Probability: Theory and Examples
Solutions Manual

The creation of this solution manual was one of the most important improvements in the second edition of Probability: Theory and Examples. The solutions are not intended to be as polished as the proofs in the book, but are supposed to give enough of the details so that little is left to the reader's imagination. It is inevitable that some of the many solutions will contain errors. If you find mistakes or better solutions send them via e-mail to rtd1@cornell.edu or via post to Rick Durrett, Dept. of Math., 523 Malott Hall, Cornell U., Ithaca NY 14853.

Rick Durrett

Contents

Appendix: Measure Theory 108

1 Laws of Large Numbers

1.1. Basic Definitions

1.1. (i) A and $B-A$ are disjoint with $B = A\cup(B-A)$ so $P(A)+P(B-A) = P(B)$ and rearranging gives the desired result.
(ii) Let $A'_n = A_n \cap A$, $B_1 = A'_1$ and for $n > 1$, $B_n = A'_n - \cup_{m=1}^{n-1} A'_m$. Since the B_n are disjoint and have union A we have using (i) and $B_m \subset A_m$

$$P(A) = \sum_{m=1}^{\infty} P(B_m) \le \sum_{m=1}^{\infty} P(A_m)$$

(iii) Let $B_n = A_n - A_{n-1}$. Then the B_n are disjoint and have $\cup_{m=1}^{\infty} B_m = A$, $\cup_{m=1}^{n} B_m = A_n$ so

$$P(A) = \sum_{m=1}^{\infty} P(B_m) = \lim_{n\to\infty} \sum_{m=1}^{n} P(B_m) = \lim_{n\to\infty} P(A_n)$$

(iv) $A_n^c \uparrow A^c$ so (iii) implies $P(A_n^c) \uparrow P(A^c)$. Since $P(B^c) = 1 - P(B)$ it follows that $P(A_n) \downarrow P(A)$.

1.2. (i) Suppose $A \in \mathcal{F}_i$ for all i. Then since each $\mathcal{F}_i$ is a σ-field, $A^c \in \mathcal{F}_i$ for each i. Suppose $A_1, A_2, \ldots$ is a countable sequence of disjoint sets that are in $\mathcal{F}_i$ for all i. Then since each $\mathcal{F}_i$ is a σ-field, $A = \cup_m A_m \in \mathcal{F}_i$ for each i.
(ii) We take the interesection of all the σ-fields containing $\mathcal{A}$. The collection of all subsets of Ω is a σ-field so the collection is not empty.

1.3. It suffices to show that if $\mathcal{F}$ is the σ-field generated by $(a_1, b_1)\times\cdots\times(a_n, b_n)$, then $\mathcal{F}$ contains (i) the open sets and (ii) all sets of the form $A_1 \times \cdots A_n$ where $A_i \in \mathcal{R}$. For (i) note that if G is open and $x \in G$ then there is a set of the form $(a_1, b_1) \times \cdots \times (a_n, b_n)$ with $a_i, b_i \in \mathbf{Q}$ that contains x and lies in G, so any open set is a countable union of our basic sets. For (ii) fix $A_2, \ldots, A_n$ and let $\mathcal{G} = \{A : A \times A_2 \times \cdots \times A_n \in \mathcal{F}\}$. Since $\mathcal{F}$ is a σ-field it is easy to see that if $\Omega \in \mathcal{G}$ then $\mathcal{G}$ is a σ-field so if $\mathcal{G} \supset \mathcal{A}$ then $\mathcal{G} \supset \sigma(\mathcal{A})$. From the last result it follows that if $A_1 \in \mathcal{R}$, $A_1 \times (a_2, b_2) \times \cdots \times (a_n, b_n) \in \mathcal{F}$. Repeating the last argument $n-1$ more times proves (ii).

1.4. It is clear that if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$. Now let A_i be a countable collection of sets. If A_i^c is countable for some i then $(\cup_i A_i)^c$ is countable. On the other hand if A_i is countable for each i then $\cup_i A_i$ is countable. To check additivity of P now, suppose the A_i are disjoint. If A_i^c is countable for some i then A_j is countable for all $j \neq i$ so $\sum_k P(A_k) = 1 = P(\cup_k A_k)$. On the other hand if A_i is countable for each i then $\cup_i A_i$ is and $\sum_k P(A_k) = 0 = P(\cup_k A_k)$.

1.5. The sets of the form $(a_1, b_1) \times \cdots \times (a_d, b_d)$ where $a_i, b_i \in \mathbf{Q}$ is a countable collection that generates $\mathcal{R}^d$.

1.6. If $B \in \mathcal{R}$ then $\{Z \in B\} = (\{X \in B\} \cap A) \cup (\{Y \in B\} \cap A^c) \in \mathcal{F}$

1.7.

$$P(\chi \geq 4) \leq (2\pi)^{-1/2} 4^{-1} e^{-8} = 3.3345 \times 10^{-5}$$

The lower bound is 15/16's of the upper bound, i.e., 3.126×10^{-5}

1.8. The intervals $(F(x-), F(x))$, $x \in \mathbf{R}$ are disjoint and each one that is nonempty contains a rational number.

1.9. Let $\hat{F}^{-1}(x) = \sup\{y : F(y) \leq x\}$ and note that $F(\hat{F}^{-1}(x)) = x$ when F is continuous. This inverse wears a hat since it is different from the one defined in the proof of (1.2). To prove the result now note that

$$P(F(X) \leq x) = P(X \leq \hat{F}^{-1}(x)) = F(\hat{F}^{-1}(x)) = x$$

1.10. If $y \in (g(\alpha), g(\beta))$ then $P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F(g^{-1}(y))$. Differentiating with respect to y gives the desired result.

1.11. If $g(x) = e^x$ then $g^{-1}(x) = \log x$ and $g'(g^{-1}(x)) = x$ so using the formula in the previous exercise gives $(2\pi)^{-1/2} e^{-(\log x)^2/2}/x$.

1.12. (i) Let $F(x) = P(X \leq x)$. $P(X^2 \leq y) = F(\sqrt{y}) - F(-\sqrt{y})$ for $y > 0$. Differentiating we see that X^2 has density function

$$(f(\sqrt{y}) + f(-\sqrt{y}))/2\sqrt{y}$$

(ii) In the case of the normal this reduces to $(2\pi y)^{-1/2} e^{-y/2}$.

1.2. Random Variables

2.1. Let $\mathcal{G}$ be the smallest σ-field containing $X^{-1}(\mathcal{A})$. Since $\sigma(X)$ is a σ-field containing $X^{-1}(\mathcal{A})$, we must have $\mathcal{G} \subset \sigma(X)$ and hence $\mathcal{G} = \{\{X \in B\} : B \in \mathcal{F}\}$ for some $\mathcal{S} \supset \mathcal{F} \supset \mathcal{A}$. However, if $\mathcal{G}$ is a σ-field then we can assume $\mathcal{F}$ is. Since $\mathcal{A}$ generates $\mathcal{S}$, it follows that $\mathcal{F} = \mathcal{S}$.

2.2. If $\{X_1 + X_2 < x\}$ then there are rational numbers r_i with $r_1 + r_2 < x$ and $X_i < r_i$ so

$$\{X_1 + X_2 < x\} = \cup_{r_1,r_2 \in \mathbf{Q}: r_1+r_2<x} \{X_1 < r_1\} \cap \{X_2 < r_2\} \in \mathcal{F}$$

2.3. Let $\Omega_0 = \{\omega : X_n(\omega) \to X(\omega)\}$. If $\omega \in \Omega_0$ it follows from the definition of continuity that $f(X_n(\omega)) \to f(X(\omega))$. Since $P(\Omega_0) = 1$ the desired result follows.

2.4. (i) If G is an open set then $f^{-1}(G)$ is open and hence measurable. Now use $\mathcal{A} =$ the collection of open sets in (2.1).
(ii) Let G be an open set and let $f(x)$ be the distance from x to the complement of G, i.e., $\inf\{|x-y| : y \in G^c\}$. f is continuous and $\{f > 0\} = G$, so we need all the open sets to make all the continuous functions measurable.

2.5. If f is l.s.c. and x_n is a sequence of points that converge to x and have $f(x_n) \le a$ then $f(x) \le a$, i.e., $\{x : f(x) \le a\}$ is closed. To argue the converse note that if $\{y : f(y) > a\}$ is open for each $a \in \mathbf{R}$ and $f(x) > a$ then it is impossible to have a sequence of points $x_n \to x$ with $f(x_n) \le a$ so $\liminf_{y \to x} f(y) \ge a$ and since $a < f(x)$ is arbitrary, f is l.s.c.
The measurability of l.s.c. functions now follows from Example 2.1. For the other type note that if f is u.s.c. then $-f$ is measurable since it is l.s.c., so $f = -(-f)$ is.

2.6. In view of the previous exercise we can show f^δ is l.s.c. by showing $\{x : f^\delta(x) > a\}$ is open for each $a \in \mathbf{R}$. To do this we note that if $f^\delta(x) > a$ then there is an $\epsilon > 0$ and a z with $|z - x| < \delta - \epsilon$ so that $f(z) > a$ but then if $|y - x| < \epsilon$ we have $f^\delta(y) > a$. A similar argument shows that $\{x : f_\delta(x) < a\}$ is open for each $a \in \mathbf{R}$ so f_δ is u.s.c. The measurability of f^0 and f_0 now follows from (2.5). The measurability of $\{f^0 = f_0\}$ follows from the fact that $f^0 - f_0$ is.

2.7. Clearly the class of $\mathcal{F}$ measurable functions contains the simple functions and by (2.5) is closed under pointwise limits. To complete the proof now it suffices to observe that any $f \in \mathcal{F}$ is the pointwise limit of the simple functions $f_n = -n \vee ([2^n f]/2^n) \wedge n$.

2.8. Clearly the collection of functions of the form $f(X)$ contains the simple functions measurable with respect to $\sigma(X)$. To show that it is closed under pointwise limits suppose $f_n(X) \to Z$, and let $f(x) = \limsup_n f_n(x)$. Since $f(X) = \limsup_n f_n(X)$ it follows that $Z = f(X)$. Since any $f(X)$ is the pointwise limit of simple functions, the desired result follows from the previous exercise.

2.9. Note that for fixed n the $B_{m,n}$ form a partition of $\mathbf{R}$ and $B_{m,n} = B_{2m,n+1} \cup B_{2m+1,n+1}$. If we write $f_n(x)$ out in binary then as $n \to \infty$ we get more digits in the expansion but don't change any of the old ones so $\lim_n f_n(x) = f(x)$ exists. Since $|f_n(X(\omega)) - Y(\omega)| \le 2^{-n}$ and $f_n(X(\omega)) \to f(X(\omega))$ for all ω, $Y = f(X)$.

1.3. Expected Value

3.1. $X - Y \ge 0$ so $E|X-Y| = E(X-Y) = EX - EY = 0$ and using (3.4) it follows that $P(|X-Y| \ge \epsilon) = 0$ for all $\epsilon > 0$.

3.2. (3.1c) is trivial if $EX = \infty$ or $EY = -\infty$. When $EX^+ < \infty$ and $EY^- < \infty$, we have $E|X|, E|Y| < \infty$ since $EX^- \le EY^-$ and $EX^+ \ge EY^+$.

To prove (3.1a) we can without loss of generality suppose $EX^-, EY^- < \infty$ and also that $EX^+ = \infty$ (for if $E|X|, E|Y| < \infty$ the result follows from the theorem). In this case, $E(X+Y)^- \le EX^- + EY^- < \infty$ and $E(X+Y)^+ \ge EX^+ - EY^- = \infty$ so $E(X+Y) = \infty = EX + EY$.

To prove (3.1b) we note that it is easy to see that if $a \neq 0$ $E(aX) = aEX$. To complete the proof now it suffices to show that if $EY = \infty$ then $E(Y+b) = \infty$, which is obvious if $b \ge 0$ and easy to prove by contradiction if $b < 0$.

3.3. Recall the proof of (5.2) in the Appendix. We let $\ell(x) \le \varphi(x)$ be a linear function with $\ell(EX) = \varphi(EX)$ and note that $E\varphi(X) \ge E\ell(X) = \ell(EX)$. If equality holds then Exercise 3.1 implies that $\varphi(X) = \ell(X)$ a.s. When φ is strictly convex we have $\varphi(x) > \ell(x)$ for $x \neq EX$ so we must have $X = EX$ a.s.

3.4. There is a linear function

$$\psi(x) = \varphi(EX_1, \ldots, EX_n) + \sum_{i=1}^{n} a_i(x_i - EX_i)$$

so that $\varphi(x) \ge \psi(x)$ for all x. Taking expected values now and using (3.1c) now gives the desired result.

3.5. (i) Let $P(X = a) = P(X = -a) = b^2/2a^2$, $P(X = 0) = 1 - (b^2/a^2)$.

(ii) As $a \to \infty$ we have $a^2 1_{(|X| \geq a)} \to 0$ a.s. Since all these random variables are smaller than X^2, the desired result follows from the dominated convergence theorem.

3.6. (i) First note that $EY = EX$ and $\text{var}(Y) = \text{var}(X)$ implies that $EY^2 = EX^2$ and since $\varphi(x) = (x+b)^2$ is a quadratic that $E\varphi(Y) = E\varphi(X)$. Applying (3.4) we have

$$P(Y \geq a) \leq E\varphi(Y)/(a+b)^2 = E\varphi(X)/(a+b)^2 = p$$

(ii) By (i) we want to find $p, b > 0$ so that $ap - b(1-p) = 0$ and $a^2 p + b^2(1-p) = \sigma^2$. Looking at the answer we can guess $p = \sigma^2/(\sigma^2 + a^2)$, pick $b = \sigma^2/a$ so that $EX = 0$ and then check that $EX^2 = \sigma^2$.

3.7. (i) Let $P(X = n) = P(X = -n) = 1/2n^2$, $P(X = 0) = 1 - 1/n^2$ for $n \geq 1$.
(ii) Let $P(X = 1 - \epsilon) = 1 - 1/n$ and $P(X = 1 + b) = 1/n$ for $n \geq 2$. To have $EX = 1$, $\text{var}(X) = \sigma^2$ we need

$$-\epsilon(1 - 1/n) + b(1/n) = 0 \qquad \epsilon^2(1 - 1/n) + b^2(1/n) = \sigma^2$$

The first equation implies $\epsilon = b/(n-1)$. Using this in the second we get

$$\sigma^2 = b^2 \frac{1}{n(n-1)} + b^2 \frac{1}{n} = \frac{b^2}{n-1}$$

3.8. Cauchy-Schwarz implies

$$\left(EY 1_{(Y>a)}\right)^2 \leq EY^2 P(Y > a)$$

The left hand side is larger than $(EY - a)^2$ so rearranging gives the desired result.

3.9. $EX_n^{2/\alpha} = n^2(1/n - 1/(n+1)) = n/(n+1) \leq 1$. If $Y \geq X_n$ for all n then $Y \geq n^\alpha$ on $(1/(n+1), 1/n)$ but then $EY \geq \sum_{n=1}^\infty n^{\alpha-1}/(n+1) = \infty$ since $\alpha > 1$.

3.10. If $g = 1_A$ this follows from the definition. Linearity of integration extends the result to simple functions, and then monotone convergence gives the result for nonnegative functions. Finally by taking positive and negative parts we get the result for integrable functions.

3.11. To see that $1_A = 1 - \prod_{i=1}^n (1 - 1_{A_i})$ note that the product is zero if and only if $\omega \in A_i$ some i. Expanding out the product gives

$$1 - \prod_{i=1}^n (1 - 1_{A_i}) = \sum_{i=1}^n 1_{A_i} - \sum_{i<j} 1_{A_i} 1_{A_j} \cdots + (-1)^n \prod_{j=1}^n 1_{A_j}$$

3.12. The first inequality should be clear. To prove the second it suffices to show

$$1_A \geq \sum_{i=1}^n 1_{A_i} - \sum_{i<j} 1_{A_i} 1_{A_j}$$

To do this we observe that if ω is in exactly m of the sets A_i then the right hand side is $m - \binom{m}{2}$ which is ≤ 1 for all $m \geq 1$. For the third inequality it suffices to show

$$1_A \leq \sum_{i=1}^n 1_{A_i} - \sum_{i<j} 1_{A_i} 1_{A_j} + \sum_{i<j<k} 1_{A_i} 1_{A_j} 1_{A_k}$$

This time if ω is in exactly m of the sets A_i then the right hand side is

$$m - \frac{m(m-1)}{2} + \frac{m(m-1)(m-2)}{6}$$

We want to show this to be ≥ 1 when $m \geq 1$. When $m \geq 5$ the third term is $\geq$ the second and this is true. Computing the value when $m = 1, 2, 3, 4$ gives $1, 1, 1, 2$ and completes the proof.

3.13. If $0 < j < k$ then $|x|^j \leq 1 + |x|^k$ so $E|X|^k < \infty$ implies $E|X|^j < \infty$. To prove the inequality note that $\varphi(x) = |x|^{k/j}$ is convex and apply (3.2) to $|X|^j$.

3.14. Jensen's inequality implies $\varphi(EX) \leq E\varphi(X)$ so the desired result follows by noting $E\varphi(X) = \sum_{m=1}^n p(m) y_m$ and

$$\varphi(EX) = \exp\left(\sum_{m=1}^n p(m) \log y_m\right) = \prod_{m=1}^n y_m^{p(m)}$$

3.15. Let $Y_n = X_n + X_1^-$. Then $Y_n \geq 0$ and $Y_n \uparrow X + X_1^-$ so the monotone convergence theorem implies $E(X_n + X_1^-) \uparrow E(X + X_1^-)$. Using (3.1a) now it follows that $EX_n + EX_1^- \uparrow EX + EX_1^-$. The assumption that $EX_1^- < \infty$ allows us to subtract EX_1^- and get the desired result.

3.16. $(y/X)1_{(X>y)} \leq 1$ and converges to 0 a.s. as $y \to \infty$ so the first result follows from the bounded convergence theorem. To prove the second result, we use our first observation to see that if $0 < y < \epsilon$

$$E(y/X; X > y) \leq P(0 < X < \epsilon) + E(y/X; X \geq \epsilon)$$

On $\{X \geq \epsilon\}$, $y/X \leq y/\epsilon \leq 1$ and $y/X \to 0$ so the bounded convergence theorem implies

$$\limsup_{y \to 0} E(y/X; X > y) \leq P(0 < X < \epsilon)$$

and the desired result follows since ϵ is arbitrary.

3.17. Let $Y_N = \sum_{n=0}^N X_n$. Using the monotone convergence theorem, the linearity of expectation, and the definition of the infinite sum of a sequence of nonnegative numbers

$$\begin{aligned} E\left(\sum_{n=0}^\infty X_n\right) &= E \lim_{N\to\infty} Y_N = \lim_{N\to\infty} EY_N \\ &= \lim_{N\to\infty} \sum_{n=0}^N EX_n = \sum_{n=0}^\infty EX_n \end{aligned}$$

3.18. Let $Y_n = |X|1_{A_n}$. Jensen's inequality and the previous exercise imply

$$\sum_{n=0}^\infty |E(X;A_n)| \le \sum_{n=0}^\infty EY_n = E\sum_{n=0}^\infty Y_n \le E|X| < \infty$$

Let $B_n = \cup_{m=0}^n A_m$, and $X_n = X1_{B_n}$. As $n \to \infty$, $X1_{B_n} \to X1_A$ and $E|X| < \infty$ so the dominated convergence theorem and the linearity of expectation imply

$$E(X;A) = \lim_{n\to\infty} E(X;B_n) = \lim_{n\to\infty} \sum_{m=0}^n E(X;A_m)$$

1.4. Independence

4.1. (i) If $A \in \sigma(X)$ then it follows from the definition of $\sigma(X)$ that $A = \{X \in C\}$ for some $C \in \mathcal{R}$. Likewise if $B \in \sigma(Y)$ then $B = \{Y \in D\}$ for some $D \in \mathcal{R}$, so using these facts and the independence of X and Y,

$$P(A\cap B) = P(X\in C, Y\in D) = P(X\in C)P(Y\in D) = P(A)P(B)$$

(ii) Conversely if $X \in \mathcal{F}$, $Y \in \mathcal{G}$ and $C, D \in \mathcal{R}$ it follows from the definition of measurability that $\{X \in C\} \in \mathcal{F}$, $\{Y \in D\} \in \mathcal{G}$. Since $\mathcal{F}$ and $\mathcal{G}$ are independent, it follows that $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$.

4.2. (i) Subtracting $P(A\cap B) = P(A)P(B)$ from $P(B) = P(B)$ shows $P(A^c \cap B) = P(A^c)P(B)$. The second and third conclusions follow by applying the first one to the pairs of independent events (B, A) and (A, B^c).
(ii) If $C, D \in \mathcal{R}$ then $\{1_A \in C\} \in \{\emptyset, A, A^c, \Omega\}$ and $\{1_B \in D\} \in \{\emptyset, B, B^c, \Omega\}$, so there are 16 things to check. When either set involved is $\emptyset$ or Ω the equality

holds, so there are only four cases to worry about and they are all covered by (i).

4.3. (i) Let $B_1 = A_1^c$ and $B_i = A_i$ for $i > 1$. If $I \subset \{1, \ldots, n\}$ does not contain 1 it is clear that $P(\cap_{i\in I} B_i) = \prod_{i\in I} P(B_i)$. Suppose now that $1 \in I$ and let $J = I - \{1\}$. Subtracting $P(\cap_{i\in I} A_i) = \prod_{i\in I} P(A_i)$ from $P(\cap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$ gives $P(A_1^c \cap \cap_{i\in J} A_i) = P(A_1^c)\prod_{i\in J} P(A_i)$.
(ii) Iterating (i) we see that if $B_i \in \{A_i, A_i^c\}$ then $B_1, \ldots, B_n$ are independent. Thus if $C_i \in \{A_i, A_i^c, \Omega\}$ $P(\cap_{i=1}^n C_i) = \prod_{i=1}^n P(C_i)$. The last equality holds trivially if some $C_i = \emptyset$, so noting $1_{A_i} \in \{\emptyset, A_i, A_i^c, \Omega\}$ the desired result follows.

4.4. Let $c_m = \int g(x_m)\,dx_m$. If some $c_m = 0$ then $g_m = 0$ and hence $f = 0$ a.e., a contradiction. Integrating over the whole space we have $1 = \prod_{m=1}^n c_m$ so each $c_m < \infty$. Let $f_m(x) = g_m(x)/c_m$ and $F_m(y) = \int_{-\infty}^y f_m(x)\,dx$ for $-\infty < x \le \infty$. Integrating over $\{x : x_m \le y_m, 1 \le m \le n\}$ we have

$$P(X_m \le y_m, 1 \le m \le n) = \prod_{m=1}^n F_m(y_m)$$

Taking $y_k = \infty$ for $k \ne m$, it follows that $F_m(y_m) = P(X_m \le y_m)$ and we have checked (4.3).

4.5. The first step is to prove the stronger condition: if $I \subset \{1, \ldots, n\}$ then

$$P(X_i = x_i, i \in I) = \prod_{i\in I} P(X_i = x_i)$$

To prove this, note that if $|I| = n-1$ this follows by summing over the possible values for the missing index and then use induction. Since $P(X_i \in S_i^c) = 0$, we can check independence by showing that if $A_i \subset S_i$ then

$$P(X_i \in A_i, 1 \le i \le n) = \prod_{i=1}^n P(X_i \in A_i)$$

To do this we let $\mathcal{A}_i$ consist of Ω and all the sets $\{X_i = x\}$ with $x \in S_i$. Clearly, $\mathcal{A}_i$ is a π-system that contains Ω. Using (4.2) it follows that $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ are independent. Since for any subset B_i of S_i, $\{X_i \in B_i\}$ is in $\sigma(\mathcal{A}_i)$ the desired result follows.

4.6. $EX_n = \int_0^1 \sin(2\pi nx)\,dx = -(2\pi n)^{-1}\cos(2\pi nx)|_0^1 = 0$. Integrating by parts twice

$$\begin{aligned} EX_mX_n &= \int_0^1 \sin(2\pi mx)\sin(2\pi nx)\,dx \\ &= \frac{m}{n}\int_0^1 \cos(2\pi mx)\cos(2\pi nx)\,dx \\ &= \frac{m^2}{n^2}\int_0^1 \sin(2\pi mx)\sin(2\pi nx)\,dx \end{aligned}$$

so if $m \neq n$, $EX_m X_n = 0$. To see that X_m and X_n are not independent note that $X_m(x) = 0$ when $x = k/2m$, $0 \leq k < 2m$ and on this set $X_n(x)$ takes on the values $V_n = \{y_0, y_1, \dots y_{2m-1}\}$. Let $[a, b] \subset [-1, 1] - V$ with $a < b$. Continuity of sin implies that if $\epsilon > 0$ is sufficiently small, we have

$$P(X_m \in [0, \epsilon], X_n \in [a, b]) = 0 < P(X_m \in [0, \epsilon])P(X_n \in [a, b])$$

4.7. (i) Using (4.9) with $z = 0$ and then with $z < 0$ and letting $z \uparrow 0$ and using the bounded convergence theorem, we have

$$\begin{aligned} P(X + Y \leq 0) &= \int F(-y)dG(y) \\ P(X + Y < 0) &= \int F(-y-)dG(y) \end{aligned}$$

where $F(-y-)$ is the left limit at $-y$. Subtracting the two expressions we have

$$P(X + Y = 0) = \int \mu(\{-y\})dG(y) = \sum_y \mu(\{-y\})\nu(\{y\})$$

since the integrand is only positive for at most countably many y.
(ii) Applying the result in (i) with Y replaced by $-Y$ and noting $\mu(\{x\}) = 0$ for all x gives the desired result.

4.8. The result is trivial for $n = 1$. If $n > 1$, let $Y_1 = X_1 + \cdots + X_{n-1}$ which is gamma$(n - 1, \lambda)$ by induction, and let $Y_2 = X_n$ which is gamma$(1, \lambda)$. Then use Example 4.3.

4.9. Suppose $Y_1 = \text{normal}(0, a)$ and $Y_2 = \text{normal}(0, b)$. Then (4.10) implies

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\sqrt{ab}} \int e^{-x^2/2a} e^{-(z-x)^2/2b} \, dx$$

Dropping the constant in front, the integral can be rewritten as

$$\begin{aligned} &\int \exp\left(-\frac{bx^2 + ax^2 - 2axz + az^2}{2ab}\right) dx \\ &\quad = \int \exp\left(-\frac{a+b}{2ab}\left\{x^2 - \frac{2a}{a+b}\, xz + \frac{a}{a+b}\, z^2\right\}\right) dx \\ &\quad = \int \exp\left(-\frac{a+b}{2ab}\left\{\left(x - \frac{a}{a+b}\, z\right)^2 + \frac{ab}{(a+b)^2}\, z^2\right\}\right) dx \end{aligned}$$

since $-\{a/(a+b)\}^2 + \{a/(a+b)\} = ab/(a+b)^2$. Factoring out the term that does not depend on x, the last integral

$$= \exp\left(-\frac{z^2}{2(a+b)}\right) \int \exp\left(-\frac{a+b}{2ab}\left(x - \frac{a}{a+b}z\right)^2\right) dx$$
$$= \exp\left(-\frac{z^2}{2(a+b)}\right) \sqrt{2\pi ab/(a+b)}$$

since the last integral is the normal density with parameters $\mu = az/(a+b)$ and $\sigma^2 = ab/(a+b)$ without its proper normalizing constant. Reintroducing the constant we dropped at the beginning,

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\sqrt{ab}} \sqrt{2\pi ab/(a+b)} \exp\left(-\frac{z^2}{2(a+b)}\right)$$

4.10. It is clear that $h(\rho(x,y))$ is symmetric and vanishes only when $x = y$. To check the triangle inequality, we note that

$$\begin{aligned} h(\rho(x,y)) + h(\rho(y,z)) &= \int_0^{\rho(x,y)} h'(u)\,du + \int_0^{\rho(y,z)} h'(u)\,du \\ &\geq \int_0^{\rho(x,y)+\rho(y,z)} h'(u)\,du \\ &\geq \int_0^{\rho(x,z)} h'(u)\,du = h(\rho(x,z)) \end{aligned}$$

the first inequality holding since h' is decreasing, the second following from the traingle inequality for ρ.

4.11. If $C, D \in \mathcal{R}$ then $f^{-1}(C), g^{-1}(D) \in \mathcal{R}$ so

$$\begin{aligned} P(f(X) \in C, g(Y) \in D) &= P(X \in f^{-1}(C), Y \in g^{-1}(D)) \\ &= P(X \in f^{-1}(C))P(Y \in g^{-1}(D)) \\ &= P(f(X) \in C)P(g(Y) \in D) \end{aligned}$$

4.12. The fact that K is prime implies that for any $\ell > 0$

$$\{\ell j \bmod K : 0 \leq j < K\} = \{0, 1, \ldots K-1\}$$

which implies that for any $m \geq 0$ we have $P(X+mY = i) = 1/K$ for $0 \leq i < K$. If $m < n$ and $\ell = n - m$ our fact implies that if $0 \leq m < n < K$ then for each

$0 \le i, j < K$ there is exactly one pair $0 \le x, y < K$ so that $x + my = i$ and $x + ny = j$. This shows

$$P(X + mY = i, X + nY = j) = 1/K^2 = P(X + mY = i)P(X + nY = j)$$

so the variables are pairwise independent.

4.13. Let X_1, X_2, X_3, X_4 be independent and take values 1 and -1 with probability 1/2 each. Let $Y_1 = X_1X_2$, $Y_2 = X_2X_3$, $Y_3 = X_3X_4$, and $Y_4 = X_4X_1$. It is easy to see that $P(Y_i = 1) = P(Y_i = -1) = 1/2$. Since $Y_1Y_2Y_3Y_4 = 1$, $P(Y_1 = Y_2 = Y_3 = 1, Y_4 = -1) = 0$ and the four random variables are not independent. To check that any three are it suffices by symmetry to consider Y_1, Y_2, Y_3. Let $i_1, i_2, i_3 \in \{-1, 1\}$

$$\begin{aligned} P(Y_1 = i_1, Y_2 = i_2, Y_3 = i_3) &= P(X_2 = i_1X_1, X_3 = i_2X_2, X_4 = i_3X_3) = 1/8 \\ &= P(Y_1 = i_1)P(Y_2 = i_2)P(Y_3 = i_3) \end{aligned}$$

4.14. Let $\mathcal{A}_1$ consist of the set $\{1, 2\}$ and $\mathcal{A}_2$ consist of the sets $\{1, 3\}$ and $\{1, 4\}$. Clearly $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent, but $\sigma(\mathcal{A}_2) =$ the set of all subsets so $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are not independent.

4.15. $\{X + Y = n\} = \cup_m \{X = m, Y = n - m\}$. The events on the right hand side are disjoint, so using independence

$$P(X + Y = n) = \sum_m P(X = m, Y = n - m) = \sum_m P(X = m)P(Y = n - m)$$

4.16. Using 4.15, some arithmetic and then the binomial theorem

$$\begin{aligned} P(X + Y = n) &= \sum_{m=0}^{n} e^{-\lambda} \frac{\lambda^m}{m!} e^{-\mu} \frac{\mu^{n-m}}{(n-m)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} \lambda^m \mu^{n-m} \\ &= e^{-(\lambda+\mu)} \frac{(\mu + \lambda)^n}{n!} \end{aligned}$$

4.17. (i) Using 4.15, some arithmetic and the observation that in order to pick k objects out of $n + m$ we must pick j from the first n for some $0 \le j \le k$ we have

$$\begin{aligned} P(X + Y = k) &= \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-(k-j)} \\ &= p^k (1-p)^{n+m-k} \sum_{j=0}^{k} \binom{n}{j} \binom{m}{k-j} \\ &= \binom{n+m}{k} p^k (1-p)^{n+m-k} \end{aligned}$$

(ii) Let $\xi_1, \xi_2, \ldots$ be independent Bernoulli(p). We will prove by induction that $S_n = \xi_1 + \cdots + \xi_n$ has a Binomial(n, p) distribution. This is trivial if $n = 1$. To do the induction step note $X = S_{n-1}$ and $Y = \xi_n$ and use (i).

4.18. (a) When $k = 0, 1, 2, 3, 4$, $P(X + Y = k) = 1/9, 2/9, 3/9, 2/9, 1/9$.
(b) We claim that the joint distribution must be

$X \backslash Y$	0	1	2
2	a	$2/9 - a$	$1/9$
1	$2/9 - a$	$1/9$	a
0	$1/9$	a	$2/9 - a$

where $0 \le a \le 2/9$. To prove this let $a_{ij} = P(X = i, Y = j)$. $P(X + Y = 0) = 1/9$ implies $a_{00} = 1/9$. Let $a_{01} = a$. $P(X = 0) = 1/3$ implies $a_{02} = 2/9 - a$. $P(X + Y = 1) = 2/9$ implies $a_{10} = 2/9 - a$. $P(Y = 0) = 1/3$ implies $a_{20} = a$. $P(X+Y = 2) = 1/3$ implies $a_{11} = 1/9$. Using the fact that the row and column sums are $1/3$ one can now fill in the rest of the table.

4.19. If we let $h(x, y) = 1_{(xy \le z)}$ in (4.7) then it follows that

$$P(XY \le z) = \iint 1_{(xy \le z)} \, dF(x) \, dG(y) = \int F(z/y) \, dG(y)$$

4.20. Let $i_1, i_2, \ldots, i_n \in \{0, 1\}$ and $x = \sum_{m=1}^n i_m 2^{-m}$

$$P(Y_1 = i_1, \ldots, Y_n = i_n) = P(\omega \in [x, x + 2^{-n})) = 2^{-n}$$

1.5. Weak Laws of Large Numbers

5.1. First note that $\text{var}(X_m)/m \to 0$ implies that for any $\epsilon > 0$ there is an $A < \infty$ so that $\text{var}(X_m) \le A + \epsilon m$. Using this estimate and the fact that $\sum_{m=1}^n m \le \sum_{m=1}^n 2m - 1 = n^2$

$$E(S_n/n - \nu_n)^2 = \frac{1}{n^2} \sum_{m=1}^n \text{var}(X_m) \le A/n + \epsilon$$

Since ϵ is arbitrary this shows the L^2 convergence of $S_n/n - \nu_n$ to 0, and convergence in probability follows from (5.3).

5.2. Let $\epsilon > 0$ and pick K so that if $k \ge K$ then $r(k) \le \epsilon$. Noting that Cauchy Schwarz implies $EX_iX_j \le (EX_i^2 EX_j^2)^{1/2} = EX_k^2 = r(0)$ and breaking the sum into $|i - j| \le K$ and $|i - j| > K$ we have

$$ES_n^2 = \sum_{1 \le i,j \le n} EX_iX_j \le n(2K + 1)r(0) + n^2\epsilon$$

Dividing by n^2 we see $\limsup E(S_n/n^2) \le \epsilon$. Since ϵ is arbitrary we have $S_n/n \to 0$ in L^2 and convergence in probability follows from (5.3).

5.3. (i) Since $f(U_1), f(U_2), \ldots$ are independent and have $E|f(U_i)| < \infty$ this follows from the weak law of large numbers, (5.8).
(ii) $P(|I_n - I| > a/n^{1/2}) \le (n/a^2)\text{var}(I_n) = \sigma^2/a^2$ where $\sigma^2 = \int f^2 - (\int f)^2$

5.4. Replacing $\log k$ by $\log n$ we see that

$$P(|X_i| > n) \le \sum_{k=n+1}^{\infty} \frac{C}{k^2 \log k} \le \frac{C}{n \log n}$$

so $nP(|X_i| > n) \to 0$ and (5.6) can be applied.

$$E|X_i| = \sum_{k=2}^{\infty} C/k \log k = \infty$$

but the truncated mean

$$\mu_n = EX_i 1_{(|X_i| \le n)} = \sum_{k=2}^{n} (-1)^k \frac{C}{k \log k} \to \sum_{k=2}^{\infty} (-1)^k \frac{C}{k \log k}$$

since the latter is an alternating series with decreasing terms (for $k \ge 3$).

5.5. $nP(X_i > n) = e/\log n \to 0$ so (5.6) can be applied. The truncated mean

$$\mu_n = EX_i 1_{(|X_i| \le n)} = \int_e^n \frac{e}{x \log x}\, dx = e \log\log x|_e^n = e \log\log n$$

so $S_n/n - e\log\log n \to 0$ in probability.

5.6. Clearly, $X = \sum_{n=1}^{X} 1 = \sum_{n=1}^{\infty} 1_{(X \ge n)}$ so taking expected values proves (i). For (ii) we consider the squares $[0,k]^2$ to get $X^2 = \sum_{n=1}^{\infty} (2n-1) 1_{(X \ge n)}$ and then take expected values to get the desired formula.

