (4.12.5), and we rephrase that solution here. Let $\mu_n = \min\{f_n, g_n\}$ and $\mu = \sum_n \mu_n$, and note that

$$d_{\mathrm{TV}}(X, Y) = \sum_n |f_n - g_n| = \sum_n \{f_n + g_n - 2\mu_n\} = 2(1 - \mu).$$

It is easy to see that

$$\tfrac{1}{2} d_{\mathrm{TV}}(X, Y) = \mathbb{P}(X \ne Y) = \begin{cases} 1 & \text{if } \mu = 0, \\ 0 & \text{if } \mu = 1, \end{cases}$$

and therefore we may assume that $0 < \mu < 1$. Let U, V, W be random variables with mass functions

$$\mathbb{P}(U = x_n) = \frac{\mu_n}{\mu}, \quad \mathbb{P}(V = x_n) = \frac{\max\{f_n - g_n, 0\}}{1 - \mu}, \quad \mathbb{P}(W = x_n) = \frac{-\min\{f_n - g_n, 0\}}{1 - \mu},$$

and let Z be a Bernoulli variable with parameter μ, independent of (U, V, W). We now choose the pair X', Y' by

$$(X', Y') = \begin{cases} (U, U) & \text{if } Z = 1, \\ (V, W) & \text{if } Z = 0. \end{cases}$$

It may be checked that X' and Y' have the same distributions as X and Y, and furthermore, $\mathbb{P}(X' \neq Y') = \mathbb{P}(Z = 0) = 1 - \mu = \frac{1}{2}d_{\mathrm{TV}}(X, Y)$.
(e) By part (d), we may find independent pairs (X_i', Y_i'), $1 \le i \le n$, having the same marginals as (X_i, Y_i), respectively, and such that $\mathbb{P}(X_i' \neq Y_i') = \frac{1}{2}d_{\mathrm{TV}}(X_i, Y_i)$. Now,

$$\begin{aligned} d_{\mathrm{TV}}\Big(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\Big) &= d_{\mathrm{TV}}\Big(\sum_{i=1}^n X_i', \sum_{i=1}^n Y_i'\Big) \\ &\le 2\mathbb{P}\Big(\sum_{i=1}^n X_i' \neq \sum_{i=1}^n Y_i'\Big) \le 2\sum_{i=1}^n \mathbb{P}(X_i' \neq Y_i') = 2\sum_{i=1}^n d_{\mathrm{TV}}(X_i, Y_i). \end{aligned}$$

17. If $X_1, X_2, \dots$ are independent variables having the Poisson distribution with parameter λ, then $S_n = X_1 + X_2 + \cdots + X_n$ has the Poisson distribution with parameter λn. Now $n^{-1}S_n \xrightarrow{\mathrm{D}} \lambda$, so that $\mathbb{E}(g(n^{-1}S_n)) \to g(\lambda)$ for all bounded continuous g. The result follows.

18. The characteristic function ψ_{mn} of

$$U_{mn} = \frac{(X_n - n) - (Y_m - m)}{\sqrt{m+n}}$$

satisfies

$$\log \psi_{mn}(t) = n\big(e^{it/\sqrt{m+n}} - 1\big) + m\big(e^{-it/\sqrt{m+n}} - 1\big) + \frac{(m-n)it}{\sqrt{m+n}} \to -\tfrac{1}{2}t^2$$

as $m, n \to \infty$, implying by the continuity theorem that $U_{mn} \xrightarrow{\mathrm{D}} N(0, 1)$. Now $X_n + Y_m$ is Poisson-distributed with parameter $m + n$, and therefore

$$V_{mn} = \sqrt{\frac{X_n + Y_m}{m+n}} \xrightarrow{\mathrm{P}} 1 \qquad \text{as } m, n \to \infty$$

by the law of large numbers and Problem (3). It follows by Slutsky's theorem (7.2.5a) that $U_{mn} / V_{mn} \xrightarrow{\mathrm{D}} N(0, 1)$ as required.

19. (a) The characteristic function of X_n is $\phi_n(t) = \exp\{i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\}$ where μ_n and σ_n^2 are the mean and variance of X_n. Now, $\lim_{n\to\infty} \phi_n(1)$ exists. However $\phi_n(1)$ has modulus $e^{-\frac{1}{2}\sigma_n^2}$, and therefore $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$ exists. The remaining component $e^{i\mu_n t}$ of $\phi_n(t)$ converges as $n \to \infty$, say $e^{i\mu_n t} \to \theta(t)$ as $n \to \infty$ where $\theta(t)$ lies on the unit circle of the complex plane. Now $\phi_n(t) \to \theta(t)e^{-\frac{1}{2}\sigma^2 t^2}$, which is required to be a characteristic function; therefore θ is a continuous function of t. Of the various ways of showing that $\theta(t) = e^{i\mu t}$ for some μ, here is one. The sequence $\psi_n(t) = e^{i\mu_n t}$ is a sequence of characteristic functions whose limit $\theta(t)$ is continuous at $t = 0$. Therefore θ is a characteristic function. However ψ_n is the characteristic function of the constant μ_n, which must converge in distribution as $n \to \infty$; it follows that the real sequence $\{\mu_n\}$ converges to some limit μ, and $\theta(t) = e^{i\mu t}$ as required.

This proves that $\phi_n(t) \to \exp\{i\mu t - \frac{1}{2}\sigma^2 t^2\}$, and therefore the limit X is $N(\mu, \sigma^2)$.
(b) Each linear combination $sX_n + tY_n$ converges in probability, and hence in distribution, to $sX + tY$. Now $sX_n + tY_n$ has a normal distribution, implying by part (a) that $sX + tY$ is normal. Therefore the joint characteristic function of X and Y satisfies

$$\begin{aligned} \phi_{X,Y}(s,t) &= \phi_{sX+tY}(1) = \exp\big\{i\mathbb{E}(sX + tY) - \tfrac{1}{2}\operatorname{var}(sX + tY)\big\} \\ &= \exp\Big\{i(s\mu_X + t\mu_Y) - \tfrac{1}{2}(s^2\sigma_X^2 + 2st\rho_{XY}\sigma_X\sigma_Y + t^2\sigma_Y^2)\Big\} \end{aligned}$$

in the natural notation. Viewed as a function of s and t, this is the joint characteristic function of a bivariate normal distribution.

When working in such a context, the technique of using linear combinations of X_n and Y_n is sometimes called the 'Cramér–Wold device'.

20. (i) Write $Y_i = X_i - \mathbb{E}(X_i)$ and $T_n = \sum_{i=1}^n Y_i$. It suffices to show that $n^{-1}T_n \xrightarrow{\text{m.s.}} 0$. Now, as $n \to \infty$,

$$\mathbb{E}(T_n^2/n^2) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{var}(X_i) + \frac{2}{n^2}\sum_{1\le i<j\le n} \operatorname{cov}(X_i, X_j) \le \frac{nc}{n^2} \to 0.$$

(ii) Let $\epsilon > 0$. There exists I such that $|\rho(X_i, X_j)| \le \epsilon$ if $|i-j| \ge I$. Now

$$\sum_{i,j=1}^n \operatorname{cov}(X_i, X_j) \le \sum_{\substack{|i-j|\le I\\ 1\le i,j\le n}} \operatorname{cov}(X_i, X_j) + \sum_{\substack{|i-j|>I\\ 1\le i,j\le n}} \operatorname{cov}(X_i, X_j) \le 2nIc + n^2\epsilon c,$$

since $\operatorname{cov}(X_i, X_j) \le |\rho(X_i, X_j)|\sqrt{\operatorname{var}(X_i)\cdot\operatorname{var}(X_j)}$. Therefore,

$$\mathbb{E}(T_n^2/n^2) \le \frac{2Ic}{n} + \epsilon c \to \epsilon c \qquad \text{as } n \to \infty.$$

This is valid for all positive ϵ, and the result follows.

21. The integral

$$\int_2^\infty \frac{c}{x\log|x|}\,dx$$

diverges, and therefore $\mathbb{E}(X_1)$ does not exist.

The characteristic function ϕ of X_1 may be expressed as

$$\phi(t) = 2c\int_2^\infty \frac{\cos(tx)}{x^2\log x}\,dx$$

whence

$$\frac{\phi(t)-\phi(0)}{2c} = -\int_2^\infty \frac{1-\cos(tx)}{x^2\log x}\,dx.$$

Now $0 \le 1-\cos\theta \le \min\{2, \theta^2\}$, and therefore

$$\left|\frac{\phi(t)-\phi(0)}{2c}\right| \le \int_2^{1/t} \frac{t^2}{\log x}\,dx + \int_{1/t}^\infty \frac{2}{x^2\log x}\,dx, \qquad \text{if } t > 0.$$

Now

$$\frac{1}{u}\int_2^u \frac{dx}{\log x} \to 0 \qquad \text{as } u \to \infty,$$

and

$$\int_u^\infty \frac{2}{x^2\log x}\,dx \le \frac{1}{\log u}\int_u^\infty \frac{2}{x^2}\,dx = \frac{2}{u\log u}, \qquad u > 1.$$

Therefore

$$\left|\frac{\phi(t)-\phi(0)}{2c}\right| = \mathrm{o}(t) \qquad \text{as } t \downarrow 0.$$

Now ϕ is an even function, and hence $\phi'(0)$ exists and equals 0. Use the result of Problem (7.11.15) to deduce that $n^{-1}\sum_1^n X_i$ converges in distribution to 0, and therefore in probability also, since 0 is constant. The X_i do not obey the strong law since they have no mean.

22. If the two points are **U** and **V** then

$$\mathbb{E}\{(U_1 - V_1)^2\} = \int_0^1 \int_0^1 (u-v)^2\,du\,dv = \tfrac{1}{6},$$

and therefore

$$\frac{1}{n}X_n^2 = \frac{1}{n}\sum_{i=1}^{n}(U_i - V_i)^2 \xrightarrow{\text{P}} \frac{1}{6} \qquad \text{as } n \to \infty,$$

by the independence of the components. It follows that $X_n/\sqrt{n} \xrightarrow{\text{P}} 1/\sqrt{6}$ either by the result of Problem (7.11.3) or by the fact that

$$\left|\frac{X_n^2}{n} - \frac{1}{6}\right| = \left|\frac{X_n}{\sqrt{n}} - \frac{1}{\sqrt{6}}\right| \cdot \left|\frac{X_n}{\sqrt{n}} + \frac{1}{\sqrt{6}}\right| \geq \frac{1}{\sqrt{6}}\left|\frac{X_n}{\sqrt{n}} - \frac{1}{\sqrt{6}}\right|.$$

23. The characteristic function of $Y_j = X_j^{-1}$ is

$$\phi(t) = \tfrac{1}{2}\int_0^1 (e^{it/x} + e^{-it/x})\,dx = \int_0^1 \cos(t/x)\,dx = |t|\int_{|t|}^{\infty} \frac{\cos y}{y^2}\,dy$$

by the substitution $x = |t|/y$. Therefore

$$\phi(t) = 1 - |t|\int_{|t|}^{\infty} \frac{1-\cos y}{y^2}\,dy = 1 - I|t| + \text{o}(|t|) \qquad \text{as } t \to 0,$$

where, integrating by parts,

$$I = \int_0^{\infty} \frac{1-\cos y}{y^2}\,dy = \int_0^{\infty} \frac{\sin u}{u}\,du = \frac{\pi}{2}.$$

It follows that $T_n = n^{-1}\sum_{j=1}^{n} X_j^{-1}$ has characteristic function

$$\phi(t/n)^n = \left(1 - \frac{\pi|t|}{2n} + \text{o}(n^{-1})\right)^n \to e^{-\frac{1}{2}\pi|t|}$$

as $t \to \infty$, whence $2T_n/\pi$ is asymptotically Cauchy-distributed. In particular,

$$\mathbb{P}\big(|2T_n/\pi| > 1\big) \to \frac{2}{\pi}\int_1^{\infty} \frac{du}{1+u^2} = \frac{1}{2} \qquad \text{as } t \to \infty.$$

24. Let m_n be a non-decreasing sequence of integers satisfying $1 \leq m_n < n$, $m_n \to \infty$, and define

$$Y_{nk} = \begin{cases} X_k & \text{if } |X_k| \leq m_n \\ \text{sign}(X_k) & \text{if } |X_k| > m_n, \end{cases}$$

noting that Y_{nk} takes the value ± 1 each with probability $\frac{1}{2}$ whenever $m_n < k \leq n$. Let $Z_n = \sum_{k=1}^{n} Y_{nk}$. Then

$$\mathbb{P}(U_n \neq Z_n) \leq \sum_{k=1}^{n} \mathbb{P}(|X_k| \geq m_n) \leq \sum_{k=m_n}^{n} \frac{1}{k^2} \to 0 \qquad \text{as } n \to \infty,$$

from which it follows that $U_n/\sqrt{n} \xrightarrow{\text{D}} N(0,1)$ if and only if $Z_n/\sqrt{n} \xrightarrow{\text{D}} N(0,1)$. Now

$$Z_n = \sum_{k=1}^{m_n} Y_{nk} + B_{n-m_n}$$

where B_{n-m_n} is the sum of $n - m_n$ independent summands each of which takes the values ± 1, each possibility having probability $\frac{1}{2}$. Furthermore

$$\left| \frac{1}{\sqrt{n}} \sum_{k=1}^{m_n} Y_{nk} \right| \le \frac{m_n^2}{\sqrt{n}}$$

which tends to 0 if m_n is chosen to be $m_n = \lfloor n^{1/5} \rfloor$; with this choice for m_n, we have that $n^{-1} B_{n-m_n} \xrightarrow{\text{D}} N(0,1)$, and the result follows.

Finally,

$$\operatorname{var}(U_n) = \sum_{k=1}^{n} \left(2 - \frac{1}{k^2}\right)$$

so that

$$\operatorname{var}\big(U_n/\sqrt{n}\big) = 2 - \frac{1}{n} \sum_{k=1}^{n} \frac{1}{k^2} \to 2.$$

25. (i) Let ϕ_n and ϕ be the characteristic functions of X_n and X. The characteristic function ψ_k of X_{N_k} is

$$\psi_k(t) = \sum_{j=1}^{\infty} \phi_j(t) \mathbb{P}(N_k = j)$$

whence

$$|\psi_k(t) - \phi(t)| \le \sum_{j=1}^{\infty} |\phi_j(t) - \phi(t)| \mathbb{P}(N_k = j).$$

Let $\epsilon > 0$. We have that $\phi_j(t) \to \phi(t)$ as $j \to \infty$, and hence for any $T > 0$, there exists $J(T)$ such that $|\phi_j(t) - \phi(t)| < \epsilon$ if $j \ge J(T)$ and $|t| \le T$. Finally, there exists $K(T)$ such that $\mathbb{P}(N_k \le J(T)) \le \epsilon$ if $k \ge K(T)$. It follows that

$$|\psi_k(t) - \phi(t)| \le 2\mathbb{P}\big(N_k \le J(T)\big) + \epsilon \mathbb{P}\big(N_k > J(T)\big) \le 3\epsilon$$

if $|t| \le T$ and $k \ge K(T)$; therefore $\psi_k(t) \to \phi(t)$ as $k \to \infty$.

(ii) Let $Y_n = \sup_{m \ge n} |X_m - X|$. For $\epsilon > 0$, $n \ge 1$,

$$\begin{aligned} \mathbb{P}\big(|X_{N_k} - X| > \epsilon\big) &\le \mathbb{P}(N_k \le n) + \mathbb{P}\big(|X_{N_k} - X| > \epsilon, N_k > n\big) \\ &\le \mathbb{P}(N_k \le n) + \mathbb{P}(Y_n > \epsilon) \to \mathbb{P}(Y_n > \epsilon) \qquad \text{as } k \to \infty. \end{aligned}$$

Now take the limit as $n \to \infty$ and use the fact that $Y_n \xrightarrow{\text{a.s.}} 0$.

26. (a) We have that

$$\frac{a(n-k,n)}{a(n+1,n)} = \prod_{i=0}^{k} \left(1 - \frac{i}{n}\right) \le \exp\left(-\sum_{i=0}^{k} \frac{i}{n}\right).$$

(b) The expectation is

$$E_n = \sum_j g\left(\frac{j-n}{\sqrt{n}}\right)\frac{n^j e^{-n}}{j!}$$

where the sum is over all j satisfying $n - M\sqrt{n} \le j \le n$. For such a value of j,

$$g\left(\frac{j-n}{\sqrt{n}}\right)\frac{n^j e^{-n}}{j!} = \frac{e^{-n}}{\sqrt{n}}\left(\frac{n^{j+1}}{j!} - \frac{n^j}{(j-1)!}\right),$$

whence E_n has the form given.

(c) Now g is continuous on the interval $[-M, 0]$, and it follows by the central limit theorem that

$$E_n \to \int_{-M}^{0} g(x)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx = \int_0^M \frac{x}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx = \frac{1-e^{-\frac{1}{2}M^2}}{\sqrt{2\pi}}.$$

Also,

$$E_n \le \frac{e^{-n}}{\sqrt{n}}a(n+1,n) \le E_n + \frac{e^{-n}}{\sqrt{n}}a(n-k,n) \le E_n + \frac{e^{-n-k^2/(2n)}}{\sqrt{n}}a(n+1,n)$$

where $k = \lfloor M\sqrt{n}\rfloor$. Take the limits as $n \to \infty$ and $M \to \infty$ in that order to obtain

$$\frac{1}{\sqrt{2\pi}} \le \lim_{n\to\infty}\left\{\frac{n^{n+\frac{1}{2}}e^{-n}}{n!}\right\} \le \frac{1}{\sqrt{2\pi}}.$$

27. Clearly

$$\mathbb{E}(R_{n+1} \mid R_0, R_1, \dots, R_n) = R_n + \frac{R_n}{n+2}$$

since a red ball is added with probability $R_n/(n+2)$. Hence

$$\mathbb{E}(S_{n+1} \mid R_0, R_1, \dots, R_n) = S_n,$$

and also $0 \le S_n \le 1$. Using the martingale convergence theorem, $S = \lim_{n\to\infty} S_n$ exists almost surely and in mean square.

28. Let $0 < \epsilon < \frac{1}{3}$, and let

$$k(t) = \lfloor \theta t\rfloor,\ m(t) = \lceil (1-\epsilon^3)k(t)\rceil,\ n(t) = \lfloor (1+\epsilon^3)k(t)\rfloor$$

and let $I_{mn}(t)$ be the indicator function of the event $\{m(t) \le M(t) < n(t)\}$. Since $M(t)/t \xrightarrow{\mathrm{P}} \theta$, we may find T such that $\mathbb{E}(I_{mn}(t)) > 1-\epsilon$ for $t \ge T$.

We may approximate $S_{M(t)}$ by the random variable $S_{k(t)}$ as follows. With $A_j = \{|S_j - S_{k(t)}| > \epsilon\sqrt{k(t)}\}$,

$$\begin{aligned}
\mathbb{P}(A_{M(t)}) &\le \mathbb{P}(A_{M(t)}, I_{mn}(t) = 1) + \mathbb{P}(A_{M(t)}, I_{mn}(t) = 0)\\
&\le \mathbb{P}\Big(\bigcup_{j=m(t)}^{k(t)-1} A_j\Big) + \mathbb{P}\Big(\bigcup_{j=k(t)}^{n(t)-1} A_j\Big) + \mathbb{P}(I_{mn}(t) = 0)\\
&\le \frac{\{k(t)-m(t)\}\sigma^2}{\epsilon^2 k(t)} + \frac{\{n(t)-k(t)\}\sigma^2}{\epsilon^2 k(t)} + \epsilon\\
&\le \epsilon(1+2\sigma^2), \qquad \text{if } t \ge T,
\end{aligned}$$

by Kolmogorov's inequality (Exercise (7.8.1) and Problem (7.11.29)). Send $t \to \infty$ to find that

$$D_t = \frac{S_{M(t)} - S_{k(t)}}{\sqrt{k(t)}} \xrightarrow{P} 0 \qquad \text{as } t \to \infty.$$

Now $S_{k(t)}/\sqrt{k(t)} \xrightarrow{D} N(0, \sigma^2)$ as $t \to \infty$, by the usual central limit theorem. Therefore

$$\frac{S_{M(t)}}{\sqrt{k(t)}} = D_t - \frac{S_{k(t)}}{\sqrt{k(t)}} \xrightarrow{D} N(0, \sigma^2),$$

which implies the first claim, since $k(t)/(\theta t) \to 1$ (see Exercise (7.2.7)). The second part follows by Slutsky's theorem (7.2.5a).

29. We have that $S_n = S_k + (S_n - S_k)$, and so, for $n \geq k$,

$$\mathbb{E}(S_n^2 I_{A_k}) = \mathbb{E}(S_k^2 I_{A_k}) + 2\mathbb{E}\{S_k(S_n - S_k)I_{A_k}\} + \mathbb{E}\{(S_n - S_k)^2 I_{A_k}\}.$$

Now $S_k^2 I_{A_k} \geq c^2 I_{A_k}$; the second term on the right side is 0, by the independence of the X's, and the third term is non-negative. The first inequality of the question follows. Summing over k, we obtain $\mathbb{E}(S_n^2) \geq c^2 \mathbb{P}(M_n > c)$ as required.

30. (i) With $S_n = \sum_{i=1}^n X_i$, we have by Kolmogorov's inequality that

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_{m+k} - S_m| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{k=m}^{m+n} \mathbb{E}(X_k^2)$$

for $\epsilon > 0$. Take the limit as $m, n \to \infty$ to obtain in the usual way that $\{S_r : r \geq 0\}$ is a.s. Cauchy convergent, and therefore a.s. convergent, if $\sum_1^\infty \mathbb{E}(X_k^2) < \infty$. It is shorter to use the martingale convergence theorem, noting that S_n is a martingale with uniformly bounded second moments.

(ii) Apply part (i) to the sequence $Y_k = X_k/b_k$ to deduce that $\sum_{k=1}^\infty X_k/b_k$ converges a.s. The claim now follows by Kronecker's lemma (see Exercise (7.8.2)).

31. (a) This is immediate by the observation that

$$e^{\lambda(\mathbf{P})} = f_{X_0} \prod_{i,j} p_{ij}^{N_{ij}}.$$

(b) Clearly $\sum_j p_{ij} = 1$ for each i, and we introduce Lagrange multipliers $\{\mu_i : i \in S\}$ and write $V = \lambda(\mathbf{P}) + \sum_i \mu_i \sum_j p_{ij}$. Differentiating V with respect to each p_{ij} yields a stationary (maximum) value when $(N_{ij}/p_{ij}) + \mu_i = 0$. Hence $\sum_k N_{ik} = -\mu_k$, and

$$\widehat{p}_{ij} = -\frac{N_{ij}}{\mu_i} = \frac{N_{ij}}{\sum_k N_{ik}}.$$

(c) We have that $N_{ij} = \sum_{r=1}^{\sum_k N_{ik}} I_r$ where I_r is the indicator function of the event that the rth transition out of i is to j. By the Markov property, the I_r are independent with constant mean p_{ij}. Using the strong law of large numbers and the fact that $\sum_k N_{ik} \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$, $\widehat{p}_{ij} \xrightarrow{\text{a.s.}} \mathbb{E}(I_1) = p_{ij}$.

32. (a) If X is transient then $V_i(n) < \infty$ a.s., and $\mu_i = \infty$, whence $V_i(n)/n \xrightarrow{\text{a.s.}} 0 = \mu_i^{-1}$. If X is persistent, then without loss of generality we may assume $X_0 = i$. Let $T(r)$ be the duration of the rth excursion from i. By the strong Markov property, the $T(r)$ are independent and identically distributed with mean μ_i. Furthermore,

$$\frac{1}{V_i(n)}\sum_{r=1}^{V_i(n)-1} T(r) \le \frac{n}{V_i(n)} \le \frac{1}{V_i(n)}\sum_{r=1}^{V_i(n)} T(r).$$

By the strong law of large numbers and the fact that $V_i(n) \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$, the two outer terms sandwich the central term, and the result follows.

(b) Note that $\sum_{r=0}^{n-1} f(X_r) = \sum_{i\in S} f(i)V_i(n)$. With Q a finite subset of S, and $\pi_i = \mu_i^{-1}$, the unique stationary distribution,

$$\begin{aligned}\left|\sum_{r=0}^{n-1}\frac{f(X_r)}{n} - \sum_i \frac{f(i)}{\mu_i}\right| &= \left|\sum_i\left(\frac{V_i(n)}{n} - \frac{1}{\mu_i}\right) f(i)\right| \\ &\le \left\{\sum_{i\in Q}\left|\frac{V_i(n)}{n} - \frac{1}{\mu_i}\right| + \sum_{i\notin Q}\left(\frac{V_i(n)}{n} + \frac{1}{\mu_i}\right)\right\}\|f\|_\infty,\end{aligned}$$

where $\|f\|_\infty = \sup\{|f(i)| : i \in S\}$. The sum over $i \in Q$ converges a.s. to 0 as $n \to \infty$, by part (a). The other sum satisfies

$$\sum_{i\notin Q}\left(\frac{V_i(n)}{n} + \frac{1}{\mu_i}\right) = 2 - \sum_{i\in Q}\left(\frac{V_i(n)}{n} + \pi_i\right)$$

which approaches 0 a.s., in the limits as $n \to \infty$ and $Q \uparrow S$.

33. (a) Since the chain is persistent, we may assume without loss of generality that $X_0 = j$. Define the times $R_1, R_2, \dots$ of return to j, the sojourn lengths $S_1, S_2, \dots$ in j, and the times $V_1, V_2, \dots$ between visits to j. By the Markov property and the strong law of large numbers,

$$\frac{1}{n}\sum_{r=1}^{n} S_r \xrightarrow{\text{a.s.}} \frac{1}{g_j}, \qquad \frac{1}{n}R_n = \frac{1}{n}\sum_{r=1}^{n} V_r \xrightarrow{\text{a.s.}} \mu_j.$$

Also, $R_n/R_{n+1} \xrightarrow{\text{a.s.}} 1$, since $\mu_j = \mathbb{E}(R_1) < \infty$. If $R_n < t < R_{n+1}$, then

$$\frac{R_n}{R_{n+1}} \cdot \frac{\sum_{r=1}^{n} S_r}{\sum_{r=1}^{n} V_r} \le \frac{1}{t}\int_0^t I_{\{X(s)=j\}}\,ds \le \frac{R_{n+1}}{R_n} \cdot \frac{\sum_{r=1}^{n+1} S_r}{\sum_{r=1}^{n+1} V_r}.$$

Let $n \to \infty$ to obtain the result.

(b) Note by Theorem (6.9.21) that $p_{ij}(t) \to \pi_j$ as $t \to \infty$. We take expectations of the integral in part (a), and the claim follows as in Corollary (6.4.22).

(c) Use the fact that

$$\int_0^t f(X(s))\,ds = \sum_{j\in S}\int_0^t I_{\{X(s)=j\}}\,ds$$

together with the method of solution of Problem (7.11.32b).

34. (a) By the first Borel–Cantelli lemma, $X_n = Y_n$ for all but finitely many values of n, almost surely. Off an event of probability zero, the sequences are identical for all large n.

(b) This follows immediately from part (a), since $X_n - Y_n = 0$ for all large n, almost surely.

(c) By the above, $a_n^{-1}\sum_{r=1}^{\infty}(X_r - Y_r) \xrightarrow{\text{a.s.}} 0$, which implies the claim.

35. Let $Y_n = X_n I_{\{|X_n| \le a\}}$. Then,

$$\sum_n \mathbb{P}(X_n \neq Y_n) = \sum_n \mathbb{P}(|X_n| > a) < \infty$$

by assumption (a), whence $\{X_n\}$ and $\{Y_n\}$ are tail-equivalent (see Problem (7.11.34)). By assumption (b) and the martingale convergence theorem (7.8.1) applied to the partial sums $\sum_{n=1}^{N}(Y_n - \mathbb{E}(Y_n))$, the infinite sum $\sum_{n=1}^{\infty}(Y_n - \mathbb{E}(Y_n))$ converges almost surely. Finally, $\sum_{n=1}^{\infty} \mathbb{E}(Y_n)$ converges by assumption (c), and therefore $\sum_{n=1}^{\infty} Y_n$, and hence $\sum_{n=1}^{\infty} X_n$, converges a.s.

36. (a) Let $n_1 < n_2 < \cdots < n_r = n$. Since the I_k take only two values, it suffices to show that

$$\mathbb{P}(I_{n_s} = 1 \text{ for } 1 \le s \le r) = \prod_{s=1}^{r} \mathbb{P}(I_{n_s} = 1).$$

Since F is continuous, the X_i take distinct values with probability 1, and furthermore the ranking of $X_1, X_2, \dots, X_n$ is equally likely to be any of the $n!$ available. Let $x_1, x_2, \dots, x_n$ be distinct reals, and write $A = \{X_i = x_i \text{ for } 1 \le i \le n\}$. Now,

$$\begin{aligned}
&\mathbb{P}(I_{n_s} = 1 \text{ for } 1 \le s \le r \mid A) \\
&\quad = \frac{1}{n!}\left\{\binom{n-1}{n_{s-1}}(n-1-n_{s-1})!\right\}\left\{\binom{n_{s-1}-1}{n_{s-2}}(n_{s-1}-1-n_{s-2})!\right\}\cdots(n_1-1)! \\
&\quad = \frac{1}{n_s}\cdot\frac{1}{n_{s-1}}\cdots\frac{1}{n_1},
\end{aligned}$$

and the claim follows on averaging over the x_i.

(b) We have that $\mathbb{E}(I_k) = \mathbb{P}(I_k = 1) = k^{-1}$ and $\operatorname{var}(I_k) = k^{-1}(1-k^{-1})$, whence $\sum_k \operatorname{var}(I_k/\log k) < \infty$. By the independence of the I_k and the martingale convergence theorem (7.8.1), $\sum_{k=1}^{\infty}(I_k - k^{-1})/\log k$ converges a.s. Therefore, by Kronecker's lemma (see Exercise (7.8.2)),

$$\frac{1}{\log n}\sum_{j=1}^{n}\left(I_j - \frac{1}{j}\right) \xrightarrow{\text{a.s.}} 0 \qquad \text{as } n \to \infty.$$

The result follows on recalling that $\sum_{j=1}^{n} j^{-1} \sim \log n$ as $n \to \infty$.

37. By an application of the three series theorem of Problem (7.11.35), the series converges almost surely.

8

Random processes

8.2 Solutions. Stationary processes

1. With $a_i(n) = \mathbb{P}(X_n = i)$, we have that

$$\begin{aligned}\operatorname{cov}(X_m, X_{m+n}) &= \mathbb{P}(X_{m+n} = 1 \mid X_m = 1)\mathbb{P}(X_m = 1) - \mathbb{P}(X_{m+n} = 1)\mathbb{P}(X_m = 1)\\ &= a_1(m)p_{11}(n) - a_1(m)a_1(m+n),\end{aligned}$$

and therefore,

$$\rho(X_m, X_{m+n}) = \frac{a_1(m)p_{11}(n) - a_1(m)a_1(m+n)}{\sqrt{a_1(m)(1-a_1(m))a_1(m+n)(1-a_1(m+n))}}.$$

Now, $a_1(m) \to \alpha/(\alpha+\beta)$ as $m \to \infty$, and

$$p_{11}(n) = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n,$$

whence $\rho(X_m, X_{m+n}) \to (1-\alpha-\beta)^n$ as $m \to \infty$. Finally,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{r=1}^{n} \mathbb{P}(X_r = 1) = \frac{\alpha}{\alpha+\beta}.$$

The process is strictly stationary if and only if X_0 has the stationary distribution.

2. We have that $\mathbb{E}(T(t)) = 0$ and $\operatorname{var}(T(t)) = \operatorname{var}(T_0) = 1$. Hence:

(a) $\rho(T(s), T(s+t)) = \mathbb{E}(T(s)T(s+t)) = \mathbb{E}\big[(-1)^{N(t+s)-N(s)}\big] = e^{-2\lambda t}$.

(b) Evidently, $\mathbb{E}(X(t)) = 0$, and

$$\begin{aligned}\mathbb{E}[X(t)^2] &= \mathbb{E}\left(\int_0^t\int_0^t T(u)T(v)\,du\,dv\right)\\ &= 2\int_{0<u<v<t} \mathbb{E}\big(T(u)T(v)\big)\,du\,dv = 2\int_{v=0}^{t}\int_{u=0}^{v} e^{-2\lambda(v-u)}\,du\,dv\\ &= \frac{1}{\lambda}\left(t - \frac{1}{2\lambda} + \frac{1}{2\lambda}e^{-2\lambda t}\right) \sim \frac{t}{\lambda} \quad \text{as } t \to \infty.\end{aligned}$$

3. We show first the existence of the limit $\lambda = \lim_{t\downarrow 0} g(t)/t$, where $g(t) = \mathbb{P}(N(t) > 0)$. Clearly,

$$\begin{aligned}g(x+y) &= \mathbb{P}\big(N(x+y) > 0\big)\\ &= \mathbb{P}\big(N(x) > 0\big) + \mathbb{P}\big(\{N(x) = 0\} \cap \{N(x+y) - N(x) > 0\}\big)\\ &\le g(x) + g(y) \qquad \text{for } x, y \ge 0.\end{aligned}$$

Such a function g is called subadditive, and the existence of λ follows by the *subadditive limit theorem* discussed in Problem (6.15.14). Note that $\lambda = \infty$ is a possibility.

Next, we partition the interval $(0, 1]$ into n equal sub-intervals, and let $I_n(r)$ be the indicator function of the event that at least one arrival lies in $\big((r-1)/n, r/n\big]$, $1 \le r \le n$. Then $\sum_{r=1}^n I_n(r) \uparrow N(1)$ as $n \to \infty$, with probability 1. By stationarity and monotone convergence,

$$\mathbb{E}(N(1)) = \mathbb{E}\left(\lim_{n\to\infty} \sum_{r=1}^{n} I_n(r)\right) = \lim_{n\to\infty} \mathbb{E}\left(\sum_{r=1}^{n} I_n(r)\right) = \lim_{n\to\infty} ng(n^{-1}) = \lambda.$$

8.3 Solutions. Renewal processes

1. See Problem (6.15.8).

2. With X a certain inter-event time, independent of the chain so far,

$$B_{n+1} = \begin{cases} X - 1 & \text{if } B_n = 0, \\ B_n - 1 & \text{if } B_n > 0. \end{cases}$$

Therefore, B is a Markov chain with transition probabilities $p_{i,i-1} = 1$ for $i > 0$, and $p_{0j} = f_{j+1}$ for $j \ge 0$, where $f_n = \mathbb{P}(X = n)$. The stationary distribution satisfies $\pi_j = \pi_{j+1} + \pi_0 f_{j+1}$, $j \ge 0$, with solution $\pi_j = \mathbb{P}(X > j)/\mathbb{E}(X)$, provided $\mathbb{E}(X)$ is finite.

The transition probabilities of B when reversed in equilibrium are

$$\widetilde{p}_{i,i+1} = \frac{\pi_{i+1}}{\pi_i} = \frac{\mathbb{P}(X > i+1)}{\mathbb{P}(X > i)}, \quad \widetilde{p}_{i0} = \frac{f_{i+1}}{\mathbb{P}(X > i)}, \qquad \text{for } i \ge 0.$$

These are the transition probabilities of the chain U of Exercise (8.3.1) with the f_j as given.

3. We have that $\rho^n u_n = \sum_{r=1}^n \rho^{n-k} u_{n-k} \rho^k f_k$, whence $v_n = \rho^n u_n$ defines a renewal sequence provided $\rho > 0$ and $\sum_n \rho^n f_n = 1$. By Exercise (8.3.1), there exists a Markov chain U and a state s such that $v_n = \mathbb{P}(U_n = s) \to \pi_s$, as $n \to \infty$, as required.

4. Noting that $N(0) = 0$,

$$\begin{aligned} \sum_{r=0}^{\infty} \mathbb{E}(N(r)) s^r &= \sum_{r=1}^{\infty} \sum_{k=1}^{r} u_k s^r = \sum_{k=1}^{\infty} u_k \sum_{r=k}^{\infty} s^r \\ &= \sum_{k=1}^{\infty} \frac{u_k s^k}{1-s} = \frac{U(s) - 1}{1-s} = \frac{F(s)U(s)}{1-s}. \end{aligned}$$

Let $S_m = \sum_{k=1}^m X_k$ and $S_0 = 0$. Then $\mathbb{P}(N(r) = n) = \mathbb{P}(S_n \le r) - \mathbb{P}(S_{n+1} \le r)$, and

$$\begin{aligned} \sum_{t=0}^{\infty} s^t \mathbb{E}\left[\binom{N(t)+k}{k}\right] &= \sum_{t=0}^{\infty} s^t \sum_{n=0}^{\infty} \binom{n+k}{k} \big(\mathbb{P}(S_n \le t) - \mathbb{P}(S_{n+1} \le t)\big) \\ &= \sum_{t=0}^{\infty} s^t \left[1 + \sum_{n=1}^{\infty} \binom{n+k-1}{k-1} \mathbb{P}(S_n \le t)\right]. \end{aligned}$$

Now,

$$\sum_{t=0}^{\infty} s^t \mathbb{P}(S_n \le t) = \sum_{t=0}^{\infty} s^t \sum_{i=0}^{t} \mathbb{P}(S_n = i) = \sum_{i=0}^{\infty} \mathbb{P}(S_n = i) \sum_{t=i}^{\infty} s^t = \frac{F(s)^n}{1-s},$$

whence, by the negative binomial theorem,

$$\sum_{t=0}^{\infty} s^t \mathbb{E}\left[\binom{N(t)+k}{k}\right] = \frac{1}{(1-s)(1-F(s))^k} = \frac{U(s)^k}{1-s}.$$

5. This is an immediate consequence of the fact that the interarrival times of a Poisson process are exponentially distributed, since this specifies the distribution of the process.

8.4 Solutions. Queues

1. We use the lack-of-memory property repeatedly, together with the fact that, if X and Y are independent exponential variables with respective parameters λ and μ, then $\mathbb{P}(X < Y) = \lambda/(\lambda+\mu)$.
(a) In this case,

$$p = \frac{1}{2}\left\{\frac{\lambda}{\lambda+\mu}\cdot\frac{\mu}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}\right\} + \frac{1}{2}\left\{\frac{\mu}{\lambda+\mu}\cdot\frac{\lambda}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}\right\} = \frac{1}{2} + \frac{2\lambda\mu}{(\lambda+\mu)^2}.$$

(b) If $\lambda < \mu$, and you pick the quicker server, $p = 1 - \left(\dfrac{\mu}{\lambda+\mu}\right)^2$.

(c) And finally, $p = \dfrac{2\lambda\mu}{(\lambda+\mu)^2}$.

2. The given event occurs if the time X to the next arrival is less than t, and also less than the time Y of service of the customer present. Now,

$$\mathbb{P}(X \le t,\ X \le Y) = \int_0^t \lambda e^{-\lambda x} e^{-\mu x}\,dx = \frac{\lambda}{\lambda+\mu}(1 - e^{-(\lambda+\mu)t}).$$

3. By conditioning on the time of passage of the first vehicle,

$$\mathbb{E}(T) = \int_0^a \bigl(x + \mathbb{E}(T)\bigr)\lambda e^{-\lambda x}\,dx + ae^{-\lambda a},$$

and the result follows. If it takes a time b to cross the other lane, and so $a+b$ to cross both, then, with an obvious notation,

(a) $$\mathbb{E}(T_a) + \mathbb{E}(T_b) = \frac{e^{a\lambda}-1}{\lambda} + \frac{e^{b\mu}-1}{\mu},$$

(b) $$\mathbb{E}(T_{a+b}) = \frac{e^{(a+b)(\lambda+\mu)}-1}{\lambda+\mu}.$$

The latter must be the greater, by a consideration of the problem, or by turgid calculation.

4. Look for a solution of the detailed balance equations

$$\mu\pi_{n+1} = \frac{\lambda(n+1)}{n+2}\pi_n, \qquad n \ge 0.$$

to find that $\pi_n = \rho^n\pi_0/(n+1)$ is a stationary distribution if $\rho < 1$, in which case $\pi_0 = -\rho/\log(1-\rho)$. Hence $\sum_n n\pi_n = \lambda\pi_0/(\mu-\lambda)$, and by the lack-of-memory property the mean time spent waiting for service is $\rho\pi_0/(\mu-\lambda)$. An arriving customer joins the queue with probability

$$\sum_{n=0}^{\infty}\frac{n+1}{n+2}\pi_n = \frac{\rho+\log(1-\rho)}{\rho\log(1-\rho)}.$$

5. By considering possible transitions during the interval $(t, t+h)$, the probability $p_i(t)$ that exactly i demonstrators are busy at time t satisfies:

$$\begin{aligned} p_2(t+h) &= p_1(t)2h + p_2(t)(1-2h) + \mathrm{o}(h), \\ p_1(t+h) &= p_0(t)2h + p_1(t)(1-h)(1-2h) + p_2(t)2h + \mathrm{o}(h), \\ p_0(t+h) &= p_0(t)(1-2h) + p_1(t)h + \mathrm{o}(h). \end{aligned}$$

Hence,

$$p_2'(t) = 2p_1(t) - 2p_2(t), \quad p_1'(t) = 2p_0(t) - 3p_1(t) + 2p_2(t), \quad p_0'(t) = -2p_0(t) + p_1(t),$$

and therefore $p_2(t) = a + be^{-2t} + ce^{-5t}$ for some constants a, b, c. By considering the values of p_2 and its derivatives at $t = 0$, the boundary conditions are found to be $a + b + c = 0$, $-2b - 5c = 0$, $4b + 25c = 4$, and the claim follows.

