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

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

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 $m = 1, 2, \ldots, n$, 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

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

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 $x \rightsquigarrow x^ne^{-x}$, 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 $v \rightsquigarrow \#\{i\colon x_i \le v\}$ has a jump of size $\#\{i\colon x_i = v\}$, 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

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

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

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

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.

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

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

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

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

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

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

$$E(Y_1 Y_2) = \frac{\gamma_1 \gamma_2}{\gamma(\gamma+1)} .$$

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

$$\mathrm{Var}(Y_1 + \cdots + Y_d) = \sum_{j=1}^{d} \sum_{i=1}^{d} \mathrm{Cov}(Y_i Y_j)$$

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 $z \leq 0$ and $F(z) = 1$ for $z \geq \frac{1}{3}$. 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

$$\{r_1 x_1 + r_2 x_2 + r_3 x_3 \colon x_i \in C_i, r_i \geq 0, r_1 + r_2 + r_3 = 1\}$$

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 $-1 \leq w \leq 1$, 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.