Section 10: Martingales

Contents

10.1 Martingales in Discrete Time

A fundamental tool in the analysis of DTMC's and continuous-time Markov processes is the notion of a martingale. Martingales also underlie the definition we will adopt for defining stochastic integrals with respect to Brownian motion. A martingale is basically a real-valued sequence that is a suitable generalization of a random walk with independent, mean-zero increments.

Definition 10.1.1 Let $(M_n : n \geq 0)$ be a sequence of real-valued random variables. Then, $(M_n : n \geq 0)$ is said to be a martingale (with respect to the sequence of random elements $(Z_n : n \geq 0)$ if:

(i) $\mathrm{E}|M_n| < \infty$ for $n \geq 0$;

(ii) for each $n \geq 0$, there exists a deterministic function $g_n(\cdot)$ such that

$$M_n = g_n(Z_0, Z_1, \ldots, Z_n);$$

(iii) $\mathrm{E}[M_{n+1}|Z_0, Z_1, \ldots, Z_n] = M_n$, for $n \geq 0$.

Remark 10.1.1 When a process $(M_n : n \geq 0)$ satisfies condition (ii), one says that $(M_n : n \geq 0)$ is adapted to $(Z_n : n \geq 0)$.

The critical component of the martingale definition is condition (iii). If we view M_n as the fortune of a gambler at time n, then condition (iii) is asserting that the gambler is involved in playing a "fair game", in which he/she has no propensity (in expectation) to either win or lose on any given gamble. As we asserted earlier, a random walk with independent mean-zero increments is a martingale. To see this, let $S_0, X_1, X_2, \ldots$ be independent random variables with finite mean, and suppose that $\mathrm{E}X_i = 0$ for $i \geq 1$. Set $Z_n = S_n = S_0 + X_1 + \cdots + X_n$. Then, conditions (i) and (ii) of Definition 10.1.1 are trivial to verify. For condition (iii), observe that

$$\mathrm{E}[S_{n+1}|S_0, \ldots, S_n] = \mathrm{E}[S_n + X_{n+1}|S_0, \ldots, S_n] = S_n + \mathrm{E}[X_{n+1}|S_0, \ldots, S_n] = S_n + \mathrm{E}X_{n+1} = S_n.$$

Martingales inherit many of the properties of mean-zero random walks. In view of the analogy with random walks, it is natural to consider the increments

$$D_i = M_i - M_{i-1}, \quad i \geq 1$$

namely, the *martingale differences*. The following proposition is a clear generalization of two of the most important properties of mean-zero random walks.

Proposition 10.1.1 Let $(M_n : n \geq 0)$ be a martingale with respect to $(Z_n : n \geq 0)$. Then,

$$\mathrm{E}M_n = \mathrm{E}M_0 \quad n \geq 0. \tag{10.1.1}$$

In addition, if $\mathrm{E}M_n^2 < \infty$ for $n \geq 0$, then

$$\mathrm{Cov}(D_i, D_j) = 0, \quad i \neq j \tag{10.1.2}$$

so that

$$\mathrm{Var}[M_n] = \mathrm{Var}[M_0] + \sum_{i=1}^{n} \mathrm{Var}[D_i]. \tag{10.1.3}$$

Proof: Relation (10.1.1) is immediate from condition (iii) of the martingale definition. For (10.1.2), note that (10.1.1) implies that $\mathrm{E}D_i = 0$, so that (10.1.2) is equivalent to asserting that $\mathrm{E}D_iD_j = 0$ for $i < j$. But

$$\mathrm{E}[D_iD_j|Z_0, \ldots, Z_{j-1}] = D_i\mathrm{E}[D_j|Z_0, \ldots, Z_{j-1}] = 0,$$

where condition (ii) of the martingale definition was used for the first equality, and condition (iii) was used for the final step. Taking expectations with respect to $(Z_0, \ldots, Z_{j-1})$, we get (10.1.2). Finally, (10.1.3) is immediate from (10.1.2). $\square$

Definition 10.1.2 A martingale $(M_n : n \geq 0)$ for which $\mathrm{E}M_n^2 < \infty$ for $n \geq 0$ is called a *square-integrable martingale.*

Before we turn to exploring further properties of martingales, let us develop some additional examples of martingales in the random walk setting.

Example 10.1.1 Let $(X_n : n \geq 1)$ be a sequence of iid mean-zero random variables with finite variance σ^2. Let $S_n = X_1 + \cdots + X_n$ and let

$$M_n = S_n^2 - n\sigma^2.$$

Then $(M_n : n \geq 0)$ is a martingale with respect to $(S_n : n \geq 0)$. The critical property to verify is (iii). Note that

$$\begin{aligned}
\mathrm{E}[M_{n+1}|S_0, \ldots, S_n] &= \mathrm{E}[(S_n + X_{n+1})^2 - (n+1)\sigma^2|S_0, \ldots, S_n] \\
&= \mathrm{E}[S_n^2 + 2S_nX_{n+1} + X_{n+1}^2 - (n+1)^2\sigma^2|S_0, \ldots, S_n] \\
&= S_n^2 + 2S_n\mathrm{E}[X_{n+1}|S_0, \ldots, S_n] + \mathrm{E}[X_{n+1}^2|S_0, \ldots, S_n] - (n+1)^2\sigma^2 \\
&= S_n^2 + \sigma^2 - (n+1)^2\sigma^2 \\
&= M_n.
\end{aligned}$$

Example 10.1.2 Let $(X_n : n \geq 1)$ be a sequence of iid random variables with common density g. Suppose that f is another density with the property that whenever $g(x) = 0$, then $f(x) = 0$. Set $L_0 = 1$ and

$$L_n = \prod_{i=1}^{n} \frac{f(X_i)}{g(X_i)}, \quad n \geq 1$$

Then, $(L_n : n \geq 0)$ is a martingale with respect to $(X_n : n \geq 1)$. Again, the critical property is verifying (iii). Here,

$$\mathrm{E}[L_{n+1}|X_1,\ldots,X_n] = \mathrm{E}\left[L_n \frac{f(X_{n+1})}{g(X_{n+1})}\middle| X_1,\ldots,X_n\right] = L_n \mathrm{E}\left[L_n \frac{f(X_{n+1})}{g(X_{n+1})}\right] = L_n \int \frac{f(x)}{g(x)} g(x)\,dx = L_n,$$

since f is a density that integrates to 1. This is known as a *likelihood ratio martingale.*