8.5 Solutions. The Wiener process

1. We might as well assume that W is standard, in that $\sigma^2 = 1$. Because the joint distribution is multivariate normal, we may use Exercise (4.7.5) for the first part, and Exercise (4.9.8) for the second, giving the answer

$$\frac{1}{8} + \frac{1}{4\pi}\left\{\sin^{-1}\sqrt{\frac{s}{t}} + \sin^{-1}\sqrt{\frac{s}{u}} + \sin^{-1}\sqrt{\frac{t}{u}}\right\}.$$

2. Writing $W(s) = \sqrt{s}X$, $W(t) = \sqrt{t}Z$, and $W(u) = \sqrt{u}Y$, we obtain random variables X, Y, Z with the standard trivariate normal distribution, with correlations $\rho_1 = \sqrt{s/u}$, $\rho_2 = \sqrt{t/u}$, $\rho_3 = \sqrt{s/t}$. By the solution to Exercise (4.9.9),

$$\mathrm{var}(Z \mid X, Y) = \frac{(u-t)(t-s)}{t(u-s)},$$

yielding $\mathrm{var}(W(t) \mid W(s), W(u))$ as required. Also,

$$\mathbb{E}\big(W(t)W(u) \mid W(s), W(v)\big) = \mathbb{E}\left\{\left[\frac{(u-t)W(s) + (t-s)W(u)}{u-s}\right] W(u) \,\middle|\, W(s), W(v)\right\},$$

which yields the conditional correlation after some algebra.

3. Whenever $a^2 + b^2 = 1$.

4. Let $\Delta_j(n) = W((j+1)t/n) - W(jt/n)$. By the independence of these increments,

$$\begin{aligned} \mathbb{E}\left(\sum_{j=0}^{n-1} \Delta_j(n)^2 - \sigma^2 t\right)^2 &= \sum_{j=0}^{n-1} \mathbb{E}\left(\Delta_j(n)^2 - \frac{t\sigma^2}{n}\right)^2 && \text{because } \mathbb{E}(\Delta_j(n)^2) = \frac{t\sigma^2}{n} \\ &= \sum_{j=0}^{n-1}\left(\frac{3t^2\sigma^4}{n^2} - \frac{2t^2\sigma^4}{n^2} + \frac{t^2\sigma^4}{n^2}\right) && \text{because } \mathbb{E}(\Delta_j(n)^4) = \frac{3t^2\sigma^4}{n^2} \\ &= \frac{2t^2\sigma^4}{n} \to 0 \qquad \text{as } n \to \infty. \end{aligned}$$

5. They all have mean zero and variance t, but only (a) has independent normally distributed increments.

8.7 Solutions to problems

1. $\mathbb{E}(Y_n) = 0$, and $\operatorname{cov}(Y_m, Y_{m+n}) = \sum_{i=0}^{r} \alpha_i \alpha_{n+i}$ for $m, n \geq 0$, with the convention that $\alpha_k = 0$ for $k > r$. The covariance does not depend on m, and therefore the sequence is stationary.

2. We have, by iteration, that $Y_n = S_n(m) + \alpha^{m+1} Y_{n-m-1}$ where $S_n(m) = \sum_{j=0}^{m} \alpha^j Z_{n-j}$. There are various ways of showing that the sequence $\{S_n(m) : m \geq 1\}$ converges in mean square and almost surely, and the shortest is as follows. We have that $\alpha^{m+1} Y_{n-m-1} \to 0$ in m.s. and a.s. as $m \to \infty$; to see this, use the facts that $\operatorname{var}(\alpha^{m+1} Y_{n-m-1}) = \alpha^{2(m+1)} \operatorname{var}(Y_0)$, and

$$\sum_m \mathbb{P}(\alpha^{m+1} Y_{n-m-1} > \epsilon) \leq \sum_m \frac{\alpha^{2(m+1)} \mathbb{E}(Y_0^2)}{\epsilon^2} < \infty, \qquad \epsilon > 0.$$

It follows that $S_n(m) = Y_n - \alpha^{m+1} Y_{n-m-1}$ converges in m.s. and a.s. as $m \to \infty$. A longer route to the same conclusion is as follows. For $r < s$,

$$\mathbb{E}\big(|S_n(s) - S_n(r)|^2\big) = \mathbb{E}\left\{\left(\sum_{j=r+1}^{s} \alpha^j Z_{n-j}\right)^2\right\} = \sum_{j=r+1}^{s} \alpha^{2j} \leq \frac{\alpha^{2r}}{1-\alpha^2},$$

whence $\{S_n(m) : m \geq 1\}$ is Cauchy convergent in mean square, and therefore converges in mean square. In order to show the almost sure convergence of $S_n(m)$, one may argue as follows. Certainly

$$\mathbb{E}\left(\sum_{j=0}^{m} |\alpha^j Z_{n-j}|\right) = \sum_{j=0}^{m} \mathbb{E}|\alpha^j Z_{n-j}| \to \sum_{j=0}^{\infty} |\alpha|^j \mathbb{E}|Z_{n-j}| \leq \sum_{j=0}^{\infty} |\alpha|^j < \infty,$$

whence $\sum_{j=0}^{\infty} \alpha^j Z_{n-j}$ is a.s. absolutely convergent, and therefore a.s. convergent also. We may express $\lim_{m\to\infty} S_n(m)$ as $\sum_{j=0}^{\infty} \alpha^j Z_{n-j}$. Also, $\alpha^{m+1} Y_{n-m-1} \to 0$ in mean square and a.s. as $m \to \infty$, and we may therefore express Y_n as

$$Y_n = \sum_{j=0}^{\infty} \alpha^j Z_{n-j} \qquad \text{a.s.}$$

It follows that $\mathbb{E}(Y_n) = \lim_{m\to\infty} \mathbb{E}(S_n(m)) = 0$. Finally, for $r > 0$, the autocovariance function is given by

$$c(r) = \operatorname{cov}(Y_n, Y_{n-r}) = \mathbb{E}\big\{(\alpha Y_{n-1} + Z_n) Y_{n-r}\big\} = \alpha c(r-1),$$

whence

$$c(r) = \alpha^{|r|} c(0) = \frac{\alpha^{|r|}}{1-\alpha^2}, \qquad r = \dots, -1, 0, 1, \dots,$$

since $c(0) = \operatorname{var}(Y_n)$.

3. If t is a non-negative integer, $N(t)$ is the number of 0's and 1's preceding the $(t+1)$th 1. Therefore $N(t) + 1$ has the negative binomial distribution with mass function

$$f(k) = \binom{k-1}{t} p^{t+1} (1-p)^{k-1-t}, \qquad k \geq t+1.$$

If t is not an integer, then $N(t) = N(\lfloor t \rfloor)$.

4. We have that

$$\mathbb{P}\big(Q(t+h) = j \mid Q(t) = i\big) = \begin{cases} \lambda h + o(h) & \text{if } j = i+1, \\ \mu i h + o(h) & \text{if } j = i-1, \\ 1 - (\lambda + \mu i)h + o(h) & \text{if } j = i, \end{cases}$$

an immigration–death process with constant birth rate λ and death rates $\mu_i = i\mu$.

Either calculate the stationary distribution in the usual way, or use the fact that birth–death processes are reversible in equilibrium. Hence $\lambda\pi_i = \mu(i+1)\pi_{i+1}$ for $i \ge 0$, whence

$$\pi_i = \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i e^{-\lambda/\mu}, \qquad i \ge 0.$$

5. We have that $\widetilde{X}(t) = R\cos(\Psi)\cos(\theta t) - R\sin(\Psi)\sin(\theta t)$. Consider the transformation $u = r\cos\psi$, $v = -r\sin\psi$, which maps $[0,\infty) \times [0, 2\pi)$ to $\mathbb{R}^2$. The Jacobian is

$$\begin{vmatrix} \dfrac{\partial u}{\partial r} & \dfrac{\partial u}{\partial \psi} \\ \dfrac{\partial v}{\partial r} & \dfrac{\partial v}{\partial \psi} \end{vmatrix} = -r,$$

whence $U = R\cos\Psi$, $V = -R\sin\Psi$ have joint density function satisfying

$$r f_{U,V}(r\cos\psi, -r\sin\psi) = f_{R,\Psi}(r,\psi).$$

Substitute $f_{U,V}(u,v) = e^{-\frac{1}{2}(u^2+v^2)}/(2\pi)$, to obtain

$$f_{R,\Psi}(r,\psi) = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2}, \qquad r > 0,\ 0 \le \psi < 2\pi.$$

Thus R and Ψ are independent, the latter being uniform on $[0, 2\pi)$.

6. A customer arriving at time u is designated *green* if he is in state A at time t, an event having probability $p(u, t-u)$. By the colouring theorem (6.13.14), the arrival times of green customers form a non-homogeneous Poisson process with intensity function $\lambda(u)p(u, t-u)$, and the claim follows.

9

Stationary processes

9.1 Solutions. Introduction

1. We examine sequences W_n of the form

$$(*)\qquad W_n = \sum_{k=0}^{\infty} a_k Z_{n-k}$$

for the real sequence $\{a_k : k \geq 0\}$. Substitute, to obtain $a_0 = 1, a_1 = \alpha, a_r = \alpha a_{r-1} + \beta a_{r-2}, r \geq 2$, with solution

$$a_r = \begin{cases} (1+r)\lambda_1^r & \text{if } \alpha^2 + 4\beta = 0, \\ \dfrac{\lambda_1^{r+1} - \lambda_2^{r+1}}{\lambda_1 - \lambda_2} & \text{otherwise,} \end{cases}$$

where λ_1 and λ_2 are the (possibly complex) roots of the quadratic $x^2 - \alpha x - \beta = 0$ (these roots are distinct if and only if $\alpha^2 + 4\beta \neq 0$).

Using the method in the solution to Problem (8.7.2), the sum in $(*)$ converges in mean square and almost surely if $|\lambda_1| < 1$ and $|\lambda_2| < 1$. Assuming this holds, we have from $(*)$ that $\mathbb{E}(W_n) = 0$ and the autocovariance function is

$$c(m) = \mathbb{E}(W_n W_{n-m}) = \alpha c(m-1) + \beta c(m-2), \qquad m \geq 1,$$

by the independence of the Z_n. Therefore W is weakly stationary, and the autocovariance function may be expressed in terms of α and β.

2. We adopt the convention that, if the binary expansion of U is non-unique, then we take the (unique) non-terminating such expansion. It is clear that X_i takes values in $\{0, 1\}$, and

$$\mathbb{P}\big(X_{n+1} = 1 \,\big|\, X_i = x_i \text{ for } 1 \leq i \leq n\big) = \tfrac{1}{2}$$

for all $x_1, x_2, \dots, x_n$; therefore the X's are independent Bernoulli random variables. For any sequence $k_1 < k_2 < \dots < k_r$, the joint distribution of $V_{k_1}, V_{k_2}, \dots, V_{k_r}$ depends only on that of $X_{k_1+1}, X_{k_1+2}, \dots$. Since this distribution is the same as the distribution of $X_1, X_2, \dots$, we have that $(V_{k_1}, V_{k_2}, \dots, V_{k_r})$ has the same distribution as $(V_0, V_{k_2-k_1}, \dots, V_{k_r-k_1})$. Therefore V is strongly stationary.

Clearly $\mathbb{E}(V_n) = \mathbb{E}(V_0) = \frac{1}{2}$, and, by the independence of the X_i,

$$\operatorname{cov}(V_0, V_n) = \sum_{i=1}^{\infty} 2^{-2i-n} \operatorname{var}(X_i) = \tfrac{1}{12}(\tfrac{1}{2})^n.$$

3. (i) For mean-square convergence, we show that $S_k = \sum_{n=0}^{k} a_n X_n$ is mean-square Cauchy convergent as $k \to \infty$. We have that, for $r < s$,

$$\mathbb{E}\{(S_s - S_r)^2\} = \sum_{i,j=r+1}^{s} a_i a_j c(i-j) \le c(0)\Big\{\sum_{i=r+1}^{s} |a_i|\Big\}^2$$

since $|c(m)| \le c(0)$ for all m, by the Cauchy–Schwarz inequality. The last sum tends to 0 as $r, s \to \infty$ if $\sum_i |a_i| < \infty$. Hence S_k converges in mean square as $k \to \infty$.

Secondly,

$$\mathbb{E}\Big(\sum_{k=1}^{n} |a_k X_k|\Big) \le \sum_{k=1}^{n} |a_k| \cdot \mathbb{E}|X_k| \le \sqrt{\mathbb{E}(X_0^2)} \sum_{k=1}^{n} |a_k|$$

which converges as $n \to \infty$ if the $|a_k|$ are summable. It follows that $\sum_{k=1}^{n} |a_k X_k|$ converges absolutely (almost surely), and hence $\sum_{k=1}^{n} a_k X_k$ converges a.s.

(ii) Each sum converges a.s. and in mean square, by part (i). Now

$$c_Y(m) = \sum_{j,k=0}^{\infty} a_j a_k c(m+k-j)$$

whence

$$\sum_m |c_Y(m)| \le c(0)\Big\{\sum_{j=0}^{\infty} |a_j|\Big\}^2 < \infty.$$

4. Clearly X_n has distribution $\boldsymbol{\pi}$ for all n, so that $\{f(X_n) : n \ge m\}$ has fdds which do not depend on the value of m. Therefore the sequence is strongly stationary.

9.2 Solutions. Linear prediction

1. (i) We have that

$$(*) \qquad \mathbb{E}\{(X_{n+1} - \alpha X_n)^2\} = (1+\alpha^2)c(0) - 2\alpha c(1),$$

which is minimized by setting $\alpha = c(1)/c(0)$. Hence $\widehat{X}_{n+1} = c(1)X_n/c(0)$.

(ii) Similarly

$$(**) \qquad \mathbb{E}\{(X_{n+1} - \beta X_n - \gamma X_{n-1})^2\} = (1+\beta^2+\gamma^2)c(0) + 2\beta(\gamma-1)c(1) - 2\gamma c(2),$$

an expression which is minimized by the choice

$$\beta = \frac{c(1)\big(c(0)-c(2)\big)}{c(0)^2 - c(1)^2}, \qquad \gamma = \frac{c(0)c(2) - c(1)^2}{c(0)^2 - c(1)^2};$$

$\widetilde{X}_{n+1}$ is given accordingly.

(iii) Substitute α, β, γ into $(*)$ and $(**)$, and subtract to obtain, after some manipulation,

$$D = \frac{\{c(1)^2 - c(0)c(2)\}^2}{c(0)\{c(0)^2 - c(1)^2\}}.$$

(a) In this case $c(0) = \frac{1}{2}$, and $c(1) = c(2) = 0$. Therefore $\widehat{X}_{n+1} = \widetilde{X}_{n+1} = 0$, and $D = 0$.
(b) In this case $D = 0$ also.

In both (a) and (b), little of substance is gained by using $\widetilde{X}_{n+1}$ in place of $\widehat{X}_{n+1}$.

2. Let $\{Z_n : n = \dots, -1, 0, 1, \dots\}$ be independent random variables with zero means and unit variances, and define the moving-average process

$$(*) \qquad X_n = \frac{Z_n + aZ_{n-1}}{\sqrt{1+a^2}}.$$

It is easily checked that X has the required autocovariance function.

By the projection theorem, $X_n - \widehat{X}_n$ is orthogonal to the collection $\{X_{n-r} : r \geq 1\}$, so that $\mathbb{E}\{(X_n - \widehat{X}_n)X_{n-r}\} = 0$, $r \geq 1$. Set $\widehat{X}_n = \sum_{s=1}^{\infty} b_s X_{n-s}$ to obtain that

$$\alpha = b_1 + b_2\alpha, \quad 0 = b_{s-1}\alpha + b_s + b_{s+1}\alpha \qquad \text{for } s \geq 2,$$

where $\alpha = a/(1+a^2)$. The unique bounded solution to the above difference equation is $b_s = (-1)^{s+1}a^s$, and therefore

$$\widehat{X}_n = \sum_{s=1}^{\infty}(-1)^{s+1}a^s X_{n-s}.$$

The mean squared error of prediction is

$$\mathbb{E}\{(X_n - \widehat{X}_n)^2\} = \mathbb{E}\left\{\left(\sum_{s=0}^{\infty}(-a)^s X_{n-s}\right)^2\right\} = \frac{1}{1+a^2}\mathbb{E}(Z_n^2) = \frac{1}{1+a^2}.$$

Clearly $\mathbb{E}(\widehat{X}_n) = 0$ and

$$\operatorname{cov}(\widehat{X}_n, \widehat{X}_{n-m}) = \sum_{r,s=1}^{\infty} b_r b_s c(m+r-s), \qquad m \geq 0,$$

so that $\widehat{X}$ is weakly stationary.

9.3 Solutions. Autocovariances and spectra

1. It is clear that $\mathbb{E}(X_n) = 0$ and $\operatorname{var}(X_n) = 1$. Also

$$\operatorname{cov}(X_m, X_{m+n}) = \cos(m\lambda)\cos\{(m+n)\lambda\} + \sin(m\lambda)\sin\{(m+n)\lambda\} = \cos(n\lambda),$$

so that X is stationary, and the spectrum of X is the singleton $\{\lambda\}$.

2. Certainly $\phi_U(t) = (e^{it\pi} - e^{-it\pi})/(2\pi it)$, so that $\mathbb{E}(X_n) = \phi_U(1)\phi_V(n) = 0$. Also

$$\operatorname{cov}(X_m, X_{m+n}) = \mathbb{E}(X_m\overline{X}_{m+n}) = \mathbb{E}\big(e^{i\{U - Vm - U + V(m+n)\}}\big) = \phi_V(n),$$

whence X is stationary. Finally, the autocovariance function is

$$c(n) = \phi_V(n) = \int e^{in\lambda}\, dF(\lambda),$$

whence F is the spectral distribution function.

3. The characteristic functions of these distributions are

(i) $$\rho(t) = e^{-\frac{1}{2}t^2},$$

(ii) $$\rho(t) = \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2}.$$

4. (i) We have that

$$\operatorname{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\sum_{j,k=1}^{n} \operatorname{cov}(X_j, X_k) = \frac{c(0)}{n^2}\int_{(-\pi,\pi]}\left(\sum_{j,k=1}^{n} e^{i(j-k)\lambda}\right) dF(\lambda).$$

The integrand is

$$\left|\sum_{j=1}^{n} e^{ij\lambda}\right|^2 = \left(\frac{e^{in\lambda}-1}{e^{i\lambda}-1}\right)\left(\frac{e^{-in\lambda}-1}{e^{-i\lambda}-1}\right) = \frac{1-\cos(n\lambda)}{1-\cos\lambda},$$

whence

$$\operatorname{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = c(0)\int_{(-\pi,\pi]}\left(\frac{\sin(n\lambda/2)}{n\sin(\lambda/2)}\right)^2 dF(\lambda).$$

It is easily seen that $|\sin\theta| \le |\theta|$, and therefore the integrand is no larger than

$$\left(\frac{\lambda/2}{\sin(\lambda/2)}\right)^2 \le (\tfrac{1}{2}\pi)^2.$$

As $n \to \infty$, the integrand converges to the function which is zero everywhere except at the origin, where (by continuity) we may assign it the value 1. It may be seen, using the dominated convergence theorem, that the integral converges to $F(0) - F(0-)$, the size of the discontinuity of F at the origin, and therefore the variance tends to 0 if and only if $F(0) - F(0-) = 0$.

Using a similar argument,

$$\frac{1}{n}\sum_{j=0}^{n-1} c(j) = \frac{c(0)}{n}\int_{(-\pi,\pi]}\left(\sum_{j=0}^{n-1} e^{ij\lambda}\right) dF(\lambda) = c(0)\int_{(-\pi,\pi]} g_n(\lambda)\, dF(\lambda)$$

where

$$g_n(\lambda) = \begin{cases} 1 & \text{if } \lambda = 0, \\ \dfrac{e^{in\lambda}-1}{n(e^{i\lambda}-1)} & \text{if } \lambda \ne 0, \end{cases}$$

is a bounded sequence of functions which converges as before to the Kronecker delta function $\delta_{\lambda 0}$. Therefore

$$\frac{1}{n}\sum_{j=0}^{n-1} c(j) \to c(0)\big(F(0) - F(0-)\big) \qquad \text{as } n \to \infty.$$

9.4 Solutions. Stochastic integration and the spectral representation

1. Let H_X be the space of all linear combinations of the X_i, and let $\overline{H}_X$ be the closure of this space, that is, H_X together with the limits of all mean-square Cauchy-convergent sequences in H_X. All members of H_X have zero mean, and therefore all members of $\overline{H}_X$ also. Now $S(\lambda) \in \overline{H}_X$ for all λ, whence $\mathbb{E}(S(\lambda) - S(\mu)) = 0$ for all λ and μ.

2. First, each Y_m lies in the space $\overline{H}_X$ containing all linear combinations of the X_n and all limits of mean-square Cauchy-convergent sequences of the same form. As in the solution to Exercise (9.4.1), all members of $\overline{H}_X$ have zero mean, and therefore $\mathbb{E}(Y_m) = 0$ for all m. Secondly,

$$\mathbb{E}(Y_m \overline{Y}_n) = \int_{(-\pi,\pi]} \frac{e^{im\lambda}e^{-in\lambda}}{2\pi f(\lambda)} f(\lambda)\, d\lambda = \delta_{mn}.$$

As for the last part,

$$\sum_{j=-\infty}^{\infty} a_j Y_{n-j} = \int_{(-\pi,\pi]} \Big(\sum_j a_j e^{-ij\lambda}\Big) \frac{e^{in\lambda}}{\sqrt{2\pi f(\lambda)}}\, dS(\lambda) = \int_{(-\pi,\pi]} e^{in\lambda}\, dS(\lambda) = X_n.$$

This proves that such a sequence X_n may be expressed as a moving average of an orthonormal sequence.

3. Let $\overline{H}_X$ be the space of all linear combinations of the X_n, together with all limits of (mean-square) Cauchy-convergent sequences of such combinations. Using the result of Problem (7.11.19), all elements in $\overline{H}_X$ are normally distributed. In particular, all increments of the spectral process are normal. Similarly, all pairs in $\overline{H}_X$ are jointly normally distributed, and therefore two members of $\overline{H}_X$ are independent if and only if they are uncorrelated. Increments of the spectral process have zero means (by Exercise (9.4.1)) and are orthogonal. Therefore they are uncorrelated, and hence independent.

9.5 Solutions. The ergodic theorem

1. With the usual shift operator τ, it is obvious that $\tau^{-1}\varnothing = \varnothing$, so that $\varnothing \in \mathcal{I}$. Secondly, if $A \in \mathcal{I}$, then $\tau^{-1}(A^c) = (\tau^{-1}A)^c = A^c$, whence $A^c \in \mathcal{I}$. Thirdly, suppose $A_1, A_2, \ldots \in \mathcal{I}$. Then

$$\tau^{-1}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \bigcup_{i=1}^{\infty} \tau^{-1}A_i = \bigcup_{i=1}^{\infty} A_i,$$

so that $\bigcup_1^\infty A_i \in \mathcal{I}$.

2. The left-hand side is the sum of covariances, $c(0)$ appearing n times, and $c(i)$ appearing $2(n-i)$ times for $0 < i < n$, in agreement with the right-hand side.

Let $\epsilon > 0$. If $\overline{c}(j) = j^{-1}\sum_{i=0}^{j-1} c(i) \to \sigma^2$ as $j \to \infty$, there exists J such that $|\overline{c}(j) - \sigma^2| < \epsilon$ when $j \ge J$. Now

$$\frac{2}{n^2}\sum_{j=1}^{n} j\overline{c}(j) \le \frac{2}{n^2}\Big\{\sum_{j=1}^{J} j\overline{c}(j) + \sum_{j=J+1}^{n} j(\sigma^2+\epsilon)\Big\} \to \sigma^2 + \epsilon$$

as $n \to \infty$. A related lower bound is proved similarly, and the claim follows since ϵ (> 0) is arbitrary.

3. It is easily seen that $S_m = \sum_{i=0}^{m} \alpha_i X_{n+i}$ constitutes a martingale with respect to the X's, and

$$\mathbb{E}(S_m^2) = \sum_{i=0}^{m} \alpha_i^2 \mathbb{E}(X_{n+i}^2) \le \sum_{i=0}^{\infty} \alpha_i^2,$$

whence S_m converges a.s. and in mean square as $m \to \infty$.

Since the X_n are independent and identically distributed, the sequence Y_n is strongly stationary; also $\mathbb{E}(Y_n) = 0$, and so $n^{-1}\sum_{i=1}^{n} Y_i \to Z$ a.s. and in mean, for some random variable Z with mean zero. For any fixed m (≥ 1), the contribution of $X_1, X_2, \dots, X_m$ towards $\sum_{i=1}^{n} Y_i$ is, for large n, no larger than

$$C_m = \left|\sum_{j=1}^{m} (\alpha_0 + \alpha_1 + \dots + \alpha_{j-1}) X_j\right|.$$

Now $n^{-1} C_m \to 0$ as $n \to \infty$, so that Z is defined in terms of the subsequence $X_{m+1}, X_{m+2}, \dots$ for all m, which is to say that Z is a tail function of a sequence of independent random variables. Therefore Z is a.s. constant, and so $Z = 0$ a.s.

9.6 Solutions. Gaussian processes

1. The quick way is to observe that c is the autocovariance function of a Poisson process with intensity 1. Alternatively, argue as follows. The sum is unchanged by taking complex conjugates, and hence is real. Therefore it equals

$$\begin{aligned}\sum_{j=1}^{n} t_j \left(|z_j|^2 + z_j \sum_{k=j+1}^{n} \bar{z}_k + \bar{z}_j \sum_{k=j+1}^{n} z_k \right) &= \sum_{j=1}^{n} t_j \left(\left|\sum_{k=j}^{n} z_k\right|^2 - \left|\sum_{k=j+1}^{n} z_k\right|^2 \right) \\ &= \sum_{j=1}^{n} (t_j - t_{j-1}) \left|\sum_{k=j}^{n} z_k\right|^2\end{aligned}$$

where $t_0 = 0$.

2. For $s, t \ge 0$, $X(s)$ and $X(s+t)$ have a bivariate normal distribution with zero means, unit variances, and covariance $c(t)$. It is standard (see Problem (4.14.13)) that $\mathbb{E}(X(s+t) \mid X(s)) = c(t)X(s)$. Now

$$\begin{aligned}c(s+t) = \mathbb{E}\big(X(0)X(s+t)\big) &= \mathbb{E}\Big\{\mathbb{E}\big(X(0)X(s+t) \mid X(0), X(s)\big)\Big\} \\ &= \mathbb{E}\big(X(0)c(t)X(s)\big) = c(s)c(t)\end{aligned}$$

by the Markov property. Therefore c satisfies $c(s+t) = c(s)c(t)$, $c(0) = 1$, whence $c(s) = c(1)^{|s|} = \rho^{|s|}$. Using the inversion formula, the spectral density function is

$$f(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} c(s) e^{-is\lambda} = \frac{1-\rho^2}{2\pi|1-\rho e^{i\lambda}|^2}, \qquad |\lambda| \le \pi.$$

Note that X has the same autocovariance function as a certain autoregressive process. Indeed, stationary Gaussian Markov processes have such a representation.

3. If X is Gaussian and strongly stationary, then it is weakly stationary since it has a finite variance. Conversely suppose X is Gaussian and weakly stationary. Then $c(s,t) = \operatorname{cov}(X(s), X(t))$ depends on $t - s$ only. The joint distribution of $X(t_1), X(t_2), \dots, X(t_n)$ depends only on the common mean and the covariances $c(t_i, t_j)$. Now $c(t_i, t_j)$ depends on $t_j - t_i$ only, whence $X(t_1), X(t_2), \dots, X(t_n)$ have the same joint distribution as $X(s + t_1), X(s + t_2), \dots, X(s + t_n)$. Therefore X is strongly stationary.

4. (a) If $s, t > 0$, we have from Problem (4.14.13) that

$$\mathbb{E}\big(X(s+t)^2 \mid X(s)\big) = X(s)^2 c(t)^2 + 1 - c(t)^2,$$

whence

$$\begin{aligned}\operatorname{cov}\big(X(s)^2, X(s+t)^2\big) &= \mathbb{E}\big(X(s)^2 X(s+t)^2\big) - 1\\ &= \mathbb{E}\Big\{X(s)^2 \mathbb{E}\big(X(s+t)^2 \mid X(s)\big)\Big\} - 1\\ &= c(t)^2 \mathbb{E}(X(s)^4) + (1 - c(t)^2)\mathbb{E}(X(s)^2) - 1 = 2c(t)^2\end{aligned}$$

by an elementary calculation.
(b) Likewise $\operatorname{cov}(X(s)^3, X(s+t)^3) = 3(3 + 2c(t)^2)c(t)$.

9.7 Solutions to problems

1. It is easily seen that $Y_n = X_n + (\alpha - \beta)X_{n-1} + \beta Y_{n-1}$, whence the autocovariance function c of Y is given by

$$c(k) = \begin{cases} \dfrac{1 + \alpha^2 - \beta^2}{1 - \beta^2} & \text{if } k = 0,\\[2ex] \beta^{|k|-1}\left\{\dfrac{\alpha(1 + \alpha\beta - \beta^2)}{1 - \beta^2}\right\} & \text{if } k \neq 0.\end{cases}$$

Set $\widehat{Y}_{n+1} = \sum_{i=0}^{\infty} a_i Y_{n-i}$ and find the a_i for which it is the case that $\mathbb{E}\{(Y_{n+1} - \widehat{Y}_{n+1})Y_{n-k}\} = 0$ for $k \geq 0$. These equations yield

$$c(k+1) = \sum_{i=0}^{\infty} a_i c(k-i), \qquad k \geq 0,$$

which have solution $a_i = \alpha(\beta - \alpha)^i$ for $i \geq 0$.

2. The autocorrelation functions of X and Y satisfy

$$\sigma_X^2 \rho_X(n) = \sigma_Y^2 \sum_{j,k=0}^{r} a_j a_k \rho_Y(n + k - j).$$

Therefore

$$\begin{aligned}\sigma_X^2 f_X(\lambda) &= \frac{\sigma_Y^2}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \sum_{j,k=0}^{r} a_j a_k \rho_Y(n + k - j)\\ &= \frac{\sigma_Y^2}{2\pi} \sum_{j,k=0}^{r} a_j a_k e^{i(k-j)\lambda} \sum_{n=-\infty}^{\infty} e^{-i(n+k-j)\lambda} \rho_Y(n + k - j)\\ &= \sigma_Y^2 |G_a(e^{i\lambda})|^2 f_Y(\lambda).\end{aligned}$$

In the case of exponential smoothing, $G_a(e^{i\lambda}) = (1-\mu)/(1-\mu e^{i\lambda})$, so that

$$f_X(\lambda) = \frac{c(1-\mu)^2 f_Y(\lambda)}{1-2\mu\cos\lambda+\mu^2}, \qquad |\lambda| < \pi,$$

where $c = \sigma_Y^2/\sigma_X^2$ is a constant chosen to make this a density function.

3. Consider the sequence $\{X_n\}$ defined by

$$X_n = Y_n - \widehat{Y}_n = Y_n - \alpha Y_{n-1} - \beta Y_{n-2}.$$

Now X_n is orthogonal to $\{Y_{n-k} : k \ge 1\}$, so that the X_n are uncorrelated random variables with spectral density function $f_X(\lambda) = (2\pi)^{-1}$, $\lambda \in (-\pi, \pi)$. By the result of Problem (9.7.2),

$$\sigma_X^2 f_X(\lambda) = \sigma_Y^2 |1-\alpha e^{i\lambda} - \beta e^{2i\lambda}|^2 f_Y(\lambda),$$

whence

$$f_Y(\lambda) = \frac{\sigma_X^2/\sigma_Y^2}{2\pi|1-\alpha e^{i\lambda}-\beta e^{2i\lambda}|^2}, \qquad -\pi < \lambda < \pi.$$

4. Let $\{X_n' : n \ge 1\}$ be the interarrival times of such a process counted from a time at which a meteorite falls. Then $X_1', X_2', \dots$ are independent and distributed as X_2. Let Y_n' be the indicator function of the event $\{X_m' = n \text{ for some } m\}$. Then

$$\begin{aligned}\mathbb{E}(Y_m Y_{m+n}) &= \mathbb{P}(Y_m = 1, Y_{m+n} = 1)\\ &= \mathbb{P}(Y_{m+n} = 1 \mid Y_m = 1)\mathbb{P}(Y_m = 1) = \mathbb{P}(Y_n' = 1)\alpha\end{aligned}$$

where $\alpha = \mathbb{P}(Y_m = 1)$. The autocovariance function of Y is therefore $c(n) = \alpha\{\mathbb{P}(Y_n' = 1) - \alpha\}$, $n \ge 0$, and Y is stationary.

The spectral density function of Y satisfies

$$f_Y(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{-in\lambda}\frac{c(n)}{\alpha(1-\alpha)} = \operatorname{Re}\left\{\frac{1}{\pi\alpha(1-\alpha)}\sum_{n=0}^{\infty} e^{in\lambda}c(n) - \frac{1}{2\pi}\right\}.$$

Now

$$\sum_{n=0}^{\infty} e^{in\lambda}Y_n' = \sum_{n=0}^{\infty} e^{i\lambda T_n'}$$

where $T_n' = X_1' + X_2' + \dots + X_n'$; just check the non-zero terms. Therefore

$$\sum_{n=0}^{\infty} e^{in\lambda}c(n) = \alpha\mathbb{E}\left\{\sum_{n=0}^{\infty} e^{i\lambda T_n'}\right\} - \frac{\alpha^2}{1-e^{i\lambda}} = \frac{\alpha}{1-\phi(\lambda)} - \frac{\alpha^2}{1-e^{i\lambda}}$$

when $e^{i\lambda} \ne 1$, where ϕ is the characteristic function of X_2. It follows that

$$f_Y(\lambda) = \frac{1}{\pi(1-\alpha)}\operatorname{Re}\left\{\frac{1}{1-\phi(\lambda)} - \frac{\alpha}{1-e^{i\lambda}}\right\} - \frac{1}{2\pi}, \qquad |\lambda| < \pi.$$

5. We have that

$$\mathbb{E}\big(\cos(nU)\big) = \int_{-\pi}^{\pi}\frac{1}{2\pi}\cos(nu)\,du = 0, \quad \mathbb{E}\big(\cos^2(nU)\big) = \int_{-\pi}^{\pi}\frac{1}{2\pi}\cos^2(nu)\,du = \tfrac{1}{2},$$

for $n \geq 1$. Also

$$\mathbb{E}\big(\cos(mU)\cos(nU)\big) = \mathbb{E}\Big\{\tfrac{1}{2}\big(\cos[(m+n)U] + \cos[(m-n)U]\big)\Big\} = 0$$

if $m \neq n$. Hence X is stationary with autocorrelation function $\rho(k) = \delta_{k0}$, and spectral density function $f(\lambda) = (2\pi)^{-1}$ for $|\lambda| < \pi$. Finally

$$\begin{aligned}\mathbb{E}\big\{\cos(mU)\cos(nU)\cos(rU)\big\} &= \tfrac{1}{2}\mathbb{E}\Big\{\big(\cos[(m+n)U] + \cos[(m-n)U]\big)\cos(rU)\Big\}\\ &= \tfrac{1}{4}\big\{\rho(m+n-r) + \rho(m-n-r)\big\}\end{aligned}$$

which takes different values in the two cases $(m, n, r) = (1, 2, 3),\ (2, 3, 4)$.

6. (a) The increments of N during any collection of intervals $\{(u_i, v_i) : 1 \leq i \leq n\}$ have the same fdds if all the intervals are shifted by the same constant. Therefore X is strongly stationary. Certainly $\mathbb{E}(X(t)) = \lambda\alpha$ for all t, and the autocovariance function is

$$c(t) = \operatorname{cov}\big(X(0), X(t)\big) = \begin{cases} 0 & \text{if } t > \alpha,\\ \lambda(\alpha - t) & \text{if } 0 \leq t \leq \alpha.\end{cases}$$

Therefore the autocorrelation function is

$$\rho(t) = \begin{cases} 0 & \text{if } |t| > \alpha,\\ 1 - |t/\alpha| & \text{if } |t| \leq \alpha,\end{cases}$$

which we recognize as the characteristic function of the spectral density $f(\lambda) = \{1 - \cos(\alpha\lambda)\}/(\alpha\pi\lambda^2)$; see Problems (5.12.27b, 28a).

(b) We have that $\mathbb{E}(X(t)) = 0$; furthermore, for $s \leq t$, the correlation of $X(s)$ and $X(t)$ is

$$\begin{aligned}\frac{1}{\sigma^2}\operatorname{cov}\big(X(s), X(t)\big) &= \frac{1}{\sigma^2}\operatorname{cov}\big(W(s) - W(s-1), W(t) - W(t-1)\big)\\ &= s - \min\{s, t-1\} - (s-1) + (s-1)\\ &= \begin{cases} 1 & \text{if } s \leq t-1,\\ s - t + 1 & \text{if } t - 1 < s \leq t.\end{cases}\end{aligned}$$

This depends on $t - s$ only, and therefore X is stationary; X is Gaussian and therefore strongly stationary also.

The autocorrelation function is

$$\rho(h) = \begin{cases} 0 & \text{if } |h| \geq 1,\\ 1 - |h| & \text{if } |h| < 1,\end{cases}$$

which we recognize as the characteristic function of the density function $f(\lambda) = (1 - \cos\lambda)/(\pi\lambda^2)$.

7. We have from Problem (8.7.1) that the general moving-average process of part (b) is stationary with autocovariance function $c(k) = \sum_{j=0}^{r} \alpha_j \alpha_{k+j}$, $k \geq 0$, with the convention that $\alpha_s = 0$ if $s < 0$ or $s > r$.

(a) In this case, the autocorrelation function is

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0,\\ \dfrac{\alpha}{1+\alpha^2} & \text{if } |k| = 1,\\ 0 & \text{if } |k| > 1,\end{cases}$$

whence the spectral density function is

$$f(\lambda) = \frac{1}{2\pi}\left(\rho(0) + e^{i\lambda}\rho(1) + e^{-i\lambda}\rho(-1)\right) = \frac{1}{2\pi}\left(1 + \frac{2\alpha\cos\lambda}{1+\alpha^2}\right), \quad |\lambda| < \pi.$$

(b) We have that

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{-ik\lambda}\rho(k) = \frac{1}{2\pi c(0)}\sum_j \alpha_j e^{ij\lambda}\sum_k \alpha_{k+j}e^{-i(k+j)\lambda} = \frac{|A(e^{i\lambda})|^2}{2\pi c(0)}$$

where $c(0) = \sum_j \alpha_j^2$ and $A(z) = \sum_j \alpha_j z^j$. See Problem (9.7.2) also.

8. The spectral density function f is given by the inversion theorem (5.9.1) as

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\rho(t)\,dt$$

under the condition $\int_0^\infty |\rho(t)|\,dt < \infty$; see Problem (5.12.20). Now

$$|f(x)| \le \frac{1}{2\pi}\int_{-\infty}^{\infty}|\rho(t)|\,dt$$

and

$$|f(x+h) - f(x)| \le \frac{1}{2\pi}\int_{-\infty}^{\infty}|e^{ith} - 1|\cdot|\rho(t)|\,dt.$$

The integrand is dominated by the integrable function $2|\rho(t)|$. Using the dominated convergence theorem, we deduce that $|f(x+h) - f(x)| \to 0$ as $h \to 0$, uniformly in x.

9. By Exercise (9.5.2), $\operatorname{var}\left(n^{-1}\sum_{j=1}^n X_j\right) \to \sigma^2$ if $C_n = n^{-1}\sum_{j=0}^{n-1}\operatorname{cov}(X_0, X_j) \to \sigma^2$. If $\operatorname{cov}(X_0, X_n) \to 0$ then $C_n \to 0$, and the result follows.

10. Let $X_1, X_2, \dots$ be independent identically distributed random variables with mean μ. The sequence X is stationary, and it is a consequence of the ergodic theorem that $n^{-1}\sum_{j=1}^n X_j \to Z$ a.s. and in mean, where Z is a tail function of $X_1, X_2, \dots$ with mean μ. Using the zero–one law, Z is a.s. constant, and therefore $\mathbb{P}(Z = \mu) = 1$.

