Brownian Models in Economics Nancy L. Stokey

CHAPTER 4: MARTINGALES

Martingales are an example of mathematics at its best, sublimely elegant and at the same time enormously useful. Initially developed to study questions that arise in gambling, the theory of martingales has subsequently been used to study a wide array of questions.

The treatment here is only an introduction, covering the key concepts and major results. Section 1 provides a formal definition and illustrates the idea with some examples. Section 2 shows how martingales can be constructed from the (stationary) transition function for a discrete-time Markov process, using an eigenvector of the transition matrix if the state space is discrete and an eigenfunction if the state space is continuous. Section 3 shows a systematic way to carry out the construction if the Markov process is the sum of i.i.d. random variables and extends the method to continuous-time processes, both Brownian motions and more general diffusions. Sub- and supermartingale are defined in section 4, and two results are proved. The first describes two important ways that submartingales arise, and the second provides an extension of Kolmogorov's inequality. The optional stopping theorem, a fundamental result for stopping times applied to martingales, is presented in section 5 and extended in section 6. Section 7 provides a statement and proof of the martingale convergence theorem.

1. Definition, examples

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $\mathbb{F} = \{\mathfrak{F}_t,\ t \geq 0\}$ a filtration contained in $\mathfrak{F}$; and $\{Z(t),\ t \geq 0\}$ a stochastic process adapted to $\mathbb{F}$. Then $[Z, \mathbb{F}]$ is a **martingale** (or Z is a **martingale with respect to** $\mathbb{F}$) if for all $t \geq 0$,

(i)

$$\mathrm{E}[\ |Z(t)|\ ] < \infty, \tag{1}$$

(ii) with probability 1,

$$\mathrm{E}[Z(t) \mid \mathfrak{F}_s] = Z(s), \quad \text{all } 0 \leq s < t. \tag{2}$$

Note that the time index t may be discrete or continuous, and the horizon may be finite or infinite. Also note that the definition involves the filtration $\mathbb{F}$ in a fundamental way. If the filtration is understood, however, one may say simply that Z is a martingale.

It is useful to consider a few examples, which illustrate some of the many ways that martingales arise. In each case, to establish that a process is a martingale it must be verified that it is adapted to an appropriate filtered space and that it satisfies (1) and (2).

The first and most obvious way to construct a martingale is as the sum of independent mean-zero random variables. Indeed, the idea originated as the stochastic process describing the wealth of a gambler playing games of chance at fair odds.

Example A (Gambler's wealth): Let $X_1, ..., X_i, ...$ be a (finite or infinite) sequence of independent random variables on the probability space $(\Omega, \mathfrak{F}, P)$, each with mean zero and with finite absolute deviation: $\mathrm{E}[X_i] = 0$ and $\mathrm{E}[|X_i|] < \infty$, all i. Let $Z_0 = 0$, and for each $k = 1, 2, ...,$ define the random variable $Z_k = \sum_{i=1}^{k} X_i$ to be the partial sum of the first k elements in the sequence. For each k let $\mathfrak{F}_k \subset \mathfrak{F}$ be the smallest $\sigma-$algebra for which $\{Z_j\}_{j=1}^{k}$ are measurable, and let $\mathbb{F} = \{\mathfrak{F}_k\}$ be the filtration consisting of this (increasing) sequence. Then the stochastic process Z is a martingale on the filtered space $(\Omega, \mathbb{F}, P)$.

To see this, first note that by construction Z is adapted to $\mathbb{F}$. Then note that for any k,

$$\begin{aligned}
\mathrm{E}\left[|Z_k|\right] &= \mathrm{E}\left[\left|\sum_{i=1}^{k} X_i\right|\right] \leq \mathrm{E}\left[\sum_{i=1}^{k} |X_i|\right] \\
&= \sum_{i=1}^{k} \mathrm{E}\left[|X_i|\right] < \infty,
\end{aligned}$$

so (1) holds. Finally, to see that (2) holds, note that for any $j < k$,

$$\begin{aligned}
\mathrm{E}\left[Z_k \mid \mathfrak{F}_j\right] &= \mathrm{E}\left[\sum_{i=1}^{k} X_i \,\middle|\, \mathfrak{F}_j\right] \\
&= \sum_{i=1}^{j} X_i + \sum_{i=j+1}^{k} \mathrm{E}\left[X_i \mid \mathfrak{F}_j\right] \\
&= Z_j,
\end{aligned}$$

where the second line uses the fact that X_i is $\mathfrak{F}_j$-measurable for $i \leq j$, and the third the fact that the X_i's are mutually independent and all have mean zero.

An analogous argument establishes that a Brownian motion without drift is a martingale. The following exercise identifies some other useful martingales connected with Brownian motions.

Exercise 4.1: a. Show that if X is $(0, \sigma^2)$ Brownian motion, then $X^2 - \sigma^2 t$ and $(X/\sigma)^2 - t$ are martingales.

b. Show that if X is (μ, σ^2) Brownian motion, then $X - \mu t$, $(X - \mu t)^2 - \sigma^2 t$, and $\left[(X - \mu t)/\sigma\right]^2 - t$ are martingales.

The key feature of a martingale is that it has increments, at every point along every sample path, with (conditional) mean zero. Thus, martingales can be created from other stochastic processes by repeatedly (in discrete time) or continuously adjusting for the (conditional) expected increment along each sample path, as in part (b) of the previous exercise. The next examples illustrate other situations where martingales arise.

Example B (Family composition): Consider a society in which each family has exactly N children. Children are born in succession (no twins), and the probability is $\pi = 1/2$ that any new addition is a girl.

To analyze various questions about family composition in this society, an appropriate probability space is needed. To construct one, note first that an outcome in this setting is a vector of length N of the form $\omega = (b, g, ..., g, b, ..., b)$ describing the sequence of births. Let Ω be the set consisting of all such vectors. Note that Ω is a discrete space containing 2^N elements. Let $\mathfrak{F}$ be the complete σ-algebra for Ω i.e., the σ-algebra consisting of all subsets. Each outcome is equally likely, so let P be the probability measure on $\mathfrak{F}$ that assigns probability $1/2^N$ to each point. Then $(\Omega, \mathfrak{F}, P)$ is a probability space.