To show why the likelihood ratio martingale arises naturally, suppose that we have observed an iid sample from a population, yielding observations $X_1, X_2, \ldots, X_n$. Assume that the underlying population is known to be iid, either with common density f or with common density g. To test the hypothesis that the X_i's have common density f (the "f-hypothesis") against the hypothesis that the X_i's have common density g (the "g-hypothesis"), the Neyman-Pearson lemma asserts that we should accept the "f-hypothesis" if the relative likelihood

$$\frac{f(X_1)\cdots f(X_n)}{g(X_1)\cdots g(X_n)} \tag{10.1.4}$$

is sufficiently large, and reject it otherwise. So, studying L_n in the case where the X_i's have common density g corresponds to studying the test statistic (10.1.4) when the "state of nature" is that the "g-hypothesis" is true. Given this interpretation, it seems natural to expect that L_n converges to zero as the sample size n goes to positive infinity. This is because for a large sample size n, it is extremely unlikely that such a sample will be better explained by the "f-hypothesis" than by the other one. The fact that L_n ought to go to zero as $n \to \infty$ is perhaps a bit surprising, given that $\mathrm{E}L_n = 1$ for $n \geq 0$.

To prove that $L_n \to 0$ almost surely as $n \to \infty$, note that

$$\log L_n = \sum_{i=1}^{n} \log\left(\frac{f(X_i)}{g(X_i)}\right).$$

Then, the strong law of large numbers guarantees that

$$\frac{1}{n}\log L_n \to \mathrm{E}\log\left(\frac{f(X_i)}{g(X_i)}\right) \quad \text{a.s. as } n \to \infty.$$

In other words,

$$\frac{1}{n}\log L_n \to \int \log\left(\frac{f(x)}{g(x)}\right) g(x)\,dx. \tag{10.1.5}$$

(The right-hand side of (10.1.5) is what is known as a *relative entropy.*) Since log is strictly concave, Jensens inequality asserts that if $f \neq g$,

$$\mathrm{E}\log\left(\frac{f(X_i)}{g(X_i)}\right) < \log \mathrm{E}\left(\frac{f(X_i)}{g(X_i)}\right) = 0 \tag{10.1.6}$$

As a consequence, not only does L_n converge to zero as $n \to \infty$ a.s, but the rate of convergence is exponentially fast. It is worth noting that this is an example of a sequence of random variables $(L_n : n \geq 0)$ for which $L_n \to 0$ a.s. and yet $\mathrm{E}L_n \nrightarrow 0$ as $n \to \infty$ (in other words, passing limits through expectations is not always valid).

Example 10.1.3 In this example, we specialize the likelihood ratio martingale a bit. Suppose that the X_i's are iid with common density g, and suppose that the moment generating function $m_X(\theta) = \mathrm{E}e^{\theta X_i}$ converges in some neighborhood of the origin. For θ within the domain of convergence of $m_X(\cdot)$, let

$$f(x) = \frac{e^{\theta x} g(x)}{m_X(\theta)},$$

or, equivalently, $f(x) = e^{\theta x - \psi(\theta)} g(x)$, where $\psi(\theta) = \log m_X(\theta)$. In this case,

$$L_n = \prod_{i=1}^{n} \frac{f(X_i)}{g(X_i)} = e^{\theta S_n - n\psi(\theta)} \tag{10.1.7}$$

The martingale $(L_n : n \geq 0)$ defined by (10.1.7) is known as an *exponential martingale.* Because the random walk $(S_n : n \geq 0)$ appears explicitly in the exponent of the martingale, $(L_n : n \geq 0)$ is well-suited to studying random walks.

Some indication of the power of this martingale should be apparent, if we explicitly display the dependence of L_n on θ as follows:

$$L_n(\theta) = e^{\theta S_n - n\psi(\theta)}$$

The defining property (iii) of a martingale asserts that

$$\mathrm{E}[L_{n+1}(\theta)|S_0, \ldots, S_n] = L_n(\theta).$$

For θ inside the domain of convergence of $m_X(\cdot)$, one can interchange the derivative and expectation, yielding

$$\mathrm{E}[L'_{n+1}(\theta)|S_0, \ldots, S_n] = L'_n(\theta).$$

In particular, $(L'_n(0) : n \geq 0)$ is a martingale. But

$$L'_n(0) = S_n - n\psi'(0).$$

It turns out that $\psi'(0) = \mathrm{E}X_1$. So, by differentiating our exponential martingale, we retrieve the random walk martingale. And by differentiating a second time, it turns out that $L''_n(0)$ is the martingale of Example 10.1.2. Through successive differentiation, we can obtain a whole infinite family of such martingales.

Exercise 10.1.1 (a) Prove that $\psi(\cdot)$ is convex.

(b) Prove that $\psi'(\cdot) = \mathrm{E}X_1$.

(c) Prove that $\psi''(0) = \mathrm{Var}[X_1]$.

(d) Prove that $L''_n(0) = (S_n - n\mu)^2 - n\sigma^2$.

(e) Compute $L'''_n(0)$.

We now turn to a fundamental result in the theory of martingales known as the *Martingale Convergence Theorem.*

Theorem 10.1.1 (Martingale Convergence Theorem in L^2) Let $(M_n : n \geq 0)$ be a martingale with respect to $(Z_n : n \geq 0)$. If $\sup_{n \geq 0} \mathrm{E}M_n^2 < \infty$, then there exists a square-integrable random variable M_∞ such that

$$\mathrm{E}[(M_n - M_0)^2] \to 0$$

as $n \to \infty$, i.e. M_n converges to M_∞ in mean square.

