12

Martingales

Summary. The general theory of martingales and submartingales has many applications. After an account of the concentration inequality for martingales, the martingale convergence theorem is proved via the upcrossings inequality. Stopping times are studied, and the optional stopping theorem proved. This leads to Wald's identity and the maximal inequality. The chapter ends with a discussion of backward martingales and continuous-time martingales. Many examples of the use of martingale theory are included.

12.1 Introduction

Random processes come in many forms, and their analysis depends heavily on the assumptions that one is prepared to make about them. There are certain broad classes of processes whose general properties enable one to build attractive theories. Two such classes are Markov processes and stationary processes. A third is the class of martingales.

(1) Definition. A sequence $Y = \{Y_n : n \geq 0\}$ is a **martingale** with respect to the sequence $X = \{X_n : n \geq 0\}$ if, for all $n \geq 0$,

(a) $\mathbb{E}|Y_n| < \infty$,

(b) $\mathbb{E}(Y_{n+1} \mid X_0, X_1, \dots, X_n) = Y_n$.

A warning note: conditional expectations are ubiquitous in this chapter. Remember that they are random variables, and that formulae of the form $\mathbb{E}(A \mid B) = C$ generally hold only 'almost surely'. We shall omit the term 'almost surely' throughout the chapter.

Here are some examples of martingales; further examples may be found in Section 7.7.

(2) Example. Simple random walk. A particle jumps either one step to the right or one step to the left, with corresponding probabilities p and q $(= 1 - p)$. Assuming the usual independence of different moves, it is clear that the position $S_n = X_1 + X_2 + \cdots + X_n$ of the particle after n steps satisfies $\mathbb{E}|S_n| \leq n$ and

$$\mathbb{E}(S_{n+1} \mid X_1, X_2, \dots, X_n) = S_n + (p - q),$$

whence it is easily seen that $Y_n = S_n - n(p - q)$ defines a martingale with respect to X. ●

(3) Example. The martingale. The following gambling strategy is called a martingale. A gambler has a large fortune. He wagers £1 on an evens bet. If he loses then he wagers £2 on the next bet. If he loses the first n plays, then he bets £2^n on the $(n+1)$th. He is bound to win sooner or later, say on the Tth bet, at which point he ceases to play, and leaves with his profit of $2^T - (1 + 2 + 4 + \cdots + 2^{T-1})$. Thus, following this strategy, he is assured an ultimate profit. This sounds like a good policy.

Writing Y_n for the accumulated gain of the gambler after the nth play (losses count negative), we have that $Y_0 = 0$ and $|Y_n| \le 1 + 2 + \cdots + 2^{n-1} = 2^n - 1$. Furthermore, $Y_{n+1} = Y_n$ if the gambler has stopped by time $n+1$, and

$$Y_{n+1} = \begin{cases} Y_n - 2^n & \text{with probability } \frac{1}{2}, \\ Y_n + 2^n & \text{with probability } \frac{1}{2}, \end{cases}$$

otherwise, implying that $\mathbb{E}(Y_{n+1} \mid Y_1, Y_2, \dots, Y_n) = Y_n$. Therefore Y is a martingale (with respect to itself).

As remarked in Example (7.7.1), this martingale possesses a particularly disturbing feature. The random time T has a geometric distribution, $\mathbb{P}(T = n) = (\frac{1}{2})^n$ for $n \ge 1$, so that the mean loss of the gambler just before his ultimate win is

$$\sum_{n=1}^{\infty} (\tfrac{1}{2})^n (1 + 2 + \cdots + 2^{n-2})$$

which equals infinity. Do not follow this strategy unless your initial capital is considerably greater than that of the casino. ●

(4) Example. De Moivre's martingale. About a century before the martingale was fashionable amongst Paris gamblers, Abraham de Moivre made use of a (mathematical) martingale to answer the following 'gambler's ruin' question. A simple random walk on the set $\{0, 1, 2, \dots, N\}$ stops when it first hits either of the absorbing barriers at 0 and at N; what is the probability that it stops at the barrier 0?

Write $X_1, X_2, \dots$ for the steps of the walk, and S_n for the position after n steps, where $S_0 = k$. Define $Y_n = (q/p)^{S_n}$ where $p = \mathbb{P}(X_i = 1)$, $p + q = 1$, and $0 < p < 1$. We claim that

$$\mathbb{E}(Y_{n+1} \mid X_1, X_2, \dots, X_n) = Y_n \quad \text{for all } n. \tag{5}$$

If S_n equals 0 or N then the process has stopped by time n, implying that $S_{n+1} = S_n$ and therefore $Y_{n+1} = Y_n$. If on the other hand $0 < S_n < N$, then

$$\begin{aligned} \mathbb{E}(Y_{n+1} \mid X_1, X_2, \dots, X_n) &= \mathbb{E}\big((q/p)^{S_n + X_{n+1}} \,\big|\, X_1, X_2, \dots, X_n\big) \\ &= (q/p)^{S_n}\big[p(q/p) + q(q/p)^{-1}\big] = Y_n, \end{aligned}$$

and (5) is proved. It follows, by taking expectations of (5), that $\mathbb{E}(Y_{n+1}) = \mathbb{E}(Y_n)$ for all n, and hence $\mathbb{E}|Y_n| = \mathbb{E}|Y_0| = (q/p)^k$ for all n. In particular Y is a martingale (with respect to the sequence X).

Let T be the number of steps before the absorption of the particle at either 0 or N. De Moivre argued as follows: $\mathbb{E}(Y_n) = (q/p)^k$ for all n, and therefore $\mathbb{E}(Y_T) = (q/p)^k$. If you are willing to accept this remark, then the answer to the original question is a simple consequence, as follows. Expanding $\mathbb{E}(Y_T)$, we have that

$$\mathbb{E}(Y_T) = (q/p)^0 p_k + (q/p)^N (1 - p_k)$$

where $p_k = \mathbb{P}(\text{absorbed at } 0 \mid S_0 = k)$. However, $\mathbb{E}(Y_T) = (q/p)^k$ by assumption, and therefore

$$p_k = \frac{\rho^k - \rho^N}{1 - \rho^N} \quad \text{where} \quad \rho = q/p$$

(so long as $\rho \neq 1$), in agreement with the calculation of Example (3.9.6).

This is a very attractive method, which relies on the statement that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$ for a certain type of random variable T. A major part of our investigation of martingales will be to determine conditions on such random variables T which ensure that the desired statements are true. ●

(6) Example. Markov chains. Let X be a discrete-time Markov chain taking values in the countable state space S with transition matrix $\mathbf{P}$. Suppose that $\psi : S \to S$ is bounded and *harmonic*, which is to say that

$$\sum_{j \in S} p_{ij} \psi(j) = \psi(i) \quad \text{for all } i \in S.$$

It is easily seen that $Y = \{\psi(X_n) : n \geq 0\}$ is a martingale with respect to X: simply use the Markov property in order to perform the calculation:

$$\mathbb{E}\big(\psi(X_{n+1}) \,\big|\, X_1, X_2, \dots, X_n\big) = \mathbb{E}\big(\psi(X_{n+1}) \,\big|\, X_n\big) = \sum_{j \in S} p_{X_n,j} \psi(j) = \psi(X_n).$$

More generally, suppose that ψ is a right eigenvector of $\mathbf{P}$, which is to say that there exists λ $(\neq 0)$ such that

$$\sum_{j \in S} p_{ij} \psi(j) = \lambda \psi(i), \qquad i \in S.$$

Then

$$\mathbb{E}\big(\psi(X_{n+1}) \,\big|\, X_1, X_2, \dots, X_n\big) = \lambda \psi(X_n),$$

implying that $\lambda^{-n} \psi(X_n)$ defines a martingale so long as $\mathbb{E}|\psi(X_n)| < \infty$ for all n. ●

Central to the definition of a martingale is the idea of conditional expectation, a subject developed to some extent in Chapter 7. As described there, the most general form of conditional expectation is of the following nature. Let Y be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having finite mean, and let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. The conditional expectation of Y given $\mathcal{G}$, written $\mathbb{E}(Y \mid \mathcal{G})$, is a $\mathcal{G}$-measurable random variable satisfying

(7) $$\mathbb{E}\big([Y - \mathbb{E}(Y \mid \mathcal{G})] I_G\big) = 0 \quad \text{for all events } G \in \mathcal{G},$$

where I_G is the indicator function of G. There is a corresponding general definition of a martingale. In preparation for this, we introduce the following terminology. Suppose that $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_0, \mathcal{F}_1, \dots\}$ is a sequence of sub-σ-fields of $\mathcal{F}$; we call $\boldsymbol{\mathcal{F}}$ a *filtration* if $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all n. A sequence $Y = \{Y_n : n \geq 0\}$ is said to be *adapted* to the filtration $\boldsymbol{\mathcal{F}}$ if Y_n is $\mathcal{F}_n$-measurable for all n. Given a filtration $\boldsymbol{\mathcal{F}}$, we normally write $\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n$ for the smallest σ-field containing $\mathcal{F}_n$ for all n.

(8) Definition. Let $\mathcal{F}$ be a filtration of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let Y be a sequence of random variables which is adapted to $\mathcal{F}$. We call the pair $(Y, \mathcal{F}) = \{(Y_n, \mathcal{F}_n) : n \geq 0\}$ a **martingale** if, for all $n \geq 0$,

(a) $\mathbb{E}|Y_n| < \infty$,

(b) $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n$.

The former definition (1) is retrieved by choosing $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$, the smallest σ-field with respect to which each of the variables $X_0, X_1, \dots, X_n$ is measurable. We shall sometimes suppress reference to the filtration $\mathcal{F}$, speaking only of a martingale Y.

Note that, if Y is a martingale with respect to $\mathcal{F}$, then it is also a martingale with respect to $\mathcal{G}$ where $\mathcal{G}_n = \sigma(Y_0, Y_1, \dots, Y_n)$. A further minor point is that martingales need not be infinite in extent: a finite sequence $\{(Y_n, \mathcal{F}_n) : 0 \leq n \leq N\}$ satisfying the above definition is also termed a martingale.

There are many cases of interest in which the martingale condition $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n$ does not hold, being replaced instead by an inequality: $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) \geq Y_n$ for all n, or $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) \leq Y_n$ for all n. Sequences satisfying such inequalities have many of the properties of martingales, and we have special names for them.

(9) Definition. Let $\mathcal{F}$ be a filtration of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let Y be a sequence of random variables which is adapted to $\mathcal{F}$. We call the pair $(Y, \mathcal{F})$ a **submartingale** if, for all $n \geq 0$,

(a) $\mathbb{E}(Y_n^+) < \infty$,

(b) $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) \geq Y_n$,

or a **supermartingale** if, for all $n \geq 0$,

(c) $\mathbb{E}(Y_n^-) < \infty$,

(d) $\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) \leq Y_n$.

Remember that $X^+ = \max\{0, X\}$ and $X^- = -\min\{0, X\}$, so that $X = X^+ - X^-$ and $|X| = X^+ + X^-$. The moment conditions (a) and (c) are weaker than the condition that $\mathbb{E}|Y_n| < \infty$. Note that Y is a martingale if and only if it is both a submartingale and a supermartingale. Also, Y is a submartingale if and only if $-Y$ is a supermartingale.

Sometimes we shall write that $(Y_n, \mathcal{F}_n)$ is a (sub/super)martingale in cases where we mean the corresponding statement for $(Y, \mathcal{F})$.

It can be somewhat tiresome to deal with sub(/super)martingales and martingales separately, keeping track of their various properties. The general picture is somewhat simplified by the following result, which expresses a submartingale as the sum of a martingale and an increasing 'predictable' process. We shall not make use of this decomposition in the rest of the chapter. Here is a piece of notation. We call the pair $(S, \mathcal{F}) = \{(S_n, \mathcal{F}_n) : n \geq 0\}$ *predictable* if S_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$. We call a predictable process $(S, \mathcal{F})$ *increasing* if $S_0 = 0$ and $\mathbb{P}(S_n \leq S_{n+1}) = 1$ for all n.

(10) Theorem. Doob decomposition. *A submartingale $(Y, \mathcal{F})$ with finite means may be expressed in the form*

(11)
$$Y_n = M_n + S_n$$

where $(M, \mathcal{F})$ is a martingale, and $(S, \mathcal{F})$ is an increasing predictable process. This decomposition is unique.

The process $(S, \mathcal{F})$ in (11) is called the *compensator* of the submartingale $(Y, \mathcal{F})$. Note that compensators have finite mean, since $0 \le S_n \le Y_n^+ - M_n$, implying that

$$\mathbb{E}|S_n| \le \mathbb{E}(Y_n^+) + \mathbb{E}|M_n|. \tag{12}$$

Proof. We define M and S explicitly as follows: $M_0 = Y_0$, $S_0 = 0$,

$$M_{n+1} - M_n = Y_{n+1} - \mathbb{E}(Y_{n+1} \mid \mathcal{F}_n), \qquad S_{n+1} - S_n = \mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) - Y_n,$$

for $n \ge 0$. It is easy to check (*exercise*) that $(M, \mathcal{F})$ and $(S, \mathcal{F})$ satisfy the statement of the theorem. To see uniqueness, suppose that $Y_n = M_n' + S_n'$ is another such decomposition. Then

$$\begin{aligned} Y_{n+1} - Y_n &= (M_{n+1}' - M_n') + (S_{n+1}' - S_n') \\ &= (M_{n+1} - M_n) + (S_{n+1} - S_n). \end{aligned}$$

Take conditional expectations given $\mathcal{F}_n$ to obtain $S_{n+1}' - S_n' = S_{n+1} - S_n$, $n \ge 0$. However, $S_0' = S_0 = 0$, and therefore $S_n' = S_n$, implying that $M_n' = M_n$. (Most of the last few statements should be qualified by 'almost surely'.) ■

Exercises for Section 12.1

1. (i) If $(Y, \mathcal{F})$ is a martingale, show that $\mathbb{E}(Y_n) = \mathbb{E}(Y_0)$ for all n.
(ii) If $(Y, \mathcal{F})$ is a submartingale (respectively supermartingale) with finite means, show that $\mathbb{E}(Y_n) \ge \mathbb{E}(Y_0)$ (respectively $\mathbb{E}(Y_n) \le \mathbb{E}(Y_0)$).

2. Let $(Y, \mathcal{F})$ be a martingale, and show that $\mathbb{E}(Y_{n+m} \mid \mathcal{F}_n) = Y_n$ for all $n, m \ge 0$.

3. Let Z_n be the size of the nth generation of a branching process with $Z_0 = 1$, having mean family size μ and extinction probability η. Show that $Z_n \mu^{-n}$ and η^{Z_n} define martingales.

4. Let $\{S_n : n \ge 0\}$ be a simple symmetric random walk on the integers with $S_0 = k$. Show that S_n and $S_n^2 - n$ are martingales. Making assumptions similar to those of de Moivre (see Example (12.1.4)), find the probability of ruin and the expected duration of the game for the gambler's ruin problem.

5. Let $(Y, \mathcal{F})$ be a martingale with the property that $\mathbb{E}(Y_n^2) < \infty$ for all n. Show that, for $i \le j \le k$, $\mathbb{E}\{(Y_k - Y_j)Y_i\} = 0$, and $\mathbb{E}\{(Y_k - Y_j)^2 \mid \mathcal{F}_i\} = \mathbb{E}(Y_k^2 \mid \mathcal{F}_i) - \mathbb{E}(Y_j^2 \mid \mathcal{F}_i)$. Suppose there exists K such that $\mathbb{E}(Y_n^2) \le K$ for all n. Show that the sequence $\{Y_n\}$ converges in mean square as $n \to \infty$.

6. Let Y be a martingale and let u be a convex function mapping $\mathbb{R}$ to $\mathbb{R}$. Show that $\{u(Y_n) : n \ge 0\}$ is a submartingale provided that $\mathbb{E}(u(Y_n)^+) < \infty$ for all n.

Show that $|Y_n|$, Y_n^2, and Y_n^+ constitute submartingales whenever the appropriate moment conditions are satisfied.

7. Let Y be a submartingale and let u be a convex non-decreasing function mapping $\mathbb{R}$ to $\mathbb{R}$. Show that $\{u(Y_n) : n \ge 0\}$ is a submartingale provided that $\mathbb{E}(u(Y_n)^+) < \infty$ for all n.

Show that (subject to a moment condition) Y_n^+ constitutes a submartingale, but that $|Y_n|$ and Y_n^2 need not constitute submartingales.

8. Let X be a discrete-time Markov chain with countable state space S and transition matrix $\mathbf{P}$. Suppose that $\psi : S \to \mathbb{R}$ is bounded and satisfies $\sum_{j \in S} p_{ij}\psi(j) \le \lambda\psi(i)$ for some $\lambda > 0$ and all $i \in S$. Show that $\lambda^{-n}\psi(X_n)$ constitutes a supermartingale.

9. Let $G_n(s)$ be the probability generating function of the size Z_n of the nth generation of a branching process, where $Z_0 = 1$ and $\operatorname{var}(Z_1) > 0$. Let H_n be the inverse function of the function G_n, viewed as a function on the interval $[0, 1]$, and show that $M_n = \{H_n(s)\}^{Z_n}$ defines a martingale with respect to the sequence Z.

12.2 Martingale differences and Hoeffding's inequality

Much of the theory of martingales is concerned with their behaviour as $n \to \infty$, and particularly with their properties of convergence. Of supreme importance is the martingale convergence theorem, a general result of great power and with many applications. Before giving an account of that theorem (in the next section), we describe a bound on the degree of fluctuation of a martingale. This bound is straightforward to derive and has many important applications.

Let $(Y, \mathcal{F})$ be a martingale. The sequence of *martingale differences* is the sequence $D = \{D_n : n \geq 1\}$ defined by $D_n = Y_n - Y_{n-1}$, so that

(1) $$Y_n = Y_0 + \sum_{i=1}^{n} D_i.$$

Note that the sequence D is such that D_n is $\mathcal{F}_n$-measurable, $\mathbb{E}|D_n| < \infty$, and

(2) $$\mathbb{E}(D_{n+1} \mid \mathcal{F}_n) = 0 \quad \text{for all } n.$$

(3) **Theorem. Hoeffding's inequality.** *Let $(Y, \mathcal{F})$ be a martingale, and suppose that there exists a sequence $K_1, K_2, \ldots$ of real numbers such that $\mathbb{P}(|Y_n - Y_{n-1}| \leq K_n) = 1$ for all n. Then*

$$\mathbb{P}(|Y_n - Y_0| \geq x) \leq 2\exp\left(-\tfrac{1}{2}x^2 \Big/ \sum_{i=1}^{n} K_i^2\right), \quad x > 0.$$

That is to say, if the martingale differences are bounded (almost surely) then there is only a small chance of a large deviation of Y_n from its initial value Y_0.