Define a filtration on this space by the sequence of births. Specifically, let $\omega_i,\ i = 1, ..., N$, denote the i^{th} component of the vector ω; for each k let $\mathfrak{F}_k \subset \mathfrak{F}$ be the smallest σ-algebra for which $\omega_1, ..., \omega_k$ are measurable; and let $\mathbb{F} = \{\mathfrak{F}_k\}$ be the filtration defined by this (increasing) sequence.

A number of different stochastic processes can be defined on the filtered space $(\Omega, \mathbb{F}, P)$. For example, let $g_0 = 0$ and for $k = 1, ..., N$, let g_k be the number of girls among the first k children. Notice that each g_k is a random variable on $(\Omega, \mathfrak{F}, P)$. Other processes can be constructed from $\{g_k\}$.

Let $X_0 = 0$ and let

$$X_k = g_k - (k - g_k), \quad k = 1, 2, ..., N,$$

be the excess of girls over boys in the first k births. Let $Y_0 = 0$ and let

$$Y_k = g_k - \frac{k}{2}, \quad k = 1, 2, ..., N,$$

be the excess of girls over $k/2$ in the first k births. Let $Z_0 = 0$, and let

$$Z_k = \frac{g_k}{k} - \frac{1}{2}, \quad k = 1, 2, ..., N,$$

be the deviation from $1/2$ of the fraction of girls in the first k births.

Exercise 4.2: a. Verify that $\{g_k\}, \{X_k\}, \{Y_k\}$ and $\{Z_k\}$ in Example B are stochastic processes on $(\Omega, \mathbb{F}, P)$.

b. Show that $\{X_k\}$ and $\{Y_k\}$ are martingales and that $\{Z_k\}$ is not.

The next example illustrates the very general principle that in a learning context, the sequence of Bayesian posteriors is martingale. The intuition for this fact is clear: if this were not so the observer would want to revise his current beliefs to incorporate the expected change next period.

Example C (Bayesian learning): Consider an experimenter trying to determine whether an urn is of type A or type B. Both types of (outwardly identical) urn contain a large number of black and white balls, but in different proportions. In the former the proportion of black balls is a and in the latter it is b, where $0 \leq a, b \leq 1$, with $a \neq b$. The urn the experimenter faces was drawn randomly from a population in which the proportion of type A urns is $p_0 \in (0, 1)$. To determine more precisely which type it is, he samples balls (with replacement), updating his beliefs by using Bayes' rule after each draw. Let $p_k, \ k = 1, 2, ...,$ denote his posterior probability after k balls have been drawn. Notice that each p_k is a random variable that takes $k + 1$ possible values. (Why?)

Exercise 4.3: a. Define an appropriate probability space for the situation described in Example C.

b. Define a filtration that makes the sequence of posteriors $\{p_k\}_{k=0}^{\infty}$ a stochastic process.

c. Show that $\{p_k\}_{k=0}^{\infty}$ is a martingale.

2. Martingales based on eigenvalues

The next two examples show how martingales can be constructed from the (stationary) transition function for a discrete time Markov process. If the state space is discrete, the construction uses an eigenvector of the transition matrix. If the state space is continuous, it use the continuous analog, an eigenfunction associated with the transition function.

Example D (Based on eigenvectors): Consider a Markov chain with state space $i \in \{1, 2, ..., I\}$ and the $I \times I$ transition matrix $\Pi = [\pi_{ij}]$, where π_{ij} is the probability of a transition from state i to state j.

Any function of the state is represented by a vector $f^T = (f_1, ..., f_I)$ containing the values for the function in states $1, ..., I$, and the vector Πf contains the conditional expected values for the function next period, given the possible states $1, ..., I$ this period. Recall that the pair (λ, v), with $\lambda \neq 0$ and $v \in \mathbf{R}^k$, are an eigenvalue and associated right eigenvector of Π if $\Pi v = \lambda v$.

Let (λ, v) be such a pair, let q_0 be a probability vector describing the distribution of the initial state, and let $\{i_k\}_{k=0}^{\infty}$ with $i_k \in \{1, 2, ..., I\}$ be the integer-valued stochastic process indicating the outcomes. Define the stochastic process $\{V_k\}_{k=0}^{\infty}$ by $V_k = v_{i_k}$, all k. Then

$$\begin{aligned}
\mathrm{E}\left[V_{k+1} \,|\, i_k\right] &= \mathrm{E}\left[v_{i_{k+1}} \,|\, i_k\right] \\
&= e_{i_k} \Pi v \\
&= \lambda v_{i_k} \\
&= \lambda V_k, \qquad k = 0, 1, ...,
\end{aligned}$$

where $e_i = (0, ..., 0, 1, 0, ..., 0)$ is a vector with a one in the i^{th} position and zeros elsewhere, and where the third line uses the eigenvector property. If $\lambda = 1$ then V is a martingale. Otherwise, the stochastic process Z defined by

$$Z_k = \lambda^{-k} V_k = \lambda^{-k} v_{i_k}, \quad k = 0, 1, ...,$$

is a martingale. In either case the filtration is the one generated by $\{i_k\}$. This filtration is (strictly) finer than the one generated by $\{v_{i_k}\}$ if v has two elements that are identical, $v_i = v_j$, for $i \neq j$.

Note that the argument in Example D continues to hold if the state space is infinite, if $I = \infty$.

The next example provides an analogous construction for Markov chains taking values in all of $\mathbf{R}$.

Example E (Based on eigenfunctions): Let F be the (stationary) transition function for a Markov chain $Y = \{Y_k\}$ with a continuous state space. That is,

$$\Pr\{Y_{k+1} \leq b \,|\, Y_k = a\} = F(b \,|\, a), \quad k = 1, 2, ...$$

Let $\lambda \neq 0$ be a real number and $\nu(\cdot)$ a function such that

$$\mathrm{E}\left[|\nu(Y_k)|\right] < \infty, \quad \text{all } k,$$

and

$$\int \nu(y) dF(y \,|\, a) = \lambda \nu(a), \quad \text{all } a.$$

Then ν is called an *eigenfunction* of F, with associated eigenvalue λ. An argument like the one above establishes that the stochastic process

$$Z_k = \lambda^{-k} \nu(Y_k), \quad k = 0, 1, 2, ...,$$

is a martingale.