Proof: The space L^2 of square-integrable random variables is a Hilbert space under the inner product $\langle X, Y\rangle = \mathrm{E}[XY]$. Since

$$\mathrm{E}M_n^2 = \mathrm{E}M_0^2 + \sum_{i=1}^{n} \mathrm{E}D_i^2,$$

it follows that $\sum_{i=1}^{\infty} \mathrm{E}D_i^2 < \infty$. For $\epsilon > 0$, choose $m = m(\epsilon)$ so that $\sum_{i=m}^{\infty} \mathrm{E}D_i^2 < \epsilon$. Then, for $n_2 > n_1 \geq m$,

$$\mathrm{E}(M_{n_2} - M_{n_1})^2 = \sum_{j=n_1+1}^{n_2} \mathrm{E}D_j^2 < \epsilon$$

so that $(M_n : n \geq 0)$ is a Cauchy sequence in L^2. Then, the completeness of L^2 yields the conclusion of the theorem. □

Actually, one does not need square integrability in order that the Martingale Convergence Theorem hold.

Theorem 10.1.2 (Martingale Convergence Theorem) Let $(M_n : n \geq 0)$ be a martingale with respect to $(Z_n : n \geq 0)$. If $\sup_{n\geq 0} \mathrm{E}|M_n| < \infty$, then there exists a finite-valued random variable M_∞ such that $M_n \to M_\infty$ a.s. as $n \to \infty$.

For a proof, see p. 233 of "Probability: Theory and Examples" 3rd ed. by R. Durrett.

We conclude this section with a brief discussion of stochastic integrals in discrete time. Let $(M)n : n \geq 0)$ be a square-integrable martingale with respect to $(Z_n : n \geq 0)$. Suppose that $(W_n : n \geq 0)$ is a sequence of random variables that is adapted to $(Z_n : n \geq 0)$. We define the stochastic integral of $(W_n : n \geq 0)$ with respect to $(M_n : n \geq 0)$ as the sequence

$$V_n = \sum_{i=1}^{n} W_{i-1}D_i = \sum_{i=1}^{n} W_{i-1}(M_i - M_{i-1}) = \sum_{i=1}^{n} W_{i-1}\Delta M_i$$

We could also have defined the stochastic integral here as $\sum_{i=1}^{n} W_i \Delta M_i$. But in that case, we would lose the nice properties listed below.

Exercise 10.1.2 Let $(M_n : n \geq 0)$ be a square-integrable martingale with respect to $(Z_n : n \geq 0)$, with $M_0 = 0$. Suppose $(W_n : n \geq 0)$ is a square-integrable sequence that is adapted to $(Z_n : n \geq 0)$.

(a) Prove that if $V_0 = 0$ and $V_n = \sum_{i=1}^{n} W_{i-1}\Delta M_i$ for $n \geq 1$, then $(V_n : n \geq 0)$ is a martingale with respect to $(Z_n : n \geq 0)$.

(b) Suppose that the martingale differences $(D_i : i \geq 1)$ are a stationary sequence of independent random variables. Show that $\mathrm{E}V_n^2 = \sigma^2 \sum_{i=0}^{n-1} \mathrm{E}W_i^2$, where $\sigma^2 = \mathrm{Var}[D_i]$.

10.2 Optional Sampling for Discrete-Time Martingales

An important property of martingales is

$$\mathrm{E}M_n = \mathrm{E}M_0 \quad n \geq 0 \tag{10.2.1}$$

The theory of "optional sampling" is concerned with extending (10.2.1) from deterministic n to random times T. As in the discussion of the strong Markov property, it is natural to restrict ourselves to stopping times. However,

$$\mathrm{E}M_T = \mathrm{E}M_0 \tag{10.2.2}$$

fails to hold for all finite-valued stopping times T.

Example 10.2.1 Let $(S_n : n \geq 0)$ be a random walk with $S_0 = 0$ and iid increments $(X_n : n \geq 1)$ defined by

$$P(X_n = 1) = P(X_n = -1) = \frac{1}{2}.$$

Put $T = \inf\{n \geq 0 : S_n = 1\}$. Since $(S_n : n \geq 0)$ is null recurrent, $T < \infty$ a.s. and $S_T = 1$. Therefore, $ES_T = 1$ and $ES_0 = 0$, and so $ES_T \neq ES_0$. Hence, the class of stopping times needs to be restricted somewhat.

Theorem 10.2.1 Let $(M_n : n \geq 0)$ be a martingale with respect to $(Z_n : n \geq 0)$. Suppose that T is a bounded random variable that is a stopping time with respect to $(Z_n : n \geq 0)$. Then $EM_T = EM_0$.

Proof: Let m be such that $P(T \leq m) = 1$. Then $M_T = M_0 + \sum_{i=1}^m D_i \mathbb{I}(T \geq i)$, and thus

$$EM_T = EM_0 + E\sum_{i=1}^m D_i \mathbb{I}(T \geq i) \tag{10.2.3}$$

Because T is a stopping time, $E[D_i \mathbb{I}(T \geq i)|Z_0, \ldots, Z_{i-1}] = \mathbb{I}(T \geq i)E[D_i|Z_0, \ldots, Z_{i-1}] = 0$, and so

$$E\sum_{i=1}^m D_i \mathbb{I}(T \geq i) = 0 \qquad \square$$

If T is a stopping time, then $T \wedge n$ is a stopping time for $n \geq 0$ (and is clearly bounded). So, optional sampling applies at $T \wedge n$ (see Theorem 10.2.1), i.e. $EM_{T\wedge n} = EM_0$ for $n \geq 0$.

If $T < \infty$ a.s., then $M_{T\wedge n} \to M_T$ a.s. as $n \to \infty$. Hence, if

$$E \lim_{n\to\infty} M_{T\wedge n} = \lim_{n\to\infty} EM_{T\wedge n}, \tag{10.2.4}$$

then (10.2.2) holds, since

$$EM_T = E \lim_{n\to\infty} M_{T\wedge n} = \lim_{n\to\infty} EM_{T\wedge n} = \lim_{n\to\infty} EM_0 = EM_0.$$

Therefore, the key to establishing (10.2.2) is (10.2.4). There are various results which one can invoke to justify (10.2.4); the most powerful of these results is the Dominated Convergence Theorem. To apply this result, we need to find a random variable W having finite mean, such that $|M_{T\wedge n}| \leq W$ for $n \geq 0$. The obvious candidate for W is

$$W = |M_0| + \sum_{i=1}^T \widetilde{D}_i, \tag{10.2.5}$$

where $\widetilde{D}_i = |D_i|$. So, if $EW < \infty$, we conclude that (10.2.4) is valid.

