Discrete-Time Martingale

March 20, 2002

The notion of *martingale* is fundamental in stochastic analysis. It was first discussed by P. Lévy. The realization of its potential and the fundamental development of the subject, however, are due to J.L. Doob.

Consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. A **filtration** $\mathsf{F} = \{\mathcal{F}_n;\ n = 0, 1, \cdots\}$ is an increasing sequence of $\mathcal{F}$; that is, $\mathcal{F}_n$ is a σ-algebra and

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots \mathcal{F}.$$

We can regard $\mathcal{F}_n$ as the collective information up to day n. The space $(\Omega, \mathcal{F}, \mathsf{F}, \mathsf{P})$ is sometimes called a **filtered probability space**. Sometimes, we need the following notation

$$\mathcal{F}_\infty \doteq \sigma(\cup_{n=0}^{\infty} \mathcal{F}_n)$$

Consider a sequence of random variable $\{X_n;\ n \geq 0\}$. We say $\{X_n\}$ is **adapted** (or, F-**adapted**), if X_n is $\mathcal{F}_n$-measurable for each n. Intuitively, we can understand this as that X_n is known at day n. The filtration generated by this $\{X_n\}$ by

$$\mathcal{F}_n^X \doteq \sigma(X_0, \cdots, X_n), \qquad \forall\ \ n \geq 0,$$

is said to be the **natural filtration**, denoted by F^X. Note that $\{X_n\}$ is always adapted to its natural filtration.

Definition: We call the process $X = (X_n, \mathcal{F}_n;\ n \geq 0)$ a **martingale** (resp. **submartingale**, **supermartingale**) if

1. $(X_n; n \geq 0)$ is F-adapted.
2. $X_n \in \mathsf{L}^1$ for each $n \geq 0$.
3. $\mathsf{E}(X_{n+1} \,|\, \mathcal{F}_n) = X_n$ for each $n \geq 0$ (resp. "$\geq$", "$\leq$").

Intuitively, if X_n represents the cumulative wealth of a gambler after n-th play and $\mathcal{F}_n$ is the cumulative information available to him at that time, then a martingale represents a *fair game*, in the sense that his (conditional) expected funture fortune is exactly his current aggregate. Similarly, a submartingale (resp. supermartingale) represents a favorable (reps. unfavorable) game.

Exercise: If $X = (X_n, \mathcal{F}_n)$ is a submartingale, then $-X = (-X_n, \mathcal{F}_n)$ is a supermartingale, and vice versa.

Exercise: If $X = (X_n, \mathcal{F}_n)$ is a martingale (i.e. submartingale, supermartingale), then

$$\mathsf{E}(X_m \,|\, \mathcal{F}_n) = X_n, \qquad \forall\; 0 \le n \le m \qquad \text{(resp. “}\ge\text{”, “}\le\text{”)}.$$

Exercise: If $X = (X_n; n \ge 0)$ is a martingale (resp. submartingale, supermartingale) with respect to some filtration F, then $X = (X_n; n \ge 0)$ is also a martingale with respect to its natural filtration F^X. (Hint: use tower property).

Below is a collection of martingales.

Example: Suppose $X = (X_n; n \ge 1)$ is a sequence of independent random variables with $\mathsf{E}(X_j) = 0$. Define

$$S_0 \equiv 0, \quad S_n \doteq \sum_{j=1}^{n} X_j$$

$$\mathcal{F}_0 \doteq \{\emptyset, \Omega\}, \quad \mathcal{F}_n \doteq \sigma(X_1, \cdots, X_n).$$

Then the process $(S_n, \mathcal{F}_n;\ n \ge 0)$ is a maringale. In particualr, if $X = (X_n; n \ge 1)$ is a sequence of iid random variables with $\mathsf{E}(X_1) = \mu$, then $(S_n - n\mu, \mathcal{F}_n; n \ge 0)$ is a martingale.

Example: Suppose $X = (X_n \ge 1)$ is a sequence of independent random variables with $\mathsf{E}(X_j) \equiv 1$. Define

$$M_0 \equiv 1, \quad M_n \doteq \prod_{j=1}^{n} X_j$$

$$\mathcal{F}_0 \doteq \{\emptyset, \Omega\}, \quad \mathcal{F}_n \doteq \sigma(X_1, \cdots, X_n).$$

Then the process $M = (M_n, \mathcal{F}_n;\ n \ge 0)$ is a maringale. In particular, if $X = (X_n; n \ge 1)$ is a sequence of iid random variables with $\mathsf{E}(X_1) = \mu \ne 0$, then $(\mu^{-n} M_n, \mathcal{F}_n; n \ge 0)$ is a martingale.

Example (Wald's martingale): Suppose $X = (X_n; n \ge 1)$ is a sequence of iid random variables with moment generating function $\phi(\theta) = \mathsf{E}\left(e^{\theta X_j}\right)$. Then

$$M_0 \equiv 1, \qquad M_n \doteq \frac{e^{\theta S_n}}{\phi^n(\theta)}; \quad \text{where } S_n = X_1 + \cdots + X_n, \ \ \forall\; n \ge 1$$

$$\mathcal{F}_0 \doteq \{\emptyset, \Omega\}, \quad \mathcal{F}_n \doteq \sigma(X_1, \cdots, X_n)$$

is a martingale.

Example (Doob's martingale): For an arbitray filtration $\mathsf{F} = (\mathcal{F}_n;\ n \ge 0)$ and every $X \in \mathsf{L}^1$, one can define the following sequence

$$X_n \doteq \mathsf{E}(X \,|\, \mathcal{F}_n), \qquad \forall\; n \ge 0.$$

Then $(X_n, \mathcal{F}_n; n \ge 0)$ is a martingale.

Example (Lévy martingale): Suppose P and Q are two probability measures on space $(\Omega, \mathcal{F})$, and $\mathsf{Q} \ll \mathsf{P}$. Let $\mathsf{F} = (\mathcal{F}_n;\ n \geq 0)$ be an arbitrary filtration. Define sequence

$$X_n \doteq \left.\frac{d\mathsf{Q}}{d\mathsf{P}}\right|_{\mathcal{F}_n}, \qquad n \geq 0.$$

Then $X = (X_n, \mathcal{F}_n; n \geq 0)$ is a martingale. Here $\frac{d\mathsf{Q}}{d\mathsf{P}}\big|_{\mathcal{G}}$ is the Radon-Nikodým derivative when Q, P are restricted on $\mathcal{G}$, for some $\mathcal{G} \subseteq \mathcal{F}$.

Below is a collection of exercises.

Exercise: Suppose $(X_n; n \geq 1)$ is a sequence of iid random variables, with $\mathsf{E}(X_1) = \mu$. Then $(S_n; n \geq 0)$ is a martingale (resp. submartingale, supermartingale) if $\mu = 0$ (resp. "$\geq$", "$\leq$").

Exercise (likelihood ratio): Let $(X_n; n \geq 1)$ be a sequence of iid random variables with common density $f(x)$. Suppose $g(x)$ is another arbitrary density function. Define the process

$$M_0 \equiv 1, \qquad M_n \doteq \prod_{j=1}^{n} \frac{g(X_j)}{f(X_j)}, \quad \forall\, n \geq 1$$

$$\mathcal{F}_0 \doteq \{\emptyset, \Omega\}, \quad \mathcal{F}_n \doteq \sigma(X_1, \cdots, X_n).$$

Then $M = (M_n, \mathcal{F}_n)$ is a martingale.

Exercise: Consider the Doob's martingale $(X_n, \mathcal{F}_n)$ with $X_n \doteq \mathsf{E}(X \mid \mathcal{F}_n)$ where $X \in \mathsf{L}^1$. Show that (X_n) is a uniformly integrable sequence.

Proof. For any $\lambda > 0$, we have

$$\begin{aligned} \int_{\{|X_n| \geq \lambda\}} |X_n| \, d\mathsf{P} &= \int_{\{|X_n| \geq \lambda\}} |\mathsf{E}(X \mid \mathcal{F}_n)| \, d\mathsf{P} \leq \int_{\{|X_n| \geq \lambda\}} \mathsf{E}(|X| \mid \mathcal{F}_n) \, d\mathsf{P} \\ &= \int_{\{|X_n| \geq \lambda\}} |X| \, d\mathsf{P} \quad \text{(since } \{|X_n| \geq \lambda\} \in \mathcal{F}_n). \end{aligned}$$

However,

$$\mathsf{P}(|X_n| \geq \lambda) \leq \frac{\mathsf{E}|X_n|}{\lambda} \leq \frac{\mathsf{E}\big(\mathsf{E}(|X| \mid \mathcal{F}_n)\big)}{\lambda} = \frac{\mathsf{E}|X|}{\lambda} \quad \to \quad 0, \quad \text{as } \lambda \to \infty.$$

The uniform integrability follows from the absolute continuity of integral. □

Proposition: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale, ϕ is a convex function. If $\phi(X_n) \in \mathsf{L}^1$ for all n, then the process

$$\phi(X) \doteq \big(\phi(X_n), \mathcal{F}_n\big); \quad n \geq 0$$

is a submartingale. Special cases include $\phi(x) = |x|^p$ for some $p \geq 1$.

This proof is an immediate consequence of conditional Jensen inequality, and hence omitted. □

1 Doob decomposition and martingale transform

A stochastic process $A = \{A_n, n \geq 0\}$ is said to be **predictable** (or **F-predictable**) if A_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 0$ (with convention $\mathcal{F}_{-1} = \mathcal{F}_0$). We say A is **increasing** if $A_0 \leq A_1 \leq \cdots$. We have the following result.

Doob's decomposition: Every submartingale $X = (X_n, \mathcal{F}_n; n \geq 0)$ can be written as

$$X = M + A$$

where $M = (M_n, \mathcal{F}_n)$ is a martingale, and $A = (A_n, \mathcal{F}_n)$ is an increasing, predictable process, with $A_0 \equiv 0$. Such decomposition is unique.

Proof. Take

$$A_0 \equiv 0, \quad A_{n+1} = A_n + \mathsf{E}(X_{n+1} \,|\, \mathcal{F}_n) - X_n, \qquad \forall\, n \geq 0.$$

This process is obviously increasing and predictable. Furthermore, let $M_n \doteq X_n - A_n$, We have

$$\mathsf{E}(M_{n+1} \,|\, \mathcal{F}_n) = \mathsf{E}(X_{n+1} - A_{n+1} \,|\, \mathcal{F}_n) = \mathsf{E}(X_{n+1} \,|\, \mathcal{F}_n) - A_{n+1} = X_n - A_n = M_n, \ \ \forall\, n \geq 0.$$

As for the uniqueness, assume $X = M' + A'$ is another decomposition. We have

$$Y \doteq M - M' = A' - A$$

is a predictable martingale with $Y_0 = 0$, which implies that $Y_n \equiv 0$ (why?). □

Martingale transform: Suppose $M = (M_n, \mathcal{F}_n)$ is a martingale, and $A = (A_n, \mathcal{F}_n)$ is a predictable process. Then the process $X = (A \bullet M, \mathcal{F}_n)$ defined as

$$(A \bullet M)_0 \equiv 0, \qquad (A \bullet M)_n \doteq \sum_{j=1}^{n} A_j (M_j - M_{j-1}); \quad \forall\, n \geq 1$$

is called the **martingale transform of A by M.**

Proposition: If the process $X = (A \bullet M)$ is integrable, then its a martingale.

Proof. It is not difficult to deduce that

$$\begin{aligned} \mathsf{E}\big((A \bullet M)_{n+1} \,|\, \mathcal{F}_n\big) &= \sum_{j=1}^{n} \mathsf{E}\big(A_j (M_j - M_{j-1}) \,|\, \mathcal{F}_n\big) + \mathsf{E}\big(A_{n+1}(M_{n+1} - M_n) \,|\, \mathcal{F}_n\big) \\ &= \sum_{j=1}^{n} A_j (M_j - M_{j-1}) + A_{n+1} \mathsf{E}(M_{n+1} - M_n \,|\, \mathcal{F}_n) = (A \bullet M)_n \end{aligned}$$

for all $n \geq 0$. □

Remark: In general, the martingale transform will not result in martingales. However, the martingale transform always lead to a so-called *local martingale*, which is a very useful (especially in the continuous-time case) concept.

Remark: Martingale transform is the discrete analogue of continuous-time *stochastic integral* $\int A\, dM$. The theory of stochastic integral is one of the greatest achievement in modern probability.

Remark: If we think of A as a betting strategy for a player in a fair game M, that is, $A_n = \{$your bet on day $t = n$, placed right before the announcement of M_n, the n-th day's randomness$\}$. A_n is predictable in the sense that you will only use the information you obtained during the previous $n-1$ days. Then $(A \bullet M)_n$ is your cumulative wealth up to the end of day $t = n$. It is always a martingale (provided the integrability condition is satisfied), in other words, *you cannot beat the system!*.

Lemma: Suppose $M = (M_n, \mathcal{F}_n)$ is a submartingale (resp. supermatingale), and $A = (A_n, \mathcal{F}_n)$ is predictable and non-negative. Then the transform $(A \bullet M)$ is still a submartingale (resp. supermartingale) provided that $(A \bullet M)$ is integrable.

The proof is left as an exercise.

2 Stopping times and basic optional sampling theorem

A F-**stopping time** is a mapping $\tau : \Omega \to \{0, 1, 2, \cdots\} \cup \{\infty\}$ such that $\{\tau \leq n\} \in \mathcal{F}_n$ for all $n = 0, 1, 2, \cdots$. It is equivalent to replace $\{\tau \leq n\}$ by $\{\tau = n\}$ in this definition (exercise). Intuitively, stopping time is one type of "non-anticipative" decision in the sense that, at $t = n$, whether or not the process is stopped only depends on the history up to time (including) $t = n$. It should not rely on any information afterwards (i.e. cannot see the future).

Remark: Sometimes we use notation

$$X_\tau \doteq \sum_{j=0}^{\infty} X_j 1_{\{\tau = j\}} + X_\infty 1_{\{\tau = \infty\}}, \quad \text{where } X_\infty \doteq \limsup_n X_n \text{ by convention.}$$

It is easy to see that X_τ is $\mathcal{F}$-measurable.

Example (Hitting time): Suppose that $X = (X_n, \mathcal{F}_n;\ n \geq 0)$ is an adapted process, and $A \in \mathcal{B}(\mathsf{R})$. Let

$$\tau(\omega) \doteq \inf\{n \geq 0;\quad X_n(\omega) \in A\} = \text{the first time the process } X \text{ hit set } A.$$

By convention, $\inf\{\emptyset\} = \infty$; that is, $\tau = \infty$ if the process X never hits set A. Obviously

$$\{\tau \leq n\} = \cup_{j=1}^{n} \{X_j \in B\} \in \mathcal{F}_n, \quad \forall\ n \geq 0.$$

Hence τ is a stopping time.

Example: Suppose $X = (X_n, \mathcal{F}_n;\ n \geq 0)$ is an adapted process, and $A \in \mathcal{B}(\mathsf{R})$. Let

$$\tau(\omega) \doteq \sup\{0 \leq n \leq 100;\quad X_n(\omega) \in A\}$$

with convention $\sup\{\emptyset\} = 0$. Then X is *NOT* a stopping time in general.

The following elementary optional sampling theorem is very useful.

Optional sampling theorem: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale (resp. submartingale, supermartingale), and τ is an arbitrary F-stopping time. Then the **stopped process**

$$X^\tau = (X_{\tau\wedge n}, \mathcal{F}_n;\ n \ge 0)$$

is also a martingale (resp. submartingale, supermartingale). In particular,

$$\mathsf{E}(X_{\tau\wedge n}) = \mathsf{E}X_0, \quad \forall\, n \ge 0, \quad \text{(resp. “}\ge\text{”, “}\le\text{”)}.$$

Proof. We should express the stopped process X^τ as a martingale transform.

$$X_{\tau\wedge n} = X_0 + \sum_{j=1}^{n} 1_{\{\tau\ge j\}}(X_j - X_{j-1}) := X_0 + \sum_{j=1}^{n} A_j(X_j - X_{j-1}) = X_0 + (A \bullet M)_n, \quad \forall\, n \ge 0.$$

Note that the process $A = (1_{\{\tau\ge n\}}, \mathcal{F}_n)$ is a predictable process (why?), and the integrability of $(A \bullet M)$ is obvious since $0 \le A_j \le 1$. The claim follows readily. □

An immediate sorollary is the following special case of optional sampling theorem.