3. The Wald martingale

An important class of Markov processes are those constructed as sequences of partial sums of i.i.d. random variables. In this case there is a systematic way to construct eigenfunctions, and the associated martingales are called **Wald martingales**. The next three examples show how they are constructed for discrete-time processes, Brownian motions, and general diffusions.

Example F (Discrete time): Let $\{X_k\}_{k=1}^{\infty}$ be a sequence of i.i.d. random variables with (common) c.d.f. G. Then the partial sums $Y_0 = 0$, and $Y_k = Y_{k-1} + X_k,\ k = 1, 2, ...$, form a Markov process with stationary increments, and the transition function is $F(b \,|\, a) = G(b - a)$, all a, b.

Suppose that for $\eta \neq 0$ the expected value

$$\lambda(\eta) = \int e^{\eta x} dG(x)$$

exists. Then the function $\nu(y; \eta) = e^{\eta y}$ is an eigenfunction of F with associated eigenvalue $\lambda(\eta)$. To see this, note that

$$\begin{aligned} \int \nu(y;\eta) dF(y \,|\, a) &= \int e^{\eta y} dG(y-a) \\ &= \int e^{\eta(a+x)} dG(x) \\ &= \nu(a;\eta)\lambda(\eta). \end{aligned}$$

Hence the argument in Example E implies that the stochastic process

$$Z_k = \lambda^{-k}(\eta)\nu(Y_k;\eta), \quad \text{all } k,$$

is a martingale.

A family of martingales can be constructed by varying the parameter η in Example F. In particular, if η can be chosen so that $\lambda(\eta) = 1 + r$, where r is a discount rate, then $\lambda(\eta)^{-k} = 1/\left(1 + r\right)^k$ plays the role of a discount factor.

The next two examples show that a similar argument can be used to construct martingales for Brownian motions and other diffusions. These martingales that will be used extensively in later chapters.

Example G (Brownian motion): Let X be a (μ, σ^2) Brownian motion. Recall (cf. Exercise 3.1) that for any $\rho \neq 0$, the stochastic process $Y(t) = \exp\{\rho X(t)\}$ is a geometric Brownian motion with parameters $\left(q(\rho),\, (\rho\sigma)^2\right)$, where

$$q(\rho) \equiv \rho\mu + \frac{1}{2}(\rho\sigma)^2 . \tag{3}$$

Consequently,

$$\begin{aligned} \mathrm{E}\{\exp[\rho X(t)]\} &= \mathrm{E}[Y(t)] \\ &= Y(0)e^{q(\rho)t} \\ &= \exp[\rho X(0) + q(\rho)t], \quad \text{all } t. \end{aligned}$$

Hence the stochastic process

$$M(t;\rho) \equiv \exp\{\rho X(t) - q(\rho)t\}, \quad \text{all } t, \tag{4}$$

is a martingale.

It follows immediately that if Y is a geometric Brownian motion with parameters $(\hat{\mu}, \sigma^2)$, then for any $\rho \neq 0$ the stochastic process

$$M(t;\rho) \equiv Y^{\rho}(t)e^{-\hat{q}(\rho)t}, \quad \text{all } t,$$

is a martingale, where

$$\hat{q}(\rho) \equiv \rho\left(\hat{\mu} - \frac{1}{2}\sigma^2\right) + \frac{1}{2}(\rho\sigma)^2 .$$

Example H (Diffusions): More generally, let X be a diffusion with stationary infinitesimal parameters $\mu(\cdot)$ and $\sigma(\cdot)$. Suppose the function $F(t,x)$ satisfies

$$F_t(t,x)+\mu(x)F_x(t,x)+\frac{1}{2}\sigma^2(x)F_{xx}(t,x)=0, \quad \text{all } t,x. \tag{5}$$

Then the stochastic process

$$M(t)=F(t,X(t)), \quad \text{all } t,\omega,$$

is a martingale. To see this, note that

$$\begin{aligned} \mathrm{E}\left[dM\right] &= \mathrm{E}\left[F_t dt + F_x\mu(X)dt + F_x\sigma(X)dW + \frac{1}{2}F_{xx}\sigma^2(x)dt\right] \\ &= 0, \end{aligned}$$

where the first line uses Ito's lemma, and the second uses (5) and the zero expectation property of stochastic integrals (Proposition 3.1).

In particular, the stochastic process

$$M(t)=e^{-rt}f(X(t)), \quad \text{all } t,\omega,$$

is a martingale if the pair (r,f) satisfy

$$-rf(x)+\mu(x)f'(x)+\frac{1}{2}\sigma^2(x)f''(x)=0, \quad \text{all } x. \tag{6}$$

For a Brownian motion, μ and σ^2 are constants, and (6) holds for any (r,f) defined by

$$r=q(\rho), \quad f(x)=e^{\rho x}, \quad \text{all } x,$$

where $\rho \neq 0$ and the function $q(\cdot)$ is defined in (3).

4. Sub- and supermartingales

Submartingales and supermartingales are defined by replacing condition (2) with an inequality. A stochastic process Z on the filtered space $(\Omega,\mathbb{F},P)$ is a **submartingale** if for all $t \geq 0$, (1) holds and

(ii) with probability 1,

$$\mathrm{E}\left[Z(t) \mid \mathfrak{F}_s\right] \geq Z(s), \quad \text{all } 0 \leq s < t. \tag{7}$$

It is a **supermartingale** if the inequality in (7) is reversed. Thus, on average a submartingale rises over time and a supermartingale falls. Notice that

Z is a submartingale if and only if $-Z$ is a supermartingale;

Z is a martingale if and only if it is both a submartingale and a supermartingale;

Z is a martingale if and only if both Z and $-Z$ are submartingales.

Exercise 4.4: Consider a modified version of Example A: suppose the gambler is playing at less than fair odds. That is, suppose that each of the X_i's has nonpositive mean, $\mathrm{E}[X_i] \leq 0$, all i. For $k = 1, 2, ...$, let $Y_k = -Z_k$ denote the gambler's net *loss* after k rounds of play. Show that Z_k is a submartingale.

Exercise 4.5: Consider a modified version of the society in Example B. Suppose that at each birth the probability of a girl is $\pi < 1/2$. Define the sequences g_k, X_k and Y_k as before. Show that X_k and Y_k are supermartingales.