Proposition 10.2.1 Suppose that there exists $c < \infty$ such that $P(D_i \leq c) = 1$ for $i \geq 1$. If $ET < \infty$, then

$$EM_T = EM_0.$$

Proof: Note that $W \le |M_0| + cT$. Since $\mathrm{E}T < \infty$, then $\mathrm{E}W < \infty$. Then, the Dominated Convergence Theorem implies that $\mathrm{E}M_{T\wedge n} \to \mathrm{E}M_T$ as $n \to \infty$, yielding the result.

Now, let's turn to an application of optional sampling.

Application 10.2.1 Let $(S_n : n \ge 0)$ be a random walk with $S_0 = 0$ and iid increments $(X_n : n \ge 1)$ defined by

$$\mathrm{P}(X_n = 1) = \mathrm{P}(X_n = -1) = \frac{1}{2}.$$

Let $T = \inf\{n \ge 0 : S_n \le -a \text{ or } S_n \ge b\}$ be the "exit time"from $[-a, b]$. Suppose that we wish to compute for $\mathrm{P}(S_T = -a)$, the probability that the random walk exits the left boundary. (This is basically the "gambler's ruin" computation for the probability of ruin.) Note that $\widetilde{D}_i = 1$ and $\mathrm{E}T < \infty$ (see Exercise 10.2.1). Hence, Proposition 10.2.1 applies and $\mathrm{E}S_T = 0$. However,

$$\mathrm{E}S_T = -a\mathrm{P}(S_T = -a) + b\mathrm{P}(S_T = b) = -a\mathrm{P}(S_T = -a) + b[1 - \mathrm{P}(S_T = -a)].$$

Therefore, $\mathrm{P}(S_T = -a) = b/(a + b)$.

Exercise 10.2.1 (a) Prove that $\mathrm{E}T < \infty$ in Application 10.2.1.

(b) Compute the value of $\mathrm{P}(S_T = -a)$ by setting up a suitable system of linear equations involving the unknowns $\mathrm{P}_x(S_T = -a)$ and solving them. (This is an alternative approach to computing the "exit" probability.)

Application 10.2.2 In this continuation of Application 10.2.1, we wish to compute $\mathrm{E}T$ (In the gambler's ruin setting, this is the mean duration of the game). Let $M_n = S_n^2 - n\sigma^2$, where $\sigma^2 = \mathrm{Var}X_i = 1$. Assuming that (10.2.2) holds,

$$\mathrm{E}S_T^2 = \sigma^2 \mathrm{E}T = \mathrm{E}T. \tag{10.2.6}$$

Solving for $\mathrm{E}S_T^2$, we have

$$\mathrm{E}S_T^2 = a^2\mathrm{P}(S_T = -a) + b^2\mathrm{P}(S_T = -b) = \frac{a^2b + ab^2}{a + b} = ab,$$

so $\mathrm{E}T = ab$. Does Proposition 10.2.1 apply? Here,

$$\widetilde{D}_i = S_i^2 - S_{i-1}^2 = (S_i + S_{i-1})X_i - 1.$$

Clearly, the $\widetilde{D}_i$ do not satisfy the hypotheses of Proposition 10.2.1, so something else is needed here.

Proposition 10.2.2 Suppose that there exists $c < \infty$ for which

$$\mathrm{E}[\widetilde{D}_i | Z_0, Z_1, \dots, Z_{i-1}] \le c \text{ on } \{T \ge i\} \text{ for } i \ge 1.$$

If $\mathrm{E}T < \infty$, then $\mathrm{E}M_T = \mathrm{E}M_0$.

Proof: Note that $\mathrm{E}W = \mathrm{E}|M_0| + \mathrm{E}\sum_{i=1}^{\infty} \widetilde{D}_i \mathbb{I}(T \ge i) = \mathrm{E}|M_0| + \sum_{i=1}^{\infty} \mathrm{E}\widetilde{D}_i \mathbb{I}(T \ge i)$. However,

$$\mathrm{E}[\widetilde{D}_i \mathbb{I}(T \ge i) | Z_0, Z_1, \dots, Z_{i-1}] = \mathbb{I}(T \ge i)\mathrm{E}[\widetilde{D}_i | Z_0, Z_1, \dots, Z_{i-1}] \le c\mathbb{I}(T \ge i).$$

Thus, $\mathrm{E}W \le \mathrm{E}|M_0| + c\sum_{i=1}^{\infty} \mathrm{E}\mathbb{I}(T \ge i) = \mathrm{E}|M0| + c\mathrm{E}T < \infty$, and consequently the Dominated Convergence Theorem applies. □

Application 5.2.2 (continued) Here,$\widetilde{D}_i \leq (|S_i|+|S_{i-1}|)+1$. So, on $\{T \geq i\}$, $\widetilde{D}_i \geq 2(|a| \vee |b|)+1$, validating the hypotheses of Proposition 10.2.2, and thus completing the desired computation.

How do we perform corresponding calculations if the random walk does not have mean zero? Specifically, suppose that $(S_n : n \geq 0)$ is a random walk with $S_0 = 0$ and iid increments $(X_n : n \geq 1)$ given by

$$P(X_n = 1) = p = 1 - P(X_n = -1).$$

Here, the key is to switch to our exponential martingale.

Application 10.2.3 Here, $m_X(\theta) = pe^{\theta} + (1-p)e^{-\theta}$, so $\psi(\theta) = \log(pe^{\theta} + (1-p)e^{\theta})$. Then, the martingale of interest is $L_n(\theta) = e^{\theta S_n - n\psi(\theta)}$. Assuming that optional sampling applies at time T, we arrive at $\mathrm{E}L_T(\theta) = 1$, or, in other words,

$$\mathrm{E}e^{\theta S_T - T\psi(\theta)} = 1. \tag{10.2.7}$$

To compute the exit probabilities from $[-a, b]$, it is desirable to eliminate the term $T\psi(\theta)$ from the exponent of (10.2.7).

Recall that ψ is convex (see Exercise 10.1.1). There exists a unique $\theta^* \neq 0$ such that $\psi(\theta^*) = 0$, given by

$$\theta^* = \log\left(\frac{1-p}{p}\right).$$

Substituting $\theta = \theta^*$ into (10.2.7), we get $\mathrm{E}e^{\theta^* S_T} = 1$. But $\mathrm{E}e\theta^* S_T = e{-}\theta^* a P(S_T = -a) + e\theta^* b P(S_T = b)$. Hence,

$$P(S_T = -a) = \frac{1 - \left(\frac{1-p}{p}\right)^b}{\left(\frac{p}{1-p}\right)^a - \left(\frac{1-p}{p}\right)^b}$$

(This is basically the probability of ruin in a gambler's ruin problem that is not fair.)