11. We have from the ergodic theorem that $n^{-1}\sum_{i=1}^n Y_i \to \mathbb{E}(Y \mid \mathcal{I})$ a.s. and in mean, where $\mathcal{I}$ is the σ-field of invariant events. The condition of the question is therefore

$$(*) \qquad \mathbb{E}(Y \mid \mathcal{I}) = \mathbb{E}(Y) \quad \text{a.s., for all appropriate } Y.$$

Suppose $(*)$ holds. Pick $A \in \mathcal{I}$, and set $Y = I_A$ to obtain $I_A = \mathbb{Q}(A)$ a.s. Now I_A takes the values 0 and 1, so that $\mathbb{Q}(A)$ equals 0 or 1, implying that $\mathbb{Q}$ is ergodic. Conversely, suppose $\mathbb{Q}$ is ergodic. Then $\mathbb{E}(Y \mid \mathcal{I})$ is measurable on a trivial σ-field, and therefore equals $\mathbb{E}(Y)$ a.s.

12. Suppose $\mathbb{Q}$ is strongly mixing. If A is an invariant event then $A = \tau^{-n}A$. Therefore $\mathbb{Q}(A) = \mathbb{Q}(A \cap \tau^{-n}A) \to \mathbb{Q}(A)^2$ as $n \to \infty$, implying that $\mathbb{Q}(A)$ equals 0 or 1, and therefore $\mathbb{Q}$ is ergodic.

13. The vector $\mathbf{X} = (X_1, X_2, \dots)$ induces a probability measure $\mathbb{Q}$ on $(\mathbb{R}^T, \mathbb{B}^T)$. Since T is measure-preserving, $\mathbb{Q}$ is stationary. Let $Y : \mathbb{R}^T \to \mathbb{R}$ be given by $Y(\mathbf{x}) = x_1$ for $\mathbf{x} = (x_1, x_2, \dots)$, and define $Y_i(\mathbf{x}) = Y(\tau^{i-1}(\mathbf{x}))$ where τ is the usual shift operator on $\mathbb{R}^T$. The vector $\mathbf{Y} = (Y_1, Y_2, \dots)$ has the same distributions as the vector $\mathbf{X}$. By the ergodic theorem for $\mathbf{Y}$, $n^{-1}\sum_{i=1}^n Y_i \to \mathbb{E}(Y \mid \mathcal{J})$ a.s. and in mean, where $\mathcal{J}$ is the invariant σ-field of τ. It follows that the limit

$$(*) \qquad Z = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i$$

exists a.s. and in mean. Now $U = \limsup_{n\to\infty}\left(n^{-1}\sum_1^n X_i\right)$ is invariant, since

$$\frac{1}{n}\left\{\sum_{i=1}^{n}\big(X_i(\omega) - X_i(T\omega)\big)\right\} = \frac{1}{n}\{X(\omega) - X(T^n\omega)\} \to 0 \qquad \text{a.s.},$$

implying that $U(\omega) = U(T\omega)$ a.s. It follows that U is $\mathcal{I}$-measurable, and it is the case that $Z = U$ a.s. Take conditional expectations of $(*)$, given $\mathcal{I}$, to obtain $U = \mathbb{E}(X \mid \mathcal{I})$ a.s.

If T is ergodic, then $\mathcal{I}$ is trivial, so that $\mathbb{E}(X \mid \mathcal{I})$ is a.s. constant; therefore $\mathbb{E}(X \mid \mathcal{I}) = \mathbb{E}(X)$ a.s.

14. If $(a, b) \subseteq [0, 1)$, then $T^{-1}(a, b) = (\frac{1}{2}a, \frac{1}{2}b) \cup (\frac{1}{2} + \frac{1}{2}a, \frac{1}{2} + \frac{1}{2}b)$, and therefore T is measurable. Secondly,

$$\mathbb{P}\big(T^{-1}(a, b)\big) = 2(\tfrac{1}{2}b - \tfrac{1}{2}a) = b - a = \mathbb{P}\big((a, b)\big),$$

so that T^{-1} preserves the measure of intervals. The intervals generate $\mathcal{B}$, and it is then standard that T^{-1} preserves the measures of all events.

Let A be invariant, in that $A = T^{-1}A$. Let $0 \le \omega < \frac{1}{2}$; it is easily seen that $T(\omega) = T(\omega + \frac{1}{2})$. Therefore $\omega \in A$ if and only if $\omega + \frac{1}{2} \in A$, implying that $A \cap [\frac{1}{2}, 1) = \frac{1}{2} + \{A \cap [0, \frac{1}{2})\}$; hence

$$\mathbb{P}(A \cap E) = \tfrac{1}{2}\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(E) \qquad \text{for } E = [0, \tfrac{1}{2}),\ [\tfrac{1}{2}, 1).$$

This proves that A is independent of both $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$. A similar proof gives that A is independent of any set E which is, for some n, the union of intervals of the form $[k2^{-n}, (k+1)2^{-n})$ for $0 \le k < 2^n$. It is a fundamental result of measure theory that there exists a sequence $E_1, E_2, \dots$ of events such that

(a) E_n is of the above form, for each n,

(b) $\mathbb{P}(A \,\triangle\, E_n) \to 0$ as $n \to \infty$.

Choosing the E_n accordingly, it follows that

$$\mathbb{P}(A \cap E_n) = \mathbb{P}(A)\mathbb{P}(E_n) \to \mathbb{P}(A)^2 \qquad \text{by independence,}$$
$$|\mathbb{P}(A \cap E_n) - \mathbb{P}(A)| \le \mathbb{P}(A \,\triangle\, E_n) \to 0.$$

Therefore $\mathbb{P}(A) = \mathbb{P}(A)^2$ so that $\mathbb{P}(A)$ equals 0 or 1.

For $\omega \in \Omega$, expand ω in base 2, $\omega = 0.\omega_1\omega_2\cdots$, and define $Y(\omega) = \omega_1$. It is easily seen that $Y(T^{n-1}\omega) = \omega_n$, whence the ergodic theorem (Problem (9.7.13)) yields that $n^{-1}\sum_{i=1}^n \omega_i \to \frac{1}{2}$ as $n \to \infty$ for all ω in some event of probability 1.

15. We may as well assume that $0 < \alpha < 1$. Let $T : [0, 1) \to [0, 1)$ be given by $T(x) = x + \alpha$ (mod 1). It is easily seen that T is invertible and measure-preserving. Furthermore $T(X)$ is uniform on $[0, 1]$, and it follows that the sequence $Z_1, Z_2, \dots$ has the same fdds as $Z_2, Z_3, \dots$, which is to say that Z is stationary. It therefore suffices to prove that T is an ergodic shift, since this will imply by the ergodic theorem that

$$\frac{1}{n}\sum_{j=1}^{n} Z_j \to \mathbb{E}(Z_1) = \int_0^1 g(u)\,du.$$

We use Fourier analysis. Let A be an invariant subset of $[0, 1)$. The indicator function of A has a Fourier series:

$$(*) \qquad I_A(x) \sim \sum_{n=-\infty}^{\infty} a_n e_n(x)$$

where $e_n(x) = e^{2\pi i n x}$ and

$$a_n = \frac{1}{2\pi}\int_0^1 I_A(x) e_{-n}(x)\,dx = \frac{1}{2\pi}\int_A e_{-n}(x)\,dx.$$

Similarly the indicator function of $T^{-1}A$ has a Fourier series,

$$I_{T^{-1}A}(x) \sim \sum_n b_n e_n(x)$$

where, using the substitution $y = T(x)$,

$$b_n = \frac{1}{2\pi}\int_0^1 I_{T^{-1}A}(x)e_{-n}(x)\,dx = \frac{1}{2\pi}\int_0^1 I_A(y)e_{-n}(T^{-1}(y))\,dy = a_n e^{-2\pi i n\alpha},$$

since $e_m(y-\alpha) = e^{-2\pi i m\alpha}e_m(y)$. Therefore $I_{T^{-1}A}$ has Fourier series

$$I_{T^{-1}A}(x) \sim \sum_n e^{-2\pi i n\alpha} a_n e_n(x).$$

Now $I_A = I_{T^{-1}A}$ since A is invariant. We compare the previous formula with that of $(*)$, and deduce that $a_n = e^{-2\pi i n\alpha}a_n$ for all n. Since α is irrational, it follows that $a_n = 0$ if $n \neq 0$, and therefore I_A has Fourier series a_0, a constant. Therefore I_A is a.s. constant, which is to say that either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

16. Let $G_t(z) = \mathbb{E}(z^{X(t)})$, the probability generating function of $X(t)$. Since X has stationary independent increments, for any n (≥ 1), $X(t)$ may be expressed as the sum

$$X(t) = \sum_{i=1}^{n}\{X(it/n) - X((i-1)t/n)\}$$

of independent identically distributed variables. Hence $X(t)$ is infinitely divisible. By Problem (5.12.13), we may write

$$(*) \qquad G_t(z) = e^{-\lambda(t)(1-A(z))}$$

for some probability generating function A, and some $\lambda(t)$.

Similarly, $X(s+t) = X(s) + \{X(s+t) - X(s)\}$, whence $G_{s+t}(z) = G_s(z)G_t(z)$, implying that $G_t(z) = e^{\mu(z)t}$ for some $\mu(z)$; we have used a little monotonicity here. Combining this with $(*)$, we obtain that $G_t(z) = e^{-\lambda t(1-A(z))}$ for some λ.

Finally, $X(t)$ has jumps of unit magnitude only, whence the probability generating function A is given by $A(z) = z$.

17. (a) We have that

$$(*) \qquad X(t) - X(0) = \{X(s) - X(0)\} + \{X(t) - X(s)\}, \qquad 0 \leq s \leq t,$$

whence, by stationarity,

$$\{m(t) - m(0)\} = \{m(s) - m(0)\} + \{m(t-s) - m(0)\}.$$

Now m is continuous, so that $m(t) - m(0) = \beta t$, $t \geq 0$, for some β; see Problem (4.14.5).
(b) Take variances of $(*)$ to obtain $v(t) = v(s) + v(t-s)$, $0 \leq s \leq t$, whence $v(t) = \sigma^2 t$ for some σ^2.

18. In the context of this chapter, a process Z is a standard Wiener process if it is Gaussian with $Z(0) = 0$, with zero means, and autocovariance function $c(s,t) = \min\{s,t\}$.

(a) $Z(t) = \alpha W(t/\alpha^2)$ satisfies $Z(0) = 0$, $\mathbb{E}(Z(t)) = 0$, and

$$\operatorname{cov}\big(Z(s), Z(t)\big) = \alpha^2 \min\{s/\alpha^2, t/\alpha^2\} = \min\{s, t\}.$$

(b) The only calculation of any interest here is

$$\begin{aligned}\operatorname{cov}\big(W(s+\alpha)-&W(\alpha), W(t+\alpha) - W(\alpha)\big)\\ &= c(s+\alpha, t+\alpha) - c(\alpha, t+\alpha) - c(s+\alpha, \alpha) + c(\alpha, \alpha)\\ &= (s+\alpha) - \alpha - \alpha + \alpha = s, \qquad s \le t.\end{aligned}$$

(c) $V(0) = 0$, and $\mathbb{E}(V(t)) = 0$. Finally, if $s, t > 0$,

$$\operatorname{cov}\big(V(s), V(t)\big) = st\operatorname{cov}\big(W(1/s), W(1/t)\big) = st \min\{1/s, 1/t\} = \min\{t, s\}.$$

(d) $Z(t) = W(1) - W(1-t)$ satisfies $Z(0) = 0$, $\mathbb{E}(Z(t)) = 0$. Also Z is Gaussian, and

$$\begin{aligned}\operatorname{cov}\big(Z(s), Z(t)\big) &= 1 - (1-s) - (1-t) + \min\{1-s, 1-t\}\\ &= \min\{s, t\}, \qquad 0 \le s, t \le 1.\end{aligned}$$

19. The process W has stationary independent increments, and $G(t) = \mathbb{E}(|W(t)|^2)$ satisfies $G(t) = t \to 0$ as $t \to 0$; hence $\int_0^\infty \phi(u)\, dW(u)$ is well defined for any ϕ satisfying

$$\int_0^\infty |\phi(u)|^2 \, dG(u) = \int_0^\infty \phi(u)^2 \, du < \infty.$$

It is obvious that $\phi(u) = I_{[0,t]}(u)$ and $\phi(u) = e^{-(t-u)} I_{[0,t]}(u)$ are such functions.

Now $X(t)$ is the limit (in mean-square) of the sequence

$$S_n(t) = \sum_{j=0}^{n-1} \big\{W((j+1)t/n) - W(jt/n)\big\}, \qquad n \ge 1.$$

However $S_n(t) = W(t)$ for all n, and therefore $S_n(t) \xrightarrow{\text{m.s.}} W(t)$ as $n \to \infty$.

Finally, $Y(s)$ is the limit (in mean-square) of a sequence of normal random variables with mean 0, and therefore is Gaussian with mean 0. If $s < t$,

$$\begin{aligned}\operatorname{cov}\big(Y(s), Y(t)\big) &= \int_0^\infty \big(e^{-(s-u)} I_{[0,s]}(u)\big)\big(e^{-(t-u)} I_{[0,t]}(u)\big)\, dG(u)\\ &= \int_0^s e^{2u-s-t}\, du = \tfrac{1}{2}(e^{s-t} - e^{-s-t}).\end{aligned}$$

Y is an Ornstein–Uhlenbeck process.

20. (a) $W(t)$ is $N(0, t)$, so that

$$\begin{aligned}\mathbb{E}|W(t)| &= \int_{-\infty}^\infty \frac{|u|}{\sqrt{2\pi t}} e^{-\frac{1}{2}(u^2/t)}\, du = \sqrt{2t/\pi},\\ \operatorname{var}(|W(t)|) &= \mathbb{E}(W(t)^2) - \frac{2t}{\pi} = t\left(1 - \frac{2}{\pi}\right).\end{aligned}$$

The process X is never negative, and therefore it is not Gaussian. It is Markov since, if $s < t$ and B is an event defined in terms of $\{X(u) : u \le s\}$, then the conditional distribution function of $X(t)$ satisfies

$$\begin{aligned}\mathbb{P}\big(X(t) \le y \mid X(s) = x, B\big) &= \mathbb{P}\big(X(t) \le y \mid W(s) = x, B\big)\mathbb{P}\big(W(s) = x \mid X(s) = x, B\big)\\ &\quad + \mathbb{P}\big(X(t) \le y \mid W(s) = -x, B\big)\mathbb{P}\big(W(s) = -x \mid X(s) = x, B\big)\\ &= \tfrac{1}{2}\Big\{\mathbb{P}\big(X(t) \le y \mid W(s) = x\big) + \mathbb{P}\big(X(t) \le y \mid W(s) = -x\big)\Big\},\end{aligned}$$

which does not depend on B.

(b) Certainly,

$$\mathbb{E}(Y(t)) = \int_{-\infty}^{\infty} \frac{e^u}{\sqrt{2\pi t}} e^{-\frac{1}{2}(u^2/t)}\,du = e^{\frac{1}{2}t}.$$

Secondly, $W(s) + W(t) = 2W(s) + \{W(t) - W(s)\}$ is $N(0, 3s + t)$ if $s < t$, implying that

$$\mathbb{E}\big(Y(s)Y(t)\big) = \mathbb{E}\left(e^{W(s)+W(t)}\right) = e^{\frac{1}{2}(3s+t)},$$

and therefore

$$\operatorname{cov}\big(Y(s), Y(t)\big) = e^{\frac{1}{2}(3s+t)} - e^{\frac{1}{2}(s+t)}, \qquad s < t.$$

$W(1)$ is $N(0, 1)$, and therefore $Y(1)$ has the log-normal distribution. Therefore Y is not Gaussian. It is Markov since W is Markov, and $Y(t)$ is a one–one function of $W(t)$.

(c) We shall assume that the random function W is a.s. continuous, a point to which we return in Chapter 13. Certainly,

$$\begin{aligned}\mathbb{E}(Z(t)) &= \int_0^t \mathbb{E}(W(u))\,du = 0,\\ \mathbb{E}\big(Z(s)Z(t)\big) &= \iint_{\substack{0\le u\le s\\ 0\le v\le t}} \mathbb{E}\big(W(u)W(v)\big)\,du\,dv\\ &= \int_{u=0}^{s}\left\{\int_{v=0}^{u} v\,dv + \int_{v=u}^{t} u\,dv\right\}du = \tfrac{1}{6}s^2(3t - s), \qquad s < t,\end{aligned}$$

since $\mathbb{E}(W(u)W(v)) = \min\{u, v\}$.

Z is Gaussian, as the following argument indicates. The single random variable $Z(t)$ may be expressed as a limit of the form

$$(*) \qquad \lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{t}{n}\right) W(it/n),$$

each such summation being normal. The limit of normal variables is normal (see Problem (7.11.19)), and therefore $Z(t)$ is normal. The limit in $(*)$ exists a.s., and hence in probability. By an appeal to (7.11.19b), pairs $(Z(s), Z(t))$ are bivariate normal, and a similar argument is valid for all n-tuples of the $Z(u)$.

The process Z is not Markov. An increment $Z(t) - Z(s)$ depends very much on $W(s) = Z'(s)$, and the collection $\{Z(u) : u \le s\}$ contains much information about $Z'(s)$ in excess of the information contained in the single value $Z(s)$.

21. Let $U_i = X(t_i)$. The random variables $A = U_1$, $B = U_2 - U_1$, $C = U_3 - U_2$, $D = U_4 - U_3$ are independent and normal with zero means and respective variances $t_1, t_2 - t_1, t_3 - t_2, t_4 - t_3$. The Jacobian of the transformation is 1, and it follows that U_1, U_2, U_3, U_4 have joint density function

$$f_{\mathbf{U}}(\mathbf{u}) = \frac{e^{-\frac{1}{2}Q}}{(2\pi)^2\sqrt{t_1(t_2 - t_1)(t_3 - t_2)(t_4 - t_3)}}$$

where

$$Q = \frac{u_1^2}{t_1} + \frac{(u_2 - u_1)^2}{t_2 - t_1} + \frac{(u_3 - u_2)^2}{t_3 - t_2} + \frac{(u_4 - u_3)^2}{t_4 - t_3}.$$

Likewise U_1 and U_4 have joint density function

$$\frac{e^{-\frac{1}{2}R}}{2\pi\sqrt{t_1(t_4 - t_1)}} \quad \text{where} \quad R = \frac{u_1^2}{t_1} + \frac{(u_4 - u_1)^2}{t_4 - t_1}.$$

Hence the joint density function of U_2 and U_3, given $U_1 = U_4 = 0$, is

$$g(u_2, u_3) = \frac{e^{-\frac{1}{2}S}}{2\pi}\sqrt{\frac{t_4 - t_1}{(t_2 - t_1)(t_3 - t_2)(t_4 - t_3)}}$$

where

$$S = \frac{u_2^2}{t_2 - t_1} + \frac{(u_3 - u_2)^2}{t_3 - t_2} + \frac{u_3^2}{t_4 - t_3}.$$

Now g is the density function of a bivariate normal distribution with zero means, marginal variances

$$\sigma_1^2 = \frac{(t_2 - t_1)(t_3 - t_2)}{t_3 - t_1}, \qquad \sigma_2^2 = \frac{(t_4 - t_3)(t_3 - t_2)}{t_4 - t_2}$$

and correlation

$$\rho = \frac{\sigma_1\sigma_2}{t_3 - t_2} = \sqrt{\frac{(t_4 - t_3)(t_2 - t_1)}{(t_4 - t_2)(t_3 - t_1)}}.$$

See also Exercise (8.5.2).

22. (a) The random variables $\{I_j(x) : 1 \le j \le n\}$ are independent, so that

$$\mathbb{E}(F_n(x)) = x, \qquad \operatorname{var}(F_n(x)) = \frac{1}{n}\operatorname{var}(I_1(x)) = \frac{x(1-x)}{n}.$$

By the central limit theorem, $\sqrt{n}\{F_n(x) - x\} \xrightarrow{\text{D}} Y(x)$, where $Y(x)$ is $N(0, x(1-x))$.

(b) The limit distribution is multivariate normal. There are general methods for showing this, and here is a sketch. If $0 \le x_1 < x_2 \le 1$, then the number M_2 $(= nF(x_2))$ of the I_j not greater than x_2 is approximately $N(nx_2, nx_2(1 - x_2))$. Conditional on $\{M_2 = m\}$, the number $M_1 = nF(x_1)$ is approximately $N(mu, mu(1 - u))$ where $u = x_1/x_2$. It is now a small exercise to see that the pair (M_1, M_2) is approximately bivariate normal with means nx_1, nx_2, with variances $nx_1(1 - x_1)$, $nx_2(1 - x_2)$, and such that

$$\mathbb{E}(M_1M_2) = \mathbb{E}\{M_2\mathbb{E}(M_1 \mid M_2)\} = \mathbb{E}(M_2^2 x_1/x_2) \sim nx_1(1 - x_2) + n^2x_1x_2,$$

whence $\operatorname{cov}(M_1, M_2) \sim nx_1(1 - x_2)$. It follows similarly that the limit of the general collection is multivariate normal with mean 0, variances $x_i(1 - x_i)$, and covariances $c_{ij} = x_i(1 - x_j)$.

(c) The autocovariance function of the limit distribution is $c(s,t) = \min\{s,t\} - st$, whereas, for $0 \le s \le t \le 1$, we have that $\operatorname{cov}(Z(s), Z(t)) = s - ts - st + st = \min\{s,t\} - st$. It may be shown that the limit of the process $\{\sqrt{n}(F_n(x) - x) : n \ge 1\}$ exists as $n \to \infty$, in a certain sense, the limit being a Brownian bridge; such a limit theorem for processes is called a 'functional limit theorem'.

10

Renewals

10.1 Solutions. The renewal equation

1. Since $\mathbb{E}(X_1) > 0$, there exists ϵ (> 0) such that $\mathbb{P}(X_1 \geq \epsilon) > \epsilon$. Let $X'_k = \epsilon I_{\{X_k \geq \epsilon\}}$, and denote by N' the related renewal process. Now $N(t) \leq N'(t)$, so that $\mathbb{E}(e^{\theta N(t)}) \leq \mathbb{E}(e^{\theta N'(t)})$, for $\theta > 0$. Let Z_m be the number of renewals (in N') between the times at which N' reaches the values $(m-1)\epsilon$ and $m\epsilon$. The Z's are independent with

$$\mathbb{E}(e^{\theta Z_m}) = \frac{\epsilon e^\theta}{1-(1-\epsilon)e^\theta}, \qquad \text{if } (1-\epsilon)e^\theta < 1,$$

whence $\mathbb{E}(e^{\theta N'(t)}) \leq \big(\epsilon e^\theta \{1-(1-\epsilon)e^\theta\}^{-1}\big)^{t/\epsilon}$ for sufficiently small positive θ.

2. Let X_1 be the time of the first arrival. If $X_1 > s$, then $W = s$. On the other hand if $X_1 < s$, then the process starts off afresh at the new starting time X_1. Therefore, by conditioning on the value of X_1,

$$\begin{aligned} F_W(x) &= \int_0^\infty \mathbb{P}(W \leq x \mid X_1 = u)\,dF(u) = \int_0^s \mathbb{P}(W \leq x-u)\,dF(u) + \int_s^\infty 1 \cdot dF(u) \\ &= \int_0^s \mathbb{P}(W \leq x-u)\,dF(u) + \{1-F(s)\} \end{aligned}$$

if $x \geq s$. It is clear that $F_W(x) = 0$ if $x < s$. This integral equation for F_W may be written in the standard form

$$F_W(x) = H(x) + \int_0^x F_W(x-u)\,d\widehat{F}(u)$$

where H and $\widehat{F}$ are given by

$$H(x) = \begin{cases} 0 & \text{if } x < s, \\ 1-F(s) & \text{if } x \geq s, \end{cases} \qquad \widehat{F}(x) = \begin{cases} F(x) & \text{if } x < s, \\ F(s) & \text{if } x \geq s. \end{cases}$$

This renewal-type equation may be solved in the usual way by the method of Laplace–Stieltjes transforms. We have that $F_W^* = H^* + F_W^* \widehat{F}^*$, whence $F_W^* = H^*/(1-\widehat{F}^*)$. If N is a Poisson process then $F(x) = 1 - e^{-\lambda x}$. In this case

$$H^*(\theta) = \int_0^\infty e^{-\theta x}\,dH(x) = e^{-(\lambda+\theta)s},$$

since H is constant apart from a jump at $x = s$. Similarly

$$\widehat{F}^*(\theta) = \int_0^s e^{-\theta x}\,dF(x) = \frac{\lambda}{\lambda+\theta}\big(1-e^{-(\lambda+\theta)s}\big),$$

so that

$$F_W^*(\theta) = \frac{(\lambda+\theta)e^{-(\lambda+\theta)s}}{\theta + \lambda e^{-(\lambda+\theta)s}}.$$

Finally, replace θ with $-\theta$, and differentiate to find the mean.

3. We have as usual that $\mathbb{P}(N(t) = n) = \mathbb{P}(S_n \le t) - \mathbb{P}(S_{n+1} \le t)$. In the respective cases,

(a) $$\mathbb{P}(N(t) = n) = \sum_{r=0}^{\lfloor t \rfloor} \frac{1}{r!}\left\{e^{-\lambda n}(\lambda n)^r - e^{-\lambda(n+1)}[\lambda(n+1)]^r\right\},$$

(b) $$\mathbb{P}(N(t) = n) = \int_0^t \left\{ \frac{\lambda^{nb} x^{nb-1}}{\Gamma(nb)} - \frac{\lambda^{(n+1)b} x^{(n+1)b-1}}{\Gamma((n+1)b)} \right\} e^{-\lambda x}\, dx.$$

4. By conditioning on X_1, $m(t) = \mathbb{E}(N(t))$ satisfies

$$m(t) = \int_0^t \big(1 + m(t-x)\big)\, dx = t + \int_0^t m(x)\, dx, \qquad 0 \le t \le 1.$$

Hence $m' = 1 + m$, with solution $m(t) = e^t - 1$, for $0 \le t \le 1$. (For larger values of t, $m(t) = 1 + \int_0^1 m(t-x)\, dx$, and a tiresome iteration is in principle possible.)

With $v(t) = \mathbb{E}(N(t)^2)$,

$$v(t) = \int_0^t \big[v(t-x) + 2m(t-x) + 1\big]\, dx = t + 2(e^t - t - 1) + \int_0^t v(x)\, dx, \quad 0 \le t \le 1.$$

Hence $v' = v + 2e^t - 1$, with solution $v(t) = 1 - e^t + 2te^t$ for $0 \le t \le 1$.

10.2 Solutions. Limit theorems

1. Let Z_i be the number of passengers in the ith plane, and assume that the Z_i are independent of each other and of the arrival process. The number of passengers who have arrived by time t is $S(t) = \sum_{i=1}^{N(t)} Z_i$. Now

$$\frac{1}{t}S(t) = \frac{N(t)}{t} \cdot \frac{S(t)}{N(t)} \to \frac{\mathbb{E}(Z_1)}{\mu} \qquad \text{a.s.}$$

by the law of the large numbers, since $N(t)/t \to 1/\mu$ a.s., and $N(t) \to \infty$ a.s.

2. We have that

$$\mathbb{E}(T_M^2) = \mathbb{E}\left\{\left(\sum_{i=1}^{\infty} Z_i I_{\{M \ge i\}}\right)^2\right\} = \sum_{i=1}^{\infty} \mathbb{E}\big(Z_i^2 I_{\{M\ge i\}}\big) + 2 \sum_{1 \le i < j < \infty} \mathbb{E}\big(Z_i Z_j I_{\{M \ge j\}}\big)$$

since $I_{\{M\ge i\}} I_{\{M \ge j\}} = I_{\{M \ge i \vee j\}}$, where $i \vee j = \max\{i, j\}$. Now

$$\mathbb{E}\big(Z_i^2 I_{\{M\ge i\}}\big) = \mathbb{E}\big(Z_i^2 I_{\{M \le i-1\}^c}\big) = \mathbb{E}(Z_i^2)\mathbb{P}(M \ge i),$$

since $\{M \le i-1\}$ is defined in terms of $Z_1, Z_2, \dots, Z_{i-1}$, and is therefore independent of Z_i. Similarly $\mathbb{E}(Z_i Z_j I_{\{M\ge j\}}) = \mathbb{E}(Z_j)\mathbb{E}(Z_i I_{\{M \ge j\}}) = 0$ if $i < j$. It follows that

$$\mathbb{E}(T_M^2) = \sum_{i=1}^{\infty} \mathbb{E}(Z_i^2)\mathbb{P}(M \ge i) = \sigma^2 \sum_{i=1}^{\infty} \mathbb{P}(M \ge i) = \sigma^2 \mathbb{E}(M).$$

3. (i) The shortest way is to observe that $N(t) + k$ is a stopping time if $k \geq 1$. Alternatively, we have by Wald's equation that $\mathbb{E}(T_{N(t)+1}) = \mu(m(t) + 1)$. Also

$$\mathbb{E}(X_{N(t)+k}) = \mathbb{E}\{\mathbb{E}(X_{N(t)+k} \mid N(t))\} = \mu, \qquad k \geq 2,$$

and therefore, for $k \geq 1$,

$$\mathbb{E}(T_{N(t)+k}) = \mathbb{E}(T_{N(t)+1}) + \sum_{j=2}^{k} \mathbb{E}(X_{N(t)+j}) = \mu\big(m(t) + k\big).$$

(ii) Suppose $p \neq 1$ and

$$\mathbb{P}(X_1 = a) = \begin{cases} p & \text{if } a = 1, \\ 1 - p & \text{if } a = 2. \end{cases}$$

Then $\mu = 2 - p \neq 1$. Also

$$\mathbb{E}(T_{N(1)}) = (1 - p)\mathbb{E}(T_0 \mid N(1) = 0) + p\mathbb{E}(T_1 \mid N(1) = 1) = p,$$

whereas $m(1) = p$. Therefore $\mathbb{E}(T_{N(1)}) \neq \mu m(1)$.

4. Let $V(t) = N(t) + 1$, and let $W_1, W_2, \ldots$ be defined inductively as follows. $W_1 = V(1)$, W_2 is obtained similarly to W_1 but relative to the renewal process starting at the $V(1)$th renewal, i.e., at time $T_{N(1)+1}$, and W_n is obtained similarly:

$$W_n = N(T_{X_{n-1}} + 1) - N(T_{X_{n-1}}) + 1, \qquad n \geq 2,$$

where $X_m = W_1 + W_2 + \cdots + W_m$. For each n, W_n is independent of the sequence $W_1, W_2, \ldots, W_{n-1}$, and therefore the W_n are independent copies of $V(1)$. It is easily seen, by measuring the time-intervals covered, that $V(t) \leq \sum_{i=1}^{\lceil t \rceil} W_i$, and hence

$$\frac{1}{t} V(t) \leq \frac{1}{t} \sum_{i=1}^{\lceil t \rceil} W_i \to \mathbb{E}(V(1)) \qquad \text{a.s. and in mean, as } t \to \infty.$$

It follows that the family $\{m^{-1} \sum_{i=1}^{m} W_i : m \geq 1\}$ is uniformly integrable (see Theorem (7.10.3)). Now $N(t) \leq V(t)$, and so $\{N(t)/t : t \geq 0\}$ is uniformly integrable also.

Since $N(t)/t \xrightarrow{\text{a.s.}} \mu^{-1}$, it follows by uniform integrability that there is also convergence in mean.

5. (a) Using the fact that $\mathbb{P}(N(t) = k) = \mathbb{P}(S_k \leq t) - \mathbb{P}(S_{k+1} \leq t)$, we find that

$$\begin{aligned} \mathbb{E}(s^{N(T)}) &= \int_0^\infty \bigg(\sum_{k=0}^{\infty} s^k \mathbb{P}(N(t) = k)\bigg) \nu e^{-\nu t}\, dt \\ &= \sum_{k=0}^{\infty} s^k \left\{ \int_0^\infty \big[\mathbb{P}(S_k \leq t) - \mathbb{P}(S_{k+1} \leq t)\big] \nu e^{-\nu t}\, dt \right\}. \end{aligned}$$

By integration by parts, $\int_0^\infty \mathbb{P}(S_k \leq t)\nu e^{-\nu t}\, dt = M(-\nu)^k$ for $k \geq 0$. Therefore,

$$\mathbb{E}(s^{N(T)}) = \sum_{k=0}^{\infty} s^k \{M(-\nu)^k - M(-\nu)^{k+1}\} = \frac{1 - M(-\nu)}{1 - sM(-\nu)}.$$

(b) In this case, $\mathbb{E}(s^{N(T)}) = \mathbb{E}(e^{\lambda T(s-1)}) = M_T(\lambda(s-1))$. When T has the given gamma distribution, $M_T(\theta) = \{\nu/(\nu - \theta)\}^b$, and

$$\mathbb{E}(s^{N(T)}) = \left(\frac{\nu}{\nu + \lambda}\right)^b \left(1 - \frac{\lambda s}{\nu + \lambda}\right)^b.$$

The coefficient of s^k may be found by use of the binomial theorem.

10.3 Solutions. Excess life

1. Let $g(y) = \mathbb{P}(E(t) > y)$, assumed not to depend on t. By the integral equation for the distribution of $E(t)$,

$$g(y) = 1 - F(t+y) + g(y)\int_0^t dF(x).$$

Write $h(x) = 1 - F(x)$ to obtain $g(y)h(t) = h(t+y)$, for $y, t \geq 0$. With $t = 0$, we have that $g(y)h(0) = h(y)$, whence $g(y) = h(y)/h(0)$ satisfies $g(t+y) = g(t)g(y)$, for $y, t \geq 0$. Now g is left-continuous, and we deduce as usual that $g(t) = e^{-\lambda t}$ for some λ. Hence $F(t) = 1 - e^{-\lambda t}$, and the renewal process is a Poisson process.

2. (a) Examine a sample path of E. If $E(t) = x$, then the sample path decreases (with slope -1) until it reaches the value 0, at which point it jumps to a height X, where X is the next interarrival time. Since X is independent of all previous interarrival times, the process is Markovian.

(b) In contrast, C has sample paths which increase (with slope 1) until a renewal occurs, at which they drop to 0. If $C(s) = x$ and, in addition, we know the entire history of the process up to time s, the time of the next renewal depends only on the length of the spent period (i.e., x) of the interarrival time in process. Hence C is Markovian.

3. (a) We have that

$$(*) \qquad \mathbb{P}\big(E(t) \leq y\big) = F(t+y) - \int_0^t G(t+y-x)\,dm(x)$$

where $G(u) = 1 - F(u)$. Check the conditions of the key renewal theorem (10.2.7): $g(t) = G(t+y)$ satisfies:

(i) $g(t) \geq 0$,

(ii) $\int_0^\infty g(t)\,dt \leq \int_0^\infty [1 - F(u)]\,du = \mathbb{E}(X_1) < \infty$,

(iii) g is non-increasing.

We conclude, by that theorem, that

$$\lim_{t\to\infty} \mathbb{P}\big(E(t) \leq y\big) = 1 - \frac{1}{\mu}\int_0^\infty g(x)\,dx = \int_0^y \frac{1}{\mu}[1 - F(x)]\,dx.$$

(b) Integrating by parts,

$$\int_0^\infty \frac{x^r}{\mu}[1 - F(x)]\,dx = \frac{1}{\mu}\int_0^\infty \frac{x^{r+1}}{r+1}\,dF(x) = \frac{\mathbb{E}(X_1^{r+1})}{\mu(r+1)}.$$

See Exercise (4.3.3).

(c) As in Exercise (4.3.3), we have that $\mathbb{E}(E(t)^r) = \int_0^\infty ry^{r-1}\mathbb{P}(E(t) > y)\,dy$, implying by $(*)$ that

$$\mathbb{E}(E(t)^r) = \mathbb{E}\big(\{(X_1 - t)^+\}^r\big) + \int_{y=0}^\infty \int_{x=0}^t ry^{r-1}\mathbb{P}(X_1 > t + y - x)\,dm(x)\,dy,$$

whence the given integral equation is valid with

$$h(u) = \int_0^\infty ry^{r-1}\mathbb{P}(X_1 > u + y)\,dy = \mathbb{E}\big(\{(X_1 - u)^+\}^r\big).$$

Now h satisfies the conditions of the key renewal theorem, whence

$$\lim_{t\to\infty} \mathbb{E}(E(t)^r) = \frac{1}{\mu}\int_0^\infty h(u)\,du = \frac{1}{\mu}\iint_{0<u,y<\infty} y^r\,dF(u+y)\,du$$
$$= \frac{1}{\mu}\int_0^\infty y^r\mathbb{P}(X_1 > y)\,dy = \frac{\mathbb{E}(X_1^{r+1})}{\mu(r+1)}.$$

4. We have that

$$\mathbb{P}\big(E(t) > y \,\big|\, C(t) = x\big) = \mathbb{P}(X_1 > y + x \mid X_1 > x) = \frac{1 - F(y+x)}{1 - F(x)},$$

whence

$$\mathbb{E}\big(E(t) \,\big|\, C(t) = x\big) = \int_0^\infty \frac{1 - F(y+x)}{1 - F(x)}\, dy = \frac{\mathbb{E}\{(X_1 - x)^+\}}{1 - F(x)}.$$

5. (a) Apply Exercise (10.2.2) to the sequence $X_i - \mu$, $1 \le i < \infty$, to obtain $\operatorname{var}(T_{M(t)} - \mu M(t)) = \sigma^2 \mathbb{E}(M(t))$.

(b) Clearly $T_{M(t)} = t + E(t)$, where E is excess lifetime, and hence $\mu M(t) = (t + E(t)) - (T_{M(t)} - \mu M(t))$, implying in turn that

$$(*) \qquad \mu^2 \operatorname{var}(M(t)) = \operatorname{var}(E(t)) + \operatorname{var}(S_{M(t)}) - 2\operatorname{cov}\big(E(t), S_{M(t)}\big),$$

where $S_{M(t)} = T_{M(t)} - \mu M(t)$. Now

$$\operatorname{var}(E(t)) \le \mathbb{E}(E(t)^2) \to \frac{\mathbb{E}(X_1^3)}{3\mu} \qquad \text{as } t \to \infty$$

if $\mathbb{E}(X_1^3) < \infty$ (see Exercise (10.3.3c)), implying that

$$(**) \qquad \frac{1}{t} \operatorname{var}(E(t)) \to 0 \qquad \text{as } t \to \infty.$$

This is valid under the weaker assumption that $\mathbb{E}(X_1^2) < \infty$, as the following argument shows. By Exercise (10.3.3c),

$$\mathbb{E}(E(t)^2) = \alpha(t) + \int_0^t \alpha(t-u)\, dm(u),$$

where $\alpha(u) = \mathbb{E}(\{(X_1 - u)^+\}^2)$. Now use the key renewal theorem together with the fact that $\alpha(t) \le \mathbb{E}(X_1^2 I_{\{X_1 > t\}}) \to 0$ as $t \to \infty$.

Using the Cauchy–Schwarz inequality,

$$\frac{1}{t}\big|\operatorname{cov}\big(E(t), S_{M(t)}\big)\big| \le \frac{1}{t}\sqrt{\operatorname{var}(E(t))\operatorname{var}(S_{M(t)})} \to 0$$

as $t \to \infty$, by part (a) and $(**)$. Returning to $(*)$, we have that

$$\frac{\mu^2}{t}\operatorname{var}(M(t)) \to \lim_{t\to\infty}\left\{\frac{\sigma^2}{t}\big(m(t) + 1\big)\right\} = \frac{\sigma^2}{\mu}.$$

10.4 Solution. Applications

1. Visualize a renewal as arriving after two stages, type 1 stages being exponential parameter λ and type 2 stages being exponential parameter μ. The 'stage' process is the flip–flop two-state Markov process of Exercise (6.9.1). With an obvious notation,

$$p_{11}(t) = \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu}.$$

Hence the excess lifetime distribution is a mixture of the exponential distribution with parameter μ, and the distribution of the sum of two exponential random variables, thus,

$$f_{E(t)}(x) = p_{11}(t)g(x) + (1 - p_{11}(t))\mu e^{-\mu x},$$

where $g(x)$ is the density function of a typical interarrival time. By Wald's equation,

$$\mathbb{E}(t + E(t)) = \mathbb{E}(S_{N(t)+1}) = \mathbb{E}(X_1)\mathbb{E}(N(t)+1) = \left(\frac{1}{\lambda} + \frac{1}{\mu}\right)(m(t)+1).$$

We substitute

$$\mathbb{E}(E(t)) = p_{11}(t)\left(\frac{1}{\lambda} + \frac{1}{\mu}\right) + (1 - p_{11}(t))\frac{1}{\mu} = \frac{1}{\mu} + \frac{p_{11}(t)}{\lambda}$$

to obtain the required expression.

10.5 Solutions. Renewal–reward processes

1. Suppose, at time s, you are paid a reward at rate $u(X(s))$. By Theorem (10.5.10), equation (10.5.7), and Exercise (6.9.11b),

$$(*) \qquad \frac{1}{t}\int_0^t I_{\{X(s)=j\}}\,ds \xrightarrow{\text{a.s.}} \frac{1}{\mu_j g_j} = \pi_j.$$

Suppose $|u(i)| \le K < \infty$ for all $i \in S$, and let F be a finite subset of the state space. Then

$$\begin{aligned}\left|\frac{1}{t}\int_0^t u(X(s))\,ds - \sum_i \pi_i u(i)\right| &= \left|\sum_i u(i)\left(\frac{1}{t}\int_0^t I_{\{X(s)=i\}}\,ds - \pi_i\right)\right| \\ &\le K\sum_{i\in F}\left|\frac{1}{t}\int_0^t I_{\{X(s)=i\}}\,ds - \pi_i\right| + K\left(\frac{t - T_t(F)}{t}\right) + K\sum_{i\notin F}\pi_i,\end{aligned}$$

where $T_t(F)$ is the total time spent in F up to time t. Take the limit as $t \to \infty$ using $(*)$, and then as $F \uparrow S$, to obtain the required result.