Proof. We begin with an elementary inequality. If $\psi > 0$, the function $g(d) = e^{\psi d}$ is convex, whence it follows that

(4) $$e^{\psi d} \leq \tfrac{1}{2}(1-d)e^{-\psi} + \tfrac{1}{2}(1+d)e^{\psi} \quad \text{if} \quad |d| \leq 1.$$

Applying this to a random variable D having mean 0 and satisfying $\mathbb{P}(|D| \leq 1) = 1$, we obtain

(5) $$\mathbb{E}(e^{\psi D}) \leq \tfrac{1}{2}(e^{-\psi} + e^{\psi}) < e^{\frac{1}{2}\psi^2},$$

by a comparison of the coefficients of ψ^{2n} for $n \geq 0$.

Moving to the proof proper, it is a consequence of Markov's inequality, Theorem (7.3.1), that

(6) $$\mathbb{P}(Y_n - Y_0 \geq x) \leq e^{-\theta x}\mathbb{E}(e^{\theta(Y_n - Y_0)})$$

for $\theta > 0$. Writing $D_n = Y_n - Y_{n-1}$, we have that

$$\mathbb{E}(e^{\theta(Y_n - Y_0)}) = \mathbb{E}(e^{\theta(Y_{n-1} - Y_0)}e^{\theta D_n}).$$

By conditioning on $\mathcal{F}_{n-1}$, we obtain

$$\begin{aligned}(7)\qquad \mathbb{E}(e^{\theta(Y_n-Y_0)} \mid \mathcal{F}_{n-1}) &= e^{\theta(Y_{n-1}-Y_0)}\mathbb{E}(e^{\theta D_n} \mid \mathcal{F}_{n-1}) \\ &\le e^{\theta(Y_{n-1}-Y_0)} \exp(\tfrac{1}{2}\theta^2 K_n^2),\end{aligned}$$

where we have used the fact that $Y_{n-1} - Y_0$ is $\mathcal{F}_{n-1}$-measurable, in addition to (5) applied to the random variable D_n/K_n. We take expectations of (7) and iterate to find that

$$\mathbb{E}(e^{\theta(Y_n-Y_0)}) \le \mathbb{E}(e^{\theta(Y_{n-1}-Y_0)}) \exp(\tfrac{1}{2}\theta^2 K_n^2) \le \exp\left(\tfrac{1}{2}\theta^2 \sum_{i=1}^{n} K_i^2\right).$$

Therefore, by (6),

$$\mathbb{P}(Y_n - Y_0 \ge x) \le \exp\left(-\theta x + \tfrac{1}{2}\theta^2 \sum_{i=1}^{n} K_i^2\right)$$

for all $\theta > 0$. Suppose $x > 0$, and set $\theta = x \big/ \sum_{i=1}^{n} K_i^2$ (this is the value which minimizes the exponent); we obtain

$$\mathbb{P}(Y_n - Y_0 \ge x) \le \exp\left(-\tfrac{1}{2}x^2 \Big/ \sum_{i=1}^{n} K_i^2\right), \qquad x > 0.$$

The same argument is valid with $Y_n - Y_0$ replaced by $Y_0 - Y_n$, and the claim of the theorem follows by adding the two (identical) bounds together. ■

(8) Example. Large deviations. Let $X_1, X_2, \dots$ be independent random variables, X_i having the Bernoulli distribution with parameter p. We set $S_n = X_1 + X_2 + \cdots + X_n$ and $Y_n = S_n - np$ to obtain a martingale Y. It is a consequence of Hoeffding's inequality that

$$\mathbb{P}(|S_n - np| \ge x\sqrt{n}) \le 2\exp(-\tfrac{1}{2}x^2/\mu) \quad \text{for} \quad x > 0,$$

where $\mu = \max\{p, 1-p\}$. This is an inequality of a type encountered already as Bernstein's inequality (2.2.4), and explored in greater depth in Section 5.11. ●

(9) Example. Bin packing. The bin packing problem is a basic problem of operations research. Given n objects with sizes $x_1, x_2, \dots, x_n$, and an unlimited collection of bins each of size 1, what is the minimum number of bins required in order to pack the objects? In the randomized version of this problem, we suppose that the objects have independent random sizes $X_1, X_2, \dots$ having some common distribution on $[0, 1]$. Let B_n be the (random) number of bins required in order to pack $X_1, X_2, \dots, X_n$ efficiently; that is, B_n is the minimum number of bins of unit capacity such that the sum of the sizes of the objects in any given bin does not exceed its capacity. It may be shown that B_n grows approximately linearly in n, in that there exists a positive constant β such that $n^{-1}B_n \to \beta$ a.s. and in mean square as $n \to \infty$. We shall not prove this here, but note its consequence:

$$(10)\qquad \frac{1}{n}\mathbb{E}(B_n) \to \beta \quad \text{as} \quad n \to \infty.$$

The next question might be to ask how close B_n is to its mean value $\mathbb{E}(B_n)$, and Hoeffding's inequality may be brought to bear here. For $i \le n$, let $Y_i = \mathbb{E}(B_n \mid \mathcal{F}_i)$, where $\mathcal{F}_i$ is the σ-field generated by $X_1, X_2, \dots, X_i$. It is easily seen that $(Y, \boldsymbol{\mathcal{F}})$ is a martingale, albeit one of finite length. Furthermore $Y_n = B_n$, and $Y_0 = \mathbb{E}(B_n)$ since $\mathcal{F}_0$ is the trivial σ-field $\{\varnothing, \Omega\}$.

Now, let $B_n(i)$ be the minimal number of bins required in order to pack all the objects *except* the ith. Since the objects are packed efficiently, we must have $B_n(i) \le B_n \le B_n(i)+1$. Taking conditional expectations given $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$, we obtain

(11)
$$\begin{aligned} \mathbb{E}\big(B_n(i) \bigm| \mathcal{F}_{i-1}\big) &\le Y_{i-1} \le \mathbb{E}\big(B_n(i) \bigm| \mathcal{F}_{i-1}\big) + 1, \\ \mathbb{E}\big(B_n(i) \bigm| \mathcal{F}_{i}\big) &\le \ Y_{i} \quad \le \mathbb{E}\big(B_n(i) \bigm| \mathcal{F}_{i}\big) + 1. \end{aligned}$$

However, $\mathbb{E}(B_n(i) \mid \mathcal{F}_{i-1}) = \mathbb{E}(B_n(i) \mid \mathcal{F}_i)$, since we are not required to pack the ith object, and hence knowledge of X_i is irrelevant. It follows from (11) that $|Y_i - Y_{i-1}| \le 1$. We may now apply Hoeffding's inequality (3) to find that

(12)
$$\mathbb{P}\big(|B_n - \mathbb{E}(B_n)| \ge x\big) \le 2\exp(-\tfrac{1}{2}x^2/n), \quad x > 0.$$

For example, setting $x = \epsilon n$, we see that the chance that B_n deviates from its mean by ϵn (or more) decays exponentially in n as $n \to \infty$. Using (10) we have also that, as $n \to \infty$,

(13)
$$\mathbb{P}(|B_n - \beta n| \ge \epsilon n) \le 2\exp\big\{-\tfrac{1}{2}\epsilon^2 n[1 + \mathrm{o}(1)]\big\}. \qquad \bullet$$

(14) Example. Travelling salesman problem. A travelling salesman is required to visit n towns but may choose his route. How does he find the shortest possible route, and how long is it? Here is a randomized version of the problem. Let $P_1 = (U_1, V_1), P_2 = (U_2, V_2), \dots, P_n = (U_n, V_n)$ be independent and uniformly distributed points in the unit square $[0, 1]^2$; that is, suppose that $U_1, U_2, \dots, U_n, V_1, V_2, \dots, V_n$ are independent random variables each having the uniform distribution on $[0, 1]$. It is required to tour these points using an aeroplane. If we tour them in the order $P_{\pi(1)}, P_{\pi(2)}, \dots, P_{\pi(n)}$, for some permutation π of $\{1, 2, \dots, n\}$, the total length of the journey is

$$d(\pi) = \sum_{i=1}^{n-1} |P_{\pi(i+1)} - P_{\pi(i)}| + |P_{\pi(n)} - P_{\pi(1)}|$$

where $|\cdot|$ denotes Euclidean distance. The shortest tour has length $D_n = \min_\pi d(\pi)$. It turns out that the asymptotic behaviour of D_n for large n is given as follows: there exists a positive constant τ such that $D_n/\sqrt{n} \to \tau$ a.s. and in mean square. We shall not prove this, but note the consequence that

(15)
$$\frac{1}{\sqrt{n}}\mathbb{E}(D_n) \to \tau \quad \text{as} \quad n \to \infty.$$

How close is D_n to its mean? As in the case of bin packing, this question may be answered in part with the aid of Hoeffding's inequality. Once again, we set $Y_i = \mathbb{E}(D_n \mid \mathcal{F}_i)$ for $i \le n$, where $\mathcal{F}_i$ is the σ-field generated by $P_1, P_2, \dots, P_i$. As before, $(Y, \boldsymbol{\mathcal{F}})$ is a martingale, and $Y_n = D_n$, $Y_0 = \mathbb{E}(D_n)$.

Let $D_n(i)$ be the minimal tour-length through the points $P_1, P_2, \dots, P_{i-1}, P_{i+1}, \dots, P_n$, and note that $\mathbb{E}(D_n(i) \mid \mathcal{F}_i) = \mathbb{E}(D_n(i) \mid \mathcal{F}_{i-1})$. The vital inequality is

$$\textbf{(16)} \qquad D_n(i) \le D_n \le D_n(i) + 2Z_i, \quad i \le n-1,$$

where Z_i is the shortest distance from P_i to one of the points $P_{i+1}, P_{i+2}, \dots, P_n$. It is obvious that $D_n \ge D_n(i)$ since every tour of all n points includes a tour of the subset $P_1, \dots, P_{i-1}, P_{i+1}, \dots, P_n$. To obtain the second inequalityof (16), we argue as follows. Suppose that P_j is the closest point to P_i amongst the set $\{P_{i+1}, P_{i+2}, \dots, P_n\}$. One way of visiting all n points is to follow the optimal tour of $P_1, \dots, P_{i-1}, P_{i+1}, \dots, P_n$, and on arriving at P_j we make a return trip to P_i. The resulting trajectory is not quite a tour, but it can be turned into a tour by not landing at P_j on the return but going directly to the next point; the resulting tour has length no greater than $D_n(i) + 2Z_i$.

We take conditional expectations of (16) to obtain

$$\begin{aligned} \mathbb{E}\big(D_n(i) \,\big|\, \mathcal{F}_{i-1}\big) \le Y_{i-1} &\le \mathbb{E}\big(D_n(i) \,\big|\, \mathcal{F}_{i-1}\big) + 2\mathbb{E}(Z_i \mid \mathcal{F}_{i-1}), \\ \mathbb{E}\big(D_n(i) \,\big|\, \mathcal{F}_{i}\big) \le Y_{i} \quad &\le \mathbb{E}\big(D_n(i) \,\big|\, \mathcal{F}_{i}\big) + 2\mathbb{E}(Z_i \mid \mathcal{F}_{i}), \end{aligned}$$

and hence

$$\textbf{(17)} \qquad |Y_i - Y_{i-1}| \le 2\max\big\{\mathbb{E}(Z_i \mid \mathcal{F}_i), \mathbb{E}(Z_i \mid \mathcal{F}_{i-1})\big\}, \quad i \le n-1.$$

In order to estimate the right side here, let $Q \in [0,1]^2$, and let $Z_i(Q)$ be the shortest distance from Q to the closest of a collection of $n-i$ points chosen uniformly at random from the unit square. If $Z_i(Q) > x$ then no point lies within the circle $C(x, Q)$ having radius x and centre at Q. Note that $\sqrt{2}$ is the largest possible distance between two points in the square. Now, there exists c such that, for all $x \in (0, \sqrt{2}]$, the intersection of $C(x, Q)$ with the unit square has area at least cx^2, uniformly in Q. Therefore

$$\textbf{(18)} \qquad \mathbb{P}\big(Z_i(Q) > x\big) \le (1 - cx^2)^{n-i}, \quad 0 < x \le \sqrt{2}.$$

Integrating over x, we find that

$$\mathbb{E}(Z_i(Q)) \le \int_0^{\sqrt{2}} (1 - cx^2)^{n-i}\, dx \le \int_0^{\sqrt{2}} e^{-cx^2(n-i)}\, dx < \frac{C}{\sqrt{n-i}}$$

for some constant C; (*exercise*). Returning to (17), we deduce that the random variables $\mathbb{E}(Z_i \mid \mathcal{F}_i)$ and $\mathbb{E}(Z_i \mid \mathcal{F}_{i-1})$ are smaller than $C/\sqrt{n-i}$, whence $|Y_i - Y_{i-1}| \le 2C/\sqrt{n-i}$ for $i \le n-1$. For the case $i = n$, we use the trivial bound $|Y_n - Y_{n-1}| \le 2\sqrt{2}$, being twice the length of the diagonal of the square.

Applying Hoeffding's inequality, we obtain

$$\textbf{(19)} \qquad \begin{aligned} \mathbb{P}(|D_n - \mathbb{E}D_n| \ge x) &\le 2\exp\left(-\frac{x^2}{2(8 + \sum_{i=1}^{n-1} 4C^2/i)}\right) \\ &\le 2\exp(-Ax^2/\log n), \quad x > 0, \end{aligned}$$

for some positive constant A. Combining this with (15), we find that

$$\mathbb{P}(|D_n - \tau\sqrt{n}| \ge \epsilon\sqrt{n}) \le 2\exp(-B\epsilon^2 n/\log n), \quad \epsilon > 0,$$

for some positive constant B and all large n. ●

(20) Example. Markov chains. Let $X = \{X_n : n \ge 0\}$ be an irreducible aperiodic Markov chain on the finite state space S with transition matrix $\mathbf{P}$. Denote by $\boldsymbol{\pi}$ the stationary distribution of X, and suppose that X_0 has distribution $\boldsymbol{\pi}$, so that X is stationary. Fix a state $s \in S$, and let $N(n)$ be the number of visits of $X_1, X_2, \dots, X_n$ to s. The sequence N is a delayed renewal process and therefore $n^{-1}N(n) \xrightarrow{\text{a.s.}} \pi_s$ as $n \to \infty$. The convergence is rather fast, as the following (somewhat overcomplicated) argument indicates.

Let $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and, for $0 < m \le n$, let $\mathcal{F}_m = \sigma(X_1, X_2, \dots, X_m)$. Set $Y_m = \mathbb{E}(N(n) \mid \mathcal{F}_m)$ for $m \ge 0$, so that $(Y_m, \mathcal{F}_m)$ is a martingale. Note that $Y_n = N(n)$ and $Y_0 = \mathbb{E}(N(n)) = n\pi_s$ by stationarity.

We write $N(m, n) = N(n) - N(m)$, $0 \le m \le n$, the number of visits to s by the subsequence $X_{m+1}, X_{m+2}, \dots, X_n$. Now

$$Y_m = \mathbb{E}\big(N(m) \bigm| \mathcal{F}_m\big) + \mathbb{E}\big(N(m,n) \bigm| \mathcal{F}_m\big) = N(m) + \mathbb{E}\big(N(m,n) \bigm| X_m\big)$$

by the Markov property. Therefore, if $m \ge 1$,

$$\begin{aligned} Y_m - Y_{m-1} &= [N(m) - N(m-1)] + \big[\mathbb{E}\big(N(m,n) \bigm| X_m\big) - \mathbb{E}\big(N(m-1,n) \bigm| X_{m-1}\big)\big] \\ &= \mathbb{E}\big(N(m-1,n) \bigm| X_m\big) - \mathbb{E}\big(N(m-1,n) \bigm| X_{m-1}\big) \end{aligned}$$

since $N(m) - N(m-1) = \delta_{X_m, s}$, the Kronecker delta. It follows that

$$\begin{aligned} |Y_m - Y_{m-1}| &\le \max_{t,u \in S} \Big|\mathbb{E}\big(N(m-1,n) \bigm| X_m = t\big) - \mathbb{E}\big(N(m-1,n) \bigm| X_{m-1} = u\big)\Big| \\ &= \max_{t,u \in S} |D_m(t,u)| \end{aligned}$$

where, by the time homogeneity of the process,

(21) $$D_m(t,u) = \mathbb{E}\big(N(n-m+1) \bigm| X_1 = t\big) - \mathbb{E}\big(N(n-m+1) \bigm| X_0 = u\big).$$

It is easily seen that

$$\mathbb{E}\big(N(n-m+1) \bigm| X_1 = t\big) \le \delta_{ts} + \mathbb{E}(T_{tu}) + \mathbb{E}\big(N(n-m+1) \bigm| X_0 = u\big),$$

where $\mathbb{E}(T_{xy})$ is the mean first-passage time from state x to state y; just wait for the first passage to u, counting one for each moment which elapses. Similarly

$$\mathbb{E}\big(N(n-m+1) \bigm| X_0 = u\big) \le \mathbb{E}(T_{ut}) + \mathbb{E}\big(N(n-m+1) \bigm| X_1 = t\big).$$

Hence, by (21), $|D_m(t,u)| \le 1 + \max\{\mathbb{E}(T_{tu}), \mathbb{E}(T_{ut})\}$, implying that

(22) $$|Y_m - Y_{m-1}| \le 1 + \mu$$

where $\mu = \max\{\mathbb{E}(T_{xy}) : x, y \in S\}$; note that $\mu < \infty$ since S is finite. Applying Hoeffding's inequality, we deduce that

$$\mathbb{P}\big(|N(n) - n\pi_s| \geq x\big) \leq 2\exp\left(-\frac{x^2}{2n(\mu+1)}\right), \quad x > 0.$$

Setting $x = n\epsilon$, we obtain

$$\mathbb{P}\left(\left|\frac{1}{n}N(n) - \pi_s\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2(\mu+1)}\right), \quad \epsilon > 0, \tag{23}$$

a large-deviation estimate which decays exponentially fast as $n \to \infty$. Similar inequalities may be established by other means, more elementary than those used above. ●

Exercises for Section 12.2

1. **Knapsack problem.** It is required to pack a knapsack to maximum benefit. Suppose you have n objects, the ith object having volume V_i and worth W_i, where $V_1, V_2, \dots, V_n, W_1, W_2, \dots, W_n$ are independent non-negative random variables with finite means, and $W_i \leq M$ for all i and some fixed M. Your knapsack has volume c, and you wish to maximize the total worth of the objects packed in it. That is, you wish to find the vector $z_1, z_2, \dots, z_n$ of 0's and 1's such that $\sum_1^n z_i V_i \leq c$ and which maximizes $\sum_1^n z_i W_i$. Let Z be the maximal possible worth of the knapsack's contents, and show that $\mathbb{P}(|Z - \mathbb{E}Z| \geq x) \leq 2\exp\{-x^2/(2nM^2)\}$ for $x > 0$.

2. **Graph colouring.** Given n vertices $v_1, v_2, \dots, v_n$, for each $1 \leq i < j \leq n$ we place an edge between v_i and v_j with probability p; different pairs are joined independently of each other. We call v_i and v_j *neighbours* if they are joined by an edge. The *chromatic number* χ of the ensuing graph is the minimal number of pencils of different colours which are required in order that each vertex may be coloured differently from each of its neighbours. Show that $\mathbb{P}(|\chi - \mathbb{E}\chi| \geq x) \leq 2\exp\{-\frac{1}{2}x^2/n\}$ for $x > 0$.