Exercise 10.2.2 Rigorously apply the optional sampling theorem in Application 10.2.3.

Application 10.2.4 Let $(S_n : n \geq 0)$ be a random walk with $S_0 = 0$ and iid increments $(X_n : n \geq 1)$ given by

$$P(X_n = 1) = p = 1 - P(X_n = -1)$$

with $p > 1/2$. This is a walk with positive drift, so that $T < \infty$ a.s. if we set $T = \inf\{n \geq 0 : S_n \geq b\}$. Our goal here is to compute the moment generating function of T, using martingale methods.

Assuming that we can invoke the optional sampling theorem at T,

$$\mathrm{E}e^{\theta S_n - n\psi(\theta)} = 1 \tag{10.2.8}$$

For T as described above, $S_T = b$ (This is a consequence of the "continuity" of the nearest-neighbor random walk. If X_i can take on values greater than or equal to 2, then S_T would not be deterministic, and this calculation becomes much harder). Relation (10.2.8) yields

$$\mathrm{E}e^{-T\psi(\theta)} = e^{-\theta b}$$

Set $\gamma = \gamma(\theta)$ so that $\theta = \psi^{-1}(\gamma)$. Then,

$$\mathrm{E}e^{-T\gamma} = e^{-\psi^{-1}(\gamma) b}$$

is the moment generating function of T (In computing $\psi^{-1}(\gamma)$, one may find multiple roots; to formally determine the appropriate root, note that the function $\mathrm{E}e^{-\gamma T}$ of the non-negative random variable T must be non-increasing in γ).

To make this result rigorous, note that if $p > 1/2$, then $\psi'(0) > 0$. The convexity of $\psi(\cdot)$ then guarantees that $\psi(\theta) > 0$ for $\theta > 0$. Consequently, for $\theta > 0$,

$$e^{\theta S_{T\wedge n} - \psi(\theta)(T\wedge n)} \leq e^{\theta S_{T\wedge n}} \leq e^{\theta b},$$

so the Dominated Convergence Theorem ensures that (10.2.8) holds for $\theta > 0$. Then for $\gamma > 0$, let $\eta = \psi^{-1}(\gamma)$ be the non-negative root of

$$\psi(\eta) = \gamma \tag{10.2.9}$$

Relation (10.2.9) yields the expression

$$\mathrm{E}e^{-\gamma T} = e^{-\psi^{-1}(\gamma)b} = e^{-\eta b},$$

where ψ^{-1} is defined as above. Note that a rigorous application of optional sampling theory has led us to the correct choice of root for the equation (10.2.9).

A similar analysis is possible for the one-sided hitting time $T = \inf\{n \geq 0 : S_n \leq -a\}$ with $a > 0$. Since $p > 1/2$, T is infinite with positive probability in this case. Again, consider the sequence $e^{\theta S_{T\wedge n} - \psi(\theta)(T\wedge n)}$. Note that if $\theta < 0$ and $\psi(\theta) > 0$, this sequence is bounded above by $e^{-\theta a}$. Hence, we may interchange limits and expectations in the expression

$$e^{-\theta a}\mathrm{E}[e^{-T\psi(\theta)}\mathbb{I}(T \leq n)] + \mathrm{E}[e^{\theta S_n - n\psi(\theta)}\mathbb{I}(T > n)] = 1,$$

thereby yielding the identity

$$\mathrm{E}[e^{-T\psi(\theta)}; T < \infty] = e^{\theta a},$$

for $\theta < 0$ satisfying $\psi(\theta) > 0$. So, for $\gamma \geq 0$, let $\eta = \psi^{-1}(\gamma)$ be the root less than or equal to $\theta^* = \log((1-p)/p) < 0$ defined by $\psi(\eta) = \gamma$. For the root defined as above, we then have

$$\mathrm{E}[e^{-\gamma T}; T < \infty] = e^{\psi^{-1}(\gamma)a}.$$

Note that by setting $\gamma = 0$, we obtain the identity

$$\mathrm{P}(T < \infty) = e^{\theta^* a}.$$

In other words, we have computed the probability that a positive drift "nearest neighbor" random walk ever drops below $-a$.

The theory of optional sampling extends beyond the martingale setting to supermartingales and submartingales.

Definition 10.2.1 Let $(M_n : n \geq 0)$ be an integrable sequence of random variables that is adapted to $(Z_n : n \geq 0)$. If for $n \geq 0$,

$$\mathrm{E}[M_{n+1}|Z_0, \ldots, Z_{n-1}] \leq M_n,$$

then $(M_n : n \geq 0)$ is said to be a *supermartingale* with respect to $(Z_n : n \geq 0)$. On the other hand, if

$$\mathrm{E}[M_{n+1}|Z_0, \ldots, Z_{n-1}] \geq M_n,$$

then $(M_n : n \geq 0)$ is said to be a *submartingale* with respect to $(Z_n : n \geq 0)$.

If M_n corresponds to the fortune of a gambler at time n, then a supermartingale indicates that the game is unfavorable to the gambler, whereas a submartingale indicates that the game is favorable.

Proposition 10.2.3 Let T be a stopping time that is adapted to $(Z_n : n \geq 0)$. If $(M_n : n \geq 0)$ is a supermartingale with respect to $(Z_n : n \geq 0)$, then

$$\mathrm{E}M_{T \wedge n} \leq \mathrm{E}M_0, \quad n \geq 0.$$

On the other hand, if $(M_n : n \geq 0)$ is a submartingale with respect to $(z_n : n \geq 0)$, then

$$\mathrm{E}M_{T \wedge n} \geq \mathrm{E}M_0, \quad n \geq 0.$$

Exercise 10.2.3 Prove Proposition 10.2.3.

Exercise 10.2.4 Let $(M_n : n \geq 0)$ be a martingale with respect to $(Z_n : n \geq 0)$. Suppose that $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function for which $\mathrm{E}|\phi(M_n)| < \infty$ for $n \geq 0$. Prove that $(\phi(M_n) : n \geq 0)$ is a submartingale with respect to $(Z_n : n \geq 0)$.

