Solutions, answers, and hints for selected problems

Complete solutions of some problems are given. Answers only are given for some other problems. For still others, only hints or partial solutions are given. Asterisks in "A Modern Approach to Probability Theory" by Fristedt and Gray identify the problems that are treated in this supplement.

For Chapter 1

1-2. Method 1: By the Binomial Theorem,

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

and, for $n > 0$,

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

Addition and then division by 2 gives

$$\sum_{\substack{k=0 \\ k \text{ even}}}^{n} \binom{n}{k} = 2^{n-1}\,.$$

The answer for positive n is $2^{n-1}/2^n = 1/2$. The answer for $n = 0$ is easily seen to equal 1.

Method 2: For $n \geq 1$ consider a sequence of length $(n-1)$. If it contains an even number of 'heads', adjoin a 'tails' to it to obtain a length-n sequence containing an even number of 'heads'. If it contains an odd number of 'heads', adjoin a 'heads' to it to obtain a length-n sequence containing an even number of 'heads'. Moreover, all length-n sequences containing an even number of 'heads' are obtained by one of the preceding two procedures. We have thus established, for $n \geq 1$, a one-to-one correspondence between the set of all length-$(n-1)$ sequences and the set of those length-n sequences that contain an even number of 'heads'. Therefore, there are 2^{n-1} length-n sequences that contain an even number of 'heads'. To treat the remaining case $n = 0$, we observe that the empty sequence, which is the only length-0 sequence, contains zero ‘heads’. Since 0 is even, there is 1 length-0 sequence containing an even number of ‘heads’.

1-4. 2^{-j}

1-10. The thirty-six points to each of which is assigned probability $\frac{1}{36}$ are the ordered pairs (r, g) for $1 \leq r \leq 6$ and $1 \leq g \leq 6$. The coordinates r and g represent the numbers showing on the red die and green die, respectively.

1-11. The set consisting of a single sample point, being the intersection of countably many events A of the form (1.2), is an event. Its probability is no larger than that of any such A. For each n and each sample point, there is such an A that has probability 2^{-n}. Thus, the probability of that sample point is no larger than 2^{-n}. Letting $n \to \infty$ we see that the probability of the sample point is 0. The process of flipping a coin until the first tails occurs terminates in a finite number of steps with probability 1.

1-12. (i) Sum the answer to Problem 4 over odd positive j to obtain $\frac{2}{3}$.

(ii) $\frac{1}{16}$

(iii) (Caution: it is common for students to use invalid reasoning in this type of problem.) We use ‘1’ and ‘0’ to denote heads and tails, respectively. Let S denote the set of finite sequences s of 1’s and 0’s terminating with 1, containing no subsequence of the form $(1, 0, 1)$ or $(1, 1, 1)$, and having the additional property that if the length of s is at least two, then the penultimate term in s is 0. For each $s \in S$, let A_s be the event consisting of those infinite sequences ω that begin with s followed by $(0, 1, 1)$, $(0, 1, 0)$, or $(1, 0, 1)$ in the next three positions, and let B_s be the event consisting of those ω that begin with s followed by $(1, 1, 1)$ or $(1, 1, 0)$ in the next three positions. Note that each A_s and B_s is a member of $\mathcal{E}$. Clearly $2P(A_s) = 3P(B_s)$.

Let $A = \bigcup_{s \in S} A_s$ and $B = \bigcup_{s \in S} B_s$. Straightforward set-theoretic arguments show that A consists of those ω in which $(1, 0, 1)$ occurs before $(1, 1, 1)$, B consists of those ω in which $(1, 1, 1)$ occurs before $(1, 0, 1)$. By writing A and B as countable unions of members of $\mathcal{E}$, we have shown that they are events. Note that in each case, these unions are taken over a family of pairwise disjoint events, from which it follows that

$$2P(A) = 2 \sum_{s \in S} P(A_s) = 3 \sum_{s \in S} P(B_s) = 3P(B).$$

Also, A and B are clearly disjoint, so

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

We will show that $P(A^c \cap B^c) = 0$, so that the above two equalities become two equations in the two unknowns $P(A)$ and $P(B)$, the solution of which gives $P(A) = \frac{3}{5}$.

To show that $P(A^c \cap B^c) = 0$ we note that $A^c \cap B^c$ is a subset of the event D_k consisting of those ω that begin with a sequence of length $3k$ having the property that, for $1 \leq j \leq k$, the sequence $(1, 1, 1)$ does not occur in positions $3j - 2$, $3j - 1$, $3j$. The number of ways of filling the first $3k$ positions of ω with 1’s and 0’s is $2^{3k} = 8^k$. The number of ways of doing it so as to obtain a member of D_k is 7^k (7 choices for positions $1, 2, 3$; 7 choices for positions $4, 5, 6$ and so forth.). Thus, $P(A^c \cap B^c) \leq P(D_k) = (\frac{7}{8})^k$. Now let $k \to \infty$ to obtain the desired conclusion, $P(A^c \cap B^c) = 0$.

(iv) $\frac{5}{8}$

1-14 Let $\mathcal{B}$ denote the Borel σ-field of $\mathbb{R}$, $\mathcal{C}$ the Borel σ-field of $\mathbb{R}^+$, and

$$\mathcal{G} = \{B \in \mathcal{B} \colon B \subseteq \mathbb{R}^+\}\,.$$

The goal is to prove $\mathcal{C} = \mathcal{G}$.

We first prove that $\mathcal{G}$ is a σ-field of subsets of $\mathbb{R}^+$. Countable unions of members of $\mathcal{B}$ are members of $\mathcal{B}$ and unions of subsets of $\mathbb{R}^+$ are subsets of $\mathbb{R}^+$. Hence, $\mathcal{G}$ is closed under countable unions. The complement in $\mathbb{R}^+$ of a member G of $\mathcal{G}$ equals $\mathbb{R}^+ \cap G^c$, where G^c denotes the complement in $\mathbb{R}$. This set is clearly a subset of $\mathbb{R}^+$ and it is also a member of $\mathcal{B}$ because it is the intersection of two members of $\mathcal{B}$. Therefore, $\mathcal{G}$ is a σ-field.

The open subsets of $\mathbb{R}^+$ have the form $\mathbb{R}^+ \cap O$, where O is open in $\mathbb{R}$. Such sets, being subsets of $\mathbb{R}^+$ and intersections of two members of $\mathcal{B}$, are members of $\mathcal{G}$. Thus, the σ-field $\mathcal{G}$ contains the σ-field generated by the collection of these open subsets—namely $\mathcal{C}$.

To show that $\mathcal{G} \subseteq \mathcal{C}$ we introduce the Borel σ-field $\mathcal{D}$ of subsets of $(-\infty, 0)$ with the relative topology and set

$$\mathcal{H} = \{C \cup D \colon C \in \mathcal{C},\ D \in \mathcal{D}\}\,.$$

We can finish the proof by showing that $\mathcal{B} \subseteq \mathcal{H}$, because $\mathcal{C}$ consists of those members of $\mathcal{H}$ which are subsets of $\mathbb{R}^+$. It is clear that $\mathcal{H}$ is closed under countable unions. The formula

$$(C \cup D)^c = (\mathbb{R}^+ \backslash C) \,\cup\, ((-\infty, 0) \backslash D)$$

for $C \subseteq \mathbb{R}^+$ and $D \subseteq (-\infty, 0)$ shows that it is closed under complementation. So $\mathcal{H}$ is a σ-field. For any open set $O \in \mathbb{R}$, the representation

$$O = (\mathbb{R}^+ \cap O) \,\cup\, ((-\infty, 0) \cap O)$$

represents O as the union of open, and therefore Borel, subsets of the spaces $\mathbb{R}^+$ and $(-\infty, 0)$. Thus, the σ-field $\mathcal{H}$ contains the σ-field generated by the collection of open subsets of $\mathbb{R}$ —namely $\mathcal{B}$.

1-16 *Hint:* It suffices to show that every open set is the union of open boxes having edges of rational length and centers with rational coordinates.

For Chapter 2

2-2. Let X be a continuous function. For any open B of the target of X, $X^{-1}(B)$ is open by continuity, and thus is an event in the domain of X. Now apply Proposition 3 with $\mathcal{E}$ equal to the collection of open subsets in the target of X.

2-3. Let B be an arbitrary measurable set in the common target of X and Y. We need to show that

$$P(\{\omega \colon X(\omega) \in B\}) = P(\{\omega \colon Y(\omega) \in B\})\,.$$

Here is the relevant calculation:

$$\begin{aligned}
&P(\{\omega \colon X(\omega) \in B\})\\
&= P(\{\omega \colon X(\omega) \in B,\, Y(\omega) = X(\omega)\}) + P(\{\omega \colon X(\omega) \in B,\, Y(\omega) \neq X(\omega)\})\\
&= P(\{\omega \colon X(\omega) \text{ and } Y(\omega) \in B,\, Y(\omega) = X(\omega)\})\,.
\end{aligned}$$

In this calculation, we used the fact that the event in the second term of the second line is contained in a null event. To complete the proof, carry out a similar calculation with the roles of X and Y reversed.

2-9. By Problem 13 of Chapter 1 and Proposition 3 we only need show that the set $A = \{\omega : X(\omega) \leq c\}$ is a Borel set for every c (or even just for every rational c). Let a equal the least upper bound of A. We will prove that every member of the interval $(0, a)$ belongs to A. Suppose $\omega_1 < a$. Since a is the least upper bound of A, there exists $\omega_2 \in A$ for which $\omega_1 < \omega_2$. Then

$$X(\omega_1) \leq X(\omega_2) \leq c\,,$$

from which it follows that $\omega_1 \in A$. Thus, A is an interval of the form $(0, a)$ or $(0, a]$ and is, therefore, Borel.

2-12. $\frac{5}{9}$

2-14. The distribution is uniform on the triangle $\{(v_1, v_2) : 0 < v_1 < v_2 < 1\}$. If B is a set for which area is defined, the value that the distribution assigns to B is twice its area, the factor of 2 arising because the triangle has area $\frac{1}{2}$. To prove that X is a random variable—*Hint:* Prove that X is continuous, or, alternatively, avoid the issue of continuity of a $\mathbb{R}^2$-valued function by first doing Problem 16 and then using it in conjunction with a proof that each coordinate function is continuous.

2-19. In case k is divisible by 4, the answer is

$$\binom{k/2}{k/4} 2^{-k}\,.$$

Otherwise, the answer is 0.

2-21. The Hausdorff distances are $\frac{1+\sqrt{2}}{2}$ between the first two; $\frac{1}{2}$ between the first and third; $\frac{2+\sqrt{2}}{2}$ between the second and third.

2-22. These are the probabilities: $\frac{2-\sqrt{2}}{2\pi}$, $\frac{1}{16}$, $\frac{\pi-2}{8\pi}$.

For Chapter 3

3-3. Fix ω. Since F is increasing, every member of $\{x\colon F(x) < \omega\}$ is less than every member of $\{x\colon F(x) \geq \omega\}$ and is thus a lower bound of $\{x\colon F(x) \geq \omega\}$. Hence $Y(\omega) \stackrel{\text{def}}{=} \sup\{x\colon F(x) < \omega\}$ is a lower bound of $\{x\colon F(x) \geq \omega\}$. Therefore $Y(\omega) \leq X(\omega)$.

To prove $Y(\omega) = X(\omega)$, suppose, for a proof by contradiction, that $Y(\omega) < X(\omega)$, and consider an $x \in (Y(\omega), X(\omega))$. Either $F(x) \geq \omega$ contradicting the defining property of $X(\omega)$ or $F(x) < \omega$ contradicting the defining property of $Y(\omega)$. Thus $Y = X$, and we will work with Y in the next paragraph.

Clearly, Y is increasing. Thus, to show left continuity we only need show $Y(\omega-) \geq Y(\omega)$ for every ω. Let $\delta > 0$. There exists $u > Y(\omega) - \delta$ for which $F(u) < \omega$. Hence there exists $\tau < \omega$ such that $F(u) < \tau$. Therefore

$$Y(\omega-) \geq Y(\tau) \geq u > Y(\omega) - \delta\,.$$

Now let $\delta \searrow 0$.

3-8. Whether a or b is finite or infinite,

$$Q((a,b)) = \int_a^b \frac{1}{\pi(1+x^2)}\,dx\,.$$

When a and b are finite this formula is also a formula for $Q([a,b])$, and similarly for $Q([a,b))$ and $Q((a,b])$ in case $a > -\infty$ or $b < \infty$, respectively. Note that the formula for $Q([a,b])$ is correct in the special case $a = b$.

3-12. Explanation for 'type' only. Suppose first that F_1 and F_2 are of the same type. Then there exist random variables X_1 and X_2 of the same type such that F_j is the distribution function of X_j. Then F_2 is also the distribution function of $aX_1 + b$ for some a and b with $a > 0$. Thus

$$F_2(x) = P(\{\omega : aX_1(\omega) + b \le x\}) = P(\{\omega : X_1(\omega) \le (x-b)/a\}) = F_1((x-b)/a)\,.$$

That is F_1 and F_2 must satisfy (3.2).

Conversely, suppose that F_2 and F_1 satisfy (3.2) for some a and b with $a > 0$. Let X_1 be a random variable with distribution function F_1. the above calculation then shows that $aX_1 + b$ is a random variable whose distribution function is F_2. Therefore F_2 is of the same type as F_1.

3-23. X is symmetric about b if and only if its distribution function F satisfies $F(x - b) = 1 - F((b-x)-)$ for all x.

For the standard Cauchy distribution

$$\begin{aligned} F(x) &= \frac{1}{2} + \frac{\arctan x}{\pi} = \frac{1}{2} + \frac{-\arctan(-x)}{\pi} \\ &= 1 - \Big(\frac{1}{2} + \frac{\arctan(-x)}{\pi}\Big) = 1 - \Big(\frac{1}{2} + \frac{\arctan((-x)-)}{\pi}\Big) = 1 - F((-x)-)\,. \end{aligned}$$

3-28. A random variable X having the Cauchy distribution of Problem 8 has density $x \rightsquigarrow \frac{1}{\pi(1+x^2)}$. For positive a and real b the continuous density of $aX + b$ is $x \rightsquigarrow \frac{a}{\pi(a^2+(x-b)^2)}$.

The density of the uniform distribution with support $[a,b]$ is $\frac{1}{b-a}$ on the interval $[a,b]$ and 0 elsewhere.

3-30.

$$\begin{aligned} \int_0^\infty ae^{-ax}\,dx = -e^{-ax}\Big|_0^\infty &= 1 \\ P(\{\omega\colon 2 \le X(\omega) \le 3\}) &= e^{-2a} - e^{-3a} \\ \text{median} &= a^{-1}\log 2 \end{aligned}$$

3-33. $g(x) = \frac{1}{2\sqrt{x}}[f(\sqrt{x}) + f(-\sqrt{x})]$ if $x > 0$ and $g(x) = 0$ if $x \le 0$.

3-34. (i)

$$\Gamma(\gamma+1) = \int_0^\infty u^\gamma e^{-u}\,du = -\int_0^\infty u^\gamma\,de^{-u}$$
$$= \int_0^\infty \gamma u^{\gamma-1}e^{-u}\,du = \gamma\Gamma(\gamma)\,.$$

(ii) An easy calculation gives $\Gamma(1) = 0!$. For an induction proof assume that $\Gamma(\gamma) = (\gamma-1)!$ for some positive integer γ. By part (i),

$$\Gamma(\gamma+1) = \gamma\Gamma(\gamma) = \gamma[(\gamma-1)!] = \gamma!\,.$$

Note that the last step in the above calculation is valid for $\gamma = 1$. That this step be valid is one of the motivations for the definition $0! = 1$.

(iii)

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty u^{-1/2}e^{-u}\,du = \int_0^\infty \sqrt{2}\,e^{-v^2/2}\,dv\,,$$

which, by Example 1 and symmetry, equals $\sqrt{\pi}$. Now use mathematical induction.

(iv)

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^\infty\int_0^\infty u^{\alpha-1}v^{\beta-1}e^{-(u+v)}\,du\,dv$$
$$= \int_0^\infty\int_v^\infty (w-v)^{\alpha-1}v^{\beta-1}e^{-w}\,dw\,dv$$
$$= \int_0^\infty\int_0^w (w-v)^{\alpha-1}v^{\beta-1}e^{-w}\,dv\,dw$$
$$= \int_0^\infty\int_0^1 w^{\alpha+\beta-1}(1-x)^{\alpha-1}x^{\beta-1}e^{-w}\,dx\,dw\,,$$

the interchange of order of integration being valid, according to a result from advanced calculus, because the integrand is continuous and nonnegative. (The validity of the interchange in integration order is also a consequence of the Fubini Theorem, to be proved in Chapter 9.) The last expression is the product of the two desired integrals.

3-40. *Hint:* For $b = 1$ and $x \geq 0$,

$$P(\{\omega\colon -\log X(\omega) \leq x\}) = P(\{\omega\colon X(\omega) \geq e^{-x}) = 1 - e^{-x}\,,$$

the standard exponential distribution function.

3-41. Denote the three distribution functions by G_i, $i = 2, 3, 4$. For each i, $G_i(y) = 0$ when $y < 0$ and $= 1$ when $y \geq 2$. For $0 \leq y < 2$:

$$G_2(y) = 1 - \sqrt{1 - \tfrac{y^2}{4}}\,;$$
$$G_3(y) = \tfrac{2}{\pi}\arcsin\tfrac{y}{2}\,;$$
$$G_4(y) = \tfrac{y^2}{4}\,.$$

For Chapter 4

4-7. $n^2p^2 + npq$ (notation of Problem 39 of Chapter 3)

4-8. 7/2

4-9. For this problem, denote the expectation operator according to Definition 1 by E_s and the expectation operator according to Definition 5 by E_p. Let X be nonnegative and simple. Thus, $E_s(X)$ and $E_p(X)$ are meaningful. Since X qualifies as an appropriate Z in the definition $E_p(X) = \sup_Z E_s(Z)$, we see that $E_p(X) \geq E_s(X)$. On the other hand, Lemma 4 implies that for all simple $Z \leq X$, $E_s(Z) \leq E_s(X)$, from which it follows immediately that $E_p(X) \leq E_s(X)$.

4-10. The random variable X defined by $X(\omega) = \frac{1}{\omega}$, defined on the probability space $((0,1], \mathcal{B}, P)$, where P denotes Lebesgue measure, has expected value ∞. This is seen by calculating $E(X_n)$ for simple random variables $X_n \leq X$ defined by $X_n(\omega) = (\lfloor X(\omega) \rfloor) \wedge n$.

4-11. We treat the case $a = b = 1$. The following calculation based on the definition of expectation for nonnegative random variables and the linearity of the expectation for simple random variables shows that $E(X) + E(Y) \leq E(X+Y)$:

$$\begin{aligned}
&E(X) + E(Y) \\
&= \sup\{E(X')\colon X' \leq X \text{ and } X' \text{ simple}\} + \sup\{E(Y')\colon Y' \leq Y \text{ and } Y' \text{ simple}\} \\
&= \sup\{E(X') + E(Y')\colon X' \leq X, Y' \leq Y \text{ and } X', Y' \text{ simple}\} \\
&= \sup\{E(X' + Y')\colon X' \leq X, Y' \leq Y \text{ and } X', Y' \text{ simple}\} \\
&\leq \sup\{E(Z)\colon Z \leq X + Y \text{ and } Z \text{ simple}\} = E(X+Y)\,.
\end{aligned}$$

To prove the opposite inequality, let Z be a simple random variable such that $Z \leq X + Y$. By the construction given in the proof of Lemma 13 of Chapter 2, we can find sequences $(X_n\colon n = 1, 2, \dots)$ and $(Y_n\colon n = 1, 2, \dots)$ of simple random variables such that for all ω and all n,

$$\begin{aligned}
X(\omega) \wedge n - \frac{1}{2^n} &\leq X_n(\omega) \leq X(\omega) \quad \text{and} \\
Y(\omega) \wedge n - \frac{1}{2^n} &\leq Y_n(\omega) \leq Y(\omega)\,.
\end{aligned}$$

It is easily checked that $X_n + Y_n \geq Z - 1/2^n$ for $n \geq \max\{Z(\omega)\colon \omega \in \Omega\}$. Thus

$$\sup_n E(X_n) + \sup_n E(Y_n) \geq E(Z)\,,$$

and the desired inequality $E(X) + E(Y) \geq E(X+Y)$ now follows from the definition of expected value.

4-14. For this problem, denote the expectation operators according to Definition 1, Definition 5, and Definition 8 by E_s, E_p, and E_g, respectively. Let X be simple (but not necessarily nonnegative). We use (4.1):

$$X = \sum_{j=1}^{n} c_j I_{C_j}\,.$$

Since $\{C_j : 1 \le j \le n\}$ is a partition,

$$X^+ = \sum_{j:c_j \ge 0} c_j I_{C_j}$$

and

$$X^- = -\sum_{j:c_j<0} c_j I_{C_j}\,.$$

For these nonnegative simple random variables we have, using Problem 9, that

$$E_p(X^+) = E_s(X^+) = \sum_{j:c_j \ge 0} c_j P(C_j)$$

and

$$E_p(X^-) = E_s(X^-) = -\sum_{j:c_j<0} c_j P(C_j)\,.$$

By these formulas and Definition 8,

$$E_g(X) = E_p(X^+) - E_p(X^-) = \sum_{j=1}^{n} c_j P(C_j) = E_s(X)\,.$$

4-21. The case where $E(X_1) = +\infty$ is easily treated, so we assume $E(X_1)$ is finite and, therefore, $P(\{\omega : |X_1(\omega)| = \infty\}) = 0$. Accordingly, except for ω belonging to some null set, we may define $Y_n(\omega) = X_n(\omega) - X_1(\omega)$ and $Y(\omega) = X(\omega) - X_1(\omega)$. For ω in the null set we set $Y_n(\omega) = Y(\omega) = 0$. Applying the Monotone Convergence Theorem to the sequence $(Y_1, Y_2, \dots)$, we deduce that $E(Y_n) \to E(Y)$ as $n \to \infty$. It follows, by property (iii) of Theorem 9, that

$$\lim_{n\to\infty} E(X_n - X_1) \to E(X - X_1)\,.$$

Since $E(X_1)$ is finite we may apply property (i) of Theorem 9 to conclude

$$\lim_{n\to\infty} [E(X_n) - E(X_1)] \to E(X) - E(X_1)\,.$$

Now add $E(X_1)$ to both sides.

4-22. $E(X) = \frac{p}{1-p}$, $E(X^2) = \frac{p(1+p)}{(1-p)^2}$ (notation of Problem 11 of Chapter 3)

4-23 $E(X) = \lambda$, $E(X^2) = \lambda + \lambda^2$ (notation of Problem 37 of Chapter 3)

4-26 The distributions of $X - b$ and $b - X$ are identical. By Theorem 15 they have the same mean. By properties (i) and (ii) of Theorem 9, these equal numbers are $E(X) - b$ and $b - E(X)$. It follows that $E(X) = b$.

4-29. b (notation of Example 1 of Chapter 3)

4-30. For standard beta distributions (that is, beta distributions with support $[0, 1]$), the answer is $\frac{\alpha}{\alpha+\beta}$ (notation of Example 3 of Chapter 3).

4-31. $E(X) = 1/k$, $E(\exp \circ X) = \infty$ if $k \le 1$ and $= \frac{k}{k-1}$ if $k > 1$

4-35. $E(X_1) = E(X_3) = \frac{4}{\pi}$, $E(X_2) = \frac{\pi}{2}$, $E(X_4) = \frac{4}{3}$

For Chapter 5

5-7. $\operatorname{Var}(X_1) = \operatorname{Var}(X_3) = \frac{2(\pi^2-8)}{\pi^2}$, $\operatorname{Var}(X_2) = \frac{32-3\pi^2}{12}$, $\operatorname{Var}(X_4) = \frac{2}{9}$

5-13. $\operatorname{Cov}(aX+b, cY+d) = ac\operatorname{Cov}(X,Y)$

5-14. An inequality, based on the fact that φ is increasing will be useful:

$$[X - E(X)]\,(\varphi \circ X) \geq [X - E(X)]\,\varphi(E(X))\,.$$

The following calculation then completes the proof:

$$\begin{aligned}
\operatorname{Cov}(X,\, \varphi \circ X) &= E\big([X - E(X)][\varphi \circ X - E(\varphi \circ X)]\big) \\
&= E\big([X - E(X)]\,(\varphi \circ X)\big) \\
&\geq E\big([X - E(X)]\,\varphi(E(X))\big) = 0\,.
\end{aligned}$$

(A slightly longer, but possibly more transparent proof, consists of first reducing the problem to the case where $E(X) = 0$ and then using the above argument for that special case.)

5-17. For Example 2 of Chapter 4, the answer is 1; for Problem 18 of Chapter 4, the answer is 0 or 1 according as n is odd or even.

5-29. $s(0,0) = s(1,1) = s(2,2) = s(3,3) = 1$, $s(1,0) = s(2,0) = s(3,0) = 0$, $s(2,1) = -1$, $s(3,1) = 2$, $s(3,2) = -3$, $s(n,k) = 0$ for $k > n$; $S(0,0) = S(1,1) = S(2,2) = S(3,3) = 1$, $S(1,0) = S(2,0) = S(3,0) = 0$, $S(2,1) = S(3,1) = 1$, $S(3,2) = 3$, $S(n,k) = 0$ for $k > n$

5-32. $\rho(1-) = 1$. Thus, if ρ is the probability generating function of a distribution Q, then $Q(\{\infty\}) = 0$. To both show that ρ is a probability generating function and calculate $Q(\{k\})$ for each $k \in \mathbb{Z}^+$ we rewrite $\rho(s)$ using partial fractions:

$$\begin{aligned}
\rho(s) &= \frac{-24}{2-s} + \frac{8}{(2-s)^2} + \frac{24}{3-s} + \frac{16}{(3-s)^2} + \frac{8}{(3-s)^3} \\
&= \frac{-12}{1-(s/2)} + \frac{2}{(1-(s/2))^2} + \frac{8}{1-(s/3)} + \frac{16/9}{(1-(s/3))^2} + \frac{8/27}{(1-(s/3))^3}\,.
\end{aligned}$$

The first two of the last five functions are equal to their power series for $|s| < 2$ and the last three for $|s| < 3$. So we can expand in power series and collect coefficients to get a power series for $\rho(s)$ that can be differentiated term-by-term to obtain the derivatives of $\rho(s)$. Thus, we only need to show that the coefficients are nonnegative in order to conclude that $\rho(s)$ is a probability generating function, and then the coefficients are the values $Q(\{k\})$.

