4

Martingales

A martingale X_n can be thought of as the fortune at time n of a player who is betting on a fair game; submartingales (supermartingales) as the outcome of betting on a favorable (unfavorable) game. There are two basic facts about martingales. The first is that you cannot make money betting on them (see Theorem 4.2.8), and in particular if you choose to stop playing at some bounded time N, then your expected winnings EX_N are equal to your initial fortune X_0. (We are supposing for the moment that X_0 is not random.) Our second fact, Theorem 4.2.11, concerns submartingales. To use a heuristic we learned from Mike Brennan, "They are the stochastic analogues of nondecreasing sequences and so if they are bounded above (to be precise, $\sup_n EX_n^+ < \infty$) they converge almost surely." As the material in Section 4.3 shows, this result has diverse applications. Later sections give sufficient conditions for martingales to converge in L^p, $p > 1$ (Section 4.4) and in L^1 (Section 4.6); study the special case of square integrable martingales (Section 4.5); and consider martingales indexed by $n \leq 0$ (Section 4.7). We give sufficient conditions for $EX_N = EX_0$ to hold for unbounded stopping times (Section 4.8). These results are quite useful for studying the behavior of random walks. Section 4.9 complements the random walk results derived from martingale arguments in Section 4.8.1 by giving combinatorial proofs.

4.1 Conditional Expectation

We begin with a definition that is important for this chapter and the next one. After giving the definition, we will consider several examples to explain it. Given are a probability space $(\Omega, \mathcal{F}_o, P)$, a σ-field $\mathcal{F} \subset \mathcal{F}_o$, and a random variable $X \in \mathcal{F}_o$ with $E|X| < \infty$. We define the **conditional expectation of** X **given** $\mathcal{F}$, $E(X|\mathcal{F})$, to be any random variable Y that has

(i) $Y \in \mathcal{F}$, i.e., is $\mathcal{F}$ measurable

(ii) for all $A \in \mathcal{F}$, $\int_A X\,dP = \int_A Y\,dP$

Any Y satisfying (i) and (ii) is said to be a **version of** $E(X|\mathcal{F})$. The first thing to be settled is that the conditional expectation exists and is unique. We tackle the second claim first but start with a technical point.

Lemma 4.1.1 *If Y satisfies (i) and (ii), then it is integrable.*

Proof Letting $A = \{Y > 0\} \in \mathcal{F}$, using (ii) twice, and then adding

$$\int_A Y\, dP = \int_A X\, dP \le \int_A |X|\, dP$$
$$\int_{A^c} -Y\, dP = \int_{A^c} -X\, dP \le \int_{A^c} |X|\, dP$$

So we have $E|Y| \le E|X|$. □

Uniqueness If Y' also satisfies (i) and (ii), then

$$\int_A Y\, dP = \int_A Y'\, dP \quad \text{for all } A \in \mathcal{F}$$

Taking $A = \{Y - Y' \ge \epsilon > 0\}$, we see

$$0 = \int_A X - X\, dP = \int_A Y - Y'\, dP \ge \epsilon P(A)$$

so $P(A) = 0$. Since this holds for all ϵ, we have $Y \le Y'$ a.s., and interchanging the roles of Y and Y', we have $Y = Y'$ a.s. Technically, all equalities such as $Y = E(X|\mathcal{F})$ should be written as $Y = E(X|\mathcal{F})$ a.s., but we have ignored this point in previous chapters and will continue to do so.

Repeating the last argument gives:

Theorem 4.1.2 *If $X_1 = X_2$ on $B \in \mathcal{F}$, then $E(X_1|\mathcal{F}) = E(X_2|\mathcal{F})$ a.s. on B.*

Proof Let $Y_1 = E(X_1|\mathcal{F})$ and $Y_2 = E(X_2|\mathcal{F})$. Taking $A = \{Y_1 - Y_2 \ge \epsilon > 0\}$, we see

$$0 = \int_{A\cap B} X_1 - X_2\, dP = \int_{A\cap B} Y_1 - Y_2\, dP \ge \epsilon P(A)$$

so $P(A) = 0$, and the conclusion follows as before. □

Existence To start, we recall ν is said to be **absolutely continuous with respect to** μ (abbreviated $\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$, and we use Theorem A.4.8:

Radon-Nikodym Theorem Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, there is a function $f \in \mathcal{F}$ so that for all $A \in \mathcal{F}$

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

f is usually denoted $d\nu/d\mu$ and called the **Radon-Nikodym derivative**.

The last theorem easily gives the existence of conditional expectation. Suppose first that $X \ge 0$. Let $\mu = P$ and

$$\nu(A) = \int_A X\, dP \quad \text{for } A \in \mathcal{F}$$

The dominated convergence theorem implies ν is a measure (see Exercise 1.5.4) and the definition of the integral implies $\nu \ll \mu$. The Radon- Nikodym derivative $d\nu/d\mu \in \mathcal{F}$ and for any $A \in \mathcal{F}$ has

$$\int_A X\, dP = \nu(A) = \int_A \frac{d\nu}{d\mu}\, dP$$

Taking $A = \Omega$, we see that $d\nu/d\mu \geq 0$ is integrable, and we have shown that $d\nu/d\mu$ is a version of $E(X|\mathcal{F})$.

To treat the general case now, write $X = X^+ - X^-$, let $Y_1 = E(X^+|\mathcal{F})$ and $Y_2 = E(X^-|\mathcal{F})$. Now $Y_1 - Y_2 \in \mathcal{F}$ is integrable, and for all $A \in \mathcal{F}$ we have

$$\begin{aligned}\int_A X\, dP &= \int_A X^+\, dP - \int_A X^-\, dP \\ &= \int_A Y_1\, dP - \int_A Y_2\, dP = \int_A (Y_1 - Y_2)\, dP\end{aligned}$$

This shows $Y_1 - Y_2$ is a version of $E(X|\mathcal{F})$ and completes the proof. □

4.1.1 Examples

Intuitively, we think of $\mathcal{F}$ as describing the information we have at our disposal - for each $A \in \mathcal{F}$, we know whether or not A has occurred. $E(X|\mathcal{F})$ is then our "best guess" of the value of X given the information we have. Some examples should help to clarify this and connect $E(X|\mathcal{F})$ with other definitions of conditional expectation.

Example 4.1.3 If $X \in \mathcal{F}$, then $E(X|\mathcal{F}) = X$; i.e., if we know X, then our "best guess" is X itself. Since X always satisfies (ii), the only thing that can keep X from being $E(X|\mathcal{F})$ is condition (i). A special case of this example is $X = c$, where c is a constant.

Example 4.1.4 At the other extreme from perfect information is no information. Suppose X is independent of $\mathcal{F}$, i.e., for all $B \in \mathcal{R}$ and $A \in \mathcal{F}$

$$P(\{X \in B\} \cap A) = P(X \in B)P(A)$$

We claim that, in this case, $E(X|\mathcal{F}) = EX$; i.e., if you don't know anything about X, then the best guess is the mean EX. To check the definition, note that $EX \in \mathcal{F}$ so (i) holds. To verify (ii), we observe that if $A \in \mathcal{F}$, then since X and $1_A \in \mathcal{F}$ are independent, Theorem 2.1.13 implies

$$\int_A X\, dP = E(X1_A) = EX\, E1_A = \int_A EX\, dP$$

The reader should note that here and in what follows the game is "guess and verify." We come up with a formula for the conditional expectation and then check that it satisfies (i) and (ii).

Example 4.1.5 In this example, we relate the new definition of conditional expectation to the first one taught in an undergraduate probability course. Suppose $\Omega_1, \Omega_2, \ldots$ is a finite or infinite partition of Ω into disjoint sets, each of which has positive probability, and let $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \ldots)$ be the σ-field generated by these sets. Then

$$E(X|\mathcal{F}) = \frac{E(X; \Omega_i)}{P(\Omega_i)} \quad \text{on } \Omega_i$$

In words, the information in Ω_i tells us which element of the partition our outcome lies in and, given this information, the best guess for X is the average value of X over Ω_i. To prove our guess is correct, observe that the proposed formula is constant on each Ω_i, so it is measurable with respect to $\mathcal{F}$. To verify (ii), it is enough to check the equality for $A = \Omega_i$, but this is trivial:

$$\int_{\Omega_i} \frac{E(X;\Omega_i)}{P(\Omega_i)}\, dP = E(X;\Omega_i) = \int_{\Omega_i} X\, dP$$

A degenerate but important special case is $\mathcal{F} = \{\emptyset, \Omega\}$, the trivial σ-field. In this case, $E(X|\mathcal{F}) = EX$.

To continue the connection with undergraduate notions, let

$$\begin{aligned} P(A|\mathcal{G}) &= E(1_A|\mathcal{G}) \\ P(A|B) &= P(A\cap B)/P(B) \end{aligned}$$

and observe that in the last example $P(A|\mathcal{F}) = P(A|\Omega_i)$ on Ω_i.

The definition of conditional expectation given a σ-field contains conditioning on a random variable as a special case. We define

$$E(X|Y) = E(X|\sigma(Y))$$

where $\sigma(Y)$ is the σ-field generated by Y.

Example 4.1.6 To continue making connection with definitions of conditional expectation from undergraduate probability, suppose X and Y have joint density $f(x,y)$, i.e.,

$$P((X,Y)\in B) = \int_B f(x,y)\, dx\, dy \quad \text{for } B \in \mathcal{R}^2$$

and suppose for simplicity that $\int f(x,y)\, dx > 0$ for all y. We claim that in this case, if $E|g(X)| < \infty$, then $E(g(X)|Y) = h(Y)$, where

$$h(y) = \int g(x) f(x,y)\, dx \Big/ \int f(x,y)\, dx$$

To "guess" this formula, note that treating the probability densities $P(Y=y)$ as if they were real probabilities

$$P(X=x|Y=y) = \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{f(x,y)}{\int f(x,y)\, dx}$$

so, integrating against the conditional probability density, we have

$$E(g(X)|Y=y) = \int g(x) P(X=x|Y=y)\, dx$$

To "verify" the proposed formula now, observe $h(Y) \in \sigma(Y)$, so (i) holds. To check (ii), observe that if $A \in \sigma(Y)$, then $A = \{\omega : Y(\omega) \in B\}$ for some $B \in \mathcal{R}$, so

$$\begin{aligned} E(h(Y);A) &= \int_B \int h(y) f(x,y)\, dx\, dy = \int_B \int g(x) f(x,y)\, dx\, dy \\ &= E(g(X)1_B(Y)) = E(g(X);A) \end{aligned}$$

Remark To drop the assumption that $\int f(x,y)\,dx > 0$, define h by

$$h(y)\int f(x,y)\,dx = \int g(x)f(x,y)\,dx$$

(i.e., h can be anything where $\int f(x,y)\,dx = 0$), and observe this is enough for the proof.

Example 4.1.7 Suppose X and Y are independent. Let φ be a function with $E|\varphi(X,Y)| < \infty$ and let $g(x) = E(\varphi(x,Y))$. We will now show that

$$E(\varphi(X,Y)|X) = g(X)$$

Proof It is clear that $g(X) \in \sigma(X)$. To check (ii), note that if $A \in \sigma(X)$, then $A = \{X \in C\}$, so using the change of variables formula (Theorem 1.6.9) and the fact that the distribution of (X,Y) is product measure (Theorem 2.1.11), then the definition of g, and change of variables again,

$$\begin{aligned}\int_A \varphi(X,Y)\,dP &= E\{\varphi(X,Y)1_C(X)\}\\ &= \int\int \phi(x,y)1_C(x)\,\nu(dy)\,\mu(dx)\\ &= \int 1_C(x)g(x)\,\mu(dx) = \int_A g(X)\,dP\end{aligned}$$

which proves the desired result. □

Example 4.1.8 (Borel's paradox) Let X be a randomly chosen point on the earth, let θ be its longitude, and φ be its latitude. It is customary to take $\theta \in [0,2\pi)$ and $\varphi \in (-\pi/2,\pi/2]$ but we can equally well take $\theta \in [0,\pi)$ and $\varphi \in (-\pi,\pi]$. In words, the new longitude specifies the great circle on which the point lies and then φ gives the angle.

At first glance it might seem that if X is uniform on the globe, then θ and the angle φ on the great circle should both be uniform over their possible values. θ is uniform but φ is not. The paradox completely evaporates once we realize that in the new or in the traditional formulation φ is independent of θ, so the conditional distribution is the unconditional one, which is not uniform since there is more land near the equator than near the North Pole.

4.1.2 Properties

Conditional expectation has many of the same properties that ordinary expectation does.

Theorem 4.1.9 *In the first two parts we assume $E|X|, E|Y| < \infty$.*
(a) Conditional expectation is linear:

$$E(aX + Y|\mathcal{F}) = aE(X|\mathcal{F}) + E(Y|\mathcal{F}) \tag{4.1.1}$$

(b) If $X \le Y$, then

$$E(X|\mathcal{F}) \le E(Y|\mathcal{F}). \tag{4.1.2}$$

(c) If $X_n \ge 0$ and $X_n \uparrow X$ with $EX < \infty$, then

$$E(X_n|\mathcal{F}) \uparrow E(X|\mathcal{F}) \tag{4.1.3}$$

Remark By applying the last result to $Y_1 - Y_n$, we see that if $Y_n \downarrow Y$ and we have $E|Y_1|, E|Y| < \infty$, then $E(Y_n|\mathcal{F}) \downarrow E(Y|\mathcal{F})$.

Proof To prove (a), we need to check that the right-hand side is a version of the left. It clearly is $\mathcal{F}$-measurable. To check (ii), we observe that if $A \in \mathcal{F}$, then by linearity of the integral and the defining properties of $E(X|\mathcal{F})$ and $E(Y|\mathcal{F})$,

$$\begin{aligned}\int_A \{aE(X|\mathcal{F}) + E(Y|\mathcal{F})\}\,dP &= a\int_A E(X|\mathcal{F})\,dP + \int_A E(Y|\mathcal{F})\,dP \\ &= a\int_A X\,dP + \int_A Y\,dP = \int_A aX + Y\,dP\end{aligned}$$

which proves (4.1.1).

Using the definition

$$\int_A E(X|\mathcal{F})\,dP = \int_A X\,dP \le \int_A Y\,dP = \int_A E(Y|\mathcal{F})\,dP$$

Letting $A = \{E(X|\mathcal{F}) - E(Y|\mathcal{F}) \ge \epsilon > 0\}$, we see that the indicated set has probability 0 for all $\epsilon > 0$, and we have proved (4.1.2).

Let $Y_n = X - X_n$. It suffices to show that $E(Y_n|\mathcal{F}) \downarrow 0$. Since $Y_n \downarrow$, (4.1.2) implies $Z_n \equiv E(Y_n|\mathcal{F}) \downarrow$ a limit Z_∞. If $A \in \mathcal{F}$, then

$$\int_A Z_n\,dP = \int_A Y_n\,dP$$

Letting $n \to \infty$, noting $Y_n \downarrow 0$, and using the dominated convergence theorem gives that $\int_A Z_\infty\,dP = 0$ for all $A \in \mathcal{F}$, so $Z_\infty \equiv 0$. □

Theorem 4.1.10 *If φ is convex and $E|X|$, $E|\varphi(X)| < \infty$, then*

$$\varphi(E(X|\mathcal{F})) \le E(\varphi(X)|\mathcal{F}) \tag{4.1.4}$$

Proof If φ is linear, the result is trivial, so we will suppose φ is not linear. We do this so that if we let $S = \{(a,b) : a,b \in \mathbf{Q},\ ax + b \le \varphi(x) \text{ for all } x\}$, then $\varphi(x) = \sup\{ax + b : (a,b) \in S\}$. See the proof of Theorem 1.6.2 for more details. If $\varphi(x) \ge ax + b$, then (4.1.2) and (4.1.1) imply

$$E(\varphi(X)|\mathcal{F}) \ge a\,E(X|\mathcal{F}) + b \quad \text{a.s.}$$

Taking the sup over $(a,b) \in S$ gives

$$E(\varphi(X)|\mathcal{F}) \ge \varphi(E(X|\mathcal{F})) \quad \text{a.s.}$$

which proves the desired result. □

Remark Here we have written a.s. by the inequalities to stress that there is an exceptional set for each a, b, so we have to take the sup over a countable set.

Theorem 4.1.11 *Conditional expectation is a contraction in L^p, $p \ge 1$.*

Proof (4.1.4) implies $|E(X|\mathcal{F})|^p \le E(|X|^p|\mathcal{F})$. Taking expected values gives

$$E(|E(X|\mathcal{F})|^p) \le E(E(|X|^p|\mathcal{F})) = E|X|^p$$ □

In the last equality, we have used an identity that is an immediate consequence of the definition (use property (ii) in the definition with $A = \Omega$).

$$E(E(Y|\mathcal{F})) = E(Y) \tag{4.1.5}$$

Conditional expectation also has properties, like (4.1.5), that have no analogue for "ordinary" expectation.

Theorem 4.1.12 *If $\mathcal{F} \subset \mathcal{G}$ and $E(X|\mathcal{G}) \in \mathcal{F}$, then $E(X|\mathcal{F}) = E(X|\mathcal{G})$.*

Proof By assumption $E(X|\mathcal{G}) \in \mathcal{F}$. To check the other part of the definition we note that if $A \in \mathcal{F} \subset \mathcal{G}$, then

$$\int_A X\,dP = \int_A E(X|\mathcal{G})\,dP \qquad \square$$

Theorem 4.1.13 *If $\mathcal{F}_1 \subset \mathcal{F}_2$, then (i) $E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(X|\mathcal{F}_1)$*
(ii) $E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$.

In words, the smaller σ-field always wins. As the proof will show, the first equality is trivial. The second is easy to prove, but in combination with Theorem 4.1.14 is a powerful tool for computing conditional expectations. I have seen it used several times to prove results that are false.

Proof Once we notice that $E(X|\mathcal{F}_1) \in \mathcal{F}_2$, (i) follows from Example 4.1.3. To prove (ii), notice that $E(X|\mathcal{F}_1) \in \mathcal{F}_1$, and if $A \in \mathcal{F}_1 \subset \mathcal{F}_2$, then

$$\int_A E(X|\mathcal{F}_1)\,dP = \int_A X\,dP = \int_A E(X|\mathcal{F}_2)\,dP \qquad \square$$

The next result shows that for conditional expectation with respect to $\mathcal{F}$, random variables $X \in \mathcal{F}$ are like constants. They can be brought outside the "integral."

Theorem 4.1.14 *If $X \in \mathcal{F}$ and $E|Y|$, $E|XY| < \infty$, then*

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F}).$$

Proof The right-hand side $\in \mathcal{F}$, so we have to check (ii). To do this, we use the usual four-step procedure. First, suppose $X = 1_B$ with $B \in \mathcal{F}$. In this case, if $A \in \mathcal{F}$

$$\int_A 1_B E(Y|\mathcal{F})\,dP = \int_{A\cap B} E(Y|\mathcal{F})\,dP = \int_{A\cap B} Y\,dP = \int_A 1_B Y\,dP$$

so (ii) holds. The last result extends to simple X by linearity. If $X, Y \geq 0$, let X_n be simple random variables that $\uparrow X$, and use the monotone convergence theorem to conclude that

$$\int_A XE(Y|\mathcal{F})\,dP = \int_A XY\,dP$$

To prove the result in general, split X and Y into their positive and negative parts. $\square$

Theorem 4.1.15 *Suppose $EX^2 < \infty$. $E(X|\mathcal{F})$ is the variable $Y \in \mathcal{F}$ that minimizes the "mean square error" $E(X - Y)^2$.*

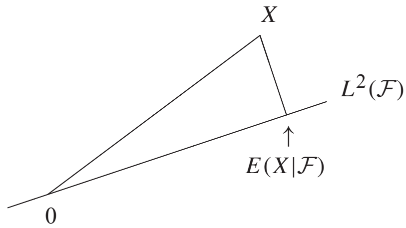

Figure 4.1 Conditional expectation as projection in L^2.

Remark This result gives a "geometric interpretation" of $E(X|\mathcal{F})$. $L^2(\mathcal{F}_o) = \{Y \in \mathcal{F}_o : EY^2 < \infty\}$ is a Hilbert space, and $L^2(\mathcal{F})$ is a closed subspace. In this case, $E(X|\mathcal{F})$ is the projection of X onto $L^2(\mathcal{F})$. That is, the point in the subspace closest to X.

Proof We begin by observing that if $Z \in L^2(\mathcal{F})$, then Theorem 4.1.14 implies

$$ZE(X|\mathcal{F}) = E(ZX|\mathcal{F})$$

($E|XZ| < \infty$ by the Cauchy-Schwarz inequality.) Taking expected values gives

$$E(ZE(X|\mathcal{F})) = E(E(ZX|\mathcal{F})) = E(ZX)$$

or, rearranging,

$$E[Z(X - E(X|\mathcal{F}))] = 0 \quad \text{for } Z \in L^2(\mathcal{F})$$

If $Y \in L^2(\mathcal{F})$ and $Z = E(X|\mathcal{F}) - Y$, then

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

since the cross-product term vanishes. From the last formula, it is easy to see $E(X - Y)^2$ is minimized when $Z = 0$. □

4.1.3 Regular Conditional Probabilities*

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ a measurable map, and $\mathcal{G}$ a σ-field $\subset \mathcal{F}$. $\mu : \Omega \times \mathcal{S} \to [0, 1]$ is said to be a **regular conditional distribution** for X given $\mathcal{G}$ if

(i) For each A, $\omega \to \mu(\omega, A)$ is a version of $P(X \in A|\mathcal{G})$.