10.3 Martingales for Discrete-Time Markov Chains

In this section, we show how the random walk martingales introduced earlier generalize to the DTMC setting. Each of the martingales constructed here will have natural analogs in the SDE context.

Let $(Y_n : n \geq 0)$ be a real-valued sequence of random variables, not necessarily Markov. A standard trick for constructing a martingale in this very general setting is to set $D_i = Y_i - \mathrm{E}[Y_i | Y_0, \ldots, Y_{i-1}]$ for $i \geq 1$. Assuming that the Y_i's are integrable, then the D_i's are martingale differences with respect to the Y_i's. Hence,

$$M_n = \sum_{i=1}^{n} [Y_i - \mathrm{E}[Y_i | Y_0, \ldots, Y_{i-1}]]$$

is a martingale. The same kind of idea works nicely in the DTMC setting. For $f : S \to \mathbb{R}$ that is bounded, note that

$$D_i = f(X_i) - \mathrm{E}[f(X_i) | X_0, \ldots, X_{i-1}] = f(X_i) - \mathrm{E}[f(X_i) | X_{i-1}] = f(X_i) - (Pf)(X_{i-1})$$

is a martingale difference with respect to $(X_i : i \geq 0)$. Hence,

$$\widetilde{M}_n = \sum_{i=1}^{n} [f(X_i) - (Pf)(X_{i-1})]$$

is a mean-zero martingale. But

$$\begin{aligned}
\widetilde{M}_n &= \sum_{i=1}^{n} [f(X_i) - (Pf)(X_{i-1})] \\
&= \sum_{i=0}^{n-1} [f(X_i) - (Pf)(X_{i-1})] + f(X_n) - f(X_0) \\
&= f(X_n) - f(X_0) - \sum_{i=0}^{n-1} (Af)(X_i)
\end{aligned}$$

It follows easily that $M_n = f(X_n) - \sum_{i=0}^{n-1}(Af)(X_i)$ is a martingale whenever f is bounded. We have proved the following result.

Proposition 10.3.1 For $f : S \to \mathbb{R}$ bounded, $M_n = f(X_n) - \sum_{i=0}^{n-1}(Af)(X_i)$ is a martingale with respect to $(X_n : n \geq 1)$.

This martingale is known as the *Dynkin martingale*. Viewing $(Af)(X_i)$ as the increment of a random walk-type process, this is clearly the DTMC analog to the random walk martingale.

Suppose that $Af = 0$. Then Proposition 10.3.1 implies that $(f(X_n) : n \geq 0)$ is a martingale with respect to $(X_n : n \geq 0)$.

Definition 10.3.1 A function $f : S \to \mathbb{R}$ for which $Af = 0$ is called a *harmonic* function.

The term "harmonic function" is widely used in the analysis literature. It refers to functions $f : \mathbb{R}^d \to \mathbb{R}$ for which $\Delta f = 0$, where

$$\Delta = \frac{\partial^2}{\partial^2 x_1^2} + \frac{\partial^2}{\partial^2 x_2^2} + \cdots + \frac{\partial^2}{\partial^2 x_d^2}.$$

(The operator Δ is known as the "Laplacian operator".) Note that if the Markov chain X corresponds (for example) to simple random walk on the lattice plane, then

$$P((x_1, y_1), (x_2, y_2)) = \begin{cases} 1/4 & \text{if } (x_2, y_2) \in \{(x_1 + 1, y_1), (x_1 - 1, y_1), (x_1, y_1 + 1), (x_1, y_1 - 1)\} \\ 0 & \text{otherwise} \end{cases}.$$

Requiring that f be harmonic in this setting forces f to satisfy

$$\frac{f(x_1 + 1, y_1) + f(x_1 - 1, y_1) + f(x_1, y_1 + 1) + f(x_1, y_1 - 1) - 4f(x_1, y_1)}{4} = 0 \qquad (10.3.1)$$

The left-hand side term turns out to be a finite-difference approximation to Δf in two dimensions. Thus, Definition 10.3.1 legitimately extends the classical notion of harmonic functions.

Proposition 10.3.2 (a) If f is a bounded function for which $Af \leq 0$, then $(f(X_n) : n \geq 0)$ is a supermartingale with respect to $(X_n : n \geq 0)$.

(b) If f is a bounded function for which $Af \geq 0$, then $(f(X_n) : n \geq 0)$ is a submartingale with respect to $(X_n : n \geq 0)$.

Exercise 10.3.1 Prove Proposition 10.3.2.

Definition 10.3.2 A function f for which $Af \leq 0$ is said to be *superharmonic*. If instead $Af \geq 0$, then f is said to be *subharmonic*.

Again, this definition extends the classical usage, which states that f is superharmonic if $\Delta f \geq 0$ and subharmonic if $\Delta f \leq 0$. It is in order to remain consistent with the classical usage that we apply the term "supermartingale" rather than "submartingale" to an unfavorable game in which M_n has a tendency to decrease in expectation.

There is a nice connection between harmonic functions and recurrence.

Exercise 10.3.2 Suppose that X is an irreducible DTMC.

(a) If X is recurrent, prove that all the bounded harmonic functions are constants. (Hint: This is easy if $|S| < \infty$. To prove the general case, use Theorem 10.1.2.)

(b) If X is transient, show that there always exists at least one non-constant bounded harmonic function.

To apply martingale theory to additive processes of the form

$$\sum_{j=0}^{n-1} g(X_j) \tag{10.3.2}$$

with X Markov, the obvious device to apply is Proposition 10.3.1. So, note that if we could find f such that

$$Af = -g \tag{10.3.3}$$

then we effectively would have our desired martingale for (10.3.2), namely

$$M_n = f(X_n) + \sum_{j=0}^{n-1} g(X_j) \tag{10.3.4}$$

(In the Markov setting, one cannot expect (10.3.2) itself to be a martingale – it just isn't. But (10.3.4) shows that it can be represented as a martingale if one adds on the "correction term" $f(X_n)$.) Because (10.3.3) plays a key role in representing (10.3.2) as a martingale, this equation has an important place in the theory of Markov processes. Equation (10.3.3) is called Poisson's equation. (In the symmetric simple random walk setting, (10.3.3) is just a finite-difference approximation to $\Delta f = -g$, which is Poisson's equation in the partial differential equations setting.)