Optional sampling theorem: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale (resp. submartingale, supermartingale), and τ is a *bounded* F-stopping time; that is, $\mathsf{P}(\tau \le K) = 1$ for some constant K. Then

$$\mathsf{E}(X_\tau) = \mathsf{E}(X_0), \quad \text{(resp. “}\ge\text{”, “}\le\text{”)}.$$

It is worth mention that the optional sampling theorem does not hold in general. Later we will see more sufficient conditions such that $\mathsf{E}(X_\tau) = \mathsf{E}(X_0)$ (resp. “$\ge$”, “$\le$”) will hold for general stopping times under some integrability conditions. The following are several examples.

Example: Let $(X_1, X_2, \cdots)$ be a sequence of iid Bernoulli random variables with $\mathsf{P}(X = \pm 1) = \frac{1}{2}$. Let

$$S_n \doteq \sum_{j=1}^{n} X_j, \quad S_0 \equiv 0; \qquad \mathcal{F}_0 \equiv \{\emptyset, \Omega\}, \quad \mathcal{F}_n = \sigma(X_1, \cdots, X_n) = \sigma(S_0, S_1, \cdots, S_n).$$

Then $S = (S_n, \mathcal{F}_n)$ is a martingale; indeed, S is a symmetric simple random walk. The hitting time

$$\tau \doteq \inf\{n \ge 0;\ \ S_n = 1\},$$

Suppose we are interested in the distribution of the hitting time τ.

The moment generating function of X_j is

$$\mathsf{E}\left(e^{\theta X_j}\right) = \frac{1}{2}\left(e^{\theta} + e^{-\theta}\right) = \cosh\theta; \quad \forall\, \theta \in \mathsf{R}.$$

It follows that the process $M = (M_n, \mathcal{F}_n)$ with

$$M_n \doteq \frac{e^{\theta S_n}}{(\cosh\theta)^n}, \qquad n \ge 0$$

is a martingale for any $\theta \in \mathsf{R}$. Now consider specifically $\theta > 0$. It follows that

$$\mathsf{E}(M_{n\wedge\tau}) = \mathsf{E}(M_0) = 1, \quad \forall\, n \geq 0,$$

thanks to the optional sampling theorem. However,

$$0 \leq M_{n\wedge\tau} \leq \frac{e^\theta}{(\cosh\theta)^{n\wedge\tau}} \leq \frac{e^\theta}{\cosh\theta}, \quad \forall\, n \geq 0.$$

We have, from DCT, that

$$\lim_n \mathsf{E}(M_{n\wedge\tau}) = \mathsf{E}(\lim_n M_{n\wedge\tau}) = 1.$$

But on set $\{\tau < \infty\}$, $\lim_n M_{n\wedge\tau} = M_\tau$, and on set $\{\tau = \infty\}$, we have

$$M_{n\wedge\tau} \leq \frac{e^\theta}{(\cosh\theta)^{n\wedge\tau}} = \frac{e^\theta}{(\cosh\theta)^{n}} \quad \to \quad 0, \qquad \text{as} \ \ n \to \infty.$$

Hence

$$1 = \mathsf{E}(\lim_n M_{n\wedge\tau}) = \mathsf{E}\left(M_\tau 1_{\{\tau<\infty\}}\right) = \mathsf{E}\left(\frac{e^\theta}{(\cosh\theta)^\tau}\cdot 1_{\{\tau<\infty\}}\right), \quad \forall\, \theta > 0.$$

Letting $\theta \to 0$, it follows from DCT that

$$\mathsf{E}(1_{\{\tau<\infty\}}) = 1, \quad \text{or} \quad \mathsf{P}(\tau < \infty) = 1.$$

Furthermore, for all $\theta > 0$,

$$\mathsf{E}\left(\frac{1}{(\cosh\theta)^\tau}\cdot 1_{\{\tau<\infty\}}\right) = \mathsf{E}\left(\frac{1}{(\cosh\theta)^\tau}\right) = e^{-\theta}$$

Let $(\cosh\theta)^{-1} = \alpha \in (0,1)$, then

$$\mathsf{E}\left(\alpha^\tau\right) = \frac{1-\sqrt{1-\alpha^2}}{\alpha}, \quad \forall\, \alpha \in (0,1).$$

It is easy to check from the table of Laplace transform that

$$\mathsf{P}(\tau = 2m) = 0, \quad \mathsf{P}(\tau = 2m-1) = (-1)^{m+1}\binom{\frac{1}{2}}{m}; \quad \forall\, m \geq 1.$$

Example: Continue with the above example. The hitting time τ, as we see in the above example, is always finite; that is, $\mathsf{P}(\tau < \infty) = 1$. Note τ is not bounded; i.e., there do not exist a number K, such that $\mathsf{P}(\tau \leq K) = 1$. It is easy to see that $S_\tau \equiv 1$, whence

$$\mathsf{E}(S_\tau) = 1 \neq 0 = \mathsf{E}(S_0).$$

Therefore the optional sampling theorem does not hold in this case.

Example (A doubling strategy): Suppose in the symmetric simple random walk, we interpret $\{X_j = 1\}$ as "success" (gain) and $\{X_j = -1\}$ as "failure" (loss) of a player at the n-th turn. Let $A_n = \{\text{player's stake at time } n\}$, for $n \geq 1$, which is non-negative, $\mathcal{F}_{n-1}$-measurable (predictable betting strategy). The total wealth process $M = (M_n, \mathcal{F}_n)$ can be written as

$$M_0 \equiv m \text{ (initial wealth)}; \; M_n = M_{n-1} + A_n X_n = M_{n-1} + A_n (S_n - S_{n-1}) = M_0 + (A \bullet S)_n, \quad \forall\, n \geq 1.$$

It is not difficult to see that M is a martingale (*"You can beat the system"*). A particular choice of betting strategy is

$$A_1 \equiv 1, \qquad A_n = \left\{ \begin{array}{ccl} 2^{n-1} & ; & \text{if } X_1 = X_2 = \cdots = X_{n-1} = -1 \\ 0 & ; & \text{otherwise.} \end{array} \right\}, \quad \forall\, n \geq 2.$$

In other words, the player double the stake after a loss and drop out of the game immediately after a win. If the first win comes at the $(n+1)$-th play, that is, on set $\{X_1 = \cdots = X_n = -1, X_{n+1} = 1\}$, then the total wealth at or after time $(n+1)$ is

$$\cdots = M_{n+2} = M_{n+1} = m - \sum_{j=1}^{n} 2^{j-1} + 2^n = m + 1.$$

Define $\tau \doteq \inf\{n \geq 1; \quad M_n = m + 1\}$. It follows that $M_{\tau+k} = 1$ for all $k \geq 0$. Furthermore, $\mathsf{P}(\tau = n) = 2^{-n}$ and $\mathsf{P}(\tau < \infty) = 1$. In this case, we have

$$\mathsf{E}(M_\tau) = m + 1 \neq m = \mathsf{E}(M_0).$$

Remark: The above betting strategy guarantees that you are always going to win. Does this mean we can beat the system? The answer is "no" in practice, because that, for such a strategy to work, you would have to be able to finance your short position, which could be as big as possible, as we will see in the next example.

Example: Continue with the above example. Suppose the game would also end if you bankrupt, that is, the game will end at

$$\sigma \doteq \inf\{n \geq 0; \quad M_n = m + 1 \text{ or } M_n \leq 0.\}$$

What will be $\mathsf{E}(M_\sigma)$? In this case, we know $\mathsf{P}(\sigma < \infty) = 1$ (why?), and for any $n \geq 0$,

$$M_{\sigma \wedge n} = \mathsf{E}(M_0) = m.$$

Letting $n \to \infty$, since $0 \leq M_{\sigma \wedge n} \leq m + 1$, we have

$$\mathsf{E}(M_\sigma) = \mathsf{E}\left(\lim_n M_{\sigma \wedge n}\right) = \mathsf{E}(M_0) = m.$$

You still cannot beat the system.

Remark: Same argument will work if we replace σ by

$$\sigma \doteq \inf\{n \geq 0; \quad M_n = 1 \text{ or } M_n \leq -K\}.$$

where K is an arbitrary positive number (the amount you are allowed to borrow). $\square$

Below is a collection of exercise.

Exercise: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale (resp. submartingale, supermartingle), and τ is a stopping time with $\mathsf{E}\tau < \infty$. If $|X_n - X_{n-1}| \leq K$, for some K and all n, then

$$\mathsf{E}(X_\tau) = \mathsf{E}(X_0) \qquad \text{(resp. } \text{“}\geq\text{”, “}\leq\text{”)}$$

Exercise: Prove the Wald's equation by optional sampling theorem: suppose $(X_1, X_2, \cdots)$ is a sequence of iid random variables with $X_j \in \mathsf{L}^1$. Let

$$S_n \doteq \sum_{j=1}^{n} X_j, \quad S_0 \equiv 0; \qquad \mathcal{F}_0 \equiv \{\emptyset, \Omega\}, \quad \mathcal{F}_n = \sigma(X_1, \cdots, X_n).$$

If τ is a F-stopping time with $\mathsf{E}\tau < \infty$, show that

$$\mathsf{E}(S_\tau) = \mathsf{E}\sum_{j=1}^{\tau} X_j = \mathsf{E}\tau \cdot \mathsf{E}X_1.$$

3 Basic convergence theorem

Suppose $X = (X_n, \mathcal{F}_n; n \geq 0)$ is a supermartingale, with $X_n - X_{n-1}$ representing your winnings per unit stake on the n-th play. Consider the following strategy: “pick two numbers $a < b$. Wait until X gets below a, then start betting one unit of stake in each play until X gets above b. Repeat.” See the following graph.

Define the following stopping times (check!):

$$\begin{aligned}
\tau_1 &\doteq \inf\{n \geq 0; \quad X_n \leq a\} \\
\tau_2 &\doteq \inf\{n \geq \tau_1; \quad X_n \geq b\} \\
&\vdots \\
\tau_{2n} &\doteq \inf\{n \geq \tau_{2n-1}; \quad X_n \leq a\} \\
\tau_{2n+1} &\doteq \inf\{n \geq \tau_{2n}; \quad X_n \geq b\} \\
&\vdots
\end{aligned}$$

with convention $\inf\{\emptyset\} = \infty$. The above betting strategy can be expressed as $A = (A_n, \mathcal{F}_n)$ with

$$A_n \doteq \left\{ \begin{array}{lll} 1 & ; & \text{if } \tau_m < k \le \tau_{m+1} \text{ for some odd } m. \\ 0 & ; & \text{otherwise.} \end{array} \right\}$$

It is not difficult to check that A is predictable. Indeed,

$$\{A_n = 1\} = \cup_{m \text{ odd}} \{\tau_m < n\} \cap \{\tau_{m+1} \ge n\} = \cup_{m \text{ odd}} \{\tau_m < n\} \cap \{\tau_{m+1} < n\}^c \in \mathcal{F}_{n-1}.$$

Suppose your initial wealth is $Y_0 \equiv 0$. Then the total wealth after n-th play is

$$Y_n = (A \bullet X)_n, \quad \forall\, n \ge 0,$$

which is a supermartingale (why?). However, if we let $U_n(a, b; \omega)$ denote the number of **upcrossings** of interval $[a, b]$ up to time n, made by sample path ω; i.e.

$$U_n(a, b; \omega) \doteq \sup\{m \ge 1;\ \ \tau_{2m} \le n\}, \qquad \text{with convention } \sup\{\emptyset\} = 0,$$

we have

$$Y_n(\omega) \ge (b - a) U_n(a, b; \omega) - \big(X_n(\omega) - a\big)^-, \qquad \forall\, \omega \in \Omega.$$

This implies the

Doob's upcrossing inequality: For a supermartingale X, we have

$$\mathsf{E} U_n(a, b) \le \frac{\mathsf{E}(X_n - a)^-}{b - a}, \qquad \forall\, n \ge 1$$

for all constants $a < b$. Here $U_n(a, b)$ is the number of upcrossings of interval $[a, b]$ by time n.

This simple inequality can be used to prove the following basic convergence theorem.

Basic martingale convergence theorem: Suppose $X = (X_n, \mathcal{F}_n)$ is a supermartingale with

$$\sup_n \mathsf{E} X_n^- < \infty.$$

Then $\lim_n X_n$ exists almost surely, and $X_\infty \doteq \limsup_n X_n = \lim_n X_n \in \mathsf{L}^1$.

Proof. Define now $U_\infty(a, b; \omega) \doteq \lim_n \uparrow U_n(a, b; \omega)$ for all $\omega \in \Omega$. Since

$$\mathsf{E}(X_n - a)^- \le \mathsf{E} X_n^- + |a| \le \sup_n \mathsf{E} X_n^- + |a|, \qquad \forall\, n$$

it follows from MCT that

$$\mathsf{E} U_\infty(a, b) = \lim_n \mathsf{E} U_n(a, b) \le \frac{\sup_n \mathsf{E} X_n^- + |a|}{b - a} < \infty.$$

Let

$$\begin{aligned} \Lambda &\doteq \{\omega \in \Omega;\ \ \lim_n X_n(\omega) \text{ does not exist on } [-\infty, \infty]\} \\ &= \Big\{\omega \in \Omega;\ \ \liminf_n X_n(\omega) \ne \limsup_n X_n(\omega)\Big\} \\ &= \cup_{\{a,b \in \mathsf{Q};\ a<b\}} \Big\{\omega \in \Omega;\ \ \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\Big\} := \cup_{\{a,b \in \mathsf{Q};\ a<b\}} \Lambda_{a,b} \end{aligned}$$

However, on set $\Lambda_{a,b}$, $U_\infty(a,b;\omega) = \infty$. It follows that $\mathsf{P}(\Lambda_{a,b} = 0$, which implies that $\mathsf{P}(\Lambda) = 0$. Hence $\lim X_n$ exists almost everywhere. It remains to show that $X_\infty \in \mathsf{L}^1$. Indeed,

$$|X_n| = X_n^+ + X_n^- = X_n + 2X_n^- \quad \Rightarrow \quad \mathsf{E}|X_n| \le \mathsf{E}X_n + 2\mathsf{E}X_n^- \le \mathsf{E}X_0 + 2\sup_n \mathsf{E}X_n^-.$$

Fatou lemma implies $X_\infty \in \mathsf{L}^1$. □

It is easy to see that X_∞ is $\mathcal{F}_\infty$-measurable. The natural question to ask now is whether $\{X_n, \mathcal{F}_n;\ n = 0, 1, \cdots, \infty\}$ is itself a supermartingale. We have the following definition.

Definition: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale (resp. supermartingale, submartingale). We say Y is a **last element**, if $Y \in \mathsf{L}^1$ is $\mathcal{F}_\infty$-measurable, and

$$(X_0, \mathcal{F}_0), \quad (X_1, \mathcal{F}_1), \quad \cdots, \quad (Y, \mathcal{F}_\infty)$$

is a martingale (resp. supermartingale, submartingale).

Theorem: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale (resp. supermartingale, submartingale). Then X has a last element if and only if $\{|X_n|\}$ (resp. $\{X_n^-\}$, $\{X_n^+\}$ is uniformly integrable, and in this case, $X_\infty = \lim_n X_n$, which exists almost surely, is indeed a **last element**; i.e. the process $X = (X_n, \mathcal{F}_n;\ n = 0, 1, \cdots, \infty)$ is a martingale (resp. supermartingale, submartingale). In particular, when X is a martingale, X is uniformly integrable if and only if there exists a integrable random variable Y such that $X_n = \mathsf{E}(Y \,|\, \mathcal{F}_n)$ for all n.

Proof. It suffices to show for the case where X is a supermartingale.

"$\Rightarrow$": suppose X has a last element Y. Then

$$\mathsf{E}(Y \,|\, \mathcal{F}_n) \le X_n \quad \Rightarrow \quad X_n^- \le \big(\mathsf{E}(Y \,|\, \mathcal{F}_n)\big)^- \le \mathsf{E}(Y^- \,|\, \mathcal{F}_n).$$

The last inequality following from conditional Jensen inequality. However, since $\big\{\mathsf{E}(Y^- \,|\, \mathcal{F}_n)\big\}$ is uniformly integrable (see the exercise on page 3), so is $\{X_n^-\}$.