(ii) For a.e. ω, $A \to \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$.

When $S = \Omega$ and X is the identity map, μ is called a **regular conditional probability**.

Continuation of Example 4.1.6 Suppose X and Y have a joint density $f(x, y) > 0$. If

$$\mu(y, A) = \int_A f(x, y)\, dx \Big/ \int f(x, y)\, dx$$

then $\mu(Y(\omega), A)$ is a r.c.d. for X given $\sigma(Y)$.

(i) in the definition follows by taking $h = 1_A$ in Example 4.1.1. To check (ii) note that the dominated convergence theorem implies that $A \to \mu(y, A)$ is a probability measure.

Regular conditional distributions are useful because they allow us to simultaneously compute the conditional expectation of all functions of X and to generalize properties of ordinary expectation in a more straightforward way.

Theorem 4.1.16 *Let $\mu(\omega, A)$ be a r.c.d. for X given $\mathcal{F}$. If $f : (S, \mathcal{S}) \to (\mathbf{R}, \mathcal{R})$ has $E|f(X)| < \infty$, then*

$$E(f(X)|\mathcal{F}) = \int \mu(\omega, dx) f(x) \quad a.s.$$

Proof 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 $f = f^+ - f^-$. □

Unfortunately, r.c.d.'s do not always exist. The first example was due to Dieudonné (1948). See Doob (1953), p. 624, or Faden (1985) for more recent developments. Without going into the details of the example, it is easy to see the source of the problem. If $A_1, A_2, \dots$ are disjoint, then (4.1.1) and (4.1.3) imply

$$P(X \in \cup_n A_n | \mathcal{G}) = \sum_n P(X \in A_n | \mathcal{G}) \quad \text{a.s.}$$

but if $\mathcal{S}$ contains enough countable collections of disjoint sets, the exceptional sets may pile up. Fortunately,

Theorem 4.1.17 *r.c.d.'s exist if $(S, \mathcal{S})$ is nice.*

Proof By definition, there is a 1-1 map $\varphi : S \to \mathbf{R}$ so that φ and φ^{-1} are measurable. Using monotonicity (4.1.2) and throwing away a countable collection of null sets, 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 $P(\varphi(X) \le q|\mathcal{G})$. Let $F(x, \omega) = \inf\{G(q, \omega) : q > x\}$. The notation may remind the reader of the proof of Theorem 3.2.12. The argument given there shows F is a distribution function. Since $G(q_n, \omega) \downarrow F(x, \omega)$, the remark after Theorem 4.1.9 implies that $F(x, \omega)$ is a version of $P(\varphi(X) \le x|\mathcal{G})$.

Now, for each $\omega \in \Omega_o$, there is a unique measure $\nu(\omega, \cdot)$ on $(\mathbf{R}, \mathcal{R})$ so that $\nu(\omega, (-\infty, x]) = F(x, \omega)$. To check that for each $B \in \mathcal{R}$, $\nu(\omega, B)$ is a version of $P(\varphi(X) \in B|\mathcal{G})$, we observe that the class of B for which this statement is true (this includes the measurability of $\omega \to \nu(\omega, B)$) is a λ-system that contains all sets of the form $(a_1, b_1] \cup \cdots (a_k, b_k]$, where $-\infty \le a_i < b_i \le \infty$, 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 $\mu(\omega, A) = \nu(\omega, B)$. □

The following generalization of Theorem 4.1.17 will be needed in Section 6.1.

Theorem 4.1.18 *Suppose X and Y take values in a nice space $(S, \mathcal{S})$ and $\mathcal{G} = \sigma(Y)$. There is a function $\mu : S \times \mathcal{S} \to [0, 1]$ so that*

(i) for each A, $\mu(Y(\omega), A)$ is a version of $P(X \in A|\mathcal{G})$

(ii) for a.e. ω, $A \to \mu(Y(\omega), A)$ is a probability measure on $(S, \mathcal{S})$.

Proof As in the proof of Theorem 4.1.17, 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 $P(\varphi(X) \le q|\mathcal{G})$. 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 Theorem 4.1.17 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 $P(\varphi(X) \le x|Y)$.

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 $-\infty \le a_i < b_i \le \infty$, 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 $\mu(y,A) = \nu(y,B)$. □

Exercises

4.1.1 **Bayes' formula.** Let $G \in \mathcal{G}$ and show that

$$P(G|A) = \int_G P(A|\mathcal{G})\,dP \Big/ \int_\Omega P(A|\mathcal{G})\,dP$$

When $\mathcal{G}$ is the σ-field generated by a partition, this reduces to the usual Bayes' formula

$$P(G_i|A) = P(A|G_i)P(G_i) \Big/ \sum_j P(A|G_j)P(G_j)$$

4.1.2 Prove **Chebyshev's inequality.** If $a > 0$, then

$$P(|X| \ge a|\mathcal{F}) \le a^{-2}E(X^2|\mathcal{F})$$

4.1.3 Imitate the proof in the remark after Theorem 1.5.2 to prove the conditional Cauchy-Schwarz inequality.

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

4.1.4 Use regular conditional probability to get the conditional Hölder inequality from the unconditional one, i.e., show that if $p,q \in (1,\infty)$ with $1/p + 1/q = 1$, then

$$E(|XY||\mathcal{G}) \le E(|X|^p|\mathcal{G})^{1/p}E(|Y|^q|\mathcal{G})^{1/q}$$

4.1.5 Give an example on $\Omega = \{a,b,c\}$ in which

$$E(E(X|\mathcal{F}_1)|\mathcal{F}_2) \ne E(E(X|\mathcal{F}_2)|\mathcal{F}_1)$$

4.1.6 Show that if $\mathcal{G} \subset \mathcal{F}$ and $EX^2 < \infty$, then

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

Dropping the second term on the left, we get an inequality that says geometrically, the larger the subspace the closer the projection is, or statistically, more information means a smaller mean square error.

4.1.7 An important special case of the previous result occurs when $\mathcal{G} = \{\emptyset, \Omega\}$. Let $\text{var}(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$. Show that

$$\text{var}(X) = E(\text{var}(X|\mathcal{F})) + \text{var}(E(X|\mathcal{F}))$$

4.1.8 Let $Y_1, Y_2, \ldots$ be i.i.d. with mean μ and variance σ^2, N an independent positive integer valued r.v. with $EN^2 < \infty$ and $X = Y_1 + \cdots + Y_N$. Show that $\text{var}(X) = \sigma^2 EN + \mu^2 \text{var}(N)$. To understand and help remember the formula, think about the two special cases in which N or Y is constant.

4.1.9 Show that if X and Y are random variables with $E(Y|\mathcal{G}) = X$ and $EY^2 = EX^2 < \infty$, then $X = Y$ a.s.

4.1.10 The result in the last exercise implies that if $EY^2 < \infty$ and $E(Y|\mathcal{G})$ has the same distribution as Y, then $E(Y|\mathcal{G}) = Y$ a.s. Prove this under the assumption $E|Y| < \infty$. Hint: The trick is to prove that $\text{sgn}(X) = \text{sgn}(E(X|\mathcal{G}))$ a.s., and then take $X = Y - c$ to get the desired result.

4.2 Martingales, Almost Sure Convergence

In this section we will define martingales and their cousins supermartingales and submartingales, and take the first steps in developing their theory. Let $\mathcal{F}_n$ be a **filtration**, i.e., an increasing sequence of σ-fields. A sequence X_n is said to be **adapted** to $\mathcal{F}_n$ if $X_n \in \mathcal{F}_n$ for all n. If X_n is sequence with

(i) $E|X_n| < \infty$,

(ii) X_n is adapted to $\mathcal{F}_n$,

(iii) $E(X_{n+1}|\mathcal{F}_n) = X_n$ for all n,

then X is said to be a **martingale** (with respect to $\mathcal{F}_n$). If in the last definition $=$ is replaced by $\le$ or $\ge$, then X is said to be a **supermartingale** or **submartingale**, respectively.

We begin by describing three examples related to random walk. Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed. Let $S_n = S_0 + \xi_1 + \cdots + \xi_n$, where S_0 is a constant. Let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \ge 1$ and take $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Example 4.2.1 (Linear martingale) If $\mu = E\xi_i = 0$, then $S_n, n \ge 0$, is a martingale with respect to $\mathcal{F}_n$.

To prove this, we observe that $S_n \in \mathcal{F}_n$, $E|S_n| < \infty$, and ξ_{n+1} is independent of $\mathcal{F}_n$, so using the linearity of conditional expectation, (4.1.1), and Example 4.1.4,

$$E(S_{n+1}|\mathcal{F}_n) = E(S_n|\mathcal{F}_n) + E(\xi_{n+1}|\mathcal{F}_n) = S_n + E\xi_{n+1} = S_n$$

If $\mu \le 0$, then the computation just completed shows $E(X_{n+1}|\mathcal{F}_n) \le X_n$, i.e., X_n is a supermartingale. In this case, X_n corresponds to betting on an unfavorable game, so there is nothing "super" about a supermartingale. The name comes from the fact that if f is superharmonic (i.e., f has continuous derivatives of order ≤ 2 and $\partial^2 f/\partial x_1^2 + \cdots + \partial^2 f/\partial x_d^2 \le 0$), then

$$f(x) \geq \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy \tag{4.2.1}$$

where $B(x,r) = \{y : |x-y| \leq r\}$ is the ball of radius r, and $|B(x,r)|$ is the volume of the ball.

If $\mu \geq 0$, then S_n is a submartingale. Applying the first result to $\xi_i' = \xi_i - \mu$, we see that $S_n - n\mu$ is a martingale.

Example 4.2.2 (Quadratic martingale) Suppose now that $\mu = E\xi_i = 0$ and $\sigma^2 = \text{var}(\xi_i) < \infty$. In this case, $S_n^2 - n\sigma^2$ is a martingale.

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

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

Example 4.2.3 (Exponential martingale) Let $Y_1, Y_2, \ldots$ be nonnegative i.i.d. random variables with $EY_m = 1$. If $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, then $M_n = \prod_{m \leq n} Y_m$ defines a martingale. To prove this note that

$$E(M_{n+1}|\mathcal{F}_n) = M_n E(X_{n+1}|\mathcal{F}_n) = Y_n$$

Suppose now that $Y_i = e^{\theta\xi_i}$ and $\phi(\theta) = Ee^{\theta\xi_i} < \infty$. $Y_i = \exp(\theta\xi)/\phi(\theta)$ has mean 1 so $EY_i = 1$ and

$$M_n = \prod_{i=1}^{n} Y_i = \exp(\theta S_n)/\phi(\theta)^n \quad \text{is a martingale.}$$

We will see many other examples below, so we turn now to deriving properties of martingales. Our first result is an immediate consequence of the definition of a supermartingale. We could take the conclusion of the result as the definition of supermartingale, but then the definition would be harder to check.

Theorem 4.2.4 *If X_n is a supermartingale, then for $n > m$, $E(X_n|\mathcal{F}_m) \leq X_m$.*

Proof The definition gives the result for $n = m + 1$. Suppose $n = m + k$ with $k \geq 2$. By Theorem 4.1.2,

$$E(X_{m+k}|\mathcal{F}_m) = E(E(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m) \leq E(X_{m+k-1}|\mathcal{F}_m)$$

by the definition and (4.1.2). The desired result now follows by induction. □

Theorem 4.2.5 *(i) If X_n is a submartingale, then for $n > m$, $E(X_n|\mathcal{F}_m) \geq X_m$.*
(ii) If X_n is a martingale, then for $n > m$, $E(X_n|\mathcal{F}_m) = X_m$.

Proof To prove (i), note that $-X_n$ is a supermartingale and use (4.1.1). For (ii), observe that X_n is a supermartingale and a submartingale. □

Remark The idea in the proof of Theorem 4.2.5 will be used many times in what follows. To keep from repeating ourselves, we will just state the result for either supermartingales or submartingales and leave it to the reader to translate the result for the other two.

Theorem 4.2.6 *If X_n is a martingale w.r.t. $\mathcal{F}_n$ and φ is a convex function with $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale w.r.t. $\mathcal{F}_n$. Consequently, if $p \geq 1$ and $E|X_n|^p < \infty$ for all n, then $|X_n|^p$ is a submartingale w.r.t. $\mathcal{F}_n$.*

Proof By Jensen's inequality and the definition

$$E(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(E(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n) \qquad \square$$

Theorem 4.2.7 *If X_n is a submartingale w.r.t. $\mathcal{F}_n$ and φ is an increasing convex function with $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale w.r.t. $\mathcal{F}_n$. Consequently, (i) If X_n is a submartingale, then $(X_n - a)^+$ is a submartingale. (ii) If X_n is a supermartingale, then $X_n \wedge a$ is a supermartingale.*

Proof By Jensen's inequality and the assumptions

$$E(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(E(X_{n+1}|\mathcal{F}_n)) \geq \varphi(X_n) \qquad \square$$

Let $\mathcal{F}_n$, $n \geq 0$ be a filtration. H_n, $n \geq 1$ is said to be a **predictable sequence** if $H_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$. In words, the value of H_n may be predicted (with certainty) from the information available at time $n-1$. In this section, we will be thinking of H_n as the amount of money a gambler will bet at time n. This can be based on the outcomes at times $1, \dots, n-1$ but not on the outcome at time n!

Once we start thinking of H_n as a gambling system, it is natural to ask how much money we would make if we used it. Let X_n be the net amount of money you would have won at time n if you had bet one dollar each time. If you bet according to a gambling system H, then your winnings at time n would be

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$$

since if at time m you have wagered \$3, the change in your fortune would be 3 times that of a person who wagered \$1. Alternatively, you can think of X_m as the value of a stock and H_m the number of shares you hold from time $m-1$ to time m.

Suppose now that $\xi_m = X_m - X_{m-1}$ have $P(\xi_m = 1) = p$ and $P(\xi_m = -1) = 1 - p$. A famous gambling system called the "martingale" is defined by $H_1 = 1$ and for $n \geq 2$,

$$H_n = \begin{cases} 2H_{n-1} & \text{if } \xi_{n-1} = -1 \\ 1 & \text{if } \xi_{n-1} = 1 \end{cases}$$

In words, we double our bet when we lose, so that if we lose k times and then win, our net winnings will be 1. To see this, consider the following concrete situation

H_n	1	2	4	8	16
ξ_n	−1	−1	−1	−1	1
$(H \cdot X)_n$	−1	−3	−7	−15	1

This system seems to provide us with a "sure thing" as long as $P(\xi_m = 1) > 0$. However, the next result says there is no system for beating an unfavorable game.

Theorem 4.2.8 *Let X_n, $n \geq 0$, be a supermartingale. If $H_n \geq 0$ is predictable and each H_n is bounded, then $(H \cdot X)_n$ is a supermartingale.*

Proof Using the fact that conditional expectation is linear, $(H \cdot X)_n \in \mathcal{F}_n$, $H_n \in \mathcal{F}_{n-1}$, and (4.1.14), we have

$$\begin{aligned} E((H \cdot X)_{n+1}|\mathcal{F}_n) &= (H \cdot X)_n + E(H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n) \\ &= (H \cdot X)_n + H_{n+1}E((X_{n+1} - X_n)|\mathcal{F}_n) \leq (H \cdot X)_n \end{aligned}$$

since $E((X_{n+1} - X_n)|\mathcal{F}_n) \leq 0$ and $H_{n+1} \geq 0$. □

Remark The same result is obviously true for submartingales and for martingales (in the last case, without the restriction $H_n \geq 0$).

We will now consider a very special gambling system: bet \$1 each time $n \leq N$ then stop playing. A random variable N is said to be a **stopping time** if $\{N = n\} \in \mathcal{F}_n$ for all $n < \infty$, i.e., the decision to stop at time n must be measurable with respect to the information known at that time. If we let $H_n = 1_{\{N \geq n\}}$, then $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$, so H_n is predictable, and it follows from Theorem 4.2.8 that $(H \cdot X)_n = X_{N \wedge n} - X_0$ is a supermartingale. Since the constant sequence $Y_n = X_0$ is a supermartingale and the sum of two supermartingales is also, we have:

Theorem 4.2.9 *If N is a stopping time and X_n is a supermartingale, then $X_{N \wedge n}$ is a supermartingale.*

Although Theorem 4.2.8 implies that you cannot make money with gambling systems, you can prove theorems with them. Suppose X_n, $n \geq 0$, is a submartingale. Let $a < b$, let $N_0 = -1$, and for $k \geq 1$ let

$$\begin{aligned} N_{2k-1} &= \inf\{m > N_{2k-2} : X_m \leq a\} \\ N_{2k} &= \inf\{m > N_{2k-1} : X_m \geq b\} \end{aligned}$$

The N_j are stopping times and $\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^c \in \mathcal{F}_{m-1}$, so

$$H_m = \begin{cases} 1 & \text{if } N_{2k-1} < m \leq N_{2k} \text{ for some } k \\ 0 & \text{otherwise} \end{cases}$$

defines a predictable sequence. $X(N_{2k-1}) \leq a$ and $X(N_{2k}) \geq b$, so between times N_{2k-1} and N_{2k}, X_m crosses from below a to above b. H_m is a gambling system that tries to take advantage of these "upcrossings." In stock market terms, we buy when $X_m \leq a$ and sell when $X_m \geq b$, so every time an upcrossing is completed, we make a profit of $\geq (b-a)$. Finally, $U_n = \sup\{k : N_{2k} \leq n\}$ is the number of upcrossings completed by time n.

Theorem 4.2.10 (Upcrossing inequality) *If X_m, $m \geq 0$, is a submartingale, then*

$$(b-a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+$$

Proof Let $Y_m = a + (X_m - a)^+$. By Theorem 4.2.7, Y_m is a submartingale. Clearly, it upcrosses $[a, b]$ the same number of times that X_m does, and we have $(b-a)U_n \leq (H \cdot Y)_n$, since each upcrossing results in a profit $\geq (b-a)$ and a final incomplete upcrossing (if there

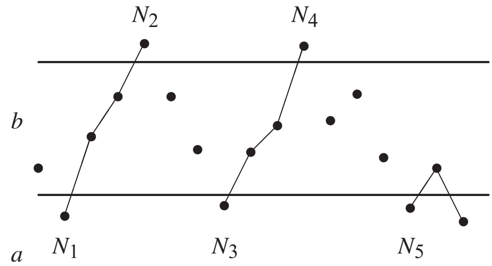

Figure 4.2 Upcrossings of (a, b). Lines indicate increments that are included in $(H \cdot X)_n$. In Y_n the points $< a$ are moved up to a.

is one) makes a nonnegative contribution to the right-hand side. It is for this reason we had to replace X_m by Y_m.

Let $K_m = 1 - H_m$. Clearly, $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$, and it follows from Theorem 4.2.8 that $E(K \cdot Y)_n \geq E(K \cdot Y)_0 = 0$, so $E(H \cdot Y)_n \leq E(Y_n - Y_0)$, proving the desired inequality. □

We have proved the result in its classical form, even though this is a little misleading. The key fact is that $E(K \cdot Y)_n \geq 0$, i.e., no matter how hard you try you can't lose money betting on a submartingale. From the upcrossing inequality, we easily get:

Theorem 4.2.11 (Martingale convergence theorem) *If X_n is a submartingale with $\sup EX_n^+ < \infty$, then as $n \to \infty$, X_n converges a.s. to a limit X with $E|X| < \infty$.*

Proof Since $(X - a)^+ \leq X^+ + |a|$, Theorem 4.2.10 implies that

$$EU_n \leq (|a| + EX_n^+)/(b - a)$$

As $n \uparrow \infty$, $U_n \uparrow U$ the number of upcrossings of $[a, b]$ by the whole sequence, so if $\sup EX_n^+ < \infty$, then $EU < \infty$ and hence $U < \infty$ a.s. Since the last conclusion holds for all rational a and b,

$$\cup_{a,b \in \mathbf{Q}} \{\liminf X_n < a < b < \limsup X_n\} \quad \text{has probability } 0$$

and hence $\limsup X_n = \liminf X_n$ a.s., i.e., $\lim X_n$ exists a.s. Fatou's lemma guarantees $EX^+ \leq \liminf EX_n^+ < \infty$, so $X < \infty$ a.s. To see $X > -\infty$, we observe that

$$EX_n^- = EX_n^+ - EX_n \leq EX_n^+ - EX_0$$

(since X_n is a submartingale), so another application of Fatou's lemma shows

$$EX^- \leq \liminf_{n \to \infty} EX_n^- \leq \sup_n EX_n^+ - EX_0 < \infty$$

and completes the proof. □

Remark To prepare for the proof of Theorem 4.7.1, the reader should note that we have shown that if the number of upcrossings of (a, b) by X_n is finite for all $a, b \in \mathbf{Q}$, then the limit of X_n exists.