2. Suppose you are paid a reward at unit rate during every interarrival time of type X, i.e., at all times t at which $M(t)$ is even. By the renewal–reward theorem (10.5.1),

$$\frac{1}{t}\int_0^t I_{\{M(s)\text{ is even}\}}\,ds \xrightarrow{\text{a.s.}} \frac{\mathbb{E}(\text{reward during interarrival time})}{\mathbb{E}(\text{length of interarrival time})} = \frac{\mathbb{E}X_1}{\mathbb{E}X_1 + \mathbb{E}Y_1}.$$

3. Suppose, at time t, you are paid a reward at rate $C(t)$. The expected reward during an interval (cycle) of length X is $\int_0^X s\,ds = \frac{1}{2}X^2$, since the age C is the same at the time s into the interval. The result follows by the renewal–reward theorem (10.5.1) and equation (10.5.7). The same conclusion is valid for the excess lifetime $E(s)$, the integral in this case being $\int_0^X (X-s)\,ds = \frac{1}{2}X^2$.

4. Suppose $X_0 = j$. Let $V_1 = \min\{n \geq 1 : X_n = j,\ X_m = k \text{ for some } 1 \leq m < n\}$, the first visit to j *subsequent to* a visit to k, and let $V_{r+1} = \min\{n \geq V_r : X_n = j,\ X_m = k \text{ for some } V_r + 1 \leq m < n\}$. The V_r are the times of a renewal process. Suppose a reward of one ecu is paid at every visit to k. By the renewal–reward theorem and equation (10.5.7),

$$(*) \qquad \pi_k = \frac{1}{\mathbb{E}(V_1 \mid X_0 = j)} \mathbb{E}\left(\sum_{m=1}^{V_1 - 1} I_{\{X_m = k\}}\right).$$

By considering the time of the first visit to k,

$$\mathbb{E}(V_1 \mid X_0 = j) = \mathbb{E}(T_k \mid X_0 = j) + \mathbb{E}(T_j \mid X_0 = k).$$

The latter expectation in $(*)$ is the mean of a random variable N having the geometric distribution $\mathbb{P}(N = n) = p(1-p)^{n-1}$ for $n \geq 1$, where $p = \mathbb{P}(T_j < T_k \mid X_0 = k)$. Since $\mathbb{E}(N) = p^{-1}$, we deduce as required that

$$\pi_k = \frac{1/\mathbb{P}(T_j < T_k \mid X_0 = k)}{\mathbb{E}(T_k \mid X_0 = j) + \mathbb{E}(T_j \mid X_0 = k)}.$$

10.6 Solutions to problems

1. (a) For any n, $\mathbb{P}(N(t) < n) \leq \mathbb{P}(T_n > t) \to 0$ as $t \to \infty$.
(b) Either use Exercise (10.1.1), or argue as follows. Since $\mu > 0$, there exists ϵ (> 0) such that $\mathbb{P}(X_1 > \epsilon) > 0$. For all n,

$$\mathbb{P}(T_n \leq n\epsilon) = 1 - \mathbb{P}(T_n > n\epsilon) \leq 1 - \mathbb{P}(X_1 > \epsilon)^n < 1,$$

so that, if $t > 0$, there exists $n = n(t)$ such that $\mathbb{P}(T_n \leq t) < 1$.

Fix t and let n be chosen accordingly. Any positive integer k may be expressed in the form $k = \alpha n + \beta$ where $0 \leq \beta < n$. Now $\mathbb{P}(T_k \leq t) \leq \mathbb{P}(T_n \leq t)^\alpha$ for $\alpha n \leq k < (\alpha+1)n$, and hence

$$m(t) = \sum_{k=1}^{\infty} \mathbb{P}(T_k \leq t) \leq \sum_{\alpha=0}^{\infty} n\mathbb{P}(T_n \leq t)^\alpha < \infty.$$

(c) It is easiest to use Exercise (10.1.1), which implies the stronger conclusion that the moment generating function of $N(t)$ is finite in a neighbourhood of the origin.

2. (i) Condition on X_1 to obtain

$$v(t) = \int_0^t \mathbb{E}\{(N(t-u)+1)^2\}\,dF(u) = \int_0^t \{v(t-u) + 2m(t-u) + 1\}\,dF(u).$$

Take Laplace–Stieltjes transforms to find that $v^* = (v^* + 2m^* + 1)F^*$, where $m^* = F^* + m^*F^*$ as usual. Therefore $v^* = m^*(1 + 2m^*)$, which may be inverted to obtain the required integral equation.
(ii) If N is a Poisson process with intensity λ, then $m(t) = \lambda t$, and therefore $v(t) = (\lambda t)^2 + \lambda t$.

3. Fix $x \in \mathbb{R}$. Then

$$\mathbb{P}\left(\frac{N(t) - (t/\mu)}{\sqrt{t\sigma^2/\mu^3}} \geq x\right) = \mathbb{P}\left(N(t) \geq (t/\mu) + x\sqrt{t\sigma^2/\mu^3}\right) = \mathbb{P}(T_{a(t)} \leq t)$$

where $a(t) = \lfloor (t/\mu) + x\sqrt{t\sigma^2/\mu^3} \rfloor$. Now,

$$\mathbb{P}(T_{a(t)} \le t) = \mathbb{P}\left(\frac{T_{a(t)} - \mu a(t)}{\sigma\sqrt{a(t)}} \le \frac{t - \mu a(t)}{\sigma\sqrt{a(t)}} \right).$$

However $a(t) \sim t/\mu$ as $t \to \infty$, and therefore

$$\frac{t - \mu a(t)}{\sigma\sqrt{a(t)}} \to -x \qquad \text{as } t \to \infty,$$

implying by the usual central limit theorem that

$$\mathbb{P}\left(\frac{N(t) - (t/\mu)}{\sqrt{t\sigma^2/\mu^3}} \ge x \right) \to \Phi(-x) \qquad \text{as } t \to \infty$$

where Φ is the $N(0, 1)$ distribution function.

An alternative proof makes use of Anscombe's theorem (7.11.28).

4. We have that, for $y \le t$,

$$\mathbb{P}\big(C(t) \ge y\big) = \mathbb{P}\big(E(t-y) > y\big) \to \lim_{u\to\infty} \mathbb{P}\big(E(u) > y\big) \qquad \text{as } t \to \infty$$
$$= \int_y^\infty \frac{1}{\mu}[1 - F(x)]\,dx$$

by Exercise (10.3.3a). The current and excess lifetimes have the same asymptotic distributions.

5. Using the lack-of-memory property of the Poisson process, the current lifetime $C(t)$ is independent of the excess lifetime $E(t)$, the latter being exponentially distributed with parameter λ. To derive the density function of $C(t)$ either solve (without difficulty in this case) the relevant integral equation, or argue as follows. Looking *backwards* in time from t, the arrival process looks like a Poisson process up to distance t (at the origin) where it stops. Therefore $C(t)$ may be expressed as $\min\{Z, t\}$ where Z is exponential with parameter λ; hence

$$f_{C(t)}(s) = \begin{cases} \lambda e^{-\lambda s} & \text{if } s \le t, \\ 0 & \text{if } s > t, \end{cases}$$

and $\mathbb{P}(C(t) = t) = e^{-\lambda t}$. Now $D(t) = C(t) + E(t)$, whose distribution is easily found (by the convolution formula) to be as given.

6. The ith interarrival time may be expressed in the form $T + Z_i$ where Z_i is exponential with parameter λ. In addition, $Z_1, Z_2, \ldots$ are independent, by the lack-of-memory property. Now

$$1 - \widetilde{F}(x) = \mathbb{P}(T + Z_1 > x) = \mathbb{P}(Z_1 > x - T) = e^{-\lambda(x-T)}, \qquad x \ge T.$$

Taking into account the (conventional) dead period beginning at time 0, we have that

$$\mathbb{P}\big(\widetilde{N}(t) \ge k\big) = \mathbb{P}\Big(kT + \sum_{i=1}^{k} Z_i \le t\Big) = \mathbb{P}\big(N(t - kT) \ge k\big), \qquad t \ge kT,$$

where N is a Poisson process.

7. We have that $\widetilde{X}_1 = L + E(L)$ where L is the length of the dead period beginning at 0, and $E(L)$ is the excess lifetime at L. Therefore, conditioning on L,

$$\mathbb{P}(\widetilde{X}_1 \le x) = \int_0^x \mathbb{P}\big(E(l) \le x - l\big)\,dF_L(l).$$

We have that

$$\mathbb{P}\big(E(t) \le y\big) = F(t+y) - \int_0^t \big\{1 - F(t+y-x)\big\}\, dm(x).$$

By the renewal equation,

$$m(t+y) = F(t+y) + \int_0^{t+y} F(t+y-x)\, dm(x),$$

whence, by subtraction,

$$\mathbb{P}\big(E(t) \le y\big) = \int_t^{t+y} \big\{1 - F(t+y-x)\big\}\, dm(x).$$

It follows that

$$\begin{aligned}\mathbb{P}\big(\widetilde{X}_1 \le x\big) &= \int_{l=0}^{x} \int_{y=l}^{x} \big\{1 - F(x-y)\big\}\, dm(y)\, dF_L(l) \\ &= \left[F_L(l) \int_l^x \big\{1 - F(x-y)\big\}\, dm(y)\right]_{l=0}^{x} + \int_0^x F_L(l)\big\{1 - F(x-l)\big\}\, dm(l)\end{aligned}$$

using integration by parts. The term in square brackets equals 0.

8. (a) Each interarrival time has the same distribution as the sum of two independent random variables with the exponential distribution. Therefore $N(t)$ has the same distribution as $\lfloor \frac{1}{2} M(t) \rfloor$ where M is a Poisson process with intensity λ. Therefore $m(t) = \frac{1}{2}\mathbb{E}(M(t)) - \frac{1}{2}\mathbb{P}(M(t) \text{ is odd})$. Now $\mathbb{E}(M(t)) = \lambda t$, and

$$\mathbb{P}\big(M(t) \text{ is odd}\big) = \sum_{n=0}^{\infty} \frac{(\lambda t)^{2n+1} e^{-\lambda t}}{(2n+1)!} = \tfrac{1}{2} e^{-\lambda t}(e^{\lambda t} - e^{-\lambda t}).$$

With more work, one may establish the probability generating function of $N(t)$.

(b) Doing part (a) as above, one may see that $\widetilde{m}(t) = m(t)$.

9. Clearly $C(t)$ and $E(t)$ are independent if the process N is a Poisson process. Conversely, suppose that $C(t)$ and $E(t)$ are independent, for each fixed choice of t. The event $\{C(t) \ge y\} \cap \{E(t) \ge x\}$ occurs if and only if $E(t-y) \ge x + y$. Therefore

$$\mathbb{P}\big(C(t) \ge y\big)\mathbb{P}\big(E(t) \ge x\big) = \mathbb{P}\big(E(t-y) \ge x + y\big).$$

Take the limit as $t \to \infty$, remembering Exercise (10.3.3) and Problem (10.6.4), to obtain that $G(y)G(x) = G(x+y)$ if $x, y \ge 0$, where

$$G(u) = \int_u^{\infty} \frac{1}{\mu}[1 - F(v)]\, dv.$$

Now $1 - G$ is a distribution function, and hence has the lack-of-memory property (Problem (4.14.5)), implying that $G(u) = e^{-\lambda u}$ for some λ. This implies in turn that $[1 - F(u)]/\mu = -G'(u) = \lambda e^{-\lambda u}$, whence $\mu = 1/\lambda$ and $F(u) = 1 - e^{-\lambda u}$.

10. Clearly N is a renewal process if N_2 is Poisson. Suppose that N is a renewal process, and write λ for the intensity of N_1, and F_2 for the interarrival time distribution of N_2. By considering the time X_1 to the first arrival of N,

$$(*) \qquad 1 - F(x) = \mathbb{P}\big(N_1(x) = N_2(x) = 0\big) = e^{-\lambda x}\big(1 - F_2(x)\big).$$

Writing E, E_i for the excess lifetimes of N, N_i, we have that

$$\mathbb{P}\big(E(t) > x\big) = \mathbb{P}\big(E_1(t) > x,\, E_2(t) > x\big) = e^{-\lambda x}\mathbb{P}\big(E_2(t) > x\big).$$

Take the limit as $t \to \infty$, using Exercise (10.3.3), to find that

$$\int_x^\infty \frac{1}{\mu}[1 - F(u)]\,du = e^{-\lambda x}\int_x^\infty \frac{1}{\mu_2}[1 - F_2(u)]\,du,$$

where μ_2 is the mean of F_2. Differentiate, and use $(*)$, to obtain

$$\frac{1}{\mu}e^{-\lambda x}[1 - F_2(x)] = \lambda e^{-\lambda x}\int_x^\infty \frac{1}{\mu_2}[1 - F_2(u)]\,du + \frac{e^{-\lambda x}}{\mu_2}[1 - F_2(x)],$$

which simplifies to give $1 - F_2(x) = c\int_x^\infty [1 - F_2(u)]\,du$ where $c = \lambda\mu/(\mu_2 - \mu)$; this integral equation has solution $F_2(x) = 1 - e^{-cx}$.

11. (i) Taking transforms of the renewal equation in the usual way, we find that

$$m^*(\theta) = \frac{F^*(\theta)}{1 - F^*(\theta)} = \frac{1}{1 - F^*(\theta)} - 1$$

where

$$F^*(\theta) = \mathbb{E}(e^{-\theta X_1}) = 1 - \theta\mu + \tfrac{1}{2}\theta^2(\mu^2 + \sigma^2) + \mathrm{o}(\theta^2)$$

as $\theta \to 0$. Substitute this into the above expression to obtain

$$m^*(\theta) = \frac{1}{\theta\mu\{1 - \frac{1}{2}\theta(\mu + \sigma^2/\mu) + \mathrm{o}(\theta)\}} - 1$$

and expand to obtain the given expression. A formal inversion yields the expression for m.

(ii) The transform of the right-hand side of the integral equation is

$$(*) \qquad \frac{1}{\mu\theta} - F_E^*(\theta) + m^*(\theta) - F_E^*(\theta)m^*(\theta).$$

By Exercise (10.3.3), $F_E^*(\theta) = [1 - F^*(\theta)]/(\mu\theta)$, and $(*)$ simplifies to $m^*(\theta) - (m^* - m^*F^* - F^*)/(\mu\theta)$, which equals $m^*(\theta)$ since the quotient is 0 (by the renewal equation).

Using the key renewal theorem, as $t \to \infty$,

$$\int_0^t [1 - F_E(t - x)]\,dm(x) \to \frac{1}{\mu}\int_0^\infty [1 - F_E(u)]\,du = \frac{\mathbb{E}(X_1^2)}{2\mu^2} = \frac{\sigma^2 + \mu^2}{2\mu^2}$$

by Exercise (10.3.3b). Therefore,

$$m(t) - \frac{t}{\mu} \to -1 + \frac{\sigma^2 + \mu^2}{2\mu^2} = \frac{\sigma^2 - \mu^2}{2\mu^2}.$$

12. (i) Conditioning on X_1, we obtain

$$m^{\mathrm{d}}(t) = F^{\mathrm{d}}(t) + \int_0^t m(t - x)\,dF^{\mathrm{d}}(x).$$

Therefore $m^{d*} = F^{d*} + m^* F^{d*}$. Also $m^* = F^* + m^* F^*$, so that

$$(*) \qquad m^{d*} = F^{d*}\left(1 + \frac{F^*}{1 - F^*}\right),$$

whence $m^{d*} = F^{d*} + m^{d*} F^*$, the transform of the given integral equation.
(ii) Arguing as in Problem (10.6.2), $v^{d*} = F^{d*} + 2m^* F^{d*} + v^* F^{d*}$ where $v^* = F^*(1+2m^*)/(1-F^*)$ is the corresponding object in the ordinary renewal process. We eliminate v^* to find that

$$v^{d*} = \frac{F^{d*}(1+2m^*)}{1-F^*} = m^{d*}(1+2m^*)$$

by $(*)$. Now invert.

13. Taking into account the structure of the process, it suffices to deal with the case $I = 1$. Refer to Example (10.4.22) for the basic notation and analysis. It is easily seen that $\beta = (\nu - 1)\lambda$. Now $\widetilde{F}(t) = 1 - e^{-\nu\lambda t}$. Solve the renewal equation (10.4.24) to obtain

$$g(t) = h(t) + \int_0^t h(t-x)\, d\widetilde{m}(x)$$

where $\widetilde{m}(x) = \nu\lambda x$ is the renewal function associated with the interarrival time distribution $\widetilde{F}$. Therefore $g(t) = 1$, and $m(t) = e^{\beta t}$.

14. We have from Lemma (10.4.5) that $p^* = 1 - F_Z^* + p^* F^*$, where $F^* = F_Y^* F_Z^*$. Solve to obtain

$$p^* = \frac{1 - F_Z^*}{1 - F_Y^* F_Z^*}.$$

15. The first locked period begins at the time of arrival of the first particle. Since all future events may be timed relative to this arrival time, we may take this time to be 0. We shall therefore assume that a particle arrives at 0; call this the 0th particle, with locking time Y_0.

We shall condition on the time X_1 of the arrival of the next particle. Now

$$\mathbb{P}(L > t \mid X_1 = u) = \begin{cases} \mathbb{P}(Y_0 > t) & \text{if } u > t, \\ \mathbb{P}(Y_0 > u)\mathbb{P}(L' > t - u) & \text{if } u \le t, \end{cases}$$

where L' has the same distribution as L; the second part is a consequence of the fact that the process 'restarts' at each arrival. Therefore

$$\mathbb{P}(L > t) = \big(1 - G(t)\big)\mathbb{P}(X_1 > t) + \int_0^t \mathbb{P}(L > t - u)\big(1 - G(u)\big) f_{X_1}(u)\, du,$$

the required integral equation.

If $G(x) = 1 - e^{-\mu x}$, the solution is $\mathbb{P}(L > t) = e^{-\mu t}$, so that L has the same distribution as the locking times of individual particles. This striking fact may be attributed to the lack-of-memory property of the exponential distribution.

16. (a) It is clear that $M(tp)$ is a renewal process whose interarrival times are distributed as $X_1 + X_2 + \cdots + X_R$ where $\mathbb{P}(R = r) = pq^{r-1}$ for $r \ge 1$. It follows that $M(t)$ is a renewal process whose first interarrival time

$$X(p) = \inf\{t : M(t) = 1\} = p \inf\{t : M(tp) = 1\}$$

has distribution function

$$\mathbb{P}\big(X(p) \le x\big) = \sum_{r=1}^{\infty} \mathbb{P}(R = r) F_r(x/p).$$

(b) The characteristic function ϕ_p of F_p is given by

$$\phi_p(t) = \sum_{r=1}^{\infty} pq^{r-1} \int_{-\infty}^{\infty} e^{ixt}\, dF_r(t/p) = \sum_{r=1}^{\infty} pq^{r-1} \phi(pt)^r = \frac{p\phi(pt)}{1 - q\phi(pt)}$$

where ϕ is the characteristic function of F. Now $\phi(pt) = 1 + i\mu pt + o(p)$ as $p \downarrow 0$, so that

$$\phi_p(t) = \frac{1 + i\mu pt + o(p)}{1 - i\mu t + o(1)} = \frac{1 + o(1)}{1 - i\mu t}$$

as $p \downarrow 0$. The limit is the characteristic function of the exponential distribution with mean μ, and the continuity theorem tells us that the process M converges in distribution as $p \downarrow 0$ to a Poisson process with intensity $1/\mu$ (in the sense that the interarrival time distribution converges to the appropriate limit).

(c) If M and N have the same fdds, then $\phi_p(t) = \phi(t)$, which implies that $\phi(pt) = \phi(t)/(p + q\phi(t))$. Hence $\psi(t) = \phi(t)^{-1}$ satisfies $\psi(pt) = q + p\psi(t)$ for $t \in \mathbb{R}$. Now ψ is continuous, and it follows as in the solution to Problem (5.12.15) that ψ has the form $\psi(t) = 1 + \beta t$, implying that $\phi(t) = (1 + \beta t)^{-1}$ for some $\beta \in \mathbb{C}$. The only characteristic function of this form is that of an exponential distribution, and the claim follows.

17. (a) Let $N(t)$ be the number of times the sequence has been typed up to the tth keystroke. Then N is a renewal process whose interarrival times have the required mean μ; we have that $\mathbb{E}(N(t))/t \to \mu^{-1}$ as $t \to \infty$. Now each epoch of time marks the completion of such a sequence with probability $(\frac{1}{100})^{14}$, so that

$$\frac{1}{t}\mathbb{E}(N(t)) = \frac{1}{t} \sum_{n=14}^{t} \left(\frac{1}{100}\right)^{14} \to \left(\frac{1}{100}\right)^{14} \qquad \text{as } t \to \infty,$$

implying that $\mu = 10^{28}$.

The problem with 'omo' is 'omomo' (i.e., appearances may overlap). Let us call an epoch of time a 'renewal point' if it marks the completion of the word 'omo', disjoint from the words completed at previous renewal points. In each appearance of 'omo', either the first 'o' or the second 'o' (but not both) is a renewal point. Therefore the probability u_n, that n is a renewal point, satisfies $(\frac{1}{100})^3 = u_n + u_{n-2}(\frac{1}{100})^2$. Average this over n to obtain

$$\left(\frac{1}{100}\right)^3 = \lim_{n \to \infty} \frac{1}{t} \sum_{n=1}^{t} \left\{ u_n + u_{n-2}\left(\frac{1}{100}\right)^2 \right\} = \frac{1}{\mu} + \frac{1}{\mu}\left(\frac{1}{100}\right)^2,$$

and therefore $\mu = 10^6 + 10^2$.

(b) (i) Arguing as for 'omo', we obtain $p^3 = u_n + pu_{n-1} + p^2 u_{n-2}$, whence $p^3 = (1 + p + p^2)/\mu$.

(ii) Similarly, $p^2 q = u_n + pqu_{n-2}$, so that $\mu = (1 + pq)/(p^2 q)$.

18. The fdds of $\{N(u) - N(t) : u \ge t\}$ depend on the distributions of $E(t)$ and of the interarrival times. In a stationary renewal process, the distribution of $E(t)$ does not depend on the value of t, whence $\{N(u) - N(t) : u \ge t\}$ has the same fdds as $\{N(u) : u \ge 0\}$, implying that X is strongly stationary.

19. We use the renewal–reward theorem. The mean time between expeditions is $B\mu$, and this is the mean length of a cycle of the process. The mean cost of keeping the bears during a cycle is $\frac{1}{2}B(B-1)c\mu$, whence the long-run average cost is $\{d + B(B-1)c\mu/2\}/(B\mu)$.

11

Queues

11.2 Solutions. M/M/1

1. The stationary distribution satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ when it exists, where $\mathbf{P}$ is the transition matrix. The equations

$$\pi_0 = \frac{\pi_1}{1+\rho}, \quad \pi_1 = \pi_0 + \frac{\pi_2}{1+\rho}, \quad \pi_n = \frac{\rho\pi_{n-1}}{1+\rho} + \frac{\pi_{n+1}}{1+\rho} \quad \text{for } n \ge 2,$$

with $\sum_{i=0}^{\infty} \pi_i = 1$, have the given solution. If $\rho \ge 1$, no such solution exists. It is slightly shorter to use the fact that such a walk is reversible in equilibrium, from which it follows that $\boldsymbol{\pi}$ satisfies

$$(*) \qquad \pi_0 = \frac{\pi_1}{1+\rho}, \quad \frac{\rho\pi_n}{1+\rho} = \frac{\pi_{n+1}}{1+\rho} \quad \text{for } n \ge 1.$$

2. (i) This continuous-time walk is a Markov chain with generator given by $g_{01} = \theta_0$, $g_{n,n+1} = \theta_n\rho/(1+\rho)$ and $g_{n,n-1} = \theta_n/(1+\rho)$ for $n \ge 1$, other off-diagonal terms being 0. Such a process is reversible in equilibrium (see Problem (6.15.16)), and its stationary distribution $\boldsymbol{\nu}$ must satisfy $\nu_n g_{n,n+1} = \nu_{n+1} g_{n+1,n}$. These equations may be written as

$$\nu_0\theta_0 = \frac{\nu_1\theta_1}{1+\rho}, \quad \frac{\nu_n\theta_n\rho}{1+\rho} = \frac{\nu_{n+1}\theta_{n+1}}{1+\rho} \quad \text{for } n \ge 1.$$

These are identical to the equations labelled $(*)$ in the previous solution, with π_n replaced by $\nu_n\theta_n$. It follows that $\nu_n = C\pi_n/\theta_n$ for some positive constant C.
(ii) If $\theta_0 = \lambda$, $\theta_n = \lambda + \mu$ for $n \ge 1$, we have that

$$1 = \sum_n \nu_n = C\left\{\frac{\pi_0}{\lambda} + \frac{1-\pi_0}{\mu+\lambda}\right\} = \frac{C}{2\lambda},$$

whence $C = 2\lambda$ and the result follows.

3. Let Q be the number of people ahead of the arriving customer at the time of his arrival. Using the lack-of-memory property of the exponential distribution, the customer in service has residual service-time with the exponential distribution, parameter μ, whence W may be expressed as $S_1 + S_2 + \cdots + S_Q$, the sum of independent exponential variables, parameter μ. The characteristic function of W is

$$\begin{aligned}\phi_W(t) &= \mathbb{E}\{\mathbb{E}(e^{itW} \mid Q)\} = \mathbb{E}\left\{\left(\frac{\mu}{\mu - it}\right)^Q\right\} \\ &= \frac{1-\rho}{1-\rho\mu/(\mu-it)} = (1-\rho) + \rho\left(\frac{\mu-\lambda}{\mu-\lambda-it}\right).\end{aligned}$$

This is the characteristic function of the given distribution. The atom at 0 corresponds to the possibility that $Q = 0$.

4. We prove this by induction on the value of $i + j$. If $i + j = 0$ then $i = j = 0$, and it is easy to check that $\pi(0; 0, 0) = 1$ and $A(0; 0, 0) = 1$, $A(n; 0, 0) = 0$ for $n \geq 1$. Suppose then that $K \geq 1$, and that the claim is valid for all pairs (i, j) satisfying $i + j = K$. Let i and j satisfy $i + j = K + 1$. The last ball picked has probability $i/(i + j)$ of being red; conditioning on the colour of the last ball, we have that

$$\pi(n; i, j) = \frac{i}{i+j}\pi(n-1; i-1, j) + \frac{j}{i+j}\pi(n+1; i, j-1).$$

Now $(i - 1) + j = K = i + (j - 1)$. Applying the induction hypothesis, we find that

$$\begin{aligned}\pi(n; i, j) = \frac{i}{i+j}\Big\{A(n-1; i-1, j) - A(n; i-1, j)\Big\} \\ + \frac{j}{i+j}\Big\{A(n+1; i, j-1) - A(n+2; i, j-1)\Big\}.\end{aligned}$$

Substitute to obtain the required answer, after a little cancellation and collection of terms. Can you see a more natural way?

5. Let A and B be independent Poisson process with intensities λ and μ respectively. These processes generate a queue-process as follows. At each arrival time of A, a customer arrives in the shop. At each arrival-time of B, the customer being served completes his service and leaves; if the queue is empty at this moment, then nothing happens. It is not difficult to see that this queue-process is $M(\lambda)/M(\mu)/1$. Suppose that $A(t) = i$ and $B(t) = j$. During the time-interval $[0, t]$, the order of arrivals and departures follows the schedule of Exercise (11.2.4), arrivals being marked as red balls and departures as lemon balls. The imbedded chain has the same distributions as the random walk of that exercise, and it follows that $\mathbb{P}\big(Q(t) = n \mid A(t) = i, B(t) = j\big) = \pi(n; i, j)$. Therefore $p_n(t) = \sum_{i,j} \pi(n; i, j)\mathbb{P}(A(t) = i)\mathbb{P}(B(t) = j)$.

6. With $\rho = \lambda/\mu$, the stationary distribution of the imbedded chain is, as in Exercise (11.2.1),

$$\widehat{\pi}_n = \begin{cases} \frac{1}{2}(1-\rho) & \text{if } n = 0, \\ \frac{1}{2}(1-\rho^2)\rho^{n-1} & \text{if } n \geq 1. \end{cases}$$

In the usual notation of continuous-time Markov chains, $g_0 = \lambda$ and $g_n = \lambda + \mu$ for $n \geq 1$, whence, by Exercise (6.10.11), there exists a constant c such that

$$\pi_0 = \frac{c}{2\lambda}(1-\rho), \quad \pi_n = \frac{c}{2(\lambda+\mu)}(1-\rho^2)\rho^{n-1} \quad \text{for } n \geq 1.$$

Now $\sum_i \pi_i = 1$, and therefore $c = 2\lambda$ and $\pi_n = (1-\rho)\rho^n$ as required. The working is reversible.

7. (a) Let $Q_i(t)$ be the number of people in the ith queue at time t, including any currently in service. The process Q_1 is reversible in equilibrium, and departures in the original process correspond to arrivals in the reversed process. It follows that the departure process of the first queue is a Poisson process with intensity λ, and that the departure process of Q_1 is independent of the current value of Q_1.

(b) We have from part (a) that, for any given t, the random variables $Q_1(t)$, $Q_2(t)$ are independent. Consider an arriving customer when the queues are in equilibrium, and let W_i be his waiting time (before service) in the ith queue. With T the time of arrival, and recalling Exercise (11.2.3),

$$\begin{aligned}\mathbb{P}(W_1 = 0,\ W_2 = 0) &> \mathbb{P}(Q_i(T) = 0 \text{ for } i = 1, 2) = \mathbb{P}(Q_1(T) = 0)\mathbb{P}(Q_2(T) = 0) \\ &= (1-\rho_1)(1-\rho_2) = \mathbb{P}(W_1 = 0)\mathbb{P}(W_2 = 0).\end{aligned}$$

Therefore W_1 and W_2 are not independent. There is a slight complication arising from the fact that T is a random variable. However, T is independent of everybody who has gone before, and in particular of the earlier values of the queue processes Q_i.

11.3 Solutions. M/G/1

1. In equilibrium, the queue-length Q_n just after the nth departure satisfies

$$(*) \qquad Q_{n+1} = A_n + Q_n - h(Q_n)$$

where A_n is the number of arrivals during the $(n+1)$th service period, and $h(m) = 1 - \delta_{m0}$. Now Q_n and Q_{n+1} have the same distribution. Take expectations to obtain

$$(**) \qquad 0 = \mathbb{E}(A_n) - \mathbb{P}(Q_n > 0),$$

where $\mathbb{E}(A_n) = \lambda d$, the mean number of arrivals in an interval of length d. Next, square $(*)$ and take expectations:

$$0 = \mathbb{E}(A_n^2) + \mathbb{E}\big(h(Q_n)^2\big) + 2\Big\{\mathbb{E}(A_n Q_n) - \mathbb{E}\big(A_n h(Q_n)\big) - \mathbb{E}\big(Q_n h(Q_n)\big)\Big\}.$$

Use the facts that A_n is independent of Q_n, and that $Q_n h(Q_n) = Q_n$, to find that

$$0 = \{(\lambda d)^2 + \lambda d\} + \mathbb{P}(Q_n > 0) + 2\big\{(\lambda d - 1)\mathbb{E}(Q_n) - \lambda d \mathbb{P}(Q_n > 0)\big\}$$

and therefore, by $(**)$,

$$\mathbb{E}(Q_n) = \frac{2\rho - \rho^2}{2(1-\rho)}.$$

2. From the standard theory, M_B satisfies $M_B(s) = M_S(s - \lambda + \lambda M_B(s))$, where $M_S(\theta) = \mu/(\mu - \theta)$. Substitute to find that $x = M_B(s)$ is a root of the quadratic $\lambda x^2 - x(\lambda + \mu - s) + \mu = 0$. For some small positive s, $M_B(s)$ is smooth and non-decreasing. Therefore $M_B(s)$ is the root given.

3. Let T_n be the instant of time at which the server is freed for the nth time. By the lack-of-memory property of the exponential distribution, the time of the first arrival after T_n is independent of all arrivals prior to T_n, whence T_n is a 'regeneration point' of the queue (so to say). It follows that the times which elapse between such regeneration points are independent, and it is easily seen that they have the same distribution.

11.4 Solutions. G/M/1

1. The transition matrix of the imbedded chain obtained by observing queue-lengths just before arrivals is

$$\mathbf{P}_A = \begin{pmatrix} 1-\alpha_0 & \alpha_0 & 0 & 0 & \dots \\ 1-\alpha_0-\alpha_1 & \alpha_1 & \alpha_0 & 0 & \dots \\ 1-\alpha_0-\alpha_1-\alpha_2 & \alpha_2 & \alpha_1 & \alpha_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_A$ may be written as

$$\pi_0 = \sum_{i=0}^{\infty} \pi_i \left(1 - \sum_{j=0}^{i} \alpha_j\right), \qquad \pi_n = \sum_{i=0}^{\infty} \alpha_i \pi_{n+i-1} \quad \text{for } n \ge 1.$$

It is easily seen, by adding, that the first equation is a consequence of the remaining equations, taken in conjunction with $\sum_0^\infty \pi_i = 1$. Therefore $\boldsymbol{\pi}$ is specified by the equation for π_n, $n \ge 1$.

The indicated substitution gives

$$\theta^n = \theta^{n-1} \sum_{i=0}^{\infty} \alpha_i \theta^i$$

which is satisfied whenever θ satisfies

$$\theta = \sum_{i=0}^{\infty} \alpha_i \theta^i = \sum_{i=0}^{\infty} \mathbb{E}\left\{ \frac{(\mu X \theta)^i e^{-\mu X}}{i!} \right\} = \mathbb{E}(e^{-\mu X} e^{\mu X \theta}) = M_X\big(\mu(\theta - 1)\big).$$

It is easily seen that $A(\theta) = M_X(\mu(\theta - 1))$ is convex and non-decreasing on $[0, 1]$, and satisfies $A(0) > 0$, $A(1) = 1$. Now $A'(1) = \mu\mathbb{E}(X) = \rho^{-1} > 1$, implying that there is a unique $\eta \in (0, 1)$ such that $A(\eta) = \eta$. With this value of η, the vector $\boldsymbol{\pi}$ given by $\pi_j = (1 - \eta)\eta^j$, $j \geq 0$, is a stationary distribution of the imbedded chain. This $\boldsymbol{\pi}$ is the unique such distribution because the chain is irreducible.

2. (i) The equilibrium distribution is $\pi_n = (1 - \eta)\eta^n$ for $n \geq 0$, with mean $\sum_{n=0}^{\infty} n\pi_n = \eta/(1 - \eta)$.
(ii) Using the lack-of-memory property of the service time in progress at the time of the arrival, we see that the waiting time may be expressed as $W = S_1 + S_2 + \cdots + S_Q$ where Q has distribution $\boldsymbol{\pi}$, given above, and the S_n are service times independent of Q. Therefore

$$\mathbb{E}(W) = \mathbb{E}(S_1)\mathbb{E}(Q) = \frac{\eta/\mu}{(1 - \eta)}.$$

3. We have that $Q(n+) = 1 + Q(n-)$ a.s. for each integer n, whence $\lim_{t\to\infty} \mathbb{P}(Q(t) = m)$ cannot exist.

Since the traffic intensity is less than 1, the imbedded chain is ergodic with stationary distribution as in Exercise (11.4.1).

11.5 Solutions. G/G/1

1. Let T_n be the starting time of the nth busy period. Then T_n is an arrival time, and also the beginning of a service period. Conditional on the value of T_n, the future evolution of the queue is independent of the past, whence the random variables $\{T_{n+1} - T_n : n \geq 1\}$ are independent. It is easily seen that they are identically distributed.

2. If the server is freed at time T, the time I until the next arrival has the exponential distribution with parameter μ (since arrivals form a Poisson process).

By the duality theory of queues, the waiting time in question has moment generating function $M_W(s) = (1 - \zeta)/(1 - \zeta M_I(s))$ where $M_I(s) = \mu/(\mu - s)$ and $\zeta = \mathbb{P}(W > 0)$. Therefore,

$$M_W(s) = \frac{\zeta\mu(1 - \zeta)}{\mu(1 - \zeta) - s} + (1 - \zeta),$$

the moment generating function of a mixture of an atom at 0 and an exponential distribution with parameter $\mu(1 - \zeta)$.

If G is the probability generating function of the equilibrium queue-length, then, using the lack-of-memory property of the exponential distribution, we have that $M_W(s) = G(\mu/(\mu - s))$, since W is the sum of the (residual) service times of the customers already present. Set $u = \mu/(\mu - s)$ to find that $G(u) = (1 - \zeta)/(1 - \zeta u)$, the generating function of the mass function $f(k) = (1 - \zeta)\zeta^k$ for $k \geq 0$.

It may of course be shown that ζ is the smallest positive root of the equation $x = M_X\big(\mu(x-1)\big)$, where X is a typical interarrival time.

3. We have that

$$1 - G(y) = \mathbb{P}(S - X > y) = \int_0^\infty \mathbb{P}(S > u + y)\,dF_X(u), \qquad y \in \mathbb{R},$$

where S and X are typical (independent) service and interarrival times. Hence, formally,

$$dG(y) = -\int_0^\infty d\mathbb{P}(S > u + y)\,dF_X(u) = dy \int_{-y}^\infty \mu e^{-\mu(u+y)}\,dF_X(u),$$

since $f_S(u+y) = e^{-\mu(u+y)}$ if $u > -y$, and is 0 otherwise.

With F as given,

$$\int_{-\infty}^x F(x-y)\,dG(y) = \iint_{\substack{-\infty<y\le x\\ -y<u<\infty}} \big\{1 - \eta e^{-\mu(1-\eta)(x-y)}\big\}\mu e^{-\mu(u+y)}\,dF_X(u)\,dy.$$

First integrate over y, then over u (noting that $F_X(u) = 0$ for $u < 0$), and the double integral collapses to $F(x)$, when $x \ge 0$.

11.6 Solution. Heavy traffic

1. Q_ρ has characteristic function

$$\phi_\rho(t) = \sum_{n=0}^\infty e^{itn}\rho^n(1-\rho) = \frac{1-\rho}{1-\rho e^{it}}.$$

Therefore the characteristic function of $(1-\rho)Q_\rho$ satisfies

$$\phi_\rho\big((1-\rho)t\big) = \frac{1-\rho}{1-\rho e^{i(1-\rho)t}} \to \frac{1}{1-it} \qquad \text{as } \rho \uparrow 1.$$

The limit characteristic function is that of the exponential distribution, and the result follows by the continuity theorem.

11.7 Solutions. Networks of queues

1. The first observation follows as in Example (11.7.4). The equilibrium distribution is given as in Theorem (11.7.14) by

$$\pi(\mathbf{n}) = \prod_{i=1}^c \frac{\alpha_i^{n_i} e^{-\alpha_i}}{n_i!}, \qquad \text{for } \mathbf{n} = (n_1, n_2, \dots, n_c) \in \mathbb{Z}^c,$$

the product of Poisson distributions. This is related to Bartlett's theorem (see Problem (8.7.6)) by defining the state A as 'being in station i at some given time'.

2. The number of customers in the queue is a birth–death process, and is therefore reversible in equilibrium. The claims follow in the same manner as was argued in the solution to Exercise (11.2.7).

3. (a) We may take as state space the set $\{0, 1', 1'', 2, 3, \dots\}$, where $i \in \{0, 2, 3, \dots\}$ is the state of having i people in the system including any currently in service, and $1'$ (respectively $1''$) is the state of having exactly one person in the system, this person being served by the first (respectively second) server. It is straightforward to check that this process is reversible in equilibrium, whence the departure process is as stated, by the argument used in Exercise (11.2.7).

(b) This time, we take as state space the set $\{0', 0'', 1', 1'', 2, 3, \dots\}$ having the same states as in part (a) with the difference that $0'$ (respectively $0''$) is the state in which there are no customers present and the first (respectively second) server has been free for the shorter time. It is easily seen that transition from $0'$ to $1''$ has strictly positive probability whereas transition from $1''$ to $0'$ has zero probability, implying that the process is not reversible. By drawing a diagram of the state space, or otherwise, it may be seen that the time-reversal of the process has the same structure as the original, with the unique change that states $0'$ are $0''$ are interchanged. Since departures in the original process correspond to arrivals in the time-reversal, the required properties follow in the same manner as in Exercise (11.2.7).

4. The total time spent by a given customer in service may be expressed as the sum of geometrically distributed number of exponential random variables, and this is easily shown to be exponential with parameter $\delta\mu$. The queue is therefore in effect a M(λ)/M($\delta\mu$)/1 system, and the stationary distribution is the geometric distribution with parameter $\rho = \lambda/(\delta\mu)$, provided $\rho < 1$. As in Exercise (11.2.7), the process of departures is Poisson.