"$\Leftarrow$": suppose $\{X_n^-\}$ is uniformly integrable, in particular, $\sup \mathsf{E}X_n^- < \infty$. It follows from the basic convergence theorem that $X_\infty \doteq \limsup_n X_n = \lim_n X_n \in \mathsf{L}^1$ almost surely. It remains to show that

$$\mathsf{E}(X_\infty \,|\, \mathcal{F}_n) \le X_n, \quad \forall \ \ n \ge 0.$$

However, for all $m \ge n$ and $A \in \mathcal{F}_n$, we have

$$\int_A X_m \, d\mathsf{P} = \int_A \mathsf{E}(X_m \,|\, \mathcal{F}_n) \, d\mathsf{P} \le \int_A X_n \, d\mathsf{P}.$$

Since $X_m = X_m^+ - X_m^-$, with $X_m^+ \to X_\infty^+$ and $X_m^- \to X_\infty^-$, it follows from Fatou lemma and DCT that

$$\lim_m \int_A X_m \, d\mathsf{P} = \lim_m \left(\int_A X_m^+ \, d\mathsf{P} - \int_A X_m^- \, d\mathsf{P} \right) \ge \int_A X_\infty^+ \, d\mathsf{P} - \int_A X_\infty^- \, d\mathsf{P} = \int_A X_\infty \, d\mathsf{P}.$$

In other words,

$$\int_A X_\infty \, d\mathsf{P} = \int_A \mathsf{E}(X_\infty \,|\, \mathcal{F}_n) \, d\mathsf{P} \le \int_A X_n \, d\mathsf{P}, \quad \forall \ A \in \mathcal{F}_n.$$

This completes the proof. □

Corollary: Suppose $X = (X_n, \mathcal{F}_n)$ is a non-negative supermartingale. Then $X_\infty \doteq \lim_n X_n$ is well-defines almost everywhere, and $X = (X_n, \mathcal{F}_n;\ n = 0, 1, \cdots, \infty)$ is a supermartingale.

Example (Consistency of Likelihood-ratio test): Suppose $X = (X_n;\ n \geq 1)$ is a sequence of iid random variables with density $f(x)$. However, there are two possibilitites, either $f(x) = p(x)$ or $f(x) = g(x)$. The question how to determine f from all the samples $X = (X_n)$. The usual likelihood-ration test goes as follows: suppose for simplicity $g(x), p(x) > 0$, for all x. Let

$$M_0 \equiv 1, \quad M_n \doteq \frac{p(X_1)}{g(X_1)} \frac{p(X_2)}{g(X_2)} \cdot \cdots \cdot \frac{p(X_n)}{g(X_n)}$$

If the ratio is bigger than a positive number, say a, then we determine that $f(x) = p(x)$, otherwise, $f(x) = g(x)$. For a fixed n, the probability of making a wrong decision is positive. However, such a hypothesis testing procedure is always *consistent* (no matter what a is), in the sense that, as $n \to \infty$, the probability of making the right decision goes to 1.

Suppose $f(x) = g(x)$ is true. Let $\mathcal{F}_0 \doteq \{\emptyset, \Omega\}$, and $\mathcal{F}_n \doteq \{X_1, \cdots, X_n\}$, then $M = (M_n, \mathcal{F}_n)$ is a martingale. It follows from basic convergence theorem that M_n converges almost surely and $M_\infty = \lim M_n \in \mathsf{L}^1$. However,

$$\log(M_n) = \sum_{j=1}^{n} \log \frac{p(X_j)}{g(X_j)}$$

is a summation of iid random variables. It follows from SLLN that

$$\frac{1}{n} \log M_n = \frac{1}{n} \sum_{j=1}^{n} \log \frac{p(X_j)}{g(X_j)} \quad \to \quad \mathsf{E} \log \frac{p(X_1)}{g(X_1)} \ \text{ as } \ n \to \infty.$$

But Jensen inequality implies that

$$\mathsf{E} \log \frac{p(X_1)}{g(X_1)} < \log \left(\mathsf{E} \frac{p(X_1)}{g(X_1)} \right) = \log \left(\int \frac{p(x)}{g(x)} g(x) dx \right) = \log \left(\int p(x) dx \right) = \log(1) = 0;$$

here the strict inequality holds since $p \neq g$. We conclude that $M_n \to 0$ or $M_\infty \equiv 0$ with probability 1. This explains the consistency. Note in this case, $\mathsf{E}M_\infty \neq \mathsf{E}M_0$, or M_∞ is not a last element. Hence the martingale M is not uniformly integrable, and it will not have any last element. □

4 General optional sampling theorem

Suppose $(\Omega, \mathcal{F}, \mathsf{P})$ is a probability space and $\mathsf{F} = (\mathcal{F}_n)$ is a filtration on it. We first define the **σ-algebra up to stopping time** τ, which is denoted by $\mathcal{F}_\tau$:

$$\mathcal{F}_\tau \doteq \{A \in \mathcal{F};\ \ A \cap \{\tau \leq n\} \in \mathcal{F}_n, \ \ \forall\ n = 0, 1, 2, \cdots, \infty\}$$

Exercise: Show that $\mathcal{F}_\tau$ is a σ-algebra, and when $\tau \equiv n$, $\mathcal{F}_\tau = \mathcal{F}_n$.

Exercise: Suppose τ, σ are both arbitrary stopping times. If $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\sigma$.

Exercise: Suppose τ, σ are both arbitrary stopping times. Show that $\sigma + \tau$, $\sigma \wedge \tau$, $\sigma \vee \tau$ are all stopping times. Furthermore,

$$\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau, \qquad \mathcal{F}_{\sigma \vee \tau} = \sigma(\mathcal{F}_\tau, \mathcal{F}_\sigma);$$

and all the following event belong to $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$:

$$\{\tau < \sigma\}, \ \ \{\tau \le \sigma\}, \ \ \{\tau > \sigma\}, \ \ \{\tau \ge \sigma\}, \ \ \{\tau = \sigma\}.$$

Exercise: Suppose $\{\tau_n\}$ is a sequence of stopping time. Then

$$\inf_n \tau_n, \quad \sup_n \tau_n, \quad \liminf_n \tau_n, \quad \limsup_n \tau_n, \quad \lim_n \tau_n \text{ (if exists)}$$

are all stopping times.

Exercise: Suppose τ is a F-stopping time. Then τ is $\mathcal{F}_\tau$-measurable. In addition, if $X = (X_n, \mathcal{F}_n)$ is adapted, then X_τ is also $\mathcal{F}_\tau$-measurable.

We have the following results.

Optional sampling theorem: Suppose $X = (X_n, \mathcal{F}_n;\ n \ge 0)$ is a martingale (resp. supermartingale, submartingale), and σ, τ are two F-stopping times, with $\mathsf{P}(\sigma \le \tau) = 1$. We have

$$\mathsf{E}(X_\tau \,|\, \mathcal{F}_\sigma) = X_\sigma \qquad \text{(resp. “}\le\text{”, “}\ge\text{”)}$$

if either of the following two conditions holds:

(a). σ and τ are both bounded; i.e. $\mathsf{P}(\sigma \le \tau \le K) = 1$ for some constant K.

(b). The process X has a last element; i.e. $\{|X_n|\}$ (resp. $\{X_n^-\}$, $\{X_n^+\}$) are uniformly integrable.

Proof. It suffices to show for the case where X is a supermartingale. We first show that $X_\tau \in \mathsf{L}^1$ (similar for X_σ). It is trivial in case (a). As for case (b), note $X_\tau = \lim_n X_{\tau\wedge n}$ (why?), which implies that

$$\mathsf{E}|X_\tau| \le \liminf_n \mathsf{E}|X_{\tau\wedge n}| = \liminf_n \left(X_{\tau\wedge n} + 2X_{\tau\wedge n}^-\right) \le \mathsf{E}(X_0) + 2\liminf_n \mathsf{E}X_{\tau\wedge n}^-.$$

However, we have $\mathsf{E}(Y \,|\, \mathcal{F}_n) \le X_n$ for all n; here $Y \in \mathsf{L}^1$ is the last element, which implies that

$$X_n^- \le \left(\mathsf{E}(Y\,|\,\mathcal{F}_n)\right)^- \le \mathsf{E}(Y^- \,|\, \mathcal{F}_n).$$

It follows that

$$\mathsf{E}X_{\tau\wedge n}^- = \sum_{j=0}^{n-1} \int_{\{\tau = j\}} X_j^- \, d\mathsf{P} + \int_{\{\tau \ge n\}} X_n^- \, d\mathsf{P} \le \sum_{j=0}^{n-1} \int_{\{\tau = j\}} Y^- \, d\mathsf{P} + \int_{\{\tau \ge n\}} Y^- \, d\mathsf{P} = \mathsf{E}Y^-,$$

and

$$\mathsf{E}|X_\tau| \le \mathsf{E}(X_0) + 2\liminf_n \mathsf{E}X_{\tau\wedge n}^- \le \mathsf{E}(X_0) + 2\mathsf{E}Y^- < \infty.$$

It remains to show that

$$\int_A X_\tau \, d\mathsf{P} \le \int_A X_\sigma \, d\mathsf{P}, \qquad \forall\ A \in \mathcal{F}_\sigma,$$

which amounts to

$$\int_{A\cap\{\sigma=n\}} X_\tau \, d\mathsf{P} \le \int_{A\cap\{\sigma=n\}} X_\sigma \, d\mathsf{P} = \int_{A\cap\{\sigma=n\}} X_n \, d\mathsf{P}, \qquad \forall\ A \in \mathcal{F}_\sigma,\ \forall\ n = 0, 1, \cdots, \infty.$$

However, let $B = A \cap \{\sigma = n\}$, we have $B \in \mathcal{F}_n$ by definition, and

$$\begin{aligned}
\int_B X_n \, d\mathsf{P} = \int_{B\cap\{\tau\ge n\}} X_n \, d\mathsf{P} &= \int_{B\cap\{\tau=n\}} X_n \, d\mathsf{P} + \int_{B\cap\{\tau\ge n+1\}} X_n \, d\mathsf{P} \\
&\ge \int_{B\cap\{\tau=n\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau\ge n+1\}} \mathsf{E}(X_{n+1} \,|\, \mathcal{F}_n) \, d\mathsf{P} \\
&= \int_{B\cap\{\tau=n\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau\ge n+1\}} X_{n+1} \, d\mathsf{P} \\
&\ge \cdots \\
&\ge \int_{B\cap\{n\le\tau\le m\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau\ge m+1\}} X_m \, d\mathsf{P}, \\
&\ge \int_{B\cap\{n\le\tau\le m\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau\ge m+1\}} \mathsf{E}(X_\infty \,|\, \mathcal{F}_m) \, d\mathsf{P} \\
&= \int_{B\cap\{n\le\tau\le m\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau\ge m+1\}} X_\infty \, d\mathsf{P} \qquad \forall\ \ m \ge n.
\end{aligned}$$

For case (a), we complete the proof by picking m big enough. As for case (b), letting $m \to \infty$, we have

$$\int_B X_n \, d\mathsf{P} \ge \int_{B\cap\{n\le\tau<\infty\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau=\infty\}} X_\infty \, d\mathsf{P} = \int_{B\cap\{n\le\tau\}} X_\tau \, d\mathsf{P} = \int_B X_\tau \, d\mathsf{P}.$$

This completes the proof. □

Another result is as follows.

Optional sampling theorem: Suppose $X = (X_n, \mathcal{F}_n;\ n \ge 0)$ is a martingale (resp. supermartingale, submartingale), and σ, τ are two F-stopping times, with $\mathsf{P}(\sigma \le \tau < \infty) = 1$. We have

$$\mathsf{E}(X_\tau \,|\, \mathcal{F}_\sigma) = X_\sigma \qquad \text{(resp. “}\le\text{”, “}\ge\text{”)},$$

provided

$$X_\sigma,\ X_\tau \in \mathsf{L}^1; \qquad \lim_{n\to\infty} \int_{\{\tau>n\}} |X_n| \, d\mathsf{P} = 0 \qquad (\text{resp. } \{X_n^-\},\ \{X_n^+\}).$$

Proof. It suffices to show for the case of supermartingale. The proof is similar to the preceding result. Note that we have

$$\begin{aligned}
\int_B X_n \, d\mathsf{P} &\ge \int_{B\cap\{n\le\tau\le m\}} X_\tau \, d\mathsf{P} + \int_{B\cap\{\tau\ge m+1\}} X_m \, d\mathsf{P} \\
&\ge \int_{B\cap\{n\le\tau\le m\}} X_\tau \, d\mathsf{P} - \int_{B\cap\{\tau\ge m+1\}} X_m^- \, d\mathsf{P}, \qquad \forall\ m \ge n.
\end{aligned}$$

Letting $m \to \infty$, we have

$$\int_B X_n \, d\mathsf{P} \ge \int_{B\cap\{n\le\tau<\infty\}} X_\tau \, d\mathsf{P} = \int_B X_\tau \, d\mathsf{P}$$

while

$$\int_{B\cap\{\tau\ge m+1\}} X_m^- \, d\mathsf{P} \le \int_{\{\tau\ge m+1\}} X_m^- \, d\mathsf{P} \quad \to \quad 0.$$

We completes the proof. $\square$

5 Martingale inequalities

First submartingale inequality: Suppose $X = (X_n, \mathcal{F}_n;\ n \ge 0)$ is a submartingale. Then

$$\mathsf{P}\left(\max_{0\le k\le n} X_k \ge \lambda\right) \le \frac{1}{\lambda}\int_{\{\max_{0\le k\le n} X_k\ge\lambda\}} X_n \, d\mathsf{P} \le \frac{1}{\lambda}\mathsf{E}X_n^+, \qquad \forall\ \lambda > 0.$$

Proof. Define a stopping time $\tau \doteq \inf\{k \ge 0; \quad X_k \ge \lambda\} \wedge n$. It follows that

$$A := \left\{\max_{0\le k\le n} X_k \ge \lambda\right\} = \{X_\tau \ge \lambda\} \in \mathcal{F}_\tau$$

However, optional sampling theorem implies that

$$\mathsf{E}(X_n \,|\, \mathcal{F}_\tau) \ge X_\tau \quad \Rightarrow \quad \lambda\mathsf{P}(A) \le \int_A X_\tau \, d\mathsf{P} \le \int_A X_n \, d\mathsf{P} \le \mathsf{E}X_n^+.$$

This completes the proof. $\square$

Second submartingale inequality: Suppose $X = (X_n, \mathcal{F}_n;\ n \ge 0)$ is a submartingale. Then

$$\mathsf{P}\left(\min_{0\le k\le n} X_k \le -\lambda\right) \le \frac{1}{\lambda}\left(\mathsf{E}X_n^+ - \mathsf{E}X_0\right), \qquad \forall\ \lambda > 0.$$

Proof. Define a stopping time $\tau \doteq \inf\{k \ge 0; \quad X_k \le -\lambda\} \wedge n$. It follows that

$$A := \left\{\min_{0\le k\le n} X_k \le -\lambda\right\} \in \mathcal{F}_\tau$$

However, optional sampling theorem implies that

$$\mathsf{E}(X_0) \le \mathsf{E}X_\tau = \int_A X_\tau \, d\mathsf{P} + \int_{A^c} X_\tau \, d\mathsf{P} \le -\lambda\mathsf{P}(A) + \int_{A^c} X_n \, d\mathsf{P} \le -\lambda\mathsf{P}(A) + \mathsf{E}X_n^+.$$

This completes the proof. $\square$

Exercise: Suppose $X = (X_n, \mathcal{F}_n)$ is a non-negative supermartingale. Show that

$$\mathsf{P}\left(\sup_{n\ge0} X_n \ge \lambda\right) \le \frac{1}{\lambda}\mathsf{E}X_0, \qquad \forall\ \lambda > 0.$$

Exercise **(Kolmogorov inequality):** Suppose $(X_1, X_2, \cdots)$ is a sequence of indepent random variables such that $\mathsf{E}X_j = 0$ for all j. Define $S_0 \equiv 0$ and $S_n \doteq \sum_{j=1}^n X_j$. Show that

$$\mathsf{P}\left(\max_{0\le k\le n} |S_k| \ge \lambda\right) \le \frac{\mathsf{E}S_n^2}{\lambda^2}, \qquad \forall\, \lambda > 0.$$

Doob's maximal inequality: Suppose $X = (X_n, \mathcal{F}_n)$ is a submartingale. Then

$$\left\|\max_{0\le k\le n} X_k^+\right\|_p \le \frac{p}{p-1}\left\|X_n^+\right\|_p, \qquad \forall\, p > 1;$$

here $\|\cdot\|_p$ denotes the L^p-norm.