An important special case of Theorem 4.2.11 is:

Theorem 4.2.12 *If $X_n \geq 0$ is a supermartingale, then as $n \to \infty$, $X_n \to X$ a.s. and $EX \leq EX_0$.*

Proof $Y_n = -X_n \leq 0$ is a submartingale with $EY_n^+ = 0$. Since $EX_0 \geq EX_n$, the inequality follows from Fatou's lemma. □

In the next section we will give several applications of the last two results. We close this one by giving two "counterexamples."

Example 4.2.13 The first shows that the assumptions of Theorem 4.2.12 (or 4.2.11) do not guarantee convergence in L^1. Let S_n be a symmetric simple random walk with $S_0 = 1$, i.e., $S_n = S_{n-1} + \xi_n$, where $\xi_1, \xi_2, \ldots$ are i.i.d. with $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. Let $N = \inf\{n : S_n = 0\}$ and let $X_n = S_{N \wedge n}$. Theorem 4.2.9 implies that X_n is a nonnegative martingale. Theorem 4.2.12 implies X_n converges to a limit $X_\infty < \infty$ that must be $\equiv 0$, since convergence to $k > 0$ is impossible. (If $X_n = k > 0$, then $X_{n+1} = k \pm 1$.) Since $EX_n = EX_0 = 1$ for all n and $X_\infty = 0$, convergence cannot occur in L^1.

Example 4.2.13 is an important counterexample to keep in mind as you read the rest of this chapter. The next one is not as important.

Example 4.2.14 We will now give an example of a martingale with $X_k \to 0$ in probability but not a.s. Let $X_0 = 0$. When $X_{k-1} = 0$, let $X_k = 1$ or -1 with probability $1/2k$ and $= 0$ with probability $1 - 1/k$. When $X_{k-1} \neq 0$, let $X_k = kX_{k-1}$ with probability $1/k$ and $= 0$ with probability $1 - 1/k$. From the construction, $P(X_k = 0) = 1 - 1/k$, so $X_k \to 0$ in probability. On the other hand, the second Borel-Cantelli lemma implies $P(X_k = 0$ for $k \geq K) = 0$, and values in $(-1, 1) - \{0\}$ are impossible, so X_k does not converge to 0 a.s.

Exercises

4.2.1 Suppose X_n is a martingale w.r.t. $\mathcal{G}_n$ and let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{G}_n \supset \mathcal{F}_n$ and X_n is a martingale w.r.t. $\mathcal{F}_n$.

4.2.2 Give an example of a submartingale X_n so that X_n^2 is a supermartingale. Hint: X_n does not have to be random.

4.2.3 Generalize (i) of Theorem 4.2.7 by showing that if X_n and Y_n are submartingales w.r.t. $\mathcal{F}_n$, then $X_n \vee Y_n$ is also.

4.2.4 Let X_n, $n \geq 0$, be a submartingale with $\sup X_n < \infty$. Let $\xi_n = X_n - X_{n-1}$ and suppose $E(\sup \xi_n^+) < \infty$. Show that X_n converges a.s.

4.2.5 Give an example of a martingale X_n with $X_n \to -\infty$ a.s. Hint: Let $X_n = \xi_1 + \cdots + \xi_n$, where the ξ_i are independent (but not identically distributed) with $E\xi_i = 0$.

4.2.6 Let $Y_1, Y_2, \ldots$ be nonnegative i.i.d. random variables with $EY_m = 1$ and $P(Y_m = 1) < 1$. By example 4.2.3 that $X_n = \prod_{m \leq n} Y_m$ defines a martingale. (i) Use Theorem 4.2.12 and an argument by contradiction to show $X_n \to 0$ a.s. (ii) Use the strong law of large numbers to conclude $(1/n) \log X_n \to c < 0$.

4.2.7 Suppose $y_n > -1$ for all n and $\sum |y_n| < \infty$. Show that $\prod_{m=1}^{\infty}(1 + y_m)$ exists.

4.2.8 Let X_n and Y_n be positive integrable and adapted to $\mathcal{F}_n$. Suppose

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

with $\sum Y_n < \infty$ a.s. Prove that X_n converges a.s. to a finite limit by finding a closely related supermartingale to which Theorem 4.2.12 can be applied.

4.2.9 **The switching principle.** Suppose X_n^1 and X_n^2 are supermartingales with respect to $\mathcal{F}_n$, and N is a stopping time so that $X_N^1 \ge X_N^2$. Then

$$Y_n = X_n^1 1_{(N>n)} + X_n^2 1_{(N \le n)} \text{ is a supermartingale.}$$
$$Z_n = X_n^1 1_{(N \ge n)} + X_n^2 1_{(N < n)} \text{ is a supermartingale.}$$

4.2.10 **Dubins' inequality.** For every positive supermartingale X_n, $n \ge 0$, the number of upcrossings U of $[a,b]$ satisfies

$$P(U \ge k) \le \left(\frac{a}{b}\right)^k E \min(X_0/a, 1)$$

To prove this, we let $N_0 = -1$ and for $j \ge 1$ let

$$N_{2j-1} = \inf\{m > N_{2j-2} : X_m \le a\}$$
$$N_{2j} = \inf\{m > N_{2j-1} : X_m \ge b\}$$

Let $Y_n = 1$ for $0 \le n < N_1$ and for $j \ge 1$

$$Y_n = \begin{cases} (b/a)^{j-1}(X_n/a) & \text{for } N_{2j-1} \le n < N_{2j} \\ (b/a)^j & \text{for } N_{2j} \le n < N_{2j+1} \end{cases}$$

(i) Use the switching principle in the previous exercise and induction to show that $Z_n^j = Y_{n \wedge N_j}$ is a supermartingale. (ii) Use $EY_{n \wedge N_{2k}} \le EY_0$ and let $n \to \infty$ to get Dubins' inequality.

4.3 Examples

In this section, we will apply the martingale convergence theorem to generalize the second Borel-Cantelli lemma and to study Polya's urn scheme, Radon-Nikodym derivatives, and branching processes. The four topics are independent of each other and are taken up in the order indicated.

4.3.1 Bounded Increments

Our first result shows that martingales with bounded increments either converge or oscillate between $+\infty$ and $-\infty$.

Theorem 4.3.1 *Let $X_1, X_2, \ldots$ be a martingale with $|X_{n+1} - X_n| \le M < \infty$. Let*

$$C = \{\lim X_n \textit{ exists and is finite}\}$$
$$D = \{\limsup X_n = +\infty \textit{ and } \liminf X_n = -\infty\}$$

Then $P(C \cup D) = 1$.

Proof Since $X_n - X_0$ is a martingale, we can without loss of generality suppose that $X_0 = 0$. Let $0 < K < \infty$ and let $N = \inf\{n : X_n \le -K\}$. $X_{n\wedge N}$ is a martingale with $X_{n\wedge N} \ge -K - M$ a.s., so applying Theorem 4.2.12 to $X_{n\wedge N} + K + M$ shows $\lim X_n$ exists on $\{N = \infty\}$. Letting $K \to \infty$, we see that the limit exists on $\{\liminf X_n > -\infty\}$. Applying the last conclusion to $-X_n$, we see that $\lim X_n$ exists on $\{\limsup X_n < \infty\}$ and the proof is complete. □

To prepare for an application of this result we need:

Theorem 4.3.2 (Doob's decomposition) *Any submartingale X_n, $n \ge 0$, can be written in a unique way as $X_n = M_n + A_n$, where M_n is a martingale and A_n is a predictable increasing sequence with $A_0 = 0$.*

Proof We want $X_n = M_n + A_n$, $E(M_n|\mathcal{F}_{n-1}) = M_{n-1}$, and $A_n \in \mathcal{F}_{n-1}$. So we must have

$$\begin{aligned} E(X_n|\mathcal{F}_{n-1}) &= E(M_n|\mathcal{F}_{n-1}) + E(A_n|\mathcal{F}_{n-1}) \\ &= M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n \end{aligned}$$

and it follows that

$$A_n - A_{n-1} = E(X_n|\mathcal{F}_{n-1}) - X_{n-1} \tag{4.3.1}$$

Since $A_0 = 0$, we have

$$A_n = \sum_{m=1}^{n} E(X_n - X_{n-1}|\mathcal{F}_{n-1}) \tag{4.3.2}$$

To check that our recipe works, we observe that $A_n - A_{n-1} \ge 0$ since X_n is a submartingale and $A_n \in \mathcal{F}_{n-1}$. To prove that $M_n = X_n - A_n$ is a martingale, we note that using $A_n \in \mathcal{F}_{n-1}$ and (4.3.1)

$$\begin{aligned} E(M_n|\mathcal{F}_{n-1}) &= E(X_n - A_n|\mathcal{F}_{n-1}) \\ &= E(X_n|\mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1} \end{aligned}$$

which completes the proof. □

To illustrate the use of this result, we do the following important example.

Example 4.3.3 Let us suppose $B_n \in \mathcal{F}_n$. Using (4.3.2)

$$M_n = \sum_{m=1}^{n} 1_{B_m} - E(1_{B_m}|\mathcal{F}_{m-1})$$

Theorem 4.3.4 (Second Borel-Cantelli lemma, II) *Let $\mathcal{F}_n$, $n \ge 0$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and A_n, $n \ge 1$ a sequence of events with $B_n \in \mathcal{F}_n$. Then*

$$\{B_n\ i.o.\} = \left\{ \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty \right\}$$

Proof If we let $X_0 = 0$ and $X_n = \sum_{m\le n} 1_{B_m}$, then X_n is a submartingale. (4.3.2) implies $B_n = \sum_{m=1}^{n} E(1_{A_m}|\mathcal{F}_{m-1})$, so if $M_0 = 0$ and

$$M_n = \sum_{m=1}^{n} 1_{A_m} - P(A_m|\mathcal{F}_{m-1})$$

for $n \geq 1$ then M_n is a martingale with $|M_n - M_{n-1}| \leq 1$. Using the notation of Theorem 4.3.1 we have:

$$\text{on } C, \quad \sum_{n=1}^{\infty} 1_{B_n} = \infty \quad \text{if and only if} \quad \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty$$

$$\text{on } D, \quad \sum_{n=1}^{\infty} 1_{B_n} = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty$$

Since $P(C \cup D) = 1$, the result follows. □

4.3.2 Polya's Urn Scheme

An urn contains r red and g green balls. At each time we draw a ball out, then replace it, and add c more balls of the color drawn. Let X_n be the fraction of green balls after the nth draw. To check that X_n is a martingale, note that if there are i red balls and j green balls at time n, then

$$X_{n+1} = \begin{cases} (j+c)/(i+j+c) & \text{with probability } j/(i+j) \\ j/(i+j+c) & \text{with probability } i/(i+j) \end{cases}$$

and we have

$$\frac{j+c}{i+j+c} \cdot \frac{j}{i+j} + \frac{j}{i+j+c} \cdot \frac{i}{i+j} = \frac{(j+c+i)j}{(i+j+c)(i+j)} = \frac{j}{i+j}$$

Since $X_n \geq 0$, Theorem 4.2.12 implies that $X_n \to X_\infty$ a.s. To compute the distribution of the limit, we observe (a) the probability of getting green on the first m draws then red on the next $\ell = n - m$ draws is

$$\frac{g}{g+r} \cdot \frac{g+c}{g+r+c} \cdots \frac{g+(m-1)c}{g+r+(m-1)c} \cdot \frac{r}{g+r+mc} \cdots \frac{r+(\ell-1)c}{g+r+(n-1)c}$$

and (b) any other outcome of the first n draws with m green balls drawn and ℓ red balls drawn has the same probability, since the denominator remains the same and the numerator is permuted. Consider the special case $c = 1$, $g = 1$, $r = 1$. Let G_n be the number of green balls after the nth draw has been completed and the new ball has been added. It follows from (a) and (b) that

$$P(G_n = m+1) = \binom{n}{m} \frac{m!\,(n-m)!}{(n+1)!} = \frac{1}{n+1}$$

so X_∞ has a uniform distribution on (0,1).

If we suppose that $c = 1$, $g = 2$, and $r = 1$, then

$$P(G_n = m+2) = \frac{n!}{m!\,(n-m)!} \frac{(m+1)!\,(n-m)!}{(n+2)!\,/2} \to 2x$$

if $n \to \infty$ and $m/n \to x$. In general, the distribution of X_∞ has density

$$\frac{\Gamma((g+r)/c)}{\Gamma(g/c)\Gamma(r/c)} x^{(g/c)-1}(1-x)^{(r/c)-1}$$

This is the **beta distribution** with parameters g/c and r/c. In Example 4.5.6 we will see that the limit behavior changes drastically if, in addition to the c balls of the color chosen, we always add one ball of the opposite color.

4.3.3 Radon-Nikodym Derivatives

Let μ be a finite measure and ν a probability measure on $(\Omega, \mathcal{F})$. Let $\mathcal{F}_n \uparrow \mathcal{F}$ be σ-fields (i.e., $\sigma(\cup \mathcal{F}_n) = \mathcal{F}$). Let μ_n and ν_n be the restrictions of μ and ν to $\mathcal{F}_n$.

Theorem 4.3.5 *Suppose $\mu_n \ll \nu_n$ for all n. Let $X_n = d\mu_n/d\nu_n$ and let $X = \limsup X_n$. Then*

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

Remark $\mu_r(A) \equiv \int_A X\, d\nu$ is a measure $\ll \nu$. Since Theorem 4.2.12 implies $\nu(X = \infty) = 0$, $\mu_s(A) \equiv \mu(A \cap \{X = \infty\})$ is singular w.r.t. ν. Thus $\mu = \mu_r + \mu_s$ gives the Lebesgue decomposition of μ (see Theorem A.4.7), and $X_\infty = d\mu_r/d\nu$, ν-a.s. Here and in the proof we need to keep track of the measure to which the a.s. refers.

Proof As the reader can probably anticipate:

Lemma 4.3.6 *X_n (defined on $(\Omega, \mathcal{F}, \nu)$) is a martingale w.r.t. $\mathcal{F}_n$.*

Proof We observe that, by definition, $X_n \in \mathcal{F}_n$. Let $A \in \mathcal{F}_n$. Since $X_n \in \mathcal{F}_n$ and ν_n is the restriction of ν to $\mathcal{F}_n$

$$\int_A X_n\, d\nu = \int_A X_n\, d\nu_n$$

Using the definition of X_n and Exercise A.4.7

$$\int_A X_n\, d\nu_n = \mu_n(A) = \mu(A)$$

the last equality holding, since $A \in \mathcal{F}_n$ and μ_n is the restriction of μ to $\mathcal{F}_n$. If $A \in \mathcal{F}_{m-1} \subset \mathcal{F}_m$, using the last result for $n = m$ and $n = m - 1$ gives

$$\int_A X_m d\nu = \mu(A) = \int_A X_{m-1} d\nu$$

so $E(X_m|\mathcal{F}_{m-1}) = X_{m-1}$. □

Since X_n is a nonnegative martingale, Theorem 4.2.12 implies that $X_n \to X$ ν-a.s. We want to check that the equality in the theorem holds. Dividing $\mu(A)$ by $\mu(\Omega)$, we can without loss of generality suppose μ is a probability measure. Let $\rho = (\mu + \nu)/2$, $\rho_n = (\mu_n + \nu_n)/2 =$ the restriction of ρ to $\mathcal{F}_n$. Let $Y_n = d\mu_n/d\rho_n$, $Z_n = d\nu_n/d\rho_n$. $Y_n, Z_n \geq 0$ and $Y_n + Z_n = 2$ (by Exercise A.4.6), so Y_n and Z_n are bounded martingales with limits Y and Z. As the reader can probably guess,

$$(*) \qquad Y = d\mu/d\rho \qquad Z = d\nu/d\rho$$

It suffices to prove the first equality. From the proof of Lemma 4.3.6, if $A \in \mathcal{F}_m \subset \mathcal{F}_n$

$$\mu(A) = \int_A Y_n \, d\rho \to \int_A Y \, d\rho$$

by the bounded convergence theorem. The last computation shows that

$$\mu(A) = \int_A Y \, d\rho \quad \text{for all } A \in \mathcal{G} = \cup_m \mathcal{F}_m$$

$\mathcal{G}$ is a π-system, so the $\pi - \lambda$ theorem implies the equality is valid for all $A \in \mathcal{F} = \sigma(\mathcal{G})$ and $(*)$ is proved.

It follows from Exercises A.4.8 and A.4.9 that $X_n = Y_n/Z_n$. At this point, the reader can probably leap to the conclusion that $X = Y/Z$. To get there carefully, note $Y + Z = 2$ ρ-a.s., so $\rho(Y = 0, Z = 0) = 0$. Having ruled out $0/0$, we have $X = Y/Z$ ρ-a.s. (Recall $X \equiv \limsup X_n$.) Let $W = (1/Z) \cdot 1_{(Z>0)}$. Using $(*)$, then $1 = ZW + 1_{(Z=0)}$, we have

(a) $$\mu(A) = \int_A Y \, d\rho = \int_A YWZ \, d\rho + \int_A 1_{(Z=0)} Y \, d\rho$$

Now $(*)$ implies $d\nu = Z \, d\rho$, and it follows from the definitions that

$$YW = X 1_{(Z>0)} = X \quad \nu\text{-a.s.}$$

the second equality holding, since $\nu(\{Z = 0\}) = 0$. Combining things, we have

(b) $$\int_A YWZ \, d\rho = \int_A X \, d\nu$$

To handle the other term, we note that $(*)$ implies $d\mu = Y \, d\rho$, and it follows from the definitions that $\{X = \infty\} = \{Z = 0\}$ μ-a.s. so

(c) $$\int_A 1_{(Z=0)} Y \, d\rho = \int_A 1_{(X=\infty)} \, d\mu$$

Combining (a), (b), and (c) gives the desired result. □

Example 4.3.7 Suppose $\mathcal{F}_n = \sigma(I_{k,n} : 0 \le k < K_n)$, where for each n, $I_{k,n}$ is a partition of Ω, and the $(n+1)$th partition is a refinement of the nth. In this case, the condition $\mu_n \ll \nu_n$ is $\nu(I_{k,n}) = 0$ implies $\mu(I_{k,n}) = 0$, and the martingale $X_n = \mu(I_{k,n})/\nu(I_{k,n})$ on $I_{k,n}$ is an approximation to the Radon-Nikodym derivative. For a concrete example, consider $\Omega = [0, 1)$, $I_{k,n} = [k2^{-n}, (k+1)2^{-n})$ for $0 \le k < 2^n$, and $\nu =$ Lebesgue measure.

Kakutani dichotomy for infinite product measures. Let μ and ν be measures on sequence space $(\mathbf{R}^{\mathbf{N}}, \mathcal{R}^{\mathbf{N}})$ that make the coordinates $\xi_n(\omega) = \omega_n$ independent. Let $F_n(x) = \mu(\xi_n \le x)$, $G_n(x) = \nu(\xi_n \le x)$. Suppose $F_n \ll G_n$ and let $q_n = dF_n/dG_n$. To avoid a problem, we will suppose $q_n > 0$, G_n-a.s.

Let $\mathcal{F}_n = \sigma(\xi_m : m \le n)$, let μ_n and ν_n be the restrictions of μ and ν to $\mathcal{F}_n$, and let

$$X_n = \frac{d\mu_n}{d\nu_n} = \prod_{m=1}^{n} q_m.$$

Theorem 4.3.5 implies that $X_n \to X$ ν-a.s. Thanks to our assumption $q_n > 0$, G_n-a.s. $\sum_{m=1}^{\infty} \log(q_m) > -\infty$is a tail event, so the Kolmogorov 0-1 law implies

$$\nu(X = 0) \in \{0, 1\} \tag{4.3.3}$$

and it follows from Theorem 4.3.5 that either $\mu \ll \nu$ or $\mu \perp \nu$. The next result gives a concrete criterion for which of the two alternatives occurs.