12.3 Crossings and convergence

Martingales are of immense value in proving convergence theorems, and the following famous result has many applications.

(1) Martingale convergence theorem. *Let $(Y, \mathcal{F})$ be a submartingale and suppose that $\mathbb{E}(Y_n^+) \leq M$ for some M and all n. There exists a random variable Y_∞ such that $Y_n \xrightarrow{\text{a.s.}} Y_\infty$ as $n \to \infty$. We have in addition that:*

(i) *Y_∞ has finite mean if $\mathbb{E}|Y_0| < \infty$, and*

(ii) *$Y_n \xrightarrow{1} Y_\infty$ if the sequence $\{Y_n : n \geq 0\}$ is uniformly integrable.*

It follows of course that any submartingale or supermartingale $(Y, \mathcal{F})$ converges almost surely if it satisfies $\mathbb{E}|Y_n| \leq M$.

The key step in the classical proof of this theorem is 'Snell's upcrossings inequality'. Suppose that $y = \{y_n : n \geq 0\}$ is a real sequence, and $[a, b]$ is a real interval. An upcrossing of $[a, b]$ is defined to be a crossing by y of $[a, b]$ in the upwards direction. More precisely, we define $T_1 = \min\{n : y_n \leq a\}$, the first time that y hits the interval $(-\infty, a]$, and

$T_2 = \min\{n > T_1 : y_n \geq b\}$, the first subsequent time when y hits $[b, \infty)$; we call the interval $[T_1, T_2]$ an *upcrossing* of $[a, b]$. In addition, let

$$T_{2k-1} = \min\{n > T_{2k-2} : y_n \leq a\}, \quad T_{2k} = \min\{n > T_{2k-1} : y_n \geq b\},$$

for $k \geq 2$, so that the upcrossings of $[a, b]$ are the intervals $[T_{2k-1}, T_{2k}]$ for $k \geq 1$. Let $U_n(a, b; y)$ be the number of upcrossings of $[a, b]$ by the subsequence $y_0, y_1, \dots, y_n$, and let $U(a, b; y) = \lim_{n\to\infty} U_n(a, b; y)$ be the total number of such upcrossings by y.

(2) Lemma. *If $U(a, b; y) < \infty$ for all rationals a and b satisfying $a < b$, then $\lim_{n\to\infty} y_n$ exists (but may be infinite).*

Proof. If $\lambda = \liminf_{n\to\infty} y_n$ and $\mu = \limsup_{n\to\infty} y_n$ satisfy $\lambda < \mu$ then there exist rationals a, b such that $\lambda < a < b < \mu$. Now $y_n \leq a$ for infinitely many n, and $y_n \geq b$ similarly, implying that $U(a, b; y) = \infty$, a contradiction. Therefore $\lambda = \mu$. ■

Suppose now that $(Y, \mathcal{F})$ is a submartingale, and let $U_n(a, b; Y)$ be the number of upcrossings of $[a, b]$ by Y up to time n.

(3) Theorem. Upcrossings inequality. *If $a < b$ then*

$$\mathbb{E}U_n(a, b; Y) \leq \frac{\mathbb{E}((Y_n - a)^+)}{b - a}.$$

Proof. Setting $Z_n = (Y_n - a)^+$, we have by Exercise (12.1.7) that $(Z, \mathcal{F})$ is a non-negative submartingale. Upcrossings by Y of $[a, b]$ correspond to upcrossings by Z of $[0, b - a]$, so that $U_n(a, b; Y) = U_n(0, b - a; Z)$.

Let $[T_{2k-1}, T_{2k}]$, $k \geq 1$, be the upcrossings by Z of $[0, b - a]$, and define the indicator functions

$$I_i = \begin{cases} 1 & \text{if } i \in (T_{2k-1}, T_{2k}] \text{ for some } k, \\ 0 & \text{otherwise.} \end{cases}$$

Note that I_i is $\mathcal{F}_{i-1}$-measurable, since

$$\{I_i = 1\} = \bigcup_k \{T_{2k-1} \leq i - 1\} \setminus \{T_{2k} \leq i - 1\},$$

an event which depends on $Y_0, Y_1, \dots, Y_{i-1}$ only. Now

$$(4) \qquad (b - a)U_n(0, b - a; Z) \leq \mathbb{E}\left(\sum_{i=1}^{n} (Z_i - Z_{i-1})I_i\right),$$

since each upcrossing of $[0, b - a]$ contributes an amount of at least $b - a$ to the summation. However

$$(5) \qquad \begin{aligned} \mathbb{E}\big((Z_i - Z_{i-1})I_i\big) &= \mathbb{E}\big(\mathbb{E}[(Z_i - Z_{i-1})I_i \mid \mathcal{F}_{i-1}]\big) = \mathbb{E}\big(I_i[\mathbb{E}(Z_i \mid \mathcal{F}_{i-1}) - Z_{i-1}]\big) \\ &\leq \mathbb{E}\big[\mathbb{E}(Z_i \mid \mathcal{F}_{i-1}) - Z_{i-1}\big] = \mathbb{E}(Z_i) - \mathbb{E}(Z_{i-1}) \end{aligned}$$

where we have used the fact that Z is a submartingale to obtain the inequality. Summing over i, we obtain from (4) that

$$(b - a)U_n(0, b - a; Z) \leq \mathbb{E}(Z_n) - \mathbb{E}(Z_0) \leq \mathbb{E}(Z_n)$$

and the lemma is proved. ■

Proof of Theorem (1). Suppose $(Y, \mathcal{F})$ is a submartingale and $\mathbb{E}(Y_n^+) \le M$ for all n. We have from the upcrossings inequality that, if $a < b$,

$$\mathbb{E}U_n(a, b; Y) \le \frac{\mathbb{E}(Y_n^+) + |a|}{b - a}$$

so that $U(a, b; Y) = \lim_{n\to\infty} U_n(a, b; Y)$ satisfies

$$\mathbb{E}U(a, b; Y) = \lim_{n\to\infty} \mathbb{E}U_n(a, b; Y) \le \frac{M + |a|}{b - a}$$

for all $a < b$. Therefore $U(a, b; Y) < \infty$ a.s. for all $a < b$. Since there are only countably many rationals, it follows that, with probability 1, $U(a, b; Y) < \infty$ for all rational a and b. By Lemma (2), the sequence Y_n converges almost surely to some limit Y_∞. We argue as follows to show that $\mathbb{P}(|Y_\infty| < \infty) = 1$. Since $|Y_n| = 2Y_n^+ - Y_n$ and $\mathbb{E}(Y_n \mid \mathcal{F}_0) \ge Y_0$, we have that

$$\mathbb{E}\big(|Y_n| \bigm| \mathcal{F}_0\big) = 2\mathbb{E}(Y_n^+ \mid \mathcal{F}_0) - \mathbb{E}(Y_n \mid \mathcal{F}_0) \le 2\mathbb{E}(Y_n^+ \mid \mathcal{F}_0) - Y_0.$$

By Fatou's lemma,

$$\text{(6)} \qquad \mathbb{E}\big(|Y_\infty| \bigm| \mathcal{F}_0\big) = \mathbb{E}\left(\liminf_{n\to\infty} |Y_n| \,\middle|\, \mathcal{F}_0\right) \le \liminf_{n\to\infty} \mathbb{E}\big(|Y_n| \bigm| \mathcal{F}_0\big) \le 2Z - Y_0$$

where $Z = \liminf_{n\to\infty} \mathbb{E}(Y_n^+ \mid \mathcal{F}_0)$. However $\mathbb{E}(Z) \le M$ by Fatou's lemma, so that $Z < \infty$ a.s., implying that $\mathbb{E}\big(|Y_\infty| \bigm| \mathcal{F}_0\big) < \infty$ a.s. Hence $\mathbb{P}\big(|Y_\infty| < \infty \bigm| \mathcal{F}_0\big) = 1$, and therefore

$$\mathbb{P}(|Y_\infty| < \infty) = \mathbb{E}\big[\mathbb{P}\big(|Y_\infty| < \infty \bigm| \mathcal{F}_0\big)\big] = 1.$$

If $\mathbb{E}|Y_0| < \infty$, we may take expectations of (6) to obtain $\mathbb{E}|Y_\infty| \le 2M - \mathbb{E}(Y_0) < \infty$. That uniform integrability is enough to ensure convergence in mean is a consequence of Theorem (7.10.3). ■

The following is an immediate corollary of the martingale convergence theorem.

(7) Theorem. *If $(Y, \mathcal{F})$ is either a non-negative supermartingale or a non-positive submartingale, then $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely.*

Proof. If Y is a non-positive submartingale then $\mathbb{E}(Y_n^+) = 0$, whence the result follows from Theorem (1). For a non-negative supermartingale Y, apply the same argument to $-Y$. ■

(8) Example. Random walk. Consider de Moivre's martingale of Example (12.1.4), namely $Y_n = (q/p)^{S_n}$ where S_n is the position after n steps of the usual simple random walk. The sequence $\{Y_n\}$ is a non-negative martingale, and hence converges almost surely to some finite limit Y as $n \to \infty$. This is not of much interest if $p = q$, since $Y_n = 1$ for all n in this case. Suppose then that $p \ne q$. The random variable Y_n takes values in the set $\{\rho^k : k = 0, \pm 1, \dots\}$ where $\rho = q/p$. Certainly Y_n cannot converge to any given (possibly random) member of this set, since this would necessarily entail that S_n converges to a finite limit (which is obviously false). Therefore Y_n converges to a limit point of the set, not lying within the set. The only such limit point which is finite is 0, and therefore $Y_n \to 0$ a.s. Hence, $S_n \to -\infty$ a.s. if $p < q$, and $S_n \to \infty$ a.s. if $p > q$. Note that Y_n does not converge in mean, since $\mathbb{E}(Y_n) = \mathbb{E}(Y_0) \neq 0$ for all n. ●

(9) Example. Doob's martingale (though some ascribe the construction to Lévy). Let Z be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}|Z| < \infty$. Suppose that $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_0, \mathcal{F}_1, \dots\}$ is a filtration, and write $\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n$ for the smallest σ-field containing every $\mathcal{F}_n$. Now define $Y_n = \mathbb{E}(Z \mid \mathcal{F}_n)$. It is easily seen that $(Y, \boldsymbol{\mathcal{F}})$ is a martingale. First, by Jensen's inequality,

$$\mathbb{E}|Y_n| = \mathbb{E}\big|\mathbb{E}(Z \mid \mathcal{F}_n)\big| \le \mathbb{E}\big\{\mathbb{E}\big(|Z| \bigm| \mathcal{F}_n\big)\big\} = \mathbb{E}|Z| < \infty,$$

and secondly

$$\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\big[\mathbb{E}(Z \mid \mathcal{F}_{n+1}) \bigm| \mathcal{F}_n\big] = \mathbb{E}(Z \mid \mathcal{F}_n)$$

since $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. Furthermore $\{Y_n\}$ is a uniformly integrable sequence, as shown in Example (7.10.13). It follows by the martingale convergence theorem that $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely and in mean.

It is actually the case that $Y_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$, so that

(10) $$\mathbb{E}(Z \mid \mathcal{F}_n) \to \mathbb{E}(Z \mid \mathcal{F}_\infty) \quad \text{a.s. and in mean.}$$

To see this, one argues as follows. Let N be a positive integer. First, $Y_n I_A \to Y_\infty I_A$ a.s. for all $A \in \mathcal{F}_N$. Now $\{Y_n I_A : n \ge N\}$ is uniformly integrable, and therefore $\mathbb{E}(Y_n I_A) \to \mathbb{E}(Y_\infty I_A)$ for all $A \in \mathcal{F}_N$. On the other hand $\mathbb{E}(Y_n I_A) = \mathbb{E}(Y_N I_A) = \mathbb{E}(Z I_A)$ for all $n \ge N$ and all $A \in \mathcal{F}_N$, by the definition of conditional expectation. Hence $\mathbb{E}(Z I_A) = \mathbb{E}(Y_\infty I_A)$ for all $A \in \mathcal{F}_N$. Letting $N \to \infty$ and using a standard result of measure theory, we find that $\mathbb{E}((Z - Y_\infty)I_A) = 0$ for all $A \in \mathcal{F}_\infty$, whence $Y_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$.

There is an important converse to these results.

(11) Lemma. *Let $(Y, \boldsymbol{\mathcal{F}})$ be a martingale. Then Y_n converges in mean if and only if there exists a random variable Z with finite mean such that $Y_n = \mathbb{E}(Z \mid \mathcal{F}_n)$. If $Y_n \xrightarrow{1} Y_\infty$, then $Y_n = \mathbb{E}(Y_\infty \mid \mathcal{F}_n)$.*

If such a random variable Z exists, we say that the martingale $(Y, \boldsymbol{\mathcal{F}})$ is *closed*.

Proof. In the light of the previous discussion, it suffices to prove that, if $(Y, \boldsymbol{\mathcal{F}})$ is a martingale which converges in mean to Y_∞, then $Y_n = \mathbb{E}(Y_\infty \mid \mathcal{F}_n)$. For any positive integer N and event $A \in \mathcal{F}_N$, it is the case that $\mathbb{E}(Y_n I_A) \to \mathbb{E}(Y_\infty I_A)$; just note that $Y_n I_A \xrightarrow{1} Y_\infty I_A$ since

$$\mathbb{E}|(Y_n - Y_\infty)I_A| \le \mathbb{E}|Y_n - Y_\infty| \to 0 \quad \text{as} \quad n \to \infty.$$

On the other hand, $\mathbb{E}(Y_n I_A) = \mathbb{E}(Y_N I_A)$ for $n \ge N$ and $A \in \mathcal{F}_N$, by the martingale property. It follows that $\mathbb{E}(Y_\infty I_A) = \mathbb{E}(Y_N I_A)$ for all $A \in \mathcal{F}_N$, which is to say that $Y_N = \mathbb{E}(Y_\infty \mid \mathcal{F}_N)$ as required. ■●

(12) Example. Zero–one law (7.3.12). Let $X_0, X_1, \dots$ be independent random variables, and let $\mathcal{T}$ be their tail σ-field; that is to say, $\mathcal{T} = \bigcap_n \mathcal{H}_n$ where $\mathcal{H}_n = \sigma(X_n, X_{n+1}, \dots)$. Here is a proof that, for all $A \in \mathcal{T}$, either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Let $A \in \mathcal{T}$ and define $Y_n = \mathbb{E}(I_A \mid \mathcal{F}_n)$ where $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Now $A \in \mathcal{T} \subseteq \mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n$, and therefore $Y_n \to \mathbb{E}(I_A \mid \mathcal{F}_\infty) = I_A$ a.s. and in mean, by (11). On the other hand $Y_n = \mathbb{E}(I_A \mid \mathcal{F}_n) = \mathbb{P}(A)$, since A ($\in \mathcal{T}$) is independent of all events in $\mathcal{F}_n$. Hence $\mathbb{P}(A) = I_A$ almost surely, which is to say that I_A is almost surely constant. However, I_A takes values 0 and 1 only, and therefore either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. ●

This completes the main contents of this section. We terminate it with one further result of interest, being a bound related to the upcrossings inequality. For a certain type of process, one may obtain rather tight bounds on the tail of the number of upcrossings.

(13) Theorem. Dubins's inequality. *Let $(Y, \mathcal{F})$ be a non-negative supermartingale. Then*

(14) $$\mathbb{P}\{U_n(a, b; Y) \ge j\} \le \left(\frac{a}{b}\right)^j \mathbb{E}\big(\min\{1, Y_0/a\}\big)$$

for $0 < a < b$ and $j \ge 0$.

Summing (14) over j, we find that

(15) $$\mathbb{E}U_n(a, b; Y) \le \frac{a}{b-a}\mathbb{E}\big(\min\{1, Y_0/a\}\big),$$

an inequality which may be compared with the upcrossings inequality (3).

Proof. This is achieved by an adaptation of the proof of the upcrossings inequality (3), and we use the notation of that proof. Fix a positive integer j. We replace the indicator function I_i by the random variable

$$J_i = \begin{cases} a^{-1}(b/a)^{k-1} & \text{if } i \in (T_{2k-1}, T_{2k}] \text{ for some } k \le j, \\ 0 & \text{otherwise.} \end{cases}$$

Next we let $X_0, X_1, \dots$ be given by $X_0 = \min\{1, Y_0/a\}$,

(16) $$X_n = X_0 + \sum_{i=1}^{n} J_i(Y_i - Y_{i-1}), \qquad n \ge 1.$$

If $T_{2j} \le n$, then

$$X_n \ge X_0 + \sum_{k=1}^{j} a^{-1}(b/a)^{k-1}(Y_{T_{2k}} - Y_{T_{2k-1}}).$$

However, $Y_{T_{2k}} \ge b$ and $Y_{T_{2k+1}} \le a$, so that

(17) $$Y_{T_{2k}} - \frac{b}{a}Y_{T_{2k+1}} \ge 0,$$

implying that

$$X_n \ge X_0 + a^{-1}(b/a)^{j-1}Y_{T_{2j}} - a^{-1}Y_{T_1}, \quad \text{if} \quad T_{2j} \le n.$$

If $Y_0 \le a$ then $T_1 = 0$ and $X_0 - a^{-1}Y_{T_1} = 0$; on the other hand, if $Y_0 > a$ then $X_0 - a^{-1}Y_{T_1} = 1 - a^{-1}Y_{T_1} > 0$. In either case it follows that

(18) $$X_n \ge (b/a)^j \quad \text{if} \quad T_{2j} \le n.$$

Now Y is a non-negative sequence, and hence $X_n \geq X_0 - a^{-1}Y_{T_1} \geq 0$ by (16) and (17). Take expectations of (18) to obtain

$$\mathbb{E}(X_n) \geq (b/a)^j \mathbb{P}\big(U_n(a,b;Y) \geq j\big), \tag{19}$$

and it remains to bound $\mathbb{E}(X_n)$ above. Arguing as in (5) and using the supermartingale property, we arrive at

$$\mathbb{E}(X_n) = \mathbb{E}(X_0) + \sum_{i=1}^{n} \mathbb{E}\big(J_i(Y_i - Y_{i-1})\big) \leq \mathbb{E}(X_0).$$

The conclusion of the theorem follows from (19). ■

(20) Example. Simple random walk. Consider de Moivre's martingale $Y_n = (q/p)^{S_n}$ of Examples (12.1.4) and (8), with $p < q$. By Theorem (13), $\mathbb{P}(U_n(a,b;Y) \geq j) \leq (a/b)^j$. An upcrossing of $[a,b]$ by Y corresponds to an upcrossing of $[\log a, \log b]$ by S (with logarithms to the base q/p). Hence

$$\mathbb{P}\big(U_n(0,r;S) \geq j\big) = \mathbb{P}\big\{U_n\big(1,(q/p)^r;Y\big) \geq j\big\} \leq (p/q)^{rj}, \quad j \geq 0.$$

Actually equality holds here in the limit as $n \to \infty$: $\mathbb{P}(U(0,r;S) \geq j) = (p/q)^{rj}$ for positive integers r; see Exercise (5.3.1). ●

Exercises for Section 12.3

1. Give a reasonable definition of a *downcrossing* of the interval $[a,b]$ by the random sequence $Y_0, Y_1, \dots$.

(a) Show that the number of downcrossings differs from the number of upcrossings by at most 1.