Proof. Without loss of generality we assume $X_n \ge 0$; indeed, if $(X_n, \mathcal{F}_n)$ is a submartingale, so is $(X_n^+, \mathcal{F}_n)$ (check!). We assume $\|X_n\|_p < \infty$, which implies that

$$\left\|\max_{0\le k\le n} X_k\right\|_p \le \|X_0 + \cdots + X_n\|_p \le \|X_0\|_p + \cdots + \|X_n\|_p \le (n+1)\|X_n\|_p < \infty,$$

since $(X_n^p, \mathcal{F}_n)$ is also a submartingale (exercise). Write $Y \doteq \max_{0\le k\le n} X_k$, we have

$$Y^p = \int_0^Y p\lambda^{p-1}\, d\lambda = p\int_0^\infty \lambda^{p-1} 1_{\{Y\ge\lambda\}}\, d\lambda$$

which implies that

$$\begin{aligned}
\|Y\|_p^p = \mathsf{E}Y^p &= p\int_0^\infty \mathsf{P}(Y \ge \lambda)\cdot\lambda^{p-1}\, d\lambda \\
&\le p\int_0^\infty \left(\int_\Omega X_n \cdot 1_{\{Y\ge\lambda\}}\, d\mathsf{P}\right)\cdot\lambda^{p-2}\, d\lambda \quad \text{(1st submartingale inequality)} \\
&= p\int_\Omega\int_0^\infty X_n \cdot 1_{\{Y\ge\lambda\}}\cdot\lambda^{p-2}\, d\lambda\, d\mathsf{P} \\
&= \frac{p}{p-1}\int_\Omega X_n\cdot Y^{p-1}\, d\mathsf{P} = \frac{p}{p-1}\mathsf{E}(X_n\cdot Y^{p-1})
\end{aligned}$$

It follows from Hölder inequality that

$$\mathsf{E}(X_n\cdot Y^{p-1}) \le (\mathsf{E}X_n^p)^{\frac{1}{p}}\cdot\left(\mathsf{E}Y^{(p-1)q}\right)^{\frac{1}{q}} = \|X_n\|_p\cdot(\|Y\|_p)^{p-1}; \quad \text{here } \frac{1}{p}+\frac{1}{q} = 1 \quad \text{or} \quad q = \frac{p}{p-1}.$$

Since $\|Y\|_p$ is finite, we have

$$\|Y\|_p \le \frac{p}{p-1}\|X_n\|_p$$

This completes the proof. □

Exercise: (Doob's maximal inequality) Suppose $X = (X_n, \mathcal{F}_n)$ is a non-negative submartingale. Then

$$\left\|\max_{0\le k\le n} X_k\right\|_1 \le \frac{e}{e-1}\left(1 + \left\|X_n\,(\log X_n)^+\right\|_1\right).$$

Proof. As before, $Y \doteq \max_{0\le k\le n} X_k$. Without loss of generality we assume $\mathsf{E}X_n < \infty$, which implies that $\mathsf{E}Y < \infty$. We have

$$\begin{aligned}\mathsf{E}Y - 1 &\le \mathsf{E}(Y-1)^+ = \int_0^\infty \mathsf{P}(Y \ge 1+t)\,dt \le \int_0^\infty \frac{1}{1+t}\int_{\{Y\ge 1+t\}} X_n\,d\mathsf{P} \\ &= \int_\Omega \int_0^\infty X_n \frac{1}{1+t}\cdot 1_{\{Y\ge 1+t\}}\,dt = \int_\Omega X_n(\log Y)^+\,d\mathsf{P} = \mathsf{E}\big(X_n(\log Y)^+\big).\end{aligned}$$

However, for every $a \ge 0, b \ge 0$, the inequality

$$a(\log b)^+ \le a\log a + \frac{b}{e}$$

holds (exercise). This implies that

$$\mathsf{E}Y - 1 \le \mathsf{E}\big(X_n(\log X_n)^+\big) + \frac{1}{e}\mathsf{E}Y \quad \Rightarrow \quad \mathsf{E}Y \le \frac{e}{e-1}\left[1 + \mathsf{E}\big(X_n(\log X_n)^+\big)\right].$$

This completes the proof. □

Exercise: Suppose $X = (X_n, \mathcal{F}_n)$ is a submartingale. Show that for all $x \in \mathsf{R}$ and $\theta > 0$,

$$\mathsf{P}\left(\max_{0\le j\le n} X_j \ge x\right) \le e^{-\theta x}\mathsf{E}e^{\theta X_n}.$$

Exercise: Suppose (X_n) is a sequence of iid $N(0,1)$ random variables, and $S_n \doteq \sum_{j=1}^n X_j$. Show that for all $\lambda > 0$,

$$\mathsf{P}\left(\max_{1\le j\le n} S_j \ge \lambda\right) \le e^{-\frac{\lambda^2}{2n}}.$$

6 Other convergence theorems and their applications

Consider the probability space $(\Omega, \mathcal{F}; \mathsf{P})$.

P. Lev́y's "upward" theorem: Suppose (X_n) is a sequence of random variables such that $|X_n| \leq Y,\ \forall n$ for some random variable $Y \in \mathsf{L}^1$, and $X_\infty \doteq \lim X_n$ exists almost surely. Then for any filtration $\mathsf{F} = (\mathcal{F}_n)$, we have

$$\lim_n \mathsf{E}\left(X_n \,|\, \mathcal{F}_n\right) = \mathsf{E}\left(X_\infty \,|\, \mathcal{F}_\infty\right) \qquad \text{a.s.;}$$

here $\mathcal{F}_\infty \doteq \sigma(\cup_n \mathcal{F}_n)$ as usual. In particular, for a fixed integrable random variable X, we have

$$\lim_n \mathsf{E}\left(X \,|\, \mathcal{F}_n\right) = \mathsf{E}\left(X \,|\, \mathcal{F}_\infty\right), \quad \text{a.s.}$$

P. Lev́y's "downward" theorem: Suppose (X_n) is a sequence of random variables such that $|X_n| \leq Y,\ \forall n$ for some random variable $Y \in \mathsf{L}^1$, and $X_\infty \doteq \lim X_n$ exists almost surely. Suppose $(\mathcal{F}_n)$ is a *decreasing* sequence of sub-σ-algebras, then

$$\lim_n \mathsf{E}\left(X_n \,|\, \mathcal{F}_n\right) = \mathsf{E}\left(X_\infty \,|\, \mathcal{F}_\infty\right) \qquad \text{a.s.;}$$

here $\mathcal{F}_\infty \doteq \cap_n \mathcal{F}_n$. In particular, for a fixed integrable random variable X, we have

$$\lim_n \mathsf{E}\left(X \,|\, \mathcal{F}_n\right) = \mathsf{E}\left(X \,|\, \mathcal{F}_\infty\right), \quad \text{a.s..}$$

Proof of the "upward" theorem: We first show for the special case where $X = X_1 = X_2 = \cdots$. Let $Y_n \doteq \mathsf{E}(X \,|\, \mathcal{F}_n)$. It follows that $(Y_n, \mathcal{F}_n)$ is a uniformly integrable martingale. Hence $Y_\infty = \lim_n Y_n$ exists and is a last element; i.e.,

$$\mathsf{E}(Y_\infty \,|\, \mathcal{F}_n) = Y_n, \qquad \forall\, n.$$

It remains to show that $\mathsf{E}(X \,|\, \mathcal{F}_\infty) = Y_\infty$; indeed, for all n and $A \in \mathcal{F}_n$, we have

$$\int_A Y_\infty \, d\mathsf{P} = \int_A Y_n \, d\mathsf{P} = \int_A X \, d\mathsf{P}.$$

An straightforward application of Dynkin system theorem (check!) yileds that for all $A \in \mathcal{F}_\infty$,

$$\int_A Y_\infty \, d\mathsf{P} = \int_A X \, d\mathsf{P}.$$

Hence $Y_\infty = \mathsf{E}(X \,|\, \mathcal{F}_\infty)$.

Now for the general case of (X_n) with $|X_n| \leq Y \in \mathsf{L}^1$. Define

$$Z_m \doteq \sup_{n \geq m} |X_n - X_\infty| \leq 2Y, \qquad \forall\, m.$$

We have $\lim_m Z_m = 0$ almost surely. Furthermore,

$$\limsup_n \mathsf{E}\left(|X_n - X_\infty| \,|\, \mathcal{F}_n\right) \leq \limsup_n \mathsf{E}\left(Z_m \,|\, \mathcal{F}_n\right) = \mathsf{E}(Z_m \,|\, \mathcal{F}_\infty), \qquad \forall\, m.$$

However, it follows from CDCT that $\lim_m \mathsf{E}(Z_m \,|\, \mathcal{F}_\infty) = 0$. Therefore,

$$\limsup_n \mathsf{E}\left(|X_n - X_\infty| \,|\, \mathcal{F}_n\right) = 0 \quad \text{(a.s.)} \quad \Rightarrow \quad \mathsf{E}(X_n \,|\, \mathcal{F}_n) - \mathsf{E}(X_\infty \,|\, \mathcal{F}_n) \quad \to \quad 0, \quad \text{as } n \to \infty.$$

But $\mathsf{E}(X_\infty \,|\, \mathcal{F}_n) \to \mathsf{E}(X_\infty \,|\, \mathcal{F}_\infty)$, we conclude the proof. □

Proof of the "downward" theorem: We only need to show for the special case where $X = X_1 = X_2 = \cdots$. The general case can be shown exactly as in the "upward" theorem. Let $Y_n = \mathsf{E}(X \,|\, \mathcal{F}_n)$ (indeed, $Y = (Y_n, \mathcal{F}_n)$ is a backward martingale). We first show that $\{Y_n\}$ converges almost surely. The proof is very similar to that of the basic convergence theorem. Fix $N \in \mathsf{N}$, define

$$Z = (Z_n, \mathcal{G}_n;\ 0 \le n \le N), \qquad \text{where } \ Z_n \doteq Y_{N-n}, \ \ \mathcal{G}_n \doteq \mathcal{F}_{N-n}.$$

It is easy to see that Z is a martingale. Upcrossing lemma implies that

$$\mathsf{E}U_N(a,b;Y) = \mathsf{E}U_N(a,b;Z) \le \frac{\mathsf{E}(Z_0 - a)^+}{b-a} \le \frac{\mathsf{E}|Z_0| + |a|}{b-a} \le \frac{\mathsf{E}|X| + |a|}{b-a}.$$

The convergence of (Y_n) can then be shown in exactly the same way.

Now assume $Y_\infty = \lim_n Y_n$. It is not difficult to see that $Y_\infty \in \mathsf{L}^1$ since $\mathsf{E}|Y_n| \le \mathsf{E}|X|$ and Fatou Lemma. Clearly Y_∞ is $\mathcal{F}_\infty$-maeasurable. Furthermore, note that (Y_n) is uniformly integrable, we have that, for all $A \in \mathcal{F}_n$,

$$\int_A Y_\infty \, d\mathsf{P} = \lim_n \int_A Y_n \, d\mathsf{P} = \lim_n \int_A X \, d\mathsf{P} = \int_A X \, d\mathsf{P}.$$

Hence $Y_\infty = \mathsf{E}(X \,|\, \mathcal{F}_\infty)$. □

These two convergence theorem can be used to prove some interesting reaults in classical probaility theory.

Kolmogorov 0-1 law: Let $(X_n; n \ge 0)$ be a sequence of independent random variables. Define

$$\mathcal{F}^n = \sigma(X_n, X_{n+1}, \cdots), \qquad \mathcal{F}^\infty \doteq \cap_n \mathcal{F}^n.$$

We call $\mathcal{F}^\infty$ the **tail σ-algebra**. Then $\forall E \in \mathcal{F}^\infty$, either $\mathsf{P}(E) = 0$ or $\mathsf{P}(E) = 1$.

Proof. Let $\mathcal{F}_n \doteq \sigma(X_0, X_1, \cdots, X_n)$. Then $\big(\mathsf{E}(1_E \,|\, \mathcal{F}_n), \mathcal{F}_n\big)$ is a martingale and it follows from upward theorem

$$\lim_n \mathsf{E}(1_E \,|\, \mathcal{F}_n) = \mathsf{E}(1_E \,|\, \mathcal{F}_\infty) = 1_E.$$

However, since $E \in \mathcal{F}^\infty$, the event E is independent of $\mathcal{F}_n$ for every n (why?). Hence

$$1_E = \lim_n \mathsf{E}(1_E \,|\, \mathcal{F}_n) = \mathsf{P}(E).$$

We completes the proof. □

Strong Law of Large Numbers (SLLN): Suppose $X = (X_1, X_2, \cdots)$ is a sequence of iid random variables with $\mathsf{E}|X_1| < \infty$. Then

$$\frac{1}{n}(X_1 + \cdots + X_n) \quad \to \quad \mathsf{E}X_1, \qquad \text{almost surely}$$

Proof. Let $S_0 \equiv 0$, and $S_n = X_1 + \cdots + X_n$. Define

$$\mathcal{F}_n \doteq \sigma(S_n, S_{n+1}, \cdots) = \sigma(S_n, X_{n+1}, X_{n+2}, \cdots)$$

Then $(\mathcal{F}_n)$ is a decreasing sequence of σ-algebras. It follows from the downward theorem that

$$\lim_n \mathsf{E}(X_1 \,|\, \mathcal{F}_n) = \mathsf{E}(X_1 \,|\, \mathcal{F}_\infty).$$

But

$$\mathsf{E}(X_1 \,|\, \mathcal{F}_n) = \mathsf{E}(X_1 \,|\, S_n, X_{n+1}, \cdots) = \mathsf{E}(X_1 \,|\, S_n) = \frac{1}{n} S_n;$$

here the last equality follows from symmetry. We have

$$\lim_n \frac{1}{n} S_n = \mathsf{E}(X_1 \,|\, \mathcal{F}_\infty).$$

Note that for every $k \in \mathsf{N}$,

$$\lim_n \frac{1}{n} S_n = \lim_n \frac{1}{n}(X_k + X_{k+1} + \cdots + X_n),$$

hence $\lim_n \frac{1}{n} S_n$ is $\mathcal{F}^k := \sigma(X_k, X_{k+1}, \cdots)$-measurable for every k, and it must be $\mathcal{F}^\infty$-measurable where $\mathcal{F}^\infty$ is the tail σ-algebra. Therefore $\lim_n \frac{1}{n} S_n = c$ for some constant c almost surely, thanks to the Kolmogorov 0-1 law. In particular,

$$\lim_n \frac{1}{n} S_n = c = \mathsf{E}(X_1 \,|\, \mathcal{F}_\infty) = \mathsf{E}\big(\mathsf{E}(X_1 \,|\, \mathcal{F}_\infty)\big) = \mathsf{E}X_1, \qquad \text{a.s.}$$

We completes the proof. □

Exercise: Show that $\frac{1}{n}(X_1 + \cdots + X_n) \to \mathsf{E}X_1$ also in L^1 sense. You might want to prove the following claim first: "if $(Y_1, Y_2, \cdots)$ is a sequence of uniformly integrable random variables, so is the sequence $\big\{\frac{1}{n}(Y_1 + \cdots + Y_n);\ n \geq 1\big\}$."

Exercise: Show that the SLLN also holds when $\mathsf{E}X_1$ is only well-defined.

7 Square integrable martingale

A martingale $X = (X_n, \mathcal{F}_n)$ is said to be **square-integrable** if $\mathsf{E}X_n^2 < \infty$ for all n. It's said to be **bounded in L^2** if $\sup_n \mathsf{E}X_n^2 < \infty$. We have the following result, whose proof is left as an exercise.

Lemma: For any square integrable martingale $X = (X_n, \mathcal{F}_n)$ and $i \leq j \leq k \leq m$, we have

$$\mathsf{E}\big[(X_j - X_i) \cdot (X_m - X_k)\big] = 0.$$

That is, the increment of square integrable martingale over non-overlapping intervals are uncorrelated (or, orthogonal). In particular, we have

$$\mathsf{E}X_n^2 = \mathsf{E}X_0^2 + \sum_{j=1}^{n} \mathsf{E}(X_j - X_{j-1})^2$$

An immediate consequence of the above result is as follows.

Proposition: A square-integrable martingale $X = (X_n, \mathcal{F}_n)$ is bounded in L^2 if and only if

$$\sum_{j=1}^{\infty} \mathsf{E}(X_j - X_{j-1})^2 < \infty.$$

And in this case, $X_\infty \doteq \lim_n X_n$ exists almost surely, and $X_n \to X_\infty$ in L^2.