Theorem 4.3.8 *$\mu \ll \nu$ or $\mu \perp \nu$, according as $\prod_{m=1}^{\infty} \int \sqrt{q_m}\, dG_m > 0$ or $= 0$.*

Proof Jensen's inequality and Exercise A.4.7 imply

$$\left(\int \sqrt{q_m}\, dG_m\right)^2 \le \int q_m\, dG_m = \int dF_m = 1$$

so the infinite product of the integrals is well defined and ≤ 1. Let

$$X_n = \prod_{m \le n} q_m(\omega_m)$$

as above, and recall that $X_n \to X$ ν-a.s. If the infinite product is 0, then

$$\int X_n^{1/2}\, d\nu = \prod_{m=1}^{n} \int \sqrt{q_m}\, dG_m \to 0$$

Fatou's lemma implies

$$\int X^{1/2}\, d\nu \le \liminf_{n \to \infty} \int X_n^{1/2}\, d\nu = 0$$

so $X = 0$ ν-a.s., and Theorem 4.3.5 implies $\mu \perp \nu$. To prove the other direction, let $Y_n = X_n^{1/2}$. Now $\int q_m\, dG_m = 1$, so if we use E to denote expected value with respect to ν, then $EY_m^2 = EX_m = 1$, so

$$E(Y_{n+k} - Y_n)^2 = E(X_{n+k} + X_n - 2X_n^{1/2} X_{n+k}^{1/2}) = 2\left(1 - \prod_{m=n+1}^{n+k} \int \sqrt{q_m}\, dG_m\right)$$

Now $|a-b| = |a^{1/2} - b^{1/2}| \cdot (a^{1/2} + b^{1/2})$, so using Cauchy-Schwarz and the fact $(a+b)^2 \le 2a^2 + 2b^2$ gives

$$\begin{aligned}
E|X_{n+k} - X_n| &= E(|Y_{n+k} - Y_n|(Y_{n+k} + Y_n)) \\
&\le \left(E(Y_{n+k} - Y_n)^2 E(Y_{n+k} + Y_n)^2\right)^{1/2} \\
&\le \left(4E(Y_{n+k} - Y_n)^2\right)^{1/2}
\end{aligned}$$

From the last two equations, it follows that if the infinite product is > 0, then X_n converges to X in $L^1(\nu)$, so $\nu(X = 0) < 1$, (4.3.3) implies the probability is 0, and the desired result follows from Theorem 4.3.5. □

4.3.4 Branching Processes

Let ξ_i^n, $i, n \geq 1$, be i.i.d. nonnegative integer-valued random variables. Define a sequence $Z_n, n \geq 0$ by $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} & \text{if } Z_n > 0 \\ 0 & \text{if } Z_n = 0 \end{cases} \tag{4.3.4}$$

Z_n is called a **Galton-Watson process**. The idea behind the definitions is that Z_n is the number of individuals in the nth generation, and each member of the nth generation gives birth independently to an identically distributed number of children. $p_k = P(\xi_i^n = k)$ is called the **offspring distribution**:

Lemma 4.3.9 *Let $\mathcal{F}_n = \sigma(\xi_i^m : i \geq 1, 1 \leq m \leq n)$ and $\mu = E\xi_i^m \in (0, \infty)$. Then Z_n/μ^n is a martingale w.r.t. $\mathcal{F}_n$.*

Proof Clearly, $Z_n \in \mathcal{F}_n$. Using Theorem 4.1.2 to conclude that on $\{Z_n = k\}$

$$E(Z_{n+1}|\mathcal{F}_n) = E(\xi_1^{n+1} + \cdots + \xi_k^{n+1}|\mathcal{F}_n) = k\mu = \mu Z_n$$

where in the second equality we used the fact that the ξ_k^{n+1} are independent of $\mathcal{F}_n$. □

Z_n/μ^n is a nonnegative martingale, so Theorem 4.2.12 implies $Z_n/\mu^n \to$ a limit a.s. We begin by identifying cases when the limit is trivial.

Theorem 4.3.10 *If $\mu < 1$, then $Z_n = 0$ for all n sufficiently large, so $Z_n/\mu^n \to 0$.*

Proof $E(Z_n/\mu^n) = E(Z_0) = 1$, so $E(Z_n) = \mu^n$. Now $Z_n \geq 1$ on $\{Z_n > 0\}$ so

$$P(Z_n > 0) \leq E(Z_n; Z_n > 0) = E(Z_n) = \mu^n \to 0$$

exponentially fast if $\mu < 1$. □

The last answer should be intuitive. If each individual on the average gives birth to less than one child, the species will die out. The next result shows that after we exclude the trivial case in which each individual has exactly one child, the same result holds when $\mu = 1$.

Theorem 4.3.11 *If $\mu = 1$ and $P(\xi_i^m = 1) < 1$, then $Z_n = 0$ for all n sufficiently large.*

Proof When $\mu = 1$, Z_n is itself a nonnegative martingale. Since Z_n is integer valued and by Theorem 4.2.12 converges to an a.s. finite limit Z_∞, we must have $Z_n = Z_\infty$ for large n. If $P(\xi_i^m = 1) < 1$ and $k > 0$, then $P(Z_n = k \text{ for all } n \geq N) = 0$ for any N, so we must have $Z_\infty \equiv 0$. □

When $\mu \leq 1$, the limit of Z_n/μ^n is 0 because the branching process dies out. Our next step is to show that if $\mu > 1$, then $P(Z_n > 0 \text{ for all } n) > 0$. For $s \in [0, 1]$, let $\varphi(s) = \sum_{k \geq 0} p_k s^k$, where $p_k = P(\xi_i^m = k)$. φ is the **generating function** for the offspring distribution p_k.

Theorem 4.3.12 *Suppose $\mu > 1$. If $Z_0 = 1$, then $P(Z_n = 0 \text{ for some } n) = \rho$ the only solution of $\varphi(\rho) = \rho$ in $[0, 1)$.*

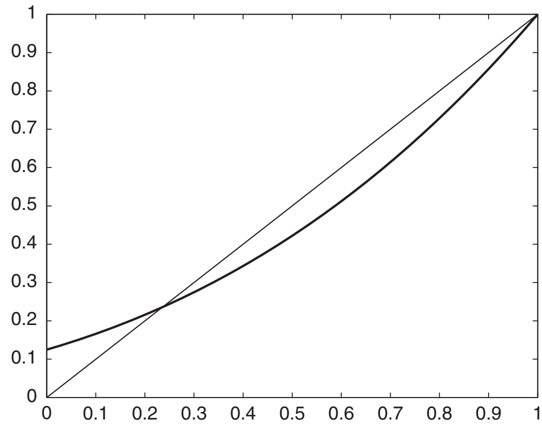

Figure 4.3 Generating function for Binomial(3,1/2).

Proof $\phi(1) = 1$ Differentiating and referring to Theorem A.5.3 for the justification gives for $s < 1$

$$\varphi'(s) = \sum_{k=1}^{\infty} k\ p_k s^{k-1} \geq 0,$$

so ϕ is increasing. We may have $\phi(s) = \infty$ when $s > 1$, so we have to work carefully.

$$\lim_{s \uparrow 1} \varphi'(s) = \sum_{k=1}^{\infty} k p_k = \mu$$

Integrating, we have

$$\phi(1) - \phi(1-h) = \int_{1-h}^{1} \phi'(s)\, ds \sim \mu h$$

as $h \to 0$, so if h is small $\phi(1-h) < 1-h$. $\phi(0) \geq 0$, so there must be a solution of $\phi(x) = x$ in $[0,1)$.

To prove uniqueness we note that for $s < 1$

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} > 0$$

since $\mu > 1$ implies that $p_k > 0$ for some $k \geq 2$. Let ρ be the smallest solution of $\phi(\rho) = \rho$ in $[0,1)$. Since $\phi(1) = 1$ and ϕ is strictly convex, we have $\phi(x) < x$ for $x \in (\rho, 1)$, so there is only one solution of $\phi(\rho) = \rho$ in $[0,1)$.

Combining the next two results will complete the proof.

(a) If $\theta_m = P(Z_m = 0)$, then $\theta_m = \sum_{k=0}^{\infty} p_k(\theta_{m-1})^k = \phi(\theta_{m-1})$

Proof of (a). If $Z_1 = k$, an event with probability p_k, then $Z_m = 0$ if and only if all k families die out in the remaining $m-1$ units of time, an independent event with probability θ_{m-1}^k. Summing over the disjoint possibilities for each k gives the desired result.

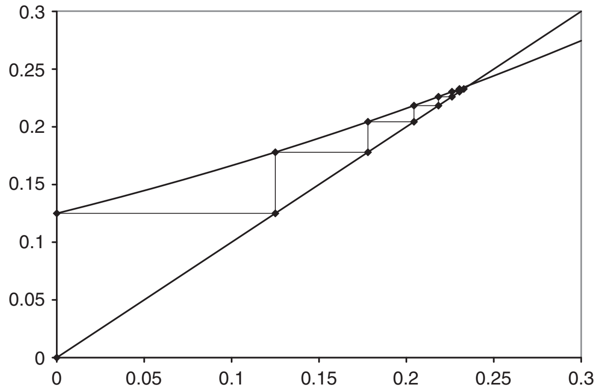

Figure 4.4 Iteration as in part (c) for the Binomial(3,1/2) generating function.

(b) As $m \uparrow \infty$, $\theta_m \uparrow \rho$.

Proof of (b). Clearly $\theta_m = P(Z_m = 0)$ is increasing. To show by induction that $\theta_m \le \rho$, we note that $\theta_0 = 0 \le \rho$, and if the result is true for $m - 1$

$$\theta_m = \varphi(\theta_{m-1}) \le \varphi(\rho) = \rho.$$

Taking limits in $\theta_m = \varphi(\theta_{m-1})$, we see $\theta_\infty = \varphi(\theta_\infty)$. Since $\theta_\infty \le \rho$, it follows that $\theta_\infty = \rho$. □

The last result shows that when $\mu > 1$, the limit of Z_n/μ^n has a chance of being nonzero. The best result on this question is due to Kesten and Stigum:

Theorem 4.3.13 *$W = \lim Z_n/\mu^n$ is not $\equiv 0$ if and only if $\sum p_k k \log k < \infty$.*

For a proof, see Athreya and Ney (1972), pp. 24–29. In the next section, we will show that $\sum k^2 p_k < \infty$ is sufficient for a nontrivial limit.

Exercises

4.3.1 Give an example of a martingale X_n with $\sup_n |X_n| < \infty$ and $P(X_n = a \text{ i.o.}) = 1$ for $a = -1, 0, 1$. This example shows that it is not enough to have $\sup |X_{n+1} - X_n| < \infty$ in Theorem 4.3.1.

4.3.2 (Assumes familiarity with finite state Markov chains.) Fine tune the example for the previous problem so that $P(X_n = 0) \to 1-2p$ and $P(X_n = -1), P(X_n = 1) \to p$, where p is your favorite number in $(0, 1/2)$, i.e., you are asked to do this for one value of p that you may choose. This example shows that a martingale can converge in distribution without converging a.s. (or in probability).

4.3.3 Let X_n and Y_n be positive integrable and adapted to $\mathcal{F}_n$. Suppose $E(X_{n+1}|\mathcal{F}_n) \le X_n + Y_n$, with $\sum Y_n < \infty$ a.s. Prove that X_n converges a.s. to a finite limit. Hint: Let $N = \inf_k \sum_{m=1}^k Y_m > M$, and stop your supermartingale at time N.

4.3.4 Let $p_m \in [0, 1)$. Use the Borel-Cantelli lemmas to show that

$$\prod_{m=1}^{\infty}(1 - p_m) = 0 \quad \text{if and only if} \quad \sum_{m=1}^{\infty} p_m = \infty.$$

4.3.5 Show $\sum_{n=2}^{\infty} P(A_n | \cap_{m=1}^{n-1} A_m^c) = \infty$ implies $P(\cap_{m=1}^{\infty} A_m^c) = 0$.

4.3.6 Check by direct computation that the X_n in Example 4.3.7 is a martingale. Show that if we drop the condition $\mu_n \ll \nu_n$ and set $X_n = 0$ when $\nu(I_{k,n}) = 0$, then $E(X_{n+1}|\mathcal{F}_n) \le X_n$.

4.3.7 Apply Theorem 4.3.5 to Example 4.3.7 to get a "probabilistic" proof of the Radon-Nikodym theorem. To be precise, suppose $\mathcal{F}$ is **countably generated** (i.e., there is a sequence of sets A_n so that $\mathcal{F} = \sigma(A_n : n \ge 1)$) and show that if μ and ν are σ-finite measures and $\mu \ll \nu$, then there is a function g so that $\mu(A) = \int_A g\, d\nu$. Before you object to this as circular reasoning (the Radon-Nikodym theorem was used to define conditional expectation!), observe that the conditional expectations that are needed for Example 4.3.7 have elementary definitions.

Bernoulli product measures For the next three exercises, suppose F_n, G_n are concentrated on $\{0, 1\}$ and have $F_n(0) = 1 - \alpha_n$, $G_n(0) = 1 - \beta_n$.

4.3.8 (i) Use Theorem 4.3.8 to find a necessary and sufficient condition for $\mu \ll \nu$. (ii) Suppose that $0 < \epsilon \le \alpha_n, \beta_n \le 1 - \epsilon < 1$. Show that in this case the condition is simply $\sum(\alpha_n - \beta_n)^2 < \infty$.

4.3.9 Show that if $\sum \alpha_n < \infty$ and $\sum \beta_n = \infty$ in the previous exercise, then $\mu \perp \nu$. This shows that the condition $\sum(\alpha_n - \beta_n)^2 < \infty$ is not sufficient for $\mu \ll \nu$ in general.

4.3.10 Suppose $0 < \alpha_n, \beta_n < 1$. Show that $\sum |\alpha_n - \beta_n| < \infty$ is sufficient for $\mu \ll \nu$ in general.

4.3.11 Show that if $P(\lim Z_n/\mu^n = 0) < 1$, then it is $= \rho$ and hence

$$\{\lim Z_n/\mu^n > 0\} = \{Z_n > 0 \text{ for all } n\} \quad \text{a.s.}$$

4.3.12 Let Z_n be a branching process with offspring distribution p_k, defined in part d of Section 4.3, and let $\varphi(\theta) = \sum p_k \theta^k$. Suppose $\rho < 1$ has $\varphi(\rho) = \rho$. Show that ρ^{Z_n} is a martingale and use this to conclude $P(Z_n = 0 \text{ for some } n \ge 1 | Z_0 = x) = \rho^x$.

4.3.13 Galton and Watson, who invented the process that bears their names, were interested in the survival of family names. Suppose each family has exactly three children but coin flips determine their sex. In the 1800s, only male children kept the family name, so following the male offspring leads to a branching process with $p_0 = 1/8$, $p_1 = 3/8$, $p_2 = 3/8$, $p_3 = 1/8$. Compute the probability ρ that the family name will die out when $Z_0 = 1$.

4.4 Doob's Inequality, Convergence in L^p, $p > 1$

We begin by proving a consequence of Theorem 4.2.9.

Theorem 4.4.1 *If X_n is a submartingale and N is a stopping time with $P(N \le k) = 1$, then*

$$EX_0 \le EX_N \le EX_k$$

Remark Let S_n be a simple random walk with $S_0 = 1$ and let $N = \inf\{n : S_n = 0\}$. (See Example 4.2.13 for more details.) $ES_0 = 1 > 0 = ES_N$, so the first inequality need not hold for unbounded stopping times. In Section 5.7 we will give conditions that guarantee $EX_0 \le EX_N$ for unbounded N.

Proof Theorem 4.2.9 implies $X_{N \wedge n}$ is a submartingale, so it follows that

$$EX_0 = EX_{N \wedge 0} \le EX_{N \wedge k} = EX_N$$

To prove the other inequality, let $K_n = 1_{\{N<n\}} = 1_{\{N \le n-1\}}$. K_n is predictable, so Theorem 4.2.8 implies $(K \cdot X)_n = X_n - X_{N \wedge n}$ is a submartingale and it follows that

$$EX_k - EX_N = E(K \cdot X)_k \ge E(K \cdot X)_0 = 0 \qquad \square$$

We will see later that Theorem 4.4.1 is very useful. The first indication of this is:

Theorem 4.4.2 (Doob's inequality) *Let X_m be a submartingale,*

$$\bar{X}_n = \max_{0 \le m \le n} X_m^+$$

$\lambda > 0$, and $A = \{\bar{X}_n \ge \lambda\}$. Then

$$\lambda P(A) \le EX_n 1_A \le EX_n^+$$

Proof Let $N = \inf\{m : X_m \ge \lambda \text{ or } m = n\}$. Since $X_N \ge \lambda$ on A,

$$\lambda P(A) \le EX_N 1_A \le EX_n 1_A$$

The second inequality follows from the fact that Theorem 4.4.1 implies $EX_N \le EX_n$ and we have $X_N = X_n$ on A^c. The second inequality is trivial, so the proof is complete. $\square$

Example 4.4.3 (Random walks) If we let $S_n = \xi_1 + \cdots + \xi_n$, where the ξ_m are independent and have $E\xi_m = 0$, $\sigma_m^2 = E\xi_m^2 < \infty$. S_n is a martingale, so Theorem 4.2.6 implies $X_n = S_n^2$ is a submartingale. If we let $\lambda = x^2$ and apply Theorem 4.4.2 to X_n, we get Kolmogorov's maximal inequality, Theorem 2.5.5:

$$P\left(\max_{1 \le m \le n} |S_m| \ge x\right) \le x^{-2} \operatorname{var}(S_n)$$

Integrating the inequality in Theorem 4.4.2 gives:

Theorem 4.4.4 (L^p maximum inequality) *If X_n is a submartingale, then for $1 < p < \infty$,*

$$E(\bar{X}_n^p) \le \left(\frac{p}{p-1}\right)^p E(X_n^+)^p$$

Consequently, if Y_n is a martingale and $Y_n^* = \max_{0 \le m \le n} |Y_m|$,

$$E|Y_n^*|^p \le \left(\frac{p}{p-1}\right)^p E(|Y_n|^p)$$

Proof The second inequality follows by applying the first to $X_n = |Y_n|$. To prove the first we will, for reasons that will become clear in a moment, work with $\bar{X}_n \wedge M$ rather than $\bar{X}_n$. Since $\{\bar{X}_n \wedge M \geq \lambda\}$ is always $\{\bar{X}_n \geq \lambda\}$ or $\emptyset$, this does not change the application of Theorem 4.4.2. Using Lemma 2.2.13, Theorem 4.4.2, Fubini's theorem, and a little calculus gives

$$
\begin{aligned}
E((\bar{X}_n \wedge M)^p) &= \int_0^\infty p\lambda^{p-1} P(\bar{X}_n \wedge M \geq \lambda)\, d\lambda \\
&\leq \int_0^\infty p\lambda^{p-1} \left(\lambda^{-1} \int X_n^+ 1_{(\bar{X}_n \wedge M \geq \lambda)}\, dP \right) d\lambda \\
&= \int X_n^+ \int_0^{\bar{X}_n \wedge M} p\lambda^{p-2}\, d\lambda\, dP \\
&= \frac{p}{p-1} \int X_n^+ (\bar{X}_n \wedge M)^{p-1}\, dP
\end{aligned}
$$

If we let $q = p/(p-1)$ be the exponent conjugate to p and apply Hölder's inequality, Theorem 1.6.3, we see that

$$
\leq \left(\frac{p}{1-p} \right) (E|X_n^+|^p)^{1/p} (E|\bar{X}_n \wedge M|^p)^{1/q}
$$

If we divide both sides of the last inequality by $(E|\bar{X}_n \wedge M|^p)^{1/q}$, which is finite thanks to the $\wedge M$, then take the pth power of each side, we get

$$
E(|\bar{X}_n \wedge M|^p) \leq \left(\frac{p}{p-1} \right)^p E(X_n^+)^p
$$

Letting $M \to \infty$ and using the monotone convergence theorem gives the desired result. □

Example 4.4.5 (There is no L^1 maximal inequality) Again, the counterexample is provided by Example 4.2.13. Let S_n be a simple random walk starting from $S_0 = 1$, $N = \inf\{n : S_n = 0\}$, and $X_n = S_{N \wedge n}$. Theorem 4.4.1 implies $EX_n = ES_{N \wedge n} = ES_0 = 1$ for all n. Using hitting probabilities for simple random walk from Theorem 4.8.7, we have

$$
P\left(\max_m X_m \geq M \right) = \frac{1}{M} \tag{4.4.1}
$$

so $E(\max_m X_m) = \sum_{M=1}^\infty P(\max_m X_m \geq M) = \sum_{M=1}^\infty 1/M = \infty$. The monotone convergence theorem implies that $E \max_{m \leq n} X_m \uparrow \infty$ as $n \uparrow \infty$.

From Theorem 4.4.4, we get the following:

Theorem 4.4.6 (L^p convergence theorem) *If X_n is a martingale with* $\sup E|X_n|^p < \infty$ *where $p > 1$, then $X_n \to X$ a.s. and in L^p.*

Proof $(EX_n^+)^p \leq (E|X_n|)^p \leq E|X_n|^p$, so it follows from the martingale convergence theorem (4.2.11) that $X_n \to X$ a.s. The second conclusion in Theorem 4.4.4 implies

$$
E\left(\sup_{0 \leq m \leq n} |X_m| \right)^p \leq \left(\frac{p}{p-1} \right)^p E|X_n|^p
$$

Letting $n \to \infty$ and using the monotone convergence theorem implies $\sup |X_n| \in L^p$. Since $|X_n - X|^p \le (2 \sup |X_n|)^p$, it follows from the dominated convergence theorem that $E|X_n - X|^p \to 0$. □

The most important special case of the results in this section occurs when $p = 2$. To treat this case, the next two results are useful.

Theorem 4.4.7 (Orthogonality of martingale increments) *Let X_n be a martingale with $EX_n^2 < \infty$ for all n. If $m \le n$ and $Y \in \mathcal{F}_m$ has $EY^2 < \infty$, then*

$$E((X_n - X_m)Y) = 0$$

and hence if $\ell < m < n$

$$E((X_n - X_m)(X_m - X_\ell)) = 0$$

Proof The Cauchy-Schwarz inequality implies $E|(X_n - X_m)Y| < \infty$. Using (4.1.5), Theorem 4.1.14, and the definition of a martingale,

$$E((X_n - X_m)Y) = E[E((X_n - X_m)Y|\mathcal{F}_m)] = E[Y E((X_n - X_m)|\mathcal{F}_m)] = 0 \qquad \square$$

Theorem 4.4.8 (Conditional variance formula) *If X_n is a martingale with $EX_n^2 < \infty$ for all n,*

$$E((X_n - X_m)^2|\mathcal{F}_m) = E(X_n^2|\mathcal{F}_m) - X_m^2.$$

Remark This is the conditional analogue of $E(X - EX)^2 = EX^2 - (EX)^2$ and is proved in exactly the same way.