5.7. Note $H(X) = \int_{-\infty}^{\infty} h(y) 1_{(X \ge y)}\, dy$ and take expected values.

5.8. Let $m(n) = \inf\{m : 2^{-m} m^{-3/2} \le n^{-1}\}$, $b_n = 2^{m(n)}$. Replacing $k(k+1)$ by $m(m+1)$ and summing we have

$$P(X_i > 2^m) \le \sum_{k=m+1}^{\infty} \frac{1}{2^k m(m+1)} = \frac{2^{-m}}{m(m+1)}$$

$$nP(X_i > b_n) \le n 2^{-m(n)}/m(n)(m(n)+1) \le (m(n)+1)^{-1/2} \to 0$$

To check (ii) in (5.5) now, we let $\bar{X} = X1_{(|X| \le 2^{m(n)})}$ and observe

$$E\bar{X}^2 \le 1 + \sum_{k=1}^{m(n)} 2^{2k} \cdot \frac{1}{2^k k(k+1)}$$

To estimate the sum divide it into $k \ge m(n)/2$ and $1 \le k < m(n)/2$ and replace k by the smallest value in each piece to get

$$\begin{aligned}
&\le 1 + \sum_{k=1}^{m(n)/2} 2^k + \frac{4}{m(n)^2} \sum_{k=m(n)/2}^{m(n)} 2^k \\
&\le 1 + 2 \cdot 2^{m(n)/2} + 8 \cdot 2^{m(n)}/m(n)^2 \le C2^{m(n)}/m(n)^2
\end{aligned}$$

Using this inequality it follows that

$$\frac{nE\bar{X}^2}{b_n^2} \le \frac{C2^{m(n)}}{m(n)^2} \cdot \frac{n}{2^{2m(n)}} \le \frac{C}{m(n)^{1/2}} \to 0$$

The last detail is to compute

$$\begin{aligned}
a_n = E(\bar{X}) &= -\sum_{k=m(n)+1}^{\infty} (2^k - 1)\frac{1}{2^k k(k+1)} \\
&= -\sum_{k=m(n)+1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1}\right) + \sum_{k=m(n)+1}^{\infty} \frac{1}{2^k k(k+1)} \\
&= -\frac{1}{m(n)+1} + \sum_{k=m(n)+1}^{\infty} \frac{1}{2^k k(k+1)} \sim -\frac{1}{m(n)} \sim -\frac{1}{\log_2 n}
\end{aligned}$$

From the definition of b_n it follows that $2^{m(n)-1} \le n/m^{3/2} \sim n/(\log_2 n)^{3/2}$ so (5.5) implies

$$\frac{S_n + n/\log_2 n}{n/(\log_2 n)^{3/2}} \to 0$$

5.9. $n\mu(s)/s \to 0$ as $s \to \infty$ and for large n we have $n\mu(1) > 1$, so we can define $b_n = \inf\{s \ge 1 : n\mu(s)/s \le 1\}$. Since $n\mu(s)/s$ only jumps up (at atoms of F), we have $n\mu(b_n) = b_n$. To check the assumptions of (5.5) now, we note that $n = b_n/\mu(b_n)$ so

$$nP(|X_k| > b_n) = \frac{b_n(1 - F(b_n))}{\mu(b_n)} = \frac{1}{\nu(b_n)} \to 0$$

since $b_n \to \infty$ as $n \to \infty$. To check (ii), we observe

$$\int_0^{b_n} \mu(x)\,dx \le b_n \mu(b_n) = b_n^2/n$$

So using (5.7) with $p = 2$

$$b_n^{-2} n E\bar{X}_{n,k}^2 \le \frac{\int_0^{b_n} 2x(1-F(x))\,dx}{\int_0^{b_n} \mu(x)\,dx} \to 0$$

since $\nu(s) \to \infty$ as $s \to \infty$. To derive the desired result now we note that $a_n = n\mu(b_n) = b_n$.

1.6. Borel-Cantelli Lemmas

6.1. Let $\epsilon > 0$. Pick N so that $P(|X| > N) \le \epsilon$, then pick $\delta < 1$ so that if $x, y \in [-N+1, N+1]$ and $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$.

$$P(|f(X_n) - f(X)| > \epsilon) \le P(|X| > N) + P(|X - X_n| > \delta)$$

so $\limsup_{n\to\infty} P(|f(X_n) - f(X)| > \epsilon) \le \epsilon$. Since ϵ is arbitrary the desired result follows.

6.2. Pick n_k so that $EX_{n_k} \to \liminf_{n\to\infty} EX_n$. By (6.2) there is a further subsequence $X_{n(m_k)}$ so that $X_{n(m_k)} \to X$ a.s. Using Fatou's lemma and the choice of n_k it follows that

$$EX \le \liminf_{k\to\infty} EX_{n(m_k)} = \liminf_{n\to\infty} EX_n$$

6.3. If $X_{n(m)}$ is a subsequence there is a further subsequence so that $X_{n(m_k)} \to X$ a.s. We have $EX_{n(m_k)} \to EX$ by (a) (3.7) or (b) (3.8). Using (6.3) it follows that $EX_n \to EX$.

6.4. Let $\varphi(z) = |z|/(1+|z|)$. (i) Since $\varphi(z) > 0$ for $z \ne 0$, $E\varphi(|X - Y|) = 0$ implies $\varphi(|X - Y|) = 0$ a.s. and hence $X = Y$ a.s. (ii) is obvious. (iii) follows by noting that Exercise 4.10 implies $\varphi(|X-Y|) + \varphi(|Y-Z|) \ge \varphi(|X-Z|)$ and then taking expected value. To check (b) note that if $X_n \to X$ in probability then since $\varphi \le 1$, Exercise 6.3 implies $d(X_n, X) = E\varphi(|X_n - X|) \to 0$. To prove the converse let $\epsilon > 0$ and note that Chebyshev's inequality implies

$$P(|X_n - X| > \epsilon) \le d(X_n, X)/\varphi(\epsilon) \to 0$$

6.5. Pick N_k so that if $m, n \geq N_k$ then $d(X_m, X_n) \leq 2^{-k}$. Given a subsequence $X_{n(m)}$ pick m_k increasing so that $n(m_k) \geq N_k$. Using Chebyshev's inequality with $\varphi(z) = z/(1+z)$ we have

$$P(|X_{n(m_k)} - X_{n(m_{k+1})}| > k^{-2}) \leq (k^2+1)2^{-k}$$

The right hand side is summable so the Borel-Cantelli lemma implies that for large k, we have $|X_{n(m_k)} - X_{n(m_{k+1})}| \leq k^{-2}$. Since $\sum_k k^{-2} < \infty$ this and the triangle inequality imply that $X_{n(m_k)}$ converges a.s. to a limit X. To see that the limit does not depend on the subsequence note that if $X_{n'(m'_k)} \to X'$ then our original assumption implies $d(X_{n(m_k)}, X_{n'(m'_k)}) \to 0$, and the bounded convergence theorem implies $d(X, X') = 0$. The desired result now follows from (6.2).

6.6. Clearly, $P(\cup_{m \geq n} A_m) \geq \max_{m \geq n} P(A_m)$. Letting $n \to \infty$ and using (iv) of (1.1), it follows that $P(\limsup A_m) \geq \limsup P(A_m)$. The result for $\liminf$ can be proved be imitating the proof of the first result or applying it to A_m^c.

6.7. Using Chebyshev's inequality we have for large n

$$P(|X_n - EX_n| > \delta EX_n) \leq \frac{\text{var}(X_n)}{\delta^2(EX_n)^2} \leq \frac{Bn^\beta}{\delta^2(a^2/2)n^{2\alpha}} = Cn^{\beta-2\alpha}$$

If we let $n_k = [k^{2/(2\alpha-\beta)}] + 1$ and $T_k = X_{n_k}$ then the last result says

$$P(|T_k - ET_k| > \delta ET_k) \leq Ck^{-2}$$

so the Borel Cantelli lemma implies $T_k/ET_k \to 1$ almost surely. Since we have $ET_{k+1}/ET_k \to 1$ the rest of the proof is the same as in the proof of (6.8).

6.8. Exercise 4.16 implies that we can subdivide X_n with large λ_n into several independent Poissons with mean ≤ 1 so we can suppose without loss of generality that $\lambda_n \leq 1$. Once we do this and notice that for a Poisson $\text{var}(X_m) = EX_m$ the proof is almost the same as that of (6.8).

6.9. The events $\{\ell_n = 0\} = \{X_n = 0\}$ are independent and have probability $1/2$, so the second Borel Cantelli lemma implies that $P(\ell_n = 0 \text{ i.o.}) = 1$. To prove the other result let $r_1 = 1$ $r_2 = 2$ and $r_n = r_{n-1} + [\log_2 n]$. Let $A_n = \{X_m = 1 \text{ for } r_{n-1} < m \leq r_n\}$. $P(A_n) \geq 1/n$, so it follows from the second Borel Cantelli lemma that $P(A_n \text{ i.o.}) = 1$, and hence $\ell_{r_n} \geq [\log_2 n]$ i.o. Since $r_n \leq n \log_2 n$ we have

$$\frac{\ell_{r_n}}{\log_2(r_n)} \geq \frac{[\log_2 n]}{\log_2 n + \log_2 \log_2 n}$$

infinitely often and the desired result follows.

6.10. Pick $\epsilon_n \downarrow 0$ and pick c_n so that $P(|X_n| > \epsilon_n c_n) \le 2^{-n}$. Since $\sum_n 2^{-n} < \infty$, the Borel-Cantelli lemma implies $P(|X_n/c_n| > \epsilon_n \text{ i.o.}) = 0$.

6.11. (i) Let $B_n = A_n^c \cap A_{n+1}$ and note that as $n \to \infty$

$$P\left(\cup_{m=n}^{\infty} A_m\right) \le P(A_n) + \sum_{m=n}^{\infty} P(B_m) \to 0$$

(ii) Let $A_n = [0, \epsilon_n)$ where $\epsilon_n \downarrow 0$ and $\sum_n \epsilon_n = \infty$. The Borel-Cantelli lemma cannot be applied but $P(A_n) \to 0$ and $P(A_n^c \cap A_{n+1}) = 0$ for all n.

6.12. Since the events A_m^c are independent

$$P\left(\cap_{m=1}^{n} A_m\right) = \prod_{m=1}^{n} (1 - P(A_m))$$

If $P(\cup_m A_m) = 1$ then the infinite product is 0, but when $P(A_m) < 1$ for all m this imples $\sum P(A_m) = \infty$ (see Lemma) and the result follows from the second Borel-Cantelli lemma.

Lemma. If $P(A_m) < 1$ for all m and $\sum_m P(A_m) < \infty$ then

$$\prod_{m=1}^{\infty} (1 - P(A_m) > 0$$

To prove this note that if $\sum_{k=1}^{n} P(A_k) < 1$ and $\pi_m = \prod_{k=1}^{m} (1 - P(A_k))$ then

$$1 - \pi_n = \sum_{m=1}^{n} \pi_{m-1} - \pi_m \le \sum_{k=1}^{n} P(A_k) < 1$$

so if $\sum_{m=M}^{\infty} P(A_m) < 1$ then $\prod_{m=M}^{\infty} (1 - P(A_m)) > 0$. If $P(A_m) < 1$ for all m then $\prod_{m=1}^{M} (1 - P(A_m)) > 0$ and the desired result follows.

6.13. If $\sum_n P(X_n > A) < \infty$ then $P(X_n > A \text{ i.o.}) = 0$ and $\sup_n X_n < \infty$. Conversely, if $\sum_n P(X_n > A) = \infty$ for all A then $P(X_n > A \text{ i.o.}) = 1$ for all A and $\sup_n X_n = \infty$.

6.14. Note that if $0 < \delta < 1$ then $P(|X_n| > \delta) = p_n$. (i) then follows immediately, and (ii) from the fact that the two Borel Cantelli lemmas imply $P(|X_n| > \delta \text{ i.o.})$ is 0 or 1 according as $\sum_n p_n < \infty$ or $= \infty$.

6.15. The answers are (i) $E|Y_i| < \infty$, (ii) $EY_i^+ < \infty$, (iii) $nP(Y_i > n) \to 0$, (iv) $P(|Y_i| < \infty) = 1$.

(i) If $E|Y_i| < \infty$ then $\sum_n P(|Y_n| > \delta n) < \infty$ for all n so $Y_n/n \to 0$ a.s. Conversely if $E|Y_i| = \infty$ then $\sum_n P(|Y_n| > n) = \infty$ so $|Y_n|/n \geq 1$ i.o.

(ii) If $EY_i^+ < \infty$ then $\sum_n P(Y_n > \delta n) < \infty$ for all n so $\limsup_{n\to\infty} Y_n^+/n \leq 0$ a.s., and it follows that $\max_{1\leq m\leq n} Y_m/n \to 0$ Conversely if $EY_i^+ = \infty$ then $\sum_n P(Y_n > n) = \infty$ so $Y_n/n \geq 1$ i.o.

(iii) $P(\max_{1\leq m\leq n} Y_m \geq \delta n) \leq nP(Y_i \geq n\delta) \to 0$. Now, if $nP(Y_i > n) \not\to 0$ we can find a $\delta \in (0,1)$, $n_k \to \infty$ and $m_k \leq n_k$ so that $m_k P(Y_i > n_k) \to \delta$. Using the second Bonferroni inequality we have

$$P\left(\max_{1\leq m\leq m_k} Y_m > n_k\right) \geq m_k P(Y_i > n_k) - \binom{m_k}{2} P(Y_i > n_k)^2 \to \delta - \delta^2/2 > 0$$

(iv) $P(|Y_n|/n > \delta) = P(|Y_n| > n\delta) \to 0$ if $P(|Y_i| < \infty) = 1$.

6.16. Note that we can pick $\delta_n \to 0$ so that $P(|X_n - X| > \delta_n) \to 0$. Let $\omega \in \Omega$ with $P(\omega) = p > 0$. For large n we have $P(|X_n - X| > \delta_n) \leq p/2$ so $|X_n(\omega) - X(\omega)| \leq \delta_n \to 0$. If $\Omega_0 = \{\omega : P(\{\omega\}) > 0\}$ then $P(\Omega_0) = 1$ so we have proved the desired result.

6.17. If m is an integer $P(X_n \geq 2^m) = 2^{-m+1}$ so taking $x_n = \log_2(Kn \log_2 n)$ and $m_n = [x_n] + 1 \leq x_n + 1$ we have $P(X_n > 2^{x_n}) \geq 2^{-x_n} = 1/Kn\log_2 n$. Since $\sum_n 1/n\log_2 n = \infty$ the second Borel Cantelli lemma implies that with probability one $X_n > 2^{x_n}$ i.o. Since K is arbitrary the desired result follows.

6.18. (i) $P(X_n \geq \log n) = 1/n$ and these events are independent so the second Borel-Cantelli implies $P(X_n \geq \log n \text{ i.o.}) = 1$. On the other hand $P(X_n \geq (1+\epsilon)\log n) = 1/n^{1+\epsilon}$ so the first Borel-Cantelli lemma implies $P(X_n \geq (1+\epsilon)\log n \text{ i.o.}) = 0$.

(ii) The first result implies that if $\epsilon > 0$ then $X_n \leq (1+\epsilon)\log n$ for large n so $\limsup_{n\to\infty} M_n/\log n \leq 1$. On the other hand if $\epsilon > 0$

$$P(M_n < (1-\epsilon)\log n) = (1 - n^{-(1-\epsilon)})^n \leq e^{-n^\epsilon}$$

which is summable so the first Borel-Cantelli lemma implies

$$P(M_n < (1-\epsilon)\log n \text{ i.o.}) = 0$$

6.19. The Borel-Cantelli lemmas imply that $P(X_m > \lambda_m \text{i.o. }) = 0$ or 1 according as $\sum_m P(X_m > \lambda_m) < \infty$ or $= \infty$. If $X_n > \lambda_n$ infinitely often then $\max_{1\leq m\leq n} X_m > \lambda_n$ infinitely often. Conversely, if $X_n \leq \lambda_n$ for large $n \geq N_0$ then for $n \geq N_1$ we will have $\max_{1\leq m\leq n} X_m \leq \lambda_n$.

6.20. Let $X_n = \sum_{k \le n} 1_{A_k}$ and $Y_n = X_n/EX_n$. Our hypothesis implies

$$\limsup_{n\to\infty} 1/EY_n^2 = \alpha$$

Letting $a = \epsilon$ in Exercise 3.8 and noting $EY_n = 1$ we have

$$P(Y_n > \epsilon) \ge (1-\epsilon)^2/EY_n^2$$

so using the definition of Y_n and Exercise 6.6 we have

$$P(A_n \text{ i.o.}) \ge P(\limsup Y_n > \epsilon) \ge \limsup P(Y_n > \epsilon) \ge (1-\epsilon)^2\alpha$$

1.7. Strong Law of Large Numbers

7.1. Our probability space is the unit interval, with the Borel sets and Lebesgue measure. For $n \ge 0$, $0 \le m < 2^n$, let $X_{2^n+m} = 1$ on $[m/2^n, (m+1)/2^n)$, 0 otherwise. Let $N(n) = 2^n + m$ on $[m/2^n, (m+1)/2^n)$. Then $X_k \to 0$ in probability but $X_{N(n)} \equiv 1$.

7.2. Let $S_n = X_1 + \cdots + X_n$, $T_n = Y_1 + \cdots + Y_n$, and $N(t) = \sup\{n : S_n + T_n \le t\}$.

$$\frac{S_{N(t)}}{S_{N(t)+1} + T_{N(t)+1}} \le \frac{R_t}{t} \le \frac{S_{N(t)+1}}{S_{N(t)} + T_{N(t)}}$$

To handle the left-hand side we note

$$\frac{S_{N(t)}}{N(t)} \cdot \frac{N(t)+1}{S_{N(t)+1} + T_{N(t)+1}} \cdot \frac{N(t)}{N(t)+1} \to EX_1 \cdot \frac{1}{EX_1 + EY_1} \cdot 1$$

A similar argument handles the right-hand side and completes the proof.

7.3. Our assumptions imply $|X_n| = U_1 \cdots U_n$ where the U_i are i.i.d. with $P(U_i \le r) = r^2$ for $0 \le r \le 1$.

$$\frac{1}{n}\log|X_n| = \frac{1}{n}\sum_{m=1}^n \log U_m \to E\log U_m$$

by the strong law of large numbers. To compute the constant we note

$$E\log U_m = \int_0^1 2r\log r\,dr = (r^2\log r - r^2/2)\Big|_0^1 = -1/2$$

7.4. (i) The strong law of large numbers implies

$$n^{-1} \log W_n \to c(p) = E \log(ap + (1-p)V_n)$$

(ii) Differentiating we have

$$c'(p) = E\left(\frac{a - V_n}{ap + (1-p)V_n}\right) \qquad c''(p) = -E\left(\frac{(a-V_n)^2}{(ap + (1-p)V_n)^2}\right) < 0$$

(iii) In order to have a maximum in $(0,1)$ we need $c'(0) > 0$ and $c'(1) < 0$, i.e., $aE(1/V_n) > 1$ and $EV_n > a$.
(iv) In this case $E(1/V) = 5/8$, $EV = 5/2$ so when $a > 5/2$ the maximum is at 1 and if $a < 8/5$ the maximum is at 0. In between the maximum occurs at the p for which

$$\frac{1}{2} \cdot \frac{a-1}{ap + (1-p)} + \frac{1}{2} \cdot \frac{a-4}{ap + 4(1-p)} = 0$$

or

$$\frac{a-1}{(a-1)p+1} = \frac{4-a}{(a-4)p+4}$$

Cross-multiplying gives

$$(a-1)(a-4)p + 4(a-1) = (4-a)(a-1)p + (4-a)$$

and solving we have $p = (5a-8)/\{2(4-a)(a-1)\}$. It is comforting to note that this is 0 when $a = 8/5$ and is 1 when $a = 5/2$.

1.8. Convergence of Random Series

8.1. It suffices to show that if $p > 1/2$ then $\limsup_{n\to\infty} S_n/n^p \le 1$ a.s., for then if $q > p$, $S_n/n^q \to 0$. (8.2) implies

$$P\left(\max_{(m-1)^\alpha \le n \le m^\alpha} |S_n| \ge m^{\alpha p}\right) \le Cm^\alpha/m^{2\alpha p}$$

When $\alpha(2p-1) > 1$ the right hand side is summable and the desired result follows from the Borel-Cantelli lemma.

8.2. $E|X|^p = \infty$ implies $\sum_{n=1}^\infty P(|X_n| > n^{1/p}) = \infty$ which in turn implies that $|X_n| \ge n^{1/p}$ i.o. The desired result now follows from

$$\max\{|S_{n-1}|, |S_n|\} \ge |X_n|/2$$

8.3. $Y_n = X_n \sin(n\pi t)/n$ has mean 0 and variance $\le 1/n^2$. Since $\sum_{n=1}^\infty \text{var}(Y_n) < \infty$ the desired result follows from (8.3).

8.4. (i) follows from (8.3) and (8.5). For (ii) let

$$P(X_n = n) = P(X_n = -n) = \sigma_n^2/2n^2 \qquad P(X_n = 0) = 1 - \sigma_n^2/n^2$$

$\sum_{n=1}^\infty \sigma_n^2/2n^2 = \infty$ implies $P(X_n \ge n \text{ i.o.}) = 1$.

8.5. To prove that (i) is equivalent to (ii) we use Kolmogorov's three series theorem (8.4) with $A = 1$ and note that if $Y_n = X_n 1_{(X_n \le 1)}$ then $\text{var}(Y_n) \le EY_n^2 \le EY_n$. To see that (ii) is equivalent to (iii) note

$$\frac{X_n}{1+X_n} \le X_n 1_{(X_n \le 1)} + 1_{(X_n > 1)} \le \frac{2X_n}{1+X_n}$$

8.6. We check the convergence of the three series in (8.4)

$$\sum_{n=1}^\infty P(|X_n| > 1) \le \sum_{n=1}^\infty E|X_n| 1_{(|X_n|>1)} < \infty$$

Let $Y_n = X_n 1_{(|X_n| \le 1)}$. $EX_n = 0$ implies $EY_n = -EX_n 1_{(|X_n|>1)}$ so

$$\sum_{n=1}^\infty |EY_n| \le \sum_{n=1}^\infty E|X_n| 1_{(|X_n|>1)} < \infty$$

Last and easiest we have

$$\sum_{n=1}^\infty \text{var}(Y_n) \le \sum_{n=1}^\infty E|X_n|^2 1_{(|X_n|\le 1)} < \infty$$

8.7. We check the convergence of the three series in (8.4).

$$\sum_{n=1}^\infty P(|X_n| > 1) \le \sum_{n=1}^\infty E|X_n|^{p(n)} < \infty$$

Let $Y_n = X_n 1_{(|X_n| \le 1)}$. If $0 < p(n) \le 1$, $|Y_n| \le |X_n|^{p(n)}$ so $|EY_n| \le E|X_n|^{p(n)}$. If $p(n) > 1$ then $EX_n = 0$ implies $EY_n = -EX_n 1_{(|X_n|>1)}$ so we again have $|EY_n| \le E|X_n|^{p(n)}$ and it follows that

$$\sum_{n=1}^\infty |EY_n| \le \sum_{n=1}^\infty E|X_n|^{p(n)} < \infty$$

Last and easiest we have

$$\sum_{n=1}^{\infty} \text{var}(Y_n) \le \sum_{n=1}^{\infty} EY_n^2 \le \sum_{n=1}^{\infty} E|X_n|^{p(n)} < \infty$$

8.8. If $E\log^+|X_1| = \infty$ then for any $K < \infty$, $\sum_{n=1}^{\infty} P(\log^+|X_n| > Kn) = \infty$, so $|X_n| > e^{Kn}$ i.o. and the radius of convergence is 0.
If $E\log^+|X_1| < \infty$ then for any $\epsilon > 0$, $\sum_{n=1}^{\infty} P(\log^+|X_n| > \epsilon n) < \infty$, so $|X_n| \le e^{\epsilon n}$ for large n and the radius of convergence is $\ge e^{-\epsilon}$. If the X_n are not $\equiv 0$ then $P(|X_n| > \delta \text{ i.o.}) = 1$ and $\sum_{n=1}^{\infty} |X_n| \cdot 1^n = \infty$.

8.9. Let $A_k = \{|S_{m,k}| > 2a, |S_{m,j}| \le 2a, m \le j < k\}$ and let $G_k = \{|S_{k,n}| \le a\}$. Since the A_k are disjoint, $A_k \cap G_k \subset \{|S_{m,n}| > a\}$, and A_k and G_k are independent

$$\begin{aligned} P(|S_{m,n}| > a) &\ge \sum_{k=m+1}^{n} P(A_k \cap G_k) \\ &= \sum_{k=m+1}^{n} P(A_k)P(G_k) \ge \min_{m<k\le n} P(G_k) \sum_{k=m+1}^{n} P(A_k) \end{aligned}$$

8.10. Let $S_{k,n} = S_n - S_k$. Convergence of S_n to S_∞ in probability and $|S_{k,n}| \le |S_k - S_\infty| + |S_\infty - S_n|$ imply

$$\min_{m\le k\le n} P(|S_{k,n}| \le a) \to 1$$

as $m, n \to \infty$. Since $P(|S_{m,n}| > a) \to 0$, $(\star)$ implies

$$P\left(\max_{m<j\le n} |S_{m,j}| > 2a\right) \to 0$$

As at the end of the proof of (8.3) this implies that with probability 1, $S_m(\omega)$ is a Cauchy sequence and converges a.s.

8.11. Let $S_{k,n} = S_n - S_k$. Convergence of S_n/n to 0 in probability and $|S_{k,n}| \le |S_k| + |S_n|$ imply that if $\epsilon > 0$ then

$$\min_{0\le k\le n} P(|S_{k,n}| \le n\epsilon) \to 1$$

as $n \to \infty$. Since $P(|S_n| > n\epsilon) \to 0$, $(\star)$ with $m = 0$ implies

$$P\left(\max_{0<j\le n} |S_j| > 2n\epsilon\right) \to 0$$

8.12. (i) Let $S_{k,n} = S_n - S_k$. Convergence of $S_n/a(n)$ to 0 in probability and $|S_{k,\ell}| \le |S_k| + |S_\ell|$ imply that if $\epsilon > 0$ then

$$\min_{2^{n-1} \le k \le 2^n} P(|S_{k,2^n}| \le \epsilon a(2^n)) \to 1$$

as $n \to \infty$. Using $(\star)$ now we see that if n is large

$$P\left(\max_{2^{n-1} < j \le 2^n} |S_{2^{n-1},j}| > 2\epsilon a(2^n) \right) \le 2P(|S_{2^{n-1},2^n}| > \epsilon a(2^n))$$

The events on the right hand side are independent and only occur finitely often (since $S_{2^n}/a(2^n) \to 0$ almost surely) so the second Borel Cantelli lemma implies that their probabilities are summable and the first Borel Cantelli implies that the event on the right hand side only occurs finitely often. Since $a(2^n)/a(2^{n-1})$ is bounded the desired result follows.

(ii) Let $a_n = n^{1/2}(\log_2 n)^{1/2+\epsilon}$. It suffices to show that $S_n/a_n \to 0$ in probability and $S(2^n)/2^{n/2}n^{1/2+\epsilon} \to 0$ almost surely. For the first conclusion we use the Chebyshev bound

$$P(|S_n/a_n| > \delta) \le ES_n^2/(\delta^2 a_n^2) = \frac{\sigma^2}{\delta^2 (\log_2 n)^{1+2\epsilon}}$$

Noting $a(2^n) = 2^{n/2}n^{1/2+\epsilon}$ we have

$$P(|S(2^n)| > \delta 2^{n/2} n^{1/2+\epsilon}) \le \sigma^2 \delta^{-2} n^{-1-2\epsilon}$$

and the desired result follows from the Borel-Cantell lemma.

1.9. Large Deviations

9.1. Taking $n = 1$ in (9.2) we see that $\gamma(a) = -\infty$ implies $P(X_1 \geq a) = 0$. If $S_n \geq na$ then $X_m \geq a$ for some $m \leq n$ so (b) implies (c). Finally if $P(S_n \geq na) = 0$ for all n then $\gamma(a) = -\infty$.

9.2. Suppose $n = km$ where $m\lambda$ is an integer.

$$P(S_n \geq n\{\lambda a + (1-\lambda)b\}) \geq P(S_{n\lambda} \geq n\lambda a)P(S_{n(1-\lambda)} \geq n(1-\lambda)b)$$

Taking $(1/n)\log$ of both sides and letting $k \to \infty$ gives

$$\gamma(\lambda a + (1-\lambda)b) \geq \lambda\gamma(a) + (1-\lambda)\gamma(b)$$

If, without loss of generality $a < b$ then letting $q_n \uparrow \lambda$ where q_n are rationals and using monotonicity extends the result to irrational λ. For a concave function f, increasing a or $h > 0$ decreases $(f(a+h) - f(a))/h$. From this observation the Lipschitz continuity follows easily.

9.3. Since $P(X \leq x_o) = 1$, $Ee^{\theta X} < \infty$ for all $\theta > 0$. Since F_θ is concentrated on $(-\infty, x_o]$ it is clear that its mean $\mu_\theta = \varphi'(\theta)/\varphi(\theta) \leq x_o$. On the other hand if $\delta > 0$, then $P(X \geq x_o - \delta) = c_\delta > 0$, $Ee^{\theta X} \geq c_\delta e^{\theta(x_o-\delta)}$, and hence

$$F_\theta(x_o - 2\delta) = \frac{1}{\varphi(\theta)}\int_{-\infty}^{x_o-2\delta} e^{\theta x}dF(x) \leq \frac{e^{x_o-2\delta)\theta}}{c_\delta e^{(x_o-\delta)\theta}} = e^{-\theta\delta}/c_\delta \to 0$$

Since $\delta > 0$ is arbitrary it follows that $\mu_\theta \to x_o$.

9.4. If we let χ have the standard normal distribution then for $a > 0$

$$P(S_n \geq na) = P(\chi \geq a\sqrt{n}) \sim (a\sqrt{n})^{-1}\exp(-a^2n/2)$$

so $(1/n)\log P(S_n \geq na) \to -a^2/2$.

9.5.

$$Ee^{\theta X} = \sum_{n=0}^{\infty} e^{-1}e^{\theta n}/n! = \exp(e^\theta - 1)$$

so $\kappa(\theta) = e^\theta - 1$, $\varphi'(\theta)/\varphi(\theta) = \kappa'(\theta) = e^\theta$, and $\theta_a = \log a$. Plugging in gives

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a\log a + a - 1$$

9.6. $1 + x \leq e^x$ with $x = \varphi(\theta) - 1$ gives $\varphi(\theta) \leq \exp(\varphi(\theta) - 1)$ To prove the other inequality, we note that

$$\varphi(\theta) - 1 = \frac{e^\theta - 2 + e^{-\theta}}{2} = \sum_{n=1}^{\infty} \frac{\theta^{2n}}{(2n)!} \leq \beta\theta^2$$

(9.3) implies $P(S_n \geq na) \leq \exp(-n\{a\theta - \beta\theta^2\})$. Taking $\theta = a/2\beta$ to minimize the upper bound the desired result follows.

9.7. Since $\gamma(a)$ is decreasing and $\geq \log P(X = x_o)$ for all $a < x_o$ we have only to show that $\limsup \gamma(a) \leq P(X = x_o)$. To do this we begin by observing that the computation for coin flips shows that the result is true for distributions that have a two point support. Now if we let $\bar{X}_i = x_o - \delta$ when $X_i \leq x_o - \delta$ and $\bar{X}_i = x_o$ when $x_o - \delta < X_i \leq x_o$ then $\bar{S}_n \geq S_n$ and hence $\bar{\gamma}(a) \geq \gamma(a)$ but $\bar{\gamma}(a) \downarrow P(\bar{X}_i = x_o) = P(x_o - \delta < X_i \leq x_o)$. Since δ is aribitrary the desired result follows.

9.8. Clearly, $P(S_n \geq na) \geq P(S_{n-1} \geq -n\epsilon)P(X_n \geq n(a+\epsilon))$. The fact that $Ee^{\theta X} = \infty$ for all $\theta > 0$ implies $\limsup_{n\to\infty}(1/n)\log P(X_n > na) = 0$, and the desired conclusion follows as in the proof of (9.6).

9.9. Let $p_n = P(X_i > (a+\epsilon)n)$. $E|X_i| < \infty$ implies

$$P\left(\max_{i\leq n} X_i > n(a+\epsilon)\right) \leq np_n \to 0$$

and hence $P(F_n) = np_n(1-p_n)^{n-1} \sim np_n$. Breaking the event F_n into disjoint pieces according to the index of the large value, and noting

$$P\left(|S_{n-1}| < n\epsilon \,\middle|\, \max_{i\leq n} X_i \leq n(a+\epsilon)\right) \to 0$$

by the weak law of large numbers and the fact that the conditioning event has a probability $\to 1$ the desired result follows.

2 Central Limit Theorems

2.1. The De Moivre-Laplace Theorem

1.1. Since $\log(1+x)/x \to 1$ as $x \to 0$, it follows that given an $\epsilon > 0$ there is a $\delta > 0$ so that if $|x| < \delta$ then $(1-\epsilon)x < \log(1+x) < (1+\epsilon)x$. From this it is easy to see that our assumptions imply

$$\sum_{j=1}^{n} \log(1 + c_{j,n}) \to \lambda$$

and the desired result follows.

1.2. Applying Stirling's formula to $n!$ we have

$$\begin{aligned}\sqrt{2\pi n} P(S_n = n+m) &= \sqrt{2\pi n} e^{-n} n^{n+m}/(n+m)! \\ &\sim \frac{n! n^m}{(n+m)!} = \left(\prod_{k=1}^{m} 1 + \frac{k}{n} \right)^{-1}\end{aligned}$$

$\sum_{k=1}^{m} k \sim m^2/2$ so if $m \sim x\sqrt{n}$, Exercise 1.1 implies the quantity in parentheses converges to $\exp(x^2/2)$.

1.3. Using (1.2) and writing $o(1)$ to denote a term that goes to 0 as $n \to \infty$ we have

$$\begin{aligned}\frac{1}{2n} \log P(S_{2n} = 2k) &= -\frac{n+k}{2n} \log\left(1 + \frac{k}{n}\right) - \frac{n-k}{2n} \log\left(1 - \frac{k}{n}\right) + o(1) \\ &\to -\frac{1+a}{2} \log(1+a) - \frac{1-a}{2} \log(1-a)\end{aligned}$$

when $k/n \to a$. Now if $k/n \ge a > 0$ we have

$$P(S_{2n} = 2k+2) = \frac{n-k}{n+k+1} P(S_{2n} = 2k) \le \frac{1-a}{1+a} P(S_{2n} = 2k)$$

and summing a geometric series we have $P(S_{2n} \geq 2k) \leq CP(S_{2n} = 2k)$.

1.4. $P(S_n = k) = e^{-n}n^k/k!$ and $k! \sim k^k e^{-k}\sqrt{2\pi k}$ so

$$P(S_n = k) \sim e^{-n+k}\left(\frac{n}{k}\right)^k / \sqrt{2\pi k}$$

and if $k/n \to a$ we have

$$\frac{1}{n}\log P(S_n = k) = -\frac{n-k}{n} - \frac{k}{n}\log\left(\frac{k}{n}\right) + o(1) \to a - 1 - a\log a$$

Now if $k/n \geq a > 1$ we have

$$P(S_n = k+1) = \frac{n}{k+1}P(S_n = k) \leq \frac{1}{a}P(S_n = k)$$

and the result follows as in Exercise 1.3.

2.2. Weak Convergence

2.1. Let $f_n(x) = 2$ if $x \in [m/2^n, (m+1)/2^n)$ and $0 \leq m < 2^n$ is an even integer.

2.2. As $n \to \infty$
(i) $P(M_n \leq yn^{1/\alpha}) = (1 - y^{-\alpha}n^{-1})^n \to \exp(-y^{-\alpha})$
(ii) $P(M_n \leq yn^{-1/\beta}) = (1 - |y|^\beta n^{-1})^n \to \exp(-|y|^\beta)$
(iii) $P(M_n \leq \log n + y) = (1 - e^{-y}n^{-1})^n \to \exp(-e^{-y})$

2.3. (i) From the asymptotic formula it follows that

$$\lim_{x\to\infty}\frac{P(X_i > x + (\theta/x))}{P(X_i > x)} = \lim_{x\to\infty}\frac{x}{x + (\theta/x)}\exp(-\theta - \{\theta^2/2x^2\}) = e^{-\theta}$$

(ii) Let $p_n = P(X_i > b_n + (x/b_n))$ and note that the definition of b_n and (i) imply $np_n \to e^{-x}$ so

$$P(b_n(M_n - b_n) \leq x) = (1 - p_n)^n \to \exp(-e^{-x})$$

(iii) By (1.4) we have

$$P(X_i > (2\log n)^{1/2}) \sim \frac{1}{(2\log n)^{1/2}} \cdot \frac{1}{n}$$

so for large n, $b_n \leq (2\log n)^{1/2}$. On the other hand

$$P(X_i > \{2\log n - 2\log\log n)\}^{1/2}) \sim \frac{1}{(2\log n)^{1/2}} \cdot \frac{\log n}{n}$$

so for large n $b_n \geq (2\log n - 2\log\log n)^{1/2}$ From (ii) we see that if $x_n \to \infty$ and $y_n \to -\infty$

$$P\left(\frac{y_n}{b_n} \leq M_n - b_n \leq \frac{x_n}{b_n}\right) \to 1$$

Taking $x_n, y_n = o(b_n)$ the desired result follows.

2.4. Let $Y_n \stackrel{d}{=} X_n$ with $Y_n \to Y_\infty$ a.s. $0 \leq g(Y_n) \to g(Y_\infty)$ so the desired result follows from Fatou's lemma.

2.5. Let $Y_n \stackrel{d}{=} X_n$ with $Y_n \to Y_\infty$ a.s. $0 \leq g(Y_n) \to g(Y_\infty)$ so the desired result follows from (3.8) in Chapter 1.

2.6. Let $x_{j,k} = \inf\{x : F(x) > j/k\}$. Since F is continuous $x_{j,k}$ is a continuity point so $F_n(x_{j,k}) \to F(x_{j,k})$. Pick N_k so that if $n \geq N_k$ then $|F_n(x_{j,k}) - F(x_{j,k})| < 1/k$ for $1 \leq j < k$. Repeating the proof of (7.4) in Chapter 1 now shows $\sup_x |F_n(x) - F(x)| \leq 2/k$ and since k is arbitrary the desired result follows.

2.7. Let $X_1, X_2, \ldots$ be i.i.d. with distribution function F and let

$$F_n(x) = n^{-1}\sum_{m=1}^{n} 1_{X_m(\omega)\leq x}$$

(7.4) implies that $\sup_x |F_n(x) - F(x)| \to 0$ with probability one. Pick a good outcome ω_0, let $x_{n,m} = X_m(\omega_0)$ and $a_{n,m} = 1/n$.

2.8. Suppose first that integer valued $X_n \Rightarrow X_\infty$. Since $k + 1/2$ is a continuity point, for each $k \in \mathbf{Z}$

$$\begin{aligned} P(X_n = k) &= F_n(k+1/2) - F_n(k-1/2) \\ &\to F(k+1/2) - F(k-1/2) = P(X_\infty = k) \end{aligned}$$

To prove the converse let $\epsilon > 0$ and find points $I = \{x_1, \ldots x_j\}$ so that $P(X_\infty \in I) \geq 1 - \epsilon$. Pick N so that if $n \geq N$ then $|P(X_n = x_i) - P(X_\infty = x_i)| \leq \epsilon/j$. Now let m be an integer, let $I_m = I \cap (-\infty, m]$, and let J_m be the integers $\leq m$ not in I_m. The triangle inequality implies that if $n \geq N$ then

$$|P(X_n \in I_m) - P(X_\infty \in I_m)| \leq \epsilon$$

The choice of $x_1, \ldots x_j$ implies $P(X_\infty \in J_m) \leq \epsilon$ while the convergence for all x_i implies that $P(X_n \in J_m) \leq 2\epsilon$ for $n \geq N$. Combining the last three inequalities implies $|P(X_n \leq m) - P(X_\infty \leq m)| \leq 3\epsilon$ for $n \geq N$. Since ϵ is arbitrary we have shown $P(X_n \leq m) \to P(X_\infty \leq m)$. Since this holds for all integers and the distribution function is constant in between integers the desired result follows.