Formulas for the geometric series and its derivatives give

$$\begin{aligned}
\rho(s) = -12\sum_{k=0}^{\infty}(\tfrac{s}{2})^k + 2\sum_{k=0}^{\infty}(k+1)(\tfrac{s}{2})^k + 8\sum_{k=0}^{\infty}(\tfrac{s}{3})^k \\
+ \tfrac{16}{9}\sum_{k=0}^{\infty}(k+1)(\tfrac{s}{3})^k + \tfrac{4}{27}\sum_{k=0}^{\infty}(k+1)(k+2)(\tfrac{s}{3})^k\,.
\end{aligned}$$

When we collect terms we get nonnegative—in fact, positive—terms, as desired:

$$Q(\{k\}) = \frac{k-5}{2^{k-1}} + \frac{4k^2+60k+272}{3^{k+3}}\,.$$

To get the mean and variance it seems best to work with $\rho(s)$ in the form originally given and use the product rule to get the first and second derivatives:

$$\rho'(s) = \frac{16}{(2-s)^3(3-s)^3} + \frac{24}{(2-s)^2(3-s)^4}$$

and

$$\rho''(s) = \frac{48}{(2-s)^4(3-s)^3} + \frac{96}{(2-s)^3(3-s)^4} + \frac{96}{(2-s)^2(3-s)^5}\,.$$

Insertion of 1 for s gives

$$\rho'(1) = \frac{7}{2} \quad \text{and} \quad \rho''(1) = 15\,.$$

Hence, the mean equals $\frac{7}{2}$ and the second moment equals $15 + \frac{7}{2} = \frac{37}{2}$. Therefore, the variance equals $\frac{74}{4} - \frac{49}{4} = \frac{25}{4}$ and the standard deviation equals $\frac{5}{2}$.

Had the problem only been to verify that ρ is a probability generating function, we could have, while calculating the first and second derivatives, seen that a straightforward induction proof would show that all derivatives are positive, and an appeal to Theorem 14 would complete the proof.

5-33. The mean is ∞ and thus the variance is undefined. The distribution Q_p corresponding to the probability generating function with parameter p satisfies $Q_p(\{\infty\}) = |1-2p|$. Also, for $0 < k = 2m < \infty$,

$$Q_p(\{2m\}) = \frac{2}{m}\binom{2m-2}{m-1}[p(1-p)]^m\,.$$

For k odd and $k=0$, $Q_p(\{k\}) = 0$.

For Chapter 6

6-6. Method 1: Using Problem 4, we get

$$\begin{aligned}(\liminf_{n\to\infty} A_n)^c &= \left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m\right)^c \\ &= \bigcap_{n=1}^{\infty}\left(\bigcap_{m=n}^{\infty} A_m\right)^c \\ &= \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m^c = \limsup_{n\to\infty} A_n^c\,.\end{aligned}$$

Method 2: We prove that the indicator functions of the two sets are equal:

$$\begin{aligned}I_{(\limsup A_n)^c} &= 1 - I_{\limsup A_n} = 1 - \limsup_n\{I_{A_n}\} \\ &= \liminf_n\{(1-I_{A_n})\} = \liminf_n\{I_{A_n^c}\} = I_{\liminf A_n^c}\,.\end{aligned}$$

6-8.

$$\begin{aligned}
\limsup_{n\to\infty}(A_n \cup B_n) &= (\limsup_{n\to\infty} A_n) \cup (\limsup_{n\to\infty} B_n)\,;\\
\liminf_{n\to\infty}(A_n \cap B_n) &= (\liminf_{n\to\infty} A_n) \cap (\liminf_{n\to\infty} B_n)\,;\\
\liminf_{n\to\infty}(A_n \setminus B_n) &= (\liminf_{n\to\infty} A_n) \setminus (\limsup_{n\to\infty} B_n)\,;\\
(\limsup_{n\to\infty} A_n) \cup (\liminf_{n\to\infty} B_n) \supseteq \liminf_{n\to\infty}(A_n \cup B_n) &\supseteq (\liminf_{n\to\infty} A_n) \cup (\liminf_{n\to\infty} B_n)\,;\\
(\liminf_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} B_n) \subseteq \limsup_{n\to\infty}(A_n \cap B_n) &\subseteq (\limsup_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} B_n)\,;\\
(\limsup_{n\to\infty} A_n) \setminus (\limsup_{n\to\infty} B_n) \subseteq \limsup_{n\to\infty}(A_n \setminus B_n) &\subseteq (\limsup_{n\to\infty} A_n) \setminus (\liminf_{n\to\infty} B_n)\,;\\
\liminf_{n\to\infty}(A_n \triangle B_n) &\subseteq (\limsup_{n\to\infty} A_n) \triangle (\liminf_{n\to\infty} B_n)\,;\\
\limsup_{n\to\infty}(A_n \triangle B_n) &\supseteq (\limsup_{n\to\infty} A_n) \triangle (\limsup_{n\to\infty} B_n)\,.
\end{aligned}$$

Problem 6 is relevant for this problem, especially the fifth equality given in the problem.

Here are some examples in which the various subset relations given above are strict. The first and seventh subset relations above are both strict in case $B_n = \emptyset \subset \Omega$ for all n and $A_n = \emptyset$ or $= \Omega$ according as n is odd or even. The second, fourth, fifth, and eighth subset relations are all strict if $A_n = B_n^c$ for all n and $A_n = \emptyset \subset \Omega$ or $= \Omega$ according as n is odd or even. The third and sixth subset relations are both strict if $A_n = B_n$ for all n and $A_n = \emptyset \subset \Omega$ or $= \Omega$ according as n is odd or even.

6-9. The middle inequality is obvious. Using the Continuity of Measure Theorem in Chapter 6, we have

$$\begin{aligned}
P(\limsup_{n\to\infty} A_n) &= P\Big(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\Big)\\
&= \lim_{n\to\infty} P\Big(\bigcup_{m=n}^{\infty} A_m\Big) \geq \limsup_{n\to\infty} P(A_n)\,,
\end{aligned}$$

thus establishing the first inequality. For the third inequality, deduce from the first inequality that $P(\limsup A_n^c) \geq \limsup P(A_n^c)$, which is equivalent to

$$1 - P\big((\limsup A_n^c)^c\big) \geq \limsup[1 - P(A_n)]\,,$$

which itself is equivalent to

$$P\big((\limsup A_n^c)^c\big) \leq \liminf P(A_n)\,.$$

By Problem 6, the event in the left side equals $\liminf A_n$, as desired.

6-13. Let $A_n = \{\omega\colon X_n(\omega) = 1\}$. By Problem 5, the event $A = \limsup_n A_n$ is that event that $\sum_n X_n = \infty$. The events A_n are pairwise negatively correlated or uncorrelated, so by the Borel-Cantelli Lemma, $P(A) = 1$ if $\sum_n P(A_n) = \infty$, and by the Borel Lemma, $P(A) = 0$ if $\sum_n P(A_n) < \infty$. The proof is now completed by noting that $P(A_n) = E(X_n)$, so that $\sum_n P(A_n) = E(\sum_n X_n)$, whether finite or infinite.

6-15. Let n denote the number of cards and C_m, for $m = 1, 2, \dots, n$, the event that card m is in position m. The i^{th} term of the formula for $P(\bigcup C_m)$ in Theorem 6 consists of the factor $(-1)^{i+1}$ and $\binom{n}{i}$ terms each of which equals the probability that each of a particular i cards are in a particular i positions. This probability equals the number of ways of placing the remaining $n-i$ cards in the remaining $n-i$ positions, divided by $n!$. We conclude that

$$P\Big(\bigcup_{m=1}^{n} C_m\Big) = \sum_{i=1}^{n}(-1)^{i+1}\binom{n}{i}\frac{(n-i)!}{n!} = \sum_{i=1}^{n}(-1)^{i+1}\frac{1}{i!}\,,$$

which approaches $1-e^{-1}$ as $n \to \infty$.

For Chapter 7

7-3. Let $\mathcal{D} = \{A\colon P(A) = Q(A)\}$. Suppose that $A \subseteq B$ are both members of $\mathcal{D}$. Then

$$P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)\,.$$

Thus, $\mathcal{D}$ is closed under proper differences. Now consider an increasing sequence $(A_1, A_2, \dots)$ of members of $\mathcal{D}$. By the Continuity of Measure Theorem, applied to both P and Q,

$$P(\lim A_n) = \lim P(A_n) = \lim Q(A_n) = Q(\lim A_n)\,.$$

Hence $\mathcal{D}$ is closed under limits of increasing sequences, and therefore $\mathcal{D}$ is a Sierpiński class. It contains $\mathcal{E}$ and so, by the Sierpiński Class Theorem it contains $\sigma(\mathcal{E})$, as desired.

7-10. The sequences $(A_n\colon n = 1, 2, \dots)$ and $(A, A, A, \dots)$ have the common limit A. By the lemma, the sequences $(R(A_n)\colon n = 1, 2, \dots)$ and $(R(A), R(A), R(A), \dots)$ have equal limits. The limit of the second of these numerical sequences is obviously $R(A)$, so $R(A)$ is also the limit of the first sequence of numbers.

7-11. Every member A of $\mathcal{E}$ is the limit of the sequence $(A, A, A, \dots)$. Thus $\mathcal{E} \subseteq \mathcal{E}_1$. It remains to prove that $\mathcal{E}_1$ is a field.

The empty set, being a member of $\mathcal{E}$, is also a member of $\mathcal{E}_1$. Let $B \in \mathcal{E}_1$. Then there exists a sequence $(B_n \in \mathcal{E}\colon n = 1, 2, \dots)$ that converges to B. By Problem 8 of Chapter 6, $B_n^c \to B^c$ as $n \to \infty$. Since $\mathcal{E}$ is a field, each B_n^c is a member of $\mathcal{E}$. Therefore $B^c \in \mathcal{E}_1$.

Let B and B_n be as in the preceding paragraph and let $C \in \mathcal{E}$. There exists a sequence $(C_n \in \mathcal{E}\colon n = 1, 2, \dots)$ that converges to C. By Problem 8 of Chapter 6, $B_n \cup C_n \to B \cup C$ as $n \to \infty$. Since $\mathcal{E}$ is a field, $B_n \cup C_n \in \mathcal{E}$ for each n. Therefore $B \cup C \in \mathcal{E}_1$.

7-17. The probability is $1 - \prod_{k=2}^{4}(1 - k^{-\beta})$. The correlation between two events A_m and A_n is easily calculated; it is 0 when $n \neq m$. Similarly, for A_m^c and A_n^c. Thus, the Borel-Cantelli Lemma may be used to calculate the probabilities of the limit supremum and limit infimum.

$$\begin{aligned}
&P(\liminf A_n) = 1 - P(\limsup A_n^c) = 0\\
&P(\limsup A_n) = 1 \text{ if } \beta \le 1 \text{ and } = 0 \text{ if } \beta > 1\\
&P(\{\omega\colon Y(\omega) = 1\}) = 1\\
&P(\{\omega\colon Z(\omega) = n\}) = n^{-\beta} \prod_{k=n+1}^{\infty} (1 - k^{-\beta}) \text{ if } n < \infty\\
&\qquad\qquad \text{and } = 1 \text{ or } = 0 \text{ according as } \beta \le 1 \text{ or } \beta > 1 \text{ if } n = \infty
\end{aligned}$$

7-24. Since $\mathcal{L}$ is in one-to-one measure-preserving correspondence with $\mathcal{S} \subset \mathbb{R}^2$, we only need show that the effect of a rotation or translation on $\mathcal{L}$ corresponds to a transformation on $\mathbb{R}^2$ having Jacobian 1, provided we identify φ with $\varphi + 2\pi$. It is clear that rotations about the origin have this property, leaving s unchanged and adding a constant to φ. Translations also have this property since they leave φ unchanged and add $-r\cos(\varphi - \theta)$ to s, where (r, θ) is the polar representation of the point to which the origin is translated.

7-25 The measure of the set of lines intersecting a line segment is twice the length of that line segment.

7-26 The measure of the set of lines intersecting a convex polygon is the perimeter of that polygon.

7-29. The expected value, whether finite or infinite, is twice the length of D divided by $2\pi r$. (It can be shown that this value is correct for arbitrary curves D contained in the interior of the circle.)

For Chapter 8

8-8. Application of the Fatou Lemma to the sequence $(g - f_n\colon n \ge 1)$ of nonnegative measurable functions gives

$$\liminf \int (g - f_n)\, d\mu \ge \int \liminf (g - f_n)\, d\mu = \int (g - \limsup f_n)\, d\mu \ge 0\,.$$

Since $\int g d\mu < \infty$, we may use linearity to obtain

$$\int g d\mu - \limsup \int f_n\, d\mu \ge \int g\, d\mu - \int \limsup f_n\, d\mu \ge 0\,.$$

Subtraction of $\int g\, d\mu$ followed by multiplication by -1 gives the last two inequalities in (8.2). The first two inequalities in (8.2) can be obtained in a similar manner using $g + f_n$, and the middle inequality in (8.2) is obvious.

Under the additional hypothesis that $\lim f_n = f$, the first and last finite quantities in (8.2) are equal, and therefore all four finite quantities are equal. Thus $\int |f|\, d\mu < \infty$ and $\int f_n\, d\mu \to \int f\, d\mu$. Applying what we have already proved to the sequence $(|f - f_n|\colon n \ge 1)$, each member of which is bounded by $2g$, we obtain

$$\lim \int |f - f_n|\, d\mu = \int \big(\lim |f - f_n|\big)\, d\mu = \int 0\, d\mu = 0\,.$$

8-12. Let $I_{t,c}$ denote the indicator function of $\{\omega\colon |X_t(\omega)| \geq c\}$.

$$E(|X_t|I_{t,c}) = E(|X_t|^{1-p} I_{t,c}|X_t|^p) \leq c^{1-p}E(|X_t|^p) \leq c^{1-p}k \to 0 \ \text{ as } c \to \infty\,.$$

8-22. By Theorem 14 the assertion to be proved can be stated as:

$$\lim_{\gamma\to\infty} \int \theta_\gamma \, d\lambda = \int \theta \, d\lambda\,,$$

where λ denotes Lebesgue measure on $\mathbb{R}$ and

$$\theta(v) = \begin{cases} e^{-v^2/2} & \text{if } v \geq 0 \\ 0 & \text{otherwise}\,. \end{cases}$$

The plan is to use the Dominated Convergence Theorem. Thus we may restrict our attention to $v \geq 0$ throughout.

We take logarithms of the integrands:

$$(\log\circ\theta_\gamma)(v) = (\gamma-1)\log(1+v\gamma^{-1/2}) - v\gamma^{1/2}\,.$$

The Taylor Formula with remainder (or an argument based on the Mean-Value Theorem) shows that $(\log\circ\theta_\gamma)(v)$ lies between

$$(\gamma-1)(v\gamma^{-1/2} - \tfrac{1}{2}v^2\gamma^{-1}) - v\gamma^{1/2}$$

and

$$(\gamma-1)(v\gamma^{-1/2} - \tfrac{1}{2}v^2\gamma^{-1} + \tfrac{1}{3}v^3\gamma^{-3/2}) - v\gamma^{1/2}\,,$$

both of which approach $-v^2/2$ as $\gamma \to \infty$. Thus, to complete the proof we only need find a dominating function having finite integral.

The integrands θ_γ are nonnegative. It is enough to show, for $\gamma \geq 1$, that $\theta_\gamma(x) \leq (1+v)e^{-v}$, since this last function of v has finite integral on $[0,\infty)$. Clearly, $\theta_\gamma(v) \leq (1+v\gamma^{-1/2})\theta_\gamma(v)$, the logarithm of which equals

$$\gamma\log(1+v\gamma^{-1/2}) - v\gamma^{1/2}\,. \tag{7.1}$$

Differentiation with respect to γ and writing x for $v\gamma^{-1/2}$ gives

$$\log(1+x) - \frac{x(2+x)}{2(1+x)}\,, \tag{7.2}$$

a function which equals 0 when $x = 0$ and is, by Problem 21, a decreasing function of x. Thus, (7.2) is nonpositive when $x \geq 0$. For $\gamma \geq 1$ [which we may assume without loss of generality], (7.1) is no larger than the value $\log(1+v) - v$ it attains when $\gamma = 1$. The exponential of this value is the desired function $(1+v)e^{-v}$. [Comment: The introduction of the factor $(1+v\gamma^{-1/2})$ in the sentence containing (7.1) was for the purpose of obtaining a decreasing function of γ.]

8-26. *Hint:* The absolute value of the integral is bounded by

$$2\sqrt[3]{n^2}\,\max\Bigl|\log\bigl(1+\frac{x-n}{n}\bigr)\Bigr|\,\max\bigl(x^n e^{-x}\bigr)\,,$$

where each maximum is over those x for which $|x-n| \leq \sqrt[3]{n^2}$. Apply the Mean-Value Theorem to the logarithmic function, standard methods of differential calculus to the function $x \rightsquigarrow x^n e^{-x}$, and the Stirling Formula to $n!$. (Note: If one works with the product of the maximum of the function $x \rightsquigarrow x^n$ and the maximum of the function $x \rightsquigarrow e^{-x}$ one does not get an inequality that is sharp enough to give the desired conclusion.)

8-35. Define a σ-finite measure ν by

$$\nu(A) = \int_A f \, d\lambda \,,$$

where λ denotes Lebesgue measure on $\mathbb{R}$, so that f is the density of ν with respect to Lebesgue measure. In particular,

$$\nu((a,b]) = \int_a^b f(x) \, dx$$

for all $a < b$. By an appropriate version of the Fundamental Theorem of Calculus,

$$\mu((a,b]) = F(b) - F(a) = \int_a^b f(x) \, dx$$

for all $a < b$. Thus, μ and ν agree on intervals of the form $(a,b]$. By the Uniqueness Theorem, they are the same measure.

For Chapter 9

9-1. Ω_1 and Ω_2 each have six members, Ω has 36 members. Each of $\mathcal{F}_1$, $\mathcal{G}_1$, $\mathcal{F}_2$ and $\mathcal{G}_2$ has $2^6 = 64$ members. $\mathcal{F}$ has 2^{36} members and $\mathcal{R}$ has 64^2 members.

9-6. $x \rightsquigarrow 1 - \lim_{\varepsilon \searrow 0} \prod_n [1 - F_n(x+\varepsilon)]$ and $\prod_n F_n$. The example $F_n = I_{[(1/n),\infty)}$ shows that one may not just set $\varepsilon = 0$ in the first of the two answers.

9-7. exponential with mean $\lambda_1\lambda_2/(\lambda_1 + \lambda_2)$

9-10. Fix $B_k \in \sigma(\mathcal{E}_k)$ for $k \in K$. For each such k there are disjoint members $A_{k,i}$, $1 \le i \le r_k$, of $\mathcal{E}_k$ such that

$$B_k = \bigcup_{i=1}^{r_k} A_{k,i} \,.$$

Hence,

$$\begin{aligned} P\Big(\bigcap_{k\in K} B_k\Big) &= P\Big(\bigcap_{k\in K}\bigcup_{i=1}^{r_k} A_{k,i}\Big) = P\Big(\bigcup_{(i_k \le r_k\colon k\in K)} \bigcap_{k\in K} A_{k,i_k}\Big) \\ &= \sum_{(i_k \le r_k\colon k\in K)} P\Big(\bigcap_{k\in K} A_{k,i_k}\Big) = \sum_{(i_k \le r_k\colon k\in K)} \prod_{k\in K} P(A_{k,i_k}) \\ &= \prod_{k\in K} \sum_{i=1}^{r_k} P(A_{k,i_k}) = \prod_{k\in K} P(B_k) \,. \end{aligned}$$

(Contrast this proof with the proof of Proposition 3.)

9-14. For each event B, let

$$\mathcal{D}_B = \{D\colon P(D \cap B) = P(D)\,P(B)\} \,.$$

Clearly each $\mathcal{D}_B$ is closed under proper differences. By continuity of measure it is also closed under monotone limits and, hence, it is a Sierpiński class.

Denote the two members of L by 1 and 2. By hypothesis, $\mathcal{E}_1 \subseteq \mathcal{D}_B$ for each $B \in \mathcal{E}_2$. By the Sierpiński Class Theorem, $\sigma(\mathcal{E}_1) \subseteq \mathcal{D}_B$ for each $B \in \mathcal{E}_2$. Therefore $\mathcal{E}_2 \subseteq \mathcal{D}_A$ for each $A \in \sigma(\mathcal{E}_1)$. Another application of the Sierpiński Class Theorem gives $\sigma(\mathcal{E}_2) \subseteq \mathcal{D}_A$ for every $A \in \sigma(\mathcal{E}_1)$, which is the desired conclusion.

9-15. The criterion is that for each finite subsequence $(A_{k_1}, \dots, A_{k_n})$,

$$P(A_{k_1} \cap \dots \cap A_{k_n}) = P(A_{k_1}) \dots P(A_{k_n}) .$$

9-23. Let us first confirm the appropriateness of the hint. Because the proposition treats x and y symmetrically, we only need prove the first of the two assertions in the proposition. To do that we need to show that $\{x\colon f(x,y) \in B\} \in \mathcal{G}$ for every measurable B in the target of f and every y. Suppose that we show that the $\mathbb{R}$-valued function $x \rightsquigarrow (I_B \circ f)(x,y)$ is measurable. Then it will follow that the inverse image of $\{1\}$ of this function is measurable. Since this inverse image equals $\{x\colon f(x,y) \in B\}$, the assertion in the hint is correct.

Since f is measurable, any function of the form $I_B \circ f$, where B is a measurable subset of the target of f, is the indicator function of some measurable set $A \in \mathcal{G} \times \mathcal{H}$. Thus, our task has become that of showing that $x \rightsquigarrow I_A(x,y)$ is measurable for each such A.

Let $\mathcal{C}$ denote the collection of sets $A \subseteq \Psi \times \Theta$ such that $x \rightsquigarrow I_A(x,y)$ is measurable for each fixed y. This class $\mathcal{C}$ contains all measurable rectangles, and the class of all measurable rectangles is closed under finite intersections. Since differences and monotone limits of measurable functions are measurable, the Sierpiński Class Theorem implies that $\mathcal{C}$ contains the indicator functions of all sets in $\mathcal{G} \times \mathcal{H}$, as desired.

9-27. The independence of X and Y is equivalent to the distribution of (X, Y) being a product measure $Q_1 \times Q_2$. By the Fubini Theorem,

$$\begin{aligned} E(|XY|) &= \int \left(\int |x|\,|y|\, Q_2(dy) \right) Q_1(dx) \\ &= \int |x| E(|Y|)\, Q_2(dx) = E(|X|)\, E(|Y|) < \infty . \end{aligned}$$

Thus we may apply the Fubini Theorem again:

$$\begin{aligned} E(XY) &= \int \left(\int xy\, Q_2(dy) \right) Q_1(dx) \\ &= \int x E(Y)\, Q_2(dx) = E(X)\, E(Y) . \end{aligned}$$

9-29. *Hint:* The crux of the matter is to show that, in the presence of independence, the existence of $E(X+Y)$ implies the existence of both $E(X)$ and $E(Y)$ and, moreover, it is not the case that one of $E(X)$ and $E(Y)$ equals ∞ and the other equals $-\infty$.

9-33. $\frac{2}{3}$

9-41. Method 1: The left side divided by the right side equals

$$\frac{\int_x^\infty e^{-u^2/2\sigma^2}\,du}{\sigma^2 x^{-1} e^{-x^2/2\sigma^2}}\,.$$

Both numerator and denominator approach 0 as $x \to \infty$; so we use the l'Hospital Rule. After differentiating we multiply throughout by $e^{x^2/2\sigma^2}$. The result is that we need to calculate the limit of

$$\frac{-1}{-\sigma^2 x^{-2} - 1}\,.$$

The limit equals 1, as desired.

Method 2: Let $\delta > 0$. For $x > \sigma/\sqrt{\delta}$,

$$\begin{aligned}\frac{1}{\sqrt{2\pi\sigma^2}} \int_x^\infty & e^{-u^2/2\sigma^2}\,du \\ &< \frac{1}{\sqrt{2\pi\sigma^2}} \int_x^\infty \Big(1 + \frac{\sigma^2}{u^2}\Big) e^{-u^2/2\sigma^2}\,du \\ &\qquad < \frac{1+\delta}{\sqrt{2\pi\sigma^2}} \int_x^\infty e^{-u^2/2\sigma^2}\,du\,.\end{aligned}$$

The expression between the two inequality signs is equal to the right side of (9.12). (The motivation behind these calculations is to replace the integrand by a slightly different integrand that has a simple antiderivative. One way to discover such an integrand is to try integration by parts along a few different paths, and, then, if, for one of these paths, the new integral is small compared with the original integral, combine it with the original integral. Of course, Method 1 is simple and straightforward, but it depends on being given the asymptotic formula in advance.)

9-42. $a_n = \sqrt{2\sigma^2 \log n}$

9-45. 0

9-47. If $x_i \le v + \delta$ for every positive δ, then $x_i \le v$; hence, the infimum that one would naturally place in (9.13), where the minimum appears, is attained and, therefore, the minimum exists. As j in the right side of (9.13) is increased, the set described there becomes smaller or stays constant and, therefore, its minimum becomes larger or stays constant. So (9.14) is true. The function $v \rightsquigarrow \sharp\{i\colon x_i \le v\}$ has a jump of size $\sharp\{i\colon x_i = v\}$, possibly 0, at each v. But the size of this jump equals the number of different values for the integer j that yield this value of v for the minimum in the right side of (9.13). Thus, (9.15) is true. The image of $\chi^{(d)}$ consists of all $y \in \mathbb{R}^d$ for which $y_1 \le y_2 \le \cdots \le y_d$. For such a y the cardinality of its inverse image equals

$$\frac{d!}{\prod_{j=1}^d (d_j!)^{1/d_j}}\,,$$

where d_j denotes the number of coordinates of y which equal y_j, including y_j itself.

To prove $\chi^{(d)}$ continuous it suffices to prove that each of its coordinate functions is uniformly continuous. Let $\varepsilon > 0$. Suppose that x and w are members of $\mathbb{R}^d$ for which $|x - w| < \varepsilon$. Then

$$\{i\colon x_i \le v\} \subseteq \{i\colon w_i \le v + \varepsilon\}\,.$$

Hence

$$\sharp\{i\colon x_i \le v\} \ge j \Longrightarrow \sharp\{i\colon w_i \le v+\varepsilon\} \ge j\,.$$

Since $[\chi^{(d)}(x)]_j$ is the smallest v for which the left side is true, we have

$$\sharp\{i\colon w_i \le [\chi^{(d)}(x)]_j + \varepsilon\} \ge j\,.$$

Therefore, $[\chi^{(d)}(w)]_j \le [\chi^{(d)}(x)]_j + \varepsilon$. The roles of x and w may be interchanged to complete the proof.

9-49. The density is $d!$ on the set of points in $[0,1]^d$ whose coordinates are in increasing order, and 0 elsewhere.

9-51. For $n = 1, 2, \dots,$

$$P(\{\omega\colon N(\omega) = n\}) = \frac{n}{(n+1)!}\,.$$

Also, $E(N) = e-1$. The support of the distribution of Z is $[0,1]$ and its density there is $z \rightsquigarrow (1-z)e^{1-z}$.

9-52. 1/16

9-53. $E(X) = \infty$ if $z \le 2$; $E(X) = \frac{\zeta(z-1)}{\zeta(z)}$ if $z > 2$. $\mathrm{Var}(X) = \infty$ if $2 < z \le 3$;

$$\mathrm{Var}(X) = \frac{\zeta(z-2)\zeta(z) - [\zeta(z-1)]^2}{\zeta(z)^2} \quad \text{if } z > 3\,.$$

The probability that X is divisible by m equals $1/m^z$ which approaches $\frac{1}{m}$ as $z \searrow 1$.