Proof Using the linearity of conditional expectation and then Theorem 4.1.14, we have

$$\begin{aligned} E(X_n^2 - 2X_nX_m + X_m^2|\mathcal{F}_m) &= E(X_n^2|\mathcal{F}_m) - 2X_m E(X_n|\mathcal{F}_m) + X_m^2 \\ &= E(X_n^2|\mathcal{F}_m) - 2X_m^2 + X_m^2 \end{aligned}$$

which gives the desired result. □

Example 4.4.9 (Branching processes) We continue the study begun at the end of the last section. Using the notation introduced there, we suppose $\mu = E(\xi_i^m) > 1$ and $\mathrm{var}\,(\xi_i^m) = \sigma^2 < \infty$. Let $X_n = Z_n/\mu^n$. Taking $m = n-1$ in Theorem 4.4.8 and rearranging, we have

$$E(X_n^2|\mathcal{F}_{n-1}) = X_{n-1}^2 + E((X_n - X_{n-1})^2|\mathcal{F}_{n-1})$$

To compute the second term, we observe

$$\begin{aligned} E((X_n - X_{n-1})^2|\mathcal{F}_{n-1}) &= E((Z_n/\mu^n - Z_{n-1}/\mu^{n-1})^2|\mathcal{F}_{n-1}) \\ &= \mu^{-2n} E((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}) \end{aligned}$$

It follows from Exercise 4.1.2 that on $\{Z_{n-1} = k\}$,

$$E((Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}) = E\left(\left(\sum_{i=1}^k \xi_i^n - \mu k\right)^2 \middle| \mathcal{F}_{n-1}\right) = k\sigma^2 = Z_{n-1}\sigma^2$$

Combining the last three equations gives

$$EX_n^2 = EX_{n-1}^2 + E(Z_{n-1}\sigma^2/\mu^{2n}) = EX_{n-1}^2 + \sigma^2/\mu^{n+1}$$

since $E(Z_{n-1}/\mu^{n-1}) = EZ_0 = 1$. Now $EX_0^2 = 1$, so $EX_1^2 = 1+\sigma^2/\mu^2$, and induction gives

$$EX_n^2 = 1 + \sigma^2 \sum_{k=2}^{n+1} \mu^{-k}$$

This shows $\sup EX_n^2 < \infty$, so $X_n \to X$ in L^2, and hence $EX_n \to EX$. $EX_n = 1$ for all n, so $EX = 1$ and X is not $\equiv 0$. It follows from Exercise 4.3.11 that $\{X > 0\} = \{Z_n > 0 \text{ for all n }\}$.

Exercises

4.4.1 Show that if $j \le k$, then $E(X_j; N = j) \le E(X_k; N = j)$ and sum over j to get a second proof of $EX_N \le EX_k$.

4.4.2 Generalize the proof of Theorem 4.4.1 to show that if X_n is a submartingale and $M \le N$ are stopping times with $P(N \le k) = 1$, then $EX_M \le EX_N$.

4.4.3 Suppose $M \le N$ are stopping times. If $A \in \mathcal{F}_M$, then

$$L = \begin{cases} M & \text{on } A \\ N & \text{on } A^c \end{cases} \quad \text{is a stopping time.}$$

4.4.4 Use the stopping times from the previous exercise to strengthen the conclusion of Exercise 4.4.2 to $X_M \le E(X_N|\mathcal{F}_M)$.

4.4.5 Prove the following variant of the conditional variance formula. If $\mathcal{F} \subset \mathcal{G}$, then

$$E\big(E[Y|\mathcal{G}] - E[Y|\mathcal{F}]\big)^2 = E(E[Y|\mathcal{G}])^2 - E(E[Y|\mathcal{F}])^2$$

4.4.6 Suppose in addition to the conditions introduced earlier that $|\xi_m| \le K$ and let $s_n^2 = \sum_{m\le n} \sigma_m^2$. Exercise 4.2.2 implies that $S_n^2 - s_n^2$ is a martingale. Use this and Theorem 4.4.1 to conclude

$$P\left(\max_{1\le m\le n} |S_m| \le x\right) \le (x+K)^2/\operatorname{var}(S_n)$$

4.4.7 The next result gives an extension of Theorem 4.4.2 to $p = 1$. Let X_n be a martingale with $X_0 = 0$ and $EX_n^2 < \infty$. Show that

$$P\left(\max_{1\le m\le n} X_m \ge \lambda\right) \le EX_n^2/(EX_n^2 + \lambda^2)$$

Hint: Use the fact that $(X_n + c)^2$ is a submartingale and optimize over c.

4.4.8 Let X_n be a submartingale and $\log^+ x = \max(\log x, 0)$.

$$E\bar{X}_n \le (1-e^{-1})^{-1}\{1 + E(X_n^+ \log^+(X_n^+))\}$$

Prove this by carrying out the following steps: (i) Imitate the proof of 4.4.2 but use the trivial bound $P(A) \le 1$ for $\lambda \le 1$ to show

$$E(\bar{X}_n \wedge M) \le 1 + \int X_n^+ \log(\bar{X}_n \wedge M)\, dP$$

(ii) Use calculus to show $a \log b \le a \log a + b/e \le a \log^+ a + b/e$.

4.4.9 Let X_n and Y_n be martingales with $EX_n^2 < \infty$ and $EY_n^2 < \infty$.

$$EX_nY_n - EX_0Y_0 = \sum_{m=1}^{n} E(X_m - X_{m-1})(Y_m - Y_{m-1})$$

4.4.10 Let X_n, $n \geq 0$, be a martingale and let $\xi_n = X_n - X_{n-1}$ for $n \geq 1$. If EX_0^2, $\sum_{m=1}^{\infty} E\xi_m^2 < \infty$, then $X_n \to X_\infty$ a.s. and in L^2.

4.4.11 Continuing with the notation from the previous problem. If $b_m \uparrow \infty$ and $\sum_{m=1}^{\infty} E\xi_m^2/b_m^2 < \infty$, then $X_n/b_n \to 0$ a.s. In particular, if $E\xi_n^2 \leq K < \infty$ and $\sum_{m=1}^{\infty} b_m^{-2} < \infty$, then $X_n/b_n \to 0$ a.s.

4.5 Square Integrable Martingales*

In this section, we will suppose

$$X_n \text{ is a martingale with } X_0 = 0 \text{ and } EX_n^2 < \infty \text{ for all } n$$

Theorem 4.2.6 implies X_n^2 is a submartingale. It follows from Doob's decomposition Theorem 4.3.2 that we can write $X_n^2 = M_n + A_n$, where M_n is a martingale, and from formulas in Theorems 4.3.2 and 4.4.8 that

$$A_n = \sum_{m=1}^{n} E(X_m^2|\mathcal{F}_{m-1}) - X_{m-1}^2 = \sum_{m=1}^{n} E((X_m - X_{m-1})^2|\mathcal{F}_{m-1})$$

A_n is called the **increasing process** associated with X_n. A_n can be thought of as a path by path measurement of the variance at time n, and $A_\infty = \lim A_n$ as the total variance in the path. Theorems 4.5.2 and 4.5.3 describe the behavior of the martingale on $\{A_n < \infty\}$ and $\{A_n = \infty\}$, respectively. The key to the proof of the first result is the following:

Theorem 4.5.1 $E\left(\sup_m |X_m|^2\right) \leq 4EA_\infty$.

Proof Applying the L^2 maximum inequality (Theorem 4.4.4) to X_n gives

$$E\left(\sup_{0\leq m\leq n} |X_m|^2\right) \leq 4EX_n^2 = 4EA_n$$

since $EX_n^2 = EM_n + EA_n$ and $EM_n = EM_0 = EX_0^2 = 0$. Using the monotone convergence theorem now gives the desired result. □

Theorem 4.5.2 $\lim_{n\to\infty} X_n$ *exists and is finite a.s. on* $\{A_\infty < \infty\}$.

Proof Let $a > 0$. Since $A_{n+1} \in \mathcal{F}_n$, $N = \inf\{n : A_{n+1} > a^2\}$ is a stopping time. Applying Theorem 4.5.1 to $X_{N\wedge n}$ and noticing $A_{N\wedge n} \leq a^2$ gives

$$E\left(\sup_n |X_{N\wedge n}|^2\right) \leq 4a^2$$

so the L^2 convergence theorem, 4.4.6, implies that $\lim X_{N\wedge n}$ exists and is finite a.s. Since a is arbitrary, the desired result follows. □

The next result is a variation on the theme of Exercise 4.4.11.

Theorem 4.5.3 *Let $f \geq 1$ be increasing with $\int_0^\infty f(t)^{-2}\,dt < \infty$. Then $X_n/f(A_n) \to 0$ a.s. on $\{A_\infty = \infty\}$.*

Proof $H_m = f(A_m)^{-1}$ is bounded and predictable, so Theorem 4.2.8 implies

$$Y_n \equiv (H \cdot X)_n = \sum_{m=1}^{n} \frac{X_m - X_{m-1}}{f(A_m)} \quad \text{is a martingale}$$

If B_n is the increasing process associated with Y_n, then

$$\begin{aligned} B_{n+1} - B_n &= E((Y_{n+1} - Y_n)^2 | \mathcal{F}_n) \\ &= E\left(\frac{(X_{n+1} - X_n)^2}{f(A_{n+1})^2} \middle| \mathcal{F}_n \right) = \frac{A_{n+1} - A_n}{f(A_{n+1})^2} \end{aligned}$$

since $f(A_{n+1}) \in \mathcal{F}_n$. Our hypotheses on f imply that

$$\sum_{n=0}^{\infty} \frac{A_{n+1} - A_n}{f(A_{n+1})^2} \leq \sum_{n=0}^{\infty} \int_{[A_n, A_{n+1})} f(t)^{-2}\,dt < \infty$$

so it follows from Theorem 4.5.2 that $Y_n \to Y_\infty$, and the desired conclusion follows from Kronecker's lemma, Theorem 2.5.9. □

Example 4.5.4 Let $\epsilon > 0$ and $f(t) = (t \log^{1+\epsilon} t)^{1/2} \vee 1$. Then f satisfies the hypotheses of Theorem 4.5.3. Let $\xi_1, \xi_2, \ldots$ be independent with $E\xi_m = 0$ and $E\xi_m^2 = \sigma_m^2$. In this case, $X_n = \xi_1 + \cdots + \xi_n$ is a square integrable martingale with $A_n = \sigma_1^2 + \cdots + \sigma_n^2$, so if $\sum_{i=1}^\infty \sigma_i^2 = \infty$, Theorem 4.5.3 implies $X_n/f(A_n) \to 0$ generalizing Theorem 2.5.11.

From Theorem 4.5.3 we get a result due to Dubins and Freedman (1965) that extends our two previous versions in Theorems 2.3.7 and 4.3.4.

Theorem 4.5.5 (Second Borel-Cantelli Lemma, III) *Suppose B_n is adapted to $\mathcal{F}_n$ and let $p_n = P(B_n | \mathcal{F}_{n-1})$. Then*

$$\sum_{m=1}^{n} 1_{B(m)} \Big/ \sum_{m=1}^{n} p_m \to 1 \quad \textit{a.s. on} \quad \left\{ \sum_{m=1}^{\infty} p_m = \infty \right\}$$

Proof Define a martingale by $X_0 = 0$ and $X_n - X_{n-1} = 1_{B_n} - P(B_n|\mathcal{F}_{n-1})$ for $n \geq 1$ so that we have

$$\left(\sum_{m=1}^{n} 1_{B(m)} \Big/ \sum_{m=1}^{n} p_m \right) - 1 = X_n \Big/ \sum_{m=1}^{n} p_m$$

The increasing process associated with X_n has

$$\begin{aligned} A_n - A_{n-1} &= E((X_n - X_{n-1})^2 | \mathcal{F}_{n-1}) \\ &= E\left((1_{B_n} - p_n)^2 \middle| \mathcal{F}_{n-1} \right) = p_n - p_n^2 \leq p_n \end{aligned}$$

On $\{A_\infty < \infty\}$, $X_n \to$ a finite limit by Theorem 4.5.2, so on $\{A_\infty < \infty\} \cap \{\sum_m p_m = \infty\}$

$$X_n \Big/ \sum_{m=1}^{n} p_m \to 0$$

$\{A_\infty = \infty\} = \{\sum_m p_m(1 - p_m) = \infty\} \subset \{\sum_m p_m = \infty\}$, so on $\{A_\infty = \infty\}$ the desired conclusion follows from Theorem 4.5.3 with $f(t) = t \vee 1$. □

Remark The trivial example $B_n = \Omega$ for all n shows we may have $A_\infty < \infty$ and $\sum p_m = \infty$ a.s.

Example 4.5.6 (Bernard Friedman's urn) Consider a variant of Polya's urn (see Section 5.3) in which we add a balls of the color drawn and b balls of the opposite color where $a \geq 0$ and $b > 0$. We will show that if we start with g green balls and r red balls, where $g, r > 0$, then the fraction of green balls $g_n \to 1/2$. Let G_n and R_n be the number of green and red balls after the nth draw is completed. Let B_n be the event that the nth ball drawn is green, and let D_n be the number of green balls drawn in the first n draws. It follows from Theorem 4.5.5 that

$$(\star) \qquad D_n \bigg/ \sum_{m=1}^{n} g_{m-1} \to 1 \quad \text{a.s. on} \quad \sum_{m=1}^{\infty} g_{m-1} = \infty$$

which always holds since $g_m \geq g/(g + r + (a+b)m)$. At this point, the argument breaks into three cases.

Case 1. $a = b = c$. In this case, the result is trivial since we always add c balls of each color.

Case 2. $a > b$. We begin with the observation

$$(*) \qquad g_{n+1} = \frac{G_{n+1}}{G_{n+1} + R_{n+1}} = \frac{g + aD_n + b(n - D_n)}{g + r + n(a+b)}$$

If $\limsup_{n\to\infty} g_n \leq x$, then $(\star)$ implies $\limsup_{n\to\infty} D_n/n \leq x$ and (since $a > b$)

$$\limsup_{n\to\infty} g_{n+1} \leq \frac{ax + b(1-x)}{a+b} = \frac{b + (a-b)x}{a+b}$$

The right-hand side is a linear function with slope < 1 and fixed point at $1/2$, so starting with the trivial upper bound $x = 1$ and iterating we conclude that $\limsup g_n \leq 1/2$. Interchanging the roles of red and green shows $\liminf_{n\to\infty} g_n \geq 1/2$, and the result follows.

Case 3. $a < b$. The result is easier to believe in this case, since we are adding more balls of the type not drawn but is a little harder to prove. The trouble is that when $b > a$ and $D_n \leq xn$, the right-hand side of $(*)$ is maximized by taking $D_n = 0$, so we need to also use the fact that if r_n is fraction of red balls, then

$$r_{n+1} = \frac{R_{n+1}}{G_{n+1} + R_{n+1}} = \frac{r + bD_n + a(n - D_n)}{g + r + n(a+b)}$$

Combining this with the formula for g_{n+1}, it follows that if $\limsup_{n\to\infty} g_n \leq x$ and $\limsup_{n\to\infty} r_n \leq y$, then

$$\limsup_{n\to\infty} g_n \leq \frac{a(1-y) + by}{a+b} = \frac{a + (b-a)y}{a+b}$$
$$\limsup_{n\to\infty} r_n \leq \frac{bx + a(1-x)}{a+b} = \frac{a + (b-a)x}{a+b}$$

Starting with the trivial bounds $x = 1$, $y = 1$ and iterating (observe the two upper bounds are always the same), we conclude as in Case 2 that both limsups are $\leq 1/2$. $\square$

Remark B. Friedman (1949) considered a number of different urn models. The previous result is due to Freedman (1965), who proved the result by different methods. The previous proof is due to Ornstein and comes from a remark in Freedman's paper.

Theorem 4.5.1 came from using Theorem 4.4.4. If we use Theorem 4.4.2 instead, we get a slightly better result.

Theorem 4.5.7 $E(\sup_n |X_n|) \le 3EA_\infty^{1/2}$.

Proof As in the proof of Theorem 4.5.2, we let $a > 0$ and let $N = \inf\{n : A_{n+1} > a^2\}$. This time, however, our starting point is

$$P\left(\sup_m |X_m| > a\right) \le P(N < \infty) + P\left(\sup_m |X_{N\wedge m}| > a\right)$$

$P(N < \infty) = P(A_\infty > a^2)$. To bound the second term, we apply Theorem 4.4.2 to $X^2_{N\wedge m}$ with $\lambda = a^2$ to get

$$P\left(\sup_{m\le n} |X_{N\wedge m}| > a\right) \le a^{-2}EX^2_{N\wedge n} = a^{-2}EA_{N\wedge n} \le a^{-2}E(A_\infty \wedge a^2)$$

Letting $n \to \infty$ in the last inequality, substituting the result in the first one, and integrating gives

$$\int_0^\infty P\left(\sup_m |X_m| > a\right) da \le \int_0^\infty P(A_\infty > a^2)\, da + \int_0^\infty a^{-2}E(A_\infty \wedge a^2)\, da$$

Since $P(A_\infty > a^2) = P(A_\infty^{1/2} > a)$, the first integral is $EA_\infty^{1/2}$. For the second, we use Lemma 2.2.13 (in the first and fourth steps), Fubini's theorem, and calculus to get

$$\int_0^\infty a^{-2}E(A_\infty \wedge a^2)\, da = \int_0^\infty a^{-2} \int_0^{a^2} P(A_\infty > b)\, db\, da$$
$$= \int_0^\infty P(A_\infty > b) \int_{\sqrt{b}}^\infty a^{-2}\, da\, db = \int_0^\infty b^{-1/2}P(A_\infty > b)\, db = 2EA_\infty^{1/2}$$

which completes the proof. □

Example 4.5.8 Let $\xi_1, \xi_2, \ldots$ be i.i.d. with $P(\xi_i = 1) = P(\xi_i = -1)$. Let $S_n = \xi_1 + \cdots + \xi_n$. Theorem 4.4.1 implies that for any stopping time N, $ES_{N\wedge n} = 0$. Using Theorem 4.5.7, we can conclude that if $EN^{1/2} < \infty$, then $ES_N = 0$. Let $T = \inf\{n : S_n = -1\}$. Since $S_T = -1$ does not have mean 0, it follows that $ET^{1/2} = \infty$.

4.6 Uniform Integrability, Convergence in L^1

In this section, we will give necessary and sufficient conditions for a martingale to converge in L^1. The key to this is the following definition. A collection of random variables X_i, $i \in I$, is said to be **uniformly integrable** if

$$\lim_{M\to\infty} \left(\sup_{i\in I} E(|X_i|; |X_i| > M)\right) = 0$$

If we pick M large enough so that the sup < 1, it follows that

$$\sup_{i \in I} E|X_i| \le M + 1 < \infty$$

This remark will be useful several times later.

A trivial example of a uniformly integrable family is a collection of random variables that are dominated by an integrable random variable, i.e., $|X_i| \le Y$, where $EY < \infty$. Our first result gives an interesting example that shows that uniformly integrable families can be very large.

Theorem 4.6.1 *Given a probability space $(\Omega, \mathcal{F}_o, P)$ and an $X \in L^1$, then $\{E(X|\mathcal{F}) : \mathcal{F}$ is a σ-field $\subset \mathcal{F}_o\}$ is uniformly integrable.*

Proof If A_n is a sequence of sets with $P(A_n) \to 0$, then the dominated convergence theorem implies $E(|X|; A_n) \to 0$. From the last result, it follows that if $\epsilon > 0$, we can pick $\delta > 0$ so that if $P(A) \le \delta$, then $E(|X|; A) \le \epsilon$. (If not, there are sets A_n with $P(A_n) \le 1/n$ and $E(|X|; A_n) > \epsilon$, a contradiction.)

Pick M large enough so that $E|X|/M \le \delta$. Jensen's inequality and the definition of conditional expectation imply

$$\begin{aligned} E(\,|E(X|\mathcal{F})|\,;\,|E(X|\mathcal{F})| > M) &\le E(\,E(|X||\mathcal{F})\,;\,E(|X||\mathcal{F}) > M) \\ &= E(\,|X|\,;\,E(|X||\mathcal{F}) > M) \end{aligned}$$

since $\{E(|X||\mathcal{F}) > M\} \in \mathcal{F}$. Using Chebyshev's inequality and recalling the definition of M, we have

$$P\{E(|X||\mathcal{F}) > M\} \le E\{E(|X||\mathcal{F})\}/M = E|X|/M \le \delta$$

So, by the choice of δ, we have

$$E(|E(X|\mathcal{F})|; |E(X|\mathcal{F})| > M) \le \epsilon \quad \text{for all } \mathcal{F}$$

Since ϵ was arbitrary, the collection is uniformly integrable. □

A common way to check uniform integrability is to use:

Theorem 4.6.2 *Let $\varphi \ge 0$ be any function with $\varphi(x)/x \to \infty$ as $x \to \infty$, e.g., $\varphi(x) = x^p$ with $p > 1$ or $\varphi(x) = x \log^+ x$. If $E\varphi(|X_i|) \le C$ for all $i \in I$, then $\{X_i : i \in I\}$ is uniformly integrable.*

Proof Let $\epsilon_M = \sup\{x/\phi(x) : x \ge M\}$. For $i \in I$

$$E(|X_i|; |X_i| > M) \le \epsilon_M E(\phi(|X_i|); |X_i| > M) \le C\epsilon_M$$

and $\epsilon_M \to 0$ as $M \to \infty$. □

The relevance of uniform integrability to convergence in L^1 is explained by:

Theorem 4.6.3 *Suppose that $E|X_n| < \infty$ for all n. If $X_n \to X$ in probability, then the following are equivalent:*
(i) $\{X_n : n \ge 0\}$ is uniformly integrable.
(ii) $X_n \to X$ in L^1.
(iii) $E|X_n| \to E|X| < \infty$.