2.9. If $X_n \to X$ in probability and g is bounded and continuous then $Eg(X_n) \to Eg(X)$ by (6.4) in Chapter 1. Since this holds for all bounded continuous functions (2.2) implies $X_n \Rightarrow X$.
To prove the converse note that $P(X_n \le c+\epsilon) \to 1$ and $P(X_n \le c-\epsilon) \to 0$, so $P(|X_n - c| > \epsilon) \to 0$, i.e., $X_n \to c$ in probability.

2.10. If $X_n \le x - c - \epsilon$ and $Y_n \le c + \epsilon$ then $X_n + Y_n \le x$ so

$$P(X_n + Y_n \le x) \ge P(X_n \le x - c - \epsilon) - P(Y_n > c + \epsilon)$$

The second probability $\to 0$. If $x - c - \epsilon$ is a continuity point of the distribution of X the first probability $\to P(X \le x - c - \epsilon)$. Letting $\epsilon \to 0$ it follows that if x is a continuity point of the distribution of $X + c$

$$\liminf_{n\to\infty} P(X_n + Y_n \le x) \ge P(X + c \le x)$$

$P(X_n + Y_n \le x) \le P(X_n \le x - c + \epsilon) + P(Y_n < c - \epsilon)$. The second probability $\to 0$. If $x - c + \epsilon$ is a continuity point of the distribution of X the first probability $\to P(X \le x - c + \epsilon)$. Letting $\epsilon \to 0$ it follows that

$$\limsup_{n\to\infty} P(X_n + Y_n \le x) \le P(X + c \le x)$$

2.11. Suppose that $x \ge 0$. The argument is similar if $x < 0$ but some details like the next inequality are different.

$$P(X_n Y_n \le x) \ge P(X_n \le x/(c+\epsilon)) - P(Y_n > c + \epsilon)$$

The second probability $\to 0$. If $x/(c+\epsilon)$ is a continuity point of the distribution of X the first probability $\to P(X \le x/(c+\epsilon))$. Letting $\epsilon \to 0$ it follows that if x is a continuity point of the distribution of cX

$$\liminf_{n\to\infty} P(X_n Y_n \le x) \ge P(cX \le x)$$

Let $\epsilon < c$. $P(X_n Y_n \le x) \le P(X_n \le x/(c-\epsilon)) + P(Y_n < c - \epsilon)$. The second probability $\to 0$. If $x/(c-\epsilon)$ is a continuity point of the distribution of X the first probability $\to P(X \le x/(c-\epsilon))$. Letting $\epsilon \to 0$ it follows that

$$\limsup_{n\to\infty} P(X_n Y_n \le x) \le P(cX \le x)$$

2.12. The spherical symmetry of the normal distribution implies that the X_n^i are unifrom over the surface of the sphere. The strong law of large numbers implies $Y_i n / \sum_{m=1}^n Y_m^2 \to Y_i$ almost surely so the desired result follows from Exercise 2.9.

2.13. $EY_n^\beta \to 1$ implies $EY_n^\beta \le C$ so (2.7) implies that the sequence is tight. Suppose $\mu_{n(k)} \Rightarrow \mu$, and let Y be a random variable with distribution μ. Exercise 2.5 implies that if $\alpha < \beta$ then $EY^\alpha = 1$. If we let $\gamma \in (\alpha, \beta)$ we have

$$EY^\gamma = 1 = (EY^\alpha)^{\gamma/\alpha}$$

so for the random variable Y^α and the convex function $\varphi(x) = (x^+)^{\gamma/\alpha}$ we have equality in Jensen's inequality and Exercise 3.3 in Chapter 1 implies $Y^\alpha = 1$ a.s.

2.14. Suppose there is a sequence of random variables with $P(|X_n| > y) \to 0$, $EX_n^2 = 1$, and $EX_n^4 \le K$. (2.7) implies that X_n is tight and hence there is a subsequence $X_{n(k)} \Rightarrow X$. Exercise 2.5 implies that $EX_{n(k)}^2 \to EX^2$ but $P(|X| \le y) = 1$ so $EX^2 \le y^2 < 1$ a contradiction.

2.15. First we check that ρ is a metric. Clearly $\rho(F,G) = 0$ if and only if $F = G$. It is also easy to see that $\rho(F,G) = \rho(G,F)$. To check the triangle inequality we note that if $G(x) \le F(x+a) + a$ and $H(x) \le G(x+b) + b$ for all x then $H(x) \le F(x+a+b) + a + b$ for all x.
Suppose now that $\epsilon_n = \rho(F_n, F) \to 0$. Letting $n \to \infty$ in

$$F(x - \epsilon_n) - \epsilon_n \le F_n(x) \le F(x + \epsilon_n) + \epsilon_n$$

we see that $F_n(x) \to F(x)$ at continuity points of F. To prove the converse let $\epsilon > 0$ and let $x_1, \ldots, x_k$ be continuity points of F so that $F(x_1) < \epsilon$, $F(x_k) > 1 - \epsilon$ and $|x_j - x_{j+1}| < \epsilon$ for $1 \le j < k$. If $F_n \Rightarrow F$ then for $n \ge N$ we have $|F_n(x_j) - F(x_j)| \le \epsilon$ for all j. To handle the other values note that if $x_j < x < x_{j+1}$ then for $n \ge N$

$$\begin{aligned} F_n(x) &\le F_n(x_{j+1}) \le F(x_{j+1}) + \epsilon \le F(x+\epsilon) + \epsilon \\ F_n(x) &\ge F_n(x_j) \ge F(x_j) - \epsilon \ge F(x - \epsilon) - \epsilon \end{aligned}$$

If $x < x_1$ then we note

$$\begin{aligned} F_n(x) &\le F_n(x_1) \le F(x_1) + \epsilon \le 2\epsilon \le F(x + 2\epsilon) + 2\epsilon \\ F_n(x) &\ge 0 \ge F(x - \epsilon) - \epsilon \end{aligned}$$

A similar argument handles $x > x_k$ and shows $\rho(F_n, F) \le 2\epsilon$ for $n \ge N$.

2.16. To prove this result we note that

$$P(Y \le x - \epsilon) - P(|X - Y| > \epsilon) \le P(X \le x) \le P(Y \le x + \epsilon) + P(|X - Y| > \epsilon)$$

2.17. If $\alpha(X, Y) = a$ then $P(|X - Y| \geq a) \geq a \geq P(|X - Y| > a)$. The worst case for the lower bound is $P(|X - Y| = a) = a$, $P(|X - Y| = 0) = 1 - a$. The worst case for the upper bound is $P(|X-Y| = a) = 1-a$, $P(|X-Y| \approx \infty) = a$.

2.3. Characteristic Functions

3.1. $\operatorname{Re}\varphi = (\varphi + \overline{\varphi})/2$ and $|\varphi|^2 = \varphi \cdot \overline{\varphi}$. If X has ch.f. φ then $-X$ has ch.f. $\overline{\varphi}$, so the desired results follow from (3.1g) and (3.1f).

3.2. (i) Using Fubini's theorem and the fact that sin is odd

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T} e^{-ita} \int e^{itx}\,\mu(dx)\,dt &= \int \frac{1}{2T}\int_{-T}^{T} e^{it(x-a)}\,dt\,\mu(dx) \\ &= \int \frac{1}{2T}\int_{-T}^{T} \cos(t(x-a))\,dt\,\mu(dx)\end{aligned}$$

Now $|\frac{1}{2T}\int_{-T}^{T} \cos(t(x-a))\,dt| \leq 1$, and as $T \to \infty$

$$\frac{1}{2T}\int_{-T}^{T} \cos(t(x-a))\,dt \to \begin{cases} 0 & x \neq a \\ 1 & x = a \end{cases}$$

so the bounded convergence theorem gives the desired result.
(ii) The periodicity follows from the fact that $e^{2\pi ni} = 1$ for any integer n. From this it follows easily that

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-itx}\varphi(t)\,dt = \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}\varphi(t)\,dt$$

To see this note that when $T = \pi n/h$ and n is an integer the integral on the left is equal to the one on the right, and we have shown in (i) that the limit on the left exists.
(iii) The first assertion follows from (3.1e). Letting $Y = X - b$ and applying (ii) to φ_Y the ch.f. of Y

$$\begin{aligned}P(X = a) = P(Y = a - b) &= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-it(a-b)}\varphi_Y(t)\,dt \\ &= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-ita}\varphi_X(t)\,dt\end{aligned}$$

3.3. If X has ch.f. φ then $-X$ has ch.f. $\overline{\varphi}$. If φ is real $\varphi = \overline{\varphi}$ so the inversion formula (3.3) implies $X \stackrel{d}{=} -X$.

3.4. By Example 3.3 and (3.1f), $X_1 + X_2$ has ch.f. $\exp(-(\sigma_1^2 + \sigma_2^2)t^2/2)$.

3.5. Examples 3.4 and 3.6 have this property since their density functions are discontinuous.

3.6. Example 3.4 implies that the X_i have ch.f. $(\sin t)/t$, so (3.1f) implies that $X_1 + \cdots + X_n$ has ch.f. $(\sin t/t)^n$. When $n \ge 2$ this is integrable so (3.3) implies that

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\sin t/t)^n e^{itx} \, dt$$

Since $\sin t$ and t are both odd, the quotient is even and we can simplify the last integral to get the indicated formula.

3.7. $X - Y$ has ch.f. $\varphi \cdot \overline{\varphi} = |\varphi|^2$. The first equality follows by taking $a = 0$ in Exercise 3.2. The second from Exercise 4.7 in Chapter 1.

3.8. Example 3.9 and (3.1f) imply that $X_1 + \cdots + X_n$ has ch.f. $\exp(-n|t|)$, so (3.1e) implies $(X_1 + \cdots + X_n)/n$ has ch.f. $\exp(-|t|)$ and hence a Cauchy distribution.

3.9. X_n has ch.f. $\varphi_n(t) = \exp(-\sigma_n^2 t^2/2)$. By taking log's we see that $\varphi_n(1)$ has a limit if and only if $\sigma_n^2 \to \sigma^2 \in [0, \infty]$. However $\sigma^2 = \infty$ is ruled out by the remark after (3.4).

3.10. Let $\varphi_n(t) = Ee^{itX_n}$ and $\psi_n(t) = Ee^{itY_n}$. Since $X_n \Rightarrow X_\infty$ and $Y_n \Rightarrow Y_\infty$, we have $\varphi_n(t) \to \varphi_\infty(t)$ and $\psi_n(t) \to \psi_\infty(t)$. $X_n + Y_n$ has ch.f. $\varphi_n(t)\psi_n(t)$ which $\to \varphi_\infty(t)\psi_\infty(t)$. Being a product of ch.f., the limit is continuous at 0 and the desired result follows from (ii) in (3.4).

3.11. (3.1f) implies $S_n = \sum_{j=1}^n X_j$ has ch.f. $u_n(t) = \prod_{j=1}^n \varphi_j(t)$. As $n \to \infty$, $S_n \to S_\infty$ a.s. So Exercise 2.9 implies $S_n \Rightarrow S_\infty$ and (i) of (3.4) implies $u_n \to u_\infty$, the ch.f. of S_∞.

3.12. By Example 3.1 and (3.1e), $\cos(t/2^m)$ is the ch.f. of a r.v. X_m with $P(X_m = 1/2^m) = P(X_m = -1/2^m) = 1/2$, so Exercise 3.11 implies $S_\infty = \sum_{m=1}^\infty X_m$ has ch.f. $\prod_{m=1}^\infty \cos(t/2^m)$. If we let $Y_m = (2^m X_m + 1)/2$ then

$$S_\infty = \sum_{m=1}^{\infty} \left(-\frac{1}{2^m} + \frac{2Y_m}{2^m} \right) = -1 + 2 \sum_{m=1}^{\infty} Y_m/2^m$$

The Y_m are i.i.d. with $P(Y_m = 0) = P(Y_m = 1) = 1/2$ so thinking about binary digits of a point chosen at random from $(0,1)$, we see $\sum_{m=1}^\infty Y_m/2^m$ is uniform on $(0,1)$. Thus S_∞ is uniform on $(-1,1)$ and has ch.f. $\sin t/t$ by Example 3.4.

3.13. A random variable with $P(X=0) = P(X=a) = 1/2$ has ch.f. $(1+e^{ita})/2$ so Exercise 3.11 implies X has ch.f.

$$\varphi(t) = \prod_{j=1}^{\infty} \left(\frac{1+e^{it2\cdot 3^{-j}}}{2} \right)$$
$$\varphi(3^k \pi) = \prod_{j=1}^{\infty} \left(\frac{1+e^{i2\pi\cdot 3^{k-j}}}{2} \right) = \prod_{m=1}^{\infty} \left(\frac{1+e^{i2\pi\cdot 3^{-m}}}{2} \right) = \varphi(\pi)$$

3.14. We prove the result by induction on n by checking the conditions of (9.1) in the appendix. To make the notation agree we write

$$\varphi^{(n)}(x) = \int (is)^n e^{ixs} \mu(ds)$$

so $f(x,s) = (is)^n e^{ixs}$. Since $|(is)^n e^{ixs}| = |s|^n$, $E|X|^n < \infty$ then (i) holds. Clearly, (ii) $\partial f/\partial x = (is)^{n+1} e^{ixs}$ is a continuous function of x. The dominated converence theorem implies

$$x \to \int (is)^{n+1} e^{ixs} \mu(ds)$$

is a continuous function so (iii) holds. Finally,

$$\int \int_{-\delta}^{\delta} |\partial f/\partial x(y+\theta, s)| \, d\theta \, \mu(ds) = \int_{-\delta}^{\delta} E|X|^{n+1} \, d\theta < \infty$$

so (iv) holds and the desired result follows from (9.1) in the Appendix.

3.15. $\varphi(t) = e^{-t^2/2} = \sum_{n=0}^{\infty} (-1)^n t^{2n}/(2^n n!)$. In this form it is easy to see that $\varphi^{(2n)}(0) = (-1)^n (2n)!/(2^n n!)$. The deisred result now follows by observing that $E|X|^n < \infty$ for all n and using the previous exercise.

3.16. (i) Let X_i be a r.v. with ch.f. φ_i. (3.1d) and (3.7) with $n=0$ imply

$$\begin{aligned} |\varphi_i(t+h) - \varphi_i(t)| &\le E|e^{ihX_i} - 1| \le E \min(h|X_i|, 2) \\ &\le E(h|X_i|; |X_i| \le h^{-1/2}) + 2P(|X_i| > h^{-1/2}) \end{aligned}$$

The first expected value is $\le h^{1/2}$, the second term goes to 0 as $h \to 0$ by tightness.
(ii) Without loss of generality we can assume the compact set is $[-K, K]$. Let $\epsilon > 0$ and pick $\delta > 0$ so that if $|h| < \delta$ then $|\varphi_i(t+h) - \varphi_i(t)| < \epsilon$ for all i. Let $m > 1/\delta$ be an integer. Since $\varphi_n \to \varphi_\infty$ pointwise we can find N large

enough so that if $n \geq N$ then $|\varphi_n(k/m) - \varphi_\infty(k/m)| < \epsilon$ for $-Km \leq k \leq Km$. Combining the two estimates shows that if $n \geq N$ then $|\varphi_n(t) - \varphi_\infty(t)| < 2\epsilon$ for $t \in [-K, K]$.
(iii) $X_n = 1/n$ has ch.f. $e^{it/n}$ that converges to 1 pointwise but not uniformly.

3.17. (i) $E\exp(itS_n/n) = \varphi(t/n)^n$. If $\varphi'(0) = ia$ then $n(\varphi(t/n) - 1) \to iat$ as $n \to \infty$ so $\varphi(t/n)^n \to e^{iat}$ the ch.f. of a pointmass at a, so $S_n/n \Rightarrow a$ and it follows from Exercise 2.9 that $S_n/n \to a$ in probability.
(ii) Conversely if $\varphi(t/n)^n \to e^{iat}$ taking logarithms shows $n \log \varphi(t/n) \to iat$ and since $\log z$ is differentiable at $z = 1$ in the complex plane it follows that $n(\varphi(t/n) - 1) \to iat$.

3.18. Following the hint and recalling cos is even we get

$$|y| = \int_0^\infty \frac{2(1 - \cos yt)}{\pi t^2}\, dt$$

Now integrate $dF(y)$ on both sides and use Fubini's theorem on the right to get the desired identity since $\operatorname{Re}\varphi(t) = \int \cos(yt)\, dF(y)$.

3.19. Since $\varphi(-t) = \overline{\varphi(t)}$, the hypothesis of (3.9) holds and it follows that $EX^2 < \infty$. Using (3.8) now it follows that $\varphi(t) = 1 + i\mu t - t^2\sigma^2/2 + o(t^2)$ and we have $EX = 0$ and $E|X|^2 = -2c$. If $\varphi(t) = 1 + o(t^2)$ then $c = 0$ and $X \equiv 0$

3.20. (3.4) shows that $Y_n \Rightarrow 0$ implies $\varphi_n(t) \to 1$ for all t. Conversely if $\varphi_n(t) \to 1$ for $|t| < \delta$ then it follows from (3.5) that the sequence Y_n is tight. Part (i) of (3.4) implies that any subsequential limit has a ch.f. that is $= 1$ on $(-\delta, \delta)$ and hence by the previous exercise must be $\equiv 1$. We have shown now that any subsequence has a further subsequence that $\Rightarrow 0$ so we have $Y_n \Rightarrow 0$ by the last paragraph of the proof of (3.4).

3.21. If S_n converges in distribution then $\varphi_n(t) = E\exp(itS_n) \to \varphi(t)$ which is a ch.f. and hence has $|\varphi(t) - 1| < 1/2$ for $t \in [-\delta, \delta]$. If $m < n$ let

$$\varphi_{m,n}(t) = E\exp(it(S_n - S_m)) = \varphi_n(t)/\varphi_m(t)$$

when $\varphi_m(t) \neq 0$. Combining our results we see that if $m, n \to \infty$ then $\varphi_{m,n} \to 1$ for $t \in [-\delta, \delta]$. Using the previous exercise now we can conclude that if $m, n \to \infty$ then $S_n - S_m \to 0$ in probability. Using Exercise 6.4 in Chapter 1 now we can conclude that there is a random variable S_∞ with $S_n \to S_\infty$ in probability.

3.22. By Polya's criterion, (3.10), it suffices to show that $\varphi(t) = \exp(-t^\alpha)$ is convex on $(0, \infty)$. To do this we note

$$\begin{aligned}
\varphi'(t) &= -\alpha t^{\alpha-1}\exp(-t^\alpha) \\
\varphi''(t) &= (\alpha^2 t^{2\alpha-2} - \alpha(\alpha-1)t^{\alpha-2})\exp(-t^\alpha)
\end{aligned}$$

which is > 0 since $\alpha \le 1$.

3.23. (3.1f) implies that $X_1 + \cdots + X_n$ has ch.f. $\exp(-n|t|^\alpha)$, so (3.1e) implies $(X_1 + \cdots + X_n)/n^{1/\alpha}$ has ch.f. $\exp(-|t|^\alpha)$.

3.24. Let $\varphi_2(t) = \varphi_1(t)$ on A, linear on each open interval that makes up A^c, and continuous. φ_2 is convex on $(0, \infty)$ and by Polya's criterion must be a ch.f. Since $e^{-|t|}$ is strictly convex we have $\{t : \varphi(t) = \varphi_1(t)\} = A$.

3.25. Let $\varphi_0(t) = (1 - |t|)^+$ and $\varphi_1(t)$ be periodic with period 2 and $= \varphi_0(t)$ on $[-1, 1]$. If X, Y, Z are independent with X and Y having ch.f. φ_0 and Z having ch.f. φ_1 then $X + Y$ and $X + Z$ both have ch.f. φ_0^2.

3.26. Let φ_X and φ_Y be the ch.f. of X and Y. Let $\delta > 0$ be such that $\varphi_X(t) \ne 0$ for $t \in [-\delta, \delta]$. If $X + Y$ and X have the same distibution then $\varphi_X(t)\varphi_Y(t) = \varphi_X(t)$ so $\varphi_Y(t) = 1$ for $t \in [-\delta, \delta]$ and hence must be $\equiv 1$ by Exercise 3.19.

3.27. $\nu_k \le E|X|^k \le \lambda^k$. Conversely if $\epsilon > 0$ then $P(|X| > \lambda - \epsilon) > 0$ so

$$\nu_k = E|X|^k \ge (\lambda - \epsilon)^k P(|X| \ge \lambda - \epsilon)$$

and $\liminf_{k\to\infty} \nu_k^{1/k} \ge \lambda - \epsilon$.

3.28. Since $\Gamma(x)$ is bounded for $x \in [1, 2]$ the identity quoted implies that $\Gamma(x) \approx [x]!$ where $f(x) \approx g(x)$ means $0 < c \le f(x)/g(x) \le C < \infty$ for all $x \ge 1$. Stirling's formula implies

$$n! \sim (n/e)^n \sqrt{2\pi n}$$

where as usual $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$. Combining this with the previous result and recalling $(n^p)^{1/n} \to 1$ shows

$$\Gamma((n + \alpha + 1)/\lambda)^{1/n} \sim \left(\frac{n + \alpha + 1}{\lambda e}\right)^{(n+\alpha+1)/\lambda n} \sim C n^{\frac{1}{\lambda}}$$

from which the desired result follows easily.

2.4. Central Limit Theorems

4.1. The mean number of sixes is $180/6 = 30$ and the standard deviation is

$$\sqrt{(180/6)(5/6)} = 5$$

so we have

$$P(S_{180} \le 24.5) = P\left(\frac{S_{180}-30}{5} \le \frac{-5.5}{5}\right)$$
$$\approx P(\chi \le -1.1) = 1 - 0.8643 = 0.1357$$

4.2. (a) Exercise 6.5 in Chapter 1 implies

$$P(S_n/\sqrt{n} \ge K \text{ i.o.}) \ge \limsup_{n\to\infty} P(S_n/\sqrt{n} \ge K) > 0$$

so Kolmogorov's 0-1 law implies $P(S_n/\sqrt{n} \ge K \text{ i.o.}) = 1$.
(b) If $S_n/\sqrt{n} \to Z$ in probability then

$$|S_{m!}/\sqrt{m!} - S_{(m+1)!}/\sqrt{(m+1)!}| \to 0 \quad \text{in probability}$$

On the other hand, the independence of $S_{m!}$ and $S_{(m+1)!} - S_{m!}$ imply

$$P\left(1 < \frac{S_{m!}}{\sqrt{m!}} < 2, \frac{S_{(m+1)!} - S_{m!}}{\sqrt{(m+1)!}} < -3\right) \to P(1 < \chi < 2)P(\chi < -3) > 0$$

so $\liminf_{m\to\infty} P(S_{m!}/\sqrt{m!} > 1, S_{(m+1)!}/\sqrt{(m+1)!} < -1) > 0$ a contradiction.

4.3. Since $Y_m = U_m + V_m$ the first inequality is obvious. The second follows from symmetry. To prove the third we note that

$$P\left(\sum_{m=1}^{n} U_m \ge K\sqrt{n}\right) \to P(\chi \ge K/\sqrt{\text{var}(U_i)})$$

If the truncation level is chosen large then $\text{var}(U_i)$ is large and the right hand side $> 2/5$, so the third inequality holds for large n.

4.4. Intuitively, since $(2x^{1/2})' = x^{-1/2}$ and $S_n/n \to 1$ in probability

$$2(\sqrt{S_n} - \sqrt{n}) = \int_n^{S_n} \frac{dx}{x^{1/2}} \approx \frac{S_n - n}{\sqrt{n}} \Rightarrow \sigma\chi$$

To make the last calulation rigorous note that when $|S_n - n| \le n^{2/3}$ (an event with probability $\to 1$)

$$\left|\int_n^{S_n} \frac{dx}{x^{1/2}} - \frac{S_n - n}{\sqrt{n}}\right| = \left|\int_n^{S_n} \frac{1}{x^{1/2}} - \frac{1}{\sqrt{n}}\,dx\right|$$
$$\le n^{2/3}\left(\frac{1}{(n-n^{2/3})^{1/2}} - \frac{1}{n^{1/2}}\right)$$
$$= n^{2/3}\int_{n-n^{2/3}}^{n} \frac{dx}{2x^{3/2}} \le \frac{n^{4/3}}{2(n-n^{2/3})^{3/2}} \to 0$$

as $n \to \infty$.

4.5. The weak law of large numbers implies $\sum_{m=1}^n X_m^2/n\sigma^2 \to 1$. $y^{-1/2}$ is continuous at 1, so (2.3) implies

$$\left(\sigma^2 n \Bigg/ \sum_{m=1}^n X_m^2\right)^{1/2} \to 1 \quad \text{in probability}$$

and Exercise 2.11 implies

$$\frac{\sum_{m=1}^n X_m}{\sigma\sqrt{n}} \left(\frac{\sigma^2 n}{\sum_{m=1}^n X_m^2}\right)^{1/2} \Rightarrow \chi \cdot 1$$

4.6. Kolmogorov's inequality ((7.2) in Chapter 1) implies

$$P\left(\sup_{(1-\epsilon)a_n \le m \le (1+\epsilon)a_n} |S_m - S_{[(1-\epsilon)a_n]}| > \delta\sigma\sqrt{a_n}\right) \le 2\epsilon/\delta^2$$

If $X_n = S_{N_n}/\sigma\sqrt{a_n}$ and $Y_n = S_{a_n}/\sigma\sqrt{a_n}$ then it follows that

$$\limsup_{n\to\infty} P(|X_n - Y_n| > \delta) \le 2\epsilon/\delta^2$$

Since this holds for all ϵ we have $P(|X_n - Y_n| > \delta) \to 0$ for each $\delta > 0$, i.e., $X_n - Y_n \to 0$ in probability. The desired conclusion follows from the converging together lemma Exercise 2.10.

4.7. $N_t/(t/\mu) \to 1$ by (7.3) in Chapter 1, so by the last exercise

$$(S_{N_t} - \mu N_t)/(\sigma^2 t/\mu)^{1/2} \Rightarrow \chi$$

In view of Exercise 2.10 we can complete the proof now by showing

$$(S_{N_t} - t)/\sqrt{t} \to 0$$

To do this, we observe that $EY_i^2 < \infty$ implies

$$\begin{aligned} P\left(\max_{1\le m\le 2t/\mu} Y_m > \epsilon\sqrt{t}\right) &\le (2t/\mu)P(Y_1 > \epsilon\sqrt{t}) \\ &\le \frac{2}{\mu\epsilon^2} E(Y_1^2; Y_1 > \epsilon\sqrt{t}) \to 0 \end{aligned}$$

by the dominated convergence theorem. Since $P(N_t + 1 \le 2t/\mu) \to 1$ and $0 \le t - S_{N_t} \le Y_{N_t+1}$, the desired result follows from Exercise 2.10.

4.8. Recall $u = [t/\mu]$. Kolmogorov's inequality implies

$$P(|S_{u+m} - (S_u + m\mu)| > t^{2/5} \text{ for some } m \in [-t^{3/5}, t^{3/5}]) \le 2 \cdot \frac{\sigma^2 t^{3/5}}{t^{4/5}} \to 0$$

as $t \to \infty$. When the event estimated in the last equation does not occur we have

$$D_t + m\mu - t^{2/5} \le S_{u+m} - t \le D_t + m\mu + t^{2/5}$$

when $m \in [-t^{3/5}, t^{3/5}]$. When

$$\begin{aligned} m &= (-D_t + 2t^{2/5})/\mu \quad S_{u+m} > t \quad \text{so} \quad N_t \le u - D_t/\mu + 2t^{2/5}/\mu \\ m &= (-D_t - 2t^{2/5})/\mu \quad S_{u+m} < t \quad \text{so} \quad N_t \ge u - D_t/\mu - 2t^{2/5}/\mu \end{aligned}$$

The last two inequalities imply (recall $u = [t/\mu]$)

$$\frac{|N_t - (t - D_t)/\mu|}{t^{1/2}} \to 0 \text{ in probability.}$$

The central limit theorem implies $D_t/\sigma\sqrt{t/\mu} \Rightarrow \chi$ and the desired result follows from Exercise 2.10.

4.9. Let $Y_m = 1$ if $X_m > 0$ and $Y_m = -1$ if $X_m < 0$. $P(X_m \ne Y_m) = m^{-2}$ so the Borel Cantelli lemma implies $P(X_m \ne Y_m \text{ i.o.}) = 0$. The ordinary central limit theorem implies $T_n = Y_1 + \cdots + Y_n$ has $T_n/\sqrt{n} \Rightarrow \chi$, so the converging together lemma, Exercise 2.10, implies $S_n/\sqrt{n} \Rightarrow \chi$.

4.10. Let $X_{n,m} = (X_m - EX_m)/\sqrt{\mathrm{var}(S_n)}$. By definition, (i) in (4.5) holds with $\sigma^2 = 1$. Since $|X_m - EX_m| \le 2M$, the sum in (ii) is 0 for large n. The desired result follows from (4.5).

4.11. Let $X_{n,m} = X_m/\sqrt{n}$. By definition (i) in (4.5) holds with $\sigma^2 = 1$. To check (ii) we note that

$$\begin{aligned} \sum_{m=1}^n E(X_{n,m}^2; |X_{n,m}| > \epsilon) &= n^{-1} \sum_{m=1}^n E(X_m^2; |X_m| > \epsilon\sqrt{n}) \\ &\le n^{-1}(\epsilon\sqrt{n})^{-\delta} \sum_{m=1}^n E(|X|^{2+\delta}) \le C(\epsilon\sqrt{n})^{-\delta} \to 0 \end{aligned}$$

The desired result now follows from (4.5).

4.12. Let $X_{n,m} = (X_m - EX_m)/\alpha_n$. By definition (i) in (4.5) holds with $\sigma^2 = 1$. To check (ii) we note that

$$\begin{aligned} \sum_{m=1}^n E(X_{n,m}^2; |X_{n,m}| > \epsilon) &= \alpha_n^{-2} \sum_{m=1}^n E((X_m - EX_m)^2; |X_m - EX_m| > \epsilon\alpha_n) \\ &\le \epsilon^{-\delta}\alpha_n^{-(2+\delta)} \sum_{m=1}^n E(|X_m - EX_m|^{2+\delta}) \to 0 \end{aligned}$$

The desired result now follows from (4.5).

4.13. (i) If $\beta > 1$ then $\sum_j P(X_j \neq 0) < \infty$ so the Borel Cantelli lemma implies $P(X_j \neq 0 \text{ i.o.}) = 0$ and $\sum_j X_j$ exists.
(ii) $EX_j^2 = j^{2-\beta}$ so $\text{var}(S_n) \sim n^{3-\beta}/(3-\beta)$. Let $X_{n,m} = X_m/n^{(3-\beta)/2}$. By definition (i) in (4.5) holds with $\sigma^2 = 1/(3-\beta)$. To check (ii) we note that when $\beta < 1$, $(3-\beta)/2 > 1$ so eventually the sum in (ii) is 0. The desired result now follows from (4.5).
(iii) When $\beta = 1$, $E\exp(itX_j) = 1 - j^{-1}(1-\cos(jt))$. So

$$E\exp(itS_n/n) = \prod_{j=1}^n \left(1 + \frac{1}{n}\,(j/n)^{-1}\{\cos(jt/n) - 1\}\right)$$

$(1-\cos(tx))/x$ is bounded for $x \le 1$ and the Riemann sums

$$\sum_{j=1}^n \frac{1}{n}\,(j/n)^{-1}\{\cos(jt/n) - 1\} \to \int_0^1 x^{-1}\{\cos(xt) - 1\}\,dx$$

so the desired result follows from Exercise 1.1.

2.6. Poisson Convergence

6.1. (i) Clearly $d(\mu,\nu) = d(\nu,\mu)$ and $d(\mu,\nu) = 0$ if and only if $\mu = \nu$. To check the triangle inequality we note that the triangle inequality for real numbers implies

$$|\mu(x) - \nu(x)| + |\nu(x) - \pi(x)| \ge |\mu(x) - \pi(x)|$$

then sum over x.
(ii) One direction of the second result is trivial. We cannot have $\|\mu_n - \mu\| \to 0$ unless $\mu_n(x) \to \mu(x)$ for each x. To prove the converse note that if $\mu_n(x) \to \mu(x)$

$$\sum_x |\mu_n(x) - \mu(x)| = 2\sum_x (\mu(x) - \mu_n(x))^+ \to 0$$

by the dominated convergence theorem.

6.2. $(\mu(x) - \nu(x))^+ \le P(X = x, X \neq Y)$ so summing over x and noting that the events on the right hand side are disjoint shows $\|\mu - \nu\|/2 \le P(XY)$. To prove the other direction note that

$$\sum_x \mu(x) \wedge \nu(x) = \sum_x \mu(x) - (\mu(x) - \nu(x))^+ = 1 - \|\mu - \nu\|/2$$

Let I_x, $x \in \mathbf{Z}$ be disjoint subintervals of $(0, 1-\|\mu-\nu\|/2)$ with length $\mu(x)\wedge\nu(x)$. Set $X = Y = x$ on I_x. Let J_x, $x \in \mathbf{Z}$, be disjoint subintervals of $(1-\|\mu-\nu\|/2, 1)$ with length $(\mu(x) - \nu(x))^+$ and set $X = x$ on J_x. Since

$$\{\mu(x) \wedge \nu(x)\} + (\mu(x) - \nu(x))^+ = \mu(x)$$

X has distribution μ. For Y, we similarly let K_x, $x \in \mathbf{Z}$, be disjoint subintervals of $(1 - \|\mu - \nu\|/2, 1)$ with length $(\nu(x) - \mu(x))^+$ and set $Y = x$ on K_x.

6.3. Let $X_{n,m} = (\tau^n_m - \tau^n_{m-1}) - 1$. The hypotheses of (6.7) hold with

$$p_{n,m} = \frac{m-1}{n}\left(1 - \frac{m-1}{n}\right) \qquad \epsilon_{n,m} = \left(\frac{m-1}{n}\right)^2$$

for $1 \le m \le k_n$. The desired result follows from (6.7) since

$$\max_{1\le m\le k_n} p_{n,m} \le k_n/n \to 0$$

$$\sum_{m=1}^{k_n} p_{n,m} \sim \frac{1}{n}\sum_{m=1}^{k_n}(m-1) \sim \frac{k_n^2}{2n} \to \frac{\lambda^2}{2}$$

$$\sum_{m=1}^{k_n} \epsilon_{n,m} = \frac{1}{n^2}\sum_{m=1}^{k_n}(m-1)^2 \sim \frac{k_n^3}{3n^2} \to 0$$

6.4. For $m \ge 1$, $\tau^n_m - \tau^n_{m-1}$ has a geometric distribution with $p = 1 - (m-1)/n$ and hence by Example 3.5 in Chapter 1 has mean $1/p = n/(n-m+1)$ and variance $(1-p)/p^2 = n(m-1)/(n-m+1)^2$.

$$\begin{aligned}\mu_{n,k} &= \sum_{m=1}^{k} \frac{n}{n-m+1} = \sum_{j=n-k+1}^{n} \frac{n}{j} \\ &\sim n\int_{n-k}^{n} \frac{dx}{x} \sim -n\ln(1-a)\end{aligned}$$

$$\begin{aligned}\sigma^2_{n,k} &= \sum_{m=1}^{k} \frac{n(m-1)}{(n-m+1)^2} = n\sum_{j=n-k+1}^{n} \frac{n-j}{j^2} \\ &= \sum_{j=n-k+1}^{n} \frac{1-j/n}{(j/n)^2} \approx n\int_{1-a}^{1} \frac{1-x}{x^2}\,dx\end{aligned}$$

Let $t_{n,m} = \tau^n_m - \tau^n_{m-1}$ and $X_{n,m} = (t_{n,m} - Et_{n,m})/\sqrt{n}$. By design $EX_{n,m} = 0$ and (i) in (4.5) holds. To check (ii) we note that if $k/n \le b < 1$ and Y is geometric with parameter $p = 1 - b$

$$\sum_{m=1}^{k} E(X^2_{n,m}; |X_{n,m}| > \epsilon) \le bnE((Y/\sqrt{n})^2; Y > \epsilon\sqrt{n}) \to 0$$

by the dominated convergence theorem.

6.5. Iterating $P(T > t + s) = P(T > t)P(T > s)$ shows

$$P(T > ks) = P(T > s)^k$$

Letting $s \to 0$ and using $P(T > 0) = 1$ it follows that $P(T > t) > 0$ for all t. Let $e^{-\lambda} = P(T > 1)$. Using

$$P(T > 2^{1-n}) = P(T > 2^{-n})^2$$

and induction shows $P(T > 2^{-n}) = \exp(-\lambda 2^{-n})$. Then using the first relationship in the proof

$$P(T > m2^{-n}) = \exp(-\lambda m 2^{-n})$$

Letting $m2^{-n} \downarrow t$, we have $P(T > t) = \exp(-\lambda t)$.

6.6. (a) $u(r + s) = P(N_{r+s} = 0) = P(N_r = 0, N_{r+s} - N_r = 0) = u(r)u(s)$ so this follows from the previous exercise.
(b) If $N(t) - N(t-) \le 1$ for all t then for large n, $\omega \notin A_n$. So $A_n \to \emptyset$ and $P(A_n) \to 0$. Since $P(A_n) = 1 - (1 - v(1/n))^n$ we must have $nv(1/n) \to 0$ i.e., (iv) holds.

6.7. We change variables $v = r(t)$ where $v_i = t_i/t_{n+1}$ for $i \le n$, $v_{n+1} = t_{n+1}$. The inverse function is

$$s(v) = (v_1 v_{n+1}, \ldots, v_n v_{n+1}, v_{n+1})$$

which has matrix of partial derivatives $\partial s_i/\partial v_j$ given by

$$\begin{pmatrix} v_{n+1} & 0 & \ldots & 0 & v_1 \\ 0 & v_{n+1} & \ldots & 0 & v_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & v_{n+1} & v_n \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

The determinant of this matrix is v_{n+1}^n so if we let $W = (V_1, \ldots, V_{n+1}) = r(T_1, \ldots, T_{n+1})$ the change of variables formula implies W has joint density

$$f_W(v_1, \ldots, v_n, v_{n+1}) = \left(\prod_{m=1}^n \lambda e^{-\lambda v_{n+1}(v_m - v_{m-1})} \right) \lambda e^{-\lambda v_{n+1}(1 - v_n)} v_{n+1}^n$$

To find the joint density of $V = (V_1, \ldots, V_n)$ we simplify the preceding formula and integrate out the last coordinate to get

$$f_V(v_1, \ldots, v_n) = \int_0^\infty \lambda^{n+1} v_{n+1}^n e^{-\lambda v_{n+1}} \, dv_{n+1} = n!$$

for $0 < v_1 < v_2 \ldots < v_n < 1$, which is the desired joint density.