(b) If $(Y, \mathcal{F})$ is a submartingale, show that the number $D_n(a,b;Y)$ of downcrossings of $[a,b]$ by Y up to time n satisfies

$$\mathbb{E}D_n(a,b;Y) \leq \frac{\mathbb{E}\{(Y_n - b)^+\}}{b-a}.$$

2. Let $(Y, \mathcal{F})$ be a supermartingale with finite means, and let $U_n(a,b;Y)$ be the number of upcrossings of the interval $[a,b]$ up to time n. Show that

$$\mathbb{E}U_n(a,b;Y) \leq \frac{\mathbb{E}\{(Y_n - a)^-\}}{b-a}.$$

Deduce that $\mathbb{E}U_n(a,b;Y) \leq a/(b-a)$ if Y is non-negative and $a \geq 0$.

3. Let X be a Markov chain with countable state space S and transition matrix $\mathbf{P}$. Suppose that X is irreducible and persistent, and that $\psi : S \to S$ is a bounded function satisfying $\sum_{j\in S} p_{ij}\psi(j) \leq \psi(i)$ for $i \in S$. Show that ψ is a constant function.

4. Let $Z_1, Z_2, \dots$ be independent random variables such that:

$$Z_n = \begin{cases} a_n & \text{with probability } \frac{1}{2}n^{-2}, \\ 0 & \text{with probability } 1 - n^{-2}, \\ -a_n & \text{with probability } \frac{1}{2}n^{-2}, \end{cases}$$

where $a_1 = 2$ and $a_n = 4\sum_{j=1}^{n-1} a_j$. Show that $Y_n = \sum_{j=1}^{n} Z_j$ defines a martingale. Show that $Y = \lim Y_n$ exists almost surely, but that there exists no M such that $\mathbb{E}|Y_n| \leq M$ for all n.

12.4 Stopping times

We are all called upon on occasion to take an action whose nature is fixed but whose timing is optional. Commonly occurring examples include getting married or divorced, employing a secretary, having a baby, and buying a house. An important feature of such actions is that they are taken in the light of the past and present, and they may not depend on the future. Other important examples arise in considering money markets. The management of portfolios is affected by such rules as: (a) sell a currency if it weakens to a predetermined threshold, (b) buy government bonds if the exchange index falls below a given level, and so on. (Such rules are often sufficiently simple to be left to computers to implement, with occasionally spectacular consequences†.)

A more mathematical example is provided by the gambling analogy. A gambler pursues a strategy which we may assume to be based upon his experience rather than his clairvoyance. That is to say, his decisions to vary his stake (or to stop gambling altogether) depend on the outcomes of the game up to the time of the decision, and no further. A gambler is able to follow the rule 'stop when ahead' but cannot be expected to follow a rule such as 'stop just before a loss'.

Such actions have the common feature that, at any time, we have sufficient information to decide whether or not to take the action *at that time*. The usual way of expressing this property in mathematical terms is as follows. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_0, \mathcal{F}_1, \dots\}$ be a filtration. We think of $\mathcal{F}_n$ as representing the information which is available at time n, or more precisely the smallest σ-field with respect to which all observations up to and including time n are measurable.

(1) Definition. A random variable T taking values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ is called a **stopping time** (with respect to the filtration $\boldsymbol{\mathcal{F}}$) if $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

Note that stopping times T satisfy

$$\{T > n\} = \{T \leq n\}^{\mathrm{c}} \in \mathcal{F}_n \quad \text{for all } n, \tag{2}$$

since $\boldsymbol{\mathcal{F}}$ is a filtration. They are not required to be finite, but may take the value ∞. Stopping times are sometimes called *Markov times*. They were discussed in Section 6.8 in the context of birth processes.

Given a filtration $\boldsymbol{\mathcal{F}}$ and a stopping time T, it is useful to introduce some notation to represent information gained up to the random time T. We denote by $\mathcal{F}_T$ the collection of all events A such that $A \cap \{T \leq n\} \in \mathcal{F}_n$ for all n. It is easily seen that $\mathcal{F}_T$ is a σ-field, and we think of $\mathcal{F}_T$ as the set of events whose occurrence or non-occurrence is known by time T.

(3) Example. The martingale (12.1.3). A fair coin is tossed repeatedly; let T be the time of the first head. Writing X_i for the number of heads on the ith toss, we have that

$$\{T = n\} = \{X_n = 1,\ X_j = 0 \text{ for } 1 \leq j < n\} \in \mathcal{F}_n$$

where $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Therefore T is a stopping time. In this case T is finite almost surely. ●

---

†At least one NYSE crash has been attributed to the use of simple online stock-dealing systems programmed to sell whenever a stock price falls to a given threshold. Such systems can be subject to feedback, and the rules have been changed to inhibit this.

(4) Example. First passage times. Let $\mathcal{F}$ be a filtration and let the random sequence X be adapted to $\mathcal{F}$, so that X_n is $\mathcal{F}_n$-measurable. For each (sufficiently nice) subset B of $\mathbb{R}$ define the *first passage time* of X to B by $T_B = \min\{n : X_n \in B\}$ with $T_B = \infty$ if $X_n \notin B$ for all n. It is easily seen that T_B is a stopping time. ●

Stopping times play an important role in the theory of martingales, as illustrated in the following examples. First, a martingale which is stopped at a random time T remains a martingale, so long as T is a stopping time.

(5) Theorem. *Let $(Y, \mathcal{F})$ be a submartingale and let T be a stopping time (with respect to $\mathcal{F}$). Then $(Z, \mathcal{F})$, defined by $Z_n = Y_{T\wedge n}$, is a submartingale.*

Here, as usual, we use the notation $x \wedge y = \min\{x, y\}$. If $(Y, \mathcal{F})$ is a martingale, then it is both a submartingale and a supermartingale, whence $Y_{T\wedge n}$ constitutes a martingale, by (5).

Proof. We may write

$$Z_n = \sum_{t=0}^{n-1} Y_t I_{\{T=t\}} + Y_n I_{\{T\geq n\}}, \tag{6}$$

whence Z_n is $\mathcal{F}_n$-measurable (using (2)) and

$$\mathbb{E}(Z_n^+) \leq \sum_{t=0}^{n} \mathbb{E}(Y_t^+) < \infty.$$

Also, from (6), $Z_{n+1} - Z_n = (Y_{n+1} - Y_n)I_{\{T>n\}}$, whence, using (2) and the submartingale property,

$$\mathbb{E}(Z_{n+1} - Z_n \mid \mathcal{F}_n) = \mathbb{E}(Y_{n+1} - Y_n \mid \mathcal{F}_n)I_{\{T>n\}} \geq 0. \qquad \blacksquare$$

One strategy open to a gambler in a casino is to change the game (think of the gambler as an investor in stocks, if you wish). If he is fortunate enough to be playing fair games, then he should not gain or lose (on average) at such a change. More formally, let $(X, \mathcal{F})$ and $(Y, \mathcal{F})$ be two martingales with respect to the filtration $\mathcal{F}$. Let T be a stopping time with respect to $\mathcal{F}$; T is the switching time from X to Y, and X_T is the 'capital' which is carried forward.

(7) Theorem. Optional switching. *Suppose that $X_T = Y_T$ on the event $\{T < \infty\}$. Then*

$$Z_n = \begin{cases} X_n & \text{if } n < T, \\ Y_n & \text{if } n \geq T, \end{cases}$$

defines a martingale with respect to $\mathcal{F}$.

Proof. We have that

$$Z_n = X_n I_{\{n<T\}} + Y_n I_{\{n\geq T\}}; \tag{8}$$

each summand is $\mathcal{F}_n$-measurable, and hence Z_n is $\mathcal{F}_n$-measurable. Also $\mathbb{E}|Z_n| \leq \mathbb{E}|X_n| + \mathbb{E}|Y_n| < \infty$. By the martingale property of X and Y,

$$\begin{aligned} Z_n &= \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)I_{\{n<T\}} + \mathbb{E}(Y_{n+1} \mid \mathcal{F}_n)I_{\{n\geq T\}} \\ &= \mathbb{E}\big(X_{n+1}I_{\{n<T\}} + Y_{n+1}I_{\{n\geq T\}} \,\big|\, \mathcal{F}_n\big), \end{aligned} \tag{9}$$

since T is a stopping time. Now

$$\textbf{(10)}\qquad \begin{aligned} X_{n+1}I_{\{n<T\}} + Y_{n+1}I_{\{n\ge T\}} &= Z_{n+1} + X_{n+1}I_{\{n+1=T\}} - Y_{n+1}I_{\{n+1=T\}} \\ &= Z_{n+1} + (X_T - Y_T)I_{\{n+1=T\}} \end{aligned}$$

whence, by (9) and the assumption that $X_T = Y_T$ on the event $\{T < \infty\}$, we have that $Z_n = \mathbb{E}(Z_{n+1} \mid \mathcal{F}_n)$, so that $(Z, \boldsymbol{\mathcal{F}})$ is a martingale. ■

'Optional switching' does not disturb the martingale property. 'Optional sampling' can be somewhat more problematical. Let $(Y, \boldsymbol{\mathcal{F}})$ be a martingale and let $T_1, T_2, \dots$ be a sequence of stopping times satisfying $T_1 \le T_2 \le \cdots < \infty$. Let $Z_0 = Y_0$ and $Z_n = Y_{T_n}$, so that the sequence Z is obtained by 'sampling' the sequence Y at the stopping times T_j. It is natural to set $\mathcal{H}_n = \mathcal{F}_{T_n}$, and to ask whether $(Z, \boldsymbol{\mathcal{H}})$ is a martingale. The answer in general is no. To see this, use the simple example when Y_n is the excess of heads over tails in n tosses of a fair coin, with $T_1 = \min\{n : Y_n = 1\}$; for this example $\mathbb{E}Y_0 = 0$ but $\mathbb{E}Y_{T_1} = 1$. The answer is, however, affirmative if the T_j are bounded.

(11) Optional sampling theorem. *Let $(Y, \boldsymbol{\mathcal{F}})$ be a submartingale.*

(a) *If T is a stopping time and there exists a deterministic $N\ (< \infty)$ such that $\mathbb{P}(T \le N) = 1$, then $\mathbb{E}(Y_T^+) < \infty$ and $\mathbb{E}(Y_T \mid \mathcal{F}_0) \ge Y_0$.*

(b) *If $T_1 \le T_2 \le \cdots$ is a sequence of stopping times such that $\mathbb{P}(T_j \le N_j) = 1$ for some deterministic real sequence N_j, then $(Z, \boldsymbol{\mathcal{H}})$, defined by $(Z_0, \mathcal{H}_0) = (Y_0, \mathcal{F}_0)$, $(Z_j, \mathcal{H}_j) = (Y_{T_j}, \mathcal{F}_{T_j})$, is a submartingale.*

If $(Y, \boldsymbol{\mathcal{F}})$ is a martingale, then it is both a submartingale and a supermartingale; Theorem (11) then implies that $\mathbb{E}(Y_T \mid \mathcal{F}_0) = Y_0$ for any bounded stopping time T, and furthermore $(Y_{T_j}, \mathcal{F}_{T_j})$ is a martingale for any increasing sequence $T_1, T_2, \dots$ of bounded stopping times.

Proof. Part (b) may be obtained without great difficulty by repeated application of part (a), and we therefore confine outselves to proving (a). Suppose $\mathbb{P}(T \le N) = 1$. Let $Z_n = Y_{T\wedge n}$, so that $(Z, \boldsymbol{\mathcal{F}})$ is a submartingale, by (5). Therefore $\mathbb{E}(Z_N^+) < \infty$ and

$$\textbf{(12)}\qquad \mathbb{E}(Z_N \mid \mathcal{F}_0) \ge Z_0 = Y_0,$$

and the proof is finished by observing that $Z_N = Y_{T\wedge N} = Y_T$ a.s. ■

Certain inequalities are of great value when studying the asymptotic properties of martingales. The following simple but powerful 'maximal inequality' is an easy consequence of the optional sampling theorem.

(13) Theorem. *Let $(Y, \boldsymbol{\mathcal{F}})$ be a martingale. For $x > 0$,*

$$\textbf{(14)}\qquad \mathbb{P}\Big(\max_{0\le m\le n} Y_m \ge x\Big) \le \frac{\mathbb{E}(Y_n^+)}{x} \quad \textit{and} \quad \mathbb{P}\Big(\max_{0\le m\le n} |Y_m| \ge x\Big) \le \frac{\mathbb{E}|Y_n|}{x}.$$

Proof. Let $x > 0$, and let $T = \min\{m : Y_m \ge x\}$ be the first passage time of Y above the level x. Then $T \wedge n$ is a bounded stopping time, and therefore $\mathbb{E}(Y_0) = \mathbb{E}(Y_{T\wedge n}) = \mathbb{E}(Y_n)$ by Theorem (11a) and the martingale property. Now $\mathbb{E}(Y_{T\wedge n}) = \mathbb{E}(Y_T I_{\{T\le n\}} + Y_n I_{\{T>n\}})$. However,

$$\mathbb{E}(Y_T I_{\{T\le n\}}) \ge x\mathbb{E}(I_{\{T\le n\}}) = x\mathbb{P}(T \le n)$$

since $Y_T \geq x$, and therefore

$$\mathbb{E}(Y_n) = \mathbb{E}(Y_{T \wedge n}) \geq x\mathbb{P}(T \leq n) + \mathbb{E}(Y_n I_{\{T>n\}}), \tag{15}$$

whence

$$x\mathbb{P}(T \leq n) \leq \mathbb{E}(Y_n I_{\{T \leq n\}}) \leq \mathbb{E}(Y_n^+)$$

as required for the first part of (14). As for the second part, just note that $(-Y, \mathcal{F})$ is a martingale, so that

$$\mathbb{P}\left(\max_{0 \leq m \leq n} \{-Y_m\} \geq x\right) \leq \frac{\mathbb{E}(Y_n^-)}{x} \quad \text{for} \quad x > 0,$$

which may be added to the first part. ∎

We shall explore maximal inequalities for submartingales and supermartingales in the forthcoming Section 12.6.

Exercises for Section 12.4

1. If T_1 and T_2 are stopping times with respect to a filtration $\mathcal{F}$, show that $T_1 + T_2$, $\max\{T_1, T_2\}$, and $\min\{T_1, T_2\}$ are stopping times also.

2. Let $X_1, X_2, \dots$ be a sequence of non-negative independent random variables and let $N(t) = \max\{n : X_1 + X_2 + \cdots + X_n \leq t\}$. Show that $N(t) + 1$ is a stopping time with respect to a suitable filtration to be specified.

3. Let $(Y, \mathcal{F})$ be a submartingale and $x > 0$. Show that

$$\mathbb{P}\left(\max_{0 \leq m \leq n} Y_m \geq x\right) \leq \frac{1}{x}\mathbb{E}(Y_n^+).$$

4. Let $(Y, \mathcal{F})$ be a non-negative supermartingale and $x > 0$. Show that

$$\mathbb{P}\left(\max_{0 \leq m \leq n} Y_m \geq x\right) \leq \frac{1}{x}\mathbb{E}(Y_0).$$

5. Let $(Y, \mathcal{F})$ be a submartingale and let S and T be stopping times satisfying $0 \leq S \leq T \leq N$ for some deterministic N. Show that $\mathbb{E}Y_0 \leq \mathbb{E}Y_S \leq \mathbb{E}Y_T \leq \mathbb{E}Y_N$.

6. Let $\{S_n\}$ be a simple random walk with $S_0 = 0$ such that $0 < p = \mathbb{P}(S_1 = 1) < \frac{1}{2}$. Use de Moivre's martingale to show that $\mathbb{E}(\sup_m S_m) \leq p/(1-2p)$. Show further that this inequality may be replaced by an equality.

7. Let $\mathcal{F}$ be a filtration. For any stopping time T with respect to $\mathcal{F}$, denote by $\mathcal{F}_T$ the collection of all events A such that, for all n, $A \cap \{T \leq n\} \in \mathcal{F}_n$. Let S and T be stopping times.
(a) Show that $\mathcal{F}_T$ is a σ-field, and that T is measurable with respect to this σ-field.
(b) If $A \in \mathcal{F}_S$, show that $A \cap \{S \leq T\} \in \mathcal{F}_T$.
(c) Let S and T satisfy $S \leq T$. Show that $\mathcal{F}_S \subseteq \mathcal{F}_T$.

12.5 Optional stopping

If you stop a martingale $(Y, \mathcal{F})$ at a fixed time n, the mean value $\mathbb{E}(Y_n)$ satisfies $\mathbb{E}(Y_n) = \mathbb{E}(Y_0)$. Under what conditions is this true if you stop after a *random* time T; that is, when is it the case that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$? The answer to this question is very valuable in studying first-passage properties of martingales (see (12.1.4) for example). It would be unreasonable to expect such a result to hold generally unless T is required to be a stopping time.

Let T be a stopping time which is finite (in that $\mathbb{P}(T < \infty) = 1$), and let $(Y, \mathcal{F})$ be a martingale. Then $T \wedge n \to T$ as $n \to \infty$, so that $Y_{T\wedge n} \to Y_T$ a.s. It follows (as in Theorem (7.10.3)) that $\mathbb{E}(Y_0) = \mathbb{E}(Y_{T\wedge n}) \to \mathbb{E}(Y_T)$ so long as the family $\{Y_{T\wedge n} : n \geq 0\}$ is uniformly integrable.

The following two theorems provide useful conditions which are sufficient for the conclusion $\mathbb{E}(Y_0) = \mathbb{E}(Y_T)$.

(1) Optional stopping theorem. *Let $(Y, \mathcal{F})$ be a martingale and let T be a stopping time. Then $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$ if:*
 (a) $\mathbb{P}(T < \infty) = 1$,
 (b) $\mathbb{E}|Y_T| < \infty$, *and*
 (c) $\mathbb{E}(Y_n I_{\{T>n\}}) \to 0$ *as* $n \to \infty$.