9-57. The distribution of the polar angle has density

$$\theta \rightsquigarrow \frac{\Gamma(2\gamma)}{4^\gamma[\Gamma(\gamma)]^2}|\sin 2\theta|^{2\gamma-1}\,.$$

The norm is a nonnegative random variable the square of which has a gamma distribution with parameter 2γ.

For Chapter 10

10-5. normal with mean $\mu_1 + \mu_2$ and standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$

10-7. $x \rightsquigarrow (1 - |x-1|) \vee 0$

10-11. probability $\frac{1}{12}$ at each of the points $\frac{k\pi}{6}$ for $-5 \le k \le 6$

10-17. For $0 \le k \le n$,

$$\begin{aligned} &P(\{\omega\colon X_{N(\omega)}(\omega) = k \text{ and } N(\omega) = n\}) = P(\{\omega\colon X_n(\omega) = k \text{ and } N(\omega) = n\}) \\ &= P(\{\omega\colon X_n(\omega) = k\})\, P(\{\omega\colon N(\omega) = n\}) = \left[\binom{n}{k} p^k (1-p)^{n-k}\right]\left[\frac{\lambda^n e^{-\lambda}}{n!}\right]. \end{aligned}$$

We sum on n:

$$\frac{(p\lambda)^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} = \frac{(p\lambda)^k e^{-p\lambda}}{k!}\,,$$

as desired.

10-21. The distribution of a single fair-coin flip is the square convolution root. If there were a cube convolution root Q, it would, by Problem 19, be supported by $\overline{\mathbb{Z}}^+$. If $Q(\{m\})$ were positive for some positive $m \in \overline{\mathbb{Z}}^+$, then $P(\{3m\})$ would also be positive, a contradiction. Thus, it would necessarily be that Q is the delta distribution δ_0, which is certainly not a cube root of P. Therefore P has no cube root.

10-30.

$$E(Y) = \frac{1}{\gamma}(\gamma_1, \dots, \gamma_d)$$
$$\mathrm{Var}(Y_i) = \frac{\gamma_i(\gamma - \gamma_i)}{\gamma^2(\gamma+1)}$$
$$\mathrm{Cov}(Y_i Y_j) = -\frac{\gamma_i\gamma_j}{\gamma^2(\gamma+1)}\,, \quad i \neq j$$

For the calculations of the above formulas one must avoid the error of treating the Dirichlet density in (10.4) as a d-dimensional density on the d-dimensional hypercube.

Here are the details of the calculation of $E(Y_1Y_2)$ under the assumption that $d \geq 4$. We replace y_d by $1 - y_1 - \cdots - y_{d-1}$ and discard the denominator $\sqrt{d}$ in (10.4) in order to obtain a density on a $(d-1)$-dimensional hypercube. (In fact, this replacement is done so often that the result of this displacement is often called the Dirichlet density.) Implicitly assuming that all variables are positive, setting

$$D = \{(y_3, \dots, y_{d-1})\colon y_3 + \cdots + y_{d-1} \leq 1\}\,,$$

and using the abbreviation $w = 1 - (y_3 + \dots y_{d-1})$, we obtain

$$E(Y_1Y_2) = \frac{\Gamma(\gamma)}{\Gamma(\gamma_1)\,\Gamma(\gamma_2)\,\Gamma(\gamma_d)} \int_D \prod_{i=3}^{d-1} \frac{y_i^{\gamma_i - 1}}{\Gamma(\gamma_i)}$$
$$\cdot \int_0^w y_2^{\gamma_2} \int_0^{w-y_2} y_1^{\gamma_1}\,(w - y_2 - y_1)^{\gamma_d - 1}\,dy_1\,dy_2\,d(y_3, \dots, y_{d-1})\,.$$

We substitute $(w-y_2)z_1$ for y_1 and then use Problem 34 of Chapter 3 for the evaluation of the innermost integral to obtain

$$E(Y_1Y_2) = \frac{\gamma_1\Gamma(\gamma)}{\Gamma(\gamma_2)\,\Gamma(\gamma_1 + \gamma_d + 1)}$$
$$\cdot \int_D \prod_{i=3}^{d-1} \frac{y_i^{\gamma_i - 1}}{\Gamma(\gamma_i)} \int_0^w y_2^{\gamma_2}\,(w - y_2)^{\gamma_1 + \gamma_d}\,dy_2\,d(y_3, \dots, y_{d-1})\,.$$

For the evaluation of the inner integral we substitute wz_2 for y_2; we get

$$E(Y_1Y_2) = \frac{\gamma_1\gamma_2\Gamma(\gamma)}{\Gamma(\gamma_2 + \gamma_1 + \gamma_d + 2)} \int_D \prod_{i=3}^{d-1} \frac{y_i^{\gamma_i - 1}}{\Gamma(\gamma_i)} w^{\gamma_2 + \gamma_1 + \gamma_d + 1}\,d(y_3, \dots, y_{d-1})\,.$$

By rearranging the constants appropriately we have come to the position of needing to calculate the integral of a Dirichlet density with parameters $\gamma_3, \dots, \gamma_{d-1}$, and $\gamma_2 +$

$\gamma_1 + \gamma_d + 2$. Since the integral of the density of any probability distribution equals 1 we obtain

$$E(Y_1Y_2) = \frac{\gamma_1\gamma_2}{\gamma(\gamma+1)}\,.$$

Since $Y_1 + \cdots + Y_d$ is a constant its variance equals 0. On the other hand, from the formula

$$\operatorname{Var}(Y_1 + \cdots + Y_d) = \sum_{j=1}^{d}\sum_{i=1}^{d} \operatorname{Cov}(Y_iY_j)$$

we see that the variance equals the sum of the entries of the covariance matrix. So, in this case, that sum is 0. But the determinant of any square matrix whose entries sum to 0 is 0, since a zero row is obtained by subtracting all the other rows from it.

10-33. Let F denote the desired distribution function. Clearly, $F(z) = 0$ for $z \le 0$ and $F(z) = 1$ for $z \ge \frac{1}{3}$. Let $z \in (0, \frac{1}{3})$. From (10.4), $1 - F(z)$ equals $2/\sqrt{3}$ times the area of those ordered triples (z_1, z_2, z_3) satisfying $z_i > z$ for $i = 1, 2, 3$ and $z_1 + z_2 + z_3 = 1$. This is the same as twice the area of those ordered pairs (z_1, z_2) such that $z_1 > z$, $z_2 > z$, and $1 - z_1 - z_2 > z$. Thus

$$1 - F(z) = 2\int_z^{1-2z}\int_z^{1-z-z_1} dz_2\, dz_1 = 1 - 6z + 9z^2\,.$$

Therefore $F(z) = 6z - 9z^2$ for $0 < z < \frac{1}{3}$.

10-36. beta with parameters $d - 1$ and 2

10-37. The distribution has support $[0, \frac{1}{2}]$ and there the distribution function is given by

$$w \rightsquigarrow \tfrac{1}{4} + 3w^2 + 3w \log \tfrac{1}{2w}\,.$$

10-40. *Hint:* For C_1, C_2, and C_3 convex compact sets, show that

$$\{r_1x_1 + r_2x_2 + r_3x_3 \colon x_i \in C_i, r_i \ge 0, r_1 + r_2 + r_3 = 1\}$$

is convex, closed, and a subset of both $(C_1 \vee C_2) \vee C_3$ and $C_1 \vee (C_2 \vee C_3)$.

10-43. $|\sin\varphi|$, $|\cos\varphi|$, $|\sin\varphi| \vee |\cos\varphi|$

10-47. For all φ and $-1 \le w \le 1$, the distribution function is

$$w \rightsquigarrow \left(\frac{\pi + w\sqrt{1-w^2} - \arccos w}{\pi}\right)^3.$$

10-48. Let A and B be two compact convex sets. Consider two arbitrary members $a_1 + b_1$ and $a_2 + b_2$ of $A + B$, where $a_i \in A$ and $b_i \in B$. Let $\kappa \in [0, 1]$. Then

$$\kappa(a_1 + b_1) + (1-\kappa)(a_2 + b_2) = [\kappa a_1 + (1-\kappa)a_2] + [\kappa b_1 + (1-\kappa)b_2]\,,$$

which, in view of the fact that A and B are convex, is the sum of a member $\kappa a_1 + (1-\kappa)a_2$ of A and a member $\kappa b_1 + (1-\kappa)b_2$ of B, and thus is itself a member of $A + B$. Thus, convexity is proved.

It remains to prove that $A + B$ is compact. Consider a sequence $(a_n + b_n \colon n = 1, 2, \dots)$, where each $a_n \in A$ and each $b_n \in B$. The sequence $((a_n, b_n) \colon n = 1, 2, \dots)$

has a subsequence $((a_{n_k}, b_{n_k})\colon k = 1, 2, \dots)$ that converges to a member (a, b) of $A \times B$, because $A \times B$ is compact. Since summation of coordinates is a continuous function on $A \times B$, the sequence $(a_{n_k} + b_{n_k})$ converges to the member $a + b$ of $A + B$. Hence, $A + B$ is compact. (By bringing the product space $A \times B$ into the argument we have avoided a proof involving a subsequence of a subsequence.)

10-52. For each φ: mean equals $\frac{4\sqrt{2}}{\pi}$ and variance equals $1 + \frac{2}{\pi} - \frac{16}{\pi^2}$

For Chapter 11

11-12. The one-point sets $\{0\}$ and $\{\pi\}$ each have probability $2^{n-1}3^{-n}$. The probability of any measurable B disjoint from each of these one-point sets is the product of $\frac{1}{2\pi}(1 - 2^n 3^{-n})$ and the Lebesgue measure of B.

11-13.

$$P\Big(\big\{\omega\colon \big(N(\omega) - 1, S_{N(\omega)-1}(\omega)\big) = (m, k)\big\}\Big) = r\binom{m}{k-m} q^{k-m} p^{2m-k}$$

for $m \le k \le 2m$ and 0 otherwise. $E(S_{N-1}) = \frac{p+2q}{r}$

11-14. for B a Borel subset of $\mathbb{R}^+$,

$$P\big(\{\omega\colon N(\omega) - 1 = m, S_{N(\omega)-1}(\omega) \in B\}\big) = Q(\{\infty\})Q^{*m}(B)\,;$$

$$E(S_{N-1}) = \frac{1}{Q(\{\infty\})} E(S_1\,;\, \{\omega\colon S_1(\omega) < \infty\})$$

11-17. Suppose that N is a stopping time. Then, for all $n \in \overline{\mathbb{Z}}^+$,

$$\{\omega\colon N(\omega) \le n\} \in \mathcal{F}_n\,,$$

which for $n = 0$ is the desired conclusion $\{\omega\colon N(\omega) = 0\} \in \mathcal{F}_0$. Suppose $0 < n < \infty$. Then

$$\{\omega\colon N(\omega) < n\} \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n\,.$$

Therefore,

$$\{\omega\colon N(\omega) = n\} = \{\omega\colon N(\omega) \le n\} \setminus \{\omega\colon N(\omega) < n\} \in \mathcal{F}_n\,.$$

We complete the proof in this direction by noting that

$$\{\omega\colon N(\omega) = \infty\} = \{\omega\colon N(\omega) \le \infty\} \setminus \bigcup_{m=0}^{\infty} \{\omega\colon N(\omega) \le m\}$$

and that all the events on the right side are members of $\mathcal{F}_\infty$.

For the converse we assume that $\{\omega\colon N(\omega) = n\} \in \mathcal{F}_n$ for all $n \in \overline{\mathbb{Z}}^+$. Then, whether $n < \infty$ or $n = \infty$,

$$\{\omega\colon N(\omega) \le n\} = \bigcup_{m \le n} \{\omega\colon N(\omega) = m\}\,.$$

All events on the right are members of $\mathcal{F}_n$ because filtrations are increasing. Therefore, the event on the left is a member of $\mathcal{F}_n$, as desired.

11-24. Let $A \in \mathcal{F}_M$. Then

$$\begin{aligned} A \cap \{\omega\colon N(\omega) \le n\} &= A \cap \big[\{\omega\colon M(\omega) \le n\} \cap \{\omega\colon N(\omega) \le n\}\big] \\ &= \big[A \cap \{\omega\colon M(\omega) \le n\}\big] \cap \{\omega\colon N(\omega) \le n\}\,, \end{aligned}$$

which, being the intersection of two members of $\mathcal{F}_n$, is a member of $\mathcal{F}_n$. Hence $A \in \mathcal{F}_N$. Therefore $\mathcal{F}_M \subseteq \mathcal{F}_N$.

11-28.

$$\sigma_1(s) = \frac{1 - \sqrt{1 - 4p(1-p)s^2}}{2(1-p)s}\,, \quad 0 \le s < 1$$

For n finite and even, the probability is 0 that n equals the hitting time of $\{1\}$. For $n = 2m - 1$, the hitting time of $\{1\}$ equals n with probability

$$\frac{1}{2m-1}\binom{2m-1}{m} p^m (1-p)^{m-1}\,.$$

The hitting time of $\{1\}$ equals ∞ with probability 0 or $(1-2p)/(1-p)$ according as $p \ge \frac{1}{2}$ or not.

If $p \ge \frac{1}{2}$, the global supremum equals ∞ with probability 1. If $p < \frac{1}{2}$, the global maximum exists a.s. and is geometrically distributed; the global maximum equals x with probability $\frac{1-2p}{1-p}(\frac{p}{1-p})^x$.

11-30. *Hint:* Use the Stirling Formula.

11-32. Let $(Z_j\colon j \ge 1)$ be a sequence of independent random variables with common distribution R (as used in the theorem). From the theorem we see that $(0, T_1, T_2, \dots)$ is distributed like a random walk with steps Z_j. Thus,

$$\begin{aligned} P(\{\omega\colon V(\omega) = k\}) &= P(\{\omega\colon Z_k(\omega) = \infty\,,\ Z_j(\omega) < \infty \text{ for } j < k\}) \\ &= P(\{\omega\colon T_1(\omega) = \infty\})\,\big[P(\{\omega\colon T_1(\omega) < \infty\})\big]^{k-1}\,. \end{aligned}$$

Set $k = 1$ to obtain the first equality in (11.6). The above calculation also shows that V is geometrically distributed unless $P(\{\omega\colon V(\omega) = \infty\}) = 1$. Thus, it only remains to prove the second equality in (11.6).

Notice that

$$V = \sum_{n=0}^{\infty} I_{\{\omega\colon S_n(\omega) = 0\}}\,.$$

Take expected values of both sides to obtain

$$E(V) = \sum_{n=0}^{\infty} Q^{*n}(\{0\})\,.$$

If the right side equals ∞, then $V = \infty$ a.s., for otherwise it would be geometrically distributed and have finite mean. If the right side is finite, then $E(V) < \infty$, and, so, V is geometrically distributed and, as for all geometrically distributed random variables with smallest value 1, $\frac{1}{E(V)} = P(\{\omega\colon V(\omega) = 1\})$.

11-40. $m!/m^m$

11-41. For $m = 3$, let Q_n denote the distribution of S_n.

$$Q_n(\{\emptyset\}) = \begin{cases} \frac{1}{4}(1 + 3^{-(n-1)}) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

$$Q_n(\{\{1\}\}) = Q_n(\{\{2\}\}) = Q_n(\{\{3\}\}) = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{1}{4}(1 + 3^{-n}) & \text{if } n \text{ is odd} \end{cases}$$

$$Q_n(\{\{1,2\}\}) = Q_n(\{\{1,3\}\}) = Q_n(\{\{2,3\}\}) = \begin{cases} \frac{1}{4}(1 - 3^{-n}) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

$$Q_n(\{\{1,2,3\}\}) = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{1}{4}(1 - 3^{-(n-1)}) & \text{if } n \text{ is odd} \end{cases}$$

11-42. probability that $\{\emptyset\}$ is hit at time n or sooner: $[1 - (\frac{1}{2})^n]^m$; probability that $\{\{1, 2, \dots, k\}\}$ is hit at the positive time $n < \infty$:

$$(\tfrac{1}{2})^{nk}\Big[\big(1 - (\tfrac{1}{2})^n\big)^{m-k} - \big(1 - (\tfrac{1}{2})^{n-1}\big)^{m-k}\Big]\,;$$

probability that hitting time of $\{1, \dots, m-1\}$ equals ∞: $(2^m - 2)/(2^m - 1)$

11-45. For $n \geq 1$ the distribution of S_n assigns equal probability to each one-point event. The sequence S is an independent sequence of random variables. For $n \geq 1$, the probability that the first return time to 0 equals n is $(\frac{1}{m})(1 - \frac{1}{m})^{n-1}$, where m is the number of members of the group.

For Chapter 12

12-10. (ii) Let $Z_n = X_1 I_{\{\omega\colon |X_1(\omega)| \leq n\}}$. Then $|Z_n(\omega)| \leq |X_1(\omega)|$ for each n and ω. Since $E(|X_1|) < \infty$ and $Z_n(\omega) \to X_1(\omega)$ for every ω for which $X_1(\omega)$ is finite, the Dominated Convergence Theorem applies to give $E(Z_n) \to E(X_1)$. Since X_1 and X_n have identical distributions, Z_n and Y_n also have identical distributions and hence the same expected value. Therefore $E(Y_n) \to E(X_1)$.

12-16. Let G denote the distribution function of $|X_1|$. Then

$$\begin{aligned}\sum_{m=1}^{\infty} P(\{\omega\colon |X_{2m}(\omega)| > 2cm\}) &= \sum_{m=1}^{\infty} [1 - G(2cm)] \\ &\geq \frac{1}{2c} \sum_{m=1}^{\infty} \int_{2cm}^{2c(m+1)} [1 - G(2cx)]\, dx \\ &= \frac{1}{2c} \int_{2c}^{\infty} [1 - G(2cx)]\, dx \\ &= \frac{1}{4c^2} \int_{4c^2}^{\infty} [1 - G(y)]\, dy\,,\end{aligned}$$

which, by Corollary 20 of Chapter 4, equals ∞, since $E(|X_1|) = \infty$. By the Borel-Cantelli Lemma, (12.1) is true.

To prove (12.2) we note that if $|X_{2m}(\omega)| > 2cm$, then $|S_{2m}(\omega)| > cm$ or $|S_{2m-1}(\omega)| > cm$ from which it follows that

$$\Big|\frac{S_{2m}(\omega)}{2m}\Big| \vee \Big|\frac{S_{2m-1}(\omega)}{2m-1}\Big| > \frac{c}{2}\,.$$

From (12.1) we see that, for almost every ω, this inequality happens for infinitely many m. Hence, 0 is the probability of the event consisting of those ω for which $S_n(\omega)/n$ converges to a number having absolute value less than $\frac{c}{2}$. Now let $c \to \infty$ through a countable sequence to conclude that (12.2) is true.

12-17. $E(S_n) = \prod_{k=1}^n E(X_k) = 2^{-n}$. An application of the Strong Law of Large Numbers to the sequence defined by $\log S_n = \sum_{k=1}^n \log X_k$ gives

$$\lim_{n\to\infty} \frac{\log S_n}{n} = E(\log X_1) = \int_0^1 \log x\, dx = -1 \text{ a.s.}$$

Since almost sure convergence implies convergence in probability, we conclude that, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(\{\omega\colon e^{-(1+\varepsilon)n} < S_n < e^{-(1-\varepsilon)n}\}) = 1\,.$$

Thus, with high probability $E(S_n)/S_n$ is very large for large n. There is some small probability that S_n is not only much larger than e^{-n}, but even much larger than 2^{-n}, and it is the contribution of this small probability to the expected value that makes $E(S_n)$ much larger (in the sense of quotients, not differences) than the typical values of S_n. The random variable S_n represents the length of the stick that has been obtained by starting with a stick of length 1 and breaking off n pieces from the stick, the length of the piece kept (or the piece broken off) at the n^{th} stage being uniformly distributed on $(0, S_{n-1})$.

12-19. $(1+p)(1-p)$, $(1+p)(1-p)^2$,

$$\frac{(1-p)^2}{1-p+p^2}\,, \quad \frac{(1+p-p^2+p^3-p^4)(1-p)}{1-p^2+2p^3-p^4}$$

12-27. Let $A \in \bigotimes_{n=1}^\infty \mathcal{G}$ and $\varepsilon > 0$. (We are only interested in exchangeable A but the first part of the argument does not use exchangeability.) By Lemma 18 of Chapter 9, there exists an integer p and a measurable subset D of $\prod_{n=1}^p \Psi$ such that $P(A \triangle B) < \varepsilon$, where

$$B = D \times \Big(\bigotimes_{n=p+1}^\infty \Psi\Big)\,.$$

Define a permutation π of $\mathbb{Z}^+ \setminus \{0\}$ by

$$\pi(n) = \begin{cases} n+p & \text{if } n \le p \\ n-p & \text{if } p < n \le 2p \\ n & \text{if } 2p < n\,. \end{cases}$$

Let $\hat\pi$ denote the corresponding permutation of Ω.

It is easy to check the following set-theoretic relation:

$$A \cap \hat\pi(A) \subseteq \big[A \triangle B\big] \cup \big[B \cap \hat\pi(B)\big] \cup \big[\hat\pi(B) \triangle \hat\pi(A)\big]\,.$$

Hence

$$P\big(A \cap \hat{\pi}(A)\big) \leq P\big(A \triangle B\big) + P\big(B \cap \hat{\pi}(B)\big) + P\big(\hat{\pi}(B) \triangle \hat{\pi}(A)\big)\,. \tag{7.3}$$

The first term on the right side of (7.3) is less than ε. Since $P(C) = P(\hat{\pi}(C))$ for any $C \in \bigotimes_{n=1}^{\infty} \mathcal{G}$,

$$P\big(\hat{\pi}(B) \triangle \hat{\pi}(A)\big) = P\big(\hat{\pi}(B \triangle A)\big) = P\big(B \triangle A\big) < \varepsilon\,.$$

Thus the third term on the right side of (7.3) is also less than ε. Therefore

$$P\big(A \cap \hat{\pi}(A)\big) < P\big(B \cap \hat{\pi}(B)\big) + 2\varepsilon \tag{7.4}$$

Now assume that A is exchangeable. Then $A \cap \hat{\pi}(A) = A$. Also, it is clear that B and $\hat{\pi}(B)$ are independent, and so

$$P(B \cap \hat{\pi}(B)) = P(B)P(\hat{\pi}(B)) = [P(B)]^2\,.$$

Another easily obtained fact is that $P(B) < P(A) + \varepsilon$. From (7.4), we therefore obtain

$$P(A) < \big(P(A) + \varepsilon\big)^2 + 2\varepsilon \leq [P(A)]^2 + 4\varepsilon + \varepsilon^2\,.$$

Algebraic manipulations give

$$P(A)[1 - P(A)] < 4\varepsilon + \varepsilon^2\,.$$

Let $\varepsilon \searrow 0$ to obtain $P(A)[1 - P(A)] = 0$, as desired.

12-30. (i) exchangeable but not tail, (ii) exchangeable and tail, (iii) neither exchangeable nor tail (but the Hewitt-Savage 0-1 Law can still be used to prove that the given event has probability 0 or 1) [Comment: the tail σ-field is a sub-σ-field of the exchangeable σ-field, so there is no random-walk example of an event that is tail but not exchangeable. This observation does not mean that the Kolmogorov 0-1 Law is a corollary of the Hewitt-Savage 0-1 Law, because there are settings where the Kolmogorov 0-1 Law applies and it is not even meaningful to speak of the exchangeable σ-field.]

12-35. $\sum P(\{\omega\colon |X_n(\omega)| > 1/n^2\}) \leq \sum(1/n^2) < \infty$. By the Borel Lemma, for almost every ω, $|X_n(\omega)| \leq (1/n^2)$ for all but finitely many n. By the comparison test for numerical series, $\sum X_n(\omega)$ converges (in fact, absolutely) for such ω.

12-40. by the Three-Series Theorem: Let b be any positive number, and define Y_n as in the theorem. By the Markov Inequality,

$$P(\{\omega\colon X_n(\omega) > b\}) \leq \frac{E(X_n)}{b} = \frac{1}{bn^2}\,.$$

Thus the series (12.8) converges. Since $0 \leq Y_n \leq X_n$, $0 \leq E(Y_n) \leq \frac{1}{n^2}$. Hence, the series (12.9) converges. Also,

$$\mathrm{Var}(Y_n) \leq E(Y_n^2) \leq bE(Y_n) \leq bE(X_n) = \frac{b}{n^2}\,.$$

Thus the series (12.10) converges. Therefore, $\sum X_n$ converges a.s. (Notice that this proof did not use the fact that the random variables are geometrically distributed.)

by Corollary 26: The distribution of X_n is geometric with parameter $(n^2+1)^{-1}$. Thus the variance is $(n^2+1)/n^4 < 2/n^2$. The series of these terms converges, as does the series of expectations. An appeal to Corollary 26 finishes the proof.

by Monotone Convergence Theorem: $E(\sum X_n) = \sum E(X_n) < \infty$. A random variable with finite expectation is finite a.s. Therefore, $\sum X_n$ is finite a.s. (Notice that for this proof, as for the proof by the Three-Series Theorem, the geometric nature of the distributions was not used.)

12-41. $\sum c_n^2 < \infty$

12-45. One place it breaks down is very early in the proof where the statement $\sum_{k=1}^{n} X_k(\omega) \neq \sum_{k=1}^{m} X_k(\omega)$ is replaced by the statement $\sum_{k=m+1}^{n} X_k(\omega) \neq 0$. These two statements are equivalent if the state space is $\mathbb{R}^d$, but if the state space is $\overline{\mathbb{R}}^+$ it is possible for the first of these two statements to be false, with both sums equal to ∞, and the second to be true.

For Chapter 13

13-15. if and only if the supports of the two uniform distributions have the same length

13-19. $k \to \frac{1-p}{1+p} p^{|k|}$; $v \rightsquigarrow \frac{(1-p)^2}{1+p^2-2p\cos v}$. [$p$ is the parameter of the (unsymmetrized) geometric distribution.]

13-30. mean equals $\sum_{k=1}^{m} k^{-1}$ and variance equals $\sum_{k=1}^{m} k^{-2}$

13-34. *Hint:* Let

$$f(\alpha) = \int_0^\infty \frac{1}{u^2+y^2}\, e^{-\alpha y}\, dy$$

and find a simple formula for $f'' + u^2 f$.

13-48.

$$\frac{2\pi}{\sqrt{a^2-b^2}} \left(\frac{a-\sqrt{a^2-b^2}}{b} \right)^{|n|}$$

13-72. yes

For Chapter 14

14-2. At any x where both F and G are continuous, $F(x) = G(x)$. The set of points where F is discontinuous is countable because F is monotone. The same is true for G. The set $\mathcal{D}$ of points where both F and G are continuous, and thus equal, is dense, because it has a countable complement. For any $y \in \mathbb{R}$, there exists a decreasing sequence $(x_1, x_2, \dots)$ in $\mathcal{D}$ such that $x_k \searrow y$ as $k \nearrow \infty$. The right continuity of F and G and the equality $F(x_k) = G(x_k)$ for each k then yield $F(y) = G(y)$.