Assume that rejoining customers go to the end of the queue, and note that the number of customers present constitutes a Markov chain. However, the composite process of arrivals is not Poisson, since increments are no longer independent. This may be seen as follows. In equilibrium, the probability of an arrival of either kind during the time interval $(t, t+h)$ is $\lambda h + \rho\mu(1-\delta)h + o(h) = (\lambda/\delta)h + o(h)$. If there were an arrival of either kind during $(t-h, t)$, then (with conditional probability $1 - O(h)$) the queue is non-empty at time t, whence the conditional probability of an arrival of either kind during $(t, t+h)$ is $\lambda h + \mu(1-\delta)h + o(h)$; this is of a larger order of magnitude than the earlier probability $(\lambda/\delta)h + o(h)$.

5. For stations r, s, we write $r \to s$ if an individual at r visits s at a later time with a strictly positive probability. Let C comprise the station j together with all stations i such that $i \to j$. The process restricted to C is an open migration process in equilibrium. By Theorem (11.7.19), the restricted process is reversible, whence the process of departures from C via j is a Poisson process with some intensity ζ. Individuals departing C via j proceed directly to k with probability

$$\frac{\lambda_{jk}\phi_j(n_j)}{\mu_j\phi_j(n_j) + \sum_{r\notin C}\lambda_{jr}\phi_j(n_j)} = \frac{\lambda_{jk}}{\mu_j + \sum_{r\notin C}\lambda_{jr}},$$

independently of the number n_j of individuals currently at j. Such a thinned Poisson process is a Poisson process also (cf. Exercise (6.8.2)), and the claim follows.

11.8 Solutions to problems

1. Although the two cases may be done together, we choose to do them separately. When $k = 1$, the equilibrium distribution $\boldsymbol{\pi}$ satisfies:

$$\begin{aligned}
\mu\pi_1 - \lambda\pi_0 &= 0, \\
\mu\pi_{n+1} - (\lambda+\mu)\pi_n + \lambda\pi_{n-1} &= 0, \qquad 1 \le n < N, \\
-\mu\pi_N + \lambda\pi_{N-1} &= 0,
\end{aligned}$$

a system of equations with solution $\pi_n = \pi_0(\lambda/\mu)^n$ for $0 \le n \le N$, where (if $\lambda \ne \mu$)

$$\pi_0^{-1} = \sum_{n=0}^{N}(\lambda/\mu)^n = \frac{1-(\lambda/\mu)^{N+1}}{1-(\lambda/\mu)}.$$

Now let $k = 2$. The queue is a birth–death process with rates

$$\lambda_i = \begin{cases} \lambda & \text{if } i < N, \\ 0 & \text{if } i \ge N, \end{cases} \qquad \mu_i = \begin{cases} \mu & \text{if } i = 1, \\ 2\mu & \text{if } i \ge 2. \end{cases}$$

It is reversible in equilibrium, and its stationary distribution satisfies $\lambda_i \pi_i = \mu_{i+1}\pi_{i+1}$. We deduce that $\pi_i = 2\rho^i \pi_0$ for $1 \le i \le N$, where $\rho = \lambda/(2\mu)$ and

$$\pi_0^{-1} = 1 + \sum_{i=1}^{N} 2\rho^i.$$

2. The answer is obtainable in either case by following the usual method. It is shorter to use the fact that such processes are reversible in equilibrium.

(a) The stationary distribution $\boldsymbol{\pi}$ satisfies $\pi_n \lambda p(n) = \pi_{n+1}\mu$ for $n \ge 0$, whence $\pi_n = \pi_0 \rho^n/n!$ where $\rho = \lambda/\mu$. Therefore $\pi_n = \rho^n e^{-\rho}/n!$.

(b) Similarly,

$$\pi_n = \pi_0 \rho^n \prod_{m=0}^{n-1} p(m) = \pi_0 \rho^n 2^{-\frac{1}{2}n(n-1)}, \qquad n \ge 0,$$

where

$$\pi_0^{-1} = \sum_{n=0}^{\infty} \rho^n \left(\tfrac{1}{2}\right)^{\frac{1}{2}n(n-1)}.$$

At the instant of arrival of a potential customer, the probability q that she joins the queue is obtained by conditioning on its length:

$$q = \sum_{n=0}^{\infty} p(n)\pi_n = \pi_0 \sum_{n=0}^{\infty} \rho^n 2^{-n-\frac{1}{2}n(n-1)} = \pi_0 \sum_{n=0}^{\infty} \rho^n 2^{-\frac{1}{2}n(n+1)} = \pi_0 \frac{1}{\rho}\{\pi_0^{-1} - 1\}.$$

3. *First method.* Let (Q_1, Q_2) be the queue-lengths, and suppose they are in equilibrium. Since Q_1 is a birth–death process, it is reversible, and we write $\widehat{Q}_1(t) = Q_1(-t)$. The sample paths of Q_1 have increasing jumps of size 1 at times of a Poisson process with intensity λ; these jumps mark arrivals at the cash desk. By reversibility, $\widehat{Q}_1$ has the same property; such increasing jumps for $\widehat{Q}_1$ are decreasing jumps for Q_1, and therefore the times of departures from the cash desk form a Poisson process with intensity λ. Using the same argument, the quantity $Q_1(t)$ together with the departures prior to t have the same joint distribution as the quantity $\widehat{Q}_1(-t)$ together with all arrivals after $-t$. However $\widehat{Q}_1(-t)$ is independent of its subsequent arrivals, and therefore $Q_1(t)$ is independent of its earlier departures.

It follows that arrivals at the second desk are in the manner of a Poisson process with intensity λ, and that $Q_2(t)$ is independent of $Q_1(t)$. Departures from the second desk form a Poisson process also.

Hence, in equilibrium, Q_1 is M(λ)/M(μ_1)/1 and Q_2 is M(λ)/M(μ_2)/1, and they are independent at any given time. Therefore their joint stationary distribution is

$$\pi_{mn} = \mathbb{P}\big(Q_1(t) = m, Q_2(t) = n\big) = (1-\rho_1)(1-\rho_2)\rho_1^m \rho_2^n$$

where $\rho_i = \lambda/\mu_i$.

Second method. The pair $(Q_1(t), Q_2(t))$ is a bivariate Markov chain. A stationary distribution $(\pi_{mn} : m, n \ge 0)$ satisfies

$$(\lambda + \mu_1 + \mu_2)\pi_{mn} = \lambda \pi_{m-1,n} + \mu_1 \pi_{m+1,n-1} + \mu_2 \pi_{m,n+1}, \qquad m, n \ge 1,$$

together with other equations when $m = 0$ or $n = 0$. It is easily checked that these equations have the solution given above, when $\rho_i < 1$ for $i = 1, 2$.

4. Let D_n be the time of the nth departure, and let $Q_n = Q(D_n+)$ be the number of waiting customers immediately after D_n. We have in the usual way that $Q_{n+1} = A_n + Q_n - h(Q_n)$, where A_n is the number of arrivals during the $(n+1)$th service time, and $h(x) = \min\{x, m\}$. Let $G(s) = \sum_{i=0}^{\infty} \pi_i s^i$ be the equilibrium probability generating function of the Q_n. Then, since Q_n is independent of A_n,

$$G(s) = \mathbb{E}(s^{A_n})\mathbb{E}(s^{Q_n - h(Q_n)})$$

where

$$\mathbb{E}(s^{A_n}) = \int_0^\infty e^{\lambda u(s-1)} f_S(u)\,du = M_S\big(\lambda(s-1)\big),$$

M_S being the moment generating function of a service time, and

$$\mathbb{E}(s^{Q_n - h(Q_n)}) = \sum_{i=0}^{m} \pi_i + \sum_{i=m+1}^{\infty} s^{i-m}\pi_i = s^{-m}\left\{G(s) + \sum_{i=0}^{m}(s^m - s^i)\pi_i\right\}.$$

Combining these relations, we obtain that G satisfies

$$s^m G(s) = M_S\big(\lambda(s-1)\big)\left\{G(s) + \sum_{i=0}^{m}(s^m - s^i)\pi_i\right\},$$

whenever it exists.

Finally suppose that $m = 2$ and $M_S(\theta) = \mu/(\mu - \theta)$. In this case,

$$G(s) = \frac{\mu\{\pi_0(s+1) + \pi_1 s\}}{\mu(s+1) - \lambda s^2}.$$

Now $G(1) = 1$, whence $\mu(2\pi_0 + \pi_1) = 2\mu - \lambda$; this implies in particular that $2\mu - \lambda > 0$. Also $G(s)$ converges for $|s| \le 1$. Therefore any zero of the denominator in the interval $[-1, 1]$ is also a zero of the numerator. There exists exactly one such zero, since the denominator is a quadratic which takes the value $-\lambda$ at $s = -1$ and the value $2\mu - \lambda$ at $s = 1$. The zero in question is at

$$s_0 = \frac{\mu - \sqrt{\mu^2 + 4\lambda\mu}}{2\lambda},$$

and it follows that $\pi_0 + (\pi_0 + \pi_1)s_0 = 0$. Solving for π_0 and π_1, we obtain

$$G(s) = \frac{1-\alpha}{1-\alpha s},$$

where $\alpha = 2\lambda/\{\mu + \sqrt{\mu^2 + 4\lambda\mu}\}$.

5. Recalling standard M/G/1 theory, the moment generating function M_B satisfies

$$(*) \qquad M_B(s) = M_S\big(s - \lambda + \lambda M_B(s)\big) = \frac{\mu}{\mu - \{s - \lambda + \lambda M_B(s)\}}$$

whence $M_B(s)$ is one of

$$\frac{(\lambda + \mu - s) \pm \sqrt{(\lambda + \mu - s)^2 - 4\lambda\mu}}{2\lambda}.$$

Now $M_B(s)$ is non-decreasing in s, and therefore it is the value with the minus sign. The density function of B may be found by inverting the moment generating function; see Feller (1971, p. 482), who has also an alternative derivation of M_B.

As for the mean and variance, either differentiate M_B, or differentiate $(*)$. Following the latter route, we obtain the following relations involving M $(= M_B)$:

$$2\lambda MM' + M + (s - \lambda - \mu)M' = 0,$$
$$2\lambda MM'' + 2\lambda(M')^2 + 2M' + (s - \lambda - \mu)M'' = 0.$$

Set $s = 0$ to obtain $M'(0) = (\mu - \lambda)^{-1}$ and $M''(0) = 2\mu(\mu - \lambda)^{-3}$, whence the claims are immediate.

6. (i) This question is closely related to Exercise (11.3.1). With the same notation as in that solution, we have that

$$(*) \qquad Q_{n+1} = A_n + Q_n - h(Q_n)$$

where $h(x) = \min\{1, x\}$. Taking expectations, we obtain $\mathbb{P}(Q_n > 0) = \mathbb{E}(A_n)$ where

$$\mathbb{E}(A_n) = \int_0^\infty \mathbb{E}(A_n \mid S = s)\, dF_S(s) = \lambda\mathbb{E}(S) = \rho,$$

and S is a typical service time. Square $(*)$ and take expectations to obtain

$$\mathbb{E}(Q_n) = \frac{\rho(1 - 2\rho) + \mathbb{E}(A_{n+1}^2)}{2(1 - \rho)},$$

where $\mathbb{E}(A_n^2)$ is found (as above) to equal $\rho + \lambda^2\mathbb{E}(S^2)$.

(ii) If a customer waits for time W and is served for time S, he leaves behind him a queue-length which is Poisson with parameter $\lambda(W + S)$. In equilibrium, its mean satisfies $\lambda\mathbb{E}(W + S) = \mathbb{E}(Q_n)$, whence $\mathbb{E}(W)$ is given as claimed.

(iii) $\mathbb{E}(W)$ is a minimum when $\mathbb{E}(S^2)$ is minimized, which occurs when S is concentrated at its mean. Deterministic service times minimize mean waiting time.

7. Condition on arrivals in $(t, t+h)$. If there are no arrivals, then $W_{t+h} \le x$ if and only if $W_t \le x+h$. If there is an arrival, and his service time is S, then $W_{t+h} \le x$ if and only if $W_t \le x + h - S$. Therefore

$$F(x; t + h) = (1 - \lambda h)F(x + h; t) + \lambda h \int_0^{x+h} F(x + h - s; t)\, dF_S(s) + o(h).$$

Subtract $F(x; t)$, divide by h, and take the limit as $h \downarrow 0$, to obtain the differential equation.

We take Laplace–Stieltjes transforms. Integrating by parts, for $\theta \le 0$,

$$\int_{(0,\infty)} e^{\theta x}\, dh(x) = -h(0) - \theta\{M_U(\theta) - H(0)\},$$
$$\int_{(0,\infty)} e^{\theta x}\, dH(x) = M_U(\theta) - H(0),$$
$$\int_{(0,\infty)} e^{\theta x}\, d\mathbb{P}(U + S \le x) = M_U(\theta)M_S(\theta),$$

and therefore

$$0 = -h(0) - \theta\{M_U(\theta) - H(0)\} + \lambda H(0) + \lambda M_U(\theta)\{M_S(\theta) - 1\}.$$

Set $\theta = 0$ to obtain that $h(0) = \lambda H(0)$, and therefore

$$H(0) = -\frac{1}{\theta} M_U(\theta)\{\lambda(M_S(\theta) - 1) - \theta\}.$$

Take the limit as $\theta \to 0$, using L'Hôpital's rule, to obtain $H(0) = 1 - \lambda\mathbb{E}(S) = 1 - \rho$. The moment generating function of U is given accordingly. Note that M_U is the same as the moment generating function of the equilibrium distribution of *actual* waiting time. That is to say, *virtual* and *actual* waiting times have the same equilibrium distributions in this case.

8. In this case U takes the values 1 and -2 each with probability $\frac{1}{2}$ (as usual, $U = S - X$ where S and X are typical (independent) service and interarrival times). The integral equation for the limiting waiting time distribution function F becomes

$$F(0) = \tfrac{1}{2}F(2), \qquad F(x) = \tfrac{1}{2}\{F(x-1) + F(x+2)\} \qquad \text{for } x = 1, 2, \dots.$$

The auxiliary equation is $\theta^3 - 2\theta + 1 = 0$, with roots 1 and $-\frac{1}{2}(1 \pm \sqrt{5})$. Only roots lying in $[-1, 1]$ can contribute, whence

$$F(x) = A + B\left(\frac{-1+\sqrt{5}}{2}\right)^x$$

for some constants A and B. Now $F(x) \to 1$ as $x \to \infty$, since the queue is stable, and therefore $A = 1$. Using the equation for $F(0)$, we find that $B = \frac{1}{2}(1 - \sqrt{5})$.

9. Q is a M(λ)/M(μ)/∞ queue, otherwise known as an immigration–death process (see Exercise (6.11.3) and Problem (6.15.18)). As found in (6.15.18), $Q(t)$ has probability generating function

$$G(s, t) = \{1 + (s-1)e^{-\mu t}\}^I \exp\{\rho(s-1)(1 - e^{-\mu t})\}$$

where $\rho = \lambda/\mu$. Hence

$$\begin{aligned}
\mathbb{E}(Q(t)) &= Ie^{-\mu t} + \rho(1 - e^{-\mu t}),\\
\mathbb{P}(Q(t) = 0) &= (1 - e^{-\mu t})^I \exp\{-\rho(1 - e^{-\mu t})\},\\
\mathbb{P}(Q(t) = n) &\to \frac{1}{n!}\rho^n e^{-\rho} \qquad \text{as } t \to \infty.
\end{aligned}$$

If $\mathbb{E}(I)$ and $\mathbb{E}(B)$ denote the mean lengths of an idle period and a busy period in equilibrium, we have that the proportion of time spent idle is $\mathbb{E}(I)/\{\mathbb{E}(I) + \mathbb{E}(B)\}$. This equals $\lim_{t\to\infty} \mathbb{P}(Q(t) = 0) = e^{-\rho}$. Now $\mathbb{E}(I) = \lambda^{-1}$, by the lack-of-memory property of the arrival process, so that $\mathbb{E}(B) = (e^\rho - 1)/\lambda$.

10. We have in the usual way that

$$(*) \qquad Q(t+1) = A_t + Q(t) - \min\{1, Q(t)\}$$

where A_t has the Poisson distribution with parameter λ. When the queue is in equilibrium, $\mathbb{E}(Q(t)) = \mathbb{E}(Q(t+1))$, and hence

$$\mathbb{P}(Q(t) > 0) = \mathbb{E}(\min\{1, Q(t)\}) = \mathbb{E}(A_t) = \lambda.$$

We have from $(*)$ that the probability generating function $G(s)$ of the equilibrium distribution of $Q(t)$ $(\equiv Q)$ is

$$G(s) = \mathbb{E}(s^{A_t})\mathbb{E}(s^{Q - \min\{1, Q\}}) = e^{\lambda(s-1)}\{\mathbb{E}(s^{Q-1} I_{\{Q \geq 1\}}) + \mathbb{P}(Q = 0)\}.$$

Also,

$$G(s) = \mathbb{E}(s^Q I_{\{Q \geq 1\}}) + \mathbb{P}(Q = 0),$$

and hence

$$G(s) = e^{\lambda(s-1)} \left\{ \frac{1}{s} G(s) + \left(1 - \frac{1}{s}\right)(1 - \lambda) \right\}$$

whence

$$G(s) = \frac{(1-s)(1-\lambda)}{1 - se^{-\lambda(s-1)}}.$$

The mean queue length is $G'(1) = \frac{1}{2}\lambda(2-\lambda)/(1-\lambda)$. Since service times are of unit length, and arrivals form a Poisson process, the mean residual service time of the customer in service at an arrival time is $\frac{1}{2}$, so long as the queue is non-empty. Hence

$$\mathbb{E}(W) = \mathbb{E}(Q) - \tfrac{1}{2}\mathbb{P}(Q > 0) = \frac{\lambda}{2(1-\lambda)}.$$

11. The length B of a typical busy period has moment generating function satisfying $M_B(s) = \exp\{s - \lambda + \lambda M_B(s)\}$; this fact may be deduced from the standard theory of M/G/1, or alternatively by a random-walk approach. Now T may be expressed as $T = I + B$ where I is the length of the first idle period, a random variable with the exponential distribution, parameter λ. It follows that $M_T(s) = \lambda M_B(s)/(\lambda - s)$. Therefore, as required,

$$(*) \qquad (\lambda - s)M_T(s) = \lambda \exp\{s - \lambda + (\lambda - s)M_T(s)\}.$$

If $\lambda \geq 1$, the queue-length at moments of departure is either null persistent or transient, and it follows that $\mathbb{E}(T) = \infty$. If $\lambda < 1$, we differentiate $(*)$ and set $s = 0$ to obtain $\lambda\mathbb{E}(T) - 1 = \lambda^2\mathbb{E}(T)$, whence $\mathbb{E}(T) = \{\lambda(1-\lambda)\}^{-1}$.

12. (a) Q is a birth–death process with parameters $\lambda_i = \lambda$, $\mu_i = \mu$, and is therefore reversible in equilibrium; see Problems (6.15.16) and (11.8.3).

(b) The equilibrium distribution satisfies $\lambda\pi_i = \mu\pi_{i+1}$ for $i \geq 0$, whence $\pi_i = (1-\rho)\rho^i$ where $\rho = \lambda/\mu$. A typical waiting time W is the sum of Q independent service times, so that

$$M_W(s) = G_Q\big(M_S(s)\big) = \frac{1-\rho}{1 - \rho\mu/(\mu - s)} = \frac{(1-\rho)(\mu - s)}{\mu(1-\rho) - s}.$$

(c) See the solution to Problem (11.8.3).

(d) Follow the solution to Problem (11.8.3) (either method) to find that, at any time t in equilibrium, the queue lengths are independent, the jth having the equilibrium distribution of M(λ)/M(μ_j)/1. The joint mass function is therefore

$$f(x_1, x_2, \dots, x_K) = \prod_{j=1}^{K} (1-\rho_j)\rho_j^{x_j}$$

where $\rho_j = \lambda/\mu_j$.

13. The size of the queue is a birth–death process with rates $\lambda_i = \lambda$, $\mu_i = \mu \min\{i, k\}$. Either solve the equilibrium equations in order to find a stationary distribution $\boldsymbol{\pi}$, or argue as follows. The process is reversible in equilibrium (see Problem (6.15.16)), and therefore $\lambda_i\pi_i = \mu_{i+1}\pi_{i+1}$ for all i. These 'balance equations' become

$$\lambda\pi_i = \begin{cases} \mu(i+1)\pi_{i+1} & \text{if } 0 \leq i < k, \\ \mu k\pi_{i+1} & \text{if } i \geq k. \end{cases}$$

These are easily solved iteratively to obtain

$$\pi_i = \begin{cases} \pi_0 \alpha^i / i! & \text{if } 0 \le i \le k, \\ \pi_0 (\alpha/k)^i k^k / k! & \text{if } i \ge k \end{cases}$$

where $\alpha = \lambda/\mu$. Therefore there exists a stationary distribution if and only if $\lambda < k\mu$, and it is given accordingly, with

$$\pi_0^{-1} = \sum_{i=0}^{k-1} \frac{\alpha^i}{i!} + \frac{k^k}{k!} \sum_{i=k}^{\infty} (\alpha/k)^i.$$

The cost of having k servers is

$$C_k = Ak + B\pi_0 \sum_{i=k}^{\infty} (i - k + 1) \frac{(\alpha/k)^i k^k}{k!}$$

where $\pi_0 = \pi_0(k)$. One finds, after a little computation, that

$$C_1 = A + \frac{B\alpha}{1-\alpha}, \qquad C_2 = 2A + \frac{2B\alpha^2}{4-\alpha^2}.$$

Therefore

$$C_2 - C_1 = \frac{\alpha^3(A-B) + \alpha^2(2B-A) - 4\alpha(A+B) + 4A}{(1-\alpha)(4-\alpha^2)}.$$

Viewed as a function of α, the numerator is a cubic taking the value $4A$ at $\alpha = 0$ and the value $-3B$ at $\alpha = 1$. This cubic has a unique zero at some $\alpha^* \in (0, 1)$, and $C_1 < C_2$ if and only if $0 < \alpha < \alpha^*$.

14. The state of the system is the number $Q(t)$ of customers within it at time t. The state 1 may be divided into two sub-states, being σ_1 and σ_2, where σ_i is the state in which server i is occupied but the other server is not. The state space is therefore $S = \{0, \sigma_1, \sigma_2, 2, 3, \dots\}$.

The usual way of finding the stationary distribution, when it exists, is to solve the equilibrium equations. An alternative is to argue as follows. If there exists a stationary distribution, then the process is reversible in equilibrium if and only if

$$(*) \qquad g_{i_1,i_2} g_{i_2,i_3} \cdots g_{i_k,i_1} = g_{i_1,i_k} g_{i_k,i_{k-1}} \cdots g_{i_2,i_1}$$

for all sequences $i_1, i_2, \dots, i_k$ of states, where $\mathbf{G} = (g_{uv})_{u,v \in S}$ is the generator of the process (this may be shown in very much the same way as was the corresponding claim for discrete-time chains in Exercise (6.5.3); see also Problem (6.15.16)). It is clear that $(*)$ is satisfied by this process for all sequences of states which do not include both σ_1 and σ_2; this holds since the terms g_{uv} are exactly those of a birth–death process in such a case. In order to see that $(*)$ holds for a sequence containing both σ_1 and σ_2, it suffices to perform the following calculation:

$$g_{0,\sigma_1} g_{\sigma_1,2} g_{2,\sigma_2} g_{\sigma_2,0} = (\tfrac{1}{2}\lambda)\lambda\mu_2\mu_1 = g_{0,\sigma_2} g_{\sigma_2,2} g_{2,\sigma_1} g_{\sigma_1,0}.$$

Since the process is reversible in equilibrium, the stationary distribution $\boldsymbol{\pi}$ satisfies $\pi_u g_{uv} = \pi_v g_{vu}$ for all $u, v \in S$, $u \ne v$. Therefore

$$\pi_u \lambda = \pi_{u+1}(\mu_1 + \mu_2), \qquad u \ge 2,$$
$$\tfrac{1}{2}\pi_0\lambda = \pi_{\sigma_1}\mu_1 = \pi_{\sigma_2}\mu_2, \quad \pi_{\sigma_1}\lambda = \pi_2\mu_2, \quad \pi_{\sigma_2}\lambda = \pi_2\mu_1,$$

and hence

$$\pi_{\sigma_1} = \frac{\lambda}{2\mu_1}\pi_0, \quad \pi_{\sigma_2} = \frac{\lambda}{2\mu_2}\pi_0, \quad \pi_u = \frac{\lambda^2}{2\mu_1\mu_2}\left(\frac{\lambda}{\mu_1+\mu_2}\right)^{u-2}\pi_0 \quad \text{for } u \ge 2.$$

This gives a stationary distribution if and only if $\lambda < \mu_1 + \mu_2$, under which assumption π_0 is easily calculated.

A similar analysis is valid if there are s servers and an arriving customer is equally likely to go to any free server, otherwise waiting in turn. This process also is reversible in equilibrium, and the stationary distribution is similar to that given above.

15. We have from the standard theory that Q_μ has as mass function $\pi_j = (1-\eta)\eta^j$, $j \ge 0$, where η is the smallest positive root of the equation $x = e^{\mu(x-1)}$. The moment generating function of $(1-\mu^{-1})Q_\mu$ is

$$M_\mu(\theta) = \mathbb{E}\big(\exp\{\theta(1-\mu^{-1})Q_\mu\}\big) = \frac{1-\eta}{1-\eta e^{\theta(1-\mu^{-1})}}.$$

Writing $\mu = 1+\epsilon$, we have by expanding $e^{\mu(\eta-1)}$ as a Taylor series that $\eta = \eta(\epsilon) = 1 - 2\epsilon + o(\epsilon)$ as $\epsilon \downarrow 0$. This gives

$$M_\mu(\theta) = \frac{2\epsilon + o(\epsilon)}{1-(1-2\epsilon)(1+\theta\epsilon)+o(\epsilon)} = \frac{2\epsilon+o(\epsilon)}{(2-\theta)\epsilon + o(\epsilon)} \to \frac{2}{2-\theta}$$

as $\epsilon \downarrow 0$, implying the result, by the continuity theorem.

16. The numbers P (of passengers) and T (of taxis) up to time t have the Poisson distribution with respective parameters πt and τt. The required probabilities $p_n = \mathbb{P}(P = T+n)$ have generating function

$$\begin{aligned}\sum_{n=-\infty}^{\infty} p_n z^n &= \sum_{n=-\infty}^{\infty}\sum_{m=0}^{\infty} \mathbb{P}(P=m+n)\mathbb{P}(T=m)z^n \\ &= \sum_{m=0}^{\infty} \mathbb{P}(T=m)z^{-m}G_P(z) \\ &= G_T(z^{-1})G_P(z) = e^{-(\pi+\tau)t}e^{(\pi z+\tau z^{-1})t},\end{aligned}$$

in which the coefficient of z^n is easily found to be that given.

17. Let $N(t)$ be the number of machines which have arrived by time t. Given that $N(t) = n$, the times $T_1, T_2, \dots, T_n$ of their arrivals may be thought of as the order statistics of a family of independent uniform variables on $[0,t]$, say $U_1, U_2, \dots, U_n$; see Theorem (6.12.7). The machine which arrived at time U_i is, at time t,

$$\left.\begin{array}{r}\text{in the } X\text{-stage}\\ \text{in the } Y\text{-stage}\\ \text{repaired}\end{array}\right\} \text{ with probability } \left\{\begin{array}{l}\alpha(t)\\ \beta(t)\\ 1-\alpha(t)-\beta(t)\end{array}\right.$$

where $\alpha(t) = \mathbb{P}(U+X>t)$ and $\beta(t) = \mathbb{P}(U+X \le t < U+X+Y)$, where U is uniform on $[0,t]$, and (X,Y) is a typical repair pair, independent of U. Therefore

$$\mathbb{P}\big(U(t)=j, V(t)=k \,\big|\, N(t)=n\big) = \frac{n!\,\alpha(t)^j\beta(t)^k(1-\alpha(t)-\beta(t))^{n-k-j}}{j!\,k!\,(n-j-k)!},$$

implying that

$$\begin{aligned}\mathbb{P}\big(U(t)=j, V(t)=k\big) &= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!}\mathbb{P}\big(U(t)=j, V(t)=k \,\big|\, N(t)=n\big) \\ &= \frac{\{\lambda t\alpha(t)\}^j e^{-\lambda t\alpha(t)}}{j!}\cdot\frac{\{\lambda t\beta(t)\}^k e^{-\lambda t\beta(t)}}{k!}.\end{aligned}$$

18. The maximum deficit M_n seen up to and including the time of the nth claim satisfies

$$M_n = \max\left\{M_{n-1}, \sum_{j=1}^{n}(K_j - X_j)\right\} = \max\{0, U_1, U_1 + U_2, \dots, U_1 + U_2 + \cdots + U_n\},$$

where the X_j are the inter-claim times, and $U_j = K_j - X_j$. We have as in the analysis of G/G/1 that M_n has the same distribution as $V_n = \max\{0, U_n, U_n + U_{n-1}, \dots, U_n + U_{n-1} + \cdots + U_1\}$, whence M_n has the same distribution as the $(n+1)$th waiting time in a M(λ)/G/1 queue with service times K_j and interarrival times X_j. The result follows by Theorem (11.3.16).

19. (a) Look for a solution to the detailed balance equations $\lambda\pi_i = (i+1)\mu\pi_{i+1}$, $0 \le i < s$, to find that the stationary distribution is given by $\pi_i = (\rho^i/i!)\pi_0$.

(b) Let p_c be the required fraction. We have by Little's theorem (10.5.18) that

$$p_c = \frac{\lambda(\pi_{c-1} - \pi_c)}{\mu} = \rho(\pi_{c-1} - \pi_c), \quad c \ge 2,$$

and $p_1 = \pi_1$, where π_s is the probability that channels $1, 2, \dots, s$ are busy in a queue M/M/s having the property that further calls are lost when all s servers are occupied.

12

Martingales

12.1 Solutions. Introduction

1. (i) We have that $\mathbb{E}(Y_m) = \mathbb{E}\{\mathbb{E}(Y_{m+1} \mid \mathcal{F}_m)\} = \mathbb{E}(Y_{m+1})$, and the result follows by induction.
(ii) For a submartingale, $\mathbb{E}(Y_m) \le \mathbb{E}\{\mathbb{E}(Y_{m+1} \mid \mathcal{F}_m)\} = \mathbb{E}(Y_{m+1})$, and the result for supermartingales follows similarly.

2. We have that

$$\mathbb{E}(Y_{n+m} \mid \mathcal{F}_n) = \mathbb{E}\big\{\mathbb{E}(Y_{n+m} \mid \mathcal{F}_{n+m-1}) \,\big|\, \mathcal{F}_n\big\} = \mathbb{E}(Y_{n+m-1} \mid \mathcal{F}_n)$$

if $m \ge 1$, since $\mathcal{F}_n \subseteq \mathcal{F}_{n+m-1}$. Iterate to obtain $\mathbb{E}(Y_{n+m} \mid \mathcal{F}_n) = \mathbb{E}(Y_n \mid \mathcal{F}_n) = Y_n$.

3. (i) $Z_n \mu^{-n}$ has mean 1, and

$$\mathbb{E}\big(Z_{n+1}\mu^{-(n+1)} \,\big|\, \mathcal{F}_n\big) = \mu^{-(n+1)}\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = \mu^{-n} Z_n,$$

where $\mathcal{F}_n = \sigma(Z_1, Z_2, \dots, Z_n)$.
(ii) Certainly $\eta^{Z_n} \le 1$, and therefore it has finite mean. Also,

$$\mathbb{E}\big(\eta^{Z_{n+1}} \,\big|\, \mathcal{F}_n\big) = \mathbb{E}\left(\eta^{\sum_1^{Z_n} X_i} \,\middle|\, \mathcal{F}_n\right) = G(\eta)^{Z_n}$$

where the X_i are independent family sizes with probability generating function G. Now $G(\eta) = \eta$, and the claim follows.

4. (i) With X_n denoting the size of the nth jump,

$$\mathbb{E}(S_{n+1} \mid \mathcal{F}_n) = S_n + \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = S_n$$

where $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Also $\mathbb{E}|S_n| \le n$, so that $\{S_n\}$ is a martingale.
(ii) Similarly $\mathbb{E}(S_n^2) = \operatorname{var}(S_n) = n$, and

$$\mathbb{E}\big(S_{n+1}^2 - (n+1) \,\big|\, \mathcal{F}_n\big) = S_n^2 + \mathbb{E}(X_{n+1}^2) + 2S_n\mathbb{E}(X_{n+1}) - (n+1) = S_n^2 - n.$$

(iii) Suppose the walk starts at k, and there are absorbing barriers at 0 and N ($\ge k$). Let T be the time at which the walk is absorbed, and make the assumptions that $\mathbb{E}(S_T) = S_0$, $\mathbb{E}(S_T^2 - T) = S_0^2$. Then the probability p_k of ultimate ruin satisfies

$$0 \cdot p_k + N \cdot (1 - p_k) = k, \qquad 0 \cdot p_k + N^2 \cdot (1 - p_k) - \mathbb{E}(T) = k^2,$$

and therefore $p_k = 1 - (k/N)$ and $\mathbb{E}(T) = k(N-k)$.

5. (i) By Exercise (12.1.2), for $r \geq i$,

$$\mathbb{E}(Y_r Y_i) = \mathbb{E}\{\mathbb{E}(Y_r Y_i \mid \mathcal{F}_i)\} = \mathbb{E}\{Y_i \mathbb{E}(Y_r \mid \mathcal{F}_i)\} = \mathbb{E}(Y_i^2),$$

an answer which is independent of r. Therefore

$$\mathbb{E}\{(Y_k - Y_j)Y_i\} = \mathbb{E}(Y_k Y_i) - \mathbb{E}(Y_j Y_i) = 0 \qquad \text{if } i \leq j \leq k.$$

(ii) We have that

$$\mathbb{E}\{(Y_k - Y_j)^2 \mid \mathcal{F}_i\} = \mathbb{E}(Y_k^2 \mid \mathcal{F}_i) - 2\mathbb{E}(Y_k Y_j \mid \mathcal{F}_i) + \mathbb{E}(Y_j^2 \mid \mathcal{F}_i).$$

Now $\mathbb{E}(Y_k Y_j \mid \mathcal{F}_i) = \mathbb{E}\{\mathbb{E}(Y_k Y_j \mid \mathcal{F}_j) \mid \mathcal{F}_i\} = \mathbb{E}(Y_j^2 \mid \mathcal{F}_i)$, and the claim follows.
(iii) Taking expectations of the last conclusion,

$$(*) \qquad 0 \leq \mathbb{E}\{(Y_k - Y_j)^2\} = \mathbb{E}(Y_k^2) - \mathbb{E}(Y_j^2), \qquad j \leq k.$$

Now $\{\mathbb{E}(Y_n^2) : n \geq 1\}$ is non-decreasing and bounded, and therefore converges. Therefore, by $(*)$, $\{Y_n : n \geq 1\}$ is Cauchy convergent in mean square, and therefore convergent in mean square, by Problem (7.11.11).

6. (i) Using Jensen's inequality (Exercise (7.9.4)),

$$\mathbb{E}(u(Y_{n+1}) \mid \mathcal{F}_n) \geq u(\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n)) = u(Y_n).$$

(ii) It suffices to note that $|x|$, x^2, and x^+ are convex functions of x; draw pictures if you are in doubt about these functions.

7. (i) This follows just as in Exercise (12.1.6), using the fact that $u\{\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n)\} \geq u(Y_n)$ in this case.
(ii) The function x^+ is convex and non-decreasing. Finally, let $\{S_n : n \geq 0\}$ be a simple random walk whose steps are +1 with probability $p\ (= 1 - q > \frac{1}{2})$ and -1 otherwise. If $S_n < 0$, then

$$\mathbb{E}\left(|S_{n+1}| \mid \mathcal{F}_n\right) = p(|S_n| - 1) + q(|S_n| + 1) = |S_n| - (p - q) < |S_n|;$$

note that $\mathbb{P}(S_n < 0) > 0$ if $n \geq 1$. The same example suffices in the remaining case.

8. Clearly $\mathbb{E}|\lambda^{-n}\psi(X_n)| \leq \lambda^{-n} \sup\{|\psi(j)| : j \in S\}$. Also,

$$\mathbb{E}(\psi(X_{n+1}) \mid \mathcal{F}_n) = \sum_{j \in S} p_{X_n, j} \psi(j) \leq \lambda \psi(X_n)$$

where $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Divide by λ^{n+1} to obtain that the given sequence is a supermartingale.

9. Since $\text{var}(Z_1) > 0$, the function G, and hence also G_n, is a strictly increasing function on [0, 1]. Since $1 = G_{n+1}(H_{n+1}(s)) = G_n(G(H_{n+1}(s)))$ and $G_n(H_n(s)) = 1$, we have that $G(H_{n+1}(s)) = H_n(s)$. With $\mathcal{F}_m = \sigma(Z_k : 0 \leq k \leq m)$,

$$\mathbb{E}(H_{n+1}(s)^{Z_{n+1}} \mid \mathcal{F}_n) = G(H_{n+1}(s))^{Z_n} = H_n(s)^{Z_n}.$$

12.2 Solutions. Martingale differences and Hoeffding's inequality

1. Let $\mathcal{F}_i = \sigma(\{V_j, W_j : 1 \le j \le i\})$ and $Y_i = \mathbb{E}(Z \mid \mathcal{F}_i)$. With $Z(j)$ the maximal worth attainable without using the jth object, we have that

$$\mathbb{E}\big(Z(j) \,\big|\, \mathcal{F}_j\big) = \mathbb{E}\big(Z(j) \,\big|\, \mathcal{F}_{j-1}\big), \qquad Z(j) \le Z \le Z(j) + M.$$

Take conditional expectations of the second inequality, given $\mathcal{F}_j$ and given $\mathcal{F}_{j-1}$, and deduce that $|Y_j - Y_{j-1}| \le M$. Therefore Y is a martingale with bounded differences, and Hoeffding's inequality yields the result.

2. Let $\mathcal{F}_i$ be the σ-field generated by the (random) edges joining pairs (v_a, v_b) with $1 \le a, b \le i$, and let $\chi_i = \mathbb{E}(\chi \mid \mathcal{F}_i)$. We write $\chi(j)$ for the minimal number of colours required in order to colour each vertex in the graph obtained by deleting v_j. The argument now follows that of the last exercise, using the fact that $\chi(j) \le \chi \le \chi(j) + 1$.

12.3 Solutions. Crossings and convergence

1. Let $T_1 = \min\{n : Y_n \ge b\}$, $T_2 = \min\{n > T_1 : Y_n \le a\}$, and define T_k inductively by

$$T_{2k-1} = \min\{n > T_{2k-2} : Y_n \ge b\}, \qquad T_{2k} = \min\{n > T_{2k-1} : Y_n \le a\}.$$

The number of downcrossings by time n is $D_n(a, b; Y) = \max\{k : T_{2k} \le n\}$.

(a) Between each pair of upcrossings of $[a, b]$, there must be a downcrossing, and *vice versa*. Hence $|D_n(a, b; Y) - U_n(a, b; Y)| \le 1$.

(b) Let I_i be the indicator function of the event that $i \in (T_{2k-1}, T_{2k}]$ for some k, and let

$$Z_n = \sum_{i=1}^{n} I_i(Y_i - Y_{i-1}), \qquad n \ge 0.$$

It is easily seen that

$$Z_n \le -(b - a)D_n(a, b; Y) + (Y_n - b)^+,$$

whence

$$(*) \qquad (b - a)\mathbb{E}D_n(a, b; Y) \le \mathbb{E}\big\{(Y_n - b)^+\big\} - \mathbb{E}(Z_n).$$

Now I_i is $\mathcal{F}_{i-1}$-measurable, since

$$\{I_i = 1\} = \bigcup_k \big(\{T_{2k-1} \le i - 1\} \setminus \{T_{2k} \le i - 1\}\big).$$

Therefore,

$$\mathbb{E}(Z_n - Z_{n-1}) = \mathbb{E}\big\{\mathbb{E}\big(I_n(Y_n - Y_{n-1}) \,\big|\, \mathcal{F}_{n-1}\big)\big\} = \mathbb{E}\big\{I_n\big(\mathbb{E}(Y_n \mid \mathcal{F}_{n-1}) - Y_{n-1}\big)\big\} \ge 0$$

since $I_n \ge 0$ and Y is a submartingale. It follows that $\mathbb{E}(Z_n) \ge \mathbb{E}(Z_{n-1}) \ge \cdots \ge \mathbb{E}(Z_0) = 0$, and the final inequality follows from $(*)$.