Poisson's equation need not have a solution for arbitrary g.

Exercise 10.3.3 Suppose that X is an irreducible transient DTMC. If g has finite support (i.e. $\{x \in S : g(x) \neq 0\}$ has finite cardinality), show that Poisson's equation has a solution.

Exercise 10.3.4 Suppose that X is an irreducible finite-state DTMC. Let π be the stationary distribution of X. Let Π be the matrix in which all rows are identical to π

(a) Prove that $\Pi P = P\Pi = \Pi^2$.

(b) Prove that $(P - \Pi)^n = P^n - \Pi$ for $n \geq 1$.

(c) Prove that if X is aperiodic, then $\sum_{n=0}^{\infty}(P - \Pi)^n$ converges absolutely.

(d) Prove that if X is aperiodic, then $(I - P + \Pi)^{-1}$ exists.

(e) Extend (d) to the periodic case.

(f) Prove that if g is such that $\pi g = 0$, then $f = (\Pi - A)^{-1}g$ solves Poisson's equation $Af = -g$.

(g) Prove that if g is such that $\pi g \neq 0$, then $Af = -g$ has no solution.

Exercise 10.3.5 We extend here the existence of solutions to Poisson's equation to infinite state irreducible positive recurrent Markov chains $X = (X_n : n \geq 0)$. Let $f : S \to \mathbb{R}$ be such that $\sum_x \pi(x)|f(x)| < \infty$. Set $f_c(x) = f(x) - \sum_y \pi(y)f(y)$, and put

$$u^*(x) = E_x \sum_{n=0}^{\tau(z)-1} f_c(X_n),$$

where $\tau(z) = \inf\{n \geq 1 : X_n = z\}$.

(a) Prove that $E_x \sum_{n=0}^{\tau(z)-1} |f_c(X_n)| < \infty$ for each $x \in S$ (so that $u^*(\cdot)$ is finite-valued).

(b) Prove that

$$u^*(x) = f_c(x) + \sum_{y \in S} P(x,y)u^*(y)$$

so that u^* is a solution of Poisson's equation.

We now turn to developing an analog to the likelihood ratio martingale that was discussed in the random walk setting. Let $X = (X_n : n \geq 0)$ be an S-valued DTMC with initial distribution ν and (one-step) transition matrix $Q = (Q(x,y) : x, y \in S)$. Suppose that we select a stochastic vector μ and transition matrix P such that

(i) $\mu(x) = 0$ whenever $\nu(x) = 0$ for $x \in S$;

(ii) $P(x,y) = 0$ whenever $Q(x,y) = 0$ for $x, y \in S$.

Proposition 10.3.3 The sequence $(L_n : n \geq 0)$ is a martingale with respect to $(X_n : n \geq 0)$, where

$$L_n = \frac{\mu(X_0)}{\nu(X_0)} \prod_{j=0}^{n-1} \frac{P(X_j, X_{j+1})}{Q(X_j, X_{j+1})}, \quad n \geq 0$$

Exercise 10.3.6 Prove Proposition 10.3.3.

We close this section with a discussion of the exponential martingale's extension to the DTMC setting. Suppose that we wish to study an additive process of the form $\sum_{j=0}^{n-1} g(X_j)$, where $(X_n : n \geq 0)$ is an irreducible finite-state DTMC. In the random walk setting, the moment generating function of the random walk played a critical role in constructing the exponential martingale. This suggests considering

$$u_n(\theta, x, y) = \mathrm{E}_x[e^{\theta \sum_{j=0}^{n-1} g(X_j)}; X_n = y]$$

for $x, y \in S$. Observe that

$$u_n(\theta, x, y) = \sum_{x_1, \ldots, xn-1} e^{\theta g(x)} P(x, x_1) e^{\theta g(x_1)} P(x_1, x_2) \cdots e^{\theta g(x_{n-1})} P(x_{n-1}, y) = K^n(\theta, x, y),$$

where $K^n(\theta, x, y)$ the $x-y$;th component of the nth power of the matrix $K(\theta)$, where

$$K(\theta, x, y) = e^{\theta g(x)} P(x,y). \tag{10.3.5}$$

Note that $K(\theta)$ is a non-negative finite irreducible matrix. Then, the Perron-Frobenius theorem for non-negative matrices implies that there exists a positive eigenvalue $\lambda(\theta)$ and corresponding positive column eigenvector $r(\theta)$ such that

$$K(\theta)r(\theta) = \lambda(\theta)r(\theta). \tag{10.3.6}$$

Let $\psi(\theta) = \log \lambda(\theta)$. We can rewrite (10.3.6) as

$$e^{-\psi(\theta)} \sum_y K(\theta, x, y) \frac{r(\theta, x)}{r(\theta, y)} = 1, \quad x \in S. \tag{10.3.7}$$

Substituting (10.3.5) into (10.3.7), we obtain

$$\sum_y e^{\theta g(x) - \psi(\theta)} P(x, y) \frac{r(\theta, x)}{r(\theta, y)} = 1$$

or equivalently,

$$\mathrm{E}_x e^{\theta g(x) - \psi(\theta)} \frac{r(\theta, X_1)}{r(\theta, X_0)} = 1$$

Proposition 10.3.4 For each $\theta \in \mathbb{R}$,

$$L_n(\theta) = e^{\theta \sum_{j=0}^{n-1} g(X_j) - n\psi(\theta)} \frac{r(\theta, X_n)}{r(\theta, X_0)}$$

is a martingale with respect to $(X_n : n \geq 0)$.

Proof: The critical verification involves showing that $\mathrm{E}[L_{n+1}(\theta)|X_0, \ldots, X_n] = L_n(\theta)$. But

$$\mathrm{E}[L_{n+1}(\theta)|X_0, \ldots, X_n] = L_n(\theta)\mathrm{E}\left[e^{\theta g(X_n) - \psi(\theta)} \frac{r(\theta, X_{n+1})}{r(\theta, X_n)} \middle| X_0, \ldots, X_n \right] = L_n(\theta). \qquad \square$$

We can rewrite this martingale as follows. Set $h(\theta, x) = \log r(\theta, x)$. Then, Proposition 10.3.4 asserts that

$$e^{h(\theta, X_n) + \theta \sum_{j=0}^{n-1} g(X_j) - n\psi(\theta)}$$

is a martingale. This exponential martingale can be used in a manner identical to the random walk setting to study $\sum_{j=0}^{n-1} g(X_j)$.