14-4. We will first show that $Q_n\{x\} \to \frac{\lambda^x}{x!} e^{-\lambda}$ for each $x \in \mathbb{Z}^+$. The factor $e^{-\lambda}$ arises as the limit of $(1-\frac{\lambda}{n})^n$. The factor λ^x already appears in the formula for $Q_n\{x\}$, and

$x!$ appears there implicitly as part of the binomial coefficient. To finish this part of the proof we need to show

$$\lim_{n\to\infty} \frac{n!}{(n-x)!n^x} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1\,.$$

The second factor obviously has the limit 1 and the first factor can be written as

$$\prod_{k=0}^{x-1} \left(1 - \frac{k}{n}\right)$$

which also has limit 1.

We will finish the proof by showing that

$$\lim_{n\to\infty} \sum_{x\le y} Q_n(x) = \sum_{x \le y} \frac{\lambda^x}{x!}\, e^{-\lambda}$$

for every $y \in \mathbb{R}$. On the left side the limit and summation can be interchanged because the summation has only finitely many nonzero terms. The desired equality then follows from the preceding paragraph.

This problem could also be done by using Proposition 8 which appears later in Chapter 14.

14-6. standard gamma distributions. For $x > 0$,

$$\lim_{\gamma\searrow 0} \frac{1}{\Gamma(\gamma)} \int_0^x u^{\gamma-1} e^{-u}\, du = 1 - \Big[\lim_{\gamma\searrow 0} \frac{1}{\Gamma(\gamma)}\Big] \Big[\lim_{\gamma\searrow 0} \int_x^\infty u^{\gamma-1} e^{-u}\, du\Big]\,.$$

The first limit in the product of two limits equals 0 and by the Dominated Convergence Theorem, the second limit equals $\int_x^\infty u^{-1} e^{-u}\, du < \infty$, a dominating function being $(u^{-1} \vee 1)e^{-u}$. We conclude that

$$\lim_{\gamma\searrow 0} \frac{1}{\Gamma(\gamma)} \int_0^x u^{\gamma-1} e^{-u}\, du = 1$$

for $x > 0$ from which convergence to the delta distribution at 0 follows (despite the fact that we did not obtain convergence to 1 at $x = 0$).

14-10. Fix $x \ge 0$ and $r > 0$. We want to show

$$\lim_{m\to\infty} \sum_{k=0}^{\lfloor mx \rfloor} \left(\frac{1}{m}\right)^r \frac{(r)_k^{\uparrow}}{k!} \left(1 - \frac{1}{m}\right)^k = \frac{1}{\Gamma(r)} \int_0^x u^{r-1} e^{-u}\, du\,,$$

which is equivalent to

$$\lim_{m\to\infty} \sum_{k=1}^{\lfloor mx \rfloor} \left(\frac{1}{m}\right)^r \frac{(r)_k^{\uparrow}}{k!} \left(1 - \frac{1}{m}\right)^k = \frac{1}{\Gamma(r)} \int_0^x u^{r-1} e^{-u}\, du\,, \tag{7.5}$$

because the term $\frac{1}{m}$, obtained by setting $k = 0$, approaches 0 as $m \to \infty$.

The sum on the left side of (7.5) can be written as

$$\int_0^x g_m(u)\, du\,,$$

where

$$g_m(u) = \begin{cases} \left(\frac{k}{m}\right)^{r-1} \frac{(r)_k^{\uparrow}}{k^{r-1}\,k!} \left(1-\frac{1}{m}\right)^k & \text{if } k-1 < mu \le k \text{ for } k = 1, 2, \dots, \lfloor mx \rfloor \\ 0 & \text{otherwise}\,; \end{cases}$$

and the right side can be written as

$$\int_0^x g(u)\,du\,,$$

where

$$g(u) = \frac{1}{\Gamma(r)}\,u^{r-1}e^{-u}\,.$$

The plan is to show that $g_m(u) \to g(u)$ as $m \to \infty$ for each u in the interval $(0, x)$ and to find a function h that has finite integral and dominates each g_m, for then the desired conclusion will follow immediately from the Dominated Convergence Theorem. We will consider the three factors in g_m separately. It is important to keep in mind that k depends on u and m and that in particular, $k \to \infty$ as $m \to \infty$ for each fixed $u \in (0, x)$, as this dependence is not explicit in the notation.

It is clear that $\left(\frac{k}{m}\right)^{r-1} \to u^{r-1}$ for $u \in (0, x)$. In case $r \le 1$, $\left(\frac{k}{m}\right)^{r-1} \le u^{r-1}$. In case $r > 1$, $\left(\frac{k}{m}\right)^{r-1} \le x^{r-1}$. Thus, we have constructed one factor of what we hope will be the dominating function h: u^{r-1} in case $r \le 1$ and the constant x^{r-1} in case $r > 1$.

The second factor in $g_m(u)$ equals

$$\frac{1}{\Gamma(r)}\,\frac{\Gamma(r+k)}{k^{r-1}\,\Gamma(k-1)}\,.$$

We use the Stirling Formula to obtain the limit:

$$\begin{aligned}
&\frac{1}{\Gamma(r)} \lim_{k\to\infty} \frac{\Gamma(r+k)}{k^{r-1}\,\Gamma(k+1)} \\
&= \frac{1}{\Gamma(r)} \lim_{k\to\infty} \frac{\sqrt{2\pi}(r+k)^{r+k-\frac{1}{2}}\,e^{-(r+k)}}{k^{r-1}\sqrt{2\pi}\,(k+1)^{k+\frac{1}{2}}\,e^{-(k+1)}} \\
&= \frac{e^{-(r-1)}}{\Gamma(r)} \lim_{k\to\infty} \left(1+\frac{r}{k}\right)^{r-1} \left(1+\frac{r-1}{k+1}\right)^{-1/2} \left(1+\frac{r-1}{k+1}\right)^{k+1} \\
&= \frac{1}{\Gamma(r)}\,.
\end{aligned}$$

The second factor in $g_m(u)$ is thus bounded as a function of k, the bound possibly depending on r. Such a constant bound will be the second factor we will use in constructing the dominating function h.

For the third factor in $g_m(u)$ we observe that

$$\left(1-\frac{1}{m}\right)^{mu+1} < \left(1-\frac{1}{m}\right)^k \le \left(1-\frac{1}{m}\right)^{mu}\,, \tag{7.6}$$

from which it follows that

$$\left(1-\frac{1}{m}\right)^k \to e^{-u}\,.$$

Moreover, (7.6) and the inequality $(1-\frac{1}{m})^m < e^{-1}$ imply that e^{-u} is a dominating function for the third factor in $g_m(u)$.

Our candidate for a dominating function $h(u)$ having finite integral is a constant multiple of $u^{r-1}e^{-u}$ in case $r \leq 1$ and a constant multiple of e^{-u} in case $r > 1$. Both these function have finite integral on the interval $[0, x]$, as desired.

For $r = 0$, each $Q_{p,r}$ is the delta distribution at 0, and, therefore, $\lim_{m\to\infty} R_m$ equals this delta distribution.

14-14. Let G denote the standard Gumbel distribution function defined in Problem 13. For $a > 0$ and $b \in \mathbb{R}$,

$$G(ax + b) = e^{-e^{-(ax+b)}} = e^{-ce^{-ax}},$$

where $c = e^{-b} > 0$.

14-16. For any real constant x,

$$\sum_{n=1}^{\infty} P[X_n > c] = \infty\,.$$

By the Borel-Cantelli Lemma, $M_n \to \infty$ a.s. as $n \to \infty$ and, hence,

$$\{\omega\colon \lim_{n\to\infty} [M_n(\omega) - \log n] \text{ exists and } > m\}$$

is a tail event of the sequence $(X_k\colon k = 1, 2, \dots)$ for every m. By the Kolmogorov 0-1 Law, the almost sure limit of $(M_n - \log n)$ must equal a constant if it exists. On the other hand, by the preceding problem the almost sure limit, if it exists, must have a Gumbel distribution. Therefore, the almost sure limit does not exist.

The sequence does not converge in probability, for if it did, there would be a subsequence that converges almost surely and the argument of the preceding paragraph would show that the distribution of the limit would have to be a delta distribution rather than a Gumbel distribution.

The preceding problem does imply that

$$\frac{M_n - \log n}{\log n} \xrightarrow{\mathcal{D}} 0 \quad \text{as } n \to \infty$$

and, therefore, that

$$\frac{M_n}{\log n} \to 1 \quad \text{in probability as } n \to \infty\,.$$

In Example 6 of Chapter 9 the stronger conclusion of almost sure convergence was obtained using calculations not needed for either this or the preceding problem.

14-22. Weibull: mean $= -\Gamma(1+\frac{1}{\alpha})$, variance $= \Gamma(1+\frac{2}{\alpha}) - [\Gamma(1+\frac{1}{\alpha})]^2$; Fréchet: mean is finite if and only if $\alpha > 1$ in which case it equals $\Gamma(1-\frac{1}{\alpha})$, variance is finite if and only if $\alpha > 2$ in which case it equals $\Gamma(1-\frac{2}{\alpha}) - [\Gamma(1-\frac{1}{\alpha})]^2$

14-35. $Q_n\{0\} = 1 - \frac{1}{n}$, $Q_n\{n^2\} = \frac{1}{n}$

14-37. We need to show

$$\lim_{z\searrow 1} Q_z(-\infty, x] = \frac{c-1}{c}$$

for all positive finite x. That is, we must show

$$\lim_{z \searrow 1} \frac{1}{\zeta(z)} \sum_{k=1}^{\lfloor c^{1/(z-1)} x \rfloor} \frac{1}{k^z} = \frac{c-1}{c} \,.$$

We may replace $\frac{1}{\zeta(z)}$ by $z-1$ because the ratio of these two functions approaches 1 as $z \searrow 1$ (as may be checked by bounding the sum that defines the Riemann zeta function by formulas involving integrals). We can bound the above sum by using:

$$\int_1^m \frac{1}{x^z}\, dz < \sum_{k=1}^m \frac{1}{k^z} < 1 + \int_1^m \frac{1}{x^z}\, dz \,;$$

that is,

$$\frac{1}{z-1}\left(1 - \frac{1}{m^{z-1}}\right) < \sum_{k=1}^m \frac{1}{k^z} < 1 + \frac{1}{z-1}\left(1 - \frac{1}{m^{z-1}}\right);$$

Replace m by $\lfloor c^{1/(z-1)} x \rfloor$, multiply by $z-1$, and let $z \searrow 1$ to obtain the desired limit $1 - \frac{1}{c}$.

14-44. Since $|\beta_n(u)| \leq 1$ for every u and n, we only need show that $1 - \Re(\beta_n(u)) \to 0$ for each u. This will follow from the hypothesis in the lemma and the inequality

$$1 - \Re\bigl(\beta(2u)\bigr) \leq 4\bigl[1 - \Re\bigl(\beta(u)\bigr)\bigr] \,,$$

which we will now prove to be valid for all characteristic functions β.

Using the positive definiteness of β we have

$$\begin{aligned} &+ \beta(0-0) z_1 \bar{z}_1 + \beta(u-0) z_1 \bar{z}_2 + \beta(2u-0) z_1 \bar{z}_3 \\ &\quad + \beta(0-u) z_2 \bar{z}_1 + \beta(u-u) z_2 \bar{z}_2 + \beta(2u-u) z_2 \bar{z}_3 \\ &\qquad + \beta(0-2u) z_3 \bar{z}_1 + \beta(u-2u) z_3 \bar{z}_2 + \beta(2u-2u) z_3 \bar{z}_3 \geq 0 \,. \end{aligned}$$

Setting $z_1 = 1$, $z_2 = -2$, $z_3 = 1$, noting that $\beta(-v) = \overline{\beta(v)}$, and using $\beta(0) = 1$, we obtain

$$6 - 8\Re\bigl(\beta(u)\bigr) + 2\Re\bigl(\beta(2u)\bigr) \geq 0 \,,$$

from which follows

$$8\bigl[1 - \Re\bigl(\beta(u)\bigr)\bigr] \geq 2\bigl[1 - \Re\bigl(\beta(2u)\bigr)\bigr] \,,$$

as desired. (Notice that the characteristic function of the standard normal distribution shows that 4 is the smallest possible constant for the inequality proved above, but it does not resolve the issue of whether $\leq$ can be replaced by $<$ for $u \neq 0$.)

14-48. The probability generating function $\rho_{p,r}$ of $Q_{p,r}$ is given by

$$\rho_{p,r}(s) = \sum_{x=0}^{\infty} (1-p)^r \binom{-r}{x} p^x s^x = (1-p)^r (1-ps)^{-r} \,.$$

Clearly, $(p,r) \rightsquigarrow \rho_{p,r}(s)$ is a continuous function on

$$\{(p,r) \colon 0 \leq p < 1 \,,\; r \geq 0\}$$

for each fixed s, so the same is true of the function $(p,r) \rightsquigarrow Q_{p,r}$.

14-49. Example 1. The moment generating function of Q_n is

$$u \rightsquigarrow \frac{1}{n+1}\sum_{k=0}^{\infty}\Big(1+\frac{1}{n}\Big)^{-k}e^{-uk/n} = \frac{1}{n+1}\cdot\frac{1}{1-\frac{e^{-u/n}}{1+\frac{1}{n}}} = \frac{1}{n\big(1-e^{-u/n}+\frac{1}{n}\big)}\,,$$

which, as $n \to \infty$, approaches, pointwise, the function $u \rightsquigarrow \frac{1}{u+1}$, the moment generating function of the exponential distribution. An appeal to Theorem 19 finishes the proof.

14-52. Let V be the constant random variable 3 and let V_n be normally distributed with mean 3 and variance n^{-2}. Let $b_n = 3$ and $a_n = n^{-1}$. Then $(V_n - b_n)/a_n$ is normally distributed with mean 0 and variance 1 for every n even though $a_n \to 0$ as $n \to \infty$.

For Chapter 15

15-1. $\sum_{k=1}^{\infty}|X_k| \le 5\sum_{k=1}^{\infty}6^{-k} = 1$. Hence the series converges absolutely a.s. and therefore, it converges a.s., in probability, and in distribution; this is true without the independence assumption. The remainder of this solution, which concerns the limiting distribution and its characteristic function does use the independence assumption. The characteristic function of X_k is the function

$$v \rightsquigarrow \frac{1}{3}\Big(\cos\frac{v}{6^k}+\cos\frac{3v}{6^k}+\cos\frac{5v}{6^k}\Big)\,.$$

Therefore the characteristic function of $\sum_{k=1}^{\infty}X_k$ is the function

$$v \rightsquigarrow \prod_{k=1}^{\infty}\Big[\frac{1}{3}\Big(\cos\frac{v}{6^k}+\cos\frac{3v}{6^k}+\cos\frac{5v}{6^k}\Big)\Big]\,. \tag{7.7}$$

A direct simplification of this formula is not easy, so we will obtain the distribution by a method that does not rely on characteristic functions.

Calculations for $n = 1, 2, 3$ lead to the conjecture that the distribution Q_n of $\sum_{k=1}^{n}X_k$ is given by

$$Q_n\{m6^{-n}\} = 6^{-n} \quad \text{for } m \text{ odd, } |m| < 6^n\,.$$

This is easily proved by induction once it is noted that

$$\frac{m}{6^n}+\frac{5}{6^{n+1}} < \frac{m+2}{6^n}-\frac{5}{6^{n+1}}\,.$$

Then it is easy to let $n \to \infty$ to conclude that the distribution of $\sum_{k=1}^{\infty}X_k$ is the uniform distribution on $(-1, 1)$.

A sidelight: we have proved that the infinite product (7.7) equals the characteristic function of the uniform distribution on $(-1, 1)$ —namely $\frac{\sin v}{v}$.

15-6. 0.10

15-9. are not (except for the delta distribution at 0 in case one regards it as a degenerate Poisson distribution)

15-14. strict type consisting of positive constants (note: negative constants constitute another strict type)

15-16. *Hint:* The function g given by

$$g(u) = \int_0^\infty \frac{1}{x^{3/2}} e^{-\frac{b}{x} - ux}\, dx$$

can be evaluated by relating $g'(u)$ to the integral that can be obtained for $g(u)$ by using the substitution $y = \frac{c}{x}$ with an appropriate c. $a = \frac{1}{\sqrt{2}}$

15-20. (i) $\left(\frac{1-p}{1-p-\varepsilon}\right)^{1-p-\varepsilon}\left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon}$ (ii) $\left(1 + \frac{\varepsilon}{E(X_1)}\right)e^{-\varepsilon/E(X_1)}$

15-28. $\sup_{\{z\colon z-n \text{ even}\}}\left|P[S_n = z] - \frac{1}{\sqrt{2\pi np(1-p)}}\exp\left(-\frac{[z-n(2p-1)]^2}{8np(1-p)}\right)\right| = o\left(n^{-1/2}\right)$ as $n \to \infty$

For Chapter 16

16-1. *Hint:* Use Example 1.

16-6. 0.309 at 0; 0.215 at ± 1; 0.093 at ± 2; 0.029 at ± 3; 0.007 at ± 4; 0.001 at ± 5; 0.000 elsewhere (Comment: Using a certain table we found values that did not come close to summing to 1, so we concluded that either that table has errors or we were reading it incorrectly. We used another table.)

16-12. Suppose that $Q_k \to Q$ as $k \to \infty$. Fix n and suppose that there exist distributions R_k such that $R_k^{*n} = Q_k$. Let β_k, and γ_k denote the characteristic functions of Q_k and R_k, respectively. Because the family $\{Q_k\colon k = 1, 2, \dots\}$ is relatively sequentially compact, the family $\{\beta_k\colon k = 1, 2, \dots\}$ is equicontinuous at 0, by Theorem 13 of Chapter 14. Thus there exists some open interval B containing 0 such that $\beta_k(u) \neq 0$ for $u \in B$ and all k. So (Problem 7 of Appendix E), $\psi_k(u) = -\log \circ \beta_k(u)$ is well-defined for $u \in B$ and all k, and the family $\{\psi_k\colon k = 1, 2, \dots\}$ is equicontinuous at 0. For $u \in B$, $\gamma_k(u) = \exp\left(-\frac{1}{n}\psi_k(u)\right)$. Hence $\{\gamma_k\colon k = 1, 2, \dots\}$ is equicontinuous at 0. By Theorem 13 of Chapter 14 the family $\{R_k\colon k = 1, 2, \dots\}$ is relatively sequentially compact, and, therefore, the sequence (R_k) contains a convergent subsequence; let R denote the limit of such a subsequence. Since the convolution of convergent sequences converges to the convolution of the limit, $R^{*n} = Q$ as desired. [Comment: For fixed n we only used $R_k^{*n} = Q_k$ for each k, rather than the full strength of infinite divisibility. If Q is infinitely divisible we can strengthen the conclusion: From the forthcoming Proposition 3 it follows that β is never 0 and therefore that R is the unique distribution whose characteristic function is $\exp \circ (\frac{1}{n} \log \circ \beta)$ and moreover, it equals the limit of the sequence (R_k).]

16-13. By Proposition 1 the product of two infinitely divisible characteristic functions is infinitely divisible. The factors we use are the characteristic function of the compound Poisson distribution corresponding to ν, as in (16.1), and the function

$$u \rightsquigarrow \exp\left(i\left[\eta - \int_{\mathbb{R}\setminus\{0\}} \chi\, d\nu\right]u - \frac{\sigma^2 u^2}{2}\right),$$

known by Problem 9 to be infinitely divisible. The product equals $\exp \circ (-\psi)$, which is, therefore, an infinitely divisible characteristic function. For $\sigma = 0$ and $\eta = \int \chi\, d\nu$, the second factor is the function $u \rightsquigarrow 1$ and thus we obtain the compound Poisson characteristic function corresponding to an arbitrary finite measure ν.

16-14. Define ν_j, $1 \leq j \leq 3$, by

$$\begin{aligned}
\nu_1(B) &= \nu\big(B \cap [-1,1]\big)\,;\\
\nu_2(B) &= \nu\big(B \cap (-\infty,-1)\big)\,;\\
\nu_3(B) &= \nu\big(B \cap (1,\infty)\big)\,.
\end{aligned}$$

Write $\psi = \sum_{j=1}^4 \psi_j$, where

$$\begin{aligned}
\psi_1(u) &= \int_{\mathbb{R}\setminus\{0\}} \big(1 - e^{iuy} + iuy\big)\,\nu_1(dy)\,;\\
\psi_2(u) &= \int_{\mathbb{R}\setminus\{0\}} \big(1 - e^{iuy}\big)\,\nu_2(dy)\,;\\
\psi_3(u) &= \int_{\mathbb{R}\setminus\{0\}} \big(1 - e^{iuy}\big)\,\nu_3(dy)\,;\\
\psi_4(u) &= iu\big(-\eta - \nu(-\infty,-1) + \nu(1,\infty)\big) + \frac{\sigma^2 u^2}{2}\,.
\end{aligned}$$

Then X has the same distribution as $\sum_{j=1}^4 X_j$, where $(X_j : 1 \leq j \leq 4)$ is an independent quadruple and, for $1 \leq j \leq 4$, X_j is infinitely divisible with characteristic function $\exp\circ(-\psi_j)$. In view of the linearity of expectation, strengthened as in Problem 29 of Chapter 9 for independent random variables, and the linearity of variance for independent random variables, we have thus replaced the original problem by four subsidiary problems—to show:

$$\begin{aligned}
E(X_1) &= 0\,, & \mathrm{Var}(X_1) &= \int_{[-1,1]\setminus\{0\}} y^2\,\nu(dy)\,;\\
E(X_2) &= \int_{(-\infty,-1)} y\,\nu(dy)\,, & \mathrm{Var}(X_2) &= \int_{(-\infty,-1)} y^2\,\nu(dy)\,;\\
E(X_3) &= \int_{(1,\infty)} y\,\nu(dy)\,, & \mathrm{Var}(X_3) &= \int_{(1,\infty)} y^2\,\nu(dy)\,;\\
E(X_4) &= \eta + \nu(-\infty,-1) - \nu(1,\infty)\,, & \mathrm{Var}(X_4) &= \sigma^2\,.
\end{aligned}$$

(Comments: In defining ψ_2 and ψ_3, but not ψ_1 we were able to split off the term involving χ. It is important that no assumptions about existence of expectations or about finiteness of either expectations or variances are being made.)

The formulas involving X_4 are the known formulas for the mean and variance of a Gaussian random variable. Standard applications of the Dominated Convergence Theorem, based on bounds from Appendix E, show that ψ_1 has derivatives of all orders, in particular orders 1 and 2, which may be calculated by differentiating under the integral sign. Thus,

$$\psi_1'(u) = \int_{[-1,1]\setminus\{0\}} (-iye^{iuy} + iy)\,\nu(dy)$$

and

$$\psi_1''(u) = \int_{[-1,1]\setminus\{0\}} y^2 e^{iuy}\,\nu(dy)\,.$$

The first and second derivatives of $\exp\circ(-\psi_1)$ exist (because those of ψ_1 do) and equal the functions $-\psi_1'\cdot(\exp\circ(-\psi_1))$ and $(-\psi_1''+(\psi_1')^2)\cdot(\exp\circ(-\psi))$. Inserting $u=0$ gives 0 for the first derivative and $\int_{[-1,1]\setminus\{0\}} y^2\,\nu(dy)$ for the second, as desired.

Turning to X_3, with the intention of skipping X_2 because its treatment is so similar to that of X_3, we note that the desired formulas are obvious in case ν_3 is the zero measure and recognize that for other ν_3 we may use Example 2. In this latter case we replace ν_3 by λR where R is a probability measure on $(1,\infty)$. In terms of the notation of Example 2 we see that X_3 has the same distribution as

$$\sum_{k=1}^{\infty} Y_k I_{[M\geq k]}\,,$$

Using the independence of each pair (Y_k, M) and monotone convergence we obtain

$$\begin{aligned} E(X_3) &= \Big(\int_{(1,\infty)} y\,R(dy)\Big)\sum_{k=1}^{\infty} P[M\geq k] \\ &= \Big(\int_{(1,\infty)} y\,R(dy)\Big)E(M) = \int_{(1,\infty)} y\,\nu(dy)\,. \end{aligned}$$

We go for the second moment rather than directly for the variance (a useful strategy when monotone convergence is being used):

$$\begin{aligned} E(X_3^2) &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} E(Y_k Y_l I_{[M\geq k\vee l]}) \\ &= 2\sum_{k=2}^{\infty}\sum_{l=1}^{k-1} E(Y_k)E(Y_l)P[M\geq k] + \sum_{k=1}^{\infty} E(Y_k^2)P[M\geq k] \\ &= 2\Big(\int_{(1,\infty)} y\,R(dy)\Big)^2 \sum_{k=2}^{\infty}(k-1)P[M\geq k] \\ &\quad + \Big(\int_{(1,\infty)} y^2\,R(dy)\Big)\sum_{k=1}^{\infty} P[M\geq k]\,. \end{aligned} \tag{7.8}$$

The second term in (7.8) is what we want to prove the variance to be, so we only need prove that the first term equals $(E(X_3))^2$. To do this we only need show that

$2\sum_{k=2}^{\infty}(k-1)P[M \geq k] = \lambda^2$, which is a consequence of the following calculation:

$$\begin{aligned}
2\sum_{k=2}^{\infty}(k-1)P[M \geq k] &= 2\sum_{k=2}^{\infty}\sum_{m=k}^{\infty}(k-1)P[M=m] \\
&= 2\sum_{m=2}^{\infty}\sum_{k=2}^{m}(k-1)P[M=m] \\
&= \sum_{m=0}^{\infty} m(m-1)P[M=m] \\
&= E(M^2) - E(M) = (\lambda^2+\lambda) - \lambda = \lambda^2 \,.
\end{aligned}$$

16-17. If $\eta = 0$ and ν is symmetric about 0, the characteristic exponent is real because the function

$$y \rightsquigarrow -\sin uy + u\chi(y)$$

is an odd function for each u. Therefore the corresponding distribution is symmetric about 0 and its characteristic exponent has the form shown.

For the converse suppose that the characteristic function is real. It follows that the characteristic exponent is real since it is continuous and equals the real number 0 at 0. Then

$$-\eta u + \int_{\mathbb{R}\setminus\{0\}} \big(-\sin uy + u\chi(y)\big)\,\nu(dy) = 0$$

for every u. Another way to get 0 is to replace η by $\eta_0 = 0$ and ν by ν_0 defined by $\nu_0(B) = \frac{1}{2}(\nu(B) + \nu(-B))$. This change, together with no change in σ also leaves the real part of the characteristic exponent unchanged. By the uniqueness of the triples in Lévy-Khinchin representations (Lemma 11) it follows that $\eta = 0$ and $\nu = \nu_0$. We are done since it is obvious that ν_0 is symmetric about 0. (Comment: Another approach is to use the measure ζ defined in Lemma 7.)