6.8. As $n \to \infty$, $T_{n+1}/n \to 1$ almost surely, so Exercise 2.11 implies

$$nV_k^n \stackrel{d}{=} nT_k/T_{n+1} \Rightarrow T_k$$

6.9. As $n \to \infty$, $T_{n+1}/n \to 1$ almost surely, so if $\epsilon > 0$ and n is large

$$n^{-1} \sum_{m=1}^{n} 1_{\{n(V_m^n - V_{m-1}^n) > x\}} \geq n^{-1} \sum_{m=1}^{n} 1_{\{T_m - T_{m-1} > x(1+\epsilon)\}} \to e^{-x(1+\epsilon)}$$

almost surely by the strong law of large numbers. A similar argument gives an upper bound of $\exp(-x(1-\epsilon))$ and the desired result follows.

6.10. Exercise 6.18 in Chapter 1 implies

$$(\log n)^{-1} \max_{1 \leq m \leq n+1} T_m - T_{m-1} \to 1$$

As $n \to \infty$, $T_{n+1}/n \to 1$ almost surely, so the desired result follows from Exercise 2.11.

6.11. Properties of the exponential distribution imply

$$P\left((n+1) \min_{1 \leq m \leq n+1} T_m - T_{m-1} > x\right) = e^{-x}$$

As $n \to \infty$, $(n+1)T_{n+1}/n^2 \to 1$ almost surely, so the desired result folllows from Exercise 2.11.

6.12. Conditioning on $N = m$, we see that if $m_0, \ldots, m_k$ add up to m then

$$\begin{aligned} P(N_0 = m_0, \ldots, N_k = m_k) &= \frac{m!}{m_0! \cdots m_k!} \cdot p_0^{m_0} \cdots p_k^{m_k} \cdot e^{-\lambda} \frac{\lambda^m}{m!} \\ &= \prod_{j=0}^{k} e^{-\lambda p_j} \frac{(\lambda p_j)^{m_j}}{m_j!} \end{aligned}$$

6.13. If the number of balls has a Poisson distribution with mean $s = n \log n - n(\log \mu)$ then the number of balls in box i, N_i, are independent with mean $s/n = \log(n/\mu)$ and hence they are vacant with probability $\exp(-s/n) = \mu/n$. Letting $X_{n,i} = 1$ if the ith box is vacant, 0 otherwise and using (6.1) it follows that the number of vacant sites converges to a Poisson with mean μ.

To prove the result for a fixed number of balls, we note that the central limit theorem implies

$$P(\text{Poisson}(s_1) < r < \text{Poisson}(s_2)) \to 1$$

Since the number of vacant boxes is decreased when the number of balls increases the desired result follows.

2.7. Stable Laws

7.1. $\log(tx)/\log t = (\log t + \log x)/\log t \to 1$ as $t \to \infty$. However, $(tx)^\epsilon/t^\epsilon = x^\epsilon$.

7.2. In the proof we showed

$$\begin{aligned} E\exp(it\hat{S}_n(\epsilon)/a_n) \to \exp\biggl(&\int_\epsilon^\infty (e^{itx}-1)\theta\alpha x^{-(\alpha+1)}\,dx \\ &+ \int_{-\infty}^{-\epsilon} (e^{itx}-1)(1-\theta)\alpha|x|^{-(\alpha+1)}\,dx\biggr) \end{aligned}$$

Since $e^{itx}-1 \sim itx$ as $x \to 0$, if we assume $\alpha < 1$ the right-hand side has a limit when $\epsilon \to 0$. Using (7.10) and (7.6) the desired result follows.

7.3. If we let $Z_m = \operatorname{sgn}(Y_m)/|Y_m|^p$, which are i.i.d., then for $x \ge 1$

$$P(|Z_m| > x) = P(|Y_m| \le x^{-1/p}) = x^{-1/p}$$

(i) When $p < 1/2$, $EZ_m^2 < \infty$ and the central limit theorem (4.1) implies

$$n^{-1/2}\sum_{m=1}^n Z_m \Rightarrow c\chi$$

(ii) When $p = 1/2$ the Z_m have the distribution considered in Example 4.8 so

$$(n\log n)^{-1/2}\sum_{m=1}^n Z_m \Rightarrow \chi$$

7.4. Let $X_1, X_2, \ldots$ be i.i.d. with $P(X_i > x) = x^{-\alpha}$ for $x \ge 1$, and let $S_n = X_1 + \cdots + X_n$. (7.7) and remarks after (7.13) imply that $(S_n - b_n)/a_n \Rightarrow Y$ where Y has a stable law with $\kappa = 1$. When $\alpha < 1$ we can take $b_n = 0$ so $Y \ge 0$.

7.5. (i) Using (3.5)

$$P(|X| > 2/u) \le u^{-1} \int_{-u}^{u} (1 - \varphi(t))\, dt$$

Using the fact that $1 - \varphi(u) \sim C|u|^\alpha$ it follows that the right hand side is $\sim C'|u|^\alpha$, and hence $P(|X| > x) \le C''|x|^{-\alpha}$ for $x \ge 1$. From the last inequality it follows that if $0 < p < \alpha$

$$\begin{aligned} E|X|^p &= \int_0^\infty p x^{p-1} P(|X| > x)\, dx \\ &\le \int_0^1 p x^{p-1}\, dx + pC'' \int_1^\infty x^{p-\alpha-1}\, dx < \infty \end{aligned}$$

(ii) Let $X_1, X_2, \ldots$ be i.i.d. with $P(X_i > x) = P(X_i < -x) = x^{-\alpha}/2$ for $x \ge 1$, and let $S_n = X_1 + \cdots + X_n$. From the convergence of $\mathcal{X}_n$ to a Poisson process we have

$$\begin{aligned} |\{m \le n : X_m > xn^{1/\alpha}\}| &\Rightarrow \text{Poisson}(x^{-\alpha}/2) \\ |\{m \le n : X_m < -n^{1/\alpha}\}| &\Rightarrow \text{Poisson}(1/2) \end{aligned}$$

Now $S_n \ge xn^{1/\alpha}$ if (i) there is at least one $X_m > xn^{1/\alpha}$ with $m \le n$, (ii) there is no $X_m < -n^{1/\alpha}$ with $m \le n$, and (iii) $\bar{S}_n(1) \ge 0$ so we have

$$\liminf_{n\to\infty} P(S_n \ge xn^{1/\alpha}) \ge \frac{x^{-\alpha}}{2} e^{-x^{-\alpha}/2} \cdot e^{-1/2} \cdot \frac{1}{2}$$

To see the inequality note that $P(i|ii) \ge P(i)$ and even if we condition on the number of $|X_m| > n^{1/\alpha}$ with $m \le n$ the distribution of $\bar{S}_n(1)$ is symmetric.

7.6. (i) Let $X_1, X_2, \ldots$ be i.i.d. with $P(X_i > x) = \theta x^{-\alpha}/2$, $P(X_i < -x) = (1-\theta)x^{-\alpha}$ for $x \ge 1$, and let $S_n = X_1 + \cdots + X_n$. (7.7) implies that $(S_n - b_n)/n^{1/\alpha} \Rightarrow Y$ and (7.15) implies $\alpha_k = \lim a_{nk}/a_n = k^{1/\alpha}$. (ii) When $\alpha < 1$ we can take $b_n = 0$ so $\beta_k = \lim_{n\to\infty}(kb_n - b_{nk})/a_n = 0$.

7.7. Let $Y_1, Y_2, \ldots$ be i.i.d. with the same distribution as Y. The previous exercise implies $n^{1/\alpha} Y \stackrel{d}{=} Y_1 + \cdots + Y_n$ so if $\psi(\lambda) = E\exp(-\lambda Y)$ then

$$\psi(\lambda)^n = \psi(n^{1/\alpha}\lambda)$$

Taking nth roots of each side we have

$$\psi(\lambda) = \psi(n^{1/\alpha}\lambda)^{1/n}$$

Setting $n = m$ and replacing λ by $\lambda m^{-1/\alpha}$ in the first equality, and then using the second gives

$$\psi(\lambda) = \psi(m^{-1/\alpha}\lambda)^m = \psi((m/n)^{-1/\alpha}\lambda)^{m/n}$$

Letting $e^{-c} = \psi(1)$ and $m/n \to \lambda^\alpha$ the desired result follows.

7.8. (i) Using the formula for the ch.f. of X and the previous exercise we have

$$E\exp(it(XY^{1/\alpha})) = E\exp(-c|t|^\alpha Y) = \exp(-c'|t|^{\alpha\beta})$$

(ii) $|W_2|$ has density $2(2\pi)^{-1/2}e^{-x^2/2}$ and $f(x) = 1/x^2$ is decreasing on $(0,\infty)$ so using Exercise 1.10 from Chapter 1, and noting $g(y) = 1/\sqrt{y}$, $g'(y) = -(1/2)y^{-3/2}$ we see that $Y = 1/|W_2|^2$ has density function

$$\frac{2}{\sqrt{2\pi}}e^{-1/2y}\cdot\frac{1}{2}y^{-3/2}$$

as claimed. Taking $X = W_1$ and $Y = 1/W_2^2$ and using (i) we see that $W_1/W_2 = XY^{1/2}$ has a symmetric stable distribution with index $2\cdot(1/2)$.

2.8. Infinitely Divisible Distributions

8.1. Suppose $Z = \text{gamma}(\alpha,\lambda)$. If $X_{n,1},\dots,X_{n,n}$ are gamma$(\alpha/n,\lambda)$ and independent then Example 4.3 in Chapter 1 implies $X_{n,1}+\cdots+X_{n,n} =_d Z$.

8.2. Suppose Z has support in $[-M,M]$. If $X_{n,1},\dots,X_{n,n}$ are independent and $Z = X_{n,1}+\cdots+X_{n,n}$ then $X_{n,1},\dots,X_{n,n}$ must have support in $[-M/n,M/n]$. So $\text{var}(X_{n,i}) \le EX_{n,i}^2 \le M^2/n^2$ and $\text{var}(Z) \le M^2/n$. Letting $n\to\infty$ we have $\text{var}(Z) = 0$.

8.3. Suppose $Z = X_{n,1}+\cdots+X_{n,n}$ where the $X_{n,i}$ are i.i.d. If φ is the ch.f. of Z and φ_n is the ch.f. of $X_{n,i}$ then $\varphi_n^n(t) = \varphi(t)$. Since $\varphi(t)$ is continuous at 0 we can pick a $\delta > 0$ so that $\varphi(t) \ne 0$ for $t \in [-\delta,\delta]$. We have supposed φ is real so taking nth roots it follows that $\varphi_n(t) \to 1$ for $t\in[-\delta,\delta]$. Using Exercise 3.20 now we conclude that $X_{n,1} \Rightarrow 0$, and (i) of (3.4) implies $\varphi_n(t)\to 1$ for all t. If $\varphi(t_0) = 0$ for some t_0 this is inconsistent with $\varphi_n^n(t_0) = \varphi(t_0)$ so φ cannot vanish.

8.4. Comparing the proof of (7.7) with the verbal description above the problem statement we see that the Lévy measure has density $1/2|x|$ for $x\in[-1,1]$, 0 otherwise.

2.9. Limit Theorems in $\mathbf{R^d}$

9.1.

$$F_i(x) = P(X_i \le x)$$
$$\lim_{n\to\infty} P(X_1 \le n, \dots, X_{i-1} \le n, X_i \le x, X_{i+1} \le n, \dots, X_d \le n)$$
$$\lim_{n\to\infty} F(n, \dots, n, x, n, \dots, n)$$

where the x is in the ith place and n's in the others.

9.2. It is clear that F has properties (ii) and (iii). To check (iv) let $G(x) = \prod_{i=1}^d F_i(x_i)$ and $H(x) = \prod_{i=1}^d F_i(x_i)(1-F_i(x_i))$. Using the notation introduced just before (iv)

$$\sum_v \operatorname{sgn}(v) G(v) = \prod_{i=1}^d F_i(b_i) - F_i(a_i)$$
$$\sum_v \operatorname{sgn}(v) H(v) = \prod_{i=1}^d \{F_i(b_i)(1 - F_i(b_i) - F_i(a_i)(1 - F_i(a_i)\}$$

To show $\sum_v \operatorname{sgn}(v)(G(v) + \alpha H(v)) \ge 0$ we note

$$\begin{aligned} F_i(b_i)(1 - F_i(b_i)) - F_i(a_i)(1 - F_i(a_i)) \\ &= \{F_i(b_i) - F_i(a_i)\}(1 - F_i(a_i)) \\ &\quad + F_i(a_i)\{(1 - F_i(b_i)) - (1 - F_i(a_i))\} \\ &= \{1 - F_i(b_i) - F_i(a_i)\}(F_i(b_i) - F_i(a_i)) \end{aligned}$$

and $|1 - F_i(b_i) - F_i(a_i)| \le 1$.

9.3. Each partial derivative kills one intergal.

9.4. If K is closed, $H = \{x : x_i \in K\}$ is closed. So

$$\limsup_{n\to\infty} P(X_{n,i} \in K) = \limsup_{n\to\infty} P(X_n \in H) \le P(X \in H) = P(X_i \in K)$$

9.5. If X has ch.f. φ then the vector $Y = (X, \dots, X)$ has ch.f.

$$\psi(t) = E\exp\left(i \sum_j t_j X\right) = \varphi\left(\sum_j t_j\right)$$

9.6. If the random variables are independent this follows from (3.1f). For the converse we note that the inversion formula implies that the joint distribution of the X_i is that of independent random variables.

9.7. Clearly, independence implies $\Gamma_{ij} = 0$ for $i \neq j$. To prove the converse note that $\Gamma_{ij} = 0$ for $i \neq j$ implies

$$\varphi_{X_1,\ldots,X_d}(t) = \prod_{j=1}^{d} \varphi_{X_j}(t_j)$$

and then use Exercise 9.6.

9.8. If $(X_1, \ldots, X_d)$ has a multivariate normal distribution then

$$\begin{aligned}\varphi_{c_1X_1+\cdots+c_dX_d}(t) &= E\exp\left(i\sum_j tc_jX_j\right)\\ &= \exp\left(-t\sum_i c_i\theta_i - \sum_i\sum_j t^2 c_i\Gamma_{ij}c_j/2\right)\end{aligned}$$

This is the ch.f. of a normal distribution with mean $c\theta^t$ and variance $c\Gamma c^t$. To prove the converse note that the assumption about the distribution of linear combinations implies

$$E\exp\left(i\sum_j c_jX_j\right) = \exp\left(-\sum_i c_i\theta_i - \sum_i\sum_j c_i\Gamma_{ij}c_j/2\right)$$

so the vector has the right ch.f.

3 Random Walks

3.1. Stopping Times

1.1. $P(X_i = 0) < 1$ rules out (i). By symmetry if (ii) or (iii) holds then the other one does as well, so (iv) is the only possibility.

1.2. The central limit theorem implies $S_n/\sqrt{n} \Rightarrow \sigma\chi$, where χ has the standard normal distribution. Exercise 6.5 from Chapter 1 implies

$$P(S_n/\sqrt{n} \geq 1 \text{ i.o.}) \geq \limsup P(S_n/\sqrt{n} \geq 1) > 0$$

So Kolmogorov's 0-1 law implies this probability is 1. This shows $\limsup S_n = \infty$ with probability 1. A similar argument shows $\liminf S_n = -\infty$ with probability one.

1.3. $\{S \wedge T = n\} = \{S = n, T \geq n\} \cup \{S \geq n, T = n\}$. The right-hand side is in $\mathcal{F}_n$ since $\{S = n\} \in \mathcal{F}_n$ and $\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$, etc. For the other result note that $\{S \vee T = n\} = \{S = n, T \leq n\} \cup \{S \leq n, T = n\}$. The right-hand side is in $\mathcal{F}_n$ since $\{S = n\} \in \mathcal{F}_n$ and $\{T \leq n\} = \cup_{m=1}^{n}\{T = m\} \in \mathcal{F}_n$, etc.

1.4. $\{S + T = n\} = \cup_{m=1}^{n-1}\{S = m, T = n - m\} \in \mathcal{F}_n$ so the result is true.

1.5. $\{Y_N \in B\} \cap \{N = n\} = \{Y_n \in B\} \cap \{N = n\} \in \mathcal{F}_n$, so $Y_N \in \mathcal{F}_N$.

1.6. If $A \in \mathcal{F}_M$ then

$$A \cap \{N = n\} = \cup_{m=1}^{n} A \cap \{M = m\} \cap \{N = n\}$$

Since $A \in \mathcal{F}_M$, $A \cap \{M = m\} \in \mathcal{F}_m \subset \mathcal{F}_n$. Thus $A \cap \{N = n\} \in \mathcal{F}_n$ and $A \in \mathcal{F}_N$.

1.7. Dividing the space into A and A^c then breaking things down according to the value of L

$$\{N = n\} = (\{L = n\} \cap A) \cup \cup_{m=1}^{n} (\{L = m\} \cap \{M = n\} \cap A^c)$$

$\{L = m\} \cap A^c$ in $\mathcal{F}_m$ whenever $A \in \mathcal{F}_m$ by the definition of $\mathcal{F}_L$. Combining this with $\{M = n\} \in \mathcal{F}_n$ proves the desired result.

1.8. (i) (1.4) implies that $P(\alpha_k < \infty) = P(\alpha < \infty)^k$ so $P(\alpha_k < \infty) \to 0$ if $P(\alpha < \infty) < 1$. (ii) (1.5) implies that $\xi_k = S_{\alpha_k} - S_{\alpha_{k-1}}$ are i.i.d., with $E\xi_k \in (0, \infty]$ so (7.2) in Chapter 1 implies $S_{\alpha_k}/k \to E\xi_1 > 0$ and $\sup_n S_n = \infty$.

1.9. By the previous exercise we get the following correspondence

$$\begin{array}{llll} P(\alpha < \infty) < 1 & P(\beta < \infty) < 1 & \sup S_n < \infty & \inf S_n > -\infty \\ P(\alpha < \infty) = 1 & P(\beta < \infty) < 1 & \sup S_n = \infty & \inf S_n > -\infty \\ P(\alpha < \infty) < 1 & P(\beta < \infty) = 1 & \sup S_n < \infty & \inf S_n = -\infty \\ P(\alpha < \infty) < 1 & P(\beta < \infty) < 1 & \sup S_n = \infty & \inf S_n = -\infty \end{array}$$

Using (1.2) now we see that the four lines correspond to (i)–(iv).

1.10. (i) A_m^n corresponds to breaking things down according to the location of the last time the minimum is attained so the A_m^n are a partition of Ω. To get the second equality we note that

$$\begin{aligned} A_m^n = \{&X_m \le 0, X_m + X_{m-1} \le 0, \dots X_m + \cdots + X_1 \le 0 \\ &X_{m+1} > 0, X_{m+1} + X_{m+2} > 0, \dots, X_{m+1} + \cdots + X_{m+n} > 0\} \end{aligned}$$

(ii) Fatou's lemma implies

$$1 \ge P(\bar\beta = \infty) \sum_{k=0}^{\infty} P(\alpha > k) = P(\bar\beta = \infty) E\alpha$$

When $P(\bar\beta = \infty) > 0$ the last inequality implies that $E\alpha < \infty$ and the desired result follows from the dominated convergence theorem. It remains to prove that if $P(\bar\beta = \infty) = 0$ then $E\alpha = \infty$. If $P(\alpha = \infty) > 0$ this is true so suppose $P(\alpha = \infty) = 0$. In this case for any fixed i $P(\alpha > n - i) \to 0$ so for any N

$$\begin{aligned} 1 &\le \liminf_{n\to\infty} \sum_{k=0}^{n-N} P(\alpha > k) P(\bar\beta > n - k) \\ &\le P(\bar\beta > N) \liminf_{n\to\infty} \sum_{k=0}^{n-N} P(\alpha > k) \end{aligned}$$

so $E\alpha \ge 1/P(\bar\beta > N)$ and since N is arbitrary the desired result follows.

1.11. (i) If $P(\bar\beta = \infty) > 0$ then $E\alpha < \infty$. In the proof of (ii) in Exercise 1.8 we observed that $S_{\alpha(k)}/k \to E\xi_1 > 0$. As $k \to \infty$, $\alpha(k)/k \to E\alpha$ so we have $\limsup S_n/n > 0$ contradicting the strong law of large numbers.

(ii) If $P(\beta = \infty) > 0$ and $P(X_i > 0) > 0$ then $P(\bar\beta = \infty) > 0$. (Take a sample path with $\beta = \infty$ and add an initial random variable that is > 0.) Thus if $P(\bar\beta = \infty) = 0$, $EX_i = 0$ and $P(X_i = 0) < 1$ then $P(\beta = \infty) = 0$. Similarly, $P(\alpha = \infty) = 0$. Now use Exercise 1.9.

1.12. Changing variables $y_k = x_1 + \cdots + x_k$ we have

$$\begin{aligned} P(T > n) &= \int_0^1 \int_0^{1-x_1} \cdots \int_0^{1-x_1 \cdots - x_{n-1}} dx_n \cdots dx_2\, dx_1 \\ &= \int \cdots \int_{0<y_1<\ldots<y_n\le 1} dy_n \cdots dy_1 = 1/n! \end{aligned}$$

since the region is $1/n!$ of the volume of $[0,1]$. From this it follows that $ET = \sum_{n=0}^\infty P(T > n) = e$ and Wald's equation implies $ES_T = ETEX_i = e/2$.

1.13. (i) The strong law implies $S_n \to \infty$ so by Exercise 1.9 we must have $P(\alpha < \infty) = 1$ and $P(\beta < \infty) < 1$. (ii) This follows from (1.4). (iii) Wald's equation implies $ES_{\alpha\wedge n} = E(\alpha \wedge n)EX_1$. The monotone convergence theorem implies that $E(\alpha \wedge n) \uparrow E\alpha$. $P(\alpha < \infty) = 1$ implies $S_{\alpha\wedge n} \to S_\alpha = 1$. (ii) and the dominated convergence theorem imply $ES_{\alpha\wedge n} \to 1$.

1.14. (i) T has a geometric distribution with success probability p so $ET = 1/p$. The first X_n that is larger than a has the distribution of X_1 conditioned on $X_1 > a$ so

$$EY_T = a + E(X - a)^+/p - c/p$$

(ii) If $a = \alpha$ the last expression reduces to α. Clearly

$$\max_{m\le n} X_m \le \alpha + \sum_{m=1}^n (X_m - \alpha)^+$$

for $n \ge 1$ subtracting cn gives the inequality in the exercise. Wald's equation implies that if $E\tau < \infty$ then

$$E\sum_{m=1}^\tau (X_m - \alpha)^+ = E\tau E(X_1 - \alpha)^+$$

Using the definition of c now we have $EY_\tau \le \alpha$.

1.15. using the definitions and then taking expected value

$$\begin{aligned} S^2_{T\wedge n} &= S^2_{T\wedge(n-1)} + (2X_nS_{n-1} + X_n^2)1_{(T\ge n)} \\ ES^2_{T\wedge n} &= ES^2_{T\wedge(n-1)} + \sigma^2 P(T \ge n) \end{aligned}$$

since $EX_n = 0$ and X_n is independent of S_{n-1} and $1_{(T \ge n)} \in \mathcal{F}_{n-1}$. [The expectation of $S_{n-1}X_n$ exists since both random variables are in L^2.] From the last equality and induction we get

$$ES_{T\wedge n}^2 = \sigma^2 \sum_{m=1}^{n} P(T \ge m)$$

$$E(S_{T\wedge n} - S_{T\wedge m})^2 = \sigma^2 \sum_{k=m+1}^{n} P(T \ge n)$$

The second equality follows from the first applied to $X_{m+1}, X_{m+2}, \ldots$. The second equality implies that $S_{T\wedge n}$ is a Cauchy sequence in L^2, so letting $n \to \infty$ in the first it follows that $ES_T^2 = \sigma^2 ET$.

3.4. Renewal Theory

4.1. Let $\bar{X}_i = X_i \wedge t$, $\bar{T}_k = \bar{X}_1 + \cdots + \bar{X}_k$, $\bar{N}_t = \inf\{k : \bar{T}_k > t\}$. Now $\bar{X}_i = X_i$ unless $X_i > t$ and $X_i > t$ implies $N_t \le i$. Now

$$t \le \bar{T}_{\bar{N}_t} \le 2t$$

the optional stopping theorem implies

$$E\bar{T}_{\bar{N}_t} = E(X_i \wedge t)E\bar{N}_t$$

and the desired result follows.

4.2. Pick $\delta > 0$ so that $P(\xi_i > \delta) = \epsilon > 0$. Let $\xi_k' = 0$ if $\xi_k \le \delta$ and $= \delta$ if $\xi_k > \delta$. Let $T_n' = \xi_1' + \cdots + \xi_n'$ and $M_t = \inf\{n : T_n' > t\}$. Clearly $T_n' \le T_n$ and so $N_t \le M_t$. M_t is the sum of $k_t = [t/\delta] + 1$ geometrics with success probability ϵ so by Example 3.5 in Chapter 1

$$\begin{aligned} EM_t &= k_t/\epsilon \\ \text{var}(M_t) &= k_t(1-\epsilon)/\epsilon^2 \\ E(M_t)^2 &= \text{var}(M_t) + (EM_t)^2 \le C(1+t^2) \end{aligned}$$

4.3. The lack of memory property of the exponential implies that the times between customers who are served is a sum of a service time with mean μ and a waiting time that is exponential with mean 1. (4.1) implies that the number of customers served up to time t, M_t satisfies $M_t/t \to 1/(1+\mu)$. (4.1) applied to the Poisson process implies $N_t/t \to 1$ a.s. so $M_t/N_t \to 1/(1+\mu)$ a.s.

4.4. Clearly if $I^\delta = \infty$ for all $\delta > 0$, h cannot be directly Riemann integrable. Suppose now that $I^\delta < \infty$ for some $\delta > 0$. Let

$$h^\delta(x) = \sup_{y \in (x-\delta, x+\delta)} h(y)$$
$$a_k^\delta = \sup_{y \in [k\delta, (k+1)\delta)} h(y)$$

If $x \in [m\delta, (m+1)\delta)$ then

$$h^\delta(x) \le a_{m-1}^\delta + a_m^\delta + a_{m+1}^\delta$$

so integrating over $[m\delta, (m+1)\delta)$ and summing over m gives

$$\int_0^\infty h^\delta(x)\,dx \le 3I^\delta < \infty$$

Now $I^\eta = \int_0^\infty a_{[x/\eta]}^\eta\,dx$, $a_{[x/\eta]}^\eta \to h(x)$ as $\eta \to 0$, and if $\eta < \delta$ then $a_{[x/\eta]}^\eta \le h^\delta(x)$ so the dominated convergence theorem implies

$$I^\eta \to \int_0^\infty h(x)\,dx$$

A similar argument shows $I_\eta \to \int_0^\infty h(x)\,dx$ and the proof is complete.

4.5. The equation comes from considering the time of the first renewal. It is easy to see using (4.10) that $h(t) = (1 - F(t))1_{(x,\infty)}$ is directly Riemann integrable whenever $\mu < \infty$ so (4.9) implies

$$H(t) \to \frac{1}{\mu} \int_0^\infty (1 - F(s))1_{(x,\infty)}(s)\,ds$$

4.6. In this case the equation is

$$H(t) = e^{-\lambda t}1_{(x,\infty)}(t) + \int_0^t H(t-s)\,\lambda e^{-\lambda s}\,ds$$

and one can check by integrating that the solution is

$$H(t) = \begin{cases} 0 & \text{if } t < x \\ e^{-\lambda x} & \text{if } t \ge x \end{cases}$$

4.7. By considering the time of the first renewal

$$H(t) = (1 - F(t+y))1_{(x,\infty)} + \int_0^t H(t-s)\,dF(s)$$

It is easy to see using (4.10) that $h(t) = 1 - F(t+y)1_{(x,\infty)}$ is directly Riemann integrable whenever $\mu < \infty$ so (4.9) implies

$$H(t) \to \frac{1}{\mu}\int_0^\infty (1 - F(y+s))1_{(x,\infty)}(s)\,ds$$

4.8. By considering ξ_1 and η_1

$$H(t) = 1 - F_1(t) + \int_0^t H(t-s)\,dF(s)$$

It follows from (4.10) that $h(t) = 1 - F_1(t)$ is directly Riemann integrable whenever $\mu_1 < \infty$ so (4.9) implies

$$H(t) \to \frac{1}{\mu}\int_0^\infty 1 - F_1(s)\,ds = \frac{\mu_1}{\mu_1 + \mu_2}$$

4.9. By considering the times of the first two renewals we see

$$H(t) = 1 - F(t) + \int_0^t H(t-s)\,dF^{2*}(s)$$

Taking $\mu_1 = \mu_2 = \mu$ in the previous exercise gives the desired result.

4.10. $V = F + V * F$ so differentiating gives the desired equality. Using (4.9) now gives

$$v(t) \to \frac{1}{\mu}\int_0^\infty f(t)\,dt = \frac{1}{\mu}$$

4.11. (i) Let $U_n = 1$ if $(X_n, \ldots, X_{n+k-1}) = (i_1, \ldots, t_k)$. Applying the strong law to the i.i.d. sequences $\{U_{i+jk}, j \geq 1\}$ for $i = 0, 1, \ldots, k-1$ shows that $N_n = \sum_{m=1}^n U_m/n \to 2^{-k}$. Since $EN_n/n \to 1/Et_2$, it follows that $Et_2 = 2^k$.
(ii) For HH we get $Et_1 = 4$ since

$$Et_1 = 1/4 + 1/4(Et_1 + 2) + 1/2(Et_1 + 1) \qquad (1/4)Et_1 = 1$$

For HT we note that if we get heads the first time then we have what we want the first time T appears so

$$Et_1 = P(H)\cdot 2 + P(T)\cdot(Et_1 + 1) \qquad (1/2)Et_1 = 3/2$$

and $Et_1 = 3$

4 Martingales

4.1. Conditional Expectation

1.1. Let $Y_i = E(X_i|\mathcal{F})$. If $A \subset B$ and $A \in \mathcal{F}$ then

$$\int_A Y_1\, dP = \int_A X_1\, dP = \int_A X_2\, dP = \int_A Y_2\, dP$$

If $A = \{Y_1 - Y_2 \geq \epsilon > 0\} \cap B$ then repeating the proof of uniqueness shows $P(A) = 0$ and $Y_1 = Y_2$ a.s. on B.

1.2. The defintion of conditional expectation implies

$$\int_B P(A|\mathcal{G})\, dP = \int_B 1_A\, dP = P(A \cap B)$$

Taking $B = G$ and $B = \Omega$ it follows that

$$\frac{\int_G P(A|\mathcal{G})\, dP}{\int P(A|\mathcal{G})\, dP} = \frac{P(G \cap A)}{P(A)} = P(G|A)$$

1.3. $a^2 1_{(|X| \geq a)} \leq X^2$ so using (1.1b) and (1.1a) gives the desired result.

1.4. (1.1b) implies $Y_M \equiv E(X_M|\mathcal{F}) \uparrow$ a limit Y. If $A \in \mathcal{F}$ then the defintion of conditional expectation implies

$$\int_A X \wedge M\, dP = \int_A Y_M\, dP$$

Using the monotone convergence theorem now gives

$$\int_A X\, dP = \int_A Y\, dP$$

1.5. (1.1b) and (1.1a) imply

$$0 \le E((X+\theta Y)^2|\mathcal{G}) = E(X^2|\mathcal{G})\theta^2 + 2E(XY|\mathcal{G})\theta + E(Y^2|\mathcal{G})$$

Now a quadratic $a\theta^2 + b\theta + c$ which is nonnegative at all rational θ must have $b^2 - 4ac \le 0$ and the desired result follows.

1.6. Let $\mathcal{F}_1 = \sigma(\{a\})$ and $\mathcal{F}_2 = \sigma(\{c\})$. Take $X(b) = 1$, $X(a) = X(c) = 0$. In this case

	a	b	c
$E(X\|\mathcal{F}_1)$	0	1/2	1/2
$E(E(X\|\mathcal{F}_1)\|\mathcal{F}_2)$	1/4	1/4	1/2

To see this is $\neq E(E(X|\mathcal{F}_2)|\mathcal{F}_1)$, we can note it is not $\in \mathcal{F}_1$.

1.7. (i) implies (ii) follows from Example 1.2. The failure of the converse follows from Example 4.2 in Chapter 1.
To prove (ii) implies (iii) we note that (1.1f), (1.3), and the assumption

$$E(XY) = EE(XY|X) = E(XE(Y|X)) = E(XEY) = EXEY$$

To see that the converse fails consider

X/Y	1	-1
1	1/4	0
0	0	1/2
-1	1/4	0

where $EX = EY = EXY = 0$ but $E(Y|X) = -1 + 2X^2$.

1.8. Let $Z = E(X|\mathcal{F}) - E(X|\mathcal{G}) \in \mathcal{F}$ and steal an equation from the proof of (1.4)

$$E\{X - E(X|\mathcal{F}) - Z\}^2 = E\{X - E(X|\mathcal{F})\}^2 + EZ^2$$

Inserting the definition of Z now gives the desired result.

1.9. $\text{var}(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$ and $E(E(X^2|\mathcal{F})) = EX^2$ we have

$$E(\text{var}(X|\mathcal{F})) = EX^2 - E(E(X|\mathcal{F})^2)$$

Since $E(E(X|\mathcal{F})) = EX$ we have

$$\text{var}(E(X|\mathcal{F})) = E(E(X|\mathcal{F})^2) - (EX)^2$$

Adding the two equations gives the desired result.

1.10. Let $\mathcal{F} = \sigma(N)$. Our first step is to prove $E(X|N) = \mu N$. Clearly (i) in the definition holds. To check (ii), it suffices to consider $A = \{N = n\}$ but in this case

$$\begin{aligned}\int_{\{N=n\}} X\,dP &= E\{(Y_1 + \cdots + Y_n)1_{(N=n)}\} \\ &= n\mu P(N = n) = \int_{\{N=n\}} \mu N\,dP\end{aligned}$$

A similar computation shows that $E(X^2|N) = \sigma^2 N + (\mu N)^2$ so

$$\text{var}(X|N) = E(X^2|N) - E(X|N)^2 = \sigma^2 N$$

and using the previous exercise we have

$$\begin{aligned}\text{var}(X) &= E(\text{var}(X|N)) + \text{var}(E(X|N)) \\ &= \sigma^2 EN + \mu^2 \text{var}(N)\end{aligned}$$

1.11. Exercise 1.8 with $\mathcal{G} = \{\emptyset, \Omega\}$ implies

$$E(Y - X)^2 + E(X - EY)^2 = E(Y - EY)^2$$

since $EY = EX$, and $EX^2 = EY^2$, $E(X - EX)^2 = E(Y - EY)^2$ and subtracting we conclude $E(Y - X)^2 = 0$.

1.12. Jensen's inequality implies

$$E(|X||\mathcal{F}) \geq |E(X|\mathcal{F})|$$

If the two expected values are equal then the two random variables must be equal almost surely, so $E(|X||\mathcal{F}) = E(X|\mathcal{F})$ a.s. on $\{E(X|\mathcal{F}) > 0\}$. Taking expected value and using the definition of conditional expectation

$$E(|X| - X; E(X|\mathcal{F}) > 0) = 0$$

This and a similar argument on $\{E(X|\mathcal{F}) < 0\}$ imply

$$\text{sgn}\,(X) = \text{sgn}\,(E(X|\mathcal{F})) \quad \text{a.s.}$$

Taking $X = Y - c$ it follows that $\text{sgn}\,(Y - c) = \text{sgn}\,(E(Y|\mathcal{G}) - c)$ a.s. for all rational c from which the desired result follows.

1.13. (i) in the definition follows by taking $h = 1_A$ in Example 1.4. To check (ii) note that the dominated convergence theorem implies that $A \to \mu(y, A)$ is a probability measure.

1.14. If $f = 1_A$ this follows from the definition. Linearity extends the result to simple f and monotone convergence to nonnegative f. Finally we get the result in general by writing $f = f^+ - f^-$.

1.15. If we fix ω and apply the ordinary Hölder inequality we get

$$\int \mu(\omega, d\omega')\,|X(\omega')Y(\omega')| \leq \left(\int \mu(\omega, d\omega')|X(\omega')|^p\right)^{1/p} \left(\int \mu(\omega, d\omega')|Y(\omega')|^q\right)^{1/q}$$

The desired result now follows from Exercise 1.14.

1.16. **Proof** As in the proof of (1.6), we find there is a set Ω_o with $P(\Omega_o) = 1$ and a family of random variables $G(q,\omega)$, $q \in \mathbf{Q}$ so that $q \to G(q,\omega)$ is nondecreasing and $\omega \to G(q,\omega)$ is a version of $P(\varphi(X) \leq q|\mathcal{G})$. Since $G(q,\omega) \in \sigma(Y)$ we can write $G(q,\omega) = H(q, Y(\omega))$. Let $F(x,y) = \inf\{G(q,y) : q > x\}$. The argument given in the proof of (1.6) shows that there is a set A_0 with $P(Y \in A_0) = 1$ so that when $y \in A_0$, F is a distribution function and that $F(x, Y(\omega))$ is a version of $P(\varphi(X) \leq x|Y)$.
Now for each $y \in A_o$, there is a unique measure $\nu(y,\cdot)$ on $(\mathbf{R}, \mathcal{R})$ so that $\nu(y, (-\infty, x]) = F(x,y))$. To check that for each $B \in \mathcal{R}$, $\nu(Y(\omega), B)$ is a version of $P(\varphi(X) \in B|Y)$, we observe that the class of B for which this statement is true (this includes the measurability of $\omega \to \nu(Y(\omega), B))$ is a λ-system that contains all sets of the form $(a_1, b_1] \cup \cdots (a_k, b_k]$ where $-\infty \leq a_i < b_i \leq \infty$, so the desired result follows from the $\pi - \lambda$ theorem. To extract the desired r.c.d. notice that if $A \in \mathcal{S}$, and $B = \varphi(A)$ then $B = (\varphi^{-1})^{-1}(A) \in \mathcal{R}$, and set $\mu(y, A) = \nu(y, B)$.