(2) Theorem. *Let $(Y, \mathcal{F})$ be a martingale and let T be a stopping time. If the Y_n are uniformly integrable and $\mathbb{P}(T < \infty) = 1$ then $Y_T = \mathbb{E}(Y_\infty \mid \mathcal{F}_T)$ and $Y_0 = \mathbb{E}(Y_T \mid \mathcal{F}_0)$. In particular $\mathbb{E}(Y_0) = \mathbb{E}(Y_T)$.*

Proof of (1). It is easily seen that $Y_T = Y_{T\wedge n} + (Y_T - Y_n)I_{\{T>n\}}$. Taking expectations and using the fact that $\mathbb{E}(Y_{T\wedge n}) = \mathbb{E}(Y_0)$ (see Theorem (12.4.11)), we find that

(3) $$\mathbb{E}(Y_T) = \mathbb{E}(Y_0) + \mathbb{E}(Y_T I_{\{T>n\}}) - \mathbb{E}(Y_n I_{\{T>n\}}).$$

The last term tends to zero as $n \to \infty$, by assumption (c). As for the penultimate term,

$$\mathbb{E}(Y_T I_{\{T>n\}}) = \sum_{k=n+1}^{\infty} \mathbb{E}(Y_T I_{\{T=k\}})$$

is, by assumption (b), the tail of the convergent series $\mathbb{E}(Y_T) = \sum_k \mathbb{E}(Y_T I_{\{T=k\}})$; therefore $\mathbb{E}(Y_T I_{\{T>n\}}) \to 0$ as $n \to \infty$, and (3) yields $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$ in the limit as $n \to \infty$. ■

Proof of (2). Since $(Y, \mathcal{F})$ is uniformly integrable, we have by Theorems (12.3.1) and (12.3.11) that the limit $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely, and $Y_n = \mathbb{E}(Y_\infty \mid \mathcal{F}_n)$. It follows from the definition (12.1.7) of conditional expectation that

(4) $$\mathbb{E}(Y_n I_A) = \mathbb{E}(Y_\infty I_A) \quad \text{for all } A \in \mathcal{F}_n.$$

Now, if $A \in \mathcal{F}_T$ then $A \cap \{T = n\} \in \mathcal{F}_n$, so that

$$\mathbb{E}(Y_T I_A) = \sum_n \mathbb{E}(Y_n I_{A\cap\{T=n\}}) = \sum_n \mathbb{E}(Y_\infty I_{A\cap\{T=n\}}) = \mathbb{E}(Y_\infty I_A),$$

whence $Y_T = \mathbb{E}(Y_\infty \mid \mathcal{F}_T)$. Secondly, since $\mathcal{F}_0 \subseteq \mathcal{F}_T$,

$$\mathbb{E}(Y_T \mid \mathcal{F}_0) = \mathbb{E}\big(\mathbb{E}(Y_\infty \mid \mathcal{F}_T) \bigm| \mathcal{F}_0\big) = \mathbb{E}(Y_\infty \mid \mathcal{F}_0) = Y_0.$$ ■

(5) Example. Markov chains. Let X be an irreducible persistent Markov chain with countable state space S and transition matrix $\mathbf{P}$, and let $\psi : S \to \mathbb{R}$ be a bounded function satisfying

$$\sum_{j \in S} p_{ij} \psi(j) = \psi(i) \quad \text{for all } i \in S.$$

Then $\psi(X_n)$ constitutes a martingale. Let T_i be the first passage time of X to the state i, that is, $T_i = \min\{n : X_n = i\}$; it is easily seen that T_i is a stopping time and is (almost surely) finite. Furthermore, the sequence $\{\psi(X_n)\}$ is bounded and therefore uniformly integrable. Applying Theorem (2), we obtain $\mathbb{E}(\psi(X_T)) = \mathbb{E}(\psi(X_0))$, whence $\mathbb{E}(\psi(X_0)) = \psi(i)$ for all states i. Therefore ψ is a constant function. ●

(6) Example. Symmetric simple random walk. Let S_n be the position of the particle after n steps and suppose that $S_0 = 0$. Then $S_n = \sum_{i=1}^n X_i$ where $X_1, X_2, \dots$ are independent and equally likely to take each of the values $+1$ and -1. It is easy to see as in Example (12.1.2) that $\{S_n\}$ is a martingale. Let a and b be positive integers and let $T = \min\{n : S_n = -a \text{ or } S_n = b\}$ be the earliest time at which the walk visits either $-a$ or b. Certainly T is a stopping time and satisfies the conditions of Theorem (1). Let p_a be the probability that the particle visits $-a$ before it visits b. By the optional stopping theorem,

$$\mathbb{E}(S_T) = (-a)p_a + b(1 - p_a), \quad \mathbb{E}(S_0) = 0; \tag{7}$$

therefore $p_a = b/(a + b)$, which agrees with the earlier result of equation (1.7.7) when the notation is translated suitably. The sequence $\{S_n\}$ is not the only martingale available. Let $\{Y_n\}$ be given by $Y_n = S_n^2 - n$; then $\{Y_n\}$ is a martingale also. Apply Theorem (1) with T given as before to obtain $\mathbb{E}(T) = ab$. ●

(8) Example. De Moivre's martingale (12.1.4). Consider now a simple random walk $\{S_n\}$ with $0 < S_0 < N$, for which each step is rightwards with probability p where $0 < p = 1-q < 1$. We have seen that $Y_n = (q/p)^{S_n}$ defines a martingale, and furthermore the first passage time T of the walk to the set $\{0, N\}$ is a stopping time. It is easily checked that conditions (1a)–(1c) of the optional stopping theorem are satisfied, and hence $\mathbb{E}((q/p)^{S_T}) = \mathbb{E}((q/p)^{S_0})$. Therefore $p_k = \mathbb{P}(S_T = 0 \mid S_0 = k)$ satisfies $p_k + (q/p)^N(1 - p_k) = (q/p)^k$, whence p_k may be calculated as in Example (12.1.4). ●

When applying the optional stopping theorem it is sometimes convenient to use a more restrictive set of conditions.

(9) Theorem. *Let $(Y, \mathcal{F})$ be a martingale, and let T be a stopping time. Then $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$ if the following hold:*

(a) $\mathbb{P}(T < \infty) = 1$, $\mathbb{E}T < \infty$, *and*

(b) *there exists a constant c such that $\mathbb{E}\big(|Y_{n+1} - Y_n| \,\big|\, \mathcal{F}_n\big) \le c$ for all $n < T$.*

Proof. By the discussion prior to (1), it suffices to show that the sequence $\{Y_{T \wedge n} : n \ge 0\}$ is uniformly integrable. Let $Z_n = |Y_n - Y_{n-1}|$ for $n \ge 1$, and $W = Z_1 + Z_2 + \cdots + Z_T$. Certainly $|Y_{T \wedge n}| \le |Y_0| + W$ for all n, and it is enough (by Example (7.10.4)) to show that $\mathbb{E}(W) < \infty$. We have that

$$W = \sum_{i=1}^{\infty} Z_i I_{\{T \ge i\}}. \tag{10}$$

Now

$$\mathbb{E}\big(Z_i I_{\{T\ge i\}} \,\big|\, \mathcal{F}_{i-1}\big) = I_{\{T\ge i\}}\mathbb{E}(Z_i \mid \mathcal{F}_{i-1}) \le c I_{\{T\ge i\}},$$

since $\{T \ge i\} = \{T \le i-1\}^{\mathrm{c}} \in \mathcal{F}_{i-1}$. Therefore $\mathbb{E}(Z_i I_{\{T\ge i\}}) \le c\mathbb{P}(T \ge i)$, giving by (10) that

$$\mathbb{E}(W) \le c\sum_{i=1}^{\infty}\mathbb{P}(T\ge i) = c\mathbb{E}(T) < \infty. \tag{11}$$

■

(12) Example. Wald's equation (10.2.9). Let $X_1, X_2, \dots$ be independent identically distributed random variables with finite mean μ, and let $S_n = \sum_{i=1}^{n} X_i$. It is easy to see that $Y_n = S_n - n\mu$ constitutes a martingale with respect to the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$. Now

$$\mathbb{E}\big(|Y_{n+1} - Y_n| \,\big|\, \mathcal{F}_n\big) = \mathbb{E}|X_{n+1} - \mu| = \mathbb{E}|X_1 - \mu| < \infty.$$

We deduce from (9) that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0) = 0$ for any stopping time T with finite mean, implying that

$$\mathbb{E}(S_T) = \mu\mathbb{E}(T), \tag{13}$$

a result derived earlier in the context of renewal theory as Lemma (10.2.9).

If the X_i have finite variance σ^2, it is also the case that

$$\mathrm{var}(Y_T) = \sigma^2\mathbb{E}(T) \qquad \text{if } \mathbb{E}(T) < \infty. \tag{14}$$

It is possible to prove this by applying the optional stopping theorem to the martingale $Z_n = Y_n^2 - n\sigma^2$, but this is not a simple application of (9). It may also be proved by exploiting Wald's identity (15), or more simply by the method of Exercise (10.2.2). ●

(15) Example. Wald's identity. This time, let $X_1, X_2, \dots$ be independent identically distributed random variables with common moment generating function $M(t) = \mathbb{E}(e^{tX})$; suppose that there exists at least one value of t ($\ne 0$) such that $1 \le M(t) < \infty$, and fix t accordingly. Let $S_n = X_1 + X_2 + \cdots + X_n$, define

$$Y_0 = 1, \quad Y_n = \frac{e^{tS_n}}{M(t)^n} \quad \text{for} \quad n \ge 1, \tag{16}$$

and let $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. It is clear that $(Y, \mathcal{F})$ is a martingale. When are the conditions of Theorem (9) valid? Let T be a stopping time with finite mean, and note that

$$\mathbb{E}\big(|Y_{n+1} - Y_n| \,\big|\, \mathcal{F}_n\big) = Y_n\mathbb{E}\left(\left|\frac{e^{tX}}{M(t)} - 1\right|\right) \le \frac{Y_n}{M(t)}\mathbb{E}\big(e^{tX} + M(t)\big) = 2Y_n. \tag{17}$$

Suppose that T is such that

$$|S_n| \le C \quad \text{for} \quad n < T, \tag{18}$$

where C is a constant. Now $M(t) \geq 1$, and

$$Y_n = \frac{e^{tS_n}}{M(t)^n} \leq \frac{e^{|t|C}}{M(t)^n} \leq e^{|t|C} \quad \text{for} \quad n < T,$$

giving by (17) that condition (9b) holds. In summary, if T is a stopping time with finite mean such that (18) holds, then

$$\textbf{(19)} \qquad \mathbb{E}\{e^{tS_T} M(t)^{-T}\} = 1 \quad \text{whenever} \quad M(t) \geq 1,$$

an equation usually called *Wald's identity*.

Here is an application of (19). Suppose the X_i have strictly positive variance, and let $T = \min\{n : S_n \leq -a \text{ or } S_n \geq b\}$ where $a, b > 0$; T is the 'first exit time' from the interval $(-a, b)$. Certainly $|S_n| \leq \max\{a, b\}$ if $n < T$. Furthermore $\mathbb{E}T < \infty$, which may be seen as follows. By the non-degeneracy of the X_i, there exist M and $\epsilon > 0$ such that $\mathbb{P}(|S_M| > a+b) > \epsilon$. If any of the quantities $|S_M|, |S_{2M} - S_M|, \dots, |S_{kM} - S_{(k-1)M}|$ exceed $a+b$ then the process must have exited $(-a, b)$ by time kM. Therefore $\mathbb{P}(T \geq kM) \leq (1-\epsilon)^k$, implying that

$$\mathbb{E}(T) = \sum_{i=1}^{\infty} \mathbb{P}(T \geq i) \leq M \sum_{k=0}^{\infty} \mathbb{P}(T \geq kM) < \infty.$$

We conclude that (19) is valid. In many concrete cases of interest, there exists θ ($\neq 0$) such that $M(\theta) = 1$. Applying (19) with $t = \theta$, we obtain $\mathbb{E}(e^{\theta S_T}) = 1$, or

$$\eta_a \mathbb{P}(S_T \leq -a) + \eta_b \mathbb{P}(S_T \geq b) = 1$$

where

$$\eta_a = \mathbb{E}(e^{\theta S_T} \mid S_T \leq -a), \qquad \eta_b = \mathbb{E}(e^{\theta S_T} \mid S_T \geq b),$$

and therefore

$$\textbf{(20)} \qquad \mathbb{P}(S_T \leq -a) = \frac{\eta_b - 1}{\eta_b - \eta_a}, \qquad \mathbb{P}(S_T \geq b) = \frac{1 - \eta_a}{\eta_b - \eta_a}.$$

When a and b are large, it is reasonable to suppose that $\eta_a \simeq e^{-\theta a}$ and $\eta_b \simeq e^{\theta b}$, giving the approximations

$$\textbf{(21)} \qquad \mathbb{P}(S_T \leq -a) \simeq \frac{e^{\theta b} - 1}{e^{\theta b} - e^{-\theta a}}, \qquad \mathbb{P}(S_T \geq b) \simeq \frac{1 - e^{-\theta a}}{e^{\theta b} - e^{-\theta a}}.$$

These approximations are of course exact if S is a simple random walk and a and b are positive integers. ●

(22) Example. Simple random walk. Suppose that $\{S_n\}$ is a simple random walk whose steps $\{X_i\}$ take the values 1 and -1 with respective probabilities p and q $(= 1 - p)$. For positive integers a and b, we have from Wald's identity (19) that

$$\textbf{(23)} \qquad e^{-at}\mathbb{E}\big(M(t)^{-T} I_{\{S_T = -a\}}\big) + e^{bt}\mathbb{E}\big(M(t)^{-T} I_{\{S_T = b\}}\big) = 1 \quad \text{if} \quad M(t) \geq 1$$

where T is the first exit time of $(-a, b)$ as before, and $M(t) = pe^t + qe^{-t}$.

Setting $M(t) = s^{-1}$, we obtain a quadratic for e^t, and hence $e^t = \lambda_1(s)$ or $e^t = \lambda_2(s)$ where

$$\lambda_1(s) = \frac{1 + \sqrt{1 - 4pqs^2}}{2ps}, \quad \lambda_2(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}.$$

Substituting these into equation (23), we obtain two linear equations in the quantities

(24) $$P_1(s) = \mathbb{E}(s^T I_{\{S_T = -a\}}), \quad P_s(s) = \mathbb{E}(s^T I_{\{S_T = b\}}),$$

with solutions

$$P_1(s) = \frac{\lambda_1^a \lambda_2^a (\lambda_1^b - \lambda_2^b)}{\lambda_1^{a+b} - \lambda_2^{a+b}}, \quad P_2(s) = \frac{\lambda_1^a - \lambda_2^a}{\lambda_1^{a+b} - \lambda_2^{a+b}},$$

which we add to obtain the probability generating function of T,

(25) $$\mathbb{E}(s^T) = P_1(s) + P_2(s), \quad 0 < s \le 1.$$

Suppose we let $a \to \infty$, so that T becomes the time until the first passage to the point b. From (24), $P_1(s) \to 0$ as $a \to \infty$ if $0 < s < 1$, and a quick calculation gives $P_2(s) \to F_b(s)$ where

$$F_b(s) = \left(\frac{1 - \sqrt{1 - 4pqs^2}}{2qs} \right)^b$$

in agreement with Theorem (5.3.5). Notice that $F_b(1) = (\min\{1, p/q\})^b$. ●

Exercises for Section 12.5

1. Let $(Y, \mathcal{F})$ be a martingale and T a stopping time such that $\mathbb{P}(T < \infty) = 1$. Show that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$ if either of the following holds:
(a) $\mathbb{E}(\sup_n |Y_{T \wedge n}|) < \infty$, (b) $\mathbb{E}(|Y_{T \wedge n}|^{1+\delta}) \le c$ for some $c, \delta > 0$ and all n.

2. Let $(Y, \mathcal{F})$ be a martingale. Show that $(Y_{T \wedge n}, \mathcal{F}_n)$ is a uniformly integrable martingale for any finite stopping time T such that either:
(a) $\mathbb{E}|Y_T| < \infty$ and $\mathbb{E}(|Y_n| I_{\{T > n\}}) \to 0$ as $n \to \infty$, or
(b) $\{Y_n\}$ is uniformly integrable.

3. Let $(Y, \mathcal{F})$ be a uniformly integrable martingale, and let S and T be finite stopping times satisfying $S \le T$. Prove that $Y_T = \mathbb{E}(Y_\infty \mid \mathcal{F}_T)$ and that $Y_S = \mathbb{E}(Y_T \mid \mathcal{F}_S)$, where Y_∞ is the almost sure limit as $n \to \infty$ of Y_n.

4. Let $\{S_n : n \ge 0\}$ be a simple symmetric random walk with $0 < S_0 < N$ and with absorbing barriers at 0 and N. Use the optional stopping theorem to show that the mean time until absorption is $\mathbb{E}\{S_0(N - S_0)\}$.

5. Let $\{S_n : n \ge 0\}$ be a simple symmetric random walk with $S_0 = 0$. Show that

$$Y_n = \frac{\cos\{\lambda[S_n - \frac{1}{2}(b - a)]\}}{(\cos \lambda)^n}$$

constitutes a martingale if $\cos \lambda \ne 0$.

Let a and b be positive integers. Show that the time T until absorption at one of two absorbing barriers at $-a$ and b satisfies

$$\mathbb{E}\big(\{\cos\lambda\}^{-T}\big) = \frac{\cos\{\frac{1}{2}\lambda(b-a)\}}{\cos\{\frac{1}{2}\lambda(b+a)\}}, \qquad 0 < \lambda < \frac{\pi}{b+a}.$$

6. Let $\{S_n : n \geq 0\}$ be a simple symmetric random walk on the positive and negative integers, with $S_0 = 0$. For each of the three following random variables, determine whether or not it is a stopping time and find its mean:

$$U = \min\{n \geq 5 : S_n = S_{n-5} + 5\}, \qquad V = U - 5, \qquad W = \min\{n : S_n = 1\}.$$

7. Let $S_n = a + \sum_{r=1}^{n} X_r$ be a simple symmetric random walk. The walk stops at the earliest time T when it reaches either of the two positions 0 or K where $0 < a < K$. Show that $M_n = \sum_{r=0}^{n} S_r - \frac{1}{3}S_n^3$ is a martingale and deduce that $\mathbb{E}\big(\sum_{r=0}^{T} S_r\big) = \frac{1}{3}(K^2 - a^2)a + a$.

8. Gambler's ruin. Let X_i be independent random variables each equally likely to take the values ± 1, and let $T = \min\{n : S_n \in \{-a, b\}\}$. Verify the conditions of the optional stopping theorem (12.5.1) for the martingale $S_n^2 - n$ and the stopping time T.

12.6 The maximal inequality

In proving the convergence of a sequence $X_1, X_2, \dots$ of random variables, it is often useful to establish an inequality of the form

$$\mathbb{P}\big(\max\{X_1, X_2, \dots, X_n\} \geq x\big) \leq A_n(x),$$

and such an inequality is sometimes called a maximal inequality. The bound $A_n(x)$ usually involves an expectation. Examples of such inequalities include Kolmogorov's inequality in the proof of the strong law of large numbers, and the Doob–Kolmogorov inequality (7.8.2) in the proof of the convergence of martingales with bounded second moments. Both these inequalities are special cases of the following maximal inequality for submartingales. In order to simplify the notation of this section, we shall write X_n^* for the maximum of the first $n+1$ members of a sequence $X_0, X_1, \dots$, so that $X_n^* = \max\{X_i : 0 \leq i \leq n\}$.

(1) Theorem. Maximal inequality.