16-20. Let X have a compound Poisson distribution with corresponding Lévy measure ν. Write $\nu = \nu_- + \nu_+$, where $\nu_-(0,\infty) = 0$ and $\nu_+(-\infty,0) = 0$. Then X has the same distribution as $X_- + X_+$, where (X_-, X_+) is an independent pair of compound Poisson random variables with corresponding Lévy measures ν_- and ν_+, the independence being a consequence of the factorization of (16.1) induced by $\nu = \nu_- + \nu_+$. If ν_- is not the zero measure, then by Problem 19 there is positive probability that $X_- < 0$ and $X_+ = 0$ and thus positive probability that $X < 0$. Therefore, ν_- must be the zero measure if $P[X \geq 0] = 1$.

16-25. The moment generating functions of a gamma distribution has the form $v \rightsquigarrow (1 + \frac{v}{a})^{-\gamma}$. Accordingly, we want to find (ξ, ν) (with $\nu\{\infty\} = 0$) such that

$$\gamma \log\Big(1 + \frac{v}{a}\Big) = \xi v + \int_{(0,\infty)} (1 - e^{-vy})\,\nu(dy)\,.$$

By letting $v \to \infty$ we see that the shift $\xi = 0$. Then differentiation of both sides, with differentiation inside the integral being justified by the Monotone Convergence

Theorem (or in some other manner), gives

$$\frac{\gamma}{a+v} = \int_{(0,\infty)} e^{-vy} y \, \nu(dy) \,.$$

It is now easy to see that the Lévy measure ν has the density $y \rightsquigarrow \gamma y^{-1} e^{-ay}$ with respect to Lebesgue measure on $(0,\infty)$.

16-33. Statement: Let $((\xi_n, \nu_n),\ n = 1, 2, \dots)$, satisfy: every $\xi_n \in \mathbb{R}^+$ and every ν_n is a Lévy measure for $\overline{\mathbb{R}}^+$. For each n, let Q_n be the infinitely divisible distribution on $\overline{\mathbb{R}}^+$ corresponding to (ξ_n, ν_n) via the relation

$$\int_{[0,\infty]} e^{-vx} \, Q_n(dx) = \exp\Big(-\xi_n v - \int_{(0,\infty]} (1 - e^{-vy}) \, \nu_n(dy) \Big) \,.$$

Then the sequence $(Q_n \colon n = 1, 2, \dots)$ converges to a distribution on $\overline{\mathbb{R}}^+$ different from the delta distribution at ∞ if and only if there exist $\xi \in \mathbb{R}^+$ and a Lévy measure ν for $\overline{\mathbb{R}}^+$ for which the following two conditions both hold:

$$\begin{aligned} \nu[x,\infty] &= \lim_{n\to\infty} \nu_n[x,\infty] \quad \text{if } 0 < x \text{ and } \nu\{x\} = 0\,; \\ \xi &= \lim_{\varepsilon \searrow 0} \limsup_{n\to\infty} \Big(\xi_n + \int_{(0,\varepsilon]} y \, \nu_n(dy) \Big) \\ &= \lim_{\varepsilon \searrow 0} \liminf_{n\to\infty} \Big(\xi_n + \int_{(0,\varepsilon]} y \, \nu_n(dy) \Big) \,. \end{aligned}$$

In case these conditions are satisfied the limit of the sequence $(Q_n \colon n \geq 1)$ is the infinitely divisible distribution with moment generating function

$$v \rightsquigarrow \exp\Big(-\xi v - \int_{(0,\infty]} (1 - e^{-vy}) \, \nu(dy) \Big) \,.$$

16-41. limiting distribution: two-sided Poisson supported by set of integral multiples of c; characteristic exponent: $u \rightsquigarrow 1 - \cos cu$.

16-42. limit exists; corresponding triple: $(0, 1, \nu)$, where ν has support $\{-1, 1\}$ and $\nu\{-1\} = \nu\{1\} = \frac{1}{2}$; characteristic exponent of the limit (not requested in the problem) is

$$u \rightsquigarrow \frac{u^2}{2} + 1 - \cos u \,.$$

16-50. *Hint:* Fix u and let $\varepsilon > 0$. By (E.2) and (E.3) of Appendix E and Lemma 20, the characteristic functions $\beta_{k,n}$ and corresponding characteristic exponents $\psi_{k,n}$ satisfy

$$\big(1 - \beta_{k,n}(u)\big) \leq \psi_{k,n}(u) \leq (1+\varepsilon)\big(1 - \beta_{k,n}(u)\big)$$

for all sufficiently large n (depending on u) and all $k \leq n$.

16-54. uan condition satisfied so Theorem 25 applicable; $u \rightsquigarrow e^{-(\log 2)u}$

16-59. Q exists and characterized by triple $(0, 0, \nu)$, where $\frac{d\nu}{d\lambda}(y) = \frac{(1-|y|)^2}{2} \vee 0$; $Q\{0\} = e^{-1/3}$

16-68. limit exists; $(0, \frac{1}{2\sqrt{3}}\sqrt{\log 2}, 0)$ is corresponding triple for its Lévy-Khinchin representation

For Chapter 17

17-3. slowly varying if $c < 1$; regularly varying of index 1 if $c = 1$; not regularly varying if $c > 1$

17-9. *Hint:* Find a bound for

$$\int_{(2^k, 2^{k+1}] \cup [-2^{k+1}, -2^k)} |s|^\beta \, R(ds) \, .$$

17-15. 1

17-17. $\frac{1}{2} + \frac{1}{\pi\alpha} \arctan(\gamma \tan \frac{\pi\alpha}{2})$ in case $\alpha \in (0,1) \cup (1,2]$; $\frac{1}{2} + \frac{1}{\pi} \arctan \frac{\xi}{k}$ with $\frac{\xi}{0} = \infty$ or $= -\infty$ according as $\xi > 0$ or $\xi < 0$ in case $\alpha = 1$; maximum value is $1 \wedge \frac{1}{\alpha}$.

17-29. in no domain of attraction

17-31. characteristic exponent of limiting distribution is $u \rightsquigarrow k_{4/3} \, |u|^{4/3}$;

$$a_n \sim 3^{3/4} \, e^{27/128} \, n^{3/4} \, e^{-(3/4)^{3/2} \sqrt{\log n}}$$

and $c_n = 0$.

17-38. in domain of attraction of stable distribution with $\alpha = 1$ and $\gamma = 1$; in domain of strict attraction of δ_1

For Chapter 18

18-5. *Hint:* Identify $\mathbf{C}[0, \infty)$ in a natural way with a closed subset of

$$\bigotimes_{n=0}^{\infty} \mathbf{C}[n, n+1] \, .$$

18-8. Let g be a continuous bounded $\mathbb{R}$-valued function on Υ. Then $g \circ h$ is a continuous bounded $\mathbb{R}$-valued function on Ψ. Therefore

$$\lim_{n \to \infty} \int_\Upsilon g \, dR_n = \lim_{n \to \infty} \int_\Psi (g \circ h) \, dQ_n = \int_\Psi (g \circ h) \, dQ = \int_\Psi g \, dR \, .$$

18-15. We first prove a related assertion—namely, the one obtained by replacing the hypothesis that A is open by the hypothesis that A is closed, in which case A is itself a Polish space by Proposition 3. If $Q(A) = 0$, this modified assertion (and also the original assertion) is clear, so assume that $Q(A) > 0$. For B a Borel subset of the Polish space A let

$$R(B) = \frac{Q(B)}{Q(A)} \, .$$

Clearly R is a probability measure. Let $\varepsilon > 0$. Corollary 18, applied to the Polish space A, shows that there exists a compact set K in the Polish space A such that $R(K) > 1 - \varepsilon$ and, thus,

$$Q(K) > (1-\varepsilon)Q(A) \geq Q(A) - \varepsilon\,.$$

The observation that, by Proposition 1 of Appendix C, K is compact in the Polish space Ψ completes the proof of the modification of the original assertion.

We return to the original assertion by now assuming that A is open in Ψ. We will prove that for every $\delta > 0$, there exists a subset C of A that is closed in Ψ and satisfies $Q(C) > Q(A) - \delta$. An application to C of the assertion proved above for closed sets then completes the proof.

Let S be a countable dense set in Ψ. It is easy to see that $S \cap A$ is a countable subset of A which, since A is open, is dense in A. For each $x \in S \cap A$, let B_x denote the closed ball centered at x whose radius is half the distance from x to A^c. It is easy to check that $A = \bigcup_{x \in S\cap A} B_x$. Replacing this union with a finite union over some finite subset of $S \cap A$ gives a closed set, a closed set whose Q-measure can, by continuity of measure, be chosen arbitrarily close to $Q(A)$, thus completing the proof.

Comment: The closed balls in the last paragraph of the proof need not be compact; this possibility is one reason the proof is so lengthy. Another reason is that an open subset of a Polish space is not necessarily a Polish space because it may not be complete. Thus, an intermediate result involving a closed subset is useful.

18-24. Let $w \in \mathbb{R}^d$. By the Classical Central Limit Theorem,

$$\Big\langle w\,, \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}} \Big\rangle = \frac{\sum_{k=1}^n \langle w, X_k\rangle - nE(\langle w, X_1\rangle)}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z_w\,,$$

where Z_w is a normally distributed $\mathbb{R}$-valued random variable having mean 0 and variance $\operatorname{Var}\langle w, X_1\rangle$. By the Cramér-Wold Device,

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}} \xrightarrow{\mathcal{D}} \text{some } Z$$

such that $\langle w, Z\rangle$ has the same distribution as Z_w for each $w \in \mathbb{R}^d$, and so we may redefine Z_w to actually equal $\langle w, Z\rangle$. Since each Z_w is normally distributed, Z itself is, by definition, normally distributed.

Let $w = (0, \dots, 0, 1, 0, \dots)$, where 1 is in the j^{th} position. Then $Z_w = Z_j$, and hence $E(Z_j) = 0$. Also,

$$\operatorname{Var} Z_j = \operatorname{Var} Z_w = \operatorname{Var}\langle w, X_1\rangle\,,$$

which equals the variance of the j^{th} coordinate of X_1. Therefore the mean vector of Z is the zero vector and the diagonal members of the covariance matrix of Z are the diagonal members of Σ.

Now let $w = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$, where 1 is in both the j^{th} and k^{th} positions. Then $Z_w = Z_j + Z_k$ and so

$$\operatorname{Var}(w, X_1) = \operatorname{Var}(Z_w) = \operatorname{Var}(Z_j) + \operatorname{Var}(Z_k) + 2\operatorname{Cov}(Z_j, Z_k)\,.$$

The left side is the sum of the variances of the j^{th} and k^{th} coordinates of X_1 and twice the covariance of the j^{th} and k^{th} coordinates. By the preceding paragraph the sum of the variances of the j^{th} and k^{th} coordinates of X_1 equals the sum $\operatorname{Var}(Z_j) + \operatorname{Var}(Z_k)$.

Thus twice the covariance of those two coordinates of X_1 must equal $2\operatorname{Cov}(Z_j, Z_k)$. Therefore the off-diagonal members of the covariance matrix of Z are the off-diagonal members of Σ.

18-26. *Hint:* Prove that $((A_\varepsilon)^c)_\varepsilon \subseteq A^c$.

18-29. first part: $\frac{|a|}{2} \wedge 1$.

For Chapter 19

19-4. The function $t \rightsquigarrow t$ is monotone (and therefore of bounded variation) on $[0,1]$ and, for each ω, the function $W(\omega, \cdot)$ is continuous. Hence (see Appendix D), we may use integration by parts to rewrite the given functional as

$$x(1) - \int_0^1 t\, dx(t) = \int_0^1 (1-t)\, dx(t)\,,$$

which in turn is the limit of Riemann-Stieltjes sums:

$$\lim_{k\to\infty} \sum_{i=1}^{k} \left(1 - \tfrac{i}{k}\right)\left(x(\tfrac{i}{k}) - x(\tfrac{i-1}{k})\right).$$

Under Wiener measure, this sum is the sum of k independent normally distributed random variables each of which has mean 0 and the i^{th} of which has variance $(1-\frac{i}{k})^2\frac{1}{k}$. Therefore the Riemann-Stieltjes sum itself is normally distributed with mean 0 and variance

$$\sum_{i=1}^{k} (1 - \tfrac{i}{k})^2 \tfrac{1}{k}\,.$$

This variance is a Riemann sum for the Riemann integral

$$\int_0^1 (1-t)^2\, dt = \tfrac{1}{3}\,.$$

By Problem 8 of Chapter 14 we see that the answer to the problem is: Gaussian with mean 0 and variance $\frac{1}{3}$.

19-8. We treat the case $m = n$; the case $m = 0$ is similar. Following along the lines of the argument in the text, but using the fact that $K(x) = 1$ is possible if $T(x) > 1$ and impossible if $x(\frac{1}{n}) < 0$, we obtain

$$\begin{aligned} &Q_n(\{x\colon K(x) = 1\}) \\ &= \frac{1}{2}\sum_{\substack{j=2\\ j\,\text{even}}}^{n} \frac{1}{j-1}\binom{j}{j/2}\frac{1}{2^j}\binom{(n-j)}{(n-j)/2}\frac{1}{2^{(n-j)}} + \frac{1}{2}\sum_{\substack{j=n+2\\ j\,\text{even}}}^{\infty} \frac{1}{j-1}\binom{j}{j/2}\frac{1}{2^j}\,, \end{aligned}$$

which, because of Lemma 12, equals

$$\frac{1}{2}\binom{n}{n/2}2^{-n} + \frac{1}{2}\sum_{\substack{j=n+2\\ j\,\text{even}}}^{\infty} \frac{1}{j-1}\binom{j}{j/2}\frac{1}{2^j}\,.$$

A straightforward induction proof that

$$\sum_{\substack{j=n+2\\ j\text{even}}}^{\infty} \frac{1}{j-1}\binom{j}{j/2}\frac{1}{2^j} = \binom{n}{n/2}2^{-n}$$

completes the proof. [For $n = 0$ (the starting value for the induction proof), the left side equals the probability—namely 1 —that the time of first return to 0 equals some finite value, and 1 is also the value of the right side when $n = 0$.]

19-11. $\frac{2+\pi}{2\pi} \approx 0.82$

19-27. We need to show that the value of the derivative of the moment generating function at 0 equals $-ab$. By definition, the derivative there equals

$$\begin{aligned}&\lim_{u\searrow 0} \frac{\sinh(a\sqrt{2u}) + \sinh(b\sqrt{2u}) - \sinh((a+b)\sqrt{2u})}{u\sinh((a+b)\sqrt{2u})}\\ &= \lim_{w\searrow 0} \frac{2[\sinh(aw) + \sinh(bw) - \sinh((a+b)w)]}{w^2\sinh((a+b)w)}\,.\end{aligned}$$

Now three applications of the l'Hospital Rule yield the desired result.

For Chapter 20

20-5. $E(X)$

20-6. Proof of (iv): By the Cauchy-Schwarz Inequality

$$E(|X-X_n|) = E(|X-X_n|1) \le \sqrt{E(|X-X_n|^2)}\sqrt{E(1^2)} = \sqrt{E((X-X_n)^2)} \to 0\,.$$

Proof of (iii), using (iv):

$$\limsup E(|X_n|) \le E(|X|) + \limsup E(|X_n - X|) = E(|X|)$$

and

$$\begin{aligned}E(|X|) &\le \liminf\big[E(|X_n|) + E(|X-X_n|)\big]\\ &\le \liminf E(|X_n|) + \limsup E(|X-X_n|) = \liminf E(|X_n|)\,,\end{aligned}$$

from which the desired conclusion follows.

20-15. By the sentence preceding the problem, $E(V_i) = 0$ for each i and $E(Z) = E(X)$. Hence, $E(X-Z) = 0$. Our task has become that of showing $E((X-Z)Y_i) = 0$ for each i. In view of the fact that each Y_i is a linear combination of 1 and the various V_j and that we have already shown that $E((X-Z)1) = 0$, we can reformulate our task as that of showing that $E(XV_j) = E(ZV_j)$ for each j.

From the definition of Z we obtain

$$E(ZV_j) = \langle X, 1\rangle E(V_j) + \sum_{i=1}^{m}\langle X, V_i\rangle E(V_iV_j) = \langle X, V_j\rangle = E(XV_j)\,.$$

For Chapter 21

21-3. By Definition 1: Clearly, $P(B \mid \mathcal{G})I_A$ is a member of $\mathbf{L}_2(\Omega, \mathcal{G}, P)$. Let $Y \in \mathbf{L}_2(\Omega, \mathcal{G}, P)$. To finish the proof we must show

$$E\Big([I_{A\cap B} - P(B \mid \mathcal{G})I_A]\, Y\Big) = 0\,.$$

That is we must show that

$$E\Big([I_B - P(B \mid \mathcal{G})]\, [I_A Y]\Big) = 0\,.$$

In view of the fact that $I_A Y$ is $\mathcal{G}$-measurable, this statement follows from the definition of $P(B \mid \mathcal{G})$.

By Proposition 2: Let $X = P(B \mid \mathcal{G})I_A$. Condition (i) of Proposition 2 is clearly satisfied by X. To check condition (ii), let $C \in \mathcal{G}$. Then we must show that

$$E(XI_C) = P((A \cap B) \cap C)\,.$$

That is, we must show that

$$E(P(B \mid \mathcal{G})I_{A\cap C}) = P(B \cap (A \cap C))\,.$$

In view of the fact that $A \cap C \in \mathcal{G}$, this last statement follows from Proposition 2 applied to $P(B \mid \mathcal{G})$.

[Comment: Notice the similarity between the two proofs. Proposition 2 says that the orthogonality condition entailed in Definition 1 need only be checked for indicator functions of members of $\mathcal{G}$ rather than for every member of $\mathbf{L}_2(\Omega, \mathcal{G}, P)$.]

21-5. The right side X of (21.1) is obviously $\sigma(C)$-measurable. To check the second condition in Proposition 2 we only have to consider the four members of $\sigma(C)$. Obviously $E(XI_\emptyset) = 0 = P(A \cap \emptyset)$. Also,

$$E(XI_C) = \frac{P(A \cap C)}{P(C)} E(I_C I_C) = P(A \cap C)$$

and similarly,

$$E(XI_{C^c}) = \frac{P(A \cap C^c)}{P(C^c)} E(I_{C^c} I_{C^c}) = P(A \cap C^c)\,.$$

Finally,

$$E(XI_\Omega) = E(XI_C) + E(XI_{C^c}) = P(A \cap C) + P(A \cap C^c) = P(A \cap \Omega)\,.$$

21-8. (ii)

$$\omega \rightsquigarrow \begin{cases} 1 & \text{if } \omega_1 + \omega_2 + \omega_3 + \omega_4 = 4 \\ \frac{1}{2} & \text{if } \omega_1 + \omega_2 + \omega_3 + \omega_4 = 2 \\ \frac{1}{6} & \text{if } \omega_1 + \omega_2 + \omega_3 + \omega_4 = 0 \\ 0 & \text{otherwise} \end{cases}$$

$\frac{1}{6}$ for the particular given ω

(v)

$$\omega \rightsquigarrow \begin{cases} \frac{1}{4} & \text{if } \omega_1 + \omega_2 + \omega_3 + \omega_4 = 0 \\ 0 & \text{otherwise} \end{cases}$$

$\frac{1}{4}$ for the particular given ω

21-9. The general formula is

$$16\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}\sum_{l=0}^{1} P(A\cap B_{i,j,k,l})I_{B_{i,j,k,l}}\,,$$

where

$$B_{i,j,k,l} = \{\psi\colon \psi_1 = 2i-1, \psi_2 = 2j-1, \psi_3 = 2k-1, \psi_4 = 2l-1\}\,.$$

(ii)

$$\omega \leadsto \begin{cases} 1 & \text{if } \omega_1+\omega_2 = 2 \\ 0 & \text{otherwise} \end{cases}$$

0 for the particular given ω

(v) same answer as problem 8

21-10. For each positive integer m and almost every ω,

$$P(\limsup A_n \mid \mathcal{G})(\omega) \le P\Big(\bigcup_{n=m}^{\infty} A_n \mid \mathcal{G}\Big)(\omega) \le \sum_{n=m}^{\infty} P(A_n \mid \mathcal{G})(\omega)\,.$$

For those ω for which the sum on the right is finite, that sum can be made arbitrarily close to 0 by choosing m sufficiently large (depending on ω). For such an ω the probability on the far left must equal 0 since it does not depend on m. This completes the proof of the first of the two assertions in the problem.

21-12. $\psi \leadsto \psi_1$

21-13. It is possible that the image of V is not a measurable subset of Ψ.

21-17. $v \leadsto v$

21-24. With Q denoting the distribution of Y and δ_x the delta distribution at x, a conditional distribution is the function

$$(\omega, B) \leadsto Q\big([X(\omega),\infty)\big)\delta_{X(\omega)}(B) + Q\big(B\cap(-\infty, X(\omega)\big)\,.$$

(Various functions are presented via this notation: one function of two variables, functions of B for various fixed values of ω, and functions of ω for various fixed values of B.)

21-25. With Q denoting any fixed distribution [for instance, the (unconditional) distribution of X and δ_c denoting the delta distribution at c, a conditional distribution is $g\circ|X|$, where

$$g(w) = \begin{cases} \frac{f(-w)}{f(-w)+f(w)}\,\delta_{-w} + \frac{f(w)}{f(-w)+f(w)}\,\delta_w & \text{if } f(-w)+f(w)\ne 0 \\ Q & \text{if } f(-w)+f(w) = 0\,. \end{cases}$$

21-30.

$$(\omega, x) \leadsto \begin{cases} \frac{1}{\lambda}\,e^{-(x-t)/\lambda} & \text{if } X(\omega)\ge t\,,\ x\ge t \\ \frac{1}{\lambda(1-e^{-t/\lambda})}\,e^{-x/\lambda} & \text{if } X(\omega)< t\,,\ 0\le x\le t \\ 0 & \text{otherwise} \end{cases}$$

21-34. The density is

$$(x_1, \dots, x_{d-1}, y) \rightsquigarrow \frac{(y - x_1 - \cdots - x_{d-1})^{\gamma_d - 1} e^{-y}}{\Gamma(\gamma_d)} \prod_{i=1}^{d-1} \frac{x_i^{\gamma_i - 1}}{\Gamma(\gamma_i)}$$

for $x_i \geq 0, y \geq x_1 + \cdots + x_{d-1}$.

Let $Y = X_1 + \cdots + X_d$. A conditional density of $(X_1, \dots, X_{d-1})$ given $\sigma(Y)$ is

$$\begin{aligned}&(\omega, (x_1, \dots, x_{d-1}))\\ &\quad\rightsquigarrow \frac{\left(1 - \frac{x_1}{Y(\omega)} - \cdots - \frac{x_{d-1}}{Y(\omega)}\right)^{\gamma_d - 1} \Gamma(\gamma_1 + \cdots + \gamma_d)}{\Gamma(\gamma_d)\,[Y(\omega)]^{d-1}} \prod_{i=1}^{d-1} \frac{\left(\frac{x_i}{Y(\omega)}\right)^{\gamma_i - 1}}{\Gamma(\gamma_i)}\end{aligned}$$

for $x_j \geq 0, x_1 + \cdots + x_{d-1} \leq Y(\omega)$ if $Y(\omega) > 0$ and $\rightsquigarrow$ the unconditional density of $(X_1, \dots, X_{d-1})$ if $Y(\omega) \leq 0$. [Note the relationship to the Dirichlet distribution which is described in an optional section of Chapter 10.]

21-44. Let Ω consist of the four points corresponding to two independent fair coins. Let $\mathcal{G}$ denote the σ-field generated by the first coin and $\mathcal{H}$ the σ-field generated by the second coin. By definition, $(\mathcal{G}, \mathcal{H})$ is an independent pair and it is clear that $\sigma(\mathcal{G}, \mathcal{H})$ consists of all subsets of Ω. Thus, any σ-field consisting of subsets of Ω is a sub-σ-field of $\sigma(\mathcal{G}, \mathcal{H})$. Let $\mathcal{K}$ be the σ-field generated by the event that exactly 1 head is flipped. Given $\mathcal{K}$ the conditional probability of any member of $\mathcal{G}$ different from $\emptyset$ and Ω equals $\frac{1}{2}$ as does the conditional of any such member of $\mathcal{H}$. But, there is no event that has conditional probability given $\mathcal{K}$ equal to $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

For Chapter 22

22-10. If X_3 were to exist so that (X_1, X_2, X_3) is exchangeable, then, since $X_1 + X_2 = 0$ with probability 1, it would follow that $X_1 + X_3 = 0$ and $X_2 + X_3 = 0$ with probability 1. By solving three equations in three unknowns it would then follow that $X_1 = 0$ with probability 1, a contradiction.

22-11. *Hint:* Apply $E(P(A \mid \mathcal{G})) = P(A)$ for various choices for A.

22-14. uniform on the set of those $\binom{n}{[n+S_n(\omega)]/2}$ sequences of ± 1's that contain $[n + S_n(\omega)]/2$ 1's and $[n - S_n(\omega)]/2$ -1's. [Comment: The answer does not depend on p.]

22-16. first term equals 1 with probability $\frac{\alpha}{\alpha+\beta}$. conditional distribution of second term given first term: equals 1 with probability $\frac{\alpha+1}{\alpha+\beta+1}$ if first term equals 1 and equals 1 with probability $\frac{\alpha}{\alpha+\beta+1}$ if first term equals 0. distribution of first two terms: equals $(1,1)$ with probability $\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$ and equals $(0,0)$ with probability $\frac{\beta(\beta+1)}{(\alpha+\beta)(\alpha+\beta+1)}$ and equals $(1,0)$ and $(0,1)$ each with probability $\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$

22-21. By exchangeability, the correlation of I_m and I_n is the same as that of I_1 and I_2 if $n \neq m$; of course, it equals 1 if $n = m$.

The correlation of I_1 and I_2 equals $\frac{c}{x_0+y_0+c}$, which approaches 0 as $(x_0, y_0) \to (\infty, \infty)$ and approaches 1 as $c \to \infty$.

For large (x_0, y_0) the knowledge of the color of a fixed number c of balls in the urn hardly influences the probability that a blue ball will be drawn. For large c, the second ball drawn is very likely to be of the same color as the first ball since after the first ball is drawn almost all the balls in the urn will have the same color as the first ball.

22-22. Using the fact that $\prod_n \frac{a+cn}{b+cn} = 0$ if $0 \le a < b$ and $0 \le c$, we have

$$\begin{aligned} P[I_n = 0 \text{ for } n > m] &= E(P[I_n = 0 \text{ for } n > m \mid \sigma(X_m, Y_m)]) \\ &= E\biggl(\prod_{n=m+1}^{\infty} \frac{Y_m + (n-m-1)c}{X_m + Y_m + (n-m-1)c}\biggr) \\ &= E(0) = 0 \end{aligned}$$

for each fixed m. Hence

$$\begin{aligned} P(\liminf\{\omega\colon I_n(\omega) = 0\}) &= P\biggl(\bigcup_{m=1}^{\infty} \bigcap_{n>m} \{\omega\colon I_n(\omega) = 0\}\biggr) \\ &\le \sum_{m=1}^{\infty} P\biggl(\bigcap_{n>m} \{\omega\colon I_n(\omega) = 0\}\biggr) \\ &= \sum_{m=1}^{\infty} 0 = 0\,, \end{aligned}$$

from which it follows that the first event in the problem has probability 1. That the second event given there also has probability 1 follows by applying the result already proved to the sequence $((1-I_n)\colon n = 1, 2, \dots)$, an application which is seen to be valid by interchanging the colors of the balls.