2. If Y is a supermartingale, then $-Y$ is a submartingale. Upcrossings of $[a, b]$ by Y correspond to downcrossings of $[-b, -a]$ by $-Y$, so that

$$\mathbb{E}U_n(a, b; Y) = \mathbb{E}D_n(-b, -a; -Y) \le \frac{\mathbb{E}\{(-Y_n + a)^+\}}{b - a} = \frac{\mathbb{E}\{(Y_n - a)^-\}}{b - a},$$

by Exercise (12.3.1). If $a, Y_n \geq 0$ then $(Y_n - a)^- \leq a$.

3. The random sequence $\{\psi(X_n) : n \geq 1\}$ is a bounded supermartingale, which converges a.s. to some limit Y. The chain is irreducible and persistent, so that each state is visited infinitely often a.s.; it follows that $\lim_{n\to\infty} \psi(X_n)$ cannot exist (a.s.) unless ψ is a constant function.

4. Y is a martingale since Y_n is the sum of independent variables with zero means. Also $\sum_1^\infty \mathbb{P}(Z_n \neq 0) = \sum_1^\infty n^{-2} < \infty$, implying by the Borel–Cantelli lemma that $Z_n = 0$ except for finitely many values of n (a.s.); therefore the partial sum Y_n converges a.s. as $n \to \infty$ to some finite limit.

It is easily seen that $a_n = 5a_{n-1}$ and therefore $a_n = 8 \cdot 5^{n-2}$, if $n \geq 3$. It follows that $|Y_n| \geq \frac{1}{2}a_n$ if and only if $|Z_n| = a_n$. Therefore

$$\mathbb{E}|Y_n| \geq \tfrac{1}{2}a_n\mathbb{P}(|Y_n| \geq \tfrac{1}{2}a_n) = \tfrac{1}{2}a_n\mathbb{P}(|Z_n| = a_n) = \frac{a_n}{2n^2}$$

which tends to infinity as $n \to \infty$.

12.4 Solutions. Stopping times

1. We have that

$$\{T_1 + T_2 = n\} = \bigcup_{k=0}^{n} \big(\{T_1 = k\} \cap \{T_2 = n-k\}\big),$$
$$\big\{\max\{T_1, T_2\} \leq n\big\} = \{T_1 \leq n\} \cap \{T_2 \leq n\},$$
$$\big\{\min\{T_1, T_2\} \leq n\big\} = \{T_1 \leq n\} \cup \{T_2 \leq n\}.$$

Each event on the right-hand side lies in $\mathcal{F}_n$.

2. Let $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$ and $S_n = X_1 + X_2 + \cdots + X_n$. Now

$$\{N(t) + 1 = n\} = \{S_{n-1} \leq t\} \cap \{S_n > t\} \in \mathcal{F}_n.$$

3. $(Y^+, \mathcal{F})$ is a submartingale, and $T = \min\{k : Y_k \geq x\}$ is a stopping time. Now $0 \leq T \wedge n \leq n$, so that $\mathbb{E}(Y_0^+) \leq \mathbb{E}(Y_{T\wedge n}^+) \leq \mathbb{E}(Y_n^+)$, whence

$$\mathbb{E}(Y_n^+) \geq \mathbb{E}\big(Y_{T\wedge n}^+ I_{\{T\leq n\}}\big) \geq x\mathbb{P}(T \leq n).$$

4. We may suppose that $\mathbb{E}(Y_0) < \infty$. With the notation of the previous solution, we have that

$$\mathbb{E}(Y_0) \geq \mathbb{E}\big(Y_{T\wedge n}\big) \geq \mathbb{E}\big(Y_{T\wedge n} I_{\{T\leq n\}}\big) \geq x\mathbb{P}(T \leq n).$$

5. It suffices to prove that $\mathbb{E}Y_S \leq \mathbb{E}Y_T$, since the other inequalities are of the same form but with different choices of pairs of stopping times. Let I_m be the indicator function of the event $\{S < m \leq T\}$, and define

$$Z_n = \sum_{m=1}^{n} I_m(Y_m - Y_{m-1}), \qquad 0 \leq n \leq N.$$

Note that I_m is $\mathcal{F}_{m-1}$-measurable, so that

$$\mathbb{E}(Z_n - Z_{n-1}) = \mathbb{E}\big\{I_n\mathbb{E}(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1})\big\} \geq 0,$$

since Y is a submartingale. Therefore $\mathbb{E}(Z_N) \geq \mathbb{E}(Z_{N-1}) \geq \cdots \geq \mathbb{E}(Z_0) = 0$. On the other hand, $Z_N = Y_T - Y_S$, and therefore $\mathbb{E}(Y_T) \geq \mathbb{E}(Y_S)$.

6. De Moivre's martingale is $Y_n = (q/p)^{S_n}$, where $q = 1 - p$. Now $Y_n \geq 0$, and $\mathbb{E}(Y_0) = 1$, and the maximal inequality gives that

$$\mathbb{P}\left(\max_{0\leq m\leq n} S_m \geq x\right) = \mathbb{P}\left(\max_{0\leq m\leq n} Y_m \geq (q/p)^x\right) \leq (p/q)^x.$$

Take the limit as $n \to \infty$ to find that $S_\infty = \sup_m S_m$ satisfies

$$(*) \qquad \mathbb{E}(S_\infty) = \sum_{x=1}^{\infty} \mathbb{P}(S_\infty \geq x) \leq \frac{p}{q-p}.$$

We can calculate $\mathbb{E}(S_\infty)$ exactly as follows. It is the case that $S_\infty \geq x$ if and only if the walk ever visits the point x, an event with probability f^x for $x \geq 0$, where $f = p/q$ (see Exercise (5.3.1)). The inequality of $(*)$ may be replaced by equality.

7. (a) First, $\varnothing \cap \{T \leq n\} = \varnothing \in \mathcal{F}_n$. Secondly, if $A \cap \{T \leq n\} \in \mathcal{F}_n$ then

$$A^c \cap \{T \leq n\} = \{T \leq n\} \setminus \big(A \cap \{T \leq n\}\big) \in \mathcal{F}_n.$$

Thirdly, if $A_1, A_2, \ldots$ satisfy $A_i \cap \{T \leq n\} \in \mathcal{F}_n$ for each i, then

$$\left(\bigcup_i A_i\right) \cap \{T \leq n\} = \bigcup_i \big(A_i \cap \{T \leq n\}\big) \in \mathcal{F}_n.$$

Therefore $\mathcal{F}_T$ is a σ-field.

For each integer m, it is the case that

$$\{T \leq m\} \cap \{T \leq n\} = \begin{cases} \{T \leq n\} & \text{if } m > n, \\ \{T \leq m\} & \text{if } m \leq n, \end{cases}$$

an event lying in $\mathcal{F}_n$. Therefore $\{T \leq m\} \in \mathcal{F}_T$ for all m.
(b) Let $A \in \mathcal{F}_S$. Then, for any n,

$$\big(A \cap \{S \leq T\}\big) \cap \{T \leq n\} = \bigcup_{m=0}^{n} \big(A \cap \{S \leq m\}\big) \cap \{T = m\},$$

the union of events in $\mathcal{F}_n$, which therefore lies in $\mathcal{F}_n$. Hence $A \cap \{S \leq T\} \in \mathcal{F}_T$.
(c) We have $\{S \leq T\} = \Omega$, and (b) implies that $A \in \mathcal{F}_T$ whenever $A \in \mathcal{F}_S$.

12.5 Solutions. Optional stopping

1. Under the conditions of (a) or (b), the family $\{Y_{T\wedge n} : n \geq 0\}$ is uniformly integrable. Now $T \wedge n \to T$ as $n \to \infty$, so that $Y_{T\wedge n} \to Y_T$ a.s. Using uniform integrability, $\mathbb{E}(Y_{T\wedge n}) \to \mathbb{E}(Y_T)$, and the claim follows by the fact that $\mathbb{E}(Y_{T\wedge n}) = \mathbb{E}(Y_0)$.

2. It suffices to prove that $\{Y_{T\wedge n} : n \geq 0\}$ is uniformly integrable. Recall that $\{X_n : n \geq 0\}$ is uniformly integrable if

$$\lim_{a\to\infty} \left\{\sup_n \mathbb{E}\left(|X_n| I_{\{|X_n|\geq a\}}\right)\right\} \to 0 \quad \text{as } a \to \infty.$$

(a) Now,

$$\begin{aligned}\mathbb{E}\left(|Y_{T\wedge n}|I_{\{|Y_{T\wedge n}|\geq a\}}\right) &= \mathbb{E}\left(|Y_T|I_{\{T\leq n,|Y_T|\geq a\}}\right) + \mathbb{E}\left(|Y_n|I_{\{T>n,|Y_n|\geq a\}}\right)\\ &\leq \mathbb{E}\left(|Y_T|I_{\{|Y_T|\geq a\}}\right) + \mathbb{E}\left(|Y_n|I_{\{T>n\}}\right) = g(a) + h(n),\end{aligned}$$

say. We have that $g(a) \to 0$ as $a \to \infty$, since $\mathbb{E}|Y_T| < \infty$. Also $h(n) \to 0$ as $n \to \infty$, so that $\sup_{n\geq N} h(n)$ may be made arbitrarily small by suitable choice of N. On the other hand, $\mathbb{E}\left(|Y_n|I_{\{|Y_n|\geq a\}}\right) \to 0$ as $a \to \infty$ *uniformly* in $n \in \{0, 1, \dots, N\}$, and the claim follows.

(b) Since Y_n^+ defines a submartingale, we have that $\sup_n \mathbb{E}(Y_{T\wedge n}^+) \leq \sup_n \mathbb{E}(Y_n^+) < \infty$, the second inequality following by the uniform integrability of $\{Y_n\}$. Using the martingale convergence theorem, $Y_{T\wedge n} \to Y_T$ a.s. where $\mathbb{E}|Y_T| < \infty$. Now

$$\mathbb{E}|Y_{T\wedge n} - Y_T| = \mathbb{E}\big(|Y_n - Y_T|I_{\{T>n\}}\big) \leq \mathbb{E}\big(|Y_n|I_{\{T>n\}}\big) + \mathbb{E}\big(|Y_T|I_{\{T>n\}}\big).$$

Also $\mathbb{P}(T > n) \to 0$ as $n \to \infty$, so that the final two terms tend to 0 (by the uniform integrability of the Y_i and the finiteness of $\mathbb{E}|Y_T|$ respectively). Therefore $Y_{T\wedge n} \xrightarrow{1} Y_T$, and the claim follows by the standard theorem (7.10.3).

3. By uniform integrability, $Y_\infty = \lim_{n\to\infty} Y_n$ exists a.s. and in mean, and $Y_n = \mathbb{E}(Y_\infty \mid \mathcal{F}_n)$.

(a) On the event $\{T = n\}$ it is the case that $Y_T = Y_n$ and $\mathbb{E}(Y_\infty \mid \mathcal{F}_T) = \mathbb{E}(Y_\infty \mid \mathcal{F}_n)$; for the latter statement, use the definition of conditional expectation. It follows that $Y_T = \mathbb{E}(Y_\infty \mid \mathcal{F}_T)$, irrespective of the value of T.

(b) We have from Exercise (12.4.7) that $\mathcal{F}_S \subseteq \mathcal{F}_T$. Now $Y_S = \mathbb{E}(Y_\infty \mid \mathcal{F}_S) = \mathbb{E}\{\mathbb{E}(Y_\infty \mid \mathcal{F}_T) \mid \mathcal{F}_S\} = \mathbb{E}(Y_T \mid \mathcal{F}_S)$.

4. Let T be the time until absorption, and note that $\{S_n\}$ is a bounded, and therefore uniformly integrable, martingale. Also $\mathbb{P}(T < \infty) = 1$ since T is no larger than the waiting time for N consecutive steps in the same direction. It follows that $\mathbb{E}(S_0) = \mathbb{E}(S_T) = N\mathbb{P}(S_T = N)$, so that $\mathbb{P}(S_T = N) = \mathbb{E}(S_0)/N$. Secondly, $\{S_n^2 - n : n \geq 0\}$ is a martingale (see Exercise (12.1.4)), and the optional stopping theorem (if it may be applied) gives that

$$\mathbb{E}(S_0^2) = \mathbb{E}(S_T^2 - T) = N^2\mathbb{P}(S_T = N) - \mathbb{E}(T),$$

and hence $\mathbb{E}(T) = N\mathbb{E}(S_0) - \mathbb{E}(S_0^2)$ as required.

It remains to check the conditions of the optional stopping theorem. Certainly $\mathbb{P}(T < \infty) = 1$, and in addition $\mathbb{E}(T^2) < \infty$ by the argument above. We have that $\mathbb{E}|S_T^2 - T| \leq N^2 + \mathbb{E}(T) < \infty$. Finally,

$$\mathbb{E}\big\{(S_n^2 - n)I_{\{T>n\}}\big\} \leq (N^2 + n)\mathbb{P}(T > n) \to 0$$

as $n \to \infty$, since $\mathbb{E}(T^2) < \infty$.

5. Let $\mathcal{F}_n = \sigma(S_1, S_2, \dots, S_n)$. It is immediate from the identity $\cos(A + \lambda) + \cos(A - \lambda) = 2\cos A \cos\lambda$ that

$$\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = \frac{\cos[\lambda(S_n + 1 - \frac{1}{2}(b-a))] + \cos[\lambda(S_n - 1 - \frac{1}{2}(b-a))]}{2(\cos\lambda)^{n+1}} = Y_n,$$

and therefore Y is a martingale (it is easy to see that $\mathbb{E}|Y_n| < \infty$ for all n).

Suppose that $0 < \lambda < \pi/(a + b)$, and note that $0 \leq |\lambda\{S_n - \frac{1}{2}(b-a)\}| < \frac{1}{2}\lambda(a+b) < \frac{1}{2}\pi$ for $n \leq T$. Now $Y_{T\wedge n}$ constitutes a martingale which satisfies

$$(*)\qquad \frac{\cos\{\frac{1}{2}\lambda(a+b)\}}{(\cos\lambda)^{T\wedge n}} \leq Y_{T\wedge n} \leq \frac{1}{(\cos\lambda)^T}.$$

If we can prove that $\mathbb{E}\{(\cos\lambda)^{-T}\} < \infty$, it will follow that $\{Y_{T\wedge n}\}$ is uniformly integrable. This will imply in turn that $\mathbb{E}(Y_T) = \lim_{n\to\infty} \mathbb{E}(Y_{T\wedge n}) = \mathbb{E}(Y_0)$, and therefore

$$\cos\{\tfrac{1}{2}\lambda(a+b)\}\mathbb{E}\{(\cos\lambda)^{-T}\} = \cos\{\tfrac{1}{2}\lambda(b-a)\}$$

as required. We have from (∗) that

$$\mathbb{E}(Y_0) = \mathbb{E}(Y_{T\wedge n}) \ge \cos\{\tfrac{1}{2}\lambda(a+b)\}\mathbb{E}\{(\cos\lambda)^{-T\wedge n}\}.$$

Now $T \wedge n \to T$ as $n \to \infty$, implying by Fatou's lemma that

$$\mathbb{E}\{(\cos\lambda)^{-T}\} \le \frac{\mathbb{E}(Y_0)}{\cos\{\frac{1}{2}\lambda(a+b)\}} = \frac{\cos\{\frac{1}{2}\lambda(a-b)\}}{\cos\{\frac{1}{2}\lambda(a+b)\}}.$$

6. (a) The occurrence of the event $\{U = n\}$ depends on $S_1, S_2, \dots, S_n$ only, and therefore U is a stopping time. Think of U as the time until the first sequence of five consecutive heads in a sequence of coins tosses. Using the renewal-theory argument of Problem (10.5.17), we find that $\mathbb{E}(U) = 62$.

(b) Knowledge of $S_1, S_2, \dots, S_n$ is insufficient to determine whether or not $V = n$, and therefore V is not a stopping time. Now $\mathbb{E}(V) = \mathbb{E}(U) - 5 = 57$.

(c) W is a stopping time, since it is a first-passage time. Also $\mathbb{E}(W) = \infty$ since the walk is *null* persistent.

7. With the usual notation,

$$\begin{aligned}\mathbb{E}(M_{m+n} \mid \mathcal{F}_m) &= \mathbb{E}\left(\sum_{r=0}^{m} S_r + \sum_{r=m+1}^{m+n} S_r - \tfrac{1}{3}(S_{m+n} - S_m + S_m)^3 \,\middle|\, \mathcal{F}_m\right)\\ &= M_m + nS_m - S_m\mathbb{E}\{(S_{m+n} - S_m)^2\}\\ &= M_m + nS_m - nS_m\mathbb{E}(X_1^2) = M_m.\end{aligned}$$

Thus $\{M_n : n \ge 0\}$ is a martingale, and evidently T is a stopping time. The conditions of the optional stopping theorem (12.5.1) hold, and therefore, by a result of Example (3.9.6),

$$a - \tfrac{1}{3}a^3 = M_0 = \mathbb{E}(M_T) = \mathbb{E}\left(\sum_{r=0}^{T} S_r\right) - \tfrac{1}{3}K^3 \cdot \frac{a}{K}.$$

8. We partition the sequence into consecutive batches of $a + b$ flips. If any such batch contains only 1's, then the game is over. Hence $\mathbb{P}(T > n(a+b)) \le \{1 - (\frac{1}{2})^{a+b}\}^n \to 0$ as $n \to \infty$. Therefore,

$$\mathbb{E}|S_T^2 - T| \le \mathbb{E}(S_T^2) + \mathbb{E}(T) \le (a+b)^2 + \mathbb{E}(T) < \infty,$$

and

$$\mathbb{E}\big[(S_T^2 - T)I_{\{T>n\}}\big] \le (a+b)^2\mathbb{P}(T > n) + \mathbb{E}(TI_{\{T>n\}}) \to 0 \qquad \text{as } n \to \infty.$$

12.7 Solutions. Backward martingales and continuous-time martingales

1. Let $s \le t$. We have that $\mathbb{E}(\eta(X(t)) \mid \mathcal{F}_s, X_s = i) = \sum_j p_{ij}(t-s)\eta(j)$. Hence

$$\frac{d}{dt}\mathbb{E}\big(\eta(X(t)) \mid \mathcal{F}_s, X_s = i\big) = \big(\mathbf{P}_{t-s}\mathbf{G}\boldsymbol{\eta}'\big)_i = 0,$$

so that $\mathbb{E}(\eta(X(t)) \mid \mathcal{F}_s, X_s = i) = \eta(i)$, which is to say that $\mathbb{E}(\eta(X(t)) \mid \mathcal{F}_s) = \eta(X(s))$.

2. Let $W(t) = \exp\{-\theta N(t) + \lambda t(1-e^{-\theta})\}$ where $\theta \ge 0$. It may be seen that $W(t \wedge T_a)$, $t \ge 0$, constitutes a martingale. Furthermore

$$|W(t \wedge T_a)| \le \exp\{\lambda(t \wedge T_a)(1-e^{-\theta})\} \uparrow \exp\{\lambda T_a(1-e^{-\theta})\} \quad \text{as } t \to \infty,$$

where, by assumption, the limit has finite expectation for sufficiently small positive θ (this fact may be checked easily). In this case, $\{W(t \wedge T_a) : t \ge 0\}$ is uniformly integrable. Now $W(t \wedge T_a) \to W(T_a)$ a.s. as $t \to \infty$, and it follows by the optional stopping theorem that

$$1 = \mathbb{E}(W(0)) = \mathbb{E}\big(W(t \wedge T_a)\big) \to \mathbb{E}(W(T_a)) = e^{-\theta a}\mathbb{E}\{e^{\lambda T_a(1-e^{-\theta})}\}.$$

Write $s = e^{-\theta}$ to obtain $s^{-a} = \mathbb{E}\{e^{\lambda T_a(1-s)}\}$. Differentiate at $s = 1$ to find that $a = \lambda\mathbb{E}(T_a)$ and $a(a+1) = \lambda^2\mathbb{E}(T_a^2)$, whence the claim is immediate.

3. Let $\mathcal{G}_m$ be the σ-field generated by the two sequences of random variables $S_m, S_{m+1}, \dots, S_n$ and $U_{m+1}, U_{m+2}, \dots, U_n$. It is a straightforward exercise in conditional density functions to see that

$$\mathbb{E}(S_m \mid \mathcal{G}_{m+1}) = \frac{m}{m+1}S_{m+1}, \quad \mathbb{E}(U_{m+1}^{-1} \mid \mathcal{G}_{m+1}) = \int_0^{U_{m+2}} \frac{(m+1)x^{m-1}}{(U_{m+2})^{m+1}}\,dx = \frac{m+1}{mU_{m+2}},$$

whence $\mathbb{E}(R_m \mid \mathcal{G}_{m+1}) = R_{m+1}$ as required. [The integrability condition is elementary.]

Let $T = \max\{m : R_m \ge 1\}$ with the convention that $T = 1$ if $R_m < 1$ for all m. As in the closely related Example (12.7.6), T is a stopping time. We apply the optional stopping theorem (12.7.5) to the backward martingale R to obtain that $\mathbb{E}(R_T \mid \mathcal{G}_n) = R_n = S_n/t$. Now, $R_T \ge 1$ on the event $\{R_m \ge 1 \text{ for some } m \le n\}$, whence

$$\frac{y}{t} = \mathbb{E}(R_T \mid S_n = y) \ge \mathbb{P}\big(R_m \ge 1 \text{ for some } m \le n \,\big|\, S_n = y\big).$$

[Equality may be shown to hold. See Karlin and Taylor 1981, pages 110–113, and Example (12.7.6).]

12.9 Solutions to problems

1. Clearly $\mathbb{E}(Z_n) \le (\mu + m)^n$, and hence $\mathbb{E}|Y_n| < \infty$. Secondly, Z_{n+1} may be expressed as $\sum_{i=1}^{Z_n} X_i + A$, where $X_1, X_2, \dots$ are the family sizes of the members of the nth generation, and A is the number of immigrants to the $(n+1)$th generation. Therefore $\mathbb{E}(Z_{n+1} \mid Z_n) = \mu Z_n + m$, whence

$$\mathbb{E}(Y_{n+1} \mid Z_n) = \frac{1}{\mu^{n+1}}\left\{\mu Z_n + m\left(1 - \frac{1-\mu^{n+1}}{1-\mu}\right)\right\} = Y_n.$$

2. Each birth in the $(n+1)$th generation is to an individual, say the sth, in the nth generation. Hence, for each r, $B_{(n+1),r}$ may be expressed in the form $B_{(n+1),r} = B_{n,s} + B_j'(s)$, where $B_j'(s)$ is the age of the parent when its jth child is born. Therefore

$$\mathbb{E}\left\{\sum_r e^{-\theta B_{(n+1),r}} \,\middle|\, \mathcal{F}_n\right\} = \mathbb{E}\left\{\sum_{s,j} e^{-\theta(B_{n,s}+B_j'(s))} \,\middle|\, \mathcal{F}_n\right\} = \sum_s e^{-\theta B_{n,s}} M_1(\theta),$$

which gives that $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n$. Finally, $\mathbb{E}(Y_1(\theta)) = 1$, and hence $\mathbb{E}(Y_n(\theta)) = 1$.

3. If $x, c > 0$, then

$$(*) \qquad \mathbb{P}\left(\max_{1\le k\le n} Y_k > x\right) \le \mathbb{P}\left(\max_{1\le k\le n} (Y_k + c)^2 > (x+c)^2\right).$$

Now $(Y_k + c)^2$ is a convex function of Y_k, and therefore defines a submartingale (Exercise (12.1.7)). Applying the maximal inequality to this submartingale, we obtain an upper bound of $\mathbb{E}\{(Y_n+c)^2\}/(x+c)^2$ for the right-hand side of $(*)$. We set $c = \mathbb{E}(Y_n^2)/x$ to obtain the result.

4. (a) Note that $Z_n = Z_{n-1} + c_n\{X_n - \mathbb{E}(X_n \mid \mathcal{F}_{n-1})\}$, so that $(Z, \mathcal{F})$ is a martingale. Let T be the stopping time $T = \min\{k : c_k Y_k \ge x\}$. Then $\mathbb{E}(Z_{T\wedge n}) = \mathbb{E}(Z_0) = 0$, so that

$$0 \ge \mathbb{E}\left\{c_{T\wedge n} Y_{T\wedge n} - \sum_{k=1}^{T\wedge n} c_k \mathbb{E}(X_k \mid \mathcal{F}_{k-1})\right\}$$

since the final term in the definition of Z_n is non-negative. Therefore

$$x\mathbb{P}(T \le n) \le \mathbb{E}\{c_{T\wedge n} Y_{T\wedge n}\} \le \sum_{k=1}^{n} c_k \mathbb{E}\{\mathbb{E}(X_k \mid \mathcal{F}_{k-1})\},$$

where we have used the facts that $Y_n \ge 0$ and $\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \ge 0$. The claim follows.

(b) Let $X_1, X_2, \dots$ be independent random variables, with zero means and finite variances, and let $Y_j = \sum_{i=1}^{j} X_i$. Then Y_j^2 defines a non-negative submartingale, whence

$$\mathbb{P}\left(\max_{1\le k\le n} |Y_k| \ge x\right) = \mathbb{P}\left(\max_{1\le k\le n} Y_k^2 \ge x^2\right) \le \frac{1}{x^2}\sum_{k=1}^{n} \mathbb{E}(Y_k^2 - Y_{k-1}^2) = \frac{1}{x^2}\sum_{k=1}^{n} \mathbb{E}(X_k^2).$$

5. The function $h(u) = |u|^r$ is convex, and therefore $Y_i(m) = |S_i - S_m|^r$, $i \ge m$, defines a submartingale with respect to the filtration $\mathcal{F}_i = \sigma(\{X_j : 1 \le j \le i\})$. Apply the HRC inequality of Problem (12.9.4), with $c_k = 1$, to obtain the required inequality.

If $r = 1$, we have that

$$(*) \qquad \mathbb{E}(|S_{m+n} - S_m|) \le \sum_{k=m+1}^{m+n} \mathbb{E}|Z_k|$$

by the triangle inequality. Let $m, n \to \infty$ to find, in the usual way, that the sequence $\{S_n\}$ converges a.s.; Kronecker's lemma (see Exercise (7.8.2)) then yields the final claim.

Suppose $1 < r \le 2$, in which case a little more work is required. The function h is differentiable, and therefore

$$h(v) - h(u) = (v-u)h'(u) + \int_0^{v-u} \{h'(u+x) - h'(u)\}\,dx.$$

Now $h'(y) = r|y|^{r-1}\mathrm{sign}(y)$ has a derivative decreasing in $|y|$. It follows (draw a picture) that $h'(u+x) - h'(u) \le 2h'(\frac{1}{2}x)$ if $x \ge 0$, and therefore the above integral is no larger than $2h(\frac{1}{2}(v-u))$. Apply this with $v = S_{m+k+1} - S_m$ and $u = S_{m+k} - S_m$, to obtain

$$\mathbb{E}(|S_{m+k+1} - S_m|^r) - \mathbb{E}(|S_{m+k} - S_m|^r) \le \mathbb{E}(Z_{m+k+1}h'(S_{m+k} - S_m)) + 2\mathbb{E}(|\tfrac{1}{2}Z_{m+k+1}|^r).$$

Sum over k and use the fact that

$$\mathbb{E}\big(Z_{m+k+1}h'(S_{m+k}-S_m)\big) = \mathbb{E}\big\{h'(S_{m+k}-S_m)\mathbb{E}(Z_{m+k+1} \mid \mathcal{F}_{m+k})\big\} = 0,$$

to deduce that

$$\mathbb{E}\big(|S_{m+n}-S_m|^r\big) \le 2^{2-r}\sum_{k=m+1}^{m+n}\mathbb{E}\big(|Z_k|^r\big).$$

The argument is completed as after $(*)$.

6. With $I_k = I_{\{Y_k=0\}}$, we have that

$$\begin{aligned}\mathbb{E}(Y_n \mid \mathcal{F}_{n-1}) &= \mathbb{E}\Big(X_nI_{n-1} + nY_{n-1}|X_n|(1-I_{n-1}) \,\Big|\, \mathcal{F}_{n-1}\Big)\\ &= I_{n-1}\mathbb{E}(X_n) + nY_{n-1}(1-I_{n-1})\mathbb{E}|X_n| = Y_{n-1}\end{aligned}$$

since $\mathbb{E}(X_n) = 0$, $\mathbb{E}|X_n| = n^{-1}$. Also $\mathbb{E}|Y_n| \le \mathbb{E}\big\{|X_n|(1+n|Y_{n-1}|)\big\}$ and $\mathbb{E}|Y_1| < \infty$, whence $\mathbb{E}|Y_n| < \infty$. Therefore $(Y, \mathcal{F})$ is a martingale.

Now $Y_n = 0$ if and only if $X_n = 0$. Therefore $\mathbb{P}(Y_n = 0) = \mathbb{P}(X_n = 0) = 1 - n^{-1} \to 1$ as $n \to \infty$, implying that $Y_n \xrightarrow{\text{P}} 0$. On the other hand, $\sum_n \mathbb{P}(X_n \ne 0) = \infty$, and therefore $\mathbb{P}(Y_n \ne 0 \text{ i.o.}) = 1$ by the second Borel–Cantelli lemma. However, Y_n takes only integer values, and therefore Y_n does not converge to 0 a.s. The martingale convergence theorem is inapplicable since $\sup_n \mathbb{E}|Y_n| = \infty$.

7. Assume that $t > 0$ and $M(t) = 1$. Then $Y_n = e^{tS_n}$ defines a positive martingale (with mean 1) with respect to $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. By the maximal inequality,

$$\mathbb{P}\left(\max_{1\le k\le n} S_k \ge x\right) = \mathbb{P}\left(\max_{1\le k\le n} Y_k \ge e^{tx}\right) \le e^{-tx}\mathbb{E}(Y_n),$$

and the result follows by taking the limit as $n \to \infty$.

8. The sequence $Y_n = \xi^{Z_n}$ defines a martingale; this may be seen easily, as in the solution to Exercise (12.1.15). Now $\{Y_n\}$ is uniformly bounded, and therefore $Y_\infty = \lim_{n\to\infty} Y_n$ exists a.s. and satisfies $\mathbb{E}(Y_\infty) = \mathbb{E}(Y_0) = \xi$.

Suppose $0 < \xi < 1$. In this case Z_1 is not a.s. zero, so that Z_n cannot converge a.s. to a constant c unless $c \in \{0, \infty\}$. Therefore the a.s. convergence of Y_n entails the a.s. convergence of Z_n to a limit random variable taking values 0 and ∞. In this case, $\mathbb{E}(Y_\infty) = 1 \cdot \mathbb{P}(Z_n \to 0) + 0 \cdot \mathbb{P}(Z_n \to \infty)$, implying that $\mathbb{P}(Z_n \to 0) = \xi$, and therefore $\mathbb{P}(Z_n \to \infty) = 1 - \xi$.

9. It is a consequence of the maximal inequality that $\mathbb{P}(Y_n^* \ge x) \le x^{-1}\mathbb{E}(Y_n I_{\{Y_n^* \ge x\}})$ for $x > 0$. Therefore

$$\begin{aligned}\mathbb{E}(Y_n^*) &= \int_0^\infty \mathbb{P}(Y_n^* \ge x)\,dx \le 1 + \int_1^\infty \mathbb{P}(Y_n^* \ge x)\,dx\\ &\le 1 + \mathbb{E}\left\{Y_n\int_1^\infty x^{-1}I_{(1,Y_n^*]}(x)\,dx\right\}\\ &= 1 + \mathbb{E}(Y_n \log^+ Y_n^*) \le 1 + \mathbb{E}(Y_n\log^+ Y_n) + \mathbb{E}(Y_n^*)/e.\end{aligned}$$

10. (a) We have, as in Exercise (12.7.1), that

$$(*) \qquad \mathbb{E}\big(h(X(t)) \,\big|\, B, X(s) = i\big) = \sum_j p_{ij}(t)h(j) \quad \text{for } s < t,$$

for any event B defined in terms of $\{X(u) : u \le s\}$. The derivative of this expression, with respect to t, is $(\mathbf{P}_t\mathbf{G}\mathbf{h}')_i$, where $\mathbf{P}_t$ is the transition semigroup, $\mathbf{G}$ is the generator, and $\mathbf{h} = (h(j) : j \ge 0)$. In this case,

$$(\mathbf{G}\mathbf{h}')_j = \sum_k g_{jk}h(k) = \lambda_j\{h(j+1) - h(j)\} - \mu_j\{h(j) - h(j-1)\} = 0$$

for all j. Therefore the left side of $(*)$ is constant for $t \ge s$, and is equal to its value at time s, i.e. $X(s)$. Hence $h(X(t))$ defines a martingale.

(b) We apply the optional stopping theorem with $T = \min\{t : X(t) \in \{0, n\}\}$ to obtain $\mathbb{E}(h(X(T))) = \mathbb{E}(h(X(0)))$, and therefore $(1 - \pi(m))h(n) = h(m)$ as required. It is necessary but not difficult to check the conditions of the optional stopping theorem.

11. (a) Since Y is a submartingale, so is Y^+ (see Exercise (12.1.6)). Now

$$\mathbb{E}(Y^+_{n+m+1} \mid \mathcal{F}_n) = \mathbb{E}\{\mathbb{E}(Y^+_{n+m+1} \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n\} \ge \mathbb{E}(Y^+_{n+m} \mid \mathcal{F}_n).$$

Therefore $\{\mathbb{E}(Y^+_{n+m} \mid \mathcal{F}_n) : m \ge 0\}$ is (a.s.) non-decreasing, and therefore converges (a.s.) to a limit M_n. Also, by monotone convergence of conditional expectation,

$$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = \lim_{m\to\infty} \mathbb{E}\{\mathbb{E}(Y^+_{n+m+1} \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n\} = \lim_{m\to\infty} \mathbb{E}(Y^+_{n+m+1} \mid \mathcal{F}_n) = M_n,$$

and furthermore $\mathbb{E}(M_n) = \lim_{m\to\infty} \mathbb{E}(Y^+_{m+n}) \le M$. It is the case that M_n is $\mathcal{F}_n$-measurable, and therefore it is a martingale.

(b) We have that $Z_n = M_n - Y_n$ is the difference of a martingale and a submartingale, and is therefore a supermartingale. Also $M_n \ge Y_n^+ \ge 0$, and the decomposition for Y_n follows.

(c) In this case Z_n is a martingale, being the difference of two martingales. Also $M_n \ge \mathbb{E}(Y_n^+ \mid \mathcal{F}_n) = Y_n^+ \ge Y_n$ a.s., and the claim follows.

12. We may as well assume that $\mu < P$ since the inequality is trivial otherwise. The moment generating function of $P - C_1$ is $M(t) = e^{t(P-\mu)+\frac{1}{2}\sigma^2t^2}$, and we choose t such that $M(t) = 1$, i.e., $t = -2(P-\mu)/\sigma^2$. Now define $Z_n = \min\{e^{tY_n}, 1\}$ and $\mathcal{F}_n = \sigma(C_1, C_2, \dots, C_n)$. Certainly $\mathbb{E}|Z_n| < \infty$; also

$$\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) \le \mathbb{E}(e^{tY_{n+1}} \mid \mathcal{F}_n) = e^{tY_n}M(t) = e^{tY_n}$$

and $\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) \le 1$, implying that $\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) \le Z_n$. Therefore $(Z_n, \mathcal{F}_n)$ is a positive supermartingale. Let $T = \inf\{n : Y_n \le 0\} = \inf\{n : Z_n = 1\}$. Then $T \wedge m$ is a bounded stopping time, whence $\mathbb{E}(Z_0) \ge \mathbb{E}(Z_{T\wedge m}) \ge \mathbb{P}(T \le m)$. Let $m \to \infty$ to obtain the result.

13. Let $\mathcal{F}_n = \sigma(R_1, R_2, \dots, R_n)$.

(a) $0 \le Y_n \le 1$, and Y_n is $\mathcal{F}_n$-measurable. Also

$$\mathbb{E}(R_{n+1} \mid R_n) = R_n + \frac{R_n}{n+r+b},$$

whence Y_n satisfies $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n$. Therefore $\{Y_n : n \ge 0\}$ is a uniformly integrable martingale, and therefore converges a.s. and in mean.

(b) In order to apply the optional stopping theorem, it suffices that $\mathbb{P}(T < \infty) = 1$ (since Y is uniformly integrable). However $\mathbb{P}(T > n) = \frac{1}{2} \cdot \frac{2}{3} \cdots \frac{n}{n+1} = (n+1)^{-1} \to 0$. Using that theorem, $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$, which is to say that $\mathbb{E}\{T/(T+2)\} = \frac{1}{2}$, and the result follows.

(c) Apply the maximal inequality.

14. As in the previous solution, with $\mathcal{G}_n$ the σ-field generated by $A_1, A_2, \dots$ and $\mathcal{F}_n$,

$$\begin{aligned}\mathbb{E}(Y_{n+1} \mid \mathcal{G}_n) &= \left(\frac{R_n + A_n}{R_n + B_n + A_n}\right)\left(\frac{R_n}{R_n + B_n}\right) + \left(\frac{R_n}{R_n + B_n + A_n}\right)\left(\frac{B_n}{R_n + B_n}\right)\\ &= \frac{R_n}{R_n + B_n} = Y_n,\end{aligned}$$

so that $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\{\mathbb{E}(Y_{n+1} \mid \mathcal{G}_n) \mid \mathcal{F}_n\} = Y_n$. Also $|Y_n| \le 1$, and therefore Y_n is a martingale.

We need to show that $\mathbb{P}(T < \infty) = 1$. Let I_n be the indicator function of the event $\{T > n\}$. We have by conditioning on the A_n that

$$\mathbb{E}(I_n \mid \mathbf{A}) = \prod_{j=0}^{n-1}\left(1 - \frac{1}{2 + S_j}\right) \to \prod_{j=0}^{\infty}\left(1 - \frac{1}{2 + S_j}\right)$$

as $n \to \infty$, where $S_j = \sum_{i=1}^{j} A_i$. The infinite product equals 0 a.s. if and only if $\sum_j (2 + S_j)^{-1} = \infty$ a.s. By monotone convergence, $\mathbb{P}(T < \infty) = 1$ under this condition. If this holds, we may apply the optional stopping theorem to obtain that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$, which is to say that

$$\mathbb{E}\left(1 - \frac{1 + A_T}{2 + S_T}\right) = \frac{1}{2}.$$

15. At each stage k, let L_k be the length of the sequence 'in play', and let Y_k be the sum of its entries, so that $L_0 = n$, $Y_0 = \sum_{i=1}^{n} x_i$. If you lose the $(k+1)$th gamble, then $L_{k+1} = L_k + 1$ and $Y_{k+1} = Y_k + Z_k$ where Z_k is the stake on that play, whereas if you win, then $L_{k+1} = L_k - 2$ and $Y_{k+1} = Y_k - Z_k$; we have assumed that $L_k \ge 2$, similar relations being valid if $L_k = 1$. Note that L_k is a random walk with mean step size -1, implying that the first-passage time T to 0 is a.s. finite, and has all moments finite. Your profits at time k amount to $Y_0 - Y_k$, whence your profit at time T is Y_0, since $Y_T = 0$.

Since the games are fair, Y_k constitutes a martingale. Therefore $\mathbb{E}(Y_{T \wedge m}) = \mathbb{E}(Y_0) \neq 0$ for all m. However $T \wedge m \to T$ a.s. as $m \to \infty$, so that $Y_{T \wedge m} \to Y_T$ a.s. Now $\mathbb{E}(Y_T) = 0 \neq \lim_{m\to\infty} \mathbb{E}(Y_{T \wedge m})$, and it follows that $\{Y_{T \wedge m} : m \ge 1\}$ is not uniformly integrable. Therefore $\mathbb{E}(\sup_m Y_{T \wedge m}) = \infty$; see Exercise (7.10.6).

16. Since the game is fair, $\mathbb{E}(S_{n+1} \mid S_n) = S_n$. Also $|S_n| \le 1 + 2 + \cdots + n < \infty$. Therefore S_n is a martingale. The occurrence of the event $\{N = n\}$ depends only on the outcomes of the coin-tosses up to and including the nth; therefore N is a stopping time.

A tail appeared at time $N - 3$, followed by three heads. Therefore the gamblers $G_1, G_2, \dots, G_{N-3}$ have forfeited their initial capital by time N, while G_{N-i} has had $i + 1$ successful rounds for $0 \le i \le 2$. Therefore $S_N = N - (p^{-1} + p^{-2} + p^{-3})$, after a little calculation. It is easy to check that N satisfies the conditions of the optional stopping theorem, and it follows that $\mathbb{E}(S_N) = \mathbb{E}(S_0) = 0$, which is to say that $\mathbb{E}(N) = p^{-1} + p^{-2} + p^{-3}$.

In order to deal with HTH, the gamblers are re-programmed to act as follows. If they win on their first bet, they bet their current fortune on *tails*, returning to heads thereafter. In this case, $S_N = N - (p^{-1} + p^{-2}q^{-1})$ where $q = 1 - p$ (remember that the game is fair), and therefore $\mathbb{E}(N) = p^{-1} + p^{-2}q^{-1}$.

17. Let $\mathcal{F}_n = \sigma(\{X_i, Y_i : 1 \le i \le n\})$, and note that T is a stopping time with respect to this filtration. Furthermore $\mathbb{P}(T < \infty) = 1$ since T is no larger than the first-passage time to 0 of either of the two single-coordinate random walks, each of which has mean 0 and is therefore persistent.