Proof *(i) implies (ii).* Let

$$\varphi_M(x) = \begin{cases} M & \text{if } x \geq M \\ x & \text{if } |x| \leq M \\ -M & \text{if } x \leq -M \end{cases}$$

The triangle inequality implies

$$|X_n - X| \leq |X_n - \varphi_M(X_n)| + |\varphi_M(X_n) - \varphi_M(X)| + |\varphi_M(X) - X|$$

Since $|\varphi_M(Y) - Y)| = (|Y| - M)^+ \leq |Y| 1_{(|Y|>M)}$, taking expected value gives

$$E|X_n - X| \leq E|\varphi_M(X_n) - \varphi_M(X)| + E(|X_n|; |X_n| > M) + E(|X|; |X| > M)$$

Theorem 2.3.4 implies that $\varphi_M(X_n) \to \varphi_M(X)$ in probability, so the first term $\to 0$ by the bounded convergence theorem. (See Exercise 2.3.5.) If $\epsilon > 0$ and M is large, uniform integrability implies that the second term $\leq \epsilon$. To bound the third term, we observe that uniform integrability implies $\sup E|X_n| < \infty$, so Fatou's lemma (in the form given in Exercise 2.3.4) implies $E|X| < \infty$, and by making M larger we can make the third term $\leq \epsilon$. Combining the last three facts shows $\limsup E|X_n - X| \leq 2\epsilon$. Since ϵ is arbitrary, this proves (ii).

(ii) implies (iii). Jensen's inequality implies

$$|E|X_n| - E|X|| \leq E||X_n| - |X|| \leq E|X_n - X| \to 0$$

(iii) implies (i). Let

$$\psi_M(x) = \begin{cases} x & \text{on } [0, M-1], \\ 0 & \text{on } [M, \infty) \\ \text{linear} & \text{on } [M-1, M] \end{cases}.$$

The dominated convergence theorem implies that if M is large, $E|X| - E\psi_M(|X|) \leq \epsilon/2$. As in the first part of the proof, the bounded convergence theorem implies $E\psi_M(|X_n|) \to E\psi_M(|X|)$, so using (iii), we get that if $n \geq n_0$

$$\begin{aligned} E(|X_n|; |X_n| > M) &\leq E|X_n| - E\psi_M(|X_n|) \\ &\leq E|X| - E\psi_M(|X|) + \epsilon/2 < \epsilon \end{aligned}$$

By choosing M larger, we can make $E(|X_n|; |X_n| > M) \leq \epsilon$ for $0 \leq n < n_0$, so X_n is uniformly integrable. □

We are now ready to state the main theorems of this section. We have already done all the work, so the proofs are short.

Theorem 4.6.4 *For a submartingale, the following are equivalent:*
(i) It is uniformly integrable.
(ii) It converges a.s. and in L^1.
(iii) It converges in L^1.

Proof *(i) implies (ii).* Uniform integrability implies $\sup E|X_n| < \infty$, so the martingale convergence theorem implies $X_n \to X$ a.s., and Theorem 4.6.3 implies $X_n \to X$ in L^1.

(ii) implies (iii). Trivial. *(iii) implies (i).* $X_n \to X$ in L^1 implies $X_n \to X$ in probability, (see Lemma 2.2.2), so this follows from Theorem 4.6.3. □

Before proving the analogue of Theorem 4.6.4 for martingales, we will isolate two parts of the argument that will be useful later.

Lemma 4.6.5 *If integrable random variables $X_n \to X$ in L^1, then*

$$E(X_n; A) \to E(X; A)$$

Proof $|EX_m 1_A - EX1_A| \le E|X_m 1_A - X1_A| \le E|X_m - X| \to 0$ □

Lemma 4.6.6 *If a martingale $X_n \to X$ in L^1, then $X_n = E(X|\mathcal{F}_n)$.*

Proof The martingale property implies that if $m > n$, $E(X_m|\mathcal{F}_n) = X_n$, so if $A \in \mathcal{F}_n$, $E(X_n; A) = E(X_m; A)$. Lemma 4.6.5 implies $E(X_m; A) \to E(X; A)$, so we have $E(X_n; A) = E(X; A)$ for all $A \in \mathcal{F}_n$. Recalling the definition of conditional expectation, it follows that $X_n = E(X|\mathcal{F}_n)$. □

Theorem 4.6.7 *For a martingale, the following are equivalent:*
(i) It is uniformly integrable.
(ii) It converges a.s. and in L^1.
(iii) It converges in L^1.
(iv) There is an integrable random variable X so that $X_n = E(X|\mathcal{F}_n)$.

Proof *(i) implies (ii).* Since martingales are also submartingales, this follows from Theorem 4.6.4. *(ii) implies (iii).* Trivial. *(iii) implies (iv).* This follows from Lemma 4.6.6. *(iv) implies (i).* This follows from Theorem 4.6.1. □

The next result is related to Lemma 4.6.6 but goes in the other direction.

Theorem 4.6.8 *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, i.e., $\mathcal{F}_n$ is an increasing sequence of σ-fields and $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. As $n \to \infty$,*

$$E(X|\mathcal{F}_n) \to E(X|\mathcal{F}_\infty) \quad \textit{a.s. and in } L^1$$

Proof The first step is to note that if $m > n$, then Theorem 4.1.13 implies

$$E(E(X|\mathcal{F}_m)|\mathcal{F}_n) = E(X|\mathcal{F}_n)$$

so $Y_n = E(X|\mathcal{F}_n)$ is a martingale. Theorem 4.6.1 implies that Y_n is uniformly integrable, so Theorem 4.6.7 implies that Y_n converges a.s. and in L^1 to a limit Y_∞. The definition of Y_n and Lemma 4.6.6 imply $E(X|\mathcal{F}_n) = Y_n = E(Y_\infty|\mathcal{F}_n)$, and hence

$$\int_A X\, dP = \int_A Y_\infty\, dP \quad \text{for all } A \in \mathcal{F}_n$$

Since X and Y_∞ are integrable, and $\cup_n \mathcal{F}_n$ is a π-system, the $\pi - \lambda$ theorem implies that the last result holds for all $A \in \mathcal{F}_\infty$. Since $Y_\infty \in \mathcal{F}_\infty$, it follows that $Y_\infty = E(X|\mathcal{F}_\infty)$. □

An immediate consequence of Theorem 4.6.8 is:

Theorem 4.6.9 (Lévy's 0-1 law) *If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then $E(1_A|\mathcal{F}_n) \to 1_A$ a.s.*

To steal a line from Chung: *"The reader is urged to ponder over the meaning of this result and judge for himself whether it is obvious or incredible."* We will now argue for the two points of view.

"It is obvious." $1_A \in \mathcal{F}_\infty$, and $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, so our best guess of 1_A given the information in $\mathcal{F}_n$ should approach 1_A (the best guess given $\mathcal{F}_\infty$).

"It is incredible." Let $X_1, X_2, \ldots$ be independent and suppose $A \in \mathcal{T}$, the tail σ-field. For each n, A is independent of $\mathcal{F}_n$, so $E(1_A|\mathcal{F}_n) = P(A)$. As $n \to \infty$, the left-hand side converges to 1_A a.s., so $P(A) = 1_A$ a.s., and it follows that $P(A) \in \{0, 1\}$, i.e., we have proved Kolmogorov's 0-1 law.

The last argument may not show that Theorem 4.6.9 is "too unusual or improbable to be possible," but this and other applications of Theorem 4.6.9 show that it is a very useful result.

A more technical consequence of Theorem 4.6.8 is:

Theorem 4.6.10 (Dominated convergence theorem for conditional expectations) *Suppose $Y_n \to Y$ a.s. and $|Y_n| \le Z$ for all n, where $EZ < \infty$. If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then*

$$E(Y_n|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty) \quad a.s.$$

Proof Let $W_N = \sup\{|Y_n - Y_m| : n, m \ge N\}$. $W_N \le 2Z$, so $EW_N < \infty$. Using monotonicity (4.1.2) and applying Theorem 4.6.8 to W_N gives

$$\limsup_{n\to\infty} E(|Y_n - Y||\mathcal{F}_n) \le \lim_{n\to\infty} E(W_N|\mathcal{F}_n) = E(W_N|\mathcal{F}_\infty)$$

The last result is true for all N and $W_N \downarrow 0$ as $N \uparrow \infty$, so (4.1.3) implies $E(W_N|\mathcal{F}_\infty) \downarrow 0$, and Jensen's inequality gives us

$$|E(Y_n|\mathcal{F}_n) - E(Y|\mathcal{F}_n)| \le E(|Y_n - Y||\mathcal{F}_n) \to 0 \quad \text{a.s. as } n \to \infty$$

Theorem 4.6.8 implies $E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty)$ a.s. The desired result follows from the last two conclusions and the triangle inequality. □

Example 4.6.11 Suppose $X_1, X_2, \ldots$ are uniformly integrable and $\to X$ a.s. Theorem 4.6.3 implies $X_n \to X$ in L^1 and combining this with Exercise 4.6.7 shows $E(X_n|\mathcal{F}) \to E(X|\mathcal{F})$ in L^1. We will now show that $E(X_n|\mathcal{F})$ need not converge a.s. Let $Y_1, Y_2, \ldots$ and $Z_1, Z_2, \ldots$ be independent r.v.'s with

$$P(Y_n = 1) = 1/n \qquad P(Y_n = 0) = 1 - 1/n$$
$$P(Z_n = n) = 1/n \qquad P(Z_n = 0) = 1 - 1/n$$

Let $X_n = Y_n Z_n$. $P(X_n > 0) = 1/n^2$, so the Borel-Cantelli lemma implies $X_n \to 0$ a.s. $E(X_n; |X_n| \ge 1) = n/n^2$, so X_n is uniformly integrable. Let $\mathcal{F} = \sigma(Y_1, Y_2, \ldots)$.

$$E(X_n|\mathcal{F}) = Y_n E(Z_n|\mathcal{F}) = Y_n EZ_n = Y_n$$

Since $Y_n \to 0$ in L^1 but not a.s., the same is true for $E(X_n|\mathcal{F})$.

Exercises

4.6.1 Let $Z_1, Z_2, \ldots$ be i.i.d. with $E|Z_i| < \infty$, let θ be an independent r.v. with finite mean, and let $Y_i = Z_i + \theta$. If Z_i is normal(0,1), then in statistical terms we have a sample from a normal population with variance 1 and unknown mean. The distribution of θ is called the **prior distribution**, and $P(\theta \in \cdot|Y_1, \ldots, Y_n)$ is called the **posterior distribution** after n observations. Show that $E(\theta|Y_1, \ldots, Y_n) \to \theta$ a.s.

In the next two exercises, $\Omega = [0, 1)$, $I_{k,n} = [k2^{-n}, (k+1)2^{-n})$, and $\mathcal{F}_n = \sigma(I_{k,n} : 0 \le k < 2^n)$.

4.6.2 f is said to be **Lipschitz continuous** if $|f(t)-f(s)| \le K|t-s|$ for $0 \le s, t < 1$. Show that $X_n = (f((k+1)2^{-n}) - f(k2^{-n}))/2^{-n}$ on $I_{k,n}$ defines a martingale, $X_n \to X_\infty$ a.s. and in L^1, and

$$f(b) - f(a) = \int_a^b X_\infty(\omega)\, d\omega$$

4.6.3 Suppose f is integrable on [0,1). $E(f|\mathcal{F}_n)$ is a step function and $\to f$ in L^1. From this it follows immediately that if $\epsilon > 0$, there is a step function g on [0,1] with $\int |f - g|\, dx < \epsilon$. This approximation is much simpler than the bare-hands approach we used in Exercise 1.4.3, but of course we are using a lot of machinery.

4.6.4 Let X_n be r.v.'s taking values in $[0, \infty)$. Let $D = \{X_n = 0 \text{ for some } n \ge 1\}$ and assume

$$P(D|X_1, \ldots, X_n) \ge \delta(x) > 0 \quad \text{a.s. on } \{X_n \le x\}$$

Use Theorem 4.6.9 to conclude that $P(D \cup \{\lim_n X_n = \infty\}) = 1$.

4.6.5 Let Z_n be a branching process with offspring distribution p_k (see the end of Section 5.3 for definitions). Use the last result to show that if $p_0 > 0$, then $P(\lim_n Z_n = 0 \text{ or } \infty) = 1$.

4.6.6 Let $X_n \in [0, 1]$ be adapted to $\mathcal{F}_n$. Let $\alpha, \beta > 0$ with $\alpha + \beta = 1$ and suppose

$$P(X_{n+1} = \alpha + \beta X_n|\mathcal{F}_n) = X_n \qquad P(X_{n+1} = \beta X_n|\mathcal{F}_n) = 1 - X_n$$

Show $P(\lim_n X_n = 0 \text{ or } 1) = 1$ and if $X_0 = \theta$, then $P(\lim_n X_n = 1) = \theta$.

4.6.7 Show that if $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $Y_n \to Y$ in L^1, then $E(Y_n|\mathcal{F}_n) \to E(Y|\mathcal{F}_\infty)$ in L^1.

4.7 Backwards Martingales

A **backwards martingale** (some authors call them reversed) is a martingale indexed by the negative integers, i.e., $X_n, n \le 0$, adapted to an increasing sequence of σ-fields $\mathcal{F}_n$ with

$$E(X_{n+1}|\mathcal{F}_n) = X_n \qquad \text{for } n \le -1$$

Because the σ-fields decrease as $n \downarrow -\infty$, the convergence theory for backwards martingales is particularly simple.

Theorem 4.7.1 *$X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. and in L^1.*

Proof Let U_n be the number of upcrossings of $[a,b]$ by $X_{-n}, \dots, X_0$. The upcrossing inequality, Theorem 4.2.10, implies $(b-a)EU_n \le E(X_0 - a)^+$. Letting $n \to \infty$ and using the monotone convergence theorem, we have $EU_\infty < \infty$, so by the remark after the proof of Theorem 4.2.11, the limit exists a.s. The martingale property implies $X_n = E(X_0|\mathcal{F}_n)$, so Theorem 4.6.1 implies X_n is uniformly integrable and Theorem 4.6.3 tells us that the convergence occurs in L^1. □

The next result identifies the limit in Theorem 4.7.1.

Theorem 4.7.2 *If $X_{-\infty} = \lim_{n\to-\infty} X_n$ and $\mathcal{F}_{-\infty} = \cap_n \mathcal{F}_n$, then $X_{-\infty} = E(X_0|\mathcal{F}_{-\infty})$.*

Proof Clearly, $X_{-\infty} \in \mathcal{F}_{-\infty}$. $X_n = E(X_0|\mathcal{F}_n)$, so if $A \in \mathcal{F}_{-\infty} \subset \mathcal{F}_n$, then

$$\int_A X_n \, dP = \int_A X_0 \, dP$$

Theorem 4.7.1 and Lemma 4.6.5 imply $E(X_n; A) \to E(X_{-\infty}; A)$, so

$$\int_A X_{-\infty} \, dP = \int_A X_0 \, dP$$

for all $A \in \mathcal{F}_{-\infty}$, proving the desired conclusion. □

The next result is Theorem 4.6.8 backwards.

Theorem 4.7.3 *If $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ as $n \downarrow -\infty$ (i.e., $\mathcal{F}_{-\infty} = \cap_n \mathcal{F}_n$), then*

$$E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_{-\infty}) \quad \text{a.s. and in } L^1$$

Proof $X_n = E(Y|\mathcal{F}_n)$ is a backwards martingale, so Theorem 4.7.1 and 4.7.2 imply that as $n \downarrow -\infty$, $X_n \to X_{-\infty}$ a.s. and in L^1, where

$$X_{-\infty} = E(X_0|\mathcal{F}_{-\infty}) = E(E(Y|\mathcal{F}_0)|\mathcal{F}_{-\infty}) = E(Y|\mathcal{F}_{-\infty})$$ □

Even though the convergence theory for backwards martingales is easy, there are some nice applications. For the rest of the section, we return to the special space utilized in Section 4.1, so we can utilize definitions given there. That is, we suppose

$$\Omega = \{(\omega_1, \omega_2, \dots) : \omega_i \in S\}$$
$$\mathcal{F} = \mathcal{S} \times \mathcal{S} \times \dots$$
$$X_n(\omega) = \omega_n$$

Let $\mathcal{E}_n$ be the σ-field generated by events that are invariant under permutations that leave $n+1, n+2, \dots$ fixed and let $\mathcal{E} = \cap_n \mathcal{E}_n$ be the exchangeable σ-field.

Example 4.7.4 (Strong law of large numbers) Let $\xi_1, \xi_2, \dots$ be i.i.d. with $E|\xi_i| < \infty$. Let $S_n = \xi_1 + \dots + \xi_n$, let $X_{-n} = S_n/n$, and let

$$\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \dots) = \sigma(S_n, \xi_{n+1}, \xi_{n+2}, \dots)$$

To compute $E(X_{-n}|\mathcal{F}_{-n-1})$, we observe that if $j, k \le n+1$, symmetry implies $E(\xi_j|\mathcal{F}_{-n-1}) = E(\xi_k|\mathcal{F}_{-n-1})$, so

$$E(\xi_{n+1}|\mathcal{F}_{-n-1}) = \frac{1}{n+1}\sum_{k=1}^{n+1} E(\xi_k|\mathcal{F}_{-n-1})$$
$$= \frac{1}{n+1}E(S_{n+1}|\mathcal{F}_{-n-1}) = \frac{S_{n+1}}{n+1}$$

Since $X_{-n} = (S_{n+1} - \xi_{n+1})/n$, it follows that

$$E(X_{-n}|\mathcal{F}_{-n-1}) = E(S_{n+1}/n|\mathcal{F}_{-n-1}) - E(\xi_{n+1}/n|\mathcal{F}_{-n-1})$$
$$= \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n+1} = X_{-n-1}$$

The last computation shows X_{-n} is a backwards martingale, so it follows from Theorems 4.7.1 and 4.7.2 that $\lim_{n\to\infty} S_n/n = E(X_{-1}|\mathcal{F}_{-\infty})$. Since $\mathcal{F}_{-n} \subset \mathcal{E}_n$, $\mathcal{F}_{-\infty} \subset \mathcal{E}$. The Hewitt-Savage 0-1 law (Theorem 2.5.4) says $\mathcal{E}$ is trivial, so we have

$$\lim_{n\to\infty} S_n/n = E(X_{-1}) \quad \text{a.s.}$$

Example 4.7.5 (Ballot theorem) Let $\{\xi_j, 1 \le j \le n\}$ be i.i.d. nonnegative integer-valued r.v.'s, let $S_k = \xi_1 + \cdots + \xi_k$, and let $G = \{S_j < j \text{ for } 1 \le j \le n\}$. Then

$$P(G|S_n) = (1 - S_n/n)^+ \tag{4.7.1}$$

Remark To explain the name, consider an election in which candidate B gets β votes and A gets $\alpha > \beta$ votes. Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. and take values 0 or 2 with probability 1/2 each. Interpreting 0's and 2's as votes for candidates A and B, we see that $G = \{A \text{ leads } B \text{ throughout the counting}\}$, so if $n = \alpha + \beta$

$$P(G|B \text{ gets } \beta \text{ votes }) = \left(1 - \frac{2\beta}{n}\right)^+ = \frac{\alpha - \beta}{\alpha + \beta}$$

the result in Theorem 4.9.2.

Proof The result is trivial when $S_n \ge n$, so suppose $S_n < n$. Computations in Example 4.7.4 show that $X_{-j} = S_j/j$ is a martingale w.r.t. $\mathcal{F}_{-j} = \sigma(S_j, \ldots, S_n)$. Let $T = \inf\{k \ge -n : X_k \ge 1\}$ and set $T = -1$ if the set is $\emptyset$. We claim that $X_T = 1$ on G^c. To check this, note that if $S_{j+1} < j+1$, then the fact that the ξ_i are nonnegative integer values implies $S_j \le S_{j+1} \le j$. Since $G \subset \{T = -1\}$ and $S_1 < 1$ implies $S_1 = 0$, we have $X_T = 0$ on G. Noting $\mathcal{F}_{-n} = \sigma(S_n)$ and using Exercise 4.4.4, we see that on $\{S_n < n\}$

$$P(G^c|S_n) = E(X_T|\mathcal{F}_{-n}) = X_{-n} = S_n/n$$

Subtracting from 1 and recalling that this computation has been done under the assumption $S_n < n$ gives the desired result. □

Example 4.7.6 (Hewitt-Savage 0-1 law) If $X_1, X_2, \ldots$ are i.i.d. and $A \in \mathcal{E}$, then $P(A) \in \{0, 1\}$.