22-24. $\frac{(m-1)!}{m^{(n-1)}}\, S\bigl((n-1), (m-1)\bigr)$

For Chapter 23

23-11. *Hint:* Use Problem 14 of Chapter 5.

23-17. Let $\omega = (0, 1]$, $\mathcal{F}$ the Borel σ-field, and P Lebesgue measure. Let $X_n = nI_{(0,1/n)}$. Then $X_n(\omega) \to 0$ for every ω and $E(X_n) = 1$, so the (unconditional) Dominated Convergence Theorem must not apply. Let

$$\mathcal{G} = \sigma\bigl((2^{-m}, 2^{-(m-1)}]\colon m = 1, 2, \dots\bigr)\,.$$

The random variable $Y(\omega) = \frac{1}{\omega}$ dominates every X_n and satisfies $E(Y \mid \mathcal{G})(\omega) = 2^m \log 2$ for $2^{-m} < \omega \le 2^{-(m-1)}$. In particular $E(Y \mid \mathcal{G})(\omega) < \infty$ for every ω. Hence the Conditional Dominated Convergence Theorem applies. We conclude that $E(X_n \mid \mathcal{G})(\omega) \to 0$ for almost every ω, a fact that we could have also obtained by directly by observing that $E(X_n \mid \mathcal{G})(\omega) = 0$ for $n > \frac{2}{\omega}$.

23-23. Problem 21 of Chapter 21

23-30. $\frac{1}{2}$ (for all b), which is larger than $\frac{1}{3}$, the (unconditional) expectation. The following paragraphs present various ways of looking at the situation.

Fix b. If, before the random experiment begins, it is understood that one will be told whether or not b is between X and Y, one will clearly want to assign a larger value to the expectation of $Y - X$ in case b is between X and Y and a smaller value otherwise.

An appropriate weighted average of these two numbers equals $\frac{1}{3}$, so, as expected, the first of these two numbers is larger than $\frac{1}{3}$.

Knowing that exactly one of two order statistics from the uniform distribution on $(0,1)$ is larger than b gives no reason for biasing one's estimate for it among the various values larger than b. Thus, the conditional mean of its excess over b is half the distance from b to 1 —namely, $\frac{1-b}{2}$. Similarly the conditional mean of the difference between b and the smaller of the two order statistics is $\frac{b}{2}$. The sum of these two conditional expectations is $\frac{1}{2}$, independently of b.

Here is a second method of getting an intuitive feel for the value $\frac{1}{2}$. Fix the number b. Pick three iid points Z_1, Z_2, and Z_3 on a circle of circumference 1. Cut the circle at Z_1 in order to straighten it into a unit interval with the counterclockwise direction on the circle corresponding locally to the direction of increase on the unit interval. Then set the smaller of Z_2 and Z_3 equal to X and the larger equal to Y. The condition that b be between X and Y is the condition that as one traverses the circle counterclockwise the contacts with either Z_2 or Z_3 alternate with the contacts with either Z_1 or b. Among such possible arrangements, there is probability $\frac{1}{2}$ that b lies in the long interval and Z_1 in the short interval determined by Z_2 and Z_3 and probability $\frac{1}{2}$ that the opposite relations hold. So the average length of the interval in which b lies is $\frac{1}{2}$.

23-33. By Problem 27 and Proposition 6, there exist choices of $E(X^+I_B \mid \mathcal{H})$ and $E(X^-I_B \mid \mathcal{H})$ such that

$$E(E(X^+ \mid \mathcal{G})I_B \mid \mathcal{H})(\omega) = E(E(X^+I_B \mid \mathcal{G}) \mid \mathcal{H})(\omega) = E(X^+I_B \mid \mathcal{H})(\omega)$$

and

$$E(E(X^- \mid \mathcal{G})I_B \mid \mathcal{H})(\omega) = E(E(X^-I_B \mid \mathcal{G}) \mid \mathcal{H})(\omega) = E(X^-I_B \mid \mathcal{H})(\omega)$$

for every sample point ω. Subtraction gives

$$\begin{aligned} &E(E(X^+ \mid \mathcal{G})I_B \mid \mathcal{H})(\omega) - E(E(X^- \mid \mathcal{G})I_B \mid \mathcal{H})(\omega) \\ &= E(XI_B \mid \mathcal{H})(\omega) \end{aligned} \tag{7.9}$$

for every ω for which the right side of (7.9) [that is, the right side of (23.9)] exists. At such an ω at least one of the two terms on the left side is finite.

We will focus on

$$A \stackrel{\text{def}}{=} \{\omega\colon E(E(X^+ \mid \mathcal{G})I_B \mid \mathcal{H})(\omega) < \infty\}\,.$$

For each $\omega \in A$,

$$\int_{[0,\infty]} x\, Z(\omega, dx) < \infty\,,$$

where Z is the conditional distribution of $E(X^+ \mid \mathcal{G})I_B$. So $E(Z(\cdot, \{\infty\})I_A) = 0$. From the definition of conditional probability we then obtain

$$P\Big(\{\omega\colon [E(X^+ \mid \mathcal{G})I_B](\omega) = \infty\} \cap A\Big) = 0\,.$$

Therefore the left side of (7.9) can be rewritten as

$$E\Big([E(X^+ \mid \mathcal{G}) - E(X^- \mid \mathcal{G})]I_B \;\Big|\; \mathcal{H}\Big)(\omega) \tag{7.10}$$

for almost every ω for which the right side of (7.9) is less than ∞. Similarly, this can be done for almost every ω for which the right side of (7.10) is greater than $-\infty$, in particular for almost every ω for which the right side of (7.9) equals ∞.

The upshot is that for almost every ω for which the right side of (7.9) exists, the left side of (7.9) can be rewritten as (7.10) in which the inside difference between two conditional expectations is not of the form $\infty - \infty$. Therefore linearity of conditional expectation may be used to complete the proof.

23-42. *Hint:* Apply the Conditional Chebyshev Inequality and then take (unconditional) expectations of both sides.

For Chapter 24

24-2. The 'if' part is obvious. For the proof of 'only if' fix n. The inequality in the problem is obviously true with equality in case $m = 0$ and it is true by definition if $m = 1$. To complete an inductive proof, let $m > 1$ and assume that

$$E(X_{n+(m-1)} \mid \mathcal{F}_n) \geq X_n \text{ a.s.}$$

Since $\mathcal{F}_n \subseteq \mathcal{F}_{n+(m-1)}$,

$$\begin{aligned} E\big(X_{n+m} \mid \mathcal{F}_n\big) &= E\big(E(X_{n+m} \mid \mathcal{F}_{n+(m-1)}) \mid \mathcal{F}_n\big) \\ &\geq E\big(X_{n+(m-1)} \mid \mathcal{F}_n\big) \geq X_n \text{ a.s.} \end{aligned}$$

24-8. We treat the real and imaginary parts simultaneously. Let $\mathcal{F}_n = \sigma(S_0, \dots, S_n)$ and denote the steps of the random walk by $X_1, X_2, \dots$. Then

$$\begin{aligned} E(Y_{n+1} \mid \mathcal{F}_n) &= \frac{1}{(\varphi(u))^{n+1}} E\big(e^{iuS_n} e^{iuX_{n+1}} \mid \mathcal{F}_n\big) \\ &= \frac{1}{(\varphi(u))^{n+1}} e^{iuS_n} E\big(e^{iuX_{n+1}} \mid \mathcal{F}_n\big) \\ &= \frac{1}{(\varphi(u))^{n}} e^{iuS_n} = Y_n\,. \end{aligned}$$

[Remark: We have proved that the real and imaginary parts of $(Y_n \colon n = 0, 1, \dots)$ are martingales with respect to the minimal filtration for the random walk, which may possibly contain larger σ-fields than the corresponding σ-fields in the minimal filtration for the sequence (Y_n).]

24-10. Proof of uniqueness: Suppose that conditions (i)-(iv) of the proposition hold as stated and that they also hold with some sequences Z and U in place of Y and V, respectively. By subtraction

$$Z_n - Y_n = V_n - U_n\,.$$

Thus $Z_n - Y_n$ is $\mathcal{F}_{n-1}$-measurable, and, hence,

$$Z_n - Y_n = E\big((Z_n - Y_n) \mid \mathcal{F}_{n-1}\big) = Z_{n-1} - Y_{n-1}\,.$$

This fact combined with $Z_0 - Y_0 = 0$, a consequence of $U_0 = V_0 = 0$, gives $Z_n = Y_n$, and therefore $U_n = V_n$ for every n.

24-11. Let $\mathcal{F}_n = \sigma(S_0, \dots, S_n)$. Then

$$\begin{aligned} E\big((S_{n+1}^2 - S_n^2) \mid \mathcal{F}_n\big) &= E\big((S_{n+1} - S_n)^2 \mid \mathcal{F}_n\big) + 2E\big((S_{n+1} - S_n)S_n \mid \mathcal{F}_n\big) \\ &\geq 0 + 2S_n E\big((S_{n+1} - S_n) \mid \mathcal{F}_n\big) = 0\,, \end{aligned}$$

as desired. [Remark: See the remark in the solution of Problem 8.] $V_n = n\,\mathrm{Var}(S_1)$.

24-20. *Hint:* $|X_n| I_{[T>n]} = X_n I_{[T>n]} \leq X_T I_{[T>n]} \leq X_T$

24-23. *Hint:* Use two relevant previous results; do not do any hard work.

24-26. The sequence $(X_n \colon n \geq 0)$, being uniformly bounded, is uniformly integrable. By Theorem 12 and the Optional Sampling Theorem, $E(X_T) \leq E(X_0) = f_0$; Clearly $E(X_T) \geq g\,P[X_T = g]$. Hence $f_0 \geq g\,P[X_T = g]$, as desired.

24-33.

$$\begin{aligned} E([S_{T_n} - \tfrac{1}{2}T_n]^2) = \mathrm{Var}(S_1)E(T_n) = 2^{-1}[1 - 2^{-n}] &\nearrow 2^{-1} = E([S_T - \tfrac{1}{2}T]^2) \\ \mathrm{Var}(S_{T_n}) = 2^{-n}[1 - 2^{-n}] &\searrow 0 = \mathrm{Var}(S_T) \\ E\big(\mathrm{Var}(S_{T_n} \mid T_n)\big) = 2^{-(n+1)} &\searrow 0 = E\big(\mathrm{Var}(S_T \mid T)\big) \end{aligned}$$

For $n > 1$, $\mathrm{Var}(S_{T_n}) < E(S_1)E(T_n)$, thus highlighting the importance of the assumption in Theorem 15 of mean 0 for the steps.

24-41. Suppose that X is a uniformly integrable martingale. By the theorem it has an almost sure limit $Y = X_\infty$ such that $(X_n \colon n \in \overline{\mathbb{Z}}+)$ is both a submartingale and a supermartingale—that is, a martingale. Hence $E(Y \mid \mathcal{F}_n) = X_n$. Moreover, Y is $\mathcal{F}_\infty$-measurable, so $E(Y \mid \mathcal{F}_\infty) = Y$.

For the converse, suppose that Y has finite expectation and

$$X_n = E(Y \mid \mathcal{F}_n)$$

for each $n \in \mathbb{Z}^+$. Take expectations of both sides to obtain $E(X_n) = E(Y)$, which is finite. For $k < n$,

$$E\big(X_n \mid \mathcal{F}_k\big) = E\big(E(Y \mid \mathcal{F}_n) \mid \mathcal{F}_k\big) = E\big(Y \mid \mathcal{F}_k\big) = X_k\,.$$

Therefore with $X_\infty = Y$, $(X_n \colon n \in \overline{\mathbb{Z}}^+)$ is a martingale with respect to the filtration $(\mathcal{G}_n \colon n \in \overline{\mathbb{Z}}^+)$, where $\mathcal{G}_n = \mathcal{F}_n$ for $n < \infty$ and

$$\mathcal{G}_\infty = \sigma(Y, \mathcal{F}_\infty)\,.$$

To prove that $\{X_n \colon n \in \mathbb{Z}^+\}$ is uniformly integrable we let $A_{n,r} = [|X_n| > r]$ and note that, for any $m > 0$,

$$\begin{aligned} E\big(|X_n|\,;\,A_{n,r}\big) &= E\big(|E(Y \;\big|\; \mathcal{F}_n)|\,;\,A_{n,r}\big) \leq E\big(E(|Y| \;\big|\; \mathcal{F}_n)\,;\,A_{n,r}\big) \\ &= E\big(|Y|\,;\,A_{n,r}\big) \leq mP(A_{n,r}) + E(|Y|\,;\,[|Y| > m])\,. \end{aligned}$$

Since, by dominated convergence, the second term approaches 0 as $m \to \infty$, we can finish the proof of uniform integrability by showing that $P(B_{n,r}) + P(C_{n,r}) \to 0$ as

$r \to \infty$ uniformly in n, where $B_{n,r} = [X_n > r]$ and $C_{n,r} = [X_n < -r]$. That this is so follows from

$$P(B_{n,r}) \le \frac{1}{r} E(X_n \,;\, B_{n,r}) = \frac{1}{r} E(Y \,;\, B_{n,r}) \le \frac{1}{r} E(|Y|)\,,$$

$$P(C_{n,r}) \le -\frac{1}{r} E(X_n \,;\, C_{n,r}) = -\frac{1}{r} E(Y \,;\, C_{n,r}) \le \frac{1}{r} E(|Y|)\,,$$

and the observation that $E(|Y|)$ is a finite number independent of r and n. From the theorem $(X_1, X_2, \dots)$ has an $\mathbf{L}_1$ and a.s. limit Z that is $\mathcal{F}_\infty$ measurable.

To prove that $Z = E(Y \mid \mathcal{F}_\infty)$ we only need show that $E((Z-Y)\,;\, D) = 0$ for every $D \in \mathcal{F}_\infty$. For $D \in \mathcal{F}_n$ we have

$$\begin{aligned} E((Z-Y)\,;\, D) &= E\Big(E((Z-Y)\,;\, D \mid \mathcal{F}_n)\Big) \\ &= E\Big(I_D E((Z-Y) \mid \mathcal{F}_n)\Big) = E(\big(I_D(X_n - X_n)\big) = 0\,, \end{aligned}$$

where I_D denotes the indicator function of D. Thus the desired equality is true for all $D \in \cup_{n=0}^\infty \mathcal{F}_n$, a collection that is closed under finite intersections, contains the entire probability space Ω, and generates $\mathcal{F}_\infty$. By linearity of expectation the set of D for which $E((Y-Z)\,;\, D) = 0$ is closed under proper differences, and, since Y and Z both have means, dominated convergence shows that it is closed under monotone limits. An appeal to the Sierpiński Class Theorem completes the proof.

24-42. The martingale $(V_n \colon n \in \mathbb{Z}^+)$, being bounded, is obviously uniformly integrable. Hence, $\lim V_n$ exists; call this limiting proportion of blue balls V_∞. From the fact that the martingale property is preserved when V_∞ is adjoined to the sequence $(V_n \colon n \in \mathbb{Z}^+)$, we conclude that the expected limiting proportion of blue balls conditioned on the contents of the urn at any particular time is the proportion of blue balls in the urn at that time.

24-45. Let Y be a $(-\infty, 0]$-valued random variable for which $E(Y) = -\infty$. Let $X_n = Y \vee (-n)$. Then $X_n(\omega) \to Y(\omega)$ for every ω. For $n = 0, 1, 2, \dots$, let $\mathcal{G}_n = \sigma(Y)$. Then $(\mathcal{G}_n \colon n = 0, 1, 2, \dots)$ is a reverse filtration to which $(X_n \colon n = 0, 1, 2, \dots)$ is adapted. Clearly $E(X_n) > -\infty$ for every n. The inequality

$$E(X_n \mid \mathcal{G}_{n+1}) = X_n \ge X_{n+1}$$

shows that $(X_0, X_1, \dots)$ is a reverse submartingale.

For Chapter 25

25-1. Define a random sequence T by $T_0 = 0$ and (25.1). Fix a finite sequence $(x_1, \dots, x_{r+s})$ such that $x_r = 1$ and let p denote the number of 1's in this sequence. Define a finite sequence $(t_0, t_1, \dots, t_p)$ by $t_0 = 0$ and

$$t_k = \inf\{m > t_{k-1} \colon x_m = 1\}\,.$$

Then the probability on the left side of (25.2) equals

$$\begin{aligned} P[T_k - T_{k-1} = t_k - t_{k-1} \text{ for } 1 \le k \le p \\ \text{and } T_{p+1} - T_p > r + s - t_p]\,, \end{aligned} \tag{7.11}$$

and, since $t_k = r$ for some k, the probability on the right side of (25.2) equals

$$\begin{aligned} &P[T_k - T_{k-1} = t_k - t_{k-1} \text{ for } k \text{ for which } t_k \leq r] \\ &\cdot P[T_k - T_{k-1} = t_k - t_{k-1} \text{ for } k \leq p \text{ for which } t_k > r \\ &\qquad \text{and } T_{p+1} - T_p > r + s - t_p]\,. \end{aligned}$$

If T is a random walk, then this product equals (7.11), and so (25.2) holds.

For the converse assume that (25.2) holds. *Hint:* To prove that T is a random walk use Proposition 3 of Chapter 11.

25-5. Since the measure generating function of R^{*k} is φ^k we have

$$\begin{aligned} \sum_{n=0}^{\infty} U\{n\} s^n &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} R^{*k}\{n\} s^n = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} R^{*k}\{n\} s^n \\ &= \sum_{k=0}^{\infty} \varphi^k(s) = \frac{1}{1-\varphi(s)} \end{aligned}$$

for $0 \leq s < 1$.

25-8. The function $s \rightsquigarrow 1 + s^2/4(1-s)$ is the measure generating function of the given sequence. Setting this function equal to $1/(1-\varphi)$ gives the formula $\varphi(s) = s^2(2-s)^{-2}$. To show that the given sequence is a potential sequence, we only need show that φ as just calculated is the measure generating function of some probability distribution on $\overline{\mathbb{Z}}^+ \setminus \{0\}$. We will do this by expanding in a power series and checking that all the coefficients are positive, that the coefficient of s^0 is 0, and that $\varphi(1-) \leq 1$. Provided that all the checks are affirmative we will at the same time get a formula for the waiting time distribution R.

Clearly $\varphi(1-) = 1$, so if it develops that there is a corresponding waiting time distribution R, then $R\{\infty\} = 0$. By the Binomial Theorem,

$$\begin{aligned} s^2(2-s)^{-2} = \frac{s^2}{4}\left(1 - \frac{s}{2}\right)^{-2} &= \frac{s^2}{4}\sum_{n=0}^{\infty}\binom{-2}{n}\left(-\frac{s}{2}\right)^n \\ &= \sum_{n=2}^{\infty}\binom{-2}{n-2}\left(-\frac{s}{2}\right)^n = \sum_{n=1}^{\infty}\frac{n-1}{2^n}s^n\,. \end{aligned}$$

Therefore $R\{n\} = (n-1)2^{-n}$ for $n = 1, 2, 3, \dots$.

25-14. *Hint:* Problem 13 may be useful.

25-15. (ii). yes; $U\{0\} = 1$, $U\{1\} = p$, $U\{n\} = p^2$ for $n \geq 2$; $R\{\infty\} = 0$,

$$R\{n\} = p\,\frac{\lambda_+^n - \lambda_-^n}{\lambda_+ - \lambda_-} - p(1-p)\,\frac{\lambda_+^{n-1} - \lambda_-^{n-1}}{\lambda_+ - \lambda_-}\,,$$

where $\lambda_\pm = \frac{1}{2}\left[1 - p \pm \sqrt{(1-p)(1+3p)}\right]$ (It may be of some interest that each $R\{n\}$ is a polynomial function of p.)

(v) no, unless $p = \frac{1}{2}$

(vii) yes; $U\{0\} = 1$, $U\{n\} = 0$ for n odd, $U\{n\} = \binom{n-1}{n/2}[p(1-p)]^{n/2}$ for $n \geq 2$ and even; measure generating function of U:

$$s \rightsquigarrow 1 + \sum_{k=1}^{\infty} \binom{2k-1}{k} [p(1-p)]^k s^{2k} = \frac{1}{2} + \frac{1}{2} \sum_{k=0}^{\infty} \binom{-1/2}{k} [-4p(1-p)s^2]^k$$
$$= \frac{1}{2}\Big[1 + \big(1 - 4p(1-p)s^2\big)^{-1/2}\Big]\,;$$

measure generating function of R:

$$s \rightsquigarrow \frac{1 - 2p(1-p)s^2 - \big(1 - 4p(1-p)s^2\big)^{1/2}}{2p(1-p)s^2} = 2\sum_{k=1}^{\infty} \binom{1/2}{k+1} [-4p(1-p)s^2]^k$$
$$= \sum_{k=1}^{\infty} \frac{1}{k+1}\binom{2k}{k} [p(1-p)]^k s^{2k}\,;$$

$R\{n\} = 0$ for n odd, $R\{n\} = \frac{2}{n+2}\binom{n}{n/2}[p(1-p)]^{n/2}$ for n even, $R\{\infty\} = \frac{|2p-1|}{p \vee (1-p)}$ [Notice that the coefficient $\frac{2}{n+2}\binom{n}{n/2}$ in the formula for $R\{n\}$, n even, is the $(n/2)^{\text{th}}$ Catalan number.]

25-20. for B a set of consecutive integers, $P(N(B) > 0) = 1 - p^{\sharp B}$, in notation of Problem 12

25-29. $\frac{\sigma^2 + \mu(\mu-1)}{2\mu}$, where μ is mean and σ^2 (possibly ∞) is variance

25-36. $R\{n\} = \frac{1}{2n-1}\binom{2n}{n}4^{-n}$, $U\{n\} = \binom{2n}{n}4^{-n}$

25-39. The solution of Problem 28 of Chapter 11 gives the measure generating function of the waiting time distribution for strict ascending ladder times:

$$\varphi^{++}(s) = \frac{1 - \sqrt{1 - 4p(1-p)s^2}}{2(1-p)s}\,.$$

The measure generating function of the waiting time distribution for weak descending ladder times can then be obtained from Theorem 22:

$$\varphi^{-}(s) = \frac{1 + 2(1-p)s - \sqrt{1 - 4p(1-p)s^2}}{2}\,.$$

It is straightforward to use the Binomial Theorem to obtain the waiting time distributions and potential measures corresponding to these two measure generating functions. The other two types of ladder times can be treated by interchanging p and $1-p$.

For Chapter 26

26-5.

$$Q_{n+1}(B) = P[X_{n+1} \in B] = E\big(P([X_{n+1} \in B] \mid \mathcal{F}_n)\big)$$
$$= E\big(\mu_{X_n}(B)\big) = \int \mu_x(B) Q_n(dx) = (Q_n T)(B)$$

$$E(f \circ X_{n+1} \mid \mathcal{F}_n) = \int f(y)\mu_{X_n}(dy) = (Tf) \circ X_n$$

26-19. Let f be the identity function on $[0,1]$. Clearly f is bounded and measurable. By Theorem 6, Y is a martingale where

$$Y_n = X_n - \sum_{k=0}^{n-1} (Gf) \circ X_k \,.$$

Solving for X_n gives a representation for X in terms of the martingale Y and a previsible sequence having the value 0 when $n = 0$. To show that this sequence is increasing, as required for a Doob decomposition, we only need show that Gf is a nonnegative function when f is the identity function. The following calculation does this:

$$Gf(x) = Tf(x) - x = E^x(X_1) - x \geq 0\,,$$

the last equality using the fact that X is a submartingale.

26-28. $x \rightsquigarrow \frac{2}{x+1}$

26-29. *Hint:* Reminder: There is one value of x that is not required to satisfy the difference equation.

26-31. $x \rightsquigarrow e^{-1} \sum_{k=x}^{\infty} \frac{1}{k!}$

26-39. Denote the two states by x and y. By the last part of Problem 38, if one of the two states is transient so is the other. Now suppose that y is null recurrent; our goal is to show that x is not positive recurrent.

By the Renewal Theorem the sequence of entries of T^n in position y along the main diagonal converges to 0 as $n \to \infty$. We will complete the proof by finding an integer k and a positive constant c such that the entry in position y along the main diagonal in T^n is larger than c times the entry in position x along the main diagonal in T^{n-k} for all $n \geq k$, for then it will follow that the sequence of entries in T^{n-k} in position x along the main diagonal will converge to 0 as $n \to \infty$, implying that x is not positive recurrent.

One way to start at y and to then be there again at time n is to first be at state x at some time r, then be at x again at some time $n - k + r$, and then be at state y at time n. By first choosing r and then k appropriately one can make the product of the probabilities of the first and third of these three tasks a positive constant c.

We omit the part of the solution treating the periodicity issue.

26-43. Suppose that, for some k, all entries of T^k are positive. For any states x and y there is positive probability of being at y at time k if the starting state is x. Hence, $\pi_{xy} > 0$. Therefore, T is irreducible. Clearly, $T^m T^k = T^{m+k}$ has only positive entries for all nonnegative integers m, and thus 1 is the greatest common divisor of the powers of T for which the upper left entry (or any other diagonal entry) is positive. Aperiodicity follows.

For the converse suppose that T is irreducible and aperiodic. The sequence of numbers in a fixed diagonal position of $T^0, T^1, T^2, \ldots$ is an aperiodic potential sequence, which, by Lemma 18 of Chapter 25, contains only finitely many zero terms. Thus, there exists an integer m such that all diagonal entries of T^m are positive. Hence, all diagonal entries of T^n are positive for $n \geq m$. Since T is irreducible, there is, for each x and y, an integer k_{xy} such that the entry in row x and column y of $T^{k_{xy}}$ is positive.

Let $k = m + \max\{k_{xy}\}$. Since T^k can be obtained by multiplying $T^{k_{xy}}$ by a power of T at least as large as m, the entry in row x and column y of T^k is positive. Thus, all entries of T^k are positive, as desired.