Exercise 4.6: Consider a modified version of the Bayesian learning in Example C. Suppose that the experimenter is mistaken about the fraction of type A urns in the population from which the urn under study was drawn. The experimenter believes that the fraction of type A urns is $\hat{p}_0$, with $0 < \hat{p}_0 < p_0$. As before the experimenter draws balls from the urn sequentially, with replacement, and updates his beliefs after each draw by using Bayes' Rule. Let $\{\hat{p}_k\}$ denote his sequence of posteriors. Show that $\{\hat{p}_k\}$ is a submartingale.

Exercise 4.7: Show that if X is a Brownian motion with positive (negative) drift, then it is a submartingale (a supermartingale).

The following theorem describes two ways that submartingales arise.

THEOREM 4.1: a. If Z is a martingale on the filtered space $(\Omega, \mathbb{F}, P)$, ϕ is a measurable convex function, and $\phi(Z(t))$ is integrable, all t, then the stochastic process $\phi(Z)$ is a submartingale.

b. If Z is a submartingale on the filtered space $(\Omega, \mathbb{F}, P)$; ϕ is a measurable, increasing, convex function; and $\phi(Z(t))$ is integrable, all t, then the stochastic process $\phi(Z)$ is a submartingale.

PROOF: In each case it must be shown that $\phi(Z)$ is adapted to $(\Omega, \mathbb{F}, P)$ and that it satisfies (1) and (7), for all t.

a. For each t, since $Z(t)$ is $\mathfrak{F}_t$-measurable and ϕ is a measurable function, it follows immediately that $\phi(Z(t))$ is also $\mathfrak{F}_t$-measurable. Hence $\phi(Z)$ is adapted to $(\Omega, \mathbb{F}, P)$. By assumption $\phi(Z(t))$ is integrable, so it satisfies (1), for all t, and since Z is a martingale it satisfies (2), for all t. Then since ϕ is convex, it follows from Jensen's inequality that with probability 1,

$$\begin{aligned} \phi(Z(s)) &= \phi(\mathrm{E}[Z(t) \mid \mathfrak{F}_s]) \\ &\leq \mathrm{E}[\phi(Z(t) \mid \mathfrak{F}_s)], \quad \text{all } 0 \leq s < t, \end{aligned} \tag{8}$$

so $\phi(Z)$ satisfies (7), for all t.

(b) Most of the argument for part (a) applies. For the last step, notice that since Z is a submartingale, it satisfies (7) for all t. Hence

$$\begin{aligned} \phi(Z(s)) &\leq \phi(\mathrm{E}[Z(t) \mid \mathfrak{F}_s]) \\ &\leq \mathrm{E}[\phi(Z(t) \mid \mathfrak{F}_s)], \quad \text{all } 0 \leq s < t, \end{aligned}$$

where the first line uses the fact that ϕ is increasing, and the second the fact that it is convex. ■

Part (a) of Theorem 4.1 implies that if Z is a martingale, then $|Z|$ and Z^2 are submartingales. In particular, if X is a (μ, σ^2) Brownian motion then $X - \mu t$ is a martingale, so $|X - \mu t|$ and $[X - \mu t]^2$ are submartingales.

The following result, an extension of Kolmogorov's inequality, provides a useful bound.

THEOREM 4.2: If $\{Z_k\}_{k=1}^n$ is a submartingale, then for any $\alpha > 0$,

$$P\left[\max_{1\leq k\leq n} |Z_k| \geq \alpha\right] \leq \frac{1}{\alpha}\mathrm{E}\left[|Z_n|\right].$$

PROOF: Fix $\alpha > 0$ and define the (disjoint) sets A_k by

$$A_k = \left\{\omega \in \Omega : Z_k \geq \alpha \quad \text{and} \quad Z_j < \alpha, \quad j = 1, 2, ..., k-1\right\}.$$

Then

$$\begin{aligned}
\mathrm{E}\left[|Z_n|\right] &\geq \sum_{k=1}^n \int_{A_k} |Z_n| \, dP \\
&\geq \sum_{k=1}^n \int_{A_k} Z_k \, dP \\
&\geq \sum_{k=1}^n \alpha \, P\left(A_k\right) \\
&= \alpha P\left(\cup_{k=1}^n A_k\right) \\
&= \alpha P\left[\max_{1\leq k\leq n} |Z_k| \geq \alpha\right],
\end{aligned}$$

where the first line uses the fact that the A_k's are disjoint, with $\cup_{k=1}^n A_k \subseteq \Omega$; the second line uses the fact that $\{Z_k\}$ is a submartingale; and the last three lines use the definition of the A_k's. ■

For an application of this result, let $\{X_i\}_{i=1}^n$ be a sequence of random variables, each with zero mean and finite variance, and let $S_k = \sum_{i=1}^k X_i$, $k = 1, ..., n$, be the sequence of their partial sums. Clearly $\{S_k\}$ is a martingale, and hence by Theorem 4.1 $Z_k = |S_k|$ is a submartingale. In this case Theorem 4.2 asserts that

$$P\left[\max_{1\leq k\leq n} |S_k| > \alpha\right] = P\left[\max_{1\leq k\leq n} S_k^2 > \alpha^2\right] \leq \frac{1}{\alpha}\mathrm{E}\left[|S_n|\right],$$

which is Kolmogorov's inequality.

5. Optional stopping theorem

The optimal stopping theorem is a powerful and extremely useful result about stopping times for martingales and submartingales. (To avoid excessive duplication, the results will not be stated separately for supermartingales.) The theorem applies to a wide class of stochastic processes, including discrete-time processes and diffusions, so it applies for all of the processes considered here. The theorem has many forms. The one below is presented for its simplicity, and its implications for some of the examples are discussed. A stronger form is presented in section 6.

For any two dates s and t, let $s \wedge t \equiv \min\{s,t\}$ denote the earlier of the two. Then, if Z is a stochastic process and T is a stopping time, let $Z\left(T \wedge t\right)$ denote the 'stopped' process defined by

$$Z\left(T \wedge t, \omega\right) = \begin{cases} Z(t,\omega), & \text{if } t < T(\omega), \\ Z\left(T(\omega),\omega\right), & \text{if } t \geq T(\omega). \end{cases}$$

Along each sample path the stopped process replaces the fluctuating path after date T with the constant value $Z(T)$.