(a) *If $(Y, \mathcal{F})$ is a submartingale, then*

$$\mathbb{P}(Y_n^* \geq x) \leq \frac{\mathbb{E}(Y_n^+)}{x} \quad \textit{for} \quad x > 0.$$

(b) *If $(Y, \mathcal{F})$ is a supermartingale and $\mathbb{E}|Y_0| < \infty$, then*

$$\mathbb{P}(Y_n^* \geq x) \leq \frac{\mathbb{E}(Y_0) + \mathbb{E}(Y_n^-)}{x} \quad \textit{for} \quad x > 0.$$

These inequalities may be improved somewhat. For example, a closer look at the proof in case (a) leads to the inequality

$$\mathbb{P}(Y_n^* \geq x) \leq \frac{1}{x}\mathbb{E}(Y_n^+ I_{\{Y_n^* \geq x\}}) \quad \text{for} \quad x > 0. \tag{2}$$

Proof. This is very similar to that of Theorem (12.4.13). Let $T = \min\{n : Y_n \geq x\}$ where $x > 0$, and suppose first that $(Y, \mathcal{F})$ is a submartingale. Then $(Y^+, \mathcal{F})$ is a non-negative submartingale with finite means by Exercise (12.1.7), and $T = \min\{n : Y_n^+ \geq x\}$ since $x > 0$. Applying the optional sampling theorem (12.4.11b) with stopping times $T_1 = T \wedge n$, $T_2 = n$, we obtain $\mathbb{E}(Y_{T\wedge n}^+) \leq \mathbb{E}(Y_n^+)$. However,

$$\begin{aligned}\mathbb{E}(Y_{T\wedge n}^+) &= \mathbb{E}(Y_T^+ I_{\{T\leq n\}}) + \mathbb{E}(Y_n^+ I_{\{T>n\}})\\ &\geq x\mathbb{P}(T \leq n) + \mathbb{E}(Y_n^+ I_{\{T>n\}})\end{aligned}$$

whence, as required,

$$\text{(3)}\qquad \begin{aligned}x\mathbb{P}(T \leq n) &\leq \mathbb{E}\big(Y_n^+(1 - I_{\{T>n\}})\big)\\ &= \mathbb{E}(Y_n^+ I_{\{T\leq n\}}) \leq \mathbb{E}(Y_n^+).\end{aligned}$$

Suppose next that $(Y, \mathcal{F})$ is a supermartingale. By optional sampling $\mathbb{E}(Y_0) \geq \mathbb{E}(Y_{T\wedge n})$. Now

$$\begin{aligned}\mathbb{E}(Y_{T\wedge n}) &= \mathbb{E}\big(Y_T I_{\{T\leq n\}} + Y_n I_{\{T>n\}}\big)\\ &\geq x\mathbb{P}(T \leq n) - \mathbb{E}(Y_n^-),\end{aligned}$$

whence $x\mathbb{P}(T \leq n) \leq \mathbb{E}(Y_0) + \mathbb{E}(Y_n^-)$. ■

Part (a) of the maximal inequality may be used to handle the maximum of a submartingale, and part (b) may be used as follows to handle its minimum. Suppose that $(Y, \mathcal{F})$ is a submartingale with finite means. Then $(-Y, \mathcal{F})$ is a supermartingale, and therefore

$$\text{(4)}\qquad \mathbb{P}\left(\min_{0\leq k\leq n} Y_k \leq -x\right) \leq \frac{\mathbb{E}(Y_n^+) - \mathbb{E}(Y_0)}{x} \quad \text{for} \quad x > 0,$$

by (1b). Using (1a) also, we find that

$$\mathbb{P}\left(\max_{0\leq k\leq n} |Y_k| \geq x\right) \leq \frac{2\mathbb{E}(Y_n^+) - \mathbb{E}(Y_0)}{x} \leq \frac{3}{x}\sup_k \mathbb{E}|Y_k|.$$

Sending n to infinity (and hiding a minor 'continuity' argument), we deduce that

$$\text{(5)}\qquad \mathbb{P}\left(\sup_k |Y_k| \geq x\right) \leq \frac{3}{x}\sup_k \mathbb{E}|Y_k|, \quad \text{for} \quad x > 0.$$

A slightly tighter conclusion is valid if $(Y, \mathcal{F})$ is a martingale rather than merely a submartingale. In this case, $(|Y_n|, \mathcal{F}_n)$ is a submartingale, whence (1a) yields

$$\text{(6)}\qquad \mathbb{P}\left(\sup_k |Y_k| \geq x\right) \leq \frac{1}{x}\sup_k \mathbb{E}|Y_k|, \quad \text{for} \quad x > 0.$$

(7) Example. Doob–Kolmogorov inequality (7.8.2). Let $(Y, \mathcal{F})$ be a martingale such that $\mathbb{E}(Y_n^2) < \infty$ for all n. Then $(Y_n^2, \mathcal{F}_n)$ is a submartingale, whence

$$\text{(8)}\qquad \mathbb{P}\left(\max_{0\leq k\leq n} |Y_k| \geq x\right) = \mathbb{P}\left(\max_{0\leq k\leq n} Y_k^2 \geq x^2\right) \leq \frac{\mathbb{E}(Y_n^2)}{x^2}$$

for $x > 0$, in agreement with (7.8.2). This is the major step in the proof of the convergence theorem (7.8.1) for martingales with bounded second moments. ●

(9) Example. Kolmogorov's inequality. Let $X_1, X_2, \dots$ be independent random variables with finite means and variances. Applying the Doob–Kolmogorov inequality (8) to the martingale $Y_n = S_n - \mathbb{E}(S_n)$ where $S_n = X_1 + X_2 + \cdots + X_n$, we obtain

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k - \mathbb{E}(S_k)| \ge x\right) \le \frac{1}{x^2}\operatorname{var}(S_n) \quad \text{for} \quad x > 0. \tag{10}$$

This powerful inequality is the principal step in the usual proof of the strong law of large numbers (7.5.1). See Problem (7.11.29) for a simple proof not using martingales. ●

The maximal inequality may be used to address the question of convergence in rth mean of martingales.

(11) Theorem. *Let $r > 1$, and let $(Y, \mathcal{F})$ be a martingale such that $\sup_n \mathbb{E}|Y_n^r| < \infty$. Then $Y_n \xrightarrow{r} Y_\infty$ where Y_∞ is the (almost sure) limit of Y_n.*

This is not difficult to prove by way of Fatou's lemma and the theory of uniform integrability. Instead, we shall make use of the following inequality.

(12) Lemma. *Let $r > 1$, and let $(Y, \mathcal{F})$ be a non-negative submartingale such that $\mathbb{E}(Y_n^r) < \infty$ for all n. Then*

$$\mathbb{E}(Y_n^r) \le \mathbb{E}((Y_n^*)^r) \le \left(\frac{r}{r-1}\right)^r \mathbb{E}(Y_n^r). \tag{13}$$

Proof. Certainly $Y_n \le Y_n^*$, and therefore the first inequality is trivial. Turning to the second, note first that

$$\mathbb{E}((Y_n^*)^r) \le \mathbb{E}\big((Y_0 + Y_1 + \cdots + Y_n)^r\big) < \infty.$$

Now, integrate by parts and use the maximal inequality (2) to obtain

$$\begin{aligned}\mathbb{E}((Y_n^*)^r) &= \int_0^\infty r x^{r-1}\mathbb{P}(Y_n^* \ge x)\,dx \le \int_0^\infty r x^{r-2}\mathbb{E}(Y_n I_{\{Y_n^*\ge x\}})\,dx \\ &= \mathbb{E}\left(Y_n \int_0^{Y_n^*} r x^{r-1}\,dx\right) = \frac{r}{r-1}\mathbb{E}\big(Y_n (Y_n^*)^{r-1}\big).\end{aligned}$$

We have by Hölder's inequality that

$$\mathbb{E}\big(Y_n (Y_n^*)^{r-1}\big) \le [\mathbb{E}(Y_n^r)]^{1/r}[\mathbb{E}((Y_n^*)^r)]^{(r-1)/r}.$$

Substituting this, and solving, we obtain

$$[\mathbb{E}(Y_n^*)^r)]^{1/r} \le \frac{r}{r-1}[\mathbb{E}(Y_n^r)]^{1/r}. \qquad \blacksquare$$

Proof of Theorem (11). Using the moment condition, $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely. Now $(|Y_n|, \mathcal{F}_n)$ is a non-negative submartingale, and hence $\mathbb{E}(\sup_k |Y_k|^r) < \infty$ by Lemma (12) and monotone convergence (5.6.12). Hence $\{Y_k^r : k \ge 0\}$ is uniformly integrable (Exercise (7.10.6)), implying by Exercise (7.10.2) that $Y_k \xrightarrow{r} Y_\infty$ as required. ■

12.7 Backward martingales and continuous-time martingales

The ideas of martingale theory find expression in several other contexts, of which we consider two in this section. The first of these concerns backward martingales. We call a sequence $\mathcal{G} = \{\mathcal{G}_n : n \geq 0\}$ of σ-fields *decreasing* if $\mathcal{G}_n \supseteq \mathcal{G}_{n+1}$ for all n.

(1) Definition. Let $\mathcal{G}$ be a decreasing sequence of σ-fields and let Y be a sequence of random variables which is adapted to $\mathcal{G}$. We call $(Y, \mathcal{G})$ a **backward** (or **reversed**) **martingale** if, for all $n \geq 0$,

(a) $\mathbb{E}|Y_n| < \infty$,

(b) $\mathbb{E}(Y_n \mid \mathcal{G}_{n+1}) = Y_{n+1}$.

Note that $\{(Y_n, \mathcal{G}_n) : n = 0, 1, 2, \dots\}$ is a backward martingale if and only if the reversed sequence $\{(Y_n, \mathcal{G}_n) : n = \dots, 2, 1, 0\}$ is a martingale, an observation which explains the use of the term.

(2) Example. Strong law of large numbers. Let $X_1, X_2, \dots$ be independent identically distributed random variables with finite mean. Set $S_n = X_1 + X_2 + \cdots + X_n$ and let $\mathcal{G}_n = \sigma(S_n, S_{n+1}, \dots)$. Then, using symmetry,

(3) $$\mathbb{E}(S_n \mid \mathcal{G}_{n+1}) = \mathbb{E}(S_n \mid S_{n+1}) = n\mathbb{E}(X_1 \mid S_{n+1}) = n\frac{S_{n+1}}{n+1}$$

since $S_{n+1} = \mathbb{E}(S_{n+1} \mid S_{n+1}) = (n+1)\mathbb{E}(X_1 \mid S_{n+1})$. Therefore $Y_n = S_n/n$ satisfies $\mathbb{E}(Y_n \mid \mathcal{G}_{n+1}) = Y_{n+1}$, whence $(Y, \mathcal{G})$ is a backward martingale. We shall see soon that backward martingales converge almost surely and in mean, and therefore there exists Y_∞ such that $Y_n \to Y_\infty$ a.s. and in mean. By the zero–one law (7.3.15), Y_∞ is almost surely constant, and hence $Y_\infty = \mathbb{E}(X_1)$ almost surely. We have proved the strong law of large numbers. ●

(4) Backward-martingale convergence theorem. *Let $(Y, \mathcal{G})$ be a backward martingale. Then Y_n converges to a limit Y_∞ almost surely and in mean.*

It is striking that no extra condition is necessary to ensure the convergence of backward martingales.

Proof. Note first that the sequence $\dots, Y_n, Y_{n-1}, \dots, Y_1, Y_0$ is a martingale with respect to the sequence $\dots, \mathcal{G}_n, \mathcal{G}_{n-1}, \dots, \mathcal{G}_1, \mathcal{G}_0$, and therefore $Y_n = \mathbb{E}(Y_0 \mid \mathcal{G}_n)$ for all n. However, $\mathbb{E}|Y_0| < \infty$, and therefore $\{Y_n\}$ is uniformly integrable by Example (7.10.13). It is therefore sufficient to prove that Y_n converges almost surely. The usual way of doing this is via an upcrossings inequality. Applying (12.3.3) to the martingale $Y_n, Y_{n-1}, \dots, Y_0$, we obtain that

$$\mathbb{E}U_n(a, b; Y) \leq \frac{E((Y_0 - a)^+)}{b - a}$$

where $U_n(a, b; Y)$ is the number of upcrossings of $[a, b]$ by the sequence $Y_n, Y_{n-1}, \dots, Y_0$. We let $n \to \infty$, and follow the proof of the martingale convergence theorem (12.3.1) to obtain the required result. ■

Rather than developing the theory of backward martingales in detail, we confine ourselves to one observation and an application. Let $(Y, \mathcal{G})$ be a backward martingale, and let T be a stopping time with respect to $\mathcal{G}$; that is, $\{T = n\} \in \mathcal{G}_n$ for all n. If T is bounded, say $\mathbb{P}(T \le N) = 1$ for some fixed N, then the sequence $Z_N, Z_{N-1}, \dots, Z_0$ defined by $Z_n = Y_{T \vee n}$ is a martingale with respect to the appropriate sequence of σ-fields (remember that $x \vee y = \max\{x, y\}$). Hence, by the optional sampling theorem (12.4.11a),

$$\mathbb{E}(Y_T \mid \mathcal{G}_N) = Y_N. \tag{5}$$

(6) Example. Ballot theorem (3.10.6). Let $X_1, X_2, \dots$ be independent identically distributed random variables taking values in $\{0, 1, 2, \dots\}$, and let $S_n = X_1 + X_2 + \cdots + X_n$. We claim that

$$\mathbb{P}\big(S_k \ge k \text{ for some } 1 \le k \le N \,\big|\, S_N = b\big) = \min\{1, b/N\}, \tag{7}$$

whenever b is such that $\mathbb{P}(S_N = b) > 0$. It is not immediately clear that this implies the ballot theorem, but look at it this way. In a ballot, each of N voters has two votes; he or she allocates both votes either to candidate A or to candidate B. Let us write X_i for the number of votes allocated to A by the ith voter, so that X_i equals either 0 or 2; assume that the X_i are independent. Now $S_k \ge k$ for some $1 \le k \le N$ if and only if B is not always in the lead. Equation (7) implies

$$\begin{aligned} \mathbb{P}\big(B \text{ always leads} \,\big|\, A \text{ receives a total of } 2a \text{ votes}\big) &\\ = 1 - \mathbb{P}\big(S_k \ge k \text{ for some } 1 \le k \le N \,\big|\, S_n = 2a\big) &\\ = 1 - \frac{2a}{N} = \frac{p-q}{p+q} & \end{aligned} \tag{8}$$

if $0 \le a < \frac{1}{2}N$, where $p = 2N - 2a$ is the number of votes received by B, and $q = 2a$ is the number received by A. This is the famous ballot theorem discussed after Corollary (3.10.6).

In order to prove equation (7), let $\mathcal{G}_n = \sigma(S_n, S_{n+1}, \dots)$, and recall that $(S_n/n, \mathcal{G}_n)$ is a backward martingale. Fix N, and let

$$T = \begin{cases} \max\{k : S_k \ge k \text{ and } 1 \le k \le N\} & \text{if this exists,} \\ 1 & \text{otherwise.} \end{cases}$$

This may not look like a stopping time, but it is. After all, for $1 < n \le N$,

$$\{T = n\} = \{S_n \ge n,\ S_k < k \text{ for } n < k \le N\},$$

an event defined in terms of $S_n, S_{n+1}, \dots$ and therefore lying in the σ-field $\mathcal{G}_n$ generated by these random variables. By a similar argument, $\{T = 1\} \in \mathcal{G}_1$.

We may assume that $S_N = b < N$, since (7) is obvious if $b \ge N$. Let $A = \{S_k \ge k$ for some $1 \le k \le N\}$. We have that $S_N < N$; therefore, if A occurs, it must be the case that $S_T \ge T$ and $S_{T+1} < T + 1$. In this case $X_{T+1} = S_{T+1} - S_T < 1$, so that $X_{T+1} = 0$ and therefore $S_T/T = 1$. On the other hand, if A does not occur then $T = 1$, and also $S_T = S_1 = 0$, implying that $S_T/T = 0$. It follows that $S_T/T = I_A$ if $S_N < N$, where I_A is the indicator function of A. Taking expectations, we obtain

$$\mathbb{E}\left(\frac{1}{T} S_T \,\middle|\, S_N = b\right) = \mathbb{P}(A \mid S_N = b) \quad \text{if} \quad b < N.$$

Finally, we apply (5) to the backward martingale $(S_n/n, \mathcal{G}_n)$ to obtain

$$\mathbb{E}\left(\frac{1}{T}S_T \,\middle|\, S_N = b\right) = \mathbb{E}\left(\frac{1}{N}S_N \,\middle|\, S_N = b\right) = \frac{b}{N}.$$

The last two equations may be combined to give (7). ●

In contrast to the theory of backward martingales, the theory of continuous-time martingales is hedged about with technical considerations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a family $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_t : t \geq 0\}$ of sub-σ-fields of $\mathcal{F}$ satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t$ whenever $s \leq t$. As before, we say that the (continuous-time) process $Y = \{Y(t) : t \geq 0\}$ is adapted to $\boldsymbol{\mathcal{F}}$ if $Y(t)$ is $\mathcal{F}_t$-measurable for all t. If Y is adapted to $\boldsymbol{\mathcal{F}}$, we call $(Y, \boldsymbol{\mathcal{F}})$ a *martingale* if $\mathbb{E}|Y(t)| < \infty$ for all t, and $\mathbb{E}(Y(t) \mid \mathcal{F}_s) = Y(s)$ whenever $s \leq t$. A random variable T taking values in $[0, \infty]$ is called a *stopping time* (with respect to the filtration $\boldsymbol{\mathcal{F}}$) if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Possibly the most important type of stopping time is the first passage time $T(A) = \inf\{t : Y(t) \in A\}$ for a suitable subset A of $\mathbb{R}$. Unfortunately $T(A)$ is not necessarily a stopping time. No problems arise if A is closed and the sample paths $\Pi(\omega) = \{(t, Y(t;\omega)) : t \geq 0\}$ of Y are continuous, but these conditions are over-restrictive. They may be relaxed at the price of making extra assumptions about the process Y and the filtration $\boldsymbol{\mathcal{F}}$. It is usual to assume in addition that:

(a) $(\Omega, \mathcal{F}, \mathbb{P})$ is complete,

(b) $\mathcal{F}_0$ contains all events A of $\mathcal{F}$ satisfying $\mathbb{P}(A) = 0$,

(c) $\boldsymbol{\mathcal{F}}$ is right-continuous in that $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \geq 0$, where $\mathcal{F}_{t+} = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$.

We shall refer to these conditions as the 'usual conditions'. Conditions (a) and (b) pose little difficulty, since an incomplete probability space may be completed, and the null events may be added to $\mathcal{F}_0$. Condition (c) is not of great importance if the process Y has right-continuous sample paths, since then $Y(t) = \lim_{\epsilon \downarrow 0} Y(t + \epsilon)$ is $\mathcal{F}_{t+}$-measurable.

Here are some examples of continuous-time martingales.