26-52. starting state of interest denoted by 0; probabilities of absorption at the absorbing states -2 and -1, respectively:

$$\sum_{k=0}^{\infty} \frac{2^{2k-1}}{(3 \cdot 2^{2k-1} - 1)(3 \cdot 2^{2k} - 1)} \quad \text{and} \quad \sum_{k=0}^{\infty} \frac{2^{2k}}{(3 \cdot 2^{2k} - 1)(3 \cdot 2^{2k+1} - 1)}$$

probability of no absorption: 1/3

26-55. We can introduce infinitely many extra transient states in order to obtain a birth-death sequence. The transition distributions μ_x are given by

$$\mu_x\{x-1\} = \frac{x}{b} \wedge 1$$
$$\mu_x\{x+1\} = \frac{b-x}{b} \vee 0\,.$$

From Problem 54 we see the relevance of the following product:

$$\prod_{z=1}^{x} \frac{\frac{b-z+1}{b} \vee 0}{\frac{z}{b} \wedge 1} = \binom{b}{x}.$$

The number r as defined in Problem 54 can now be calculated:

$$r = \sum_{x=0}^{\infty} \binom{b}{x} = \sum_{x=0}^{b} \binom{b}{x} = 2^b\,.$$

The equilibrium distribution Q for the Ehrenfest urn sequence is given by

$$Q\{x\} = \frac{1}{2^b}\binom{b}{x}, \quad 0 \le x \le b\,.$$

For Chapter 27

27-2. μ denotes De Finetti measure; for $i = 1, 2, 3$, $\mu\{y_i\} = \frac{1}{3}$, where $y_i\{1\} = y_i\{6-i\} = \frac{1}{2}$

27-4. De Finetti measure equals delta measure at uniform distribution on $\{x \in \mathbb{Z} \colon 1 \le x \le 12\}$

27-6. Yes. By letting p_i equal the value assigned to the one-point set $\{i\}$ by a probability measure on $\{1, 2, 3, 4\}$, the probability measure itself is represented by an ordered 4-tuple (p_1, p_2, p_3, p_4) The De Finetti measure assigns probability

$\frac{1}{512}$ to $(1, 0, 0, 0)$ and to each of the other 3 permutations thereof;
$\frac{1}{128}$ to $(\frac{3}{4}, \frac{1}{4}, 0, 0)$ and to each of the other 11 permutations thereof;
$\frac{3}{128}$ to $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$ and to each of the other 11 permutations thereof;
$\frac{3}{256}$ to $(\frac{1}{2}, \frac{1}{2}, 0, 0)$ and to each of the other 5 permutations thereof;
$\frac{35}{64}$ to $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

27-10. μ denotes De Finetti measure; $\mu\{m/n\} = P[Z_1 + \cdots + Z_n = m]$.

27-15. $P[X_1 = X_2 = 1] = P[X_1 = X_2 = 0] = \frac{5n-6}{12(n-1)}$
$P[X_1 = -X_2 = 1] = P[X_1 = -X_2 = -1] = \frac{n}{12(n-1)}$

27-31. $\alpha+$ the numbers of 1's, β + the number of 0's

27-32. conditional distribution of (Y, X_{m+1}, X_{m+2}) has density with respect to $\mu \times \gamma \times \gamma$, where γ denotes counting measure on $\{0, 1\}$; density is

$$(p, z_1, z_2) \rightsquigarrow \frac{p^{(X_1+\cdots+X_m+z_1+z_2)}\,(1-p)^{(m+2)-(X_1+\cdots+X_m+z_1+z_2)}}{\int_{[0,1]} x^{(X_1+\cdots+X_m)}\,(1-x)^{m-(X_1+\cdots+X_m)}\,\mu(dx)}\,;$$

integration in p and z_2 gives conditional density with respect to γ of X_{m+1}:

$$z_1 \rightsquigarrow \frac{\int_{[0,1]} p^{(X_1+\cdots+X_m+z_1)}\,(1-p)^{(m+1)-(X_1+\cdots+X_m+z_1)}\,\mu(dp)}{\int_{[0,1]} x^{(X_1+\cdots+X_m)}\,(1-x)^{m-(X_1+\cdots+X_m)}\,\mu(dx)}\,;$$

multiplication by z_1 and integration in z_1 give the conditional expectation of X_{m+1}:

$$\frac{\int_{[0,1]} p^{(X_1+\cdots+X_m+1)}\,(1-p)^{m-(X_1+\cdots+X_m)}\,\mu(dp)}{\int_{[0,1]} x^{(X_1+\cdots+X_m)}\,(1-x)^{m-(X_1+\cdots+X_m)}\,\mu(dx)}\,,$$

which equals

$$\frac{X_1 + \cdots + X_m + 1}{m+2}$$

in case μ is the standard uniform distribution.

27-39. density of each of X_1 and X_2: $x \rightsquigarrow \frac{1}{2}e^{-x} + \frac{1}{4}e^{-x/2}$; density of (X_1, X_2): $(x_1, x_2) \rightsquigarrow \frac{1}{4}\Big(e^{-x_1-x_2/2} + e^{-x_2-x_1/2}\Big)$; De Finetti measure assigns probability 1 to the set of uniform two-point distributions, the density of the two points being $\{y_1, y_2\} \rightsquigarrow \frac{1}{2}\Big(e^{-y_1-y_2/2} + e^{-y_2-y_1/2}\Big)$, $0 < y_1 < y_2$.

27-47. conditional distribution of reciprocal of mean of Y given $(X_1, \ldots, X_m)$ is gamma with main parameter $m+1$ and scaling parameter $1 + \sum_{j=1}^{m} X_j$

27-52. The stick-breaking random walk breaks a stick into random pieces in such a way that, say, the sizes of the first three pieces determines how much of the stick is left for pieces 4, 5, ..., to share but gives no information about the relative sizes of these pieces. Certain information about $(X_1, \ldots, X_m)$ might, for example, give information about the sizes of pieces 1, 2, and 3, without giving information about the relative sizes of the remaining pieces. (Comment: The authors of this book find this explanation to be neither complete nor satisfactory, but it is the best that they could do.)

27-55. The formula is trivial when $k = 0$; it is $1 = 1/1$. Assume it is true for k and multiply both sides by

$$P[X_{k+1} = x_{k+1} \mid X_1 = x_1, \ldots, X_k = x_k] = \frac{c + \gamma_{x_{k+1}}}{k + \sum_{i=1}^{d} \gamma_i}\,,$$

where c equals the number of x_j, $j \le k$, for which $x_j = x_{k+1}$. The result follows.

For Chapter 28

28-4. It suffices to prove that

$$\begin{aligned} &P[(X_m, X_{m+k}, \dots, X_{m+(d-1)k}) \in A] \\ &= P[(X_{m+k}, X_{m+2k}, \dots, X_{m+dk}) \in A] \end{aligned} \tag{7.12}$$

for every positive integer d and every Borel set $A \subseteq \Psi^d$, where Ψ denotes the common target of the X_j. Set

$$B = \{(x_0, x_1, \dots, x_{m+(d-1)k}) \in \Psi^{m+(d-1)k+1} \colon (x_m, x_{m+k}, \dots, x_{m+(d-1)k}) \in A\}\,.$$

Then the left side of (7.12) equals

$$P[(X_0, X_1, \dots, X_{m+(d-1)k}) \in B]$$

and the right side equals

$$P[(X_k, X_{k+1}, \dots, X_{m+dk}) \in B]\,.$$

These are equal by Problem 3.

28-6. *Hint:* From the given sequence obtain the desired joint distributions of every finite set of random variables. Use this information to construct a sequence $(Y_0, Y_{-1}, Y_{-2}, \dots)$ using Theorem 3 of Chapter 22. Then treat $(\dots, Y_{-2}, Y_{-1}, Y_0)$ as a single random object and use it as the first member of a random sequence to be constructed using Theorem 3 of Chapter 22 again, with the next members being $Y_1, Y_2, \dots$.

28-21. Let A be an set for which $R(A) \neq S(A)$. By Problem 18,

$$(I_A\,,\, I_A \circ \tau\,,\, I_A \circ \tau^2\,,\, \dots)$$

is ergodic. By the Birkhoff Ergodic Theorem the sequence

$$\Big(\frac{1}{n}\sum_{k=0}^{n-1} I_A \circ \tau^k \colon n = 1, 2, \dots\Big)$$

converges to $R(A)$ with R-probability 1 and also to $S(A)$ with S-probability 1. Since $S(A) \neq R(A)$, these two events are disjoint, and thus the mutual singularity is established.

28-23. Suppose first that a is rational, say p/q in lowest terms with q positive. Then the following set is easily seen to be shift-invariant and have Lebesgue measure $\frac{1}{2}$:

$$\{x \in [0,1) \colon x \in [\tfrac{p}{q}, \tfrac{2p+1}{2q}) \text{ for some } p\}\,.$$

Now suppose that a is irrational. Rotation through angle $2\pi a$ generates a shift transformation on $[0,1)^\infty$. It is clear that any shift-invariant distribution is determined by the initial distribution on $[0,1)$, but it may be that some choices for that distribution do not yield a shift-invariant measure on $[0,1)^\infty$. In fact, we will prove that the only initial distribution that does yield a shift-invariant measure on $[0,1)^\infty$ is Lebesgue measure.

For every $n \in \mathbb{Z}^+$ and 'left-closed, right-open subinterval' J of $[0,1)$, possibly with 'wrap-around', any shift-invariant measure assigns the same value to J and the interval

J_{na} obtained by adding na to each endpoint of $J \mod n$. For any left-closed, right-open interval K having the same length as J, a sequence $(n_k \in \mathbb{Z}^+ \colon k = 1, 2, \dots)$ can be chosen so that

$$K = \lim_{k \to \infty} J_{n_k\, a}\,.$$

Hence all open intervals having the same length have the same measure, and therefore the only initial distribution that yields a shift-invariant distribution is Lebesgue measure.

Since there is only one shift-invariant distribution, that distribution is extremal and by Theorem 4, therefore ergodic. The Weyl Equidistribution Theorem is then an immediate consequence of the Birkhoff Ergodic Theorem.

28-25. $Q\{i\}T(i,j)$

28-28. Suppose that X is strongly mixing and consider any $A \in \mathcal{T}$. For each n there exists B_n such that $A = \tau^{-n}(B_n)$. As $n \to \infty$,

$$\begin{aligned}\Big|Q(A) - [Q(A)]^2\Big| &= \Big|Q(A \cap \tau^{-n}(B_n)) - Q(A)Q(\tau^{-n}(B_n))\Big| \\ &= \Big|Q(A \cap \tau^{-n}(B_n)) - Q(A)Q(B_n)\Big| \to 0\,.\end{aligned}$$

Therefore $Q(A)$, being a solution of $\Big|Q(A) - [Q(A)]^2\Big| = 0$, equals 0 or 1, as desired.

For the converse we assume that $\mathcal{T}$ is 0-1 trivial and fix a member A of $\mathcal{H}$. Then for all $B \in \mathcal{H}$ and all positive integers n,

$$\begin{aligned}\Big|Q(A \cap \tau^{-n}(B)) - Q(A)Q(B)\Big| &= \Big|E_Q\big(I_A I_{\tau^{-n}(B)} - Q(A) I_{\tau^{-n}(B)}\big)\Big| \\ &= \Big|E_Q\big([Q(A \mid \mathcal{H}_n) - Q(A)]\, I_{\tau^{-n}(B)}\big)\Big| \\ &\le E_Q\Big(\Big|Q(A \mid \mathcal{H}_n) - Q(A)\Big|\Big)\,, \end{aligned} \tag{7.13}$$

where E_Q denotes expectation based on the distribution Q and

$$\mathcal{H}_n = \{\tau^{-n}(C) \colon C \in \mathcal{H}\}\,.$$

To finish the proof we only need show that (7.13) approaches 0 as $n \to \infty$, the uniformity in B resulting from the fact that (7.13) does not depend on B. By the Bounded Convergence Theorem, we only need show

$$\lim_{n \to \infty} \Big[Q(A \mid \mathcal{H}_n) - Q(A)\Big] = 0\,.$$

By the Reverse Martingale Convergence Theorem, this limit does exist and equals $Q(A \mid \mathcal{T}) - Q(A)$, a random variable which has mean 0 and which, since $\mathcal{T}$ is 0-1 trivial, is a.s. constant. Therefore it must equal 0 a.s. as desired.

28-30. NEEDS TO BE DONE.

28-45. Let X denote a stationary Gaussian sequence with correlation function $(m,n) \rightsquigarrow \rho^{|m-n|}$. The result is obvious if $\rho = \pm 1$, so we assume $|\rho| < 1$. Following the hint, the conditional distribution of X_n given $(X_0, X_1, \dots, X_{n-1})$ is Gaussian with a constant variance and mean

$$(7.14)\qquad \begin{pmatrix}\rho & \rho^2 & \dots & \rho^n\end{pmatrix}\begin{pmatrix}1 & \rho & \dots & \rho^{n-1}\\ \rho & 1 & \dots & \rho^{n-2}\\ \vdots & \vdots & & \vdots\\ \rho^{n-1} & \rho^{n-2} & \dots & 1\end{pmatrix}^{-1}\begin{pmatrix}X_{n-1}\\ X_{n-2}\\ \vdots\\ X_0\end{pmatrix}.$$

Since the first matrix is a row matrix that is a multiple of the first column of the matrix whose inverse is in the formula, the matrix product (7.14) is some multiple of X_{n-1}, and this is all that is needed to show that X is Markov.

For Chapter 29

29-5. Let $r = \sum_{j=1}^m r_j$.

$$P[X\{1\}=k] = \frac{\binom{r_1}{k}\binom{r-r_1}{n-k}}{\binom{r}{n}}$$

$$P[X\{1\}=k, X\{2\}=l] = \frac{\binom{r_1}{k}\binom{r_2}{l}\binom{r-r_1-r_2}{n-k-l}}{\binom{r}{n}}$$

$$P[X\{1\}=k, X\{2\}=l, X\{3\}=m] = \frac{\binom{r_1}{k}\binom{r_2}{l}\binom{r_3}{m}\binom{r-r_1-r_2-r_3}{n-k-l-m}}{\binom{r}{n}}$$

29-8. $P[X(B)=z] = \frac{(\sharp B)^z(n-\sharp B)^{r-z}}{n^r}\binom{r}{z}$, $0 \le z \le r$. Thus the distribution of $X(B)$ is binomial with parameters $\frac{\sharp B}{n}$ and r.

29-13. Let $(V_n : n \ge 0)$ be a renewal sequence. Define a random measure X on $\mathbb{Z}^+$ by $X\{n\} = V_n$. Clearly X is a point process and its intensity measure equals the potential measure of the renewal sequence.

29-18. We use the formula for the probability that a Poisson random variable equals 0. For $v \ge 0$,

$$P[V \ge v] = P[X(\{0,1,\dots,v-1\}) = 0] = e^{-v}.$$

Then

$$P[V = v] = P[V \ge v] - P[V \ge (v+1)] = e^{-v} - e^{-(v+1)} = (1-e^{-1})e^{-v}.$$

29-23. Write

$$Y \cup \{0\} = \{0 = Y_0 < Y_1 < Y_2 < \dots\},$$

and let $(S_0 = 0, S_1, S_2, \dots)$ be a random walk having exponentially distributed steps with mean c^{-1}. For an arbitrary positive integer n we will show that $(Y_1,\dots,Y_n)$ and $(S_1,\dots,S_n)$ have the same distribution, thereby finishing the proof. We will verify that the distribution of each of these random vectors has the same density with respect to n-dimensional Lebesgue measure—namely,

$$(7.15)\qquad (y_1,\dots,y_n) \rightsquigarrow \begin{cases} c^n e^{-cy_n} & \text{if } 0 < y_1 < \dots < y_n\\ 0 & \text{otherwise}.\end{cases}$$

To check that this is the correct density for $(Y_1,\dots,Y_n)$ we integrate it over a set of the form $\prod_{i=1}^n [u_i,v_i)$, where

$$0=v_0<u_1<v_1<u_2<\cdots<u_n<v_n=\infty\,.$$

We get

$$\begin{aligned}
e^{-cu_n}\prod_{i=1}^{n-1} c(v_i-u_i) &= \Big(\prod_{i=1}^{n-1} c(v_i-u_i)e^{-c(v_i-u_i)}\Big)\Big(\prod_{i=1}^{n} e^{-(u_i-v_{i-1})}\Big)\\
&= \Big(\prod_{i=1}^{n-1} P[\sharp(Y\cap[u_i,v_i))=1]\Big)\Big(\prod_{i=1}^{n} P[\sharp(Y\cap[v_{i-1},u_i))=0]\Big)\\
&= P[Y_i\in[u_i,v_i) \text{ for } 1\le i\le n]\,,
\end{aligned}$$

as desired.

We know that the density of $\big((S_1-S_0)\,,\,(S_2-S_1)\,,\,\dots\,,\,(S_n-S_1)\big)$ is

$$(x_1,\dots,x_n)\rightsquigarrow\begin{cases}\prod_{i=1}^n ce^{-cx_i} & \text{if each } x_i>0\\ 0 & \text{otherwise}\,.\end{cases}$$

We can get the density of $(S_1,\dots,S_n)$ by using the linear transformation $y_k=x_1+\cdots+x_k$, $1\le k\le n$, the Jacobian of which equals 1. The result is the desired density (7.15).

29-24. *Hint:* One approach is to start with sequences U and V having the desired properties and then use Problem 23 to show that $\{(U_n,V_n)\colon n=1,2,\dots\}$ is a Poisson point process with intensity measure $\lambda\times\mu$.

29-26. c^{-3}

29-29. π, $\frac{\pi}{2}$

29-34. $h\rightsquigarrow\frac{1}{n}\sum_{i=1}^n h(i)$ for $r=1$; $h\rightsquigarrow\frac{1}{n}\big[\sum_{i=1}^n[h(i)]^{-1}\big]\prod_{j=1}^n h(j)$ for $r=n-1$

29-39. $h\rightsquigarrow\exp\big(-\sum_{\psi\in\Psi}(1-h(\psi))\big)$, where Ψ is the countable set

29-43. The probability generating functional of $X+Y$ is

$$\begin{aligned}
h\rightsquigarrow E\Big(\prod_{\psi\in\Psi}[h(\psi)]^{(X+Y)(\{\psi\})}\Big) &= E\Big(\prod_{\psi\in\Psi}[h(\psi)]^{X(\{\psi\})}[h(\psi)]^{Y(\{\psi\})}\Big)\\
&= E\Big(\Big[\prod_{\psi\in\Psi}[h(\psi)]^{X(\{\psi\})}\Big]\Big[\prod_{\psi\in\Psi}[h(\psi)]^{Y(\{\psi\})}\Big]\Big)\\
&= E\Big(\prod_{\psi\in\Psi}[h(\psi)]^{X(\{\psi\})}\Big)\,E\Big(\prod_{\psi\in\Psi}[h(\psi)]^{Y(\{\psi\})}\Big)\,,
\end{aligned}$$

which is the product of the probability generating functionals of X and Y.

29-50. Suppose that $X_n\xrightarrow{\mathcal{D}}X$ as $n\to\infty$; that is, $Q_n\to Q$, where Q_n and Q denote the distributions of X_n and X, respectively. Let h be in the domain of the probability generating functional of Q (and thus of each Q_n). Assume first that h is bounded below by a positive constant. Then the function

$$\pi \rightsquigarrow \int \log(1/h)\, d\pi$$

is continuous, and thus the same is true for the function

$$\pi \rightsquigarrow e^{-\int \log(1/h)\, d\pi}\,. \tag{7.16}$$

For this latter function it is straightforward to remove the assumption that h be bounded below by a positive constant (of course, using the conventions $\infty \cdot 0 = 0$ and $e^{-\infty} = 0$). That

$$\int e^{-\int \log(1/h)\, d\pi}\, dQ_n \to \int e^{-\int \log(1/h)\, d\pi}\, dQ$$

follows from the continuity of the function (7.16). That the limiting probability generating functional has the property described in the theorem is a consequence of Proposition 16 which says that all probability generating functionals have a more general property.

For the converse suppose that $\mathfrak{F}$ is the limit of a sequence of probability generating functionals corresponding to a sequence $(Q_n \colon n = 1, 2, \dots)$ of distributions of point processes in a locally compact Polish space Ψ, and that $\mathfrak{F}$ satisfies the condition in the theorem. Let C be any compact subset of Ψ. By using Lemma 1 one can show that there exists a compact set B such that every point of C is an interior point of B and that therefore there exists a continuous $[(1 - \frac{1}{m}), 1]$-valued function h_m such that $h_m(\psi) = 1 - \frac{1}{m}$ for $\psi \in C$ and $h_m(\psi) = 1$ for $\psi \in B^c$.

Let $\varepsilon > 0$ Since $\mathfrak{F}(h_m) \to 1$ as $m \to \infty$, we can fix m so that for all n

$$\begin{aligned} Q_n\{\pi \colon \pi(C) \leq z\} &\geq \int_{\{\pi \colon \pi(C) \leq z\}} \prod_{\psi} [h_m(\psi)]^{\pi(\{\psi\})}\, Q_n(d\pi) \\ &> 1 - \tfrac{\varepsilon}{2} - \int_{\{\pi \colon \pi(C) > z\}} \prod_{\psi} [h_m(\psi)]^{\pi(\{\psi\})}\, Q_n(d\pi) \\ &\geq 1 - \tfrac{\varepsilon}{2} - \left(1 - \tfrac{1}{m}\right)^{z+1}, \end{aligned}$$

which is larger than $1 - \varepsilon$ for sufficiently large z. By Theorem 19, every subsequence of (Q_n) has a convergent subsequence. By the first paragraph of this proof, $\mathfrak{F}$ is the probability generating functional of every subsequential limit. By Theorem 14 all subsequential limits are identical. Therefore, the sequence (Q_n) itself converges to a limit whose probability generating functional is $\mathfrak{F}$.

For Chapter 30

30-1. $I_{[2+(1/n),\infty)}$ right-continuous; pointwise limit $I_{(2,\infty)}$ not right-continuous at 2.

30-10. The moment generating function is

$$u \rightsquigarrow E\Big(e^{-uY_t}\Big) = E\Big(\exp\Big(-u \sum_{x\in[0,\infty]} xX\big((0,t]\times\{x\}\big)\Big)\Big)$$
$$= E\left(\prod_{\substack{s\in(0,t]\\ x\in[0,\infty]}} \big[e^{-ux}\big]^{X\{(s,x)\}}\right). \tag{7.17}$$

For calculating (7.17), we may replace $(0,t]$ by $[0,t]$. The function $(s,x) \rightsquigarrow e^{-ux}$ is a continuous function on the compact set $[0,t]\times[0,\infty]$, taking the value 0 at (s,∞) if $u>0$ and the value 1 there if $u=0$. Therefore we may apply Proposition 15 of Chapter 29 to conclude that (7.17) equals

$$\exp\Big(-\int_{[0,t]\times[0,\infty]} (1-e^{-ux})\,\kappa(\lambda\times Q)(d(s,x))\Big)$$
$$= \exp\Big(-\kappa t\int_{[0,\infty]} (1-e^{-ux})\,Q(dx)\Big)\,.$$

We could have treated the problem as a single-variable problem by working with the Poisson point process X_t, the restriction of X to $(0,t]\times[0,\infty]$.

In view of Remark 1, Q might be a probability measure on $(0,\infty]$, which is not compact. We could handle this setting, by adjoining 0 to $(0,\infty]$ and specifying $Q\{0\} = 0$, or by approximating $x \rightsquigarrow e^{-ux}$ by continuous functions that equal 1 for small x.

It is not possible to treat characteristic functions by adjoining $\pm\infty$ to $\mathbb{R}$ in order to obtain compactness, because one will then lose continuity. Approximation of the functions $x \rightsquigarrow e^{ivx}$ by functions that are continuous everywhere and constant for large x is a method that works. By then going to the limit one obtains the characteristic function of Y_t:

$$v \rightsquigarrow \exp\Big(-\kappa t\int_{[0,\infty]} (1-e^{ivx})\,Q(dx)\Big)\,.$$

30-13. $1-e^{-t\nu[y,\infty]}$

30-16. Set

$$\widetilde{R}_y(B) = R(\{v\in\mathbf{D}^+[0,1]\colon yv\in B\})\,,$$

and let $0=t_0\le t_1<t_2<\cdots<t_d=1$. The proof relies on showing that

$$P[Z_{t_i}-Z_{t_{i-1}}\le b_i \text{ for } 1\le i\le d]$$
$$= \int_{(0,\infty)} \widetilde{R}_y(\{z\in\mathbf{D}^+[0,1]\colon z_{t_i}-z_{t_{i-1}}\le b_i \text{ for } 1\le i\le d\})\,ae^{-ay}\,dy \tag{7.18}$$

for positive numbers b_i.

The left side of (7.18) equals

$$\prod_{i=1}^d \int_0^{b_i} \frac{a^{(t_i-t_{i-1})}\,\theta_i^{(t_i-t_{i-1})-1}\,e^{-a\theta_i}}{\Gamma(t_i-t_{i-1})}\,d\theta_i$$
$$= a\prod_{i=1}^d \int_0^{b_i} \frac{\theta_i^{(t_i-t_{i-1})-1}\,e^{-a\theta_i}}{\Gamma(t_i-t_{i-1})}\,d\theta_i\,. \tag{7.19}$$

The right side of (7.18) equals

$$\int_0^\infty R(\{v \in \mathbf{D}^+[0,1] \colon v_{t_i} - v_{t_{i-1}} \leq \frac{b_i}{y} \text{ for } 1 \leq i \leq d\})\, ae^{-ay}\, dy\,.$$

From Problem 15, we can rewrite this expression in terms of a Dirichlet distribution:

$$\begin{aligned} a \int_{\substack{0<\rho_i \leq b_i/y\,,\, i\leq(d-1) \\ 0<1-\rho_1-\cdots-\rho_{d-1}\leq b_d/y}} \int_0^\infty e^{-ay} &\frac{(1-\rho_1-\cdots-\rho_{d-1})^{(t_d-t_{d-1})-1}}{\Gamma(t_d-t_{d-1})} \\ &\cdot \prod_{i=1}^{d-1} \frac{\rho_i^{(t_i-t_{i-1})-1}}{\Gamma(t_i-t_{i-1})}\, dy\, d\rho_{d-1}\, \ldots\, d\rho_1\,. \end{aligned} \tag{7.20}$$

For $1 \leq i \leq d-1$, let $\theta_i = y\rho_i$, and also let $\theta_d = y(1-\rho_1-\cdots-\rho_{d-1})$. The Jacobian of this transformation is $\frac{d(\theta_1,\ldots,\theta_{d-1},\theta_d)}{d(\rho_1,\ldots,\rho_{d-1},y)} = y^{d-1}$; hence this change of variables turns (7.20) into (7.19), as desired.

30-25. negative binomial with parameters $1/\big(1+E(Y_1)\big)$ and $\tau E(Z_1)$

30-31. (iii): Let $\varepsilon > 0$, and denote the distribution of Z_t by Q_t. Then for $\varepsilon t < 1$,

$$\begin{aligned} \big(1-e^{-1}\big) P[Z_t > \varepsilon t] &\leq \int_{(\varepsilon t,\infty)} \big(1-e^{-x/(\varepsilon t)}\big)\, Q_t(dx) \\ &\leq 1 - \exp\Big(-t \int_{(0,1]} \big(1-e^{-y/(\varepsilon t)}\big)\, \nu(dy)\Big) \\ &\leq t \int_{(0,1]} \big(1-e^{-y/(\varepsilon t)}\big)\, \nu(dy)\Big) \\ &\leq \varepsilon^{-1} \int_{(0,\varepsilon t]} y\, \nu(dy) + t \int_{(\varepsilon t,1]} \nu(dy)\,. \end{aligned} \tag{7.21}$$

The first term in (7.21) goes to 0 as $t \searrow 0$. To treat the second term, let $\delta > 0$ and choose $r \in (0,1)$ so that $\int_{(0,r]} s\, \nu(ds) < \delta$. Then as $t \searrow 0$,

$$t \int_{(\varepsilon t,1]} \nu(dy) \leq \int_{(\varepsilon t,r]} s\, \nu(ds) + t\nu(r,1] \to \int_{(0,r]} s\, \nu(ds) < \delta\,.$$

Since δ is an arbitrary positive number, it follows that

$$\lim_{t\searrow 0}\ t \int_{(\varepsilon t,1]} \nu(dy) = 0\,,$$

as desired. *Hint:* for (vi): For any $t \in (0,1]$ there exists a nonnegative integer n such that $t > 2^{-n-1}$ and

$$\frac{Z_t}{t} \leq 2^{n+1} Z_{2^{-n}}\,.$$

30-32. The carelessness might be ignoring the term 'almost' in the phrase ' a.s.'.