4.2. Martingales, Almost Sure Convergence

2.1. Since $X_n \in \mathcal{G}_n$ and $n \to \mathcal{G}_n$ is increasing $\mathcal{F}_n = \sigma(X_1, \ldots, X_n) \subset \mathcal{G}_n$. To check that X_n is a martingale note that $X_n \in \mathcal{F}_n$, while (1.2) implies

$$E(X_{n+1}|\mathcal{F}_n) = E(E(X_{n+1}|\mathcal{G}_n)|\mathcal{F}_n) = E(X_n|\mathcal{F}_n) = X_n$$

2.2. The fact that f is continuous implies it is bounded on bounded sets and hence $E|f(S_n)| < \infty$. Using various definitions now, we have

$$\begin{aligned} E(f(S_{n+1})|\mathcal{F}_n) &= E(f(S_n + \xi_{n+1})|\mathcal{F}_n) \\ &= \frac{1}{|B(0,1)|} \int_{B(S_n,1)} f(y)\,dy \leq f(S_n) \end{aligned}$$

2.3. Let $a_n \geq 0$ be decreasing. Then $X_n = -a_n$ is a submartingale but $X_n = a_n^2$ is a supermartingale.

2.4. Suppose $P(\xi_i = -1) = 1 - \epsilon_i$, $P(\xi_i = (1-\epsilon_i)/\epsilon_i) = \epsilon_i$. Pick $\epsilon_i > 0$ so that $\sum_i \epsilon_i < \infty$, e.g. $\epsilon_i = i^{-2}$. $P(\xi_i \neq -1 \text{ i.o.}) = 0$ so $X_n/n \to -1$ and $X_n \to -\infty$.

2.5. $A_n = \sum_{m=1}^n P(B_m|\mathcal{F}_{m-1})$.

2.6. Since $(S_n + \xi_{n+1})^2 = S_n^2 + 2S_n\xi_{n+1} + \xi_{n+1}^2$ and ξ_{n+1} is independent of $\mathcal{F}_n$, we have

$$\begin{aligned}
E(S_{n+1}^2 - s_{n+1}^2|\mathcal{F}_n) &= S_n^2 + 2S_n E(\xi_{n+1}|\mathcal{F}_n) + E(\xi_{n+1}^2|\mathcal{F}_n) - s_{n+1}^2 \\
&= S_n^2 + 0 + \sigma_{n+1}^2 - s_{n+1}^2 = S_n^2 - s_n^2
\end{aligned}$$

2.7. Clearly, $X_n^{(k)} \in \mathcal{F}_n$. The independence of the ξ_i, (4.8) in Chapter 1 and the triangle inequality imply $E|X_n^{(k)}| < \infty$. Since $X_{n+1}^{(k)} = X_n^{(k)} + X_n^{(k-1)}\xi_{n+1}$ taking conditional expectation and using (1.3) gives

$$E(X_{n+1}^{(k)}|\mathcal{F}_n) = X_n^{(k)} + X_n^{(k-1)} E(\xi_{n+1}|\mathcal{F}_n) = X_n^{(k)}$$

2.8. Clearly, $X_n \vee Y_n \in \mathcal{F}_n$. Since $|X_n \vee Y_n| \leq |X_n| + |Y_n|$, $E|X_n \vee Y_n| < \infty$. Use monotonicity (1.1b) and the defintion of supermartingale

$$\begin{aligned}
E(X_{n+1} \vee Y_{n+1}|\mathcal{F}_n) &\geq E(X_{n+1}|\mathcal{F}_n) \geq X_n \\
E(X_{n+1} \vee Y_{n+1}|\mathcal{F}_n) &\geq E(Y_{n+1}|\mathcal{F}_n) \geq Y_n
\end{aligned}$$

From this it follows that $E(X_{n+1} \vee Y_{n+1}|\mathcal{F}_n) \geq X_n \vee Y_n$.

2.9. (i) Clearly $X_n \in \mathcal{F}_n$, and $E|X_n| < \infty$. Using (1.3) now we have

$$E(X_{n+1}|\mathcal{F}_n) = X_n E(Y_{n+1}|\mathcal{F}_n) = X_n$$

since Y_{n+1} is independent of $\mathcal{F}_n$ and has $EY_n = 1$.
(ii) (2.11) implies $X_n \to X_\infty < \infty$ a.s. We want to show that if $P(Y_m = 1) < 1$ then $X_\infty = 0$ a.s. To do this let ϵ be chosen so that $P(|Y_m - 1| > \epsilon) = \eta > 0$. Now if $\delta > 0$

$$P(|X_{n+1} - X_n| \geq \delta\epsilon) \geq P(X_n \geq \delta) P(|Y_{n+1} - 1| > \epsilon)$$

The almost sure convergence of $X_n \to X_\infty$ implies the left hand side $\to 0$ so $P(X_n \geq \delta) \to 0$. This shows $X_n \to 0$ in probability, so $X_\infty = 0$ a.s.

(iii) Applying Jensen's inequality with $\varphi(x) = \log x$ to $Y_i \vee \delta$ and then letting $\delta \to 0$ we have $E \log Y_i \in [-\infty, 0]$. Applying the strong law of large numbers, (7.3) in Chapter 1, to $-\log Y_i$ we have

$$\frac{1}{n} \log X_n = \frac{1}{n} \sum_{m=1}^{n} \log Y_m \to E \log Y_1 \in [-\infty, 0]$$

2.10. Our first step is to prove

Lemma. When $|y| \le 1/2$, $y - y^2 \le \log(1+y) \le y$.

Proof $1 + y \le e^y$ implies $\log(1+y) \le y$ for all y. Expanding $\log(1+y)$ in power series gives

$$\log(1+y) = y - \frac{y^2}{2} + \frac{y^3}{3} - \frac{y^4}{4} + \cdots$$

When $|y| \le 1/2$

$$\left| -\frac{y^2}{2} + \frac{y^3}{3} - \frac{y^4}{4} + \cdots \right| \le \frac{y^2}{2} \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \right) = y^2$$

which completes the proof. □

Now if $\sum_{m=1}^{\infty} |y_m| < \infty$ we have $|y_m| \le 1/2$ for $m \ge M$ so $\sum_{m=1}^{\infty} y_m^2 < \infty$ and if $N \ge M$ the lemma implies

$$\sum_{m=N}^{\infty} y_m - y_m^2 \le \sum_{m=N}^{\infty} \log(1+y_m) \le \sum_{m=N}^{\infty} y_m$$

The last inequality shows $\sum_{m=N}^{\infty} \log(1+y_m) \to 0$ as $N \to \infty$, so $\prod_{m=1}^{\infty}(1+y_m)$ exists.

2.11. Let $W_n = X_n / \prod_{m=1}^{n-1}(1+Y_m)$. Clearly $W_n \in \mathcal{F}_n$, $E|W_n| \le E|X_n| < \infty$. Using (1.3) now and the definition gives

$$\begin{aligned} E(W_{n+1}|\mathcal{F}_n) &= \frac{1}{\prod_{m=1}^{n}(1+Y_m)} E(X_{n+1}|\mathcal{F}_n) \\ &\le \frac{1}{\prod_{m=1}^{n-1}(1+Y_m)} X_n = W_n \end{aligned}$$

Thus W_n is a nonnegative supermartingale and (2.11) implies that $W_n \to W_\infty$ a.s. The assumption $\sum_m Y_m < \infty$ implies that $\prod_{m=1}^{n-1}(1+Y_m) \to \prod_{m=1}^{\infty}(1+Y_m)$, so $X_n \to W_\infty \prod_{m=1}^{\infty}(1+Y_m)$ a.s.

2.12. Let S_n be the random walk from Exercise 2.2. That exercise implies $f(S_n) \geq 0$ is a supermartingale, so (2.11) implies $f(S_n)$ converges to a limit almost surely. If f is continuous and noncontant then there are constants $\alpha < \beta$ so that $G = \{f < \alpha\}$ and $H = \{f > \beta\}$ are nonempty open sets. Since the random walk S_n has mean 0 and finite variance, (2.7) and (2.8) in Chapter 3 imply that S_n visits G and H infinitely often. This implies

$$\liminf f(S_n) \leq \alpha < \beta \leq \limsup f(S_n)$$

a contradiction which implies f must be constant.

2.13. Using the definition of Y_{n+1}, the inequality $X^1_N \geq X^2_N$, the fact that $\{N \leq n\} \in \mathcal{F}_n$ (and hence $\{N > n\} \in \mathcal{F}_n$), and finally the supermartingale property we have

$$\begin{aligned}
E(Y_{n+1}|\mathcal{F}_n) &= E(X^1_{n+1}1_{(N>n+1)} + X^2_{n+1}1_{(N\leq n+1)}|\mathcal{F}_n) \\
&\leq E(X^1_{n+1}1_{(N>n)} + X^2_{n+1}1_{(N\leq n)}|\mathcal{F}_n) \\
&= E(X^1_{n+1}|\mathcal{F}_n)1_{(N>n)} + E(X^2_{n+1}|\mathcal{F}_n)1_{(N\leq n)} \\
&\leq X^1_n 1_{(N>n)} + X^2_n 1_{(N\leq n)} = Y_n
\end{aligned}$$

2.14. (i) To start we note that $Z^1_n \equiv 1$ is clearly a supermartingale. For the induction step we have to consider two cases $k = 2j$ and $k = 2j+1$. In the case $k = 2j$ we use the previous exercise with $X^1 = Z^{2j-1}$, $X^2 = (b/a)^{j-1}(X_n/a)$, and $N = N_{2j-1}$. Clearly these are supermartingales. To check the other condition we note that since $X_N \leq a$ we have $X^1_N = (b/a)^{j-1} \geq X^2_N$.
In the case $k = 2j+1$ we use the previous exercise with $X^1 = Z^{2j}$ and $X^2 = (b/a)^j$, and $N = N_{2j}$. Clearly these are supermartingales. To check the other condition we note that since $X_N \geq b$ we have $X^1_N \geq (b/a)^j = X^2_N$.
(ii) Since $Z^{2k} - n$ is a supermartingale, $EY_0 \geq EY_{n\wedge N_{2k}}$. Letting $n \to \infty$ and using Fatou's lemma we have

$$E(min(X_0/a, 1) = EY_0 \geq E(Y_{N_{2k}}; N_{2k} < \infty) = (b/a)^k P(U \geq k)$$

4.3. Examples

3.1. Let $N = \inf\{n : X_n > M\}$. $X_{N\wedge n}$ is a submartingale with

$$X^+_{N\wedge n} \leq M + \sup_n \xi^+_n$$

so $\sup_n EX^+_{N\wedge n} < \infty$. (2.10) implies $X_{N\wedge n} \to$ a limit so X_n converges on $\{N = \infty\}$. Letting $M \to \infty$ and recalling we have assumed $\sup_n \xi_n^+ < \infty$ gives the desired conclusion.

3.2. Let $U_1, U_2, \ldots$ be i.i.d. uniform on (0,1). If $X_n = 0$ then $X_{n+1} = 1$ if $U_{n+1} \geq 1/2$, $X_{n+1} = -1$ if $U_{n+1} < 1/2$. If $X_n \neq 0$ then $X_{n+1} = 0$ if $U_{n+1} > n^{-2}$, while $X_{n+1} = n^2 X_n$ if $U_{n+1} < n^{-2}$. [We use the sequence of uniforms because it makes it clear that "the decisions at time $n+1$ are independent of the past."] $\sum_n 1/n^2 < \infty$ so the Borel Cantelli lemma implies that eventually we just go from 0 to ± 1 and then back to 0 again, so $\sup |X_n| < \infty$.

3.3. Modify the previous example so that if $X_n = 0$ then $X_{n+1} = 1$ on $U_{n+1} > 3/4$, $X_{n+1} = -1$ if $U_{n+1} < 1/4$, $X_{n+1} = 0$ otherwise. The previous argument shows that eventually X_n is indistiguishable from the Markov chain with transition matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1 & 0 \end{pmatrix}$$

This chain converges to its stationary distribution which assigns mass 2/3 to 0 and 1/6 each to -1 and 1.

3.4. Let $W_n = X_n - \sum_{m=1}^{n-1} Y_m$. Clearly $W_n \in \mathcal{F}_n$ and $E|W_n| < \infty$. Using the linearity of conditional expectation, $\sum_{m=1}^n Y_m \in \mathcal{F}_n$, and the defintion we have

$$\begin{aligned} E(W_{n+1}|\mathcal{F}_n) &\leq E(X_{n+1}|\mathcal{F}_n) - \sum_{m=1}^{n} Y_m \\ &\leq X_n - \sum_{m=1}^{n-1} Y_m = W_n \end{aligned}$$

Let M be a large number and $N = \inf\{k : \sum_{m=1}^k Y_m > M\}$. Now $W_{N\wedge n}$ is a supermartingale by (2.8) and

$$W_{N\wedge n} = X_{N\wedge n} - \sum_{m=1}^{(N\wedge n)-1} Y_m$$

so applying (2.11) to $M + W_{N\wedge n}$ we see that $\lim_{n\to\infty} W_{N\wedge n}$ exists and hence $\lim_{n\to\infty} W_n$ exists on $\{N = \infty\} \subset \{\sum_m Y_m \leq M\}$. As $M \uparrow \infty$ the right hand side $\uparrow \Omega$, so the proof is complete.

3.5. Let $X_m \in \{0, 1\}$ be independent with $P(X_m = 1) = p_m$. Then

$$\prod_{m=1}^{\infty} (1 - p_m) = P(X_m = 0 \text{ for all } m \geq 1)$$

(i) If $\sum_{m=1}^{\infty} p_m = \infty$ then $P(X_m = 1 \text{ i.o.}) = 1$ so the product is 0. (ii) If $\sum_{m=1}^{\infty} p_m < \infty$ then $\sum_{m=M}^{\infty} p_m < 1$ for large M, so $P(X_m = 0 \text{ for all } m \geq M) > 0$ and since $p_m < 1$ for all m, $P(X_m = 0 \text{ for all } m \geq 1) > 0$.

3.6. Let $p_1 = P(A_1)$ and $p_n = P(A_n | \cap_{m=1}^{n-1} A_m^c)$.

$$\prod_{m=1}^{n} (1 - p_m) = P(\cap_{m=1}^{n} A_m^c)$$

so letting $n \to \infty$ and using (i) of Exercise 3.5 gives the desired result.

3.7. Suppose $I_{k,n} = I_{1,n+1} \cup \cdots \cup I_{m,n+1}$. If $\nu(I_{j,n+1}) > 0$ for all j we have by using the various definitions that

$$\begin{aligned}\int_{I_{k,n}} X_{n+1}\, dP &= \sum_{j=1}^{m} \frac{\mu(I_{j,n+1})}{\nu(I_{j,n+1})} \nu(I_{j,n+1}) \\ &= \mu(I_{k,n}) = \frac{\mu(I_{k,n})}{\nu(I_{k,n})} \nu(I_{k,n}) = \int_{I_{k,n}} X_n\, dP\end{aligned}$$

If $\nu(I_{j,n+1}) = 0$ for some j then the first sum should be restricted to the j with $\nu(I_{j,n+1}) > 0$. If $\mu \ll \nu$ the second $=$ holds but in general we have only $\leq$.

3.8. If μ and ν are σ-finite we can find a sequence of sets $\Omega_k \uparrow \Omega$ so that $\mu(\Omega_k)$ and $\nu(\Omega_k)$ are $< \infty$ and $\nu(\Omega_1) > 0$. By restricting our attention to Ω_k we can assume that μ and ν are finite measures and by normalizing that that ν is a probability measure. Let $\mathcal{F}_n = \sigma(\{B_m : 1 \leq m \leq n\})$ where $B_m = A_m \cap \Omega_k$. Let μ_n and ν_n be the restrictions of μ and ν to $\mathcal{F}_n$, and let $X_n = d\mu_n/d\nu_n$. (3.3) implies that $X_n \to X$ ν-a.s. where

$$\mu(A) = \int_A X\, d\nu + \mu(A \cap \{X = \infty\})$$

Since $X < \infty$ ν a.s. and $\mu \ll \nu$, the second term is 0, and we have the desired Radon Nikodym derivative.

3.9. (i) $\int \sqrt{q_m}\, dG_m = \sqrt{(1-\alpha_m)(1-\beta_m)} + \sqrt{\alpha_m \beta_m}$ so the necessary and sufficient condition is

$$\prod_{m=1}^{\infty} \sqrt{(1-\alpha_m)(1-\beta_m)} + \sqrt{\alpha_m \beta_m} > 0$$

(ii) Let $f_p(x) = \sqrt{(1-p)(1-x)} + \sqrt{px}$. Under our assumptions on α_m and β_m, Exercise 3.5 implies $\prod_{m=1}^{\infty} f_{\beta_m}(\alpha_m) > 0$ if and only if $\sum_{m=1}^{\infty} 1 - f_{\beta_m}(\alpha_m) < \infty$.

Our task is then to show that the last condition is equivalent to $\sum_{m=1}^{\infty}(\alpha_m - \beta_m)^2 < \infty$. Differentiating gives

$$\begin{aligned} f_p'(x) &= \frac{1}{2}\sqrt{\frac{p}{x}} - \frac{1}{2}\sqrt{\frac{1-p}{1-x}} \\ f_p''(x) &= -\frac{1}{4}\frac{\sqrt{p}}{x^{3/2}} - \frac{1}{4}\frac{\sqrt{1-p}}{(1-x)^{3/2}} < 0 \end{aligned}$$

If $\epsilon \le x, p \le 1-\epsilon$ then

$$-A \equiv -\frac{\sqrt{\epsilon}}{2(1-\epsilon)^{3/2}} \ge f_p''(x) \ge -\frac{\sqrt{1-\epsilon}}{2\epsilon^{3/2}} \equiv -B$$

We have $f_p'(p) = 0$ so integrating gives

$$\begin{aligned} 0 \ge f_p(x) - f_p(p) &= \int_p^x f_p'(y)\,dy \\ &= \int_p^x \int_p^y f_p''(z)\,dz\,dy \ge -B(x-p)^2/2 \end{aligned}$$

A similar argument establishes an upper bound of $-A(x-p)^2/2$ so using $f_p(p) = 1$ we have

$$A(x-p)^2/2 \le 1 - f_p(x) \le B(x-p)^2/2$$

3.10. The Borel Cantelli lemmas imply that when $\sum \alpha_n < \infty$ μ concentrates on points in $\{0,1\}^N$ with finitely many ones while $\sum \beta_n = \infty$ implies ν concentrates on points in $\{0,1\}^N$ with infinitely many ones.

3.11. Let $U_1, U_2, \ldots$ be i.i.d. uniform on $(0,1)$. Let $X_n = 1$ if $U_n < \alpha_n$ and 0 otherwise. Let $Y_n = 1$ if $U_n < \beta_n$ and 0 otherwise. Then $X_1, X_2, \ldots$ are independent with distribution F_n and $Y_1, Y_2, \ldots$ are independent with distribution G_n. If $\sum |\alpha_n - \beta_n| < \infty$ then for large N $\sum_{n \ge N} |\alpha_n - \beta_n| < 1$ which implies $P(X_n = Y_n \text{ for } n \ge N) > 0$. Since $0 < \alpha_n < \beta_n < 1$ it follows that $P(X_n = Y_n \text{ for } n \ge 1) > 0$. This shows that the measures μ and ν induced by the sequences $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ are not mutually singular so by the Kakutani dichotomy they must be absolutely continuous.

3.12. Let $\theta = P(\lim Z_n/\mu^n = 0)$. By considering what happens at the first step we see

$$\theta = \sum_{k=0}^{\infty} p_k \theta^k = \varphi(\theta)$$

Since we assumed $\theta < 1$, it follows from (b) in the proof of (3.10)) that $\theta = \rho$. It is clear that

$$\{Z_n > 0 \text{ for all } n\} \supset \{\lim Z_n/\mu^n > 0\}$$

Since each set has probability $1 - \rho$ they must be equal a.s.

3.13. ρ is a root of

$$x = \frac{1}{8} + \frac{3}{8}x + \frac{3}{8}x^2 + \frac{1}{8}x^3$$

Subtracting x from each side and the multiplying by 8 this becomes

$$0 = x^3 + 3x^2 - 5x + 1 = (x-1)(x^2 + 4x - 1)$$

The quadratic has roots $-2 \pm \sqrt{5}$ so $\rho = \sqrt{5} - 2$.

4.4. Doob's inequality, L^p convergence when $p > 1$

4.1. Since $\{N = k\}$, using $X_j \le E(X_k|\mathcal{F}_j)$ and the definition of conditional expectation gives that

$$E(X_N; N = j) = E(X_j; N = j) \le E(X_k; N = j)$$

Summing over j now we have $EX_N \le EX_k$.

4.2. Let $K_n = 1_{M<n\le N}$. $\{M < n \le N\} = \{M \le n-1\} \cap \{N < n\}^c$ so K_n is predictable. $Y_n = (K \cdot X)_n = X_{N\wedge n} - X_{M\wedge n}$ is a submartingale. Taking $n = k$ and $n = 0$ we have $EX_N - EX_M \ge 0$.

4.3. Exercise 1.7 in Chapter 3 implies that for $A \in \mathcal{F}_M$

$$L = \begin{cases} M & \text{on } A \\ N & \text{on } A^c \end{cases}$$

is a stopping time. Using Exercise 4.2 now gives $EX_L \le EX_N$. Since $L = M$ on A and $L = N$ on A^c, subtracting $E(X_N; A^c)$ from each side and using the definition of conditional expectation gives

$$E(X_M; A) \le E(X_M; A) = E(E(X_N|\mathcal{F}_M); A)$$

Since this holds for all $A \in \mathcal{F}_M$ it follows that $X_M \le E(X_N|\mathcal{F}_M)$.

4.4. Let $A = \{\max_{1\le m\le n} |S_m| > x\}$ and $N = \inf\{m : |S_m| > x \text{ or } m = n\}$. Since N is a stopping time with $P(N \le n) = 1$, (4.1) implies

$$0 = E(S_N^2 - s_N^2) \le (x+K)^2 P(A) + (x^2 - \text{var}(S_n))P(A^c)$$

since on A, $|S_N| \le x + K$ and and on A^c, $S_N^2 = S_n^2 \le x^2$. Letting $P(A) = 1 - P(A^c)$ and rearranging we have

$$(x+K)^2 \ge (\text{var}(S_n) - x^2 + (x+K)^2)P(A^c) \ge \text{var}(S_n)P(A^c)$$

4.5. If $-c < \lambda$ then using an obvious inequality, then (4.1) and the fact $EX_n = 0$

$$\begin{aligned} P\left(\max_{1\le m\le n} X_m \ge \lambda\right) &\le P\left(\max_{1\le m\le n} (X_n + c)^2 \ge (c+\lambda)^2\right) \\ &\le \frac{E(X_n+c)^2}{(c+\lambda)^2} = \frac{EX_n^2 + c^2}{(c+\lambda)^2} \end{aligned}$$

To optimize the bound we differentiate with respect to c and set the result to 0 to get

$$\begin{aligned} -2\frac{EX_n^2 + c^2}{(c+\lambda)^3} + \frac{2c}{(c+\lambda)^2} &= 0 \\ 2c(c+\lambda) - 2(EX_n^2 + c^2) &= 0 \end{aligned}$$

so $c = EX_n^2/\lambda$. Plugging this into the upper bound and then multiplying top and bottom by λ^2

$$\frac{EX_n^2 + (EX_n^2/\lambda)^2}{\left(\frac{EX_n^2}{\lambda} + \lambda\right)^2} = \frac{(\lambda^2 + EX_n^2)(EX_n^2)}{(EX_n^2 + \lambda^2)^2}$$

4.6. Since X_n^+ is a submartingale and x^p is increasing and convex it follows that

$$(X_m^+)^p \le \{E(X_n^+|\mathcal{F}_m)\}^p \le E((X_n^+)^p|\mathcal{F}_m)$$

Taking expected value now we have $E(X_m^+)^p < \infty$ and it follows that

$$E\bar{X}_n^p \le \sum_{m=1}^n E(X_m^+)^p < \infty$$

4.7. Arguing as in the proof of (4.3)

$$\begin{aligned} E(\bar{X}_n \wedge M) &\le 1 + \int_1^\infty P(\bar{X}_n \wedge M \ge \lambda)\, d\lambda \\ &\le 1 + \int_1^\infty \lambda^{-1} \int X_n^+ 1_{(\bar{X}_n \wedge M \ge \lambda)}\, dP \\ &\le 1 + \int X_n^+ \int_1^{\bar{X}_n \wedge M} \lambda^{-1}\, d\lambda\, dP \\ &= 1 + \int X_n^+ \log(\bar{X}_n \wedge M)\, dP \end{aligned}$$

(ii) $a \log b \le a \log a + b/e \le a \log^+ a + b/e$

Proof The second inequality is trivial. To prove the first we note that it is trivial if $b < a$. Now for fixed a the maximum value of $(a\log b - a\log a)/b$ for $b \ge a$ occurs when

$$0 = \left(\frac{a\log b - a\log a}{b}\right)' = \frac{a}{b^2} - \frac{a\log b - a\log a}{b^2}$$

i.e., when $b = ae$. In this case the ratio $= 1/e$. □

(iii) To complete the proof of (4.4) now we use the Lemma to get

$$E(\bar{X}_n \wedge M) \le 1 + E(X_n^+ \log^+ X_n^+) + E(\bar{X}_n \wedge M)/e$$

Since $E(\bar{X}_n \wedge M)/e < \infty$ we can subtract this from both sides and then divide by $(1 - e^{-1})$ to get

$$E(\bar{X}_n \wedge M) \le (1 + e^{-1})^{-1}(1 + EX_n^+ \log^+ X_n^+)$$

Letting $M \to \infty$ and using the dominated convergence theorem gives the desired result. □

4.8. (4.6) implies that $E(X_m Y_{m-1}) = E(X_{m-1}Y_{m-1})$. Interchanging X and Y, we have $E(X_{m-1}Y_m) = E(X_{m-1}Y_{m-1})$, so

$$\begin{aligned}
E(X_m - X_{m-1})&(Y_m - Y_{m-1}) \\
&= EX_mY_m - EX_mY_{m-1} - EX_{m-1}Y_m - EX_{m-1}Y_{m-1} \\
&= EX_mY_m + (-2+1)EX_{m-1}Y_{m-1}
\end{aligned}$$

Summing over $m - 1$ to n now gives the desired result.

4.9. Taking $X = Y$ in the previous exercise

$$EX_n^2 = EX_0^2 + \sum_{m=1}^{n} E\xi_m^2$$

So our assumptions imply $\sup_n EX_n^2$ and (4.5) implies $X_n \to X_\infty$ in L^2.

4.10. Applying the previous exercise to the martingale $Y_n = \sum_{m=1}^n \xi_m/b_m$ we have $Y_m \to Y_\infty$ a.s and in L^2, so Kronecker's lemma ((8.5) in Chapter 1) implies $(X_n - X_0)/b_n \to 0$ a.s.

4.11. $S_{N\wedge n}$ is a martingale with increasing process $\sigma^2(N \wedge n)$. If $EN^{1/2} < \infty$ then $E\sup_n |S_{N\wedge n}| < \infty$. (4.1) implies that $ES_{N\wedge n} = 0$. Letting $n \to \infty$ and using the dominated convergence theorem, $ES_N = 0$.

4.5. Uniform Integrability, Convergence in $\mathbf{L^1}$

5.1. Let $\epsilon_M = \sup\{x/\varphi(x) : x \geq M\}$. For $i \in I$

$$E(|X_i|; |X_i| > M) \leq \epsilon_M E(\varphi(|X_i|); |X_i| > M) \leq C\epsilon_M$$

and $\epsilon_M \to 0$ as $M \to \infty$.

5.2. Let $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. (5.5) implies that

$$E(\theta|\mathcal{F}_n) \to E(\theta|\mathcal{F}_\infty)$$

To complete the proof it suffices to show that $\theta \in \mathcal{F}_\infty$. To do this we observe that the strong law implies $(Y_1 + \cdots + Y_n)/n \to \theta$ a.s.

5.3. Let $a_{n,k} = \{f((k+1)2^{-n}) - f(k2^{-n})\}/2^{-n}$. Since $I_{k,n} = I_{2k,n+1} \cup I_{2k+1,n+1}$, it follows from Example 1.3 that on $I_{k,n}$

$$E(X_{n+1}|\mathcal{F}_n) = \frac{a_{2k,n+1} + a_{2k+1,n}}{2} = a_{k,n} = X_n$$

Since $0 \leq X_n \leq K$ it is uniformly integrable, so (5.5) implies $X_n \to X_\infty$ a.s. and in L^1, and (5.5) implies $X_n = E(X_\infty|\mathcal{F}_n)$. This implies that

$$(*) \qquad f(b) - f(a) = \int_a^b X_\infty(\omega)\, d\omega$$

holds when $a = k2^{-n}$ and $b = (k+1)2^{-n}$. Adding a finite number of these equations we see $(*)$ holds when $a = k2^{-n}$ and $b = m2^{-n}$ where $m > k$. Taking limits and using the fact that f is continuous and $|X(\omega)| \leq K$ we have $(*)$ for all a and b.

5.4. $E(f|\mathcal{F}_n)$ is uniformly integrable so it converges a.s. and in L^1 to $E(f|\mathcal{F}_\infty)$, which is $= f$ since $f \in \mathcal{F}_\infty$.

5.5. On $\{\liminf_{n\to\infty} X_n \leq M\}$, $X_n \leq M+1$ i.o. so

$$P(D|X_1, \ldots, X_n) \geq \delta(M+1) > 0 \quad \text{i.o.}$$

Since the right hand side $\to 1_D$, we must have

$$D \supset \{\liminf_{n\to\infty} X_n \leq M\}$$

Letting $M \to \infty$, we have $D \supset \{\liminf_{n\to\infty} X_n < \infty\}$ a.s.

5.6. If $p_0 > 0$ then $P(Z_{n+1} = 0|Z_1, \dots, Z_n) \geq p_0^k$ on $\{Z_n \leq k\}$ so Exercise 5.5 gives the desired result.

5.7.

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= X_n(\alpha + \beta X_n) + (1 - X_n)\beta X_n \\ &= \alpha X_n + \beta X_n = X_n \end{aligned}$$

so X_n is a martingale. $0 \leq X_n \leq 1$ so (5.5) implies $X_n \to X_\infty$ a.s. and in L^1. When $X_n = x$, X_{n+1} is either $\alpha + \beta x$ or βx so convergence to $x \in (0,1)$ is impossible. The constancy of martingale expectation and the bounded convergence theorem imply

$$\theta = EX_0 = EX_n \to EX_\infty$$

Since $X_\infty \in \{0,1\}$ it follows that $P(X_\infty = 1) = \theta$ and $P(X_\infty = 0) = 1 - \theta$.

5.8. The trinagle inequality implies

$$E|E(Y_n|\mathcal{F}_n) - E(Y|\mathcal{F})| \leq E|E(Y_n|\mathcal{F}_n) - E(Y|\mathcal{F}_n)| + E|E(Y|\mathcal{F}_n) - E(Y|\mathcal{F})|$$

Jensen's inequality and (1.1f) imply

$$E|E(Y_n|\mathcal{F}_n) - E(Y|\mathcal{F}_n)| \leq EE(|Y_n - Y|\,|\mathcal{F}_n) = E|Y_n - Y| \to 0$$

since $Y_n \to Y$ in L^1. For the other term we note (5.6) implies

$$E|E(Y|\mathcal{F}_n) - E(Y|\mathcal{F})| \to 0$$

4.6. Backwards Martingales

6.1. The L^p maximal inequality (4.3) implies

$$E\left(\sup_{-n \leq m \leq 0} |X_m|^p\right) \leq \left(\frac{p}{p-1}\right)^p E|X_0|^p$$

Letting $n \to \infty$ it follows that $\sup_m |X_m| \in L^p$. Since $|X_n - X_{-\infty}|^p \leq 2\sup|X_n|^p$ it follows from the dominated convergence theorem that $X_n \to X_{-\infty}$ in L^p.

6.2. Let $W_N = \sup\{|Y_n - Y_m| : n, m \leq -N\}$. $W_N \leq 2Z$ so $EW_N < \infty$. Using monotonicity (1.1b) and applying (6.3) to W_N gives

$$\limsup_{n \to -\infty} E(|Y_n - Y_{-\infty}||\mathcal{F}_n) \leq \lim_{n \to -\infty} E(W_N|\mathcal{F}_n) = E(W_N|\mathcal{F}_{-\infty})$$

The last result is true for all N and $W_N \downarrow 0$ as $N \uparrow \infty$, so (1.1c) implies $E(W_N|\mathcal{F}_{-\infty}) \downarrow 0$, and Jensen's inequality gives us

$$|E(Y_n|\mathcal{F}_n) - E(Y_{-\infty}|\mathcal{F}_n)| \le E(|Y_n - Y_{-\infty}||\mathcal{F}_n) \to 0 \quad \text{a.s. as } n \to -\infty$$

(6.2) implies $E(Y_{-\infty}|\mathcal{F}_n) \to E(Y_{-\infty}|\mathcal{F}_{-\infty})$ a.s. The desired result follows from the last two conclusions and the triangle inequality.

6.3. By exchangeability all outcomes with m 1's and $(n-m)$ 0's have the same probability. If we call this r then by counting the number of outcomes in the two events we have

$$P(S_n = m) = \binom{n}{m} r$$
$$P(X_1 = 1, \ldots, X_k = 1, S_n = m) = \binom{n-k}{m-k} r$$

Dividing the first equation by the second gives the desired result.

6.4. Exchangeability implies

$$0 \le \binom{n}{2}^{-1} E(X_1 + \cdots + X_n)^2 = 2E(X_1X_2) + \binom{n}{2}^{-1} nEX_1^2$$

Letting $n \to \infty$ now gives the desired result.

6.5. Let $\varphi(x,y) = (x-y)^2$ and define $A_n(\varphi)$ as in (6.5). We have $A_n(\varphi) = E(\varphi(X_1, X_2)|\mathcal{E}_n)$ so it follows from (6.3) that

$$E(\varphi(X_1, X_2)|\mathcal{E}_n) \to E(\varphi(X_1, X_2)|\mathcal{E}) = E\varphi(X_1, X_2)$$

since $\mathcal{E}$ is trivial.

4.7. Optional Stopping Theorems

7.1. Let $N = \inf\{n : X_n \ge \lambda\}$. (7.6) implies $EX_0 \ge EX_N \ge \lambda P(N < \infty)$.

7.2. Writing T instead of T_1 and using (4.1) we have

$$E(S_{T\wedge n} - (p-q)(T \wedge n))^2 = \sigma^2 E(T \wedge n)$$

Letting $n \to \infty$ and using Fatou's lemma

$$E(1 - (p-q)T)^2 \le \sigma^2 ET < \infty$$

so $ET^2 < \infty$. Expanding out the square in the first equation now we have

$$X_{T\wedge n} = S_{T\wedge n}^2 - 2S_{T\wedge n}(p-q)(T\wedge n) + (p-q)^2(T\wedge n)^2 - \sigma^2(T\wedge n)$$

Now $1 \geq S_{T\wedge n} \geq \min_m S_m$ and Example 7.1 implies $E(\min_m S_m)^2 < \infty$, so using the Cauchy Schwarz inequality for the second term we see that each of the four terms is dominated by an integrable random variable so letting $n \to \infty$ and using dominated convergence

$$0 = 1 - 2(p-q)ET + (p-q)^2 ET^2 - \sigma^2 ET$$

Recalling $ET = 1/(p-q)$ and solving gives

$$ET^2 = \frac{1}{(p-q)^2} + \frac{\sigma^2}{(p-q)^3}$$

so $\text{var}(T) = ET^2 - (ET)^2 = \sigma^2/(p-q)^3$.

7.3. (i) Using (4.1) we have

$$0 = ES_{T\wedge n}^2 - (T\wedge n)$$

As $n \to \infty$, $ES_{T\wedge n}^2 \to a^2$ by bounded convergence, and $E(T\wedge n) \uparrow ET$ by monotone convergence so $ET = a^2$.
(ii) Since $\xi_n = \pm 1$ with equal probability, $\xi_n^2 = \xi_n^4 = 1$, and

$$\begin{aligned}
E(S_n^3\xi_{n+1}|\mathcal{F}_n) &= S_n^3 E(\xi_{n+1}|\mathcal{F}_n) = 0\\
E(S_n\xi_{n+1}^3|\mathcal{F}_n) &= S_n E(\xi_{n+1}^3|\mathcal{F}_n) = 0\\
E(S_n\xi_{n+1}|\mathcal{F}_n) &= S_n E(\xi_{n+1}|\mathcal{F}_n) = 0
\end{aligned}$$

Substituting $S_{n+1} = S_n + \xi_{n+1}$, expanding out the powers and using the last three identities

$$\begin{aligned}
&E\Big((S_n+\xi_{n+1})^4 - 6(n+1)(S_n+\xi_{n+1})^2 + b(n+1)^2 + c(n+1)\Big|\mathcal{F}_n\Big)\\
&= S_n^4 + 6S_n^2 + 1 - 6(n+1)S_n^2 - 6(n+1) + bn^2 + b(2n+1) + cn + c\\
&\qquad\qquad = S_n^4 - 6nS_n^2 + bn^2 + cn + (2b-6)n + (b+c-5) = Y_n
\end{aligned}$$

if $b = 3$ and $c = 2$. Using (4.1) now

$$3E(T\wedge n)^2 = E\{6(T\wedge n)S_{T\wedge n}^2 - S_{T\wedge n}^4 - 2(T\wedge n)\}$$

Letting $n \to \infty$, using the monotone convergence theorem on the left and the dominated convergence theorem on the right.

$$3ET^2 = 6a^2 ET - a^4 - 2ET$$

Recalling $ET = a^2$ gives $ET^2 = (5a^4 - 2a^2)/3$.

7.4. (i) Using (1.3) and the fact that ξ_{n+1} is independent of $\mathcal{F}_n$

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= \exp(\theta S_n - (n+1)\psi(\theta))E(\exp(\theta\xi_{n+1})|\mathcal{F}_n) \\ &= \exp(\theta S_n - n\psi(\theta)) \end{aligned}$$

(ii) As shown in Section 1.9, $\psi'(\theta) = \varphi'(\theta)/\varphi(\theta)$ and if we let

$$dF_\theta = (e^{\theta x}/\varphi(\theta))\, dF$$

then we have

$$\frac{d}{d\theta}\frac{\varphi'(\theta)}{\varphi(\theta)} = \frac{\varphi''(\theta)}{\varphi(\theta)} - \left(\frac{\varphi'(\theta)}{\varphi(\theta)}\right)^2 = \int x^2\, dF_\theta(x) - \left(\int x\, dF_\theta(x)\right)^2 > 0$$

since the last expression is the variance of F_θ, and this distribution is nondegenerate if ξ_i is not constant.