Proof. It follows from the preceding lemma that

$$\sup_n \mathsf{E}X_n^2 = \mathsf{E}X_0^2 + \sum_{j=1}^{\infty} \mathsf{E}(X_j - X_{j-1})^2.$$

The first part of this proposition is then trivial. And when X is bounded in L^2, $\{X_n\}$ is uniformly integrable. Therefore,

$$X_\infty \doteq \lim_n X_n$$

exists almost surely. As to the L^2 convergence of X_n, note that the orthogonality implies

$$\mathsf{E}(X_j - X_n)^2 = \sum_{k=n}^{j-1} \mathsf{E}(X_{k+1} - X_k)^2,$$

whence by Fatou Lemma,

$$\mathsf{E}(X_\infty - X_n)^2 \le \liminf_j \mathsf{E}(X_j - X_n)^2 \le \liminf_j \sum_{k=n}^{j-1} \mathsf{E}(X_{k+1} - X_k)^2 = \sum_{k=n}^{\infty} \mathsf{E}(X_{k+1} - X_k)^2.$$

Letting $n \to \infty$, we conclude that $X_n \to X_\infty$ in L^2. □

7.1 The quadratic variation process

Suppose $X = (X_n, \mathcal{F}_n)$ is a square-integrable martingale. Then the process $(X_n^2, \mathcal{F}_n)$ is a submartingale, and it follows from Doob decomposition

$$X^2 = M + A$$

where $M = (M_n, \mathcal{F}_n)$ is a martingale and $A = (A_n, \mathcal{F}_n)$ is a non-decreasing, predictable process with $A_0 \equiv 0$. Such a decomposition is unique.

Definition: The process A is said to be the **quadratic variation process** of X, denoted by $\langle X \rangle$. In other words, $\langle X \rangle$ is the unique non-decreasing, predictable process such that $\langle X \rangle_0 \equiv 0$, and $\left(X_n^2 - \langle X \rangle_n, \mathcal{F}_n\right)$ is a martingale. We also define

$$\langle X \rangle_\infty \doteq \lim_n \langle X \rangle_n = \lim_n A_n.$$

Remark: Suppose $X = (X_n, \mathcal{F}_n)$ is a square-integrable martingale, so is $Y = (X_n - X_0, \mathcal{F}_n)$. Note Y is null at $n = 0$, and $\langle Y \rangle = \langle X \rangle$.

Lemma: A square-integrable martingale X is bunded in L^2 if and only if $\mathsf{E}\langle X\rangle_\infty < \infty$. Indeed,

$$\mathsf{E}\langle X\rangle_\infty = \sum_{j=1}^{\infty} \mathsf{E}(X_j - X_{j-1})^2 = \lim_n \mathsf{E}X_n^2 - \mathsf{E}X_0^2 = \sup_n \mathsf{E}X_n^2 - \mathsf{E}X_0^2.$$

This proof of the lemma is left as an exercise.

Lemma: For every $n \geq 1$, we have

$$\mathsf{E}\left((X_n - X_{n-1})^2 \,|\, \mathcal{F}_{n-1}\right) = \langle X\rangle_n - \langle X\rangle_{n-1}.$$

Proof. Indeed,

$$\begin{aligned} \mathsf{E}\left((X_n - X_{n-1})^2 \,|\, \mathcal{F}_{n-1}\right) &= \mathsf{E}(X_n^2 \,|\, \mathcal{F}_{n-1}) - X_{n-1}^2 = \mathsf{E}(M_n + \langle X\rangle_n \,|\, \mathcal{F}_{n-1}) - (M_{n-1} + \langle X\rangle_{n-1}) \\ &= \langle X\rangle_n - \langle X\rangle_{n-1}. \end{aligned}$$

This completes the proof. □

Example: Suppose $X = (X_1, X_2, \cdots)$ is a sequence of independent L^2-random variables with zero-mean and $\mathsf{E}X_j^2 := \sigma_j^2$. Define

$$S_0 \equiv 0, \quad S_n \equiv \sum_{j=1}^{n} X_n; \qquad \mathcal{F}_0 \doteq \{\emptyset, \Omega\}, \quad \mathcal{F}_n \doteq \sigma(X_1, \cdots, X_n).$$

Then $S = (S_n, \mathcal{F}_n; n \geq 0)$ is a square-integrable martingale. Clearly,

$$\langle S\rangle_n - \langle S\rangle_{n-1} = \mathsf{E}\left((S_n - S_{n-1})^2 \,|\, \mathcal{F}_{n-1}\right) = \mathsf{E}(X_n^2 \,|\, \mathcal{F}_{n-1}) = \mathsf{E}X_n^2 = \sigma_n^2, \quad \forall\, n \geq 1.$$

It follows from induction that

$$\langle S\rangle_n = \sum_{j=1}^{n} \sigma_j^2.$$

An immediate consequence is that

$$\sum_{j=1}^{\infty} \sigma_j^2 < \infty \qquad \Rightarrow \qquad \sum_{j=1}^{\infty} X_j \quad \text{exists almost surely.}$$ □

The convergence of a square-integrable martingale $X = (X_n, \mathcal{F}_n)$ is related to the finiteness of $\langle X\rangle_\infty$. We have the following convergence result.

Theorem: Suppose $X = (X_n, \mathcal{F}_n)$ is a square-integrable martingale.

1. $\lim_n X_n$ exists in R almost surely on the set $\{\langle X\rangle_\infty < \infty\}$.
2. The converse is also true when X has bounded increment; that is, $\langle X\rangle_\infty < \infty$ almost surely on set $\{\lim_n X_n \text{ exists in } \mathsf{R}\}$, provided

$$|X_{n+1} - X_n| \leq C, \qquad \text{almost surely } \ , \forall\, n \geq 0$$

for some constant $C \geq 0$.

Proof. Without loss of generality, we assume $X_0 \equiv 0$; otheriwse, just replace X_n by $X_n - X_0$. For an arbitrary $K > 0$, define $\tau \doteq \inf\{n \geq 0; \quad A_{n+1} = \langle X \rangle_{n+1} \geq K\}$. Since $\langle X \rangle$ is a predictable process, τ is a stopping time. It follows from optional sampling theorem that the stopped process $X^\tau = (X_{\tau\wedge n}, \mathcal{F}_n)$ is also a square-integrable martingale. Furthermore, we claim that

$$\langle X^\tau \rangle = \langle X \rangle^\tau, \quad \text{ or equivalently } \quad \langle X^\tau \rangle_n = \langle X \rangle_{\tau\wedge n}, \qquad \forall\ n \geq 0;$$

indeed, it is not difficult to verify that $(\langle X \rangle_{\tau\wedge n}; \mathcal{F}_n)$ is a predictable, non-decreasing process (exercise), and

$$\left\{ (X^\tau)_n^2 - \langle X \rangle_n^\tau = (X^2 - \langle X \rangle)_{\tau\wedge n}; \mathcal{F}_n \right\}$$

is also a martingale, thanks to optional sampling theorem again. Now that $\langle X^\tau \rangle_n = \langle X \rangle_{\tau\wedge n} \leq K$ for all n, we have $\langle X^\tau \rangle_n \leq K$. It follows that $\lim_n X_n^\tau = \lim_n X_{\tau\wedge n}$ exists almost surely.

"(1)". On set $\{\langle X \rangle_\infty < K\}$, we have $\tau = \infty$, whence $\lim_n X_n^\tau = \lim_n X_{\tau\wedge n} = \lim_n X_n$ exists almost surely. Since K is arbitrary, we conclude that $\lim_n X_n$ exists almost surely on $\{\langle X \rangle_\infty < \infty\}$.

"(2)". Assume that X has bounded increment. We want to show that

$$\mathsf{P}\left(\{\langle X \rangle_\infty = \infty\} \cap \{\lim_n X_n \text{ exists in } \mathsf{R}\} \right) = 0.$$

It suffices to show that

$$\mathsf{P}\left(\{\langle X \rangle_\infty = \infty\} \cap \{\sup_n |X_n| < \infty\} \right) = 0,$$

or equivalently

$$\mathsf{P}\left(\{\langle X \rangle_\infty = \infty\} \cap \{\sup_n |X_n| \leq K\} \right) = 0, \quad \forall\ \ K \geq 0.$$

Set $T \doteq \inf\{n \geq 0; \ |X_n| > K\}$, which is a stopping time. Since $\{\sup_n |X_n| \leq K\} = \{T = \infty\}$, it suffices to show that for all $K \geq 0$,

$$\mathsf{P}\left(\{\langle X \rangle_\infty = \infty\} \cap \{T = \infty\} \right) = 0.$$

However, for every $n \geq 0$, we have

$$\begin{aligned} \mathsf{E}\langle X \rangle_{T\wedge n} &= \mathsf{E}X_{T\wedge n}^2 = \mathsf{E}\big(X_{T\wedge n-1} + (X_{T\wedge n} - X_{T\wedge n-1})\big)^2 \leq 2\left(\mathsf{E}X_{T\wedge n-1}^2 + \mathsf{E}(X_{T\wedge n} - X_{T\wedge n-1})^2\right) \\ &\leq 2(K^2 + C^2) \end{aligned}$$

It follows from MCT that $\mathsf{E}\langle X \rangle_T < \infty$, which implies $\mathsf{P}\left(\{\langle X \rangle_\infty = \infty\} \cap \{T = \infty\} \right) = 0$. □

Example: Suppose $X = (X_1, X_2, \cdots)$ is a sequence of independent L^2-random variables with zero-mean and $\mathsf{E}X_j^2 := \sigma_j^2$. We have

1. if $\sum_{j=1}^\infty \sigma_j^2 < \infty$, then $\sum_{j=1}^\infty X_j$ converges almost surely.
2. if $\sum_{j=1}^\infty X_j$ converges almost surely, then $\sum_{j=1}^\infty \sigma_j^2 < \infty$, provided that $|X_j| \leq K$, $\forall\ j$, for some constant K.

Exercise: Suppose $X = (X_1, X_2, \cdots)$ is a sequence of iid random variables with $\mathsf{P}(X = \pm 1) = \frac{1}{2}$. Let (a_n) be a sequence of real numbers. Then

$$\sum_{n=1}^\infty a_n X_n \quad \text{converges almost surely} \qquad \Leftrightarrow \qquad \sum_{n=1}^\infty a_n^2 < \infty.$$

7.2 Some classical convergence results

This convergence result can be used to prove some classical convergence results. But let us recall some useful results from analysis.

Kronecker Lemma: Let $(b_n; n \geq 1)$ be a sequence of non-decreasing, strictly positive real numbers with $b_n \uparrow \infty$, and $\{x_n\}$ is a sequence of real numbers. We have

$$\sum_{n=1}^{\infty} \frac{x_n}{b_n} \quad \text{converges} \quad \Rightarrow \quad \frac{1}{b_n} \sum_{j=1}^{n} x_j \quad \rightarrow \quad 0.$$

Proof. Let $v_n = \sum_{j=1}^n x_j / b_j$, and denote $v_\infty = \lim_n v_n$. We have

$$\begin{aligned} \sum_{j=1}^{n} x_j &= \sum_{j=1}^{n} (v_j - v_{j-1}) b_j = \sum_{j=1}^{n} v_j b_j - \sum_{j=1}^{n} v_{j-1} b_j = v_n b_n + \sum_{j=1}^{n-1} v_j b_j - \sum_{j=0}^{n-1} v_j b_{j+1} \\ &= v_n b_n + \sum_{j=0}^{n-1} v_j (b_j - b_{j+1}) \quad \text{(here } b_0 = 0\text{)}. \end{aligned}$$

Hence

$$\frac{1}{b_n} \sum_{j=1}^{n} x_j = v_n - \frac{1}{b_n} \sum_{j=0}^{n-1} (b_{j+1} - b_j) v_j.$$

We complete the proof with the following exercise. □

Exercise: Complete the proof by showing

$$\lim_n \frac{1}{b_n} \sum_{j=0}^{n-1} (b_{j+1} - b_j) v_j = v_\infty.$$

SLLN with variance constraints: Suppose $X = (X_n; \mathcal{F}_n)$ is a sequence of independent random variables with $\mathsf{E} X_j = 0, \forall\, j$ and

$$\sum_n \frac{1}{n^2} \mathsf{E} X_n^2 < \infty.$$

Then $\frac{S_n}{n} \rightarrow 0$ almost surely.

Proof. Thanks to the Kronecker lemma, it suffices to show that

$$M_n \doteq \sum_n \frac{1}{n} X_n \quad \text{converges almost surely.}$$

But this is already shown in the preceding section. □

Strong law of large numbers: Suppose $X = (X_1, X_2, \cdots)$ is a sequence of iid integrable random variable with $\mathsf{E} X_1 = \mu$. Then

$$\frac{1}{n} \sum_{j=1}^{n} X_j \quad \rightarrow \quad \mathsf{E} X_1, \qquad \text{almost surely.}$$

Proof. Let $Y_n \doteq X_n 1_{\{|X_n| \le n\}}$ be a truncation of X_n for all n. It is not difficult to see that

$$\lim_n \mathsf{E} Y_n = \mathsf{E} X_1.$$

However,

$$\sum_{n=1}^{\infty} \mathsf{P}(Y_n \neq X_n) = \sum_{n=1}^{\infty} \mathsf{P}(|X_n| > n) = \sum_{n=1}^{\infty} \mathsf{P}(|X_1| > n).$$

Note that for any random variable Z, $\mathsf{E}|Z| < \infty$ if and only if $\sum_{n=1}^{\infty} \mathsf{P}(|Z| > n) < \infty$ (exercise). Hence

$$\sum_{n=1}^{\infty} \mathsf{P}(Y_n \neq X_n) < \infty \qquad \Rightarrow \qquad \mathsf{P}(Y_n \neq X_n, \text{ i.o.}) = 0,$$

by Borel-Cantelli Lemma. Hence

$$\frac{1}{n}\sum_{j=1}^{n} X_j \quad \text{converges to } \mathsf{E}X_1 \qquad \Leftrightarrow \qquad \frac{1}{n}\sum_{j=1}^{n} Y_j \quad \text{converges to } \mathsf{E}X_1.$$

It is sufficient to show then

$$\frac{1}{n}\sum_{j=1}^{n} (Y_j - \mathsf{E}Y_j) \quad \to \quad 0 \qquad \text{almost surely,}$$

But this follows from SLLN with variance constraints if we can show that

$$\sum_{j=1}^{\infty} \frac{1}{j^2} \mathrm{Var} Y_j < \infty.$$

Indeed,

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{1}{j^2}\mathrm{Var}Y_j &\le \sum_{j=1}^{\infty} \frac{1}{j^2}\mathsf{E}Y_j^2 = \sum_{j=1}^{\infty} \frac{1}{j^2}\mathsf{E}\left(X_j^2 \cdot 1_{\{|X_j| \le j\}}\right) \\ &= \mathsf{E}\left(\sum_{j=1}^{\infty} \frac{1}{j^2} X_1^2 \cdot \sum_{k=0}^{j-1} 1_{\{k < |X_1| \le k+1\}}\right) \le \mathsf{E}\left(\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} \frac{1}{j^2}(k+1)^2 \cdot 1_{\{k<|X_1|\le k+1\}}\right) \\ &\le 2\mathsf{E}\left(\sum_{k=0}^{\infty}(k+1)\cdot 1_{\{k<|X_1|\le k+1\}}\right) \le 2\left(1 + \mathsf{E}|X_1|\right); \end{aligned}$$

here we have used the inequality

$$\sum_{j=k+1}^{\infty} \frac{1}{j^2} \le \sum_{j=k+1}^{\infty} \frac{2}{j(j+1)} = \sum_{j=k+1}^{\infty}\left(\frac{2}{j} - \frac{2}{j+1}\right) = \frac{2}{k+1}, \qquad \forall\, k \ge 0.$$

This completes the proof. □

Kolmogorov's three-series theorem: Let $X = (X_1, X_2, \cdots)$ be a sequence of independent random variables. Define the truncation $X_n^{(K)} \doteq X_n 1_{\{|X_n| \le K\}}$. Then $\sum_{n=1}^{\infty} X_n$ converges almost surely, if and only if

1. $\sum_{n=1}^{\infty} \mathsf{P}(|X_n| > K) < \infty$;
2. $\sum_{n=1}^{\infty} \mathsf{E}\left(X_n^{(K)}\right)$ converges;
3. $\sum_{n=1}^{\infty} \mathrm{Var}\left(X_n^{(K)}\right) < \infty$

for some (and hence for all) positive real number K.

Proof. We should first introduce the following lemma:

Lemma: Suppose (X_n) is a squence of independent random variables, uniformly bounded by K, or $\mathsf{P}(|X_n| \leq K) = 1$, for all n. Then $\sum_n X_n$ converges in R almost surely, if and only if $\sum_n \mathsf{E}X_n$ converges in R and $\sum_n \mathrm{Var}X_n < \infty$.