Let $\sigma_1^2 = \operatorname{var}(X_1)$ and $\sigma_2^2 = \operatorname{var}(Y_1)$. We have that $U_n - U_0$ and $V_n - V_0$ are sums of independent summands with means 0 and variances σ_1^2 and σ_2^2 respectively. It follows by considering the martingales $(U_n - U_0)^2 - n\sigma_1^2$ and $(V_n - V_0)^2 - n\sigma_2^2$ (see equation (12.5.14) and Exercise (10.2.2)) that

$$\mathbb{E}\{(U_T - U_0)^2\} = \sigma_1^2\mathbb{E}(T), \quad \mathbb{E}\{(V_T - V_0)^2\} = \sigma_2^2\mathbb{E}(T).$$

Applying the same argument to $(U_n + V_n) - (U_0 + V_0)$, we obtain

$$\mathbb{E}\big\{(U_T + V_T - U_0 - V_0)^2\big\} = \mathbb{E}(T)\mathbb{E}\{(X_1 + Y_1)^2\} = \mathbb{E}(T)(\sigma_1^2 + 2c + \sigma_2^2).$$

Subtract the two earlier equations to obtain

$$(*) \qquad \mathbb{E}\big\{(U_T - U_0)(V_T - V_0)\big\} = c\mathbb{E}(T)$$

if $\mathbb{E}(T) < \infty$. Now $U_T V_T = 0$, and in addition $\mathbb{E}(U_T) = U_0$, $\mathbb{E}(V_T) = V_0$, by Wald's equation and the fact that $\mathbb{E}(X_1) = \mathbb{E}(Y_1) = 0$. It follows that $-\mathbb{E}(U_0 V_0) = c\mathbb{E}(T)$ if $\mathbb{E}(T) < \infty$, in which case $c < 0$.

Suppose conversely that $c < 0$. Then $(*)$ is valid with T replaced throughout by the bounded stopping time $T \wedge m$, and hence

$$0 \le \mathbb{E}(U_{T\wedge m} V_{T\wedge m}) = \mathbb{E}(U_0 V_0) + c\mathbb{E}(T \wedge m).$$

Therefore $\mathbb{E}(T \wedge m) \le \mathbb{E}(U_0 V_0)/(2|c|)$ for all m, implying that $\mathbb{E}(T) = \lim_{m\to\infty} \mathbb{E}(T \wedge m) < \infty$, and so $\mathbb{E}(T) = -\mathbb{E}(U_0 V_0)/c$ as before.

18. Certainly $0 \le X_n \le 1$, and in addition X_n is measurable with respect to the σ-field $\mathcal{F}_n = \sigma(R_1, R_2, \dots, R_n)$. Also $\mathbb{E}(R_{n+1} \mid R_n) = R_n - R_n/(52 - n)$, whence $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$. Therefore X_n is a martingale.

A strategy corresponds to a stopping time. If the player decides to call at the stopping time T, he wins with (conditional) probability X_T, and therefore $\mathbb{P}(\text{wins}) = \mathbb{E}(X_T)$, which equals $\mathbb{E}(X_0)$ $(= \frac{1}{2})$ by the optional stopping theorem.

Here is a trivial solution to the problem. It may be seen that the chance of winning is the same for a player who, after calling "Red Now", picks the card placed at the bottom of the pack rather than that at the top. The bottom card is red with probability $\frac{1}{2}$, irrespective of the strategy of the player.

19. (a) A sum s of money in week t is equivalent to a sum $s/(1+\alpha)^t$ in week 0, since the latter sum may be invested now to yield s in week t. If he sells in week t, his discounted costs are $\sum_{n=1}^{t} c/(1+\alpha)^n$ and his discounted profit is $X_t/(1+\alpha)^t$. He wishes to find a stopping time for which his mean discounted gain is a maximum.

Now

$$-\sum_{n=1}^{T}(1+\alpha)^{-n}c = \frac{c}{\alpha}\{(1+\alpha)^{-T} - 1\},$$

so that $\mu(T) = \mathbb{E}\{(1+\alpha)^{-T} Z_T\} - (c/\alpha)$.

(b) The function $h(y) = \alpha y - \int_y^\infty \mathbb{P}(Z_n > y)\,dy$ is continuous and strictly increasing on $[0, \infty)$, with $h(0) = -\mathbb{E}(Z_n) < 0$ and $h(y) \to \infty$ as $y \to \infty$. Therefore there exists a unique γ (> 0) such that $h(\gamma) = 0$, and we choose γ accordingly.

(c) Let $\mathcal{F}_n = \sigma(Z_1, Z_2, \dots, Z_n)$. We have that

$$\mathbb{E}\big(\max\{Z_n, \gamma\}\big) = \gamma + \int_\gamma^\infty [1 - G(y)]\,dy = (1+\alpha)\gamma$$

where $G(y) = \mathbb{P}(Z_n \le y)$. Therefore $\mathbb{E}(V_{n+1} \mid \mathcal{F}_n) = (1+\alpha)^{-n}\gamma \le V_n$, so that $(V_n, \mathcal{F}_n)$ is a non-negative supermartingale.

Let $\mu(\tau)$ be the mean gain of following the strategy 'accept the first offer exceeding $\tau - (c/\alpha)$'. The corresponding stopping time T satisfies $\mathbb{P}(T = n) = G(\tau)^n(1 - G(\tau))$, and therefore

$$\begin{aligned}\mu(\tau) + (c/\alpha) &= \sum_{n=0}^{\infty} \mathbb{E}\{(1+\alpha)^{-T} Z_T I_{\{T=n\}}\}\\ &= \sum_{n=0}^{\infty}(1+\alpha)^{-n} G(\tau)^n(1 - G(\tau))\mathbb{E}(Z_1 \mid Z_1 > \tau)\\ &= \frac{1+\alpha}{1+\alpha - G(\tau)}\left\{\tau(1 - G(\tau)) + \int_{\tau}^{\infty}(1 - G(y))\,dy\right\}.\end{aligned}$$

Differentiate with care to find that the only value of τ lying in the support of Z_1 such that $\mu'(\tau) = 0$ is the value $\tau = \gamma$. Furthermore this value gives a maximum for $\mu(\tau)$. Therefore, amongst strategies of the above sort, the best is that with $\tau = \gamma$. Note that $\mu(\gamma) = \gamma(1+\alpha) - (c/\alpha)$.

Consider now a general strategy with corresponding stopping time T, where $\mathbb{P}(T < \infty) = 1$. For any positive integer m, $T \wedge m$ is a bounded stopping time, whence $\mathbb{E}(V_{T\wedge m}) \le \mathbb{E}(V_0) = \gamma(1+\alpha)$. Now $|V_{T\wedge m}| \le \sum_{i=0}^{\infty} |V_i|$, and $\sum_{i=0}^{\infty} \mathbb{E}|V_i| < \infty$. Therefore $\{V_{T\wedge m} : m \ge 0\}$ is uniformly integrable. Also $V_{T\wedge m} \to V_T$ a.s. as $m \to \infty$, and it follows that $\mathbb{E}(V_{T\wedge m}) \to \mathbb{E}(V_T)$. We conclude that $\mu(T) = \mathbb{E}(V_T) - (c/\alpha) \le \gamma(1+\alpha) - (c/\alpha) = \mu(\gamma)$. Therefore the strategy given above is optimal.

(d) In the special case, $\mathbb{P}(Z_1 > y) = (y-1)^{-2}$ for $y \ge 2$, whence $\gamma = 10$. The target price is therefore 9, and the mean number of weeks before selling is $G(\gamma)/(1 - G(\gamma)) = 80$.

20. Since G is convex on $[0, \infty)$ wherever it is finite, and since $G(1) = 1$ and $G'(1) < 1$, there exists a unique value of η (> 1) such that $G(\eta) = \eta$. Furthermore, $Y_n = \eta^{Z_n}$ defines a martingale with mean $\mathbb{E}(Y_0) = \eta$. Using the maximal inequality (12.6.6),

$$\mathbb{P}\big(\sup_n Z_n \ge k\big) = \mathbb{P}\big(\sup_n Y_n \ge \eta^k\big) \le \frac{1}{\eta^{k-1}}$$

for positive integers k. Therefore

$$\mathbb{E}\big(\sup_n Z_n\big) \le \sum_{k=1}^{\infty} \frac{1}{\eta - 1}.$$

21. Let M_n be the number present after the nth round, so $M_0 = K$, and $M_{n+1} = M_n - X_{n+1}$, $n \ge 1$, where X_n is the number of matches in the nth round. By the result of Problem (3.11.17), $\mathbb{E}X_n = 1$ for all n, whence

$$\mathbb{E}(M_{n+1} + n + 1 \mid \mathcal{F}_n) = M_n + n,$$

where $\mathcal{F}_n$ is the σ-field generated by $M_0, M_1, \dots, M_n$. Thus the sequence $\{M_n + n\}$ is a martingale. Now, N is clearly a stopping time, and therefore $K = M_0 + 0 = \mathbb{E}(M_N + N) = \mathbb{E}N$.

We have that

$$\begin{aligned}&\mathbb{E}\{(M_{n+1} + n + 1)^2 + M_{n+1} \mid \mathcal{F}_n\}\\ &\quad= (M_n + n)^2 - 2(M_n + n)\mathbb{E}(X_{n+1} - 1) + M_n + \mathbb{E}\{(X_{n+1} - 1)^2 - X_{n+1} \mid \mathcal{F}_n\}\\ &\quad\le (M_n + n)^2 + M_n,\end{aligned}$$

where we have used the fact that

$$\operatorname{var}(X_{n+1} \mid \mathcal{F}_n) = \begin{cases} 1 & \text{if } M_n > 1,\\ 0 & \text{if } M_n = 1.\end{cases}$$

Hence the sequence $\{(M_n + n)^2 + M_n\}$ is a supermartingale. By an optional stopping theorem for supermartingales,

$$K^2 + K = M_0^2 + M_0 \ge \mathbb{E}\{(M_N + N)^2 + M_N\} = \mathbb{E}(N^2),$$

and therefore $\mathrm{var}(N) \le K$.

22. In the usual notation,

$$\begin{aligned}\mathbb{E}\big(M(s+t) \mid \mathcal{F}_s\big) &= \mathbb{E}\left(\int_0^s W(u)\,du + \int_s^{s+t} W(u)\,du - \tfrac{1}{3}\{W(s+t) - W(s) + W(s)\}^3 \,\Big|\, \mathcal{F}_s\right)\\ &= M(s) + tW(s) - W(s)\mathbb{E}\big([W(s+t) - W(s)]^2 \mid \mathcal{F}_s\big) = M(s)\end{aligned}$$

as required. We apply the optional stopping theorem (12.7.12) with the stopping time $T = \inf\{u : W(u) \in \{a, b\}\}$. The hypotheses of the theorem follow easily from the boundedness of the process for $t \in [0, T]$, and it follows that

$$\mathbb{E}\left(\int_0^T W(u)\,du - \tfrac{1}{3}W(T)^3\right) = 0.$$

Hence the required area A has mean

$$\mathbb{E}(A) = \mathbb{E}\left(\int_0^T W(u)\,du\right) = \frac{1}{3}\mathbb{E}(W(T)^3) = \frac{1}{3}a^3\left(\frac{-b}{a-b}\right) + \frac{1}{3}b^3\left(\frac{a}{a-b}\right).$$

[We have used the optional stopping theorem twice actually, in that $\mathbb{E}(W(T)) = 0$ and therefore $\mathbb{P}(W(T) = a) = -b/(a-b)$.]

23. With $\mathcal{F}_s = \sigma(W(u) : 0 \le u \le s)$, we have for $s < t$ that

$$\mathbb{E}(R(t)^2 \mid \mathcal{F}_s) = \mathbb{E}\big(|W(s)|^2 + |W(t) - W(s)|^2 + 2W(s)\cdot(W(t) - W(s)) \,\big|\, \mathcal{F}_s\big) = R(s)^2 + (t-s),$$

and the first claim follows. We apply the optional stopping theorem (12.7.12) with $T = \inf\{u : |W(u)| = a\}$, as in Problem (12.9.22), to find that $0 = \mathbb{E}(R(T)^2 - T) = a^2 - \mathbb{E}(T)$.

24. We apply the optional stopping theorem to the martingale $W(t)$ with the stopping time T to find that $\mathbb{E}(W(T)) = -a(1 - p_b) + bp_b = 0$, where $p_b = \mathbb{P}(W(T) = b)$. By Example (12.7.10), $W(t)^2 - t$ is a martingale, and therefore, by the optional stopping theorem again,

$$\mathbb{E}\big((W(T)^2 - T\big) = a^2(1 - p_b) + b^2 p_b - \mathbb{E}(T) = 0,$$

whence $\mathbb{E}(T) = ab$. For the final part, we take $a = b$ and apply the optional stopping theorem to the martingale $\exp[\theta W(t) - \frac{1}{2}\theta^2 t]$ to obtain

$$\mathbb{E}\big(\exp[\theta W(T) - \tfrac{1}{2}\theta^2 T]\big) = \{e^{-b\theta}(1 - p_b) + e^{b\theta}p_b\}\mathbb{E}(e^{-\frac{1}{2}\theta^2 T}) = 1,$$

on noting that the conditional distribution of T given $W(T) = b$ is the same as that given $W(T) = -b$. Therefore, $\mathbb{E}(e^{-\frac{1}{2}\theta^2 T}) = 1/\cosh(b\theta)$, and the answer follows by substituting $s = \frac{1}{2}\theta^2$.

13

Diffusion processes

13.3 Solutions. Diffusion processes

1. It is easily seen that

$$\mathbb{E}\{X(t+h) - X(t) \mid X(t)\} = (\lambda - \mu)X(t)h + o(h),$$
$$\mathbb{E}(\{X(t+h) - X(t)\}^2 \mid X(t)) = (\lambda + \mu)X(t)h + o(h),$$

which suggest a diffusion approximation with instantaneous mean $a(t, x) = (\lambda - \mu)x$ and instantaneous variance $b(t, x) = (\lambda + \mu)x$.

2. The following method is not entirely rigorous (it is an argument of the following well-known type: it is valid when it works, and not otherwise). We have that

$$\frac{\partial M}{\partial t} = \int_{-\infty}^{\infty} e^{\theta y} \frac{\partial f}{\partial t} \, dy = \int_{-\infty}^{\infty} \{\theta a(t, y) + \tfrac{1}{2}\theta^2 b(t, y)\} e^{\theta y} f \, dy,$$

by using the forward equation and integrating by parts. Assume that $a(t, y) = \sum_n \alpha_n(t) y^n$, $b(t, y) = \sum_n \beta_n(t) y^n$. The required expression follows from the 'fact' that

$$\int_{-\infty}^{\infty} e^{\theta y} y^n f \, dy = \frac{\partial^n}{\partial \theta^n} \int_{-\infty}^{\infty} e^{\theta y} f \, dy = \frac{\partial^n M}{\partial \theta^n}.$$

3. Using Exercise (13.3.2) or otherwise, we obtain the equation

$$\frac{\partial M}{\partial t} = \theta m M + \tfrac{1}{2}\theta^2 M$$

with boundary condition $M(0, \theta) = 1$. The solution is $M(t) = \exp\{\frac{1}{2}\theta(2m + \theta)t\}$.

4. Using Exercise (13.3.2) or otherwise, we obtain the equation

$$\frac{\partial M}{\partial t} = -\theta \frac{\partial M}{\partial \theta} + \tfrac{1}{2}\theta^2 M$$

with boundary condition $M(0, \theta) = 1$. The characteristics of the equation are given by

$$\frac{dt}{1} = \frac{d\theta}{\theta} = \frac{2\, dM}{\theta^2 M},$$

with solution $M(t, \theta) = e^{\frac{1}{4}\theta^2} g(\theta e^{-t})$ where g is a function satisfying $1 = e^{\frac{1}{4}\theta^2} g(\theta)$. Therefore $M = \exp\{\frac{1}{4}\theta^2(1 - e^{-2t})\}$.

5. Fix $t > 0$. Suppose we are given $W_1(s)$, $W_2(s)$, $W_3(s)$, for $0 \le s \le t$. By Pythagoras's theorem, $R(t+u)^2 = X_1^2 + X_2^2 + X_3^2$ where the X_i are independent $N(W_i(t), u)$ variables. Using the result of Exercise (5.7.7), the conditional distribution of $R(t+u)^2$ (and hence of $R(t+u)$ also) depends only on the value of the non-centrality parameter $\theta = R(t)^2$ of the relevant non-central χ^2 distribution. It follows that R satisfies the Markov property. This argument is valid for the n-dimensional Bessel process.

6. By the spherical symmetry of the process, the conditional distribution of $R(s+a)$ given $R(s) = x$ is the same as that given $W(s) = (x, 0, 0)$. Therefore, recalling the solution to Exercise (13.3.5),

$$\begin{aligned}
&\mathbb{P}\big(R(s+a) \le y \mid R(s) = x\big) \\
&\quad = \int_{\substack{(u,v,w):\\ u^2+v^2+w^2 \le y^2}} \frac{1}{(2\pi a)^{3/2}} \exp\left\{-\frac{(u-x)^2+v^2+w^2}{2a}\right\} du\,dv\,dw \\
&\quad = \int_{\rho=0}^{y}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \frac{1}{(2\pi a)^{3/2}} \exp\left\{-\frac{\rho^2 - 2\rho x\cos\theta + x^2}{2a}\right\} \rho^2 \sin\theta\, d\theta\, d\phi\, d\rho \\
&\quad = \int_0^y \frac{\rho/x}{\sqrt{2\pi a}} \left\{\exp\left(-\frac{(\rho-x)^2}{2a}\right) - \exp\left(-\frac{(\rho+x)^2}{2a}\right)\right\} d\rho,
\end{aligned}$$

and the result follows by differentiating with respect to y.

7. Continuous functions of continuous functions are continuous. The Markov property is preserved because $g(\cdot)$ is single-valued with a unique inverse.

8. (a) Since $\mathbb{E}(e^{\sigma W(t)}) = e^{\frac{1}{2}\sigma^2 t}$, this is not a martingale.

(b) This is a Wiener process (see Problem (13.12.1)), and is certainly a martingale.

(c) With $\mathcal{F}_t = \sigma(W(s) : 0 \le s \le t)$ and $t, u > 0$,

$$\begin{aligned}
\mathbb{E}\left\{(t+u)W(t+u) - \int_0^{t+u} W(s)\,ds \;\middle|\; \mathcal{F}_t\right\} &= (t+u)W(t) - \int_0^t W(s)\,ds - \int_t^{t+u} W(t)\,ds \\
&= tW(t) - \int_0^t W(s)\,ds,
\end{aligned}$$

whence this is a martingale. [The integrability condition is easily verified.]

9. (a) With $s < t$, $S(t) = S(s)\exp\{a(t-s) + b(W(t) - W(s))\}$. Now $W(t) - W(s)$ is independent of $\{W(u) : 0 \le u \le s\}$, and the claim follows.

(b) $S(t)$ is clearly integrable and adapted to the filtration $\boldsymbol{\mathcal{F}} = (\mathcal{F}_t)$ so that, for $s < t$,

$$\mathbb{E}\big(S(t) \mid \mathcal{F}_s\big) = S(s)\mathbb{E}\big(\exp\{a(t-s) + b(W(t) - W(s))\} \mid \mathcal{F}_s\big) = S(s)\exp\{a(t-s) + \tfrac{1}{2}b^2(t-s)\},$$

which equals $S(s)$ if and only if $a + \frac{1}{2}b^2 = 0$. In this case, $\mathbb{E}(S(t)) = \mathbb{E}(S(0)) = 1$.

10. Either find the instantaneous mean and variance, and solve the forward equation, or argue directly as follows. With $s < t$,

$$\mathbb{P}\big(S(t) \le y \mid S(s) = x\big) = \mathbb{P}\big(bW(t) \le -at + \log y \mid bW(s) = -as + \log x\big).$$

Now $b(W(t) - W(s))$ is independent of $W(s)$ and is distributed as $N(0, b^2(t-s))$, and we obtain on differentiating with respect to y that

$$f(t, y \mid s, x) = \frac{1}{y\sqrt{2\pi b^2(t-s)}} \exp\left(-\frac{(\log(y/x) - a(t-s))^2}{2b^2(t-s)}\right), \qquad x, y > 0.$$

13.4 Solutions. First passage times

1. Certainly X has continuous sample paths, and in addition $\mathbb{E}|X(t)| < \infty$. Also, if $s < t$,

$$\mathbb{E}\big(X(t) \mid \mathcal{F}_s\big) = X(s)e^{\frac{1}{2}\theta^2(t-s)}\mathbb{E}\big(e^{i\theta\{W(t)-W(s)\}} \mid \mathcal{F}_s\big) = X(s)e^{\frac{1}{2}\theta^2(t-s)}e^{-\frac{1}{2}\theta^2(t-s)} = X(s)$$

as required, where we have used the fact that $W(t) - W(s)$ is $N(0, t-s)$ and is independent of $\mathcal{F}_s$.

2. Apply the optional stopping theorem to the martingale X of Exercise (13.4.1), with the stopping time T, to obtain $\mathbb{E}(X(T)) = 1$. Now $W(T) = aT + b$, and therefore $\mathbb{E}(e^{\psi T + i\theta b}) = 1$ where $\psi = ia\theta + \frac{1}{2}\theta^2$. Solve to find that

$$\mathbb{E}(e^{\psi T}) = e^{-i\theta b} = \exp\left\{-b\big(\sqrt{a^2 - 2\psi} + a\big)\right\}$$

is the solution which gives a moment generating function.

3. We have that $T \le u$ if and only if there is no zero in $(u, t]$, an event with probability $1 - (2/\pi)\cos^{-1}\{\sqrt{u/t}\}$, and the claim follows on drawing a triangle.

13.5 Solution. Barriers

1. Solving the forward equation subject to the appropriate boundary conditions, we obtain as usual that

$$f^{\mathrm{r}}(t, y) = g(t, y \mid d) + e^{-2md}g(t, y \mid -d) - \int_{-\infty}^{-d} 2me^{2mx}g(t, y \mid x)\,dx$$

where $g(t, y \mid x) = (2\pi t)^{-\frac{1}{2}}\exp\{-(y - x - mt)^2/(2t)\}$. The first two terms tend to 0 as $t \to \infty$, regardless of the sign of m. As for the integral, make the substitution $u = (x - y - mt)/\sqrt{t}$ to obtain, as $t \to \infty$,

$$-\int_{-\infty}^{-(d+y+mt)/\sqrt{t}} 2me^{2my}\frac{e^{-\frac{1}{2}u^2}}{\sqrt{2\pi}}\,du \to \begin{cases} 2|m|e^{-2|m|y} & \text{if } m < 0, \\ 0 & \text{if } m \ge 0. \end{cases}$$

13.6 Solutions. Excursions and the Brownian bridge

1. Let $f(t, x) = (2\pi t)^{-\frac{1}{2}}e^{-x^2/(2t)}$. It may be seen that

$$\mathbb{P}\big(W(t) > x \,\big|\, Z, W(0) = 0\big) = \lim_{w \downarrow 0}\mathbb{P}\big(W(t) > x \,\big|\, Z, W(0) = w\big)$$

where $Z = \{\text{no zeros in } (0, t]\}$; the small missing step here may be filled by conditioning instead on the event $\{W(\epsilon) = w, \text{ no zeros in } (\epsilon, t]\}$, and taking the limit as $\epsilon \downarrow 0$. Now, if $w > 0$,

$$\mathbb{P}\big(W(t) > x, Z \,\big|\, W(0) = w\big) = \int_x^{\infty}\{f(t, y - w) - f(t, y + w)\}\,dy$$

by the reflection principle, and

$$\mathbb{P}\big(Z \,\big|\, W(0) = w\big) = 1 - 2\int_w^{\infty} f(t, y)\,dy = \int_{-w}^{w} f(t, y)\,dy$$

by a consideration of the minimum value of W on $(0, t]$. It follows that the density function of $W(t)$, conditional on $Z \cap \{W(0) = w\}$, where $w > 0$, is

$$h_w(x) = \frac{f(t, x-w) - f(t, x+w)}{\int_{-w}^{w} f(t, y)\,dy}, \qquad x > 0.$$

Divide top and bottom by $2w$, and take the limit as $w \downarrow 0$:

$$\lim_{w \downarrow 0} h_w(x) = -\frac{1}{f(t, 0)} \frac{\partial f}{\partial x} = \frac{x}{t} e^{-x^2/(2t)}, \qquad x > 0.$$

2. It is a standard exercise that, for a Wiener process W,

$$\mathbb{E}\{W(t) \mid W(s) = a, W(1) = 0\} = a\left(\frac{1-t}{1-s}\right),$$
$$\mathbb{E}\{W(s)^2 \mid W(0) = W(1) = 0\} = s(1-s),$$

if $0 \le s \le t \le 1$. Therefore the Brownian bridge B satisfies, for $0 \le s \le t \le 1$,

$$\mathbb{E}\big(B(s)B(t)\big) = \mathbb{E}\{B(s)\mathbb{E}\big(B(t) \mid B(s)\big)\} = \frac{1-t}{1-s}\mathbb{E}(B(s)^2) = s(1-t)$$

as required. Certainly $\mathbb{E}(B(s)) = 0$ for all s, by symmetry.

3. $\widehat{W}$ is a zero-mean Gaussian process on $[0, 1]$ with continuous sample paths, and also $\widehat{W}(0) = \widehat{W}(1) = 0$. Therefore $\widehat{W}$ is a Brownian bridge if it has the same autocovariance function as the Brownian bridge, that is, $c(s, t) = \min\{s, t\} - st$. For $s < t$,

$$\operatorname{cov}\big(\widehat{W}(s), \widehat{W}(t)\big) = \operatorname{cov}\big(W(s) - sW(1), W(t) - tW(1)\big) = s - ts - st + st = s - st$$

since $\operatorname{cov}(W(u), W(v)) = \min\{u, v\}$. The claim follows.

4. Either calculate the instantaneous mean and variance of $\widetilde{W}$, or repeat the argument in the solution to Exercise (13.6.3). The only complication in this case is the necessity to show that $\widetilde{W}(t)$ is a.s. continuous at $t = 1$, i.e., that $u^{-1}W(u-1) \to 0$ a.s. as $u \to \infty$. There are various ways to show this. Certainly it is true in the limit as $u \to \infty$ through the integers, since, for integral u, $W(u-1)$ may be expressed as the sum of $u-1$ independent $N(0, 1)$ variables (use the strong law). It remains to fill in the gaps. Let n be a positive integer, let $x > 0$, and write $M_n = \max\{|W(u) - W(n)| : n \le u \le n+1\}$. We have by the stationarity of the increments that

$$\sum_{n=0}^{\infty} \mathbb{P}(M_n \ge nx) = \sum_{n=0}^{\infty} \mathbb{P}(M_1 \ge nx) \le 1 + \frac{\mathbb{E}(M_1)}{x} < \infty,$$

implying by the Borel–Cantelli lemma that $n^{-1}M_n \le x$ for all but finitely many values of n, a.s. Therefore $n^{-1}M_n \to 0$ a.s. as $n \to \infty$, implying that

$$\lim_{u \to \infty} \frac{1}{u+1}|W(u)| \le \lim_{n \to \infty} \frac{1}{n}\{|W(n)| + M_n\} \to 0 \qquad \text{a.s.}$$

5. In the notation of Exercise (13.6.4), we are asked to calculate the probability that W has no zeros in the time interval between $s/(1-s)$ and $t/(1-t)$. By Theorem (13.4.8), this equals

$$1 - \frac{2}{\pi}\cos^{-1}\sqrt{\frac{s(1-y)}{t(1-s)}} = \frac{2}{\pi}\cos^{-1}\sqrt{\frac{t-s}{t(1-s)}}.$$

13.7 Solutions. Stochastic calculus

1. Let $\mathcal{F}_s = \sigma(W_u : 0 \le u \le s)$. Fix $n \ge 1$ and define $X_n(k) = |W_{kt/2^n}|$ for $0 \le k \le 2^n$. By Jensen's inequality, the sequence $\{X_n(k) : 0 \le k \le 2^n\}$ is a non-negative submartingale with respect to the filtration $\mathcal{F}_{kt/2^n}$, with finite variance. Hence, by Exercise (4.3.3) and equation (12.6.2), $X_n^* = \max\{X_n(k) : 0 \le k \le 2^n\}$ satisfies

$$\begin{aligned}\mathbb{E}(X_n^{*2}) &= 2\int_0^\infty x\mathbb{P}(X_n^* > x)\,dx \le 2\int_0^\infty \mathbb{E}(W_t^+ I_{\{X_n^* \ge x\}})\,dx = 2\mathbb{E}\left\{W_t^+ \int_0^{X_n^*} dx\right\}\\ &= 2\mathbb{E}(W_t^+ X_n^*) \le 2\sqrt{\mathbb{E}(W_t^2)\mathbb{E}(X_n^{*2})} \qquad \text{by the Cauchy–Schwarz inequality.}\end{aligned}$$

Hence $\mathbb{E}(X_n^{*2}) \le 4\mathbb{E}(W_t^2)$. Now X_n^{*2} is monotone increasing in n, and W has continuous sample paths. By monotone convergence,

$$\mathbb{E}\left(\max_{s\le t} |W_s|^2\right) = \lim_{n\to\infty} \mathbb{E}(X_n^{*2}) \le 4\mathbb{E}(W_t^2).$$

2. See the solution to Exercise (8.5.4).

3. (a) We have that

$$I_1(n) = \tfrac{1}{2}\left\{\sum_{j=0}^{n-1}(V_{j+1}^2 - V_j^2) - \sum_{j=0}^{n-1}(V_{j+1} - V_j)^2\right\}.$$

The first summation equals W_t^2, by successive concellation, and the mean-square limit of the second summation is t, by Exercise (8.5.4). Hence $\lim_{n\to\infty} I_1(n) = \frac{1}{2}W_t^2 - \frac{1}{2}t$ in mean square.

Likewise, we obtain the mean-square limits:

$$\lim_{n\to\infty} I_2(n) = \tfrac{1}{2}W_t^2 + \tfrac{1}{2}t, \qquad \lim_{n\to\infty} I_3(n) = \lim_{n\to\infty} I_4(n) = \tfrac{1}{2}W_t^2.$$

4. Clearly $\mathbb{E}(U(t)) = 0$. The process U is Gaussian with autocovariance function

$$\mathbb{E}\big(U(s)U(s+t)\big) = \mathbb{E}\Big(\mathbb{E}\big(U(s)U(s+t) \mid \mathcal{F}_s\big)\Big) = e^{-\beta s}e^{-\beta(s+t)}\mathbb{E}\big[W(e^{2\beta s})^2\big] = e^{-\beta t}.$$

Thus U is a stationary Gaussian Markov process, namely the Ornstein–Uhlenbeck process. [See Example (9.6.10).]

5. Clearly $\mathbb{E}(U_t) = 0$. For $s < t$,

$$\begin{aligned}\mathbb{E}(U_s U_{s+t}) &= \mathbb{E}(W_s W_t) + \beta^2\mathbb{E}\left(\int_{u=0}^{s}\int_{v=0}^{t} e^{-\beta(s-u)}W_u e^{-\beta(t-v)}W_v\,du\,dv\right)\\ &\quad - \mathbb{E}\left(W_t\beta\int_0^s e^{-\beta(s-u)}W_u\,du\right) - \mathbb{E}\left(W_s\int_0^t e^{-\beta(t-v)}W_v\,dv\right)\\ &= s + \beta^2 e^{-\beta(s+t)}\int_{u=0}^{s}\int_{v=0}^{t} e^{\beta(u+v)}\min\{u,v\}\,du\,dv\\ &\quad - \beta\int_0^s e^{-\beta(s-u)}\min\{u,t\}\,du - \int_0^t e^{-\beta(t-v)}\min\{s,v\}\,dv\\ &= \frac{e^{2\beta s}-1}{2\beta}e^{-\beta(s+t)}\end{aligned}$$

after prolonged integration. By the linearity of the definition of U, it is a Gaussian process. From the calculation above, it has autocovariance function $c(s, s+t) = (e^{-\beta(t-s)} - e^{-\beta(t+s)})/(2\beta)$. From this we may calculate the instantaneous mean and variance, and thus we recognize an Ornstein–Uhlenbeck process. See also Exercise (13.3.4) and Problem (13.12.4).

13.8 Solutions. The Itô integral

1. (a) Fix $t > 0$ and let $n \geq 1$ and $\delta = t/n$. We write $t_j = jt/n$ and $V_j = W_{t_j}$. By the absence of correlation of Wiener increments, and the Cauchy–Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\left(\left|\int_0^t W_s\,ds - \sum_{j=0}^{n-1} V_{j+1}(t_{j+1}-t_j)\right|^2\right) &= \mathbb{E}\left(\left|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(V_{j+1}-W_s)\,ds\right|^2\right)\\
&= \sum_{j=0}^{n-1}\mathbb{E}\left(\left|\int_{t_j}^{t_{j+1}}(V_{j+1}-W_s)\,ds\right|^2\right)\\
&\leq \sum_{j=0}^{n-1}\left\{(t_{j+1}-t_j)\int_{t_j}^{t_{j+1}}\mathbb{E}(|V_{j+1}-W_s|^2)\,ds\right\}\\
&= \sum_{j=0}^{n-1}\frac{1}{2}(t_{j+1}-t_j)^3 = \sum_{j=0}^{n-1}\frac{1}{2}\left(\frac{t}{n}\right)^3 \to 0 \quad \text{as } n\to\infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_0^t s\,dW_s &= \lim_{n\to\infty}\sum_{j=0}^{n-1} t_j(V_{j+1}-V_j) = \lim_{n\to\infty}\sum_{j=0}^{n-1}\big(t_{j+1}V_{j+1} - t_jV_j - (t_{j+1}-t_j)V_{j+1}\big)\\
&= \lim_{n\to\infty}\left(tW_t - \sum_{j=0}^{n-1}V_{j+1}(t_{j+1}-t_j)\right) = tW_t - \int_0^t W_s\,ds.
\end{aligned}$$

(b) As $n \to \infty$,

$$\begin{aligned}
\sum_{j=0}^{n-1} V_j^2(V_{j+1}-V_j) &= \tfrac{1}{3}\sum_{j=0}^{n-1}\{V_{j+1}^3 - V_j^3 - 3V_j(V_{j+1}-V_j)^2 - (V_{j+1}-V_j)^3\}\\
&= \tfrac{1}{3}W_t^3 - \sum_{j=0}^{n-1}\Big[V_j(t_{j+1}-t_j) + V_j\{(V_{j+1}-V_j)^2 - (t_{j+1}-t_j)\}\Big] - \tfrac{1}{3}\sum_{j=0}^{n-1}(V_{j+1}-V_j)^3\\
&\to \tfrac{1}{3}W_t^3 - \int_0^t W(s)\,ds + 0 + 0.
\end{aligned}$$

The fact that the last two terms tend to 0 in mean square may be verified in the usual way. For example,

$$\begin{aligned}
\mathbb{E}\left(\left|\sum_{j=0}^{n-1}(V_{j+1}-V_j)^3\right|^2\right) &= \sum_{j=0}^{n-1}\mathbb{E}[(V_{j+1}-V_j)^6]\\
&= 6\sum_{j=0}^{n-1}(t_{j+1}-t_j)^3 = 6\sum_{j=0}^{n-1}\left(\frac{t}{n}\right)^3 \to 0 \qquad \text{as } n\to\infty.
\end{aligned}$$

(c) It was shown in Exercise (13.7.3a) that $\int_0^t W_s \, dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t$. Hence,

$$\mathbb{E}\left(\left[\int_0^t W_s \, dW_s\right]^2\right) = \tfrac{1}{4}\{\mathbb{E}(W_t^4) - 2t\mathbb{E}(W_t^2) + t^2\} = \tfrac{1}{2}t^2,$$

and the result follows because $\mathbb{E}(W_t^2) = t$.

2. Fix $t > 0$ and $n \geq 1$, and let $\delta = t/n$. We set $V_j = W_{jt/n}$. It is the case that $X_t = \lim_{n\to\infty} \sum_j V_j(t_{j+1} - t_j)$. Each term in the sum is normally distributed, and all partial sums are multivariate normal for all $\delta > 0$, and hence also in the limit as $\delta \to 0$. Obviously $\mathbb{E}(X_t) = 0$. For $s \leq t$,

$$\begin{aligned}\mathbb{E}(X_s X_t) &= \int_0^t \int_0^s \mathbb{E}(W_u W_v) \, du \, dv = \int_0^t \int_0^s \min\{u, v\} \, du \, dv \\ &= \int_0^s \tfrac{1}{2}u^2 \, du + \int_0^s u(t-u) \, du = s^2\left(\frac{t}{2} - \frac{s}{6}\right).\end{aligned}$$

Hence $\text{var}(X_t) = \frac{1}{3}t^3$, and the autocovariance function is

$$\rho(X_s, X_t) = 3\sqrt{\frac{s}{t}}\left(\frac{1}{2} - \frac{s}{6t}\right).$$

3. By the Cauchy–Schwarz inequality, as $n \to \infty$,

$$\mathbb{E}\big[\{\mathbb{E}(X_n \mid \mathcal{G}) - \mathbb{E}(X \mid \mathcal{G})\}^2\big] \leq \mathbb{E}\big[\mathbb{E}\{(X_n - X)^2 \mid \mathcal{G}\}\big] = \mathbb{E}[(X_n - X)^2] \to 0.$$

4. We square the equation $\|I(\psi_1 + \psi_2)\|_2 = \|\psi_1 + \psi_2\|$ and use the fact that $\|I(\psi_i)\|_2 = \|\psi_i\|$ for $i = 1, 2$, to deduce the result.

5. The question permits us to use the integrating factor $e^{\beta t}$ to give, formally,

$$e^{\beta t} X_t = \int_0^t e^{\beta s} \frac{dW_s}{ds} \, ds = e^{\beta t} W_t - \beta \int_0^t e^{\beta s} W_s \, ds$$

on integrating by parts. This is the required result, and substitution verifies that it satisfies the given equation.

6. Find a sequence $\boldsymbol{\phi} = (\phi^{(n)})$ of predictable step functions such that $\|\phi^{(n)} - \psi\| \to 0$ as $n \to \infty$. By the argument before equation (13.8.9), $I(\phi^{(n)}) \xrightarrow{\text{m.s.}} I(\psi)$ as $n \to \infty$. By Lemma (13.8.4), $\|I(\phi^{(n)})\|_2 = \|\phi^{(n)}\|$, and the claim follows.

13.9 Solutions. Itô's formula

1. The process Z is continuous and adapted with $Z_0 = 0$. We have by Theorem (13.8.11) that $\mathbb{E}(Z_t - Z_s \mid \mathcal{F}_s) = 0$, and by Exercise (13.8.6) that

$$\mathbb{E}\big([Z_t - Z_s]^2 \mid \mathcal{F}_s\big) = \mathbb{E}\left(\int_s^t \frac{X_u^2 + Y_u^2}{R_u^2} \, du\right) = t - s.$$

The first claim follows by the Lévy characterization of a Wiener process (12.7.10).

We have in n dimensions that $R^2 = X_1^2 + X_2^2 + \cdots + X_n^2$, and the same argument yields that $Z_t = \sum_i \int_0^t (X_i/R)\, dX_i$ is a Wiener process. By Example (13.9.7) and the above,

$$d(R^2) = 2\sum_{i=1}^{n} X_i\, dX_i + n\, dt = 2R\sum_{i=1}^{n} \frac{X_i}{R}\, dX_i + n\, dt = 2R\, dW + n\, dt.$$

2. Applying Itô's formula (13.9.4) to $Y_t = W_t^4$ we obtain $dY_t = 4W_t^3\, dW_t + 6W_t^2\, dt$. Hence,

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\int_0^t 4W_s^3\, dW_s\right) + \mathbb{E}\left(\int_0^t 6W_s^2\, ds\right) = 6\int_0^t s\, ds = 3t^2.$$

3. Apply Itô's formula (13.9.4) to obtain $dY_t = W_t\, dt + t\, dW_t$. Cf. Exercise (13.8.1).

4. Note that $X_1 = \cos W$ and $X_2 = \sin W$. By Itô's formula (13.9.4),

$$\begin{aligned} dY = d(X_1 + iX_2) &= dX_1 + i\, dX_2 = d(\cos W) + i\, d(\sin W) \\ &= -\sin W\, dW - \tfrac{1}{2}\cos W\, dt + i\cos W\, dW - \tfrac{1}{2}\sin W\, dt. \end{aligned}$$

5. We apply Itô's formula to obtain:
(a) $(1+t)\, dX = -X\, dt + dW$,
(b) $dX = -\frac{1}{2}X\, dt + \sqrt{1 - X^2}\, dW$,
(c) $d\begin{pmatrix} X \\ Y \end{pmatrix} = -\dfrac{1}{2}\begin{pmatrix} X \\ Y \end{pmatrix} dt + \begin{pmatrix} 0 & -a/b \\ b/a & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix} dW.$

13.10 Solutions. Option pricing

1. (a) We have that

$$\begin{aligned} \mathbb{E}\big((ae^Z - K)^+\big) &= \int_{\log(K/a)}^{\infty} (ae^z - K)\frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{(z-\gamma)^2}{2\tau^2}\right) dz \\ &= \int_{\alpha}^{\infty} (ae^{\gamma+\tau y} - K)\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\, dy \quad \text{where } y = \frac{z-\gamma}{\tau},\ \alpha = \frac{\log(K/a) - \gamma}{\tau} \\ &= ae^{\gamma+\frac{1}{2}\tau^2}\int_{\alpha}^{\infty}\frac{e^{-\frac{1}{2}(y-\tau)^2}}{\sqrt{2\pi}}\, dy - K\Phi(-\alpha) \\ &= ae^{\gamma+\frac{1}{2}\tau^2}\Phi(\tau - \alpha) - K\Phi(-\alpha). \end{aligned}$$

(b) We have that $S_T = ae^Z$ where $a = S_t$ and, under the relevant conditional $\mathbb{Q}$-distribution, Z is normal with mean $\gamma = (r - \frac{1}{2}\sigma^2)(T - t)$ and variance $\tau^2 = \sigma^2(T - t)$. The claim now follows by the result of part (a).