The optimal stopping theorem assets that if Z is a (sub)martingale and T is a stopping time, then the stopped process $Z(T \wedge t)$ is also a (sub)martingale. Moreover, the expected value of the stopped process at any date t (is bounded below by) is equal to the expectation of the initial value $Z(0)$ and (is bounded above by) is equal to the expected value of the unstopped process at t. Finally, if the stopping time is bounded, then the expected terminal value for the stopped process (is greater than) is equal to the expected value at the initial date.

Theorem 4.3 (Optional stopping theorem): Let Z be a (sub)martingale on the filtered space $(\Omega, \mathbb{F}, P)$ and T a stopping time. Then

(a) $Z(T \wedge t)$ is also a (sub)martingale, and it satisfies

$$\mathrm{E}\left[Z(0)\right] \quad (\leq) = \quad \mathrm{E}\left[Z(T \wedge t)\right] \quad (\leq) = \quad \mathrm{E}\left[Z(t)\right], \quad \text{all } t; \tag{9}$$

(b) if there exists $N < \infty$ such that $0 \leq T(\omega) \leq N$, all ω, then

$$\mathrm{E}\left[Z(0)\right] \quad (\leq) = \quad \mathrm{E}\left[Z(T)\right] \quad (\leq) = \quad \mathrm{E}\left[Z(N)\right]. \tag{10}$$

See Appendix B for a proof.

The intuition for part (a) of this result is clear. For a martingale, the constant value along the sample path after date $T(\omega)$ in the stopped process is equal to the expected value of the original process at all subsequent dates, $\mathrm{E}[Z(t) \mid Z(T)] = Z(T)$, all $t \geq T(\omega)$. Consequently, replacing the original (fluctuating) path with the (constant) stopped value does not change expectations taken at earlier dates. For a submartingale, the same reasoning produces the stated inequalities. If the stopping time T is uniformly bounded, then (10) follows immediately: set $t = N$ in (9) and note that $T \wedge N = T$.

To see more concretely what Theorem 4.3 says, it is useful to look again at the examples in section 1.

Example A' (Gambler's wealth): Recall the gambler. We saw in Example A that if the game has fair odds and the gambler simply plays without stopping, the stochastic process $\{Z_k\}$ describing his net gain is a martingale. Hence for any fixed k, his expected net gain after k rounds of play is zero, $\mathrm{E}[Z_k] = 0,\ k = 1, 2, ...$ Theorem 4.3 says something even stronger.

Suppose that the gambler has a 'system' that involves stopping when he is ahead. Any such system defines a stopping time T. We must now distinguish between potential rounds and rounds actually played. Part (a) of Theorem 4.3 says that if the gambler uses his system, then his sequence of net gains, $\{Z_{T \wedge k}\}_{k=1}^{\infty}$, is still a martingale. Hence his expected net gain after k potential rounds is zero, as it would have been if he had simply played all k rounds. That is,

$$0 = Z_0 = \mathrm{E}\left[Z_{T \wedge k}\right] = \mathrm{E}\left[Z_k\right], \quad k = 1, 2, ...$$

Part (b) says that if his stopping rule puts a finite upper bound N on the number of rounds actually played, then his expected net gain at the end of play is also zero. That is, no stopping rule can alter the martingale property of the net gains, and none can lead to an expected net gain (or loss).

Similarly, as shown in Exercise 4.4, if the gambler is playing at unfair odds the stochastic process $\{Y_k\} = \{-Z_k\}$ describing his net loss is a submartingale: his expected losses increase with each round of play. Part (a) of Theorem 4.3 says that the same is true if he uses a stopping rule: his sequence of net losses $\{Y_{T\wedge k}\}$ is still a submartingale. Its expected value after k rounds of potential play is bounded below by zero and bounded above by the expected value of the net loss if all k rounds are actually played. Thus, using a stopping rule can reduce his expect net loss but cannot produce an expected gain. Part (b) says that if there is a finite bound on the stopping time, then his expected net loss at the end of play is bounded below by zero and bounded above by the expected loss he would incur if he simply played all N rounds.

Example B' (Family composition): Recall the model of family composition. Suppose parents value sons over daughters and would like to tilt their expected family composition toward sons. Suppose that every family has at least one child and that there is an upper bound N on family size. The only tool parents have at their disposal is their decision when to stop having children. For example, they might use the rule: keep having children until a son is born or there are N daughters, and then stop.

We saw in Example B that if the probability of a girl is $\pi = 1/2$ at each birth, and if the family has N children, then the stochastic processes X_k describing the excess of girls over boys after k births, and Y_k describing the excess of girls over $k/2$ after k births, are martingales. Part (a) of Theorem 4.3 implies that if the family uses a stopping rule T, with $T \leq N$, then the corresponding stopped processes are also martingales. Part (b) implies that the expected values for these two variables are zero for completed families under the stopping rule, just as they are if every family has N children.

Similarly, as shown in Exercise 4.5, if the probability of a girl is $\pi < 1/2$ at each birth and the family has N children, then X_k and Y_k are supermartingales. Part (a) of Theorem 4.3 then implies that if families use a stopping rule T, the corresponding stopped sequences are also supermartingales. Parts (a) and (b) together imply that

$$\mathrm{E}\left[X_N\right] \leq \mathrm{E}\left[X_T\right] \leq \mathrm{E}\left[X_0\right] = 0,$$

and similarly for Y_k. That is, by either measure family composition is tilted towards boys (because of the uneven birth ratio) and is more heavily tilted for the unstopped process.

Exercise 4.8: Consider the family composition example with $\pi = 1/2$.

(a) Suppose all families have exactly N children, and let G be the fraction of girls in completed families. What are the possible values for the random variable G? What is the probability distribution over these outcomes?

(b) Suppose families use the stopping rule: continue having children until a son is born or there are N daughters, and then stop. Let H be the fraction of girls in completed families. What are the possible values for the random variable H? What is the probability distribution over these outcomes? What is average family size in this society?

Example C' (Bayesian learning): Recall the example of Bayesian learning. Suppose that an outside observer is watching the experiments, and that the experimenter wants to manipulate the observer's beliefs. For concreteness suppose that he would like to convince her that the urn is likely to be of type A. The only tool the experimenter can use to manipulate the observer's beliefs is a stopping rule. If the experimenter draws a ball, the outcome is seen by the observer. But the experimenter can stop sampling at any time, i.e., he can choose a stopping rule T.