Proof. The direction "⇐" is clear. As for the direction "⇒": consider an independent *copy* of (X_n), say $(\tilde{X}_n)$; i.e. (X_n) and $(\tilde{X}_n)$ are independent, and the law of (X_n) is the same as that of $(\tilde{X}_n)$. If $\sum_n X_n$ converges almost surely, so is $(\tilde{X}_n)$ (why?). Therefore, $\sum_n (X_n - \tilde{X}_n)$ converges almost surely. But $(X_n - \tilde{X}_n)$ is a sequence of uniformly bounded, independent random variables with zero mean. It follows that

$$\sum_n \mathrm{Var}(X_n - \tilde{X}_n) < \infty \qquad \text{or} \qquad \sum_n \mathrm{Var}X_n < \infty.$$

This in turn implies that

$$\sum_n (X_n - \mathsf{E}X_n)$$

converges, or $\sum_n \mathsf{E}X_n$ converges. □

"⇒": Assume that $\sum_n X_n$ converges almost surely. Fix any $K > 0$. Since $lim_n X_n = 0$ almost surely, we have

$$\mathsf{P}(|X_n| > K, \ \text{i.o.}) = 0 \qquad \Rightarrow \qquad \sum_n \mathsf{P}(|X_n| > K) = 0,$$

thanks to Borel-Cantelli lemma. This implies that

$$\mathsf{P}(X_n \neq X_n^{(K)}, \ \text{i.o.}) = 0 \qquad \Rightarrow \qquad \sum_n X_n^{(K)} \ \text{converges almost surely.}$$

Now (2) and (3) are implied by the lemma.

"⇐": It follows from the lemma that $\sum_n X_n^{(K)}$ converges almost surely. But Borel-Cantelli lemma implies that $\mathsf{P}(|X_n| > K, \ \text{i.o.}) = 0$, which implies $\sum_n X_n$ converges almost surely. □

7.3 A SLLN for square-integrable martingales

We have the following result.

Theorem: Suppose $X = (X_n, \mathcal{F}_n)$ be a square-integrable martingale with $X_0 \equiv 0$. Then

$$\lim_n \frac{X_n}{\langle X \rangle_n} = \left\{ \begin{array}{ccc} 0 & ; & \text{almost surely on set } \{\langle X \rangle_\infty = \infty\} \\ \frac{1}{\langle X \rangle_\infty} \lim_n X_n & ; & \text{almost surely on set } \{\langle X \rangle_\infty < \infty\} \end{array} \right\}$$

Proof. We only need to prove for the set $\{\langle X\rangle_\infty = \infty\}$, and it suffices to show that

$$\lim_n \frac{X_n}{1+\langle X\rangle_n} = 0, \qquad \text{almost surely,}$$

or by Kronecker lemma, to show that

$$M_n \doteq \sum_{k=1}^n \frac{X_k - X_{k-1}}{1+\langle X\rangle_k} \qquad \text{converges almost surely.}$$

But $M = (M_n, \mathcal{F}_n)$ is a martingale, since the process $\langle X\rangle$ is predictable and $\frac{1}{1+\langle X\rangle}$ is always bounded by 1. M is clearly square-integrable, and it is sufficent to show that $\langle M\rangle_\infty < \infty$ almost surely. However,

$$\begin{aligned}
\langle M\rangle_n - \langle M\rangle_{n-1} &= \mathsf{E}\left((M_n - M_{n-1})^2 \,|\, \mathcal{F}_{n-1}\right) = \mathsf{E}\left(\frac{1}{(1+\langle X\rangle_n)^2}\cdot (X_n - X_{n-1})^2 \,|\, \mathcal{F}_{n-1}\right)\\
&= \frac{1}{(1+\langle X\rangle_n)^2}\cdot \mathsf{E}\left((X_n - X_{n-1})^2 \,|\, \mathcal{F}_{n-1}\right) = \frac{\langle X\rangle_n - \langle X\rangle_{n-1}}{(1+\langle X\rangle_n)^2}\\
&\le \frac{1}{1+\langle X\rangle_{n-1}} - \frac{1}{1+\langle X\rangle_n}; \qquad\qquad \forall\, n.
\end{aligned}$$

By induction, we have

$$\langle M\rangle_n \le \frac{1}{1+\langle X\rangle_0} - \frac{1}{1+\langle X\rangle_n} \le 1 \quad \Rightarrow \quad \langle M\rangle_\infty \le 1.$$

This completes the proof. □

Exercise: Suppose $f : [0,\infty) \to [0,\infty)$ is a non-decreasing function such that

$$\int_0^\infty \frac{1}{\big(1+f(u)\big)^2}\, du < \infty.$$

Show that for all square-integrable martingale $X = (X_n, \mathcal{F}_n)$ with $X_0 \equiv 0$,

$$\frac{X_n}{f(\langle X\rangle_n)} \quad \to \quad 0$$

almost surely on set $\{\langle X\rangle_\infty = \infty\}$.

8 Applications of martingale theory

8.1 Law of Iterated Logarithm

This is a very deep result that describes precisely the growth of the random walk on real line. Below we give a proof for the special case where $X_j \sim N(0,1)$ with the help of martingale theory.

Law of Iterated Logarithm: Suppose $X = (X_n; n \geq 1)$ is a sequence iid random varaibles with $\mathsf{E}X_1 = 0$, $\mathsf{E}(X_1^2) = \sigma^2$. Set $S_n \doteq \sum_{j=1}^n X_j$ for all $n \geq 1$. Then we have

$$\limsup_n \frac{S_n}{\sqrt{2\sigma^2 n \log\log n}} = 1, \quad \liminf_n \frac{S_n}{\sqrt{2\sigma^2 n \log\log n}} = -1$$

Proof. We need the following result.

Lemma: Suppose (X_n) is a sequence of iid $N(0,1)$ random variables, and $S_n \doteq \sum_{j=1}^n X_j$. Then for all $\lambda > 0$,

$$\mathsf{P}\left(\max_{1\leq j\leq n} S_j \geq \lambda\right) \leq e^{-\frac{\lambda^2}{2n}}.$$

Proof. For any $\theta > 0$, we have

$$\mathsf{P}\left(\max_{1\leq k\leq n} S_k \geq \lambda\right) = \mathsf{P}\left(\max_{1\leq k\leq n} e^{\theta S_k} \geq e^{\theta\lambda}\right).$$

But $\{e^{\theta S_n}\}$ is a submartingale since $\{S_n\}$ is a martingale and the function $\phi(x) = e^{\theta x}$ is convex. It follows from submartingale inequality that

$$\mathsf{P}\left(\max_{1\leq k\leq n} S_k \geq \lambda\right) \leq e^{-\theta\lambda}\mathsf{E}\left(e^{\theta S_n}\right) = e^{-\theta\lambda + \frac{n}{2}\theta^2}.$$

This inequality is true for any $\theta > 0$. The best θ we can choose is $\theta^* = \frac{\lambda}{n}$, which implies that

$$\mathsf{P}\left(\max_{1\leq k\leq n} S_k \geq \lambda\right) \leq e^{-\frac{\lambda^2}{2n}}. \qquad \square$$

Set $h(n) \doteq \sqrt{2n\log\log n}$. Without loss of generality we assume $\sigma = 1$. It suffices to show the lim sup, and lim inf is obtained the same way.

1. *Upper bound*: Fix an arbitrary $M > 1$. For an arbitrary sequence of positive numbers $\{\lambda_n\}$ we have

   $$\mathsf{P}\left(\max_{1\leq k\leq M^n} S_k \geq \lambda_n\right) \leq \exp\left\{-\frac{\lambda_n^2}{2M^n}\right\}.$$

   Choose $\lambda_n = Mh(M^{n-1})$, we have

   $$\frac{\lambda_n^2}{2M^n} = \frac{M^2 h^2(M^{n-1})}{2M^n} = \frac{M^2 \cdot 2M^{n-1}\log\log(M^{n-1})}{2M^n} = M\log\log M^{n-1}.$$

   Therefore,

   $$\sum_n \mathsf{P}\left(\max_{1\leq k\leq M^n} S_k \geq \lambda_n\right) \leq \sum_n (n-1)^{-M}\cdot(\log M)^{-M} < \infty.$$

   It follows from Borel-Cantelli lemma that

   $$\mathsf{P}\left(\max_{1\leq k\leq M^n} S_k \geq \lambda_n, \quad \text{i.o.}\right) = 0.$$

In other words, there exists $(\tilde{\Omega}) \subset \Omega$ such that $\mathsf{P}(\tilde{\Omega}) = 1$, and $\forall \omega \in \tilde{\Omega}$,

$$\max_{1 \le k \le M^n} S_k \le \lambda_n, \qquad \text{for all } n \text{ big enough, say } n \ge N(\omega).$$

Now for any m, there exists n such that $M^{n-1} \le m \le M^n$, and $\forall\ \omega \in \tilde{\Omega}$, we have

$$S_m(\omega) \le \max_{1 \le k \le M^n} S_k(\omega) \le \lambda_n = Mh(M^{n-1}) \le Mh(m), \quad \text{for all } m \text{ big enough.}$$

It follows that

$$\limsup_m \frac{S_m(\omega)}{h(m)} \le M, \qquad \forall\ \omega \in \tilde{\Omega}.$$

Since $M > 1$ is arbitrary, we have

$$\limsup_m \frac{S_m(\omega)}{h(m)} \le M, \qquad \text{almost surely.}$$

2. *Lower bound*: Fix an interger $\ell > 1$, and $0 < \epsilon < 1$, a positive real number. Let

$$F_n \doteq \left\{ S_{\ell^{n+1}} - S_{\ell^n} \ge (1-\epsilon) h(\ell^{n+1} - \ell^n) \right\} = \left\{ \frac{S_{\ell^{n+1}} - S_{\ell^n}}{\sqrt{\ell^{n+1} - \ell^n}} \ge (1-\epsilon) \frac{h(\ell^{n+1} - \ell^n)}{\sqrt{\ell^{n+1} - \ell^n}} \right\} := \mathsf{P}(Z \ge y_n);$$

here Z is a standard normal, and

$$y_n = (1-\epsilon)\sqrt{2 \log\log(\ell^{n+1} - \ell^n)} = (1-\epsilon)\sqrt{2\log(n \log \ell + \log(\ell - 1))}.$$

It follows fron the exercise below that

$$\mathsf{P}(F_n) \quad \sim \quad \frac{1}{y_n}\phi(y_n) = \frac{1}{\sqrt{2\pi} y_n} e^{-\frac{x^2}{2}} \quad \sim \quad \frac{\text{const}}{(n \log \ell)^{-(1-\epsilon)^2} \cdot \sqrt{2 \log n}};$$

whence $\sum \mathsf{P}(F_n) = \infty$, and it follows from Borel-Cantelli Lemma that $\mathsf{P}(F_n, \ \text{i.o.}) = 1$, or, for almost every $\omega \in \Omega$,

$$S_{\ell^{n+1}} - S_{\ell^n} \ge (1-\epsilon) h(\ell^{n+1} - \ell^n) \qquad \text{i.o.}$$

However, since we already know that

$$\limsup_n \frac{S_n}{h(n)} \le 1 \qquad \Rightarrow \qquad \liminf_n \frac{S_n}{h(n)} \ge -1 \quad \text{(by symmetry).}$$

Hence for almost every $\omega \in \Omega$, we have

$$S_m(\omega) \ge -2h(\omega), \qquad \text{for } m \text{ big enough.}$$

It follows that

$$S_{\ell^{n+1}} \ge S_{\ell^n} + (1-\epsilon)h(\ell^{n+1} - \ell^n) \ge -2h(\ell^n) + (1-\epsilon)h(\ell^{n+1} - \ell^n) \qquad \text{i.o.}$$

or

$$\frac{S_{\ell^{n+1}}}{h(\ell^{n+1})} \ge -2\frac{h(\ell^n)}{h(\ell^{n+1})} + (1-\epsilon)\frac{h(\ell^{n+1} - \ell^n)}{h(\ell^{n+1})}.$$

It follows that (check!)

$$\frac{S_{\ell^{n+1}}}{h(\ell^{n+1})} \ge -\frac{2}{\sqrt{\ell}} + (1-\epsilon)\sqrt{1 - \frac{1}{\ell}}$$

almost surely. Letting $\ell \to \infty$ and $\epsilon \to 0$, we completes the proof. □

Exercise: Suppose Z is a standard normal, then for any $x > 0$, we have

$$\left(x + \frac{1}{x}\right) \phi(x) \le \mathsf{P}(Z \ge x) \le \frac{1}{x}\phi(x);$$

here ϕ is the density of standard nomral; that is,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

(Hint: Use $\phi'(x) = -x\phi(x)$ and $\left(\frac{1}{x}\phi(x)\right)' = -\left(1 + \frac{1}{x^2}\right)\phi(x)$, then integrate from x to ∞.)

8.2 Stochastic approximation

Background: Let $h : \mathsf{R} \to \mathsf{R}$ be some generic function in some quite general class (say, monotone or continuous), and for any $\theta \in \mathsf{R}$, one can measure $h(\theta)$. The measure may or may not involve measurement error. The goal is to find the root of the equation

$$h(\theta) = 0.$$

We should assume that θ^* is the unique solution, and unknown.

If there is no measurement error, then good algorithms exist so that the convergence to θ^* is (better than) geometrically fast. For example, the Newton-Raphson algorithm:

$$\theta_{n+1} = \theta_n - \frac{h(\theta_n)}{h'(\theta_n)}; \qquad \forall\, n \ge 0.$$

And (under certain conditions) the algorithm will converge to θ^* no matter what the starting point θ_0 is.

However, if there is some non-negligible measurement error associated with the quantity $h(\theta)$, one has to be satisfied with some slower algorithm, but would require the algorithm to be highly robust. For example, suppose that θ_n is the estimate of θ^* at $t = n$, then at $t = n+1$ we will observe $y_{n+1} = H(\theta_n, X_{n+1})$; here (X_n) is a sequence of independent random variables, and

$$H(\theta, x) = h(\theta) + M(\theta, x) \quad \text{ with } \quad \mathsf{E}M(\theta, X_n) = 0, \ \ \forall\, \theta \in \mathsf{R},\ \forall\, n \ge 0.$$

The term $M(\theta, X_n)$ is the "measurement error". The goal is to find a sequence of estimates (θ_n) such that $\theta_n \to \theta^*$ with probability 1 no matter what the initial guess θ_0 is.

This subsection considers a special *adaptive* stochastic approximation algorithm: Robbins-Monro algorithm, which takes the recursive form

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, X_{n+1}), \qquad \forall\, n;$$

here (γ_n) is a sequence of non-negative real numbers. The requirement of the algorithm put on some constraints on the sequence of constants (γ_n): (1) the sequence $(\to_n)$ should converge to 0 so that θ_{n+1} and θ_n are getting closer; (2) the sequence of $(\to_n)$ should converge to 0 too fast; say, if $\sum \gamma_n < \infty$, and H is bounded, then

$$\sum_n |\theta_{n+1} - \theta_n| = \sum_n \gamma_n |H| \le K \sum_n \gamma_n < \infty.$$

If θ_0 is very far away to θ_*, the sequence (θ_n) could converge to a wrong point and never get close to the real θ^*.

Theorem (Robbins-Monro): Assume that for any non-negative function f, we have

$$\mathsf{E}\left[f(X_{n+1}) \,|\, \mathcal{F}_n\right] = \mathsf{E}\left[f(X_{n+1}) \,|\, \theta_n\right] = \int f(x)\, d\mu_{\theta_n}(x), \quad \text{almost surely;}$$

here $\mathcal{F}_n \doteq \sigma(X_n, \theta_n, X_{n-1}, \theta_{n-1}, \cdots)$. In other words, knowing the past, the conditional distribution of X_{n+1} depends only on θ_n, and is μ_{θ_n} (a probability measure on R). Further assume the "growth condition" that

$$\sigma^2(\theta) \doteq \int |H(\theta, x)|^2 \, d\mu_\theta(x) \leq C(1 + |\theta|^2)$$

for some $C > 0$, which implies $h(\theta) = \int H(\theta, x)\, d\mu_\theta(x)$ is well-defined.

Suppose the *stability condition*

$$\inf_{\epsilon \leq |\theta - \theta^*| \leq \frac{1}{\epsilon}} (\theta - \theta^*) h(\theta) > 0, \qquad \forall \ 0 < \epsilon < 1$$

for some $\theta^* \in \mathsf{R}$. Then for any sequence of non-negative numbers (γ_n) such that

$$\sum_n \gamma_n^2 < \infty, \qquad \sum_n \gamma_n = \infty,$$

we have $\theta \to \theta^*$ almost surely, regardless of θ_0.

Proof. We need the following result on "almost supermartingales" of Robbins & Siegmund (1971).