2. (a) Set $\xi(t, S) = \xi(t, S_t)$ and $\psi(t, S) = \psi(t, S_t)$, in the natural notation. By Theorem (13.10.15), we have $\psi_x = \psi_t = 0$, whence $\psi(t, x) = c$ for all t, x, and some constant c.
(b) Recall that $dS = \mu S\, dt + \sigma S\, dW$. Now,

$$d(\xi S + \psi e^{rt}) = d(S^2 + \psi e^{rt}) = (\sigma S)^2\, dt + 2S\, dS + e^{rt}\, d\psi + \psi re^{rt}\, dt,$$

by Example (13.9.7). By equation (13.10.4), the portfolio is self-financing if this equals $S\,dS + \psi r e^{rt}\,dt$, and thus we arrive at the SDE $e^{rt}\,d\psi = -S\,dS - \sigma^2 S^2\,dt$, whence

$$\psi(t, S) = -\int_0^t e^{-ru} S_u\,dS_u - \sigma^2 \int_0^t e^{-ru} S_u^2\,du.$$

(c) Note first that $Z_t = \int_0^t S_u\,du$ satisfies $dZ_t = S_t\,dt$. By Example (13.9.8), $d(S_t Z_t) = Z_t\,dS_t + S_t^2\,dt$, whence

$$d(\xi S + \psi e^{rt}) = Z_t\,dS_t + S_t^2\,dt + e^{rt}\,d\psi + re^{rt}\,dt.$$

Using equation (13.10.4), the portfolio is self-financing if this equals $Z_t\,dS_t + \psi r e^{rt}\,dt$, and thus we require that $e^{rt}\,d\psi = -S_t^2\,dt$, which is to say that

$$\psi(t, S) = -\int_0^t e^{-ru} S_u^2\,du.$$

3. We need to check equation (13.10.4) remembering that $dM_t = 0$. Each of these portfolios is self-financing.

(a) This case is obvious.

(b) $d(\xi S + \psi) = d(2S^2 - S^2 - t) = 2S\,dS + dt - dt = \xi\,dS$.

(c) $d(\xi S + \psi) = -S - t\,dS + S = \xi\,dS$.

(d) Recalling Example (13.9.8), we have that

$$d(\xi S + \psi) = d\left(S_t \int_0^t S_s\,ds - \int_0^t S_s^2\,ds\right) = S_t^2\,dt + dS_t \int_0^t S_s\,ds - S_t^2\,dt = \xi\,dS_t.$$

4. The time of exercise of an American call option must be a stopping time for the filtration $(\mathcal{F}_t)$. The value of the option, if exercised at the stopping time τ, is $V_\tau = (S_\tau - K)^+$, and it follows by the usual argument that the value at time 0 of the option exercised at τ is $\mathbb{E}_{\mathbb{Q}}(e^{-r\tau} V_\tau)$. Thus the value at time 0 of the American option is $\sup_\tau\{\mathbb{E}_{\mathbb{Q}}(e^{-r\tau} V_\tau)\}$, where the supremum is taken over all stopping times τ satisfying $\mathbb{P}(\tau \le T) = 1$. Under the probability measure $\mathbb{Q}$, the process $e^{-rt} V_t$ is a martingale, whence, by the optional stopping theorem, $\mathbb{E}_{\mathbb{Q}}(e^{-r\tau} V_\tau) = V_0$ for all stopping times τ. The claim follows.

5. We rewrite the value at time 0 of the European call option, possibly with the aid of Exercise (13.10.1), as

$$e^{-rT}\mathbb{E}\Big(\big(S_0 \exp\{rT - \tfrac{1}{2}\sigma^2 T + \sigma\sqrt{T}N\} - K\big)^+\Big) = \mathbb{E}\Big(\big(S_0 \exp\{-\tfrac{1}{2}\sigma^2 T + \sigma\sqrt{T}N\} - Ke^{-rT}\big)^+\Big),$$

where N is an $N(0,1)$ random variable. It is immediate that this is increasing in S_0 and r and is decreasing in K. To show monotonicity in T, we argue as follows. Let $T_1 < T_2$ and consider the European option with exercise date T_2. In the corresponding American option we are allowed to exercise the option at the earlier time T_1. By Exercise (13.10.4), it is never better to stop earlier than T_2, and the claim follows.

Monotonicity in σ may be shown by differentiation.

13.11 Solutions. Passage probabilities and potentials

1. Let H be a closed sphere with radius R ($> |w|$), and define $p_R(r) = \mathbb{P}\big(G \text{ before } H \bigm| |W(0)| = r\big)$. Then p_R satisfies Laplace's equation in $\mathbb{R}^d$, and hence

$$\frac{d}{dr}\left(r^{d-1}\frac{dp_R}{dr}\right) = 0$$

since p_R is spherically symmetric. Solve subject to the boundary equations $p_R(\epsilon) = 1$, $p_R(R) = 0$, to obtain

$$p_R(r) = \frac{r^{2-d} - R^{2-d}}{\epsilon^{2-d} - R^{2-d}} \to (\epsilon/r)^{d-2} \qquad \text{as } R \to \infty.$$

2. The electrical resistance R_n between 0 and the set Δ_n is no smaller than the resistance obtained by, for every $i = 1, 2, \dots$, 'shorting out' all vertices in the set Δ_i. This new network amounts to a linear chain of resistances in series, points labelled Δ_i and Δ_{i+1} being joined by a resistance if size N_i^{-1}. It follows that

$$R(G) = \lim_{n\to\infty} R_n \ge \sum_{i=0}^{\infty} \frac{1}{N_i}.$$

By Theorem (13.11.18), the walk is persistent if $\sum_i N_i^{-1} = \infty$.

3. Thinking of G as an electrical network, one may obtain the network H by replacing the resistance of every edge e lying in G but not in H by ∞. Let 0 be a vertex of H. By a well known fact in the theory of electrical networks, $R(H) \ge R(G)$, and the result follows by Theorem (13.11.19).

13.12 Solutions to problems

1. (a) $T(t) = \alpha W(t/\alpha^2)$ has continuous sample paths with stationary independent increments, since W has these properties. Also $T(t)/\alpha$ is $N(0, t/\alpha^2)$, whence $T(t)$ is $N(0, t)$.

(b) As for part (a).

(c) Certainly V has continuous sample paths on $(0, \infty)$. For continuity at 0 it suffices to prove that $tW(t^{-1}) \to 0$ a.s. as $t \downarrow 0$; this was done in the solution to Exercise (13.6.4).

If (u, v), (s, t) are disjoint time-intervals, then so are (v^{-1}, u^{-1}), (t^{-1}, s^{-1}); since W has independent increments, so has V. Finally,

$$V(s+t) - V(s) = tW((s+t)^{-1}) - s\{W(s^{-1}) - W((s+t)^{-1})\}$$

is $N(0, \beta)$ if $s, t > 0$, where

$$\beta = \frac{t^2}{s+t} + s^2\left(\frac{1}{s} - \frac{1}{s+t}\right) = t.$$

2. Certainly W is Gaussian with continuous sample paths and zero means, and it is therefore sufficient to prove that $\text{cov}(W(s), W(t)) = \min\{s, t\}$. Now, if $s \le t$,

$$\text{cov}\big(W(s), W(t)\big) = \frac{\text{cov}(X(r^{-1}(s)), X(r^{-1}(t)))}{v(r^{-1}(s))v(r^{-1}(t))} = \frac{u(r^{-1}(s))v(r^{-1}(t))}{v(r^{-1}(s))v(r^{-1}(t))} = r(r^{-1}(s)) = s$$

as required.

If $u(s) = s$, $v(t) = 1 - t$, then $r(t) = t/(1-t)$, and $r^{-1}(w) = w/(1+w)$ for $0 \le w < \infty$. In this case $X(t) = (1-t)W(t/(1-t))$.

3. Certainly U is Gaussian with zero means, and $U(0) = 0$. Now, with $s_t = e^{2\beta t} - 1$,

$$\begin{aligned}\mathbb{E}\{U(t+h) \mid U(t) = u\} &= e^{-\beta(t+h)}\mathbb{E}\{W(s_{t+h}) \mid W(s_t) = ue^{\beta t}\}\\ &= ue^{-\beta(t+h)}e^{\beta t} = u - \beta uh + \mathrm{o}(h),\end{aligned}$$

whence the instantaneous mean of U is $a(t,u) = -\beta u$. Secondly, $s_{t+h} = s_t + 2\beta e^{2\beta t}h + \mathrm{o}(h)$, and therefore

$$\begin{aligned}\mathbb{E}\{U(t+h)^2 \mid U(t) = u\} &= e^{-2\beta(t+h)}\mathbb{E}\{W(s_{t+h})^2 \mid W(s_t) = ue^{\beta t}\}\\ &= e^{-2\beta(t+h)}\big(u^2e^{2\beta t} + 2\beta e^{2\beta t}h + \mathrm{o}(h)\big)\\ &= u^2 - 2\beta h(u^2 - 1) + \mathrm{o}(h).\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}\{|U(t+h) - U(t)|^2 \mid U(t) = u\} &= u^2 - 2\beta h(u^2-1) - 2u(u - \beta uh) + u^2 + \mathrm{o}(h)\\ &= 2\beta h + \mathrm{o}(h),\end{aligned}$$

and the instantaneous variance is $b(t,u) = 2\beta$.

4. Bartlett's equation (see Exercise (13.3.4)) for $M(t,\theta) = \mathbb{E}(e^{\theta V(t)})$ is

$$\frac{\partial M}{\partial t} = -\beta\theta\frac{\partial M}{\partial \theta} + \tfrac{1}{2}\sigma^2\theta^2 M$$

with boundary condition $M(\theta, 0) = e^{\theta u}$. Solve this equation (as in the exercise given) to obtain

$$M(t,\theta) = \exp\left\{\theta ue^{-\beta t} + \frac{1}{2}\theta^2\cdot\frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right\},$$

the moment generating function of the given normal distribution. Now $M(t,\theta) \to \exp\{\frac{1}{2}\theta^2\sigma^2/(2\beta)\}$ as $t \to \infty$, whence by the continuity theorem $V(t)$ converges in distribution to the $N(0, \frac{1}{2}\sigma^2/\beta)$ distribution.

If $V(0)$ has this limit distribution, then so does $V(t)$ for all t. Therefore the sequence $(V(t_1), \dots, V(t_n))$ has the same joint distribution as $(V(t_1+h), \dots, V(t_n+h))$ for all $h, t_1, \dots, t_n$, whenever $V(0)$ has this normal distribution.

In the stationary case, $\mathbb{E}(V(t)) = 0$ and, for $s \le t$,

$$\operatorname{cov}\big(V(s), V(t)\big) = \mathbb{E}\{V(s)\mathbb{E}(V(t) \mid V(s))\} = \mathbb{E}\{V(s)^2e^{-\beta(t-s)}\} = c(0)e^{-\beta|t-s|}$$

where $c(0) = \operatorname{var}(V(s))$; we have used the first part here. This is the autocovariance function of a stationary Gaussian Markov process (see Example (9.6.10)). Since all such processes have autocovariance functions of this form (i.e., for some choice of β), all such processes are stationary Ornstein–Uhlenbeck processes.

The autocorrelation function is $\rho(s) = e^{-\beta|s|}$, which is the characteristic function of the Cauchy density function

$$f(x) = \frac{1}{\beta\pi\{1 + (x/\beta)^2\}}, \qquad x \in \mathbb{R}.$$

5. Bartlett's equation (see Exercise (13.3.2)) for M is

$$\frac{\partial M}{\partial t} = \alpha\theta\frac{\partial M}{\partial \theta} + \tfrac{1}{2}\beta\theta^2\frac{\partial M}{\partial \theta},$$

subject to $M(0,\theta) = e^{\theta d}$. The characteristics satisfy

$$\frac{dM}{0} = \frac{dt}{1} = -\frac{2\,d\theta}{2\alpha\theta + \beta\theta^2}.$$

The solution is $M = g(\theta e^{\alpha t}/(\alpha + \frac{1}{2}\beta\theta))$ where g is a function satisfying $g(\theta/(\alpha + \frac{1}{2}\beta\theta)) = e^{\theta d}$. The solution follows as given.

By elementary calculations,

$$\mathbb{E}(D(t)) = \left.\frac{\partial M}{\partial \theta}\right|_{\theta=0} = de^{\alpha t},$$
$$\mathbb{E}(D(t)^2) = \left.\frac{\partial^2 M}{\partial \theta^2}\right|_{\theta=0} = \frac{\beta d}{\alpha}e^{\alpha t}(e^{\alpha t} - 1) + d^2 e^{2\alpha t},$$

whence $\text{var}(D(t)) = (\beta d/\alpha)e^{\alpha t}(e^{\alpha t} - 1)$. Finally

$$\mathbb{P}(D(t) = 0) = \lim_{\theta\to-\infty} M(t,\theta) = \exp\left\{\frac{2\alpha d e^{\alpha t}}{\beta(1 - e^{\alpha t})}\right\}$$

which converges to $e^{-2\alpha d/\beta}$ as $t \to \infty$.

6. The equilibrium density function $g(y)$ satisfies the (reduced) forward equation

$$(*)\qquad -\frac{d}{dy}(ag) + \frac{1}{2}\frac{d^2}{dy^2}(bg) = 0$$

where $a(y) = -\beta y$ and $b(y) = \sigma^2$ are the instantaneous mean and variance. The boundary conditions are

$$\beta y g + \tfrac{1}{2}\sigma^2\frac{dg}{dy} = 0, \qquad y = -c, d.$$

Integrate $(*)$ from $-c$ to y, using the boundary conditions, to obtain

$$\beta y g + \tfrac{1}{2}\sigma^2\frac{dg}{dy} = 0, \qquad -c \le y \le d.$$

Integrate again to obtain $g(y) = Ae^{-\beta y^2/\sigma^2}$. The constant A is given by the fact that $\int_{-c}^{d} g(y)\,dy = 1$.

7. First we show that the series converges uniformly (along a subsequence), implying that the limit exists and is a continuous function of t. Set

$$Z_{mn}(t) = \sum_{k=m}^{n-1}\frac{\sin(kt)}{k}X_k, \quad M_{mn} = \sup\{|Z_{mn}(t)| : 0 \le t \le \pi\}.$$

We have that

$$(*)\qquad M_{mn}^2 \le \sup_{0\le t\le\pi}\left|\sum_{k=m}^{n-1}\frac{e^{ikt}}{k}X_k\right|^2 \le \sum_{k=m}^{n-1}\frac{X_k^2}{k^2} + 2\sum_{l=1}^{n-m-1}\left|\sum_{j=m}^{n-l-1}\frac{X_j X_{j+l}}{j(j+l)}\right|.$$

The mean value of the final term is, by the Cauchy–Schwarz inequality, no larger than

$$2\sum_{l=1}^{n-m-1}\sqrt{\mathbb{E}\left(\left|\sum_{j=m}^{n-l-1}\frac{X_jX_{j+l}}{j(j+l)}\right|^2\right)}=2\sum_{l=1}^{n-m-1}\sqrt{\sum_{j=m}^{n-l-1}\frac{1}{j^2(j+l)^2}}\le 2(n-m)\sqrt{\frac{n-m}{m^4}}.$$

Combine this with $(*)$ to obtain

$$\mathbb{E}(M_{m,2m})^2\le\mathbb{E}(M_{m,2m}^2)\le\frac{3}{\sqrt{m}}.$$

It follows that

$$\mathbb{E}\left(\sum_{n=1}^{\infty}M_{2^{n-1},2^n}\right)\le\sum_{n=1}^{\infty}\frac{6}{2^{n/2}}<\infty,$$

implying that $\sum_{n=1}^{\infty}M_{2^{n-1},2^n}<\infty$ a.s. Therefore the series which defines W converges uniformly with probability 1 (along a subsequence), and hence W has (a.s.) continuous sample paths.

Certainly W is a Gaussian process since $W(t)$ is the sum of normal variables (see Problem (7.11.19)). Furthermore $\mathbb{E}(W(t))=0$, and

$$\operatorname{cov}\big(W(s),W(t)\big)=\frac{st}{\pi}+\frac{2}{\pi}\sum_{k=1}^{\infty}\frac{\sin(ks)\sin(kt)}{k^2}$$

since the X_i are independent with zero means and unit variances. It is an exercise in Fourier analysis to deduce that $\operatorname{cov}(W(s),W(t))=\min\{s,t\}$.

8. We wish to find a solution $g(t,y)$ to the equation

$$(*)\qquad \frac{\partial g}{\partial t}=\frac{1}{2}\frac{\partial^2 g}{\partial y^2},\qquad |y|<b,$$

satisfying the boundary conditions

$$g(0,y)=\delta_{y0}\quad\text{if } |y|\le b,\qquad g(t,y)=0\quad\text{if } |y|=b.$$

Let $g(t,y\mid d)$ be the $N(d,t)$ density function, and note that $g(\cdot,\cdot\mid d)$ satisfies $(*)$ for any 'source' d. Let

$$g(t,y)=\sum_{k=-\infty}^{\infty}(-1)^k g(t,y\mid 2kb),$$

a series which converges absolutely and is differentiable term by term. Since each summand satisfies $(*)$, so does the sum. Now $g(0,y)$ is a combination of Dirac delta functions, one at each multiple of $2b$. Only one such multiple lies in $[-b,b]$, and hence $g(y,0)=\delta_{d0}$. Also, setting $y=b$, the contributions from the sources at $-2(k-1)b$ and $2kb$ cancel, so that $g(t,b)=0$. Similarly $g(t,-b)=0$, and therefore g is the required solution.

Here is an alternative method. Look for the solution to $(*)$ of the form $e^{-\lambda_n t}\sin\{\frac{1}{2}n\pi(y+b)/b\}$; such a sine function vanishes when $|y|=b$. Substitute into $(*)$ to obtain $\lambda_n=n^2\pi^2/(8b^2)$. A linear combination of such functions has the form

$$g(t,y)=\sum_{n=1}^{\infty}a_n e^{-\lambda_n t}\sin\left(\frac{n\pi(y+b)}{2b}\right).$$

We choose the constants a_n such that $g(0, y) = \delta_{y0}$ for $|y| < b$. With the aid of a little Fourier analysis, one finds that $a_n = b^{-1}\sin(\frac{1}{2}n\pi)$.

Finally, the required probability equals the probability that W^{a} has been absorbed by time t, a probability expressible as $1 - \int_{-b}^{b} f^{\mathrm{a}}(t, y)\,dy$. Using the second expression for f^{a}, this yields

$$\frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}e^{-\lambda_n t}\sin^3(\tfrac{1}{2}n\pi).$$

9. Recall that $U(t) = e^{-2mD(t)}$ is a martingale. Let T be the time of absorption, and assume that the conditions of the optional stopping theorem are satisfied. Then $\mathbb{E}(U(0)) = \mathbb{E}(U(T))$, which is to say that $1 = e^{2ma}p_a + e^{-2mb}(1 - p_a)$.

10. (a) We may assume that $a, b > 0$. With

$$p_t(b) = \mathbb{P}\big(W(t) > b,\ F(0, t)\ \big|\ W(0) = a\big),$$

we have by the reflection principle that

$$\begin{aligned} p_t(b) &= \mathbb{P}\big(W(t) > b \mid W(0) = a\big) - \mathbb{P}\big(W(t) < -b \mid W(0) = a\big) \\ &= \mathbb{P}\big(b - a < W(t) < b + a \mid W(0) = 0\big), \end{aligned}$$

giving that

$$\frac{\partial p_t(b)}{\partial b} = f(t, b + a) - f(t, b - a)$$

where $f(t, x)$ is the $N(0, t)$ density function. Now, using conditional probabilities,

$$\mathbb{P}\big(F(0, t) \mid W(0) = a,\ W(t) = b\big) = -\frac{1}{f(t, b - a)}\frac{\partial p_t(b)}{\partial b} = 1 - e^{-2ab/t}.$$

(b) We know that

$$\mathbb{P}(F(s, t)) = 1 - \frac{2}{\pi}\cos^{-1}\{\sqrt{s/t}\} = \frac{2}{\pi}\sin^{-1}\{\sqrt{s/t}\}$$

if $0 < s < t$. The claim follows since $F(t_0, t_2) \subseteq F(t_1, t_2)$.

(c) Remember that $\sin x = x + o(x)$ as $x \downarrow 0$. Take the limit in part (b) as $t_0 \downarrow 0$ to obtain $\sqrt{t_1/t_2}$.

11. Let $M(t) = \sup\{W(s) : 0 \le s \le t\}$ and recall that $M(t)$ has the same distribution as $|W(t)|$. By symmetry,

$$\mathbb{P}\Big(\sup_{0\le s\le t}|W(s)| \ge w\Big) \le 2\mathbb{P}\big(M(t) \ge w\big) = 2\mathbb{P}\big(|W(t)| \ge w\big).$$

By Chebyshov's inequality,

$$\mathbb{P}\big(|W(t)| \ge w\big) \le \frac{\mathbb{E}(W(t)^2)}{w^2} = \frac{t}{w^2}.$$

Fix $\epsilon > 0$, and let

$$A_n(\epsilon) = \big\{|W(s)|/s > \epsilon \text{ for some } s \text{ satisfying } 2^{n-1} < s \le 2^n\big\}.$$

Note that

$$A_n(\epsilon) \subseteq \Big\{\sup_{2^{n-1}<s\le 2^n}|W(s)| \ge 2^{2n/3}\Big\} \subseteq \Big\{\sup_{0\le s\le 2^n}|W(s)| \ge 2^{2n/3}\Big\}$$

for all large n, and also

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\sup_{0 \le s \le 2^n} |W(s)| \ge 2^{2n/3} \right) \le \sum_{n=1}^{\infty} \frac{2^{n+1}}{2^{4n/3}} < \infty.$$

Therefore $\sum_n \mathbb{P}(A_n(\epsilon)) < \infty$, implying by the Borel–Cantelli lemma that (a.s.) only finitely many of the $A_n(\epsilon)$ occur. Therefore $t^{-1}W(t) \to 0$ a.s. as $t \to \infty$. Compare with the solution to the relevant part of Exercise (13.6.4).

12. We require the solution to Laplace's equation $\nabla^2 p = 0$, subject to the boundary condition

$$p(\mathbf{w}) = \begin{cases} 0 & \text{if } \mathbf{w} \in H, \\ 1 & \text{if } \mathbf{w} \in G. \end{cases}$$

Look for a solution in polar coordinates of the form

$$p(r, \theta) = \sum_{n=0}^{\infty} r^n \{a_n \sin(n\theta) + b_n \cos(n\theta)\}.$$

Certainly combinations having this form satisfy Laplace's equation, and the boundary condition gives that

$$(*) \qquad H(\theta) = b_0 + \sum_{n=1}^{\infty} \{a_n \sin(n\theta) + b_n \cos(n\theta)\}, \qquad |\theta| < \pi,$$

where

$$H(\theta) = \begin{cases} 0 & \text{if } -\pi < \theta < 0, \\ 1 & \text{if } 0 < \theta < \pi. \end{cases}$$

The collection $\{\sin(m\theta), \cos(m\theta) : m \ge 0\}$ are orthogonal over $(-\pi, \pi)$. Multiply through $(*)$ by $\sin(m\theta)$ and integrate over $(-\pi, \pi)$ to obtain $\pi a_m = \{1 - \cos(\pi m)\}/m$, and similarly $b_0 = \frac{1}{2}$ and $b_m = 0$ for $m \ge 1$.

13. The joint density function of two independent $N(0, t)$ random variables is $(2\pi t)^{-1} \exp\{-(x^2 + y^2)/(2t)\}$. Since this function is unchanged by rotations of the plane, it follows that the two coordinates of the particle's position are independent Wiener processes, regardless of the orientation of the coordinate system. We may thus assume that l is the line $x = d$ for some fixed positive d.

The particle is bound to visit the line l sooner or later, since $\mathbb{P}(W_1(t) < d \text{ for all } t) = 0$. The first-passage time T has density function

$$f_T(t) = \frac{d}{\sqrt{2\pi t^3}} e^{-d^2/(2t)}, \qquad t > 0.$$

Conditional on $\{T = t\}$, $D = W_2(T)$ is $N(0, t)$. Therefore the density function of D is

$$f_D(u) = \int_0^{\infty} f_{D|T}(u \mid t) f_T(t)\, dt = \int_0^{\infty} \frac{d}{2\pi t^2} e^{-(u^2+d^2)/(2t)}\, dt = \frac{d}{\pi(u^2 + d^2)}, \qquad u \in \mathbb{R},$$

giving that D/d has the Cauchy distribution.

The angle $\Theta = \widehat{\mathrm{POR}}$ satisfies $\theta = \tan^{-1}(D/d)$, whence

$$\mathbb{P}(\Theta \le \theta) = \mathbb{P}(D \le d \tan\theta) = \frac{1}{2} + \frac{\theta}{\pi}, \qquad |\theta| < \tfrac{1}{2}\pi.$$

14. By an extension of Itô's formula to functions of two Wiener processes, $U = u(W_1, W_2)$ and $V = v(W_1, W_2)$ satisfy

$$dU = u_x\,dW_1 + u_y\,dW_2 + \tfrac{1}{2}(u_{xx} + u_{yy})\,dt,$$
$$dV = v_x\,dW_1 + v_y\,dW_2 + \tfrac{1}{2}(v_{xx} + v_{yy})\,dt,$$

where u_x, v_{yy}, etc, denote partial derivatives of u and v. Since ϕ is analytic, u and v satisfy the Cauchy–Riemann equations $u_x = v_y$, $u_y = -v_x$, whence u and v are harmonic in that $u_{xx} + u_{yy} = v_{xx} + v_{yy} = 0$. Therefore,

$$dU = u_x\,dW_1 + u_y\,dW_2, \quad dV = -u_y\,dW_1 + u_x\,dW_2.$$

The matrix $\begin{pmatrix} u_x & u_y \\ -u_y & u_x \end{pmatrix}$ is an orthogonal rotation of $\mathbb{R}^2$ when $u_x^2 + u_y^2 = 1$. Since the joint distribution of the pair (W_1, W_2) is invariant under such rotations, the claim follows.

15. One method of solution uses the fact that the reversed Wiener process $\{W(t-s) - W(t) : 0 \le s \le t\}$ has the same distribution as $\{W(s) : 0 \le s \le t\}$. Thus $M(t) - W(t) = \max_{0 \le s \le t}\{W(s) - W(t)\}$ has the same distribution as $\max_{0 \le u \le t}\{W(u) - W(0)\} = M(t)$. Alternatively, by the reflection principle,

$$\mathbb{P}\big(M(t) \ge x,\ W(t) \le y\big) = \mathbb{P}(W(t) \ge 2x - y) \quad \text{for } x \ge \max\{0, y\}.$$

By differentiation, the pair $M(t)$, $W(t)$ has joint density function $-2\phi'(2x - y)$ for $y \le x$, $x \ge 0$, where ϕ is the density function of the $N(0, t)$ distribution. Hence $M(t)$ and $M(t) - W(t)$ have the joint density function $-2\phi'(x + y)$. Since this function is symmetric in its arguments, $M(t)$ and $M(t) - W(t)$ have the same marginal distribution.

16. The Lebesgue measure $\Lambda(Z)$ is given by

$$\Lambda(Z) = \int_0^\infty I_{\{W(t)=u\}}\,du,$$

whence by Fubini's theorem (cf. equation (5.6.13)),

$$\mathbb{E}(\Lambda(Z)) = \int_0^\infty \mathbb{P}\big(W(t) = u\big)\,dt = 0.$$

17. Let $0 < a < b < c < d$, and let $M(x, y) = \max_{x \le s \le y} W(s)$. Then

$$M(c,d) - M(a,b) = \max_{c \le s \le d}\big\{W(s) - W(c)\big\} + \big\{W(c) - W(b)\big\} - \max_{a \le s \le b}\big\{W(s) - W(b)\big\}.$$

Since the three terms on the right are independent and continuous random variables, it follows that $\mathbb{P}\big((M(c,d) = M(a,b)\big) = 0$. Since there are only countably many rationals, we deduce that $\mathbb{P}\big((M(c,d) = M(a,b) \text{ for all rationals } a < b < c < d\big) = 1$, and the result follows.

18. The result is easily seen by exhaustion to be true when $n = 1$. Suppose it is true for all $m \le n - 1$ where $n \ge 2$.

(i) If $s_n \le 0$, then (whatever the final term of the permutation) the number of positive partial sums and the position of the first maximum depend only on the remaining $n - 1$ terms. Equality follows by the induction hypothesis.

(ii) If $s_n > 0$, then

$$A_r = \sum_{k=1}^{n} A_{r-1}(k),$$

where $A_{r-1}(k)$ is the number of permutations with x_k in the final place, for which exactly $r-1$ of the first $n-1$ terms are strictly positive. Consider a permutation $\pi = (x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}}, x_k)$ with x_k in the final place, and move the position of x_k to obtain the new permutation $\pi' = (x_k, x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}})$. The first appearance of the maximum in π' is at its rth place if and only if the first maximum of the reduced permutation $(x_{i_1}, x_{i_2}, \dots, x_{i_{n-1}})$ is at its $(r-1)$th place. [Note that $r = 0$ is impossible since $s_n > 0$.] It follows that

$$B_r = \sum_{k=1}^{n} B_{r-1}(k),$$

where $B_{r-1}(k)$ is the number of permutations with x_k in the final place, for which the first appearance of the maximum is at the $(r-1)$th place.

By the induction hypothesis, $A_{r-1}(k) = B_{r-1}(k)$, since these quantities depend on the $n-1$ terms excluding x_k. The result follows.

19. Suppose that $S_m = \sum_{j=1}^{m} X_j$, $0 \le m \le n$, are the partial sums of n independent identically distributed random variables X_j. Let A_n be the number of strictly positive partial sums, and R_n the index of the first appearance of the value of the maximal partial sum. Each of the $n!$ permutations of $(X_1, X_2, \dots, X_n)$ has the same joint distribution. Consider the kth permutation, and let I_k be the indicator function of the event that exactly r partial sums are positive, and let J_k be the indicator function that the first appearance of the maximum is at the rth place. Then, using Problem (13.12.18),

$$\mathbb{P}(A_n = r) = \frac{1}{n!}\sum_{k=1}^{n!} \mathbb{E}(I_k) = \frac{1}{n!}\sum_{k=1}^{n!} \mathbb{E}(J_k) = \mathbb{P}(R_n = r).$$

We apply this with $X_j = W(jt/n) - W((j-1)t/n)$, so that $S_m = W(mt/n)$. Thus $A_n = \sum_j I_{\{W(jt/n)>0\}}$ has the same distribution as

$$R_n = \min\{k \ge 0 : W(kt/n) = \max_{0 \le j \le n} W(jt/n)\}.$$

By Problem (13.12.17), $R_n \xrightarrow{\text{a.s.}} R$ as $n \to \infty$. By Problem (13.12.16), the time spent by W at zero is a.s. a null set, whence $A_n \xrightarrow{\text{a.s.}} A$. Hence A and R have the same distribution. We argue as follows to obtain that that L and R have the same distribution. Making repeated use of Theorem (13.4.6) and the symmetry of W,

$$\begin{aligned}
\mathbb{P}(L < x) &= \mathbb{P}\Big(\sup_{x \le s \le t} W(s) < 0\Big) + \mathbb{P}\Big(\inf_{x \le s \le t} W(s) > 0\Big) \\
&= 2\mathbb{P}\Big(\sup_{x \le s \le t} \{W(s) - W(x)\} < -W(x)\Big) = 2\mathbb{P}\big(|W(t) - W(x)| < W(x)\big) \\
&= \mathbb{P}\big(|W(t) - W(x)| < |W(x)|\big) \\
&= \mathbb{P}\Big(\sup_{x \le s \le t} \{W(s) - W(x)\} < \sup_{0 \le s \le x} \{W(s) - W(x)\}\Big) = \mathbb{P}(R \le x).
\end{aligned}$$

Finally, by Problem (13.12.15) and the circular symmetry of the joint density distribution of two independent $N(0,1)$ variables U, V,

$$\mathbb{P}\big(|W(t) - W(x)| < |W(x)|\big) = \mathbb{P}\big((t-x)V^2 \le xU^2\big) = \mathbb{P}\left(\frac{V^2}{U^2 + V^2} \le \frac{x}{t}\right) = \frac{2}{\pi}\sin^{-1}\sqrt{\frac{x}{t}}.$$

20. Let

$$T_x = \begin{cases} \inf\{t \le 1 : W(t) = x\} & \text{if this set is non-empty,} \\ 1 & \text{otherwise,} \end{cases}$$

and similarly $V_x = \sup\{t \le 1 : W(t) = x\}$, with $V_x = 1$ if $W(t) \ne x$ for all $t \in [0, 1]$. Recall that U_0 and V_0 have an arc sine distribution as in Problem (13.12.19). On the event $\{U_x < 1\}$, we may write (using the re-scaling property of W)

$$U_x = T_x + (1 - T_x)\widetilde{U}_0, \quad V_x = T_x + (1 - T_x)\widetilde{V}_0,$$

where $\widetilde{U}_0$ and $\widetilde{V}_0$ are independent of U_x and V_x, and have the above arc sine distribution. Hence U_x and V_x have the same distribution. Now T_x has the first passage distribution of Theorem (13.4.5), whence

$$f_{T_x,\widetilde{U}_0}(\tau, \phi) = \left\{ \frac{x}{\sqrt{2\pi\tau^3}} \exp\left(-\frac{x^2}{2\tau}\right) \right\} \left\{ \frac{1}{\pi\sqrt{\phi(1-\phi)}} \right\}.$$

Therefore,

$$f_{T_x,U_x}(t, u) = f_{T_x,\widetilde{U}_0}\left(t, \frac{u-t}{1-t}\right) \cdot \frac{1}{1-t},$$

and

$$f_{U_x}(u) = \int_0^u f_{T_x,U_x}(t, u)\, du = \frac{1}{\pi\sqrt{u(1-u)}} \exp\left(-\frac{x^2}{2u}\right), \quad 0 < x < 1.$$

21. Note that V is a martingale, by Theorem (13.8.11). Fix t and let $\psi_s = \text{sign}(W_s)$, $0 \le s \le t$. We have that $\|\psi\| = \sqrt{t}$, implying by Exercise (13.8.6) that $\mathbb{E}(V_t^2) = \|I(\psi)\|_2^2 = t$. By a similar calculation, $\mathbb{E}(V_t^2 \mid \mathcal{F}_s) = V_s^2 + t - s$ for $0 \le s \le t$. That is to say, $V_t^2 - t$ defines a martingale, and the result follows by the Lévy characterization theorem of Example (12.7.10).

22. The mean cost per unit time is

$$\mu(T) = \frac{1}{T}\left\{ R + C\int_0^T \mathbb{P}\big(|W(t)| \ge a\big)\, dt \right\} = \frac{1}{T}\left\{ R + 2C\int_0^T \big(1 - \Phi(a/\sqrt{t})\big)\, dt \right\}.$$

Differentiate to obtain that $\mu'(T) = 0$ if

$$R = 2C\left\{ \int_0^T \Phi(a/\sqrt{t})\, dt - T\Phi(a/\sqrt{T}) \right\} = aC\int_0^T t^{-1}\phi(a/\sqrt{t})\, dt,$$

where we have integrated by parts.

23. Consider the portfolio with $\xi(t, S_t)$ units of stock and $\psi(t, S_t)$ units of bond, having total value $w(t, S_t) = x\xi(t, x) + e^{rt}\psi(t, S_t)$. By assumption,

$$(*) \qquad (1 - \gamma)x\xi(t, x) = \gamma e^{rt}\psi(t, x).$$

Differentiate this equation with respect to x and substitute from equation (13.10.16) to obtain the differential equation $(1 - \gamma)\xi + x\xi_x = 0$, with solution $\xi(t, x) = h(t)x^{\gamma-1}$, for some function $h(t)$. We substitute this, together with $(*)$, into equation (13.10.17) to obtain that

$$h' - h(1 - \gamma)(\tfrac{1}{2}\gamma\sigma^2 + r) = 0.$$

It follows that $h(t) = A\exp\{(1 - \gamma)(\frac{1}{2}\gamma\sigma^2 + r)t\}$, where A is an absolute constant to be determined according to the size of the initial investment. Finally, $w(t, x) = \gamma^{-1}x\xi(t, x) = \gamma^{-1}h(t)x^{\gamma}$.

24. Using Itô's formula (13.9.4), the drift term in the SDE for U_t is

$$\big(-u_1(T - t, W) + \tfrac{1}{2}u_{22}(T - t, W)\big)\, dt,$$

where u_1 and u_{22} denote partial derivatives of u. The drift function is identically zero if and only if $u_1 = \frac{1}{2}u_{22}$.

Bibliography

A man will turn over half a library to make one book. Samuel Johnson

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions with formulas, graphs and mathematical tables*. Dover, New York.

Billingsley, P. (1995). *Probability and measure* (3rd edn). Wiley, New York.

Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, MA, reprinted by SIAM, 1992.

Chung, K. L. (1974). *A course in probability theory* (2nd edn). Academic Press, New York.

Cox, D. R. and Miller, H. D. (1965). *The theory of stochastic processes*. Chapman and Hall, London.

Doob, J. L. (1953). *Stochastic processes*. Wiley, New York.

Feller, W. (1968). *An introduction to probability theory and its applications,* Vol. 1 (3rd edn). Wiley, New York.

Feller, W. (1971). *An introduction to probability theory and its applications,* Vol. 2 (2nd edn). Wiley, New York.

Grimmett, G. R. and Stirzaker, D. R. (2001). *Probability and random processes,* (3rd edn). Oxford University Press, Oxford.

Grimmett, G. R. and Welsh, D. J. A. (1986). *Probability, an introduction.* Clarendon Press, Oxford.

Hall, M. (1983). *Combinatorial theory* (2nd edn). Wiley, New York.

Harris, T. E. (1963). *The theory of branching processes*. Springer, Berlin.

Karlin, S. and Taylor, H. M. (1975). *A first course in stochastic processes* (2nd edn). Academic Press, New York.

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press, New York.

Laha, R. G. and Rohatgi, V. K. (1979). *Probability theory*. Wiley, New York.

Loève, M. (1977). *Probability theory,* Vol. 1 (4th edn). Springer, Berlin.

Loève, M. (1978). *Probability theory,* Vol. 2 (4th edn). Springer, Berlin.

Moran, P. A. P. (1968). *An introduction to probability theory*. Clarendon Press, Oxford.

Stirzaker, D. R. (1994). *Elementary probability*. Cambridge University Press, Cambridge.

Stirzaker, D. R. (1999). *Probability and random variables*. Cambridge University Press, Cambridge.

Williams, D. (1991). *Probability with martingales*. Cambridge University Press, Cambridge.

Index

Abbreviations used in this index: c.f. characteristic function; distn distribution; eqn equation; fn function; m.g.f. moment generating function; p.g.f. probability generating function; pr. process; r.v. random variable; r.w. random walk; s.r.w. simple random walk; thm theorem.

E

This is a revised, updated and greatly expanded version of the authors' *Probability and Random Processes: Problems and Solutions*, first published in 1992. The 1000+ exercises contained within are not merely drill problems but have been chosen to illustrate the concepts, illuminate the subject and both inform and entertain the student. A broad range of subjects is covered, including: elementary aspects of probability and random variables; sampling; Markov chains; convergence; stationary processes; renewals; queues; martingales; diffusion; mathematical finance and the Black–Scholes model. A bibliography and extensive index conclude the volume.

This text is intended to serve students as a companion for elementary, intermediate or advanced courses in probability and random processes. It will also be useful for anyone needing a source for large numbers of problems in these areas. In particular this text acts as a companion to the authors' *Probability and Random Processes: Third Edition* (OUP, 2001).

FROM A REVIEW OF *PROBABILITY AND RANDOM PROCESSES: PROBLEMS AND SOLUTIONS*

A wealth of interesting teaching material at all levels.
P. A. L. Embrechts *Short Book Reviews*

OXFORD
UNIVERSITY PRESS

ISBN 0-19-857221-2

9 780198 572213

www.oup.com