Consider first the case where the observer's prior is p_0, the (correct) *ex ante* probability that the urn is type A. As the sampling progresses her sequence of posteriors is the stochastic process $\{p_k\}$ described in Example C. As shown there, this process is a martingale. Part (a) of Theorem 4.3 implies that if the experimenter uses a stopping rule T, the stochastic process $\{p_{k \wedge T}\}$ describing the observer's posteriors is still a martingale. Thus, even a devious experimenter cannot design a sampling rule that destroys the martingale property of the posteriors. (Neither can a clumsy one.)

Part (b) of Theorem 4.3 implies that if there is an upper bound N on the sample size under the stopping rule, then the expected value of the observer's posterior after sampling has stopped is equal to her prior before sampling begins, regardless of the stopping rule chosen by the experimenter. That is, $\mathrm{E}[p_T] = p_0$, for any stopping rule.

If the observer's prior is incorrect, with $0 < \hat{p}_0 < p_0$, then her sequence of posteriors as sampling proceeds is the stochastic process $\{\hat{p}_k\}$ in Exercise 4.6. We saw there that this process is a submartingale. In this case part (a) of Theorem 4.3 implies that if the experimenter uses a stopping rule T, the process $\{\hat{p}_{k \wedge T}\}$ describing the observer's posteriors is still a submartingale, and it satisfies the inequalities in (9). If the sample size is uniformly bounded above by N, part (b) put bounds on the observer's final beliefs $\hat{p}_T$.

Now suppose the experimenter can choose the stopping rule, and is interested in convincing the observer that the urn is likely to be of type A. That is, he wants to choose a stopping rule T that maximizes the observer's final posterior, $\mathrm{E}[\hat{p}_T]$. Since $\{\hat{p}_k\}$ is a submartingale, part (b) of Theorem 4.3 implies that the experimenter can do no better than to set $T = N$, always using the largest allowable sample. The idea behind this result is clear: since the observer's initial beliefs are biased downward, providing her with more information moves them, on average, further upward (closer to the true value).

If the observer's initial prior is biased in the other direction, if $p_0 < \hat{p}_0 < 1$, then the stochastic process $\{\hat{p}_k\}$ describing her posteriors is a supermartingale. In this case an analogous argument shows that an experimenter who wants to maximize $\mathrm{E}[\hat{p}_T]$ should choose the stopping rule $T = 0$, giving the observer no information.

6. Optional stopping theorem, extended

Part (b) of Theorem 4.3 requires that the stopping time T be uniformly bounded. Here the result is extended to cases where $T < \infty$ but there is no uniform upper bound on the stopping time. An example is then used to illustrate what can go wrong if a key assumption does not hold.

The main idea is as follows. Let $S_t \subseteq \Omega$ denote the set where the process has stopped by date t:

$$S_t = \{\omega \in \Omega : T(\omega) \leq t\}.$$

Then the expected value of the stopped process can be written as the sum of two parts:

$$\begin{aligned}
\mathrm{E}\left[Z_0\right] &= \mathrm{E}\left[Z_{T\wedge t}\right] \\
&= \int_{\Omega} Z_{T\wedge t}(\omega) dP(\omega) \\
&= \int_{S_t} Z_T\left(\omega\right) dP(\omega) + \int_{S_t^c} Z_t\left(\omega\right) dP(\omega), \quad \text{all } t,
\end{aligned}$$

where the first line uses part (a) of Theorem 4.3. To insure that the expression in the last line converges to $\mathrm{E}[Z_T]$ as $t \to \infty$, several restrictions are needed. First, the probability accounted for by the first term must converges to one, $\lim_{t\to\infty} P(S_t) = 1$. This holds if (and only if) $\Pr\{T < \infty\} = 1$. In addition, to insure that the first term has a well defined limit, the positive and negative parts of Z_T must have bounded integrals, i.e., Z_T must be integrable. Finally, the second term must converge to zero as $t \to \infty$. The next result states these requirements more formally.

Theorem 4.4 (Extension of optional stopping theorem): Let Z be a (sub)martingale on the filtered space $(\Omega, \mathbb{F}, P)$ and T a stopping time. If

$$\begin{aligned} &\text{(i)} \quad \Pr\{T < \infty\} = 1, \\ &\text{(ii)} \quad \mathrm{E}\left[|Z(T)|\right] < \infty, \\ &\text{(iii)} \quad \lim_{t\to\infty} \mathrm{E}\left[\left|Z(t)\, I_{\{T>t\}}\right|\right] = 0, \end{aligned}$$

then

$$\mathrm{E}\left[Z(0)\right] \quad (\leq) = \quad \mathrm{E}\left[Z(T)\right].$$

See Appendix B for a proof.

The following example illustrates why condition (iii) is needed. Consider a gambler playing at fair odds in a game where the bet is \$1 in the first round and is doubled at ever subsequent round. That is, at each round $k = 1, 2, ...$, the gambler wins or loses $X_k = \pm 2^{k-1}$, and the probability of winning is 1/2. Let $Z_0 = 0$ and $Z_k = \sum_{i=1}^{k} X_i,\ k = 1, 2, ...$, so $\{Z_k\}$ is the sequence of his net gains. Clearly $\{Z_k\}$ is a martingale, and $\mathrm{E}[Z_k] = 0$, $k = 1, 2, ...$

Suppose the gambler uses the following system to try to win: he stops playing when his net gain is positive and continues otherwise. Under this stopping rule T the evolution of his net wealth, $\{Z_{T\wedge k}\}$, is described by

	0	1	2	3	4	5	6
1		1/2	3/4	7/8	15/16	31/32	63/64
0	1						
-1		1/2					
-3			1/4				
-7				1/8			
-15					1/16		
-31						1/32	...

Each column represents a time period $k = 0, 1, 2, ...$, and each row represents a possible value for $Z_{T\wedge k} \in \left\{1, 0, -1, -3, -7, ..., 1-2^k, ...\right\}$. The entries in column k constitute the probability vector for the outcomes $Z_{T\wedge k}$. Notice that the stopped process $\{Z_{T\wedge k}\}$ satisfies

$$\mathrm{E}\left[Z_{T\wedge k}\right] = 0 = Z_0, \quad k = 0, 1, 2, ...,$$

in accord with part (a) of Theorem 4.3.