Lemma: Let (b_n), (c_n) be non-negative sequence of real numbers so that

$$\sum_n b_n < \infty, \qquad \sum_n c_n < \infty.$$

Let $Z = (Z_n, \mathcal{F}_n)$ and $D = (D_n, \mathcal{F}_n)$ be non-negative, adapted process such that

$$\mathsf{E}(Z_{n+1} \,|\, \mathcal{F}_n) \leq (1 + b_n) Z_n + c_n - D_n, \quad \text{almost surely,} \quad \forall \ n.$$

Then we have almost surely

$$\sum_n D_n < \infty, \qquad \lim_n Z_n \ \text{ exists in } \mathsf{R}.$$

Let $T_n \doteq \theta_n - \theta_n^*$, and $Z_n \doteq |T_n|^2$. It follows that

$$Z_{n+1} = (\theta_{n+1} - \theta^*)^2 = (\theta_{n+1} - \theta_n + \theta_n - \theta_*)^2 = \big(T_n - \gamma_{n+1} H(\theta_n, X_{n+1})\big)^2,$$

or

$$Z_{n+1} = Z_n - 2\gamma_{n+1} T_n H(\theta_n, X_{n+1}) + \gamma_{n+1}^2 H^2(\theta_n, X_{n+1}),$$

and

$$\begin{aligned}
\mathsf{E}(Z_{n+1}^2 \,|\, \mathcal{F}_n) &= Z_n - 2\gamma_{n+1}T_n h(\theta_n) + \gamma_{n+1}^2\sigma^2(\theta_n) \\
&\le Z_n - 2\gamma_{n+1}T_n h(\theta_n) + C \cdot \gamma_{n+1}^2(1+|\theta_n|^2) \\
&= Z_n - 2\gamma_{n+1}T_n h(\theta_n) + C \cdot \gamma_{n+1}^2(1+|T_n+\theta^*|^2) \\
&\le Z_n - 2\gamma_{n+1}T_n h(\theta_n) + C \cdot \gamma_{n+1}^2(2Z_n+c) \qquad \text{(for some constant } c\text{)} \\
&\le Z_n(1+C\gamma_{n+1}^2) + \bar{C}\gamma_{n+1}^2 - 2\gamma_{n+1}T_n h(\theta_n) \\
&:= Z_n(1+b_n) + c_n - D_n
\end{aligned}$$

Note here $D = (D_n, \mathcal{F}_n)$ is non-negative. By assumption $\sum b_n < \infty$ and $\sum c_n < \infty$, hence it follows from the lemma that

$$Z_n \to Z \qquad \text{and} \qquad \sum_n D_n < \infty; \qquad \text{almost surely.}$$

It remains to show that $Z = 0$ almost surely. Indeed, suppose $Z_n(\omega) > 0$ for some $\omega \in \Omega$, then for some $\epsilon \in (0,1)$, $\epsilon \le |T_n(\omega)| \le \epsilon^{-1}$ for all big enough n. This implies that

$$\liminf_n T_n(\omega)h\big(\theta_n(\omega)\big) = \liminf_n (\theta_n - \theta^*)h\big(\theta_n(\omega)\big) > 0,$$

which in turn implies that

$$\infty = \sum_n \gamma_{n+1}T_n(\omega)h\big(\theta_n(\omega)\big) = \frac{1}{2}\sum_n D_n(\omega).$$

We conclude that $\mathsf{P}(Z > 0) = 0$, or $Z = 0$ almost surely. □

Exercise: Prove the lemma used in the preceding proof.

Proof. We first show the convergence of (Z_n). Note that (D_n) is non-negative, whence

$$\mathsf{E}(Z_{n+1} \,|\, \mathcal{F}_n) \le (1+b_n)Z_n + c_n \le e^{b_n}Z_n + c_n,$$

which implies that $U = (U_n, \mathcal{F}_n)$ is a supermartingale; here

$$U_n \doteq e^{-B_{n-1}}Z_n - C_n, \qquad \text{with} \quad B_n \doteq \sum_{j=0}^{n} b_j, \quad C_n \doteq \sum_{j=0}^{n-1} e^{-B_j}c_j.$$

However, for all $n \ge 0$, almost surely.

$$U_n^- \le C_n \le \sum_{j=1}^{\infty} c_j < \infty.$$

It follows from the basic convegence theorem that (U_n) converges almost surely. But by assumption, (C_n) and (B_n) are clearly convergent. Therefore, (Z_n) converges almost surely.

It remains to show $\sum D_n < \infty$ almost surely. However, it is not difficult to verify that

$$\bar{D}_n \doteq e^{-B_n}D_n \le U_n - \mathsf{E}(U_{n+1} \,|\, \mathcal{F}_n),$$

and that $\sum D_n < \infty$ is equivalent to $\sum \bar{D}_n < \infty$. Doob'd decomposition yields that $U = M - A$ for some martingale M and predictable, non-decreasing process A, which implies that

$$\bar{D}_n \leq (M_n - A_n) - \mathsf{E}(M_{n+1} - A_{n+1} \,|\, \mathcal{F}_n) = A_{n+1} - A_n.$$

Therefore, it suffices to show that $A_\infty < \infty$ almost surely. But

$$\mathsf{E}A_n = \mathsf{E}M_n - \mathsf{E}U_n \leq \mathsf{E}M_n + \mathsf{E}U_n^- = \mathsf{E}M_0 + \mathsf{E}U_n^- \leq \mathsf{E}M_0 + \sum_{j=1}^{\infty} c_j.$$

This completes the proof. □

8.3 Markov chains

Consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and a sequence of random variables $X = (X_n, \mathcal{F}_n)$; here $\mathsf{F} = (\mathcal{F}_n)$ is the natural filtration generated by X, or $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$. The process X is said to be a **Markov chain** if

$$\mathsf{P}(X_{n+1} \in B \,|\, X_0, X_1, \cdots, X_n) = \mathsf{P}(X_{n+1} \in B \,|\, \mathcal{F}_n) = \mathsf{P}(X_{n+1} \in B \,|\, X_n), \quad \text{almost surely}$$

holds for every $B \in \mathcal{B}(\mathsf{R})$ and $n \geq 0$.

Proposition: Show that X is a Markov chain if and only if either one of the following two conditions holds:

1. $\mathsf{P}(A \,|\, X_0, X_1, \cdots, X_n) = \mathsf{P}(A \,|\, X_n)$ almost surely, $\forall\, A \in \sigma(X_{n+1}, X_{n+2}, \cdots)$.
2. $\mathsf{P}(A \cap B \,|\, X_n) = \mathsf{P}(A \,|\, X_n) \cdot \mathsf{P}(B \,|\, X_n)$ almost surely, $\forall\, A \in \sigma(X_{n+1}, X_{n+2}, \cdots)$ and $B \in \mathcal{F}_n$. That is, "future is inpendent of past conditional on present".

Exercise: Give a proof of this proposition.

Let $P_{n+1}(x; B) \doteq \mathsf{P}(X_{n+1} \in B \,|\, X_n = x)$; and P_n is said to be the *n-th stage transition probability".*

Strong Markov property: Let τ be an arbitrary F-stopping time. We have

$$\mathsf{P}(X_{\tau+1} \in B \,|\, \mathcal{F}_\tau) = P_{\tau+1}(X_\tau; B), \quad \text{almost surely on set } \{\tau < \infty\}.$$

Proof. It suffices to show that

$$\mathsf{P}\left(1_{\{\tau<\infty\}} \cdot 1_{\{X_{\tau+1} \in B\}} \,|\, \mathcal{F}_\tau\right) = 1_{\{\tau<\infty\}} \mathsf{P}(X_{\tau+1} \in B \,|\, \mathcal{F}_\tau) = P_{\tau+1}(X_\tau; B) 1_{\{\tau<\infty\}}, \quad \text{almost surely.}$$

However, $P_{\tau+1}(X_\tau; B) 1_{\{\tau<\infty\}}$ is clearly $\mathcal{F}_\tau$-measurable. We only need to show that, $\forall A \in \mathcal{F}_\tau$,

$$\int_A P_{\tau+1}(X_\tau; B) 1_{\{\tau<\infty\}} \, d\mathsf{P} = \int_A 1_{\{\tau<\infty\}} \cdot 1_{\{X_{\tau+1} \in B\}} \, d\mathsf{P} = \mathsf{P}(A \cap \{\tau < \infty\} \cap \{X_{\tau+1} \in B\}).$$

Indeed, since $A \cap \{\tau = n\} \in \mathcal{F}_n$, we have

$$\begin{aligned}
\int_A P_{\tau+1}(X_\tau; B) 1_{\{\tau<\infty\}}\, d\mathsf{P} &= \sum_{n=0}^{\infty} \int_{A\cap\{\tau=n\}} \mathsf{P}_{n+1}(X_n; B)\, d\mathsf{P} \\
&= \sum_{n=0}^{\infty} \mathsf{P}\left(A \cap \{\tau = n\} \cap \{X_{n+1} \in B\}\right) \\
&= \mathsf{P}(A \cap \{\tau < \infty\} \cap \{X_{\tau+1} \in B\}).
\end{aligned}$$

This completes the proof. □

A Markov chain X is said to be **time-homogeneous** if $P_n(x; \cdot)$ is independent of n, and we should denote $P = P_1 = P_2 = \cdots$, and by $P^{(n)}(x; B) \doteq \mathsf{P}(X_n \in B \,|\, X_0 = x)$, $\forall n \geq 0$. In particular, $\mathsf{P}^{(1)} = \mathsf{P}$ and $\mathsf{P}^{(0)}(x; B) = 1_B(x)$.

In the rest of the section, we exclusively consider time-homogeneous markov chains with a **countable state space** S for easy illustration. The transition probability $P = (P_{ij})$ where

$$P_{ij} \doteq \mathsf{P}(X_{n+1} = j \,|\, X_n = i), \qquad \forall\, i, j \in S.$$

Definition: A function $f : S \to [0, \infty)$ is said to be **harmonic** (resp. superharmonic, subharmonic) for $P(\cdot;\cdot)$ if

$$f(i) = \sum_{j\in S} P_{ij} f(j), \qquad \forall\, i \in S \qquad (\text{resp. “}\geq\text{”, “}\leq\text{”})$$

Lemma: Suppose $X = (X_n, \mathcal{F}_n)$ is a time-homogeneous Markov chain with transition probability $P(\cdot;\cdot)$. If f is P-harmonic (resp. superharmonic, subharmonic), then $\big(f(X_n), \mathcal{F}_n\big)$ is a martingale (resp. supermartingale, submartingale).

Proof: Observe that

$$\mathsf{E}\big(f(X_{n+1}) \,|\, \mathcal{F}_n\big) = \mathsf{E}\big(f(X_{n+1}) \,|\, X_n\big) = \sum_{j\in S} P_{ij} f(j) \Bigg|_{i=X_n}.$$

The result follows readily. □

Example: Let $B \subseteq S$. The function

$$g(i) \doteq \mathsf{P}(X_n \in B, \ \text{ i.o.} \,|\, X_0 \equiv i), \qquad \forall\, i \in S;$$

is harmonic.

Proof. It follows from time-homogeneity and Markov property

$$g(j) = \mathsf{P}(X_n \in B, \ \text{ i.o.} \,|\, X_1 = j) = \mathsf{P}(X_n \in B, \ \text{ i.o.} \,|\, X_1 = j, X_0 = i), \qquad \forall\, i \in S.$$

It follows that

$$\begin{aligned}\sum_j P_{ij} g(j) &= \sum_j \mathsf{P}(X_1 = j \,|\, X_0 = i) \cdot \mathsf{P}(X_n \in B, \ \text{ i.o.} \,|\, X_1 = j, X_0 = i) \\ &= \sum_j \mathsf{P}(X_1 = j, \ X_n \in B, \ \text{ i.o.} \,|\, X_0 = i) = \mathsf{P}(X_n \in B, \ \text{ i.o.} \,|\, X_0 = i) \\ &= g(i).\end{aligned}$$

This completes the proof. □

Example: Let $B \subseteq S$ be given. With $T_B \doteq \inf\{n \geq 0; \ X_n \in B\}$, define

$$\phi(i) \doteq \mathsf{P}(T_B < \infty \,|\, X_0 \equiv i); \qquad \forall\, i \in S.$$

That is, $\phi(i)$ is the probability of hitting B when the Markov chain starts from i. Then the function ϕ is superharmonic; whence $\big(\phi(X_n), \mathcal{F}_n\big)$ is a supermartingale. Furthermore, ϕ is harmonic on B^c; that is,

$$\phi(i) = \sum_j P_{ij} \phi(j), \qquad \forall \ \ i \in B^c,$$

whence

$$\mathsf{E}\big(\phi(X_{n+1}) \,|\, \mathcal{F}_n\big) = \phi(X_n), \qquad \text{almost surely on set } \{X_n \in B^c\}.$$

Proof. It follows from time-homogeneity and Markov property that

$$\phi(j) = \mathsf{P}\left(\cup_{k=1}^{\infty}\{X_k \in B\} \,\big|\, X_1 = j\right) = \mathsf{P}\left(\cup_{k=1}^{\infty}\{X_k \in B\} \,\big|\, X_1 = j, X_0 = i\right), \quad \forall\, i \in S.$$

It follows that

$$\begin{aligned}\sum_j P_{ij} \phi(j) &= \sum_j \mathsf{P}(X_1 = j \,|\, X_0 = i) \cdot \mathsf{P}\left(\cup_{k=1}^{\infty}\{X_k \in B\} \,\big|\, X_1 = j, X_0 = i\right) \\ &= \sum_j \mathsf{P}(X_1 = j, \ \cup_{k=1}^{\infty}\{X_k \in B\} \,|\, X_0 = i) = \mathsf{P}(\cup_{k=1}^{\infty}\{X_k \in B\} \,|\, X_0 = i) \\ &\leq \mathsf{P}(\cup_{k=0}^{\infty}\{X_k \in B\} \,|\, X_0 = i) = g(i).\end{aligned}$$

This proves the supermartingale property. As for the martingale property on B^c, just observe the above inequality is indeed an equality when $X_0 = i \in B^c$.

Theorem: An irreducible Markov chain $X = (X_n, \mathcal{F}_n)$ with countable state space S and transition probability matrix $P = (P_{ij})$ is recurrent if and only if every non-negative P-superharmonic function is a constant.

Proof. "⇒". Suppose X is irreducible, recurrent and $h \geq 0$ is a superharmonic function. The process $h(X) = \big(h(X_n), \mathcal{F}_n\big)$ is then a non-negative supermartingale. Therefore, for any F-stopping time τ, we have

$$\mathsf{E}h(X_\tau) \leq \mathsf{E}(X_0), \qquad \text{or} \qquad h(i) \geq \mathsf{E}\big(h(X_\tau) \,|\, X_0 = i\big), \quad \forall\, i \in S.$$

thanks to optional sampling theorem. In particular, fix an arbitrary $j \in S$, let $\tau \doteq \inf\{n \geq 1;\;\; X_n = j\}$. Since X is irreducible and recurrent, τ is almost surely finite, and

$$h(i) \geq \mathsf{E}\big(h(X_\tau) \,|\, X_0 = i\big) = \mathsf{E}\big(h(j) \,|\, X_0 = i\big) = h(j).$$

But i, j are arbitrary, hence h is a constant.

"$\Leftarrow$". Fix an arbitrary $k \in S$. Let $\tau_k \doteq \inf\{n \geq 1;\;\; X_n = k\}$, and

$$f(i) \doteq \mathsf{P}(\tau_k < \infty \,|\, X_0 = i), \qquad \forall\, i \in S.$$

Then f is a non-negative superharmonic function. Indeed,

$$\begin{aligned} f(i) &= \mathsf{P}(\tau_k < \infty \,|\, X_0 = i) = \sum_j \mathsf{P}(\tau_k < \infty,\; X_1 = j \,|\, X_0 = i) \\ &= \sum_{j \neq k} P_{ij} f(j) + P_{ik} \geq \sum_j P_{ij} f(j); \end{aligned}$$

here the last inequality follows from $0 \leq f \leq 1$. By assumption, the function f must be a constant, say $f(j) \equiv A \in [0, 1]$. It follows that

$$A = A \sum_{j \neq k} P_{ij} + P_{ik} = (1 - P_{ik})A + P_{ik}, \qquad \forall\, i \in S.$$

Suppose $A \neq 1$, then $P_{ik} \equiv 0$ for all $i \in S$. Now we can construct a superharmonic function

$$\phi(j) = 0, \quad \forall\, j \neq k, \qquad \phi(k) = 1;$$

indeed, $\forall\, i \in S$,

$$\sum_j P_{ij}\phi(j) = \sum_{j \neq k} P_{ij}\phi(j) + P_{ik}\phi(k) \equiv 0 \leq \phi(i).$$

This is a contradiction. Hence $A \equiv 1$, and the chain is irreducible and recurrent. □

8.4 Branching process

The number of offsprings for a typical individual in a population is denoted by an integer-valued random variable Z with

$$\mathsf{P}(Z = j) = p_j, \qquad j = 0, 1, 2, \cdots; \quad 0 < p_0 < 1, \;\; \sum_{j=0}^{\infty} p_j = 1$$

and

$$0 < \mu = \mathsf{E}Z = \sum_{j=0}^{\infty} j p_j < \infty.$$

Consider the stochastic recurrsion (branching process) $X = (X_n)$, the size of the n-the generation, starting with $X_0 \equiv 1$. We should assume

$$X_{n+1} = \sum_{k=1}^{X_n} Z_k^{(n+1)}, \qquad \forall\, n \geq 0;$$

here $Z^{(n+1)} = \{Z_1^{(n+1)}, Z_2^{(n+1)}, \cdots\}$ is a sequence of iid copies of Z, and we assume the independence between generations; i.e. $\{Z^{(1)}, Z^{(2)}, \cdots\}$ are independent. The intepretation is that $Z_k^{(n+1)}$ is the number of offsprings from the k-th member of the n-the generation.