30-35. in case $\alpha < 1$, 0 or ∞ according as $\beta < \frac{1}{\alpha}$ or $\beta \geq \frac{1}{\alpha}$; in case $\alpha = 1$, 0 if and only if $\beta < 1$, and ∞ if and only if $\beta > 1$

For Chapter 31

31-2. For the last assertion one may for each ω, view Q as a probability measure on $(\mathbf{D}([0,\infty),\Psi),\mathcal{H})$. Then Q_t is the distribution of the Ψ-valued random variable $\varphi \rightsquigarrow \varphi_t$ defined on the probability space $(\mathbf{D}([0,\infty),\Psi),\mathcal{H},Q)$. Since $\varphi_u \to \varphi_t$ as $u \searrow t$ and almost sure (in this case sure) convergence implies convergence in distribution, $Q_u \to Q_t$ as $u \searrow t$ (for each ω, not just the requested 'a.s.').

31-3. (i) for all $x \in \Psi$, $\mu_{x,t} \to \mu_{x,0} = \delta_x$ as $t \searrow 0$; (ii) for all Borel $A \subseteq \Psi$ and $t \geq 0$, the function $x \rightsquigarrow \mu_{x,t}(A)$ is measurable; (iii) for all Borel $A \subseteq \Psi$, $s,t \geq 0$, and $x \in \Psi$,

$$\mu_{x,s+t}(A) = \int_\Psi \mu_{y,s}(A)\,\mu_{x,t}(dy)\,.$$

31-9. Let R_t denote the distribution of the Lévy process at time t. Then

$$T_t f(x) = \int f(x+y)\,R_t(dy)\,.$$

Let R denote the distribution of the Lévy process. Then the corresponding Markov family $(Q^x : x \in \mathbb{R})$ is defined by

$$Q^x(B) = R\big(\{\varphi\colon [t \rightsquigarrow (x+\varphi_t)] \in B\}\big)\,.$$

31-15. $Gf(x) = \kappa \int (f(y) - f(x))\,Q(dy)$, in the notation of Example 1 of Chapter 30.

31-21. Suppose that Q_0 is an equilibrium distribution for $\widetilde{T}$. Then

$$Q_0 T_t = e^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^k}{k!}Q_0\widetilde{T}^k = e^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^k}{k!}Q_0 = Q_0\,.$$

Hint: for converse: Use Problem 16.

31-23. *Hint:* Let f be the indicator function of the one-point set $\{y\}$ and use Theorem 14.

31-25.

$$\begin{aligned}
p_{00}(t) &= (q_{01}+q_{10})^{-1}\big[q_{10} + q_{01}\exp[-(q_{01}+q_{10})t]\big]\\
p_{01}(t) &= (q_{01}+q_{10})^{-1}\,q_{01}\big(1-\exp[-(q_{01}+q_{10})t]\big)\\
p_{10}(t) &= (q_{01}+q_{10})^{-1}\,q_{10}\big(1-\exp[-(q_{01}+q_{10})t]\big)\\
p_{12}(t) &= (q_{01}+q_{10})^{-1}\big(q_{01} + q_{10}\exp[-(q_{01}+q_{10})t]\big)
\end{aligned}$$

The limits at ∞ of both p_{00} and p_{10} are the same: $(q_{01}+q_{10})^{-1}q_{10}$, the value the equilibrium distribution assigns to $\{0\}$. The limits at ∞ of both p_{01} and p_{11} are the same: $(q_{01}+q_{10})^{-1}q_{01}$, the value the equilibrium distribution assigns to $\{1\}$.

31-28. The solution to Problem 23 involves applying Theorem 14 to the indicator functions of one-point sets. When the rates are unbounded, such functions may not be in the domain of the generator. For example, let the state space be $\mathbb{Z}^+$, let the transition rates q_{xy} have the property that $q_{x0} \to \infty$ as $x \to \infty$, and let f be the indicator function of $\{0\}$. Then f is not in the domain of the infinitesimal generator because the limit in the definition does not exist boundedly, and Theorem 14 does not apply. Nevertheless, it can be shown that (31.12) holds whenever the state space is countable, even in the case of unbounded rates.

31-29. Let M be the largest member of the support of ρ, x_0 the initial state, and U_n the time of the n^{th} jump. The construction ensures that $X_{U_n} \le x_0 + (M-1)n$. Therefore, conditioned on $\mathcal{F}_{U_{n-1}}$, $U_n - U_{n-1}$ is exponential with mean at least $1/\gamma(x_0 + (M-1)(n-1))$. An inductive argument based on this fact shows that for each n, the distribution function of U_n is bounded above by the distribution function of the sum of n independent exponentially distributed random variables with means $1/\gamma x_0$, $1/(\gamma(x_0+M-1))$, ... ,$1/(\gamma(x_0+(M-1)(n-1))$. Such a sum of exponentially distributed random variables diverges almost surely as $n \to \infty$ by the Kolmogorov Three-Series Theorem. It follows that $U_n \to \infty$ a.s. as $n \to \infty$.

31-36. $\frac{d\nu}{d\lambda}(y) = b(1-b)ce^{-(1-b)cy}$, $\nu\{\infty\} = 0$; equilibrium distribution assigns value $(1-b)b^x$ to x; jump-rate function is

$$x \rightsquigarrow \begin{cases} cb^{-1} & \text{if } x = 0 \\ c & \text{if } x > 0\,; \end{cases}$$

transition probabilities from x to $x-1$ equal 1 for $x > 0$ and from 0 to $x > 0$ equal $(1-b)b^{x-1}$; transition rates from x to $x-1$ equal c for $x > 0$ and from 0 to $x > 0$ equal $c(1-b)b^x$ and all others equal 0

For Chapter 32

32-1.

$$(\xi,\eta) \rightsquigarrow \begin{cases} b_x(\xi) & \text{if } \eta = \xi^x \\ d_x(\xi) & \text{if } \eta = {}_x\xi \\ j_{xy}(\xi) & \text{if } \eta = {}_x\xi^y \\ 0 & \text{otherwise} \end{cases}$$

32-7. For $\xi \in \Xi$, let $X^{(\xi)}$ be the process defined in the construction with initial state ξ. The discussion in the paragraph following the proof of Theorem 2 shows that for each time $t \ge 0$, the function $\xi \rightsquigarrow X_t^{(\xi)}$ is almost surely a continuous function. It follows from the Bounded Convergence Theorem that, for any continuous function $f\colon \Xi \to \mathbb{R}$, the function $\xi \rightsquigarrow E(f \circ X_t^{(\xi)})$ is continuous. Thus, the transition semigroup is Feller.

32-10. Here is one way to make a correct 'if and only if' statement: Let G and $G^{(k)}$ be as in the first sentence of Problem 9. The 'if' statement is: If $G^{(k)}f \to Gf$ pointwise as $k \to \infty$ for all $f \in \mathfrak{F}$, then $X_t^{(k)} \to X_t$ as $k \to \infty$, uniformly for t in bounded subsets of $[0,\infty)$ and for all choices of initial states $\xi^{(k)}$ and ξ such that $\xi^{(k)} \to \xi$. The 'only if' statement is: If there exists a function $f \in \mathfrak{F}$ and a state η such that $G^{(k)}f(\eta)$ does not converge to $Gf(\eta)$ as $k \to \infty$, then there exists a time $t > 0$ and a sequence of initial states $\xi^{(k)}$ converging to a state ξ as $k \to \infty$ such that $X_t^{(k)}$ does not converge to X_t as $k \to \infty$. (In this second statement, we may take $\xi^{(k)} = \xi = \eta$ for all k and let t be any sufficiently small positive time.)

To prove the 'if' statement, it is enough to show that for any site x and any time $t \geq 0$, there exists a nonnegative random variable K that is almost surely finite such that $X_s^{(k)}(x) = X_s(x)$ for all $k \geq K$ and $s \in [0,t]$. This last statement is a slightly stronger version of the statement made in the paragraph immediately following the proof of Theorem 2. To prove this stronger statement, first note that since $G^{(k)}f \to Gf$ for $f \in \mathfrak{F}$ as $k \to \infty$, each rate in the system with infinitesimal generator $G^{(k)}$ converges uniformly as $k \to \infty$ to the corresponding rate in the system with infinitesimal generator G. Now consider the construction of $X^{(k)}$ and X using the universal coupling. Let A be as in the statement following the proof of Theorem 2 and let K be large enough so that $\xi^{(k)}$ agrees with ξ at sites in A for $k \geq K$. We can also choose K large enough so that the rates of $G^{(k)}$ at sites in A are uniformly as close as we like to the corresponding rates of G when $k \geq K$. A simple modification of the proof of Theorem 2 shows that we can thereby make the probability arbitrarily close to 1 that the processes $X^{(k)}$ and X take the same values at x at all times in $[0,t]$. Further details are left to the reader.

The hypothesis in the 'only if' statement implies that there exists a site x such that at least one of the rates at x for the process with infinitesimal generator G is not the pointwise limit as $k \to \infty$ of the corresponding rates for the processes with infinitesimal generators $G^{(k)}$. It follows that there exist arbitrarily large integers k and a state η such that the process with infinitesimal generator $G^{(k)}$ and initial state η will not behave the same at the site x as the process with infinitesimal generator G and initial state η, at least for short time periods. Once again, the details are left to the reader.

32-13. (This problem is incorrectly stated in the book. The statement is not true for the contact process with threshold birth rates. Also, a stronger statement is proved for the contact process with sexual reproduction in Problem 12. So the problem should only be done for the contact process of Example 1.) For finite sets $A \subseteq \mathbb{Z}^d$, let

$$f_A(\xi) = \sum_{x \in A} \xi(x)\,.$$

Direct calculation shows that if ξ is a state with only finitely many occupied sites, then

$$Gf_A(\xi) \leq (1-\delta) f_A(\xi)\,, \tag{7.22}$$

provided A is chosen large enough to include all x such that $\xi(x) = 1$.

Let ξ_0 be a state with only finitely many occupied sites, and let (X_t) be the interacting particle system with initial state ξ_0 and infinitesimal generator G. For each finite set $A \subseteq \mathbb{Z}^d$, define a random time σ_A by

$$\sigma_A = \inf\{t \geq 0\colon X_t(x) = 1 \text{ for some } x \notin A\}\,.$$

Also, let

$$\tau = \inf\{t \geq 0\colon X_t = \bar{0}\}\,.$$

Since the interacting particle system is a solution to the martingale problem for G, it follows from (7.22) and the Optional Sampling Theorem that for any time $t \geq 0$,

$$E(f_A(X_{t\wedge\sigma_A\wedge\tau})) - f_A(\xi_0) \leq E\Big(\int_0^{t\wedge\sigma_A\wedge\tau} (1-\delta) f_A(X_s)\, ds\Big)\,.$$

Since $\delta > 1$, the integrand on the right side is bounded above by $(1-\delta)$ for all $s < \sigma_A$, so

$$E(f_A(X_{t\wedge\sigma_A\wedge\tau})) - f_A(\xi_0) \le (1-\delta)E(t\wedge\sigma_A\wedge\tau)\,,$$

from which it follows immediately that

$$f_A(\xi_0) \ge (\delta-1)E(t\wedge\sigma_A\wedge\tau)\,.$$

We leave it to the reader to check that $\sigma_A \nearrow \infty$ a.s. as $A \nearrow \mathbb{Z}^d$. Thus, after first letting $A \nearrow \mathbb{Z}^d$ and then letting $t \nearrow \infty$, we have by the Monotone Convergence Theorem that

$$\sum_{x\in\mathbb{Z}^d} \xi_0(x) \ge (\delta-1)E(\tau)\,.$$

Since ξ_0 has only finitely many occupied sites, the left side of this inequality is finite. It follows that τ has finite expectation, and hence that τ is finite almost surely, as desired.

32-16. It is easily checked that for each site x, the process $(X_t(x), t \ge 0)$ is a pure-jump Markov process with state space $\{0,1\}$, transition rates $q_{01} = 1$ and $q_{10} = 2^{|x|}$, and initial state 0. It follows from Problem 25 of Chapter 31 that

$$P[X_t(x) = 1] < 2^{-|x|}\,.$$

By the Borel Lemma, $\sum_x X_t(x)$ is finite a.s. Thus, for any fixed time t, the number of occupied sites at time t is finite a.s.

For the second part of the problem, we fix $t \in (0,\infty)$. We know from the previous part of the problem that at any given time s there are infinitely many vacant sites. Since the birth rates are all equal to 1 at vacant sites, it is not hard to show that, with probability 1, infinitely many births occur during every time interval of positive length. In particular, infinitely many births occur with probability 1 during the time interval $(0,t)$. Let

$$x_1 = \min\{x > 0\colon \text{ there is a birth at } x \text{ during } (0,t)\}\,.$$

Let U_1 be the time of the first birth at x_1 and V_1 the time of the first death at x_1.

We now proceed by induction. We assume that random sites $x_1,\dots,x_n$ have been defined for some $n \ge 1$, with corresponding random times $U_1,\dots,U_n$ and $V_1,\dots,V_n$, where for each $k = 2,\dots,n$, U_k is the time of the first birth at x_k after time U_{k-1}, and V_k is the time of the first death at x_k after time U_k. Note that $U_1 < U_2 < \cdots < U_n$. As part of the induction, we also assume that $U_n < t\wedge V_1\wedge\cdots\wedge V_n$. This assumption implies that the time interval $(U_n, t\wedge V_1\wedge\cdots\wedge U_n)$ has positive length, so the following random site is almost surely defined:

$$x_{n+1} = \min\{x > 0\colon \text{ there is a birth at } x \text{ during } (U_n, t\wedge V_1\wedge\cdots\wedge V_n)\,.$$

Let U_{n+1} be the time of the first birth at x_{n+1} after time U_n, and V_{n+1} the time of the first death at x_{n+1} after time U_{n+1}. Note that our construction ensures that $U_{n+1} < t\wedge V_1\wedge\cdots\wedge V_{n+1}$, as required by the assumption made in the inductive step.

Let $U = \lim_{n\to\infty} U_n$. Our construction of U shows that U is defined almost surely, and that when it is defined, $U \le t$. This construction also shows that $X_{U-}(x_n) = 1$ for all $n = 1,2,\dots$. Our construction of the process (X_t) shows that, with probability 1, at most one death can occur at time U, so infinitely many sites are occupied at time U, as desired.

32-18. $j_{xy}(\xi) = \rho\{y-x\}$ if $\xi(y) = 0$; $d_x(\xi) = \sum_y \xi(y)\rho\{y-x\}$; other rates are 0

32-24. TO BE DONE

For Chapter 33

33-2. *Hint:* Let $(\mathcal{F}_t : t \geq 0)$ denote the minimal filtration of the Wiener process W. Square both sides of (33.1) and then take expectations. Six terms result on the right side. The following calculation shows that one of them is equal to 0:

$$E\Big(Z_{n\varepsilon}a(Z_{n\varepsilon})(W_{(n+1)\varepsilon} - W_{n\varepsilon})\Big) = E\Big(E\big(Z_{n\varepsilon}a(Z_{n\varepsilon})(W_{(n+1)\varepsilon} - W_{n\varepsilon}) \bigm| \mathcal{F}_{n\varepsilon}\big)\Big)$$
$$= E\Big(Z_{n\varepsilon}a(Z_{n\varepsilon})E\big((W_{(n+1)\varepsilon} - W_{n\varepsilon}) \bigm| \mathcal{F}_{n\varepsilon}\big)\Big) = 0\,.$$

Similarly,

$$E\big(b(Z_{n\varepsilon})\varepsilon a(Z_{n\varepsilon})(W_{(n+1)\varepsilon} - W_{n\varepsilon})\big) = 0\,.$$

The following calculation is relevant for another of the six terms:

$$E\Big([a(Z_\varepsilon)]^2(W_{(n+1)\varepsilon} - W_{n\varepsilon})^2\Big) = E\Big(E\big([a(Z_\varepsilon)]^2(W_{(n+1)\varepsilon} - W_{n\varepsilon})^2 \bigm| \mathcal{F}_{n\varepsilon}\big)\Big)$$
$$= E\Big([a(Z_\varepsilon)]^2E\big((W_{(n+1)\varepsilon} - W_{n\varepsilon})^2 \bigm| \mathcal{F}_{n\varepsilon}\big)\Big) = \varepsilon E\Big([a(Z_\varepsilon)]^2\Big)\,.$$

33-5. yes

33-12. $d(e^{\alpha W}) = \alpha e^{\alpha W}\,dW + \frac{1}{2}\alpha^2 e^{\alpha W}\,dt$

33-15. TO BE DONE

33-17. For $z \in \mathbb{R}$, let $Z^{(z)}$ denote the solution of (33.19) with initial state z, and let $(T_t, t \geq 0)$ denote the corresponding transition semigroup. Since $T_t f(z) = E(f \circ Z_t^{(z)})$, the Bounded Convergence Theorem implies that it is enough to show that for each $t \geq 0$ and $z \in \mathbb{R}$, $\lim_{y\to z} Z_t^{(y)} = Z_t^{(z)}$ a.s. In the proof of Theorem 7 it is shown that each random variable $Z_t^{(y)}$ is the limit in probability of random variables $Z_t^{(y,\varepsilon)}$ as $\varepsilon \searrow 0$. From the definitions it is apparent that $y \rightsquigarrow Z^{(y,\varepsilon)}$ is almost surely a continuous function for each $\varepsilon > 0$. Thus, it is enough to show that

$$\lim_{\varepsilon,\eta\searrow 0}\sup_{y\in\mathbb{R}}|Z^{(y,\varepsilon)} - Z^{(y,\eta)}| = 0\,. \tag{7.23}$$

Noting that the estimates used in the proof of Theorem 7 do not depend on the initial value y, we see that, with minor modifications, the argument in that proof can be used to give (7.23).

33-29. $Gf = \frac{1}{2}\Delta f$ for sufficiently nice functions $f\colon \mathbb{R}^d \to \mathbb{R}$. For bounded continuous functions f having bounded continuous first, second, and third partial derivatives, this fact can be proved by direct computation, using the second degree Taylor polynomial approximation of f with remainder.

For Appendix A

A-2. The derivative $x \rightsquigarrow 1 - \cos x$ is positive for $-2\pi < x < 0$ and also for $0 < x <$

2π. A theorem of calculus says that a continuous function on a closed interval that has a positive derivative at all interior points of that interval is strictly increasing on the closed interval. Therefore the given function is strictly increasing on the interval $[-2\pi, 0]$ and on the interval $[0, 2\pi]$. By the preceding problem it is strictly increasing on the interval $[-2\pi, 2\pi]$. (Notice that the argument can be extended to prove that the given function is strictly increasing on $\mathbb{R}$.)

For Appendix B

B-1. Proof that a closed subset of a compact set is compact. Let B be a closed subset of a compact set C, and let $\mathcal{O}$ be an open covering of B. Consider $\mathcal{O} \cup \{B^c\}$, the collection obtained by adjoining the complement of B to the collection $\mathcal{O}$. This collection is an open covering of C. It contains a finite subcovering of C. The members of $\mathcal{O}$ in this finite subcovering of C constitute a finite subcovering (from $\mathcal{O}$) of B.

B-5. The 'only if' part is trivial. We will prove the contrapositive of the 'if part', so suppose that the sequence does not converge to y. Then there exists $\varepsilon > 0$ and an infinite subsequence $(x_{n_k} : k = 1, 2, \dots)$ of (x_n) such that $\rho(x_{n_k}, y) > \varepsilon$ for all k. No further subsequence of this subsequence can converge to y because the distance between y and every member of that further subsequence would be greater than ε.

For Appendix C

C-5. Suppose that $x \in \partial B$. Case 1, $x \in B$: Every neighborhood of x contains a member of B —namely x itself. If some neighborhood did not contain a member of B^c, then x would be a member an open subset of that neighborhood which itself would be a subset of B. Hence x would belong to the interior of B and thus not to ∂B.

Case 2, $x \notin B$: Now we must show that every neighborhood of x contains a member of B. If there were some neighborhood lying entirely inside B^c, there would be an open subset of that neighborhood containing x and having the same property. The complement of that open set would be a closed set containing B and thus containing the closure of B. Therefore x would not belong to ∂B.

For the converse suppose that every neighborhood of x contains at least one point of B and least one point in B^c. First we observe that x cannot be a member of the interior of B, for, if it were, this interior would be a neighborhood of x that contains no member of B^c. To finish the proof we must show that x belongs to the closure of B. If it did not, the complement of the closure of B would be a neighborhood of x containing no point of B, which is a contradiction.

C-6. *Hint:* Avoid doing work similar to that needed for the preceding problem.

C-9. $[a, b)$, both open and closed whether $b < \infty$ or $b = \infty$; $(a, b]$, neither open nor closed whether $a > -\infty$ or $a = -\infty$; $[a, b]$ closed but not open; (a, b) open but not closed whether a and b are finite or infinite; $[a, a]$ is only compact interval

C-10. Closure under arbitrary unions: clearly yes if all sets in the union belong to $\mathcal{O}$; if one of the sets in the union contains ∞ and has a complement that is a compact subset C of Ω, the union will contain ∞ and have a complement that is closed subset of the compact subset C of Ω. An appeal to Proposition 1 completes this portion of the proof.

Closure under finite intersections: clearly yes if one of the sets in the intersection does not contain ∞; if all do contain ∞, then so does the intersection and the complement of the intersection is the union of a finite number of compact subsets of Ω. The definition of compactness shows that a finite union of compact sets is compact.

Compactness: An open covering must have at least one set that contains ∞. Take any such set O. The remaining sets in the open covering cover the compact complement of O. Thus there is a finite subcovering of this complement. Adjoin O to this finite subcovering to obtain a finite subcovering of Ω^*.

C-14. The closed interval $[0,1]$ of $\mathbb{R}$ with the usual topology is not open in that topology, but it is an open subset of the topological space $[0,1]$ with the relative topology.

Now assume that $\Psi \in \mathcal{O}$ and that $O \subset \Psi$ is open in the relative topology on Ψ. Then $O = A \cap \Psi$ for some $A \in \mathcal{O}$. Hence, O, the intersection of two members of $\mathcal{O}$, is itself a member of $\mathcal{O}$.

For Appendix D

D-1. 30

D-2. $\frac{3}{11}$

D-14. According to Theorem 4 we only need prove that f is Riemann-Stieltjes integrable with respect to g, and for doing that, Proposition 2 says that we only need prove that f is bounded and fg' is Riemann integrable.

Suppose that f is unbounded. For each m there exists $x_m \in [a,b]$ such that $|f(x_m)| > m$. Let x denote a limit of a subsequence of (x_m). It cannot be that infinitely many members of the subsequence equal x. If infinitely many members are larger than x, then $f(x+)$ does not exist. If infinitely many members are smaller than x, then $f(x-)$ does not exist. Therefore the assumption that f is unbounded leads to a contradiction, and hence f is bounded.

For future use we show that for each $\delta > 0$, there exists only finitely many x such that

$$f(x-) \vee f(x) \vee f(x+) > \delta + f(x-) \wedge f(x) \wedge f(x+)\,.$$

If there were infinitely many, then at the limit y of a convergence sequence of distinct such x, either $f(y+)$ or $f(y-)$ would fail to exist.

Turning to the proof of Riemann integrability of fg', we let $\varepsilon > 0$. For each $x \in [a,b]$ let J_x be an open interval in $[a,b]$ such that

- $x \in J_x$,
- $|f(y) - f(x+)| < \frac{\varepsilon}{4(b-a)}$ if $x < y \in J_x$,
- $|f(y) - f(x-)| < \frac{\varepsilon}{4(b-a)}$ if $x > y \in J_x$.

(Reminder: Intervals in $[a,b]$ including the endpoint a or b can be open in the relative topology of $[a,b]$. Alternatively, we could have let J_a and J_b be open intervals in $\mathbb{R}$ containing members outside the interval $[a,b]$.) Since $[a,b]$ is compact there exists a finite collection of intervals J_x whose union equals $[a,b]$. Let $\widehat{P}$ be the point partition of

$[a,b]$ consisting of the endpoints of the intervals in this finite collection and the points midway between two consecutive endpoints.

For each point x for which

$$f(x-) \vee f(x) \vee f(x+) > \tfrac{\varepsilon}{4(b-a)} + f(x-) \wedge f(x) \wedge f(x+)\,,$$

of which there are only finitely many—say q —introduce a close interval $K_x \subseteq [a,b]$ containing x as an interior point and having length less than $\frac{\varepsilon}{4qs}$, where s denotes the supremum of $|f(x)g'(x)|$ for $x \in [a,b]$. Let P denote the point partition of $[a,b]$ obtained by adjoining the endpoints of each such K_x to $\widehat{P}$.

Consider any refinement P' of P. For any Riemann sum of fg' corresponding to P', the total contribution arising from intervals lying in the various K_x is less than $\varepsilon/4$. The contributions to any two such Riemann sums arising from other intervals differ by less than $3\varepsilon/4$. Thus any two Riemann sums of any refinement of P differ by less than ε.

Now a straightforward argument using a sequence of refinements corresponding to a decreasing sequence (ε_k) gives a Cauchy sequence of Riemann sums. Then the above argument can be used again to show that the limit of this Cauchy sequence is the value of the Riemann integral, and thus in particular, that the Riemann integral of fg' exists.

Comment: For those whose definition of Riemann integrals involves upper and lower integrals and sums rather than Riemann sums, the above argument can be shortened a bit. We have not adopted the 'upper-lower' approach because it does not generalize nicely to the Riemann-Stieltjes setting.

For Appendix E

E-4. We consider the real part of $\exp\circ\lambda$:

$$(\Re \circ \exp\circ\lambda) = (\exp\circ\Re\circ\lambda)\cdot(\cos\circ\Im\circ\lambda)\,.$$

Using the Product Rule and Chain Rule for $\mathbb{R}$-valued functions we obtain

$$\begin{aligned}(\Re\circ\beta)' &= (\exp\circ\Re\circ\lambda)\cdot(\Re\circ\lambda)'\cdot(\cos\circ\Im\circ\lambda)\\ &\qquad -(\exp\circ\Re\circ\lambda)\cdot(\sin\circ\Im\circ\lambda)\cdot(\Im\circ\lambda)'\\ &= (\Re\circ\lambda')\cdot(\Re\circ\exp\circ\lambda) - (\Im\circ\lambda')\cdot(\Im\circ\exp\circ\lambda)\\ &= \Re\circ\big(\lambda'\cdot(\exp\circ\lambda)\big)\,,\end{aligned}$$

as desired. We omit the similar calculation relevant for the imaginary part.

E-9. no