(iii)
$$\begin{aligned} \sqrt{X_n^\theta} &= \exp((\theta/2)S_n - (n/2)\psi(\theta)) \\ &= X_n^{\theta/2}\exp(n\{\psi(\theta/2) - \psi(\theta)/2\}) \end{aligned}$$

Strict convexity and $\psi(0) = 0$ imply $\psi(\theta/2) - \psi(\theta)/2 < 0$. $X_n^{\theta/2}$ is martingale with $X_0^{\theta/2} = 1$ so

$$E\sqrt{X_n^\theta} = \exp(n\{\psi(\theta/2) - \psi(\theta)/2\}) \to 0$$

as $n \to \infty$ and it follows that $X_n^\theta \to 0$ in probability.

7.5. If $\theta \geq 0$ then $\varphi(\theta) \geq (e^\theta + e^{-\theta})/2 \geq 1$ so $\psi(\theta) = \ln\varphi(\theta) \geq 0$ and

$$X_{n\wedge T} = \exp(\theta S_{T\wedge n} - (T\wedge n)\psi(\theta)) \leq e^\theta$$

Using (4.1), letting $n \to \infty$ and using the bounded convergence theorem we have

$$1 = EX_{T\wedge n} \to E\exp(\theta S_T - T\psi(\theta))$$

or since $S_T = 1$, $1 = e^\theta E\varphi(\theta)^{-T}$.
(ii) Setting $\varphi(s) = pe^\theta + qe^{-\theta} = 1/s$ and $x = e^{-\theta}$ we have $qsx^2 - x + ps = 0$. Solving gives

$$Es^T = x = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2qs}$$

$s \to Es^T$ is continuous on $[0,1]$ and $\to 0$ as $s \to 0$ so the $-$ root is always the right choice.

7.6. $X_{T\wedge n}$ is bounded so the optional stopping theorem and Chebyshev's inequality imply $1 = EX_T \geq e^{\theta_o a} P(S_T \leq a)$.

7.7. Let η be Normal$(0,\sigma^2)$.

$$\begin{aligned} Ee^{\theta\xi_1} = Ee^{\theta(c-\mu-\eta)} &= e^{\theta(c-\mu)} \int e^{-\theta x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\, dx \\ &= \exp(\theta(c-\mu) + \theta^2\sigma^2/2) \cdot \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x+\theta\sigma^2)^2/2\sigma^2}\, dx \\ &= \exp(\theta(c-\mu) + \theta^2\sigma^2/2) \end{aligned}$$

since the integral is the total mass of a normal density with mean $-\theta\sigma^2$ and variance σ^2. Taking $\theta_o = 2(\mu - c)/\sigma^2$ we have $\varphi(\theta_o) = 1$. Applying the result in Exercise 7.6 to $S_n - S_0$ with $a = -S_0$, we have the desired result.

7.8. Using Exercise 1.1 in Chapter 4, the fact that the ξ_j^{n+1} are independent of $\mathcal{F}_n$, the definition of φ, and the definition of ρ, we see that on $\{Z_n = k\}$

$$E(\rho^{Z_{n+1}}|\mathcal{F}_n) = E(\rho^{\xi_1^{n+1}+\cdots+\xi_k^{n+1}}|\mathcal{F}_n) = \varphi(\rho)^k = \rho^k = \rho^{Z_n}$$

so ρ^{Z_n} is a martingale. Let $N = \inf\{n : Z_n = 0\}$. (4.1) implies $\rho^x = E_x(\rho^{Z_{N\wedge n}})$. Exercise 5.6 implies that $Z_n \to \infty$ on $N = \infty$ so letting $n \to \infty$ and using the bounded convergence theorem gives the desired result.

5 Markov Chains

5.1. Definitions and Examples

1.1. Exercise 1.1 of Chapter 4 implies that on $Z_n = i > 0$

$$P(Z_{n+1} = j|\mathcal{F}_n) = P\left(\sum_{m=1}^{i} \xi_m^{n+1} \middle| \mathcal{F}_n \right) = p(i,j)$$

since the ξ_m^{n+1} are independent of $\mathcal{F}_n$.

1.2. $p^2(1,2) = p(1,3)p(3,2) = (0.9)(0.4) = 0.36$. To get from 2 to 3 in three steps there are three ways 2213, 2113, 2133, so

$$p^3(2,3) = (.7)(.9)(.1 + .3 + .6) = .63$$

1.3. This is correct for $n = 0$. For the inductive step note

$$\begin{aligned}
P_\mu(X_{n+1} = 0) = P_\mu(X_n = 0)(1-\alpha) + P_\mu(X_n = 1)\beta \\
&= (1-\alpha)\left\{ \frac{\beta}{\beta+\alpha} + (1-\alpha-\beta)^n \left(\mu(0) - \frac{\beta}{\alpha+\beta} \right) \right\} \\
&+ \beta \left\{ \frac{\alpha}{\beta+\alpha} - (1-\alpha-\beta)^n \left(\mu(0) - \frac{\beta}{\alpha+\beta} \right) \right\} \\
&= \frac{\beta}{\beta+\alpha} + (1-\alpha-\beta)^{n+1} \left(\mu(0) - \frac{\beta}{\alpha+\beta} \right)
\end{aligned}$$

1.4. The transition matrix is

	HH	HT	TH	TT
HH	1/2	1/2	0	0
HT	0	0	1/2	1/2
TH	1/2	1/2	0	0
TT	0	0	1/2	1/2

Since X_n and X_{n+2} are independent $p^2(i,j) = 1/4$ for all i and j.

1.5.

	AA,AA	AA,Aa	AA,aa	Aa,Aa	Aa,aa	aa,aa
AA,AA	1	0	0	0	0	0
AA,Aa	1/4	1/2	0	1/4	0	0
AA,aa	0	0	0	1	0	0
Aa,Aa	1/16	1/4	1/8	1/4	1/4	1/16
Aa,aa	0	0	0	1/4	1/2	1/4
aa,aa	0	0	0	0	0	1

1.6. This is a Markov chain since the probability of adding a new value at time $n+1$ depends on the number of values we have seen up to time n. $p(k, k+1) = 1 - k/N$, $p(k,k) = k/N$, $p(i,j) = 0$ otherwise.

1.7. X_n is not a Markov chain since $X_{n+1} = X_n + 1$ with probability 1/2 when $X_n = S_n$ and with probability 0 when $X_n > S_n$.

1.8. Let $i_1, \ldots, i_n \in \{-1, 1\}$ and $N = |\{m \le n : i_m = 1\}|$.

$$P(X_1 = i_1, \ldots, X_n = i_n) = \int \theta^N (1-\theta)^{n-N}\, d\theta$$
$$P(X_1 = i_1, \ldots, X_n = i_n, X_{n+1} = 1) = \int \theta^{N+1} (1-\theta)^{n-N}\, d\theta$$

Now $\int_0^1 x^m (1-x)^k \, dx = m!k!/(m+k+1)!$ so

$$P(X_{n+1} = 1 | X_1 = i_1, \ldots, X_n = i_n) = \frac{(S_n+1)!/(n+2)!}{S_n!/(n+1)!} = \frac{S_n+1}{n+2}$$

(ii) Since the conditional expectation is only a function of S_n, (1.1) implies that S_n is a Markov chain.

5.2. Extensions of the Markov Property

2.1. Using the hint, $1_A \in \mathcal{F}_n$, the Markov property (2.1), then $E_\mu(1_B|X_n) \in \sigma(X_n)$

$$\begin{aligned} P_\mu(A \cap B | X_n) &= E_\mu(E_\mu(1_A 1_B | \mathcal{F}_n)|X_n) \\ &= E_\mu(1_A E_\mu(1_B|\mathcal{F}_n)|X_n) \\ &= E_\mu(1_A E_\mu(1_B|X_n)|X_n) \\ &= E_\mu(1_A|X_n) E_\mu(1_B|X_n) \end{aligned}$$

2.2. Let $A = \{x : P_x(D) \ge \delta\}$. The Markov property and the definition of A imply $P(D|X_n) \ge \delta$ on $\{X_n \in A\}$. so (2.3) implies

$$P(\{X_n \in A \text{ i.o.}\} - \{X_n = a \text{ i.o.}\}) = 0$$

Since $\delta > 0$ is arbitrary the desired result follows.
(ii) Under the assumptions of Exercise 5.5 in Chapter 4, $h(X_n) \to 0$ implies $X_n \to \infty$.

2.3. Clearly, $P_x(X_n = y) = \sum_{m=1}^n P_x(T_y = m, X_n = y)$. When $m = n$, $P_x(T_y = n, X_n = y) = P_x(T_y = n) = P_x(T_y = n)p^0(y,y)$. To handle $m < n$ note that the Markov property implies

$$P_x(X_n = y|\mathcal{F}_n) = P_x(1_{(X_{n-m}=y)} \circ \theta_m|\mathcal{F}_m) = P_{X_m}(X_{n-m} = y)$$

Integrating over $\{T_y = m\} \in \mathcal{F}_m$ where $X_m = y$ and using the definition of conditional expectation we have

$$\begin{aligned} P_x(T_y = m, X_n = y) &= E_x(1_{\{X_n=y\}}; T_y = m) \\ &= P_x(T_y = m)P_y(X_{n-m} = y) = P_x(T_y = m)p^{n-m}(y,y) \end{aligned}$$

2.4. Let $T = \inf\{m \ge k : X_m = x\}$. Imitating the proof in Exercise 2.3 it is easy to show that

$$P_x(X_m = x) = \sum_{\ell=k}^{m} P_x(T = \ell)p^{m-\ell}(x,x)$$

Summing from $m = k$ to $n + k$, using Fubini's theorem to interchange the sum, then using the trivial inequalities $p^j(x,x) \ge 0$ and $P_x(T \le n+k) \le 1$ we have

$$\begin{aligned} \sum_{m=k}^{n+k} P_x(X_m = x) &= \sum_{m=k}^{n+k}\sum_{\ell=k}^{m} P_x(T = \ell)p^{m-\ell}(x,x) \\ &= \sum_{\ell=k}^{n+k}\sum_{m=\ell}^{n+k} P_x(T = \ell)p^{m-\ell}(x,x) \\ &\le \sum_{\ell=k}^{n+k} P_x(T = \ell)\sum_{j=0}^{n} p^j(x,x) \\ &\le \sum_{j=0}^{n} p^j(x,x) = \sum_{m=0}^{n} P_x(X_m = x) \end{aligned}$$

2.5. Since $P_x(\tau_C < \infty) > 0$ there is an $n(x)$ so that $P_x(\tau_C \le n(x)) > 0$. Let

$$N = \max_{x \in S-C} n(x) < \infty \qquad \epsilon = \min_{x \in S-C} P_x(\tau_C \le N) > 0$$

The Markov property implies

$$P_x(\tau_C \circ \theta_{(k-1)N} > N | \mathcal{F}_{(k-1)N}) = P_{X_{(k-1)N}}(\tau_C > N)$$

Integrating over $\{\tau_C > (k-1)N\}$ using the definition of conditional probability and the bound above we have

$$\begin{aligned} P_x(\tau_C > kN) &= E_x\left(1_{(\tau_C \circ \theta_{(k-1)N} > N)}; \tau_C > (k-1)N\right) \\ &= E_x\left(P_x(\tau_C \circ \theta_{(k-1)N} > N | \mathcal{F}_{(k-1)N}); \tau_C > (k-1)N\right) \\ &= E_x\left(P_{X_{(k-1)N}}(\tau_C > N); \tau_C > (k-1)N\right) \\ &\le (1-\epsilon) P_x(\tau_C > (k-1)N) \end{aligned}$$

from which the result follows by induction.

2.6. (i) If $x \notin A \cup B$ then $1_{(\tau_A < \tau_B)} \circ \theta_1 = 1_{(\tau_A < \tau_B)}$. Taking expected value we have

$$\begin{aligned} P_x(\tau_A < \tau_B) &= E_x(1_{(\tau_A < \tau_B)} \circ \theta_1) \\ &= E_x(E_x(1_{(\tau_A < \tau_B)} \circ \theta_1 | \mathcal{F}_1)) = E_x h(X_1) \end{aligned}$$

(ii) To simplify typing we will write T for $\tau_{A \cup B}$. On $\{T > n\} \in \mathcal{F}_n$ we have $X_{(n+1)\wedge T} = X_{n+1}$ so using Exercise 1.1 in Chapter 4, the Markov property and (i) we have

$$\begin{aligned} E(h(X_{n \wedge T}) | \mathcal{F}_n) &= E(h(X_{n+1}) | \mathcal{F}_n) = E(h(X_1) \circ \theta_n | \mathcal{F}_n) \\ &= E_{X_n} h(X_1) = h(X_n) = h(X_{n \wedge T}) \end{aligned}$$

On $\{T \le n\} \in \mathcal{F}_n$ we have $X_{(n+1)\wedge T} = X_{n \wedge T} \in \mathcal{F}_n$ so using Exercise 1.1 in Chapter 4 we have

$$E(h(X_{(n+1)\wedge T}) | \mathcal{F}_n) = E(h(X_{n \wedge T}) | \mathcal{F}_n) = h(X_{n \wedge T})$$

(iii) Exercise 2.5 implies that $T = \tau_{A \cup B} < \infty$ a.s. Since $S - (A \cup B)$ is finite, any solution h is bounded, so the martingale property and the bounded convergence theorem imply

$$h(x) = E_x h(X_{n \wedge T}) \to E_x h(X_T) = P_x(\tau_A < \tau_B)$$

2.7. (i) $0 = E_0 X_n$ and $X_n \ge 0$ imply $P_0(X_n = 0) = 1$. Similarly. $N = E_N X_n$ and $X_n \le N$ imply $P_N(X_n = N) = 1$.

(ii) Exercise 2.5 implies $P_x(\tau_0 \wedge \tau_N < \infty) = 1$. The martingale property and the bounded convergence theorem imply

$$\begin{aligned} x &= E_x(X(\tau_0 \wedge \tau_N \wedge n)) \\ &\to E_x X(\tau_0 \wedge \tau_N) = N P_x(\tau_N < \tau_0) \end{aligned}$$

2.8. (i) corresponds to sampling with replacement from a population with i 1's and $(N-i)$ 0's so the expected number of 1's in the sample is i.
(ii) corresponds to sampling without replacement from a population with $2i$ 1's and $2(N-i)$ 0's so the expected number of 1's in the sample is i.

2.9. The expected number of A's in each offspring $=2$ times the fraction of A's in its parents, so the number of A's is a martingale. Using Exercise 2.7 we see that the absorption probabilities from a starting point with k A's is $k/4$.

2.10. (i) If $x \notin A$, $\tau_A \circ \theta_1 = \tau_A - 1$. Taking expected value gives

$$\begin{aligned} g(x) - 1 = E_x(\tau_A - 1) &= E_x(\tau_A \circ \theta_1) \\ &= E_x E_x(\tau_A \circ \theta_1 | \mathcal{F}_1) = E_x g(X_1) \end{aligned}$$

(ii) On $\{\tau_A > n\} \in \mathcal{F}_n$, $g(X_{(n+1)\wedge\tau_A}) + (n+1) \wedge \tau_A = g(X_{n+1}) + (n+1)$, so using Exercise 1.1 in Chapter 4, and (i) we have

$$\begin{aligned} &E_x(g(X_{(n+1)\wedge\tau_A}) + (n+1) \wedge \tau_A | \mathcal{F}_n) = E_x(g(X_{n+1}) + (n+1) | \mathcal{F}_n) \\ &\quad = E(g(X_{n+1})|\mathcal{F}_n) + (n+1) = g(X_n) - 1 + (n+1) = g(X_n) + n \end{aligned}$$

On $\{\tau_A \le n\} \in \mathcal{F}_n$, $g(X_{(n+1)\wedge\tau_A}) + (n+1) \wedge \tau_A = g(X_{n\wedge\tau_A}) + (n \wedge \tau_A)$, so using Exercise 1.1 in Chapter 4, we have

$$\begin{aligned} E_x(g(X_{(n+1)\wedge\tau_A}) + ((n+1) \wedge \tau_A) | \mathcal{F}_n) &= E_x(g(X_{n\wedge\tau_A}) + (n \wedge \tau_A) | \mathcal{F}_n) \\ &= g(X_{n\wedge\tau_A}) + (n \wedge \tau_A) \end{aligned}$$

(iii) Exercise 2.5 implies $P_y(\tau_A > kN) \le (1-\epsilon)^k$ for all $y \notin A$ so $E_y \tau_A < \infty$. Since $S - A$ is finite any solution is bounded. Using the martingale property, the bounded and monotone convergence theorems

$$g(x) = E_x(g(X_{n\wedge\tau_A}) + (n \wedge \tau_A)) \to E_x \tau_A$$

2.11. In this case the equation $(*)$ becomes

$$\begin{aligned} g(H,H) &= 0 \\ g(H,T) &= 1 + \frac{1}{2} g(T,H) + \frac{1}{2} g(T,T) \\ g(T,H) &= 1 + \frac{1}{2} g(H,T) \\ g(T,T) &= 1 + \frac{1}{2} g(T,H) + \frac{1}{2} g(T,T) \end{aligned}$$

Comparing the second and fourth equations we see $g(H,T) = g(T,T)$. Using this in the second equation and rearranging the third gives

$$g(T,T) = 2 + g(T,H)$$
$$g(H,T) = 2g(T,H) - 2$$

Noticing that the left-hand sides are equal and solving gives $g(T,H) = 4$, $g(H,T) = g(T,T) = 6$, and

$$EN_1 = \frac{1}{4}(4 + 6 + 6) = 4$$

2.12. (ii) We claim that

$$P(I_j = 1 | I_{j+1} = i_{j+1}, \ldots, I_k = i_k) = 1/j$$

To prove this note that if $n = \inf\{m > j : i_m = 1\}$ then the conditioning event tells us that when the chain left n it jumped to at least as far as j. Since the original jump distribution was uniform on $\{1, \ldots, n-1\}$ the conditional distribution is uniform on $\{1, \ldots, j\}$.

5.3. Recurrence and Transience

3.1. $v_k = v_1 \circ \theta_{R_{k-1}}$. Let v be one of the countably many possible values for the v_i. Since $X(R_{k-1}) = y$ a.s., the strong Markov property implies

$$P_y(v_1 \circ \theta_{R_{k-1}} = v | \mathcal{F}_{R_{k-1}}) = P_y(v_1 = v)$$

This implies v_k is independent of $\mathcal{F}_{R_{k-1}}$ and hence of $v_1, \ldots, v_{k-1}$

3.2. (i) follows from Exercise 2.3. To prove (ii) note that using (i), Fubini's theorem, and then changing variables in the inner sum gives

$$\begin{aligned}
u(s) - 1 &= \sum_{n=1}^{\infty} u_n s^n = \sum_{n=1}^{\infty} \sum_{m=1}^{n} f_m u_{n-m} s^n \\
&= \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} f_m s^m u_{n-m} s^{n-m} \\
&= \sum_{m=1}^{\infty} f_m s^m \sum_{k=0} u_k s^k = f(s)u(s)
\end{aligned}$$

3.3. If $h(x) = (1-x)^{-1/2}$. Differentiating we have

$$\begin{aligned}
h'(x) &= \frac{1}{2}(1-x)^{-3/2} \\
h''(x) &= \frac{1}{2} \cdot \frac{3}{2} \cdot (1-x)^{-5/2} \\
h^{(m)}(x) &= \frac{(2m)!}{m!4^m}(1-x)^{-(2m+1)/2}
\end{aligned}$$

Recalling $h(x) = \sum_{m=0}^{\infty} h^{(m)}(0)/m!$ we have

$$u(s) = \sum_{m=0}^{\infty} \binom{2m}{m} p^m q^m s^{2m} = (1 - 4pqs^2)^{-1/2}$$

so using Exercise 3.2 $f(s) = 1 - 1/u(s) = 1 - (1-4pqs^2)^{1/2}$.
(iii) Setting $s = 1$, we have $P_0(T_0 < \infty) = 1 - (1-4pq)^{1/2}$.

3.4. The strong Markov property implies

$$P_x(T_z \circ \theta_{T_y} < \infty | \mathcal{F}_{T_y}) = P_y(T_z < \infty) \quad \text{on } \{T_y < \infty\}$$

Integrating over $\{T_y < \infty\}$ and using the definition of conditional expectation

$$\begin{aligned}
P_x(T_z < \infty) &\geq P_x(T_z \circ \theta_{T_y} < \infty) \\
&= E_x(P_x(T_z \circ \theta_{T_y} < \infty | \mathcal{F}_{T_y}); T_y < \infty) \\
&= E_x(P_y(T_z < \infty); T_y < \infty) \\
&= P_x(T_y < \infty) P_y(T_z < \infty)
\end{aligned}$$

3.5. $\rho_{xy} > 0$ for all x, y so the chain is irreducible. The desired result now follows from (3.5).

3.6. (i) Using (3.7) we have

$$P_{20}(T_{40} < T_0) = \frac{\sum_{m=0}^{19} (20/18)^m}{\sum_{m=0}^{39} (20/18)^m} = \frac{(20/18)^{19} - (20/18)^{-1}}{(20/18)^{39} - (20/18)^{-1}}$$

Multiplying top and bottom by 20/18 and calculating that $(20/18)^{20} = 8.225$, $(8.225)^2 = 67.654$ we have

$$P_{20}(T_{40} < T_0) = \frac{7.225}{66.654} = 0.1084$$

(ii) Using (4.1), rearranging and then using the monotone and dominated convergence theorems we have

$$\begin{aligned}
E_{20}\left(X_{T\wedge n} + \frac{2}{38}(T \wedge n)\right) &= 20 \\
E_{20}(T \wedge n) &= 380 - 19E_{20}(X_{T\wedge n}) \\
E_{20}T = 380 - 19 \cdot 40 \cdot P_{20}(T_{40} < T_0) &= 297.6
\end{aligned}$$

3.7. Let $\tau = \inf\{n > 0 : X_n \in F\}$, $\epsilon = \inf\{\varphi(x) : x \in F\}$, and pick y so that $\varphi(y) < \epsilon$. Our assumptions imply that $Y_n = \varphi(X_{n\wedge\tau})$ is supermartingale. Using (4.1) in Chapter 4 now we see that

$$\varphi(y) \geq E_y \varphi(X_{n\wedge\tau}) \geq \epsilon P_y(\tau < n)$$

Letting $n \to \infty$ we see that $P_y(\tau < \infty) \leq \varphi(y)/\epsilon < 1$.

3.8. Writing $p_x = 1/2 + c_x/x$ we have

$$\begin{aligned}
E_x X_1^\alpha - x^\alpha &= ((x+1)^\alpha - x^\alpha)\, p_x + ((x-1)^\alpha - x^\alpha)\, q_x \\
&= \frac{1}{2}\left((x+1)^\alpha - 2x^\alpha + (x-1)^\alpha\right) \\
&\quad + \frac{c_x}{x}\left(\{(x+1)^\alpha - x^\alpha\} + \{x^\alpha - (x-1)^\alpha\}\right)
\end{aligned}$$

A little calculus shows

$$\begin{aligned}
(x+1)^\alpha - x^\alpha &= \int_0^1 \alpha(x+y)^{\alpha-1}\, dy \sim \alpha x^{\alpha-1} \\
(x+1)^\alpha - 2x^\alpha + (x-1)^\alpha &= \int_0^1 \alpha\{(x+y)^{\alpha-1} - (x-1+y)^{\alpha-1}\}\, dy \\
&\sim \alpha(\alpha-1)x^{\alpha-2}
\end{aligned}$$

This implies that when x is large

$$E_x X_1^\alpha - x^\alpha \approx \alpha x^{\alpha-2}\left\{\frac{\alpha-1}{2} + 2C\right\}$$

If $C < 1/4$ then by taking α close to 0 we can make this < 0. When $C > 1/4$ we take $\alpha < 0$, so we want the quantity inside the brackets to be > 0 which again is possible for α close enough to 0.

3.9. If $f \geq 0$ is superharmonic then $Y_n = f(X_n)$ is a supermartingale so Y_n converges a.s. to a limit Y_∞. If X_n is recurrent then for any x, $X_n = x$ i.o., so $f(x) = Y_\infty$ and f is constant.

Conversely, if the chain is transient then $P_y(T_y < \infty) < 1$ for some x and $f(x) = P_x(T_y < \infty)$ is a nonconstant superharmonic function.

3.10. $E_x X_1 = px + \lambda < x$ if $\lambda < (1-p)x$.

5.4. Stationary Measures

4.1. The symmetric form of the Markov property given in Exercise 2.1 implies that for any initial distribution Y_m is a Markov chain. To compute its transition probability we note

$$\begin{aligned} P_\mu(Y_{m+1} = y | Y_m = x) &= \frac{P_\mu(Y_m = x, Y_{m+1} = y)}{P_\mu(Y_m = x)} \\ &= \frac{P_\mu(X_{n-(m+1)} = y) P_\mu(X_{n-m} = x | X_{n-(m+1)} = y)}{P_\mu(X_{n-m} = x)} \\ &= \frac{\mu(y)p(y,x)}{\mu(x)} \end{aligned}$$

4.2. In order for the chain to visit j before returning to 0, it must jump to j or beyond, which has probability $\sum_{k=j}^\infty f_{k+1}$ and in this case it will visit j exactly once.
(ii) Plugging into the formula we have

$$\begin{aligned} q(i, i+1) &= \frac{\mu(i+1)}{\mu(i)} = P(\xi > i+1 | \xi > i) \\ q(i, 0) &= \frac{\mu(0)p(0,i)}{\mu(i)} = P(\xi = i+1 | \xi > i) \end{aligned}$$

which a little thought reveals is the transition probability for the age of the item in use at time n.

4.3. Since the stationary distribution is unique up to constant multiples

$$\frac{\mu_y(z)}{\mu_y(y)} = \frac{\mu_x(z)}{\mu_x(y)}$$

Since $\mu_y(y) = 1$ rearranging gives the desired equality.

4.4. Since simple random walk is recurrent, (4.4) implies that the stationary measure $\mu(x) \equiv 1$ is unique up to constant multiples. If we do the cycle trick starting from 0, the resulting stationary measure has $\mu_0(0) = 1$ and $\mu_0(k) =$ the expected number of visits to k before returning to 0, so $\mu_0(k) = 1$.

4.5. If we let $a = P_x(T_y < T_x)$ and $b = P_y(T_x < T_y)$ then the number of visits to y before we return to x has $P_x(N_y = 0) = 1 - a$ and $P_x(N_k = j) = a(1-b)^{j-1}b$ for $j \geq 1$, so $EN_k = a/b$. In the case of random walks when $x = 0$ we have $a = b = 1/2|y|$.

4.6. (i) Iterating shows that

$$q^n(x, y) = \frac{\mu(y)p^n(y, x)}{\mu(x)}$$

Given x and y there is an n so that $p^n(y, x) > 0$ and hence $q^n(x, y) > 0$. Summing over n and using (3.3) we see that all states are recurrent under q.
(ii) Dividing by $\mu(y)$ and using the defintion of q we have

$$h(y) = \frac{\nu(y)}{\mu(y)} \geq \sum_x q(y, x)\frac{\nu(x)}{\mu(x)}$$

so h is nonnegative superharmonic, and Exercise 3.9 implies that it must be constant.

4.7. By (4.7) the renewal chain is positive recurrent if and only if $E_0 T_0 < \infty$ but $X_1 = k$ implies $T_0 = k + 1$ so $E_0 T_0 = \sum_k k f_k$.

4.8. Let $n = \inf\{m : p^m(x, y) > 0\}$ and pick $x_1, \ldots, x_{n-1} \neq x$ so that

$$p(x, x_1)p(x_1, x_2) \cdots p(x_{n-1}, y) > 0$$

The Markov property implies

$$\begin{aligned} E_x T_x &\geq E_x(T_x; X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = y) \\ &\geq p(x, x_1)p(x_1, x_2) \cdots p(x_{n-1}, y) E_y T_x \end{aligned}$$

so $E_y T_x < \infty$.

4.9. If p is recurrent then any stationary distribution is a constant multiple of μ and hence has infinite total mass, so there cannot be a stationary distribution.

4.10. This is a random walk on a graph, so $\mu(i) =$ the degree of i defines a stationary measure. With a little patience we can compute the degrees for the upper 4×4 square in the chessboard to be

2	3	4	4
3	4	6	6
4	6	8	8
4	6	8	8

Adding up these numbers we get 84 so the total mass of μ is 336. Thus if π is the stationary distribution and c is a corner then $\pi(c) = 2/336$ and (4.6) implies $E_c T_c = 168$.

4.11. Using (4.1) from Chapter 4 it follows that

$$x \geq E_x\left(X_{n\wedge\tau} + \epsilon(n\wedge\tau)\right) \geq \epsilon E_x(\tau\wedge n)$$

Letting $n \to \infty$ and using the monotone convergence theorem the desired result follows.

4.12. The Markov property and the result of the previous exercise imply that

$$E_0 T_0 - 1 = \sum_x p(0,x) E_x \tau \leq \sum_x p(0,x)\frac{x}{\epsilon} = \frac{1}{\epsilon} E_x X_1 < \infty$$

5.5. Asymptotic Behavior

5.1. Making a table of the number of black and white balls in the two urns

	L	R
black	n	$b-n$
white	$m-n$	$m-(b-n)$

we can read off the transition probability. If $0 \leq n \leq b$ then

$$\begin{aligned} p(n,n+1) &= \frac{m-n}{m}\cdot\frac{b-n}{m} \\ p(n,n-1) &= \frac{n}{m}\cdot\frac{m+n-b}{m} \\ p(n,n) &= 1 - p(n,n-1) - p(n,n+1) \end{aligned}$$

5.2. $\{1,7\}$, $\{2,3\}$, $\{4,5,6\}$.

5.3. Let Z be a bounded invariant random variable and $h(x) = E_x Z$. The invariance of Z and the Markov property imply

$$E_\mu(Z|\mathcal{F}_n) = E_\mu(Z\circ\theta_n|\mathcal{F}_n) = h(X_n)$$

so $h(X_n)$ is martingale and h is a bounded harmonic function.
Conversely if h is bounded and harmonic then $h(X_n)$ is a bounded martingale. (2.10) in Chapter 4 implies $Z = \lim_{n\to\infty} h(X_n)$ exists. Z is shift invariant since $Z\circ\theta = \lim_{n\to\infty} h(X_{n+1})$. (5.5) in Chapter 4 implies $h(X_n) = E(Z|\mathcal{F}_n)$.

5.4. (i) ξ_m corresponds to the number of customers that have arrived minus the one that was served. It is easy to see that the $M/G/1$ queue satisfies $X_{n+1} = (X_n + \xi_{m+1})^+$ and the new defintion does as well.
(ii) When $X_{m-1} = 0$ and $\xi_m = -1$ the random walk reaches a new negative minimum so

$$|\{m \le n : X_{m-1} = 0, \xi_m = -1\}| = \left(\min_{m\le n} S_m\right)^-$$

The desired result follows once we show that

$$n^{-1} \min_{m\le n} S_m \to E\xi_m = \mu - 1$$

To do this note that the strong law of large numbers implies that $S_n/n \to \mu - 1$. This implies that

$$\limsup_{n\to\infty} n^{-1} \min_{m\le n} S_m \le \mu - 1$$

To argue the other inequality, note that if $\epsilon > 0$ and $n \ge N$ then $S_n \ge (\mu - 1)(1+\epsilon)n$. When n is large the minimum no longer comes from the $n \le N$ and we have

$$n^{-1} \min_{m\le n} S_m \ge (\mu - 1)(1 + \epsilon)$$

5.5. (i) The fact that the V_k^f are i.i.d. follows from Exercise 3.1, while (4.3) implies

$$E|V_k^f| \le EV_k^{|f|} = \int |f(y)|\, \mu_x(dy)$$

(ii) The strong law of large numbers implies

$$\frac{1}{m}\sum_{k=1}^m V_k^f \to EV_k^f \quad \text{a.s.}$$

Taking $m = K_n$ and noting that the renewal theorem implies $K_n/n \to 1/E_xT_x$ a.s. the desired result follows.
(iii) From Exercise 6.14 in Chapter 1 we see that if $EV_n^{|f|} < \infty$ then

$$\frac{1}{n} \max_{1\le m\le n} V_m^{|f|} \to 0 \quad \text{a.s.}$$

It is easy to see that $K_n \le n$ and

$$\left|\sum_{m=1}^n f(X_m) - \sum_{m=1}^{K_n} V_m^f\right| \le \max_{1\le m\le n} V_m^{|f|}$$

and the desired result follows.

5.6. (i) From (ii) in Exercise 5.5, we know that $K_n/n \to 1/E_xT_x$. Since the V_k^f are i.i.d. with $EV_k^f = 0$ and $E(V_k^f)^2 < \infty$ the desired result follows from Exercise 4.7 in Chapter 2.
(ii) $E(V_k^{|f|})^2 < \infty$ implies that for any $\epsilon > 0$

$$\sum_k P\left\{(V_k^{|f|})^2 > \epsilon^2 k\right\} < \infty$$

so the Borel-Cantelli lemma implies $P(V_k^{|f|} > \epsilon\sqrt{k} \text{ i.o.}) = 0$. From this it follows easily that

$$n^{-1/2} \sup_{k \le n} V_k^{|f|} \to 0 \quad \text{a.s.}$$

Since $K_n \le n$ the desired result follows easily.

5.7. Let $S_k = \sum_{m=1}^{T_y^k} 1_{(X_m = z)}$. Applying (i) in Exercise 5.5 to $f = 1_{\{z\}}$ then using the strong law and (4.4) we have

$$S_k/k \to \mu_y(z) = \frac{m(z)}{m(y)} \quad \text{a.s.}$$

If $T_y^k \le n < T_y^{k+1}$ then

$$\frac{S_k}{k} \le \frac{N_n(z)}{k} \le \frac{S_{k+1}}{k+1} \cdot \frac{k+1}{k}$$

Letting $k \to \infty$ now we get the desired result.

5.8. (i) Breaking things down according to the value of J

$$\begin{aligned} P_x(X_m = z) &= \bar{p}_m(x,z) + \sum_{j=1}^{m-1} P_x(X_m = z, J = j) \\ &= \bar{p}_m(x,z) + \sum_{j=1}^{m-1} P_x(X_j = y, X_{j+1} \ne y, \ldots, X_{m-1} \ne y, X_m = z) \end{aligned}$$

If we let $A_k = \{X_1 \ne y, \ldots, X_{k-1} \ne y, X_k = z\}$ then using the definition of A_k, the definition of conditional expectation, the Markov property, and the definitions of p^j and $\bar{p}^{m-j}$

$$\begin{aligned} P_x(X_j = y, &X_{j+1} \ne y, \ldots, X_{m-1} \ne y, X_m = z) = E_x(1_{A_{m-j}} \circ \theta_j; X_j = y) \\ &= E_x(E_x(1_{A_{m-j}} \circ \theta_j | \mathcal{F}_j); X_j = y) \\ &= E_x(P_y(A_{m-j}); X_j = y) = p^j(x,y)\bar{p}_{m-j}(y,z) \end{aligned}$$

Combining this with the first equality and summing over m

$$\sum_{m=1}^{n} p^m(x,z) = \sum_{m=1}^{n} \bar{p}_m(x,z) + \sum_{m=1}^{n}\sum_{j=1}^{m-1} p^j(x,y)\bar{p}_{m-j}(y,z)$$

Interchanging the order of the last two sums and changing variables $k = m - j$ gives the desired formula.
(ii) $P_y(T_x < T_y)\sum_{m=1}^{\infty} \bar{p}_m(x,z) \le \mu_y(z) < \infty$ and recurrence implies that $\sum_{m=1}^{\infty} p^m(x,y) = \infty$ so we have

$$\sum_{m=1}^{n} \bar{p}_m(x,z) \Big/ \sum_{m=1}^{n} p^m(x,y) \to 0$$

To handle the second term let $a_j = p^j(x,y)$, $b_m = \sum_{k=1}^{m} \bar{p}_k(y,z)$ and note that $b_m \to \mu_y(z)$ and $a_m \le 1$ with $\sum_{m=1}^{\infty} a_m = \infty$ so

$$\sum_{j=1}^{n-1} a_j b_{n-j} \Big/ \sum_{m=1}^{n} a_m \to \mu_y(z)$$

To prove the last result let $\epsilon > 0$, pick N so that $|b_m - \mu_y(z)| < \epsilon$ for $m \ge N$ and then divide the sum into $1 \le j \le n - N$ and $n - N < j < n$.

5.9. By aperiodicity we can pick an N_x so that for all $n \ge N_x$ $p^n(x,x) > 0$. By irreducibility there is an $n(x,y)$ so that $p^{n(x,y)}(x,y) > 0$. Let

$$N = \max\{N_x, n(x,y) : x, y \in S\} < \infty$$

by the finiteness of S.

$$p^{2N}(x,y) \ge p^{n(x,y)}(x,y)p^{2N-n(x,y)}(y,y) > 0$$

since $2N - n(x,y) \ge N$.

5.10. If $\epsilon = \inf p(x,y) > 0$ and there are N states then

$$P(X_{n+1} = Y_{n+1} | X_n = x, Y_n = y) = \sum_z p(x,z)p(y,z) \ge \epsilon^2 N$$

so $P(T > n+1 | T > n) \le (1 - \epsilon^2 N)$ and we have $P(T > n) \le (1 - \epsilon^2 N)^n$.

5.11. To couple X_{n+m} and Y_{n+m} we first run the two chains to time n. If $X_n = Y_n$ an event with probability $\ge 1 - \alpha_n$ then we can certainly arrange things so that $X_{n+m} = Y_{n+m}$. On the other hand it follows from the definition of α_m that

$$P(X_{n+m} \ne Y_{n+m} | X_n = k, Y_n = \ell) \le \alpha_m$$

5.6. General State Space

6.1. As in (3.2)

$$\sum_{n=1}^{\infty} \bar{p}^n(\alpha, \alpha) = \sum_{k=1}^{\infty} P_\alpha(R_k < \infty)$$
$$= \sum_{k=1}^{\infty} P_\alpha(R < \infty)^k = \frac{P_\alpha(R < \infty)}{1 - P_\alpha(R < \infty)}$$

6.2. By Example 6.1, without loss of generality $A = \{a\}$ and $B = \{b\}$. Let $R = \{x : \rho_{bx} > 0\}$. If α is recurrent then b is recurrent, so if $x \in R$ then (3.4) implies x is recurrent. (i) implies $\rho_{xb} > 0$. If y is another point in R then Exercise 3.4 implies $\rho_{xy} \geq \rho_{xb}\rho_{by} > 0$ so R is irreducible. Let $T = S - R$. If $z \in T$ then $\rho_{bz} = 0$ but $\rho_{zb} > 0$ so z is transient by remarks after Example 3.1.