Throughout this section, the filtration $\mathsf{F} = (\mathcal{F}_n)$ is the natural filtration generated by X, or $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$. First of all, $X = (X_n, \mathcal{F}_n)$ is a Markov chain; and

$$\mathsf{E}(X_{n+1} \,|\, \mathcal{F}_n) = \mathsf{E}(X_{n+1} \,|\, X_n) = \mu X_n.$$

It is not difficult to check that the process $M = (M_n, \mathcal{F}_n)$ with $M_n \doteq \mu^{-n} X_n$, is a martingale. It follows from the basic convergence theorem that

$$M_n \;\to\; M_\infty, \qquad \text{almost surely for some } \mathcal{F}_\infty\text{-measurable random variable } M_\infty.$$

We also define the *time-to-extinction*

$$T \doteq \inf\{n \geq 0; \quad X_n = 0\},$$

which is a stopping time and could take value ∞. The *extinction probability* is defined as

$$\pi \doteq \mathsf{P}(T < \infty).$$

There are several interesting questions: (1) what is the value of extinction probability π; when will $\pi = 1$? (2) when $\pi < 1$, what can we say about the distribution of M_∞? (obviously, when $\pi = 1$, $M_\infty = 0$ with probability one).

Theorem: The extinction probability π is the smallest positive root of the equation

$$s = \phi(s), \qquad \text{where } \; \phi(s) \doteq \mathsf{E}(s^Z) = \sum_{j=0}^{\infty} p_j s^j = p_0 + \sum_{j=1}^{\infty} p_j s^j; \qquad \forall\; 0 \leq s \leq 1.$$

In particular, if $\mu > 1$, then $0 < \pi < 1$; if $\mu \leq 1$, then $\pi \equiv 1$.

Remark: It follows immediately that, when $\mu \leq 1$, $M_\infty \equiv 0$ and M is not a uniformly integrable martingale, since $1 = \mathsf{E}M_0 \neq \mathsf{E}M_\infty = 0$.

Remark: The function ϕ is non-decreasing and convex since function s^j is convex for all $j \in \mathsf{N}_0$. Furthermore, $\phi(1) = 1$, and it is not difficult to check that

$$\phi'(1) = \sum_{j=1}^{\infty} j p_j = \mu.$$

We have the following graph:

Proof. We first compute the moment generating function of X_n, denoted by $\phi_n(s) \doteq \mathsf{E}(s^{X_n})$. Clearly $\phi_0(s) = s$, and recurrisively we have

$$\phi_{n+1}(s) \doteq \mathsf{E}(s^{X_{n+1}}) = \mathsf{E}\left(\mathsf{E}(s^{X_{n+1}} \,|\, X_n)\right) = \mathsf{E}\left(\phi^{X_n}(s)\right) = \phi_n\big(\phi(s)\big); \quad \forall\, n \geq 0.$$

It is not difficult to conclude that

$$\phi_{n+1}(s) = \phi \circ \phi \circ \cdots \circ \phi \qquad n \text{ times}.$$

Define the extinction probability at the n-th generation

$$\pi_n \doteq \mathsf{P}(T \leq n) = \mathsf{P}(X_n = 0), \qquad \forall\, n \geq 0.$$

It is easy to see that (π_n) is non-decreasing and

$$\pi = \lim_n \uparrow \pi_n.$$

However, note $\phi_{n+1} = \phi \circ \phi_n$, we have

$$\pi_{n+1} = \phi_{n+1}(0) = \phi\big(\phi_n(0)\big) = \phi(\pi_n), \quad \forall\, n \geq 0.$$

Letting $n \to \infty$, it yields that

$$\pi = \phi(\pi).$$

It remains to show that π is the smallest positive root; indeed,

$$\begin{array}{rcl} \pi_1 & = & p_0 \leq \text{any positive root of } s = \phi(s), \text{ say } \nu. \\ \pi_2 & = & \phi(p_0) \leq \phi(\nu) = \nu \\ \vdots & \vdots & \vdots \\ \pi & \leq & \nu, \ \text{ an arbitrary positive root of } s = \phi(s). \end{array}$$

This completes the proof. □

As for the distribution of M_∞ in case $\mu > 1$ we have the following result.

Proposition: Assume $Z \in \mathsf{L}^2$, $\mathsf{E}Z = \mu > 1$ and let $\sigma^2 = \mathrm{Var}(Z)$. Then

$$\mathrm{Var} M_\infty = \frac{\sigma^2}{\mu(\mu - 1)}.$$

(Hint: Show that M is a martingale bounded in L^2 and hence $M_n \to M_\infty$ in L^2.)

Remark: The condition $Z \in \mathsf{L}^2$ is a sufficient condition for M_∞ to be non-trivial. It can be shown that $\mathsf{E}Z > 1$ and $\mathsf{E}(Z \log Z) < \infty$ is indeed the sufficient and necessary condition for the non-triviality of M_∞ (otherwise, $M_\infty = 0$ with probability one); see Athreya and Ney (1972).

Example: Suppose Z is geometrically distributed with

$$\mathsf{P}(Z = j) = pq^j, \qquad \forall\, j \geq 0;$$

where $0 < p < 1$ and $p + q = 1$. Compute the extinction probability π, and the distribution of M_∞. (Hint: For the distribution of M_∞, show that

$$\mathsf{E}(e^{-\lambda M_\infty}) = \frac{p\lambda + q - p}{q\lambda + q - p} = \pi + \int_0^\infty (1 - \pi)^2 e^{-(1-\pi)x} \cdot e^{-\lambda x}\, dx.$$

Then derive the distribution of M_∞.)

8.5 Mathematical finance

In *binomial asset pricing model*, stock prices are modelled in discrete time, assuming that at each time step, the stock price will change to one of two possible values. See the following figure.

Even though the dynamics of the binomial pricing model is too simple compared to the real-world stock price movement, it provides a good approximation to continuous-time models with sufficiently many time steps. Besides its advantange of computational tractability, it also help illustrate the idea of "arbitrage pricing" and "risk-neutral pricing".

A Preliminary Example: Consider the following one-period binomial pricing model. The movement of the stock price is indicated in the following figure.

Suppose we buy a share of *European call option* at day 0 with strike price $K = 14$ dollars and expiration time day 1. That is, the payoff from exercising the option is

$$Y_t = (S_1 - K)^+ = \max\{S_1 - K, 0\}.$$

Remark: The holder of the call-option has the *right*, not the obligation (hence the name "option") to buy a share of stock at the strike price . A European put-option gives the holder the right to sell a share of stock with strike price K; that is, the payoff is

$$Y_t = (K - S_1)^+ = \max\{K - S_1, 0\}.$$

Remark: The name "European" means that the option can only be exercised at the "expriation date". On the contrast, "American" options can be exercised at any time before or at the expiration date. The price (value) of the American option is obviously higher than its European counterpart.

We also assume that the interest rate is zero, that is, \$1 today is worth \$1 tomorrow. Now the question is: what should be the price of this call-option (say, p) at day 0?

Arbitrage Free Principle: Suppose at day 0, we construct a portfolio by adding x share of stocks besides the call-option. The value of x is yet to be determined. This portfolio is worth $10x+p$

at time day 0. Now at day1, the portfolio is worth either $20x + 6$ (if the stock price goes up to \$20) or $5x$ (if the stock price comes down to \$5). However, if we pick x so that

$$20x + 6 = 5x \qquad \Rightarrow \qquad x = -\frac{2}{5} = -0.4,$$

This portfolio yields a riskless payoff of $20x + 6 = 5x = -2$ dollar at day 1. The arbitrage free principle says that

$$10x + p = -2 \qquad \Rightarrow \qquad -4 + p = -2 \qquad \Rightarrow \qquad p = 2.$$

That is, the option is worth \$2 at day 0.

Remark: We did not specify real-life probability for the two possible outcomes. In other words, whatever the real probabilities are, the option price is always \$2.

Delta δ: The quantity $x = -0.4$ is called "delta" in lots of occasions. Its role is to "hedge" away the risk ("delta-hedging").

Risk-neutral Probability: The risk-neutral probability is an (artificial) probability measure, under which the expected return of the option payoff equals the option price. In this example, one can find the risk-neutral probability by solving the equations

$$a + b = 1, \quad 6a + 0b = 2 \qquad \Rightarrow \qquad a = \frac{1}{3}, \;\; b = \frac{2}{3}.$$

Remark: Under the risk-neutral probability, the expected stock return is

$$\frac{1}{3} \cdot 20 + \frac{2}{3} \cdot 5 = 10 = S_0(1 + r).$$

This is indeed a general phenomenon (Martingale property).

Remark: No matter how the option payoff changes, we always get the *same* risk-neutral probability. For example, the strike price is now $K = 15$. Constructing portfolio with value $10x + p$ at day 0, with riskless value

$$20x + 5 = 5x \qquad \Rightarrow \qquad x = -\frac{1}{3} \qquad \Rightarrow \qquad 10x + p = -\frac{5}{3} \qquad \Rightarrow \qquad p = \frac{5}{3},$$

which equal the expected return under probability measure $(\frac{1}{3}, \frac{2}{3})$

$$\frac{1}{3} \cdot 5 + \frac{2}{3} \cdot 0 = \frac{5}{3}.$$

General Pricing Formulae: A general one-period binomila pricing model is indicated by the following figure.

Suppose the interest rate is r. That is, 1 dollar at day 0 is worth $R = 1 + r$ dollar at day 1. We also assume that

$$d < R < u.$$

Still let p denote the price of the option. Then we can construct a portfolio with x stocks such that it has a riskless return, or,

$$uS \cdot x + C_u = dS \cdot x + C_d \qquad \Rightarrow \qquad x = -\frac{C_u - C_d}{uS - dS}.$$

The price of the option is

$$(S \cdot x + p)R = uS \cdot x + C_u \qquad \Rightarrow \qquad p = \frac{1}{R}\left(\frac{R-d}{u-d} \cdot C_u + \frac{u-R}{u-d} \cdot C_d\right)$$

The risk-neutral probability P^* is

$$\left(\frac{R-d}{u-d}, \frac{u-R}{u-d}\right),$$

which is independent of the option payoff. The Option price can be written as

$$p = \frac{1}{R}\mathsf{E}^{\mathsf{P}^*}[C].$$

That is, the option price is the discounted expected payoff under risk neutral probability.

Furthermore, the discounted expected value of the stock at day 1 is

$$\frac{1}{R}\left(\frac{R-d}{u-d} \cdot uS + \frac{u-R}{u-d} \cdot dS\right) = S,$$

or

$$\mathsf{E}^{\mathsf{P}^*}\left[R^{-1}S_1 \,|S_0\right] = S_0.$$

This indeed says that the discounted stock price is a martingale under risk-neutral measure.

Replication: Let p be the price for the option with payoff C. Then with initial wealth p, one can construct a portfolio so that its value at day 1 completely replicates the option payoff. Indeed, put

$$x = \frac{C_u - C_d}{u - d}$$

amount of money into the stock and the rest $p - x$ amount of money into the bank. Such a portfolio will generate

$$uS \cdot \frac{x}{S} + (p-x)R = (u-R)x + pR = \frac{u-R}{u-d}(C_u - C_d) + \left(\frac{R-d}{u-d} \cdot C_u + \frac{u-R}{u-d} \cdot C_d\right) = C_u$$

when the stock price is uS at day 1, and generate

$$dS \cdot \frac{x}{S} + (p-x)R = (d-R)x + pR = \frac{d-R}{u-d}(C_u - C_d) + \left(\frac{R-d}{u-d} \cdot C_u + \frac{u-R}{u-d} \cdot C_d\right) = C_d$$

when the stock price becomes dS at day 1.

In general, we have multi-period binomial pricing model. We should conclude with the following concrete example.

Assuming the interest rate is zero, the risk-neutral probability $(q, 1-q)$ is such that

$$S_{\text{now}} = qS_{\text{up}} + (1-q)S_{\text{down}} \qquad \Rightarrow \qquad q = \frac{1}{2}.$$

We have the following tree.

9 A collection of exercises

Exercise (Ballot problem): Suppose $(X_1, X_2, \cdots)$ is a sequence of iid, non-negative integer valued L^1-random variables. Let $S_n \equiv \sum_{j=1}^n X_j$, then

$$\mathsf{P}\left(S_j < j,\ \forall\, 1 \le j \le n \,\middle|\, S_n\right) = \left(1 - \frac{S_n}{n}\right)^+.$$

(Hint: Let $Y_j = (n-j)^{-1}S_{n-j}$ and $\mathcal{F}_j \doteq \sigma(S_n, S_{n-1}, \cdots, S_{n-j})$ for all $0 \le j \le n-1$. Then $Y = (Y_j, \mathcal{F}_j)$ is a martingale).

Exercise: Suppose τ is a F-stopping time such that for some integer $K \ge 1$ and some $\epsilon > 0$, we have, for every n:

$$\mathsf{P}(\tau \le n + K \mid \mathcal{F}_n) > \epsilon, \qquad \text{almost surely.}$$

Show that $\mathsf{E}\tau < \infty$.

Kakutani's theorem: Let $(X_1, X_2, \cdot)$ be independent non-negative random variables with $\mathsf{E}X_j \equiv 1$ for all j. Define $M_0 = 1$ and $M_n = X_1 X_2 \cdots X_n$. Then $M = (M_n, \mathcal{F}_n^X)$ is a martingale and $M_\infty = \lim_n M_n$ exists almost surely. The following statements are equivalent:

1. $\mathsf{E}M_\infty = 1$.
2. M is uniformly integrable.
3. $\prod a_n > 0$ where $a_n \doteq \mathsf{E}\sqrt{X_n} \in (0, 1]$.
4. $\sum (1 - a_n) < \infty$.

If either of the above statements fails, then $M_\infty = 0$ almost surely.

Exercise (Polýa's urn): At time $t = 0$, an urn contains b black balls and w whitel balls. At each time $t = n$ $(n = 1, 2, \cdots)$ a ball is randomly chosen from the urn, and is put bakc to the urn, together with another ball of the same color. Just after time $t = n$, there are $(n + b + w)$ balls in the urn, of which $(b + B_n)$ are black; here B_n is the number of black balls chosen by time $t = n$. Define

$$M_n \doteq \frac{b + B_n}{n + b + w} = \text{proportion of black balls in urn right after time } t = n.$$

Show that $M = (M_n, \mathcal{F}_n^B)$ is a martingale; here F^B is the natural filtration generated by $B = (B_n)$. Let $M_\infty = \lim_n M_n$, which always exists by basic convergence theorem. Find the distribution of M_∞ in case of $b = w = 1$.

Exercise: Let $X = (X_n, \mathcal{F}_n)$ be a martingale with $\mathsf{E}X_n = 0$ and $\mathsf{E}X_n^2 < \infty$ for all n. Show that

$$\mathsf{P}\left(\max_{0 \le j \le n} X_j > \lambda\right) \le \frac{\mathsf{E}X_n^2}{\mathsf{E}X_n^2 + \lambda^2}, \qquad \forall\, \lambda > 0.$$

(Hint: For every $c > 0$, $(X_n + c)^2$ is a submartingale.)

Exercise: Suppose $X = (X_n, \mathcal{F}_n)$ is a martingale such that $\sup_n |X_{n+1} - X_n| \in \mathsf{L}^1$. Let

$$\begin{aligned} A_1 &\doteq \{\omega \in \Omega; \quad \lim_n X_n(\omega) \ \text{ exists in } \mathsf{R}\} \\ A_2 &\doteq \{\omega \in \Omega; \quad \limsup_n X_n(\omega) = +\infty, \quad \liminf_n X_n(\omega) = -\infty.\} \end{aligned}$$

Show that $\mathsf{P}(A_1) + \mathsf{P}(A_2) = 1$. In other words, (X_n) either converges or oscillates greatly.