It is clear that for any $\varepsilon > 0$ there exists $n > 1$ such that $\Pr\{T > n\} < \varepsilon$. Hence $\Pr\{T < \infty\} = 1$. In addition, there is no problem integrating the constant function Z_T. But clearly

$$\mathrm{E}\left[Z_T\right] = 1 \neq 0 = Z_0.$$

To see why Theorem 4.4 does not apply, note that the gambler's total loss conditional on continued betting, multiplied by the probability that he is still betting, is

$$\begin{aligned} \mathrm{E}\left[Z_k I_{\{T>k\}}\right] &= \frac{1}{2^k}\left[\sum_{i=1}^{k} X_i\right] \\ &= 2^{-k}\left[-2^0 - 2^1 - ... - 2^{k-1}\right] \\ &= -\sum_{i=1}^{k} 2^{-i} \rightarrow -1 \text{ as } k \rightarrow \infty. \end{aligned}$$

(This fact can also be read directly from the Table above.) Hence condition (iii) of Theorem 4.4 does not hold: the gambler's losses increase fast enough to offset the declining probability that he is still playing.

Exercise 4.9: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables, each taking values ± 1 with equal probability. Let $Z_k = \sum_{i=1}^{k} X_i$ be the sequence of their partial sums. Fix an integer $M \geq 1$ and let T be the stopping time defined as the first time the process reaches M. By Theorem 4.3 the stopped process $Z_{T \wedge k}$ satisfies

$$0 = Z_0 = \mathrm{E}\left[Z_{T \wedge k}\right] = \mathrm{E}\left[Z_k\right], \qquad \text{all } k \geq 0.$$

In addition, it is a fact that $\Pr\{T < \infty\} = 1$. But clearly

$$\mathrm{E}\left[Z_T\right] = M \neq 0 = \mathrm{E}\left[Z_0\right].$$

Hence Theorem 4.4 does not hold.

Show that $\{Z_k\}$ violates the condition

$$\lim_{k \to \infty} \mathrm{E}\left[\left|Z_k I_{\{T>k\}}\right|\right] = 0.$$

The next result is a further extension of the optional stopping results in Theorems 4.3 and 4.4. The new feature here is that the expectation is conditioned on the information at a stopping time rather than a fixed date.

THEOREM 4.5: If $\{Z_k\}_{k=1}^{n}$ is a (sub)martingale and τ_1, τ_2 are stopping times with $1 \leq \tau_1 \leq \tau_2 \leq n$, then

$$\mathrm{E}\left[Z_{\tau_2} \mid \mathfrak{F}_{\tau_1}\right] \quad (\geq) = \quad Z_{\tau_1}.$$

For a proof see Billingsley (1995, Theorem 35.2). Theorem 4.5 will be used to establish a preliminary result that in turn will be used to prove the martingale convergence theorem.

7. Martingale convergence theorem

The martingale convergence theorem is one of the most famous results for stochastic processes. It has a wide array of uses, so its fame is justified. In this section it is stated and is proved for discrete time processes. A few of its implications are then discussed.

The proof draws on a preliminary result, a crossing property. If $\{Z_k\}$ is a submartingale, then in expectation it is nondecreasing. Fix any two numbers $\alpha < \beta$. An **upcrossing** occurs whenever the sample path rises from α to β. To count the upcrossings, define the sequence of stopping times $\{\tau_i\}$ as follows. For $i = 1$,

$$\tau_1 = \begin{cases} \min\{k \geq 1 : Z_k \leq \alpha\}, & \text{if } Z_k \leq \alpha, \quad \text{some } k \geq 1, \\ n, & \text{otherwise.} \end{cases}$$

Thereafter, if i is even,

$$\tau_i = \begin{cases} \min\{k > i-1 : Z_k \geq \beta\}, & \text{if } Z_k \geq \beta, \quad \text{some } k > i-1, \\ n, & \text{otherwise;} \end{cases}$$

and if i is odd,

$$\tau_i = \begin{cases} \min\{k > i-1 : Z_k \leq \alpha\}, & \text{if } Z_k \leq \alpha, \quad \text{some } k > i-1, \\ n, & \text{otherwise.} \end{cases}$$

Define M by $\tau_M = n$, and let U be the number of upcrossings. Figure 4.1 depicts a sample path $Z_k(\omega)$ with two upcrossings. The following results bounds the expected number of upcrossings.

LEMMA 4.6: If $\{Z_k\}_{k=1}^{n}$ is a submartingale, then

$$\mathrm{E}[U] \leq \frac{\mathrm{E}[|Z_n|] + |\alpha|}{\beta - \alpha}. \tag{11}$$

PROOF: Define $Y_k = \max\{\alpha, Z_k\}$. Figure 4.1 shows how Y_k, the solid line, alters the sample path for Z_k. Since Y_k is an increasing convex function of Z_k, by Theorem 4.1 the stochastic process $\{Y_k\}$ is also a submartingale. Moreover, $\{Y_k\}$ and $\{Z_k\}$ have the same stopping times $\{\tau_{i_k}\}$ and the same number of upcrossings, call it U, between α and β.

Note that

$$Y_n = Y_{\tau_1} + \sum_{i=2}^{M} \left[Y_{\tau_i} - Y_{\tau_{i-1}}\right].$$

Let Σ_e be the sum for i even, and Σ_o be the sum for i odd. Then take expectations and use the fact that $Y_{\tau_1} = \alpha$ to find that

$$\mathrm{E}\left[Y_n - \alpha\right] = \mathrm{E}\left[\Sigma_e\right] + \mathrm{E}\left[\Sigma_o\right].$$

Since Y_k is a submartingale and the τ_j's are stopping times, Theorem 4.5 implies that

$$\mathrm{E}\left[\Sigma_o\right] = \mathrm{E}\left[\sum_{j=3,5,\ldots}^{M} \left(Y_{\tau_j} - Y_{\tau_{j-1}}\right)\right] \geq 0.$$

The same holds for Σ_e, but something stronger is needed. The term Σ_e includes all the upcrossings plus, possibly, a remainder term. That is,

$$\Sigma_e \geq \begin{cases} (\beta - a)\,U + \left[Y_{\tau_M} - Y_{\tau_{M-1}}\right], & \text{if } M \text{ is even and } Y_{\tau_M} < \beta, \\ (\beta - a)\,U, & \text{otherwise.} \end{cases}$$