(9) Example. Poisson process. Let $\{N(t) : t \geq 0\}$ be a Poisson process with intensity λ, and let $\mathcal{F}_t$ be the σ-field generated by $\{N(u) : 0 \leq u \leq t\}$. It is easily seen that

$$\begin{aligned} U(t) &= N(t) - \lambda t, \\ V(t) &= U(t)^2 - \lambda t, \\ W(t) &= \exp[-\theta N(t) + \lambda t(1 - e^{-\theta})], \end{aligned}$$

constitute martingales with respect to $\boldsymbol{\mathcal{F}}$.

There is a converse statement. Suppose $N = \{N(t) : t \geq 0\}$ is an integer-valued non-decreasing process such that, for all θ,

$$W(t) = \exp[-\theta N(t) + \lambda t(1 - e^{-\theta})]$$

is a martingale. Then, if $s < t$,

$$\begin{aligned} \mathbb{E}\big(\exp\{-\theta[N(t) - N(s)]\} \mid \mathcal{F}_s\big) &= \mathbb{E}\left(\frac{W(t)}{W(s)} \exp[-\lambda(t-s)(1 - e^{-\theta})] \,\middle|\, \mathcal{F}_s\right) \\ &= \exp[-\lambda(t-s)(1 - e^{-\theta})] \end{aligned}$$

by the martingale condition. Hence N has independent increments, $N(t) - N(s)$ having the Poisson distribution with parameter $\lambda(t-s)$. ●

(10) Example. Wiener process. Let $\{W(t) : t \geq 0\}$ be a standard Wiener process with continuous sample paths, and let $\mathcal{F}_t$ be the σ-field generated by $\{W(u) : 0 \leq u \leq t\}$. It is easily seen that $W(t)$, $W(t)^2 - t$, and $\exp[\theta W(t) - \frac{1}{2}\theta^2 t]$ constitute martingales with respect to $\boldsymbol{\mathcal{F}}$. Conversely it may be shown that, if $W(t)$ and $W(t)^2 - t$ are martingales with continuous sample paths, and $W(0) = 0$, then W is a standard Wiener process; this is sometimes called 'Lévy's characterization theorem'. ●

Versions of the convergence and optional stopping theorems are valid in continuous time.

(11) Convergence theorem. *Let $(Y, \boldsymbol{\mathcal{F}})$ be a martingale with right-continuous sample paths. If $\mathbb{E}|Y(t)| \leq M$ for some M and all t, then $Y_\infty = \lim_{t\to\infty} Y(t)$ exists almost surely. If, in addition, $(Y, \boldsymbol{\mathcal{F}})$ is uniformly integrable then $Y(t) \xrightarrow{1} Y_\infty$.*

Sketch proof. For each $m \geq 1$, the sequence $\{(Y(n2^{-m}), \mathcal{F}_{n2^{-m}}) : n \geq 0\}$ constitutes a discrete-time martingale. Under the conditions of the theorem, these martingales converge as $n \to \infty$. The right-continuity property of Y may be used to fill in the gaps. ■

(12) Optional stopping theorem. *Let $(Y, \boldsymbol{\mathcal{F}})$ be a uniformly integrable martingale with right-continuous sample paths. Suppose that S and T are stopping times such that $S \leq T$. Then $\mathbb{E}(Y(T) \mid \mathcal{F}_S) = Y(S)$.*

The idea of the proof is to 'discretize' Y as in the previous proof, use the optional stopping theorem for uniformly integrable discrete-time martingales, and then pass to the continuous limit.

Exercises for Section 12.7

1. Let X be a continuous-time Markov chain with finite state space S and generator $\mathbf{G}$. Let $\eta = \{\eta(i) : i \in S\}$ be a root of the equation $\mathbf{G}\eta' = \mathbf{0}$. Show that $\eta(X(t))$ constitutes a martingale with respect to $\mathcal{F}_t = \sigma(\{X(u) : u \leq t\})$.

2. Let N be a Poisson process with intensity λ and $N(0) = 0$, and let $T_a = \min\{t : N(t) = a\}$, where a is a positive integer. Assuming that $\mathbb{E}\{\exp(\psi T_a)\} < \infty$ for sufficiently small positive ψ, use the optional stopping theorem to show that $\text{var}(T_a) = a\lambda^{-2}$.

3. Let $S_m = \sum_{r=1}^{m} X_r$, $m \leq n$, where the X_r are independent and identically distributed with finite mean. Denote by $U_1, U_2, \dots, U_n$ the order statistics of n independent variables which are uniformly distributed on $(0, t)$, and set $U_{n+1} = t$. Show that $R_m = S_m/U_{m+1}$, $0 \leq m \leq n$, is a backward martingale with respect to a suitable sequence of σ-fields, and deduce that

$$\mathbb{P}\big(R_m \geq 1 \text{ for some } m \leq n \,\big|\, S_n = y\big) \leq \min\{y/t, 1\}.$$

12.8 Some examples

(1) Example. Gambling systems. In practice, gamblers do not invariably follow simple strategies, but they vary their manner of play according to a personal system. One way of expressing this is as follows. For a given game, write $Y_0, Y_1, \dots$ for the sequence of capitals obtained by wagering one unit on each play; we allow the Y_i to be negative. That is to say, let Y_0 be the initial capital, and let Y_n be the capital after n gambles each involving a unit stake. Take as filtration the sequence $\mathcal{F}$ given by $\mathcal{F}_n = \sigma(Y_0, Y_1, \dots, Y_n)$. A general betting strategy would allow the gambler to vary her stake. If she bets S_n on the nth play, her profit is $S_n(Y_n - Y_{n-1})$, since $Y_n - Y_{n-1}$ is the profit resulting from a stake of one unit. Hence the gambler's capital Z_n after n plays satisfies

$$Z_n = Z_{n-1} + S_n(Y_n - Y_{n-1}) = Y_0 + \sum_{i=1}^{n} S_i(Y_i - Y_{i-1}), \tag{2}$$

where Y_0 is the gambler's initial capital. The S_n must have the following special property. The gambler decides the value of S_n in advance of the nth play, which is to say that S_n depends only on $Y_0, Y_1, \dots, Y_{n-1}$, and therefore S_n is $\mathcal{F}_{n-1}$-measurable. That is, $(S, \mathcal{F})$ must be a predictable process.

The sequence Z given by (2) is called the *transform* of Y by S. If Y is a martingale, we call Z a *martingale transform*.

Suppose $(Y, \mathcal{F})$ is a martingale. The gambler may hope to find a predictable process $(S, \mathcal{F})$ (called a *system*) for which the martingale transform Z (of Y by S) is no longer a martingale. She hopes in vain, since all martingale transforms have the martingale property. Here is a version of that statement.

(3) Theorem. *Let $(S, \mathcal{F})$ be a predictable process, and let Z be the transform of Y by S. Then:*

(a) *if $(Y, \mathcal{F})$ is a martingale, then $(Z, \mathcal{F})$ is a martingale so long as $\mathbb{E}|Z_n| < \infty$ for all n,*

(b) *if $(Y, \mathcal{F})$ is a submartingale and in addition $S_n \geq 0$ for all n, then $(Z, \mathcal{F})$ is a submartingale so long as $\mathbb{E}(Z_n^+) < \infty$ for all n.*

Proof. From (2),

$$\begin{aligned}\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) - Z_n &= \mathbb{E}\big[S_{n+1}(Y_{n+1} - Y_n) \,\big|\, \mathcal{F}_n\big] \\ &= S_{n+1}\big[\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) - Y_n\big].\end{aligned}$$

The last term is zero if Y is a martingale, and is non-negative if Y is a submartingale and $S_{n+1} \geq 0$. ■

A number of special cases are of value.

(4) *Optional skipping.* At each play, the gambler either wagers a unit stake or skips the round; S equals either 0 or 1.

(5) *Optional stopping.* The gambler wagers a unit stake on each play until the (random) time T, when she gambles for the last time. That is,

$$S_n = \begin{cases} 1 & \text{if } n \leq T, \\ 0 & \text{if } n > T, \end{cases}$$

and $Z_n = Y_{T\wedge n}$. Now $\{T = n\} = \{S_n = 1,\ S_{n+1} = 0\} \in \mathcal{F}_n$, so that T is a stopping time. It is a consequence of (3) that $(Y_{T\wedge n}, \mathcal{F}_n)$ is a martingale whenever Y is a martingale, as established earlier.

(6) *Optional starting.* The gambler does not play until the $(T+1)$th play, where T is a stopping time. In this case $S_n = 0$ for $n \le T$. ●

(7) Example. Likelihood ratios. Let $X_1, X_2, \dots$ be independent identically distributed random variables with common density function f. Suppose that it is known that $f(\cdot)$ is either $p(\cdot)$ or $q(\cdot)$, where p and q are given (different) densities; the statistical problem is to decide which of the two is the true density. A common approach is to calculate the *likelihood ratio*

$$Y_n = \frac{p(X_1)p(X_2)\cdots p(X_n)}{q(X_1)q(X_2)\cdots q(X_n)}$$

(assume for neatness for $q(x) > 0$ for all x), and to adopt the strategy:

(8) $$\text{decide } p \text{ if } Y_n \ge a, \quad \text{decide } q \text{ if } Y_n < a,$$

where a is some predetermined positive level.

Let $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. If $f = q$, then

$$\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n \mathbb{E}\left(\frac{p(X_{n+1})}{q(X_{n+1})}\right) = Y_n \int_{-\infty}^{\infty} \frac{p(x)}{q(x)} q(x)\,dx = Y_n$$

since p is a density function. Furthermore

$$\mathbb{E}|Y_n| = \int_{\mathbb{R}^n} \frac{p(x_1)p(x_2)\cdots p(x_n)}{q(x_1)q(x_2)\cdots q(x_n)} q(x_1)\cdots q(x_n)\,dx_1\cdots dx_n = 1.$$

It follows that $(Y, \mathcal{F})$ is a martingale, under the assumption that q is the common density function of the X_i. By an application of the convergence theorem, the limit $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely under this assumption. We may calculate Y_∞ explicitly as follows:

$$\log Y_n = \sum_{i=1}^{n} \log\left(\frac{p(X_i)}{q(X_i)}\right),$$

the sum of independent identically distributed random variables. The logarithm function is concave, so that

$$\mathbb{E}\left(\log\left(\frac{p(X_1)}{q(X_1)}\right)\right) < \log\left(\mathbb{E}\left(\frac{p(X_1)}{q(X_1)}\right)\right) = 0$$

by Jensen's inequality, Exercise (5.6.1). Applying the strong law of large numbers (7.5.1), we deduce that $n^{-1}\log Y_n$ converges almost surely to some point in $[-\infty, 0)$, implying that $Y_n \xrightarrow{\text{a.s.}} Y_\infty = 0$. (This is a case when the sequence Y_n does not converge to Y_∞ in mean, and $Y_n \ne \mathbb{E}(Y_\infty \mid \mathcal{F}_n)$.)

The fact that $Y_n \xrightarrow{\text{a.s.}} 0$ tells us that $Y_n < a$ for all large n, and hence the decision rule (8) gives the correct answer (that is, that $f = q$) for all large n. Indeed the probability that the outcome of the decision rule is ever in error satisfies $\mathbb{P}(Y_n \ge a \text{ for any } n \ge 1) \le a^{-1}$, by the maximal inequality (12.6.6). ●

(9) Example. Epidemics. A village contains $N+1$ people, one of whom is suffering from a fatal and infectious illness. Let $S(t)$ be the number of susceptible people at time t (that is, living people who have not yet been infected), let $I(t)$ be the number of infectives (that is, living people with the disease), and let $D(t) = N + 1 - S(t) - I(t)$ be the number of dead people. Assume that $(S(t), I(t), D(t))$ is a (trivariate) Markov chain in continuous time with transition rates

$$(s, i, d) \to \begin{cases} (s-1, i+1, d) & \text{at rate } \lambda si, \\ (s, i-1, d+1) & \text{at rate } \mu i; \end{cases}$$

that is to say, some susceptible becomes infective at rate λsi, and some infective dies at rate μi, where s and i are the numbers of susceptibles and infectives. This is the model of (6.12.4) with the introduction of death. The three variables always add up to $N+1$, and therefore we may suppress reference to the dead, writing (s, i) for a typical state of the process. Suppose we can find $\boldsymbol{\psi} = \{\psi(s, i) : 0 \le s + i \le N + 1\}$ such that $\mathbf{G}\boldsymbol{\psi}' = \mathbf{0}$, where $\mathbf{G}$ is the generator of the chain; think of $\boldsymbol{\psi}$ as a row vector. Then the transition semigroup $\mathbf{P}_t = e^{t\mathbf{G}}$ satisfies

$$\mathbf{P}_t\boldsymbol{\psi}' = \boldsymbol{\psi}' + \sum_{n=1}^{\infty} \frac{1}{n!} t^n \mathbf{G}^n \boldsymbol{\psi}' = \boldsymbol{\psi}',$$

whence it is easily seen (Exercise (12.7.1)) that $Y(t) = \psi(S(t), I(t))$ defines a continuous-time martingale with respect to the filtration $\mathcal{F}_t = \sigma\big(\{S(u), I(u) : 0 \le u \le t\}\big)$.

Now $\mathbf{G}\boldsymbol{\psi}' = \mathbf{0}$ if and only if

(10) $$\lambda si\psi(s-1, i+1) - (\lambda si + \mu i)\psi(s, i) + \mu i\psi(s, i-1) = 0$$

for all relevant i and s. If we look for a solution of the form $\psi(s, i) = \alpha(s)\beta(i)$, we obtain

(11) $$\lambda s\alpha(s-1)\beta(i+1) - (\lambda s + \mu)\alpha(s)\beta(i) + \mu\alpha(s)\beta(i-1) = 0.$$

Viewed as a difference equation in the $\beta(i)$, this suggests setting

(12) $$\beta(i) = B^i \quad \text{for some } B.$$

With this choice and a little calculation, one finds that

(13) $$\alpha(s) = \prod_{k=s+1}^{N} \left(\frac{\lambda Bk - \mu(1-B)}{\lambda B^2 k} \right)$$

will do. With such choices for α and β, the process $\psi(S(t), I(t)) = \alpha(S(t))\beta(I(t))$ constitutes a martingale.

Two possibilities spring to mind. Either everyone dies ultimately (that is, $S(t) = 0$ before $I(t) = 0$) or the disease dies off before everyone has caught it (that is, $I(t) = 0$ before $S(t) = 0$). Let $T = \inf\{t : S(t)I(t) = 0\}$ be the time at which the process terminates. Clearly T is a stopping time, and therefore

$$\mathbb{E}\big(\psi(S(T), I(T))\big) = \psi(S(0), I(0)) = \alpha(N)\beta(1) = B,$$

which is to say that

$$(14)\qquad \mathbb{E}\left(B^{I(T)} \prod_{k=S(T)+1}^{N} \left(\frac{\lambda Bk - \mu(1-B)}{\lambda B^2 k} \right) \right) = B$$

for all B. From this equation we wish to determine whether $S(T) = 0$ or $I(T) = 0$, corresponding to the two possibilities described above.

We have a free choice of B in (14), and we choose the following values. For $1 \le r \le N$, define $B_r = \mu/(\lambda r + \mu)$, so that $\lambda r B_r - \mu(1 - B_r) = 0$. Substitute $B = B_r$ in (14) to obtain

$$(15)\qquad \mathbb{E}\left(B_r^{S(T)-N} \prod_{k=S(T)+1}^{N} \left(\frac{k-r}{k} \right) \right) = B_r$$

(remember that $I(T) = 0$ if $S(T) \neq 0$). Put $r = N$ to get $\mathbb{P}(S(T) = N) = B_N$. More generally, we have from (15) that $p_j = \mathbb{P}(S(T) = j)$ satisfies

$$(16)\qquad p_N + \frac{N-r}{NB_r} p_{N-1} + \frac{(N-r)(N-r-1)}{N(N-1)B_r^2} p_{N-2} + \cdots + \frac{(N-r)!\,r!}{N!\,B_r^{N-r}} p_r = B_r,$$

for $1 \le r \le N$. From these equations, $p_0 = \mathbb{P}(S(T) = 0)$ may in principle be calculated. ●

(17) Example. Our final two examples are relevant to mathematical analysis. Let $f : [0, 1] \to \mathbb{R}$ be a (measurable) function such that

$$(18)\qquad \int_0^1 |f(x)|\,dx < \infty;$$

that is, f is integrable. We shall show that there exists a sequence $\{f_n : n \ge 0\}$ of step functions such that $f_n(x) \to f(x)$ as $n \to \infty$, except possibly for an exceptional set of values of x having Lebesgue measure 0.

Let X be uniformly distributed on $[0, 1]$, and define X_n by

$$(19)\qquad X_n = k2^{-n} \quad \text{if} \quad k2^{-n} \le X < (k+1)2^{-n}$$

where k and n are non-negative integers. It is easily seen that $X_n \uparrow X$ as $n \to \infty$, and furthermore $2^n(X_n - X_{n-1})$ equals the nth term in the binary expansion of X.

Define $Y = f(X)$ and $Y_n = \mathbb{E}(Y \mid \mathcal{F}_n)$ where $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$. Now $\mathbb{E}|f(X)| < \infty$ by (18), and therefore $(Y, \boldsymbol{\mathcal{F}})$ is a uniformly integrable martingale (see Example (12.3.9)). It follows that

$$(20)\qquad Y_n \to Y_\infty = \mathbb{E}(Y \mid \mathcal{F}_\infty) \quad \text{a.s. and in mean,}$$

where $\mathcal{F}_\infty = \sigma(X_0, X_1, X_2, \ldots) = \sigma(X)$. Hence $Y_\infty = \mathbb{E}(f(X) \mid X) = f(X)$, and in addition

$$(21)\qquad Y_n = \mathbb{E}(Y \mid \mathcal{F}_n) = \mathbb{E}(Y \mid X_0, X_1, \ldots, X_n) = \int_{X_n}^{X_n + 2^{-n}} f(u) 2^n\,du = f_n(X)$$

where $f_n : [0, 1] \to \mathbb{R}$ is the step function defined by

$$f_n(x) = 2^n \int_{x_n}^{x_n+2^{-n}} f(u)\, du,$$

x_n being the number of the form $k2^{-n}$ satisfying $x_n \le x < x_n + 2^{-n}$. We have from (20) that $f_n(X) \to f(X)$ a.s. and in mean, whence $f_n(x) \to f(x)$ for almost all x, and furthermore

$$\int_0^1 |f_n(x) - f(x)|\, dx \to 0 \quad \text{as} \quad n \to \infty. \qquad \bullet$$

(**22**) **Example.** This time let $f : [0, 1] \to \mathbb{R}$ be Lipschitz continuous, which is to say that there exists C such that

(**23**) $$|f(x) - f(y)| \le C|x - y| \quad \text{for all } x, y \in [0, 1].$$

Lipschitz continuity is of course somewhere between continuity and differentiability: Lipschitz-continuous functions are necessarily continuous but need not be differentiable (in the usual sense). We shall see, however, that there must exist a function g such that

$$f(x) - f(0) = \int_0^x g(u)\, du, \quad x \in [0, 1];$$

the function g is called the *Radon–Nikodým derivative* of f (with respect to Lebesgue measure).