6.3. Suppose that the chain is recurrent when (A, B) is used. Since $P_x(\tau_{A'} < \infty) > 0$ we have $P_\alpha(\tau_{A'} < \infty) > 0$ and (6.4) implies $P_\alpha(\bar{X}_n \in A' \text{ i.o.}) = 1$. (ii) of the definition for (A', B') and (2.3) now give the desired result.

6.4. If $E\xi_n \leq 0$ then $P(S_n \leq 0 \text{ for some } n) = 1$ and hence

$$P(W_n = 0 \text{ for some } n) = 1$$

If $E\xi_n > 0$ then $S_n \to \infty$ a.s. and $P(S_n > 0 \text{ for all } n) > 0$, so we have $P(W_n > 0 \text{ for all } n) > 0$.

6.5. $V_1 = \theta V_0 + \xi_1$. Let N be chosen large enough so that $E|\xi_1| \leq (1 - \theta)N$. If $|x| \geq N$ then

$$E_x|V_1| \leq \theta|x| + E|\xi_1| \leq |x|$$

Using (3.9) now with $\varphi(x) = |x|$ we see that $P_x(|V_n| \leq N) = 1$ for any x. From this and the Markov property it follows that $P_x(|V_n| \leq N \text{ i.o.}) = 1$. Since

$$\inf_{y:|y| \leq N} P_y(\bar{V}_2 = \alpha) > 0$$

it follows from (2.3) that V_n is recurrent.

6.6.

$$P_x(V_1 < \gamma x) \leq P_x(\xi_1 < (\gamma - \theta)x) \leq \frac{E|\xi|}{(\theta - \gamma)x}$$

If x is large $\sum_{n=1}^{\infty} \frac{E|\xi|}{(\theta-\gamma)x\gamma^{n-1}} < 1$ so $P_x(V_n \geq \gamma^n x \text{ for all } n) > 0$.

6.7. Let $\mathcal{F}_n = \sigma(Y_0, \ldots, Y_n)$. In this case $|Y_n| = \beta\sqrt{|Y_{n-1}|}|\chi_n|$ where χ_n is a standard normal independent of $\mathcal{F}_{n-1}$ The sign of Y_n is independent of $\mathcal{F}_{n-1}$ and $|Y_n|$ so it is enough to look at the behavior of $|Y_n|$. Taking logs and iterating we have

$$\begin{aligned}\log|Y_n| &= \log(\beta|\chi_n|) + 2^{-1}\log|Y_{n-1}| \\ &= \log(\beta|\chi_n|) + 2^{-1}\log(\beta|\chi_{n-1}|) + 2^{-2}\log|Y_{n-2}| \\ &= \sum_{m=0}^{n-1} 2^{-m}\log(\beta|\chi_{n-m}|) + 2^{-n}\log|Y_0|\end{aligned}$$

Since $E\log(\beta|\chi|) < \infty$ it is easy to see from this representation that $\log|Y_n| \Rightarrow$ a limit independent of Y_0. Using $P(|Y_n| \le K \text{ i.o.}) \ge \limsup P(|Y_n| \le K)$ and (2.3) now it follows easily that Y_n is recurrent for any β.

6.8. Let $T_0 = 0$ and $T_n = \inf\{m \ge T_{n-1} + k : X_m \in G_{k,\delta}\}$. The definition of $G_{k,\delta}$ implies

$$P(T_n < T_\alpha | T_{n-1} < T_\alpha) \le (1-\delta)$$

so if we let $N = \sup\{n : T_n < T_\alpha\}$ then $EN \le 1/\delta$. Since we can only have $X_m \in G_{k,\delta}$ when $T_n \le m < T_n + k$ for some $n \ge 0$ it follows that

$$\bar{\mu}(G_{k,\delta}) \le k\left(1 + \frac{1}{\delta}\right) \le 2k/\delta$$

Assumption (i) implies $S \subset \cup_{k,m} G_{k,1/m}$ so $\bar{\mu}$ is σ-finite.

6.9. If $\lambda(C) = 0$ then $P_\alpha(\bar{X}_n \in C) = 0$ for all n so $P_\alpha(\bar{X}_n \in C, R > n) = 0$ for all n and $\bar{\mu}(C) = 0$. To prove the converse note that if $\bar{\mu}(C) = 0$ then $P_\alpha(\bar{X}_n \in C, R > n) = 0$ for all n. Now if $P_\alpha(\bar{X}_m \in C) > 0$ and we let M be the smallest m for which this holds we have

$$P_\alpha(\bar{X}_M \in C) = P_\alpha(X_M \in C, R > M) = 0$$

a contradiction so $P_\alpha(\bar{X}_m \in C) = 0$ for all m and $\lambda(C) = 0$.

6.10. The almost sure convergence of the sum follows from Exercise 8.8 in Chapter 1. The sum Z is a stationary distribution since obviously $\xi + \theta Z =_d Z$.

6.11. To prepare for the proof we note that by considering the time of the first visit to α

$$\sum_{n=1}^{\infty} \bar{p}^n(x,\alpha) = P_x(T_\alpha < \infty)\sum_{m=0}^{\infty} \bar{p}^m(\alpha,\alpha) \le \sum_{m=0}^{\infty} \bar{p}^m(\alpha,\alpha)$$

Let $\bar{\pi} = \pi\bar{p}$. By (6.6) this is a stationary probability measure for $\bar{p}$. Irreducibility and the fact that $\bar{\pi} = \bar{\pi}\bar{p}^n$ imply that $\bar{\pi}(\alpha) > 0$ so using our preliminary

$$\infty = \sum_{n=1}^{\infty} \bar{\pi}(\alpha) = \int \bar{\pi}(dx) \sum_{n=1}^{\infty} \bar{p}^n(x,\alpha) \leq \sum_{m=0}^{\infty} \bar{p}^m(\alpha,\alpha)$$

and the recurrence follows from Exercise 6.1.

6.12. Induction implies

$$V_n = \xi_n + \theta\xi_{n-1} + \cdots + \theta^{n-1}\xi_1 + \theta^n V_0$$

$Y_n = \theta^n Y_0 \to 0$ in probability and

$$X_n \stackrel{d}{=} \xi_0 + \theta\xi_1 + \cdots + \theta^{n-1}\xi_{n-1} \to \sum_{n=0}^{\infty} \theta^n \xi_n$$

So the converging together lemma, 2.10 in Chapter 2 implies

$$V_n \Rightarrow \sum_{n=0}^{\infty} \theta^n \xi_n$$

6.13. (i) See the solution of 5.4.
(ii) $S_n - m_n = \max_{0 \leq k \leq n} S_n - S_k$
(iii) $\max(S_0, S_1, \ldots, S_n) =_d \max(S'_0, S'_1, \ldots, S'_n)$ As $n \to \infty$,

$$\max(S_0, S_1, \ldots, S_n) \to \max(S_0, S_1, S_2 \ldots) \quad \text{a.s.}$$

6.14. Let F be the distribution of Y.

$$\begin{aligned} P(X - Y > x) &= \int_0^{\infty} P(X > x + y)\, dF(y) \\ &= e^{-\lambda x} \int_0^{\infty} P(X > y)\, dF(y) = a e^{-\lambda x} \end{aligned}$$

6 Ergodic Theorems

6.1. Definitions and Examples

1.1. If $A \in \mathcal{I}$ then $\varphi^{-1}(A^c) = (\varphi^{-1}A)^c = A^c$ so $A^c \in \mathcal{I}$. If $A_n \in \mathcal{I}$ are disjoint then

$$\varphi^{-1}(\cup_n A_n) = \cup_n \varphi^{-1}(A_n) = \cup_n A_n$$

so $\cup_n A_n \in \mathcal{I}$. To prove the second claim note that the set of invariant random variables contains the indicator functions 1_A with $A \in \mathcal{I}$ and is closed under pointwise limits, so all $X \in \mathcal{I}$ are invariant. To prove the other direction note that if X is invariant and $B \in \mathcal{R}$ then

$$\{\omega : X(\omega) \in B\} = \{\omega : X\varphi(\omega) \in B\} = \varphi^{-1}(\{\omega : X(\omega) \in B\}$$

so $\{\omega : X(\omega) \in B\} \in \mathcal{I}$.

1.2. (i) $\varphi^{-1}(B) = \cup_{n=1}^{\infty}\varphi^{-n}(A) \subset \cup_{n=0}^{\infty}\varphi^{-n}(A) = B$.
(ii) $\varphi^{-1}(C) = \cap_{n=1}^{\infty}\varphi^{-n}(B) = C$ since $\varphi^{-1}(B) \subset B$.
(iii) We claim that if A is almost invariant then $A = B = C$ a.s.
To see that $P(A\Delta B) = 0$ we begin by noting that φ is measure preserving so

$$\begin{aligned} P(\varphi^{-n}(A)\Delta\varphi^{-(n+1)}(A)) &= P(\varphi^{-1}[\varphi^{-(n-1)}(A)\Delta\varphi^{-n}(A)]) \\ &= P(\varphi^{-(n-1)}(A)\Delta\varphi^{-n}(A)) \end{aligned}$$

Since $P(A\Delta\varphi^{-1}(A)) = 0$ it follows by induction that

$$P(\varphi^{-n}(A)\Delta\varphi^{-(n+1)}(A)) = 0$$

for all $n \geq 0$. Using the triangle inequality $P(A\Delta C) \leq P(A\Delta B) + P(B\Delta C)$ it follows that $P(A\Delta\varphi^{-n}(A)) = 0$. Since this holds for all $n \geq 1$ and is trivial for $n = 0$ we have

$$P(A\Delta B) \leq \sum_{n=0}^{\infty} P(A\Delta\varphi^{-n}(A)) = 0$$

To see that $P(B\Delta C) = 0$ note that $B - \varphi^{-1}(B) \subset A - \varphi^{-1}(A)$ has measure 0, and φ is measure preserving so induction implies $P(\varphi^{-n}(B) - \varphi^{-(n+1)}(B)) = 0$ and we have

$$P\left(B - \cap_{n=1}^{\infty} \varphi^{-n}(B)\right) = 0$$

This shows $P(B - C) = 0$. Since $B \supset C$ the desired conclusion follows.

Conversely, if C is strictly invariant and $P(A\Delta C) = 0$ then

$$P(\varphi^{-1}A\Delta C) = P(\varphi^{-1}(A\Delta C)) = P(A\Delta C) = 0$$

so $P(\varphi^{-1}A\Delta A) \le P(\varphi^{-1}A\Delta C) + P(C\Delta A) = 0$.

1.3. Let $\Omega = \{0, 1\}$, $\mathcal{F}$ = all subsets, P assign mass 1/2 to each point, $T(\omega) = 1 - \omega$ preserves P and clearly there are no invariant sets other than $\emptyset$ and Ω. However T^2 is the identity and is not ergodic.

1.4. (i) Since all the x_m are distinct, for some $m < n \le N$ we must have $|x_m - x_n| \le 1/N$. Define $k_j \in \mathbf{Z}$ so that $j\theta = k_j + x_j$. By considering two cases $x_m < x_n$ and $x_m > x_n$ we see that either $x_{n-m} = |x_n - x_m|$ or $x_{n-m} = 1 - |x_n - x_m|$. In these two cases we have, for $k < N$,

$$x_{k(n-m)} = k|x_n - x_m| \quad \text{and} \quad x_{k(n-m)} = 1 - k|x_n - x_m|$$

respectively. This shows that the orbit comes within $1/N$ of any point. Since N is arbitrary, the desired result follows.
(ii) Let $\delta > 0$ and $\epsilon = \delta P(A)$. Applying Exercise 3.1 to the algebra $\mathcal{A}$ of finite disjoint unions of intervals $[u, v)$, it follows that there is $B \in \mathcal{A}$ so that $P(A\Delta B) < \epsilon$ and hence $P(B) \le P(A) + \epsilon$. If $B = +_{i=1}^{m}[u_i, v_i)$ and $A \cap [u_i, v_i) \le (1 - \delta)|v_i - u_i|$ for all i then

$$P(A) \le (1 - \delta)P(B) \le (1 - \delta)(P(A) + \epsilon) \le (1 - \delta^2)P(A)$$

a contradiction, so we must have $A \cap [u_i, v_i) \ge (1 - \delta)|v_i - u_i|$ for some i.
(iii) Let A be invariant and $\delta > 0$. It follows from (ii) that there is an interval $[a, b)$ so that $|A \cap [a, b)| \ge (1-\delta)(b-a)$. If $1/(n+1) < b-a < 1/n$ then there are $y_1, \ldots, y_n$ so that $B_k = ([a, b] + y_k)$ mod 1 are disjoint. Since the x_n are dense, we can find n_k so that $B_k = ([a, b] + x_{n_k})$ mod 1 are disjoint. The invariance of A implies that $(A + x_n)$ mod $1 \subset A$. Since $|A \cap [a, b]| > (1 - \delta)(b - a)$, it follows that

$$|A| \ge n(b - a)(1 - \delta) \ge \frac{n}{n + 1}(1 - \delta)$$

Since n and δ are arbitrary the desired result follows.

1.5. If $f(x) = \sum_k c_k e^{2\pi kx}$ then

$$f(\varphi(x)) = \sum_k c_k e^{2\pi i 2kx}$$

The uniqueness of the Fourier coefficients implies $c_k = c_{2k}$. Iterating we see $c_k = c_{2^j k}$, so if $c_k \neq 0$ for some $k \neq 0$ then we cannot have $\sum_k c_k^2 < \infty$

1.6. From the definition it is clear that

$$\begin{aligned}\mu \circ \varphi^{-1}[a,b] &= \sum_{n=1}^{\infty} \mu\left(\left[\frac{1}{n+b}, \frac{1}{n+a}\right]\right) \\ &= \sum_{n=1}^{\infty} \ln\left(\frac{n+a+1}{n+a}\right) - \ln\left(\frac{n+b+1}{n+b}\right)\end{aligned}$$

since $\int_u^v dx/(1+x) = \ln(1+v) - \ln(1+u)$. If we replace ∞ by N the sum is

$$\ln\left(\frac{N+a+1}{N+b+1}\right) + \ln(1+b) - \ln(1+a)$$

As $N \to \infty$ the right-hand side converges to $\mu([a,b])$.

1.7. To check stationarity, we let $j > n$ and note that for any i, $Z_i, Z_{i+1}, \dots, Z_{i+j}$ consists of a partial block with a length that is uniformly distributed on $1, \dots n$, then a number of full blocks of length n and then a partial block n.

To check ergodicity we note that the tail σ-field of the Z_m is contained in that of the block process, which is trivial since it is i.i.d.

6.2. Birkhoff's Ergodic Theorem

2.1. Let X_M' and X_M'' be defined as in the proof of (2.1). The bounded convergence theorem implies

$$E\left|\frac{1}{n}\sum_{m=0}^{n-1} X_M'(\varphi^m\omega) - E(X_M'|\mathcal{I})\right|^p \to 0$$

Writing $\|Z\|_p = (E|Z|^p)^{1/p}$ and using the triangle inequality

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{m=0}^{n-1} X_M''(\varphi^m\omega) - E(X_M''|\mathcal{I})\right\|_p \\ &\qquad \le \left\|\frac{1}{n}\sum_{m=0}^{n-1} X_M''(\varphi^m\omega)\right\|_p + \|E(X_M''|\mathcal{I})\|_p \\ &\qquad \le \frac{1}{n}\sum_{m=0}^{n-1} \|X_M''(\varphi^m\omega)\|_p + \|E(X_M''|\mathcal{I})\|_p \le 2\|X_M''\|_p\end{aligned}$$

since $E|X''_M(\varphi^m\omega)|^p = E|X''_M|^p$ and $E|E(X''_M|\mathcal{I})|^p \le E|X''_M|^p$ by (1.1e) in Chapter 4.

2.2. (i) Let $h_M(\omega) = \sup_{m\ge M} |g_m(\omega) - g(\omega)|$.

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} g_m(\varphi^m\omega) \le \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1}(g+h_N)(\varphi^m\omega)$$
$$= E(g+h_M|\mathcal{I})$$

since $g_m \le g + h_M$ for all $m \ge M$. $h_M \downarrow 0$ as $M \uparrow \infty$ and h_0 is integrable, so (1.1c) in Chapter 4 implies $E(g+h_M|\mathcal{I}) \downarrow E(g|\mathcal{I})$.

(ii) The triangle inequality and the convergence of $g_m \to g$ in L^1 imply

$$E\left|\frac{1}{n}\sum_{m=0}^{n-1} g_m(\varphi^m\omega) - \frac{1}{n}\sum_{m=0}^{n-1} g(\varphi^m\omega)\right| \le \frac{1}{n}\sum_{m=0}^{n-1} E|g_m - g| \to 0$$

The ergodic theorem implies

$$E\left|\frac{1}{n}\sum_{m=0}^{n-1} g(\varphi^m\omega) - E(g|\mathcal{I})\right| \to 0$$

Combining the last two results and using the triangle inequality gives the desired result.

2.3. Let X'_M and X''_M be defined as in the proof of (2.1). The result for bounded random variables implies

$$\frac{1}{n}\sum_{m=0}^{n-1} X'_M(\varphi^m\omega) \to E(X'_M|\mathcal{I})$$

Using (2.3) now on X''_M we get

$$P\left(\sup_n \left|\frac{1}{n}\sum_{m=0}^{n-1} X''_M(\varphi^m\omega)\right| > \alpha\right) \le \alpha^{-1}E|X''_M|$$

As $M \uparrow \infty$, $E|X''_M| \downarrow 0$. A trivial special case of (5.9) in Chapter 4 implies $E(X'_M|\mathcal{I}) \to E(X|\mathcal{I})$ so

$$P\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m\omega) > E(X|\mathcal{I}) + 2\alpha\right) = 0$$

Since the last result holds for any $\alpha > 0$ the desired result follows.

6.3. Recurrence

3.1. Counting each point visited at the last time it is visited in $\{1, \dots, n\}$

$$ER_n = \sum_{m=1}^{n} P(S_{m+1} - S_m \neq 0, \dots, S_n - S_m \neq 0) = \sum_{m=1}^{n} g_{m-1}$$

3.2. When $P(X_i > 1) = 0$

$$\left\{1, \dots, \max_{m \le n} S_m\right\} \subset R_n \subset \left\{\min_{m \le n} S_m, \dots, \max_{m \le n} S_m\right\}$$

If $EX_i > 0$ then $S_n/n \to EX_i > 0$ so $S_n \to \infty$ and $\min_{m \le n} S_m > -\infty$ a.s. To evaluate the limit of $\max_{m \le n} S_m/n$ we observe that for any K

$$\begin{aligned} \lim_{n\to\infty} \frac{S_n}{n} &\le \liminf_{n\to\infty} \left(\max_{1 \le k \le n} |S_k|/n\right) \le \limsup_{n\to\infty} \left(\max_{1 \le k \le n} |S_k|/n\right) \\ &= \limsup_{n\to\infty} \left(\max_{K \le k \le n} |S_k|/n\right) \le \left(\max_{k \ge K} |S_k|/k\right) \end{aligned}$$

3.3. $\varphi(\theta) = E\exp(\theta X_i)$ is convex, $\varphi(\theta) \to \infty$ as $\theta \to -\infty$ and the left derivative at 0 has $\varphi'(0) = EX_i > 0$ so there is a unique $\theta < 0$ so that $\varphi(\theta) = 1$. Exercise 7.4 in Chapter 4 implies that $\exp(\theta S_n)$ is a martingale. (4.1) in Chapter 4 implies $1 = E\exp(\theta S_{N \wedge n})$. Since $\exp(\theta S_{N \wedge n}) \le e^{-\theta}$ and $S_n \to \infty$ as $n \to \infty$ the bounded convergence theorem implies $1 = e^{-\theta} P(N < \infty)$.

3.4. It suffices to show

$$E\left(\sum_{1 \le m \le T_1} 1_{(X_m \in B)}; X_0 \in A\right) = P(X_0 \in B)$$

To do this we observe that the left hand side is

$$\begin{aligned} &\sum_{m=1}^{\infty} P(X_0 \in A, X_1 \notin A, \dots, X_{m-1} \notin A, X_m \in B) \\ &= \sum_{m=1}^{\infty} P(X_{-m} \in A, X_{-m+1} \notin A, \dots, X_{-1} \notin A, X_0 \in B) = P(X_0 \in B) \end{aligned}$$

3.5. First note that (3.3) implies $\bar{E}T_1 = 1/P(X_0 = 1)$, so the right hand side is $P(X_0 = 1, T_1 \geq n)$. To compute the left now we break things down according to the position of the first 1 to the left of 0 and use translation invariance to conclude $P(T_1 = n)$ is

$$\begin{aligned}
&= \sum_{m=0}^{\infty} P(X_{-m} = 1, X_j = 0 \text{ for } j \in (-m, n), X_n = 1) \\
&= \sum_{m=0}^{\infty} P(X_0 = 1, X_j = 0 \text{ for } j \in (0, m+n), X_{m+n} = 1) \\
&= P(X_0 = 1, T_1 \geq n)
\end{aligned}$$

6.6. A Subadditive Ergodic Theorem

6.1. (1.3) implies that the stationary sequences in (ii) are ergodic. Exercise 3.1 implies $EX_{0,n} = \sum_{m=1}^{n} P(S_1 \neq 0, \ldots, S_n \neq 0)$. Since $P(S_1 \neq 0, \ldots, S_n \neq 0)$ is decreasing it follows easily that $EX_{0,n}/n \to P($ no return to 0 $)$.

6.2. (a) $EL_1 = P(X_1 = Y_1) = 1/2$. To compute EL_2 let $N_2 = |\{i \leq 2 : X_i = Y_i\}$ and note that $L_2 - N_2 = 0$ unless (X_1, X_2, Y_1, Y_2) is $(1,0,0,1)$ or $(0,1,1,0)$. In these two cases which have probability 1/16 each $L_2 - N_2 = 1$ so $EL_2 = EN_2 + 1/8 = 9/8$ so $EL_2/2 = 9/16$

(b) The expected number of sequences of length K is $\binom{n}{K}^2 2^{-K}$. Taking $K = an$ using Stirling's formula $m! \sim m^m e^{-m}\sqrt{2\pi m}$ without the term under the square root we have that the above

$$\approx \frac{n^{2n} 2^{-an}}{(an)^{2an}((1-a)n)^{2(1-a)n}} = (a^{2a}(1-a)^{2(1-a)}2^a)^{-n}$$

From the last computation it follows that

$$\frac{1}{n}\log\left(\binom{n}{na}^2 2^{-na}\right) \to -2a\log a - 2(1-a)\log(1-a) - a\log 2$$

When $a = 1$ the right hand side is $-\log 2 < 0$. By continuity it is also negative for a close to 1.

6.7. Applications

7.1. It is easy to see that

$$E(X_1+Y_1)=\int_0^\infty P(X_1+Y_1>t)\,dt=\int_0^\infty e^{-t^2/2}\,dt=\sqrt{\pi/2}$$

Symmetry implies $EX_1 = EY_1 = \sqrt{\pi/8}$. The law of large numbers implies $X_n/n, Y_n/n \to \sqrt{\pi/8}$. Since $(X_1,Y_1),(X_2,Y_2),\ldots$ is increasing the desired results follows.

7.2. Since there are $\binom{n}{k}$ subsets and each is in the correct order with probability $1/k!$ we have

$$EJ_k^n \le \binom{n}{k}\Big/ k! \le \frac{n^k}{(k!)^2} \approx \frac{\sqrt{n}^{2k}}{k^{2k}e^{-2k}}$$

where in the last equality we have used Stirling's formula without the $\sqrt{k}$ term. Letting $k=\alpha\sqrt{n}$ we have

$$\frac{1}{\sqrt{n}}\log EJ_k^n \to -2\alpha\log\alpha + 2\alpha < 0$$

when $\alpha > e$.

7.3. It is immediate from the definition that $EY_1 = 1$. Grouping the individuals in generation $n+1$ according to their parents in generation n and using $EY_1 = 1$ it is easy to see that this is a martingale. Since Y_n is a nonnegative martingale $Y_n \to Y < \infty$. However, if $\exp(-\theta a)/\mu\varphi(\theta) = b > 1$ and $X_{0,n} \le an$ then $Y_n \ge b^n$ so this cannot happen infinitely often.

7.4. Let k_m be the integer so that $t(k_m,-m)=a_m$. Let $X_{m,n}$ be the amount of time it takes water starting from $(k_m,-m)$ to reach depth n. It is clear that $X_{0,m}+X_{m,n} \ge X_{0,n}$ Since $EX_{0,1}^+ < \infty$ and $X_{m,n} \ge 0$ (iv) holds. (6.1) implies that $X_{0,n}/n \to X$ a.s. To see that the limit is constant, enumerate the edges in some order (e.g., take each row in turn from left to right) $e_1, e_2, \ldots$ and observe that X is measurable with respect to the tail σ-field of the i.i.d. sequence $\tau(e_1), \tau(e_2), \ldots$.

7.5. (i) a_1 is the minimum of two mean one exponentials so it is a mean $1/2$ exponential. (ii) Let S_n be the sum of n independent mean 1 exponentials. Results in Section 1.9 imply that for $a < 1$

$$\frac{1}{n}\log P(S_n \le na) \to -a + 1 + \log a$$

Since there are 2^n paths down to level n, we see that if $f(a) = \log 2 - a + 1 + \log a < 0$ then $\gamma \le a$. Since f is continuous and $f(1) = \log 2$ this must hold for some $a < 1$.

7 Brownian Motion

7.1. Definition and Construction

1.1. Let $\mathcal{A} = \{A = \{\omega : (\omega(t_1), \omega(t_2), \ldots) \in B\} : B \in \mathcal{R}^{\{1,2,\ldots\}}\}$. Clearly, any $A \in \mathcal{A}$ is in the σ-field generated by the finite dimensional sets. To complete the proof, we only have to check that $\mathcal{A}$ is a σ-field. The first and easier step is to note if $A = \{\omega : (\omega(t_1), \omega(t_2), \ldots) \in B\}$ then $A^c = \{\omega : (\omega(t_1), \omega(t_2), \ldots) \in B^c\} \in \mathcal{A}$. To check that $\mathcal{A}$ is closed under countable unions, let $A_n = \{\omega : (\omega(t_1^n), \omega(t_2^n), \ldots) \in B_n\}$, let $t_1, t_2, \ldots$ be an ordering of $\{t_m^n : n, m \geq 1\}$ and note that we can write $A_n = \{\omega : (\omega(t_1), \omega(t_2), \ldots) \in E_n\}$ so $\cup_n A_n = \{\omega : (\omega(t_1), \omega(t_2), \ldots) \in \cup_n E_n\} \in \mathcal{A}$.

1.2. Let $A_n = \{\omega :$ there is an $s \in [0,1]$ so that $|B_t - B_s| \leq C|t-s|^\gamma$ when $|t-s| \leq k/n\}$. For $1 \leq i \leq n-k+1$ let

$$\begin{aligned} Y_{i,n} &= \max\left\{ \left| B\left(\frac{i+j}{n}\right) - B\left(\frac{i+j-1}{n}\right) \right| : j = 0, 1, \ldots k-1 \right\} \\ B_n &= \{ \text{ at least one } Y_{i,n} \text{ is } \leq (2k-1)C/n^\gamma \} \end{aligned}$$

Again $A_n \subset B_n$ but this time if $\gamma > 1/2 + 1/k$

$$\begin{aligned} P(B_n) &\leq nP(|B(1/n)| \leq (2k-1)C/n^\gamma)^k \\ &\leq nP(|B(1)| \leq (2k-1)Cn^{1/2-\gamma})^k \\ &\leq C' n^{k(1/2-\gamma)+1} \to 0 \end{aligned}$$

1.3. The first step is to observe that the scaling relationship (1.2) implies

$$(\star) \qquad \Delta_{m,n} \stackrel{d}{=} 2^{-n/2} \Delta_{1,0}$$

while the definition of Brownian motion shows $E\Delta_{1,0}^2 = t$, and $E(\Delta_{1,0}^2 - t)^2 = C < \infty$. Using ($\star$) and the definition of Brownian motion, it follows that if $k \neq m$ then $\Delta_{k,n}^2 - t2^{-n}$ and $\Delta_{m,n}^2 - t2^{-n}$ are independent and have mean 0 so

$$E\left(\sum_{1\le m\le 2^n} (\Delta_{m,n}^2 - t2^{-n})\right)^2 = \sum_{1\le m\le 2^n} E\left(\Delta_{m,n}^2 - t2^{-n}\right)^2 = 2^n C 2^{-2n}$$

where in the last equality we have used ($\star$) again. The last result and Chebyshev's inequality imply

$$P\left(\left|\sum_{1\le m\le 2^n} \Delta_{m,n}^2 - t\right| \ge 1/n\right) \le Cn^2 2^{-n}$$

The right hand side is summable so the Borel Cantelli lemma (see e.g. (6.1) in Chapter 1 of Durrett (1991)) implies

$$P\left(\left|\sum_{m\le 2^n} \Delta_{m,n}^2 - t\right| \ge 1/n \text{ infinitely often}\right) = 0$$

7.2. Markov Property, Blumenthal's 0-1 Law

2.1. Let $Y = 1_{(T_0>t)}$ and note that $T_0 \circ \theta_1 = R - 1$ so $Y \circ \theta_1 = 1_{(R>1+t)}$. Using the Markov property gives

$$P_x(R > 1 + t|\mathcal{F}_1) = P_{B_1}(T_0 > t)$$

Taking expected value now and recalling $P_x(B_1 = y) = p_1(x, y)$ gives

$$P_x(R > 1 + t) = \int p_1(x, y) P_y(T_0 > t)\, dy$$

2.2. Let $Y = 1_{(T_0>1-t)}$ and note that $Y \circ \theta_t = 1_{(L\le t)}$. Using the Markov property gives

$$P_0(L \le t|\mathcal{F}_t) = P_{B_t}(T_0 > 1 - t)$$

Taking expected value now and recalling $P_0(B_t = y) = p_t(0, y)$ gives

$$P_0(L \le t) = \int p_t(0, y) P_y(T_0 > 1 - t)\, dy$$

2.3. (2.6), (2.7), and the Markov property imply that with probablity one there are times $s_1 > t_1 > s_2 > t_2 \ldots$ that converge to a so that $B(s_n) = B(a)$ and $B(t_n) > B(a)$. In each interval (s_{n+1}, s_n) there is a local maximum.

2.4. (i) $Z \equiv \limsup_{t\downarrow 0} B(t)/f(t) \in \mathcal{F}_0^+$ so (2.7) implies $P_0(Z > c) \in \{0, 1\}$ for each c, which implies that Z is constant almost surely.
(ii) Let $C < \infty$, $t_n \downarrow 0$, $A_N = \{B(t_n) \geq C\sqrt{t_n} \text{ for some } n \geq N\}$ and $A = \cap_N A_N$. A trivial inequality and the scaling relation (1.2) implies

$$P_0(A_N) \geq P_0(B(t_N) \geq C\sqrt{t_N}) = P_0(B(1) \geq C) > 0$$

Letting $N \to \infty$ and noting $A_N \downarrow A$ we have $P_0(A) \geq P_0(B_1 \geq C) > 0$. Since $A \in \mathcal{F}_0^+$ it follows from (2.7) that $P_0(A) = 1$, that is, $\limsup_{t\to 0} B(t)/\sqrt{t} \geq C$ with probability one. Since C is arbitrary the proof is complete.

7.3. Stopping Times, Strong Markov Property

3.1. If $m2^{-n} < t \leq (m+1)2^{-n}$ then $\{S_n < t\} = \{S < m2^{-n}\} \in \mathcal{F}_{m2^{-n}} \subset \mathcal{F}_t$.

3.2. Since constant times are stopping times the last three statements follow from the first three.
$\{S \wedge T \leq t\} = \{S \leq t\} \cup \{T \leq t\} \in \mathcal{F}_t$.
$\{S \vee T \leq t\} = \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t$
$\{S + T < t\} = \cup_{q,r\in\mathbf{Q}: q+r<t} \{S < q\} \cap \{T < r\} \in \mathcal{F}_t$

3.3. Define R_n by $R_1 = T_1$, $R_n = R_{n-1} \vee T_n$. Repeated use of Exercise 3.2 shows that R_n is a stopping time. As $n \uparrow \infty$ $R_n \uparrow \sup_n T_n$ so the desired result follows from (3.3).

Define S_n by $S_1 = T_1$, $S_n = S_{n-1} \wedge T_n$. Repeated use of Exercise 3.2 shows that S_n is a stopping time. As $n \uparrow \infty$ $S_n \downarrow \inf_n T_n$ so the desired result follows from (3.2).

$\limsup_n T_n = \inf_n \sup_{m\geq n} T_m$ and $\liminf_n T_n = \sup_n \inf_{m\geq n} T_m$ so the last two results follow easily from the first two.

3.4. First if $A \in \mathcal{F}_S$ then

$$A \cap \{S < t\} = \cup_n (A \cap \{S \leq t - 1/n\}) \in \mathcal{F}_t$$

On the other hand if $A \cap \{S < t\} \in \mathcal{F}_t$ and the filtration is right continuous then

$$A \cap \{S \leq t\} = \cap_n (A \cap \{S < t + 1/n\}) \in \cap_n \mathcal{F}_{t+1/n} = \mathcal{F}_t$$

3.5. $\{R \leq t\} = \{S \leq t\} \cap A \in \mathcal{F}_t$ since $A \in \mathcal{F}_S$

3.6. (i) Let $r = s \wedge t$.

$$\{S < t\} \cap \{S < s\} = \{S < r\} \in \mathcal{F}_r \subset \mathcal{F}_s$$
$$\{S \le t\} \cap \{S \le s\} = \{S \le r\} \in \mathcal{F}_r \subset \mathcal{F}_s$$

This shows $\{S < t\}$ and $\{S \le t\}$ are in $\mathcal{F}_S$. Taking complements and interestions we get $\{S \ge t\}$, $\{S > t\}$, and $\{S = t\}$ are in $\mathcal{F}_S$.

(ii) $\{S < T\} \cap \{S < t\} = \cup_{q<t} \{S < q\} \cap \{T > q\} \in \mathcal{F}_t$ by (i), so $\{S < T\} \in \mathcal{F}_S$. $\{S < T\} \cap \{T < t\} = \cup_{q<t} \{S < q\} \cap \{q < T < t\} \in \mathcal{F}_t$ by (i), so $\{S < T\} \in \mathcal{F}_T$. Here the unions were taken over rational q. Interchanging the roles of S and T we have $\{S > T\}$ in $\mathcal{F}_S \cap \mathcal{F}_T$. Taking complements and interestions we get $\{S \ge T\}$, $\{S \le T\}$, and $\{S = T\}$ are in $\mathcal{F}_S \cap \mathcal{F}_T$.

3.7. If $A \in \mathcal{R}$ then

$$\{B(S_n) \in A\} \cap \{S_n \le t\} = \cup_{0 \le m \le 2^n t} \{S_n = m/2^n\} \cap \{B(m/2^n) \in A\} \in \mathcal{F}_t$$

by (i) of Exercise 3.6. This shows $\{B(S_n) \in A\} \in \mathcal{F}_{S_n}$ so $B(S_n) \in \mathcal{F}_{S_n}$. Letting $n \to \infty$ and using (3.6) we have $B_S = \lim_n B(S_n) \in \cap_n \mathcal{F}_{S_n} = \mathcal{F}_S$.

7.4. Maxima and Zeros

4.1. (i)Let $Y_s(\omega) = 1$ if $s < t$ and $u < \omega(t-s) < v$, 0 otherwise. Let

$$\bar{Y}_s(\omega) = \begin{cases} 1 & \text{if } s < t,\ 2a - v < \omega(t-s) < 2a - u \\ 0 & \text{otherwise} \end{cases}$$

Symmetry of the normal distribution implies $E_a Y_s = E_a \bar{Y}_s$, so if we let $S = \inf\{s < t : B_s = a\}$ and apply the strong Markov property then on $\{S < \infty\}$

$$E_x(Y_S \circ \theta_S | \mathcal{F}_S) = E_a Y_S = E_a \bar{Y}_S = E_x(\bar{Y}_S \circ \theta_S | \mathcal{F}_S)$$

Taking expected values now gives the desired result.
(ii) Letting $M_t = \max_{0 \le s \le t} B_s$ we can rewrite (4.7) as

$$P_0(M_t > a, u < B_t < v) = P_0(2a - v < B_t < 2a - u)$$

Letting the interval (u, v) shrink to x we see that

$$P_0(M_t > a, B_t = x) = P_0(B_t = 2a - x) = \frac{1}{\sqrt{2\pi t}} e^{-(2a-x)^2/2t}$$

Differentiating with respect to a now we get the joint density

$$P_0(M_t = a, B_t = x) = \frac{2(2a-x)}{\sqrt{2\pi t^3}} e^{-(2a-x)^2/2t}$$

4.2. We begin by noting symmetry and Exercise 2.1 imply

$$\begin{aligned} P_0(R \le 1+t) &= 2\int_0^\infty p_1(0,y) \int_0^t P_y(T_0 = s)\, ds\, dy \\ &= \int_0^t 2\int_0^\infty p_1(0,y) P_y(T_0 = s)\, dy\, ds \end{aligned}$$

by Fubini's theorem, so the integrand gives the density $P_0(R = 1+t)$. Since $P_y(T_0 = t) = P_0(T_y = t)$, (4.7) gives

$$\begin{aligned} P_0(R = 1+t) &= 2\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{1}{\sqrt{2\pi t^3}} y e^{-y^2/2t}\, dy \\ &= \frac{1}{2\pi t^{3/2}} \int_0^\infty y e^{-y^2(1+t)/2t}\, dy = \frac{1}{2\pi t^{3/2}} \frac{t}{(1+t)} \end{aligned}$$

7.5. Martingales

5.1. It follows from (5.6) that

$$\cosh(\theta B_t) e^{-\theta^2 t/2} = \frac{1}{2}\left\{\exp(\theta B_t - \theta^2 t/2) + \exp(-\theta B_t - (-\theta)^2 t/2)\right\}$$

is a martingale. (5.1) and this imply

$$1 = E_0\left(\cosh(B_{T\wedge t}) e^{-\theta^2 (T\wedge t)/2}\right)$$

Letting $t \to \infty$ and using the bounded convergence theorem we have

$$1 = \cosh(a) E_0\left(e^{-\theta^2 T/2}\right)$$

5.2. It follows from (5.1) and (5.6) that

$$1 = E_0 \exp(\theta B_{\tau\wedge t} - \theta^2(\tau \wedge t)/2)$$

$\theta = b + \sqrt{b^2 + 2\lambda}$ is the larger root of $\theta b - \theta^2/2 = -\lambda$ and $B_{T\wedge t} \le a + b(T \wedge t)$ so using the bounded convergence theorem we have

$$1 = E_0\left(\exp(\theta(a + b\tau) - \theta^2\tau/2); \tau < \infty\right)$$

Substituting in the value of θ and rearranging gives the desired result.