10.4 The Strong Law for Martingales

As for sums of independent mean zero rv's, we expect that in great generality,

$$n^{-1} \sum_{i=1}^{n} D_i \to 0 \quad \text{a.s.} \tag{10.4.1}$$

as $n \to \infty$. This is easy to establish if we weaken the a.s. convergence to convergence in probability, since

$$\begin{aligned} P(|n^{-1} \sum_{i=1}^{n} D_i| > a) &\leq \frac{E(n^{-1} \sum_{i=1}^{n} D_i)^2}{\epsilon^2} \\ &= \frac{1}{n^2 \epsilon^2} \sum_{i=1}^{n} ED_i^2, \end{aligned}$$

so that if

$$\sup_{n\geq 1} ED_n^2 < \infty, \tag{10.4.2}$$

it clearly follows that

$$n^{-1}\sum_{i=1}^{n} D_i \xrightarrow{p} 0 \tag{10.4.3}$$

as $n \to \infty$. To prove (10.4.3) to a.s. convergence, we need to apply the Martingale Convergence Theorem. Since $(n^{-1}\sum_{i=1}^{n} D_i : n \geq 1)$ is not a martingale, we need to use something to "bridge the gap" between $n^{-1}\sum_{i=1}^{n} D_i$ and the world of martingales. The appropriate "bridge" is Kronecker's lemma.

Kronecker's Lemma: If $(x_n : n \geq 1)$ and $(a_n : n \geq 1)$ are two real-valued sequences for which $(a_n : n \geq 1)$ is non-negative and increasing to infinity, then the existence of a finite-valued z such that

$$\sum_{j=1}^{n}\left(\frac{x_j}{a_j}\right) \to z$$

as $n \to \infty$ implies that

$$\frac{1}{a_n}\sum_{j=1}^{n} x_j \to 0$$

as $n \to \infty$.

To apply this result in our martingale setting, let

$$\tilde{M}_n = \sum_{j=1}^{n}\left(\frac{D_j}{j}\right)$$

and observe that $(\tilde{M}_n : n \geq 0)$ is a martingale for which

$$\begin{aligned} E|\tilde{M}_n| &\leq \sqrt{E\left(\sum_{j=1}^{n} D_j/j\right)^2} \\ &= \sqrt{\sum_{j=1}^{\infty} ED_j^2/j^2}, \end{aligned}$$

so that in the presence of (10.4.2), the Martingale Convergence Theorem can be applied, yielding the conclusion that there exists a finite-valued $\tilde{M}_\infty$ for which

$$\tilde{M}_n \to \tilde{M}_\infty \quad \text{a.s.}$$

as $n \to \infty$. An application of Kronecker's lemma "path-by-path" then yields

$$n^{-1}\sum_{i=1}^{n} D_i \to 0$$

a.s. as $n \to \infty$.

Exercise 10.4.1 Use the above argument to prove the strong law

$$n^{-1}\sum_{i=0}^{n-1} f(X_i) \to \sum_z \pi(z) f(z) \quad \text{a.s.}$$

as $n \to \infty$ for a given finite-state irreducible Markov chain (with equilibrium distribution $\pi = (\pi(x) : x \in S)$).

10.5 The Central Limit Theorem for Martingales

We discuss here general conditions under which

$$n^{-1/2}\sum_{i=1}^{n} D_i \Rightarrow \sigma N(0,1)$$

as $n \to \infty$ in discrete time or under which

$$t^{-1/2} M(t) \Rightarrow \sigma N(0,1) \tag{10.5.1}$$

as $n \to \infty$ in continuous time. Since discrete-time martingales are just a special case of continuous time martingales, we focus on (10.5.1).

Note that

$$M(t) = M(0) + \sum_{i=1}^{n} (M(it/n) - M((i-1)t/n))$$

so that in the presence of square integrability

$$EM^2(t) = EM^2(0) + E\sum_{i=1}^{n} (M(it/n) - M((i-1)t/n))^2.$$

For a given square-integrable martingale $(M(t)t \geq 0)$ define the *quadratic variation* of M to be

$$[M](t) = \lim_{n\to\infty} \sum_{i=0}^{n} (M(it/n) - M((i-1)t/n))^2.$$

Theorem 10.5.1 Let $(M(t) : t \geq 0)$ be a square-integrable martingale with right continuous paths with left limits. If either:

$$\frac{1}{\sqrt{t}} E \sup_{0\leq s\leq t} |M(s) - M(s-)| \to 0$$

and

$$\frac{1}{t}[M](t) \xrightarrow{p} \sigma^2$$

as $t \to \infty$, or

$$\frac{1}{t} E \sup_{0\leq s\leq t} |M(s) - M(s-)|^2 \to 0,$$

$$\frac{1}{t} E \sup_{0\leq s\leq t} |\langle M\rangle(s) - \langle M\rangle(s-)| \to 0$$

and

$$\frac{1}{t}\langle M\rangle(t) \xrightarrow{p} \sigma^2$$

as $t \to \infty$, then

$$t^{-1/2}M(t) \Rightarrow \sigma N(0,1)$$

as $t \to \infty$.

Remark 10.5.1 Note that Markov jump processes have right continuous paths with left limits, so this result applies in the Markov jump process setting.

Remark 10.5.2 When specialized to discrete time,

$$[M](n) = M^2(0) + \sum_{i=1}^{n} D_i^2$$

and

$$\langle M\rangle(n) = \sum_{i=1}^{n} E[D_i^2|Z_0, \ldots, Z_{i-1}]$$

Exercise 10.5.1 Use the Martingale CLT to prove that there exists σ for which

$$n^{-1/2}\left(\sum_{i=0}^{n-1} f(X_i) - n\sum_{z} \pi(z)f(z)\right) \Rightarrow \sigma N(0,1)$$

as $n \to \infty$, provided that $(X_n : n \geq 0)$ is a finite state irreducible Markov chain.