The key to the new proof is:

Lemma 4.7.7 *Suppose $X_1, X_2, \ldots$ are i.i.d. and let*

$$A_n(\varphi) = \frac{1}{(n)_k} \sum_i \varphi(X_{i_1}, \ldots, X_{i_k})$$

where the sum is over all sequences of distinct integers $1 \le i_1, \ldots, i_k \le n$ and

$$(n)_k = n(n-1)\cdots(n-k+1)$$

is the number of such sequences. If φ is bounded, $A_n(\varphi) \to E\varphi(X_1, \ldots, X_k)$ a.s.

Proof $A_n(\varphi) \in \mathcal{E}_n$, so

$$\begin{aligned} A_n(\varphi) = E(A_n(\varphi)|\mathcal{E}_n) &= \frac{1}{(n)_k} \sum_i E(\varphi(X_{i_1}, \ldots, X_{i_k})|\mathcal{E}_n) \\ &= E(\varphi(X_1, \ldots, X_k)|\mathcal{E}_n) \end{aligned}$$

since all the terms in the sum are the same. Theorem 4.7.3 with $\mathcal{F}_{-m} = \mathcal{E}_m$ for $m \ge 1$ implies that

$$E(\varphi(X_1, \ldots, X_k)|\mathcal{E}_n) \to E(\varphi(X_1, \ldots, X_k)|\mathcal{E})$$

We want to show that the limit is $E(\varphi(X_1, \ldots, X_k))$. The first step is to observe that there are $k(n-1)_{k-1}$ terms in $A_n(\varphi)$ involving X_1 and φ is bounded, so if we let $1 \in i$ denote the sum over sequences that contain 1

$$\frac{1}{(n)_k} \sum_{1 \in i} \varphi(X_{i_1}, \ldots, X_{i_k}) \le \frac{k(n-1)_{k-1}}{(n)_k} \sup \phi \to 0$$

This shows that

$$E(\varphi(X_1, \ldots, X_k)|\mathcal{E}) \in \sigma(X_2, X_3, \ldots)$$

Repeating the argument for $2, 3, \ldots, k$ shows

$$E(\varphi(X_1, \ldots, X_k)|\mathcal{E}) \in \sigma(X_{k+1}, X_{k+2}, \ldots)$$

Intuitively, if the conditional expectation of a r.v. is independent of the r.v., then

(a) $$E(\varphi(X_1, \ldots, X_k)|\mathcal{E}) = E(\varphi(X_1, \ldots, X_k))$$

To show this, we prove:

(b) If $EX^2 < \infty$ and $E(X|\mathcal{G}) \in \mathcal{F}$ with X independent of $\mathcal{F}$, then $E(X|\mathcal{G}) = EX$.

Proof Let $Y = E(X|\mathcal{G})$ and note that Theorem 4.1.11 implies $EY^2 \le EX^2 < \infty$. By independence, $EXY = EX\,EY = (EY)^2$ since $EY = EX$. From the geometric interpretation of conditional expectation, Theorem 4.1.15, $E((X-Y)Y) = 0$, so $EY^2 = EXY = (EY)^2$ and $\text{var}\,(Y) = EY^2 - (EY)^2 = 0$. □

(a) holds for all bounded φ, so $\mathcal{E}$ is independent of $\mathcal{G}_k = \sigma(X_1, \ldots, X_k)$. Since this holds for all k, and $\cup_k \mathcal{G}_k$ is a π-system that contains Ω, Theorem 2.1.6 implies $\mathcal{E}$ is independent of $\sigma(\cup_k \mathcal{G}_k) \supset \mathcal{E}$, and we get the usual 0-1 law punch line. If $A \in \mathcal{E}$, it is independent of itself, and hence $P(A) = P(A \cap A) = P(A)P(A)$, i.e., $P(A) \in \{0, 1\}$. □

Example 4.7.8 (de Finetti's Theorem) A sequence $X_1, X_2, \ldots$ is said to be **exchangeable** if for each n and permutation π of $\{1, \ldots, n\}$, $(X_1, \ldots, X_n)$ and $(X_{\pi(1)}, \ldots, X_{\pi(n)})$ have the same distribution.

Theorem 4.7.9 (de Finetti's Theorem) *If $X_1, X_2, \ldots$ are exchangeable then conditional on $\mathcal{E}$, $X_1, X_2, \ldots$ are independent and identically distributed.*

Proof Repeating the first calculation in the proof of Lemma 4.7.7 and using the notation introduced there shows that for any exchangeable sequence:

$$\begin{aligned} A_n(\varphi) = E(A_n(\varphi)|\mathcal{E}_n) &= \frac{1}{(n)_k} \sum_i E(\varphi(X_{i_1}, \ldots, X_{i_k})|\mathcal{E}_n) \\ &= E(\varphi(X_1, \ldots, X_k)|\mathcal{E}_n) \end{aligned}$$

since all the terms in the sum are the same. Again, Theorem 4.7.3 implies that

$$A_n(\varphi) \to E(\varphi(X_1, \ldots, X_k)|\mathcal{E}) \tag{4.7.2}$$

This time, however, $\mathcal{E}$ may be nontrivial, so we cannot hope to show that the limit is $E(\varphi(X_1, \ldots, X_k))$.

Let f and g be bounded functions on $\mathbf{R}^{k-1}$ and $\mathbf{R}$, respectively. If we let $I_{n,k}$ be the set of all sequences of distinct integers $1 \le i_1, \ldots, i_k \le n$, then

$$\begin{aligned} (n)_{k-1} A_n(f)\, n A_n(g) &= \sum_{i \in I_{n,k-1}} f(X_{i_1}, \ldots, X_{i_{k-1}}) \sum_m g(X_m) \\ &= \sum_{i \in I_{n,k}} f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_{i_k}) \\ &\quad + \sum_{i \in I_{n,k-1}} \sum_{j=1}^{k-1} f(X_{i_1}, \ldots, X_{i_{k-1}}) g(X_{i_j}) \end{aligned}$$

If we let $\varphi(x_1, \ldots, x_k) = f(x_1, \ldots, x_{k-1}) g(x_k)$, note that

$$\frac{(n)_{k-1} n}{(n)_k} = \frac{n}{(n-k+1)} \quad \text{and} \quad \frac{(n)_{k-1}}{(n)_k} = \frac{1}{(n-k+1)}$$

then rearrange, we have

$$A_n(\varphi) = \frac{n}{n-k+1} A_n(f) A_n(g) - \frac{1}{n-k+1} \sum_{j=1}^{k-1} A_n(\varphi_j)$$

where $\varphi_j(x_1, \ldots, x_{k-1}) = f(x_1, \ldots, x_{k-1}) g(x_j)$. Applying (4.7.2) to φ, f, g, and all the φ_j gives

$$E(f(X_1, \ldots, X_{k-1}) g(X_k)|\mathcal{E}) = E(f(X_1, \ldots, X_{k-1})|\mathcal{E}) E(g(X_k)|\mathcal{E})$$

It follows by induction that

$$E\left(\prod_{j=1}^k f_j(X_j) \middle| \mathcal{E} \right) = \prod_{j=1}^k E(f_j(X_j)|\mathcal{E}) \qquad \square$$

When the X_i take values in a nice space, there is a regular conditional distribution for $(X_1, X_2, \ldots)$ given $\mathcal{E}$, and the sequence can be represented as a mixture of i.i.d. sequences. Hewitt and Savage (1956) call the sequence **presentable** in this case. For the usual measure theoretic problems, the last result is not valid when the X_i take values in an arbitrary measure space. See Dubins and Freedman (1979) and Freedman (1980) for counterexamples.

The simplest special case of Theorem 4.7.9 occurs when the $X_i \in \{0, 1\}$. In this case:

Theorem 4.7.10 *If $X_1, X_2, \ldots$ are exchangeable and take values in $\{0, 1\}$, then there is a probability distribution on $[0, 1]$ so that*

$$P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \int_0^1 \theta^k (1-\theta)^{n-k}\, dF(\theta)$$

This result is useful for people concerned about the foundations of statistics (see Section 3.7 of Savage (1972)), since from the palatable assumption of symmetry one gets the powerful conclusion that the sequence is a mixture of i.i.d. sequences. Theorem 4.7.10 has been proved in a variety of different ways. See Feller, Vol. II (1971), pp. 228–229 for a proof that is related to the moment problem. Diaconis and Freedman (1980) have a nice proof that starts with the trivial observation that the distribution of a finite exchangeable sequence X_m, $1 \le m \le n$ has the form $p_0 H_{0,n} + \cdots + p_n H_{n,n}$, where $H_{m,n}$ is "drawing without replacement from an urn with m ones and $n - m$ zeros." If $m \to \infty$ and $m/n \to p$, then $H_{m,n}$ approaches product measure with density p. Theorem 4.7.10 follows easily from this, and one can get bounds on the rate of convergence.

Exercises

4.7.1 Show that if a backwards martingale has $X_0 \in L^p$, the convergence occurs in L^p.

4.7.2 Prove the backwards analogue of Theorem 4.6.10. Suppose $Y_n \to Y_{-\infty}$ a.s. as $n \to -\infty$ and $|Y_n| \le Z$ a.s., where $EZ < \infty$. If $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$, then $E(Y_n|\mathcal{F}_n) \to E(Y_{-\infty}|\mathcal{F}_{-\infty})$ a.s.

4.7.3 Prove directly from the definition that if $X_1, X_2, \ldots \in \{0, 1\}$ are exchangeable

$$P(X_1 = 1, \ldots, X_k = 1 | S_n = m) = \binom{n-k}{n-m} \Big/ \binom{n}{m}$$

4.7.4 If $X_1, X_2, \ldots \in \mathbf{R}$ are exchangeable with $EX_i^2 < \infty$, then $E(X_1 X_2) \ge 0$.

4.7.5 Use the first few lines of the proof of Lemma 4.7.7 to conclude that if $X_1, X_2, \ldots$ are i.i.d. with $EX_i = \mu$ and $\operatorname{var}(X_i) = \sigma^2 < \infty$, then

$$\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} (X_i - X_j)^2 \to 2\sigma^2$$

4.8 Optional Stopping Theorems

In this section, we will prove a number of results that allow us to conclude that if X_n is a submartingale and $M \le N$ are stopping times, then $EX_M \le EX_N$. Example 4.2.13 shows that this is not always true, but Exercise 4.4.2 shows this is true if N is bounded, so our attention will be focused on the case of unbounded N.

Theorem 4.8.1 *If X_n is a uniformly integrable submartingale, then for any stopping time N, $X_{N\wedge n}$ is uniformly integrable.*

As in Theorem 4.2.5, the last result implies one for supermartingales with $\geq$ and one for martingales with $=$. This is true for the next two theorems as well.

Proof X_n^+ is a submartingale, so Theorem 4.4.1 implies $EX_{N\wedge n}^+ \leq EX_n^+$. Since X_n^+ is uniformly integrable, it follows from the remark after the definition that

$$\sup_n EX_{N\wedge n}^+ \leq \sup_n EX_n^+ < \infty$$

Using the martingale convergence theorem (4.2.11) now gives $X_{N\wedge n} \to X_N$ a.s. (here $X_\infty = \lim_n X_n$) and $E|X_N| < \infty$. With this established, the rest is easy. We write

$$\begin{aligned} E(|X_{N\wedge n}|; |X_{N\wedge n}| > K) &= E(|X_N|; |X_N| > K, N \leq n) \\ &\quad + E(|X_n|; |X_n| > K, N > n) \end{aligned}$$

Since $E|X_N| < \infty$ and X_n is uniformly integrable, if K is large, then each term is $< \epsilon/2$. □

From the last computation in the proof of Theorem 4.8.1, we get:

Theorem 4.8.2 *If $E|X_N| < \infty$ and $X_n 1_{(N>n)}$ is uniformly integrable, then $X_{N\wedge n}$ is uniformly integrable and hence $EX_0 \leq EX_N$.*

Theorem 4.8.3 *If X_n is a uniformly integrable submartingale, then for any stopping time $N \leq \infty$, we have $EX_0 \leq EX_N \leq EX_\infty$, where $X_\infty = \lim X_n$.*

Proof Theorem 4.4.1 implies $EX_0 \leq EX_{N\wedge n} \leq EX_n$. Letting $n \to \infty$ and observing that Theorem 4.8.1 and 4.6.4 imply $X_{N\wedge n} \to X_N$ and $X_n \to X_\infty$ in L^1 gives the desired result. □

The next result does not require uniform integrability.

Theorem 4.8.4 *If X_n is a nonnegative supermartingale and $N \leq \infty$ is a stopping time, then $EX_0 \geq EX_N$, where $X_\infty = \lim X_n$, which exists by Theorem 4.2.12.*

Proof Using Theorem 4.4.1 and Fatou's Lemma,

$$EX_0 \geq \liminf_{n\to\infty} EX_{N\wedge n} \geq EX_N$$ □

The next result is useful in some situations.

Theorem 4.8.5 *Suppose X_n is a submartingale and $E(|X_{n+1} - X_n||\mathcal{F}_n) \leq B$ a.s. If N is a stopping time with $EN < \infty$, then $X_{N\wedge n}$ is uniformly integrable and hence $EX_N \geq EX_0$.*

Proof We begin by observing that

$$|X_{N\wedge n}| \leq |X_0| + \sum_{m=0}^{\infty} |X_{m+1} - X_m| 1_{(N>m)}$$

To prove uniform integrability, it suffices to show that the right-hand side has finite expectation for then $|X_{N\wedge n}|$ is dominated by an integrable r.v. Now, $\{N > m\} \in \mathcal{F}_m$, so

$$E(|X_{m+1} - X_m|; N > m) = E(E(|X_{m+1} - X_m||\mathcal{F}_m); N > m) \leq BP(N > m)$$

and $E\sum_{m=0}^{\infty} |X_{m+1} - X_m| 1_{(N>m)} \leq B \sum_{m=0}^{\infty} P(N > m) = BEN < \infty$. □

4.8.1 Applications to Random Walks

Let $\xi_1, \xi_2, \ldots$ be i.i.d., $S_n = S_0 + \xi_1 + \cdots + \xi_n$, where S_0 is a constant, and let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. We will now derive some result by using the three martingales from Section 4.2.

Linear martingale If we let $\mu = E\xi_i$, then $X_n = S_n - n\mu$ is a martingale. (See Example 4.2.1.)

Using the linear martingale with Theorem 4.8.5 gives:

Theorem 4.8.6 (Wald's equation) *If $\xi_1, \xi_2, \ldots$ are i.i.d. with $E\xi_i = \mu$, $S_n = \xi_1 + \cdots + \xi_n$ and N is a stopping time with $EN < \infty$, then $ES_N = \mu EN$.*

Proof Let $X_n = S_n - n\mu$ and note that $E(|X_{n+1} - X_n||\mathcal{F}_n) = E|\xi_i - \mu|$. □

Quadratic martingale Suppose $E\xi_i = 0$ and $E\xi_i^2 = \sigma^2 \in (0, \infty)$. Then $X_n = S_n - n\sigma^2$ is a martingale. (See Example 4.2.2.)

Exponential martingale Suppose that $\phi(\theta) = E\exp(\theta\xi_i) < \infty$. Then $X_n = \exp(\theta S_n)/\phi(\theta)^n$ is martingale. (See Example 4.2.3.)

Theorem 4.8.7 (Symmetric simple random walk) *refers to the special case in which $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. Suppose $S_0 = x$ and let $N = \min\{n : S_n \notin (a, b)\}$. Writing a subscript x to remind us of the starting point*

$$(a) \qquad P_x(S_N = a) = \frac{b - x}{b - a} \qquad P_x(S_N = b) = \frac{x - a}{b - a}$$

(b) $E_0 N = -ab$ and hence $E_x N = (b - x)(x - a)$.

Let $T_x = \min\{n : S_n = x\}$. Taking $a = 0$, $x = 1$ and $b = M$ we have

$$P_1(T_M < T_0) = \frac{1}{M} \qquad P_1(T_M < T_0) = \frac{M - 1}{M}$$

The first result proves (4.4.1). Letting $M \to \infty$ in the second we have $P_1(T_0 < \infty) = 1$.

Proof (a) To see that $P(N < \infty) = 1$ note that if we have $(b - a)$ consecutive steps of size $+1$, we will exit the interval. From this it follows that

$$P(N > m(b - a)) \leq (1 - 2^{-(b-a)})^m$$

so $EN < \infty$.

Clearly, $E|S_N| < \infty$ and $S_n 1_{\{N>n\}}$ are uniformly integrable, so using Theorem 4.8.2, we have

$$x = ES_N = aP_x(S_N = a) + b[1 - P_x(S_N = a)]$$

Rearranging, we have $P_x(S_N = a) = (b-x)/(b-a)$ subtracting this from 1, $P_x(S_N = b) = (x-a)/(b-a)$.
(b) The second result is an immediate consequence of the first.

Using the stopping theorem for the bounded stopping time $N \wedge n$, we have

$$0 = E_0 S_{N\wedge n}^2 - E_0(N \wedge n)$$

The monotone convergence theorem implies that $E_0(N \wedge N) \uparrow E_0 N$. Using the bounded convergence theorem and the result of (a) with $x = 0$ implies

$$\begin{aligned} E_0 S_{N\wedge n}^2 &\to a^2 \frac{b}{b-a} + b^2 \frac{-a}{b-a} \\ &= -ab \left[\frac{-a}{b-a} + \frac{b}{b-a} \right] = -ab \end{aligned}$$

which completes the proof. □

Remark The reader should study the technique in this proof of (b) because it is useful in a number of situations. We apply Theorem 4.4.1 to the bounded stopping time $T_b \wedge n$, then let $n \to \infty$, and use an appropriate convergence theorem.

Theorem 4.8.8 *Let S_n be symmetric random walk with $S_0 = 0$ and let $T_1 = \min\{n : S_n = 1\}$.*

$$Es^{T_1} = \frac{1 - \sqrt{1-s^2}}{s}$$

Inverting the generating function we find

$$P(T_1 = 2n-1) = \frac{1}{2n-1} \cdot \frac{(2n)!}{n!\,n!} 2^{-2n} \tag{4.8.1}$$

Proof We will use the exponential martingale $X_n = \exp(\theta S_n)/\phi(\theta)^n$ with $\theta > 0$. The remark after Theorem 4.8.7 implies $P_0(T_1 < \infty) = 1$.

$$\phi(\theta) = E \exp(\theta \xi_i) = (e^{\theta} + e^{-\theta})/2.$$

$\phi(0) = 1$, $\phi'(0) = 0$, and $\phi''(\theta) > 0$, so if $\theta > 0$, then $\phi(\theta) > 1$. This implies $X_{N\wedge n} \in [0, e^{\theta}]$ and it follows from the bounded convergence theorem that

$$1 = EX(T_1) \quad \text{and hence} \quad e^{-\theta} = E(\phi(\theta)^{-T_1})$$

To convert this into the formula for the generating function we set

$$\phi(\theta) = \frac{e^{\theta} + e^{-\theta}}{2} = 1/s.$$

Letting $x = e^{\theta}$ and doing some algebra, we want $x + x^{-1} = 2/s$ or

$$sx^2 - 2x + s = 0$$

The quadratic equation implies

$$x = \frac{2 \pm \sqrt{4-4s^2}}{2s} = \frac{1 \pm \sqrt{1-s^2}}{s}$$

Since $Es^{T_1} = \sum_{k=1}^{\infty} s^k P(T_1 = k)$, we want the solution that is 0 when $s = 0$, which is $(1 - \sqrt{1-s^2})/s$.

To invert the generating function we we use Newton's binomial formula

$$(1+t)^a = 1 + \binom{a}{1} t + \binom{a}{2} t^2 + \binom{a}{3} t^3 + \cdots$$

where $\binom{x}{r} = x(x-1)\ldots(x-r+1)/r!$. Taking $t = -s^2$ and $a = 1/2$, we have

$$\begin{aligned}
\sqrt{1-s^2} &= 1 - \binom{1/2}{1} s^2 + \binom{1/2}{2} s^4 - \binom{1/2}{3} s^6 + \cdots \\
\frac{1-\sqrt{1-s^2}}{s} &= \binom{1/2}{1} s - \binom{1/2}{2} s^3 + \binom{1/2}{3} s^5 + \cdots
\end{aligned}$$

The coefficient of s^{2n-1} is

$$\begin{aligned}
(-1)^{n-1} \frac{(1/2)(-1/2)\cdots(3-2n)/2}{n!} &= \frac{1 \cdot 3 \cdots (2n-3)}{n!} \cdot 2^{-n} \\
&= \frac{1}{2n-1} \frac{(2n)!}{n!\,n!} 2^{-2n}
\end{aligned}$$

which completes the proof. □

Theorem 4.8.9 (Asymmetric simple random walk) *refers to the special case in which $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q \equiv 1 - p$ with $p \neq q$.*

(a) If $\varphi(y) = \{(1-p)/p\}^y$, then $\varphi(S_n)$ is a martingale.