Since Y is a submartingale, $\mathrm{E}\left[Y_{\tau_M} - Y_{\tau_{M-1}} | Y_{\tau_{M-1}}\right] \geq 0$. Hence

$$\mathrm{E}\left[\Sigma_e\right] \geq (\beta - \alpha)\,\mathrm{E}\left[U\right],$$

and summing the two pieces gives

$$\begin{aligned} \mathrm{E}\left[|Y_n|\right] + |\alpha| &\geq \mathrm{E}\left[Y_n - \alpha\right] \\ &= \mathrm{E}\left[\Sigma_e\right] + \mathrm{E}\left[\Sigma_o\right] \\ &\geq (\beta - a)\,\mathrm{E}\left[U\right]. \quad \blacksquare \end{aligned}$$

Since Σ_o appears to be the sum of downcrossings, the conclusion that $\mathrm{E}[\Sigma_o] \geq 0$ may seem surprising. The idea is that if i is odd, having reached $Z_{\tau_{i-1}} \geq \beta$, there are two possibilities for τ_i. The process can fall back to α, or it can stay above α, in which case $\tau_i = \tau_M = n$. Since $\{Z_k\}$ is a submartingale, the expected increment, conditional on any state, is nonnegative. Hence conditional on $Z_{\tau_{i-1}}$, in expectation these two possibilities contribute a positive increment. In the sum Σ_o the positive contribution appears in the remainder terms, terms of the form $\left[Z_{\tau_M} - Z_{\tau_{M-1}}\right]$, which in expectation outweigh the others.

The final result is the famous Martingale Convergence Theorem. It says that if Z is a submartingale and is bounded in a certain sense, then with probability one it converges to a random variable Z^*. Thus, the result has two parts. First, for a.e. $\hat{\omega} \in \Omega$, the sample path $Z(t, \hat{\omega})$ converges pointwise. That is, there exists $Z^*(\omega) = \lim_{t\to\infty} Z(t, \hat{\omega})$. In addition, Z^* is integrable, so

$$\lim_{t\to\infty} \mathrm{E}\left[|Z_t - Z^*|\right] = 0.$$

THEOREM 4.7 (MARTINGALE CONVERGENCE): Let Z be a submartingale with

$$\sup_{t\geq 0} \mathrm{E}\left[|Z(t)|\right] = D < +\infty.$$

Then $Z(t) \to Z^*$ with probability 1, where Z^* is a random variable with $\mathrm{E}[|Z^*|] \leq D$.

PROOF: (For discrete time) Let $\{Z_k\}$ be a submartingale and fix any $\alpha < \beta$. By Lemma 4.6 the expected number of upcrossings of $Z_1, ..., Z_n$, call it U_n satisfies

$$\mathrm{E}\left[U_n\right] \leq \frac{\mathrm{E}\left[|Z_n|\right] + |\alpha|}{\beta - \alpha}.$$

The random variable U_n is nondecreasing and is bounded above by $(D + |\alpha|) / (\beta - \alpha)$, so it follows from the Monotone Convergence Theorem that $\lim_{n\to\infty} U_n$ is integrable, and consequently is finite-valued a.e.

For each ω, define

$$\begin{aligned} Z^{\sup}(\omega) &= \limsup_{k\to\infty} Z_k(\omega) \qquad \text{and} \\ Z_{\inf}(\omega) &= \liminf_{k\to\infty} Z_k(\omega). \end{aligned}$$

If $Z_{\inf}(\omega) < \alpha < \beta < Z^{\sup}(\omega)$, then $U_n(\omega)$ would diverge as $n \to \infty$. Hence

$$P\left[Z_{\inf} < \alpha < \beta < Z^{\sup}\right] = 0.$$

But the set where $Z_{\inf} < Z^{\sup}$ can be written as

$$\left[Z_{\inf} < Z^{\sup}\right] = \cup \left[Z_{\inf} < \alpha < \beta < Z^{\sup}\right],$$

where the union is over all rational pairs $\alpha < \beta$. Since each set on the right side has probability zero, so does their sum. Hence $P\left[Z_{\inf} < Z^{\sup}\right] = 0$.

That is $P\left[Z_{\inf} = Z^{\sup}\right] = 1$. Call their common value Z^*. By Fatou's lemma

$$\mathrm{E}\left[|Z^*|\right] \leq \liminf_{k\to\infty} \mathrm{E}\left[|Z_k|\right] \leq D,$$

so Z^* is integrable. Hence Z^* is finite with probability one. ■

It is essential for the result that $\mathrm{E}[|Z_k|]$ be bounded by some finite D. To see why, consider the following example. Let $\{X_i\}$ be a sequence of i.i.d. random variables taking values ±1 with equal probability, and let $Z_k = \sum_{i=1}^{k} X_i$ be the sequence of partial sums. Clearly $\{Z_k\}$ is a martingale. For large k, the distribution of Z_k is well approximated by a Normal with mean zero and a standard deviation that grows like $\sqrt{k}$. Hence $Z_k \nrightarrow Z^*$, for any Z^*. Theorem 4.7 does not apply here: for any finite D there exists k sufficiently large so that $\mathrm{E}[|Z_k|] > D$.

Exercise 4.9: For each of the following examples either (i) show that Theorem 4.7 applies and calculate Z^*, or (ii) explain why Theorem 4.7 does not apply.

a. Let $\{Z_k\} = \{p_k\}$ be the sequence of posteriors for the Bayesian learning problem in Example C, where the prior p_0 is correct.

b. Let $\{Z_k\} = \{\hat{p}_k\}$ be the sequence of posteriors for the Bayesian learning problem in Exercise 4.6, where the prior $\hat{p}_0 < p_0$ is too low.

c. Let $\{Z_k\}$ be the sequence of net gains defined in the previous section, for the gambler who stops if he wins and doubles his bet if he loses.

d. Let $Z = M(t; \rho)$ be a Wald martingale of the type defined in (4).

8. Notes

Karlin and Taylor (1975, Ch. 6 and 7) provide an excellent introductory discussion of stopping times and martingales, with many helpful examples. Breiman (1968) and Billingsley (1995) also have excellent discussions, including treatments of the martingale convergence theorem and various forms of the optional stopping theorem.