As in the last example, let X be uniformly distributed on $[0, 1]$, define X_n by (19), and let

(**24**) $$Z_n = 2^n\big[f(X_n + 2^{-n}) - f(X_n)\big].$$

It may be seen as follows that $(Z, \mathcal{F})$ is a martingale (with respect to the filtration $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$). First, we check that $\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = Z_n$. To this end note that, conditional on $X_0, X_1, \dots, X_n$, it is the case that X_{n+1} is equally likely to take the value X_n or the value $X_n + 2^{-n-1}$. Therefore

$$\begin{aligned}\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) &= \tfrac{1}{2}2^{n+1}\big[f(X_n + 2^{-n-1}) - f(X_n)\big] \\ &\qquad + \tfrac{1}{2}2^{n+1}\big[f(X_n + 2^{-n}) - f(X_n + 2^{-n-1})\big] \\ &= 2^n\big[f(X_n + 2^{-n}) - f(X_n)\big] = Z_n.\end{aligned}$$

Secondly, by the Lipschitz continuity (23) of f, it is the case that $|Z_n| \le C$, whence $(Z, \mathcal{F})$ is a bounded martingale.

Therefore Z_n converges almost surely and in mean to some limit Z_∞, and furthermore $Z_n = \mathbb{E}(Z_\infty \mid \mathcal{F}_n)$ by Lemma (12.3.11). Now Z_∞ is $\mathcal{F}_\infty$-measurable where $\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n = \sigma(X_0, X_1, X_2, \dots) = \sigma(X)$, which implies that Z_∞ is a function of X, say $Z_\infty = g(X)$. As in equation (21), the relation

$$Z_n = \mathbb{E}\big(g(X) \,\big|\, X_0, X_1, \dots, X_n\big)$$

becomes

$$f(X_n + 2^{-n}) - f(X_n) = \int_{X_n}^{X_n+2^{-n}} g(u)\, du.$$

This is an ('almost sure') identity for X_n, which has positive probability of taking any value of the form $k2^{-n}$ for $0 \le k < 2^n$. Hence

$$f\big((k+1)2^{-n}\big) - f(k2^{-n}) = \int_{k2^{-n}}^{(k+1)2^{-n}} g(u)\, du,$$

whence, by summing,

$$f(x) - f(0) = \int_0^x g(u)\, du$$

for all x of the form $k2^{-n}$ for some $n \ge 1$ and $0 \le k < 2^n$. The corresponding result for general $x \in [0, 1]$ is obtained by taking a limit along a sequence of such 'dyadic rationals'. ●

12.9 Problems

1. Let Z_n be the size of the nth generation of a branching process with immigration in which the mean family size is μ ($\ne 1$) and the mean number of immigrants per generation is m. Show that

$$Y_n = \mu^{-n}\left\{Z_n - m\frac{1-\mu^n}{1-\mu}\right\}$$

defines a martingale.

2. In an age-dependent branching process, each individual gives birth to a random number of offspring at random times. At time 0, there exists a single progenitor who has N children at the subsequent times $B_1 \le B_2 \le \cdots \le B_N$; his family may be described by the vector $(N, B_1, B_2, \dots, B_N)$. Each subsequent member x of the population has a family described similarly by a vector $(N(x), B_1(x), \dots, B_{N(x)}(x))$ having the same distribution as $(N, B_1, \dots, B_N)$ and independent of all other individuals' families. The number $N(x)$ is the number of his offspring, and $B_i(x)$ is the time between the births of the parent and the ith offspring. Let $\{B_{n,r} : r \ge 1\}$ be the times of births of individuals in the nth generation. Let $M_n(\theta) = \sum_r e^{-\theta B_{n,r}}$, and show that $Y_n = M_n(\theta)/\mathbb{E}(M_1(\theta))^n$ defines a martingale with respect to $\mathcal{F}_n = \sigma(\{B_{m,r} : m \le n,\ r \ge 1\})$, for any value of θ such that $\mathbb{E}M_1(\theta) < \infty$.

3. Let $(Y, \mathcal{F})$ be a martingale with $\mathbb{E}Y_n = 0$ and $\mathbb{E}(Y_n^2) < \infty$ for all n. Show that

$$\mathbb{P}\left(\max_{1\le k\le n} Y_k > x\right) \le \frac{\mathbb{E}(Y_n^2)}{\mathbb{E}(Y_n^2) + x^2}, \qquad x > 0.$$

4. Let $(Y, \mathcal{F})$ be a non-negative submartingale with $Y_0 = 0$, and let $\{c_n\}$ be a non-increasing sequence of positive numbers. Show that

$$\mathbb{P}\left(\max_{1\le k\le n} c_k Y_k \ge x\right) \le \frac{1}{x}\sum_{k=1}^{n} c_k \mathbb{E}(Y_k - Y_{k-1}), \qquad x > 0.$$

Such an inequality is sometimes named after subsets of Hájek, Rényi, and Chow. Deduce Kolmogorov's inequality for the sum of independent random variables. [Hint: Work with the martingale $Z_n = c_n Y_n - \sum_{k=1}^n c_k \mathbb{E}(X_k \mid \mathcal{F}_{k-1}) + \sum_{k=1}^n (c_{k-1} - c_k) Y_{k-1}$ where $X_k = Y_k - Y_{k-1}$.]

5. Suppose that the sequence $\{X_n : n \geq 1\}$ of random variables satisfies $\mathbb{E}(X_n \mid X_1, X_2, \dots, X_{n-1}) = 0$ for all n, and also $\sum_{k=1}^{\infty} \mathbb{E}(|X_k|^r)/k^r < \infty$ for some $r \in [1, 2]$. Let $S_n = \sum_{i=1}^{n} Z_i$ where $Z_i = X_i/i$, and show that

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_{m+k} - S_m| \geq x\right) \leq \frac{1}{x^r}\mathbb{E}\big(|S_{m+n} - S_m|^r\big), \qquad x > 0.$$

Deduce that S_n converges a.s. as $n \to \infty$, and hence that $n^{-1}\sum_1^n X_k \xrightarrow{\text{a.s.}} 0$. [Hint: In the case $1 < r \leq 2$, prove and use the fact that $h(u) = |u|^r$ satisfies $h(v) - h(u) \leq (v-u)h'(u) + 2h((v-u)/2)$. Kronecker's lemma is useful for the last part.]

6. Let $X_1, X_2, \dots$ be independent random variables with

$$X_n = \begin{cases} 1 & \text{with probability } (2n)^{-1}, \\ 0 & \text{with probability } 1 - n^{-1}, \\ -1 & \text{with probability } (2n)^{-1}. \end{cases}$$

Let $Y_1 = X_1$ and for $n \geq 2$

$$Y_n = \begin{cases} X_n & \text{if } Y_{n-1} = 0, \\ nY_{n-1}|X_n| & \text{if } Y_{n-1} \neq 0. \end{cases}$$

Show that Y_n is a martingale with respect to $\mathcal{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$. Show that Y_n does not converge almost surely. Does Y_n converge in any way? Why does the martingale convergence theorem not apply?

7. Let $X_1, X_2, \dots$ be independent identically distributed random variables and suppose that $M(t) = \mathbb{E}(e^{tX_1})$ satisfies $M(t) = 1$ for some $t > 0$. Show that $\mathbb{P}(S_k \geq x \text{ for some } k) \leq e^{-tx}$ for $x > 0$ and such a value of t, where $S_k = X_1 + X_2 + \cdots + X_k$.

8. Let Z_n be the size of the nth generation of a branching process with family-size probability generating function $G(s)$, and assume $Z_0 = 1$. Let ξ be the smallest positive root of $G(s) = s$. Use the martingale convergence theorem to show that, if $0 < \xi < 1$, then $\mathbb{P}(Z_n \to 0) = \xi$ and $\mathbb{P}(Z_n \to \infty) = 1 - \xi$.

9. Let $(Y, \mathcal{F})$ be a non-negative martingale, and let $Y_n^* = \max\{Y_k : 0 \leq k \leq n\}$. Show that

$$\mathbb{E}(Y_n^*) \leq \frac{e}{e-1}\Big\{1 + \mathbb{E}\big(Y_n(\log Y_n)^+\big)\Big\}.$$

[Hint: $a \log^+ b \leq a \log^+ a + b/e$ if $a, b \geq 0$, where $\log^+ x = \max\{0, \log x\}$.]

10. Let $X = \{X(t) : t \geq 0\}$ be a birth–death process with parameters λ_i, μ_i, where $\lambda_i = 0$ if and only if $i = 0$. Define $h(0) = 0$, $h(1) = 1$, and

$$h(j) = 1 + \sum_{i=1}^{j-1} \frac{\mu_1 \mu_2 \cdots \mu_i}{\lambda_1 \lambda_2 \cdots \lambda_i}, \qquad j \geq 2.$$

Show that $h(X(t))$ constitutes a martingale with respect to the filtration $\mathcal{F}_t = \sigma(\{X(u) : 0 \leq u \leq t\})$, whenever $\mathbb{E}h(X(t)) < \infty$ for all t. (You may assume that the forward equations are satisfied.)

Fix n, and let $m < n$; let $\pi(m)$ be the probability that the process is absorbed at 0 before it reaches size n, having started at size m. Show that $\pi(m) = 1 - \{h(m)/h(n)\}$.

11. Let $(Y, \mathcal{F})$ be a submartingale such that $\mathbb{E}(Y_n^+) \le M$ for some M and all n.

(a) Show that $M_n = \lim_{m\to\infty} \mathbb{E}(Y_{n+m}^+ \mid \mathcal{F}_n)$ exists (almost surely) and defines a martingale with respect to $\mathcal{F}$.

(b) Show that Y_n may be expressed in the form $Y_n = X_n - Z_n$ where $(X, \mathcal{F})$ is a non-negative martingale, and $(Z, \mathcal{F})$ is a non-negative supermartingale. This representation of Y is sometimes termed the 'Krickeberg decomposition'.

(c) Let $(Y, \mathcal{F})$ be a martingale such that $\mathbb{E}|Y_n| \le M$ for some M and all n. Show that Y may be expressed as the difference of two non-negative martingales.

12. Let $£Y_n$ be the assets of an insurance company after n years of trading. During each year it receives a total (fixed) income of $£P$ in premiums. During the nth year it pays out a total of $£C_n$ in claims. Thus $Y_{n+1} = Y_n + P - C_{n+1}$. Suppose that $C_1, C_2, \dots$ are independent $N(\mu, \sigma^2)$ variables and show that the probability of ultimate bankruptcy satisfies

$$\mathbb{P}\big(Y_n \le 0 \text{ for some } n\big) \le \exp\left\{-\frac{2(P-\mu)Y_0}{\sigma^2}\right\}.$$

13. Pólya's urn. A bag contains red and blue balls, with initially r red and b blue where $rb > 0$. A ball is drawn from the bag, its colour noted, and then it is returned to the bag together with a new ball of the same colour. Let R_n be the number of red balls after n such operations.

(a) Show that $Y_n = R_n/(n+r+b)$ is a martingale which converges almost surely and in mean.

(b) Let T be the number of balls drawn until the first blue ball appears, and suppose that $r = b = 1$. Show that $\mathbb{E}\{(T+2)^{-1}\} = \frac{1}{4}$.

(c) Suppose $r = b = 1$, and show that $\mathbb{P}(Y_n \ge \frac{3}{4} \text{ for some } n) \le \frac{2}{3}$.

14. Here is a modification of the last problem. Let $\{A_n : n \ge 1\}$ be a sequence of random variables, each being a non-negative integer. We are provided with the bag of Problem (12.9.13), and we add balls according to the following rules. At each stage a ball is drawn from the bag, and its colour noted; we assume that the distribution of this colour depends only on the current contents of the bag and not on any further information concerning the A_n. We return this ball together with A_n new balls of the same colour. Write R_n and B_n for the numbers of red and blue balls in the urn after n operations, and let $\mathcal{F}_n = \sigma(\{R_k, B_k : 0 \le k \le n\})$. Show that $Y_n = R_n/(R_n + B_n)$ defines a martingale. Suppose $R_0 = B_0 = 1$, let T be the number of balls drawn until the first blue ball appears, and show that

$$\mathbb{E}\left(\frac{1 + A_T}{2 + \sum_{i=1}^{T} A_i}\right) = \frac{1}{2},$$

so long as $\sum_n \big(2 + \sum_{i=1}^{n} A_i\big)^{-1} = \infty$ a.s.

15. Labouchere system. Here is a gambling system for playing a fair game. Choose a sequence $x_1, x_2, \dots, x_n$ of positive numbers.

Wager the sum of the first and last numbers on an evens bet. If you win, delete those two numbers; if you lose, append their sum as an extra term x_{n+1} $(= x_1 + x_n)$ at the right-hand end of the sequence.

You play iteratively according to the above rule. If the sequence ever contains one term only, you wager that amount on an evens bet. If you win, you delete the term, and if you lose you append it to the sequence to obtain two terms.

Show that, with probability 1, the game terminates with a profit of $\sum_1^n x_i$, and that the time until termination has finite mean.

This looks like another clever strategy. Show that the mean size of your largest stake before winning is infinite. (When Henry Labouchere was sent down from Trinity College, Cambridge, in 1852, his gambling debts exceeded $£6000$.)

16. Here is a martingale approach to the question of determining the mean number of tosses of a coin before the first appearance of the sequence HHH. A large casino contains infinitely many gamblers $G_1, G_2, \ldots$, each with an initial fortune of \$1. A croupier tosses a coin repeatedly. For each n, gambler G_n bets as follows. Just before the nth toss he stakes his \$1 on the event that the nth toss shows heads. The game is assumed fair, so that he receives a total of $\$p^{-1}$ if he wins, where p is the probability of heads. If he wins this gamble, then he *repeatedly* stakes his entire current fortune on heads, at the same odds as his first gamble. At the first subsequent tail he loses his fortune and leaves the casino, penniless. Let S_n be the casino's profit (losses count negative) after the nth toss. Show that S_n is a martingale. Let N be the number of tosses before the first appearance of HHH; show that N is a stopping time and hence find $\mathbb{E}(N)$.

Now adapt this scheme to calculate the mean time to the first appearance of the sequence HTH.

17. Let $\{(X_k, Y_k) : k \geq 1\}$ be a sequence of independent identically distributed random vectors such that each X_k and Y_k takes values in the set $\{-1, 0, 1, 2, \ldots\}$. Suppose that $\mathbb{E}(X_1) = \mathbb{E}(Y_1) = 0$ and $\mathbb{E}(X_1 Y_1) = c$, and furthermore X_1 and Y_1 have finite non-zero variances. Let U_0 and V_0 be positive integers, and define $(U_{n+1}, V_{n+1}) = (U_n, V_n) + (X_{n+1}, Y_{n+1})$ for each $n \geq 0$. Let $T = \min\{n : U_n V_n = 0\}$ be the first hitting time by the random walk (U_n, V_n) of the axes of $\mathbb{R}^2$. Show that $\mathbb{E}(T) < \infty$ if and only if $c < 0$, and that $\mathbb{E}(T) = -\mathbb{E}(U_0 V_0)/c$ in this case. [Hint: You might show that $U_n V_n - cn$ is a martingale.]

18. The game 'Red Now' may be played by a single player with a well shuffled conventional pack of 52 playing cards. At times $n = 1, 2, \ldots, 52$ the player turns over a new card and observes its colour. Just once in the game he must say, just before exposing a card, "Red Now". He wins the game if the next exposed card is red. Let R_n be the number of red cards remaining face down after the nth card has been turned over. Show that $X_n = R_n/(52 - n)$, $0 \leq n < 52$, defines a martingale. Show that there is no strategy for the player which results in a probability of winning different from $\frac{1}{2}$.

19. A businessman has a redundant piece of equipment which he advertises for sale, inviting "offers over £1000". He anticipates that, each week for the foreseeable future, he will be approached by one prospective purchaser, the offers made in week $0, 1, \ldots$ being $£1000X_0, £1000X_1, \ldots$, where $X_0, X_1, \ldots$ are independent random variables with a common density function f and finite mean. Storage of the equipment costs $£1000c$ per week and the prevailing rate of interest is α (> 0) per week. Explain why a sensible strategy for the businessman is to sell in the week T, where T is a stopping time chosen so as to maximize

$$\mu(T) = \mathbb{E}\left\{(1+\alpha)^{-T} X_T - \sum_{n=1}^{T} (1+\alpha)^{-n} c\right\}.$$

Show that this problem is equivalent to maximizing $\mathbb{E}\{(1+\alpha)^{-T} Z_T\}$ where $Z_n = X_n + c/\alpha$.

Show that there exists a unique positive real number γ with the property that

$$\alpha\gamma = \int_{\gamma}^{\infty} \mathbb{P}(Z_n > y)\, dy,$$

and that, for this value of γ, the sequence $V_n = (1+\alpha)^{-n} \max\{Z_n, \gamma\}$ constitutes a supermartingale. Deduce that the optimal strategy for the businessman is to set a target price τ (which you should specify in terms of γ) and sell the first time he is offered at least this price.

In the case when $f(x) = 2x^{-3}$ for $x \geq 1$, and $c = \alpha = \frac{1}{90}$, find his target price and the expected number of weeks he will have to wait before selling.

20. Let Z be a branching process satisfying $Z_0 = 1$, $\mathbb{E}(Z_1) < 1$, and $\mathbb{P}(Z_1 \geq 2) > 0$. Show that $\mathbb{E}(\sup_n Z_n) \leq \eta/(\eta - 1)$, where η is the largest root of the equation $x = G(x)$ and G is the probability generating function of Z_1.

21. Matching. In a cloakroom there are K coats belonging to K people who make an attempt to leave by picking a coat at random. Those who pick their own coat leave, the rest return the coats and try again at random. Let N be the number of rounds of attempts until everyone has left. Show that $\mathbb{E}N = K$ and $\text{var}(N) \le K$.

22. Let W be a standard Wiener process, and define

$$M(t) = \int_0^t W(u)\,du - \tfrac{1}{3}W(t)^3.$$

Show that $M(t)$ is a martingale, and deduce that the expected area under the path of W until it first reaches one of the levels a (> 0) or b (< 0) is $-\frac{1}{3}ab(a+b)$.

23. Let $W = (W_1, W_2, \dots, W_d)$ be a d-dimensional Wiener process, the W_i being independent one-dimensional Wiener processes with $W_i(0) = 0$ and variance parameter $\sigma^2 = d^{-1}$. Let $R(t)^2 = W_1(t)^2 + W_2(t)^2 + \cdots + W_d(t)^2$, and show that $R(t)^2 - t$ is a martingale. Deduce that the mean time to hit the sphere of $\mathbb{R}^d$ with radius a is a^2.

24. Let W be a standard one-dimensional Wiener process, and let $a, b > 0$. Let T be the earliest time at which W visits either of the two points $-a$, b. Show that $\mathbb{P}(W(T) = b) = a/(a+b)$ and $\mathbb{E}(T) = ab$. In the case $a = b$, find $\mathbb{E}(e^{-sT})$ for $s > 0$.