(b) If we let $T_z = \inf\{n : S_n = z\}$, then for $a < x < b$

$$P_x(T_a < T_b) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(a)} \qquad P_x(T_b < T_a) = \frac{\varphi(x) - \varphi(a)}{\varphi(b) - \varphi(a)} \tag{4.8.2}$$

For the last two parts suppose $1/2 < p < 1$.

(c) If $a < 0$, then $P(\min_n S_n \le a) = P(T_a < \infty) = \{(1-p)/p\}^{-a}$.

(d) If $b > 0$, then $P(T_b < \infty) = 1$ and $ET_b = b/(2p-1)$.

Proof Since S_n and ξ_{n+1} are independent, Example 4.1.7 implies that on $\{S_n = m\}$,

$$\begin{aligned}
E(\varphi(S_{n+1})|\mathcal{F}_n) &= p \cdot \left(\frac{1-p}{p}\right)^{m+1} + (1-p)\left(\frac{1-p}{p}\right)^{m-1} \\
&= \{1 - p + p\} \left(\frac{1-p}{p}\right)^m = \varphi(S_n)
\end{aligned}$$

which proves (a).

Let $N = T_a \wedge T_b$. Since $\varphi(S_{N \wedge n})$ is bounded, Theorem 4.8.2 implies

$$\varphi(x) = E\varphi(S_N) = P_x(T_a < T_b)\varphi(a) + [1 - P_x(T_a < T_a)]\varphi(b)$$

Rearranging gives the formula for $P_x(T_a < T_b)$, subtracting from 1 gives the one for $P_x(T_b < T_a)$.

Letting $b \to \infty$ and noting $\varphi(b) \to 0$ gives the result in (c), since $T_a < \infty$ if and only if $T_a < T_b$ for some b. To start to prove (d) we note that $\varphi(a) \to \infty$ as $a \to -\infty$, so $P(T_b < \infty) = 1$. For the second conclusion, we note that $X_n = S_n - (p-q)n$ is a martingale. Since $T_b \wedge n$ is a bounded stopping time, Theorem 4.4.1 implies

$$0 = E\left(S_{T_b \wedge n} - (p-q)(T_b \wedge n)\right)$$

Now $b \geq S_{T_b \wedge n} \geq \min_m S_m$ and (c) implies $E(\inf_m S_m) > -\infty$, so the dominated convergence theorem implies $ES_{T_b \wedge n} \to ES_{T_b}$ as $n \to \infty$. The monotone convergence theorem implies $E(T_b \wedge n) \uparrow ET_b$, so we have $b = (p-q)ET_b$. □

Example 4.8.10 A gambler is playing roulette and betting \$1 on black each time. The probability she wins \$1 is 18/38, and the probability she loses \$1 is 20/38. Calculate the probability that starting with \$20 she reaches \$40 before losing her money.

$(1-p)/p) = 20/18$, so using (4.8.2) we have

$$\begin{aligned} P_{20}(T_{40} < T_0) &= \frac{(10/9)^{20} - 1}{(10/9)^{40} - 1} \\ &= \frac{8.225 - 1}{67.655 - 1} = 0.1083 \end{aligned}$$

Exercises

4.8.1 Generalize Theorem 4.8.2 to show that if $L \leq M$ are stopping times and $Y_{M \wedge n}$ is a uniformly integrable submartingale, then $EY_L \leq EY_M$ and

$$Y_L \leq E(Y_M | \mathcal{F}_L)$$

4.8.2 If $X_n \geq 0$ is a supermartingale, then $P(\sup X_n > \lambda) \leq EX_0/\lambda$.

4.8.3 Let $S_n = \xi_1 + \cdots + \xi_n$ where the ξ_i are independent with $E\xi_i = 0$ and $\text{var}\,(X_i) = \sigma^2$. $S_n^2 - n\sigma^2$ is a martingale. Let $T = \min\{n : |S_n| > a\}$. Use Theorem 4.8.2 to show that $ET \geq a^2/\sigma^2$.

4.8.4 *Wald's second equation.* Let $S_n = \xi_1 + \cdots + \xi_n$, where the ξ_i are independent with $E\xi_i = 0$ and $\text{var}\,(\xi_i) = \sigma^2$. Use the martingale from the previous problem to show that if T is a stopping time with $ET < \infty$, then $ES_T^2 = \sigma^2 ET$.

4.8.5 *Variance of the time of gambler's ruin.* Let $\xi_1, \xi_2, \ldots$ be independent with $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q = 1 - p$, where $p < 1/2$. Let $S_n = S_0 + \xi_1 + \cdots + \xi_n$ and let $V_0 = \min\{n \geq 0 : S_n = 0\}$. Theorem 4.8.9 tells us that $E_x V_0 = x/(1-2p)$. The aim of this problem is to compute the variance of V_0. If we let $Y_i = \xi_i - (p-q)$ and note that $EY_i = 0$ and

$$\text{var}\,(Y_i) = \text{var}\,(X_i) = EX_i u^2 - (EX_i)^2$$

then it follows that $(S_n - (p-q)n)^2 - n(1-(p-q)^2)$ is a martingale. (a) Use this to conclude that when $S_0 = x$ the variance of V_0 is

$$x \cdot \frac{1 - (p-q)^2}{(q-p)^3}$$

(b) Why must the answer in (a) be of the form cx?

4.8.6 *Generating function of the time of gambler's ruin.* Continue with the set-up of the previous problem. (a) Use the exponential martingale and our stopping theorem to conclude that if $\theta \le 0$, then $e^{\theta x} = E_x(\phi(\theta)^{-V_0})$. (b) Let $0 < s < 1$. Solve the equation $\phi(\theta) = 1/s$, then use (a) to conclude

$$E_x(s^{V_0}) = \left(\frac{1-\sqrt{1-4pqs^2}}{2ps}\right)^x$$

(c) Why must the answer in (b) be of the form $f(s)^x$?

4.8.7 Let S_n be a symmetric simple random walk starting at 0, and let $T = \inf\{n : S_n \notin (-a,a)\}$, where a is an integer. Find constants b and c so that $Y_n = S_n^4 - 6nS_n^2 + bn^2 + cn$ is a martingale, and use this to compute ET^2.

4.8.8 Let $S_n = \xi_1 + \cdots + \xi_n$ be a random walk. Suppose $\varphi(\theta_o) = E\exp(\theta_o\xi_1) = 1$ for some $\theta_o < 0$ and ξ_i is not constant. In this special case of the exponetial martingale $X_n = \exp(\theta_o S_n)$ is a martingale. Let $\tau = \inf\{n : S_n \notin (a,b)\}$ and $Y_n = X_{n\wedge T}$. Use Theorem 4.8.2 to conclude that $EX_\tau = 1$ and $P(S_\tau \le a) \le \exp(-\theta_o a)$.

4.8.9 Continuing with the set-up of the previous problem, suppose the ξ_i are integer valued with $P(\xi_i < -1) = 0$, $P(\xi_i = -1) > 0$, and $E\xi_i > 0$. Let $T = \inf\{n : S_n = a\}$ with $a < 0$. Use the martingale $X_n = \exp(\theta_o S_n)$ to conclude that $P(T < \infty) = \exp(-\theta_o a)$.

4.8.10 Consider a favorable game in which the payoffs are -1, 1, or 2 with probability 1/3 each. Use the results of the previous problem to compute the probability we ever go broke (i.e, our winnings W_n reach \$0) when we start with \$$i$.

4.8.11 Let S_n be the total assets of an insurance company at the end of year n. In year n, premiums totaling $c > 0$ are received and claims ζ_n are paid, where ζ_n is Normal(μ, σ^2) and $\mu < c$. To be precise, if $\xi_n = c - \zeta_n$, then $S_n = S_{n-1} + \xi_n$. The company is ruined if its assets drop to 0 or less. Show that if $S_0 > 0$ is nonrandom, then

$$P(\text{ ruin }) \le \exp(-2(c-\mu)S_0/\sigma^2)$$

4.9 Combinatorics of Simple Random Walk*

In the last section we proved some results for simple random walks using martingales. In this section we will delve deeper into their properties using combinatorial arguments. We will not be using martingales, but this section is a good transition to the study of Markov chains in the next chapter.

To facilitate discussion, we will think of the sequence $S_0, S_1, S_2, \ldots, S_n$ as being represented by a polygonal line with segments $(k-1, S_{k-1}) \to (k, S_k)$. A **path** is a polygonal line that is a possible outcome of simple random walk. To count the number of paths from (0,0) to (n,x), it is convenient to introduce a and b defined by: $a = (n+x)/2$ is the number of positive steps in the path and $b = (n-x)/2$ is the number of negative steps. Notice that

$n = a + b$ and $x = a - b$. If $-n \le x \le n$ and $n - x$ is even, the a and b defined above are nonnegative integers, and the number of paths from (0,0) to (n, x) is

$$N_{n,x} = \binom{n}{a} \tag{4.9.1}$$

Otherwise, the number of paths is 0.

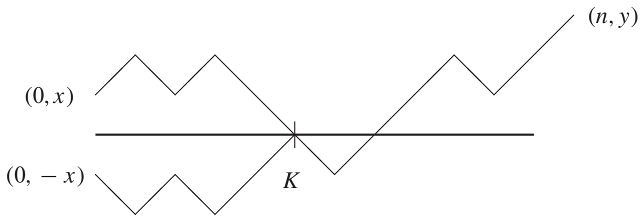

Figure 4.5 Reflection Principle

Theorem 4.9.1 (Reflection principle) *If $x, y > 0$, then the number of paths from $(0, x)$ to (n, y) that are 0 at some time is equal to the number of paths from $(0, -x)$ to (n, y).*

Proof Suppose $(0, s_0), (1, s_1), \ldots, (n, s_n)$ is a path from $(0, x)$ to (n, y). Let $K = \inf\{k : s_k = 0\}$. Let $s'_k = -s_k$ for $k \le K$, $s'_k = s_k$ for $K \le k \le n$. Then (k, s'_k), $0 \le k \le n$, is a path from $(0, -x)$ to (n, y). Conversely, if $(0, t_0), (1, t_1), \ldots, (n, t_n)$ is a path from $(0, -x)$ to (n, y), then it must cross 0. Let $K = \inf\{k : t_k = 0\}$. Let $t'_k = -t_k$ for $k \le K$, $t'_k = t_k$ for $K \le k \le n$. Then (k, t'_k), $0 \le k \le n$, is a path from $(0, -x)$ to (n, y) that is 0 at time K. The last two observations set up a one-to-one correspondence between the two classes of paths, so their numbers must be equal. □

From Theorem 4.9.1 we get a result first proved in 1878.

Theorem 4.9.2 (The Ballot Theorem) *Suppose that in an election candidate A gets α votes, and candidate B gets β votes where $\beta < \alpha$. The probability that throughout the counting A always leads B is $(\alpha - \beta)/(\alpha + \beta)$.*

Proof Let $x = \alpha - \beta$, $n = \alpha + \beta$. Clearly, there are as many such outcomes as there are paths from (1,1) to (n, x) that are never 0. The reflection principle implies that the number of paths from (1,1) to (n, x) that are 0 at some time the number of paths from (1,-1) to (n, x), so by (4.9.1) the number of paths from (1,1) to (n, x) that are never 0 is

$$\begin{aligned} N_{n-1,x-1} - N_{n-1,x+1} &= \binom{n-1}{\alpha-1} - \binom{n-1}{\alpha} \\ &= \frac{(n-1)!}{(\alpha-1)!\,(n-\alpha)!} - \frac{(n-1)!}{\alpha!\,(n-\alpha-1)!} \\ &= \frac{\alpha - (n-\alpha)}{n} \cdot \frac{n!}{\alpha!\,(n-\alpha)!} = \frac{\alpha-\beta}{\alpha+\beta} N_{n,x} \end{aligned}$$

since $n = \alpha + \beta$, this proves the desired result. □

Using the ballot theorem, we can compute the distribution of the time to hit 0 for simple random walk.

Lemma 4.9.3 $P(S_1 \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2n} = 0)$.

Proof $P(S_1 > 0, \ldots, S_{2n} > 0) = \sum_{r=1}^{\infty} P(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r)$. From the proof of Theorem 4.9.2, we see that the number of paths from (0,0) to $(2n, 2r)$ that are never 0 at positive times (= the number of paths from (1,1) to $(2n, 2r)$ that are never 0) is

$$N_{2n-1,2r-1} - N_{2n-1,2r+1}$$

If we let $p_{n,x} = P(S_n = x)$, then this implies

$$P(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r) = \frac{1}{2}(p_{2n-1,2r-1} - p_{2n-1,2r+1})$$

Summing from $r = 1$ to ∞ gives

$$P(S_1 > 0, \ldots, S_{2n} > 0) = \frac{1}{2} p_{2n-1,1} = \frac{1}{2} P(S_{2n} = 0)$$

Symmetry implies $P(S_1 < 0, \ldots, S_{2n} < 0) = (1/2)P(S_{2n} = 0)$, and the proof is complete. □

Let $R = \inf\{m \geq 1 : S_m = 0\}$. Combining Lemma 4.9.3 with the central limit theorem for the binomial distribution, Theorem 3.1.2, gives

$$P(R > 2n) = P(S_{2n} = 0) \sim \pi^{-1/2} n^{-1/2} \tag{4.9.2}$$

Since $P(R > x)/\, P(|R| > x) = 1$, it follows from Theorem 3.8.8 that R is in the domain of attraction of the stable law with $\alpha = 1/2$ and $\kappa = 1$. This implies that if R_n is the time of the nth return to 0, then $R_n/n^2 \Rightarrow Y$, the indicated stable law. In Example 3.8.5, we considered $\tau = T_1$, where $T_x = \inf\{n : S_n = x\}$. Since $S_1 \in \{-1, 1\}$ and $T_1 =_d T_{-1}$, $R =_d 1 + T_1$, and it follows that $T_n/n^2 \Rightarrow Y$, the same stable law. In Example 8.1.12 we will use this observation to show that the limit has the same distribution as the hitting time of 1 for Brownian motion, which has a density given in (7.4.6).

From (4.9.2) we get

$$\begin{aligned}
P(T_1 = 2n - 1) = P(R = 2n) &= P(R > 2n - 2) - P(R > 2n) \\
&= \frac{(2n-2)!}{(n-1)!\,(n-1)!} 2^{-2(n-1)} - \frac{(2n)!}{n!\,n!} 2^{-2n} \\
&= \frac{(2n)!}{n!\,n!} 2^{-2n} \left(\frac{n \cdot n}{(2n-1) \cdot 2n} \cdot 4 - 1 \right) \\
&= \frac{1}{2n-1} \frac{(2n)!}{n!\,n!} 2^{-2n} \sim \frac{1}{2} \pi^{-1/2} n^{-3/2}
\end{aligned}$$

This completes our discussion of visits to 0. We turn now to the arcsine laws. The first one concerns

$$L_{2n} = \sup\{m \leq 2n : S_m = 0\}$$

It is remarkably easy to compute the distribution of L_{2n}.

Lemma 4.9.4 *Let* $u_{2m} = P(S_{2m} = 0)$. *Then* $P(L_{2n} = 2k) = u_{2k} u_{2n-2k}$.

Proof $P(L_{2n} = 2k) = P(S_{2k} = 0, S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0)$, so the desired result follows from Lemma 4.9.3. □

Theorem 4.9.5 (Arcsine law for the last visit to 0) *For* $0 < a < b < 1$,

$$P(a \le L_{2n}/2n \le b) \to \int_a^b \pi^{-1}(x(1-x))^{-1/2}\,dx$$

To see the reason for the name, substitute $y = x^{1/2}$, $dy = (1/2)x^{-1/2}\,dx$ in the integral to obtain

$$\int_{\sqrt{a}}^{\sqrt{b}} \frac{2}{\pi}(1-y^2)^{-1/2}\,dy = \frac{2}{\pi}\{\arcsin(\sqrt{b}) - \arcsin(\sqrt{a})\}$$

Since L_{2n} is the time of the last zero before $2n$, it is surprising that the answer is symmetric about 1/2. The symmetry of the limit distribution implies

$$P(L_{2n}/2n \le 1/2) \to 1/2$$

In gambling terms, if two people were to bet \$1 on a coin flip every day of the year, then with probability $1/2$, one of the players will be ahead from July 1 to the end of the year, an event that would undoubtedly cause the other player to complain about his bad luck.

Proof of Theorem 4.9.5. From the asymptotic formula for u_{2n}, it follows that if $k/n \to x$, then

$$nP(L_{2n} = 2k) \to \pi^{-1}(x(1-x))^{-1/2}$$

To get from this to the desired result, we let $2na_n =$ the smallest even integer $\ge 2na$, let $2nb_n =$ the largest even integer $\le 2nb$, and let $f_n(x) = nP(L_{2n} = k)$ for $2k/2n \le x < 2(k+1)/2n$, so we can write

$$P(a \le L_{2n}/2n \le b) = \sum_{k=na_n}^{nb_n} P(L_{2n} = 2k) = \int_{a_n}^{b_n+1/n} f_n(x)\,dx$$

Our first result implies that uniformly on compact sets

$$f_n(x) \to f(x) = \pi^{-1}(x(1-x))^{-1/2}$$

The uniformity of the convergence implies

$$\sup_{a_n \le x \le b_n+1/n} f_n(x) \to \sup_{a \le x \le b} f(x) < \infty$$

if $0 < a \le b < 1$, so the bounded convergence theorem gives

$$\int_{a_n}^{b_n+1/n} f_n(x)\,dx \to \int_a^b f(x)\,dx \qquad \square$$

The next result deals directly with the amount of time one player is ahead.

Theorem 4.9.6 (Arcsine law for time above 0) *Let* π_{2n} *be the number of segments* $(k-1, S_{k-1}) \to (k, S_k)$ *that lie above the axis (i.e., in* $\{(x,y) : y \ge 0\}$*), and let* $u_m = P(S_m = 0)$.

$$P(\pi_{2n} = 2k) = u_{2k}u_{2n-2k}$$

and consequently, if $0 < a < b < 1$

$$P(a \le \pi_{2n}/2n \le b) \to \int_a^b \pi^{-1}(x(1-x))^{-1/2}\,dx$$

Remark Since $\pi_{2n} =_d L_{2n}$, the second conclusion follows from the proof of Theorem 4.9.5. The reader should note that the limiting density $\pi^{-1}(x(1-x))^{-1/2}$ has a minimum at $x = 1/2$, and $\to \infty$ as $x \to 0$ or 1. An equal division of steps between the positive and negative side is therefore the least likely possibility, and completely one-sided divisions have the highest probability.

Proof Let $\beta_{2k,2n}$ denote the probability of interest. We will prove $\beta_{2k,2n} = u_{2k}u_{2n-2k}$ by induction. When $n = 1$, it is clear that

$$\beta_{0,2} = \beta_{2,2} = 1/2 = u_0 u_2$$

For a general n, first suppose $k = n$. From the proof of Lemma 4.9.3, we have

$$\begin{aligned}
\frac{1}{2}u_{2n} &= P(S_1 > 0, \ldots, S_{2n} > 0) \\
&= P(S_1 = 1, S_2 - S_1 \geq 0, \ldots, S_{2n} - S_1 \geq 0) \\
&= \frac{1}{2}P(S_1 \geq 0, \ldots, S_{2n-1} \geq 0) \\
&= \frac{1}{2}P(S_1 \geq 0, \ldots, S_{2n} \geq 0) = \frac{1}{2}\beta_{2n,2n}
\end{aligned}$$

The next to last equality follows from the observation that if $S_{2n-1} \geq 0$, then $S_{2n-1} \geq 1$, and hence $S_{2n} \geq 0$.

The last computation proves the result for $k = n$. Since $\beta_{0,2n} = \beta_{2n,2n}$, the result is also true when $k = 0$. Suppose now that $1 \leq k \leq n-1$. In this case, if R is the time of the first return to 0, then $R = 2m$ for some m with $0 < m < n$. Letting $f_{2m} = P(R = 2m)$ and breaking things up according to whether the first excursion was on the positive or negative side gives

$$\beta_{2k,2n} = \frac{1}{2}\sum_{m=1}^{k} f_{2m}\beta_{2k-2m,2n-2m} + \frac{1}{2}\sum_{m=1}^{n-k} f_{2m}\beta_{2k,2n-2m}$$

Using the induction hypothesis, it follows that

$$\beta_{2k,2n} = \frac{1}{2}u_{2n-2k}\sum_{m=1}^{k} f_{2m}u_{2k-2m} + \frac{1}{2}u_{2k}\sum_{m=1}^{n-k} f_{2m}u_{2n-2k-2m}$$

By considering the time of the first return to 0, we see

$$u_{2k} = \sum_{m=1}^{k} f_{2m}u_{2k-2m} \qquad u_{2n-2k} = \sum_{m=1}^{n-k} f_{2m}u_{2n-2k-2m}$$

and the desired result follows. □

Exercises

4.9.1 Let $a \in S$, $f_n = P_a(T_a = n)$, and $u_n = P_a(X_n = a)$. (i) Show that $u_n = \sum_{1\leq m\leq n} f_m u_{n-m}$. (ii) Let $u(s) = \sum_{n\geq 0} u_n s^n$, $f(s) = \sum_{n\geq 1} f_n s^n$, and show $u(s) = 1/(1 - f(s))$.