5.3. (i) $T_a = \sigma$ when $T_a < T_b$ and $T_a = \sigma + T_a \circ \theta_\sigma$ when $T_b < T_a$. Using the defintion of conditional expectation and (1.3) in Chapter 4 we have

$$\begin{aligned} E_x\left(e^{-\lambda T_a}; T_b < T_a\right) &= E_x\left(E_x\left(e^{-\lambda(\sigma + T_a \circ \theta_\sigma}\middle| \mathcal{F}_\sigma\right); T_b < T_a\right) \\ &= E_x\left(e^{-\lambda\sigma} E_x\left(e^{-\lambda T_a} \circ \theta_\sigma \middle| \mathcal{F}_\sigma\right); T_b < T_a\right) \end{aligned}$$

Since $B_\sigma = b$ on $T_b < T_a$, the strong Markov property implies

$$E_x\left(e^{-\lambda T_a} \circ \theta_\sigma \middle| \mathcal{F}_\sigma\right) = E_b\left(e^{-\lambda T_a}\right)$$

and completes the proof of the formula.

(ii) Letting $u = E_x(e^{-\lambda\sigma}; T_a < T_b)$ and $v = E_x(e^{-\lambda\sigma}; T_b < T_a)$ then using (4.4) we can write the equations as

$$\begin{aligned} \exp(-(x-a)\sqrt{2\lambda}) &= u + v\exp(-(b-a)\sqrt{2\lambda}) \\ \exp(-(b-x)\sqrt{2\lambda}) &= v + u\exp(-(b-a)\sqrt{2\lambda}) \end{aligned}$$

Multiplying the first equation by $\exp((b-a)\sqrt{2\lambda})$ and subtracting the second gives

$$\sinh((b-x)\sqrt{2\lambda}) = \sinh((b-a)\sqrt{2\lambda})u$$

One can solve for v in a similar way.

5.4. (5.1) and (5.8) imply

$$E(B(U\wedge t)^4 - 6(U\wedge t)B(U\wedge t)^2) = -3E(U\wedge t)^2$$

By putting (a, b) inside a larger symmetric interval and using (5.5) we get $EU < \infty$. Letting $t \to \infty$, using the dominated convergence theorem on the left hand side, and the monotone convergence theorem on the right gives $E(B_U^4 - 6UB_U^2) = -3EU^2$ so using Cauchy-Schwarz

$$EU^2 \le 2EUB_U^2 \le 2\left(EU^2\right)^{1/2}\left(EB_U^4\right)^{1/2}$$

and it follows that $EU^2 \le 4EB_U^4$.

5.5. $p_t(x,y) = (2\pi t)^{-1/2} e^{-(y-x)^2/2t}$. Differentiating gives

$$\begin{aligned}
\frac{\partial p_t}{\partial t} &= -\frac{1}{2}(2\pi)^{-1/2} t^{-3/2} e^{-(y-x)^2/2t} + (2\pi t)^{-1/2} e^{-(y-x)^2/2t} \frac{(y-x)^2}{2t^2} \\
\frac{\partial p_t}{\partial y} &= (2\pi t)^{-1/2} e^{-(y-x)^2/2t} \cdot -\frac{(y-x)}{t} \\
\frac{\partial^2 p_t}{\partial y^2} &= (2\pi t)^{-1/2} e^{-(y-x)^2/2t} \frac{(y-x)^2}{t^2} + (2\pi t)^{-1/2} e^{-(y-x)^2/2t} \cdot -\frac{1}{t}
\end{aligned}$$

so

$$\partial p_t/\partial t = (1/2)\partial^2 p_t \partial y^2$$

To check the second claim note that

$$\begin{aligned}
\frac{\partial}{\partial t}(p_t(x,y)u(t,y)) &= u(t,y)\frac{\partial}{\partial t} p_t(x,y) + p_t(x,y)\frac{\partial}{\partial t} u(t,y) \\
&= u(t,y)\frac{1}{2}\frac{\partial^2}{\partial y^2} p_t(x,y) + p_t(x,y)\frac{\partial}{\partial t} u(t,y)
\end{aligned}$$

Integrating by parts twice in the first term results in

$$\int p_t(x,y)\left(\frac{1}{2}\frac{\partial^2}{\partial y^2} u(t,y) + \frac{\partial}{\partial t} u(t,y)\right) dy = 0$$

5.6. If we let $u(t,x) = x^6 - atx^4 + bt^2x^2 - ct^3$ then

$$\begin{aligned}
\frac{\partial u}{\partial x} &= 6x^5 - 4atx^3 + 2bt^2 x \\
\frac{\partial^2 u}{\partial x^2} &= 30x^4 - 12atx^2 + 2bt^2 \\
\frac{\partial u}{\partial t} &= -ax^4 + 2btx^2 - 3ct^2
\end{aligned}$$

To have $\partial u/\partial t = -\frac{1}{2}\partial^2 u/\partial x^2$ we need

$$-a + 15 = 0 \qquad 2b - 6a = 0 \qquad -3c + b = 0$$

i.e., $a = 15$, $b = 45$, $c = 15$. Using (5.1) we have

$$E\left(B^6_{T\wedge t} - 15(T\wedge t)B^4_{T\wedge t} + 45(T\wedge t)^2 B^2_{T\wedge t}\right) = 15E(T\wedge t)^3$$

From (5.5) and (5.9) we know $ET = a^2$ and $ET^2 = 5a^4/3 < \infty$. Using the dominated convergence theorem on the left and the monotone convergence theorem on the right, we have

$$a^6\left(1 - 15 + 45 \cdot \frac{5}{3}\right) = 15ET^3$$

so $ET^3 = 61/15$.

5.7. $u(t,x) = (1+t)^{-1/2}\exp(x_t^2/(1+t)) = (2\pi)^{1/2} p_{1+t}(0, ix)$ where $i = \sqrt{-1}$ so $\partial u/\partial t + (1/2)\partial^2 u/\partial x^2 = 0$ and Exercise 5.5 implies $u(t, B_t)$ is a martingale. Being a nonnegative martingale it must converge to a finite limit a.s. However, if we let $x_t = B_t/((1+t)\log(1+t))^{1/2}$ then

$$(1+t)^{-1/2}\exp(B_t^2/(1+t)) = (1+t)^{-1/2}\exp(x_t^2 \log(1+t))$$

so we cannot have $x_t^2 \geq 1/2$ i.o.

7.6. Donsker's Theorem

6.1. Exercise 5.4 implies $ET_{u,v}^2 \leq C\int x^4\,\mu_{u,v}(dx)$ so using a computation after (6.2)

$$E\left(T_{U,V}^2\right) \leq CE\int x^4\,\mu_{U,V}(dx) = CEX^4$$

6.2. $\varphi(\omega) = \max_{0\leq s\leq 1}\omega(s) - \min_{0\leq s\leq 1}\omega(s)$ is continuous so $(*)$ implies

$$\frac{1}{\sqrt{n}}\left(\max_{0\leq m\leq n} S_m - \min_{0\leq m\leq n} S_m\right) \Rightarrow \max_{0\leq s\leq 1} B_s - \min_{0\leq s\leq 1} B_s$$

6.3. (i) Clearly $(1/n)\sum_{m=1}^n B(m/n) - B((m-1)/n)$ has a normal distribution. The sums converges a.s. and hence in distribuiton to $\int_0^1 B_t\,dt$, so by Exercise 3.9 the integral has a normal distribution. To compute the variance, we write

$$\begin{aligned} E\left(\int_0^1 B_t\,dt\right)^2 &= E\left(\int_0^1\int_0^1 B_s B_t\,dt\,ds\right) \\ &= 2\left(\int_0^1\int_s^1 E(B_s B_t)\,dt\,ds\right) \\ &= 2\int_0^1\int_s^1 s\,dt\,ds \\ &= 2\int_0^1 s(1-s)\,ds = 2\left(\frac{s^2}{2} - \frac{s^3}{3}\right)\Bigg|_0^1 = \frac{1}{3} \end{aligned}$$

(ii) Let $X_{n,m} = (n+1-m)X_m/n^{3/2}$. $EX_{n,m} = 0$ and

$$\sum_{m=1}^n EX_{n,m}^2 = n^{-3}\sum_{k=1}^n j^2 \to 1/3$$

To check (ii) in (4.5) in Chapter 2 now, we observe that if $1 \le m \le n$

$$E\Big(\Big(n^{-3/2}(n+1-m)X_m/n^{3/2}\Big)^2; \Big|n^{-3/2}(n+1-m)X_m/n^{3/2}\Big| > \epsilon\Big)$$
$$\le \frac{1}{n}E\left(X_1^2; |X_1| > \epsilon\sqrt{n}\right)$$

so the sum in (ii) is $\le E(X_1^2; |X_1| > \epsilon\sqrt{n}) \to 0$ by dominated convergence.

7.7. CLT's for Dependent Variables

7.1. On $\{\zeta_n = i\}$ we have

$$E(X_{n+1}|\mathcal{G}_n) = \int x\,dH_i(x) = 0$$
$$E(X_{n+1}^2|\mathcal{G}_n) = \int x^2\,dH_i(x) = \sigma_i^2$$

The ergodic theorem for Markov chains, Example 2.2 in Chapter 6 (or Exercise 5.2 in Chapter 5) implies that

$$n^{-1}\sum_{m=1}^n \sigma^2(X_m) \to \sum_x \sigma^2(x)\pi(x) \quad \text{a.s.}$$

7.2. Let $\mu = P(\eta_n = 1)$ and let $X_n = \eta_n - 1/4$. Since X_n is 1-dependent, the formula in Example 7.1 implies $\sigma^2 = EX_0^2 + 2E(X_0X_1)$. $EX_0^2 = \text{var}(\eta_0) = (1/4)(3/4)$ since η_0 is Bernoulli(1/4). For the other term we note

$$EX_0X_1 = E\left[(\eta_0 - 1/4)(\eta_1 - 1/4)\right] = -1/16$$

since $EZ_0Z_1 = 0$ and $EZ_i = 1/4$. Combining things we have $\sigma^2 = 2/16$.
To identify Y_0 we use the formula from the proof and the fact that X_1 is independent of $\mathcal{F}_{-1}$, to conclude

$$Y_0 = X_0 - E(X_0|\mathcal{F}_{-1}) + E(X_1|\mathcal{F}_0) - EX_1$$
$$= 1_{(\xi_0=H,\xi_1=T)} - \frac{1}{2}1_{(\xi_0=H)} + \frac{1}{2}1_{(\xi_1=H)} - 1/4$$

7.3. The Markov property implies

$$E(X_0|\mathcal{F}_{-n}) = \sum_j p^{n-1}(\zeta_{-n}, j)\mu_j$$

Since Markov chain is irreducible with a finite state space, combining Exercise 5.10 with fact that $\sum_i \pi(i)\mu_i = 0$ shows there are constants $0 < \gamma, C < \infty$ so that

$$\sup_i \left| \sum_j p^{n-1}(i,j)\mu_j \right| \le Ce^{-\gamma n}$$

7.8. Empirical Distributions, Brownian Bridge

8.1. Exercise 4.1 implies that

$$P\left(\max_{0\le t\le 1} B_t > b, -\epsilon < B_1 < \epsilon\right) = P(2b - \epsilon < B_1 < 2b + \epsilon)$$

Since $P(|B_1| < \epsilon) \sim 2\epsilon \cdot (2\pi)^{-1/2}$ it follows that

$$P\left(\max_{0\le t\le 1} B_t > b \middle| -\epsilon < B_1 < \epsilon\right) \to e^{-(2b^2)/2}$$

7.9. Laws of the Iterated Logarithm

9.1. Letting $f(t) = 2(1+\epsilon)\log\log\log t$ and using a familar formula from the proof of (9.1)

$$P_0(B_{t_k} > (t_k f(t_k))^{1/2}) \sim \kappa f(t_k)^{-1/2} \exp(-(1+\epsilon)\log k)$$

The right-hand side is summable so

$$\limsup_{k\to\infty} B_{t_k}/(2t_k \log\log\log t_k)^{1/2} \le 1$$

For a bound in the other direction take $g(t) = 2(1-\epsilon)\log\log\log t$ and note that

$$P_0(B_{t_k} - B_{t_{k-1}} > ((t_k - t_{k-1})g(t_k))^{1/2}) \sim \kappa g(t_k)^{-1/2} \exp(-(1-\epsilon)\log k)$$

The sum of the right-hand side is ∞ and the events on the left are independent so

$$P_0\left(B_{t_k} - B_{t_{k-1}} > ((t_k - t_{k-1})g(t_k))^{1/2} \text{ i.o.}\right) = 1$$

Combining this with the result for the lim sup and noting $t_{k-1}/t_k \to 0$ the desired result follows easily.

9.2. $E|X_i|^\alpha = \infty$ implies $\sum_{m=1}^\infty P(|X_i| > Cn^{1/\alpha}) = \infty$ for any C. Using the second Borel-Cantelli now we see that $\limsup_{n\to\infty} |X_n|/n^{1/\alpha} \geq C$, i.e., the $\limsup = \infty$. Since $\max\{|S_n|, |S_{n-1}|\} \geq |X_n|/2$ it follows that $\limsup_{n\to\infty} S_n/n^{1/\alpha} = \infty$.

9.3. (9.1) implies that

$$\limsup_{n\to\infty} S_n/(2n \log\log n)^{1/2} = 1 \qquad \liminf_{n\to\infty} S_n/(2n \log\log n)^{1/2} = -1$$

so the limit set is contained in $[-1, 1]$. On the other hand

$$\sum_{m=1}^{\infty} P(X_n > \epsilon\sqrt{n}) < \infty$$

for any ϵ so $X_n/\sqrt{n} \to 0$. This shows that the differences $(S_{n+1} - S_n)/\sqrt{n} \to 0$ so as $S_n/(2n \log\log n)^{1/2}$ wanders back and forth between 1 and -1 it fills up the entire interval.

Appendix: Measure Theory

A.1. Lebesgue-Stieltjes Measures

1.1. (i) If $A, B \in \cup_i \mathcal{F}_i$ then $A, B \in \mathcal{F}_n$ for some n, so A^c, $A \cup B \in \mathcal{F}_n$.
(ii) Let $\Omega = [0,1)$, $\mathcal{F}_n = \sigma(\{[m/2^n, (m+1)/2^n), 0 \le m < 2^n\}$. $\sigma(\cup_i \mathcal{F}_i) =$ the Borel subsets of $[0,1)$ but $[0,1/3) \notin \cup_i \mathcal{F}_i$.

1.2. If A has asymptotic density θ then A^c has asymptotic density $1-\theta$. However, $\mathcal{A}$ is not closed under unions. To prove this note that if A has the property that $|\{2k-1, 2k\} \cap A| = 1$ for all integers k then A has asymptotic density $1/2$. Let A consist of the odd integers between $(2k-1)!$ and $(2k)!$ and the even integers between $(2k)!$ and $(2k+1)!$. Let $B = 2\mathbf{Z}$. Then

$$\limsup_{n\to\infty} |(A \cup B) \cap \{1, 2, \ldots n\}|/n = 1$$
$$\liminf_{n\to\infty} |(A \cup B) \cap \{1, 2, \ldots n\}|/n = 1/2$$

1.3. (i) $B = A + (B - A)$ so $\mu(B) = \mu(A) + \mu(B - A) \ge \mu(A)$.
(ii) Let $A'_n = A_n \cap A$, $B_1 = A'_1$ and for $n > 1$, $B_n = A'_n - \cup_{m=1}^{n-1} (A'_m)^c$. Since the B_n are disjoint and have union A we have using (i) and $B_m \subset A_m$

$$\mu(A) = \sum_{m=1}^{\infty} \mu(B_m) \le \sum_{m=1}^{\infty} \mu(A_m)$$

(iii) Let $B_n = A_n - A_{n-1}$. Then the B_n are disjoint and have $\cup_{m=1}^{\infty} B_m = A$, $\cup_{m=1}^{n} B_m = A_n$ so

$$\mu(A) = \sum_{m=1}^{\infty} \mu(B_m) = \lim_{n\to\infty} \sum_{m=1}^{n} \mu(B_m) = \lim_{n\to\infty} \mu(A_n)$$

(iv) $A_1 - A_n \uparrow A_1 - A$ so (iii) implies $\mu(A_1 - A_n) \uparrow \mu(A_1 - A)$. Since $\mu(A_1 - B) = \mu(A_1) - \mu(B)$ it follows that $\mu(A_n) \downarrow \mu(A)$.

1.4. $\mu(\mathbf{Z}) = 1$ but $\mu(\{n\}) = 0$ for all n and $\mathbf{Z} = \cup_n \{n\}$ so μ is not countably additive on $\sigma(\mathcal{A})$.

1.5. By fixing the sets in coordinates $2, \ldots, d$ it is easy to see $\sigma(\mathcal{R}_o^d) \supset \mathcal{R} \times \mathcal{R}_o \times \mathcal{R}_o$ and iterating gives the desired result.

A.2. Carathéodary's Extension Theorem

2.1. Let $\mathcal{C} = \{\{1,2\},\{2,3\}\}$. Let μ be counting measure. Let $\nu(A) = 2$ if $2 \in A$, 0 otherwise.

A.3. Completion, etc

3.1. By (3.1) there are $A_i \in \mathcal{A}$ so that $\cup_i A_i \supset B$ and $\sum_i \mu(A_i) \le \mu(B) + \epsilon/2$. Pick I so that $\sum_{i>I} \mu(A_i) < \epsilon/2$, and let $A = \cup_{i \le I} A_i$. Since $B \subset \cup_i A_i$, we have $B - A \subset \cup_{i>I} A_i$ and hence $\mu(B - A) \le \mu(\cup_{i>I} A_i) \le \epsilon/2$. To bound the other difference we note that $A - B \subset (\cup_i A_i) - B$ and $\cup_i A_i \supset B$ so $\mu(A - B) \le \mu(\cup_i A_i) - \mu(B) \le \epsilon/2$.

3.2. (i) For each rational r, let $E_r = r +' D_q$. The E_r are disjoint subsets of $(0,1]$, so $\sum_r \mu(E_r) \le 1$ but we have $\mu(E_r) = \mu(D_q)$, so $\mu(D_q) = 0$.
(ii) By translating A we can suppose without loss of generality that $\mu(A \cap (0,1]) > 0$. For each rational q let $A_q = A \cap B_q$. If every A_q is measurable then $\mu(A_q) = 0$ by (i) and $\mu(A \cap (0,1]) = \sum_q \mu(A_q) = 0$ a contradicition.

3.3. Write the rotated rectangle B as $\{(x,y) : a \le x \le b, f(x) \le y \le g(x)\}$ where f and g are piecewise linear. Subdividing $[a,b]$ into n equal pieces, using the upper Riemann sum for g and the lower Riemann sum for f, then letting $n \to \infty$ we conclude that $\lambda^*(B) = \lambda(A)$.

(ii) By covering D with the appropriate rotations and translations of sets used to cover C, we conclude $\lambda^*(D) \le \lambda^*(C)$. Interchanging the roles of C and D proves that equality holds.

A.4. Integration

4.1. Let $A_\delta = \{x : f(x) > \delta\}$ and note that $A_\delta \uparrow A_0$ as $\delta \downarrow 0$. If $\mu(A_0) > 0$ then $\mu(A_\delta) > 0$ for some $\delta > 0$. If $\mu(A_\delta) > 0$ then $\mu(A_\delta \cap [-m,m]) > 0$ for some m. Letting $h(x) = \delta$ on $A_\delta \cap [-m,m]$ and 0 otherwise we have

$$\int f\, d\mu \ge \int h\, d\mu = \delta \mu(A_\delta \cap [-m,m]) > 0$$

a contradiction.

4.2. Let $g = \sum_{m=1}^{\infty} \frac{m}{2^n} 1_{E_{n,m}}$ Since $g \le f$, (iv) in (4.5) implies

$$\sum_{m=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) = \int g\,d\mu \le \int f\,d\mu$$

$$\limsup_{n\to\infty} \sum_{m=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) \le \int f\,d\mu$$

For the other inequality let h be the class used to define the integral. That is, $0 \le h \le f$, h is bounded, and $H = \{x : h(x) > 0\}$ has $\mu(H) < \infty$.

$$g + \frac{1}{2^n} 1_H \ge f 1_H \ge h$$

so using (iv) in (4.5) again we have

$$\frac{1}{2^n}\mu(H) + \sum_{m=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) \ge \int h\,d\mu$$

Letting $n \to \infty$ now gives

$$\liminf_{n\to\infty} \sum_{m=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) \ge \int h\,d\mu$$

Since h is an aribitrary member of the defining class the desired result follows.

4.3. Since

$$\int |g - (\varphi - \psi)|\,d\mu \le \int |g^+ - \varphi|\,d\mu + \int |g^- - \psi|\,d\mu$$

it suffices to prove the result when $g \ge 0$. Using Exercise 4.2, we can pick n large enough so that if $E_{n,m} = \{x : m/2^n \le f(x) < (m+1)/2^n\}$ and $h(x) = \sum_{m=1}^{\infty}(m/2^n) 1_{E_{n,m}}$ then $\int g - h\,d\mu < \epsilon/2$. Since

$$\sum_{n=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) = \int h\,d\mu \le \int g\,d\mu < \infty$$

we can pick M so that $\sum_{m>M} \frac{m}{2^n} \mu(E_{n,m}) < \epsilon/2$. If we let

$$\varphi = \sum_{m=1}^{M} \frac{m}{2^n} 1_{E_{n,m}}$$

then $\int |g - \varphi|\, d\mu = \int g - h\, d\mu + \int h - \varphi\, d\mu < \epsilon$.

(ii) Pick A_m that are finite unions of open intervals so that $|A_m \Delta E_{n,m}| \le \epsilon M^{-2}$ and let

$$q(x) = \sum_{m=1}^{M} \frac{m}{2^n} 1_{A_m}$$

Now the sum above is $= \sum_{j=1}^{k} c_j 1_{(a_{j-1},a_j)}$ almost everywhere (i.e., except at the end points of the intervals) for some $a_0 < a_1 < \cdots < a_k$ and $c_j \in \mathbf{R}$.

$$\int |\varphi - q|\, d\mu \le \sum_{m=1}^{M} \frac{m}{2^n} \mu(A_m \Delta E_{n,m}) \le \frac{\epsilon}{2^n}$$

(iii) To make the continuous function replace each $c_j 1_{(a_{j-1},a_j)}$ by a function r_j that is 0 on $(a_{j-1}, a_j)^c$, c_j on $[a_{j-1} + \delta_j, a_j - \delta_j]$, and linear otherwise. If we let $r(x) = \sum_{j=1}^{k} r_j(x)$ then

$$\int |q(x) - r(x)| = \sum_{j=1}^{k} \delta_j c_j < \epsilon$$

if we take $\delta_j c_j < \epsilon/k$.

4.4. Suppose $g(x) = c1_{(a,b)}(x)$. In this case

$$\int g(x) \cos nx\, dx = c \int_a^b \cos nx\, dx = \frac{c}{n} \sin nx \Big|_a^b$$

so the absolute value of the integral is smaller than $2|c|/n$ and hence $\to 0$. Linearity extends the last result to step functions. Using Exercise 4.3 we can approximate g by a step function q so that $\int |g - q|\, dx < \epsilon$. Since $|\cos nx| \le 1$ the triangle inequality implies

$$\left| \int g(x) \cos nx\, dx \right| \le \left| \int q(x) \cos nx\, dx \right| + \int |g(x) - q(x)|\, dx$$

so the $\limsup$ of the left hand side $< \epsilon$ and since ϵ is arbitrary the proof is complete.

4.5. (a) does not imply (b): let $f(x) = 1_{[0,1]}$. This function is continuous at $x \ne 0$ and 1 but if $g = f$ a.e. then g will be discontinuous at 0 and 1.
(b) does not imply (a): $f = 1_{\mathbf{Q}}$ where $\mathbf{Q}$ = the rationals is equal a.e. to the continuous function that is $\equiv 0$. However $1_{\mathbf{Q}}$ is not continuous anywhere.

4.6. Let $E_m^n = \{\omega : x_{m-1}^n \le f(x) < x_m^n\}$, $\psi_n = x_{m-1}^n$ on E_m^n and $\varphi_n = x_m^n$ on E_m^n. $\psi_n \le f \le \varphi_n \le \psi_n + \text{mesh}(\sigma_n)$ so (iv) in (4.7) implies

$$\int \psi_n \, d\mu \le \int f \, d\mu \le \int \varphi_n \, d\mu \le \int \psi_n \, d\mu + \text{mesh}(\sigma_n)\mu(\Omega)$$

It follows from the last inequality that if we have a sequence of partitions with $\text{mesh}(\sigma_n) \to 0$ then

$$\bar{U}(\sigma_n) = \int \psi_n \, d\mu, \;\; \bar{L}(\sigma_n) = \int \varphi_n \, d\mu, \; \to \int f \, d\mu$$

A.5. Properties of the Integral

5.1. If $|g| \le M$ a.e. then $|fg| \le M|f|$ a.e. and (iv) in (4.7) implies

$$\int |fg| \, d\mu \le M \int |f| \, d\mu = M\|f\|_1$$

Taking the inf over M now gives the desired result.

5.2. If $\mu(\{x : |f(x)| > M\}) = 0$ then $\int |f|^p \, d\mu \le M^p$ so $\limsup_{p\to\infty} \|f\|_p \le M$. On the other hand if $\mu(\{x : |f(x)| > N\}) = \delta > 0$ then $\int |f|^p \, d\mu \ge \delta N^p$ so $\liminf_{p\to\infty} \|f\|_p \ge N$. Taking the inf over M and sup over N gives the desired result.

5.3. Since $|f+g| \le |f| + |g|$ we have

$$\begin{aligned} \int |f+g|^p \, dx &\le \int |f|\,|f+g|^{p-1} \, dx + \int |g|\,|f+g|^{p-1} \, dx \\ &\le \|f\|_p \| \, |f+g|^{p-1}\|_q + \|g\|_p \| \, |f+g|^{p-1}\|_q \end{aligned}$$

Now $q = p/(p-1)$ so

$$\| \, |f+g|^{p-1}\|_q = \left(\int |f+g|^p \, dx\right)^{1/q} = \|f+g\|_q^{p-1}$$

and dividing each side of the first display by $\|f+g\|_q^{p-1}$ gives the desired result.
(ii) Since $|f+g| \le |f| + |g|$, (iv) and (iii) of (4.7) imply that

$$\int |f+g| \, dx \le \int |f| + |g| \, dx \le \int |f| \, dx + \int |g| \, dx$$

It is easy to see that if $\mu\{x : |f(x)| \geq M\} = 0$ and $\mu\{x : |g(x)| \geq N\} = 0$ then $\mu\{x : |f(x) + g(x)| \geq M + N\} = 0$. Taking the inf over M and N we have

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$$

5.4. If σ_n is a sequence of partitions with $\text{mesh}(\sigma_n) \to 0$ then $f^{\sigma_n}(x) \to f(x)$ at all points of continuity of f so the bounded convergence theorem implies

$$U(\sigma_n) = \int_{[a,b]} f^{\sigma_n}(x)\, dx \to \int_{[a,b]} f(x)\, dx$$

A similar argument to applies to the lower Riemann sum and completes the proof.

5.5. If $0 \leq (g_n + g_1^-) \uparrow (g + g_1^-)$ then the monotone convergence theorem implies

$$\int g_n - g_1^-\, d\mu \uparrow \int g - g_1^-\, d\mu$$

Since $\int g_1^-\, d\mu < \infty$ we can add $\int g_1^-\, d\mu$ to both sides and use (ii) of (4.5) to get the desired result.

5.6. $\sum_{m=0}^n g_m \uparrow \sum_{m=0}^\infty g_m$ so the monotone convergence theorem implies

$$\begin{aligned}\int \sum_{m=0}^\infty g_m\, d\mu &= \lim_{n\to\infty} \int \sum_{m=0}^n g_m\, d\mu \\ &= \lim_{n\to\infty} \sum_{m=0}^n \int g_m\, d\mu = \sum_{m=0}^\infty \int g_m\, d\mu\end{aligned}$$

5.7. (i) follows from the monotone convergence theorem.
(ii) Let $f = |g|$ and pick n so that

$$\int |g|\, d\mu - \int |g| \wedge n\, d\mu < \frac{\epsilon}{2}$$

Then let $\delta < \epsilon/(2n)$. Now if $\mu(A) < \delta$

$$\int_A g\, d\mu \leq \int |g| - (|g| \wedge n)\, d\mu + \int_A |g| \wedge n\, d\mu < \frac{\epsilon}{2} + \mu(A)n < \epsilon$$

5.8. $\sum_{m=0}^n f1_{E_m} \to f1_E$ and is dominated by the integrable function $|f|$, so the dominated convergence theorem implies

$$\int_E f\, d\mu = \lim_{n\to\infty} \sum_{m=0}^n \int_{E_m} f\, d\mu$$

5.9. If $x_n \to c \in (a,b)$ then $f1_{[a,x_n]} \to f1_{[a,c]}$ a.e. and is dominated by $|f|$ so the dominated convergence theorem implies $g(x_n) \to g(c)$.

5.10. First suppose $f \geq 0$. Let $\varphi_n(x) = m/2^n$ on $\{x : m/2^n \leq f(x) < (m+1) < 2^n\}$ for $1 \leq m < n2^n$ and 0 otherwise. As $n \uparrow \infty$, $\varphi_n(x) \uparrow f(x)$ so so the dominated convergence theorem implies $\int |f - \varphi_n|^p \, d\mu \to 0$. To extend to the general case now, let φ_n^+ approximate f^+, let φ_n^- approximate f^-, and let $\varphi = \varphi^+ - \varphi^-$ and note that

$$\int |f - \varphi| \, d\mu = \int |f^+ - \varphi_n^+| \, d\mu + \int |f^- - \varphi_n^-| \, d\mu$$

5.11. Exercise 5.6 implies $\int \sum |f_n| \, d\mu = \sum_n \int |f_n| \, d\mu < \infty$ so $\sum |f_n| < \infty$ a.e.,

$$g_n = \sum_{m=1}^{n} f_m \to g = \sum_{m=1}^{\infty} f_m \quad \text{a.e.}$$

and the dominated convergence theorem implies $\int g_n \, d\mu \to \int g \, d\mu$. To finish the proof now we notice that (iv) of (4.7) implies

$$\int g_n \, d\mu = \sum_{m=1}^{n} \int f_m \, d\mu$$

and we have $\sum_{m=1}^{\infty} \left| \int f_m \, d\mu \right| \leq \sum_{m=1}^{\infty} \int |f_m| \, d\mu < \infty$ so

$$\sum_{m=1}^{n} \int f_m \, d\mu \to \sum_{m=1}^{\infty} \int f_m \, d\mu$$

A.6. Product Measure, Fubini's Theorem

6.1. The first step is to observe $\mathcal{A} \times \mathcal{B}_o \subset \sigma(\mathcal{A}_o \times \mathcal{B}_o)$ so $\sigma(\mathcal{A}_o \times \mathcal{B}_o) = \mathcal{A} \times \mathcal{B}$. Since $\mathcal{A}_o \times \mathcal{B}_o$ is closed under intersection, uniqueness follows from (2.2).

6.2. $|f| \geq 0$ so

$$\int |f| \, d(\mu_1 \times \mu_2) = \int_X \int_Y |f(x,y)| \, \mu_2(dy) \, \mu_1(dx) < \infty$$

This shows f is integrable and the result follows from (6.2).

6.3. Let $Y = [0, \infty)$, $\mathcal{B}$ = the Borel subsets, and λ = Lebesgue measure. Let $f(x, y) = 1_{\{(x,y):0<y<g(x)\}}$, and observe

$$\begin{aligned}
\int f \, d(\mu \times \lambda) &= (\mu \times \lambda)(\{(x, y) : 0 < y < g(x)\}) \\
\int_X \int_Y f(x, y) \, dy \, \mu(dx) &= \int_X g(x) \, \mu(dx) \\
\int_Y \int_X f(x, y) \, \mu(dx) \, dy &= \int_0^\infty \mu(g(x) > y) \, dy
\end{aligned}$$

6.4. (i) Let $f(x, y) = 1_{(a<x\le y\le b)}$ and observe

$$\begin{aligned}
(\mu \times \nu)(\{(x, y) : a < x \le y \le b\}) &= \int f \, d(\mu \times \nu) \\
&= \int \int f \, d\mu \, d\nu = \int_{(a,b]} \{F(y) - F(a)\} dG(y)
\end{aligned}$$

(ii) Using (i) twice we have

$$\begin{aligned}
&\int_{(a,b]} \{F(y) - F(a)\} dG(y) + \int_{(a,b]} \{G(y) - G(a)\} dF(y) \\
&\quad = F(a)\{G(b) - G(a)\} + G(a)\{F(b) - F(a)\} \\
&\quad + (\mu \times \nu)((a, b] \times (a, b]) + (\mu \times \nu)(\{(x, x) : a < x \le b\})
\end{aligned}$$

The third term is $(F(b) - F(a))(G(b) - G(a))$ so the sum of the first three is $F(b)G(b) - F(a)G(a)$.
(iii) If $F = G$ is continuous then the last term vanishes.

6.5. Let $f(x, y) = 1_{\{(x,y):x<y\le x+c\}}$.

$$\begin{aligned}
\int \int f(x, y) \, \mu(dy) \, dx &= \int F(x + c) - F(x) \, dx \\
\int \int f(x, y) \, dx \, \mu(dy) &= c \int \mu(dy) = c\mu(\mathbf{R})
\end{aligned}$$

so the desired result follows from (6.2).

6.6. We begin by observing that

$$\int_0^a \int_0^\infty |e^{-xy} \sin x| \, dy \, dx = \int_0^a \frac{|\sin x|}{x} \, dx < \infty$$

since $\sin x/x$ is bounded on $[0,a]$. So Exercise 6.2 implies $e^{-xy}\sin x$ is integrable in the strip. Removing the absolute values from the last computation gives the left hand side of the desired formula. To get the right hand side integrate by parts twice:

$$f(x) = e^{-xy} \quad f'(x) = -ye^{-xy} \quad g'(x) = \sin x \quad g(x) = -\cos x$$
$$\int_0^a e^{-xy}\,\sin x\,dx = -e^{-ay}\cos a + 1 - \int_0^a ye^{-xy}\cos x\,dx$$
$$f(x) = -ye^{-xy} \quad f'(x) = y^2e^{-xy} \quad g'(x) = \cos x \quad g(x) = \sin x$$
$$-\int_0^a ye^{-xy}\,\cos x\,dx = -ye^{-ay}\sin a - \int_0^a y^2e^{-xy}\sin x\,dx$$

Rearranging gives

$$\int_0^a e^{-xy}\,\sin x\,dx = \frac{1}{1+y^2}(1 - e^{-ay}\cos a - ye^{-ay}\sin a)$$

Integrating and recalling $d\tan^{-1}(y)/dy = 1/(1+y^2)$ gives the displayed equation. To get the bound note $\int_0^\infty e^{-ay}\,dy = 1/a$ and $\int_0^\infty ye^{-ay}\,dy = 1/a^2$.

A.8. Radon-Nikodym Theorem

8.1. If $\mu(\{A \cap \{x : f(x) < 0\}) = 0$ then for $B \subset A$

$$\int_B f\,d\mu = \int_{B\cap\{x:f(x)>0\}} f\,d\mu \ge 0$$

If $E = A \cap \{x : f(x) < -\epsilon\}$ has positive measure for some $\epsilon > 0$ then

$$\int_E f\,d\mu \le \int_E -\epsilon\,d\mu < 0$$

so A is not positive.

8.2. Let μ be the uniform distribution on the Cantor set, C, defined in Example 1.7 of Chapter 1. $\mu(C^c) = 0$ and $\lambda(C) = 0$ so the two measures are mutually singular.

8.3. If $F \subset E$ then since $(A \cup B)^c$ is a null set.

$$\alpha(F) = \alpha(F\cap A) + \alpha(F\cap B) \le \alpha(E\cap A) = \alpha_+(E)$$

8.4. Suppose $\nu_r^1 + \nu_s^1$ and $\nu_r^2 + \nu_s^2$ are two decompositions. Let A_i be so that $\nu_s^i(A_i) = 0$ and $\mu(A_i^c) = 0$. Clearly $\mu(A_1^c \cup A_2^c) = 0$. The fact that $\nu_r^i \ll \mu$ implies $\nu_r^i(A_1^c \cup A_2^c) = 0$. Combining this with $\nu_s^1(A_1) = 0 = \nu_s^2(A_2)$ it follows that

$$\begin{aligned}\nu_r^1(E) &= \nu_r^1(E \cap A_1 \cap A_2) = \mu(E \cap A_1 \cap A_2) \\ &= \nu_r^2(E \cap A_1 \cap A_2) = \nu_r^2(E)\end{aligned}$$

This shows $\nu_r^1 = \nu_r^2$ and it follows that $\nu_s^1 = \nu_s^2$.

8.5. Since $\mu_2 \perp \nu$, there is an A with $\mu_2(A) = 0$ and $\nu(A^c) = 0$. $\mu_1 \ll \mu_2$ implies $\mu_1(A) = 0$ so $\mu_1 \perp \nu$.

8.6. Let $g_i = d\nu_i/d\mu$. The definition implies $\nu_i(B) = \int_B g_i \, d\mu$ so

$$(\nu_1 + \nu_2)(B) = \int_B (g_1 + g_2) \, d\mu$$

and the desired result follows from uniqueness.

8.7. If $F = 1_A$ this follows from the definition. Linearity gives the result for simple functions; monotone convergence the result for nonnegative functions.

8.8. Let $f = (d\pi/d\nu)1_B$ in Exercise 8.7 to get

$$\int_B \frac{d\pi}{d\nu} \cdot \frac{d\nu}{d\mu} d\mu = \int_B \frac{d\pi}{d\mu} d\nu = \pi(B)$$

where the second equality follows from a second application of Exercise 8.7.

8.9. Letting $\pi = \mu$ in Exercise 8.8 we have

$$1 = \frac{d\mu}{d\nu} \cdot \frac{d\nu}{d\mu}$$